AMOUNT OF DNA IN PLASMA AND CANCER RISK: A PROSPECTIVE STUDY

Emmanuelle Gormally[1], Pierre Hainaut[1], Elodie Caboux[1], Luisa Airoldi[2], Herman Autrup[3], Christian Malaveille[1], Alison Dunning[4], Seymour Garte[5], Giuseppe Matullo[6], Kim Overvad[7,8], Anne Tjonneland[9], Francoise Clavel-Chapelon[10], Paolo Boffetta[11], Heiner Boeing[12], Antonia Trichopoulou[13], Domenico Palli[14], Vittorio Krogh[15], Rosario Tumino[16], Salvatore Panico[17], H. Bas Bueno-De-Mesquita[18], Petra H. Peeters[19], Eiliv Lund[20], Carlos A. Gonzalez[21], Carmen Martinez[22], Miren Dorronsoro[23], Aurelio Barricarte[24], M. Jose Tormo[25], J. Ramón Quiros[26], Goran Berglund[27], Goran Hallmans[28], Nicholas E. Day[29], Timothy J. Key[30], Fabrizio Veglia[6], Marco Peluso[31], Teresa Norat[1], Rodolfo Saracci[1], Rudolf Kaaks[1], Elio Riboli[1] and Paolo Vineis[6,32*]

[1]*International Agency for Research on Cancer, Lyon, France*
[2]*Istituto Mario Negri, Milan, Italy*
[3]*Department of Occupational and Environmental Medicine, Aarhus, Denmark*
[4]*Department of Oncology, University of Cambridge, Cambridge, United Kingdom*
[5]*Genetics Research Institute, Milan, Italy*
[6]*Institute for Scientific Interchange Foundation, Turin, Italy*
[7]*Department of Clinical Epidemiology, Aalborg Hospital and Aarhus University Hospital, Aarhus, Denmark*
[8]*Department of Epidemiology and Social Medicine, University of Aarhus, Aarhus, Denmark*
[9]*Institute of Cancer Epidemiology, Danish Cancer Society, Copenhagen, Denmark*
[10]*INSERM U521, Institut Gustave Roussy, Villejuif, France*
[11]*Division of Clinical Epidemiology, Deutsches Krebsforschungszentrum, Heidelberg, Germany*
[12]*German Institute of Human Nutrition, Potsdam-Rehbücke, Germany*
[13]*Department of Hygiene and Epidemiology, Medical School, University of Athens, Athens, Greece*
[14]*Molecular and Nutritional Epidemiology Unit, Centro per lo Studio e la Prevenzione Oncologica, Scientific Institute of Tuscany, Florence, Italy*
[15]*Department of Epidemiology, National Cancer Institute, Milan, Italy*
[16]*Ragusa Cancer Registry, Azienda Ospedaliera "Civile M.P. Arezzo," Ragusa, Italy*
[17]*Department of Clinical and Experimental Medicine, Università Federico II, Naples, Italy*
[18]*Center for Nutrition and Health, National Institute for Public Health and the Environment, Bilthoven, The Netherlands*
[19]*Julius Center for Health Sciences and Primary Care, University Medical Center, Utrecht, The Netherlands*
[20]*Institute of Community Medicine, University of Tromso, Tromso, Norway*
[21]*Department of Epidemiology, Catalan Institute of Oncology, Barcelona, Spain*
[22]*Andalusian School of Public Health, Granada, Spain*
[23]*Department of Public Health of Guipuzkoa, San Sebastian, Spain*
[24]*Public Health Institute, Navarra, Spain*
[25]*Consejería de Sanidad y Consumo, Murcia, Spain*
[26]*Dirección General de Salud Pública, Consejería de Salud y Servicios Sanitarios Asturias, Oviedo, Spain*
[27]*Malmö Diet and Cancer Study, Lund University, Malmö, Sweden*
[28]*Department of Nutritional Research, University of Umeå, Umeå, Sweden*
[29]*M.R.C. Dunn Human Nutrition Unit, Cambridge, United Kingdom*
[30]*Cancer Research U.K. Epidemiology Unit, University of Oxford, Oxford, United Kingdom*
[31]*Laboratorio di Biologia Molecolare, Centro per lo Studio e la Prevenzione Oncologica, Florence, Italy*
[32]*Cancer Epidemiology Unit, University of Turin and Centro Prevenzione Oncologica-Piemonte, Turin, Italy*

Levels of plasma DNA concentrations in cancer patients have been shown to be higher than the plasma DNA concentrations found in healthy subjects. The value of plasma DNA levels for development of neoplastic or pulmonary disease was evaluated in a large prospective study. Plasma samples ($n = 1,184$) were analyzed from 776 controls, 359 cases of cancer (lung, bladder, oral cavity, pharynx, larynx, leukemia) and 49 deaths from chronic obstructive pulmonary disease (COPD), including never smokers and ex-smokers, from 9 countries across Europe. The amount of plasma DNA was variable across the European Prospective Investigation into Cancer and Nutrition (EPIC) centers. High DNA concentrations in some centers might be due to the type of population recruited and/or the treatment of the samples. An elevated and statistically significant odds ratio (OR) was found for COPD deaths (OR = 2.53; 95% CI = 1.06–6.02), while nonsignificant increased ORs were present for oral cancers, cancers of the pharynx and larynx and leukemia. When the analyses were stratified by time since recruitment (below or above 36 months), the increased ORs were limited to the more recent period of recruitment, *i.e.,* a time elapsed between blood drawing and disease onset lower than 36 months. This was particularly true for COPD deaths (OR = 12.7; 95% CI = 1.57–103) and leukemia (OR = 2.37; 95% CI = 1.20–4.67).
© *2004 Wiley-Liss, Inc.*

Key words: *plasmatic DNA; cancer; prospective studies; molecular repidemiology*

Tumoral materials such as cells, DNA, RNA and proteins can be recovered from blood, urine, feces, pancreatic juice and sputum of cancer patients.[1,2] In blood, the amount of DNA in cancer patients is sometimes 100- to 1,000-fold higher than in healthy subjects. The detection of genetic and epigenetic modifications in tumors and matched plasma or serum DNA has demonstrated that part of this DNA is of tumoral origin.[3] However, the exact nature and

origin of this DNA are still controversial. While early studies have indicated that most of the DNA in the plasma is in the form of large quasigenome-size segments, more recent works have clearly detected the presence of DNA laddering, a hallmark of apoptosis. So far, no correlations between tumor site, stage and size, prognosis and amounts of plasma DNA have been clearly demonstrated.[4–7] In previous investigations, cancer patients had an average of 180 ng of DNA per milliliter of blood, with amounts ranging from 0 to 5,000 ng/ml, while healthy individuals had an average of 13 ng DNA per ml, ranging from 0 to 100 ng/ml.[3] This observation raises the possibility that plasma DNA levels may be indicative of disease status. However, the relationship between amounts of plasma DNA and cancer has only been considered according to a cross-sectional design, while prospective studies are needed to evaluate the value of plasma DNA with respect to etiology and prognosis of cancer at different sites.

In this study, we have evaluated the predictive value of plasma DNA levels in a large longitudinal study, European Prospective Investigation into Cancer and Nutrition (EPIC).[8] The main advantage of the longitudinal design is that at the time blood was collected, participants were free of cancer.

MATERIAL AND METHODS

Selection of subjects and collection of specimens: the EPIC/Gen-Air prospective study

EPIC is a multicenter European study coordinated by the International Agency for Research on Cancer (Lyon, France), in which more than 500,000 healthy volunteers were recruited in 22 centers from 9 European countries (Sweden, Denmark, The Netherlands, United Kingdom, France, Germany, Spain, Italy, Greece). Norway joined EPIC after the Gen-Air study was started. The cohort includes subjects of both sexes, mostly in the age range of 35–74 at recruitment. Recruitment took place between 1993 and 1999. Detailed dietary and lifestyle histories are available from self-administered questionnaires, plus a 24-hr dietary recall through person-to-person interview for an 8% subsample of the EPIC cohort, anthropologic measurements and a 40 ml blood sample. All questionnaire information collected in the initial study is available in a computerized format. Informed consent was obtained from all participants. In this context, EPIC researchers are identifying all incident cases of cancer and deaths occurring in the cohort. Cases are identified through cancer registries and vital statistics in all centers, plus hospital discharge records in some centers. There is a detailed protocol for the identification of all cancers and deaths, with virtually 100% completeness. Procedure are monitored by an *ad hoc* EPIC Endpoint Committee.

The Gen-Air nested study identified all cases of lung, bladder, pharynx, larynx and oral cancers and all leukemias and deaths from chronic obstructive pulmonary disease (COPD) within EPIC. Cases are matched 1:2 with controls having the same age at recruitment ($\pm$ 5 years), time since blood drawing, sex, country and smoking status. All cases and controls are never smokers or ex-smokers for at least 10 years. Three hundred fifty-nine cancer cases (including 89 bladder, 82 lung, 28 oral, 31 pharynx or larynx cancers, 129 leukemias), 49 COPD deaths and 776 controls were identified at the end of 2001 and were included in the present analyses. We were not able to match 40 cases. Mean follow-up was 89 months (range, 51–123 months).

Extraction and analysis of plasma DNA

For 1,184 participants, plasma DNA was extracted from samples of 300 μl of plasma and diluted with 200 μl of water using a standard method based on affinity purification (QiaGen, Hilden, Germany) that we have extensively validated in previous studies.[9]

DNA extracted from plasma was quantified with the PicoGreen dsDNA Quantification Reagent Kit (Molecular Probes, Leiden, The Netherlands). PicoGreen specifically binds to dsDNA and after excitation at 485 nm the complex dsDNA/PicoGreen fluorescence is detected at 538 nm. A calibration curve ranging between 0.2 and 0.0005 ng/μl using DNA at a known concentration (100 μg/ml) provided in the kit was included in each experiment. The concentrations of the samples were calculated twice at 2 different dilutions. Five microliters and 10 μl of DNA were diluted in a 200 μl reaction mixture containing 1 $\times$ TE (10 mM Tris-HCl, 1 mM EDTA, pH 7.5) and PicoGreen (1:400). After 5 min of incubation, the reaction was excited and the fluorescence was recorded with a fluoremeter (Labsystem, Helsinki, Finland). Plasma DNA concentrations were estimated using the calibration curve. An average of the 2 concentrations obtained with the 2 dilutions was calculated. If the concentrations at 1 in 5 and 1 in 10 dilutions differed, the concentrations were measured a second

TABLE I – DISTRIBUTION OF PLASMA DNA CONCENTRATIONS BY CENTERS WITH AT LEAST 5 SUBJECTS

Center	Number of subjects	Arithmetic mean	Geometric mean	Range
Ile-de-France	6	17	16	11–25
Northeast of France	5	18	18	15–23
Florence	47	23	16	2.9–170
Varese	52	23	15	1–183
Ragusa	9	23	30	10–46
Turin	41	190	100	5–1,128
Asturias	25	55	27	1–491
Granada	26	30	25	2.8–97
Murcia	16	29	26	11–62
Navarra	31	80	53	1.3–315
San Sebastian	17	31	24	6.8–73
Cambridge	292	44	31	1.5–905
Oxford	78	316	106	5.2–10,974
Bilthoven	13	38	29	13–169
Utrecht	66	75	59	7–287
Greece	49	76	47	2–612
Heidelberg	67	44	16	0–601
Potsdam	102	33	25	3–304
Umeåa	119	30	17	3–984
Aarhus	29	29	17	0.2–126
Copenhagen	55	44	18	3–1,159

DNA concentrations are expressed in ng/ml.

TABLE II – PLASMA DNA CONCENTRATIONS FOR CONTROLS, COPD DEATHS AND THE DIFFERENT TYPES OF TUMORS

	n	Mean		Median	Range	Standard deviation	p-value[1]
		Arithmetic	Geometric				
Controls	776	67	28	26	0–10	405	
COPD deaths	49	85	47	41	9–708	134	0.005
Cancers							
Bladder	89	73	31	26	3–1,588	186	0.31
Leukemia	129	72	36	34	0–1,159	127	0.008
Lung	82	65	29	25	2–1,128	143	0.64
Oral	28	62	32	28	5–520	104	0.42
Pharynx/larynx	31	86	30	33	1–1,568	277	0.57

DNA concentrations are expressed in ng/ml.– [1]Univariate comparison with controls after log transformation (non-parametric test)

time. If the results were still discordant, the sample was not included in the analysis.

Statistical analyses

Arithmetic and geometric means, medians and standard deviations were computed for cases and controls. Nonparametric ANOVA was used for comparisons after log transformation. The latter was justified by the skewed distribution of the values of plasma DNA. Logistic regression (based on Cox proportional hazards model) was used to estimate odds ratios (ORs) of overall and specific disease status for the amount of plasma DNA by categorizing plasma DNA as a dichotomous variable (below/above median value in controls). Both conditional (for matched pairs) and unconditional models were fitted. Multivariate analyses were performed with the SAS package.

RESULTS

The amount of DNA was highly variable across EPIC centers ($p < 0.0001$). Table I gives a summary of the statistics by center. The highest concentrations were found in samples coming from Oxford (mean, 316 ng/ml, based on 78 subjects) and Turin (mean, 190 ng/ml, based on 41 subjects). Geometric means showed a similar variation between centers.

TABLE III – UNIVARIATE AND MULTIVARIATE ANALYSIS OF PLASMA DNA AMOUNT

	F-value	DF	*p*-value
Univariate analysis			
Controls only			
Center	11.23	22	< 0.0001
Age	5.21	1	0.023
Gender	0.52	1	0.47
Multivariate analysis			
Cases and controls (model including age and center)			
Center	16.6	23	< 0.0001
Age	1.56	1	0.21
All death and tumours	2.3	6	0.03
Single disease			
COPD deaths			0.002
Bladder cancer			0.44
Leukemia			0.028
Lung cancer			0.83
Oral cancer			0.57
Pharynx-larynx cancer			0.65

As a continuous variable (logarithm transformation, dependent variable) by center, age (continuous variable), gender and time between blood drawing and diagnosis (for cases only). n = 1,184.

Overall summary statistics of the concentrations of plasma DNA isolated from the 1,184 subjects are shown in Table II. The average amount of plasma DNA in the various pathology groups ranged from 62 to 86 ng/ml, with considerable variations as illustrated by high standard deviations. Comparison with controls in univariate analysis indicated a significant difference for COPD deaths ($p = 0.005$) and leukemia ($p = 0.008$).

Table III shows that plasma DNA increased significantly with increasing age of the subjects (average increase of 0.6 ng/ml per year of age), but the association disappeared in multivariate analysis. No association was found with sex (Table III) or smoking status.

In multivariate analysis (Table III), the amount of plasma DNA, as a continuous variable, was apparently associated with overall cancer onset, and also with death from COPD, after adjustment for age and center. The only single type of tumor showing a statistically significant association with the amount of DNA was leukemia. Inclusion of smoking status did not affect estimates. After exclusion of the 2 centers with the largest amount of DNA, there was still a weak association with all cancers ($p = 0.05$) and with deaths from COPD ($p = 0.05$). However, a strong heterogeneity between centers remained in multivariate analysis ($p < 0.0001$).

Table IV shows the results of fitting conditional logistic regression models for matched pairs. ORs are adjusted by matching variables (age, sex, time since blood drawing, smoking status) and, in addition, by center. An elevated and statistically significant OR is shown for COPD deaths, while nonsignificant ORs are present for oral cancers, cancers of pharynx and larynx and leukemia. Exclusion of Oxford and Turin changed the estimates slightly. When the analyses were stratified by time since recruitment (below or above 36 months), the increased ORs were limited to the more recent period of recruitment, *i.e.*, a time elapsed between blood drawing and disease onset lower than 36 months (Table IV). The strong association with COPD below 36 months is particularly noteworthy (OR = 12.7); also, the association with leukemia is stronger (OR = 2.37) and statistically significant below 36 months.

DISCUSSION

There is clear evidence that cancer patients have larger amounts of cell-free plasma DNA than healthy subjects.[3] Although previous studies suggest that elevated plasma DNA levels may predict neoplastic disease, the limitation of these studies is their cross-sectional nature, *i.e.*, the fact that plasma DNA was measured in patients whose cancer had already been diagnosed.[4,5] To our knowledge, our study is the first prospective investigation that considers the relationship of the amount of DNA in plasma with the risk of cancer or pulmonary disease.

TABLE IV – ODDS, RATIOS (CONDITIONAL LOGISTIC REGRESSION MODEL) AND 95% CONFIDENCE INTERVALS FOR THE AMOUNT OF PLASMA DNA

Disease	OR	95% CI	After exclusion of Oxford and Turin OR	95% CI
COPD	2.53	1.06–6.02	2.0	0.76–5.24
Bladder cancer	1.0	0.53–1.77	0.70	0.33–1.51
Leukemia	1.27	0.74–2.17	1.25	0.72–2.18
Lung cancer	0.66	0.31–1.36	0.48	0.20–1.17
Oral cancer	1.79	0.62–5.15	1.78	0.62–5.14
Pharynx-larynx cancer	1.48	0.50–4.34	1.38	0.41–4.68
All cancers	1.05	0.76–1.46	1.02	0.72–1.43
Stratification by time since recruitment	**Less than 36 months**		**More than 36 months**	
COPD	12.7	1.57–103	1.34	0.46–3.88
Bladder cancer	0.98	0.49–1.96	0.94	0.29–3.07
Leukemia	2.37	1.20–4.67	0.27	0.09–0.86
Lung cancer	0.80	0.32–2.02	0.47	0.14–1.60
Oral, pharynx-larynx cancer	1.09	0.40–2.95	2.77	0.83–9.24
All cancers	1.30	0.87–1.95	0.70	0.39–1.26

Dichotomous variable, above vs. below median. The models include center as a covariate.

A statistically significant association was found between DNA amount and death from COPD. An association with leukemia was found when using plasma DNA as a continuous variable, but not in logistic models with plasma DNA dichotomized. No significant association was found for other cancers tested. Interestingly, the associations were stronger when analyses were limited to cases that arose less than 36 months after blood drawing. This is consistent with the interpretation that plasmatic DNA comes from the disease process itself. An odds ratio as high as 12.7 was found for COPD deaths, but the OR for leukemia also was particularly high among the cases detected during the first 36 months of follow up (OR = 2.37).

Considerable variation between centers was noted. The high DNA concentrations in Oxford and Turin, in particular, might be due to the type of population recruited and/or the treatment of the samples. In the centers with the highest plasma DNA concentrations, blood samples might have been prepared and frozen many hours after blood drawing, probably allowing some lysis of white blood cells to occur and the release of their DNA in the plasma. In Oxford, samples were posted from general practices throughout the United Kingdom to one central laboratory. In Turin, however, most of the samples were processed in a few hours, and only a small minority were centrifuged after 24 hr. Therefore, the reasons for center variability are not completely clear.

The association between death from COPD and DNA amount indicates that long-lasting inflammatory processes are likely associated with increased cell death/apoptosis that leads to DNA release a few years or months before death. This observation points to the fact that many pathologic processes other than cancer can result in the release of intact or fragmented DNA in blood. In a recent case-control study of the size of DNA fragments in the plasma of subjects with or without chronic liver disease including liver cancer, we found that plasma DNA from patients with liver cirrhosis had a tendency to contain fragmented apoptotic-type DNA, whereas the plasma DNA from tumor patients essentially consisted of large genome-size segments (data not shown). The large coefficient of variation, *i.e.*, interindividual variation, is worth noting. Though our method did not specifically measure human DNA, microorganisms such as viruses are present in the plasma at levels that are too low to be a significant source of DNA contamination.[10]

In conclusion, this prospective study suggests that the amount of DNA in plasma collected from healthy individuals is associated with COPD deaths and, more weakly, with leukemia. This observation is interesting from a scientific point of view but cannot be used for clinical purposes. This conclusion is compatible with the one drawn by Chang *et al.*[5] in a case-control study of 330 subjects, including healthy subjects, patients with various neoplastic diseases and patients with nonneoplastic diseases. It would be worthwhile to develop further studies focusing on the presence of large genome-size fragments in the plasma, since the occurrence of the latter is more likely to reflect a process of DNA release that is preferentially associated with cancer.

REFERENCES

1. Hu YC, Sidransky D, Ahrendt SA. Molecular detection approaches for smoking associated tumors. Oncogene 2002;21:7289–97.
2. Sidransky D. Advances in cancer detection. Sci Am 1996;275:104–9.
3. Anker P, Stroun M. Circulating DNA in plasma or serum. Medicina (B Aires) 2000;60:699–702.
4. Beau-Faller M, Gaub MP, Schneider A, Ducrocq X, Massard G, Gasser B, Chenard MP, Kessler R, Anker P, Stroun M, Weitzenblum E, Pauli G, Wihlm JM, Quoix E, Oudet P. Plasma DNA microsatellite panel as sensitive and tumor-specific marker in lung cancer patients. Int J Cancer 2003;105:361–70.
5. Chang HW, Lee SM, Goodman SN, Singer G, Cho SK, Sokoll LJ, Montz FJ, Roden R, Zhang Z, Chan DW, Kurman RJ, Shih I. Assessment of plasma DNA levels, allelic imbalance, and CA 125 as diagnostic tests for cancer. J Natl Cancer Inst 2002;94:1697–703.
6. Holdenrieder S, Stieber P, Forg T, Kuhl M, Schulz L, Busch M, Schalhorn A, Seidel D. Apoptosis in serum of patients with solid tumours. Anticancer Res 1999;19:2721–4.
7. Sozzi G, Conte D, Mariani L, Lo VS, Roz L, Lombardo C, Pierotti MA, Tavecchio L. Analysis of circulating tumor DNA in plasma at diagnosis and during follow-up of lung cancer patients. Cancer Res 2001;61:4675–8.
8. Riboli E, Kaaks R. The EPIC project: rationale and study design. Int J Epidemiol 1997;26(Suppl 1):S6–14.
9. Kirk GD, Camus-Randon AM, Mendy M, Goedert JJ, Merle P, Trepo C, Brechot C, Hainaut P, Montesano R. Ser-249 p53 mutations in plasma DNA of patients with hepatocellular carcinoma from the gambia. J Natl Cancer Inst 2000;92:148–53.
10. Gallagher A, Armstrong AA, MacKenzie J, Shield L, Khan G, Lake A, Proctor S, Taylor P, Clements GB, Jarrett RF. Detection of Epstein-Barr virus (EBV) genomes in the serum of patients with EBV-associated Hodgkin's disease. Int J Cancer 1999;84:442–8.