

Chromosomal aberration frequency in lymphocytes predicts the risk of cancer: results from a pooled cohort study of 22 358 subjects in 11 countries

Stefano Bonassi^{1*}, Hannu Norppa², Marcello Ceppi¹, Ulf Strömberg³, Roel Vermeulen⁴, Ariana Znaor⁵, Antonina Cebulska-Wasilewska⁶, Eleonora Fabianova⁷, Alexandra Fucic⁸, Sarolta Gundy⁹, Inger-Lise Hansteen¹⁰, Lisbeth E.Knudsen¹¹, Juozas Lazutka¹², Pavel Rossner¹³, Radim J.Sram¹³ and Paolo Boffetta¹⁴

¹Unit of Molecular Epidemiology, National Cancer Research Institute, Genoa, Italy, ²New Technologies and Risks, Work Environment Development, Finnish Institute of Occupational Health, Topeliuksenkatu 41 aA, FI-00250 Helsinki Finland, ³Department of Occupational and Environmental Medicine, Lund University, S-221 85 Lund, Sweden, ⁴Environmental and Occupational Health Division, Institute for Risk Assessment Sciences, University Utrecht, Utrecht, PO Box 80176, 3503 TD Utrecht, The Netherlands, ⁵Croatian National Cancer Registry, Croatian National Institute of Public Health, Rockefellerova 7, 10000 Zagreb, Croatia, ⁶Department of Radiation and Environmental Biology, The Henryk Niewodniczanski Institute of Nuclear Physics, PAN, Radzikowskiego Street 152, 31-342 Kraków, Poland, ⁷Department of Occupational Health, Specialized State Health Institute, Banska Bystrica, Slovakia, ⁸Institute for Medical Research and Occupational Health, 10 000 Zagreb, Ksaverska c 2 Croatia, ⁹Department of Diagnostic Onco-cytogenetics, National Institute of Oncology, Ráth Gy. u. 7-9, Budapest 1122, Hungary, ¹⁰Department of Laboratory Medicine, Section of Medical Genetics, Telemark Hospital, Skien, Norway, ¹¹Environmental Health Institute of Public Health, University of Copenhagen, Øster Farimagsgade 5, Po Box 2099, 1014 Copenhagen K, Denmark, ¹²Vilnius University, Universiteto 3, 01513 Vilnius, Lithuania, ¹³Laboratory of Genetic Ecotoxicology, c/o Institute of Experimental Medicine Academy of Sciences of the Czech Republic, Videnska 1083, 142 20 Prague 4, Czech Republic and ¹⁴Genetics and Epidemiology Cluster, International Agency for Research on Cancer, 150 cours Albert Thomas, 69008 Lyon, France

*To whom correspondence should be addressed. Tel: +390 10 5600924;
Fax: +390 10 5600501;
Email: stefano.bonassi@istge.it

Mechanistic evidence linking chromosomal aberration (CA) to early stages of cancer has been recently supported by the results of epidemiological studies that associated CA frequency in peripheral lymphocytes of healthy individuals to future cancer incidence. To overcome the limitations of single studies and to evaluate the strength of this association, a pooled analysis was carried out. The pooled database included 11 national cohorts and a total of 22 358 cancer-free individuals who underwent genetic screening with CA for biomonitoring purposes during 1965–2002 and were followed up for cancer incidence and/or mortality for an average of 10.1 years; 368 cancer deaths and 675 incident cancer cases were observed. Subjects were classified within each laboratory according to tertiles of CA frequency. The relative risk (RR) of cancer was increased for subjects in the medium [RR = 1.31, 95% confidence interval (CI) = 1.07–1.60] and in the high (RR = 1.41; 95% CI = 1.16–1.72) tertiles when compared with the low tertile. This increase was mostly driven by chromosome-type aberrations. The presence of ring chromosomes increased the RR to 2.22 (95% CI = 1.34–3.68). The strongest association was found for stomach cancer [RR_{medium} = 1.17 (95% CI = 0.37–3.70), RR_{high} = 3.13 (95% CI = 1.17–8.39)]. Exposure to carcinogens did not modify the effect of CA levels on overall cancer risk. These results reinforce the evidence of a link between CA

frequency and cancer risk and provide novel information on the role of aberration subclass and cancer type.

Introduction

Lifestyle factors, environmental exposures and genetic background interact in the development of many cancers. The use of phenotypic markers that integrate the impact of environmental factors and genetic processes would be of great value in risk assessment and surveillance of populations occupationally or environmentally exposed to carcinogens. To date, the strongest mechanistic and epidemiological evidence for such markers is available for chromosome damage (1–3).

The first modern description of a specific chromosome abnormality in a human tumour, i.e. the Philadelphia chromosome in chronic myeloid leukaemia, was published almost 50 years ago (4), but only with the introduction of banding techniques, the major role of chromosomal rearrangements in carcinogenesis was fully recognized (5). Subsequent observations that many genes affected by chromosomal rearrangements were involved in critical stages in cell growth, development or survival focused the interest of researchers on how the rearrangements alter the function of these target genes. The model linking chromosome breakage and rearrangement to the activation/deactivation of genes relevant to early stages of carcinogenesis was further refined in the 1990s, when the use of the fluorescence *in situ* hybridization technique in irradiated cells contributed to understanding the role of double-strand breaks in the formation of chromosomal aberrations (CAs) and the implications of homology-dependent and homology-independent pathways of double-strand break repair (1,6).

Although mechanistic evidence supporting the role of chromosomal alterations in early stages of cancer has been available for a long time, the first epidemiological data showing that the frequency of CA in peripheral lymphocytes may predict cancer incidence in human populations were published in early 1990s when two small cohort studies were performed in healthy subjects scored for CA in the Nordic countries and Italy (7–9). Later analyses based on those cohorts suggested that the cancer risk predictivity of CA is not modified by occupational exposure to carcinogens or tobacco smoking, supporting the hypothesis that the association between CA and cancer might be independent of exposure (10). Other cohort studies published thereafter have generally confirmed the results of the first reports, with some discrepancy in the strength of the associations between CA and cancer risk and in the independence from exposure to carcinogens (11,12).

The aim of this pooled analysis, based on original data of 11 national cohort studies from Europe already published, was to provide a semi-quantitative estimate of the association between CA and cancer risk, a measure that could be used for risk assessment. Other goals were to estimate the association with specific cancer sites and major subclasses of CA and the role of exposure to carcinogenic agents as potential effect modifiers of the association between CA and cancer risk.

Materials and methods

The study design of all cohort studies included in the pooled analysis was similar to the original approach used by the Nordic study group in the first studies published in the early 1990s (13). In brief, cancer-free subjects who underwent CA screening for biomonitoring purposes—mostly due to occupational exposure to carcinogens and mutagens—were followed up from the date of the first CA test until the date of death, cancer diagnosis, emigration, 85th

Abbreviations: CA, chromosomal aberration; CI, confidence interval; IR, ionizing radiation; RR, relative risk.

birthday or the end of follow-up (2000–2003, depending on the country), whichever occurred first. To standardize for interlaboratory variation, all subjects were classified within each laboratory according to tertiles of CA frequency as follows: *low* (1–33 percentile), *medium* (34–66 percentile) or *high* (67–100 percentile). The approach based on tertiles was the most efficient according to the Akaike information criteria (score = 0.70 for total CA and all cancers) when compared in the pooled database with other methods of standardization such as *z* scores (score = 0.72) and deviation from the mean (score = 0.72) (14).

Information on cancer incidence and mortality was collected through linkage with national/local cancer and cause of death registries. In Hungary and Poland, an active system of follow-up for mortality and cancer incidence was set up (12).

CA frequency was measured in cultured peripheral blood lymphocytes using the classification criteria of Savage (15). Only individuals with at least 100 scored metaphases were included in the database (>100 metaphases were scored for 15% of subjects and >200 for 1.3%). Gaps were not included. The culture time was 48 or 72 h. Tertiles for CA data based on 48 h culture time (89.7% of the whole series) were evaluated separately from those with 72 h culture time. More detail about the study design and the quality procedures implemented can be found in the original papers (7–12). The Italian database included an additional unpublished cohort of 391 subjects, among whom 9 cancer cases occurred at the end of the follow-up.

Pooled analysis

After obtaining approval by local institutional review boards, individual cytogenetic and epidemiological data from the original cohorts were collected and entered in a pooled database. The information gathered in each cohort included demographic data of the subject, the laboratory performing the assay, information on exposure to genotoxic agents and smoking habits. CA data included the date of the test; culture time; the number of cells scored and the number of chromatid breaks, dicentrics, chromosome breaks, chromatid exchanges, ring chromosomes, marker chromosomes and aberrant cells. Total CA frequency was defined as the number of cells with aberrations, excluding gaps, per 100 cells. Chromosome-type aberrations included chromosome-type breaks, ring chromosomes, marker chromosomes and dicentrics, whereas chromatid-type aberrations included chromatid-type breaks and chromatid exchanges. The criteria for inclusion of individuals in the pooled analysis were valid demographic data, at least 15 years of age and no cancer history at the time of the CA test. Whenever a subject was analysed more than once, only the first test was considered for frequency classification.

The pooled databases included a total of 22 358 subjects from 11 countries and 45 laboratories screened for CA between 1965 and 2002 and followed up for a median period of 10.1 years. The Czech data accounted for 48.3% of the total persons-years. Details about the pooled cohort are reported in Table 1.

Non-melanoma skin cancer was not systematically registered in the original cohorts, and therefore, subjects suffering from this cancer were excluded from the pooled analysis. This change contributed to small differences observed in the number of subjects (24 subjects overall) and in the national risk estimates with respect to published original data.

Information on occupational exposure at the time of the CA test was available for most subjects, although the quality of this information was heterogeneous among the cohorts. A common classification scheme based on the exposure groups identified by Tinnerberg *et al.* (16) was adopted. The most frequent exposure was to ionizing radiation (IR), with 2974 subjects, followed by exposure to cytostatic drugs (2521 subjects) and polycyclic aromatic hydrocarbon (2004 subjects). The group of non-exposed subjects included as controls in some of the studies consisted of 3662 subjects. No data on occupational exposure were available for cohorts from Sweden and Finland.

Information on smoking status at the time of the test was available for 18 766 subjects (83.9% of the total), among whom, 9657 were classified as ever smokers (43.2%). The amount of cigarettes smoked per day was registered only in a few cohorts and the quality of these data was not comparable among countries; therefore, this information was not used in the statistical analysis.

Statistical analysis

The pooled effect of CA frequency on cancer rates was estimated with a negative binomial regression analysis (17). All estimates were adjusted for the concomitant effect of age (5 years age-class), gender, time since test (0–4, 5–9, 10–14, 15–19 and ≥20 years), smoking habit (ever/never/former) and occupational exposure to carcinogens (yes/no). Relative risks (RRs) were estimated according to different levels of CA frequency together with its 95% confidence intervals (CIs). The presence of a linear trend of RRs was tested with the χ^2 trend test.

To take into account the different magnitude of cancer rates in each country, a random intercept model was fitted to the data to estimate the variance component due to the between-countries variability. A random slope model was performed to verify if the CA effect on cancer rates varied between countries

Table 1. Selected characteristics of the study population and frequency of cells with CA

Country	Laboratories	Subjects	Cancer deaths	Cancer cases	Person-years ^a	Median follow-up (years) ^b	Period of test	CA percentile		CTA ^b percentile		CSA ^c percentile		Age mean (SD)	Males (%)	Exposed (%)	Ever smokers (%)
								33 rd	67 th	33 rd	67 th	33 rd	67 th				
Croatia	1	1320	9	24	11 148	7.5	1982–2000	2.0	4.0	1.0	3.0	0.0	0.0	44.4 (9.7)	55.8	100.0	44.8
Czech Republic	15	11 991	172	363	114 712	9.2	1975–1999	1.5	3.0	1.0	2.0	0.0	0.0	47.5 (11.7)	55.4	84.5	42.2
Denmark	1	197	—	6	2456	12.8	1987	1.0	2.0	0.0	1.0	0.0	0.0	50.7 (10.8)	100.0	66.0	61.9
Finland	2	557	—	39	10 189	18.6	1974–1988	1.0	3.0	0.6	2.0	0.0	0.0	54.6 (9.6)	69.5	—	43.8
Hungary	1	840	29	46	6823	6.0	1978–2001	1.0	2.0	0.0	2.0	0.0	0.0	45.3 (12.5)	64.8	25.5	40.2
Italy	12	1964	99	—	31 818	14.9	1965–1995	1.3	4.0	0.7	2.0	0.0	0.0	56.3 (12.9)	72.4	62.1	38.4
Lithuania	1	812	14	24	7478	9.0	1981–2002	2.0	4.0	1.0	2.0	1.0	2.0	48.7 (9.4)	78.4	91.6	57.9
Norway	2	471	—	48	7936	16.3	1970–1988	1.0	2.0	0.0	1.0	0.0	0.0	57.6 (14.4)	80.0	65.8	44.2
Poland	2	456	16	22	5012	11.7	1981–2001	0.7	1.8	0.0	0.4	0.0	0.0	55.1 (14.5)	84.2	57.5	61.6
Slovakia	4	2994	29	55	27 135	8.9	1985–2001	1.0	2.0	1.0	2.0	0.0	0.0	47.8 (10.5)	57.8	76.4	43.8
Sweden	4	756	—	48	12 740	18.5	1970–1987	1.0	2.0	0.0	1.0	0.0	0.0	54.6 (11.7)	72.9	—	35.8
Total	45	22 358	368	675	237 447	10.1	1965–2002	1.0	3.0	1.0	2.0	0.0	0.0	48.2 (11.7)	60.8	74.2	43.2

^aAt the end of cancer incidence follow-up (except for Italy since mortality data only were available).

^bChromatid-type CAs.

^cChromosome-type CAs.

(18). These models allowed us to estimate the amount of heterogeneity present in the data and accordingly provided suitable estimates of the standard errors of the combined effects. A sensitivity analysis was performed to examine the variations in the pooled effect estimates due to the exclusion of one country at a time. Similarly, other sensitivity analyses were performed to evaluate the potential impact of an unbalanced distribution of main cohort characteristics such as culture time or the number of cells scored.

The potential effect modification of exposure to carcinogens on the main association between CA frequency and cancer was explored by means of the likelihood ratio test, i.e. comparing two hierarchical regression models for each major exposure category identified by the exposure matrix, with and without the interaction term, respectively (18).

To test the possibility that pre-clinical stages of cancer may have influenced CA frequency, we repeated the analysis excluding the first 2 years of follow-up and also tested a possible effect modification due to time since test.

Additional analyses were performed by country, by type of CA including subclasses and for cancer sites with at least 10 observed cases.

Kaplan–Meier product-limit method was applied to estimate the elapsed time without cancer at different CA levels. The proportional hazard assumption was not rejected based on the approach suggested by Grambsch *et al.* (19) [$\chi^2(2 \text{ df}) = 1.14, P = 0.57$].

The computer softwares MLwiN (20) and STATA (21) were used for the statistical analysis.

Results

An overall number of 675 incident cancers and 368 cancer deaths were registered at the end of the follow-up. The two end points were analysed independently. The most frequent cancer sites were lung (99 incident cases and 100 deaths), colon and rectum (88 and 53), bladder and kidney (76 and 28), breast (76 and 11) and the lympho-haematopoietic system (46 and 30).

The results on the association between overall cancer incidence and CA frequency in the whole database are reported in Table II. The RR was increased for subjects in the medium (RR = 1.31, 95% CI = 1.07–1.60) and in the high (RR = 1.41, 95% CI = 1.16–1.72) tertiles of the total CA distribution when compared with the low tertile (test for linear trend, $P = 0.001$). An increase in RR was seen within the subclass of chromosome-type aberrations, i.e. 1.29 (95% CI = 1.05–1.59) and 1.42 (95% CI = 1.17–1.71) for medium and high

Table II. RRs of cancer incidence by type of CAs (All tumours - ICD IX 140–208)

Type of CA	Subjects	Cases	Person-years	RR	95% CI
Total CAs					
Low ^a	6399	152	63 210	1	—
Medium	7456	240	71 956	1.31	1.07–1.60
High	6438	283	70 463	1.41	1.16–1.72
Chromatid-type aberrations					
Low ^a	6026	185	60 523	1	—
Medium	7146	226	69 696	1.10	0.90–1.35
High	6870	252	72 106	1.12	0.92–1.36
Chromosome-type aberrations					
Low ^a	8521	195	84 273	1	—
Medium	5366	207	55 311	1.29	1.05–1.59
High	6150	261	62 702	1.42	1.17–1.71
Dicentric chromosomes					
No ^a	7288	253	78 632	1	—
Yes	976	52	10 432	1.32	0.96–1.80
Ring chromosomes					
No ^a	7542	255	78 058	1	—
Yes	188	20	1941	2.22	1.34–3.68
Chromatid exchanges					
No ^a	4621	158	51 466	1	—
Yes	498	25	5550	1.30	0.83–2.01
Marker chromosomes					
No ^a	4849	163	53 121	1	—
Yes	270	20	3895	1.39	0.84–2.30

ICD, International Classification of Diseases.

RRs were adjusted for country, gender, age, time since test, job exposure and smoking habit.

^aReference category.

tertiles (test for linear trend, $P < 0.001$), respectively, but not for chromatid-type aberrations. Among the subclasses of CAs, the RR for the presence of ring chromosomes was 2.22 (95% CI = 1.34–3.68), whereas the non-significant RRs of other types of chromosomal abnormalities (dicentrics, chromatid exchanges and markers) ranged between 1.30 and 1.39.

A similar pattern of risk was found for mortality. Subjects in the medium tertile of total CA distribution experienced an RR of 1.36 (95% CI = 1.04–1.81, deaths = 136) and those in the high tertile an RR of 1.35 (95% CI = 1.03–1.80, deaths = 154). The RR for the two main classes of CA followed the same pattern observed for cancer incidence. The presence of marker chromosomes entailed a mortality RR of 1.71 (95% CI = 0.97–3.02).

Although increased risks were observed in all cohorts but Finland (with the highest RRs in Hungary and in Norway), country-specific results showed a certain degree of heterogeneity, and the relationship between RRs of medium and high CA tertiles was equivocal in some countries (Figure 1). However, fitting a random slope model to the data did not provide strong evidence of between-country heterogeneity in CA effect [$\chi^2 = 1.92$ (2 df), $P = 0.38$]. Furthermore, to evaluate if the overall association between CA and cancer risk was driven by a specific country, a sensitivity analysis was performed, sequentially removing a country from the analysis. The most extreme difference in RRs was obtained by removing the Czech Republic (increase to $RR_{\text{medium}} = 1.32, RR_{\text{high}} = 1.64$) and Hungary (decrease to $RR_{\text{medium}} = 1.25, RR_{\text{high}} = 1.29$) from the pooled analyses. In all cases, RRs remained significantly increased.

No covariate tested as potential effect modifier, i.e. gender, smoking habit, age, time from test and occupational exposure, was found to significantly modify the effect of CA levels on overall cancer risk in the whole cohort. In the analysis excluding the first 2 years of observation, the results on the association between CA and cancer incidence were not modified [597 cancer cases, $RR_{\text{medium}} = 1.33$ (95% CI = 1.07–1.66), $RR_{\text{high}} = 1.46$ (95% CI = 1.18–1.81)]. Similarly, the analyses stratified by median duration of follow-up did not yield different results in the two periods (data not shown).

In the analysis of specific types of cancer, the strongest association was found for stomach cancer [32 cases, $RR_{\text{medium}} = 1.17$ (95% CI = 0.37–3.70), $RR_{\text{high}} = 3.13$ (95% CI = 1.17–8.39); test for linear trend,

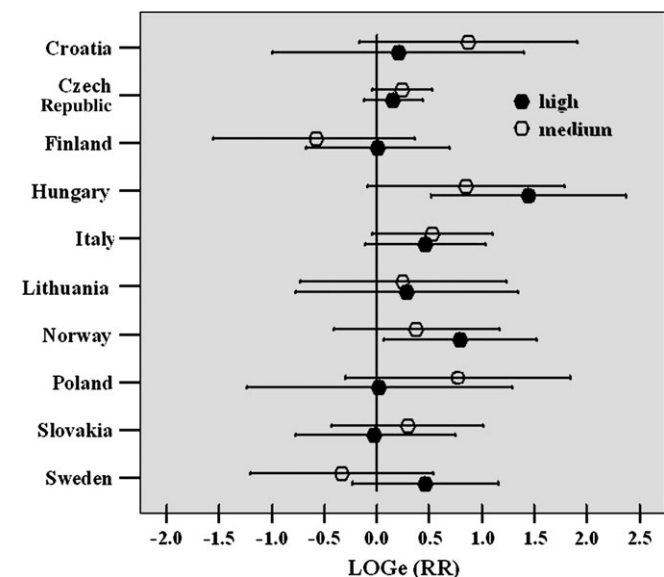


Fig. 1. Country-specific RR of cancer by frequency of CAs. RR was adjusted for gender, age, time since test, job exposure and smoking habit and was expressed in log scale to improve the readability of the figure. Symbols indicate RR by CA category, bars represent CIs. Reference category is the low tertile. RRs from Denmark could not be estimated because no cancer cases occurred in the reference group.

Table III. Cancer-specific RRs for frequency of CAs

Cancer site (ICD IX)	CA tertile						
	Low ^a		Medium		High		
	Cases	Cases	RR	95% CI	Cases	RR	95% CI
All cancers (140–208)	152	240	1.31	1.07–1.60	283	1.41	1.16–1.72
Mouth and pharynx (140–149)	7	5	0.55	0.17–1.74	15	1.47	0.60–3.63
Stomach (151)	5	7	1.17	0.37–3.70	20	3.13	1.17–8.39
Colon and rectum (153–154)	20	31	1.24	0.71–2.18	37	1.30	0.75–2.24
Liver and biliary ducts (155–156)	4	4	0.85	0.21–3.42	11	2.13	0.68–6.74
Pancreas (157)	3	6	1.66	0.42–6.66	7	1.69	0.44–6.58
Lung (162)	22	28	0.98	0.56–1.71	49	1.37	0.83–2.27
Melanoma (172)	3	6	1.74	0.43–7.01	9	2.65	0.71–9.92
Breast (174)	22	31	1.23	0.71–2.13	23	0.95	0.53–1.72
Female genital tract (179–183)	11	24	1.81	0.88–3.69	17	1.36	0.63–2.91
Prostate (185)	10	9	0.82	0.33–2.01	26	1.92	0.91–4.02
Bladder and kidney (188–189)	19	35	1.50	0.86–2.63	22	0.81	0.44–1.50
Lymphohaematopoietic system (200–208)	11	18	1.40	0.66–2.97	17	1.29	0.60–2.76

ICD, International Classification of Diseases.

RRs were adjusted for country, gender, age, time since test, job exposure and smoking habit.

^aReference category.

$P = 0.009$], whereas the results for other cancer sites were only suggestive of an association (Table III). Analysis based on cancer mortality showed comparable results except for neoplasms of the lymphohaematopoietic system, for which an association with CA frequency [30 cases, $RR_{\text{medium}} = 2.39$ (95% CI = 0.77–7.45), $RR_{\text{high}} = 2.61$ (95% CI = 0.85–8.02)] was suggested. No country-specific effect was found for any cancer site, but this analysis was hampered by small numbers.

As a whole, subjects exposed to genotoxic agents did not present a significant increase of cancer risk when compared with non-exposed subjects [$RR = 1.08$ (95% CI = 0.86–1.36)]; however, some exposure groups showed increased risks for specific cancer sites, such as lung cancer in workers exposed to IR [$RR = 2.71$ (95% CI = 1.14–6.43)]. The association of cancer risk with CA frequency was particularly evident in workers exposed to IR [2974 subjects, $RR_{\text{medium}} = 2.01$ (95% CI = 1.15–3.51), $RR_{\text{high}} = 1.81$ (95% CI = 1.03–3.20)] and at a lower extent in groups exposed to polyvinyl chloride and plastics [945 subjects, $RR_{\text{medium}} = 1.62$ (95% CI = 0.53–4.96), $RR_{\text{high}} = 1.73$ (95% CI = 0.58–5.19)], metals [662 subjects, $RR_{\text{medium}} = 3.10$ (95% CI = 0.64–14.97), $RR_{\text{high}} = 2.75$ (95% CI = 0.59–12.80)] and welding fumes [392 subjects, $RR_{\text{medium}} = 2.01$ (95% CI = 0.42–9.51), $RR_{\text{high}} = 2.63$ (95% CI = 0.56–12.28)]. The RRs among subjects not exposed to occupational carcinogens were 1.39 (medium, 95% CI = 0.86–2.25) and 1.54 (high, 95% CI = 0.90–2.64). Each occupational category was tested as a possible effect modifier of the association between CA and overall cancer risk, and none of the interaction terms were significant. The presence of effect modification due to occupational exposure to carcinogens was tested also for subclasses of CA and for major groups of cancer. All interactions tested turned out to be non-significant.

A similar analysis was performed to test the effect modification of cigarette smoking. When compared with non-smokers, ever smokers experienced an increased risk of all cancers ($RR = 1.53$, 95% CI = 1.28–1.83) and especially lung cancer ($RR = 6.65$, 95% CI = 3.17–13.95). The association between CA and risk of cancer was stronger in ever smokers [316 cases, $RR_{\text{medium}} = 1.42$ (95% CI = 1.04–1.94), $RR_{\text{high}} = 1.54$ (95% CI = 1.14–2.08); test for linear trend, $P = 0.008$] than in non-smokers [208 cases, $RR_{\text{medium}} = 1.30$ (95%

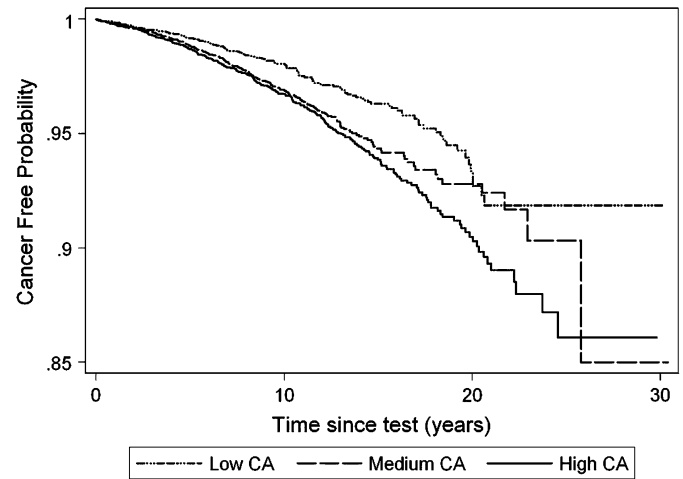


Fig. 2. Kaplan–Meier curves for total cancer incidence (International Classification of Diseases IX 140–208) tertile of CA frequency based on pooled data from 11 European cohorts. Cancer-free probability refers to time from CA test to the first cancer diagnosis.

CI = 0.93–1.82), $RR_{\text{high}} = 1.18$ (95% CI = 0.83–1.68)], but the interaction was not statistically significant ($P = 0.53$).

The association between cancer risk and CA frequency was consistent over the follow-up time, as shown by the Kaplan–Meier curves of the probability to be cancer free (Figure 2), which indicated a significantly worse outcome in subjects in the medium and high CA tertiles (log-rank test $P < .001$). Survival curves based on cancer mortality showed similar results (data not shown).

Discussion

The results of the present study, which included data from all CA cohorts published so far, confirm that CA frequency measured in peripheral blood lymphocytes of healthy subjects is associated with cancer risk, although the strength of this conclusion is somewhat weakened by the lack of a clear dose–response in most national cohorts. The independence of these findings from the time elapsed between cytogenetic testing and cancer diagnosis supports the assumption that increased CA levels in high-risk subjects are not attributable to clinically undetected cancers.

The pooled approach applied in this study gave several advantages over a meta-analysis of published data, including the possibility to perform interaction tests and subgroup analyses. Nevertheless, pooling data of biomarkers such as CA, which are characterized by variability of protocols, equipment and methods, may pose a serious question of heterogeneity. To address this issue, data from each laboratory were categorized in three frequency categories, an approach which according to the Akaike information criteria was more efficient than other methods of standardization.

A possible source of misclassification is the use of a single CA measure to classify subjects according to CA frequency. Due to the availability of a repeated measure for 11 909 subjects in the Czech cohort (99.3%), we could evaluate the agreement between the two measures in this database. 76.1% of these subjects were classified in the same category for both tests, whereas for 11.6 and 12.2%, the second test would have resulted in a higher or lower class compared with the first test, respectively. To evaluate if the choice of the test may have influenced the strength or the direction of the association, we compared risk estimates based on tertiles calculated with the first and the second assay. The results were very similar, i.e. $RR_{\text{medium}} = 1.27$ and 1.24 and $RR_{\text{high}} = 1.17$ and 1.21, respectively, showing that the classification based on a single measure of CA—at least in the Czech cohort—did not affect the estimate of the association between CA and cancer.

The interpretation of the study findings at country level is complex. Although a dose–response was present only in Hungary and Norway (test for linear trend, $P = 0.002$ and $P = 0.030$, respectively), the risk estimates were above unity in most cohorts, confirming the importance of a pooled design to reach the adequate statistical power to test this hypothesis.

The large size of the pooled database allowed us to investigate in greater detail hypotheses not properly tested in the original studies, in particular, whether the frequency of chromosome-type and chromatid-type aberrations has a different predictive value for cancer risk. The question is of great relevance since the induction of these aberrations reflects different mechanisms. In peripheral lymphocytes, chromosome-type CAs are induced by IR and other S-independent clastogens, causing DNA double-strand breaks, whereas chromatid-type CAs are produced by S-dependent agents such as many chemical mutagens. The published evidence is contrasting since earlier Nordic–Italian studies found that both subclasses are equally predictive of cancer risk (22), whereas the larger cohorts from the Czech Republic (11) and Central Eastern European countries (12) suggest that chromosome-type CA may show a stronger association. The results of the pooled estimates reflect the prevailing effect found in these latter cohorts. However, the limited improvement over the results based on total CA does not support the use of chromosome-type aberrations as the most predictive biomarker of risk.

The increased cancer risk observed in subjects bearing ring chromosomes (20 cases) and the borderline increase observed for marker chromosomes (limited to mortality) are not circumstantiated enough to be considered for future risk assessment protocols, and chance of error measurement remain plausible explanations. However, these findings should further be considered, given the mechanistic evidence linking chromosomal rearrangements to early stages of cancer.

Information about subclasses of chromosome- and chromatid-type aberrations was available only for a part of the cohort. To check if this subgroup was representative of the whole cohort, we re-evaluated cancer risks associated with total CA frequency in the same subset of subjects. The results were in agreement with a random selection of the population, showing risks very similar to those of the whole cohort [$RR_{\text{medium}} = 1.40$ (95% = 1.01–1.92), $RR_{\text{high}} = 1.67$ (95% CI = 1.22–2.28)].

The hypothesis that CA frequency is associated with specific cancers has been investigated since the first cohort studies. A stronger association was initially described with lung cancer and haematological malignancies (8), whereas more recently, based on larger numbers, both the Czech cohort (11) and the Central and Eastern European country cohort (12) reported an increased risk of stomach cancer. The analysis of this pooled databases showed the association of CA frequency with stomach cancer as the only cancer site significantly associated with CA frequency (Table III). The hypothesis of a spurious association due to the number of comparisons cannot be ruled out; however, there is supportive mechanistic evidence. The presence of chromosome instability among the most relevant events in intestinal and diffuse stomach cancers has extensively been described (23,24) and linked to the metabolisms of agents involved in stomach carcinogenesis, such as folic acid and vitamin B12 (25,26). Interestingly, a recently published cohort study linking the frequency of micronuclei in lymphocytes of healthy subjects to the risk of cancer reported stomach cancer among the sites more specifically associated with micronuclei frequency (27). A recent report indicated that *Helicobacter pylori* infection is associated with an increased level of micronuclei in peripheral lymphocytes (28).

Another open issue is the role of exposures to genotoxic and carcinogenic agents experienced by study subjects at the time of CA testing. A nested case–control study carried out within the Nordic–Italian databases to specifically test the presence of interaction of CA frequency with major occupational exposure to carcinogens and with cigarette smoking did not find any difference in the risks (10). However, this independence was later put in doubt by the observation that in a subcohort of Czech miners exposed to radon, subjects with high levels of CA showed a much higher risk of cancer than those in other

occupational groups of the same cohort (29). In the present study, the non-homogeneous quality of exposure assessment among studies limited the possibility to test by the presence of interaction. However, with an exploratory approach, study subjects were grouped by genotoxic exposure and risks were estimated within each subcohort. The small increase of cancer risk in the whole group of workers exposed to carcinogens when compared with non-exposed workers can be explained by exposure misclassification. The reliability of the design was confirmed by the high risks found for well-known associations between carcinogenic agents and tumour sites, such as those reported previously for lung cancer in exposure to IR. No clear evidence of effect modification was found for any exposure, although there was a suggestion of a higher risk associated with CA frequency for subjects exposed to IR. The estimate of cancer risk associated with smoking habit was more reliable, as supported by the close to expected RRs found in smokers for all cancers and lung cancer. However, also for this exposure, results of the interaction tests—although affected by a limited statistical power—are in agreement with the hypothesis that exposure is not an effect modifier of the association between CA and cancer.

The results of this study represent the synthesis of research conducted over two decades aiming at validating biomarkers of chromosomal damage as predictors of cancer occurrence in groups of healthy individuals. One of the indirect results of this effort was the establishment of an international database of over 20 000 subjects screened for CA, which can be followed up to provide further and more reliable data regarding the use of this biomarker to identify high-risk populations.

Funding

European Union (Cytogenetic Biomarkers and Human Cancer Risk: QLK4-CT-2000-00628 and QLK4-CT-2002-02831) to S.B., H.N., U.S., A.C.-W., E.F., A.F., S.G., I.-L.H., L.E.K., J.L., P.R., R.J.S. and P.B.; Environmental Cancer Risk, Nutrition and Individual Susceptibility Network of Excellence (513943) to R.V. and P.B.; the Italian Association for Cancer Research to S.B.; the Italian Space Agency to S.B.; the National Cancer Institute of the USA to R.V., A.Z. and A.F.; the Finnish Work Environment Fund to H.N.; and the Czech Ministry of the Environment (VaV-SL/5/160/05) to P.R. and R.J.S.

Acknowledgements

The Cancer Registry of Norway is not responsible for the analysis or interpretation of the Norwegian cancer data presented. We wish to acknowledge the important contribution to this research by the late Lars Hagmar, Lund University, Sweden, who authored seminal cohort studies on the link between CA and cancer.

Conflict of Interest Statement: None declared.

References

- Mitelman, F. *et al.* (2004) Fusion genes and rearranged genes as a linear function of chromosome aberrations in cancer. *Nat. Genet.*, **36**, 331–334.
- Mitelman, F. *et al.* (2007) The impact of translocations and gene fusions on cancer causation. *Nat. Rev. Cancer*, **7**, 233–245.
- Bonassi, S. *et al.* (2004) Chromosomal aberrations and risk of cancer in humans: an epidemiological perspective. *Cytogenet. Genome Res.*, **104**, 376–382.
- Nowell, P.C. *et al.* (1961) Chromosome studies in human leukemia. II. Chronic granulocytic leukemia. *J. Natl Cancer Inst.*, **27**, 1013–1021.
- Holmquist, G.P. (1992) Chromosome bands, their chromatin flavors, and their functional features. *Am. J. Hum. Genet.*, **51**, 17–37.
- Khanna, K.K. *et al.* (2001) DNA double-strand breaks, signaling, repair and the cancer connection. *Nat. Genet.*, **27**, 247–254.
- Hagmar, L. *et al.* (1994) Cancer risk in humans predicted by increased levels of chromosome aberrations in lymphocytes: Nordic Study Group on the Health Risk of Chromosome Damage. *Cancer Res.*, **54**, 2919–2922.

8. Bonassi,S. *et al.* (1995) Are chromosome aberrations in circulating lymphocytes predictive of a future cancer onset in humans? Preliminary results of an Italian cohort study. *Cancer Genet. Cytogenet.*, **79**, 133–135.
9. Hagmar,L. *et al.* (1998) Chromosomal aberrations in lymphocytes predict human cancer: a report from the European Study Group on Cytogenetic Biomarkers and Health (ESCH). *Cancer Res.*, **58**, 4117–4121.
10. Bonassi,S. *et al.* (2000) Chromosomal aberrations in lymphocytes predict human cancer independently of exposure to carcinogens. *Cancer Res.*, **60**, 1619–1625.
11. Rossner,P. *et al.* (2005) Chromosomal aberration in lymphocytes of healthy subjects and risk of cancer. *Environ. Health Perspect.*, **113**, 517–520.
12. Boffetta,P. *et al.* (2007) Chromosomal aberrations and cancer risk: results of a cohort study from Central Europe. *Am. J. Epidemiol.*, **165**, 36–43.
13. Nordic Study Group on the Health Risk of Chromosome Damage. (1990) An inter-Nordic prospective study on cytogenetic endpoints and cancer risk. *Cancer Genet. Cytogenet.*, **45**, 85–92.
14. Agresti,A. (1990) *Categorical Data Analysis*. John Wiley & Sons Inc, New York, NY, 251.
15. Savage,J.R. (1976) Classification and relationships of induced chromosomal structural changes. *J. Med. Genet.*, **13**, 103–122.
16. Tinnerberg,H. *et al.* (2003) Retrospective exposure assessment and quality control in an international multicentre case control study. *Ann. Occup. Hyg.*, **47**, 37–47.
17. Lindsey,J.K. (1995) *Modelling Frequency and Count Data*. Oxford University Press, Oxford, UK.
18. Leyland,A.H. *et al.* (2001) *Multilevel Modelling of Health Statistics*. John Wiley & Sons Ltd, Chichester, UK.
19. Grambsch,P.M. *et al.* (1994) Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika*, **81**, 515–526.
20. Rasbash,J. *et al.* (2001) MLwiN version 1.10. Multilevel Models Project, Institute of Education, University of London, London, UK.
21. StataCorp. (2004) *Stata Statistical Software: Release 8*. StataCorp LP, College Station, TX.
22. Hagmar,L. *et al.* (2004) Impact of types of lymphocyte chromosomal aberrations on human cancer risk: results from Nordic and Italian cohorts. *Cancer Res.*, **64**, 2258–2263.
23. Tahara,E. (2004) Genetic pathways of two types of gastric cancer. In Buffler,P., Rice,J. and Baan,R. (eds.) *Mechanisms of Carcinogenesis: Contributions of Molecular Epidemiology*. IARC Scientific Publications No. 157, IARC, Lyon. 327–349.
24. Tsafir,D. *et al.* (2006) Relationship of gene expression and chromosomal abnormalities in colorectal cancer. *Cancer Res.*, **66**, 2129–2137.
25. Blount,B.C. *et al.* (1997) Folate deficiency causes uracil misincorporation into human DNA and chromosome breakage, implications for cancer and neuronal damage. *Proc. Natl Acad. Sci. USA*, **94**, 3290–3295.
26. Kimura,M. *et al.* (2004) Methylenetetrahydrofolate reductase C677T polymorphism, folic acid and riboflavin are important determinants of genome stability in cultured human lymphocytes. *J. Nutr.*, **134**, 48–56.
27. Bonassi,S. *et al.* (2007) An increased micronucleus frequency in peripheral blood lymphocytes predicts the risk of cancer in humans. *Carcinogenesis*, **28**, 625–631.
28. Suárez,S. *et al.* (2007) Increased frequency of micronuclei in peripheral blood lymphocytes of subjects infected with *Helicobacter pylori*. *Mutat. Res.*, **626**, 162–170.
29. Smerhovsky,Z. *et al.* (2001) Risk of cancer in an occupationally exposed cohort with increased level of chromosomal aberrations. *Environ. Health Perspect.*, **109**, 41–45.

*Received January 18, 2008; revised March 5, 2008;
accepted March 6, 2008*