INSTITUTE OF HUMAN GENETICS OF THE UNIVERSITY HOSPITAL OF ULM

MEDICAL DIRECTOR: PROF DR. C. KUBISCH

CHARACTERIZATION OF CELLULAR DNA REPAIR CAPACITY BY MOLECULAR AND CYTOGENETIC APPROACHES: SUITABILITY AS INTERMEDIATE PHENOTYPE OF CANCER

DISSERTATION

FOR THE DOCTORAL DEGREE OF HUMAN BIOLOGY (DR. BIOL. HUM.)

FACULTY OF MEDICINE, ULM UNIVERSITY

SUBMITTED BY

HARALD MARTIN SUROWY

BORN IN MEMMINGEN, GERMANY

2011

Acting Dean:	Prof. Dr. Thomas Wirth
1. Correspondent:	Prof. Dr. Walther Vogel
2. Correspondent:	PD Dr. Marcus Cronauer
Day of Promotion:	October 21st, 2011

Table of Contents

LIST OF ABBREVIATIONS III

1 INTRODUCTION 1

1.1 Genetic Predisposition to Cancer 1

1.1.1 Cancer Heritability and Susceptibility 1

1.1.2 DNA Repair Disorders and Cancer Risk 3

1.2 Cellular DNA Repair Capacity and Cancer 6

1.3 Assays Used for the Measure of Cellular DNA Repair Capacity from Peripheral Blood Samples 7

1.3.1 Micronucleus Assay 7

1.3.2 Sister Chromatid Exchange Assay 8

1.3.3 Mitotic Delay Assay 10

1.3.4 Further Measures of DNA Repair Capacity 10

1.4 Aims of the Study 11

2 PROBANDS, MATERIALS AND METHODS 13

2.1 Probands 13

2.1.1 Cohorts with DNA Repair and Genotype Data 13

2.1.2 Cohorts with Genotype Data Only 14

2.2 Materials 16

2.2.1 Sources of Peripheral Blood Lymphocytes and DNA 16

2.2.2 Cell Lines 16

2.2.3 Laboratory Material and Resources 16

2.3 Experimental Methods 21

2.3.1 Cell Culture Procedures 21

2.3.2 Micronucleus Assay 22

2.3.3 Sister Chromatid Exchange Assay 26

2.3.4 Mitotic Delay Assay 29

2.3.5 Isolation of DNA from Whole Blood Samples 32

2.3.6 Polymerase Chain Reaction (PCR) 33

2.3.7 Genotyping of DNA Polymorphisms 34

2.4 Statistical Procedures 39

2.4.1 Heritability Estimates 39

2.4.2 Linear Regression and Variance Analyses 41

2.4.3 Differences of DNA Repair Assay Results Between Groups 42

2.4.4 Association Studies 44

3 RESULTS 48

3.1 Differences of Cellular DNA Repair Capacity Between Breast Cancer Cases and Controls 49

3.2 The Impact of Low-Frequency Coding Variants in DNA Repair Genes on Cellular DNA Repair Capacity 56

3.2.1 Gene and Variant Selection 56
3.2.2 Genotyping Results 58
3.2.3 Association of Breast Cancer With Variants in the Selected DNA Repair Genes 60
3.2.4 Association Between Genotypes and DNA Repair Capacity 69
3.2.5 Further Analyses of SNP 09 73

3.3 Characterization of Assays on Cellular DNA Repair Capacity 75

3.3.1 Heritability of Measures of Cellular DNA Repair Capacity 75
3.3.2 Correlation Between the DNA Repair Assays 85
3.3.3 Correlation Between Lymphoblastoid Cell Lines and Peripheral Blood Lymphocytes 86
3.3.4 Study of the Intrinsic Variability of the Micronucleus Assay 88

4 DISCUSSION 97

4.1 Association Studies 97

4.1.1 Associations Between Cellular DNA Repair Capacity and Sporadic Breast Cancer 98
4.1.2 Coding Variants in DNA Repair Genes, Breast Cancer Risk and Cellular DNA Repair Capacity 101

4.2 Characteristics of DNA Repair Capacity Measures 111

4.2.1 Heritability of Cellular DNA Repair Capacity 111
4.2.2 Correlations Between the Measured DNA Repair End Points 115
4.2.3 Lymphoblastoid Cell Lines as Surrogates for Peripheral Blood Samples 116
4.2.4 Intrinsic Variability of the Micronucleus Assay 118

4.3 Conclusions and Prospects 121

5 SYNOPSIS 123

6 LITERATURE 125

ATTACHMENT 143

List of Abbreviations

a	adenine (DNA base)
A	alanine (amino acid)
aa	amino acids
ac	autoclaved
ACE model	additive genetic, common environment, unique environment model
ANOVA	analysis of variance
aqua dd.	double distilled water
AUC	area under the curve
BNC	binucleated cells
bp	base pair
BPDE	Benzo[a]pyrenediepoxide
BrdU	5-Bromo-2'-deoxyuridine
c	cytosine (DNA base)
C	cysteine (amino acid)
CEU	Utah Residents with Northern and Western European Ancestry
CI	confidence interval
Coef. Var.	coefficient of variation
CP	control pairs
Cyt-B	Cytochalasin B
DAPI	4',6-Diamidino-2-phenylindol-dihydrochloride
dbSNP	single nucleotide polymorphisms database (NCBI)
DMSO	Dimethyl sulfoxide
DNA	deoxyribonucleic acid
dNTP	deoxynucleoside triphosphate
DSB	double strand breaks
DZ	dizygous twin pairs
E	glutamic acid (amino acid)
EBV	Epstein-Barr Virus
EDTA	Ethylendiamide tetraacetate
f	female
F	phenylalanine (amino acid)
FA	Fanconi Anemia
FCS	fetal calf serum
g	guanine (DNA base)
G	glycine (amino acid)
G_1 phase	first gap phase (cell cycle)
G_1-S	border of the first gap and DNA synthesis phases (cell cycle)
G_2 phase	second gap phase (cell cycle)
G_2/M	second gap and mitosis phases (cell cycle)
G_2-S	border of the second gap and DNA synthesis phases (cell cycle)
G_2/S	ratio of cells in the second gap and the DNA synthesis phases (cell cycle)
G_2/S-b	baseline ratio of cells in the second gap and the DNA synthesis phases (cell cycle)
G_2/S-i	induced ratio of cells in the second gap and the DNA synthesis phases (cell cycle)
GWAS	genome-wide association study
Gy	Gray (radiation dose)
H	histidine (amino acid)
h^2	heritability
HCR	host-cell reactivation
HGVS	Human Genome Variation Society
HR	homologous recombination
HUGO	Human Genome Organization
HWE	Hardy-Weinberg Equilibrium
I	isoleucine (amino acid)
I-A, I-B	Investigator A, Investigator B
ICC	intraclass correlation coefficient
L	leucine (amino acid)
LCL	lymphoblastoid cell lines
LD	linkage disequilibrium
m	male
M	molar
M phase	mitosis phase (cell cycle)
MAF	minor allele frequency
Max.	maximum
MD	mitotic delay
Min.	minimum
MN	micronucleus, micronuclei
MN-b	baseline micronucleus assay result (frequency)
MN-i	induced micronucleus assay result (frequency)
mRNA	messenger ribonucleic acid
MSA	mean square error among twin pairs
MSW	mean square error within twin pairs
MZ	monozygous twin pairs
n	number of probands

N	asparagine (amino acid)
NCBI	National Center for Biotechnology Information, USA
NN	wild-type genotype
Nr.	Number
NV	heterozygous genotype
OD	optical density
OR	odds ratio
p	short arm (chromosome)
P	proline (amino acid)
P value	probability value (significance)
P-trend	Armitage's test for trend, probability value (significance)
P1, P2, P3	Preparator 1, 2, 3
PAR	population attributable risk
PBL	peripheral blood lymphocytes
PBS	phosphate buffered saline
PCR	polymerase chain reaction
pH	potentia hydrogenii
q	long arm (chromosome)
Q	glutamine (amino acid)
r	correlation coefficient
r^2	regression coefficient
R	arginine (amino acid)
RNA	ribonucleic acid
ROC	receiver operating characteristic
S	serine (amino acid)
S phase	DNA synthesis phase (cell cycle)
SCE	sister chromatid exchange
SCE-b	baseline sister chromatid exchange assay result (frequency)
SCE-i	induced sister chromatid exchange assay result (frequency)
SD	standard deviation
SDS	Sodium dodecylsulfate
SNP	single nucleotide polymorphism
SSC	saline sodium citrate
STC	short-time cultures
Std. Error	standard error
t	thymine (DNA base)
T	threonine (amino acid)
TRIS	Trishydroxy methylaminomethane
V	valine (amino acid)
VV	homozygous variant genotype
V_{CP}	within-pair variance of control pairs
V_{DZ}	within-pair variance of dizygous twin pairs
V_{MZ}	within-pair variance of monozygous twin pairs
VAR	variant allele
vs.	versus
W	tryptophan (amino acid)
WT	wild-type allele
x	-fold

In the manuscript, the official gene symbols provided by the Human Genome Organization Gene Nomenclature Consortium are used in italics. The official full names of the genes are listed in *Attachment A*.

1 Introduction

With an incidence of 12.7 million new cases and 7.6 million deaths in 2008, cancer is the third leading cause of death worldwide, after cardiovascular diseases and infections [248]. The most common cancers are lung, breast, colorectal, stomach and prostate cancer (see *Figure 01*). In Germany, approximately 480,000 newly diagnosed cancers and 212,000 cancer-related deaths were reported in 2008 [66].

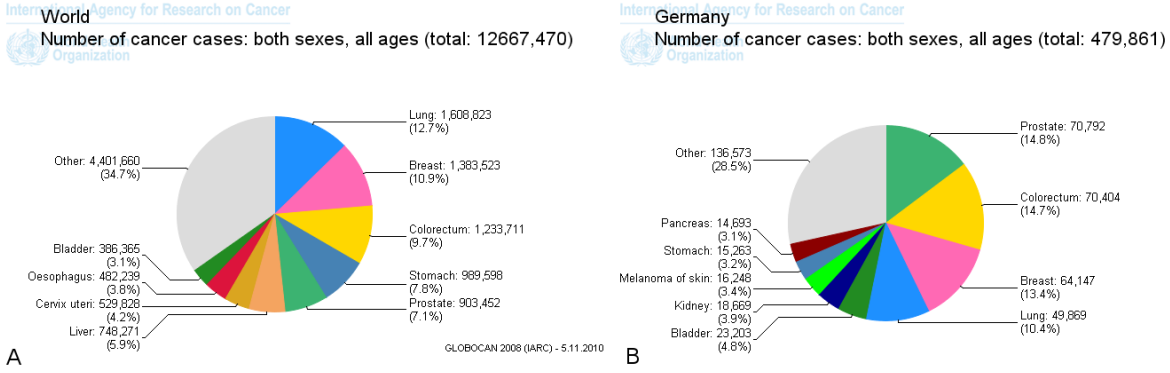

Figure 01: Incidence of the most common cancers worldwide (A) and in Germany (B). Graphs were generated with GLOBOCAN 2008 provided by the International Agency for the Research on Cancer, World Health Organization, (http://globocan.iarc.fr, [66], accession date: November 05, 2010.)

1.1 Genetic Predisposition to Cancer

1.1.1 Cancer Heritability and Susceptibility

Aside from clearly heritable cancer syndromes such as the Li-Fraumeni syndrome [131], familial clustering of cancer cases has also been described for most common cancers [76,85]. For example 10 % of all breast and colorectal cancers are estimated to have a familial background. Thus, a contribution of genetic factors to cancer formation is evident. Large studies on the co-occurrence of cancer in families and twin pairs revealed a significant degree of heritability, which is defined as the proportion of the variance of a trait that is attributable to genetic factors, opposite to environmental factors. Prostate cancer (42 %), colorectal cancer (35 %) and breast cancer (27 %) were found to be the common cancers with the highest degrees of heritability [132], followed by numerous other cancers with a significant contribution of heritable factors to cancer risk [48].

Linkage analyses in families resulted in the identification of the highly penetrating susceptibility genes for familial breast and ovarian cancer, *BRCA1* and *BRCA2* [154,247], which belong to the

most prominent and most thoroughly characterized cancer susceptibility genes. Mutation analyses revealed that germline mutations in *BRCA1* are found in 43 % and of *BRCA2* in 10 % of the families with a high risk of breast and ovarian cancer in the German population [111,152]. Recently, highly germline mutations of the *RAD51C* gene have been identified in further 1.3 % of those families [153]. Other high risk susceptibility genes have been described for colorectal and endometrial cancer [124,166] as well as melanoma [107]. In many other cancers, e.g. prostate cancer, no high risk susceptibility genes have been found at all.

A second class of cancer susceptibility genes which confer a low to moderate risk increase have been found by gene-centered association studies and mutation screening. The genes *ATM*, *BARD1*, *BRIP1*, *CHEK2*, *NBN* and *PALB2* are well-characterized examples of such susceptibility genes for breast cancer [25,151,175,180,197,214]. *BARD1* has additionally been associated to neuroblastoma susceptibility [39], *CHEK2* to prostate, kidney and thyroid cancer susceptibility [47] and *PALB2* has also been identified as a susceptibility gene for pancreatic cancer [106]. The *NBN* gene is associated with an increased risk of prostate cancer, too [46]. Biallelic mutations of *ATM*, *NBN* and *PALB2* or *BRIP1* result in the genetic disorders Ataxia Telangiectasia, Nijmegen Breakage Syndrome and Fanconi Anemia, respectively (see *1.1.2*). Many other putative susceptibility genes have been identified, however in most cases the risk increase is marginal and studies often produce contradictory results for different populations, indicating a very minor contribution of these genes to cancer causation [144].

Recently, genome-wide association studies (GWAS) have yielded numerous further variants associated with an increased risk for various types of cancer. In contrast to gene-centered studies, GWAS do not rely on a pre-selection of genes after linkage results or a functional involvement in cancer formation. The risk loci of breast and ovarian cancer, prostate cancer and colorectal cancer comprise approximately half of all associations found by GWAS [89,99]. In several instances, two or more unrelated GWAS could replicate significant associations in the same genomic regions, to a part also independently in different populations. Furthermore, independent associations of the same region to multiple cancers have been found, for example in the chromosomal region 8q24 which contains common risk variants of prostate, breast, colorectal and bladder cancer [191,224,227,243]. However, the associated alleles produced by GWAS are almost exclusively common variants and the increase in cancer risk is very low, with a simulated median odds ratio of ~ 1.2 [99]. In addition, most of the genome-wide associated variants are situated in intronic or intergenic regions and tag haplotype blocks which are supposed to carry the true disease allele in a coding sequence or genic region [54,88,89]. This circumstance makes the clarification of the respective functional effects of the associated loci in cancer development difficult, and implies the

need for further analyses by genotyping or sequencing analyses in the vicinity of the associated variant.

Taken together, the known cancer susceptibility genes and loci only explain a minority of up to 20 % of the underlying genetic variation in cancer risk. The major proportion of the reported heritability of common cancers is still missing, which is especially in the case of GWAS a currently discussed issue [113,167]. Thus, the image is consolidating that cancer is a genetically heterogeneous disease which involves many risk loci that interact in a complex manner. A further degree of complexity is added by the fact that the genetically determined variation of cancer represents only the basis which modulates the altogether larger effects of the specific individual influences of environmental factors [48,70,132].

As the genetic background of most familial cancer cases and of the vast majority of sporadic cancer cases still remains unexplained, a more comprehensive characterization of individual cancer risk is necessary to elucidate the influence of genetic factors to cancer formation.

1.1.2 DNA Repair Disorders and Cancer Risk

Numerous genetic disorders with recessive inheritance, especially those which include chromosomal instability as a cellular phenotype, also mean an increased risk of cancer for the affected individuals as one of their symptoms. Often, the susceptibility is also present in the healthy heterozygous carriers of one mutated allele of the respective disease gene. Therefore, many genes causing such disorders are found among the high or moderate risk cancer susceptibility genes.

The common denominator of a large number of genes involved in cancer susceptibility and chromosomal instability disorders is their involvement in DNA repair processes. While by the year 2001 about 70 human DNA repair genes had been characterized [184], over 150 were known in 2005 [246]. Currently, the Entrez Gene database of the National Center for Biotechnology Information (NCBI), National Institutes of Health, USA [142] (accessed November 10th, 2010), lists 468 gene products with a connection to human DNA repair. This high number of genes underlines that the cellular DNA repair system constitutes a large network of proteins with a high degree of complexity that comprises a significant portion of the estimated 20,000-25,000 human genes [98].

The processes induced upon the occurrence of DNA damage can broadly be attributed to several functional levels. Besides the various pathways available for the recognition and removal of the specific DNA lesions, those include cell cycle alterations, chromatin remodeling, the induction of apoptosis and changes in transcription. Also pathways and factors responsible for the general

maintenance of chromosomal stability, such as chromatid cohesion and the conservation of chromosome ends, are recognized as processes involved in DNA repair. As there are numerous overlaps and interconnections between the single factors and pathways, they form a cellular DNA repair network (for reviews, see [192] and [100]).

The factors involved in the repair of DNA damage depend on the various types of DNA lesions and on many other circumstances, for example the cell cycle phase in which the damage occurs. The direct repair of few of particular DNA lesions such as thymine dimers can be facilitated by specialized proteins. The mismatch repair system detects improper base pairing between the DNA strands and thus acts in a proof-reading function. Mutations in the mismatch repair genes cause the Lynch syndrome, which is classified as a familial cancer syndrome [139].
Structural aberrations of the DNA, such as base adducts, disrupted nucleotides and single strand breaks are removed via the base excision repair and nucleotide excision repair systems. The different complementation groups of Xeroderma Pigmentosum are caused by a homozygous loss-of-function of genes which are all involved in the nucleotide excision repair pathway (*XP* and *ERCC* genes), however defects in some of these genes can also be the cause for other syndromes, including the Cockayne Syndrome, Trichothiodystrophy and the Cerebro-Oculo-Facio-Skeletal Syndrome [119]. The Xeroderma Pigmentosum Variant phenotype is caused by defects in the *POLH* gene, a polymerase involved in a translesion synthesis mechanism [148] which allows the replication fork to overcome such structural aberrations during the DNA replication.
Biallelic mutations in genes which are involved in the maintenance of telomeres can be the cause of dyskeratosis congenita [41], while the Li-Fraumeni syndrome, a familial cancer syndrome, is attributable to mutations in the *TP53* gene which is involved in the activation of apoptosis upon DNA damage [209].

Double strand breaks (DSB) constitute the most dramatic type of DNA lesion, and a large spectrum of genetic disorders with an increased incidence in cancer, also in heterozygous carriers, is caused by mutations in genes involved in their repair. DSB can lead to chromosomal aberrations or loss of chromosomal material in the daughter cells, if not or improperly repaired [161]. DSB can arise by the direct actions of mutagens such as ionizing radiation, but also the repair processes of other DNA lesions can result in DSB.
DSB are also generated during the repair of inter-strand DNA cross-links via the Fanconi Anemia (FA) pathway [116]. All of the 13 complementation groups of FA are caused by biallelic mutations in one of the corresponding 13 genes, the *FANC* genes, which concertedly function in the removal of inter-strand DNA-cross-links [155]. Many FA patients develop various types of childhood cancer

[221] and leukemia [2,179]. Recently, mutations in the *RAD51C* gene have been shown to result in a FA-like disorder [236]. A substantial part of the *FANC* genes have also been identified as cancer susceptibility genes. As the genes were mostly identified before an implication in the FA pathway became evident, alternative nomenclatures exist. *FANCD1* is the alias of the *BRCA2* gene, which interacts closely with *PALB2* (*FANCN*) [250], and *BRIP1* (*FANCJ*) forms a heterodimer with *BRCA1* [38]. Mutations in *FANCD2*, *BRCA2*, *BRIP1, PALB2* and *RAD51C* are associated with an increased risk of familial or sporadic breast cancer [18,153,175,197,247] and other cancers.

The genes mutated in the Nijmegen Breakage Syndrome (*NBN*), Ataxia Telangiectasia (*ATM*) and Seckel-Syndrome *(ATR*) function in the recognition of DNA double strand breaks [163,194,220]. Both Ataxia Telangiectasia and Nijmegen Breakage Syndrome patients have high risk to develop lymphomas and other malignancies in addition to the typical phenotype, and cells defective of *NBN* and *ATM* feature chromosomal instability and increased sensitivity to ionizing radiation. [221,232]. Heterozygous carriers of *NBN* mutations also exhibit increased radiosensitivity as well as susceptibility to familial breast and prostate cancer [25,46,83]. Polymorphisms in *NBN* have been associated to sporadic breast cancer [137]. Impaired DNA repair and increased susceptibility to cancer, especially breast cancer, are also present in heterozygous carriers of *ATM* mutations [123,180,219], and associations of common ATM polymorphisms to breast and colorectal cancer have been reported [84,145].

Various, in part overlapping pathways are available for the removal of DSB lesions [150], the two best characterized of which are the non-homologous end-joining and the homologous recombination (HR) pathway. Non-homologous end-joining comprises the processing and subsequent ligation of the DSB ends, and although the pathway is error-prone and can result in a loss of information it is available during all stages of the cell cycle. Defects in this pathway lead to genetic defect syndromes. Mutations of the *LIG4* gene result in the Lig4 Syndrome [162], and polymorphisms in this gene have also been associated to lung and ovarian cancer [169,190]. Mutations in the *LIG4* gene and in the *DCLRE1C* gene can also lead to Severe Combined Immunodeficiency with Radiosensitivity [35,157]. In contrast, the HR mechanism relies on the recombination with the homologous template on the sister chromatid. Thus the pathway is only available during the DNA synthesis phase (S phase) and second gap phase (G_2 phase) of the cell cycle, but the genetic information is preserved [202]. Both Bloom's as well as the Werner's Syndromes are attributable to mutations in DNA helicase genes (*BLM* and *WRN,* respectively) which are involved in HR [56,256]. Bloom's Syndrome as well as Werner's Syndrome patients have a strongly increased risk to develop many types of cancers [73,77] and Bloom's Syndrome patients display an exceptionally high baseline frequency of sister chromatid exchanges [226]. Associations of polymorphisms in *BLM* and *WRN* to cancer risk have been reported [34,52].

1.2 Cellular DNA Repair Capacity and Cancer

Loss-of-function mutations, but also coding variants in DNA repair genes have been associated with an increased risk for cancer. Implied from the large degree of variation in the human genome, many of such potentially variant sites are present in a particular gene in the human population. Thus, each individual carries a unique and complex pattern of genetic variations in DNA repair genes. The impact of a particular variant on the function of a protein on the molecular level is difficult to characterize, especially if the functionality of the gene product is very slightly altered or if only one of the functions of the protein in a specialized context is affected. This circumstance is further complicated by the various interactions of factors in the DNA repair network, which does also lead to an interaction of the effects of variants in the DNA repair genes.
On the other hand, the individual background of variations in DNA repair genes determines the overall cellular DNA repair capacity of an individual.
This capacity can be assessed with assays which visualize different end points of DNA repair processes that take place in a cell. In contrast to the analysis of particular molecular mechanisms, for example the presence of a distinct enzymatic activity, these assays allow a characterization of the overall ability of the cells of an individual to react on DNA damage. An impaired cellular DNA repair capacity implies a larger effect size of a normal level of damage input as well as of damage inflicted by mutagens. Accordingly, many studies have shown associations between an increased risk of cancer and an impaired cellular DNA repair capacity, often measured after the induction of DNA damage. Thus, the cellular DNA repair capacity is discussed as a biomarker for cancer risk in general [22,50].
Also the influence of polymorphisms in DNA repair genes on the cellular DNA repair capacity has been shown [19,105]. In contrast to the cancer risk loci identified by GWAS, the association of variant alleles to a reduced DNA repair capacity implies also a validation of the functional relevance. Thus, the genes and loci found in DNA repair studies represent ideal candidates for cancer susceptibility studies [42]. As the efficiency of cellular DNA repair capacity is unaffected by the influence o environmental factors which can also modulate the risk of cancer, the DNA repair phenotype reflects the genetic background better than cancer itself. This led to the notion that the cellular DNA repair capacity could serve as an intermediate phenotype, between cancer and its complex genetic background [50].

1.3 Assays Used for the Measure of Cellular DNA Repair Capacity from Peripheral Blood Samples

Most measures of cellular DNA repair capacity are closely related to measures of chromosomal damage which are used in mutagenicity studies. These rely on dose-response curves to measure the induction of DNA damage by different agents or environmental mutagens. By inflicting a standardized amount of DNA damage, or no damage at all, variations of these assays enable the measurement of inter-individual differences in the response to a defined damage burden, which in turn represent the individual capability to repair the inflicted damage. Various manifestations of DNA damage and thus end points of the DNA repair network can be assessed, and either the failure or the success of DNA repair processes can be visualized by different assays. As many cancer-susceptibility genes and disorders with a high cancer incidence are associated with chromosomal instability, assays which visualize processes concerning DSB are favored to assess the degree of this instability in the context of cancer.

1.3.1 Micronucleus Assay

Chromatin material which is not incorporated in the cell nucleus is enveloped with a nucleic membrane [160,241] and can be observed as one or more separate micronuclei (MN) within a cell. As MN can arise from acentric chromosomal fragments and from chromosomes lagging during the mitotic cell division [134,178], they can be utilized as surrogate markers of chromosomal damage inflicted by clastogenic or aneugenic agents [81,195] and thus represent the failure of the cellular double strand break repair system. The induction of a cytokinesis block with Cytochalasin B (Cyt-B) results in the formation of two main nuclei. Cyt-B inhibits the polymerization of actin filaments and thus cell division, but not the reconstitution of the daughter nuclei [181]. When performed on lymphocytes, the scoring of MN within these binucleated cells (BNC) allows the standardization on cells that have completed one cell cycle, whereas Cyt-B itself does not induce MN in the concentrations applied for BNC induction [217].

MN frequencies of human peripheral blood lymphocytes are influenced by numerous intrinsic and environmental factors, including age, gender, physical activity, and several dietary and nutrition factors as well as by potentially damaging lifestyle factors, e.g. smoking habit and alcohol consume (see [20,60] for reviews). Furthermore, MN frequencies show a linear increase with the exposure dose of ionizing radiation or mutagenic chemical agents, e.g. methylmethanesulfonate [64,102]. Thus, the MN assay is currently a standard method applied in biomonitoring and

mutagenicity studies for the evaluation of DNA damage inflicted *in vivo* or *in vitro,* especially of DSB inflicted by clastogenic or aneugenic agents. The damage-induced MN frequency is thought to be better suited for the measurement of inter-individual variability in the response to DNA damage than the baseline MN frequency. For this purpose test procedures with a standardized mutagenic challenge, e.g. ionizing radiation, are applied. Contrary to baseline MN frequencies, the induced MN frequencies are less affected by environmental factors and age [11,233].

Some types of cancer, including breast cancer, have been associated with an elevated baseline MN frequency [28,95,158]. This association has also been observed for the radiation-induced MN frequency. About 30 % of breast cancer patients exhibit an increased radiosensitivity when compared to controls [9,196], and the average induced MN frequency of groups of breast cancer patients is significantly increased compared to control groups [11,233]. However, this could not be confirmed for all types of cancer, e.g. prostate cancer [235]. In patients with a familial predisposition to breast cancer due to mutations in the *BRCA1* or *BRCA2* genes, elevated radiation-induced MN frequencies have been observed compared to controls [8,186]. The same holds true for *BRCA1*- or *BRCA2*-deficient human cell lines. It remains controversial if *BRCA1* or *BRCA2* mutation carriers can generally be distinguished from healthy controls before the onset of breast cancer [118,239].

The influence of the genetic background of DNA repair has been shown in single studies which found significant associations of functional polymorphisms within DNA repair genes and elevated MN frequencies [94]. These include genes which have been associated with an increased cancer risk, like *ATM* [79] and the *XP* and *XRCC* genes [3,4,149]. Associations to variants in genes involved in the metabolism of the respective mutagens have also been shown for chemically induced MN frequencies [115,141].

Presently, some studies seek to evolve the standard micronucleus assay into a more comprehensive so-called “cytome” assay by scoring additional cytogenetic features like nucleoplasmic bridges or cellular buds aside from micronuclei. The aim is to increase the discriminating potential in either biomonitoring as well as in cancer risk studies [59].

1.3.2 Sister Chromatid Exchange Assay

Sister chromatid exchanges (SCEs) arise from reciprocal exchanges of DNA sections on replicating chromosomes and are thus markers for successful homologous recombination events between sister chromatids [205,245]. Because HR events between sister chromatids result by definition in identical sequences, the quantity of SCEs is difficult to measure on a molecular level. However,

upon the incorporation of the base substitute 5-Bromodeoxyuridine (BrdU) during the DNA replication, SCEs can be visualized and quantified with the cytogenetic SCE assay in metaphases during the second mitosis [171,212].

DNA repair by HR can occur only during the S and G_2 phases of the cell cycle, when the replicated sequence is available to serve as the substrate for an error- and loss-free repair of DNA breaks [7]. During this process, the recombination with the sister chromatid is preferred over the recombination with the homologous chromosome, probably due to the smaller spatial distance [104]. Sister chromatid recombination occurs mainly during DNA replication. When a replication fork encounters an unrepaired single-strand break on either the leading or the lagging strand, the resulting polar, one-ended DSB induces the activation of the HR machinery which uses the sister chromatid as main template [45]. As single-strand nicks can arise during the removal of DNA adducts or inter-strand cross-links, this fact also provides an explanation for the inducibility of SCEs by alkylating and cross-linking agents [121] and the insensitivity of the SCE assay to mutagens with primarily clastogenic properties which cause DSB directly, such as ionizing radiation [156]. Therefore, the SCE assay is used as a mutagenicity test for the mentioned classes of agents [208,238], which is applied in the characterization of potential environmental or occupational mutagens [146,189]. A number of studies have confirmed that significant influences of age, gender as well as of nutrition and lifestyle factors such as smoking, coffee consume on baseline SCE frequencies are present [15,261].

Some studies indicate altered SCE frequencies in cancer patients compared to controls [12,159]. Furthermore, patients which are carriers of *BRCA2* mutations exhibit higher SCE frequencies than *BRCA2*-negative family members [112]. However, other studies concluded that an increased SCE frequency is likely due to cancer treatment by radiotherapy or chemotherapy [6].

Several genes involved in HR repair of DSBs and other DNA repair pathways are involved in the formation of SCEs [185,200]. Loss of function of the *BLM* helicase, which has a crucial role during classical DSB repair by HR, results in the Bloom Syndrome which includes dramatically increased baseline SCE counts in patient's metaphases [226] but not in heterozygous carriers [120]. Coding polymorphisms in several DNA repair genes could be associated with altered baseline or induced SCE frequencies [40,128]. Also polymorphisms mutagen-metabolizing enzymes are associated with differences in the chemically induced SCE frequency [21].

1.3.3 Mitotic Delay Assay

Depending on the type of the inflicted DNA damage and on the cell cycle stage, diverse cellular checkpoint systems are activated which inhibit the progression through the cell cycle and provide the time for the repair of the DNA lesions [32]. Upon exposure to ionizing radiation, cells complete the S phase and accumulate in the G2 phase [165] to allow for the repair of the inflicted DNA DSBs by HR before entering mitosis [108]. The efficiency of the induction of this mitotic delay (MD) is dependent on the irradiation dose [254] and it can easy be visualized by flow cytometry when chromatin staining is used to quantify the amount of cells in the different phases of the cell cycle [172]. As the MD can also be induced by ultraviolet radiation and numerous chemical agents [68,96], the MD assay is a simple and fast method for the quantification of cell cycle alterations due to exposure to genotoxic mutagens. Only few environmental influences or effects of nutrition factors, e.g. caffeine [188] on the MD assay have been reported.

Some studies have identified an increased cell cycle delay in the G_2 phase in breast cancer patients compared to controls [91,122]. The repression of DNA repair genes, to a major part of genes which are involved in the second gap and mitosis phases (G_2/M) checkpoint system and in the recognition of DNA damage, induces altered mitotic delay indices *in vitro*. Examples are the checkpoint kinases *CHEK1* and *CHEK2* [176,253] but also genes which are involved in HR, such as *BRCA1, ATM or WEE1* [78,187,252].

The connection to an impaired repair of DSB is known from heterozygous carriers of *BRCA1* mutations, which show no reduction of chromosome breaks after the induction of the mitotic delay [57]. Also functional polymorphisms of the DNA repair genes *APEX1* and *XRCC1* are associated with increased MD as well as with an increased risk of cancer [91,92].

1.3.4 Further Measures of DNA Repair Capacity

Numerous additional proposed measures of the cellular DNA repair capacity exist, which are mostly also derived from mutagenicity assays, such as the host-cell reactivation (HCR) assay. Cell lines are transfected with a previously damaged plasmid DNA that contains a reporter gene. The resulting reporter gene activity is thought to represent DNA repair processes which take place in the cell, depending from the amount and type of damage inflicted and the repair pathways necessary to produce a fully and error-free repaired reporter plasmid. It has been applied to study the association of the cellular DNA repair capacity for damage inflicted by DNA adducts to bladder, breast or lung cancer [133,177,242]. The HCR is less sensitive to DSB damage induced by ionizing radiation [244].

The comet assay quantifies DNA single strand breaks which are visualized by gel electrophoresis of single cells [206]. With modifications, it can also be used for the detection of DNA double strand breaks and DNA inter-strand cross-links [164].
Other tests are based on the immunostaining of proteins which are thought to aggregate at sites of DNA damage and subsequently perform DNA repair (repair foci). Such a molecular approach which is applied for the assessment of the repair of DNA DSBs is the detection and quantification of phosphorylated histone H2AX [198]. DSBs induce the phosphorylation to gamma-H2AX which forms foci at the locations of the breaks [183]. However, it remains controversial if the number of gamma-H2AX foci can indeed exclusively indicate the presence or the repair of DSB [31,147].

1.4 Aims of the Study

Genetic variation in DNA repair genes has been associated to an increased occurrence of cancer. Besides few loci which confer a high or moderate risk increase when mutated, many common and rare low-risk variants have been identified. Isolated studies have also reported associations of an impaired cellular DNA repair capacity to cancer and also to mutations and variants in DNA repair genes. Thus, the cellular DNA repair capacity is discussed as an intermediate phenotype between cancer and the involved genetic factors.
The aim of the present study was to elucidate if the cellular DNA repair capacity can be associated with the occurrence of breast cancer on the one hand and with variants in DNA repair genes on the other hand. To complete this framework, the association of breast cancer to missense variants in DNA repair genes should also be assessed. The three types of associations were to be analyzed in a single study on the same set of probands. By this means should be clarified if the assessment of the cellular DNA repair capacity may contribute to identify individuals with an increased cancer risk and the underlying genetic factors at once.
Several measures which are used to represent the cellular DNA repair capacity were selected in such a way that different end points of cellular DNA repair processes were covered. The baseline and the radiation induced micronucleus assays represent unrepaired double strand breaks, while the baseline and the Benzo[a]pyrenediepoxide induced sister chromatid exchange assays visualize homologous recombination events. The mitotic delay assay gives information about the induction of a cell cycle delay in the G_2 phase upon damage induction.

A further aim of the present study was the analysis of several barely characterized properties of the used measures of DNA repair capacity. The applied assays are required to reflect predominantly genetic factors, in contrast to variations of environmental influences. The degree of genetic influence on the outcome of the applied assays should be analyzed by estimating their heritability in a twin study. The information if the used assays display independent or redundant end points of cellular DNA repair processes was assessed by studying the correlations between these measures on an individual basis. The applicability of lymphoblastoid cell lines to serve as basis of the assays instead of peripheral blood lymphocytes should also be studied. This was evaluated by correlating the individual results of the DNA repair assays on both types of cells in a small subset of probands. The large intrinsic variability of the micronucleus assay impairs its reproducibility. The origins of this variability should be analyzed by studying several factors which could contribute to the variability of the MN frequency in one individual sample.

The three types of associations and the stated characteristics should be studied to assess the question if the cellular DNA repair capacity can serve as an intermediate phenotype of cancer.

2 Probands, Materials and Methods

2.1 Probands

The present study includes data from patients with familial breast and ovarian cancer and healthy relatives, from patients with sporadic breast and ovarian cancer and from healthy and unrelated control individuals. Patients with borderline status or other cancers, mainly endometrium carcinoma and cervix carcinoma, were excluded. Informed consent was obtained from all participants according to the Institutional Review Board at the University of Ulm. The study includes data from six sample cohorts, as described below *(*see *2.1.1*, *2.1.2*). From the probands of four of these cohorts, fresh samples of peripheral blood could be obtained, on which DNA repair assays were performed in parallel and from which also genomic DNA for the genotyping analyses was isolated. DNA repair data and genotype data were available from probands of these cohorts. From the probands of the two other cohorts, only DNA samples were available and only genotype data could be generated. Within the six cohorts, the availability of gender and age information was heterogeneous. The sample cohorts and descriptions are summarized in *Table 01* on P*age 15.*

2.1.1 Cohorts with DNA Repair and Genotype Data

Sporadic breast and ovarian cancer cases and controls from the University Hospital of Freiburg ("Freiburg" cohort)

Sporadic breast and ovarian cancer cases with negative or unknown familial histories were recruited from the Clinic of Obstetrics and Gynecology of the University Hospital of Freiburg, Germany, where also the blood sampling procedure was performed. The samples were sent by mail to the Institute of Human Genetics of the University Hospital of Ulm, Germany. Age-matched controls were recruited among persons who received medical care for reasons not related to cancer and among hospital personnel. DNA isolation (see *2.3.5*), genotyping (see *2.3.6*) and DNA repair assays (see *2.3.2*-*2.3.4*) were performed at the Institute of Human Genetics, Ulm, Germany. Information about age was available for most individuals, gender information for all individuals. The 109 cases and 169 controls were exclusively of female gender.

Sporadic breast and ovarian cancer cases and controls from the University Hospital of Ulm ("Ulm" cohort)

Further sporadic breast and ovarian cancer cases were recruited from the Clinic of Obstetrics and Gynecology of the University Hospital of Ulm, Germany. After blood withdrawal, the samples

were transferred to the Institute of Human Genetics where DNA isolation, genotyping and DNA repair assays were performed. Age-matched controls were recruited from patients who were treated in the clinic for other reasons than cancer and from hospital personnel. Information about age and gender was available for most individuals. The 181 cases were of female gender, the 101 controls included 3 male individuals.

Additional controls for DNA repair analyses ("Repair Controls" cohort)

Students and personnel of the University of Ulm, Germany could be recruited as healthy control individuals for DNA repair and genotyping studies. Blood sampling, DNA isolation, genotyping and DNA repair assays were performed at the Institute of Human Genetics, Ulm, Germany. Age and gender information was available for most individuals. The 115 controls included 25 male individuals.

Twin Pairs ("Twin" cohort)

For the heritability study, 42 monozygous and 10 dizygous healthy twin pairs could be recruited. The same gender was mandatory for both individuals of a dizygous pair. The twins mostly reported themselves to live separately for more than one year at the time of blood sampling. For all pairs, the zygosity status was determined by analyses of molecular genetic markers as described in section *2.3.7.3.* Whereas all monozygous twin pairs were confirmed, three of the self-reported dizygous pairs turned out to be, in fact, homozygous and had to be re-categorized. This underlines the importance of a genetic validation of a phenotypic assessment of the twin status. Peripheral blood of the participants was sampled either at the Institute of Human Genetics in Ulm, or at the participant's family doctor and sent by mail. For 43 of the 52 pairs, the blood samples from both individuals of a pair could be obtained on the same day, and the time interval did not exceed 17 days for any twin pair. DNA isolation, genotyping and DNA repair assays were performed at the Institute of Human Genetics, Ulm, Germany. Age and gender information was available for all pairs, 14 of the 52 pairs were male. One randomly chosen individual of each twin pair was included as a healthy population control in genotyping analyses.

2.1.2 Cohorts with Genotype Data Only

Breast and ovarian cancer cases and healthy relatives from high risk families ("Familial" cohort)

DNA samples of cases and healthy individuals from high-risk families with positive histories of breast or ovarian cancer were kindly provided by Prof. Deißler, Clinic of Gynecology and

Obstetrics, University Hospital of Ulm, Germany. All individuals previously underwent predictive testing for mutations in the *BRCA1* and *BRCA2* genes by gene sequencing analyses. Genotyping was performed at the Institute of Human Genetics, Ulm, Germany. Information about age and gender was available for all individuals. In total, 444 individuals from 355 families were included in this study. Of these families, 238 were *BRCA*-negative, 84 were *BRCA*-positive and 33 had unclassified variants either in the *BRCA1* or the *BRCA2* gene.

Population controls ("Population Controls" cohort)

Population controls were generated from anonymized DNA samples originating from parenthood analyses at the Institute of Human Genetics of Ulm, Germany. Only unrelated individuals of Central European origin were included. Information about age, gender or any diseases was effaced due to the anonymization. This cohort comprised 819 samples.

Table 01: Summarized description of the sample cohorts with numbers of samples in each group.

Cohort		Freiburg	Ulm	Repair Controls	Twin	Familial	Population Controls	TOTAL
Total Numbers (n) of samples	Total	278	282	115	104	444	819	**2042**
	Controls	169	101	115	104	142	819	**1450**
	Cases	109	181	-	-	302	-	**592**
n of samples with *genotype* information	Total	273	226	113	104	444	819	**1976**
	Controls	166	74	113	104	142	819	**1415**
	Cases	107	152	-		302	-	**561**
n of samples with *DNA* [a] *repair* assay information	Total	172	189	115	101	-	-	**577**
	Controls	101	85	115	100	-	-	**402**
	Cases	71	104	-	-	-	-	**175**
n of samples with *age* information	Total	110	172	103	104	442	-	**931**
	Controls	63	71	103	104	141	-	**482**
	Cases	47	101	-	-	301	-	**449**
***Mean age* in years ± standard deviation**	Total	49.3 *± 15.0*	53.4 *± 14.2*	23.2 *± 4.1*	27.1 *± 11.1*	53.2 *± 14.4*	-	**46.6 *± 17.6***
	Controls	44.7 *± 14.3*	46.3 *± 14.9*	23.2 *± 4.1*	27.1 *± 11.1*	43.6 *± 13.2*	-	**36.3 *± 15.4***
	Cases	55.4 *± 12.7*	58.2 *± 11.5*	-	-	57.6 *± 12.7*	-	**57.6 *± 12.4***
n of samples with gender information	Total m/f [b]	0/278	3/274	25/85	28/76	27/417	-	**83/1130**
	Controls m/f	0/169	3/98	25/85	28/76	19/123	-	**75/551**
	Cases m/f	0/109	0/176	-	-	8/294	-	**8/579**

[a] DNA: deoxyribonucleic acid [b] m: male, f: female

2.2 Materials

2.2.1 Sources of Peripheral Blood Lymphocytes and DNA

As a prerequisite for the determination of cellular DNA repair capacity with the respective applicable assays, whole blood cultures were set up from peripheral blood samples which were obtained from the probands mentioned above (see *2.1*). Samples that were collected directly at the Institute of Human Genetics in Ulm or at the Clinic of Obstetrics and Gynecology in Ulm ("Ulm", "Repair Controls" and "Twin" cohorts) were mostly setup on the same or the following day. The shipping time of 1 to 7 days of the samples from the Clinic of Obstetrics and Gynecology in Freiburg ("Freiburg" cohort) and those samples from the "Twin" cohort that were collected by the individuals' family doctor resulted in the according delay of culture setup.
DNA was also isolated from those blood samples. Further DNA samples were available from the Institute of Human Genetics ("Population Controls" cohort) or kindly provided by the Clinic of Obstetrics and Gynecology in Ulm ("Familial" cohort).

2.2.2 Cell Lines

Immortalized lymphoblastoid cell lines (LCL) from peripheral whole blood samples of 9 monozygous and 3 dizygous twin pairs (24 individuals in total) were generated at the Institute of Virology, University Hospital of Ulm, Germany. The lymphoblastoid cells originate from B-lymphocytes which are transformed by transfection with the Epstein-Barr-Virus.

2.2.3 Laboratory Material and Resources

Chemicals

Acetic acid	AppliChem, Darmstadt, Germany
Ammonium chloride (NH_4Cl)	Merck, Darmstadt, Germany
Benzo[a]pyrenediepoxide (BPDE)	Biochemisches Institut für Umweltcarcinogene, Großhansdorf, Germany
5-Bromo-2'-deoxyuridine (BrdU), minimum 99 %	Sigma-Aldrich, St. Louis, USA
Cleaning Solution for Flow Systems	Partec, Münster, Germany
Colcemide (Demecolcin), 10 µg/ml	Biochrom, Berlin, Germany
Cytochalasin-B from *Helminthosporium dermatioideum*	Sigma-Aldrich, St. Louis, USA
4',6-Diamidino-2-phenylindoldihydrochloride (DAPI)	Sigma-Aldrich, St. Louis, USA

DAPI Staining Solution	Partec, Münster, Germany
Dimethyl sulphoxide (DMSO)	Serva, Heidelberg, Germany
Disodiumhydrogenphosphate-2-hydrate (Na_2HPO_4*2 H_2O)	Merck, Darmstadt, Germany
Ethylendiamide tetraacetate (EDTA)	Fluka Chemie, Neu-Ulm, Germany
Ethanol, absolute	Carl Roth, Karlsruhe, Germany
Eukitt	Kindler, Freiburg, Germany
Eva Green	Biotium Inc., Hayward, USA
Fetal Calf Serum (FCS)	Biochrom, Berlin, Germany
Gentamycin	Biochrom, Berlin, Germany
Giemsas Azur-Eosin-Methylenblaulösung (Giemsa)	Merck, Darmstadt, Germany
Methanol	Sigma-Aldrich, St. Louis, USA
Oligonucleotides, custom	Biomers, Ulm, Germany
Phytohaemagglutinin (M-form)	Invitrogen, Life Technologies, Carlsbad, USA
Potassium chloride (KCl)	Merck, Darmstadt, Germany
Potassium hydrogencarbonate ($KHCO_3$)	Merck, Darmstadt, Germany
Potassium dihydrogenphosphate (KH_2PO_4)	Merck, Darmstadt, Germany
Proteinase K	Sigma-Aldrich, St. Louis, USA
Sodium chloride (NaCl)	AppliChem, Darmstadt, Germany
Sodium chloride (NaCl) 0.9 %	B.Braun, Melsungen, Germany
Sodium dodecylsulfate (SDS)	Serva, Heidelberg, Germany
Trisodium citrate dihydrate (Na_3-Citrate * 2 H_2O)	Merck, Darmstadt, Germany
Trishydroxy methylaminomethane (TRIS)	Sigma-Aldrich, St. Louis, USA
VectaShield® Mounting Medium for Fluorescence	Vector Labs, Burlingame, USA

Culture Media

RPMI-1640 with L-Glutamine and 25 mM Hepes	Invitrogen, Life Technologies, Carlsbad, USA
RPMI-1640 with GlutaMax	Invitrogen, Life Technologies, Carlsbad, USA

Buffers and Solutions

Cytochalasin-B (Cyt-B) solution	5 mg	Cytochalasin-B
	10 ml	DMSO
DAPI solution	20 ml	SSC buffer (20x)
	ad 100 ml	aqua dd.
	100 µl	DAPI (1 mg/ml)
Fixative 1	1 part	acetic acid
	5 parts	methanol
	6 parts	NaCl (0.9%)

Fixative 2	1 part 5 parts	acetic acid methanol
Fixative S	1 part 3 parts	acetic acid methanol
Giemsa staining solution	92 ml 8 ml filtrated	Sorenson buffer Giemsa
Ibraimov solution, modified	3 ml 5 ml 92 ml	methanol acetic acid aqua dd.
Lysis buffer	8.28 g 1.0 g 34 mg ad 1000 ml pH ad 7.4 (autoclaved, ac.)	NH_4Cl $KHCO_3$ EDTA aqua dd.
Phosphate buffered saline (PBS) buffer	8 g 0.2 g 0.2 g 1.42 g ad 1000 ml pH ad 7.2 (ac.)	NaCl KCl KH_2PO_4 * 2 H_2O Na_2HPO_{4*} aqua dd.
Potassium chloride (KCl) solution, 0.4 %	2 g ad 500 ml	KCl aqua dd. (ac.)
Potassium chloride (KCl) solution, 0.56 %	2.8 g ad 500 ml	KCl aqua dd. (ac.)
Saline-EDTA buffer	4.39 g 8.41 g ad 1000 ml pH ad 8.0	NaCl ETDA aqua dd.
Saline sodium citrate (SSC) buffer, 20x	3 M 0.3 M pH ad 7.0 (ac.)	NaCl Na_3-Citrate * 2 H_2O
Saturated sodium chloride (NaCl) solution	35.06 g ad 100 ml	NaCl aqua dd.
Sodium dodecylsulfate (SDS) solution, 20%	20 g ad 100 ml	SDS aqua dd.

Sorenson solution A	0.33 M	$Na_2HPO_4 * 2 H_2O$ (ac.)
Sorenson solution B	0.33 M	KH_2PHO_4 (ac.)
Sorenson buffer	5 ml 5 ml ad 100 ml	Sorenson A Sorenson B aqua dd.
TRIS-EDTA buffer	121 mg 37 mg ad 100 ml pH ad 8.0	TRIS EDTA aqua dd.

Reagent Systems

Blood & Cell Culture DNA mini Kit	Qiagen, Hilden, Germany
NucleoSpin® Blood Quick Pure	Machery-Nagel, Düren, Germany
PowerPlex 16®	Promega, Madison, USA
TaqMan® AmpliTaq Gold Master Mix	Applied Biosystems, Life Technologies, Carlsbad, USA
Taqman® Genotyping Master Mix	Applied Biosystems, Life Technologies, Carlsbad, USA
Taqman® Custom Genotyping Assays	Applied Biosystems, Life Technologies, Carlsbad, USA
Taqman® SNP Predesigned Genotyping Assays	Applied Biosystems, Life Technologies, Carlsbad, USA

Laboratory Equipment

ABI-PRISM® 3100 Genetic Analyzer	Applied Biosystems, Life Technologies, Carlsbad, USA
Axioplan (fluorescence microscope)	Carl Zeiss, Oberkochen, Germany
Axioplan 2 (fluorescence microscope)	Carl Zeiss, Oberkochen, Germany
Fast Real-Time PCR System 7900 HT	Applied Biosystems, Life Technologies, Carlsbad, USA
Biofuge pico	Heraeus Instruments, Hanau, Germany
Flow cytometer CCA	Partec, Münster, Germany
Flue	w.r.t. Laborbau, Stadtlohn, Germany
Gammacell 2000 Cs-137 source, 44.8 Tera-Becquerel on January 30th, 2007	Molsgaard Medical, Heorsholm, Denmark
GeneAmp 9600 PCR System	Perkin & Elmer, Waltham, USA
Incubator 37 °C	Heraeus Instruments, Hanau, Germany
Incubator 37 °C, 5 % CO_2	Heraeus Instruments, Hanau, Germany
Light Microscope	Carl Zeiss, Oberkochen, Germany
Megafuge 1.0R	Thermo Fisher Scientific, Waltham, USA
Nanodrop Photometer	Implen, Munich, Germany
Packard Multiprobe II	Perkin & Elmer, Waltham, USA

Thermomixer 5436	Eppendorf, Hamburg, Germany
Uniflow UVUB 1200	Nunc, Wiesbaden Germany
Vortex Genie 2	Bender & Hobein, Zurich, Switzerland

Plastic ware and other laboratory commodities were obtained from Nunc, Roskilde, Denmark, and from Sarstedt, Nümbrecht, Germany to a major part. MicroAmp Optical 384-Well Reaction Plates and MicroAmp Optical Adhesive Films for use in the Fast Real-Time PCR System 7900 HT were obtained from Applied Biosystems, Life Technologies, Carlsbad, USA.

Software

ABI 7900 HT Sequence Detection Systems (SDS) Versions 2.3 rev. C and 2.4.	Applied Biosystems, Life Technologies, Carlsbad, USA
ABI-PRISM® 3100 Genetic Analyser Data Collection Software v. 1.1	Applied Biosystems, Life Technologies, Carlsbad, USA
ABI-PRISM® GeneScan v. 3.7	Applied Biosystems, Life Technologies, Carlsbad, USA
ABI-PRISM® Genotyper v. 2.0	Applied Biosystems, Life Technologies, Carlsbad, USA
Applied Biosystems High Resolution Melting (HRM) Software v. 2.0	Applied Biosystems, Life Technologies, Carlsbad, USA
Haploview v. 4.2	Broad Institute, Cambridge, USA
Metafer Ikaros Karyotyping System v. 5.0	Metasystems, Altlussheim, Germany
Metafer 4 – Msearch versions 2.6.9, 3.1.2 and 3.6.4	Metasystems, Altlussheim, Germany
Microsoft Office Excel 2007	Microsoft, Redmont, USA
Microsoft Office Access 2007	Microsoft, Redmont, USA
PASW v. 18	SPSS Inc., Chicago, USA
SAS v. 9.1	SAS Institute Inc., Cary, USA
StatView v. 5.0.1	SAS Institute Inc., Cary, USA
WinMDI v. 2.9	Flow Cytometry Core Facility, The Scripps Research Institute, La Jolla, USA

Internet Resources

http://www.expasy.org	Expert Protein Analysis System, Swiss Institute of Bioinformatics, Switzerland [71]
http://www.genenames.org	HUGO Gene Nomenclature Committee
http://hapmap.ncbi.nlm.nih.gov/	International HapMap Project [97]
http://www.hgvs.org	Human Genome Variation Society

http://www.ncbi.nlm.nih.gov	National Center for Biotechnology Information (NCBI), National Institutes of Health, USA
http://www.ncbi.nlm.nih.gov/biosystems	NCBI BioSystems database [72]
http://www.ncbi.nlm.nih.gov/gene	NCBI Entrez Gene database [142]
http://www.ncbi.nlm.nih.gov/pubmed	NCBI PubMed literature database
http://www.ncbi.nlm.nih.gov/snp	NCBI single nucleotide polymorphism (SNP) database [201]
http://genetics.bwh.harvard.edu/pph2/	Polymorphism Phenotyping v2 [1]
http://ihg2.helmholtz-muenchen.de/cgi-bin/hw/hwa1.pl	Tests for deviation from Hardy-Weinberg Equilibrium , Institute of Human Genetics, Technical University Munich
http://www.uniprot.org/uniprot	UniProt-Knowledgebase/SwissProt database [24,229]

2.3 Experimental Methods

2.3.1 Cell Culture Procedures

Set up of short-time whole blood cultures

Blood from probands was withdrawn into 7.5 ml monovettes with heparin as anticoagulant, and 500 µl of whole blood were added to 5 ml of growth medium in flat bottom culture tubes. The horizontal incubation on the flat surface provides better growth conditions for the lymphocytes compared to the standard cylindrical cell culture tubes.

The growth medium for whole blood cultures contained RPMI-1640 with L-Glutamine and 25 mM Hepes as CO_2 buffer, supplemented with 20 % fetal calf serum, 1 % of the antibiotic gentamycin and 1% phytohaemagglutinin (M-form) for stimulation of lymphocyte growth. 5 ml of the growth medium were portioned into flat bottom culture tubes and stored at -20 °C until thawing at 37 °C before the addition of whole blood. The different DNA repair assays commenced immediately and the cultures were treated according to the specific requirements of each test.

Generally, a total of six cultures were set up per individual. One culture each was set up for the baseline and the induced micronucleus assay as well as for the baseline and the induced sister chromatid exchange assay. Two further cultures were used for the mitotic delay assay, one with and one without damage induction. For twin pairs, the cultures for all assays were set up in duplicate, resulting in twelve cultures per individual.

For numerous probands from the “Freiburg” and “Ulm” cohorts, only the radiation induced micronucleus assay was performed due to low volumes of the available blood samples or experimental reasons.

Set up and culture of lymphoblastoid cell lines

B-Lymphocytes can be transformed into permanent lymphoblastoid cell lines (LCL) by transfection with the Epstein-Barr-Virus [74]. This procedure was kindly conducted by the Institute of Virology, University of Ulm, Germany. LCL lines were established from 7.5 ml of peripheral whole-blood samples from the respective probands.

As growth medium for lymphoblastoid cell lines (LCL) served RPMI-1640 GlutaMax medium, supplemented with 10 % FCS and 1 % gentamycin, which was stored at 4 °C. LCL cultures are suspension cultures and were kept in 15 ml growth medium in 75 ml (T_{75}) cell culture flasks which were incubated vertically at 37 °C and 5 % CO_2. Passaging proceeded upon the depletion of the growth medium, which was accompanied by a color change of the pH indicator. Typically each 3 – 4 days 5 ml of a proliferating culture were transferred into a new T_{75} culture flask with 10 ml fresh growth medium.

One day past passaging, aliquots were taken and further used as each of the respective DNA repair assays required.

2.3.2 Micronucleus Assay

With each of the applied DNA repair assays, an individual cellular DNA repair capacity was assessed by measuring different DNA repair outcomes in the presence of a defined amount of damage and after a defined time to repair this damage.

Micronuclei indicate the failure of the cellular DNA double strand break repair system before the entry into mitosis. Acentric chromosomal fragments resulting from unrepaired DNA double strand breaks which are not incorporated into one of the daughter nuclei become enveloped with a nuclear membrane and are visible as additional micronuclei (MN) that can easily be counted.

One culture was used to assess the MN frequency resulting from baseline damage. The MN frequency after the induction of DNA double strand break damage by exposure to 2 Gray (Gy) of ionizing radiation immediately after culture setup was assessed in a separate culture. After a defined time (44 h) provided to both lymphocyte cultures for the stimulation of lymphocyte growth and to repair the present damage before the entry into the first mitosis, a cell cycle block was inflicted by addition of Cytochalasin B (Cyt-B). This agent inhibits the polymerization of actin

filaments and thus permits the reconstitution of the daughter nuclei but prohibits cell division. Thus, dividing cells that have completed one mitosis within 24 h after Cyt-B addition are visible as binucleated cells (BNC, see *Figure 02*). By scoring MN only in BNC, the MN frequency is standardized on cells with an active cell cycle and thus an active DNA repair.

After slide preparation and chromatin staining with 4'-6-Diamindino-2-phenylindoldihydro-chloride (DAPI), MN scoring in BNC was performed with an automated imaging system and an individual baseline or induced MN frequency was determined. These frequencies indicate the amount of unrepaired DNA DSB damage and are thus measures of the individual cellular DNA repair capacity.

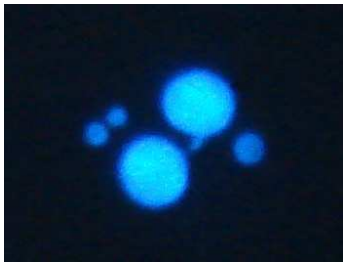

Figure 02: **Binucleated cell with three micronuclei and a nucleoplasmic bridge between the two nuclei. Chromatin was stained with 4'-6-Diamindino-2-phenylindoldihydrochloride (DAPI).** Own image.

Irradiation

Immediately after culture setup, the whole blood culture for the induced MN assay was irradiated with 2 Gy of ionizing radiation from a Caesium-137 source at a dose rate of 1 Gy/min at room temperature. The culture for the baseline assay was left untreated. The MN assay commenced subsequently by incubating the cultures at 37 °C.

Whole-blood culture and Lymphocyte Harvest

Following culture setup and, if appropriate irradiation, the cultures were incubated for 44 h at 37 °C. At this point of time Cyt-B, dissolved in dimethyl sulphoxide, was added to a final concentration of 6 µg / ml, resulting in the formation of binucleated lymphocytes. After further 24 h at 37 °C, lymphocytes were harvested by standard procedures. In brief, after centrifugation for 4 min at 2000 rpm and room temperature, the growth medium was removed and hypotonic treatment commenced by adding of 5 ml 0.56 % KCl at 4 °C. The culture was resuspendend, immediately centrifuged, the supernatant subsequently removed, and 5 ml of Fixative 1 at -20 °C were added. After 10 min incubation at room temperature, subsequent centrifugation and removal of the supernatant, 5 ml of Fixative 2 at -20 °C were added. The procedure was repeated

three times without incubation between addition of Fixative 2 and centrifugation. The cell suspensions were stored in the culture tube at 4 °C least for one night.

Slide Preparation and staining

Slides were prepared by dropping the cell suspension on slides that were previously wetted and wiped to generate a thin water film. The organic Fixative 2 displaces the water film, resulting in the swelling and spreading of the cells without disrupting them. With a light microscope, the dried slides were visually controlled for size and distribution of the binucleated cells, as the subsequent automated scoring of micronuclei relies on the recognition of BNC with a defined range of morphologic parameters. Chromatin was stained with 1 µg / ml DAPI. The slides were incubated for 10 min in the DAPI solution containing 100 µl DAPI (1 mg / ml) in 100 ml 4-fold (4x) SSC buffer, subsequently washed and dried and afterwards sealed with VectaShield®.

Micronucleus counts

For the scoring of micronuclei in BNC, an automated imaging system was used. The system allows the fast and reproductive scoring of a large number of cells on up to eight slides at once, and thus avoids the problems of the variability introduced by different human scorers and by the limited number of BNC that can be scored by human scores in a reasonable amount of time. It is based on the Axioplan 2 fluorescence microscope with a charge coupled device camera. The Metafer Msearch 4 software scans the complete slide area and each image is analyzed for the presence of BNC and micronuclei in their vicinity. The classifiers comprise defined parameters concerning the morphology of BNC and MN as well as the distance of the two nuclei of the BNC and the distance of MN to these. The parameters were modeled according to the scoring criteria of the Human Micronucleus Project [62,234]. The parameters of the classifier used for the scoring of binucleated cells and micronuclei from short-time whole-blood cultures (“STC”) are listed in detail in *Attachment B*. All scored BNC with two or more MN were visually re-examined to remove obvious artifacts. The system was updated twice during the course of the study (Metafer MSearch 4 versions 2.6.9, 3.1.2 and 3.6.1), each update leading to slight differences in scoring performance. Because of these and other study-related differences of MN frequencies between the cohorts, they were normalized on the respective control means of each cohort (see *3.1*).

At least two slides were generated from each cell suspension. The resulting micronucleus frequency was determined from the sum of all scored BNC and MN on all slides and is given as MN per 1000 BNC. The Metafer Msearch system scans the complete slide area, resulting in a several fold variation in the number of scored BNC per individual due to differences in culture

growth and the density of the cells on the slides. An average number of 3854 BNC was scored for each MN frequency value.

Modifications for the MN assay on lymphoblastoid cell lines

2 ml of a growing lymphoblastoid cell line (LCL) were added to 3 ml RPMI-1640 GlutaMax medium, supplemented with 10 % FCS and 1 % gentamycin in a 25 ml cell culture flask. The incubation started subsequently at 37 °C with 5 % CO_2, as the growth medium contained no CO_2 buffer. Immediately after culture setup, the damage-induced cultures were irradiated with 2 Gy of ionizing radiation as described. As no time for the induction of lymphocyte growth was required, Cyt-B was added to a final concentration of 6 µg / ml after 24 h of incubation. After additional 24 h, culture harvest, slide preparation and micronuclei scoring followed as described.

Procedures for the study of the reproducibility of the MN assay

A study to elucidate the origin of the intrinsic test variability inherent to the micronucleus assay was performed. *Figure 03* on *Page 26* gives an overview about the procedure.

One sample of peripheral blood from a healthy proband aged 66 years was used for the study on the baseline micronucleus frequency and one further sample from the same person for the induced MN frequency. For each type of assay, the two investigators (I-A and I-B) each set up 20 whole-blood cultures from the blood samples (40 cultures in total) within a time span of 10 minutes and performed the micronucleus assay as described.

From the induced cultures, two slides were prepared from each of the 20 cultures of investigator A by one preparator (I-A/P1), as did another preparator for the cultures of investigator B (I-B/P2), resulting in a total of 80 slides. Each slide was scored twice with the automated imaging system, resulting in 160 counts with different histories of assay investigator (1-2), culture (1-20), slide (1-2) and count (1-2).

For the study of the variability of the baseline MN frequency, the cultures were set up by two investigators and a total of 80 slides (2 per culture) were generated by two preparators as described for the induced assay (I-A/P1 and I-B/P2). In addition, the two slide preparators generated two further slides per culture from the same cell suspensions in a cross-over experiment two weeks later, whereat preparator 2 generated the slides from the 20 cultures from investigator A (I-A/P2), and reverse (I-B/P1), totaling again in 80 slides. One additional week later, a further slide preparator generated another two slides from the cell suspensions of the 20 cultures of investigator A (I-A/P3) and the 20 cultures of investigator B (I-B/P3) that were used in the two previous rounds. This resulted in three rounds of slide preparation, in each of which 80 slides were generated. Each slide was again counted twice. Thus, 480 counts with different

histories of assay investigator (1-2), culture (1-20), slide preparation round (1-3), slide (1-2) and count (1-2) were available. *Figure 03 (A)* provides a basic overview of the study on the baseline MN frequency, *Figure 03 (B)* of the study on the radiation induced MN frequency.

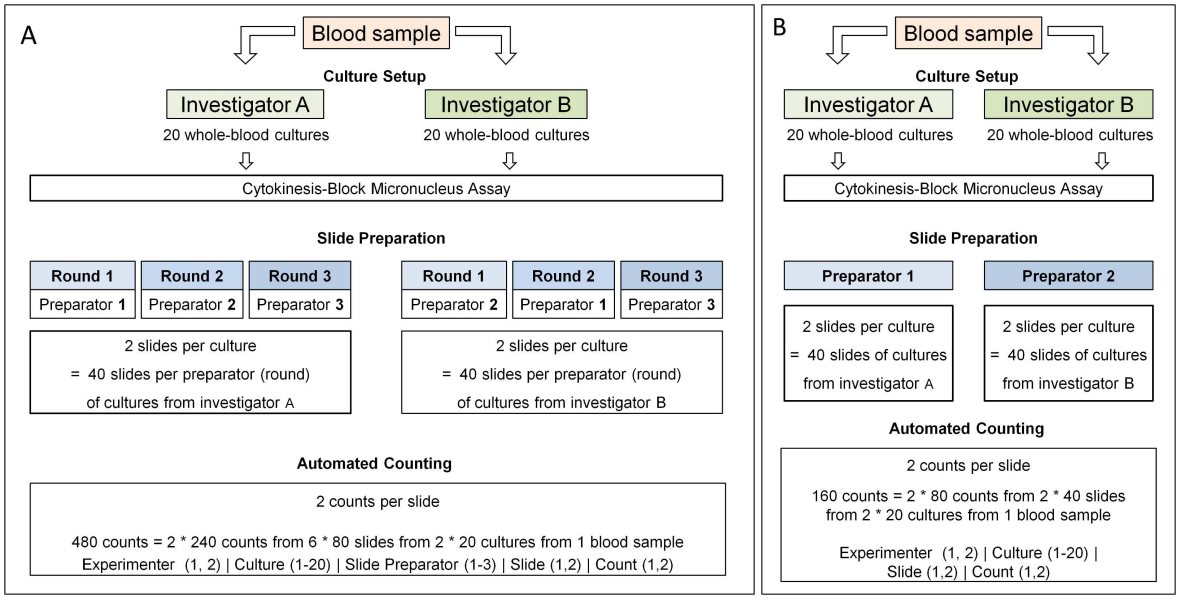

Figure 03: Overview of the variability study procedure for the baseline (A) and the radiation induced (B) micronucleus frequency.

2.3.3 Sister Chromatid Exchange Assay

Theoretical Background

This test is based on the visualization of homologous recombination events between sister chromatids by differential staining of the chromatids of metaphase chromosomes. The staining is achieved by incorporation of the thymine substitute 5-Bromo-2-deoxyuridine (BrdU) into the DNA, which is added to the cultures immediately after culture setup. After the replication, the DNA in each chromatid consists of one substituted and one not substituted strand. After the second replication, the chromosomes consist of one half and one fully substituted chromatid. The altered properties of the BrdU-substituted chromatin manifest after ultraviolet irradiation and subsequent Giemsa staining of chromosomes in the second metaphase, where the fully substituted chromatids exhibit a weaker staining compared to half substituted chromatids. Thus the sister chromatids can be distinguished and recombination events between them can be seen as exchanges of the respective staining intensity (see *Figure 04* on *Page 27)*.

In addition to the measurement of the baseline sister chromatid exchange (SCE) frequency, SCE were also induced by adding Benzo[a]pyrenediepoxide (BPDE) at culture setup. BPDE is a

carcinogen that forms covalent DNA adducts which have to be removed during DNA replication. This process enhances the formation of SCE.

Metaphase arrest was achieved by adding Colcemide prior to culture harvest. Colcemide inhibits the polymerization of tubulin and thus blocks the formation of the spindle apparatus. The cultures were subsequently harvested and metaphase preparations were generated. After differential staining, SCE frequencies of untreated and BPDE-induced cultures were counted. These frequencies do provide a measure for the individual DNA repair capacity by indicating the amount of sister chromatid recombination events induced by the damage present.

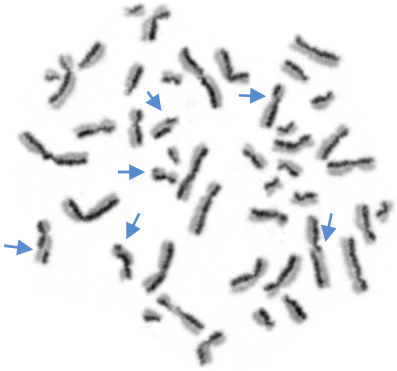

Figure 04: **Metaphase chromosomes with differential stained chromatids.** Switches of dark and light staining between the chromatids (indicated by blue arrows) are sister chromatid exchanges. Own image.

BPDE treatment

For DNA damage induction for the induced SCE assay, BPDE was added to a final concentration of 0.5 µM immediately after culture setup and the cultures were incubated at 37 °C for 30 min. To obtain measures for the repair of the induced DNA damage rather than for the level of induction itself, BPDE was then removed together with the growth medium after centrifugation for 2 min at 800 rpm at room temperature. The remaining culture sample was transferred into a new flat bottom tube containing fresh growth medium which was previously warmed to 37 °C.

Whole-blood culture and lymphocyte harvest

Following culture setup and, if appropriate, mutagenic treatment, BrdU was added to both untreated as well as the BPDE induced cultures to a final concentration of 20 µg / ml. The cultures were incubated for 68 h at 37 °C. To increase metaphase yield, Colcemide was added to a final concentration of 0.1 µg / ml and the cultures were incubated for additional 2 h.

Lymphocytes were harvested by standard procedures. In brief, after centrifugation for 4 min at 2000 rpm and room temperature, the growth medium was removed and 5 ml 0.4 % KCl at 37 °C were added for hypotonic treatment. The culture was resuspended and incubated for further 30

min at 37 °C, subsequently centrifuged and the supernatant was removed. Five ml modified Ibraimov solution at 4 °C were added, the culture was resuspendend and immediately centrifuged. The procedure was repeated once with 5 ml methanol at -20 °C and twice with Fixative S at -20 °C. The cell suspensions were stored in the culture tube over night at 4 °C.

Differential staining and metaphase counts

Metaphase preparations were generated by dropping the cell suspension onto wet slides. The organic fixative S displaces the water and in the process the metaphase chromosomes are spread. At least two slides were prepared from each culture. The slides were visually controlled for metaphase spread with a light microscope and dried over night at 37 °C. The slides were then incubated for 30 min in Sorenson buffer under UV light (254 nm) at room temperature and subsequently incubated for 1 h in 2x SSC buffer at 60 °C. Afterwards, the cells were differently stained by incubation in Sorenson buffer with 8 % Giemsa for 15 min at room temperature. The slides were washed, dried for 2 h at room temperature and subsequently sealed with Eukitt. At least two slides were prepared from each culture.

Metaphases were evaluated with the Ikaros Karyotyping System 5.0 software. The metaphases were photographed with a charge coupled device camera mounted on a Axioplan fluorescence microscope and displayed on the screen, which enabled the visual scoring of SCEs. Only those metaphases with a complete karyotype were taken into account, and SCEs were scored on at least 20 completely differentially stained metaphases per culture. The average number of SCEs per metaphase was reported as the individual SCE frequency.

Modifications for the SCE assay on lymphoblastoid cell lines

Two ml of a growing lymphoblastoid cell line (LCL) were added to 3 ml RPMI-1640 GlutaMax medium, supplemented with 10 % FCS and 1 % gentamycin, either in a 25 ml cell culture flask or, when required for DNA damage induction in a cell culture tube. The incubation proceeded at 37 °C with 5 % CO_2.

For the damage induction, immediately after culture setup BPDE was added to a final concentration of 0.5 µM and the culture was incubated for 30 min at 37 °C and 5 % CO_2. BPDE was subsequently removed together with the growth medium after centrifugation for 2 min at 800 rpm and room temperature. The remaining culture sample was then transferred to a 25 ml cell culture flask containing 5 ml growth medium. The baseline culture was left untreated after culture setup and incubated at 37 °C and 5 % CO_2. Subsequently, BrdU was added to both cultures to a final concentration of 20 µg / ml. The cultures were subsequently incubated for only 44 h before the addition of colcemide, because in contrast to the whole blood cultures no time for growth

stimulation of lymphocytes was required. Culture harvest, slide preparation, staining and SCE counts were done as described for whole-blood cultures.

2.3.4 Mitotic Delay Assay

This assay visualizes changes in the cell cycle upon the infliction of DNA double strand breaks. These induce a cell cycle arrest prior to the entry into mitosis. Because cells complete the DNA synthesis to facilitate the DNA double strand break repair, the damage-induced cultures show an accumulation of cells in the G_2 phase and a depletion of cells in S phase, compared to an untreated control culture. Here, DNA damage was inflicted by 2 Gy of ionizing irradition in a proliferating whole-blood culture after 54 hours of incubation. The mitotic delay was determined 18 hours later by comparing the ratio of cells in G_2 phase and cells in S phase of the damage-induced culture to the same ratio of the untreated culture.

For this purpose, the cells were stained with DAPI and analyzed using a flow cytometer. The DNA content of single cells in a suspended culture, which equals to the intensity of their staining, is measured and assigned to one of 1024 channels. Cells in the the G_2 phase have twice the amount of chromatin as cells in the first gap phase (G_1 phase) of the cell cycle, and the chromatin content of cells in the S phase is in between. The number of cells in each channel is counted, resulting in a characteristic DNA histogram which allows to distinguish the cell cycle phases and to determine the fraction of cells in a respective phase (see *Figure 05* on *Page 31*).

By indicating the efficiency of the induction of a cell cycle arrest upon DNA damage and its subsequent repair, this assay does provide a measure for the individual cellular DNA repair capacity.

Whole-blood culture and staining

Following culture setup and 54 h hours of incubation at 37 °C, one whole blood culture was irradiated with 2 Gy of ionizing radiation from a Cs-137 source at a dose rate of 1 Gy / min at room temperature, while the other culture was left untreated. The MD assay continued subsequently by incubating both cultures at 37 °C for additional 18 h, after which both cultures were harvested for flow cytometric analyses. From each culture, an aliquot of 500 µl was added to 1.5 ml Partec DAPI staining solution at 4 °C in flow cytometer tubes. Besides DAPI, the staining solution contains detergents which increase the permeability of the cell and nuclear membranes and thus allow the dye access to the chromatin of the otherwise intact cells. To ensure the

complete and uniform staining of the chromatin, the tubes from both cultures were subsequently sealed with parafilm and kept at 4 °C at least for one night.

Flow cytometry

The Partec CCA Flow cytometer was used to analyze the cell cycle distribution of the untreated and the damage-induced culture. It measures the staining intensity and thus chromatin content of each cell and assigns it to one of the 1024 available channels, and a DNA histogram from the number of scored cells in the respective intensity channels is generated. Beam focus and gain level were optimized to generate the most accurate DNA histograms with minimized widths of the G_1 and G_2 phase peaks. Upper and lower intensity thresholds were defined to exclude artifacts like cell conglomerates (very high staining intensity) or fragmented nuclei (very low staining intensity). Detailed preferences are shown in *Attachment C*. Prior to the measurement, the samples were thoroughly resuspended. For each culture, at least 20,000 cells were scored and classified according to their staining intensity and the channel data generated by the flow cytometer was stored in files. Before and after each measure, the Partec Cleaning Solution For Fluorescence Systems was utilized to remove remaining cells from the capillaries.

DNA histogram analysis

With the WinMDI 2.9 software the resulting files were converted into text files, which were in turn imported into the Microsoft Office Excel software, with each row representing one of the 1024 channels of the DNA histogram and displaying the count of cells in the respective channel. The borders of the G_1 and S phases as well as of the S and G_2 phases were defined by calculating them from the G_1 and G_2 phase maxima, since the base widths of the peaks cannot be assessed directly because they overlap with the S phase (see *Figure 05*).

First, the channel with the maximum cell count value of the G_1 peak (G_1 max) was determined. Subsequently the channels with the lowest count above the half of the G_1 maximum value ($G_1/2$) were defined on both sides of the G_1 peak, one towards the S phase ($G_1/2$-direct) and the other one towards the sub G_1 side ($G_1/2$-reverse). The range of channels between the $G_1/2$-direct channel and the $G_1/2$-reverse channel was determined, which resembles the width of the G_1 peak at the half of its height. This range was added to the channel number of the G_1 max channel and the so defined channel (G_1-S) was interpreted as the border of the G_1 and S phases, since the true base width of the G_1 peak cannot be determined directly from the histogram.

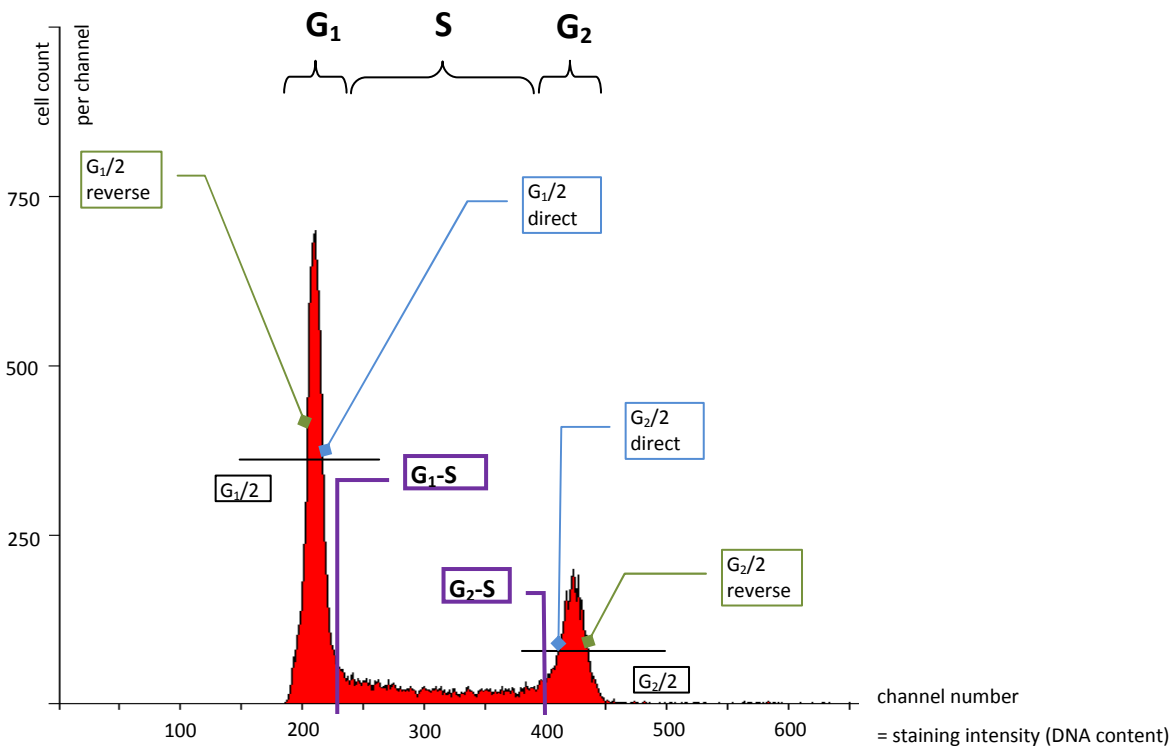

Figure 05: **Histogram of the distribution of the deoxyribonucleic acid (DNA) content of a short-time whole blood culture**. The cell counts are shown on the ordinate, the DNA content on the abscissa. The first gap (G_1), DNA synthesis (S) and second gap (G_2) phases of the cell cycle are indicated, as well as the hallmarks used for the calculation of the borders between G_1 and S phases and between S and G_2 phases, as these cannot be determined directly. $G_1/2$ and $G_2/2$: the half cell count of the maximum count of the G_1 or G_2 phases, respectively. $G_1/2$ reverse and $G_2/2$ direct: channels with the lowest cell count above the half of the maximum counts of the G_1 phase. $G_2/2$ direct and $G_2/2$ reverse: channels with the lowest count above the half of the maximum count of the G_2 phase. G_1-S: border of the G_1 and S phases. G_2-S: border of the G_2 and S phases.

The procedure was repeated at the G_2 peak, however in the last step, the range of channels was subtracted from the G_2 peak channel, resulting in the G_2-S channel that was interpreted as the border of the S and G_2 phases.

With this information, and the cell counts in each channel, the number of cells in the S and G_2 phases was calculated for each culture.

Mitotic delay index

The mitotic delay (MD) index was generated from the ratios of cells in the S and G_2 phases of the cell cycle in the untreated (G_2/S-b) and the damage-induced culture (G_2/S-i). It is calculated as ratio from these ratios of the damage-induced and the untreated culture: ([G2/S-i] / [G2/S-b]).

Modifications for the MD assay on lymphoblastoid cell lines

Two ml of a growing lymphoblastoid cell line (LCL) were added to 3 ml RPMI-1640 GlutaMax medium, supplemented with 10 % FCS and 1 % gentamycin in a 25 ml cell culture flask. As the

cells were already proliferating, the damage induced culture was immediately irradiated with 2 Gy, while the other culture was left untreated. The incubation started subsequently at 37 °C with 5 % CO_2 for 18 h. After this time, culture harvest, staining and flow cytometric analysis followed as described for the whole-blood cultures.

2.3.5 Isolation of DNA from Whole Blood Samples

Experimental procedures

For the purpose of DNA extraction, whole blood samples from monovettes with EDTA as coagulant were used if available. Often only one blood sample in monovettes with Heparin as anticoagulant could be obtained for the use in the DNA repair assays and DNA extraction. After the cultures needed for the DNA repair assays were set up, the remainder of the blood sample was washed with PBS buffer and centrifuged at 1500 rpm for 10 min at room temperature. Subsequently the sample was treated as samples from monovettes with EDTA as anticoagulant.

The samples were added to 30 ml lysis buffer at 4 °C and incubated for 15 min on ice, thus achieving the lysis of erythrocyte membranes. The lymphocytes were pelletized by subsequent centrifugation for 10 min at 1000 rpm at 4 °C, the supernatant was completely removed and the pellet was resuspended in 10 ml lysis buffer at 4 °C. This step was repeated until all traces of hemoglobin were removed and the samples appeared white. After centrifugation, the lymphocyte pellet was resuspended for proteolysis in 5 ml saline-EDTA buffer containing 50 µl (4 U) proteinase K and 250 µl SDS (20 %) and incubated for several hours at 55 °C or overnight at 37 °C.

Subsequently, 1.8 ml saturated NaCl (6 M) were added and the sample was vortexed for 15 seconds and centrifuged min at 4000 rpm for 15 min at room temperature. The supernatant was transferred into a new tube and two volumes absolute ethanol at -20 °C were added, resulting in the precipitation of the DNA. The precipitate was retrieved from the supernatant and subsequently washed with 70 % ethanol. Finally, the DNA was suspended in 0.2 – 1 ml TRIS-EDTA buffer, depending on the approximate retrieved amount.

When only small amounts of whole-blood sample were available, commercial reagent systems (kits) were used to isolate DNA. These kits use variations of the above described techniques and are based on the specific binding of DNA to membranes with defined surface properties. For less than 2 ml of available whole-blood sample, the Nucleospin Blood Quick Pure Kit (Machery-Nagel) and for less than 200 µl the Blood & Cell Culture DNA mini Kit (Qiagen) were used after the manufacturer's instructions.

Quantification of DNA concentrations

To obtain information about the concentration and the quality of the isolated DNA samples they were diluted as necessary (typically 1:100) and their extinction was assessed in a spectral photometer at a wavelength of 260 nm, which represents the specific absorption maximum of nucleic acids. With the extinction coefficient of DNA (0.02 / µg*cm^2) and the cuvette diameter of 1 cm, a measured optical density (OD) at 260 nm (OD_{260}) of 1 represents a DNA concentration of 50 ng / µl according to the Lambert-Beer law. An additional measurement at 280 nm (OD_{280}) is informative for the degree of contamination by proteins. A quotient of OD_{260} / OD_{280} of 1.8 – 2.0 was deemed ideal.

2.3.6 Polymerase Chain Reaction (PCR)

The basis for many molecular genetic analyses is the specific amplification of a defined short stretch of the genomic DNA, which is achieved with a polymerase chain reaction (PCR). This method is based on a cyclic amplification of the target amplicon sequence based on the semiconservative replication of DNA. Each amplification cycle includes three steps.

Initially, the double stranded DNA is denatured by heating to 92-96 °C for usually 30 s, resulting in two single strands. Subsequently, a pair of synthetic short oligonucleotides, the primers, anneal to complementary sections in the genome which are defined by the respective primer sequences. These are chosen in such a way that each primer does bind on one of the two template strands, with the target amplicon sequence in between. The annealing temperature and time depend on the specific sequences and self-binding properties of each primer pair and are usually 50-60 °C for 30 s. Finally, a heat-stable DNA polymerase performs the elongation of the complimentary new strand based on the sequence of the template strand, starting from the primers by incorporation of matching desoxyribonucleotides (dNTPs). The elongation time depends on the length and the sequence of the amplicon, whereas the elongation temperature is determined by the optimum of the applied DNA polymerase, usually at 60-72 °C for ca. 30-60 s.

As each amplification cycle generates two identical copies of one DNA template, the PCR leads to an exponential synthesis of the amplicon which is delimited by the primer sequences. A typical PCR program comprises 30 – 40 amplification cycles, preceded by the initial denaturation of genomic DNA at 94 °C for approximately 3 min and followed by a final elongation of incomplete synthesis products, usually at 72 °C for 10 min.

The PCR method is applied in the context of numerous molecular genetic analyses, including genotyping analyses, and many variations of the general protocol are applied as required for each

analysis. In this work, the PCR was applied in the context of genotyping procedures as part of the TaqMan allelic discrimination and the High Resolution Melting methods as well as for microsatellite analyses.

2.3.7 Genotyping of DNA Polymorphisms

In the context of the association studies, the study population (described in *2.1*) was analyzed for the presence of known genetic variants in 15 DNA repair genes. These variants comprised 22 single nucleotide polymorphisms (SNPs) and one deletion of 21 base pairs. Detailed information of the variants and genes is shown in *Tables 10* and *11*. As more than 2000 samples of genomic DNA were available, high-throughput targeted genotyping methods were applied, which are faster and less expensive compared to DNA sequencing. The SNPs were genotyped with allelic discrimination assays based on the TaqMan® principle, while high resolution melting was applied to genotype the deletion. Both methods include a real-time PCR step and were performed with the Fast Real-Time PCR System 7900 HT (Applied Biosystems).

The experiments were conducted in a 384 well reaction plate format. Each plate contained at least 1 % (4 wells) of non-template negative controls and 1 % of sample duplicates as internal controls. Failing proband samples were repeated once. The pipetting of the respective master mixes and the DNA samples into the 384 well plates was performed with the help of the Packard Multiprobe II pipetting system (Perkin-Elmer).

2.3.7.1 TaqMan Allelic Discrimination Assay

Theoretical Background

The TaqMan principle relies on a conventional PCR, with the addition of two additional oligonucleotides, the probes, for the analysis of specific genotypes. Each of the two probes (20 – 26 nucleotides) is complimentary to only one of the alleles of the SNP. The two probes carry two different fluorescent dyes at the 5’ end, while the 3’ ends carry a quencher which inhibits the emission of a fluorescence signal by the fluorescence resonance energy transfer effect. Thus the intact probe does not emit a fluorescence signal. The annealing step of the PCR enables the hybridization of the conventional primers at the end of the amplicon while the fluophore-labeled probes hybridize to the single stranded DNA templates at the SNP site. During the subsequent elongation, which starts at the conventional PCR primers, the DNA polymerase encounters the

hybridized probe and cleaves it due its 5'-3'-exonuclease activity. In the process, both the fluorescence dye and the quencher are released and disjoined, disrupting the fluorescence resonance energy transfer and thus leading to a fluorescence signal.
The discrimination between the three different possible genotypes relies on the quantification of the intensities of the two different fluorescence signals, which are proportional to the amount of the respective probes that were disintegrated during the PCR reaction. If a proband carries a homozygous genotype, only one of the two probes can hybridize to the SNP allele, leading to the release of only one of the fluorescent dyes. A heterozygous genotype allows the hybridization of both probes, thus leading to the release of both dyes.

Experimental procedures

The composition of the reaction mix for one sample is shown in *Table 02.*

Table 02: Composition of the reaction mix for the TaqMan® allelic discrimination assay. DNA: deoxyribonucleic acid.

Component	Volume
TaqMan® Genotyping Master Mix	2.25 µl
TaqMan® Genotyping Assay	0.11 µl
H_2O	1.14 µl
DNA template (25 ng / µl)	1.00 µl
Total:	**4.50 µl**

The Taqman® Genotyping Master Mix contains the DNA polymerase, dNTPs, the reaction buffer and a fluorescent dye (ROX) as internal quantification standard. The specific Taqman® Genotyping Assays for each SNP include the two conventional PCR primer oligonucleotides for the PCR as well as the two fluorescence-labeled probe oligonucleotides for the allelic discrimination. These assays were either available as predesigned assays for the majority of the studied SNPs or custom-made upon request for the remaining SNP, as shown in *Table 03* on *Page 36*. Further information about the SNPs and their genomic location are shown in *Tables 10* and *11* on *Pages 57* and *58*.

All components excluding the DNA templates were combined in master mixes, of which 3.5 µl were distributed in the required wells of the 384-well reaction plates. Subsequently, 1.0 µl of the specific DNA templates (25 ng / µl) was added to each well and the plates were sealed with transparent adhesive plate seals (Applied Biosystems).

The PCR conditions were set according to the manufacturer's protocol with the ABI 7900 HT Sequence Detection Systems software and are shown in *Table 04* on *Page 36*.

Table 03: Catalogue numbers of the selected variants from the National Center of Biotechnology Information (NCBI) single nucleotide polymorphism database (dbSNP) and the catalogue numbers of TaqMan® Genotyping Assays by Applied Biosystems, which were used for the genotyping of the variants. Also the assay type is given.

NCBI dbSNP Catalogue Number [a)]	Applied Biosystems Assay Catalogue Number	Assay Type
rs34099167	AHN0YPO	custom
rs34540355	AHO9WVW	custom
rs72478986	AH5H85G	custom
rs28997576	AHLI146	custom
rs17857107	C__25744255_10	predesigned
rs13044759	C___2507695_10	predesigned
rs61376834	C__25985607_10	predesigned
rs7763565	C__25965787_10	predesigned
rs3087399	C__15793609_10	predesigned
rs3810813	C__25643861_10	predesigned
rs3827530	C__25643853_10	predesigned
rs714181	C___2264851_10	predesigned
rs14422	C__25943405_10	predesigned
rs2304103	C__15973568_10	predesigned
rs3816032	C__25971684_10	predesigned
rs4888444	C__22272621_10	predesigned
rs35123152	C__25957264_10	predesigned
rs2229361	C___1665421_10	predesigned
rs2305868	C__16191916_10	predesigned
rs4986764	C___2547422_10	predesigned
rs35087820	C__25615205_10	predesigned
rs3218786	C__25940731_10	predesigned

[a)] For detailed information on the selected variants, see also *3.2.1* and *Table 11* on *Page 58*

Table 04: Thermal cycler program for the TaqMan allelic discrimination assay. PCR: polymerase chain reaction.

Step		Temperature	Duration	Cycles
Pre read		60 °C	2 min	-
Real-time PCR	Initial denaturation	95 °C	10 min	-
	Cycle-denaturation	92 °C	15 s	45 cycles
	Annealing / Elongation	60 °C	60 s	
Post read		60 °C	2 min	-

After a pre-read step to determine the background fluorescence level followed the real-time PCR step. As the PCR polymerase has a temperature optimum reaching from 60 °C to 72 °C, both the annealing and the elongation step were jointly performed at 60 °C. In the final post-read step, the fluorescence intensities of the released dyes were quantified. These data was used to determine the individual allele constellations of the proband with the Applied Biosystems Sequence Detection Systems software. Microsoft Office Excel and Access software was used for linking the genotype data with other data of the individual probands.

2.3.7.2 High Resolution Melting of PCR Products

Theoretical Background

The high resolution melting method includes a real-time PCR in the presence of a fluorescent dye that binds highly specific to DNA-double strands but not to single stranded DNA. Upon binding, a fluorescence signal is emitted, which allows to quantify the accumulation of the defined amplicon during the real-time PCR run. In turn, if the PCR product is subsequently melted by gradually increasing the temperature, the fluorescent dye is released upon the abrupt separation of the DNA strands and the fluorescence signal decreases within a narrow temperature range.

This melting curve is specific for each PCR product, and sequence variations such as SNPs, insertions or deletions lead to characteristic deviations in the melting curves, which can be visualized and measured. Usually, both homozygous states as well as the heterozygous state of a polymorphism are distinguishable from each other.

Due to technical issues, the resolution of the melting curves of the applied Fast Real-Time PCR System 7900 HT is not fully suited for the detection of SNPs and thus only the 21 base pair deletion, whose heterozygous or homozygous presence leads to large shifts in the melting curves, was genotyped with this assay. The High Resolution Melting analyses were performed on a PCR amplicon of 78 bp in length, which included the region of the deletion.

Experimental Procedures

The reaction mix for the high resolution melting analysis of the 21-base-pair deletion is shown in *Table 05* on *Page 38* and additional information about the polymorphism is shown in *Tables 10* and *11* on *Pages 57* and *58.* The mix contains the TaqMan® AmpliTaq Gold Master Mix, which includes the DNA polymerase and the reaction buffer. The Eva Green dye is a new generation fluorescent dye whose binding affinity to double stranded DNA, in contrast to DNA single strands, is higher than for the widely used SYBR-Green dye. Thus, the amount of double stranded DNA can be measured with greater reliability, which also leads to a greater resolution of the melting curves.

All components excluding the DNA templates were joined in master mixes, of which 19 µl were distributed in the required wells of the 384-well reaction plates. 1.0 µl of the specific DNA templates (50 ng / µl) was added to each well and the plates were sealed with a transparent adhesive plate seal.

The PCR conditions were set according to the manufacturer's recommendations with the ABI 7900 HT Sequence Detection Systems software and are shown in *Table 06* on *Page 38.*

Table 05: Reaction mix for the high resolution melting analysis of the 21-base-pair deletion. DNA: deoxyribonucleic acid.

Component	Volume
TaqMan® AmpliTaq Gold Master Mix	10.0 µl
Eva Green dye, 20x	1.0 µl
H_2O	5.6 µl
Forward primer: gtgccctcagaaaatatacca, 5 µM [a]	1.2 µl
Reverse primer: ctccctgatgtaccaccaac, 5 µM	1.2 µl
DNA template (50 ng / µl)	1.0 µl
Total:	**20.0 µl**

[a]) a: adenine, c: cytosine, g: guanine, t: thymine

Table 06: Thermal cycler program for the high resolution melting analysis. PCR: polymerase chain reaction.

Step		Temperature	Duration	Cycles
Real-time PCR	Initial denaturation	95 °C	10 min	-
	Cycle-denaturation	95 °C	15 s	45 x
	Annealing	60 °C	30 s	
	Elongation	72 °C	30 s	
	Final Elongation	72 °C	5 min	-
Melting Curve	Denaturation	95 °C	2 min	
	Re-Annealing	20 °C	1 min	
	Melting start point	55 °C	15 s ramp rate: 1 %	
	Melting end point	95 °C	15 s	

The melting curves were analyzed with the Applied Biosystems High Resolution Melting Software version 2.0 according to the manufacturer's recommendations. In derivative plot, which displays the first negative derivative of the melting curve (intensity change per temperature unit), the pre- and the post-melt regions were set on the borders of the melting transition region and the software was used to determine the respective allele constellations of the individual samples. The data was exported to the Microsoft Office Excel and Access programs for combination with other individual proband data.

2.3.7.3 Microsatellite Analyses in Twin Pairs

To determine the zygote status of twin pairs, the lengths of defined microsatellite polymorphisms were analyzed. Microsatellites are short repeats of 2 – 4 nucleotides, some of which posses a large number of different alleles (repeat lengths) in the population. The extremely low probability for a particular allele constellation when several defined microsatellites are analyzed, allows the genetic identification of an individual and the assessment of kinships, including parenthood and

identity analyses. Monozygous twin pairs have identical genotypes and thus identical allele constellations, while dizygous twin pairs are supposed to have numerous different microsatellite alleles.

A PCR that encompasses the microsatellite enables the determination of the repeat length, as each repeat unit lengthens the amplicon. Genomic DNA was isolated from the twins as described (see *2.3.5*), and the commercial PowerPlex® 16 system (Promega) was used according to the manufacturer's instructions. The reagent system enables the analysis of 16 defined microsatellites in a multiplex PCR reaction. 0.5 µl of genomic DNA (50 ng / µl) were added to the reagents of the system and deionized water was added to a final volume of 25 µl. The amplification reaction was performed in a GeneAmp 9600 thermal cycler (Perkin & Elmer), according to the recommendations of the manufacturer. Specific amplicons of the respective microsatellites were identified by a capillary gel electrophoresis on an ABI PRISM® 3100 Genetic Analyser (Applied Biosystems). This was possible because the Powerplex system contains differently fluorescence-labeled primers and additionally uses differences in the length of all possible amplicons of a respective microsatellite. The allele constellations of the two individuals of each twin pair were analyzed with the ABI GeneScan and Genotyper softwares. If the constellations were identical, the pair was considered as monozygous, otherwise as dizygous.

2.4 Statistical Procedures

Descriptive statistics were generated with the Statview 5.1 software and statistical analyses were performed with the StatView 5.1 and PASW 18 software packages. For additional analyses concerning heritability estimates, the SAS 9.1 software package was used. Haplotypes were estimated using the Haploview 4.2 software.

2.4.1 Heritability Estimates

Heritability estimates allow a quantification of the genetic influence on the observed results of the respective DNA repair assays. The heritability (h^2) of a trait is defined as the proportion of its overall variation which is explained by genetic factors, with $0 \leq h^2 \leq 1$ (0: absence of genetic influence, 1: exclusive genetic determination), in contrast to environmental influences. One way to estimate heritability is in by a twin study, which uses the interpretation that monozygous twin

pairs (MZ) share all their genetic components (ratio 1.0), while dizygous twin pairs (DZ) share only half of them (ratio 0.5) and unrelated, randomly paired control individuals (CP) share no genetic components (ratio 0.0). Different approaches can be applied to estimate the heritability of each type of DNA repair assay, which do rely on comparisons of the specific variances of the respective assay results within each of the sample groups (MZ, DZ or CP). In the comparisons of MZ or DZ to controls, the 76 control individuals were used to generate 38 age- and gender-matched control pairs (see *Table 21* on *Page 76*) with the help of a randomization procedure. This procedure was repeated 100 times, and the randomization run with the median heritability was chosen. For each of the possible comparisons of CP to MZ or DZ in each of the evaluated DNA repair assays, the randomization process was performed independently.
As some individuals lacked results from baseline and induced sister chromatid exchange assays as well as from the mitotic delay assay, the number of available MZ, DZ and CP differs for these assays (see *Tables 24 – 28* on *Pages 81 – 84*).

The intraclass correlation coefficients (ICC) provide basic information about the degree of similarity of the particular DNA repair assay results between the individuals of each twin or control pair, in comparison to the entire results within the respective sample group (MZ, DZ and CP). For each assay and in each sample group, the ICC was calculated from the mean square errors among all twin pairs (MSA) and the mean square errors within each twin pair (MSW). It is given by

$$r = \frac{\mathrm{MSA} - \mathrm{MSW}}{\mathrm{MSA} + \mathrm{MSW}} \qquad \text{(Equation 01)}$$

Biometric modeling of variance components using the SAS 9.1 software package (SAS) according to Feng *et al*. [65] was the first approach to estimate heritability. Age, gender and culture setup delay were included as covariates. The applied ACE model (additive genetic, common environment, unique environment model) splits the total observed variants into three components after which it is named: (1) The individually modeled additive genetic component (A), with a genotypic correlation coefficient of 1.0 for MZ, 0.5 for DZ and 0.0 for CP, representing the shared genetic components of each group (see above); (2) The component representing common environmental influences (C) on both individuals of a pair, which was thus modeled identically for the individuals of each pair, regardless of the affiliation to MZ, DZ or CP; (3) The individual unique environmental or residual component (E) which represents non-shared environmental influences and the variation due to intrinsic or technical factors, and was thus modeled individually for each proband of MZ, DZ and CP.

These parameters of the single components were modeled in such a way that the true observed data structure of the respective DNA repair assay results was most closely resembled.

A further, more traditional method to obtain heritability estimates is the direct comparison of the summarized within-pair variances between MZ (V_{MZ}) and DZ (V_{DZ}) twins or between one of those and CP (V_{CP}) [237]. The estimates are respectively given as

$$h^2 = \frac{(V_{DZ} - V_{MZ})}{V_{DZ}} \quad , \qquad (\textit{Equation.02})$$

$$h^2 = \frac{(V_{CP} - V_{MZ})}{V_{CP}} \quad \text{and} \qquad (\textit{Equation 03})$$

$$h^2 = \frac{(V_{CP} - V_{DZ})}{V_{CP}} \qquad (\textit{Equation 04})$$

Additionally, the 95 % confidence intervals of the heritability estimates were determined and furthermore h^2 was tested against the null hypothesis of the absence of genetic factors that contribute to the observed variances (probability, P-values). As the calculation of confidence intervals and P-values are based on different approaches, they may occasionally seem inconsistent [237].

2.4.2 Linear Regression and Variance Analyses

For most individuals with DNA repair data, information about age and gender was also available. The influences of age as well as the delay between withdrawal of the blood sample and culture setup (culture setup delay) on the outcomes of DNA repair assays were evaluated by linear regression, the influence of gender was assessed by logistic regression. Furthermore, an analysis of variance (ANOVA) comprising all three factors was performed. The applied DNA repair assays were performed in parallel from the same blood sample for each individual, and linear regression was applied to elicit the degree of relationships between the individual results of these assays. Additionally, the individual correlations between the DNA repair measures generated from peripheral blood samples and those generated from the corresponding lymphoblastoid cell lines from the same blood sample of the same individual were assessed.

Regression relies on the estimation of a function for the relationship of a dependent variable to an independent predictor variable by minimizing the error squares. The regression coefficient r^2 equals the proportion of the variability of the dependent variable that is explained by the

independent variable. Linear regression estimates a linear function with a quantitative dependent variable, while in logistic regression a nominal dependent variable is used for modeling.
The null hypothesis of no relation between two variables was tested with a t-test. The resulting P-value represents the probability that the variation explained by the model is resulting from chance. The threshold for a significant deviation was generally set to $\alpha = 0.05$.

The ANOVA analysis performed with the StatView 5.1 software comprised the respective DNA repair test as dependent variable and age, gender and culture setup delay as independent factors. The single factors as well as all interactions between them were evaluated for a significant contribution to the observed variability of the dependent variable with a t-test ($\alpha = 0.05$).

For the variance analyses concerning baseline or induced micronucleus frequencies, the SAS 9.1 software package was applied to assess the proportion of variation attributable to the different histories of the micronucleus frequency counts. The linear ANOVA model included the factors slide set (or investigator for the induced MN frequency), culture number, slide number and count number (see *2.3.2*). These factors were ordered in a hierarchical fashion, and the total variance as well as the variances that the respective factors contributed is displayed.

2.4.3 Differences of DNA Repair Assay Results Between Groups

The means and distributions of the DNA repair tests were compared between breast cancer cases and controls or between carriers of different genotypes of the analyzed polymorphisms. For the study on the variability of the baseline and the induced micronucleus frequency, the means of different slide sets were compared.

The unpaired t-test provides the likelihood by which the observed differences between the means of two groups may occur by chance, which is represented by the P-value. The threshold was set as $\alpha = 0.05$, meaning a 5 % probability that the different means would occur if there was no true difference between the two groups.
The nonparametric Kolmogorov-Smirnov test evaluates if the distribution of a variable is similar for two different groups. The Wilcoxon signed rank test is a paired nonparametric test which tests if the mean differences between pairs of values equals zero. The likelihood that the observed differences may occur by chance is represented by the P-value, the threshold for significance was also set as $\alpha = 0.05$.

Box plots (see *Figure 06*) provide a graphical representation of the test results displaying the 10th, 25th, 50th, 75th and 90th percentile of a quantitative variable, which are in this context the individual results of the respective DNA repair assays.

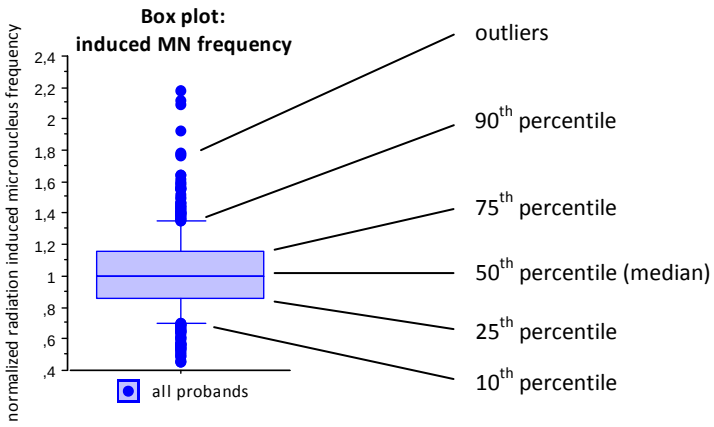

Figure 06: Description of the box plot as graphical representation of the normalized induced micronucleus (MN) frequency in all available probands as an example for a quantitative variable.

The box is bounded by the 25th and 75th percentiles, which are also called the lower and upper quartile. Thus, the box comprises the inter-quartile range. The median (50th percentile) is shown within the box. The two whiskers end at the 10th and 90th percentile, respectively. All outlying values above or below these percentiles are represented as dots. With box plots, the distribution of a given continuous variable can be evaluated, and their distribution can be compared between two or more groups.

Receiver operating characteristic curves (ROC curves, see *Figure 07* on *Page 44*) are used to evaluate the quality of a classification of subjects into a positive and a negative group based on a quantitative variable [263]. In the context of this study, the classification as breast cancer case (positive) or as control (negative) based on the results of the respective DNA repair assays was evaluated. Previously it has to be defined if larger or smaller values of the test variable indicate the positive classification.

To generate a ROC curve, all possible individual values of the test variable are successively selected as thresholds for the discrimination into positive cases (breast cancer cases) and negative cases (controls). For each resulting classification based on the particular threshold, the sensitivity, also called the true positive rate, is plotted against the false positive rate (1-specificity). The true positive rate is the number of cancer patients which were, based on the threshold, correctly classified positives, divided by the number of all cancer patients. The false positive rate is the number of controls which were falsely classified as positives, divided by the number of all

controls. The PASW Statistics 18 software package was used for the generation of ROC curves, which are shown as a blue line in the corresponding diagrams.

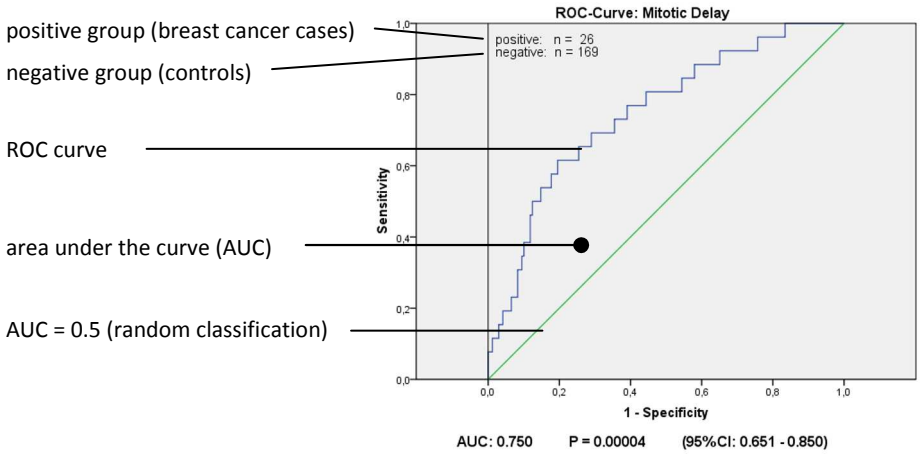

Figure 07: Exemplary description of the receiver operating characteristic (ROC) curve of the mitotic delay assay. The area under the curve (AUC) gives information about the performance of a classification of probands into the positive and a negative group based on a quantitative variable, for example the results of a DNA repair assay. P: Probability value; 95%CI: 95 % confidence interval.

The probability that a random assay result which is classified as positive (= breast cancer case) exceeds the result of a random negative case (= control) is represented by the area under the ROC curve (AUC). AUC = 0.5 would state a random classification, and is represented by the green lines in the diagrams. The more the ROC curve deviates from this line towards the left upper corner, the more accurate is the classification based on the selected variable. The calculation of the AUC includes 95 % confidence intervals and the significance (threshold $\alpha = 0.05$) of the deviation from a random classification (P-value).

2.4.4 Association Studies

The significance of a deviation from Hardy-Weinberg-Equilibrium, the allelic and genotypic odds ratios as well as the corresponding P trend value were calculated according the procedures stated in [193]. These procedures are implemented in the internet resource http://ihg2.helmholtz-muenchen.de/cgi-bin/hw/hwa1.pl (accession date: October 10, 2010). Allelic and genotypic odds ratios as well as their 95 % confidence intervals were calculated using the software package StatView 5.1 (SAS) by performing an unconditional logistic regression analysis.

Hardy-Weinberg-Equilibrium

For a given polymorphism with defined allele frequencies, the genotype distribution is fixed in an ideal population, meaning the population is large and free from the influences of migration or mating preferences, and the polymorphism is not subject to selection processes. For an autosomal locus with two alleles and their corresponding allele frequencies p and q follows that

$p^2 + 2 * p * q + q^2 = 1$ and (*Equation 05*)

$p + q = 1$, (*Equation 06*)

where p^2 is the frequency of homozygous carriers of the p allele, q^2 is the frequency of individuals homozygous for the q allele and 2*p*q is the frequency of the heterozygous individuals. Deviations of the observed genotype distributions from the expected numbers are indicators for genotyping errors, relationships between study probands or the admixture of genetically divergent subpopulations. If a deviation from the HWE is present, the results of association studies may not be valid.

With respect to the low minor allele frequencies, the frequencies of genotypes were tested for a deviation from Hardy-Weinberg-Equilibrium (HWE) with Fisher's exact test.

Odds Ratios

The odds ratio (OR) allows to compare the distribution of a risk factor between groups of cases and controls. Generally it can be deduced from the following fourfold table (*Table 07*):

Table 07: Fourfold table for the calculation of odds ratios.

	Number of Cases	**Number of Controls**
Exposed (Risk allele or genotype present)	a	b
Not exposed (Risk allele or genotype not present)	c	d

by the equation

$$\mathrm{OR} = \frac{a * d}{c * b} \quad \text{(Equation 07)}$$

The resulting OR can generally be interpreted as the factor by which the chance to develop a disease increases if a certain exposition was given, in the present context the chance to develop cancer if the variant allele or genotype is inherent (exposed).

The allelic odds ratio was determined by using the respective allele frequencies p and q in cases and controls, where p is the variant allele:

$$\text{allelic OR} = \frac{p\,(\text{cases}) * q\,(\text{controls})}{q\,(\text{cases}) * p\,(\text{controls})} \qquad (\textit{Equation 08})$$

Because the evaluated polymorphisms had, with one exception, low minor allele frequencies (0.02 – 0.15), only very few individuals homozygous for the respective minor alleles were expected. Thus, the genotypic odds ratio between wild type (11) and heterozygous individuals (12) was additionally calculated:

$$\text{genotypic OR} = \frac{n12\,(\text{cases}) * n11\,(\text{controls})}{n11\,(\text{cases}) * n12\,(\text{controls})} \qquad (\textit{Equation 09})$$

The corresponding 95 % confidence intervals were calculated for both types of odds ratios.

The significance of the resulting odds ratios was determined with Armitage's test for trend (P-trend). This chi-squared test takes into account if the relative increase in the presence of a defined allele between the three genotypes also leads to a linear shift of the proportions of cases and controls between the respective genotypes, especially if none of the strict Mendelian models is applicable. If a certain allele of a genotyped variant is associated with cancer, the proportion of cancer cases is expected to be the smallest in individuals homozygous for the wild type allele, larger in heterozygous individuals and largest in homozygous carriers of the suspected disease allele.

Estimation of the Population Attributable Risk from case-control studies

The population attributable risk (PAR) can be interpreted as the proportion of all disease cases in the study population which is related to a specific exposition, here the presence of the variant allele of a genotyped polymorphism. It can be directly estimated from the fourfold table (*Table 07* on *Page 45*) used for the estimation of odds ratios [222], with the heterozygous and homozygous carriers of the variant allele serving as the exposed group and individuals homozygous for the wild type allele as the not exposed group:

$$\text{PAR} = 1 - \frac{c * (b + d)}{d * (a + c)} \qquad (\textit{Equation 10})$$

Haplotype Estimates

If in the studied population certain alleles of two variant loci are more frequently situated on the same chromosome than is expected under the assumption of a free recombination between the two loci, a linkage disequilibrium (LD) is present between these loci. Two possibilities for the

origin of the different allele constellations on one chromosome, which are called haplotypes, exist, namely a subsequent occurrence of polymorphisms due to mutations or by recombinational events between the two loci. In the former case, a complete LD between the two variants is observed, while it is less pronounced in the latter case.

Only the observed genotype frequencies are available to analyze if two or more variants are in LD. Some observed genotype combinations in the single individuals are not conclusive, for example if both loci are heterozygous and could be constituted by more than two haplotype phases. Therefore, the most probable combination is estimated by the overall haplotype frequency in the proband cohort. In the genes in which two or more variants were genotyped in the context of the present study, the haplotype frequencies were estimated using the Haploview software version 4.2 [17].

3 Results

Mutations and variations in many DNA repair genes have been associated to the occurrence of breast cancer. The unique pattern of variation in DNA repair genes defines the individual cellular ability to react on DNA damage. Differences in this overall cellular DNA repair capacity have on the one hand been associated to the incidence of breast cancer and on the other hand to the presence of coding variants in DNA repair genes. The aim of this study was to elucidate if the cellular DNA repair capacity, coding variants in selected DNA repair genes and breast cancer can mutually be associated with each other in a single study on the same set of probands.

The applied measures of cellular DNA repair capacity included the baseline and the induced micronucleus assay (MN-b and MN-i, respectively) to visualize chromosomal fragments resulting from unrepaired DNA double strand breaks. Also the baseline and the induced sister chromatid exchange assay (SCE-b and SCE-i, respectively) were applied, which indicate recombinational repair events between sister chromatids. The mitotic delay assay (MD) quantifies the delay of the cell cycle in G_2 phase upon DNA damage induction. These assays were based on short-time lymphocyte cultures from peripheral blood samples.

The assays originate from mutagenicity tests used in human biomonitoring studies. Thus it was evaluated as a prerequisite, to which degree these assays indeed reflect variations in the individual genetic background, in contrast to individual exposures to environmental factors. This was performed by estimating their heritability in the context of a twin study. Correlation analyses were performed to study if the DNA repair assay results reflect independent end points of the cellular DNA repair network. From a small subset of probands lymphoblastoid cell lines (LCL) were generated additionally to evaluate their use as a surrogate for peripheral blood lymphocytes (PBL) as the basis of the applied DNA repair assays. The correlation between the respective assay results from LCL and PBL was studied on an individual basis.

The intrinsic test variability of the MN assays emerged as a factor that limits the reproducibility of individual results. To identify the origin of this variability, factors which potentially contribute to the variability of the MN assay results from one individual blood sample were studied.

3.1 Differences of Cellular DNA Repair Capacity Between Breast Cancer Cases and Controls

All applied tests were evaluated with regard to an association between sporadic breast cancer and altered DNA repair assay results. Test results of sporadic breast cancer cases and controls were available from the "Ulm" cohort and additional controls from the "Repair Control" and "Twin" cohorts. Only one randomly chosen individual of each twin pair was included as control to avoid relationship. Additional data of cases and controls for the induced micronucleus assay were available from the "Freiburg" cohort, but not for the other tests.

In the baseline and the induced MN frequency, differences between the "Freiburg", "Ulm" and the other two cohorts ("Repair Controls", "Twin") were present in controls. Therefore, the baseline as well as the induced MN frequency was normalized on the respective mean of the controls in each cohort. The differences were in part attributable to differences between the versions of the automated system that was used for the scoring of MN (see *2.3.2*). Also the origin of the blood samples contributed to the observed differences, possibly via the conditions in sample transportation and sample handling until culture setup (see *2.1.1*). For example, blood samples from the "Freiburg" cohort arrived exclusively by mail.

In the group of controls as well as in the breast cancer cases group, the results of all tests were compatible with a normal distribution and could be analyzed using linear regression, t-test and ANOVA.
In controls, linear regression analyses revealed few statistically significant but very weak influences on the results of the five assays. Culture setup delay was correlated to the baseline MN frequency with nominal significance ($P = 0.02$, $r^2 = 0.05$), and age was highly significant correlated to the MD index ($P = 0.0006$, $r^2 = 0.07$). However, the very small correlation coefficients ($r^2 < 0.1$) do not support a major influence of these factors on the assays. None of the tests was significantly influenced by gender. An additional ANOVA also revealed no significant influences of the single or combined factors on any of the assay results. Thus, no adjustments were considered for these factors in further analyses.

The focus of the analysis was if controls and sporadic breast cancer cases exhibit significant differences in the results of the respective assays (Student's T-test). *Table 08* on *Page 50* summarizes these data for all applied tests.

Table 08: Differences between sporadic breast cancer cases and controls in the five DNA repair assays: the baseline and the induced micronucleus assay (MN-b and MN-i, respectively), the baseline and the induced sister chromatid exchange assay (SCE-b and SCE-i, respectively) and the mitotic delay assay (MD). Given are the numbers of probands, the means and standard deviations (SD) in cases and controls as well as the significance of the difference (probability value (P)). DNA: deoxyribonucleic acid.

		MN-b (normalized)	MN-i (normalized)	SCE-s	SCE-i	MD
Numbers	**Controls**	162	315	132	128	169
	Cases	35	161	14	14	26
Means ± SD	**Controls**	1.00 ± 0.43	0.99 ± 0.23	7.38 ± 1.71	14.41 ± 3.16	3.88 ± 0.92
	Cases	0.99 ± 0.43	1.08 ± 0.30	7.74 ± 1.79	14.24 ± 3.16	3.05 ± 0.81
Difference of means		P = 0.93	**P = 0.0003**	P = 0.46	P = 0.85	**P < 0.0001**

Figures 08 – 12 on *Pages 50 – 53* provide additional illustrations for the differences of the respective assay results between the groups of controls and sporadic breast cancer patients. The box-plots (*Figures 08 – 12 A*) visualize the distributions of values within the groups of controls (blue) and sporadic breast cancer patients (red).

Receiver operating characteristic curves (ROC-Curves, *Figures 08 – 12 B*) are used to evaluate the performance of a classification of breast cancer cases and controls (see *2.4.3*) based on the respective assays. The area under the curve (AUC) represents the probability that a random assay result which is classified as positive (= case) exceeds the result of a random negative case (= control).

No difference between sporadic breast cancer cases (n = 35) and controls (n = 162) was seen with the baseline MN assay (*Table 08* and *Figure 08*).

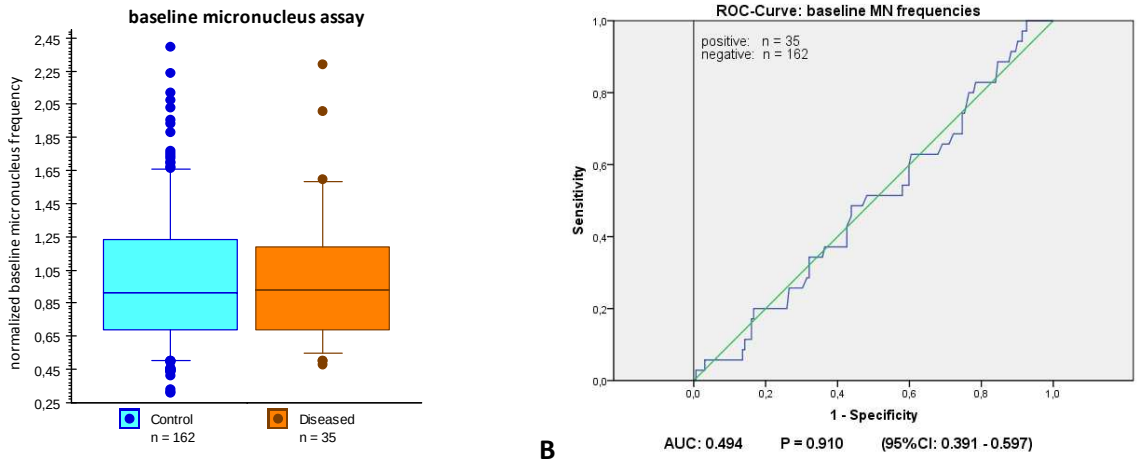

Figure 08: Differences of breast cancer patients and controls in the baseline micronucleus (MN) assay: Box-plots (A) and receiver operating characteristic (ROC) curve (B). n: number of probands; AUC: area under the curve; P: Probability value; 95%CI: 95 % confidence interval.

The means showed nearly no differences (P = 0.93). The distributions of the results within the two groups were also highly similar (P = 0.99, Kolmogorov-Smirnov test), as visualized by the box plots with nearly identical medians, 75th and 25th percentiles. Thus, the baseline micronucleus frequency did not exhibit general differences between sporadic breast cancer cases and controls. Accordingly, the ROC curve did not significantly deviate from the random classification (AUC = 0.49, P = 0.91).

In contrast, the radiation induced MN frequency differed significantly between sporadic breast cancer cases (n = 161) and controls (n = 315; see *Table 08* and *Figure 09*).

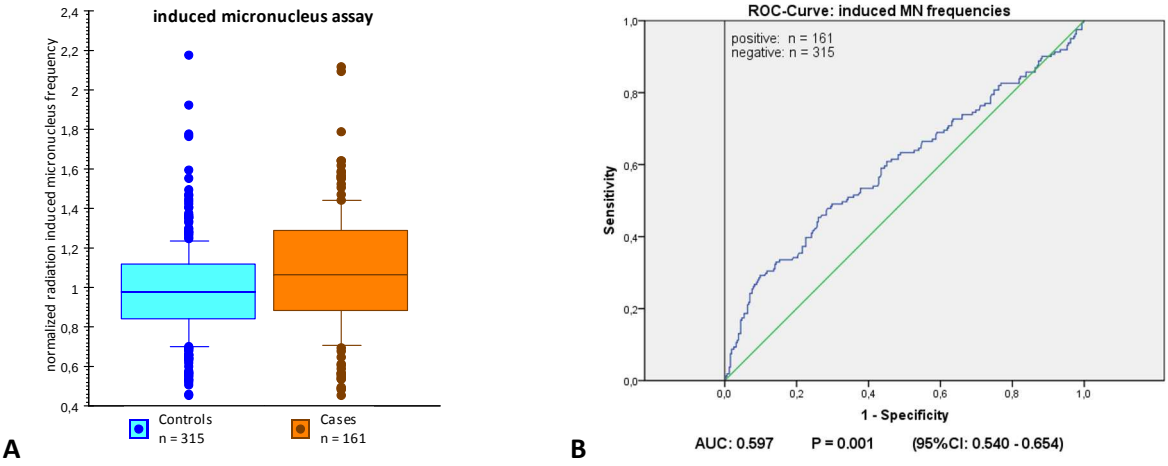

Figure 09: **Differences of breast cancer patients and controls in the induced micronucleus (MN) assay: Box-plots (A) and receiver operating characteristic (ROC) curve (B).** n: number of probands; AUC: area under the curve; P: Probability value; 95%CI: 95 % confidence interval.

The means of the two groups differed highly significantly (P = 0.0003), and the distribution of the results of breast cancer cases was also highly significantly shifted towards a higher induced MN frequency (P = 0.0007). This is underlined by the box plot which shows the differences of the medians as well as of the 75th and 90th percentiles. The 10th and 25th percentiles of the induced MN frequency were similar, indicating that those sporadic breast cancer cases with increased radiation induced micronucleus frequencies might comprise a more radiosensitive subgroup of all cases. The ROC curve deviated significantly from a random classification (P = 0.001) and indicated a probability of 60 % that a breast cancer case exceeds the radiation induced MN frequency of a control (AUC = 0.60).

Indicated from *Table 08* and *Figure 10*, the baseline sister chromatid exchange assay allowed no discrimination between sporadic breast cancer patients (n = 14) and controls (n = 132).

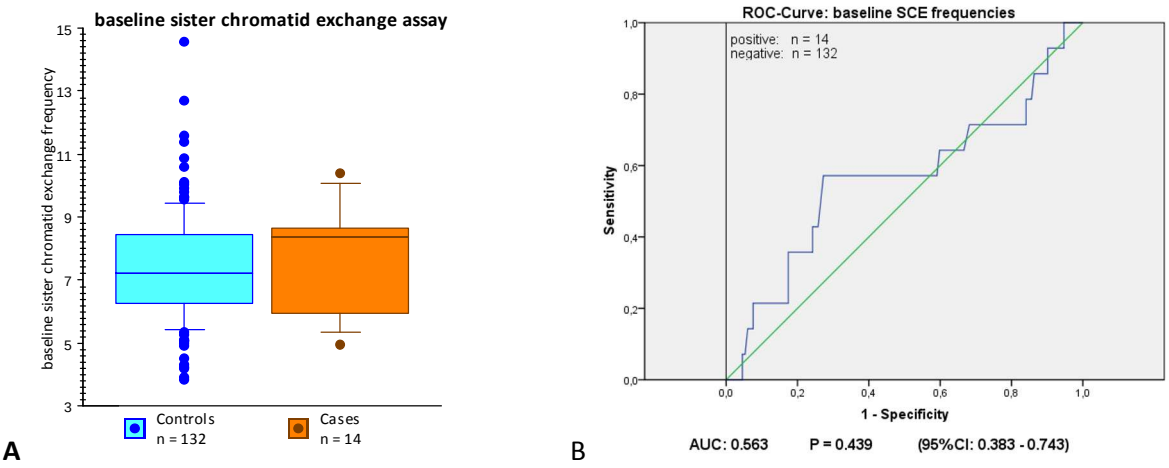

Figure 10: **Differences of breast cancer patients and controls in the baseline sister chromatid exchange (SCE) assay: Box-plots (A) and receiver operating characteristic (ROC) curve (B).** n: number of probands; AUC: area under the curve; P: Probability value; 95%CI: 95 % confidence interval.

The means did not significantly differ (P = 0.46), the same holds true for the distributions (P = 0.21). However the box plots reveal that, while the 75th and 25th percentiles were similar between breast cancer cases and controls, the median of cases is clearly higher. The ROC-Curve did also show a random classification of breast cancer cases on controls based on the baseline SCE frequency (AUC 0.56, P = 0.43). However, the number of breast cancer cases for which assay results were available was very small, which might impede the yield of clearly significant differences.

Differences between sporadic breast cancer patients (n = 14) and controls (n = 128) were not present in the induced sister chromatid exchange assay (see *Table 08* and *Figure 11*).

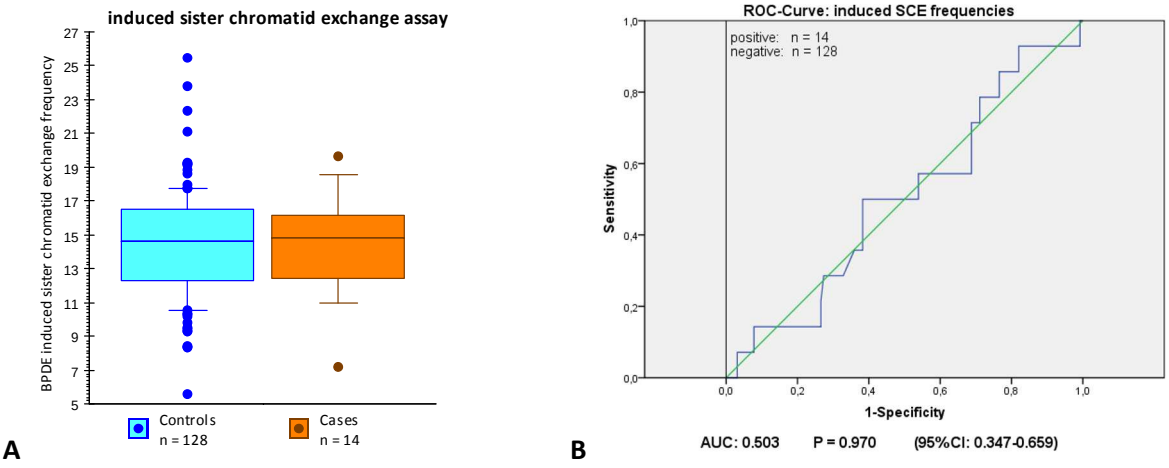

Figure 11: **Differences of breast cancer patients and controls in the induced sister chromatid exchange (SCE) assay: Box-plots (A) and receiver operating characteristic (ROC) curve (B).** n: number of probands; AUC: area under the curve; P: Probability value; 95%CI: 95 % confidence interval. BPDE: Benzo[a]pyrenediepoxide.

Neither means (P = 0.85) nor distributions (P = 0.99) differed significantly, as underlined by the box plots which show highly similar medians and the comparable 75th and 25th quartiles between the two groups. Accordingly, the ROC curve indicated that a classification of breast cancer cases and controls based on BPDE induced sister chromatid exchange frequencies is random (AUC = 0.50, P = 0.97). As in the baseline SCE assay, the number of breast cancer cases with available results is also very low.

As demonstrated in *Table 08* and *Figure 12*, the MD assay yielded a significant discrimination between breast cancer cases (n = 26) and controls (n = 169).

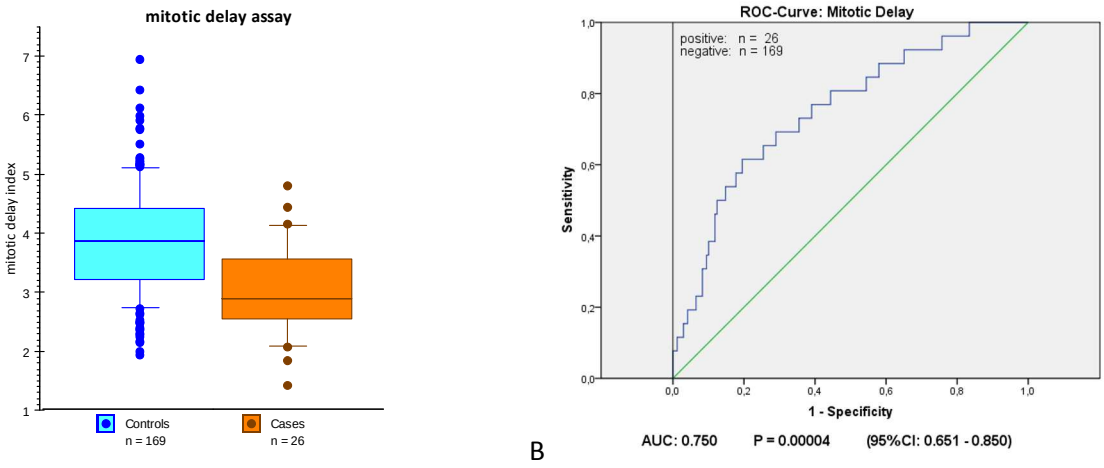

Figure 12: Differences of breast cancer patients and controls in the mitotic delay (MD) assay: Box-plots (A) and receiver operating characteristic (ROC) curve (B). For the generation of the ROC curve, lower MD index values were classified as positive (see *2.4.3*). n: number of probands; AUC: area under the curve; P: Probability value; 95%CI: 95 % confidence interval.

Breast cancer cases showed a considerably smaller mean MD index than controls (P < 0.0001) resulting in a significantly shifted distribution (P = 0.0007). This is illustrated by the box plots, which display the differences of the medians as well as of the 75th and 25th percentiles, with the median of the cases clearly below the 25th percentile of the controls. In addition the 10th and 90th percentiles also differed between the two groups, and the range of cases is narrower than those of controls. The maximum MD index of cases is lower than the 90th percentile of controls. Taken together, it can be stated that sporadic breast cancer cases exhibited in general a decreased mitotic delay index compared to controls. The ROC curve also indicated a highly significant deviation from a random classification (P = 0.00004), with a 75 % probability that the MD index of a breast cancer case is lower than the MD of a control (AUC = 0.75).

The MD index is constituted of the baseline and the radiation induced G_2/S ratios. The contribution of these two factors to the observed discrimination between sporadic breast cancer cases and controls was considered separately. *Table 09* summarizes the data.

Table 09: Differences of baseline and induced ratios of cells in the second gap and the deoxyribonucleic acid (DNA) synthesis phases of the cell cycle (G_2/S-b and G2/S-i, respectively) between sporadic breast cancer cases and controls. Given are: the numbers of cases and controls, the means and standard deviations (SD); the significance of the difference (probability value (P)).

		G_2/S-b	G_2/S-i
Numbers	Controls	169	169
	Cases	26	26
Means ± SD	Controls	0.34 ± 0.05	1.30 ± 0.29
	Cases	0.41 ± 0.07	1.22 ± 0.36
Difference of means		**P < 0.0001**	P = 0.25

The shift of sporadic cancer cases (n = 26) towards a lower MD index compared to controls (n = 169) was nearly exclusively attributable to the clearly increased baseline G_2/S ratio of cancer cases.

Figure 13 provides additional visualizations for the baseline G_2/S ratio.

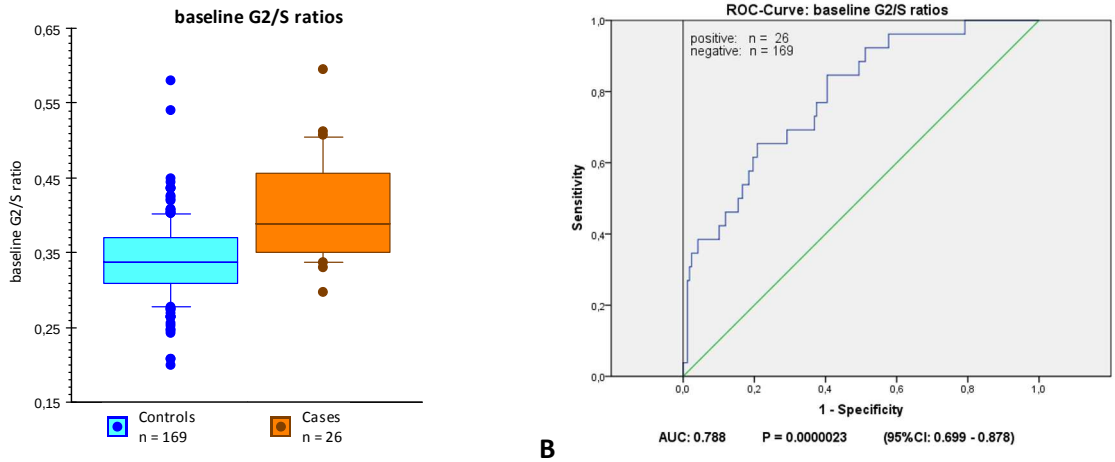

Figure 13: Differences of the ratio of untreated cells in the second gap (G_2) and the deoxyribonucleic acid (DNA) synthesis (S) phases of the cell cycle in breast cancer cases and controls: Box-plots (A) and receiver operating characteristic (ROC) curve (B). n: number of probands; AUC: area under the curve; P: Probability value; 95%CI: 95 % confidence interval.

The means (P < 0.0001) and the distributions (P = 0.0003) of the baseline G_2/S ratio differed highly significantly between the two groups. The box plots revealed the highly different medians, the 75th and 25th as well as the 90th and 10th percentiles. The 75th percentile of the controls is lower than the median of the cases. Whereas the maximum baseline G_2/S ratio value of cases is similar

to that of controls, the minimum baseline G_2/S ratio in cases is between the 25th and 10th percentile of controls. Thus, sporadic breast cancer cases exhibit a general and highly significant accumulation of cells in G_2 phase compared to S phase when untreated cultures are studied. The ROC curve of the baseline G_2/S ratios deviated highly significantly from a random classification (p = 0.000002), with a 79 % probability that the G_2/S ratio of a sporadic breast cancer case exceed that of a control (AUC = 0.79)

Figure 14 provides the visualizations for the radiation induced G_2/S ratio.

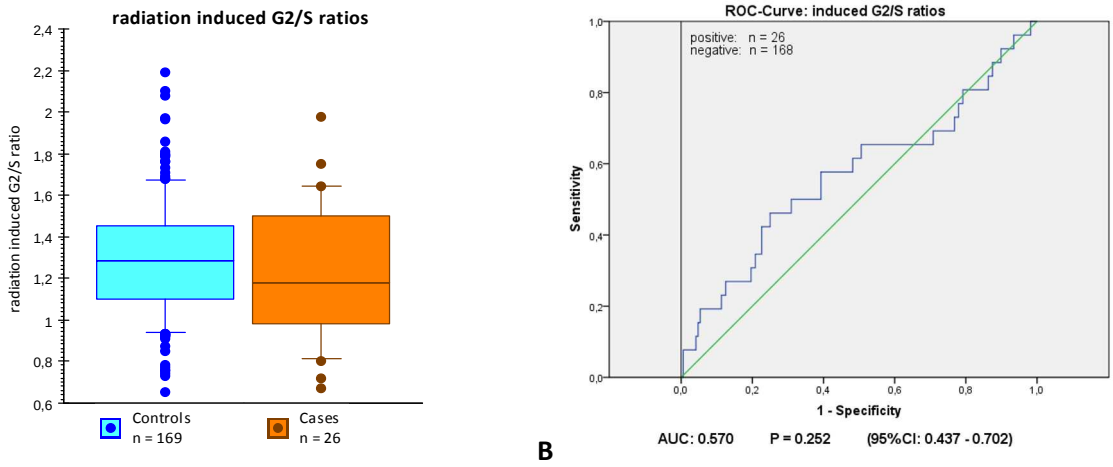

Figure 14: Differences of the ratio of irradiated cells in the second gap (G_2) and the deoxyribonucleic acid (DNA) synthesis (S) phases of the cell cycle in breast cancer cases and controls: Box-plots (A) and receiver operating characteristic (ROC) curve (B). In the ROC curve, lower values were classified as positive (see *2.4.3*). n: number of probands; AUC: area under the curve; P: Probability value; 95%CI: 95 % confidence interval.

In contrast, neither the means nor the distributions of the G_2/S ratio in irradiated cells differ significantly between cancer cases and controls (p = 0.25 and p = 0.26), and while the box plot depicts a lower median in cases, the 75th and 25th percentiles of cases exceed those of controls. In summary, the induced G_2/S ratios of cases exhibit a small and not significant decrease compared to controls. In the ROC curve the deviation from a random classification was not significant for the induced G_2/S ratios (AUC = 0.57, p = 0.25).

In summary, the strongest difference between sporadic breast cancer cases and controls was seen with the mitotic delay assay, followed by the induced MN frequency, while there were no significant differences with the baseline MN frequency as well as with the baseline and induced SCE frequency. The results for the baseline G_2/S ratios provide an even better discrimination and classification of breast cancer patients and controls than the mitotic delay, while there are only marginal differences in radiation induced G_2/S ratios.

3.2 The Impact of Low-Frequency Coding Variants in DNA Repair Genes on Cellular DNA Repair Capacity

3.2.1 Gene and Variant Selection

For the present study, candidate DNA repair genes were selected and rare coding variants in these genes were genotyped. Associations between the genotypes and breast cancer risk as well as the influence of the genotypes on the cellular DNA repair capacity were studied. The selection of DNA repair genes was performed with the PubMed literature database (http://www.ncbi.nlm.nih.gov/pubmed/) of the National Center for Biotechnology Information (NCBI) and the NCBI BioSystems database (http://www.ncbi.nlm.nih.gov/biosystems, accession date January 29th, 2010) [72], provided by the National Institutes of Health of the United States of America. In the BioSystems database, the functional system "DNA Repair" in *Homo sapiens* (BioSystems ID 105837) was selected. The genes included in this system were evaluated by an extensive search of literature within the PubMed database, which was also applied to search and evaluate further genes with known relations to those provided by the BioSystems database.
The focus of candidate gene selection was on less well-studied DNA repair genes, meaning genes which were only recently involved or not yet thoroughly characterized in the context of DNA repair and cancer. Additionally, the set of candidate genes was chosen in such a way that different pathways of the cellular DNA repair network were covered. The genes were evaluated for possible contributions to the DNA repair end points measured by the applied DNA repair assays and for possible associations to cancer risk.

The NCBI Gene and NCBI Single Nucleotide Polymorphisms (dbSNP) databases (http://www.ncbi.nlm.nih.gov/gene/ and http://www.ncbi.nlm.nih.gov/snp, dbSNP build 130, NCBI build 36.3, accession date January 29th, 2010) [142,201] were used to further evaluate the selected candidate genes for the presence of variants with a possible impact on the function of the proteins. Only missense variants with reported minor allele frequencies (MAF) between 0.02 and 0.16 in the European population were chosen for genotyping. Population data was available from the International HapMap project (http://hapmap.ncbi.nlm.nih.gov/, HapMap Data Phase III, Genome Browser Release #2 / Phase 1 & 2, February 2009, on NCBI build 36, dbSNP build 126, accessed January 29th, 2010) for the CEU (Utah Residents with Northern and Western European Ancestry) population [97]. If a gene contained more than one applicable infrequent missense variant, each one was genotyped. Additional genes and variants were included due to supporting

previous reports concerning a possible association to DNA repair (*ESCO*, *POLI*) or cancer (*BARD1*, *BRIP1*). In total, 23 variants in 15 genes were genotyped. These comprised 22 non-synonymous single nucleotide polymorphisms (SNPs) and one 21-base pair in-frame deletion, resulting in the loss of 7 amino acids. The selected genes as well as their genomic location and pathways in which they are involved are shown in *Table 10.*

Table 10: Information about the selected deoxyribonucleic acid (DNA) repair genes. Given are: the official Gene Symbols and Gene Names provided by the Human Genome Organization Gene Nomenclature Consortium (see also *attachment A*); the location in the genome (consisting of the chromosome number, the location on the long (p) or short arm (q) of the chromosome, and the chromosome band (Giemsa staining); the identification number in the Gene database of the National Center of Biotechnology Information (NCBI); the pathways in which the genes are involved.

Official Gene Symbol [a]	Official Gene Name [a]	Genomic Location	NCBI Gene Identifier	Functional Involvement
NEK1	*NIMA (never in mitosis gene a)-related kinase 1*	4q33	4750	G_2/M [c] checkpoint
NINL	*ninein-like*	20p11.22-p11.1	22981	G_2/M checkpoint
TIMELESS	*timeless homolog (Drosophila*)	12q12-q13	8914	G_2/M checkpoint
FBXO5	*F-box protein 5*	6q25.2	26271	G_2/M checkpoint
REV1	*REV1 homolog (S. cerevisiae)*	2q11.1-q11.2	51455	Polymerase
SLX4	*SLX4 structure-specific endonuclease subunit homolog (S. cerevisiae)*	16p13.3	84464	Homologous recombination
C17ORF70	*chromosome 17 open reading frame 70*	17q25.3	80233	Fanconi anemia
C19ORF40	*chromosome 19 open reading frame 40*	19q13.11	91442	Fanconi anemia
TERF2IP	*telomeric repeat binding factor 2, interacting protein*	16q23.1	54386	Telomere stability
WRAP53	*WD repeat containing, antisense to TP53*	17p13.1	55135	Telomere stability
HLTF	*helicase-like transcription factor*	3q25.1-q26.1	6596	Transcription factor
BARD1 [b]	*BRCA1 associated RING domain 1*	2q34-q35	580	Homologous recombination
BRIP1 [b]	*BRCA1 interacting protein C-terminal helicase 1*	17q22.2	83990	Fanconi anemia
ESCO1 [b]	*establishment of cohesion 1 homolog 1 (S. cerevisiae)*	18q11.2	114799	Sister chromatid cohesion
POLI [b]	*polymerase (DNA directed) iota*	18q21.1	11201	Polymerase

[a] Official gene symbols and names provided by the Human Genome Organization

[b] These genes were selected due to supporting previous reports

[c] G2/M: second gap and mitosis phases of the cell cycle

The genotyped variants are described in *Table 11* on *Page 58*. For better readability and comprehensibility of this part of the manuscript, each genotyped variant was given a serial number. The 21-bp deletion is not a SNP, but was nevertheless given the serial number SNP 19 for simplicity reasons.

Table 11: Information about the variants in deoxyribonucleic acid (DNA) repair genes which were chosen for genotyping, including the National Center for Biotechnology Information (NCBI) single nucleotide polymorphism database (dbSNP) catalogue numbers; and the single nucleotide polymorphism (SNP) names provided by the Human Genome Variation Society (HGVS) The resulting amino acid changes in the one-letter amino acid code with the wild-type amino acid followed by the amino acid residue in the gene product followed by the variant amino acid.

Official Gene Symbol [a]	SNP Nr.	NCBI dbSNP catalogue number	HGVS SNP name (including position and nucleotide change)[b]	Amino Acid [c] Change
NEK1	**01**	rs34099167	NM_012224.1:c.2171A>G	E724G
	02	rs34540355	NM_012224.1:c.1388C>T	A463V
NINL	**03**	rs17857107	NM_025176.4:c.4097G>A	R1366H
	04	rs13044759	NM_025176.4:c.826C>T	R275W
TIMELESS	**05**	rs61376834	NM_003920.2:c.3050T>C	I1017T
	06	rs72478986	NM_003920.2:c.385G>T	A128S
FBXO5	**07**	rs7763565	NM_012177.2:c.490C>T	L163F
REV1	**08**	rs3087399	NM_001037872.1:c.1118A>G	N373S
SLX4	**09**	rs3810813	NM_032444.2:c.3812C>T	S1271F
	10	rs3827530	NM_032444.2:c.3662C>T	A1221V
	11	rs714181	NM_032444.2:c.3365C>T	P1122L
C17ORF70	**12**	rs14422	NM_025161.4:c.1996G>A	A665T
C19ORF40	**13**	rs2304103	NM_152266.3:c.473C>T	S158L
	14	rs3816032	NM_152266.3:c.575T>C	I192T
TERF2IP	**15**	rs4888444	NM_018975.2:c.970A>G	L323E
WRAP53	**16**	rs35123152	NM_018081.1:c.1480G>C	E483Q
HLTF	**17**	rs2229361	NM_139048.2:c.2456G>A	R819H
	18	rs2305868	NM_139048.2:c.932A>G	N311S
BARD 1	**19**[d]	rs28997575	NM_000465.2:c.1075_1095del21	365delLPECSSP
	20	rs28997576	NM_000465.2:c.1670G>T	C557S
BRIP1	**21**	rs4986764	NM_032043.1:c.2755C>T	P918S
ESCO1	**22**	rs35087820	NM_052911.2:c.572A>G	N191S
POLI	**23**	rs3218786	NM_007195.2:c.1595T>C	F532S

[a] Official gene symbols provided by the Human Genome Organization.

[b] SNP name provided by the Human Genome Variation Society (www.hgvs.org). The code includes the accession number of the reference sequence of the messenger ribonucleic acid (mRNA) in the NCBI database, ending at the colon. The position of the SNP within the coding region (indicated by the "c.") is given by subsequent the number, where the first nucleotide of the start codon ATG is defined as position 1. The nucleotide exchange is described after the number (wild type > variant). del: deletion

[c] One-letter amino acid code: alanine (A), cysteine (C), glutamic acid (E), phenylalanine (F), glycine (G), histidine (H), isoleucine (I), leucine (L), asparagine (N), proline (P), glutamine (Q), arginine (R), serine (S), threonine (T), valine (V), tryptophan (W).

[d] This 21 base-pair in-frame deletion is not a SNP but was awarded the serial number SNP 19 for simplicity reasons. The deletion spans the positions 1075 to 1095 of the coding sequence, resulting in the loss of 7 amino acids in the protein.

3.2.2 Genotyping Results

Twenty-three variant loci were genotyped in all probands from which DNA samples had been obtained (n = 2014). Raw data underwent quality control procedures which are recommended for high-throughput genotyping approaches. These include negative controls in order to visualize cross sample contamination and duplicate samples for the validation of the genotyping

procedure. Failing samples were repeated once. Twenty-one samples yielded no results in any of the analyses, and 17 further samples were excluded due to the failure of three or more independent genotyping analyses, indicating impaired quality of the respective DNA sample. The genotyping efficiency in the remaining samples (n = 1976) was between 98.7 % and 100 % for all variants.

The minor allele of the variant rs35123152 in the *WRAP53* gene (SNP 16) was neither present in cases nor in controls, although a minor allele frequency in the European population of 0.09 was expected. Thus, the data for this variant was excluded from any further analyses.

The MAF of all variants in the total of the included 1189 healthy controls from the cohorts without familial breast cancer background are shown in *Table 12* in comparison to the MAF expected from population data (HapMap CEU, see *3.2.1*).

Table 12: Expected and observed minor allele frequencies (MAF) of the genotyped single nucleotide polymorphisms (SNP). Given are: the SNP number; the MAF in the Population of the Utah Residents with Northern and Western European Ancestry (CEU) of the International HapMap database; the MAF observed in the genotyped controls (n: number of probands).

SNP Number [a]	MAF expected in the HapMap CEU population	MAF observed in controls (n = 1189)
01	0.12	0.10
02	0.07	0.06
03	0.09	0.11
04	0.04	0.04
05	0.08	0.05
06	0.05	0.01
07	0.07	0.06
08	0.10	0.14
09	0.05	0.05
10	0.13	0.06
11	0.16	0.08
12	0.10	0.11
13	0.05	0.04
14	0.12	0.12
15	0.05	0.04
16	0.09	0.00[b]
17	0.02	0.04
18	0.02	0.02
19	0.03	0.02
20	0.03	0.02
21	0.38	0.42
22	0.03	0.03
23	0.01	0.03

[a)] Consult *Table 11* on *Page 58* for background information on the genotyped variants.
[b)] The minor allele was not present in any of the samples.

3.2.3 Association of Breast Cancer With Variants in the Selected DNA Repair Genes

3.2.3.1 Genotype Associations

Prior to each association test, the genotype frequencies of each SNP in cases and controls were evaluated for deviation from the Hardy-Weinberg-Equilibrium (HWE) with Fisher's exact test. No deviations were present.

Association analyses were performed for all of the 22 successfully genotyped variants. The odds ratios (OR) of allele frequency differences between cases and controls (allelic OR) were determined, as well as the OR of the frequency of heterozygous variant carriers to non-carriers (heterozygous OR) between cases and controls. The 95 % confidence intervals (95 % CI) were also calculated. The significance of the associations of the variants to breast cancer was determined with Armitage's test for trend (P-trend), which takes all three possible genotypes into account. The nominal significance threshold was set to $\alpha = 0.05$. Significance values were subsequently corrected for multiple testing of the 22 variants by applying the Bonferroni correction method (threshold: 0.05/22 = 0.00227) and by permutation testing (10^4 permutations) of the allelic association using the Haploview 4.2 software. In addition, the population attributable risk was determined after Taylor *et al.* [222].

Association analyses using all cohorts

The cases group comprised the sporadic breast cancer cases from the "Ulm" and "Freiburg" cohorts. Additionally, in the "Familial" cohort, one affected index person was defined in each family and the affected index cases from BRCA-negative high-risk families were included.

The control group comprised all samples from the "Population Controls" and "Repair Controls" cohorts, the control samples from the "Freiburg" and "Ulm" cohorts, and additionally one randomly chosen individual from each twin pair from the "Twin" cohort.

Age information was available for all cases from the "Familial" cohort and for a minority of cases from the "Freiburg" and "Ulm" cohorts (see *Table 01* on *Page 15*).

Table 13 on *Page 61* summarizes minor allele frequencies, tests for deviations from the HWE, allelic and heterozygous OR with 95 % confidence intervals and the significance of the associations evaluated by the P-trend test. The raw genotype counts are shown in *Attachment D*.

Table 13: Association data of the analyzed variants in all breast and ovarian cancer cases and controls: Single nucleotide polymorphism number (SNP Nr.); minor allele frequencies (MAF); significance of a deviation from the Hardy-Weinberg-Equilibrium (HWE) indicated by the probability value (P). Shown are the allelic and genotypic odds ratios (OR) of the allele frequencies in cases and controls, as well as their 95 % confidence intervals (95 % CI). The significance of the associations is given by the probability value of Armitage's test for trend (P-trend). n: number of probands.

SNP Nr.[a]	**Controls** n = 1189		**Cases** n = 435		**Allelic** (WT vs. VAR)[b]		**Genotypic** (NN vs. NV)[c]		**P-trend value**
	MAF	HWE	MAF	HWE	OR	95 % CI	OR	95 % CI	
1	0.10	P=0.12	0.09	P=1.00	1.08	[0.84-1.39]	1.15	[0.87-1.52]	P=0.54
2	0.06	P=0.59	0.05	P=1.00	0.92	[0.65-1.29]	0.94	[0.66-1.35]	P=0.62
3	0.11	P=0.54	0.10	P=0.60	0.96	[0.74-1.24]	0.98	[0.74-1.30]	P=0.76
4	0.04	P=0.26	0.05	P=1.00	1.34	[0.94-1.93]	1.30	[0.89-1.89]	P=0.10
5	0.05	P=0.07	0.05	P=0.61	0.96	[0.67-1.37]	0.96	[0.66-1.38]	P=0.82
6	0.01	P=1.00	0.01	P=1.00	1.05	[0.51-2.19]	1.05	[0.50-2.20]	P=0.89
7	0.06	P=0.80	0.05	P=1.62	0.82	[0.59-1.16]	0.88	[0.61-1.24]	P=0.26
8	0.14	P=0.07	0.12	P=0.09	0.86	[0.68-1.09]	0.90	[0.69-1.16]	P=0.20
9	**0.05**	**P=0.75**	**0.07**	**P=0.72**	**1.55**	**[1.12-2.13]**	**1.63**	**[1.17-2.28]**	**P=0.006**
10	0.06	P=0.80	0.05	P=1.00	0.85	[0.60-1.19]	0.86	[0.60-1.24]	P=0.84
11	0.08	P=0.70	0.07	P=0.27	0.88	[0.66-1.19]	0.84	[0.61-1.16]	P=0.42
12	0.11	P=0.46	0.10	P=0.41	0.88	[0.68-1.14]	0.93	[0.64-1.15]	P=0.36
13	0.04	P=0.73	0.05	P=0.07	1.09	[0.75-1.57]	0.95	[0.64-1.41]	P=0.66
14	0.12	P=0.27	0.12	P=0.64	0.96	[0.76-1.22]	0.97	[0.73-1.27]	P=0.75
15	0.04	P=0.73	0.05	P=1.00	1.03	[0.71-1.50]	1.06	[0.72-1.55]	P=0.86
16	0.00	-	0.00	-	-	-	-	-	-
17	0.04	P=1.00	0.03	P=0.30	0.71	[0.45-1.11]	0.66	[0.41-1.05]	P=0.13
18	0.02	P=1.00	0.02	P=1.00	0.77	[0.43-1.33]	0.76	[0.43-1.33]	P=0.34
19	0.02	P=1.00	0.02	P=1.00	1.01	[0.57-1.81]	1.01	[0.57-1.82]	P=0.96
20	0.02	P=0.32	0.02	P=0.13	1.02	[0.57-1.82]	0.93	[0.50-1.72]	P=0.95
21	0.42	P=0.86	0.42	P=0.69	1.00	[0.85-1.17]	0.97	[0.76-1.25]	P=0.99
22	0.03	P=1.00	0.04	P=1.00	1.22	[0.79-1.88]	1.27	[0.81-1.97]	P=0.35
23	0.03	P=0.40	0.03	P=0.07	0.93	[0.60-1.44]	0.79	[0.50-1.27]	P=0.75

[a] Consult *Table 11* on *Page 58* for background information on the genotyped variants.

[b] Wild-type allele (WT) versus variant (VAR) allele.

[c] Wild-type genotype (NN) versus heterozygous genotype (NV).

The minor allele of SNP 09, which is the coding variant rs3810813 located in the *SLX4* gene, was significantly overrepresented (P-trend = 0.006) in breast cancer cases compared to controls, with an allelic OR of 1.55 (95 % CI: 1.12 – 2.13) and a genotypic OR of 1.63 (95 % CI: 1.17 – 2.28). The significance of the association did not withstand Bonferroni correction for multiple testing of the 22 variants (threshold: P = 0.05/22 = 0.0022) and permutation testing (allelic association after 10^4 replicates: P = 0.06).

The variant results in a change from Serine to Phenylalanine on amino acid position 1271 in the Slx4 Protein. Slx4 is involved in the homologous recombination pathway of DNA double strand

break repair. It binds several DNA damage recognition and repair proteins and is part of a complex which facilitates the resolving of Holliday Junction.

No further genotyped variants were associated to breast cancer in the analysis of all cohorts.

Stratification on young breast cancer cases

To find variants specifically associated with early-onset breast and ovarian cancer, additional analyses were performed within each group by stratification of cases by age ≤ 60 years and ≤ 40 years.

When considering only breast and ovarian cancer cases aged 60 years or younger, the minor allele of SNP 09 was even stronger overrepresented in cases (MAF = 0.10, n = 167) compared to controls (n = 1189), than if all cases are included. The allelic OR was 2.25 (95 % CI: 1.51 – 3.36) and the genotypic OR was 2.38 (95 % CI: 1.55 – 3.68), with very high significance (P-trend = 0.00005). When only cases aged 40 years and younger were considered, the minor allele frequency of SNP 09 in those cases (MAF = 0.17, n = 23) was further increased compared to controls (n = 1189) and also compared to all cases and to cases ≤ 60 years. The association was also highly significant (P-trend = 0.00011), with an allelic OR of 4.18 (95 % CI: 1.91 – 9.17) and a genotypic OR of 5.32 (95 % CI: 2.21 – 12.8). Both P-trend values withstood the correction for multiple testing of the 22 variants in the three analyses (all cancer cases, cases ≤ 60 years and cases ≤ 40 years) with the Bonferroni method (threshold: P = 0.05 / (22*3) = 0.00075) and permutation testing (P = 0.0021 for cases ≤ 60 and P = 0.011 for cases ≤ 40).

SNP 10 represents a further variant in the *SLX4* gene, rs3827530, and the minor allele showed a borderline significant (P-trend = 0.04) underrepresentation in cases aged 60 years and younger (MAF = 0.03) compared to controls (MAF = 0.06), with an allelic OR of 0.51 (95 % CI: 0.51 – 0.96). However, the genotypic OR of 0.54 (95 % CI: 0.29 – 1.02) had a confidence interval that includes equality (OR = 1.00).

The minor allele of SNP 21, rs4986764 in the *BRIP1* gene, was underrepresented in cases aged ≤ 40 years (MAF = 0.26) compared to controls (MAF = 0.42), however only with borderline significance (P-trend = 0.03) that does not withstand Bonferroni correction or permutation testing (p = 0.54). The allelic OR was 0.49 (95 % CI: 0.25 – 0.96) and the genotypic OR was 0.59 (95 % CI: 0.25 – 1.37) and also includes equality (OR = 1.00) in its 95 % confidence interval.

In the remaining variants, no age effects were observed.

Summary of the association study

Table 14 gives an overview of the results of the genotype association study, with each cell displaying the calculated allelic odds ratios (numbers) and the significance of the association (field colors).

Table 14: Visualization of the association study results and age stratifications. Blue fields indicate that no significant association was present, green fields indicate a nominally significant association (probability value (P) < 0.05) and orange fields a global significant association that withstands correction for multiple testing with the Bonferroni method in all age groups (threshold: P < 0.00075) and permutation testing. Numbers indicate the allelic odds ratios of the respective minor allele frequencies in cases and controls. SNP: single nucleotide polymorphism.

SNP Number [a]	All	Age ≤ 60 years	Age ≤ 40 years
01	1.08	1.28	1.82
02	0.92	1.19	1.96
03	0.96	1.05	1.51
04	1.34	1.21	1.09
05	0.96	0.99	0.84
06	1.05	0.82	0.96
07	0.82	0.84	0.67
08	0.86	0.88	0.44
09	**1.55**	**2.25**	**4.18**
10	0.85	**0.51**	0.34
11	0.88	0.63	0.51
12	0.88	0.89	1.21
13	1.09	1.00	0.48
14	0.96	0.72	0.51
15	1.03	0.79	0.23
17	0.71	0.97	1.67
18	0.77	1.40	1.85
19	1.01	0.99	2.45
20	1.02	0.83	1.21
21	1.00	0.95	**0.49**
22	1.22	1.02	1.50
23	0.93	1.05	1.27

[a] Consult *Table 11* on *Page 58* for background information on the genotyped variants.

The minor allele of the variant SNP 09, which is rs3810813 in the *SLX4* gene, was strongly associated with breast cancer when all cases were compared all controls. The OR was ~ 1.6. The corresponding population attributable risk (PAR), which was calculated from the frequencies of carriers in non-carriers in the breast cancer cases in control groups, is 5.6 %. The OR and thus the PAR were increased in cases ≤ 60 years (OR ~ 2.3, PAR = 11.5 %) and strongly increased in cases ≤ 40 years (OR ~ 4.5; PAR = 27.6 %).

In cases up to 60 years of age, the minor allele SNP 10, which is the second genotyped variant in the *SLX4* gene (rs3827530), was underrepresented in cases with a nominal significance (OR ~ 0.5). The minor allele of SNP 21, the common variant rs4986764 in the *BRIP* gene is underrepresented in cases aged ≤ 40 years (OR ~ 0.5) with nominal significance.

Evaluation of SNP 09 in sample subgroups

To elucidate which of the subgroups contributed to the association of SNP 09 to breast cancer, sporadic breast cancer cases and BRCA-negative familial breast cancer cases were separately analyzed.

The "Freiburg" and "Ulm" cohorts both included sporadic breast cancer cases and age-matched controls, and were studied together to reach sufficient group sizes. Here, the minor allele of SNP 09 was not overrepresented in sporadic breast cancer cases (n = 246, MAF = 0.06) compared to controls (n = 229, MAF = 0.05, P-trend = 0.69). Whereas there was also no significant association in cases aged ≤ 60 years (n = 61, MAF = 0.08, P-trend = 0.19), the minor allele was significantly overrepresented in cases aged ≤ 40 years (n = 10, MAF = 0.20, P-trend = 0.005), with an allelic OR of 4.69 (95 % CI: 1.45 – 15.1) and a genotypic OR of 6.51 (95 % CI: 1.70 – 24.9).

The BRCA-negative breast cancer cases from the "Familial" cohort (n = 189) were compared to the remaining controls from the "Population Controls", "Repair Controls" and "Twin" cohorts (n = 962). Here, the minor allele was highly significant overrepresented (P-trend = 0.00038) in cases (MAF = 0.09) in comparison to the controls (MAF = 0.05), with an allelic OR of 2.06 (95 % CI: 1.37 – 3.09) and a genotypic OR of 2.14 (95 % CI: 1.38 – 3.30). The overrepresentation of the minor allele was also significant in cases ≤ 60 years (n = 106, MAF = 0.11, P-trend = 0.00005) with an allelic OR of 2.57 (95 % CI: 1.60 – 4.13) and a genotypic OR of 2.66 (95 % CI: 1.58 – 4.47), and also in cases ≤ 40 years (n = 13, MAF = 0.15, P-trend = 0.011) with an allelic OR of 3.66 (95 % CI: 1.24 – 10.8) and a genotypic OR of 4.46 (95 % CI: 1.35 – 14.8).

Thus, the reproducible association of SNP 09 to the incidence of cancer was observed in the two independent subgroups.

3.2.3.2 Haplotype Analyses

The focus of this analysis was if the minor alleles of the rare variants which are situated in the same gene are situated on the same haplotype, and if particular haplotypes show an association to an increased risk of breast cancer. This was especially important concerning the associations of SNPs 09 and 10 in *SLX4*. Whereas the minor allele of SNP 09 was significantly overrepresented in breast cancer cases, SNP 10 showed on the contrary a borderline significant overrepresentation of the major allele in cases ≤ 60 years.

Two variants were genotyped in *NEK1, NINL, TIMELESS, C19ORF40, HLTF* and *BARD1* and three variants in *SLX4*. The haplotypes, as well as their respective frequencies in all probands (total) and split by cases and controls as well as the significance of the association of a haplotype to an increased breast cancer risk are shown in *Table 15* on *Page 66.*

Aside from *C19ORF40* (SNPS 13 and 14) and *SLX4 (SNPs 10 and 11)*, the minor alleles of the SNPs located in one gene are not situated on the same haplotypes, and no association of a haplotype to increased cancer risk is present. In the *C19ORF40* gene, the comparably less frequent minor allele of SNP 13 is exclusively present in combination with the minor allele of SNP 14, but none of the haplotypes shows an association to breast cancer.

The minor allele of SNP 09 in the *SLX4* gene is exclusive for one haplotype only, while the minor alleles of SNP 10 and SNP 11 constitute two further haplotypes. As such, the less frequent minor allele of SNP 10 is always present in combination with the minor allele of SNP 11.

Only the *SLX4* haplotype carrying the minor allele of SNP 09 was significantly associated with increased risk of breast cancer ($p = 0.0033$, after correction for multiple testing by permutation testing $p = 0.06$). This allele is tied to the major allele of SNP 10, which was also associated with breast cancer risk. Thus both alleles tag the same risk locus, which is located in a linkage disequilibrium block that includes the *SLX4* gene (see *Figure 15*). The low significance of the association of the major allele of SNP 10 is explained by its very high frequency (0.94), compared to that of the minor allele of SNP 09 (0.05), since only a small fraction of the carriers of the major allele of SNP 10 are also carriers of the true risk locus. The latter is much better tagged by the minor allele of SNP 09.

Table 15: Haplotype constellations in the genes with two or three genotyped variants. The official gene symbols as well as the respective single nucleotide polymorphism (SNP) numbers (Nr) constituting the haplotypes (SNP A, SNP B, SNP C) are given (see *Table 11*). WT (blue fields): wild-type (major) allele; VAR (red fields): variant (minor) allele. The respective haplotype frequencies in all probands (total) and split by cases and controls as well as the significance of an association to an increased breast cancer risk, indicated by the probability value (P) are shown. n: number of probands.

Official Gene Symbol [a]	SNP A (Nr.)	SNP B (Nr.)	SNP C (Nr.)	Haplotype Frequency			Significance (P Value)
				Total (n = 1624)	Cases (n = 435)	Controls (n = 1189)	
NEK1	**01**	**02**					
	WT	WT		0.838	0.834	0.840	P = 0.68
	VAR	WT		0.106	0.111	0.104	P = 0.56
	WT	VAR		0.056	0.055	0.056	P = 0.89
NINL	**03**	**04**					
	WT	WT		0.852	0.846	0.854	P = 0.58
	VAR	WT		0.104	0.101	0.105	P = 0.77
	WT	VAR		0.043	0.052	0.040	P = 0.16
TIMELESS	**05**	**06**					
	WT	WT		0.940	0.939	0.941	P = 0.87
	VAR	WT		0.049	0.049	0.049	P = 0.94
	WT	VAR		0.011	0.011	0.011	P = 0.81
SLX4	**09**	**10**	**11**				
	WT	WT	WT	0.867	0.854	0.871	P = 0.20
	WT	VAR	VAR	0.058	0.053	0.061	P = 0.41
	VAR	WT	WT	0.053	0.072	0.046	**P = 0.0033**
	WT	WT	VAR	0.022	0.021	0.022	P = 0.84
C19ORF40	**13**	**14**					
	WT	WT		0.882	0.885	0.881	P = 0.76
	WT	VAR		0.073	0.069	0.074	P = 0.60
	VAR	VAR		0.045	0.046	0.045	P = 0.86
HLTF	**17**	**18**					
	WT	WT		0.942	0.953	0.938	P = 0.11
	VAR	WT		0.037	0.029	0.040	P = 0.15
	WT	VAR		0.021	0.018	0.022	P = 0.50
BARD1	**19**	**20**					
	WT	WT		0.963	0.963	0.962	P = 0.92
	WT	VAR		0.019	0.018	0.019	P = 0.92
	VAR	WT		0.019	0.018	0.019	P = 0.97

[a] Official gene symbols provided by the Human Genome Organization.

The phased haplotype display of the *SLX4* gene region in the HapMap CEU population is shown in *Figure 15* on *Page 67*.

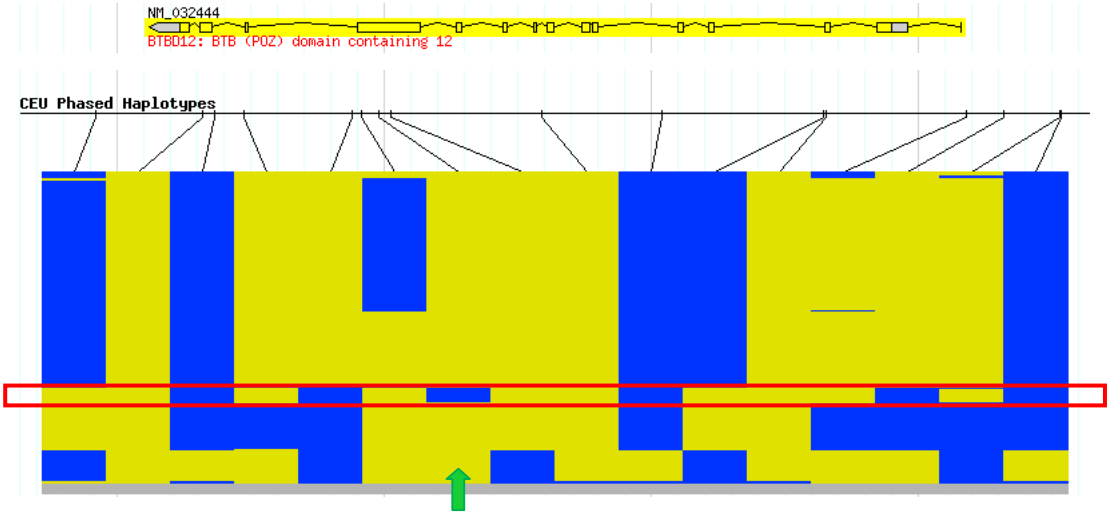

Figure 15: Phased haplotype display (lower panel) of rs3810813 (SNP 09, green arrow) and further polymorphisms in the *SLX4* gene region (upper panel) in the HapMap CEU population consisting of 242 haplotypes of 121 individuals. The both alleles of each SNP are shown in blue and yellow color. The minor allele of SNP 09 tags a long haplotype (red frame) which exceeds the gene in both directions. Source: http://hapmap.ncbi.nlm.nih.gov/ (accessed 11.01.2011). *BTBD12* is an alternative name of the *SLX4* gene. NM_032444 is the accession number of the reference sequence of the gene in the National Center for Biotechnology Information (NCBI) database. SNP: single nucleotide polymorphism. CEU: Utah Residents with Northern and Western European Ancestry.

It revealed that no other known polymorphism within this region provides a splitting of the haplotype tagged by the minor allele of SNP 09 and no further variant with a possibly stronger association to cancer risk could be defined.

Therefore it was evaluated, if the minor allele of SNP 09 (rs3810813) itself could have an impact on the structure and function of the Slx4 protein. This was done by studying the effects of the polymorphism *in-silico* with the help of several databases and tools which are accessible online.

3.2.3.3 *In-silico* Studies of the Impact of the Minor Allele of SNP 09 on the Slx4 Function

As stated, the presence of the minor allele of SNP 09 leads to an amino acid change from serine to phenylalanine at amino acid residue 1271 in the Slx4 protein. According to the Expert Protein Analysis System Server (ExPASy, http://www.expasy.org, accessed 11.01.2011, [71]) of the Swiss Institute of Bioinformatics, Switzerland, which hosts amongst others the UniProt-Knowledgebase/Swiss-Prot database (http://www.uniprot.org/uniprot, accessed 11.01.2011,

[24,229]), the residue change is classified as a natural variant (catalogue number VAR_019327) in the Slx4 protein (catalogue number Q8IY92).
It is located within a region necessary for the interaction with Plk1 and the Terf2-Terf2ip telomere binding complex, and at the border to a region necessary for the interaction with the Mus81-Eme1 heterodimer. These interacting proteins are closely related to the Slx4 function in the homologous recombination system. The physiochemical properties of the change are from a small polar amino acid (Serine) to a large and aromatic amino acid (Phenylalanine). The Blosum score [86], which states the logarithmic likelihood ratio of the occurrence of two particular amino acids at the same residue by chance (range: -4 to + 11), is -2, indicating the residue change is comparably rare.

The alignment to ortholog Slx4 proteins of other species revealed that the serine residue is evolutionary conserved among mammals and also in the yeasts *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe.* In *S. cerevisiae*, this residue is classified as a potential phosphorylation target of *ATM* or *ATR*. The residue is not conserved in *Gallus gallus* (Chicken) and *Drosophila melanogaster* (fruit fly).

The effect of the residue change on the structure of the Slx4 protein was assessed with the Polymorphism Phenotyping v2 (PolyPhen-2) software tool (http:/genetics.bwh.harvard.edu/ggi/pph2/index.shtml, version 2.0.23r349, accessed 11.01.2011) [1]. The software calculates the probability for a damaging effect from a variety of factors and sources, including the amino acid sequence and structural features of the protein and information from known damaging alleles which cause human Mendelian diseases. In the “HumanDiv” model, which is suitable for the evaluation of rare alleles at loci involved in complex phenotypes, it was classified as probably damaging, which is the highest possible classification category. The residue change reached a probability of 0.975 (range 0 – 1) for a damaging effect, with a sensitivity of 0.7 and a specificity of 0.94. In the other model, “HumanVar”, which is more suitable for the diagnosis of Mendelian diseases but less suited for complex phenotypes, the probability for a damaging effect is 0.966 with a sensitivity of 0.58 and a specificity of 0.90.

Taken together, these results indicate a probable impact of the amino acid residue change on the function of the Slx4 protein.

3.2.4 Association Between Genotypes and DNA Repair Capacity

The influence of the variants on the outcome of the baseline and the induced micronucleus assay (MN-b and MN-i, respectively), the baseline and the induced sister chromatid exchange assay (SCE-b and SCE-i, respectively) and the mitotic delay assay (MD) was studied by assuming that a possible alteration of the cellular DNA repair capacity by one of the variants constitutes a dominant effect. The DNA repair assay results of the carriers of one or two variant alleles of the respective SNPs were compared to those of non-carriers and evaluated for the presence of significant differences. These analyses are summarized in *Table 16.*

Table 16: Association of the presence of the minor alleles of the genotyped polymorphisms with differences in the results of the respective deoxyribonucleic acid (DNA) repair assays. The assays are the baseline and the induced micronucleus assay (MN-b and MN-i, respectively), the baseline and the induced sister chromatid exchange assay (SCE-b and SCE-i, respectively) and the mitotic delay assay (MD). The significance is indicated by the probability value (P). n: number of probands. SNP: single nucleotide polymorphism.

SNP Number [a]	MN-b	MN-i	SCE-b	SCE-i	MD
	n = 190	n = 436	n = 145	n = 141	n = 186
01	P = 0.80	P = 0.14	P = 0.59	P = 0.27	P = 0.60
02	P = 0.51	P = 0.81	P = 0.30	P = 0.61	P = 0.19
03	P = 0.17	P = 0.80	P = 0.17	P = 0.13	P = 0.44
04	P = 0.36	P = 0.15	P = 0.79	P = 0.14	**P = 0.02**
05	P = 0.89	P = 0.54	P = 0.97	P = 0.41	P = 0.35
06	P = 0.30	P = 0.26	P = 0.17	P = 0.64	P = 0.86
07	P = 0.19	P = 0.37	P = 0.41	P = 0.58	P = 0.53
08	P = 0.67	P = 0.98	P = 0.48	P = 0.68	P = 0.33
09	P = 0.77	**P = 0.006**	P = 0.35	P = 0.78	P = 0.59
10	P = 0.21	P = 0.32	P = 0.92	P = 0.37	P = 0.77
11	P = 0.15	P = 0.87	P = 0.52	P = 0.32	P = 0.89
12	P = 0.23	P = 0.64	P = 0.80	P = 0.78	P = 0.93
13	P = 0.22	P = 0.87	P = 0.48	P = 0.62	P = 0.74
14	P = 0.42	P = 0.27	P = 0.39	P = 0.39	P = 0.31
15	P = 0.33	P = 0.56	P = 0.65	P = 0.78	P = 0.57
16	-	-	-	-	-
17	P = 0.53	P = 0.29	P = 0.68	P = 0.68	P = 0.71
18	P = 0.60	P = 0.26	P = 0.87	P = 0.83	P = 0.57
19	P = 0.48	P = 0.61	P = 0.15	P = 0.68	P = 0.20
20	P = 0.87	P = 0.42	P = 0.89	P = 0.67	P = 0.87
21	P = 0.11	P = 0.42	P = 0.60	P = 0.83	P = 0.63
22	P = 0.32	P = 0.72	P = 0.61	P = 0.86	P = 0.57
23	P = 0.23	P = 0.59	P = 0.99	P = 0.90	P = 0.96

[a] See *Table 11* on *Page 58* for background information on the variants.

The outcome of the radiation induced MN assay was significantly influenced by the presence of the minor allele of SNP 09 (rs3810813) in the *SLX4* gene, as was the MD assay by the presence of the minor allele of SNP 04 (rs13044759) in the *NINL* gene.

Table 17 and *Figure 16* provide further details for the influence of SNP 09 on the radiation induced MN frequency for cases and controls.

Table 17: Differences between non-carriers and carriers of the variant allele of SNP 09 (rs3810813) in the radiation induced micronucleus (MN) assay, split by healthy controls (Controls) and sporadic breast cancer cases (Cases). Shown are the number of probands (n), the means and standard deviations (SD) of each proband group as well as the significance of differences between non-carriers and carriers of the minor allele, indicated by the probability value (P). SNP: single nucleotide polymorphism.

SNP 09 and induced MN assay	Controls		Cases	
	Non-carriers	Carriers	Non-carriers	Carriers
n	272	30	118	16
Mean ± SD	0.97 ± 0.22	1.01 ± 0.17	1.02 ± 0.29	1.24 ± 0.36
Difference	P = 0.30		**P = 0.008**	

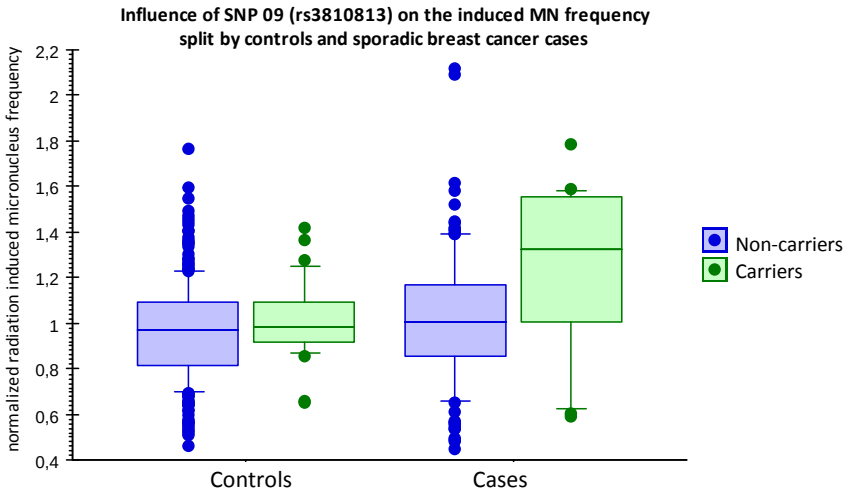

Figure 16: **Differences of the radiation induced micronucleus (MN) frequency in carriers of a variant allele of SNP 09 (rs3810813) compared to non-carriers.** The analysis is split by healthy controls (Controls) and sporadic breast cancer cases (Cases). SNP: single nucleotide polymorphism.

Carriers of this variant (n = 46) exhibited a significantly increased mean radiation induced MN frequency compared to non-carriers (n = 390), as shown in *Table 16*.

Considering controls and sporadic breast cancer cases separately revealed that this increase was nearly exclusively due to a strong increase of the induced MN frequency in the breast cancer

cases which carry the minor allele. The median of the carriers is close to the 90th percentile of controls. In controls, there was no difference between carriers and non-carriers. However, the box plots revealed that the 25th percentile is comparably higher in carriers of the variant, indicating that also in controls the carriers are generally not found among those individuals with a low induced MN frequency. Notably, the same variant was highly significantly associated with an increased risk of breast cancer compared to controls (see *3.2.3.1*). The *SLX4* gene is involved in the homologous recombination of DNA double strand breaks. A failure in the repair of these lesions is denoted with the MN assay.

Table 18 and *Figure 17* provide details for the influence of SNP 04 on the mitotic delay.

Table 18: Differences between non-carriers and carriers of the variant allele of SNP 04 (rs13044759) in the mitotic delay (MD) assay, split by healthy controls (Controls) and sporadic breast cancer cases (Cases). Shown are the number of probands (n), the means and standard deviations (SD) of each proband group as well as the significance of differences between non-carriers and carriers of the minor allele, indicated by the probability value (P). SNP: single nucleotide polymorphism.

SNP 04 and MD index	Controls		Cases	
	Non-carriers	Carriers	Non-carriers	Carriers
n	149	15	18	5
Mean ± SD	3.81 ± 0.90	4.56 ± 0.98	3.01 ± 0.92	3.31 ± 0.62
Difference	**P = 0.003**		P = 0.51	

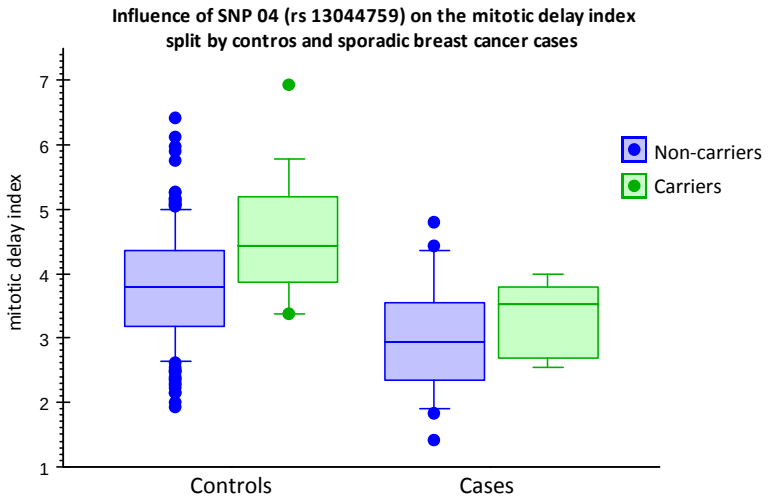

Figure 17: Differences of the mitotic delay index in carriers of a variant allele of SNP 04 (rs13044759) compared to non-carriers. The analysis is split by healthy controls (Controls) and sporadic breast cancer cases (Cases). SNP: single nucleotide polymorphism.

A significantly increased mean MD index was found in carriers of this variant (n = 20) compared to non-carriers (n = 167) (see *Table 16*), thus the G_2 delay upon DNA damage induction is stronger in

carriers compared to non-carriers. A separate consideration of controls and sporadic breast cancer cases revealed that the increase of the MD index in carriers of the variant compared to carriers of the wild type genotype was significant only in controls, while it was not in sporadic breast cancer cases. The lack of significance is probably due to the low numbers in cases, with 18 carriers of the wild type genotype and only 5 carriers of the variant. This is underlined by the box plots, which revealed that in controls as well as in cases, the median of the MD index in carriers is similar to the 75th percentile of the MD index in the non-carriers. These results indicate that the variant is generally associated with an increased G_2 delay in both controls and sporadic breast cancer cases. Notably, the *NINL* gene is involved in the G_2/M checkpoint, and differences in the G_2 delay upon DNA damage induction are denoted with the MD assay.

The overall increase in the induced MN frequency that was associated with the variant allele of SNP 09 was due to a strong effect in sporadic breast cancer cases only. Therefore, effects of the other genotyped polymorphisms on the results of the DNA repair assays were also analyzed separately in cases and in controls.

Only borderline significant associations were found when breast cancer cases and controls were separately studied. These are summarized *Table 19.*

Table 19: Association between the genotyped variants and differences in the outcome of deoxyribonucleic acid (DNA) repair assays which reached nominal significance only in the sporadic breast cancer cases or the controls group. Shown are: the proband group; the number (Nr.) of the associated single nucleotide polymorphism (SNP); the assay results (of the baseline and the induced sister chromatid exchange assay (SCE-b and SCE-i respectively) as well as of the baseline micronucleus assay (MN-b)); number of probands (n) as well as the means and standard deviations (SD) for carriers and for non-carriers of the respective variants; the significance of the association given by the probability value (P).

Group	SNP Nr.	Assay	Carriers		Non-Carriers		Significance (P Value)
			n	Means ± SD	n	Means ± SD	
Cases	21	SCE-b	9	6.39 ± 1.0	4	9.36 ± 1.65	P = 0.02
	21	SCE-i	9	13.12 ± 2.87	4	17.15 ± 2.37	P = 0.03
	22	MN-b	2	1.70 ± 0.87	30	0.95 ± 0.38	P = 0.02
Controls	21	MN-b	100	0.96 ± 0.43	56	1.10 ± 0.43	P = 0.045

Three of the four associations were seen with the SNP 21 in the *BRIP1* gene, which is the only common variant studied (see *3.2.1*) and one with SNP 19, the in-frame deletion in *BARD1*. All associations have a very borderline significance and the number of sporadic breast cancer cases is low. Thus, they are likely to be influenced by chance.

3.2.5 Further Analyses of SNP 09

DNA repair assay results in discordant twin pairs

Of the ten dizygous twin pairs, four were discordant for the presence of the minor allele of SNP 09 (rs3810813 in the *SLX4* gene), despite the low MAF of this variant. Therefore, the results of the DNA repair assays were compared between carriers and non-carriers of each pair, as presented in *Table 20.* The statistical difference of the means was assessed with the nonparametric paired Wilcoxon test.

Table 20: Deoxyribonucleic acid (DNA) repair assay results in carriers (WT) and non-carriers (VAR) of the minor allele of SNP 09 in discordant dizygous twin pairs, as well as the significance of a difference between the means (Wilcoxon-test), indicated by the probability value (P). Higher values are displayed in italics. The assays are the baseline and the induced micronucleus assay (MN-b and MN-i, respectively), the baseline and the induced sister chromatid exchange assay (SCE-b and SCE-i, respectively) and the mitotic delay assay (MD). Normalized MN frequencies (see *3.1*) are shown. n: number of probands.

	MN-b		**MN-i**		**SCE-b**		**SCE-i**		**MD**	
n	4		4		3		3		4	
Genotype	WT	VAR	WT	VAR	WT	VAR	WT	VAR	WT	VAR
Pair 1	0.41	*0.87*	1.05	*1.08*	8.90	*9.16*	13.05	*14.00*	*3.37*	2.40
Pair 2	*0.79*	0.42	0.82	*0.97*	*5.25*	4.20	5.27	*8.38*	3.36	*3.98*
Pair 3	*1.05*	0.80	0.72	*1.01*	4.50	*7.45*	9.30	*13.33*	3.37	*4.22*
Pair 4	0.70	*0.90*	*0.67*	0.66	-	-	-	-	4.93	*5.16*
Difference	P = 0.99		P = 0.14		P = 0.59		P = 0.11		P = 0.71	

In those individuals of the four dizygous twin pairs which carry the minor allele of SNP 09, the means of the baseline and the induced MN frequency, of the baseline and the induced SCE frequency, and of the mitotic delay are increased compared to non-carriers. However, due to the small number of discordant dizygous twin pairs and the large standard deviations, no statistical significance is reached.

Frequency of the minor allele in *BRCA*-positive breast cancer families

In breast cancer families with the presence of a mutation in *BRCA1* or *BRCA2*, one affected index person was defined. The allele frequency of the minor allele of SNP 09 in the *BRCA*-positive index patients (n = 84, MAF = 0.065) was not significantly different from that of the controls (n = 1189, MAF = 0.05), P-trend = 0.31.

For some *BRCA*-positive families, only a single DNA sample of an unaffected *BRCA*-positive individual was available (n = 15). In this group, the minor allele frequency was clearly increased

(MAF = 0.20) compared to controls, which was statistically significant (P trend = 0.00014), with an allelic OR of 4.97 (95 % CI: 1.990 – 12.386) and a genotypic OR of 6.65 (95 % CI: 2.324 – 19.049).
In the families with multiple available DNA samples in which also carriers of the minor allele of SNP 09 were present (n = 9), the allele showed no noticeable co-segregation with the cancer incidence, as in six of the families the carriers were exclusively unaffected women.

3.3 Characterization of Assays on Cellular DNA Repair Capacity

The baseline and the induced micronucleus assay (MN-b and MN-i, respectively), the baseline and the induced sister chromatid exchange assay (SCE-b and SCE-i, respectively) and the mitotic delay assay (MD) were performed in parallel from same peripheral blood samples of the population controls for the DNA repair analysis cohort ("Repair Controls"), of the twin pairs cohort ("Twin") and of a subgroup of probands from the "Ulm" cohort. The induced MN assay was additionally performed on the "Freiburg" cohort and on further samples from the "Ulm" cohort.

3.3.1 Heritability of Measures of Cellular DNA Repair Capacity

To evaluate if inter-individual differences which are observed with a certain assay are applicable as a measure for the individual cellular DNA repair capacity, the respective assay is required to reflect the variability which is introduced by differences in the genetic background. Heritability estimates provide a measure, to which proportion the genetic factors contribute to this observed inter-individual variability of the respective DNA repair assay results, in contrast to the variation caused by individually exclusive environmental factors.
Here, the heritability of the respective assays was estimated in a twin study. This type of analysis is based on the comparison of the respective within-pair differences of the assay results between monozygous twin pairs, dizygous twin pairs and randomly paired controls. The following results concerning the baseline and the induced micronucleus assay have been published [216].

Monozygous (MZ) and dizygous (DZ) twin pairs from the "Twin" cohort as well as random control pairs (CP), generated from age- and gender-matched probands from the "Repair Control" cohort, were used for the generation of heritability estimates. The DNA repair assays failed in one of the MZ pairs and in one individual of a further MZ pair, thus only 40 MZ pairs could be included. Sample characteristics concerning age and gender are shown in *Table 21* on *Page 76*.

Mean values and standard deviation (SD) of the baseline and the induced micronucleus frequency, of the baseline and the induced sister chromatid exchange frequency and of the mitotic delay index are summarized in *Table 22* on *Page 76* for MZ, DZ and CP.

Table 21: Characteristics of the probands from the "Twin" cohort, split by monozygous (MZ) and dizygous (DZ) twin pairs, and age and gender-matched control pairs (CP) from the "Repair Control" cohort used for the generation of heritability estimates**.** Mean age and standard deviation (SD) is displayed in years (y). n: number of probands.

Group	n individuals	Mean age (y) ± SD	Gender (male/female)	n pairs
MZ	80	27,9 ± 11,7	(16 / 64)	40
DZ	20	29,4 ± 12,2	(10 / 10)	10
CP [a]	76	22,4 ± 1,2	(14 / 62)	38[a]

[a)] Control pairs were repeatedly generated from control individuals with a randomization procedure (see *2.4.1*).

Table 22: Mean values and standard deviations (SD) of the baseline and the induced micronucleus assay (MN-b and MN-i, respectively), the baseline and the induced sister chromatid exchange assay (SCE-b and SCE-i, respectively) and the mitotic delay assay (MD) from the probands used for the generation of heritability estimates, split by monozygous (MZ) and dizygous (DZ) twin pairs and control pairs (CP). n: number of probands.

Group	n[a]	Mean Values ± SD				
		MN-b [b]	MN-i [b]	SCE-b	SCE-i	MD
MZ	80	47.86 ± 21.40	246.56 ± 54.42	7.12 ± 1.62	13.85 ± 2.53	3.89 ± 0.94
DZ	20	40.49 ± 14.01	232.05 ± 50.56	7.50 ± 2.15	12.68 ± 4.10	3.95 ± 1.16
CP	76	58.35 ± 23.67	229.24 ± 34.97	7.46 ± 1.83	14.58 ± 2.87	3.97 ± 0.80

[a)] Total numbers. Numbers of probands or pairs with available results deviate slightly for each assay (see Tables 3.04 to 3.08).
[a)] MN values were not normalized on the group means because there were no differences between the "Repair Controls" and "Twin" cohorts.

The mean values of all DNA repair assays show only slight deviations between MZ, DZ or CP sample groups. The baseline and the radiation induced MN frequency were not normalized on the cohort's means because there were no differences between the "Twin" and the "Repair Controls" cohort.

Influence of age, gender and delay of the culture setup

Influences of age and gender on the repair assay results were mostly absent from the sample groups and reached a statistically significant influence on only a few results. In each case the effect was very small, with $r^2 \leq 0.2$. A borderline influence of gender was seen on the induced MN frequency in DZ ($r^2 = 0.21$, $P = 0.05$). In MZ pairs, age was weakly correlated with the baseline and the induced MN frequency ($r^2 = 0.05$, $P = 0.04$ and $r^2 = 0.11$, $P = 0.003$, respectively) and with the MD index ($r^2 = 0.09$, $P = 0.008$). A further borderline association of age with the baseline MN frequency was seen in CP ($r^2 = 0.05$, $P = 0.05$). The delay of the culture setup of the samples which were sent by mail, in comparison to those collected locally, showed a weak influence only on the baseline and the induced MN frequency in MZ ($r^2 = 0.1$, $P = 0.006$ and $r^2 = 0.09$ and $P = 0.006$). Despite the weakness of these influences, the calculations of intraclass correlations and

heritability estimates were nevertheless adjusted for these effects. Adjusted and not adjusted model yielded only negligible differences, thus only the data for the adjusted model is shown.

Within-pair differences of DNA repair measures

The within pair differences of MZ, DZ and CP are shown for the baseline MN frequency *Figure 18 (A)* and for the radiation induced MN frequency in *Figure 18 (B)*.

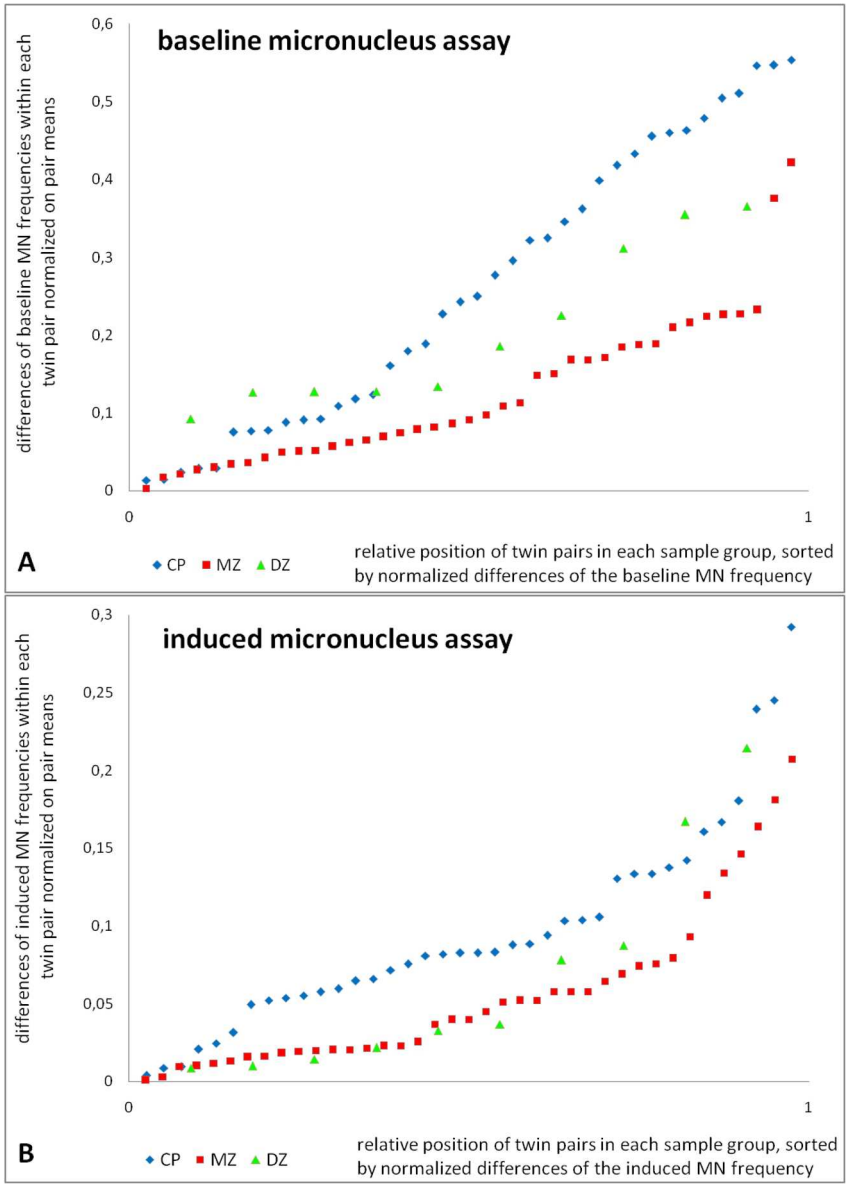

Figure 18: **Within-pair differences of the baseline (A) and the induced (B) micronucleus (MN) frequency in monozygous twin pairs (MZ, red squares), dizygous twin pairs (DZ, green triangles) and control pairs (CP, blue diamonds).** The pairs of each group are sorted according to normalized within-pair differences of the baseline (A) or the induced (B) MN frequency (ordinates) and distributed evenly (abscissas).

The within pair differences in MZ are clearly smaller than in CP for both types of the MN assay. The DZ pairs show intermediate within pair differences, although those DZ pairs with small differences overlap with CP in the case of the baseline MN frequency and with MZ in the case of the induced MN frequency. This overlap might be due to the comparably small numbers of DZ pairs (n = 10) and chance coincidences of matching assay results within a DZ pair.

Within-pair differences of the baseline and the BPDE induced sister chromatid exchange frequency in MZ, DZ and CP are visualized in *Figure 19 (A)* and (*B)*.

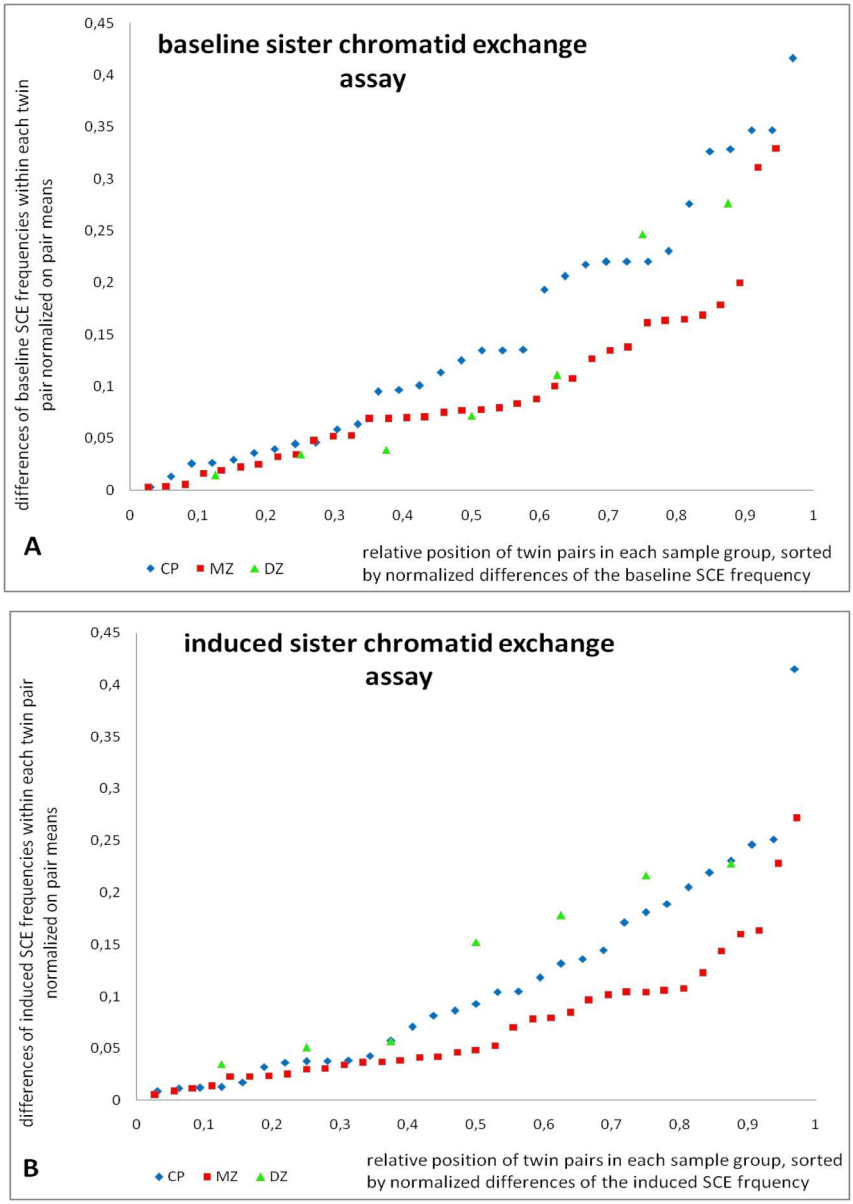

Figure 19: **Within-pair differences of the baseline (A) and the induced (B) sister chromatid exchange (SCE) frequency in monozygous twin pairs (MZ, red squares), dizygous twin pairs (DZ, green triangles) and control pairs (CP, blue diamonds).** The pairs of each group are sorted according to normalized within-pair differences of the baseline (A) or the induced (B) SCE frequency (ordinates) and distributed evenly (abscissas).

For both types of the SCE assay, the within-pair differences of MZ were smaller than those of CP. The differences of DZ showed no clear separation, they overlap with those of MZ and CP in the baseline SCE assay, and in the induced SCE assay they are largely comparable with the CP differences. However, only 7 out of the 10 DZ pairs could be used for this evaluation (see Tables *21* and *22*). This low number is due to the comparably high failure rate of the SCE assays in this study, which occasionally lead to missing values in one of the individual pair samples. Aside from DZ pairs the number of MZ pairs was also reduced. Generally, CP showed weakly increased within-pair differences compared to MZ in both assays. The differences of DZ are unsteady and vary between those of MZ or DZ.

Within-pair differences of MZ, DZ and CP in the mitotic delay assay are visualized in *Figure 20.*

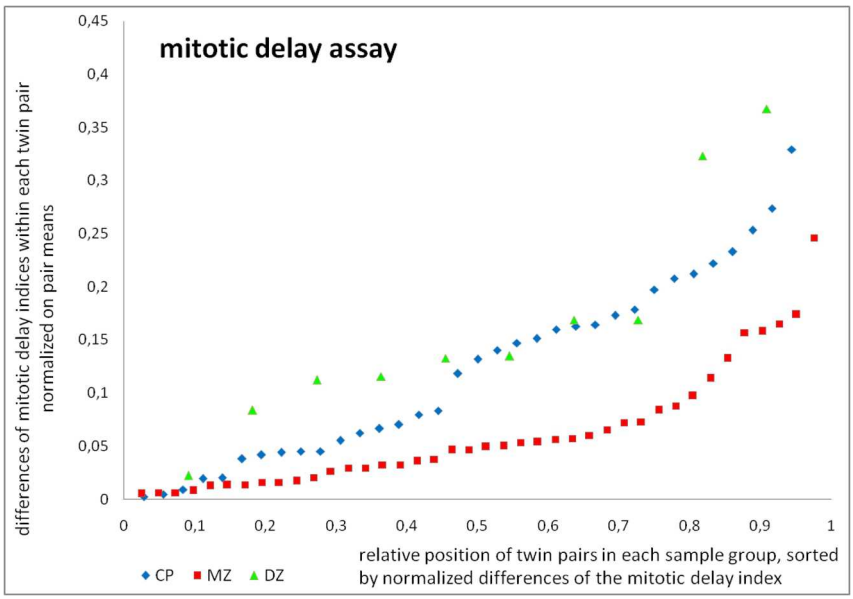

Figure 20: **Within-pair differences of the mitotic delay (MD) index in monozygous twin pairs (MZ, red squares), dizygous twin pairs (DZ, green triangles) and control pairs (CP, blue diamonds).** The pairs of each group are sorted according to normalized within-pair differences of the MD index (ordinates) and distributed evenly (abscissas).

The mitotic delay assay revealed a clear distinction of the within-pair differences of the MZ and CP sample groups. The differences of DZ are comparable to those of CP, despite the lack of shared genetic components in the latter. The DZ pairs with small within-pair differences overlap with CP, and two DZ pairs are among the pairs with the largest within-pair differences.

In general, the within-pair differences increase with a decreasing degree of shared genetic components. The gap between MZ pairs, which share all genetic components, and CP pairs, which share none, is very clear in the baseline and the induced MN assay and in the MD assay, while it is

less pronounced in the baseline and the induced SCE assay. The within-pair differences of DZ are varying but are largely situated in between those of MZ and CP, which was expected from the intermediate ratio of shared genetic components. The only exception was the MD assay, where the DZ differences were larger than those of CP.

Intraclass correlation coefficients

The different intraclass correlation coefficients (ICC) of MZ, DZ and CP for the single assays are shown in *Table 23*. The ICC gives a measure of the correlation of the assay results within the pairs of the respective groups.

Table 23: Intraclass correlation coefficients of monozygous (MZ) and dizygous (DZ) twin pairs and control pairs (CP) for the baseline and the induced micronucleus assay (MN-b and MN-i, respectively), the baseline and the induced sister chromatid exchange assay (SCE-b and SCE-i, respectively) and the mitotic delay assay (MD). n: number of probands. r: correlation coefficient.

Group	n[a]	Intraclass Correlation Coefficients				
		MN-b	MN-i	SCE-b	SCE-i	MD
MZ	80	r = 0.63	r = 0.79	r = 0.17	r = 0.51	r = 0.78
DZ	20	r = 0.31	r = 0.74	r = 0.00	r = 0.43	r = 0.00
CP	76	r = 0.00	r = 0.00	r = 0.03	r = 0.00	r = 0.00

[a] Total numbers. Numbers of probands or pairs with available results deviate slightly for each assay (see Tables 3.04 to 3.08)

The correlation is clearly present in MZ twin pairs, with the exception of the very weak correlation of the baseline SCE frequency. The induced MN frequency and the mitotic delay index show the strongest correlations in MZ. Whereas the correlation in DZ twin pairs is quite strong for the induced MN frequency, it is comparably lower for the baseline MN frequency. The same holds true for the SCE assay, but the correlation is moderately present in the induced and completely absent from the baseline SCE frequency. DZ pairs do also not show any correlation for MD. There is a lack of correlation for any of the assay results in CP.

Heritability estimates of DNA repair assays

Heritability estimates were generated by biometric modeling (ACE model) and direct variance comparison of within-pair differences. The h^2 value can be interpreted as the percentage to which the observed variability of a trait, here the respective DNA repair assay results, can be attributed to heritable genetic factors.

Table 24 on *Page 81* shows the heritability estimates for the baseline MN frequency.

Table 24: Heritability estimates (h^2) of the baseline micronucleus (MN) frequency determined with biometric modeling and direct variance comparison between monozygous twin pairs (MZ), dizygous twin pairs (DZ) and randomly paired controls (CP). Also shown are the significance, indicated by the probability value (P-value) and the 95 % confidence intervals (95 % CI). n: number of probands (pairs); vs.: versus.

Heritability estimate	Comparison	Baseline MN frequency			
		n (pairs)	h^2	P-value	[95 % CI][a]
biometric modeling	**MZ vs. CP**	**38 | 36**	**0.68**	**< 0.0001**	**[0.28 – 1]**
	MZ vs. DZ	38 | 10	0.08	0.43	[0 – 0.95]
	DZ vs. CP	10 | 36	-[b]	-	-
Direct Variance Comparison	**MZ vs. CP**	38 | 36	**0.72**	**0.0011**	**[0.52 – 0.92]**
	MZ vs. DZ	38 | 10	0.04	0.25	[0 – 1]
	DZ vs. CP	10 | 36	0.49	*0.013*	[0 – 1]

[a] Statistical test and confidence interval are based on different approaches and their results may appear sometimes inconsistent.
[b] The model could not be fitted on the observed data.

When MZ were compared to CP, both heritability estimates yielded a high degree of heritability of the baseline MN frequency, with $h^2 \sim 0.7$. The null hypothesis, assuming no heritable contribution to the observed variability of baseline MN frequencies, was clearly rejected, as displayed by the highly significant P-values and the confidence intervals which exclude $h^2 = 0$. In the ACE model, the variance components were 68.5 % for the additive genetic component (A), 0 % for the common environment component (C) and 31.5 % for the unique environment (E) component. The direct comparison of within-pair variances yielded a similar result.

Due to the small numbers of DZ that could be recruited (n = 10), and because the within-pair differences of some DZ pairs are smaller than those of most MZ pairs, heritability estimates are inconclusive for the MZ to DZ and DZ to CP comparisons. The ACE model could not be fitted on the observed data when DZ were compared to CP.

As CP comprise the whole variability of the population and could be recruited in a sufficient number, the MZ to DZ comparison is probably the most robust one.

Table 25 on *Page 82* shows the heritability estimates for the radiation induced MN frequency.

The radiation induced MN frequency showed a similar degree of heritability as the baseline MN frequency in both approaches when MZ were compared to CP, with $h^2 \sim 0.6$ and also a clear rejection of the absence of a heritable contribution to the observed variability. The ACE model yielded estimated proportions of 68.0 % for the additive genetic component, 3.5 % for the common environmental and 28.5 % for the unique environmental component as contributors to

the observed variability. Heritability estimates by direct variance comparison yielded a similar result.

For the same reasons that apply for the baseline MN frequency, the MZ to DZ and DZ to CP comparisons were also inconclusive, and the ACE model could not be fitted on the observed data structure when DZ were compared to CP. Thus, the MZ to CP comparison can again be interpreted as the most robust heritability estimate.

Table 25: Heritability estimates (h^2) of the induced micronucleus (MN) frequency determined with biometric modeling and direct variance comparison between monozygous twin pairs (MZ), dizygous twin pairs (DZ) and randomly paired controls (CP). Also shown are the significance, indicated by the probability value (P-value) and the 95 % confidence intervals (95 % CI). n: number of probands (pairs); vs.: versus.

Heritability estimate	Comparison	Induced MN frequency			
		n pairs	h^2	P-value	[95 % CI][a]
biometric modeling	**MZ vs. CP**	38 | 36	**0.68**	**0.006**	**[0 – 1]**
	MZ vs. DZ	38 | 10	0.08	0.37	[0 – 0.56]
	DZ vs. CP	10 | 36	-[b]	-	-
Direct Variance Comparison	**MZ vs. CP**	38 | 36	**0.57**	**< 0.0001**	**[0.28 – 0.87]**
	MZ vs. DZ	38 | 10	0.16	0.76	[0 – 1]
	DZ vs. CP	10 | 36	0.71	0.40	*[0.40 – 1]*

[a] Statistical test and confidence interval are based on different approaches and their results may appear sometimes inconsistent.
[b] The model could not be fitted on the observed data.

The heritability estimates for the baseline SCE frequency are shown in *Table 26.*

Table 26: Heritability estimates (h^2) of the baseline sister chromatid exchange (SCE) frequency determined with biometric modeling and direct variance comparison between monozygous twin pairs (MZ), dizygous twin pairs (DZ) and randomly paired controls (CP). Also shown are the significance, indicated by the probability value (P-value) and the 95 % confidence intervals (95 % CI). n: number of probands (pairs); vs.: versus.

Heritability estimate	Comparison	Baseline SCE frequency			
		n pairs	h^2	P-value	[95 % CI][a]
biometric modeling	MZ vs. CP	36 | 32	0.24	0.11	[0 – 0.63]
	MZ vs. DZ	36 | 7	0.00	0.10	[0 – 1]
	DZ vs. CP	7 | 32	-[b]	-	-
Direct Variance Comparison	MZ vs. CP	36 | 32	0.24	0.49	[0 – 0.83]
	MZ vs. DZ	36 | 7	0.08	0.71	[0 – 1]
	DZ vs. CP	7 | 32	0.17	0.52	[0 – 1]

[a] Statistical test and confidence interval are based on different approaches and their results may appear sometimes inconsistent.
[b] The model could not be fitted on the observed data.

When MZ were compared to CP, both approaches yielded an identical but low degree of estimated heritability for the baseline SCE frequency with $h^2 = 0.24$. The null hypthesis, assuming no heritabile contribution to the observed variances, is not rejected and the confidence intervals do also include the absence of heritability. This indicates some uncertainity and there is no evidence that the baseline SCE frequency is influenced by genetic factors. Like in the estimates of the baseline or the induced MN frequency, the number of DZ pairs (n = 7) is not sufficient to generate reliable heritability estimates with any of the two approaches, if DZ are compared to MZ or CP. As the confidence intervals range from 0 to 1, the null hypothesis, assuming no heritable contribution to the observed variability, cannot be rejected.

Table 27 shows the heritability estimates for the BPDE induced SCE frequency.

Table 27: Heritability estimates (h^2) of the induced sister chromatid exchange (SCE) frequency, determined with biometric modeling and direct variance comparison between monozygous twin pairs (MZ), dizygous twin pairs (DZ) and randomly paired controls (CP). Also shown are the significance, indicated by the probability value (P-value) and the 95 % confidence intervals (95 % CI). n: number of probands (pairs); vs.: versus.

Heritability estimate	Comparison	Induced SCE frequency			
		n pairs	h^2	P-value	[95 % CI][a]
biometric modeling	MZ vs. CP	35 | 31	0.20	0.13	[0 – 0.56]
	MZ vs. DZ	35 | 7	0.35	0.38	[0 – 1]
	DZ vs. CP	7 | 31	-[b]	-	-
Direct Variance Comparison	**MZ vs. CP**	**35 | 31**	**0.48**	**0.027**	**[0.07 – 0.89]**
	MZ vs. DZ	35| 7	0.36	0.83	[0 – 1]
	DZ vs. CP	7 | 31	0.19	0.31	[0 – 1]

[a] Statistical test and confidence interval are based on different approaches and their results may appear sometimes inconsistent.
[b] The model could not be fitted on the observed data.

In the ACE model, a heritability of $h^2 \sim 0.2$ was estimated for the induced SCE frequency, however no significant contribution of the heritable factors could be affirmed when MZ were compared to DZ. In contrast, the direct variance comparison yielded an intermediate degree of heritability, with $h^2 \sim 0.5$, supported by the rejection of the absence of genetic factors as well as by the confidence intervals.

Again, both approaches did not yield conclusive heritability estimates for MZ to DZ or DZ to CP comparisons.

The heritability estimates for the MD index are shown in *Table 28*.

Table 28: Heritability estimates (h^2) of the mitotic delay (MD) index, determined with biometric modeling and direct variance comparison between monozygous twin pairs (MZ), dizygous twin pairs (DZ) and randomly paired controls (CP). Also shown are the significance, indicated by the probability value (P-value) and the 95 % confidence intervals (95 % CI). n: number of probands (pairs); vs.: versus.

Heritability estimate	Comparison	MD index			
		n pairs	h^2	P-value	[95 % CI][a]
biometric modeling	**MZ vs. CP**	**40 | 35**	**0.38**	**0.0048**	**[0 – 0.95]**
	MZ vs. DZ	**40 | 10**	**0.80**	**< 0.0001**	**[0.22 – 1]**
	DZ vs. CP	35 | 10	-[b]	-	-
Direct Variance Comparison	**MZ vs. CP**	**40 | 35**	**0.79**	**< 0.0001**	**[0.64 – 0.94]**
	MZ vs. DZ	**40 | 10**	**0.83**	**0.0064**	**[0.50 – 1]**
	DZ vs. CP	10 | 35	0.00	1.00	[0 – 1]

[a] Statistical test and confidence interval are based on different approaches and their results may appear sometimes inconsistent.
[b] The model could not be fitted on the observed data.

In the ACE model as well as in the direct variance comparison, the heritability estimates from MZ to DZ as well as from the MZ to CP comparisons were conclusive and revealed a high degree of heritability for the comparison of MZ to CP (h^2 ~ 0.4) and an even higher degree if MZ were compared to DZ (h^2 ~ 0.8). Each of the estimates is supported by the rejection of an absence of contributing heritable factors and by the confidence intervals. In the ACE model, the proportions of the variance components in the MZ to CP comparison were 38.2 % for the additive genetic component (A), 37.4 % for the common environment component (C) and 24.4 % for the unique environment and residual component (E). The proportions were 79.6 % (A), 0 % (C) and 20.4 % (E) for the MZ to DZ comparison.

In summary, the highest degree of heritability was estimated for the MD assay (h^2 ~ 0.8), followed by the baseline MN frequency (h^2 ~ 0.7) and the induced MN frequency (h^2 ~ 0.6), which all show a major contribution of genetic factors to the observed variability of the results. When MZ were compared to CP, both biometric modeling and the direct comparison of within-pair variances resulted in similar heritability estimates with a significant rejection of the absence of heritable components. The number of DZ pairs was generally too small to yield conclusive heritability estimates, with the exception of the MZ to DZ comparison for the MD assay.
In contrast, the induced SCE frequency only yielded a conclusive heritability estimate (h^2 ~ 0.5) for one of the approaches when MZ were compared to CP, while no conclusive estimate could be obtained for the baseline SCE frequency.

3.3.2 Correlation Between the DNA Repair Assays

The individual correlation between the respective DNA repair assay results was studied to obtain information about the relationships between the single assays if these are performed in parallel on the same set of probands. These were the baseline and the induced MN frequency (MN-b and MN-i, respectively), the baseline and the induced SCE frequency (SCE-b and SCE-i, respectively) and mitotic delay index (MD). Also the correlation of the two components of the MD, the baseline and induced ratios of cells in G_2 to cells in S phase (G_2/S-b and G_2/S-i) to each other and to the results of the other assays was studied. For an individual blood sample, all assays were performed in parallel. The probands comprised the "Twin" and "Repair Controls" cohort and a subgroup of the "Ulm" cohort from which enough blood could be obtained to perform all tests in parallel on the same samples. This was not always possible and therefore the sample number available for the respective comparisons differed between n = 177 and n = 268. Regression coefficients and their significance were calculated. As the study comprised 21 comparisons, Bonferroni correction for multiple testing was applied (threshold: P = 0.05/21 = 0.0023). This correction is a conservative approach in this context, because the comparisons are not independent.

Table 29 on *Page 86* summarizes the results of the study.

Letting aside the G_2/S ratios constituting the MD value, only the baseline and the induced SCE frequency showed a significant correlation, which was however weak ($r^2 = 0.24$). The baseline MN frequency and the induced SCE frequency exhibited a nominally significant but negligible correlation. Taken together, the results of the baseline and induced MN assays, the baseline and induced SCE assay and the MD assay are largely independent and thus contribute non-redundant information on the individual cellular DNA repair capacity.

When the two G_2/S ratio components of the mitotic delay assay are considered separately, it is obvious that both are highly significant associated with the final MD value, which depends more on the induced G_2/S ratio ($r^2 = 0.56$) than from the baseline ratio ($r^2 = 0.18$). However, G_2/S-b and G_2/S-i are also highly significantly correlated to each other, although this correlation is tiny ($r^2 = 0.07$). Furthermore, G_2/S-b shows a highly significant and G_2/S-i a nominally significant albeit very weak correlation to the induced SCE frequency ($r^2 = 0.06$ and $r^2 = 0.03$). As the baseline and the induced G_2/S ratios are not independent and are components of the mitotic delay index, they were not further considered separately.

Table 29: Correlation between the baseline and induced micronucleus frequency (MN-b and MN-i, respectively), the baseline and induced sister chromatid exchange frequency (SCE-b and SCE-i, respectively) and the mitotic delay index (MD) between the assay types on peripheral blood lymphocytes (PBL) from the same individual blood samples. The two components of the mitotic delay index, the baseline and the induced ratios of cells in the second gap and deoxyribonucleic acid (DNA) synthesis phases of the cell cycle (G_2/S-b and G_2/S-i, respectively) were also evaluated. Given are the numbers of probands, the regression coefficients and the significance of the correlation indicated by the probability value (P). n: number of probands.

End point 1	versus (vs.)	End Point 2	n	Regression Coefficient (r^2)	Significance (P-value)
MN-b	vs.	MN-i	268	0.01	P = 0.11
SCE-b	vs.	**SCE-i**	**177**	**0.24**	**P < 0.0001***
MN-b	vs.	SCE-b	183	0.001	P = 0.65
MN-b	**vs.**	**SCE-i**	**177**	**0.04**	P = 0.008
MN-b	vs.	MD	248	0.001	P = 0.86
MN-i	vs.	SCE-b	183	0.04	P = 0.4
MN-i	vs.	SCE-i	177	0.001	P = 0.65
MN-i	vs.	MD	248	0.008	P = 0.17
SCE-b	vs.	MD	187	0.01	P = 0.15
SCE-i	vs.	MD	181	0.001	P = 0.95
G_2/S-b	vs.	**G_2/S-i**	**268**	**0.07**	**P < 0.0001***
G_2/S-b	**vs.**	**MD**	**268**	**0.18**	**P < 0.0001***
G_2/S-i	**vs.**	**MD**	**268**	**0.56**	**P < 0.0001***
MN-b	vs.	G_2/S-b	248	0.01	P = 0.10
MN-b	vs.	G_2/S-i	248	0.01	P = 0.11
MN-i	vs.	G_2/S-b	248	0.001	P = 0.83
MN-i	vs.	G_2/S-i	248	0.008	P = 0.15
SCE-b	vs.	G_2/S-b	187	0.001	P = 0.64
SCE-b	vs.	G_2/S-i	187	0.02	P = 0.06
SCE-i	**vs.**	**G_2/S-b**	**181**	**0.06**	**P = 0.001***
SCE-i	**vs.**	**G_2/S-i**	**181**	**0.03**	P = 0.022

* The shown nominal P-value is smaller than the Bonferroni threshold for correction for multiple testing for 21 comparisons (threshold: p = 0.05/21 = 0.0023)

3.3.3 Correlation Between Lymphoblastoid Cell Lines and Peripheral Blood Lymphocytes

The use of peripheral blood for the assays implies difficulties regarding both the amount of available sample material as well as the time of sample arrival. As this may impair reproducibility, the applicability of lymphoblastoid cell lines (LCL) as a surrogate and as a readily available source

of material was assessed. LCL were generated by transfection of peripheral blood samples from healthy probands (n = 24) with the Epstein-Barr virus (see *2.2.2*). The baseline and the induced MN assay, the baseline and the induced SCE assay as well as the MD assay were performed both on PBL as well as on LCL originating from the same blood sample. The results were evaluated for individual correlation.

Table 30 summarizes the information about the means, ranges, coefficients of variation, and correlation of the different test results. There is a general lack of significant correlation between any assay results of LCL and PBL on an individual basis.

The mean baseline and induced MN and SCE frequency as well as the MD index are invariably lower in LCL compared to PBL, especially concerning SCE frequencies. Whereas the ranges of the individual measurements of the baseline and the induced MN frequency and the MD index overlap between the respective assays on the two cell types, this is not the case with SCE frequencies. On the other hand, the coefficient of variation is ubiquitously larger in LCL, especially in induced MN and baseline SCE frequencies.

Table 30: Comparison of the baseline and the induced micronucleus frequency (MN-b and MN-i, respectively), the baseline and the induced sister chromatid exchange frequency (SCE-b and SCE-i, respectively) and the mitotic delay index (MD) between peripheral blood lymphocytes (PBL) and lymphoblastoid cell lines (LCL) from the same blood sample of each proband. Given are the numbers of probands (n), the means, standard deviations (SD), ranges, coefficients of variants, regression coefficients (r^2) and the significance (probability values (P))

Assay	Cell Type	n	Assay Results			Individual Correlation	
			Mean ± SD	Range	Coefficient of variation	r^2	Significance
MN-b	PBL	24	50.0 ± 24.1	18-105	0.48	r^2 = 0.02	P = 0.53
	LCL	24	48.1 ± 24.8	13-98	0.52		
MN-i	PBL	24	237.8 ± 58.8	124-339	0.25	r^2 = 0.004	P = 0.76
	LCL	24	213.5 ± 88.2	98-405	0.41		
SCE-b	PBL	16	7.52 ± 0.99	6.15-9.16	0.13	r^2 = 0.06	P = 0.38
	LCL	23	4.13 ± 1.05	2.65-6.00	0.25		
SCE-i	PBL	16	13.83 ± 1.64	12.19-18.75	0.12	r^2 = 0.03	P = 0.55
	LCL	20	8.39 ± 1.52	6.70-11.90	0.18		
MD	PBL	24	3.41 ± 0.68	2.39-4.93	0.20	r^2 = 0.09	P = 0.15
	LCL	24	2.04 ± 0.50	1.19-3.64	0.25		

In LCL alone, there were a number of significant and moderate correlations between some MN and SCE assay results (MN-b vs. MN-i: r^2 = 0.41, P = 0.007; SCE-b vs. SCE-i: r^2 = 0.25, P = 0.02; MN-b vs. SCE-b: r^2 = 0.25, P = 0.02; MN-i vs. SCE-b: r^2 = 0.30, P = 0.007; MN-i vs. SCE-i: r^2 = 0.41, P = 0.003). MN-b and SCE-i were not correlated in LCL, and neither was the MD to any other assay.

3.3.4 Study of the Intrinsic Variability of the Micronucleus Assay

During the heritability study, it became evident that the intrinsic test variability of the micronucleus assay is of the same order of magnitude as the within-pair variability of monozygous twin pairs. Thus, the variability of the micronucleus frequency was studied by analyzing multiple preparations derived from one individual (see *2.3.2*). Both the baseline and the induced MN frequency were studied, each on one peripheral blood sample from the same individual. The sample was used by two investigators to set up 20 whole-blood cultures each. Per culture, two slides were prepared, and each of the slides was scored twice, resulting in 40 slides and 80 counts per a single slide set of one investigator. In the case of the baseline MN frequency, three rounds of slide preparations were performed by different preparators from the cultures of each investigator. For the induced MN frequency, one round of slide preparations was performed, also by different preparators (see *2.3.2* and *Figure 03* on *Page 26*). The MN frequency is given as counted micronuclei per 1000 scored BNC.

3.3.4.1 Variation of the Baseline Micronucleus Frequency

Properties of the dataset

A total of 1000 binucleated cells (BNC) is the recommended threshold used for a valid micronucleus frequency for one individual by the Human Micronucleus Consortium. In this analysis, more than 80 % of the 480 single counts were based on more than 1000 BNC. As an automated imaging system was used, the number of scored BNC depended on culture growth and slide density. The BNC counts were highest for the first preparation round, with nearly all counts based above 1000 BNC, and lowest for the third round with a major proportion of the counts based on counts between 600 and 1000 BNC. In total 1,140,480 BNC from one individual blood sample were scored for the presence of micronuclei in this analysis of the baseline MN frequency.

The properties of the slide sets for the baseline micronucleus frequency are presented in *Table 31* on *Page 89,* while *Figure 21* on *Page 89* provides a visualization of the results as box blots.

There were three rounds of slide preparations for this study (see *Figure 03 A*). In the first round (“original”), two slide preparators (P1 and P2) generated two sets of slides from the cell suspensions of the cultures of Investigator A (I-A) and investigator B (I-B), as described (sets I-A/P1 and I-B/P2). In the second round (“cross over”), the preparators were switched, so that each one produced the slides from the cultures of the other investigator (sets I-A/P2 and I-B/P1). In the

third round ("preparator 3"), a third preparator generated two further slide sets from the cultures of both investigators (sets I-A/P3 and I-B/P3).

Table 31: Descriptive statistics for the baseline micronucleus (MN) frequency as obtained in the six series of slides from the single blood sample. I-A and I-B denote the cultures set up by the respective investigator (A and B) and P1-P3 the preparator (1, 2 or 3) who generated the slides. Shown are the means, standard deviations (SD), standard errors (Std. Error), number of counts in the slide set, minimum (Min.) and maximum (Max.) micronucleus frequency as well as their ratio (Ratio Max/Min), variance, coefficient of variation (Coef. Var.), and sum squares for each of the slide sets.

	Baseline MN frequency									
	Mean	SD	Std. Error	Counts	Min.	Max.	Ratio Max/Min	Variance	Coef. Var.	Sum of Squares
Total	45.65	22.63	1.03	480	11.54	142.67	12.4	512.33	0.5	1245762.87
I-A/P1	51.26	9.48	1.06	80	33.62	77.8	2.3	89.94	0.19	217309.55
I-B/P2	83.85	25.75	2.88	80	36.14	142.67	3.9	663.16	0.31	614816.27
I-A/P2	39.85	8.92	1.00	80	23.04	71.37	3.1	79.51	0.22	133307.42
I-B/P1	32.21	8.23	0.92	80	11.65	52.66	4.5	67.74	0.26	88345.53
I-A/P3	34.37	9.32	1.04	80	14.12	66.06	4.7	86.85	0.27	101372.53
I-B/P3	32.38	9.25	1.03	80	11.54	58.3	5.1	85.54	0.29	90611.58

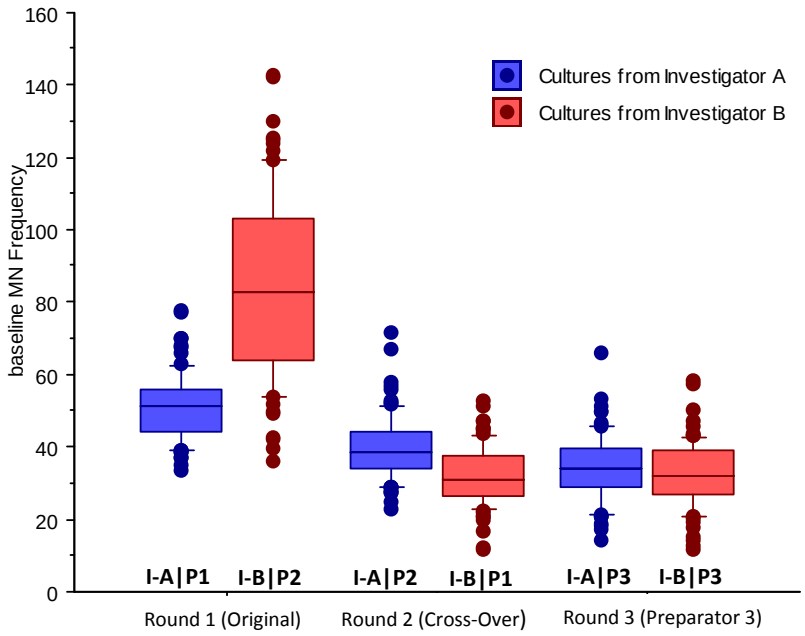

Figure 21: **Box plots of the baseline micronucleus (MN) frequency counts, divided by the six slide sets.** The plots are grouped according to the slide preparation round (1-3). I-A (blue) and I-B (red) denote the cultures set up by investigator A or B and P1, P2 and P3 the respective slide preparators.

It is obvious that the slide sets from the first round, and most predominantly the set I-B/P2, differ from the sets generated in the second and the third rounds (see *Figure 21*). Whereas the means of the slide sets generated by preparator 3 (I-A/P3 and I-B/P3) and slide set I-B/P1 showed no significant differences, the means of the other slide sets (I-A/P1, I-B/P2 and I-A/P2) differed

significantly from this group and from each other (P < 0.0001). The single micronucleus frequencies cover a wide range of values, which is however somewhat enlarged by slide set I-B/P2. Within a slide set, this amounts to a ratio of 2.1 to 5.3 between the smallest and the largest value. Letting aside the outstanding slide set I-B/P2, the slide sets show similar absolute ranges of MN frequencies, with a difference of 41 to 52 between the smallest and the largest frequencies.

To elucidate the origins of the differences between the slide sets, the analysis focused on the differences of single cultures between the preparation rounds. The means of the 40 cultures (20 cultures per investigator) showed no generalized linear shift towards a lower baseline micronucleus frequency from the first to the last round of slide preparations. The relations of the single culture means to each other within the respective preparation rounds differ between the three rounds and follow no discernible pattern. This is visualized in *Figure 22* for the means of the cultures of investigator A as determined in each preparation round.

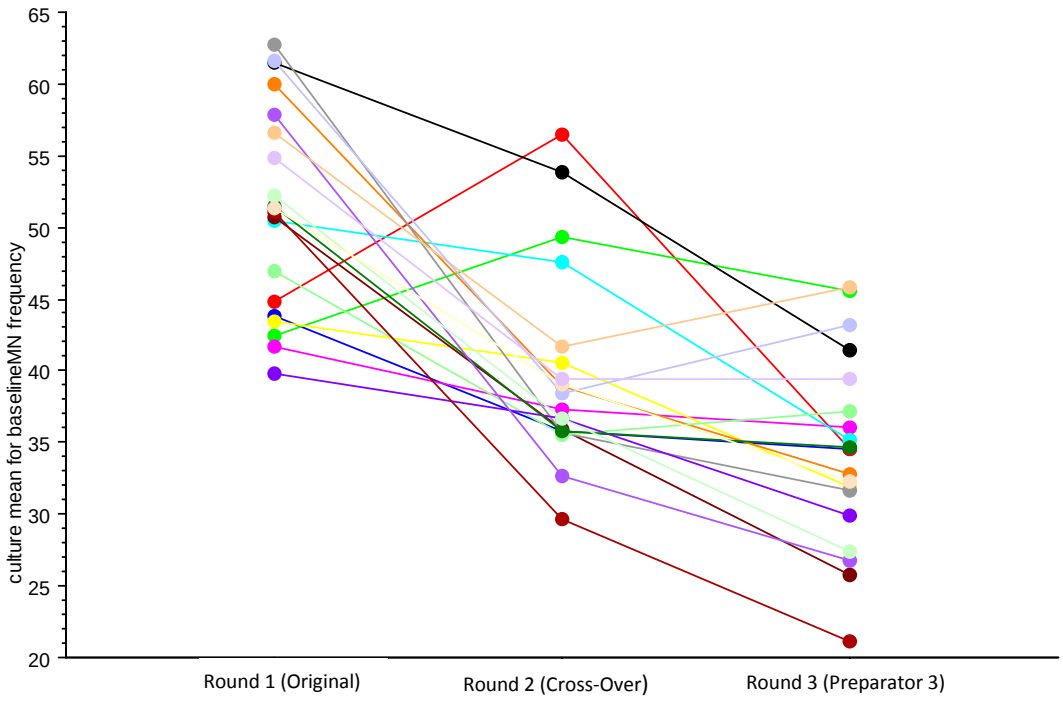

Figure 22: **Baseline micronucleus (MN) frequency means of the cultures of investigator A as determined in each preparation round.** Dots with the same colors denote the mean values of the same cultures which are connected with lines between the single preparation rounds. No clear direction or visible pattern in the shift of culture means from the first to the third round is recognizable.

Analysis of the variability factors

The results of the variability factor analyses (see *2.4.2*) of the dataset from the baseline MN frequency are presented in *Table 32* on *Page 91.*

Table 32: Variance analysis for the baseline micronucleus (MN) frequency as obtained in the six series of slides from the single blood sample. I-A and I-B denote the cultures set up by the respective investigator. P1, P2 and P3 denote the preparator who generated the slides. Variances of the respective factors as well as their proportion to the respective total variance are shown. All analyses were performed without and with correction for the number of scored binucleated cells (BNC) on which a single baseline micronucleus frequency count is based.

	Baseline MN frequency					
		Variance				
	total	**slide set**	**culture**	**slide**	**count**	**residual**
All measurements	*582.33 (100 %)*	396 (68.1 %)	115 (19.8 %)	57 (9.7 %)	0.8 (0.2 %)	12.8 (2.2 %)
All measurements BNC-corrected	*531.68 (100 %)*	399 (75.0 %)	77 (14.4 %)	43 (8.0 %)	0.9 (0.2 %)	12.7 (2.4 %)
All measurements except I-B/P2	*146.0 (100 %)*	61.7 (42.3 %)	17.4 (11.9 %)	52.1 (35.7 %)	1.0 (0.7 %)	13.8 (9.4 %)
All measures except I-B/P2 corrected for BNC	*130.8 (100 %)*	62.2 (47.6 %)	15.1 (11.5 %)	39.0 (29.8 %)	1.0 (0.8 %)	13.5 (10.3 %)
I-A/P1	*90.9 (100 %)*	-	13.9 (15.3 %)	73.3 (80.6 %)	0.0	3.7 (4.1 %)
I-A/P1 BNC corrected	*62.5 (100 %)*	-	13.2 (21.1 %)	46.1 (73.8 %)	0.0	3.2 (5.1 %)
I-B/P2	*694.4 (100 %)*	-	606.0 (87.3 %)	80.9 (11.6 %)	0.0	7.5 (1.1 %)
I-B/P2 BNC corrected	*455.4 (100 %)*	-	384.0 (84.3 %)	62.5 (13.7 %)	0.1 (0.02 %)	8.8 (1.9 %)
I-A/P2	*85.2 (100 %)*	-	24.2 (28.4 %)	37.9 (44.5 %)	3.8 (4.5 %)	19.3 (22.6 %)
I-A/P2 BNC corrected	*76.3 (100 %)*	-	26.4 (34.6 %)	27.3 (35.8 %)	3.1 (4.1 %)	19.5 (25.5%)
I-B/P1	*65.7 (100 %)*	-	21.5 (32.7 %)	39.2 (59.7)	0.2 (0.3 %)	4.8 (7.3 %)
I-B/P1 BNC corrected	*25.4 (100 %)*	-	4.9 (19.3 %)	15.8 (62.2 %)	0.3 (1.2 %)	4.4 (17.3 %)
I-A/P3	*89.0 (100 %)*	-	3.6 (4.0 %)	68.1 (76.5 %)	2.1 (2.4 %)	15.2 (17.1 %)
I-A/P3 BNC corrected	*88.3 (100 %)*	-	9.3 (10.5 %)	61.7 (69.9 %)	2.0 (2.3 %)	15.3 (17.3 %)
I-B/P3	*71.3 (100 %)*	-	23.5 (33.0 %)	42.6 (59.7 %)	0.6 (0.8 %)	4.6 (6.5 %)
I-B/P3 BNC corrected	*71.5 (100 %)*	-	21.8 (30.5 %)	44.7 (62.5 %)	0.6 (0.8 %)	4.4 (6.2 %)

Because the first round of slide preparations, especially slide set I-B/P2, showed a moderate correlation of the baseline MN frequency to the number of binucleated cells scored in the respective counts, also BNC-number corrected analyses are shown.

In the analysis combining all of the six slide sets, the main proportion of variance occurs between the six slide sets (68.1 %), also after correction for the number of scored BNC (75.0 %). The variation introduced by the differences between the cultures from a particular investigator does only weakly contribute to the overall variance (19.8 %, after correction for BNC 14.4 %). The same holds true for the contribution of duplicate slides from a particular culture (9.7 %, after correction for BNC 8.0 %). The repeated scoring of slides has a negligible influence (0.2 %), and only a very minor proportion of the total variance could not be attributed to one of the factors (~ 2.5 %).

If the outstanding slide set I-B/P2 is excluded, the remaining variance introduced by the different slide sets (~ 45 %) is of the same proportion as variance introduced by the culture (~ 12 %) and slide (~ 33 %) factors together. Here, the correction for BNC does alter the result only slightly. As the means of the two slide sets generated by the same preparator (preparator 3) are not statistically different, the differences between the slide sets of the two investigators in the two other rounds are not attributable to general differences of the cultures between the investigators, which may have been introduced during culture setup. This complies with the expectations, because all cultures were set up in a narrow time frame of 10 minutes.

The distribution of the variance between the factors “culture” and “slide” within the single slide sets is differing, while the repeated counts have again only a very minor influence. In most of the slide sets, the proportion of variance introduced by differences between the cultures is approximately 25 %, while the difference explained by differences of the duplicate slides of each culture is ~ 65 %. In contrast, these proportions are 85 % and 12 % in the I-B/P2 set. Correcting for BNC has only a minor influence on the results.
The order in which the culture were set up by the two investigators was known, however the ordering of the results did not reveal any recognizable pattern. Thus there are no discernible factors affecting the culture setup.

3.3.4.2 Variation of the Induced Micronucleus Frequency

The variation of the induced MN frequency was assessed in a single run only. Here, more than 90 % of the 160 single counts were based on more than 1000 BNC, and in total 455,658 BNC originating from one individual blood sample were scored for the presence of micronuclei.

Table 33 and *Figure 23* present the results of the two slide sets from the cultures of the two investigators.

Table 33: Descriptive statistics for the induced micronucleus (MN) frequency as obtained in the two series of slides from the single blood sample. I-A and I-B denote the cultures set up by the respective investigator (A or B). Shown are the means, standard deviations (SD), standard errors (Std. Error), number of counts in the slide set, minimum (Min.) and maximum (Max.) micronucleus frequency as well as their ratio (Ratio Max/Min), variants, coefficient of variation (Coef. Var.) and sum squares for each of the slide sets.

	Induced MN frequency									
	Mean	SD	Std. Error	Counts	Min.	Max.	Ratio Max/Min	Variance	Coef. Var.	Sum of Squares
Total	212.04	47.22	3.73	160	92.49	390.31	4.22	2229.72	0.22	7548251.13
I-A	214.19	56.64	6.33	80	92.49	390.31	4.22	3208.10	0.26	3923741.12
I-B	209.89	35.62	3.98	80	125.34	298.04	2.38	1268.79	0.17	3624510.01

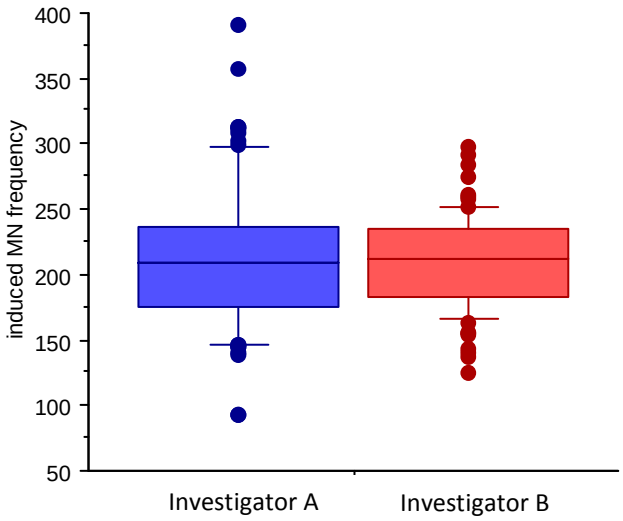

Figure 23: Box plots of the induced micronucleus (MN) frequency, divided by the two slide sets of investigator A and B.

The means of the slide sets of investigator A and B did not significantly differ (P = 0.57). However, the range of the measurements differed considerably between the two groups, with a 4.2-fold minimum to maximum ratio for the slide set of investigator A and a 2.4-fold ratio for investigator B.

The results of the analysis of the contributing variability factors in the dataset from the induced MN frequency are presented in *Table 34.* Only minor influences of the number of scored BNC, on which the single MN frequency counts are based, were present. Nevertheless the BNC-corrected results are also presented.

Table 34: Variance analysis for the induced micronucleus (MN) frequency as obtained in the six series of slides from the single blood sample. Investigator A and Investigator B denote the cultures set up by the respective investigator. Variances of the respective factors as well as their proportion to the respective total variance are shown. All analyses were performed without and with correction for the number of scored binucleated cells (BNC) on which a single induced micronucleus frequency count is based.

	Induced MN frequency					
		Variance				
	total	**slide set**	**culture**	**slide**	**count**	**residual**
All measurements	2345.5	0	1258 (53.6 %)	1055 (45.0 %)	6.5 (0.3 %)	26.0 (1.1 %)
All measurements BNC-corrected	2251.3	0	1236 (54.9 %)	991 (44.0 %)	7.2 (0.3 %)	27.1 (1.2 %)
Investigator A	3473.6	n.a.	2448 (70.5 %)	983 (28.3 %)	14.3 (0.4 %)	28.3 (0.8 %)
BNC corrected	3403.5	n.a.	2289 (67.2 %)	1065 (31.3 %)	19.1 (0.6 %)	30.4 (0.9 %)
Investigator B	1300.3	n.a.	150 (11.6 %)	1128 (86.7 %)	1.2 (0.1 %)	21.1 (1.6 %)
BNC corrected	1199.5	n.a.	261 (21.8 %)	917 (76.4 %)	0.1 (0.01 %)	21.4 (1.8 %)

In this analysis, there is no variation introduced from differences between the slide sets of the two investigators. The total variance is almost equally divided into the variance introduced by differences between the cultures of each investigator (~ 54 %) and between the slides from each culture (44 %). The repeated counting of the slides shows almost no contribution to the total variance (0.3 %), while only a very small proportion of the total variance is not attributable to one of the factors (~ 1.2 %).

The proportions of the culture and slide factors differ considerably between the slide sets of the two investigators. In the set of investigator A, approximately 68 % of the variance is attributable to the culture and ~ 30 % to the slide, while in the set of investigator B these are 16 % and 82 %, respectively.

No recognizable patterns were revealed by ordering the results according to the order at culture setup.

3.3.4.3 Minimum Necessary Number of Cultures and Slides

As there was a considerable degree of variation below the slide set or investigator level, which was mostly attributable to the culture and slide factors, the minimum number of cultures and slides that are necessary to obtain constant mean MN frequencies from one individual was evaluated. As stated, from each culture two slides were generated, each of which was scored twice, resulting in four different MN counts per culture. The means of the slide sets of investigator A and B were compared when considering only one of the counts per culture or the two counts from one of the two slides per culture. In addition, the slide sets were divided into groups of ten or five cultures, which were compared to each other.

Concerning baseline MN frequencies, the analysis was performed in the slide sets generated by preparator 3 in the third round (I-A/P3 and I-B/P3). They comprise the two most homogenous and similar slide sets in contrast to the sets generated in the first or second rounds of slide preparation, and their means do not significantly differ between the cultures of the two investigators ($P = 0.18$).

When the two slides prepared from each culture (40 counts in total) were separately compared between the two slide sets, neither the means of all counts from the first slides ($P = 0.24$) nor from the second slides ($P = 0.42$) showed significant differences between the two investigators. When only one count of one slide from each culture (20 counts) was considered for the comparison, the means of the respective counts did also not significantly differ between the cultures of the two investigators (slide1/count1: $P = 0.33$; slide1/count2: $P = 0.50$; slide2/count1: $P = 0.60$; slide2/count2: $P = 0.54$).

Next, the 20 cultures of each investigator were divided into two groups of 10 cultures, totaling in four groups (two groups per investigator) with 40 counts available for each of the groups. The means of the groups were compared to each other, and of the six possible comparisons, one resulted in a borderline significant difference ($P = 0.052$).

When four groups, each comprising five cultures (20 counts), were generated for each of the two investigators, the resulting 28 possible comparisons of the eight groups yielded one significant difference of two culture group means ($P = 0.01$). When this comparison was separately performed between the two slides of each culture, the difference could equally be attributed to

both slides, with each slide yielding two significant ($P < 0.05$) and two borderline significant ($P < 0.10$) differences between two of the culture groups.

For the induced MN frequency, the same grouping procedures were applied. When the two slides of each culture (40 counts) where considered separately, both comparisons revealed no significant differences of the means between the two investigators (slide1: $P = 0.49$; slide2: $P = 0.14$). The same holds true for the four comparisons if only one count of one slide (20 counts) was included (slide1/count1: $P = 0.77$; slide1/count2: $P = 0.55$; slide2/count1: $P = 0.29$; slide2/count2: $P = 0.32$).

When four groups comprising ten cultures (40 counts) were generated (two groups per investigator), none of the six possible comparisons between the groups resulted in a significant difference of the group means. When eight groups, each comprising five cultures (20 counts), were generated (four from Investigator A and four from Investigator B) of the 28 possible comparisons of the group means, 7 showed a significant difference ($P < 0.05$). Four of these were comparisons between two culture groups of investigator A, two between two groups of investigator B and one between one of the groups of each investigator. Four further comparisons came close to significance ($P < 0.10$). Both slides of each culture contributed equally to the differences between the cultures, with six significant differences when only slide 1 was considered and five when considering only slide 2.

Concerning both the baseline as well as the induced MN frequency, significantly different means between groups of five cultures were present, with a high probability of occurrence especially for the induced MN frequency. This cannot be attributed to the expectation of false positive results resulting from multiple testing. A borderline significant difference between the means of the baseline MN frequency was also present when groups of ten cultures were compared.

If both of the slides of one culture are separately considered, the number of significant differences is increased compared to the differences when the whole cultures are considered. Thus, in this setup, a minimum number of ten cultures with two slides each was necessary to obtain constant mean MN frequencies from one individual.

4 Discussion

The association between genetic variation and the risk of cancer is well established. For many common cancers, the known susceptibility genes and risk loci do explain only a minor part of the observed heritability. As a considerable fraction of cancer susceptibility genes is involved in the cellular DNA repair network, the overall cellular DNA repair capacity is discussed as an intermediate phenotype between cancer and the involved genetic factors. This intermediate phenotype might be of use to identify new cancer risk loci which have not been found with gene-centered or genome-wide association studies.

The purpose of the present study was to perform a characterization of the cellular DNA repair capacity in the context of breast cancer. The focus was on the question if the cellular DNA repair capacity can contribute to identify the mechanisms which underlie the associations between DNA repair genes and cancer risk. This aspect was elucidated by assessing if the cellular DNA repair capacity, coding variants in selected DNA repair genes and the occurrence of cancer can be associated with one another in one study on the same set of probands.

The assays which are used as measures of the cellular DNA repair capacity are mostly derived from mutagenicity tests applied in human biomonitoring. As such they have primarily been developed to measure the effects of environmental mutagens. The performances of these assays concerning DNA repair capacity measurement have barely been characterized, despite their frequent use in this context. Thus, particular characteristics of the applied assays were studied in addition. These included heritability estimates for each of the single assays and the study of the relationships between the different end points of DNA repair which are visualized by the assays. Furthermore, the use of lymphoblastoid cell lines as surrogates for peripheral blood lymphocytes was assessed in a small subset of probands. Because large test-related variations in the results of the micronucleus assay became evident during the course of the study, the sources of these variations were also investigated.

4.1 Association Studies

Three independent associations were studied, including the association of cellular DNA repair capacity to cancer. Several assays which are used as measures of cellular DNA repair capacity have been reported to reveal differences between cancer cases and controls. In the present study, the

baseline and the radiation induced micronucleus (MN) assay, the baseline and the radiation induced sister chromatid exchange assay (SCE) and the mitotic delay assay (MD) were used. Also other assays which were not applied in this study of the cellular DNA repair capacity have been reported to discriminate between cancer patients and controls, for example the comet assay [257] and the host-cell reactivation assay [177].
Associations between infrequent coding variants in selected DNA repair genes were the second type of association that was studied, while the third type were associations between these coding variants and the measures of cellular DNA repair capacity.

4.1.1 Associations Between Cellular DNA Repair Capacity and Sporadic Breast Cancer

During the present study, a total of five measures of the cellular DNA repair capacity were studied in sporadic breast cancer cases and age-matched controls. None of the tests allowed a prediction of the group affiliation on an individual basis, but the results of two of the assays were significantly different between cancer cases and controls. While cancer cases displayed an increased radiation induced MN frequency, the MD index was decreased compared to controls (see *3.1*). These data provide information about the mechanisms which are involved in the occurrence of sporadic breast cancer in the studied sample group.

The largest difference between sporadic breast cancer cases and controls was present in the MD index. The cases show a clearly smaller MD index, and a correct classification of cases based on this assay has a probability of ~ 75 %, as indicated by the ROC curve.
This result implies that the genetic factors which underlie the MD are involved in breast cancer risk. The reduced MD index observed in this study represents a less efficient induction of the G_2/M cell cycle checkpoint. It indicates the failure to recognize the damage or to activate the cell cycle checkpoint. Thus the cells enter mitosis without having properly repaired the DNA damage.
On the other hand, an increased MD index would suggest an impaired repair of the inflicted damage, because the cells need more time until the lesions are removed. For example, Fanconi Anemia is characterized by an accumulation of cells in the G_2 phase upon damage induction with mitomycin-C, compared to control cells. This effect is caused by defects in the inter-strand cross-link repair [155]. The underlying factors of both an increased as well as a decreased MD index lead to chromosomal instability and thus an elevated risk of cancer.
No standardized procedures exist to calculate the cell cycle delay in the G_2 phase upon damage induction. The MD index used in the present study considers only those cells which indeed enter

the cell cycle. The number of cells in the G_2 phase was standardized on the number of cells in the S phase. By calculating the ratio of the two G_2/S ratios of the untreated as well as in the damage-induced cultures, the G_2 delay is standardized on the normal cell cycle distribution without damage induction. In contrast, many other studies on the G_2 delay included the G_1 phase in their calculations of an index, which might include a large fraction of cells which cannot be stimulated to enter the cell cycle, such as granulocytes [67]. As this is not the case with the MD index used here, it can be interpreted as a comparably robust measure of the damage-induced delay of the cell cycle in the G_2 phase.

A significant reduction of the G_2 delay was seen for breast and lung cancer cases which were compared to controls [87,260]. Differences between breast cancer patients and controls were also seen in further studies [91,122], albeit cancer cases displayed an increased delay, which cannot be attributed to differences in the calculations. This is an apparent contradiction. An unlikely hypothesis might be that the underlying factors of cancer risk are different between the studied populations, and either result in a less efficient activation of the G_2/M checkpoint or in an impaired double strand break repair.

Interestingly, the association of a decreased MD with sporadic breast cancer was nearly exclusively attributable to an increased baseline G_2/S ratio in cancer cases. This accumulation of cells in G_2 phase could be due to an increased effect size of DNA damaging agents present in the environment and would thus explain the increased risk of cancer.

Also the radiation induced MN frequency of breast cancer cases was significantly increased compared to controls, albeit the difference was considerably smaller than in the MD assay. The probability of a correct classification of breast cancer cases based on the induced MN frequency was 60 %, as indicated by the ROC curve.

Thus, the cells of breast cancer cases have a higher number of unrepaired chromosome breaks at the entry into mitosis after the induction of DNA damage. This indicates an impaired DNA double strand break repair capacity in breast cancer patients, either concerning the recognition or the repair of breaks, thereby leading to an increased chromosomal instability.

An association between alterations in the radiation induced MN frequency and the risk of cancer has been observed in earlier studies. In familial or unselected breast cancer patients, an elevated radiation induced MN frequency was reported [9,13,53,233], while it was decreased in cervix cancer patients [13]. The MN frequency which was induced by nitrosamine, a tobacco carcinogen, was elevated in lung cancer patients [55].

In contrast, no differences were present in the baseline MN frequency. Other studies have reported associations of an elevated baseline MN frequency to breast cancer, as well as to other

cancers [13,55,233]. There are also negative results, for example no conclusive evidence was found for an association of altered baseline or induced MN frequencies to prostate cancer [235]. Most of these studies were designed as retrospective case-control studies, which also holds true for the present study. To date, a prospective study has been conducted only for the baseline MN frequency. This study with more than 6500 probands that were followed over a period of up to 25 years found that individuals who developed cancer had a significantly increased baseline MN frequency at the time of sampling. Besides age, no significant influences of epidemiological factors could be detected [28]. For the induced MN frequency after damage induction, such a study is yet missing to supplement the large number of retrospective studies with positive results.
The different results of the baseline and the induced MN frequency are in concordance with the additional observations that the two measures are not correlated (see *3.3.2*). On the other hand, the baseline as well as the induced MN frequency exhibited the same high heritability (see *3.3.1*). A possible interpretation is that the underlying genetic factors are different for these two measures. However, it is very unlikely that only those factors which influence the radiation induced MN frequency do also imply an increased risk of cancer. On the other hand, the activation of some factors involved in DNA repair could depend on the induction of a sufficient amount of DNA damage and would thus be less involved in the processes which are represented by the baseline MN frequency. Accordingly, many DNA repair genes have been shown to influence the damage induced MN frequency. Genetic factors affecting the baseline MN frequency could be primarily involved in the metabolism and mutagens.
A potential difference of the baseline MN frequency between cases and controls could possibly be occluded by the variation introduced by exogenic factors, e.g. in the sample handling. Due to the manifold increase in the MN counts upon damage induction, these factors have comparably less influence on the induced MN frequency. In addition, upon damage induction the DSB repair system is under full stress and the effect size of an impairment is likely to be larger than under normal conditions.

Differences between sporadic breast cancer cases and controls were neither present in the baseline nor in the BPDE induced SCE frequency. This indicates that the genetic factors involved in the outcome of these assays do not contribute to cancer risk.
An increased baseline SCE frequency compared to that of controls has been reported in breast cancer patients and their first degree relatives [43]. In contrast, another study reported an increase of the baseline SCE frequency only in breast cancer patients which had received chemotherapy [6]. This response to cancer treatment might be a confounding factor in other studies also, and is in accordance with the stronger reflection of exogenic influences, such as

mutagen exposure, compared to the MN and MD assays as indicated by the additional information from the heritability estimates.

Influences of age and gender and culture setup delay are mainly described as influences on the baseline MN frequency [20]. As influences of these factors were largely absent for all of the applied tests, they were not considered as potential confounding effects in the present study.

The results of the present study show that cancer patients and controls exhibit differences in their cellular DNA repair capacity. For such studies, appropriate tests have to be used, which do indeed represent differences in the factors underlying the occurrence of cancer. Here, the induced micronucleus assay and the mitotic delay assay were suited for these analyses. Of course, no individual prediction on cancer risk can be achieved with the tests that were used here or described in the literature. The assessment of the cellular DNA repair capacity nevertheless provides a tool which can be integrated into a more comprehensive characterization of the personal risk of cancer, e.g. in combination with the assessment of mutagenic influences. Furthermore, it can contribute to identify the molecular genetic origin of the differences between controls and cancer.

4.1.2 Coding Variants in DNA Repair Genes, Breast Cancer Risk and Cellular DNA Repair Capacity

A relationship between genetic variants and an increased risk to develop cancer has frequently been reported in studies which used gene-centered or genome-wide approaches. Also influences of genetic variation on the outcome of different measures of the cellular DNA repair capacity have been described. The present study included such association studies which focused on genes that are involved in DNA repair. The genes included in the present study were selected after extensive review of the literature according to a known or putative function in the DNA repair processes whose end points are visualized by the applied DNA repair assays. This was done in such a way that different functional systems of the DNA repair network were represented (see *3.2.1*). A further focus was set on genes which were either involved in cancer formation or if an association to the occurrence of cancer was reported.

Sporadic breast cancer cases as well as familial cases without a mutation in the high susceptibility genes *BRCA1* or *BRCA2 w*ere included in the present study. As the genetic background of these cases is largely unexplained, genes that were already well studied and characterized in the

context of cellular DNA repair capacity or cancer, for example the *BRCA* or *ATM* genes, were excluded. The selection focused genes that were only recently reported to be involved in this context. With one exception, only missense variants with a minor allele frequency (MAF) between 0.02 and 0.16 were chosen for genotyping.

A variant which reduces the efficiency of the gene product affects the DNA repair capacity and cancer risk. Common variants often mediate only a small risk increase. They do most likely not affect the protein function and probably tag the haplotypes which carry the true risk alleles. This is known from the numerous intronic, intergenic and synonymous common variants which have been found by genome-wide association studies [54]. On the other hand, the high risk in families with hereditary cancer is mostly due to the presence of rare private mutations which abrogate the function of the affected allele. Genome-wide approaches for the identification of common variants and rare mutations are established, but still only a minor fraction of the heritability of cancer is explained. In the gene-centered approach used in the present study, human genome databases were utilized as described to identify low-frequency variants in the most suitable candidate DNA repair genes. These intermediate variants can be expected to have a higher probability for a direct functional effect than common variants. If the sample size is sufficiently large, they also allow successful associations because of their higher frequency compared to rare private variants.

A total of 22 variants in 14 genes could successfully be genotyped. The minor allele of the variant SNP 16, which is rs35123152 in *WRAP53* gene, was not present in the genotyped population although an allele frequency of 0.09 was expected. This failure was most probably due to a defective genotyping assay which was pre-designed by the manufacturer.

One genotyped variant in the *SLX4* gene (SNP 09) was significantly associated to an elevated incidence of breast cancer on the one side, as well as to an increased radiation-induced micronucleus frequency on the other side (see *3.2.3* and *3.2.4*). A variant in the *NINL* gene (SNP 04) displayed a significant association to an elevated mitotic delay index. Notably, both of these associated DNA repair assays were also associated with breast cancer, underlining the general concordance of the results obtained with the different association studies.

The genotyped variant in *BRIP1* (SNP 21) showed a borderline association to breast cancer. None of the further genotyped variants in the other genes showed an association to either the increased occurrence of breast cancer or the altered outcome of DNA repair assays. The baseline MN frequency as well as the baseline and the induced SCE frequency were not associated to any of the genotyped variants. Isolated studies on the associations of these assays to genetic variants have been performed, yielding both positive and negative results [115,128,149].

Associations of *SLX4* to cancer risk and increased radiation induced micronucleus frequency

Association to cancer risk

The minor allele of the variant SNP 09, which is rs3810813 located in the *SLX4* gene, was clearly associated with an increased risk of breast cancer, with an odds ratio of ~ 1.6. Due to the frequency of the associated allele of 0.05, the corresponding population attributable risk is 5.6 %. The association was above the significance threshold after correction for the multiple testing of the 22 variants in three age groups (66 tests).

When the analysis was stratified for young breast cancer cases aged ≤ 60 years and ≤ 40 years, the association was highly significant. The odds ratios increased with younger age of the cases, up to ~ 4.2 in breast cancer cases ≤ 40 years, albeit only 23 cases were present in this young age group. The accumulation of the minor allele in young cases indicates that this variant in the *SLX4* gene is a risk locus for early-onset breast cancer. Due to the comparably high allele frequency, it would explain a considerable portion of these young cases, with an estimated population attributable risk of > 25 %.

The same holds true when *BRCA*-negative familial and sporadic breast cancer cases are considered separately. The association was highly significant in the familial cases. This suggests that the *SLX4* gene could play a role in families with hereditary breast cancer, which is supported by the increased frequency of variant carriers in young familial cases. Nevertheless, this association was also seen in an independent group of sporadic breast cancer cases and the increase of the frequency of variant carriers in younger cases is comparable in both groups. These data establish this variant as cancer risk locus.

In index cases from *BRCA*-positive familial breast cancer families, no increase in the minor allele frequency was present. As mutations in the *BRCA* genes confer a high susceptibility to breast cancer, they are most likely causative for the disease in those individuals and thus no overrepresentation of another potential susceptibility allele was expected in this group.

Due to the lack of large families, co-segregation of the variant with the occurrence of cancer cannot be evaluated.

Haplotype analyses of *SLX4* revealed that the minor allele of the variant rs3810813 is in complete linkage disequilibrium with the other two genotyped variants. The haplotype which carries the minor allele of SNP 09 exclusively carries the major alleles of the SNPs 10 and 11 and is also significantly associated to the increased risk of breast cancer. Thus, the borderline association of the major allele of SNP 10 (rs3827530) to increased breast cancer risk that was seen in cases ≤ 60 does tag the true risk locus which is much better represented by the minor allele of SNP 09. As the

carriers of this allele only comprise a small minority of the carriers of the major allele of SNP 10, the association is much weaker for the latter variant.

Analyses with the HapMap database revealed that the haplotype which his tagged by the minor allele of SNP 09 exceeds the gene sequence in both directions and none of the genotyped polymorphisms did split this haplotype further. Therefore, no other variant with a possibly stronger association to cancer risk can be defined from these data. In turn, this would mean that the true risk locus could be situated anywhere in the gene. The possibility, that the variant has nevertheless itself an effect on the gene product and thus mediates the elevated occurrence of cancer, is discussed below.

The *SLX4* gene is required for the repair of DNA cross-links and double strand breaks and has a wide range of coordinative activities in the cellular DNA repair network. It is a member of a complex which is able to resolve Holliday Junctions at homologous recombination repair sites. Furthermore, it serves as a scaffold for the binding of several structure-specific endonucleases involved in HR or cross-link repair, and also binds to the mismatch repair and telomere binding complexes. Cells in which the *SLX4* expression is abolished display an increased sensitivity to cross-linking or adductive agents [58,218]. Thus, an impaired or less effective gene product most likely influences the cellular DNA repair system. This influence would manifest in increased numbers of not or improperly repaired DNA double strand breaks. This would either be due to the failure to resolve Holliday Junctions or as an indirect effect of stalled replication forks as a result of an impaired cross-link repair. With respect to these potential tumor suppressor properties, the involvement of *SLX4* in breast cancer susceptibility is highly probable, although no studies have yet been published concerning cancer.

The results of the present study were unexpectedly supported by the very recent findings that biallelic mutations in *SLX4* cause Fanconi Anemia [114,213]. Thus, *SLX4* is also a member of the Fanconi Anemia pathway. These studies confirm that this pathway and homologous recombination system form a combined network that functions in the cellular double strand break and cross-link repair. *SLX4* has not been associated to cancer before, and the results presented here provide the first and conclusive evidence that this gene is, a susceptibility gene for breast cancer, like many other *FANC* genes [155].

Association of rs3810813 to DNA repair capacity

Also the third type of association was found, namely between this variant in *SLX4* and an increased radiation induced MN frequency (see *3.2.4*). The same variant is associated with an

increased incidence of breast cancer. Furthermore, also an elevated induced MN frequency is associated with breast cancer.

Thus, carriers of this variant display an impaired double strand break repair upon damage induction, which can directly explain the increased cancer incidence. In turn, this could best be explained by assuming that the associated variant has functional consequences for the resulting gene product.

There is only one observation that cannot fully be fitted into this framework. The mean MN frequencies of carriers and non-carriers of the variant do not differ significantly in controls. One interpretation might be that the genotyped variant does tag the causative locus which is much rarer than the minor allele of SNP 09, and that all carriers of this causative allele are affected. Hence, the functional effect of the allele could be detected only in cancer cases. An effect of the genotyped variant itself might nevertheless be present, as the carriers are infrequent among individuals with a low induced MN frequency in controls. This issue cannot be resolved at the present state. Further studies are needed which may include additional functional tests on a cellular level and an extensive search for individual mutations on the risk haplotype.

Another hint for a potential functional effect of the variants came from the fact that four of the ten dizygous twin pairs were discordant for the presence of the minor allele of SNP 09, although this is a very small sample (see *3.2.5*). In a pair wise comparison, the results of the induced MN frequency and the induced SCE frequency were clearly increased in carriers of the variant, compared to non-carriers, however due to the small sample size no significant difference was yielded. It remains speculative if these differences display a true effect of the presence of the minor allele on the cellular DNA repair capacity.

The damage induced MN frequency has also previously been associated to variations in DNA repair genes, which does support the results of the present study. It is increased in heterozygous carriers of mutations in the *ATM* gene [79]. The repression of genes involved in various DNA repair processes leads to increased radiation induced MN frequencies [259]. Also heterozygous carriers of *BRCA1* and *BRCA2* mutations exhibit an increased induced MN frequency [8,225]. Conflicting findings concerning the association to *BRCA* mutations exist, these could however be due to experimental differences in the MN assay and do not support a negative result [118,239]. For chemically induced MN frequencies, associations to polymorphisms in genes involved in the metabolism of chemical mutagens have also been shown [115,130,141].

A coding variant in *SLX4* could be associated with an impaired cellular DNA repair capacity, which was visualized by an increased induced MN frequency. This elevated frequency was also associated to an elevated occurrence of sporadic breast cancer. Whereas *SLX4* is involved in the

double strand break repair via the homologous recombination pathway, the MN assay measures the failure of double strand break repair. This general concordance of the results leads to the conclusion that the assessment of the cellular DNA repair capacity can identify on the one hand individuals with an increased susceptibility to cancer and can contribute on the other hand to the identification of the underlying genetic variations.

Possible functional impact of rs3810813 on SLX4

Prompted by the combined association to cancer and to an increased radiation induced MN frequency, a characterization of the possible functional impact of the minor allele of SNP 09 on the Slx4 protein was performed *in silico*. The aim was to evaluate if the variant itself may be responsible for the associations or if it tags a true risk locus elsewhere in the gene.

The variant results in an amino acid change from serine to phenylalanine on the residue position 1271 in the protein. This residue is not located in one of the known functional domains. It is situated in a region which is necessary for the interaction with the telomere binding complex Terf2-Terf2IP, suggesting an involvement of *SLX4* in telomere maintenance [26]. The region also mediates the binding to the protein kinase Plk1, which most likely phosphorylates Slx4 and thus regulates its activity [218]. Furthermore, the residue is located on the border of a region necessary for the interaction with the structure-specific heterodimeric endonuclease Eme1/Mus81. This complex is required for the generation of double strand breaks at stalled recombination forks and for their subsequent repair via homologous recombination with the sister chromatid [185].

The serine residue is conserved in rodents and yeast, suggesting the residue could be of relevance for the protein function. It is followed by a glutamine residue which is also conserved in mammals and in the yeast *Saccharomcyes cerevisiae*. Serine-Glutamine (SQ) motifs are common phosphorylation sites, and the residue has been characterized as a potential phosphorylation substrate for *ATM* or *ATR* in *S. cerevisiae*. Thus, the serine residue may be necessary for the regulation of the protein activity by these kinases. A residue change to phenylalanine would disrupt the phosphorylation site and may inhibit the regulation of the Slx4 activity. On the other hand, the Glutamine is not conserved in the yeast *Schizosaccharomyces pombe*. In addition, the size of the protein is different in *S. pompe* (419 amino acids (aa)) and *S. cerevisiae* (748 aa), compared to the human protein (1834 aa). This indicates that the human protein features domains and structures that are not present in yeast, which could result in erroneous alignments of the protein sequences. Furthermore, the serine residue is not conserved in *Gallus gallus* (chicken) and *Drosophila melanogaster* (fruit fly). This suggests that the motif may be functional only in mammals.

The residue change from Serine to Phenylalanine has probably a damaging effect on the structure of the human Slx4 protein, as determined with the Polyphen 2 program. The probability is 97.5 %, with a false positive estimate of 6 % and was determined with the model which is suited to evaluate the impact of variant alleles involved in complex diseases, such as cancer. Thus, besides the possible function as a phosphorylation target, the minor allele of the variant probably results in structural alterations which might impede the interaction with some of the binding partners of Slx4.

These data suggest a high probability that the variant is the true effect locus. The Slx4 protein is an integral part of the repair processes which are induced by inter-strand cross-links and double strand breaks. The increased risk of cancer in carriers of the minor allele of the variant rs3810813 of *SLX4* as well as the reduced capacity for the repair of DNA double strand beaks indicate an impaired function of the variant Slx4 protein. Therefore, the possible impact of the residue change on the structure and a potential disruption of a phosphorylation site both support the interpretation, that the genotyped variant could itself be the cause of this reduced efficiency.

NINL and increased mitotic delay

The *NINL* gene yielded a positive association result, while no associations were observed in the other genes which are involved the G_2/M cycle checkpoint. *NINL* has not yet been associated to cancer risk, and also the two rare variants included in the present study were not associated to breast cancer susceptibility.

In the present study, carriers of the minor allele of one of the variants, SNP 04 which is rs13044759 exhibited a significantly increased mitotic delay index (see *3.2.4*). The increase of the MD index suggests that the variant directly influences the gene function regardless of the status as control or sporadic breast cancer case. Breast cancer cases had shown a significantly reduced mitotic delay compared to controls, indicating a less effective induction of the cell cycle checkpoint (see *4.1.1*). Therefore this increased delay means a more efficient induction and thus a protective effect. As carriers of the variant do not show a decreased occurrence of cancer, the size of the effect is probably small.

The variant results in a change of the amino acid residue from arginine to tryptophan on position 275 in the Nlp protein. Nlp is a direct interaction partner of Brca1 and this interaction is necessary for the localization of the gene product to the centrosome. The loss of *NINL* function results in an impaired cell cycle arrest before the entry into mitosis, leading to a defective segregation of chromosomes and to genetic instability [103]. On the other hand, the gene product is oncogenic

and causes the spontaneous occurrence of breast, ovary and testicle tumors when overexpressed in mice [199]. Furthermore it is overexpressed in human ovarian tumors [173]. This indicates that the adherence to a correct dose of the *NINL* gene product is important in order to prevent negative effects on the cell.

An influence of variation in DNA repair genes on the G_2 delay has been described before. Functional variants in two genes involved in the base excision repair, *APE* and *XRCC1*, have been associated to a prolonged G_2 delay upon exposure to ionizing irradiation [92]. As the MD assay is largely unaffected by age or environmental factors, genetic variations are likely to play a major role in the observed variability.

In the present study, a coding variant within the *NINL* gene was associated with differences in the mitotic delay assay. This assay also revealed differences between sporadic breast cancer cases and controls. In contrast to *SLX4*, the minor allele of the variant was not itself associated with breast cancer. But while *NINL* contributes to the activation of the G_2/M cell cycle checkpoint, the MD index reflects the G_2 delay.

The variant in *SLX4* was associated to breast cancer as well as to an increased induced MN frequency, and the latter was also associated to breast cancer. The additional concordance in the *NINL* gene substantiates the conclusion the cellular DNA repair capacity can be utilized in the identification of the genetic variations that underlie cancer risk.

BRIP1 and cancer susceptibility

As stated, many of the genes of the Fanconi Anemia pathway have also been associated to an increased risk of breast cancer or other cancers. Three genes which were previously known to be involved in this pathway were included in the present study, namely *BRIP1*, *C17ORF70* and *C19ORF40*. Of these genes, only SNP 21 in *BRIP1* displayed a very borderline association in breast cancer cases aged 40 years or younger (see *3.2.3*). The odds ratio of ~ 0.5 indicated a reduced cancer incidence in carriers of the minor allele. It is the only common variant included in the present study.

The variant allele of SNP 21, which is rs4986764, leads to a change of the amino acid from proline to serine at residue 919 of the Brip1 protein. The direct interaction of Brip1 with Brca1 is necessary for the regulation of the DNA repair pathway choice [251].

SNP 21 is also the only variant included in the study for which an association to breast cancer had previously been published. While one study reported an increased occurrence of breast cancer in cases up to 50 years of age [203], other studies could not confirm this association [69,231]. The

data from the literature suggest that the variant only tags one or several risk loci in the *BRIP1* gene, which may not be present in particular population. Other variations and truncating mutations in *BRIP1* have been associated to breast and ovarian cancer [197,204], but the associations were marginal. Recently, a recurrent truncating mutation in *BRIP1* was also associated with prostate cancer susceptibility [117].
In addition, no differences between carriers and non-carriers of the variant were seen with any of the applied DNA repair capacity measures. This contributes to the notion that the studied variant in the *BRIP1* gene does not affect the cellular DNA repair capacity.

In summary, it cannot be excluded that the association in young breast cancer cases tags a true but rare risk locus which is situated on the same haplotype with the common major allele of the genotyped variant.

Negative results

Rare variants in two other studied genes which are involved in the Fanconi Anemia pathway, *C17ORF70* and *C19ORF40*, were not associated to breast cancer or to differences in the applied DNA repair assays. They are required for the activation of the FA pathway upon cross-link damage [93,135], and *C19ORF40* is also involved in a cell cycle checkpoint activation which is independent from the FA pathway and mediated via *ATR* [44]. Defects of the gene products might result in alterations of the mitotic delay or in an increased number of stalled replication forks due to unrepaired cross-link damage. During their resolution, double strand breaks are generated, which could be visualized in the MN assays and could also manifest in an elevated SCE frequency.
Also the studied variants in the other genes which are involved in the G_2/M checkpoint did not show positive association results. *NEK1* is essential for the activation of the cell cycle arrest after damage induction by ionizing irradiation. [170]. *TIMELESS* is involved in the maintenance of genome stability during replication as it stabilizes replication forks and is involved in the cohesion of sister chromatids. [129]. The gene *FBXO5* thus contributes to the maintenance of the cell cycle arrest in the G_2 phase [126]. Functional effects of the genotyped variants in those genes would also be expected to manifest in differences of the cell cycle delay.
Besides *SLX4*, one further gene involved in the homologous recombination pathway was studied. *BARD1* directly interacts with *BRCA1*, and is involved in many aspects of the cellular DNA repair processes [78]. Common *BARD1* variants have been associated with an increased risk of breast cancer [109,210], however in numerous other studies as well as in the present study these associations could not be replicated [101,230]. Since the positive results came mostly from

studies which included Northern European populations, these findings suggest a population-specific contribution to cancer risk, in which the common polymorphisms might tag risk alleles which are only present in particular populations. Recently, truncating and missense mutations have been found in a minority of hereditary breast cancer families without mutations in the *BRCA* genes [49], underlining the general involvement of *BARD1* in cancer susceptibility.

In some of the studied functional groups of DNA repair genes, no member showed an association of coding missense variants to breast cancer risk or to the results of the applied DNA repair assays. This indicates that either the respective pathways do not contribute to the development of breast cancer, or that the missense variants have no effect on the protein function. Further, they do not tag a causative locus elsewhere in these genes.

These included the genes involved in the maintenance of telomere ends. While *TERF2IP* is a member of the telomere binding complex, *WRAP53* is involved in the maturation of the telomerase RNA [26,143]. Deficits of the gene products might lead to repair processes at the telomere ends which could induce translocations.

Also polymorphisms of genes involved in the fidelity of the DNA replication were not associated to an increased risk of breast cancer and showed no influence on any of the DNA repair assays. *REV1* facilitates the bypass of inter-strand cross-links and abasic sites and has been associated to cervix cancer [80]. *POLI* is involved in the translesion synthesis via an alternative base-pairing mechanism and is associated to an increased susceptibility of tumor formation in mouse models for lung and colorectal cancer [125,211]. The transcription factor *HTLF* is involved in the regression and restart of replication forks at DNA damage sites [23]. *ESCO1* is required for sister chromatid cohesion during the S phase and also upon DNA double strand break damage in the G_2 and M phases [228,258]. Recently, the studied variants in *POLI* and *ESCO1* were also associated with the presence of the *TMPRSS2-ERG* fusion gene which is commonly found in human prostate carcinomas [138]. A reduced efficiency of these gene products would result in an increased number of stalled replication forks or impair the error-free homologous recombination pathway, thus leading to an increased number of not or improperly repaired DSB.

4.2 Characteristics of DNA Repair Capacity Measures

Another aspect of the present study was the analysis of several characteristics of the applied DNA repair assays. These assays are derived from mutagenicity assays, but despite their frequent use in studies on DNA repair, their performance as measures of the cellular DNA repair capacity has not been well characterized. The degree of influence of genetic factors on the DNA repair assay results was analyzed by estimating their heritability in a twin study. The question if the applied DNA repair assays represent independent or redundant end points of cellular DNA repair processes was assessed by studying the correlation of the results of the different assays, which were performed in parallel on the same probands. Also the use of lymphoblastoid cell lines as surrogates for peripheral blood lymphocytes was evaluated by comparing the assay results on both types of cells in a small subset of the probands. The large test-related variability of the MN assay impairs its reproducibility, which became evident also during the course of the present study. The factors which possibly contribute to this variability were analyzed to elicit its origin.

4.2.1 Heritability of Cellular DNA Repair Capacity

The assays which were applied as measures of the individual DNA repair capacity in the present study show a considerable inter-individual variation. This can either be caused by genetic factors or by environmental and assay-related influences. The applied assays are derived from tests used in the context of mutagenicity studies and some are known to be influenced by exogenic factors [60].

To elucidate to which degree the observed variations originate from differences in the individual genetic background, the heritability of the baseline and the radiation induced MN frequency, of the baseline and the BPDE induced SCE frequency, and of the MD index were estimated by a twin study. The variation of the respective assay results of monozygous twin pairs (MZ), dizygous twin pairs (DZ), and randomly paired unrelated control individuals (CP) were compared. Whereas MZ have an identical genotype, DZ have half of the genetic variations in common, and CP share no genetic variations. These differences are utilized in heritability estimates to assess the influence of genetic factors on the observed variation of the assay results. Two approaches were used here, the biometric modeling of variance components (ACE model) and the direct comparison of within-pair variances. Both approaches yielded similar heritability estimates for most of the respective assay results.

Heritability estimates

High estimates of heritability (h^2 ~ 0.6 – 0.8) were obtained for the baseline and the radiation induced MN frequency as well as for the mitotic delay (see *3.3.1*). The absence of heritable factors as contributors to the observed variability could be excluded. These results indicate that genetic factors have a strong influence on the observed inter-individual variability of these assays and contribute to a major part to their outcome.

In the ACE model, the residual variance component comprises the variability introduced by unique environmental and test-related factors. This component was nearly identical in both types of MN frequency. This result is surprising, because due to the induction of the majority of MN by the irradiation the induced MN frequency is thought to be less influenced by environmental factors. Possibly, this reduction of variability comes along with an increased assay variation caused by the exposure to ionizing irradiation, which could originate from potential inconsistencies in the exact irradiation dose.

In contrast, heritability estimates of the baseline SCE and the induced SCE frequency do not support a major contribution of genetic factors to the observed variability. The induced SCE frequency shows a comparably higher heritability estimate, which is in line with the expectation of a reduced environmental influence compared to the baseline frequency.

Intraclass correlation coefficients (ICC) are used as an alternative measure for the influence of genetic factors, and can also be calculated on data from twin pairs. The results were comparable to the heritability estimates. In the MZ group, strong correlations ($r > 0.6$) were present for the baseline and the induced MN frequency as well as for the MD index. On the other hand the correlations were weak (≤ 0.5) for the baseline and the induced SCE frequency. ICC were zero or close to zero for all assays in CP pairs. In DZ pairs, they were either zero or smaller than those of MZ pairs. Thus, also the ICC indicate a major genetic influence on the baseline and the induced MN frequency as well as on the MD index, but not on the baseline or the induced SCE frequency.

The effects of age, gender and delay of the culture setup on the results of the respective assays were in some instances statistically significant within the MZ, DZ or CP sample groups. However they were weak and thus negligible. In addition, these influences are not relevant for the heritability estimates because they are identical within each pair. The gender distribution of CP closely represents that of MZ, and CP are of similar age as MZ pairs. Nevertheless, the analyses were adjusted for these factors, yielding only slightly different results, as expected.

There were slight differences between the mean assay results between the three sample groups, but the standard deviations were always of the same orders of magnitude. These differences

might originate in the slight variations in age or gender distribution between the groups. As only within-pair variances are compared between the groups, the heritability estimates do not depend on the group means.

Properties of the MZ to DZ and MZ to CP comparisons

A problem posed the unexpectedly low number of DZ pairs that could be recruited. This circumstance led to inconclusive results when MZ were compared to DZ, and the DZ group is likely to be subject to random effects, for example the matching of assay results in a DZ pair by chance. Interestingly, three out of the self-reported thirteen DZ pairs turned out to be in fact monozygous after the molecular analysis and had to be re-categorized. This wrong pair classification of over 20 % of all recruited DZ pairs based on the phenotypic appearance emphasizes the importance of a molecular genetic verification of the twin status in such studies. The MZ to DZ comparison is an underestimation of heritability, because it does not comprise the full variability that is present in the population. Furthermore, the low number of DZ pairs could have reduced the overall within-pair variance. The heritability estimates would also yield low and inconclusive results if the within-pair variability of the MZ or the DZ group is of the same order of magnitude as the intrinsic test variability of the respective assay.

Therefore CP, who do not share a common background of genetic variation, were also included. The MZ to CP comparison could result in an overestimation of the true heritability. The contribution of the common environment shared by the two individuals of a pair cannot be clearly distinguished from the genetic influence in this analysis. However, the procedures to estimate heritability from twin studies were originally designed in the context of studies on psychological or behavioral traits, in which a common childhood environment might possess an influence [29]. In contrast, the observed assay results are determined by biological factors and environmental influences which act on peripheral blood lymphocytes. Thus, environmental differences that occurred a long time ago during childhood could hardly contribute to the observed differences in the results of the respective assays. Furthermore, in the case of the damage induced MN and SCE frequency as well as in the MD assay, the damage is inflicted *in vitro* after the blood sampling. Therefore, a common environment is expected to only negligibly reduce the total variance if the two individuals of a pair do not presently share the same living environment. Nearly all of the twin pairs included in this study reported themselves to live separately. These expectations are in close accordance with the observations. The biometric modeling of the variance components (ACE model) revealed that an absent or an only very minor influence of the common environment component fits best on the observed data structure of the baseline and induced MN frequency. A low degree of influence of the common environment fits best on the data structure of the MD

assay. Thus, these heritability estimates from the MZ to CP comparison are in this context resistant to an overestimation.

Relations to heritability estimates of other traits

The heritability estimates for the baseline and the induced MN frequency and for the MD index were higher than the known heritability of most cancers, for example prostate cancer (h^2 ~ 0.42), colorectal cancer (h^2 ~ 0.35) and breast cancer (h^2 ~ 0.27) [132]. They are also higher than the estimates for many of the less frequent cancers [48]. This increased heritability is to be expected if the applied measures of cellular DNA repair capacity are indeed intermediate phenotypes between a complex disease, in this case cancer, and the underlying genetic factors. This is due to the fact that they are closer related to the involved genetic factors than the cancer disease itself.

The heritability of aspects of the cellular DNA repair has been estimated in other studies. These produced results which were compatible to those of the present study. A twin study on the baseline and a chemically induced MN frequency reported similar estimates as the present study, although the sample size was comparably small [223]. Some twin studies with small sample sizes reported a lack of genetic influence on the baseline or the chemically induced SCE frequency [5,82], while an estimate of h^2 ~ 0.3 was obtained for the baseline SCE frequency in a large study with 188 MZ and 88 DZ pairs [90]. A recent twin study on radiation induced cell cycle alterations showed a strong correlation in MZ pairs and reported a conclusive heritability estimate of h^2 ~ 0.7 [67].

In some twin studies, other end points of cellular DNA repair processes were analyzed. These included the number of chromosome aberrations in metaphases after the treatment with different mutagens, as well as apoptotic response assays, A major contribution of genetic factors to the observed variability was observed in each case [37,67,249]. The same holds true for the number of chromosome and chromatid breaks after damage induction with ionizing radiation [30]. The comparably high heritability of chromosomal damage after irradiation was also seen earlier in a study with families of breast cancer patients [182].

The heritability estimates which were obtained in the present and other studies suggest that the cellular DNA repair capacity, if measured with suitable assays, does reflect the genetic background to a major part. It can be interpreted as an intermediate phenotype between DNA repair and cancer. Thus, an impaired DNA repair capacity may be useful as a predictor of cancer risk and furthermore for the identification of the DNA repair gene variants which underlie the risk increase.

4.2.2 Correlations Between the Measured DNA Repair End Points

The study of correlations between the applied DNA repair assays can provide an insight to which degree the single assays measure the end points of different or redundant DNA repair processes. This kind of information is important to evaluate the combined use of these assays, because a comprehensive characterization of the cellular DNA repair capacity requires the assessment of independent end points. In the present study, the baseline and the radiation induced MN frequency, the baseline and the BPDE induced SCE frequency, and the MD index were evaluated as measures of cellular DNA repair capacity. These and further measures are applied in many studies as indicators of chromosomal damage or DNA repair capacity, and two or more of them are often used in parallel on the same set of probands. In these studies, the results were correlated to numerous quantitative genetic and exogenic variables. However, data on the correlation between the results of the applied assays on an individual basis are rarely presented, and often only correlations between the means of different subgroups are shown.

In the present study the correlation between the results of different assays that were performed on the same control probands was studied on an individual basis. Generally, no or negligible correlations were present between the different assays (see *3.3.2*). A weak correlation was present between the baseline and the BPDE induced SCE frequency. A correlation between the baseline and a diepoxybutane-induced SCE frequency was noticed before, with a similar strength [14]. The heritability estimates indicate that both types of SCE frequency are primarily influenced by exogenic factors. The weak correlation indicates that at least some of these factors influence both the baseline as well as the induced SCE frequency. Because the correlation was small, the baseline and the induced SCE frequency can still be regarded as largely independent.

Interestingly, the baseline and the radiation induced MN frequency showed also no correlation, although both types of MN frequency exhibited a similar high degree of heritability. This suggests that the involved genetic factors are different for the two types of MN frequency. The irradiation directly induces DNA double strand breaks after the set up of the culture, thus results of the induced MN assay primarily depend on the genetic factors involved in the removal of these lesions. In contrast, the lesions visualized by the baseline MN frequency originate from intrinsic factors or from exposure to mutagens. Accordingly, the baseline MN frequency is also influenced by genes involved in mutagen sensitivity [94]. The induction of double strand breaks by exposure to ionizing radiation might on the other hand lead to an increased variability of the MN assay which could also explain the missing correlation.

In the context of further, isolated studies, some of the applied measures have been correlated to each other and to the results of other DNA repair or mutagenicity assays. In accordance with the present study, no or only weak correlations were reported [15,140,174]. This contributes to the picture of a general lack of correlation between the measured end points of cellular DNA repair processes.

The mitotic delay index was defined here by the two G_2/S ratios of the cultures with and without damage induction by ionizing irradiation. Of course, both G_2/S ratios exhibited highly significant and comparably strong correlations to the final MD value. Both G_2/S ratios also showed a significant but weak correlation to each other, indicating that at least some of the factors which influence the cell cycle have an effect both in untreated cells and upon the induction of damage. There were no relevant correlations to any of the other assays.

In summary, the applied measures of DNA repair capacity are not or only very weakly correlated to each other as determined in parallel on the same probands. Thus, they reflect largely different end points of the cellular DNA repair network and cannot substitute each other. They complement each other and can be used to independently characterize different aspects of the cellular DNA repair capacity.

4.2.3 Lymphoblastoid Cell Lines as Surrogates for Peripheral Blood Samples

The DNA repair assays used here were performed on fresh peripheral blood samples of the probands, which are limited both in their temporal and quantitative availability. If experiments have to be repeated in the context of long time studies or due to experimental failures, a further withdrawal of blood may not be possible or acceptable. To avoid this problem, immortalized lymphoblastoid cell lines (LCL) are frequently used instead of peripheral blood lymphocytes. LCL are considered to represent the original donor. In spite of their widespread use, the results of LCL have only rarely been compared to those of peripheral blood lymphocytes (PBL). In the present study, the DNA repair assay results on LCL and PBL of the same probands (n = 24) were evaluated for a correlation on an individual basis.

The results indicate that LCL are not applicable as individual surrogates for PBL for the use in the evaluated DNA repair assays (see *3.3.3*). No correlations were present between the respective assay results, indicating that the immortalization of cells by transfection with the Epstein-Barr virus induces large differences compared to the normal PBL of the donor.

LCL displayed numerous moderate correlations between some of the DNA repair assays. In the corresponding PBL samples of the 24 probands, only the baseline and the induced MN frequency were also correlated (r^2 = 0.33, P = 0.003), but weaker as in LCL. This indicates that in LCL the general independence of the assay results is not reflected and the end points of cellular DNA repair capacity are more related than in PBL. The weak correlation of the MN frequencies and the lack of correlation of SCE frequencies in the corresponding PBL samples of the 24 probands, are not in line to the results from the rest of the PBL samples. This is likely due to the comparably small size of the sample group used for the comparison to LCL, which is only a small fraction of all available PBL samples.

Few previous studies on the cellular DNA repair capacity have also used LCL, albeit with different results regarding their applicability as surrogates for PBL. While differences in the induction of MN by exposure to ionizing radiation were seen between breast cancer cases or *BRCA* mutation carriers and controls when PBL were used, these were not present in LCL [10,207]. In a further study, LCL of *BRCA1* and *BRCA2* mutation carriers as well controls exhibited considerable variations of the radiation induced MN frequency, telomere length and cell viability within each group. This compromised the prediction of the *BRCA* mutation status. [136]. On the other hand, LCL derived from sisters discordant for breast cancer were reported to exhibit a significant difference in the number of remaining BPDE adducts after time for repair [110].
The variability of the LCL in comparison to PBL originates most likely in the reprogramming and immortalization of the lymphocytes by the EBV transfection [215]. EBV-transfected cell clines exhibit differential expression of micro-RNAs [75,255] and differences in the promoter methylation status of a large number of genes [33] compared to PBL. As a result of these differences in gene expression, LCL cannot be expected to reflect genetic and environmental influences acting on the donors in the same way as PBL.

Cryopreserved isolated lymphocytes have also been investigated as a surrogate lymphocyte source. Results which exclude the use of cryopreserved lymphocytes instead of fresh peripheral blood lymphocytes have been reported [36]. In contrast, one study stated that the baseline MN frequency was highly correlated between the cryopreserved and freshly isolated lymphocytes [262]. This conclusion is not convincing, as the regression coefficient is only 0.23 and the correlation explains only one quarter of the observed variability between the cryopreserved and fresh lymphocytes of the donor. This would mean that also the cryopreservation introduces a large amount of variability which might lead to false results.

4.2.4 Intrinsic Variability of the Micronucleus Assay

It is known that the baseline and the induced MN frequency are influenced by many factors, which include genetic variations and exogenic factors (see [20] and [60] for reviews). Also a large degree of test-related variability is present, which poses a big problem to the reproducibility of results obtained with the MN assay [240]. Numerous large studies have been conducted to elicit the origin of this variability [16,27,61,63,127]. They attributed the variations mostly to differences in the experimental procedures between the laboratories and to variations in the scoring of MN. However, a large degree of variation still remains unexplained and is ascribed to the intrinsic test variability. Within the present work, this variability was also noticed, especially during the twin study where multiple cultures per individual were used. However, the origin of this intrinsic test variability within one laboratory has barely been analyzed.

The present study on the variability of the MN frequency was designed in such a way that most of the reported sources of variability were excluded. The inter-individual variability due to environmental or genetic factors as well as the intra-individual variability over time was excluded by the use of one single blood sample for each of the two types of MN assay. The assay was performed in a way to minimize potential sources of variation. The same culture media and solutions were used, and the experimental procedures, including the preparation of slides, were conducted in one series.

The scorer variability was largely eliminated by the use of an automated imaging system, the effectiveness of which was verified by the repeated scoring of all slides. The remaining variability factors were the two experimenters that conducted the assays, the lymphocyte culture and the slide preparation (see *3.3.4*).

In most of the studies, micronuclei were scored visually, and the differences between cultures and slides have been attributed to the variability introduced by the human scorers. This manual scoring of MN in binucleated cells is time and personnel consuming as well as prone to uncontrollable variation introduced by the human factor. The use of automated imaging systems for the scoring of the microscope slides that are also applied in the visual scoring does significantly reduce the amount of variation introduced by human scorers and increase the sensitivity of the assay as well as the reproducibility of the results [51,168,234]. Such a system was applied in the present study (see *2.3.2*).

The results show still a considerable degree of variability, with the differences between the smallest and the largest MN frequency within a homogeneous series of slides value ranging from

2 to 5-fold. In the baseline MN frequency, this variability is equally distributed between the investigator on the one side and the factors “culture” and “slide” on the other side. As such, the variance introduced by the two different slides that were generated from one culture is approximately twice the variance which is introduced by differences between the cultures within a slide set.

In the induced MN frequency, the variability is exclusively introduced by differences between cultures and slides. The slide sets of the two investigators show considerable differences. Whereas the absolute variability introduced by the slide factor is quite constant, the variance introduced by the culture factor is about 2-fold larger in investigator A, whereas it is only one fifth in investigator B. These results do not contradict the previously mentioned studies, as here a refined structure of the variance distribution is provided. The large inter-laboratory study carried out by the Human Micronucleus Consortium was based on cultures from one blood sample which were exposed to different doses of ionizing radiation. Slides from these cultures were sent out to be manually scored by the different laboratories. A model which could explain the major part of the observed variability attributed a large portion of the variance to the different radiation doses. However without the contribution of the exposure to ionizing irradiation, less than 50 % of the remaining variability was explained. Due to the setup of the mentioned study, differences between the cultures could not be assessed and the variation that was observed between the laboratories could only be ascribed to the manual slide scoring [61]. In the present study, the variation could be attributed to differences between cultures and slide preparations to the major part.

These variations cannot be overcome by increasing the number of binucleated cells which are scored from one slide or from multiple slides generated from one culture. The mean MN frequency was very similar between the two investigators. Exceptions were the slide sets from the first round of preparations in the study of the baseline MN frequency, which are discussed below.

An attempt was made to estimate the minimum number of parallel cultures necessary to obtain a “true” MN frequency value. For this analysis, the homogenous slide sets of the induced MN frequency and of the third preparation round of the baseline MN frequency were selected. In this setup, the statistical differences between randomly chosen subsets of cultures disappeared when a total of ten cultures were included in each subset. This number might be reduced to no less than five cultures, if a certain error from the true mean is considered acceptable.

The variability does not pose a problem if the MN assay is applied in mutagenicity studies, because these rely on the generation of dose-response curves upon the exposure to a particular

mutagen. However, in case-control studies the observed variability will most likely occlude influences with small effect size and may produce false-negative results, if the sample size is not sufficient to achieve enough discriminating power. This could explain some part of the many seemingly contradictory results in which the MN assay was used to assess the influence of epidemiological or genetic factors.

The observed differences between the cultures also influence the assessment of the MN frequency at different time points and thus can explain the observed intra-individual variability [240].

The source of these variations within the culture and slide level is hard to determine and due to the set up of the present study, it cannot be attributed to a particular step in the protocol.

For the baseline MN frequency, different rounds of slide preparation yielded highly significant differences between the slide sets. There is a systematic shift towards lower mean MN frequencies in the three consecutive slide sets of each investigator (see *Figure 03* and *Figure 22*). This shift is on the other hand not systematic when means of the single cultures are considered. The same holds true for the single counts or slides. The origins of this effect are unclear, and no reports of similar observations have been found in the literature.

The three rounds of slide preparations were conducted at different times after slide preparation. As such, the cells that were used for the third round spent much more time in the fixative (3 weeks) than the cells used for the first round (3 days). If the time in the fixative is interpreted as the determining variable, this effect would be expected.

The most outstanding slide set (I-B/P2) showed a moderate correlation of the MN frequency to the corresponding number of counted BNC, indicating an effect due to dense cultures or slides. The inclusion of the BNC numbers into the variance model resulted in an overall reduced variability of the baseline MN frequency and did eliminate some part of the variance of the outstanding slide set. However, it had only a weak effect on the other slide sets. This suggests that more steps of the protocol than previously thought have to be considered as possible sources of the intrinsic test variability of both the baseline as well the induced MN frequency, e.g. the time span in which the cells are kept in the fixative.

Further studies have to be conducted to elicit the sources of the observed variations of the MN frequency and to give an insight into the origin of systematic errors. They imply the risk of biased results if two groups are subject to large or opposed shifts due to subtle differences in the protocols.

4.3 Conclusions and Prospects

The overall cellular DNA repair capacity has been suggested and repeatedly studied as a potential intermediate phenotype of cancer. As many cancer susceptibility genes are DNA repair genes, the cellular DNA repair capacity may help to identify individuals with an increased cancer risk due to the presence of risk alleles in these genes. The aim of the present study was to characterize the relationship between genotypes of variants in DNA repair genes, the cellular DNA repair capacity and the occurrence of breast cancer.

Assays which measure different end points of DNA repair processes were applied. All processes induced by the presence and the removal of the specific lesions influence these end points, and thus variations in any of these processes affect the results of the applied DNA repair assays. Also the effects of further factors, for example differences in the epigenetic regulation of the involved genes, can influence the outcome of the DNA repair assays. The combination of different assays provides a more comprehensive characterization of the cellular ability to repair DNA damage than the independent study of the various distinct molecular mechanisms which are involved in the removal of a particular lesion.

The baseline and the induced micronucleus assay as well as the mitotic delay assay were determined by genetic factors to a major part, as revealed by the heritability estimates in a twin study. The heritability estimates were notably higher than for any of the common cancers, which is exactly what would be expected from an intermediate phenotype. In addition, the different DNA repair assays were found to be independent.

The observed concordance between the results of the different association studies which were performed on the same set of probands can conclusively support the view that the cellular DNA repair capacity is suitable as an intermediate phenotype of cancer, if studied with appropriate assays. Sporadic breast cancer cases displayed an elevated induced micronucleus frequency and a decreased mitotic delay index. Altered results of these two DNA repair assays were also associated to missense variants in genes which influence the corresponding DNA repair end points. *NINL* contributes to the activation of the G_2/M cell cycle check point and was associated with a decreased MD index, while *SLX4*, which is involved in the homologous recombination pathway, was associated with an increased induced MN frequency. Notably, the same variant in *SLX4* was also associated with an increased occurrence of familial and independently also of sporadic breast cancer. Especially in young cases the association yielded high odds ratios in both

groups. This is the first evidence that *SLX4*, which has very recently been established as a member of the Fanconi Anemia pathway, is a breast cancer susceptibility gene.
With regard to the outcome of the DNA repair assays and the functions of the associated genes, the data from the present study suggest that the increased risk of breast cancer is mediated by an impaired recognition or repair of radiation-induced DNA double strand breaks.

For the further use of the cellular DNA repair capacity as an intermediate phenotype of cancer, the origins of the intrinsic variations of DNA repair assays have to be identified and eliminated. Especially the MN assay, which is already frequently applied in the search for the genetic factors that underlie cancer risk, would be much more efficient and conclusive if its variability could be considerably reduced.
The associations which were seen in the present study between the cellular DNA capacity, the genotypes in DNA repair genes and the occurrence of breast cancer, have to be verified with a replication study in an independent cohort of breast cancer cases and controls. The associations to increased cancer risk have been in part already been verified in the present study by the separate comparison of sporadic and familial breast cancer cases with different cohorts of controls.
The association of the cellular DNA repair capacity to the presence of genetic variants could only be studied in sporadic breast cancer cases. This information is yet missing for the familial breast cancer cases, which exhibited the clearer association of genetic variants and breast cancer risk. A cohort of large families would also enable to test the co-segregation of the cellular DNA repair capacity with genetic variants. Also the co-segregation of the associated *SLX4* variant with the occurrence of cancer could be studied in such a cohort.
A further possible approach would be a genome-wide association study with different end points of cellular DNA repair processes as phenotypes. Such a study could contribute to uncover the numerous underlying genetic factors which are suggested from the high heritability estimates for some of the measures of cellular DNA repair capacity that were used in the present study. Whole-exome sequencing of cancer cases which display an impaired cellular DNA repair capacity could contribute to identify the potentially causative genetic variants.

5 Synopsis

Variants in many deoxyribonucleic acid (DNA) repair genes have been associated with an increased risk to develop cancer. Some studies have also associated an impaired cellular DNA repair capacity to an elevated occurrence of cancer or to the presence of particular genetic polymorphisms. The purpose of this study was to elucidate if all three aspects, namely the cellular DNA repair capacity, genotypes of DNA repair genes and the occurrence of breast cancer, can be associated with each other on the same set of probands in a single study.

The study population comprised different cohorts with sporadic or familial breast cancer cases and controls. Assays applicable for the quantification of different end points of cellular DNA repair processes were performed in parallel on peripheral blood samples of the probands: the baseline and the radiation induced micronucleus (MN) assay, the baseline and the Benzo[a]pyrenediepoxide induced sister chromatid exchange (SCE) assay and the mitotic delay (MD) assay. Fifteen DNA repair genes with a possible involvement in one of the studied DNA repair end points were chosen after extensive literature study. A total of 23 missense variants in these genes were genotyped.

Three types of association studies were performed on the same set of probands. The first type of association was between the occurrence of breast cancer and cellular DNA repair capacity. Altered results in two of the used DNA repair tests, an increased induced MN frequency and a decreased MD index, were significantly associated with the occurrence of breast cancer.

In the second type of association study, coding variants in DNA repair genes were associated with the occurrence of cancer. The minor allele of the variant rs3810813 in the *SLX4* gene, which results in a serine to phenylalanine residue change at position 1271 in the Slx4 protein, was significantly overrepresented in breast cancer cases (n = 435) compared to controls (n = 1189). The odds ratio was ~ 1.6 and the corresponding population attributable risk was 5.6 %. The minor allele frequency (MAF) displayed an increase with younger age of the cases. The overrepresentation was significant in familial breast cancer cases without a mutation in the *BRCA1* or *BRCA2* genes. It was also present in an independent group of sporadic breast cancer cases and reached significance when only young cases aged ≤ 40 years were considered. The *in-silico* analysis revealed that a probable functional impact of the variant on the function of the Slx4 protein. The common variant rs4986764 in the *BRIP1* gene showed a borderline association of the major allele to breast cancer in cases ≤ 40 years.

The third type of association was seen again with the minor allele of variant rs3810813 in *SLX4.* It was also associated with an increase of the induced MN frequency, albeit this was significant in

breast cancer cases only. Also the minor allele of the variant rs13044759 in the *NINL* gene, which is involved in the establishment of the cell cycle checkpoint in the second gap (G_2) phase, was associated with an increased MD index.

An accompanying aspect of the present study was the analysis of rarely characterized features of the DNA repair capacity measures as a prerequisite for their use in the definition as an intermediate phenotype of cancer. The heritability of the DNA repair assays was estimated in a twin study and indicated a major influence of genetic factors on the baseline MN frequency (70 %), on the induced MN frequency (60 %) and on the MD index (80 %). The results of the single assays were not or, in the case of the baseline and the induced SCE frequency, only weakly correlated. Lymphoblastoid cell lines (LCL) were generated from a subset of probands (n=24) and showed no correlations to the individual measures which were obtained from peripheral blood lymphocytes (PBL). In an analysis on the reproducibility of the MN assay, its intrinsic variability could be attributed to differences between cultures originating from one blood sample and to differences between the slide preparations from the same culture, although unexplained systematic effects were observed concerning the baseline MN frequency.

All measured end points are largely independent and can be used to characterize different aspects of the cellular DNA repair capacity. LCL are not applicable as surrogates for PBL albeit they would ensure the constant availability of lymphocyte material. The MN assays and the MD assay reflect the influence of genetic factors to a major part. Thus, the observed concordance between the results concerning the associations between coding variants in DNA repair genes, breast cancer incidence and altered outcomes of the induced MN and the MD assay supports the view that the cellular DNA repair capacity can be utilized as an intermediate phenotype of breast cancer. As *SLX4* is involved in the repair of DNA double strand breaks and the MN assay measures the presence of chromosomal fragments, these concordant results suggest a negative influence of the associated variant on the cellular capacity to repair double strand breaks. The increased MD index in carriers of the associated variant in *NINL* indicates a positive effect on the efficiency of the induction of the cell cycle delay, as the MD index was reduced in breast cancer cases.

The cellular DNA repair capacity could be used as the phenotype in large-scale studies to uncover the complex genetic origin of cancer risk.

6 Literature

1. Adzhubei I A, Schmidt S, Peshkin L, Ramensky V E, Gerasimova A, Bork P, Kondrashov A S, Sunyaev S R: A method and server for predicting damaging missense mutations. Nat.Methods, 7: 248-249 (2010)

2. Alter B P: Cancer in Fanconi anemia, 1927-2001. Cancer, 97: 425-440 (2003)

3. Angelini S, Kumar R, Carbone F, Bermejo J L, Maffei F, Cantelli-Forti G, Hemminki K, Hrelia P: Inherited susceptibility to bleomycin-induced micronuclei: correlating polymorphisms in GSTT1, GSTM1 and DNA repair genes with mutagen sensitivity. Mutat.Res., 638: 90-97 (2008)

4. Angelini S, Kumar R, Carbone F, Maffei F, Forti G C, Violante F S, Lodi V, Curti S, Hemminki K, Hrelia P: Micronuclei in humans induced by exposure to low level of ionizing radiation: influence of polymorphisms in DNA repair genes. Mutat.Res., 570: 105-117 (2005)

5. Anton-Guirgis H, Culver B D, Kurosaki T, Elston R: A study of multiple biological markers in twins. Acta Genet.Med.Gemellol.(Roma), 34: 153-165 (1985)

6. Aristei C, Stracci F, Guerrieri P, Anselmo P, Armellini R, Rulli A, Barberini F, Latini P, Menghini A R: Frequency of sister chromatid exchanges and micronuclei monitored over time in patients with early-stage breast cancer: results of an observational study. Cancer Genet.Cytogenet., 192: 24-29 (2009)

7. Aylon Y, Liefshitz B, Kupiec M: The CDK regulates repair of double-strand breaks by homologous recombination during the cell cycle. EMBO J., 23: 4868-4875 (2004)

8. Baeyens A, Thierens H, Claes K, Poppe B, De Ridder L, Vral A: Chromosomal radiosensitivity in BRCA1 and BRCA2 mutation carriers. Int.J.Radiat.Biol., 80: 745-756 (2004)

9. Baeyens A, Thierens H, Claes K, Poppe B, Messiaen L, De Ridder L, Vral A: Chromosomal radiosensitivity in breast cancer patients with a known or putative genetic predisposition. Br.J.Cancer, 87: 1379-1385 (2002)

10. Baeyens A, Thierens H, Vandenbulcke K, De Ridder L, Vral A: The use of EBV-transformed cell lines of breast cancer patients to measure chromosomal radiosensitivity. Mutagenesis, 19: 285-290 (2004)

11. Baeyens A, Van Den B R, Makar A, Thierens H, De Ridder L, Vral A: Chromosomal radiosensitivity in breast cancer patients: influence of age of onset of the disease. Oncol.Rep., 13: 347-353 (2005)

12. Baltaci V, Kayikcioglu F, Alpas I, Zeyneloglu H, Haberal A: Sister chromatid exchange rate and alkaline comet assay scores in patients with ovarian cancer. Gynecol.Oncol., 84: 62-66 (2002)

13. Ban S, Konomi C, Iwakawa M, Yamada S, Ohno T, Tsuji H, Noda S, Matui Y, Harada Y, Cologne J B, Imai T: Radiosensitivity of peripheral blood lymphocytes obtained from patients with cancers of the breast, head and neck or cervix as determined with a micronucleus assay. J.Radiat.Res., 45: 535-541 (2004)

14. Barale R, Barrai I, Sbrana I, Migliore L, Marrazzini A, Scarcelli V, Bacci E, Di Sibio A, Tessa A, Cocchi L: Monitoring human exposure to urban air pollutants. Environ.Health Perspect., 101 Suppl 3: 89-95 (1993)

15. Barale R, Chelotti L, Davini T, Del Ry S, Andreassi M G, Ballardin M, Bulleri M, He J, Baldacci S, Di Pede F, Gemignani F, Landi S: Sister chromatid exchange and micronucleus frequency in human lymphocytes of 1,650 subjects in an Italian population: II. Contribution of sex, age, and lifestyle. Environ.Mol.Mutagen., 31: 228-242 (1998)

16. Barale R, Marrazzini A, Bacci E, Di Sibio A, Tessa A, Cocchi L, Scarcelli V, Lubrano V, Vassalle C, Landi S: Sister chromatid exchange and micronucleus frequency in human lymphocytes of 1,650 subjects in an Italian population: I. Contribution of methodological factors. Environ.Mol.Mutagen., 31: 218-227 (1998)

17. Barrett J C, Fry B, Maller J, Daly M J: Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics, 21: 263-265 (2005)

18. Barroso E, Milne R L, Fernandez L P, Zamora P, Arias J I, Benitez J, Ribas G: FANCD2 associated with sporadic breast cancer risk. Carcinogenesis, 27: 1930-1937 (2006)

19. Bartsch H, Dally H, Popanda O, Risch A, Schmezer P: Genetic risk profiles for cancer susceptibility and therapy response. Recent Results Cancer Res., 174: 19-36 (2007)

20. Battershill J M, Burnett K, Bull S: Factors affecting the incidence of genotoxicity biomarkers in peripheral blood lymphocytes: impact on design of biomonitoring studies. Mutagenesis, 23: 423-437 (2008)

21. Bernardini S, Hirvonen A, Jarventaus H, Norppa H: Influence of GSTM1 and GSTT1 genotypes on sister chromatid exchange induction by styrene in cultured human lymphocytes. Carcinogenesis, 23: 893-897 (2002)

22. Berwick M, Vineis P: Markers of DNA repair and susceptibility to cancer in humans: an epidemiologic review. J.Natl.Cancer Inst., 92: 874-897 (2000)

23. Blastyak A, Hajdu I, Unk I, Haracska L: Role of double-stranded DNA translocase activity of human HLTF in replication of damaged DNA. Mol.Cell.Biol., 30: 684-693 (2010)

24. Boeckmann B, Bairoch A, Apweiler R, Blatter M C, Estreicher A, Gasteiger E, Martin M J, Michoud K, O'Donovan C, Phan I, Pilbout S, Schneider M: The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. Nucleic Acids Res., 31: 365-370 (2003)

25. Bogdanova N, Feshchenko S, Schurmann P, Waltes R, Wieland B, Hillemanns P, Rogov Y I, Dammann O, Bremer M, Karstens J H, Sohn C, Varon R, Dork T: Nijmegen Breakage Syndrome mutations and risk of breast cancer. Int.J.Cancer, 122: 802-806 (2008)

26. Bombarde O, Boby C, Gomez D, Frit P, Giraud-Panis M J, Gilson E, Salles B, Calsou P: TRF2/RAP1 and DNA-PK mediate a double protection against joining at telomeric ends. EMBO J., 29: 1573-1584 (2010)

27. Bonassi S, Fenech M, Lando C, Lin Y P, Ceppi M, Chang W P, Holland N, Kirsch-Volders M, Zeiger E, Ban S, Barale R, Bigatti M P, Bolognesi C, Jia C, Di Giorgio M, Ferguson L R, Fucic A, Lima O G, Hrelia P, Krishnaja A P, Lee T K, Migliore L, Mikhalevich L, Mirkova E, Mosesso P, Muller W U, Odagiri Y, Scarffi M R, Szabova E, Vorobtsova I, Vral A, Zijno A: HUman MicroNucleus project: international database comparison for results with the cytokinesis-block micronucleus assay in human lymphocytes: I. Effect of laboratory protocol, scoring criteria, and host factors on the frequency of micronuclei. Environ.Mol.Mutagen., 37: 31-45 (2001)

28. Bonassi S, Znaor A, Ceppi M, Lando C, Chang W P, Holland N, Kirsch-Volders M, Zeiger E, Ban S, Barale R, Bigatti M P, Bolognesi C, Cebulska-Wasilewska A, Fabianova E, Fucic A, Hagmar L, Joksic G, Martelli A, Migliore L, Mirkova E, Scarfi M R, Zijno A, Norppa H, Fenech M: An increased micronucleus frequency in peripheral blood lymphocytes predicts the risk of cancer in humans. Carcinogenesis, 28: 625-631 (2007)

29. Boomsma D, Busjahn A, Peltonen L: Classical twin studies and beyond. Nat.Rev.Genet., 3: 872-882 (2002)

30. Borgmann K, Haeberle D, Doerk T, Busjahn A, Stephan G, Dikomey E: Genetic determination of chromosomal radiosensitivities in G0- and G2-phase human lymphocytes. Radiother.Oncol., 83: 196-202 (2007)

31. Bouquet F, Muller C, Salles B: The loss of gammaH2AX signal is a marker of DNA double strand breaks repair only at low levels of DNA damage. Cell.Cycle, 5: 1116-1122 (2006)

32. Branzei D, Foiani M: Regulation of DNA repair throughout the cell cycle. Nat.Rev.Mol.Cell Biol., 9: 297-308 (2008)

33. Brennan E P, Ehrich M, Brazil D P, Crean J K, Murphy M, Sadlier D M, Martin F, Godson C, McKnight A J, van den Boom D, Maxwell A P, Savage D A: Comparative analysis of DNA methylation profiles in peripheral blood leukocytes versus lymphoblastoid cell lines. Epigenetics, 4: 159-164 (2009)

34. Broberg K, Huynh E, Schlawicke Engstrom K, Bjork J, Albin M, Ingvar C, Olsson H, Hoglund M: Association between polymorphisms in RMI1, TOP3A, and BLM and risk of cancer, a case-control study. BMC Cancer, 9: 140 (2009)

35. Buck D, Moshous D, de Chasseval R, Ma Y, le Deist F, Cavazzana-Calvo M, Fischer A, Casanova J L, Lieber M R, de Villartay J P: Severe combined immunodeficiency and microcephaly in siblings with hypomorphic mutations in DNA ligase IV. Eur.J.Immunol., 36: 224-235 (2006)

36. Burrill W, Levine E L, Hindocha P, Roberts S A, Scott D: The use of cryopreserved lymphocytes in assessing inter-individual radiosensitivity with the micronucleus assay. Int.J.Radiat.Biol., 76: 375-382 (2000)

37. Camplejohn R S, Hodgson S, Carter N, Kato B S, Spector T D: Heritability of DNA-damage-induced apoptosis and its relationship with age in lymphocytes from female twins. Br.J.Cancer, 95: 520-524 (2006)

38. Cantor S B, Bell D W, Ganesan S, Kass E M, Drapkin R, Grossman S, Wahrer D C, Sgroi D C, Lane W S, Haber D A, Livingston D M: BACH1, a novel helicase-like protein, interacts directly with BRCA1 and contributes to its DNA repair function. Cell, 105: 149-160 (2001)

39. Capasso M, Devoto M, Hou C, Asgharzadeh S, Glessner J T, Attiyeh E F, Mosse Y P, Kim C, Diskin S J, Cole K A, Bosse K, Diamond M, Laudenslager M, Winter C, Bradfield J P, Scott R H, Jagannathan J, Garris M, McConville C, London W B, Seeger R C, Grant S F, Li H, Rahman N, Rappaport E, Hakonarson H, Maris J M: Common variations in BARD1 influence susceptibility to high-risk neuroblastoma. Nat.Genet., 41: 718-723 (2009)

40. Capp J P, Boudsocq F, Bergoglio V, Trouche D, Cazaux C, Blanco L, Hoffmann J S, Canitrot Y: The R438W polymorphism of human DNA polymerase lambda triggers cellular sensitivity to camptothecin by compromising the homologous recombination repair pathway. Carcinogenesis, 31: 1742-1747 (2010)

41. Carroll K A, Ly H: Telomere dysfunction in human diseases: the long and short of it!. Int.J.Clin.Exp.Pathol., 2: 528-543 (2009)

42. Carvajal-Carmona L G: Genetic dissection of intermediate phenotypes as a way to discover novel cancer susceptibility alleles. Curr.Opin.Genet.Dev., 20: 308-314 (2010)

43. Cefle K, Ucur A, Guney N, Ozturk S, Palanduz S, Tas F, Asoglu O, Bayrak A, Muslumanoglu M, Aydiner A: Increased sister chromatid exchange frequency in young women with breast cancer and in their first-degree relatives. Cancer Genet.Cytogenet., 171: 65-67 (2006)

44. Collis S J, Ciccia A, Deans A J, Horejsi Z, Martin J S, Maslen S L, Skehel J M, Elledge S J, West S C, Boulton S J: FANCM and FAAP24 function in ATR-mediated checkpoint signaling independently of the Fanconi anemia core complex. Mol.Cell, 32: 313-324 (2008)

45. Cortes-Ledesma F, Tous C, Aguilera A: Different genetic requirements for repair of replication-born double-strand breaks by sister-chromatid recombination and break-induced replication. Nucleic Acids Res., 35: 6560-6570 (2007)

46. Cybulski C, Gorski B, Debniak T, Gliniewicz B, Mierzejewski M, Masojc B, Jakubowska A, Matyjasik J, Zlowocka E, Sikorski A, Narod S A, Lubinski J: NBS1 is a prostate cancer susceptibility gene. Cancer Res., 64: 1215-1219 (2004)

47. Cybulski C, Gorski B, Huzarski T, Masojc B, Mierzejewski M, Debniak T, Teodorczyk U, Byrski T, Gronwald J, Matyjasik J, Zlowocka E, Lenner M, Grabowska E, Nej K, Castaneda J, Medrek K, Szymanska A, Szymanska J, Kurzawski G, Suchy J, Oszurek O, Witek A, Narod S A, Lubinski J: CHEK2 is a multiorgan cancer susceptibility gene. Am.J.Hum.Genet., 75: 1131-1135 (2004)

48. Czene K, Lichtenstein P, Hemminki K: Environmental and heritable causes of cancer among 9.6 million individuals in the Swedish Family-Cancer Database. Int.J.Cancer, 99: 260-266 (2002)

49. De Brakeleer S, De Greve J, Loris R, Janin N, Lissens W, Sermijn E, Teugels E: Cancer predisposing missense and protein truncating BARD1 mutations in non-BRCA1 or BRCA2 breast cancer families. Hum.Mutat., 31: E1175-85 (2010)

50. Decordier I, Loock K V, Kirsch-Volders M: Phenotyping for DNA repair capacity. Mutat.Res., 705: 107-129 (2010)

51. Decordier I, Papine A, Plas G, Roesems S, Vande Loock K, Moreno-Palomo J, Cemeli E, Anderson D, Fucic A, Marcos R, Soussaline F, Kirsch-Volders M: Automated image analysis of cytokinesis-blocked micronuclei: an adapted protocol and a validated scoring procedure for biomonitoring. Mutagenesis, 24: 85-93 (2009)

52. Ding S L, Yu J C, Chen S T, Hsu G C, Shen C Y: Genetic variation in the premature aging gene WRN: a case-control study on breast cancer susceptibility. Cancer Epidemiol.Biomarkers Prev., 16: 263-269 (2007)

53. Djuzenova C S, Muhl B, Fehn M, Oppitz U, Muller B, Flentje M: Radiosensitivity in breast cancer assessed by the Comet and micronucleus assays. Br.J.Cancer, 94: 1194-1203 (2006)

54. Easton D F, Eeles R A: Genome-wide association studies in cancer. Hum.Mol.Genet., 17: R109-R115 (2008)

55. El Zein R A, Fenech M, Lopez M S, Spitz M R, Etzel C J: Cytokinesis-blocked micronucleus cytome assay biomarkers identify lung cancer cases amongst smokers. Cancer Epidemiol.Biomarkers Prev., 17: 1111-1119 (2008)

56. Ellis N A, Groden J, Ye T Z, Straughen J, Lennon D J, Ciocci S, Proytcheva M, German J: The Bloom's syndrome gene product is homologous to RecQ helicases. Cell, 83: 655-666 (1995)

57. Febrer E, Mestres M, Caballin M R, Barrios L, Ribas M, Gutierrez-Enriquez S, Alonso C, Ramon y Cajal T, Francesc Barquinero J: Mitotic delay in lymphocytes from BRCA1 heterozygotes unable to reduce the radiation-induced chromosomal damage. DNA Repair (Amst), 7: 1907-1911 (2008)

58. Fekairi S, Scaglione S, Chahwan C, Taylor E R, Tissier A, Coulon S, Dong M Q, Ruse C, Yates J R,3rd, Russell P, Fuchs R P, McGowan C H, Gaillard P H: Human SLX4 is a Holliday junction resolvase subunit that binds multiple DNA repair/recombination endonucleases. Cell, 138: 78-89 (2009)

59. Fenech M: Cytokinesis-block micronucleus cytome assay. Nat.Protoc., 2: 1084-1104 (2007)

60. Fenech M, Bonassi S: The effect of age, gender, diet and lifestyle on DNA damage measured using micronucleus frequency in human peripheral blood lymphocytes. Mutagenesis, 26: 43-49 (2011)

61. Fenech M, Bonassi S, Turner J, Lando C, Ceppi M, Chang W P, Holland N, Kirsch-Volders M, Zeiger E, Bigatti M P, Bolognesi C, Cao J, De Luca G, Di Giorgio M, Ferguson L R, Fucic A, Lima O G, Hadjidekova V V, Hrelia P, Jaworska A, Joksic G, Krishnaja A P, Lee T K, Martelli A, McKay M J, Migliore L, Mirkova E, Muller W U, Odagiri Y, Orsiere T, Scarfi M R, Silva M J, Sofuni T, Surralles J, Trenta G, Vorobtsova I, Vral A, Zijno A, HUman MicroNucleus project: Intra- and inter-laboratory variation in the scoring of micronuclei and nucleoplasmic bridges in binucleated human lymphocytes. Results of an international slide-scoring exercise by the HUMN project. Mutat.Res., 534: 45-64 (2003)

62. Fenech M, Chang W P, Kirsch-Volders M, Holland N, Bonassi S, Zeiger E: HUMN project: detailed description of the scoring criteria for the cytokinesis-block micronucleus assay using isolated human lymphocyte cultures. Mutat.Res., 534: 65-75 (2003)

63. Fenech M, Holland N, Chang W P, Zeiger E, Bonassi S: The HUman MicroNucleus Project--An international collaborative study on the use of the micronucleus technique for measuring DNA damage in humans. Mutat.Res., 428: 271-283 (1999)

64. Fenech M, Morley A A: Cytokinesis-block micronucleus method in human lymphocytes: effect of in vivo ageing and low dose X-irradiation. Mutat.Res., 161: 193-198 (1986)

65. Feng R, Zhou G, Zhang M, Zhang H: Analysis of twin data using SAS. Biometrics, 65: 584-589 (2009)

66. Ferlay J, Shin H R, Bray F, Forman D, Mathers C, Parkin D M: Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. Int.J.Cancer, 127: 2893-2917 (2010)

67. Finnon P, Robertson N, Dziwura S, Raffy C, Zhang W, Ainsbury L, Kaprio J, Badie C, Bouffler S: Evidence for significant heritability of apoptotic and cell cycle responses to ionising radiation. Hum.Genet., 123: 485-493 (2008)

68. Franchitto A, Pichierri P, Mosesso P, Palitti F: Caffeine effect on the mitotic delay induced by G2 treatment with UVC or mitomycin C. Mutagenesis, 13: 499-505 (1998)

69. Frank B, Hemminki K, Meindl A, Wappenschmidt B, Sutter C, Kiechle M, Bugert P, Schmutzler R K, Bartram C R, Burwinkel B: BRIP1 (BACH1) variants and familial breast cancer risk: a case-control study. BMC Cancer, 7: 83 (2007)

70. Galvan A, Ioannidis J P, Dragani T A: Beyond genome-wide association studies: genetic heterogeneity and individual predisposition to cancer. Trends Genet., 26: 132-141 (2010)

71. Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel R D, Bairoch A: ExPASy: The proteomics server for in-depth protein knowledge and analysis. Nucleic Acids Res., 31: 3784-3788 (2003)

72. Geer L Y, Marchler-Bauer A, Geer R C, Han L, He J, He S, Liu C, Shi W, Bryant S H: The NCBI BioSystems database. Nucleic Acids Res., 38: D492-D496 (2010)

73. German J: Bloom's syndrome. XX. The first 100 cancers. Cancer Genet.Cytogenet., 93: 100-106 (1997)

74. Gerper P, Whang-Peng J, Monroe J H: Transformation and chromosome changes induced by Epstein-Barr virus in normal human leukocyte cultures. Proc.Natl.Acad.Sci.U.S.A., 63: 740-747 (1969)

75. Godshalk S E, Bhaduri-McIntosh S, Slack F J: Epstein-Barr virus-mediated dysregulation of human microRNA expression. Cell.Cycle, 7: 3595-3600 (2008)

76. Goldgar D E, Easton D F, Cannon-Albright L A, Skolnick M H: Systematic population-based assessment of cancer risk in first-degree relatives of cancer probands. J.Natl.Cancer Inst., 86: 1600-1608 (1994)

77. Goto M, Miller R W, Ishikawa Y, Sugano H: Excess of rare cancers in Werner syndrome (adult progeria). Cancer Epidemiol.Biomarkers Prev., 5: 239-246 (1996)

78. Greenberg R A, Sobhian B, Pathania S, Cantor S B, Nakatani Y, Livingston D M: Multifactorial contributions to an acute DNA damage response by BRCA1/BARD1-containing complexes. Genes Dev., 20: 34-46 (2006)

79. Gutierrez-Enriquez S, Fernet M, Dork T, Bremer M, Lauge A, Stoppa-Lyonnet D, Moullan N, Angele S, Hall J: Functional consequences of ATM sequence variants for chromosomal radiosensitivity. Genes Chromosomes Cancer, 40: 109-119 (2004)

80. He X, Ye F, Zhang J, Cheng Q, Shen J, Chen H: REV1 genetic variants associated with the risk of cervical carcinoma. Eur.J.Epidemiol., 23: 403-409 (2008)

81. Heddle J A: A rapid in vivo test for chromosomal damage. Mutat.Res., 18: 187-190 (1973)

82. Hedner K, Iselius L, Mitelman F, Norden A, Pero R W: A twin study of sister chromatid exchanges in human lymphocytes following carcinogen exposure and DNA repair incubation. Cytogenet.Cell Genet., 38: 189-191 (1984)

83. Heikkinen K, Rapakko K, Karppinen S M, Erkko H, Knuutila S, Lundan T, Mannermaa A, Borresen-Dale A L, Borg A, Barkardottir R B, Petrini J, Winqvist R: RAD50 and NBS1 are breast cancer susceptibility genes associated with genomic instability. Carcinogenesis, 27: 1593-1599 (2006)

84. Heikkinen K, Rapakko K, Karppinen S M, Erkko H, Nieminen P, Winqvist R: Association of common ATM polymorphism with bilateral breast cancer. Int.J.Cancer, 116: 69-72 (2005)

85. Hemminki K, Vaittinen P, Kyyronen P: Age-specific familial risks in common cancers of the offspring. Int.J.Cancer, 78: 172-175 (1998)

86. Henikoff S, Henikoff J G: Amino acid substitution matrices from protein blocks. Proc.Natl.Acad.Sci.U.S.A., 89: 10915-10919 (1992)

87. Hill J W, Tansavatdi K, Lockett K L, Allen G O, Takita C, Pollack A, Hu J J: Validation of the cell cycle G(2) delay assay in assessing ionizing radiation sensitivity and breast cancer risk. Cancer.Manag.Res., 1: 39-48 (2009)

88. Hindorff L A, Sethupathy P, Junkins H A, Ramos E M, Mehta J P, Collins F S, Manolio T A: Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. Proc.Natl.Acad.Sci.U.S.A., 106: 9362-9367 (2009)

89. Hindorff L, Junkins H, Hall P, Mehta J, Manolio T: A Catalog of Published Genome-Wide Association Studies. Available at: www.genome.gov/gwastudies. Accessed November 11, 2010

90. Hirsch B A, Sentz K K, McGue M: Genetic and environmental influences on baseline SCE. Environ.Mol.Mutagen., 20: 2-11 (1992)

91. Hu J J, Smith T R, Miller M S, Lohman K, Case L D: Genetic regulation of ionizing radiation sensitivity and breast cancer risk. Environ.Mol.Mutagen., 39: 208-215 (2002)

92. Hu J J, Smith T R, Miller M S, Mohrenweiser H W, Golden A, Case L D: Amino acid substitution variants of APE1 and XRCC1 genes associated with ionizing radiation sensitivity. Carcinogenesis, 22: 917-922 (2001)

93. Huang M, Kim J M, Shiotani B, Yang K, Zou L, D'Andrea A D: The FANCM/FAAP24 complex is required for the DNA interstrand crosslink-induced checkpoint response. Mol.Cell, 39: 259-268 (2010)

94. Iarmarcovai G, Bonassi S, Botta A, Baan R A, Orsiere T: Genetic polymorphisms and micronucleus formation: a review of the literature. Mutat.Res., 658: 215-233 (2008)

95. Iarmarcovai G, Ceppi M, Botta A, Orsiere T, Bonassi S: Micronuclei frequency in peripheral blood lymphocytes of cancer patients: a meta-analysis. Mutat.Res., 659: 274-283 (2008)

96. Imai M, Matsuno R, Komura J, Ono T, Yamamoto K: Induction of mitosis delay, apoptosis and aneuploidy in human cells by phenyl hydroquinone, an Ames test-negative carcinogen. Genes Genet.Syst., 84: 73-79 (2009)

97. International HapMap Consortium: The International HapMap Project. Nature, 426: 789-796 (2003)

98. International Human Genome Sequencing Consortium: Finishing the euchromatic sequence of the human genome. Nature, 431: 931-945 (2004)

99. Ioannidis J P, Castaldi P, Evangelou E: A compendium of genome-wide associations for cancer: critical synopsis and reappraisal. J.Natl.Cancer Inst., 102: 846-858 (2010)

100. Jackson S P, Bartek J: The DNA-damage response in human biology and disease. Nature, 461: 1071-1078 (2009)

101. Jakubowska A, Cybulski C, Szymanska A, Huzarski T, Byrski T, Gronwald J, Debniak T, Gorski B, Kowalska E, Narod S A, Lubinski J: BARD1 and breast cancer in Poland. Breast Cancer Res.Treat., 107: 119-122 (2008)

102. Jenssen D, Ramel C: Factors affecting the induction of micronuclei at low doses of X-rays, MMS and dimethylnitrosamine in mouse erythroblasts. Mutat.Res., 58: 51-65 (1978)

103. Jin S, Gao H, Mazzacurati L, Wang Y, Fan W, Chen Q, Yu W, Wang M, Zhu X, Zhang C, Zhan Q: BRCA1 interaction of centrosomal protein Nlp is required for successful mitotic progression. J.Biol.Chem., 284: 22970-22977 (2009)

104. Johnson R D, Jasin M: Sister chromatid gene conversion is a prominent double-strand break repair pathway in mammalian cells. EMBO J., 19: 3398-3407 (2000)

105. Jones I M, Thomas C B, Xi T, Mohrenweiser H W, Nelson D O: Exploration of methods to identify polymorphisms associated with variation in DNA repair capacity phenotypes. Mutat.Res., 616: 213-220 (2007)

106. Jones S, Hruban R H, Kamiyama M, Borges M, Zhang X, Parsons D W, Lin J C, Palmisano E, Brune K, Jaffee E M, Iacobuzio-Donahue C A, Maitra A, Parmigiani G, Kern S E, Velculescu V E, Kinzler K W, Vogelstein B, Eshleman J R, Goggins M, Klein A P: Exomic sequencing identifies PALB2 as a pancreatic cancer susceptibility gene. Science, 324: 217 (2009)

107. Kamb A, Shattuck-Eidens D, Eeles R, Liu Q, Gruis N A, Ding W, Hussey C, Tran T, Miki Y, Weaver-Feldhaus J: Analysis of the p16 gene (CDKN2) as a candidate for the chromosome 9p melanoma susceptibility locus. Nat.Genet., 8: 23-26 (1994)

108. Kao G D, McKenna W G, Yen T J: Detection of repair activity during the DNA damage-induced G2 delay in human cancer cells. Oncogene, 20: 3486-3496 (2001)

109. Karppinen S M, Barkardottir R B, Backenhorn K, Sydenham T, Syrjakoski K, Schleutker J, Ikonen T, Pylkas K, Rapakko K, Erkko H, Johannesdottir G, Gerdes A M, Thomassen M, Agnarsson B A, Grip M, Kallioniemi A, Kere J, Aaltonen L A, Arason A, Moller P, Kruse T A, Borg A, Winqvist R: Nordic collaborative study of the BARD1 Cys557Ser allele in 3956 cancer cases: enrichment in familial BRCA1/BRCA2 mutation-negative breast cancer but not in other malignancies. J.Med.Genet., 43: 856-862 (2006)

110. Kennedy D O, Agrawal M, Shen J, Terry M B, Zhang F F, Senie R T, Motykiewicz G, Santella R M: DNA repair capacity of lymphoblastoid cell lines from sisters discordant for breast cancer. J.Natl.Cancer Inst., 97: 127-132 (2005)

111. Kiechle M, Gross E, Schwarz-Boeger U, Pfisterer J, Jonat W, Gerber W D, Albacht B, Fischer B, Schlegelberger B, Arnold N: Ten novel BRCA1 and BRCA2 mutations in breast and/or ovarian cancer families from northern Germany. Hum.Mutat., 16: 529-530 (2000)

112. Kim M K, Zitzmann S, Westermann F, Arnold K, Brouwers S, Schwab M, Savelyeva L: Increased rates of spontaneous sister chromatid exchange in lymphocytes of BRCA2+/- carriers of familial breast cancer clusters. Cancer Lett., 210: 85-94 (2004)

113. Kim S T, Cheng Y, Hsu F C, Jin T, Kader A K, Zheng S L, Isaacs W B, Xu J, Sun J: Prostate cancer risk-associated variants reported from genome-wide association studies: meta-analysis and their contribution to genetic Variation. Prostate, 70: 1729-1738 (2010)

114. Kim Y, Lach F P, Desetty R, Hanenberg H, Auerbach A D, Smogorzewska A: Mutations of the SLX4 gene in Fanconi anemia. Nat.Genet., 43: 142-146 (2011)

115. Kirsch-Volders M, Mateuca R A, Roelants M, Tremp A, Zeiger E, Bonassi S, Holland N, Chang W P, Aka P V, Deboeck M, Godderis L, Haufroid V, Ishikawa H, Laffon B, Marcos R, Migliore L, Norppa H, Teixeira J P, Zijno A, Fenech M: The effects of GSTM1 and GSTT1 polymorphisms on micronucleus frequencies in human lymphocytes in vivo. Cancer Epidemiol.Biomarkers Prev., 15: 1038-1042 (2006)

116. Knipscheer P, Raschle M, Smogorzewska A, Enoiu M, Ho T V, Scharer O D, Elledge S J, Walter J C: The Fanconi anemia pathway promotes replication-dependent DNA interstrand cross-link repair. Science, 326: 1698-1701 (2009)

117. Kote-Jarai Z, Jugurnauth S, Mulholland S, Leongamornlert D A, Guy M, Edwards S, Tymrakiewitcz M, O'Brien L, Hall A, Wilkinson R, Al Olama A A, Morrison J, Muir K, Neal D, Donovan J, Hamdy F, Easton D F, Eeles R, UKGPCS Collaborators, British Association of Urological Surgeons' Section of Oncology: A recurrent truncating germline mutation in the BRIP1/FANCJ gene and susceptibility to prostate cancer. Br.J.Cancer, 100: 426-430 (2009)

118. Kotsopoulos J, Chen Z, Vallis K A, Poll A, Ainsworth P, Narod S A: DNA repair capacity as a possible biomarker of breast cancer risk in female BRCA1 mutation carriers. Br.J.Cancer, 96: 118-125 (2007)

119. Kraemer K H, Patronas N J, Schiffmann R, Brooks B P, Tamura D, DiGiovanna J J: Xeroderma pigmentosum, trichothiodystrophy and Cockayne syndrome: a complex genotype-phenotype relationship. Neuroscience, 145: 1388-1396 (2007)

120. Kuhn E M, Therman E: No increased chromosome breakage in three Bloom's syndrome heterozygotes. J.Med.Genet., 16: 219-222 (1979)

121. Lasne C, Gu Z W, Venegas W, Chouroulinkov I: The in vitro micronucleus assay for detection of cytogenetic effects induced by mutagen-carcinogens: comparison with the in vitro sister-chromatid exchange assay. Mutat.Res., 130: 273-282 (1984)

122. Lavin M F, Bennett I, Ramsay J, Gardiner R A, Seymour G J, Farrell A, Walsh M: Identification of a potentially radiosensitive subgroup among patients with breast cancer. J.Natl.Cancer Inst., 86: 1627-1634 (1994)

123. Lavin M F, Le Poidevin P, Bates P: Enhanced levels of radiation-induced G2 phase delay in ataxia telangiectasia heterozygotes. Cancer Genet.Cytogenet., 60: 183-187 (1992)

124. Leach F S, Nicolaides N C, Papadopoulos N, Liu B, Jen J, Parsons R, Peltomaki P, Sistonen P, Aaltonen L A, Nystrom-Lahti M: Mutations of a mutS homolog in hereditary nonpolyposis colorectal cancer. Cell, 75: 1215-1225 (1993)

125. Lee G H, Matsushita H: Genetic linkage between Pol iota deficiency and increased susceptibility to lung tumors in mice. Cancer.Sci., 96: 256-259 (2005)

126. Lee J, Kim J A, Barbier V, Fotedar A, Fotedar R: DNA damage triggers p21WAF1-dependent Emi1 down-regulation that maintains G2 arrest. Mol.Biol.Cell, 20: 1891-1902 (2009)

127. Lee T K, O'Brien K, Eaves G S, Christie K I, Varga L: Effect of blood storage on radiation-induced micronuclei in human lymphocytes. Mutat.Res., 444: 201-206 (1999)

128. Lei Y C, Hwang S J, Chang C C, Kuo H W, Luo J C, Chang M J, Cheng T J: Effects on sister chromatid exchange frequency of polymorphisms in DNA repair gene XRCC1 in smokers. Mutat.Res., 519: 93-101 (2002)

129. Leman A R, Noguchi C, Lee C Y, Noguchi E: Human Timeless and Tipin stabilize replication forks and facilitate sister-chromatid cohesion. J.Cell.Sci., 123: 660-670 (2010)

130. Leopardi P, Marcon F, Caiola S, Cafolla A, Siniscalchi E, Zijno A, Crebelli R: Effects of folic acid deficiency and MTHFR C677T polymorphism on spontaneous and radiation-induced micronuclei in human lymphocytes. Mutagenesis, 21: 327-333 (2006)

131. Li F P, Fraumeni J F,Jr: Soft-tissue sarcomas, breast cancer, and other neoplasms. A familial syndrome?. Ann.Intern.Med., 71: 747-752 (1969)

132. Lichtenstein P, Holm N V, Verkasalo P K, Iliadou A, Kaprio J, Koskenvuo M, Pukkala E, Skytthe A, Hemminki K: Environmental and heritable factors in the causation of cancer--analyses of cohorts of twins from Sweden, Denmark, and Finland. N.Engl.J.Med., 343: 78-85 (2000)

133. Lin J, Kadlubar F F, Spitz M R, Zhao H, Wu X: A modified host cell reactivation assay to measure DNA repair capacity for removing 4-aminobiphenyl adducts: a pilot study of bladder cancer. Cancer Epidemiol.Biomarkers Prev., 14: 1832-1836 (2005)

134. Lindberg H K, Falck G C, Jarventaus H, Norppa H: Characterization of chromosomes and chromosomal fragments in human lymphocyte micronuclei by telomeric and centromeric FISH. Mutagenesis, 23: 371-376 (2008)

135. Ling C, Ishiai M, Ali A M, Medhurst A L, Neveling K, Kalb R, Yan Z, Xue Y, Oostra A B, Auerbach A D, Hoatlin M E, Schindler D, Joenje H, de Winter J P, Takata M, Meetei A R, Wang W: FAAP100 is essential for activation of the Fanconi anemia-associated DNA damage response pathway. EMBO J., 26: 2104-2114 (2007)

136. Lovelock P K, Wong E M, Sprung C N, Marsh A, Hobson K, French J D, Southey M, Sculley T, Pandeya N, Brown M A, Chenevix-Trench G, Spurdle A B, McKay M J, kConFab Investigators: Prediction of BRCA1 and BRCA2 mutation status using post-irradiation assays of lymphoblastoid cell lines is compromised by inter-cell-line phenotypic variability. Breast Cancer Res.Treat., 104: 257-266 (2007)

137. Lu J, Wei Q, Bondy M L, Li D, Brewster A, Shete S, Yu T K, Sahin A, Meric-Bernstam F, Hunt K K, Singletary S E, Ross M I, Wang L E: Polymorphisms and haplotypes of the NBS1 gene are associated with risk of sporadic breast cancer in non-Hispanic white women. Carcinogenesis, 27: 2209-2216 (2006)

138. Luedeke M, Linnert C M, Hofer M D, Surowy H M, Rinckleb A E, Hoegel J, Kuefer R, Rubin M A, Vogel W, Maier C: Predisposition for TMPRSS2-ERG fusion in prostate cancer by variants in DNA repair genes. Cancer Epidemiol.Biomarkers Prev., 18: 3030-3035 (2009)

139. Lynch H T, Lynch P M, Lanspa S J, Snyder C L, Lynch J F, Boland C R: Review of the Lynch syndrome: history, molecular genetics, screening, differential diagnosis, and medicolegal ramifications. Clin.Genet., 76: 1-18 (2009)

140. MacGregor J T, Wehr C M, Hiatt R A, Peters B, Tucker J D, Langlois R G, Jacob R A, Jensen R H, Yager J W, Shigenaga M K, Frei B, Eynon B P, Ames B N: 'Spontaneous' genetic damage in man: evaluation of interindividual variability, relationship among markers of damage, and influence of nutritional status. Mutat.Res., 377: 125-135 (1997)

141. Maffei F, Carbone F, Angelini S, Forti G C, Norppa H, Hrelia P: Micronuclei frequency induced by bleomycin in human peripheral lymphocytes: correlating BLHX polymorphism with mutagen sensitivity. Mutat.Res., 639: 20-26 (2008)

142. Maglott D, Ostell J, Pruitt K D, Tatusova T: Entrez Gene: gene-centered information at NCBI. Nucleic Acids Res., 35: D26-D31 (2007)

143. Mahmoudi S, Henriksson S, Weibrecht I, Smith S, Soderberg O, Stromblad S, Wiman K G, Farnebo M: WRAP53 is essential for Cajal body formation and for targeting the survival of motor neuron complex to Cajal bodies. PLoS Biol., 8: e1000521 (2010)

144. Maier C, Haeusler J, Herkommer K, Vesovic Z, Hoegel J, Vogel W, Paiss T: Mutation screening and association study of RNASEL as a prostate cancer susceptibility gene. Br.J.Cancer, 92: 1159-1164 (2005)

145. Maillet P, Chappuis P O, Vaudan G, Dobbie Z, Muller H, Hutter P, Sappino A P: A polymorphism in the ATM gene modulates the penetrance of hereditary non-polyposis colorectal cancer. Int.J.Cancer, 88: 928-931 (2000)

146. Major J, Jakab M G, Tompa A: Genotoxicological monitoring of 175 subjects living in the green belts, inner town or near chemical industrial estates in Greater Budapest agglomeration, Hungary. Mutat.Res., 412: 9-16 (1998)

147. Marti T M, Hefner E, Feeney L, Natale V, Cleaver J E: H2AX phosphorylation within the G1 phase after UV irradiation depends on nucleotide excision repair and not DNA double-strand breaks. Proc.Natl.Acad.Sci.U.S.A., 103: 9891-9896 (2006)

148. Masutani C, Kusumoto R, Yamada A, Dohmae N, Yokoi M, Yuasa M, Araki M, Iwai S, Takio K, Hanaoka F: The XPV (xeroderma pigmentosum variant) gene encodes human DNA polymerase eta. Nature, 399: 700-704 (1999)

149. Mateuca R A, Roelants M, Iarmarcovai G, Aka P V, Godderis L, Tremp A, Bonassi S, Fenech M, Berge-Lefranc J L, Kirsch-Volders M: hOGG1(326), XRCC1(399) and XRCC3(241) polymorphisms influence micronucleus frequencies in human lymphocytes in vivo. Mutagenesis, 23: 35-41 (2008)

150. McVey M, Lee S E: MMEJ repair of double-strand breaks (director's cut): deleted sequences and alternative endings. Trends Genet., 24: 529-538 (2008)

151. Meijers-Heijboer H, van den Ouweland A, Klijn J, Wasielewski M, de Snoo A, Oldenburg R, Hollestelle A, Houben M, Crepin E, van Veghel-Plandsoen M, Elstrodt F, van Duijn C, Bartels C, Meijers C, Schutte M, McGuffog L, Thompson D, Easton D, Sodha N, Seal S, Barfoot R, Mangion J, Chang-Claude J, Eccles D, Eeles R, Evans D G, Houlston R, Murday V, Narod S, Peretz T, Peto J, Phelan C, Zhang H X, Szabo C, Devilee P, Goldgar D, Futreal P A, Nathanson K L, Weber B, Rahman N, Stratton M R, CHEK2-Breast Cancer Consortium: Low-penetrance susceptibility to breast cancer due to CHEK2(*)1100delC in noncarriers of BRCA1 or BRCA2 mutations. Nat.Genet., 31: 55-59 (2002)

152. Meindl A: Comprehensive analysis of 989 patients with breast or ovarian cancer provides BRCA1 and BRCA2 mutation profiles and frequencies for the German population. Int.J.Cancer, 97: 472-480 (2002)

153. Meindl A, Hellebrand H, Wiek C, Erven V, Wappenschmidt B, Niederacher D, Freund M, Lichtner P, Hartmann L, Schaal H, Ramser J, Honisch E, Kubisch C, Wichmann H E, Kast K, Deissler H, Engel C, Muller-Myhsok B, Neveling K, Kiechle M, Mathew C G, Schindler D, Schmutzler R K, Hanenberg H: Germline mutations in breast and ovarian cancer pedigrees establish RAD51C as a human cancer susceptibility gene. Nat.Genet., 42: 410-414 (2010)

154. Miki Y, Swensen J, Shattuck-Eidens D, Futreal P A, Harshman K, Tavtigian S, Liu Q, Cochran C, Bennett L M, Ding W: A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. Science, 266: 66-71 (1994)

155. Moldovan G L, D'Andrea A D: How the fanconi anemia pathway guards the genome. Annu.Rev.Genet., 43: 223-249 (2009)

156. Morgan W F, Crossen P E: X irradiation and sister chromatid exchange in cultured human lymphocytes. Environ.Mutagen., 2: 149-155 (1980)

157. Moshous D, Callebaut I, de Chasseval R, Corneo B, Cavazzana-Calvo M, Le Deist F, Tezcan I, Sanal O, Bertrand Y, Philippe N, Fischer A, de Villartay J P: Artemis, a novel DNA double-strand break repair/V(D)J recombination protein, is mutated in human severe combined immune deficiency. Cell, 105: 177-186 (2001)

158. Murgia E, Ballardin M, Bonassi S, Rossi A M, Barale R: Validation of micronuclei frequency in peripheral blood lymphocytes as early cancer risk biomarker in a nested case-control study. Mutat.Res., 639: 27-34 (2008)

159. Norppa H, Bonassi S, Hansteen I L, Hagmar L, Stromberg U, Rossner P, Boffetta P, Lindholm C, Gundy S, Lazutka J, Cebulska-Wasilewska A, Fabianova E, Sram R J, Knudsen L E, Barale R, Fucic A: Chromosomal aberrations and SCEs as biomarkers of cancer risk. Mutat.Res., 600: 37-45 (2006)

160. Norppa H, Falck G C: What do human micronuclei contain?. Mutagenesis, 18: 221-233 (2003)

161. Obe G, Pfeiffer P, Savage J R, Johannes C, Goedecke W, Jeppesen P, Natarajan A T, Martinez-Lopez W, Folle G A, Drets M E: Chromosomal aberrations: formation, identification and distribution. Mutat.Res., 504: 17-36 (2002)

162. O'Driscoll M, Cerosaletti K M, Girard P M, Dai Y, Stumm M, Kysela B, Hirsch B, Gennery A, Palmer S E, Seidel J, Gatti R A, Varon R, Oettinger M A, Neitzel H, Jeggo P A, Concannon P: DNA ligase IV mutations identified in patients exhibiting developmental delay and immunodeficiency. Mol.Cell, 8: 1175-1185 (2001)

163. O'Driscoll M, Ruiz-Perez V L, Woods C G, Jeggo P A, Goodship J A: A splicing mutation affecting expression of ataxia-telangiectasia and Rad3-related protein (ATR) results in Seckel syndrome. Nat.Genet., 33: 497-501 (2003)

164. Olive P L, Banath J P: The comet assay: a method to measure DNA damage in individual cells. Nat.Protoc., 1: 23-29 (2006)

165. Olivieri G, Micheli A: Mitotic delay and repair in human lymphocytes. Mutat.Res., 122: 65-72 (1983)

166. Papadopoulos N, Nicolaides N C, Wei Y F, Ruben S M, Carter K C, Rosen C A, Haseltine W A, Fleischmann R D, Fraser C M, Adams M D: Mutation of a mutL homolog in hereditary colon cancer. Science, 263: 1625-1629 (1994)

167. Park J H, Wacholder S, Gail M H, Peters U, Jacobs K B, Chanock S J, Chatterjee N: Estimation of effect size distribution from genome-wide association studies and implications for future discoveries. Nat.Genet., 42: 570-575 (2010)

168. Patino-Garcia B, Hoegel J, Varga D, Hoehne M, Michel I, Jainta S, Kreienberg R, Maier C, Vogel W: Scoring variability of micronuclei in binucleated human lymphocytes in a case-control study. Mutagenesis, 21: 191-197 (2006)

169. Pearce C L, Near A M, Van Den Berg D J, Ramus S J, Gentry-Maharaj A, Menon U, Gayther S A, Anderson A R, Edlund C K, Wu A H, Chen X, Beesley J, Webb P M, Holt S K, Chen C, Doherty J A, Rossing M A, Whittemore A S, McGuire V, DiCioccio R A, Goodman M T, Lurie G, Carney M E, Wilkens L R, Ness R B, Moysich K B, Edwards R, Jennison E, Kjaer S K, Hogdall E, Hogdall C K, Goode E L, Sellers T A, Vierkant R A, Cunningham J M, Schildkraut J M, Berchuck A, Moorman P G, Iversen E S, Cramer D W, Terry K L, Vitonis A F, Titus-Ernstoff L, Song H, Pharoah P D, Spurdle A B, Anton-Culver H, Ziogas A, Brewster W, Galitovskiy V, Chenevix-Trench G, Australian Cancer Study, Australian Ovarian Cancer Study Group: Validating genetic risk associations for ovarian cancer through the international Ovarian Cancer Association Consortium. Br.J.Cancer, 100: 412-420 (2009)

170. Pelegrini A L, Moura D J, Brenner B L, Ledur P F, Maques G P, Henriques J A, Saffi J, Lenz G: Nek1 silencing slows down DNA repair and blocks DNA damage-induced cell cycle arrest. Mutagenesis, 25: 447-454 (2010)

171. Perry P, Wolff S: New Giemsa method for the differential staining of sister chromatids. Nature, 251: 156-158 (1974)

172. Pozarowski P, Darzynkiewicz Z: Analysis of cell cycle by flow cytometry. Methods Mol.Biol., 281: 301-311 (2004)

173. Qu D, Qu H, Fu M, Zhao X, Liu R, Sui L, Zhan Q: Increased expression of Nlp, a potential oncogene in ovarian cancer, and its implication in carcinogenesis. Gynecol.Oncol., 110: 230-236 (2008)

174. Radack K, Martin V, Wones R, Buncher R, Pinney S, Mandell K: Intercorrelations and sources of variability in three mutagenicity assays: a population-based study. Mutat.Res., 350: 295-306 (1996)

175. Rahman N, Seal S, Thompson D, Kelly P, Renwick A, Elliott A, Reid S, Spanova K, Barfoot R, Chagtai T, Jayatilake H, McGuffog L, Hanks S, Evans D G, Eccles D, Breast Cancer Susceptibility Collaboration (UK), Easton D F, Stratton M R: PALB2, which encodes a BRCA2-interacting protein, is a breast cancer susceptibility gene. Nat.Genet., 39: 165-167 (2007)

176. Rainey M D, Black E J, Zachos G, Gillespie D A: Chk2 is required for optimal mitotic delay in response to irradiation-induced DNA damage incurred in G2 phase. Oncogene, 27: 896-906 (2008)

177. Ramos J M, Ruiz A, Colen R, Lopez I D, Grossman L, Matta J L: DNA repair and breast carcinoma susceptibility in women. Cancer, 100: 1352-1357 (2004)

178. Rao X, Zhang Y, Yi Q, Hou H, Xu B, Chu L, Huang Y, Zhang W, Fenech M, Shi Q: Multiple origins of spontaneously arising micronuclei in HeLa cells: direct evidence from long-term live cell imaging. Mutat.Res., 646: 41-49 (2008)

179. Reid S, Schindler D, Hanenberg H, Barker K, Hanks S, Kalb R, Neveling K, Kelly P, Seal S, Freund M, Wurm M, Batish S D, Lach F P, Yetgin S, Neitzel H, Ariffin H, Tischkowitz M, Mathew C G, Auerbach A D, Rahman N: Biallelic mutations in PALB2 cause Fanconi anemia subtype FA-N and predispose to childhood cancer. Nat.Genet., 39: 162-164 (2007)

180. Renwick A, Thompson D, Seal S, Kelly P, Chagtai T, Ahmed M, North B, Jayatilake H, Barfoot R, Spanova K, McGuffog L, Evans D G, Eccles D, Breast Cancer Susceptibility Collaboration (UK), Easton D F, Stratton M R, Rahman N: ATM mutations that cause ataxia-telangiectasia are breast cancer susceptibility alleles. Nat.Genet., 38: 873-875 (2006)

181. Ridler M A, Smith G F: The response of human cultured lymphocytes to cytochalasin B. J.Cell.Sci., 3: 595-602 (1968)

182. Roberts S A, Spreadborough A R, Bulman B, Barber J B, Evans D G, Scott D: Heritability of cellular radiosensitivity: a marker of low-penetrance predisposition genes in breast cancer?. Am.J.Hum.Genet., 65: 784-794 (1999)

183. Rogakou E P, Pilch D R, Orr A H, Ivanova V S, Bonner W M: DNA double-stranded breaks induce histone H2AX phosphorylation on serine 139. J.Biol.Chem., 273: 5858-5868 (1998)

184. Ronen A, Glickman B W: Human DNA repair genes. Environ.Mol.Mutagen., 37: 241-283 (2001)

185. Roseaulin L, Yamada Y, Tsutsui Y, Russell P, Iwasaki H, Arcangioli B: Mus81 is essential for sister chromatid recombination at broken replication forks. EMBO J., 27: 1378-1387 (2008)

186. Rothfuss A, Schutz P, Bochum S, Volm T, Eberhardt E, Kreienberg R, Vogel W, Speit G: Induced micronucleus frequencies in peripheral lymphocytes as a screening test for carriers of a BRCA1 mutation in breast cancer families. Cancer Res., 60: 390-394 (0228)

187. Rowley R, Hudson J, Young P G: The wee1 protein kinase is required for radiation-induced mitotic delay. Nature, 356: 353-355 (1992)

188. Rowley R, Zorch M, Leeper D B: Effect of caffeine on radiation-induced mitotic delay: delayed expression of G2 arrest. Radiat.Res., 97: 178-185 (1984)

189. Sahin A, Tatar A, Oztas S, Seven B, Varoglu E, Yesilyurt A, Ayan A K: Evaluation of the genotoxic effects of chronic low-dose ionizing radiation exposure on nuclear medicine workers. Nucl.Med.Biol., 36: 575-578 (2009)

190. Sakiyama T, Kohno T, Mimaki S, Ohta T, Yanagitani N, Sobue T, Kunitoh H, Saito R, Shimizu K, Hirama C, Kimura J, Maeno G, Hirose H, Eguchi T, Saito D, Ohki M, Yokota J: Association of amino acid substitution polymorphisms in DNA repair genes TP53, POLI, REV1 and LIG4 with lung cancer risk. Int.J.Cancer, 114: 730-737 (2005)

191. Salinas C A, Kwon E, Carlson C S, Koopmeiners J S, Feng Z, Karyadi D M, Ostrander E A, Stanford J L: Multiple independent genetic variants in the 8q24 region are associated with prostate cancer risk. Cancer Epidemiol.Biomarkers Prev., 17: 1203-1213 (2008)

192. Sancar A, Lindsey-Boltz L A, Unsal-Kacmaz K, Linn S: Molecular mechanisms of mammalian DNA repair and the DNA damage checkpoints. Annu.Rev.Biochem., 73: 39-85 (2004)

193. Sasieni P D: From genotypes to genes: doubling the sample size. Biometrics, 53: 1253-1261 (1997)

194. Savitsky K, Bar-Shira A, Gilad S, Rotman G, Ziv Y, Vanagaite L, Tagle D A, Smith S, Uziel T, Sfez S, Ashkenazi M, Pecker I, Frydman M, Harnik R, Patanjali S R, Simmons A, Clines G A, Sartiel A, Gatti R A, Chessa L, Sanal O, Lavin M F, Jaspers N G, Taylor A M, Arlett C F, Miki T, Weissman S M, Lovett M, Collins F S, Shiloh Y: A single ataxia telangiectasia gene with a product similar to PI-3 kinase. Science, 268: 1749-1753 (1995)

195. Schmid W: Chemical mutagen testing on in vivo somatic mammalian cells. Agents Actions, 3: 77-85 (1973)

196. Scott D, Barber J B, Levine E L, Burrill W, Roberts S A: Radiation-induced micronucleus induction in lymphocytes identifies a high frequency of radiosensitive cases among breast cancer patients: a test for predisposition?. Br.J.Cancer, 77: 614-620 (1998)

197. Seal S, Thompson D, Renwick A, Elliott A, Kelly P, Barfoot R, Chagtai T, Jayatilake H, Ahmed M, Spanova K, North B, McGuffog L, Evans D G, Eccles D, Breast Cancer Susceptibility Collaboration (UK), Easton D F, Stratton M R, Rahman N: Truncating mutations in the Fanconi anemia J gene BRIP1 are low-penetrance breast cancer susceptibility alleles. Nat.Genet., 38: 1239-1241 (2006)

198. Sedelnikova O A, Rogakou E P, Panyutin I G, Bonner W M: Quantitative detection of (125)IdU-induced DNA double-strand breaks with gamma-H2AX antibody. Radiat.Res., 158: 486-492 (2002)

199. Shao S, Liu R, Wang Y, Song Y, Zuo L, Xue L, Lu N, Hou N, Wang M, Yang X, Zhan Q: Centrosomal Nlp is an oncogenic protein that is gene-amplified in human tumors and causes spontaneous tumorigenesis in transgenic mice. J.Clin.Invest., 120: 498-507 (2010)

200. Sharma S, Brosh R M,Jr: Human RECQ1 is a DNA damage responsive protein required for genotoxic stress resistance and suppression of sister chromatid exchanges. PLoS One, 2: e1297 (2007)

201. Sherry S T, Ward M H, Kholodov M, Baker J, Phan L, Smigielski E M, Sirotkin K: dbSNP: the NCBI database of genetic variation. Nucleic Acids Res., 29: 308-311 (2001)

202. Shrivastav M, De Haro L P, Nickoloff J A: Regulation of DNA double-strand break repair pathway choice. Cell Res., 18: 134-147 (2008)

203. Sigurdson A J, Hauptmann M, Chatterjee N, Alexander B H, Doody M M, Rutter J L, Struewing J P: Kin-cohort estimates for familial breast cancer risk in relation to variants in DNA base excision repair, BRCA1 interacting and growth factor genes. BMC Cancer, 4: 9 (2004)

204. Song H, Ramus S J, Kjaer S K, Hogdall E, Dicioccio R A, Whittemore A S, McGuire V, Hogdall C, Jacobs I J, Easton D F, Ponder B A, Dunning A M, Gayther S A, Pharoah P D: Tagging single nucleotide polymorphisms in the BRIP1 gene and susceptibility to breast and ovarian cancer. PLoS One, 2: e268 (2007)

205. Sonoda E, Sasaki M S, Morrison C, Yamaguchi-Iwai Y, Takata M, Takeda S: Sister chromatid exchanges are mediated by homologous recombination in vertebrate cells. Mol.Cell.Biol., 19: 5166-5169 (1999)

206. Speit G, Hanelt S, Helbig R, Seidel A, Hartmann A: Detection of DNA effects in human cells with the comet assay and their relevance for mutagenesis. Toxicol.Lett., 88: 91-98 (1996)

207. Speit G, Trenz K: Chromosomal mutagen sensitivity associated with mutations in BRCA genes. Cytogenet.Genome Res., 104: 325-332 (2004)

208. Speit G, Vogel W, Mehnert K: Do the frequencies of sister chromatid exchanges in endoreduplicated mitoses provide a measure for lesion persistence and repair?. Chromosoma, 91: 369-371 (1985)

209. Srivastava S, Zou Z Q, Pirollo K, Blattner W, Chang E H: Germ-line transmission of a mutated p53 gene in a cancer-prone family with Li-Fraumeni syndrome. Nature, 348: 747-749 (1990)

210. Stacey S N, Sulem P, Johannsson O T, Helgason A, Gudmundsson J, Kostic J P, Kristjansson K, Jonsdottir T, Sigurdsson H, Hrafnkelsson J, Johannsson J, Sveinsson T, Myrdal G, Grimsson H N, Bergthorsson J T, Amundadottir L T, Gulcher J R, Thorsteinsdottir U, Kong A, Stefansson K: The BARD1 Cys557Ser variant and breast cancer risk in Iceland. PLoS Med., 3: e217 (2006)

211. Starr T K, Allaei R, Silverstein K A, Staggs R A, Sarver A L, Bergemann T L, Gupta M, O'Sullivan M G, Matise I, Dupuy A J, Collier L S, Powers S, Oberg A L, Asmann Y W, Thibodeau S N, Tessarollo L, Copeland N G, Jenkins N A, Cormier R T, Largaespada D A: A transposon-based genetic screen in mice identifies genes altered in colorectal cancer. Science, 323: 1747-1750 (2009)

212. Stetka D G,Jr, Spahn M C: SCEs are induced by replication of BrdU-substituted DNA templates, but not by incorporation of BrdU into nascent DNA. Mutat.Res., 140: 33-42 (1984)

213. Stoepker C, Hain K, Schuster B, Hilhorst-Hofstee Y, Rooimans M A, Steltenpool J, Oostra A B, Eirich K, Korthof E T, Nieuwint A W, Jaspers N G, Bettecken T, Joenje H, Schindler D, Rouse J, de Winter J P: SLX4, a coordinator of structure-specific endonucleases, is mutated in a new Fanconi anemia subtype. Nat.Genet., (2011)

214. Stratton M R, Rahman N: The emerging landscape of breast cancer susceptibility. Nat.Genet., 40: 17-22 (2008)

215. Sugimoto M, Tahara H, Ide T, Furuichi Y: Steps involved in immortalization and tumorigenesis in human B-lymphoblastoid cell lines transformed by Epstein-Barr virus. Cancer Res., 64: 3361-3364 (2004)

216. Surowy H, Rinckleb A, Luedeke M, Stuber M, Wecker A, Varga D, Maier C, Hoegel J, Vogel W: Heritability of baseline and induced micronucleus frequencies. Mutagenesis, 26: 111-117 (2011)

217. Surralles J, Antoccia A, Creus A, Degrassi F, Peris F, Tanzarella C, Xamena N, Marcos R: The effect of cytochalasin-B concentration on the frequency of micronuclei induced by four standard mutagens. Results from two laboratories. Mutagenesis, 9: 347-353 (1994)

218. Svendsen J M, Smogorzewska A, Sowa M E, O'Connell B C, Gygi S P, Elledge S J, Harper J W: Mammalian BTBD12/SLX4 assembles a Holliday junction resolvase and is required for DNA repair. Cell, 138: 63-77 (2009)

219. Swift M, Morrell D, Massey R B, Chase C L: Incidence of cancer in 161 families affected by ataxia-telangiectasia. N.Engl.J.Med., 325: 1831-1836 (1991)

220. Tauchi H, Matsuura S, Isomura M, Kinjo T, Nakamura A, Sakamoto S, Kondo N, Endo S, Komatsu K, Nakamura Y: Sequence analysis of an 800-kb genomic DNA region on chromosome 8q21 that contains the Nijmegen breakage syndrome gene, NBS1. Genomics, 55: 242-247 (1999)

221. Taylor A M, Metcalfe J A, Thick J, Mak Y F: Leukemia and lymphoma in ataxia telangiectasia. Blood, 87: 423-438 (1996)

222. Taylor J W: Simple estimation of population attributable risk from case-control studies. Am.J.Epidemiol., 106: 260 (1977)

223. Tedeschi B, Cicchetti R, Argentin G, Caporossi D, Pittaluga M, Parisi P, Vernole P: Aphidicolin and bleomycin induced chromosome damage as biomarker of mutagen sensitivity: a twin study. Mutat.Res., 546: 55-64 (2004)

224. Tenesa A, Farrington S M, Prendergast J G, Porteous M E, Walker M, Haq N, Barnetson R A, Theodoratou E, Cetnarskyj R, Cartwright N, Semple C, Clark A J, Reid F J, Smith L A, Kavoussanakis K, Koessler T, Pharoah P D, Buch S, Schafmayer C, Tepel J, Schreiber S, Volzke H, Schmidt C O, Hampe J, Chang-Claude J, Hoffmeister M, Brenner H, Wilkening S, Canzian F, Capella G, Moreno V, Deary I J, Starr J M, Tomlinson I P, Kemp Z, Howarth K, Carvajal-Carmona L, Webb E, Broderick P, Vijayakrishnan J, Houlston R S, Rennert G, Ballinger D, Rozek L, Gruber S B, Matsuda K, Kidokoro T, Nakamura Y, Zanke B W, Greenwood C M, Rangrej J, Kustra R, Montpetit A, Hudson T J, Gallinger S, Campbell H, Dunlop M G: Genome-wide association scan identifies a colorectal cancer susceptibility locus on 11q23 and replicates risk loci at 8q24 and 18q21. Nat.Genet., 40: 631-637 (2008)

225. Trenz K, Rothfuss A, Schutz P, Speit G: Mutagen sensitivity of peripheral blood from women carrying a BRCA1 or BRCA2 mutation. Mutat.Res., 500: 89-96 (2002)

226. Tsuji H, Kojima T: Presence of abnormally high incidences of sister chromatid exchanges in three successive cell cycles in Bloom's syndrome lymphocytes. Chromosoma, 93: 87-93 (1985)

227. Turnbull C, Ahmed S, Morrison J, Pernet D, Renwick A, Maranian M, Seal S, Ghoussaini M, Hines S, Healey C S, Hughes D, Warren-Perry M, Tapper W, Eccles D, Evans D G, Breast Cancer Susceptibility Collaboration (UK), Hooning M, Schutte M, van den Ouweland A, Houlston R, Ross G, Langford C, Pharoah P D, Stratton M R, Dunning A M, Rahman N, Easton D F: Genome-wide association study identifies five new breast cancer susceptibility loci. Nat.Genet., 42: 504-507 (2010)

228. Unal E, Heidinger-Pauli J M, Koshland D: DNA double-strand breaks trigger genome-wide sister-chromatid cohesion through Eco1 (Ctf7). Science, 317: 245-248 (2007)

229. UniProt Consortium: The Universal Protein Resource (UniProt) in 2010. Nucleic Acids Res., 38: D142-D148 (2010)

230. Vahteristo P, Syrjakoski K, Heikkinen T, Eerola H, Aittomaki K, von Smitten K, Holli K, Blomqvist C, Kallioniemi O P, Nevanlinna H: BARD1 variants Cys557Ser and Val507Met in breast cancer predisposition. Eur.J.Hum.Genet., 14: 167-172 (2006)

231. Vahteristo P, Yliannala K, Tamminen A, Eerola H, Blomqvist C, Nevanlinna H: BACH1 Ser919Pro variant and breast cancer risk. BMC Cancer, 6: 19 (2006)

232. van der Burgt I, Chrzanowska K H, Smeets D, Weemaes C: Nijmegen breakage syndrome. J.Med.Genet., 33: 153-156 (1996)

233. Varga D, Hoegel J, Maier C, Jainta S, Hoehne M, Patino-Garcia B, Michel I, Schwarz-Boeger U, Kiechle M, Kreienberg R, Vogel W: On the difference of micronucleus frequencies in peripheral blood lymphocytes between breast cancer patients and controls. Mutagenesis, 21: 313-320 (2006)

234. Varga D, Johannes T, Jainta S, Schuster S, Schwarz-Boeger U, Kiechle M, Patino G B, Vogel W: An automated scoring procedure for the micronucleus test by image analysis. Mutagenesis, 19: 391-397 (2004)

235. Varga D, Michel I, Patino-Garcia B, Paiss T, Vogel W, Maier C: Radiosensitivity detected by the micronucleus test is not generally increased in sporadic prostate cancer patients. Cytogenet.Genome Res., 111: 41-45 (2005)

236. Vaz F, Hanenberg H, Schuster B, Barker K, Wiek C, Erven V, Neveling K, Endt D, Kesterton I, Autore F, Fraternali F, Freund M, Hartmann L, Grimwade D, Roberts R G, Schaal H, Mohammed S, Rahman N, Schindler D, Mathew C G: Mutation of the RAD51C gene in a Fanconi anemia-like disorder. Nat.Genet., 42: 406-409 (2010)

237. Vogel F, Motulsky A G: Appendix 5: heritability estimates from twin data. In: Human genetics: problems and approaches, Bd 3rd rev. ed., Springer, Berlin, New York, S. 769-773 (1997)

238. Vogel W, Bauknecht T: Differential chromatid staining by in vivo treatment as a mutagenicity test system. Nature, 260: 448-449 (1976)

239. Vogel W, Surowy H: Reduced DNA repair in BRCA1 mutation carriers undetectable before onset of breast cancer?. Br.J.Cancer, 97: 1184-1186 (2007)

240. Vral A, Thierens H, Baeyens A, De Ridder L: The micronucleus and G2-phase assays for human blood lymphocytes as biomarkers of individual sensitivity to ionizing radiation: limitations imposed by intraindividual variability. Radiat.Res., 157: 472-477 (2002)

241. Walker J A, Boreham D R, Unrau P, Duncan A M: Chromosome content and ultrastructure of radiation-induced micronuclei. Mutagenesis, 11: 419-424 (1996)

242. Wang L, Wei Q, Shi Q, Guo Z, Qiao Y, Spitz M R: A modified host-cell reactivation assay to measure repair of alkylating DNA damage for assessing risk of lung adenocarcinoma. Carcinogenesis, 28: 1430-1436 (2007)

243. Wang M, Wang M, Zhang W, Yuan L, Fu G, Wei Q, Zhang Z: Common genetic variants on 8q24 contribute to susceptibility to bladder cancer in a Chinese population. Carcinogenesis, 30: 991-996 (2009)

244. Wei Q, Spitz M R, Gu J, Cheng L, Xu X, Strom S S, Kripke M L, Hsu T C: DNA repair capacity correlates with mutagen sensitivity in lymphoblastoid cell lines. Cancer Epidemiol.Biomarkers Prev., 5: 199-204 (1996)

245. Wolff S, Perry P: Differential Giemsa staining of sister chromatids and the study of chromatid exchanges without autoradiography. Chromosoma, 48: 341-353 (1974)

246. Wood R D, Mitchell M, Lindahl T: Human DNA repair genes, 2005. Mutat.Res., 577: 275-283 (2005)

247. Wooster R, Bignell G, Lancaster J, Swift S, Seal S, Mangion J, Collins N, Gregory S, Gumbs C, Micklem G: Identification of the breast cancer susceptibility gene BRCA2. Nature, 378: 789-792 (1995)

248. World Health Organization: The global burden of disease: 2004 update, WHO Press, Geneva, Switzerland, S. 7-26 (2008)

249. Wu X, Spitz M R, Amos C I, Lin J, Shao L, Gu J, de Andrade M, Benowitz N L, Shields P G, Swan G E: Mutagen sensitivity has high heritability: evidence from a twin study. Cancer Res., 66: 5993-5996 (2006)

250. Xia B, Sheng Q, Nakanishi K, Ohashi A, Wu J, Christ N, Liu X, Jasin M, Couch F J, Livingston D M: Control of BRCA2 cellular and clinical functions by a nuclear partner, PALB2. Mol.Cell, 22: 719-729 (2006)

251. Xie J, Litman R, Wang S, Peng M, Guillemette S, Rooney T, Cantor S B: Targeting the FANCJ-BRCA1 interaction promotes a switch from recombination to poleta-dependent bypass. Oncogene, 29: 2499-2508 (2010)

252. Xu B, Kim S T, Lim D S, Kastan M B: Two molecularly distinct G(2)/M checkpoints are induced by ionizing irradiation. Mol.Cell.Biol., 22: 1049-1059 (2002)

253. Yarden R I, Pardo-Reoyo S, Sgagias M, Cowan K H, Brody L C: BRCA1 regulates the G2/M checkpoint by activating Chk1 kinase upon DNA damage. Nat.Genet., 30: 285-289 (2002)

254. Yi P N, Evans H H, Beer J Z, Rha C K: Relationships between mitotic delay and the dose rate of X radiation. Radiat.Res., 140: 387-392 (1994)

255. Yin Q, McBride J, Fewell C, Lacey M, Wang X, Lin Z, Cameron J, Flemington E K: MicroRNA-155 is an Epstein-Barr virus-induced gene that modulates Epstein-Barr virus-regulated gene expression pathways. J.Virol., 82: 5295-5306 (2008)

256. Yu C E, Oshima J, Fu Y H, Wijsman E M, Hisama F, Alisch R, Matthews S, Nakura J, Miki T, Ouais S, Martin G M, Mulligan J, Schellenberg G D: Positional cloning of the Werner's syndrome gene. Science, 272: 258-262 (1996)

257. Zhang C, Naftalis E, Euhus D: Carcinogen-Induced DNA Double Strand Break Repair in Sporadic Breast Cancer. J.Surg.Res., 135: 120-128 (2006)

258. Zhang J, Shi X, Li Y, Kim B J, Jia J, Huang Z, Yang T, Fu X, Jung S Y, Wang Y, Zhang P, Kim S T, Pan X, Qin J: Acetylation of Smc3 by Eco1 is required for S phase sister chromatid cohesion in both human and yeast. Mol.Cell, 31: 143-151 (2008)

259. Zhang Y, Rohde L H, Emami K, Hammond D, Casey R, Mehta S K, Jeevarajan A S, Pierson D L, Wu H: Suppressed expression of non-DSB repair genes inhibits gamma-radiation-induced cytogenetic repair and cell cycle arrest. DNA Repair (Amst), 7: 1835-1845 (2008)

260. Zheng Y L, Loffredo C A, Alberg A J, Yu Z, Jones R T, Perlmutter D, Enewold L, Krasna M J, Yung R, Shields P G, Harris C C: Less efficient g2-m checkpoint is associated with an increased risk of lung cancer in African Americans. Cancer Res., 65: 9566-9573 (2005)

261. Zijno A, Andreoli C, Leopardi P, Marcon F, Rossi S, Caiola S, Verdina A, Galati R, Cafolla A, Crebelli R: Folate status, metabolic genotype, and biomarkers of genotoxicity in healthy subjects. Carcinogenesis, 24: 1097-1103 (2003)

262. Zijno A, Saini F, Crebelli R: Suitability of cryopreserved isolated lymphocytes for the analysis of micronuclei with the cytokinesis-block method. Mutagenesis, 22: 311-315 (2007)

263. Zweig M H, Campbell G: Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. Clin.Chem., 39: 561-577 (1993)

Attachment

A: Official gene symbols and full names according to the Human Genome Organization (HUGO) Gene Nomenclature Consortium.

APEX1	APEX nuclease (multifunctional DNA repair enzyme) 1
ATM	ataxia telangiectasia mutated
ATR	ataxia telangiectasia and Rad3 related
BARD1	BRCA1-associated RING domain 1
BLM	Bloom syndrome, RecQ helicase-like
BRCA1	breast cancer 1, early onset
BRCA2	breast cancer 2, early onset
BRIP1	BRCA1 interacting protein C-terminal helicase 1
C17ORF70	chromosome 17 open reading frame 70
C19ORF40	chromosome 19 open reading frame 40
CHEK1	CHK1 checkpoint homolog (S. pombe)
CHEK2	CHK2 checkpoint homolog (S. pombe)
DCLRE1C	DNA cross-link repair 1C
EME1	essential meiotic endonuclease 1 homolog 1 (S. pombe)
ERG	v-ets erythroblastosis virus E26 oncogene homolog (avian)
ERCC genes	excision repair cross-complementing rodent repair deficiency, complementation groups
ESCO1	establishment of cohesion 1 homolog 1 (S. cerevisiae)
FANC genes	Fanconi anemia complementation groups
FBXO5	F-box protein 5
H2AX	H2A histone family, member X
HLTF	helicase-like transcription factor
LIG4	ligase IV, DNA, ATP-dependent
MUS81	MUS81 endonuclease homolog (S. cerevisiae)
NBN	nibrin
NEK1	NIMA (never in mitosis gene a)-related kinase 1
NINL	ninein-like
PALB2	Partner and localizer of BRCA2
PLK1	polo-like kinase 1
POLH	polymerase (DNA directed) eta
POLI	polymerase (DNA directed) iota
RAD51C	RAD51 homolog C (S. cerevisiae)
REV1	REV1 homolog (S. cerevisiae)
SLX4	SLX4 structure-specific endonuclease subunit homolog (S. cerevisiae)
TERF2	telomeric repeat binding factor 2
TERF2IP	telomeric repeat binding factor 2, interacting protein
TIMELESS	timeless homolog (Drosophila)
TMPRSS2	transmembrane protease, serine 2
TP53	tumor protein p53
XP genes	xeroderma pigmentosum, complementation groups
WRN	Werner syndrome, RecQ helicase-like
WEE1	WEE1 homolog (S. pombe)
WRAP53	WD repeat containing, antisense to TP53
XRCC genes	X-ray repair complementing defective repair in Chinese hamster cells

B: Classifier for the scoring of binucleated cells and micronuclei in short-time whole blood cultures with the Metafer MSearch 4 software

Classifier Name:	STC
Classifier Description:	short-time whole-blood cultures

Nuclei parameters

Nuclei: Image Processing Operations:	Sharpen(3,4)
Nuclei: Object Threshold (in %):	18
Nuclei: Minimum Area (in 1/100 µm²):	6000
Nuclei: Maximum Area (in 1/100 µm²):	50000
Nuclei: Maximum Relative Concavity Depth (in 1/1000):	90
Nuclei: Maximum Aspect Ratio (in 1/1000) :	1500
Nuclei: Maximum Distance (in 1/10 µm):	250
Nuclei: Maximum Area Asymmetry (in %):	90
Nuclei: Region of Interest Radius (in 1/10 µm):	300
Nuclei: Maximum Object Area in ROI (in 1/100 µm²):	5000

Micronuclei parameters

Micronuclei: Image Processing Operations:	MedianV(3) MedianH(3) Average(3,1) Sharpen(5,5)
Micronuclei: Object Threshold (in %):	5
Micronuclei: Minimum Area (in 1/100 µm²):	400
Micronuclei: Maximum Area (in 1/100 µm²):	6000
Micronuclei: Maximum Relative Concavity Depth (in 1/1000):	1000
Micronuclei: Maximum Aspect Ratio (in 1/1000):	1720
Micronuclei: Maximum Distance (in 1/10 µm):	400

C: Parameters for the analysis of the cell cycle with the Partec CCA Flow Cytometer

Channel resolution:	1024
Photomultiplier gain:	380 – 420
Total cell count maximum:	25,000
Total count volume maximum:	5 ml
Lower Level (channel):	40
Upper Level (channel):	999
Flow Rate (cells / s):	< 100
Speed (µl / s):	0.5-3 (optimal flow rates were varying between samples)

D: Raw genotype counts in cases and controls

Table 35: Absolute counts of non-carriers (NN), heterozygous carriers (NV) and homozygous carriers (VV) of the genotyped single nucleotide polymorphisms number (SNP Nr.) in the controls and cases used in the association tests. The respective genotyping efficiency is shown. n: number of probands.

SNP Nr.	Controls (n = 1189)			Cases (n = 435)			Genotyping Efficiency %
	NN	NV	VV	NN	NV	VV	
1	960	211	18	343	87	5	100,00
2	1053	129	5	389	45	1	99,88
3	948	230	11	349	83	3	100,00
4	1094	95	0	390	44	1	100,00
5	1066	122	0	392	43	0	99,94
6	1163	26	0	425	10	0	100,00
7	1043	141	5	389	46	0	100,00
8	877	297	15	332	101	2	100,00
9	1078	108	3	373	61	1	100,00
10	1047	137	5	390	44	1	100,00
11	1002	178	9	375	56	4	100,00
12	936	227	17	354	74	6	99,38
13	1084	104	1	396	36	3	100,00
14	924	244	21	341	87	7	100,00
15	1084	104	1	395	40	0	100,00
16	1189	0	0	435	0	0	100,00
17	1090	93	1	410	23	1	99,63
18	1132	57	0	418	16	0	99,94
19	1136	43	0	417	16	0	99,26
20	1147	41	1	420	14	1	100,00
21	405	574	208	150	207	78	99,88
22	1120	68	1	403	31	0	99,94
23	1089	81	0	407	24	2	98,71

Ich möchte herzlich den folgenden Personen danken:

Den **Patienten**, **Kontrollprobanden** und **Zwillingen** für ihre Teilnahme am DNA-Reparatur Projekt.

Prof Dr. Walther Vogel für die Ermöglichung und die Betreuung der Doktorarbeit am Institut für Humangenetik sowie für den Einsatz für das Gelingen dieser Arbeit, die fachliche und persönliche Unterstützung, und die anregenden Gespräche und Diskussionen während meiner Zeit als Doktorand.

Herr PD. Dr. Markus Cronauer für die freundliche Übernahme der Zweitgutachterschaft für die Doktorarbeit.

Frau PD Dr. Christiane Maier für die Unterstützung und die Anregungen, die sie mir während dieser der Doktorarbeit stets zukommen ließ, und dafür dass sie immer ein offenes Ohr für meine Anliegen hatte. Dies hat wesentlich zum Gelingen beigetragen.

Herrn Prof. Dr. Josef Högel für seine Geduld und seine Bereitschaft, mir die Geheimnisse der Mathematik im Allgemeinen und der Statistik im Speziellen näher zu bringen, sowie den übrigen Arbeitsgruppenleitern des Institutes für Humangenetik für ihre zahlreichen Anregungen und Vorschläge.

Meinen langjährigen Doktoranden-Mitstreitern **Manuel Lüdeke**, **Antje Rinckleb** und **Silvia Kastler** für ihre Freundschaft, den Spaß und die vielen Unternehmungen, sowie für so manche fachliche Anregung.

Antje Kollak, Margot Brugger, Bärbel Weber, Irina Wiest, Petra Reutter und Birgit Schmoll für die hervorragende technische Unterstützung bei den diversen Projekten, Regina Heidenreich für die Hilfe in Verwaltungsangelegenheiten, Petra Schütz und Heidrun Götz für die Hilfestellung in der Zellkultur, und allen hier nicht genannten Personen, die mir stets mit Rat und Tat zur Seite standen.

Den medizinischen Doktoranden und biologischen Diplomanden, die ich während der Bearbeitung der vorliegenden Doktorarbeit betreut habe, sowie den zahlreichen Praktikanten für die hervorragende Zusammenarbeit.

Den übrigen Doktoranden und Diplomanden der Humangenetik, die während meiner Zugehörigkeit zum Institut für die schöne Zeit und den kollegialen Zusammenhalt in- und außerhalb des Labors gesorgt haben.

Meinem Zwillingsbruder Oliver dafür, dass er mir stets einen Anlaufpunkt jenseits des akademischen Tellerrandes bot, für die Ermutigungen und für so manche lange Nächte.

Meiner Mutter, für die umfassende Unterstützung während meiner Zeit als Student und Doktorand, die das Zustandekommen dieser Doktorarbeit erst ermöglicht hat, sowie bei meiner Familie die mich stets auf meinem Wege unterstützt hat.

Meiner Partnerin, ohne die ich nicht einmal ein halber Mensch wäre, dafür dass sie immer ein offenes Ohr für meine Anliegen hatte, ich stets die kleinen Erfolge und Ernüchterungen mit ihr teilen konnte und dass sie mir besonders während der Zeit des Schreibens so großartig zur Seite stand.