*Mutagenesis*, 2025, **40**, 30–38
https://doi.org/10.1093/mutage/geae006
Advance access publication 29 February 2024
**Review**

OXFORD

# Review

# Genetic and environmental associations of nonspecific chromosomal aberrations

**Kari Hemminki[1,2,*], Yasmeen Niazi[3,4], Ludmila Vodickova[1,5,6], Pavel Vodicka[1,5,6], Asta Försti[3,4]**

[1]Faculty of Medicine and Biomedical Center in Pilsen, Charles University, Pilsen 32300, Czech Republic
[2]Division of Cancer Epidemiology, German Cancer Research Centre (DKFZ), Heidelberg 69120, Germany
[3]Hopp Children's Cancer Center (KiTZ), Heidelberg, Germany
[4]Division of Pediatric Neurooncology, German Cancer Research Center (DKFZ), German Cancer Consortium (DKTK), Heidelberg, Germany
[5]Department of Molecular Biology of Cancer, Institute of Experimental Medicine of the Czech Academy of Sciences, Videnska 1083, Prague 14200, Czech Republic
[6]Institute of Biology and Medical Genetics, Charles University, Albertov 4, Prague 12800, Czech Republic

[*]Corresponding author. Faculty of Medicine and Biomedical Center in Pilsen, Charles University in Prague, Pilsen 32300, Czech Republic.
E-mail: k.hemminki@dkfz.de

## Abstract

Nonspecific structural chromosomal aberrations (CAs) are found in around 1% of circulating lymphocytes from healthy individuals but the frequency may be higher after exposure to carcinogenic chemicals or radiation. CAs have been used in the monitoring of persons exposed to genotoxic agents and radiation. Previous studies on occupationally exposed individuals have shown associations between the frequency of CAs in peripheral blood lymphocytes and subsequent cancer risk. The cause for CA formation is believed to be unrepaired or insufficiently repaired DNA double-strand breaks or other DNA damage, and additionally telomere shortening. CAs include chromosome (CSAs) and chromatid type aberrations (CTAs). In the present review, we first describe the types of CAs, the conventional techniques used for their detection and some aspects of interpreting the results. We then focus on germline genetic variation in the frequency and type of CAs measured in a genome-wide association study in healthy individuals in relation to occupational and smoking-related exposure compared to nonexposed referents. The associations (at $P < 10^{-5}$) on 1473 healthy individuals were broadly classified in candidate genes from functional pathways related to DNA damage response/repair, including *PSMA1*, *UBR5*, *RRM2B*, *PMS2P4*, *STAG3L4*, *BOD1*, *COPRS*, and *FTO;* another group included genes related to apoptosis, cell proliferation, angiogenesis, and tumorigenesis, *COPB1, NR2C1, COPRS, RHOT1*, *ITGB3, SYK,* and *SEMA6A;* a third small group mapped to genes *KLF7, SEMA5A* and *ITGB3* which were related to autistic traits, known to manifest frequent CAs. Dedicated studies on 153 DNA repair genes showed associations for some 30 genes, the expression of which could be modified by the implicated variants. We finally point out that monitoring of CAs is so far the only method of assessing cancer risk in healthy human populations, and the use of the technology should be made more attractive by developing automated performance steps and incorporating artificial intelligence methods into the scoring.

**Keywords:** chromosomal damage; cancer; genetics; DNA repair; double-strand break

## Introduction

Hanahan and Weinberg have described the hallmarks of cancer and defined genome instability and mutation as one of the underlying mechanisms expediting the acquisition of such hallmarks [1]. Genomic instability leads to chromosomal aberrations (CAs) which are established risk factors of cancer [2–4]. CAs are biological endpoints that reflect the effect of mutagens on the genome [5]. CAs, in individuals exposed to genotoxins environmentally or due to lifestyle choices, have been used in cancer biology since the latter half of the last century after the discovery of the normal karyotype of somatic cells in 1956 by Tjio and Levan, discovery of the Philadelphia chromosome in 1960 by Novell and Hungerford and of chromosome banding in 1968 by Caspersson [6–9]. There are only a few direct methods of measuring the extent of human exposure to genotoxins. Cytogenetic analysis of peripheral blood lymphocytes for CAs is one of the methods that enables the direct measurement of gross changes occurring in DNA due to genotoxins by analyzing changes in chromosomes within the cell under a light microscope [10]. It has been estimated that the attributable proportion of high frequencies of CAs for overall cancer risk may be as high as 0.25, indicating that CAs in peripheral blood lymphocytes (PBLs) could be used as a surrogate endpoint of cancer risk in human biomonitoring [11,12]. Importantly, it has been shown that the association of CAs with cancer risk is independent of exposure assessment [2]. Much of the supporting evidence for the predictive value of CAs was generated already 2 decades ago and many cellular and molecular techniques have been introduced since then. Yet no other biomarker has emerged for general cancer risk assessment applicable to healthy subjects from the general population with a reasonable

Received 21 September 2023; accepted 28 February 2024

attributable proportion. Application of CA biomonitoring in healthy subjects has been technically demanding and it has not undergone improvements that would enable large-scale studies. Thus, the applications in *in vivo* monitoring of exposure to genotoxic chemicals have become fewer [13–15]. However, in the area of biological exposure monitoring to ionizing radiation analysis of CAs, particularly dicentric chromosomes and translocations, has remained a standard protocol [16–18].

In the present review, we focus on germline genetic variation influencing the frequency of unspecific CAs in healthy individuals in relation to occupational chemical and smoking-related exposure [19]. We specifically focus on the genome-wide association studies that we have conducted summarizing the results. We are not reviewing the largely old literature on biomonitoring of CA in occupational and environmental settings. We finally discuss what is known about genetic make-up influencing CA frequency, and any prospects to revitalize a future application of unspecific CAs in exposure monitoring and individual cancer risk assessment. As background information, we start by describing the types of CAs and the techniques for their measurement. Else the review is limited to genetic and environmental associations of nonspecific CAs and will not cover multiple causes of CAs including the novel findings related to telomere dysfunction, nuclear envelope fragile compartmentalization and the role of micronuclei [20–22].

## Types of chromosomal aberrations

CAs can be classified in many ways, considering the number of chromosomes (**numerical CAs**) or chromosomal structure (**structural CAs**). Numerical changes include changes in the number of chromosomes, aneuploidy and polyploidy. Structural CAs involve rearrangements through displacement, loss or gain of chromosomal segments, in which the normal sequence integrity has been disrupted. Structural CAs can involve only a few nucleotides or much larger segments [23]. These include translocations, deletions, insertions, inversions, breaks, sister chromatid exchange, micronuclei, chromothripsis, and changes in telomere length. Structural CAs can be divided into **symmetrical (or stable)** or **asymmetrical (or unstable).** Asymmetrical CAs include dicentric chromosomes, fragment, and ring chromosomes which lead to uneven segregation in mitosis (thus the term unstable) [16–18].

### Clonal CAs

are specific recurrent aberrations including translocations and inversions that have gained a growth advantage. These can be detected by molecular cytogenetic techniques such as FISH (fluorescent *in situ* hybridization) and DNA and RNA sequencing techniques [9,24]. They have been defined in the literature as "chromosome aberration which can be detected at least twice within 20 to 40 randomly examined mitotic figures" [25]. With this criterion, the frequency of clonal CAs should be higher than 5%–10% but usually, the researchers consider CAs as clonal when the frequencies are >30% [25]. These clonal CAs are strongly associated with distinct tumor types and are an initial event in oncogenesis [4]. By 2019, the number of known fusion genes in the Mitelman Database of Chromosome Aberrations and Gene Fusions in Cancer exceeded 20 000, the rapid increase in the numbers has been due the application of DNA/RNA sequencing techniques [26]. The current number (August 2023) is 70 000 (Mitelman Database Chromosome Aberrations and Gene Fusions in Cancer (isb-cgc.org)). Recurrent rearrangements have an important clinical role in disease classification and prognostic assessment, which started in hematological neoplasms but has later been extended to patients with solid tumors [27,28]. Recent data suggests that fusion genes arise as a stochastic process and many of them have no clinical meaning which is in line that fusion genes are found also in healthy tissue [26]. However, even if the random formation of fusion genes may be the overall rule, genetic control has been shown for translocation 11;14, which is common in hematological neoplasms [29].

### Nonclonal/nonspecific CAs

can be distinguished at metaphase with the help of classical cytogenetic techniques as nonspecific aberrations. These CAs do not clonally expand and they may remain in the cells for their lifetime [11]. Nonclonal CAs have been ignored in favor of disease-related clonal CAs, and for a long time, these were considered to be genetic noise. However, nonclonal CAs make up for a great majority of chromosomal changes and have long been observed in normal and disease conditions and are important to measure system instability. As a new form of genetic information, i.e. system inheritance, nonclonal CAs are not background noise but rather a basis for heterogeneity in genome and precondition in different diseases including cancer [18].

### Chromosome type aberrations (CSAs)

are aberrations that involve both chromatids of a chromosome (Fig. 1a). CSAs are formed in peripheral blood lymphocytes (PBLs) in resting G0 phase by the agents that produce double-strand breaks. These agents include ionizing radiations and chemical clastogens, such as alkylating agents, that have an effect similar to ionizing radiations. In G0–G1 lymphocytes, the formation of CSAs is mainly due to the repair of double-strand breaks (DSBs) by nonhomologous DNA end joining, or nonconservative homologous recombination repair. CSAs are the result of incompletely repaired DSBs or unrepaired DSBs. When cultured PBLs go through DNA synthesis and chromosomes are duplicated, the aberrations formed earlier are doubled and CSAs can be seen in metaphase. These include dicentric chromosomes with difragments, ring chromosomes with difragments and abnormal chromosomes [23,24].

### Chromatid-type aberrations (CTAs)

reflect changes in only one chromatid of a chromosome (Fig. 1b) and are induced by S-phase-dependent chemical agents such as ethylene oxide and UV light. CTAs are formed by homologous recombination from base modifications and single-strand breaks that are enzymatically converted to double-strand breaks or by incomplete or failed DNA repair. Most of the CTAs are chromatid breaks. Chromatid exchanges are the result of the dislocation of chromosomal material to another chromosome or within the chromosome and are different from sister chromatid exchanges, see Fig. 1c [12,30].

Monitoring of CAs in exposure to radiation has been instructive in many ways. Dicentric chromosomes and

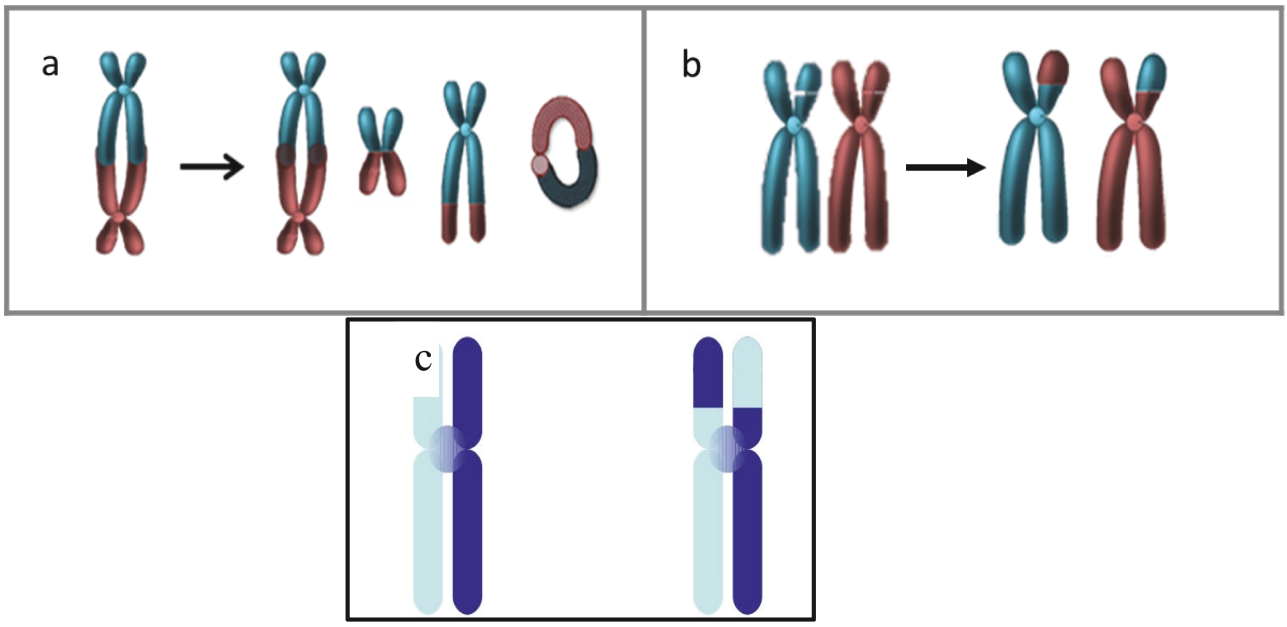


**Figure 1.** Types of chromosomal aberrations. (a) CSA, dicentric and ring chromosomes; (b) CTA, gaps, and exchanges; (c) sister chromatid exchange.

reciprocal translocations have been considered "a gold standard" for radiation bio-dosimetry. They are very rare in unirradiated persons. Dicentric chromosomes increase in a linear fashion depending on the dose (when exposure is X- and gamma-rays) and linear-quadratic manner (alpha-particles and neutrinos) [12]. Reciprocal translocations persist for the lifetime of PBLs, about 6 years, while unstable dicentrics disappear with a half-life of 2–3 years or faster [12].

## Conventional cytogenetic analysis

Conventional cytogenetic techniques are used for the analysis of karyotype and nonspecific CAs. Essentially the same technology has been used since the correct counting of human chromosomes in 1956 by Tjio and Levan [9,31]. The method enables the identification of numerical and structural aberrations present in metaphase cells under a microscope. Despite the fact that more sensitive molecular cytogenetics techniques have been developed and are being used in modern cytogenetics analyses, conventional cytogenetics still has its importance in many contexts. It is the most common and easily accessible assay.

In the conventional karyotype analysis, cultured PBLs are stimulated by phytohemagglutinin to enter mitosis and arrested in the metaphase by colcemid after which the chromosomes are harvested and a single cell suspension is produced (Fig. 2). Following fixation, the dehydrated swollen metaphase cells are mounted onto the glass slides. Staining is done with Giemsa stain followed by the trypsin treatment. Conventional karyotyping uses the G-banding method and it can generate up to 1000 bands per haploid human genome [32]. Every band has been assigned a specific designation to represent its location on each chromosome by the international system for human cytogenetic nomenclature (ISCN) [33]. For nonspecific CAs scoring microscopic analysis usually covers 100 mitoses in blinded analysis in coded slides.

## Genome-wide association studies (GWASs)

### Study population

In 2019, we published a GWAS based on genotyping data of 1473 healthy individuals from the Czech Republic and Slovakia, as described in detail [19]. The exposed population ($N$ = 607) had experienced genotoxic exposure due to their occupation and/or smoking habits. There were 40.5% female and 59.5% male individuals with a median age of 43 years. A majority of the individuals were exposed to small organic compounds and/or they were smokers (66%). The reference group comprised 866 individuals (67.3% female and 32.7% male) with median age of 43 years, without any known genotoxic exposure.

### Genotyping

Genotyping of the study population was done using Illumina HumanOmniExpressExome8v1.3 chip arrays, comprising nearly 1 million SNPs throughout the genome. After imputation for common variants across the genome over 10 million genotyped and imputed variants with a minor allele frequency of over 5% fulfilled the typical GWAS quality control, described in [19].

### Population stratification

One of the quality control measures is shown here because limited data are available on the population stratification of the Czech and Slovak populations. Principal component analysis (PCA) is shown in Fig. 3 on population stratification of the current study population with respect to the European (CEU), Han/Chinese and Yoruba populations. The Czech and Slovak genotypes match well with the Europeans but show also outliers. As the differences in ancestry may be a source of bias, the outliers were removed from the analysis [34,35]. Owing to the vast differences between the genetics of the three ancestral groups the first two principal components are adequate to stratify the individuals from these populations.

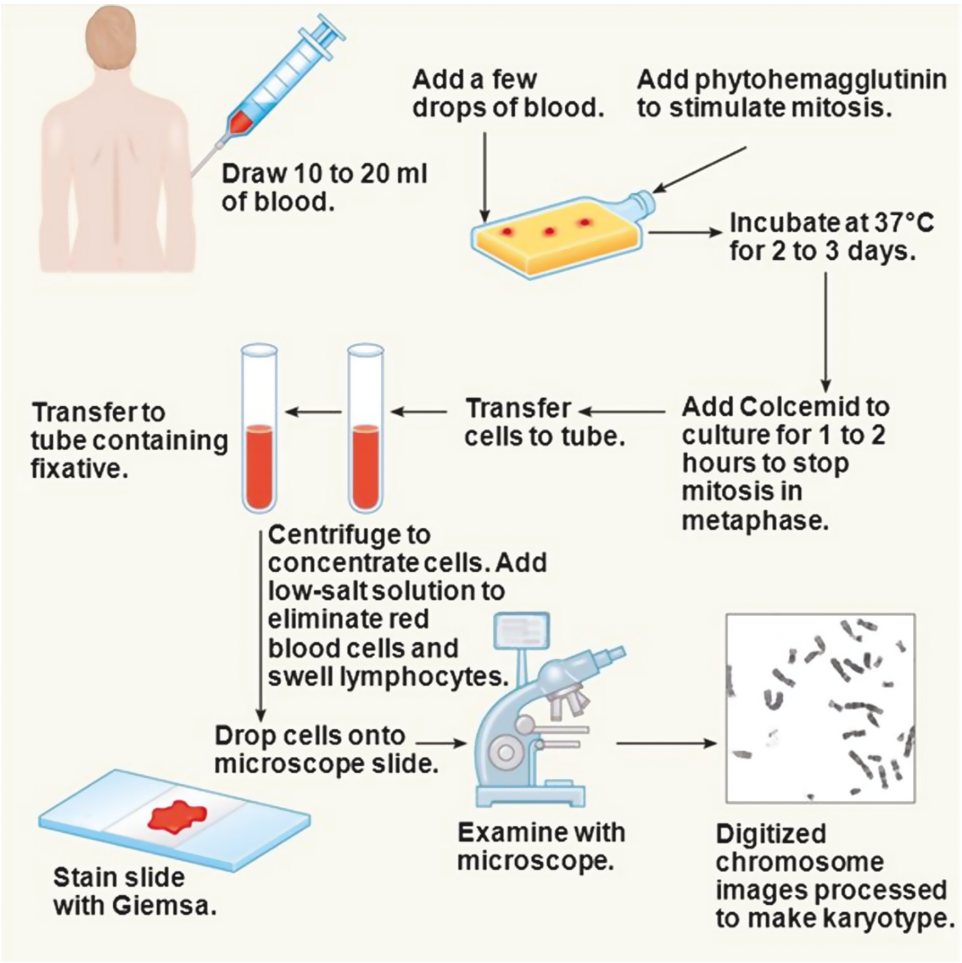


**Figure 2.** Karyotype analysis (from Cummings. eBook: Human Heredity: Principles and Issues, 11E. ©2016 Brooks/Cole, a part of Cengage, Inc. Reproduced by permission. www.cengage.com/permissions).

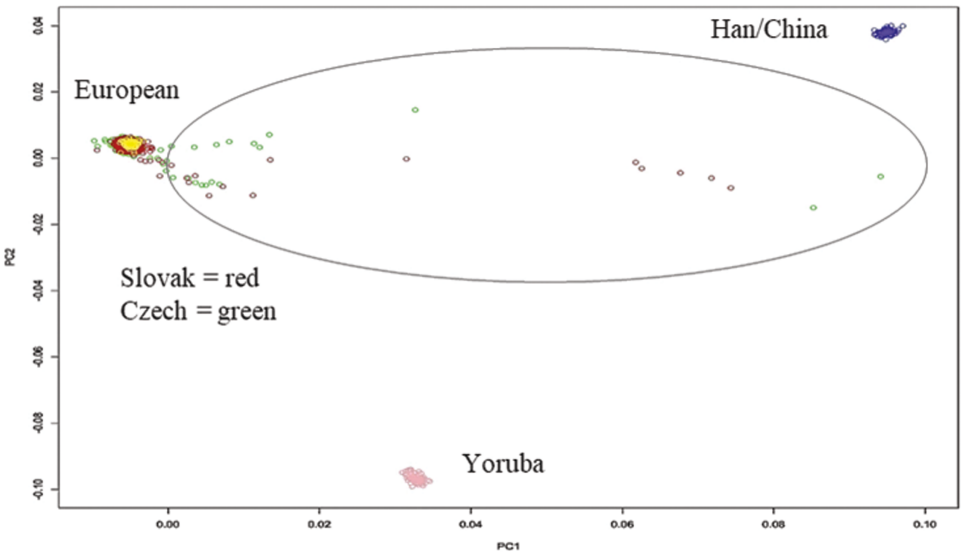


**Figure 3.** Population stratification with principal component analysis (PCA). The Czech and Slovak populations are shown together with the European (CEU), Han/Chinese, and Yoruba populations. The circle shows the Czech and Slovak individuals not matching with the European population; they were excluded from the analyses.

## Association analyses

### Chromosomal aberrations in exposed and reference populations

A genetic association study was published by dividing the whole population into the unexposed reference group and the exposed group as described above based on the assumption that chemical exposures may involve unique pathways in CA induction [19]. The distribution of CAs showed different modes in the two populations, with the highest frequency at two CAs in the exposed population compared to zero in the referent population (Fig. 4).

### Association of variants with CAs in exposed and reference populations

Association analysis was run with both logistic and linear regression models. We collected significant associations ($P < 10^{-5}$) from the two populations in Fig. 5 (**19**). The color code shows from which population the results were collected. In the reference group significant associations were observed in variants in genes functionally related to DNA damage response/repair, *PSMA1* (proteasome 20S subunit alpha 1), *PMS2P4* (PMS1 homolog 2, mismatch repair system component pseudogene 4), *RRM2B* (ribonucleotide reductase regulatory TP53 inducible subunit M2B) and *UBR5* (ubiquitin protein ligase E3 component N-recognin 5). *RRM2B* is also related to chromatin modulation as are the genes *STAG3L4* (STAG3 cohesin complex component like 4 (pseudogene)), and *BOD1* (biorientation of chromosomes in cell division 1). Additionally, associations were found with 2 loci related to tumor progression/suppression ontology, including *COPB1* (COPI coat complex subunit beta 1) and *NR2C1* (nuclear receptor subfamily 2 group C member 1). Among the exposed

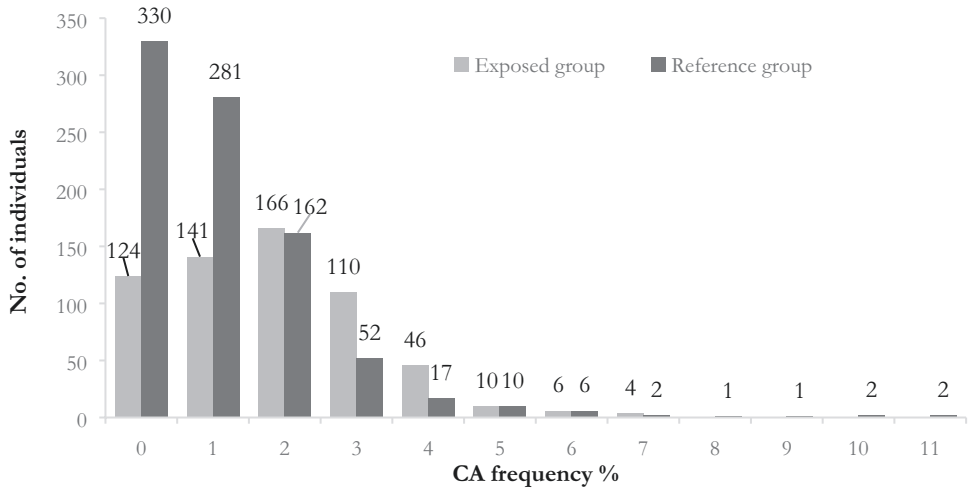


**Figure 4.** Distribution of CAs in the study population.

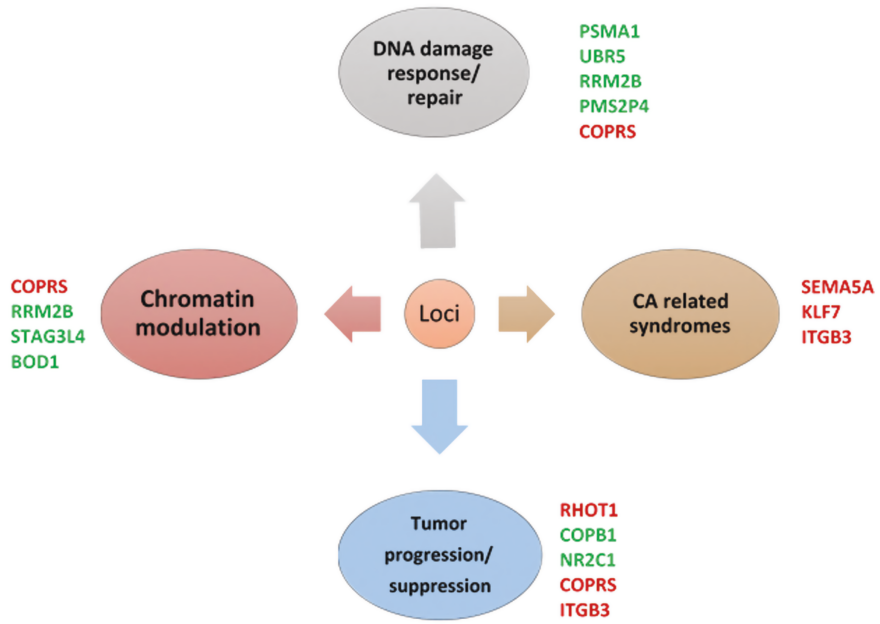


**Figure 5.** Summary of the GWAS results from exposed and reference groups according to gene functions. Reference group genes = green; exposed group genes = red. Note that some genes are included in multiple functional pathways. CA, chromosomal aberration.

group, a relevant functional classification was CA-related syndromes, including genes *SEMA5A* (semaphorin 5A), *KLF7* (KLF transcription factor 7), and *ITGB3* (integrin subunit beta 3), which was also classified as tumor progression/suppression ontology. In that category were also *RHOT1* (ras homolog family member T1) and *COPRS* (coordinator of PRMT5 and differentiation stimulator); the gene ontology classification for this gene is histone binding which may explain it diverse functions (see Fig. 5).

A meta-analysis of the GWASs from the exposed and reference populations identified further three loci, potentially representing loci predisposing to CAs independent of exposure. These included a gene implicated in DNA damage response/repair, *FTO* (FTO alpha-ketoglutarate dependent dioxygenase), and two genes related to tumor progression/suppression, *SYK* (spleen associated tyrosine kinase) and *SEMA6A* (semaphorin 6A).

In conclusion, the results suggested a number of genes associated with CA frequency, as summarized in Fig. 5. These were found among genes broadly functionally related to DNA damage response/repair and chromatin modulation (*PSMA1, UBR5, RRM2B, PMS2P4, STAG3L4, BOD1, COPRS,* and *FTO*) as a common group. The second group was related to apoptosis, cell proliferation, angiogenesis and tumorigenesis (*COPB1, NR2C1, COPRS, RHOT1*, *ITGB3, SYK,* and *SEMA6A*). Three different loci (mapped to genes *KLF7, ITGB3,* and *SEMA5A*) were directly or indirectly related to autism/autistic traits, conditions linked to CAs [19]. These results suggest a complex interaction of various genetic factors responsible for the inter-individual differences in CA frequency in the presence and absence of evident exposure to genotoxins. Further functional studies are warranted to unravel the mechanism behind these interactions, yet these results help narrow down the important genes and pathways behind them.

### Variants related to DNA repair genes and their expression

Considering the likely mechanistic role of DNA repair in the formation of CAs, as supported also by the above GWAS, we decided to concentrate on candidate gene analysis on 153 known DNA repair genes the results of which have been reported [36,37]. Gene boundaries (start and end positions) were acquired from UCSC genome browser hg19 assembly [38] and all single nucleotide polymorphisms (SNPs) within gene boundaries including 100 kb upstream and downstream extended window were analyzed in order to include also potential regulatory elements to the analysis. All loci with SNPs with *P*-value of $5 \times 10^{-3}$ or below according to the additive model in the above GWAS were investigated further. This level of significance was selected to identify potential associations above the background level of the GWAS, considering adjustment for multiple comparisons by the Binomial Sequential Goodness of Fit test. The top SNPs were investigated for their predicted functional role in modifying the expression of the respective DNA repair gene. This functional analysis was carried out utilizing online bioinformatics resources.

### Results from the reference population of 866 individuals

These persons were nonsmokers and had no known occupational exposure to genotoxic substances [36]. Significant associations with CAs were found for 22 DNA repair genes. Nucleotide excision repair pathway genes showed the most associations with six genes. Among the associated genes were several in which mutations manifest CA phenotype, including Fanconi anemia, Werner syndrome and Bloom syndrome and genes that are important in maintaining genome stability, as well as *PARP2* (Poly (ADP-ribose) polymerase 2) and mismatch repair genes. Replication proteins encoding genes *RPA2* and *RPA3* may participate in telomere maintenance through the synthesis of the C strand of telomeres. Errors in NHEJ1 (nonhomologous end joining factor 1) function may lead to translocations. The results showed plausible genetic rationale for the formation of CAs in the healthy nonsmoking populations [36].

In the same study, the associated SNPs were annotated using GTEx for the expression quantitative trait loci (eQTLs) targeting the linked DNA repair gene [36]. Table 1 shows tissues where eQTL data were generated, normalized effect size (defined as the slope of the linear regression of the alternative allele compared to the reference allele in the human reference genome) and *P*-value. eQTL data were available for most of the tagged genes (18/22) and for *MSH4* and *MSH3* data were available on two different tagging SNPs. Close to one-half of the data were generated in whole blood. Some of the normalized effect sizes were high and highly significant such as 0.69 for *DUT* ($4.3 \times 10^{-55}$) and –0.45 for *RPA3* ($2.5 \times 10^{-22}$).

### Results from the exposed population of 607 individuals

A total of 14 genes were associated with CAs and five of them were shared with the results from the above reference population [37]. For the base excision repair pathway, the implicated genes *PARP1* and *PARP2* encode poly(ADP-ribosyl) transferases with an important role in maintaining genome stability through diverse mechanisms. Similar functions are also known for *GTF2H* (general transcription factor IIH subunits 4 and 5), Fanconi anemia pathway genes, and *PMS2,* a mismatch repair gene. Most of the implicated SNPs (10/14) were eQTLs influencing the expression of the target repair gene. For some the significance levels were very high, including eQTLs for *MGMT* ($1.40 \times 10^{-33}$) and for *PARP1* ($9.00 \times 10^{-23}$) [37].

In conclusion, the results on DNA repair genes from the two populations suggest pathways with mechanistic rationale for the formation of CAs and highlight the role of the implicated SNPs in regulating the expression of the target DNA repair genes.

## Genetic contribution to nonspecific chromosomal aberrations

The described studies constitute some of the first efforts to understand the genetics basis of nonspecific CAs and the first genome-wide studies. They add to the previous genetic studies on nonspecific CAs which had focused on individual genes among DNA repair, mitotic checkpoint, and metabolic genes [36,37,39–45]. The results indicate that many individual genes are likely to contribute to the formation of CAs, in agreement with the first published GWAS results on CAs [44]. There was no evidence that a single gene would show a strong association with nonspecific CAs. This is opposite to the demonstrated association of cyclin D1 splice site variant on translocation *t*(11;14) which exerts clinical influence in many hematological malignancies [29]. However, a contributing

**Table 1.** *In silico* properties of associated variants according to GTEx (eQTL) [36].

| Repair pathway | Chr | Tagging SNP | Gene | eQTL | | *P*-values |
|---|---|---|---|---|---|---|
| | | | | Tissue | Normalized effect size[a] | |
| NER | 6 | rs13202019 | *GTF2H5* | Whole blood | 0.17 | 2.20E – 08 |
| | 6 | rs1052693 | *GTF2H4* | Muscle—Skeletal | –0.27 | 1.40E – 13 |
| | 7 | rs12702634 | *RPA3* | Sun exposed skin | –0.45 | 2.50E – 22 |
| | 5 | rs2130756 | *CDK7* | Whole blood | 0.1 | 2.90E – 06 |
| | 1 | rs10794509 | RPA2 | Whole blood | –0.35 | 3.3E – 19 |
| MMR | 1 | rs143170391 | *MSH4* | Cultured fibroblasts | –0.24 | 1.60E – 04 |
| | 1 | rs12732495 | *MSH4* | Cultured fibroblasts | –0.29 | 3.70E – 05 |
| | 7 | rs62456189 | *PMS2* | Whole blood | 0.22 | 6.00E – 07 |
| | 5 | rs443611 | *MSH3* | Liver | 0.36 | 4.90E – 13 |
| | 5 | rs32980 | *MSH3* | Whole blood | 0.28 | 3.60E – 09 |
| BER | 4 | rs10013040 | *NEIL3* | Testis | 0.23 | 5.50E – 06 |
| | 14 | rs4981148 | *PARP2* | Cultured fibroblasts | 0.1 | 2.20E – 06 |
| | 16 | rs2541632 | *MPG* | Adrenal glands | –0.0.47 | 6.30E – 08 |
| Fanconi anemia | 15 | rs113771463 | *RAD51* | Cultured fibroblasts | 0.067 | 3.10E – 08 |
| NHEJ | 8 | rs2293982 | *RRM2B* | Whole blood | 0.057 | 9.90E – 05 |
| | 15 | rs3784618 | *DUT* | Whole blood | 0.69 | 4.30E – 55 |
| | 2 | rs7572601 | *NHEJ1* | Whole blood | –0.15 | 1.10E – 06 |
| | 15 | rs8034371 | *BLM* | Heart—left ventricle | –0.2 | 1.80E – 08 |
| Ubiquitination | 3 | rs7641235 | *RAD18* | Cultured fibroblasts | –0.12 | 3.90E – 08 |
| DNA polymerases | 5 | rs17672542 | *POLK* | Testis | 0.16 | 7.70E – 06 |

[a]The normalized effect size of the eQTLs is defined as the slope of the linear regression and is computed as the effect of the alternative allele relative to the reference allele in the human reference genome. NER, nucleotide excision repair; MMR, mismatch excision repair; BER, base excision repair, NHEJ, nonhomologous end-joining.

factor is certainly the imprecise scoring of nonspecific CAs. Stochastic factors play a role in how the microscopic fields are selected for scoring 100 metaphases. For example, in the referent population, the most common class was zero CAs, but if the number of scored metaphases would be 10- or 100-fold higher than 100, a much more precise frequency could be arrived at.

It is relevant to point out that even if the frequency of CAs was higher in the exposed population they were found also in the reference population without known occupational or lifestyle (smoking) exposures. The significant associations were broadly classified into variants in genes functionally related to DNA damage response/repair and chromatin modulation commonly found in the reference population; genes related to known syndromes expressing CAs which we mainly found in the exposed population; genes related to cell proliferation and tumorigenesis were found in both groups. We have no valid explanation for the preferential presentation of the SNPs in the exposed or reference population but both are consistent with the mechanistic rationale that the variants in the large machinery of genes maintaining genomic integrity influence CA frequencies. A functionally important novel finding was that the large majority of SNPs among the 153 analyzed DNA repair genes were eQTLs with significant changing expression of the target repair genes.

## Prospects of biomonitoring for cancer risk

CAs in PBLs are biomarkers of genotoxic exposure but their application in human biomonitoring has been conducted in only some countries, particularly in some Eastern European countries during the socialist rule. However, now the activity has ceased even in these countries, as elsewhere, in spite of convincing evidence that CA frequency is predictive of future cancer risk [2]. Alternative techniques applicable to the healthy population are not available even though many resources have been invested to develop such techniques.

Why this is the case is difficult to understand or explain. Are genotoxic occupational exposures better controlled by stricter regulation of the allowed concentrations for these chemicals or has the role of occupational health changed away from disease prevention to diagnosing and treatment? This is in contrast to biomonitoring of radiation exposure, for which CAs techniques are routinely used [12,13,16]. Radiation exposure monitoring is simple and cancer causation by ionizing radiation may be easier to understand than cancer induced by chemical carcinogens. Societal responsibility for human health appears to be higher for radioactive substances than for occupational and environmental chemical carcinogens. Additionally, technology for nonspecific CAs has been only minimally automated (for metaphase search) while multicolor FISH (fluorescence *in situ* hybridization) techniques, although quite complicated, have been partially automated for CA scoring [18]. It is difficult to believe that artificial intelligence methods would not work for nonspecific CAs as they are working for delicate structural changes on complex histological slides [46]. Such methods have been developed now for micronuclei analysis, which can be accurately scored in .3 s per image and further developed to exceed manual detection rate some 20-fold [47,48].

We cannot predict what the future of organized biomonitoring for cancer will be but it should be sobering to consider what the main causes of cancer are. According to

the twin studies, environmental factors are the main causes of all major types of cancer [49,50]. Thus, nonspecific CAs in blood may reflect the person's response to environmental factors (" exposome") but additionally, they are able to integrate these into her genetic make-up. Analysis of genetic variants (SNPs) considers only germline genetic makeup and if we believe that cancer is indeed mainly an environmental disease, an SNP test gauging only genetic factors will have a markedly reduced power to predict cancer risk. Nevertheless, over more than a decade, combinations of SNPs have been used in the assessment of cancer risk [51]. SNP panels have been devised for "genetic risk scores" by pooling risk SNPs for various cancers. Initially, a few SNPs were combined to the score but now these may include 100 or more SNPs with minute extra risks. Although such polygenic risk scores reflect the human genetic background of low-risk genes, the clinical utility and application of these are debated and so far limited [52]. However, private companies market SNP-based detection from blood samples directly to potential customers who would like to receive predictions of their risks for cancer or any of numerous other diseases (https://lifesciences.tecan.com/genetic-testing-innovations-in-genomics-and-ngs). This is done without any regard for how large the genetic component for the particular disease may be. Although such commercial activity may not be health-promoting, it reflects people's health concerns and need-to-know about the possible diseases that they may face. Reliable prediction tools are urgently needed but, outside hereditary diseases, they need to combine environmental fingerprints on top of the genetic make-up. Suffice it to summarize in finishing that the promise of CA monitoring for genotoxic risk should be exploited by scaling up the processing and adding type-specific filters which may tell about their formation.

## Data availability statement

For original data beyond the publications please contact the authors.

## Acknowledgments

Supported by the European Union's Horizon 2020 research and innovation programme, grant no. 856620, the Cooperatio Program, research area SURG and National Institute for Cancer Research—NICR (Programme EXCELES, ID Project No. LX22NPO5102), funded by the European Union—Next Generation EU, further by the Ministry of Health of the Czech Republic AZV NU21–03-00506 and by GACR grants 21-27902 and 23-05609S and UNCE/MED/006 University Center of Clinical and Experimental Surgery.

## Conflict of interest statement

The authors declare no potential conflicts of interest.

## References

1. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell* 2011;**144**:646–74.
2. Bonassi S, Norppa H, Ceppi M, *et al*. Chromosomal aberration frequency in lymphocytes predicts the risk of cancer: results from a pooled cohort study of 22 358 subjects in 11 countries. *Carcinogenesis* 2008;**29**:1178–83.
3. Mitelman F, Johansson B, Mertens F. The impact of translocations and gene fusions on cancer causation. *Nat Rev Cancer* 2007;**7**:233–45.
4. Mertens F, Johansson B, Fioretos T, *et al*. The emerging complexity of gene fusions in cancer. *Nat Rev Cancer* 2015;**15**:371–81.
5. Carrano AV, Natarajan AT. International Commission for Protection Against Environmental Mutagens and Carcinogens. ICPEMC publication no. 14. Considerations for population monitoring using cytogenetic techniques. *Mutat Res* 1988;**204**:379–406.
6. Ford CE, Hamerton JL. The chromosomes of man. *Nature* 1956;**178**:1020–3.
7. Arnason U. 50 years after—examination of some circumstances around the establishment of the correct chromosome number of man. *Hereditas* 2006;**143**:202–11.
8. Nowell PC. Discovery of the Philadelphia chromosome: a personal perspective. *J Clin Invest* 2007;**117**:2033–5.
9. Trask BJ. Human cytogenetics: 46 chromosomes, 46 years and counting. *Nat Rev Genet* 2002;**3**:769–78.
10. Ashby J. Comparison of techniques for monitoring human exposure to genotoxic chemicals. *Mutat Res* 1988;**204**:543–51.
11. Hagmar L, Strömberg U, Tinnerberg H, *et al*. Epidemiological evaluation of cytogenetic biomarkers as potential surrogate end-points for cancer. *IARC Sci Publ* 2004:207–15.
12. Albertini RJ, Anderson D, Douglas GR, *et al*. IPCS guidelines for the monitoring of genotoxic effects of carcinogens in humans. International Programme on Chemical Safety. *Mutat Res* 2000;**463**:111–72.
13. Ceppi M, Smolkova B, Staruchova M, *et al*. Genotoxic effects of occupational exposure to glass fibres—a human biomonitoring study. *Mutat Res Genet Toxicol Environ Mutagen* 2023;**885**:503572.
14. Farkas G, Kocsis ZS, Székely G, *et al*. Spontaneous chromosomal aberrations in lymphocytes and development of tumor in hospital workers. *Anticancer Res* 2022;**42**:1059–64.
15. Kadlcikova D, Musilova P, Hradska H, *et al*. Chromosomal damage in occupationally exposed health professionals assessed by two cytogenetic methods. *Arch Environ Occup Health* 2023;**78**:158–69.
16. Anderson RM. Cytogenetic biomarkers of radiation exposure. *Clin Oncol (R Coll Radiol)* 2019;**31**:311–8.
17. Vinnikov VA, Belyakov O. Radiation exposure biomarkers in the practice of medical radiology: cooperative research and the role of the International Atomic Energy Agency (IAEA) Biodosimetry/Radiobiology Laboratory. *Health Phys* 2020;**119**:83–94.
18. Balajee AS, Livingston GK, Escalona MB, *et al*. Cytogenetic follow-up studies on humans with internal and external exposure to ionizing radiation. *J Radiol Prot* 2021;**41**.
19. Niazi Y, Thomsen H, Smolkova B, *et al*. Distinct pathways associated with chromosomal aberration frequency in a cohort exposed to genotoxic compounds compared to general population. *Mutagenesis* 2019;**34**:323–30.
20. Cleal K, Baird DM. Catastrophic endgames: emerging mechanisms of telomere-driven genomic instability. *Trends Genet* 2020;**36**:347–59.
21. Mammel AE, Hatch EM. Genome instability from nuclear catastrophe and DNA damage. *Semin Cell Dev Biol* 2022;**123**:131–9.
22. Guo X, Dai X, Wu X, *et al*. Small but strong: mutational and functional landscapes of micronuclei in cancer genomes. *Int J Cancer* 2021;**148**:812–24.
23. Janssen A, van der Burg M, Szuhai K, *et al*. Chromosome segregation errors as a cause of DNA damage and structural chromosome aberrations. *Science* 2011;**333**:1895–8.
24. Urakami K, Shimoda Y, Ohshima K, *et al*. Next generation sequencing approach for detecting 491 fusion genes from human cancer. *Biomed Res* 2016;**37**:51–62.
25. Heng HH, Regan SM, Liu G, *et al*. Why it is crucial to analyze non clonal chromosome aberrations or NCCAs? *Mol Cytogenet* 2016;**9**:15.

26. Johansson B, Mertens F, Schyman T, *et al*. Most gene fusions in cancer are stochastic events. *Genes Chromosomes Cancer* 2019;**58**:607–11.
27. Greaves M. A causal mechanism for childhood acute lymphoblastic leukaemia. *Nat Rev Cancer* 2018;**18**:471–84.
28. Ni H, Wang XT, Fang R, *et al*. Gene fusions in human cancers: a review focused on diagnostic biomarkers, method selections, and treatments. *Crit Rev Oncog* 2017;**22**:403–10.
29. Weinhold N, Johnson DC, Chubb D, *et al*. The CCND1 G870A polymorphism is a risk factor for *t*(11;14)(q13;q32) multiple myeloma. *Nature Genet* 2013;**45**:522–5.
30. Pfeiffer P, Goedecke W, Obe G. Mechanisms of DNA double-strand break repair and their potential to induce chromosomal aberrations. *Mutagenesis* 2000;**15**:289–302.
31. Dolan M. Conventional and molecular cytogenetics in cancer. In: Yousef GM, Jothy S (eds.), *Molecular Testing in Cancer*. New York: SpringerLink, 2014.
32. Smith K. Chapter 38—basic cytogenetic techniques: culturing, slide making, and G banding. In: Celis JE (ed.), *Cell Biology (Third Edition)*. Burlington: Academic Press, 2006, 381–5.
33. Stevens-Kroef M, Simons A, Rack K, Hastings RJ. Cytogenetic nomenclature and reporting. In: Wan TSK (ed.), *Cancer Cytogenetics: Methods and Protocols*. New York: Springer, 2017, 303–9.
34. Tian C, Gregersen PK, Seldin MF. Accounting for ancestry: population substructure and genome-wide association studies. *Hum Mol Genet* 2008;**17**:R143–50.
35. Frazer KA, Ballinger DG, Cox DR, *et al*; International HapMap Consortium. A second generation human haplotype map of over 3.1 million SNPs. *Nature* 2007;**449**:851–61.
36. Niazi Y, Thomsen H, Smolkova B, *et al*. DNA repair gene polymorphisms and chromosomal aberrations in healthy, non-smoking population. *DNA Repair (Amst)* 2021;**101**:103079.
37. Niazi Y, Thomsen H, Smolkova B, *et al*. DNA repair gene polymorphisms and chromosomal aberrations in exposed populations. *Front Genet* 2021;**12**:691947.
38. Rosenbloom KR, Armstrong J, Barber GP, *et al*. The UCSC genome browser database: 2015 update. *Nucleic Acids Res* 2015;**43**:D670–81.
39. Forsti A, Frank C, Smolkova B, *et al*. Genetic variation in the major mitotic checkpoint genes associated with chromosomal aberrations in healthy humans. *Cancer Lett* 2016;**380**:442–6.
40. Vodicka P, Musak L, Frank C, *et al*. Interactions of DNA repair gene variants modulate chromosomal aberrations in healthy subjects. *Carcinogenesis* 2015;**36**:1299–306.
41. Vodicka P, Naccarati A, Vodickova L, *et al*. Do GST polymorphisms modulate the frequency of chromosomal aberrations in healthy subjects? *Environ Health Perspect* 2009;**117**:A384–5.
42. Hemminki K, Frank C, Forsti A, *et al*. Metabolic gene variants associated with chromosomal aberrations in healthy humans. *Genes Chromosomes Cancer* 2015;**54**:260–6.
43. Hemminki K, Musak L, Vymetalkova V, *et al*. Cyclin D1 splice site variant triggers chromosomal aberrations in healthy humans. *Leukemia* 2014;**28**:721–2.
44. Niazi Y, Thomsen H, Smolkova B, *et al*. Genetic variation associated with chromosomal aberration frequency: a genome-wide association study. *Environ Mol Mutagen* 2019;**60**:17–28.
45. Niazi Y, Thomsen H, Smolkova B, *et al*. Impact of genetic polymorphisms in kinetochore and spindle assembly genes on chromosomal aberration frequency in healthy humans. *Mutat Res Genet Toxicol Environ Mutagen* 2020;**858–860**:503253.
46. Kimura K, Ai T, Horiuchi Y, *et al*. Automated diagnostic support system with deep learning algorithms for distinction of Philadelphia chromosome-negative myeloproliferative neoplasms using peripheral blood specimen. *Sci Rep* 2021;**11**:3367.
47. Shen X, Chen Y, Li C, *et al*. Rapid and automatic detection of micronuclei in binucleated lymphocytes image. *Sci Rep* 2022;**12**:3913.
48. Shen X, Ma T, Li C, *et al*. High-precision automatic identification method for dicentric chromosome images using two-stage convolutional neural network. *Sci Rep* 2023;**13**:2124.
49. Lichtenstein P, Holm N, Verkasalo P, *et al*. Environmental and heritable factors in the causation of cancer. *N Engl J Med* 2000;**343**:78–85.
50. Mucci LA, Hjelmborg JB, Harris JR, *et al*; Nordic Twin Study of Cancer (NorTwinCan) Collaboration. Familial risk and heritability of cancer among twins in Nordic countries. *JAMA* 2016;**315**:68–76.
51. Zheng SL, Sun J, Wiklund F, *et al*. Cumulative association of five genetic variants with prostate cancer. *N Engl J Med* 2008;**358**:910–9.
52. Sud A, Horton RH, Hingorani AD, *et al*. Realistic expectations are key to realising the benefits of polygenic scores. *BMJ* 2023;**380**:e073149.