



# DNA Repair Gene Polymorphisms and Chromosomal Aberrations in Exposed Populations

Yasmeen Niazi<sup>1,2,3\*</sup>, Hauke Thomsen<sup>1,4</sup>, Bozena Smolkova<sup>5</sup>, Ludmila Vodickova<sup>6,7,8</sup>, Sona Vodenkova<sup>6</sup>, Michal Kroupa<sup>6,8</sup>, Veronika Vymetalkova<sup>6,7,8</sup>, Alena Kazimirova<sup>9</sup>, Magdalena Barancokova<sup>9</sup>, Katarina Volkovova<sup>9</sup>, Marta Staruchova<sup>9</sup>, Per Hoffmann<sup>10,11</sup>, Markus M. Nöthen<sup>10</sup>, Maria Dusinska<sup>12</sup>, Ludovit Musak<sup>13</sup>, Pavel Vodicka<sup>6,7,8</sup>, Kari Hemminki<sup>1,8,14\*†</sup> and Asta Försti<sup>1,2,3†</sup>

## OPEN ACCESS

### Edited by:

Stefano Gnan,  
Institut Curie, France

### Reviewed by:

Vladimir Druzhinin,  
Kemerovo State University, Russia  
Alexander V. Rubanovich,  
Vavilov Institute of General Genetics,  
Russian Academy of Sciences, Russia

### \*Correspondence:

Yasmeen Niazi  
y.niazi@kitz-heidelberg.de  
Kari Hemminki  
K.Hemminki@dkfz-heidelberg.de

†These authors share senior  
authorship

### Specialty section:

This article was submitted to  
Human and Medical Genomics,  
a section of the journal  
Frontiers in Genetics

Received: 07 April 2021

Accepted: 30 April 2021

Published: 16 June 2021

### Citation:

Niazi Y, Thomsen H, Smolkova B,  
Vodickova L, Vodenkova S,  
Kroupa M, Vymetalkova V,  
Kazimirova A, Barancokova M,  
Volkovova K, Staruchova M,  
Hoffmann P, Nöthen MM, Dusinska M,  
Musak L, Vodicka P, Hemminki K and  
Försti A (2021) DNA Repair Gene  
Polymorphisms and Chromosomal  
Aberrations in Exposed Populations.  
Front. Genet. 12:691947.  
doi: 10.3389/fgene.2021.691947

<sup>1</sup> Department of Molecular Genetic Epidemiology, German Cancer Research Center (DKFZ), Heidelberg, Germany, <sup>2</sup> Hopp Children's Cancer Center (KITZ), Heidelberg, Germany, <sup>3</sup> Division of Pediatric Neurooncology, German Cancer Research Center (DKFZ), German Cancer Consortium (DKTK), Heidelberg, Germany, <sup>4</sup> GeneWerk GmbH, Heidelberg, Germany, <sup>5</sup> Department of Molecular Oncology, Cancer Research Institute, Biomedical Research Center of the Slovak Academy of Sciences, Bratislava, Slovakia, <sup>6</sup> Department of Molecular Biology of Cancer, Institute of Experimental Medicine, Czech Academy of Sciences, Prague, Czechia, <sup>7</sup> First Faculty of Medicine, Institute of Biology and Medical Genetics, Charles University, Prague, Czechia, <sup>8</sup> Faculty of Medicine and Biomedical Center in Pilsen, Charles University in Prague, Prague, Czechia, <sup>9</sup> Department of Biology, Faculty of Medicine, Slovak Medical University, Bratislava, Slovakia, <sup>10</sup> Institute of Human Genetics, School of Medicine and University Hospital Bonn, University of Bonn, Bonn, Germany, <sup>11</sup> Division of Medical Genetics, Department of Biomedicine, University of Basel, Basel, Switzerland, <sup>12</sup> Health Effects Laboratory, Department of Environmental Chemistry, NILU-Norwegian Institute for Air Research, Kjeller, Norway, <sup>13</sup> Jessenius Faculty of Medicine, Biomedical Center Martin, Comenius University in Bratislava, Bratislava, Slovakia, <sup>14</sup> Division of Cancer Epidemiology, German Cancer Research Centre (DKFZ), Heidelberg, Germany

DNA damage and unrepaired or insufficiently repaired DNA double-strand breaks as well as telomere shortening contribute to the formation of structural chromosomal aberrations (CAs). Non-specific CAs have been used in the monitoring of individuals exposed to potential carcinogenic chemicals and radiation. The frequency of CAs in peripheral blood lymphocytes (PBLs) has been associated with cancer risk and the association has also been found in incident cancer patients. CAs include chromosome-type aberrations (CSAs) and chromatid-type aberrations (CTAs) and their sum CA<sub>tot</sub>. In the present study, we used data from our published genome-wide association studies (GWASs) and extracted the results for 153 DNA repair genes for 607 persons who had occupational exposure to diverse harmful substances/radiation and/or personal exposure to tobacco smoking. The analyses were conducted using linear and logistic regression models to study the association of DNA repair gene polymorphisms with CAs. Considering an arbitrary cutoff level of  $5 \times 10^{-3}$ , 14 loci passed the threshold, and included 7 repair pathways for CTA, 4 for CSA, and 3 for CA<sub>tot</sub>; 10 SNPs were eQTLs influencing the expression of the target repair gene. For the base excision repair pathway, the implicated genes *PARP1* and *PARP2* encode poly(ADP-ribosyl) transferases with multiple regulatory functions. *PARP1* and *PARP2* have an important role in maintaining genome stability through diverse mechanisms. Other candidate genes with known roles for CSAs included *GTF2H* (general transcription factor IIIH

subunits 4 and 5), Fanconi anemia pathway genes, and *PMS2*, a mismatch repair gene. The present results suggest pathways with mechanistic rationale for the formation of CAs and emphasize the need to further develop techniques for measuring individual sensitivity to genotoxic exposure.

**Keywords:** chromosomal aberrations, association study, DNA repair, exposure, polymorphism

## INTRODUCTION

Human cancers are often associated with chromosomal instability with complex numerical and structural chromosomal aberrations (CAs), which may be causative events in the process of malignant transformation (Futreal et al., 2004; Rajagopalan and Lengauer, 2004; Mitelman et al., 2007; Burrell et al., 2013). Structural CAs may be specific, such as translocations and inversions, or non-specific, such as chromatid breaks, fragmented or missing parts of chromosomes, and fusions resulting in dicentric and ring chromosomes (Bignold, 2009). The former are often recurrent and they are currently analyzed by molecular cytogenetic methods while the latter are scored by classical cytogenetic techniques, which are able to recognize chromosome-type aberrations (CSAs) and chromatid-type aberrations (CTAs) according to morphological changes (Hagmar et al., 2004). CTAs are formed due to insufficiently repaired double-strand breaks (DSBs) during the late S or G2 phase of the cell cycle (Natarajan and Palitti, 2008; Bignold, 2009; Durante et al., 2013), whereas CSAs are the result of direct DNA damage due to radiation, chemical mutagens, or shortening of telomeres during the G0/G1 phase (Albertini et al., 2000; Jones et al., 2012). Non-specific CAs have been used in the monitoring of populations occupationally exposed to potential carcinogenic chemicals and radiation and an increased frequency of CAs in peripheral blood lymphocytes (PBLs) has been associated with cancer risk and the association has also been found in incident cancer patients (Rossner et al., 2005; Vodicka et al., 2010; Vodenkova et al., 2015).

Unrepaired or insufficiently repaired DSBs, as well as telomerase dysfunction, represent the mechanistic bases for the formation of structural CAs (Natarajan and Palitti, 2008; Bignold, 2009; Durante et al., 2013; Vodicka et al., 2018; Srinivas et al., 2020). However, even other types of DNA repair pathways may contribute to CA formation as these are found in inherited syndromes manifesting DNA repair gene mutations (Rahman, 2014). Eukaryotic cells have four conserved but distinct pathways of DSB repair: non-homologous DNA end joining (NHEJ), alternate end joining (a-EJ), homologous recombination (HR), and single-strand annealing (SSA) (Sung, 2018). In non-malignant cells, the majority of DSBs are removed *via* either NHEJ or HR, with minor contribution of a-EJ and SSA. Repair *via* HR may be error-free while the three other DSB repairs are error-prone, particularly the rare a-EJ and SSA. Repair errors emerge as mutations and CAs with smaller or larger DNA sequence losses. The role of telomerase dysfunction has been emerging more recently, with growing evidence that shorter telomeres are associated with increased frequency of CAs, particularly of the CSA type (Li et al., 2013;

Hemminki et al., 2015). Telomeres become shorter at each round of replication and critically shortened telomeres may be poorly end-capped and may be recognized as DSBs by repair machinery that may result as CAs (Maser and DePinho, 2002; Meeker et al., 2004; Gostissa et al., 2011; Jones and Jallepalli, 2012; Maciejowski et al., 2015). It has been shown that telomere shortening is associated with a decreased capacity to repair DSBs in multiple types of cancer (Kroupa et al., 2017).

In the present study, we used data from our published genome-wide association studies (GWASs) (Niazi et al., 2018, 2019) and extracted the results for 153 DNA repair genes to find out the association between CA frequency and DNA repair pathways. The population was occupationally exposed to diverse harmful substances/radiation and/or personally exposed to tobacco smoking. The analyses were conducted for the types of CAs (CA<sub>tot</sub>, CSAs, and CTAs) using linear and logistic regression models.

## MATERIALS AND METHODS

Our cohort comprised 607 individuals recruited from the Czech and Slovak Republics. The subjects were investigated for chromosomal abnormalities in previous occupational exposure-related epidemiological studies or as regular medical monitoring in factories with exposure to genotoxic compounds. These studies involved individuals with defined exposure to small organic compounds, heavy metals, radiations, and asbestos and other mineral fibers as well as unexposed controls (Vodicka et al., 2004a,b; Dusinska et al., 2004a,b, 2012; Musak et al., 2008; Kazimirova et al., 2009). Prior to blood sampling, study subjects were informed according to the rules of Helsinki declaration and written approval was obtained. Ethics Committees of the Slovak Medical University, the Jessenius Faculty of Medicine, the Comenius University Bratislava, the Institute for Clinical and Experimental Medicine in Slovakia, and the Thomayer Hospital and the General University Hospital in the Czech Republic approved the study design.

The study population (**Table 1**) contained about 60% males and 40% females. All individuals included in the study were either exposed to genotoxic compounds due to their occupation and/or they were smokers. About half of the individuals (52.1%) had a history of occupational exposure to genotoxic organic compounds while 12.7% were exposed to heavy metals, mineral fibers, and low levels of radiations. All subjects filled a questionnaire listing beside the type of job and periods of exposure other exogenous factors such as smoking, radiation exposure, and dietary dispositions. About 66% of the individuals included in the study were smokers. Age of the participants

**TABLE 1** | Descriptive attributes of the study cohort and exposure-based distribution.

Study cohort		Covariate effect ( <i>P</i> -value) <sup>e</sup>	
Age (years)	Median	43	0.56
	Range	19–80	
Gender (%)	Females	40.5	0.05
	Males	59.5	
Smoking status (%)	Smokers	66.1	5.55E-05
	Non-smokers	33.9	
Occupational exposure (n)	Small organic compounds	316	2.42E-05
	Heavy metals	6	
	Radiation (pilots)	6	
	Asbestos	19	
	Stone wool	28	
	Glass fibers	18	
	Others <sup>a</sup>	214	
No. of individuals with	High CATot freq <sup>b</sup>	342	
	Low CATot freq	265	
	High CTA freq <sup>c</sup>	345	
	Low CTA freq	262	
	High CSA freq <sup>d</sup>	321	
	Low CSA freq	286	

<sup>a</sup>Office workers and blood donors who were reported as smokers.

<sup>b</sup>High CATot freq =  $\geq 2$  CAs/100 cells.

<sup>c</sup>High CTA freq =  $\geq 1$  CA/100 cells.

<sup>d</sup>High CSA freq =  $\geq 1$  CA/100 cells.

<sup>e</sup>*P*-value indicates the association of the covariates (age, gender, smoking status, and occupational exposure) with CAs.

ranged from 19 to 80 years with a median age of 43 years. Cytogenetic analysis was done in PBLs that were stimulated to grow and cultured for 48 h (Vodicka et al., 2010). About 100 mitoses per person were evaluated to score the frequency of CSAs and CTAs and they were summed up to CATot (i.e., CSA + CTA = CATot).

For GWAS genotyping, Illumina HumanOmniExpressExome8v1.3 chip arrays were used and the quality control (QC) criteria were implemented according to the predetermined benchmarks (Niazi et al., 2018, 2019). Samples were included on the basis of successful genotyping  $\geq 95\%$ . Duplicates and related individuals were excluded by identity-by-state (IBS) score. Population outliers determined by the principal component analysis were removed. After prephasing with SHAPEIT v2.12 (Delaneau et al., 2011), imputation was performed using UK10K (Walter et al., 2015) and 1,000 genomes (phase 3, October 2014) (1000 Genomes Project Consortium, Auton et al., 2015) as reference panels with IMPUTE2 v2.3.2 software (Howie et al., 2011). Prior to analysis, SNPs were filtered according to call rate ( $< 95\%$ ), Hardy–Weinberg equilibrium (HWE) ( $P < 1.0 \times 10^{-5}$ ), minor allele frequency (MAF) ( $< 0.05$ ), and imputation quality (Info  $< 0.70$ ).

Association analysis between CA frequency and SNPs in DNA repair genes was conducted using PLINK version 1.90b3.30 (Purcell et al., 2007) using logistic (binary) and linear regression

analyses on three phenotypes CATot, CSAs, and CTAs. For binary logistic regression analysis, individuals were divided into high and low CA frequency groups. For CATot analysis, individuals with  $\geq 2\%$  CAs were included in the high-frequency group, while for CSAs and CTAs, the threshold for inclusion into the high-frequency group was  $\geq 1\%$  (Dusinska et al., 2004a; Vodicka et al., 2010). The analyses were adjusted for gender, age, smoking status, and occupational exposure. GWAS summary statistics were then used for our gene-based study that included a list of 170 DNA repair genes (Wood et al., 2001, 2005; Friedberg et al., 2006; Lange et al., 2011; **Table 2**). For these genes, coordinates were extracted from UCSC genome browser's hg19 assembly, which gave a list of genes with chromosome number and transcription start and end position. Genes on the X chromosome were excluded from the analysis as well as those with no match found in NCBI RefSeq list, leaving 153 genes for the analysis (Rosenbloom et al., 2015). On the basis of the gene coordinates, a region including the gene of interest with 100 kb upstream and 100 kb downstream regions was selected, and all the SNPs in this window were analyzed. In total, about 40,000 SNPs from the repair genes' regions were analyzed, with about 2000 independent loci among them as determined by using PLINK's linkage disequilibrium-based pruning. These regions were plotted in LocusZoom (Pruim et al., 2010) and SNPs with *P*-value  $5 \times 10^{-3}$  or below were further studied for their capacity to influence the functional aspects of the corresponding DNA repair genes. This threshold was set to only select the SNPs above the background level of association in the analysis. *In silico* tools utilized in this analysis were Haploreg, GTex, and RegulomDB 2.0 (Ward and Kellis, 2012; Boyle et al., 2012; GTex Consortium., 2013). These were used to ascertain linkage disequilibrium (LD) between the SNPs from the same locus identified by different phenotypic analysis as well as location [intergenic, 3' and 5' untranslated regions (UTRs)], intronic or expression quantitative trait locus (eQTL, minimal *P*-value of  $10^{-5}$ ), and effect (synonymous, missense, and non-sense) of the genetic variation. Regulome DB version2.0 provided chromatin state, information about changed motifs, transcription factors, and DNase accessibility.

## RESULTS

We identified 14 independent loci associated with CA frequency from six different analyses (two regression models, namely, linear and logistic for each of the three phenotypes, CATot, CSA, and CTA) below the applied cutoff, *P*-value  $5 \times 10^{-3}$ ; note that the *REV3L* (REV3 like, DNA-directed polymerase zeta catalytic subunit) SNP was detected by both the linear and logistic models in CTA analysis (**Table 3**). All the SNPs that remained after cutoff *P*-value  $5 \times 10^{-3}$  in all phenotypes' logistic and linear models are given in **Supplementary Material**. If one would consider the analysis of 153 genes, and assume one association per gene, the Bonferroni type of corrected significance level would have a *P*-value of  $3.2 \times 10^{-4}$ . SNPs that remained significantly associated with CAs after applying this criterion are indicated in bold in **Table 3**. Among CATot

**TABLE 2** | Total 153 studied genes grouped based on DNA repair type (genes where the SNPs were associated with CAs are in bold letters).

Base excision repair (BER)	Other BER and strand break joining factors	Poly(ADP-ribose) polymerase (PARP) enzymes that bind to DNA	Direct reversal of damage	Repair of DNA-protein crosslinks	Mismatch excision repair (MMR)	Nucleotide excision repair (NER)	Nucleotide excision repair (NER)	Homologous recombination (HR)	Non-homologous end-joining (NHEJ)	Fanconi anemia	DNA polymerases (catalytic subunits)	Editing and processing nucleases	Ubiquitination and modification	Chromatin structure	Genes defective in diseases associated with sensitivity to DNA damaging agents	Other identified genes with known or suspected DNA repair function	Other conserved DNA damage response genes
UNG	APEX1 (APE1)	<b>PARP1</b> (ADPRT)	<b>MGMT</b>	TDP1	MSH2	CDK7	XPC	RAD51	XRCC6 (Ku70)	FANCA	POLB	FEN1 (DNase IV)	UBE2B (RAD6B)	H2AFX (H2AX)	BLM	DCLRE1A (SNM1)	ATR
SMUG1	LIG3	<b>PARP2</b> (ADPRTL2)	ALKBH2 (ABH2)		MSH3	CCNH	RAD23B	DMC1	XRCC5 (Ku80)	<b>FANCC</b>	POLG	FAN1 (MTMR15)	RAD18	CHAF1A (CAF1)	WRN	DCLRE1B (SNM1B)	<b>MDC1</b>
MBD4	XRCC1	<b>PARP3</b> (ADPRTL3)	ALKBH3 (DEPC1)		MSH6	MINAT1	RAD23A	XRCC2	PRKDC	BRCA2 (FANCD1)	POLD1	TREX1 (DNase III)	SHPRH	SETMAR (METNASE)	RECQL4	RECQL (RECQ1)	RAD1
TDG	PNKP				MLH1	ERCC5 (XPG)	XPA	XRCC3	LIG4	<b>FANCD2</b>	POLE	EXO1 (HEX1)	HLTF (SMARCA3)		ATM	RECQL5	RAD9A
OGG1	APLF (C2ORF13)				<b>PMS2</b>	ERCC1	DDB1	RAD52	<b>XRCC4</b>	FANCE	PCNA	APTX (apratatin)	RNF168		TTDN1 (C7orf11)	HELQ (HEL308)	HUS1
MUTYH (MYH)					MSH4	ERCC4 (XPF)	DDB2 (XPE)	RAD54L	DCLRE1C (Artemis)	FANCF	<b>REV3L (POLZ)</b>	SPO11	RNF8			RDM1 (RAD52B)	RAD17 (RAD24)
NTHL1 (NTH1)					MSH5	LIG1	RPA1	RAD54B	NHEJ1 (XLF, Cernunnos)	FANCG (XRCC9)	MAD2L2 (REV7)	FLJ35220 (ENDOV)	RNF4				CHEK1
MPG					MLH3	ERCC8 (CSA)	RPA2	BRCA1	NUDT1 (MTH1)	FANCI (KIAA1794)	POLH		UBE2V2 (MMS2)				CHEK2
NEIL1					PMS1	ERCC6 (CSB)	RPA3	RAD50	DUT	<b>BRIP1 (FANCI-J)</b>	POLI (RAD30B)		UBE2N (UBC13)				TP53
NEIL2						XAB2 (HCNP)	ERCC3 (XPB)	NBN (NBS1)	RRM2B (p53R2)	FANCL	POLQ						<b>TP53BP1 (53BP1)</b>
<b>NEIL3</b>						MMS19L (MMS19)	ERCC2 (XPD)	RBBP8 (CtIP)		FANCM	POLK (DINB1)						ATRIP
						GTF2H1	GTF2H3	MUS81			PALB2 (FANCD1)						TOPBP1
						GTF2H2	<b>GTF2H4</b>				RAD51C (FANCO)						CLK2
							<b>GTF2H5 (TTDA)</b>				FAAP24 (C19orf40)		POLN (POL4P)				PER1

**TABLE 3 |** SNP associations with  $P$ -value  $\leq 5 \times 10^{-3}$  from logistic and linear regression analyses of three CA types (CA<sub>tot</sub>, CTA, and CSA).

CA <sub>tot</sub> -logistic	DNA repair Gene	Type of DNA repair	SNP	CHR	BP	A1	OR	95% CI	P	In silico
	<i>GTF2H4</i>	Nucleotide excision repair (NER)	<b>rs3130780</b>	<b>6</b>	<b>30874308</b>	<b>T</b>	<b>1.89</b>	<b>1.36–2.64</b>	<b>1.77E-04</b>	1.7 kb 5' of <i>GTF2H4</i>
	<i>PARP1</i>	Base excision repair (BER) PARP enzymes	rs1341334	1	226605024	G	1.56	1.21–2.00	5.16E-04	9.2 kb 5' of <i>PARP1</i> /eQTL
CA <sub>tot</sub> -linear	DNA repair Gene	Type of DNA repair	SNP	CHR	BP	A1	Beta	95% CI	P	In silico
	<i>MGMT</i>	Direct reversal of DNA damage	rs12247555	10	131370520	C	0.09	0.03–0.15	2.78E-03	Intronic/eQTL
CTA-logistic	DNA repair Gene	Type of DNA repair	SNP	CHR	BP	A1	OR	95% CI	P	In silico
	<i>NEIL3</i>	Base excision repair (BER)	rs10009807	4	178229925	A	0.69	0.54–0.89	4.62E-03	1.1 kb 5' of <i>NEIL3</i> /histone marks
	<i>REV3L</i>	DNA polymerases (catalytic subunits)	rs7742724	6	111839019	A	1.63	1.23–2.16	6.42E-04	eQTL
	<i>BRIP1</i>	Fanconi anemia	rs17542001	17	59915590	C	0.64	0.48–0.85	1.86E-03	Intronic/eQTL
CTA-linear	DNA repair Gene	Type of DNA repair	SNP	CHR	BP	A1	Beta	95% CI	P	In silico
	<i>FANCC</i>	Fanconi anemia	rs13292454	9	97995075	A	0.16	0.05–0.28	3.93E-03	Intronic
	<i>MDC1</i>	Conserved DNA damage response genes	rs3094090	6	30669956	C	0.19	0.07–0.32	2.33E-03	Intronic
	<i>REV3L</i>	DNA polymerases (catalytic subunits)	<b>rs7742724</b>	<b>6</b>	<b>111839019</b>	<b>A</b>	<b>0.16</b>	<b>0.09–0.24</b>	<b>2.47E-05</b>	eQTL
	<i>XRCC4</i>	Non-homologous end-joining (NHEJ)	rs301286	5	82602955	C	-0.14	(-0.22)–(-0.05)	1.42E-03	Intronic/histone marks
	<i>TP53BP1</i>	Conserved DNA damage response genes	rs28702649	15	43648629	T	0.12	0.05–0.19	1.31E-03	eQTL
CSA-logistic	DNA repair Gene	Type of DNA repair	SNP	CHR	BP	A1	OR	95% CI	P	In silico
	<i>GTF2H5</i>	Nucleotide excision repair (NER)	<b>rs1744178</b>	<b>6</b>	<b>158496856</b>	<b>T</b>	<b>1.74</b>	<b>1.31–2.31</b>	<b>1.50E-04</b>	eQTL
	<i>PARP2</i>	Base excision repair (BER) PARP enzymes	rs2318861	14	20758949	G	0.54	0.38–0.78	9.85E-04	eQTL
CSA-linear	DNA repair Gene	Type of DNA repair	SNP	CHR	BP	A1	Beta	95% CI	P	In silico
	<i>FANCD2</i>	Fanconi anemia	rs61429272	3	10037320	C	0.15	0.05–0.24	4.13E-03	eQTL
	<i>PMS2</i>	Mismatch excision repair (MMR)	rs12702464	7	6041506	C	-0.13	(-0.21)–(-0.04)	4.64E-03	Intronic/eQTL

ORs (in logistic regression analysis), Beta values (in linear regression analysis), and their corresponding P-values and in silico predictions are shown. SNP associations that survived Bonferroni correction for multiple testing are marked in bold. SNP single nucleotide polymorphism; CHR chromosome; OR odds ratios; A1 The allele for which beta and OR are calculated.

associated loci, SNP rs3130780 from logistic regression analysis had a P-value that was significant according to such a correction (OR 1.89, 95% CI 1.36–2.64, P-value  $1.77 \times 10^{-4}$ ). The SNP is located 1.7 kb 5' to *GTF2H4* (general transcription factor IIH subunit 4), which belongs to the nucleotide excision repair (NER) pathway. In the same analysis for gene *PARP1* [poly(ADP-ribose polymerase 1), rs1341334 at 1q42.12 with OR 1.56, 95% CI 1.21–2.00, and P-value  $5.16 \times 10^{-4}$  also

came close to the significance threshold. From the linear regression analysis for CA<sub>tot</sub>, no significant association was identified and the only SNP above the background level was an intronic variant in *MGMT* (O6-methylguanine-DNA methyltransferase) gene. This gene is involved in the direct reversal of DNA damage.

A SNP marking the gene *REV3L* was found to be associated with the CTA phenotype in both linear and binary logistic

regression models. This variant, rs7742724, exhibited a notable association in the linear model with  $\beta$  0.16, 95% CI 0.09–0.24, and  $P$ -value  $2.47 \times 10^{-5}$  and a similar tendency in the binary model but with an elevated  $P$ -value of  $6.42 \times 10^{-4}$ . The other variants from the CTA analysis had  $P$ -values ranging from  $1.31 \times 10^{-3}$  to  $4.62 \times 10^{-3}$ . These included intronic SNPs in the genes *BRIP1* (BRCA1 interacting protein C-terminal helicase 1), *FANCC* (Fanconi anemia complementation group C), *MDC1* (mediator of DNA damage checkpoint 1), and *XRCC4* (X-ray repair cross-complementing protein 4) as well as in *TP53BP1* (TP53-binding protein 1). *BRIP1* and *FANCC* belong to the Fanconi anemia pathway while *MDC1* and *TP53BP1* are conserved DNA damage response genes. *XRCC4* is an NHEJ gene. A variant rs10009807 located at 1.1 kb 5' to *NEIL3* (nei like DNA glycosylase 3), which is a base excision repair (BER) pathway gene, was also among the associations identified from the CTA analysis.

The CSA group presented a total of four associations from both models; SNP at 6q25.3 and eQTL to *GTF2H5* (general transcription factor IIIH subunit 5) was the best candidate with OR 1.74, 95% CI 1.31–2.31, and  $P$ -value  $1.50 \times 10^{-4}$ . *GTF2H5* is a member of the NER pathway. rs2318861, an eQTL SNP, for *PARP2* [poly(ADP-ribose) polymerase 2], which ADP-ribosylates DNA by acting on terminal phosphates at DNA strand breaks, had a  $P$ -value of  $9.85 \times 10^{-4}$ . In the linear regression analysis for CSAs, two associations included a Fanconi anemia gene *FANCD2* (Fanconi anemia complementation group D2) and an intronic variant for mismatch repair (MMR) pathway gene *PMS2* (*PMS1* homolog 2, mismatch repair system component).

In **Table 4**, the candidate SNPs were annotated using RegulomeDB and GTex; some eQTL data were also retrieved using Haploreg. SNPs linked to genes *PARP1*, *NEIL3*, *FANCC*, *XRCC4*, and *FANCD2* show DNase accessibility in blood and all the selected variants were located in either the region of strong transcription or in transcription start sites (TSSs) and enhancers in blood and many other tissues. The eQTLs summarized in **Table 4** each target the linked DNA repair gene. The SNP linked to *PARP1* was a strong eQTL in cultured fibroblasts at  $9 \times 10^{-23}$ , and the one linked to the *MGMT* gene was a strong eQTL in the whole blood at  $1.4 \times 10^{-33}$ . The SNPs associated with *MDC1*, *GTF2H5*, *PARP2*, and *FANCD2* were eQTLs in the whole blood/cultured fibroblasts. The SNP for *PMS2* was an eQTL in the aorta, whereas those linked to genes *REV3L* and *BRIP1* were eQTLs in the brain and the tibial nerve, respectively.

## DISCUSSION

Genetic variation can be the cause of inter-individual differences in susceptibility to CAs and susceptibility to cancer (Vodicka et al., 2004a). In a previous analysis of 11 DNA repair genes in mixed population of occupationally exposed and unexposed individuals, associations with CAs were found with *XPD* and *RAD54L* polymorphisms (Vodicka et al., 2015). In the present study, SNPs from a total of 153 DNA repair genes were tested on an exposed population of 607 individuals. It can be expected that DNA repair is more critical in persons exposed to high apparent exposure vs. background environmental exposure and

the distribution of CAs has been shown to be skewed to higher damage levels in the exposed population (Niazi et al., 2019). This population difference together with a more stringent significance threshold in the present study might be the reason for the different outcomes of these two studies. While CSAs and CTAs are assumed to be independent markers of damage, arising at different phases of the cell cycle, CATot is a composite measure as the sum of CSAs and CTAs. For the presentation of the results, we selected an arbitrary cutoff level of  $5 \times 10^{-3}$ , which appeared to be stringent as only 14 SNPs passed the threshold and, with one exception, these results from the linear and logistic models were different. We considered the Bonferroni type of adjusted  $P$ -value as  $3.2 \times 10^{-4}$  based on the 153 genes tested; even though many more independent LD regions were considered, the sample size for rarer SNPs afforded a limited power (with a MAF of 10%, only six homozygous variants were to be expected). Credibility to the findings is supported by the chromatin state and eQTL data. All SNPs were located at a site of strong transcription, enhancer, or TSS, and five SNPs were located at DNase-accessible sites in blood. Ten of the 14 candidate SNPs influenced the expression of the target DNA repair gene, and for 5, the data were obtained from whole blood.

Most positive associations at  $P$ -value below  $5 \times 10^{-3}$  were found for CTAs ( $N = 7$ ), followed by four for CSAs and three for CATot. SNPs for the NER pathway emerged two times, for BER and Fanconi anemia repair pathways three times, while SNPs in other pathways were unique. For the BER pathway, the implicated SNPs were eQTLs to the target genes *PARP1* and *PARP2*, two homologs encoding chromatin-associated enzymes, poly(ADP-ribosyl)transferases, which modify various nuclear proteins by poly(ADP-ribosylation) with multiple downstream regulatory functions (Azarm and Smith, 2020). *PARP1* may be associated with xeroderma pigmentosum, complementation group A through interactions with XPA, and the related susceptibility to skin cancer. *PARP1* is involved in the synthesis of telomere C-strand (Azarm and Smith, 2020). *PARP2* has partially overlapping biochemical functions with *PARP1*. *PARP1* and *PARP2* function in both single- and double-strand DNA repair, and they have an important role in maintaining genome stability through diverse mechanisms. PARP inhibitors are being used as anticancer agents in *BRCA1/2* mutated cancers (Boussios et al., 2020).

The NER pathway genes *GTF2H4* and *GTF2H5* encode different subunits of general transcription factor IIIH and both associations were highly significant (Rimel and Taatjes, 2018; Hill and Theos, 2019; Kolesnikova et al., 2019). The proteins share structural and functional homology and they are associated with NER enzymes XPB and XPD (Kolesnikova et al., 2019). Syndromes associated with *GTF2H4/5* include xeroderma pigmentosum, complementation groups B and D, Cockayne syndrome, and trichothiodystrophy, all of which are characterized by extreme sensitivity to ultraviolet radiation and development of other sun-related problems such as excessive freckling and skin cancer. These syndromes as well as Fanconi anemia germline mutations display genomic instability and CAs (Chan and Ngeow, 2017; Hill and Theos, 2019). The related SNP targeted *GTF2H5* as an eQTL.

**TABLE 4** | Regulome DB 2.0/GTex *in silico* analysis of associated variants.

Chr	SNP	Gene	Accessibility (DNase and FAIRE) in tissues/cell types	Chromatin state	Motifs	eQTL			ChIP
						Tissue	Normalized effect size*	P-value	
6	rs3130780	<i>GTF2H4</i>	–	Blood + 102	BCL6, MEF2A	–	–	–	ZNF792, MIXL1
1	rs1341334	<i>PARP1</i>	Blood (K562) + 9 other tissues	Enhancer in blood	BCL6, NANOG	Cultured fibroblasts	–0.31	9.00E-23	ZFX,ZNF770
10	rs12247555	<i>MGMT</i>	–	Blood + 16	–	Whole blood	–0.29	1.40E-33	–
4	rs10009807	<i>NEIL3</i>	Blood, B cells, T cells	Enhancer in blood + 30	–	–	–	–	POLR2A
17	rs17542001	<i>BRIP1</i>	–	Blood + 62	–	Nerve–tibial	–0.41	7.30E-15	–
9	rs13292454	<i>FANCC</i>	Blood + 3 others	Blood + 48	FOXJ3	–	–	–	–
6	rs3094090	<i>MDC1</i>	Tibial nerve	Blood + 123	IRF3	Whole blood	0.13	6.40E-06	ZNF664
6	rs7742724	<i>REV3L</i>	Mammary glands	Blood + 23	–	Brain–Cerebellum	0.28	1.40E-06	SMARCA4 #
5	rs301286	<i>XRCC4</i>	Blood + 6 others	Active TSS in blood + 6	–	–	–	–	STAT5A#, STAT3#, TBP# + 10
15	rs28702649	<i>TP53BP1</i>	Placenta, H9, OCI-LY7	Blood + 109	SOX1	Cultured fibroblasts	–0.13	4.10E-08	CTCF#, RAD21, ZBTB33#
6	rs1744178	<i>GTF2H5</i>	H7-hESC, Lower leg	Blood + 118	–	*Whole blood	*	2.8783E-17	–
14	rs2318861	<i>PARP2</i>	–	Blood + 124	–	*Whole blood	*	6.61974E-07	ZBTB40
3	rs61429272	<i>FANCD2</i>	Naïve B cell + 9 other tissues	Blood + 99	–	Whole blood	–0.17	2.70E-09	–
7	rs12702464	<i>PMS2</i>	–	Blood + 124	–	Artery–aorta	–0.44	4.50E-12	–

\*eQTL values from Westra et al. (2013).

#In blood.

In summary, the present study on DNA repair gene polymorphisms in a healthy population with occupational and personal genotoxic exposures revealed SNP associations with CA frequency at the *P*-value level of  $5 \times 10^{-3}$  within 14 different genes, many of which with key roles in maintaining genomic integrity and thus plausibly associated with mechanisms leading to CAs. More than half of the implicated SNPs were eQTLs to the target DNA repair genes. Although the recent interest in measuring random CAs has decreased because of cumbersome techniques, the present results suggest that the results may have understandable mechanistic links. If the current techniques cannot be improved, there will be a need to provide alternative approaches for measuring individual sensitivity to genotoxic exposure that may lead to increased risk of cancer.

## DATA AVAILABILITY STATEMENT

All the relevant data presented in the study is given in **Supplementary Material**, further inquiries can be directed to the corresponding author/s.

## ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Ethics Committees of the Slovak Medical University, the Jessenius Faculty of Medicine, the Comenius University Bratislava, the Institute for Clinical and Experimental Medicine in Slovakia and the Thomayer Hospital and the General University Hospital in the Czech Republic. The patients/participants provided their written informed consent to participate in this study.

## AUTHOR CONTRIBUTIONS

AF, KH, YN designed the study. HT and YN analyzed the data. YN and BS performed the genotyping. PV, LV, LM, MD, and BS provided the samples. SV, MK, VV, KV, AK, and MS collected subject information and performed cytogenetic analysis. PH and MN were responsible for the GWAS. AF, KH, PV, LV, MD, and BS critically revised the manuscript. YN wrote the first draft. KH and AF supervised the study. All authors contributed to manuscript revision, read, and approved the submitted version.

## FUNDING

In the Czech Republic, the KH was supported by the European Union's Horizon 2020 Research and Innovation Programme, Grant No. 856620 (Chaperon), the National Science Foundation (18-09709S, 19-10543S), Charles University in Prague (PROGRE Q 28), Medical Faculty in Pilsen, Charles University in Prague, National Sustainability Programme I (Nr.LO 1503), Charles University Research Centre program (UNCE/MED/006), and European Commission contracts (QLK4-CT-1999-01629, ERBICI 15-CT96-1012 and CIPA-CT94-0129). In Slovakia, support came from Slovak Grant Agency (APVT-21 013202, APVT-21-017704), Ministry of Health, Slovak Republic (2005/43-SZU-21, 2006/07-SZU-02 MZ SR, 2005/42-SZU-20), SZU and Competence Center for Research

## REFERENCES

- 1000 Genomes Project Consortium, Auton, A., Abecasis, G. R., Altshuler, D. M., Durbin, R. M., Abecasis, G. R., et al. (2015). A global reference for human genetic variation. *Nature* 526:68. doi: 10.1038/nature15393
- Albertini, R. J., Anderson, D., Douglas, G. R., Hagmar, L., Hemminki, K., Merlo, F., et al. (2000). IPCS guidelines for the monitoring of genotoxic effects of carcinogens in humans. International Programme on Chemical Safety. *Mutat. Res.* 463, 111–172. doi: 10.1016/s1383-5742(00)00049-1
- Azarm, K., and Smith, S. (2020). Nuclear PARPs and genome integrity. *Genes Dev.* 34, 285–301. doi: 10.1101/gad.334730.119
- Bignold, L. P. (2009). Mechanisms of clastogen-induced chromosomal aberrations: a critical review and description of a model based on failures of tethering of DNA strand ends to strand-breaking enzymes. *Mutat. Res.* 681, 271–298. doi: 10.1016/j.mrrev.2008.11.004
- Boussios, S., Abson, C., Moschetta, M., Rassy, E., Karathanasi, A., Bhat, T., et al. (2020). Poly (ADP-Ribose) polymerase inhibitors: Talazoparib in ovarian cancer and beyond. *Drugs R D* 20, 55–73.
- Boyle, A. P., Hong, E. L., Hariharan, M., Cheng, Y., Schaub, M. A., Kasowski, M., et al. (2012). Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res.* 22, 1790–1797. doi: 10.1101/gr.137323.112
- Burrell, R. A., McClelland, S. E., Endesfelder, D., Groth, P., Weller, M. C., Shaikh, N., et al. (2013). Replication stress links structural and numerical cancer chromosomal instability. *Nature* 494, 492–496. doi: 10.1038/nature11935
- Chan, S. H., and Ngeow, J. (2017). Germline mutation contribution to chromosomal instability. *Endocr. Relat. Cancer* 24, T33–T46.
- Delaneau, O., Marchini, J., and Zagury, J. F. (2011). A linear complexity phasing method for thousands of genomes. *Nat. Methods* 9, 179–181. doi: 10.1038/nmeth.1785
- Durante, M., Bedford, J. S., Chen, D. J., Conrad, S., Cornforth, M. N., Natarajan, A. T., et al. (2013). From DNA damage to chromosome aberrations: joining the break. *Mutat. Res.* 756, 5–13. doi: 10.1016/j.mrgentox.2013.05.014
- Dusinska, M., Barancokova, M., Kazimirova, A., Harrington, V., Volkovova, K., Staruchova, M., et al. (2004a). Does occupational exposure to mineral fibres cause DNA or chromosome damage? *Mutat. Res.* 553, 103–110. doi: 10.1016/j.mrfmmm.2004.06.029
- Dusinska, M., Collins, A., Kazimirova, A., Barancokova, M., Harrington, V., Volkovova, K., et al. (2004b). Genotoxic effects of asbestos in humans. *Mutat. Res.* 553, 91–102. doi: 10.1016/j.mrfmmm.2004.06.027
- Dusinska, M., Staruchova, M., Horska, A., Smolkova, B., Collins, A., Bonassi, S., et al. (2012). Are glutathione S transferases involved in DNA damage signalling? interactions with DNA damage and repair revealed from molecular epidemiology studies. *Mutat. Res.* 736, 130–137. doi: 10.1016/j.mrfmmm.2012.03.003
- and Development in the Field of Diagnostics and Therapy of Oncological Diseases Slovakia (ITMS code: 26220220153), project Biomedical Center Martin, Slovakia co-financed from EU sources (ITMS code: 26220220187), the project “Carcinogenic and toxic metals in working environment” co-financed by EU sources and the European Regional Development Fund, Slovakia (ITMS: 26220220111), and the Research and Development Support Agency Slovakia (APVV-15-0217).

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2021.691947/full#supplementary-material>



- Mitelman, F., Johansson, B., and Mertens, F. (2007). The impact of translocations and gene fusions on cancer causation. *Nat. Rev. Cancer* 7, 233–245.
- Musak, L., Soucek, P., Vodickova, L., Naccarati, A., Halasova, E., Polakova, V., et al. (2008). Chromosomal aberrations in tire plant workers and interaction with polymorphisms of biotransformation and DNA repair genes. *Mutat. Res.* 641, 36–42.
- Natarajan, A. T., and Palitti, F. (2008). DNA repair and chromosomal alterations. *Mutat. Res.* 657, 3–7.
- Niazi, Y., Thomsen, H., Smolkova, B., Vodickova, L., Vodenkova, S., Kroupa, M., et al. (2018). Genetic variation associated with chromosomal aberration frequency: a genome-wide association study. *Environ. Mol. Mutagen.* 60, 17–28.
- Niazi, Y., Thomsen, H., Smolkova, B., Vodickova, L., Vodenkova, S., Kroupa, M., et al. (2019). Distinct pathways associated with chromosomal aberration frequency in a cohort exposed to genotoxic compounds compared to general population. *Mutagenesis* 34, 323–330.
- Pruim, R. J., Welch, R. P., Sanna, S., Teslovich, T. M., Chines, P. S., Gliedt, T. P., et al. (2010). LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics* 26, 2336–2337.
- Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., et al. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81, 559–575.
- Rahman, N. (2014). Realizing the promise of cancer predisposition genes. *Nature* 505, 302–308.
- Rajagopalan, H., and Lengauer, C. (2004). Aneuploidy and cancer. *Nature* 432, 338–341.
- Rimel, J. K., and Taatjes, D. J. (2018). The essential and multifunctional TFIIF complex. *Protein Sci.* 27, 1018–1037.
- Rosenbloom, K. R., Armstrong, J., Barber, G. P., Casper, J., Clawson, H., Diekhans, M., et al. (2015). The UCSC genome browser database: 2015 update. *Nucleic Acids Res.* 43, D670–D681.
- Rossner, P., Boffetta, P., Ceppi, M., Bonassi, S., Smerhovsky, Z., Landa, K., et al. (2005). Chromosomal aberrations in lymphocytes of healthy subjects and risk of cancer. *Environ. Health Perspect.* 113, 517–520.
- Srinivas, N., Rachakonda, S., and Kumar, R. (2020). Telomeres and telomere length: a general overview. *Cancers (Basel)* 12:558.
- Sung, P. (2018). Introduction to the thematic minireview series: DNA double-strand break repair and pathway choice. *J. Biol. Chem.* 293, 10500–10501.
- Vodenkova, S., Polivkova, Z., Musak, L., Smerhovsky, Z., Zoubkova, H., Sytarova, S., et al. (2015). Structural chromosomal aberrations as potential risk markers in incident cancer patients. *Mutagenesis* 30, 557–563.
- Vodicka, P., Kumar, R., Stetina, R., Musak, L., Soucek, P., Haufroid, V., et al. (2004a). Markers of individual susceptibility and DNA repair rate in workers exposed to xenobiotics in a tire plant. *Environ. Mol. Mutagen.* 44, 283–292.
- Vodicka, P., Musak, L., Frank, C., Kazimirova, A., Vymetalkova, V., Barancokova, M., et al. (2015). Interactions of DNA repair gene variants modulate chromosomal aberrations in healthy subjects. *Carcinogenesis* 36, 1299–1306.
- Vodicka, P., Musak, L., Vodickova, L., Vodenkova, S., Catalano, C., Kroupa, M., et al. (2018). Genetic variation of acquired structural chromosomal aberrations. *Mutat. Res. Genet. Toxicol. Environ. Mutagen.* 836(Pt A), 13–21.
- Vodicka, P., Polivkova, Z., Sytarova, S., Demova, H., Kucerova, M., Vodickova, L., et al. (2010). Chromosomal damage in peripheral blood lymphocytes of newly diagnosed cancer patients and healthy controls. *Carcinogenesis* 31, 1238–1241.
- Vodicka, P., Tuimala, J., Stetina, R., Kumar, R., Manini, P., Naccarati, A., et al. (2004b). Cytogenetic markers, DNA single-strand breaks, urinary metabolites, and DNA repair rates in styrene-exposed lamination workers. *Environ. Health Perspect.* 112, 867–871.
- Walter, K., Min, J. L., Huang, J., Crooks, L., Memari, Y., McCarthy, S., et al. (2015). The UK10K project identifies rare variants in health and disease. *Nature* 526, 82–90.
- Ward, L. D., and Kellis, M. (2012). HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Res.* 40, D930–D934.
- Westra, H.-J., Peters, M. J., Esko, T., Yaghootkar, H., Schurmann, C., Kettunen, J., et al. (2013). Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nat. Genet.* 45, 1238–1243. doi: 10.1038/ng.2756
- Wood, R. D., Mitchell, M., and Lindahl, T. (2005). Human DNA repair genes, 2005. *Mutat. Res.* 577, 275–283.
- Wood, R. D., Mitchell, M., Sgouros, J., and Lindahl, T. (2001). Human DNA repair genes. *Science* 291, 1284–1289.

**Conflict of Interest:** HT was an employee at GeneWerk GmbH, which had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Niazi, Thomsen, Smolkova, Vodickova, Vodenkova, Kroupa, Vymetalkova, Kazimirova, Barancokova, Volkovova, Staruchova, Hoffmann, Nöthen, Dusinska, Musak, Vodicka, Hemminki and Försti. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.