American Journal of Epidemiology
Vol. 175, No. 6
DOI: 10.1093/aje/kwr430
Advance Access publication:
February 20, 2012


## Original Contribution

# Association Between Total Number of Deaths, Diabetes Mellitus, Incident Cancers, and Haplotypes in Chromosomal Region 8q24 in a Prospective Study

**Simonetta Guarrera, Fulvio Ricceri, Silvia Polidoro, Carlotta Sacerdote, Alessandra Allione, Fabio Rosa, Floriana Voglino, Rossana Critelli, Alessia Russo, Paolo Vineis, and Giuseppe Matullo***

* Correspondence to Dr. Giuseppe Matullo, Department of Genetics, Biology and Biochemistry, University of Torino, 19 Via Santena, 10126 Torino, Italy (e-mail: giuseppe.matullo@unito.it).

*Initially submitted December 22, 2010; accepted for publication August 10, 2011.*

The 8q24 region is a gene desert, although chromosomal aberrations and somatic amplification involving this region, including translocations involving the protooncogene *c-MYC*, have been frequently reported in people with cancer. To investigate the role of variants in 8q24 region, the authors analyzed data from a prospective study ($n = 10{,}372$ participants who were followed for 11 years) in which a large number of health events (>1,500) occurred (1993–1998). They genotyped all subjects for 5 candidate single nucleotide polymorphisms (rs672888, rs1447295, rs9642880, rs16901979, and rs6983267) that were identified in previous genome-wide scans. Although significant associations with individual single nucleotide polymorphisms were small in magnitude, the authors observed higher increases in the risks of different types of cancer with specific haplotypes, particularly when subjects were homozygous for the haplotype: for breast cancer and homozygotes for haplotype CAGCT, hazard ratio = 3.40, 95% confidence interval: 1.24, 9.21; for prostate cancer and grouped rare haplotypes, hazard ratio = 7.43, 95% confidence interval: 3.00, 18.37; and for brain cancer and homozygotes for haplotype CGGCT, hazard ratio = 13.48, 95% confidence interval: 3.00, 59.53. Significant associations were also observed between haplotypes and deaths from cardiovascular diseases and cerebrovascular diseases; the most stable association was between homozygotes for haplotypes CGTCG and CAGCT and total deaths in men (hazard ratio = 3.5, 95% confidence interval: 1.8, 6.9, and hazard ratio = 2.8, 95% confidence interval: 1.3, 6.4, respectively). In conclusion, the authors have observed a strong pleiotropic effect of the 8q24 region in a large prospective study. This observation can shed light on the mechanisms underlying reported associations between 8q24 variants and disparate chronic diseases.

genetic pleiotrophy; haplotypes; prospective studies

Abbreviations: CI, confidence interval; FPRP, false-positive report probability; HR, hazard ratio; LD, linkage disequilibrium; SNP, single nucleotide polymorphism; TCF7L2, transcription factor 7-like 2.

***Editor's note:*** *An invited commentary on this article appears on page 488.*

Single nucleotide polymorphisms (SNPs) in the 8q24 chromosome region have been associated with prostate cancer in genome-wide association studies (1–5). Other cancers were subsequently associated with this region, including colorectal cancer (6–10), breast cancer (11, 12), pancreatic cancer (13), bladder cancer (14–16), thyroid cancer (17), smoking-related cancers (15), chronic lymphocytic leukemia (18), testicular germ tumors (19), and gliomas (20). Investigators have also found relations between other conditions, such as diabetes mellitus (21), age-related hearing impairment (22), schizophrenia (23, 24), bipolar disorder (25), and cleft lip and palate (26, 27) and SNPs in this region. Although a wide range of diseases was observed to be associated with a variety of SNPs, one SNP in particular (rs6983267) was associated with multiple outcomes. The 8q24 region is a gene desert. Although some transcripts have been described in this

region, such as *ψSRRM1* (serine/arginine repetitive matrix 1 pseudogene 1), *FAM84B* (family with sequence similarity 84, member B), *LOC727677* (uncharacterized LOC727677), *POU5F1P1* (POU class 5 homeobox 1B), and *PVT1* (Pvt1 oncogene (non-protein coding)), the only known strong candidate gene at the 3′ end of the region is the protooncogene *c-MYC* (c-myc myelocytomatosis viral oncogene homolog (avian)). Region 8q24 contains fragile sites that have been linked to translocations involving *MYC* (28). Also, it has been proposed that this region is involved in the regulation of *MYC*, but the evidence is conflicting (29, 30). Finally, a role for microRNA has been suggested (31–33).

To our knowledge, all studies of this region conducted so far have been cross-sectional case-control studies in which usually only a single or a few diseases were considered, thus limiting investigation of gene pleiotropism. Also, the choice of controls can determine the quality of the study and the interpretation of the results. We used a cohort of more than 10,000 subjects (10,372 genotyped) to look at the multiple outcomes associated with 5 SNPs (rs672888, rs1447295, rs9642880, rs16901979, and rs6983267) in the 8q24 region, including all incident cancers, all causes of death, incident and prevalent cases of diabetes mellitus, and incident and prevalent cases of cardiovascular disease. In addition, we have extensive data on anthropometric measurements, dietary habits, smoking habits, and occupations of participants in our cohort. The 5 selected SNPs were all either associated with diseases (bladder, prostate, or colorectal cancer) or reported to be tagSNPs for a previously reported disease-associated SNP (e.g., rs672888 has been reported to tag the breast cancer-associated SNP rs13281615; $R^2 = 0.967$). Analysis of the whole set of prospective data in conjunction with genetic markers can shed light on the functional role of the 8q24 region.

## MATERIALS AND METHODS

### Subjects

Healthy subjects were recruited from the Italian branch of the European Prospective Investigation Into Cancer and Nutrition, a multicenter European study in which investigators recruited more than 520,000 healthy volunteers from 10 European countries (34). The Turin cohort comprises 10,604 healthy volunteers enrolled in the Turin area (Northern Italy) between 1993 and 1998 (35).

The cohort includes participants of both sexes (6,046 men and 4,558 women), most of whom were 35–74 years of age at recruitment. Detailed dietary and lifestyle histories were collected for each volunteer through the use of a self-administered questionnaire; 24-hour dietary recalls collected through person-to-person interviews and anthropologic measurements were also available. At enrollment, all participants signed informed consent forms and agreed to provide detailed information on their dietary and lifestyle habits, as well as to have their health status followed-up throughout life. Follow-up information on vital status and cancer incidence was obtained annually from the municipality and the local cancer registry. After vital status updates, death certificates were obtained from the mortality registries. In addition, we performed a follow up to identify cases of diabetes mellitus and cardiovascular disease. We identified incident cases of diabetes, defined as new diagnoses made after the cohort inception. We excluded prevalent cases of diabetes (those present at recruitment). The diagnoses of diabetes were confirmed by different sources of information, including self-reported diagnosis in the baseline questionnaire, use of diabetes-specific medication, and linkage to the regional diabetes registry and to the regional hospital discharge record database. Only persons with confirmed case statuses (positive in 2 or more sources of information) were included in this analysis. Subjects affected by cardiovascular and cerebrovascular diseases were identified through record linkage to the regional hospital discharge record database using a standardized procedure. Clinical records were obtained for each identified case to confirm the diagnosis.

### DNA extraction and genotyping

A 30–40 mL peripheral blood sample was collected from all volunteers and fractionated, aliquoted, and properly stored the day of collection. Buffy coats for DNA extraction were available for 10,372 of the 10,604 participants. DNA was extracted from 400 μL of buffy coat using a QIAsymphony SP instrument (QIAGEN, Hilden, Germany) according to manufacturer protocols and stored in liquid nitrogen until use.

Polymorphisms were selected from those previously found to be significantly associated with at least 1 disease in prior studies (rs672888, rs1447295, rs9642880, rs16901979, and rs6983267), taking into account linkage disequilibrium (LD) (see Web Figure 1, available at http://aje.oxfordjournals.org/, for SNP location and a LD plot from HapMap data). A 5′ nuclease assay with MGB TaqMan Probes on a 7900HT Fast Real-Time PCR System (AppliedBiosystems, Foster City, California) with premade assays (TaqMan SNP Genotyping Assays, AppliedBiosystems) was used to genotype the selected markers. All genotypes automatically assigned by the software were inspected by an operator to check cluster quality and were manually edited or removed when appropriate.

### Haplotype reconstruction

We performed a LD analysis of rs672888, rs1447295, rs9642880, rs16901979, and rs6983267 in our samples (36) (Web Figure 2). Because no pairs of SNPs reached the threshold of $R^2 > 0.80$, we used all 5 SNPs in the haplotype reconstruction process.

We imputed the phase of the haplotype using a Bayesian method in which the Bayesian a priori was based on an approximation to the coalescent and the inference based on the Markov chain Monte Carlo approach. The analysis of LD was performed with Haploview, version 4.1 (Broad Institute, Cambridge, Massachusetts) (36); the phases of haplotypes were inferred using PHASE, version 2.1 (University of Washington, Seattle, Washington) (37, 38).

### Statistical analyses

For each SNP and pathology (incident cancer, incident cardiovascular disease, incident diabetes, and death), we calculated hazard ratios and the corresponding 95% confidence

**Table 1.** Per-Allele Hazard Ratios for the Association Between Single Nucleotide Polymorphism rs6983267 in Chromosome Region 8q24 and Cancer Incidence and Total Mortality, by Cancer Type and Sex, Italy, 1993–1998

| Type of Cancer by Sex | No. of Incidences | Hazard Ratio | 95% Confidence Interval |
|---|---|---|---|
| Men and women | | | |
| All cancer types | 859 | 1.19 | 1.08, 1.31 |
| All cancer types except skin | 763 | 1.21 | 1.09, 1.34 |
| Head and neck | 10 | 0.18 | 0.05, 0.64 |
| Stomach | 21 | 1.57 | 0.84, 2.91 |
| Colorectal | 119 | 1.23 | 0.96, 1.60 |
| Pancreatic | 23 | 1.23 | 0.68, 2.22 |
| Liver and cholecystic, primary | 16 | 0.91 | 0.45, 1.83 |
| Lung, primary | 60 | 1.72 | 1.19, 2.48 |
| Mesothelioma | 8 | 0.63 | 0.22, 1.74 |
| Melanoma | 37 | 1.38 | 0.87, 2.20 |
| Breast | 144 | 1.09 | 0.86, 1.39 |
| Kidney | 17 | 0.95 | 0.48, 1.87 |
| Bladder | 50 | 1.47 | 0.99, 2.19 |
| Brain | 26 | 1.81 | 1.02, 3.20 |
| Thyroid | 17 | 1.48 | 0.74, 2.95 |
| Lymphoma | 38 | 0.84 | 0.53, 1.33 |
| Leukemia | 24 | 1.35 | 0.76, 2.40 |
| Total mortality | 304 | 1.00 | 0.85, 1.18 |
| Men | | | |
| All cancer types | 487 | 1.34 | 1.19, 1.52 |
| All cancer types except skin | 434 | 1.37 | 1.20, 1.57 |
| Head and neck | 8 | 0.16 | 0.03, 1.80 |
| Stomach | 16 | 1.10 | 0.54, 2.18 |
| Colorectal | 75 | 1.50 | 1.09, 2.08 |
| Pancreatic | 11 | 1.59 | 0.68, 3.71 |
| Liver and cholecystic, primary | 8 | 0.64 | 0.23, 1.78 |
| Lung, primary | 49 | 2.13 | 1.40, 3.23 |
| Mesothelioma | 6 | 0.54 | 0.17, 1.80 |
| Melanoma | 23 | 1.30 | 0.73, 2.32 |
| Breast | 3 | 2.18 | 0.40, 11.84 |
| Prostate | 113 | 1.47 | 1.12, 1.91 |
| Kidney | 14 | 1.10 | 0.52, 2.30 |
| Bladder | 41 | 1.53 | 0.99, 2.38 |
| Brain | 13 | 2.92 | 1.23, 6.94 |
| Thyroid | 6 | 5.30 | 1.17, 24.10 |
| Lymphoma | 21 | 1.10 | 0.60, 2.01 |
| Leukemia | 16 | 0.83 | 0.41, 1.68 |
| Total mortality | 197 | 1.12 | 0.92, 1.37 |
| Women | | | |
| All cancer types | 372 | 1.01 | 0.88, 1.18 |
| All cancer types except skin | 329 | 1.03 | 0.89, 1.21 |
| Head and neck | 2 | | |
| Stomach | 5 | 9.48 | 1.18, 76.37 |
| Colorectal | 44 | 0.86 | 0.58, 1.37 |
| Pancreatic | 12 | 1.00 | 0.43, 2.28 |
| Liver and cholecystic, primary | 8 | 1.28 | 0.47, 3.50 |
| Lung, primary | 11 | 0.67 | 0.28, 1.59 |
| Mesothelioma | 2 | 0.92 | 0.12, 6.70 |
| Melanoma | 14 | 1.54 | 0.71, 3.36 |
| Breast | 141 | 1.08 | 0.84, 1.38 |
| Endometrium | 17 | 0.84 | 0.42, 1.70 |
| Ovarian | 17 | 0.71 | 0.31, 1.62 |
| Kidney | 3 | 0.46 | 0.79, 2.61 |
| Bladder | 9 | 1.19 | 0.46, 3.08 |
| Brain | 13 | 1.17 | 0.53, 2.58 |
| Thyroid | 11 | 0.81 | 0.34, 1.92 |
| Lymphoma | 17 | 0.60 | 0.30, 1.20 |
| Leukemia | 8 | 4.51 | 1.27, 16.10 |
| Total mortality | 107 | 0.81 | 0.62, 1.07 |

intervals using the Cox proportional hazards model. We performed calculations for 4 different genetic models: dominant (homozygous wild-type genotype (AA) vs. heterozygous genotype (AB) plus homozygous variant genotype (BB)); codominant (AA vs. AB and AA vs. BB); recessive (AA plus AB vs. BB); and per-allele (dose-effect: 0, 1, and 2 variant alleles). We also computed prevalence odds ratios for prevalent diabetes and the corresponding 95% confidence intervals for the 4 models. Moreover, we performed per-haplotype analyses to test whether the results would be more informative than those of the genotype analyses.

We performed crude analyses and calculated adjusted hazard ratios and odds ratios first by sex and age (using quartiles among controls) and then by smoking status (current smokers, never smokers, and former smokers who had stopped smoking at least 1 year prior). Subjects with missing values for the variables for which we adjusted the models were excluded. We also conducted analyses stratified by sex.

To account for multiple comparisons, we estimated the false-positive report probability (FPRP) based on the Wacholder method (39). FPRP values for the significant SNP and haplotype associations, which are reported in Tables 1, 2, and 3, were calculated together with the statistical power to detect odds ratios of 1.2, 1.5, and 2.0 (Web Table 1) and odds ratios of 5.0 and 10.0 (Web Table 2). We allowed for different levels of prior probability ranging from 0.5 to 0.0001 and, given the many comparisons, we considered a cut point of 0.2

**Table 2.** Diabetes Incidence by Selected Single Nucleotide Polymorphisms, Italy, 1993–1998

| Single Nucleotide Polymorphism | Odds Ratio[a] | 95% Confidence Interval | *P* Value |
|---|---|---|---|
| rs672888 ($n = 449$) | | | |
| Per allele | 1.20 | 1.05, 1.38 | 0.007 |
| Recessive | 1.34 | 1.07, 1.68 | 0.05 |
| Dominant | 1.23 | 1.00, 1.52 | 0.01 |
| rs1447295 ($n = 451$) | | | |
| Per allele | 0.70 | 0.51, 0.96 | 0.008 |
| Recessive | NA | | NA |
| Dominant | 0.68 | 0.49, 0.93 | 0.01 |
| rs9642880 ($n = 449$) | | | |
| Per allele | 1.00 | 0.87,1.14 | 0.95 |
| Recessive | 0.97 | 0.76, 1.23 | 0.78 |
| Dominant | 1.01 | 0.83, 1.25 | 0.89 |
| rs16901979 ($n = 446$) | | | |
| Per allele | 0.73 | 0.48, 1.10 | 0.11 |
| Recessive | 1.19 | 0.15, 9.10 | 0.87 |
| Dominant | 0.70 | 0.46, 1.08 | 0.09 |
| rs6983267 ($n = 451$) | | | |
| Per allele | 1.01 | 0.88, 1.15 | 0.91 |
| Recessive | 1.01 | 0.81, 1.26 | 0.92 |
| Dominant | 1.01 | 0.81, 1.26 | 0.94 |

[a] Adjusted for age and sex.

for FPRP, although we could not exclude higher thresholds. All analyses were performed using SAS, version 9.1 (SAS Institute, Inc., Cary, North Carolina) and Stata, version 10.1 (StataCorp LP, College Station, Texas).

## RESULTS

In a median follow-up of 11.24 years, we identified more than 1,500 health-related outcomes, including 773 newly diagnosed cases of cancer (all cancer sites except skin), 451 incident cases of diabetes, and 342 deaths, all of which were diagnosed before December 2006. Table 1 shows results of tests of the association between rs6983267 and any disease. Considering both sexes together, a statistically significant increased risk was observed for all cancer sites, including (hazard ratio (HR) = 1.19, 95% confidence interval (CI): 1.08, 1.31) or excluding (HR = 1.21, 95% CI: 1.09, 1.34) skin cancer. An increased risk was also found for lung (HR = 1.72, 95% CI: 1.19, 2.48) and brain (HR = 1.81, 95% CI: 1.02, 3.20) cancers, whereas a protective effect was obtained for head and neck cancer (HR = 0.18, 95% CI: 0.05, 0.64). Among men, statistically significant associations were seen with all cancers (HR = 1.34, 95% CI: 1.19, 1.52), colorectal cancer (HR = 1.50, 95% CI: 1.09, 2.08), prostate cancer (HR = 1.47, 95% CI: 1.12, 1.91), lung cancer (HR = 2.13, 95% CI: 1.40, 3.23), brain cancer (HR = 2.92, 95% CI: 1.23, 6.94), and thyroid cancer (HR = 5.30, 95% CI: 1.17, 24.10). However, none of these diseases (including total cancers) showed evidence of an association in women, although the smaller number of cancer outcomes in women limited statistical power. On the other hand, we observed a statistically significant increased risk in women for stomach cancer (HR = 2.92, 95% CI: 1.23, 6.94) and leukemia (HR = 2.92, 95% CI: 1.23, 6.94), although the numbers of cases were quite small (5 and 8 cases, respectively).

Despite a borderline association between rs6983267 and total mortality in men, no specific cause of death was associated with this or any other SNP. Other SNPs in the 8q24 region were associated with a different pattern of diseases. In particular, rs672888 and rs1447295 showed statistically significant associations with incident diabetes (Table 2). In addition, rs672888 was associated with thyroid cancer in both men ($n = 17$; HR = 2.34, 95% CI: 1.17, 4.72) and women ($n = 11$; HR = 2.78, 95% CI: 1.13, 6.79).

Table 3 shows the haplotype reconstruction obtained using PHASE, version 2.1, and Table 4 shows the statistically significant associations found with haplotypes when at least 15 cases were observed (full results are shown in Web Tables 3 and 4). Although associations with individual SNPs had smaller effects in general, we observed higher increases in risk for certain haplotypes and specific cancer types, in particular when subjects were homozygous for a haplotype. For breast cancer, diplotype 5/5 (CAGCT) yielded an estimated hazard ratio of 3.40 (95% CI: 1.24, 9.21); for prostate cancer, grouped rare haplotypes yielded an estimated hazard ratio of 7.43 (95% CI: 3.00, 18.37); and for brain cancer, diplotype 7/7 (CGGCT) yielded an estimated hazard ratio of 28.61 (95% CI: 5.9, 138).

Significant associations were also observed between haplotypes and all deaths from cardiovascular diseases (for diplotype 4/4 (CGTCG), HR = 7.88, 95% CI: 2.81, 22.02) and cerebrovascular diseases (for diplotype 8/8 (CGTCT), HR = 10.26, 95% CI: 1.34, 78.40). The most stable associations were between diplotypes 4/4 and 5/5 and the total number of deaths in men (based on 217 observations; HR = 3.5, 95% CI: 1.8, 6.9, and HR = 2.8, 95% CI: 1.3, 6.4, respectively).

We also considered the distribution of SNPs and haplotypes (Web Tables 5 and 6) by anthropometric measurements, smoking status, and dietary habits. We did not find any evidence of associations between SNPs in the 8q24 region and personal characteristics of the subjects. The exceptions were weak associations between rs672888 and glycemic load ($P = 0.02$) and between rs1447295 and intake of carbohydrates ($P = 0.008$) (Web Table 5), both of which were in line with the association between these SNPs and diabetes. Another significant association was found between rs16901979 and consumption of animal protein ($P = 0.007$).

### Power and false positives

The FPRP results (Web Table 1) showed an acceptable level of false reports considering the association between rs6983267 and all cancer sites in men, with very low prior probability (<0.0001) and an expected odds ratio of 1.5 (power = 96%). For the other significant SNP associations, the power of this study was high enough (>80%) to explain only the association of rs6983267 with colorectal and prostate cancers in men (prior probabilities of 0.25 and 0.1, respectively), with

**Table 3.** Haplotype Reconstruction Using PHASE ,Version 2.1[a], Italy, 1993–1998

| Haplotype | rs16901979 | rs672888 | rs6983267 | rs1447295 | rs9642880 | Frequency (SE) |
|---|---|---|---|---|---|---|
| | | | *Common Haplotypes* | | | |
| 1 | C | A | T | C | G | 0.150 (0.002) |
| 2 | C | A | G | C | G | 0.141 (0.001) |
| 3 | C | A | T | C | T | 0.114 (0.002) |
| 4 | C | G | T | C | G | 0.110 (0.002) |
| 5 | C | A | G | C | T | 0.107 (0.002) |
| 6 | C | G | G | C | G | 0.107 (0.002) |
| 7 | C | G | G | C | T | 0.086 (0.002) |
| 8 | C | G | T | C | T | 0.086 (0.002) |
| 9 | C | A | G | A | G | 0.011 (0.001) |
| | | | *Rare Haplotypes* | | | |
| 10 | C | A | T | A | G | 0.009 (0.0009) |
| 11 | C | A | G | A | T | 0.009 (0.0007) |
| 12 | C | G | T | A | G | 0.009 (0.0008) |
| 13 | C | A | T | A | T | 0.007 (0.0006) |
| 14 | C | G | G | A | G | 0.007 (0.0007) |
| 15 | C | G | T | A | T | 0.006 (0.0005) |
| 16 | C | G | G | A | T | 0.006 (0.0006) |
| 17 | A | G | T | C | G | 0.005 (0.0004) |
| 18 | A | A | G | C | G | 0.005 (0.0004) |
| 19 | A | A | T | C | G | 0.004 (0.0004) |
| 20 | A | A | G | C | T | 0.004 (0.0004) |
| 21 | A | G | T | C | T | 0.004 (0.0004) |
| 22 | A | G | G | C | G | 0.004 (0.0004) |
| 23 | A | G | G | C | T | 0.003 (0.0004) |
| 24 | A | A | T | C | T | 0.003 (0.0004) |
| 25 | A | G | T | A | G | 0.0008 (0.0003) |
| 26 | A | G | T | A | T | 0.0007 (0.0003) |
| 27 | A | A | G | A | T | 0.0004 (0.0002) |
| 28 | A | A | G | A | G | 0.0004 (0.0002) |
| 29 | A | G | G | A | G | 0.0004 (0.0001) |
| 30 | A | G | G | A | T | 0.0003 (0.0001) |
| 31 | A | A | T | A | T | 0.0003 (0.0001) |
| 32 | A | A | T | A | G | 0.0003 (0.0001) |

Abbreviation: SE, standard error.
[a] The sum of the frequencies is not exactly 1 because of approximations.

an expected odds ratio of 2. As far as haplotype associations were concerned, FPRP results showed no association that was adequately powered to detect an odds ratio <2.0 (Web Table 1); however, the observed odds ratios were generally much higher than 2 (most were in the range of 4–15 or higher), so we further explored FPRP values for odds ratios of 5.0, 10.0, and 15.0 (Web Table 2). The associations with sufficient power were between haplotype 4 (CGTCG) and haplotype 5 (CAGCT) (2 vs. 0 copies for both) and all causes of death, for an expected odds ratio of 5, whereas the associations of haplotype 2 (CAGCG) with head and neck cancer, haplotype 5 (CAGCT) with breast cancer, and haplotype 4 (CGTCG) with cardiovascular disease mortality had a very low prior probability, yielding a statistical power in the range of 0.67–0.74, for an expected odds ratio of 10.

## DISCUSSION

The purpose of the present analysis was to investigate the whole range of health effects of inherited variants in the 8q24 region, which was found to be significantly associated with different diseases in previous genome-wide association studies. To the best of our knowledge, no prospective study has been published thus far. In 10,372 subjects followed for

**Table 4.** Analysis by Haplotype, Italy, 1993–1998

| | Sex | No. of Cases | 1 Copy vs. 0 Copies | | 2 Copies vs. 0 Copies | |
|---|---|---|---|---|---|---|
| | | | HR[a] | 95% CI | HR[a] | 95% CI |
| Cancer type and haplotype | | | | | | |
| Brain, haplotype 2 | Men and women | 26 | 2.07 | 0.92, 3.61 | 13.38 | 3.00, 59.53 |
| Prostate, rare haplotypes | Men | 114 | 1.70 | 1.11, 2.59 | 7.43 | 3.00, 18.37 |
| Breast, haplotype 5 | Women | 142 | 1.38 | 0.91, 2.08 | 3.40 | 1.24, 9.21 |
| Cause of death and haplotype | | | | | | |
| Cardiovascular disease, haplotype 4 | Men and women | 47 | 0.53 | 017, 1.71 | 7.88 | 2.81, 22.02 |
| Cerebrovascular disease, haplotype 8 | Men and women | 16 | 1.24 | 0.29, 5.51 | 10.26 | 1.34, 78.40 |
| All causes, haplotype 4 | Men | 217 | 1.01 | 0.69, 1.56 | 3.50 | 1.80, 6.86 |
| All causes, haplotype 5 | Men | 217 | 1.10 | 0.78, 1.57 | 2.84 | 1.27, 6.43 |
| Cardiovascular disease, haplotype 4 | Men | 26 | 0.34 | 0.47, 2.56 | 13.60 | 4.67, 39.67 |

Abbreviations: CI, confidence interval; HR, hazard ratio.
[a] Only statistically significant hazard ratios in at least 1 sex with at least 15 cases of disease are shown.

11 years, we found a number of associations between variants in the 8q24 region and various cancer types, diabetes, and death. Although associations with individual SNPs generally had smaller effects, we observed very strong associations between haplotypes and cancers of head and neck, breast, prostate, and brain and between haplotypes and deaths from cardiovascular and neurologic diseases and diabetes, with relative risks usually greater than 10. Therefore, we observed a very strong pleiotropism in the region.

The most stable and striking association was between individuals carrying 2 copies of the 4 and 5 haplotypes (e.g., diplotypes 4/4 and 5/5) and total deaths in men (based on 217 observations; HR = 3.5, 95% CI: 1.8, 6.9, and HR = 2.8, 95% CI: 1.3, 6.4, respectively). If these associations are confirmed, a large proportion of deaths in men could be explained by these genetic variants, although the reasons for the sex specificity are far from clear.

The 8q24 region is a gene desert. However, multiple enhancer elements are present within this region, and they can regulate transcription of *MYC* (40). In particular, it has been reported that the region harboring the rs6983267 variant (prostate/colorectal risk region 3) is a transcriptional enhancer. Alleles at rs6983267 differentially bind transcription factor 7-like 2 (TCF7L2) (12, 20), as well as the beta-catenin/transcription factor 4 complex as measured in colon cancer cells (41). Moreover, the region physically interacts with the *MYC* protooncogene through a chromatin loop (29, 40). It has also been demonstrated that 1 enhancer element physically interacts with the *MYC* promoter via transcription factor 4 binding and acts in an allele-specific manner to regulate *MYC* expression (42). The allelic status seems not to interfere with the creation of the chromatin loop leading to the interaction of the enhancer region containing rs6983267 with the promoter region of *c-MYC*. However, an allele-specific regulation of *c-MYC* gene expression has been shown, including an enhanced expression of the *c-MYC* allele in *cis* with the rs6983267-G allele (41). Tuupanen et al. (43) provided evidence that the common predisposition to colorectal cancer associated with the 8q24 region arises from enhanced responsiveness to Wnt signaling.

Although these data provide support for a biologic mechanism underlying this non-protein-coding risk variant, Prokunina-Olsson et al. (30) recently reported a strong correlation between the expression of *MYC* and a unique splicing form of TCF7L2 in noncancer colon samples but did not report any correlation with genotypes of rs6983267 or interaction of rs6983267 with TCF7L2 expression. These findings suggest that some splicing forms of TCF7L2 may be functionally important for regulation of *MYC* expression in colon tissue, but this regulation is not directly dependent on rs6983267. A negative association has been reported between several SNPs at region 8q24 and a role in *c-MYC* amplification or chromosomal fragility (44).

A role for microRNA in the 8q24 region has been reported in epigenetic mechanisms (31–33). Huppi et al. (31) investigated whether the translocated region in 8q24 may contain microRNAs. Seven microRNAs were identified through computational analysis and experimental verification. High-level expression of one of them (hsa-miR-1204 precursor) was also seen in several epithelial cancer cell lines with *MYC/PVT1* coamplification, suggesting a potentially broad role for these microRNAs in tumorigenesis. In another study on medulloblastoma, Lu et al. (32) identified a novel amplification at the 8q24.22–q24.23 region that was independent of *MYC* amplification at 8q24.21, suggesting involvement of 2 microRNAs: hsa-miR-30b and hsa-miR-30d. Finally, Pomerantz et al. (33) investigated whether the 8q24 region contained unannotated transcribed microRNA and *cis*-acting enhancers by using next-generation sequencing in histologically normal radical prostatectomy tissue. They did not find evidence for significant microRNA transcription within the 8q24 prostate cancer risk loci. Likewise, no convincing association between RNA expression and risk allele status was detected in either histologically normal or tumor tissue.

In the present study, we found that most associations were sex-specific, with more associations in men than in women. These findings are difficult to interpret, but it is unlikely they were due to distortions related to design because this was a prospective study in which we made internal comparisons. One possible explanation could be related to sex-specific epigenetic regulations of the 8q24 region, such as methylation, or the presence of hormonal responsive elements. In a recent study (45), Jia et al. characterized a 5-megabase chromatin segment encompassing all the risk regions for RNA expression, histone modifications, and locations occupied by RNA polymerase II and androgen receptors. This led to identification of several transcriptional enhancers that were verified using reporter assays. Two enhancers in one risk region were occupied by androgen receptors and responded to androgen treatment; one contained a SNP (rs11986220) within a FoxA1 binding site, with the prostate cancer risk allele facilitating both stronger FoxA1 binding and stronger androgen responsiveness. The presence of androgen-responsive elements in the region could explain sex differences in the risks of different types of cancer.

Few investigators have studied methylation differences in tumors with 8q24 alterations. Advanced-stage prostate cancers have been reported to frequently harbor chromosome 8 alterations and hypomethylation of LINE-1 retrotransposons (46), and a specific hypomethylation of a CCGG site in exon 3 of *c-MYC* has been reported in human myeloma cell lines, leading to microRNA levels of *c-MYC* that are 30-50-fold higher than those in normal peripheral blood lymphocytes (47).

Further important insights into functional mechanisms involved in the 8q24 region can be found by better exploring the haplotype structure of the region. We observed striking associations between haplotypes and different cancers and other diseases, yielding strong evidence of pleiotropism in the region. In the present study, rare haplotypes were grouped to have enough power to detect signals. Further studies using bigger samples are needed to more precisely identify the specific haplotypes involved.

Recent studies have shown that different haplotypes were associated with different cancer sites. For instance, a protective effect has been found for an 8q24 haplotype (5 SNPs) and prostate cancer in a white population (48), and an increased risk has been observed between different haplotypes in the 8q24 region and breast, prostate, and colorectal cancers (49), as well as papillary thyroid carcinoma (17). Unfortunately, not all of the studies analyzed the same SNPs, so it is not possible to make direct comparisons of haplotypes because of the different degrees of LD across populations (e.g., at least 3 subregions of 8q24 have been independently associated with prostate cancer risk, the most centromeric of which appears to be population-specific) (50). Moreover, published data have suggested that multiple interacting SNPs within 8q24, as well as different regions on chromosome 8 far beyond this 8q24 candidate region, may confer an increased risk of cancer. An example for prostate cancer was reported by Beuten et al. (48), who showed that SNPs within the 8q24 region were in high LD (log of the odds $>3$) with SNPs located within the C1r/C1s, uEGF, and BMP1 gene (*CUB*) and the Sushi multiple domains 1 gene (*CSMD1*) in both whites and Hispanics. This gene is located on the short arm of chromosome 8 within the 8p23 region. They also found high LD (log of the odds $>3$) between SNPs in region 8q24 and SNPs within the b-defensin-1 gene (*DEFB1*) located at 8p23 in whites, within the CUB and Sushi multiple domains 3 gene (*CSMD3*) at 8p23 in Hispanics, and within the pleckstrin and Sec7 domain-containing 3 gene (*PSD3*) located at 8p22 in both whites and Hispanics. High LD (log of the odds $>5$) was also noticed between SNPs in region 8q24 and SNPs within the *TMEM75* gene downstream of the region investigated.

Finally, we did not find any obvious association between SNPs or haplotypes in the 8q24 region and personal characteristics of the subjects, including anthropometric measures, smoking status, and diet. The exceptions were weak associations between rs672888 and the glycemic load ($P = 0.02$) and between rs1447295 and the intake of carbohydrates ($P = 0.008$). The latter observations, though weak, were in line with the association of both SNPs with diabetes. The role of 8q24 in diabetes onset has been observed repeatedly before, and the correspondence between the same SNPs related to diabetes and some dietary behaviors is worth of further investigation.

Our findings suggest the importance of better investigation of the pleiotropic effect of 8q24 that also considers its possible interaction with regions far beyond on the same chromosome and genetic regions on other chromosomes involved in different diseases. In particular, the discovery of links between SNPs and cancer, cardiovascular disease, and diabetes could be extremely useful in the attempt to disentangle mechanisms common to different chronic diseases and the metabolic syndrome, potentially leading to common preventive measures and/or therapies.

## ACKNOWLEDGMENTS

Author affiliations: Human Genetics Foundation, Torino, Italy (Simonetta Guarrera, Fulvio Ricceri, Silvia Polidoro, Alessandra Allione, Fabio Rosa, Floriana Voglino, Rossana Critelli, Alessia Russo, Paolo Vineis, Giuseppe Matullo); Department of Genetics, Biology, and Biochemistry, University of Torino, Torino, Italy (Fulvio Ricceri, Carlotta Sacerdote, Giuseppe Matullo); Centro di Prevenzione-Piemonte, Torino, Italy (Carlotta Sacerdote); and Medical Research Council-Health Protection Agency Centre for Environment and Health, St. Mary's Campus, Imperial College London, London, United Kingdom (Paolo Vineis).

This work was supported by grants from the Associazione Italiana per le Ricerche sul Cancro, Italy (to Giuseppe Matullo) and the Environmental Cancer Risk Nutrition and Individual Susceptibility project (to Giuseppe Matullo and Paolo Vineis), a network of excellence operating within the European Union Sixth Framework Program, Priority 5: "Food Quality and Safety" (contract No.513943). This work was also supported by grants from the Compagnia di San Paolo, Turin, Italy (to Paolo Vineis and Giuseppe Matullo) and the Progetto Integrato Oncologia of the Italian government (to Paolo Vineis and Giuseppe Matullo).

Simonetta Guarrera and Fulvio Ricceri contributed equally to this work. Paolo Vineis designed and coordinated the prospective study and wrote the manuscript; Simonetta Guarrera, Silvia Polidoro, Alessandra Allione, Rossana Critelli, and Alessia Russo performed laboratory analyses; Giuseppe Matullo coordinated the laboratory analyses; and Fulvio Ricceri, Floriana Voglino, Fabio Rosa, and Carlotta Sacerdote performed the statistical analyses. All contributed to the writing of the manuscript.

Conflict of interest: none declared.

## REFERENCES

1. Gudmundsson J, Sulem P, Gudbjartsson DF, et al. Genome-wide association and replication studies identify four variants associated with prostate cancer susceptibility. *Nat Genet*. 2009; 41(10):1122–1126.
2. Gudmundsson J, Sulem P, Manolescu A, et al. Genome-wide association study identifies a second prostate cancer susceptibility variant at 8q24. *Nat Genet*. 2007;39(5): 631–637.
3. Kiemeney LA. Words of wisdom. Re: Genome-wide association study of prostate cancer identifies a second risk locus at 8q24. *Eur Urol*. 2007;52(3):920–921.
4. Thomas G, Jacobs KB, Yeager M, et al. Multiple loci identified in a genome-wide association study of prostate cancer. *Nat Genet*. 2008;40(3):310–315.
5. Yeager M, Orr N, Hayes RB, et al. Genome-wide association study of prostate cancer identifies a second risk locus at 8q24. *Nat Genet*. 2007;39(5):645–649.
6. Neklason DW, Kerber RA, Nilson DB, et al. Common familial colorectal cancer linked to chromosome 7q31: a genome-wide analysis. *Cancer Res*. 2008;68(21):8993–8997.
7. Tenesa A, Farrington SM, Prendergast JG, et al. Genome-wide association scan identifies a colorectal cancer susceptibility locus on 11q23 and replicates risk loci at 8q24 and 18q21. *Nat Genet*. 2008;40(5):631–637.
8. Tomlinson I, Webb E, Carvajal-Carmona L, et al. A genome-wide association scan of tag SNPs identifies a susceptibility variant for colorectal cancer at 8q24.21. CORGI Consortium. *Nat Genet*. 2007;39(8):984–988.
9. Tomlinson IP, Webb E, Carvajal-Carmona L, et al. A genome-wide association study identifies colorectal cancer susceptibility loci on chromosomes 10p14 and 8q23.3. CORGI Consortium. *Nat Genet*. 2008;40(5):623–630.
10. Zanke BW, Greenwood CM, Rangrej J, et al. Genome-wide association scan identifies a colorectal cancer susceptibility locus on chromosome 8q24. *Nat Genet*. 2007;39(8):989–994.
11. Antoniou AC, Sinilnikova OM, McGuffog L, et al. Common variants in LSP1, 2q35 and 8q24 and breast cancer risk for BRCA1 and BRCA2 mutation carriers. Kathleen Cuningham Foundation Consortium for Research into Familial Breast Cancer. *Hum Mol Genet*. 2009;18(22):4442–4456.
12. Wokolorczyk D, Gliniewicz B, Sikorski A, et al. A range of cancers is associated with the rs6983267 marker on chromosome 8. *Cancer Res*. 2008;68(23):9982–9986.
13. Couch FJ, Wang X, McWilliams RR, et al. Association of breast cancer susceptibility variants with risk of pancreatic cancer. *Cancer Epidemiol Biomarkers Prev.* 2009;18(11): 3044–3048.
14. Wang M, Wang M, Zhang W, et al. Common genetic variants on 8q24 contribute to susceptibility to bladder cancer in a Chinese population. *Carcinogenesis*. 2009;30(6):991–996.
15. Park SL, Chang SC, Cai L, et al. Associations between variants of the 8q24 chromosome and nine smoking-related cancer sites. *Cancer Epidemiol Biomarkers Prev.* 2008;17(11):3193–3202.
16. Kiemeney LA, Thorlacius S, Sulem P, et al. Sequence variant on 8q24 confers susceptibility to urinary bladder cancer. *Nat Genet*. 2008;40(11):1307–1312.
17. He H, Nagy R, Liyanarachchi S, et al. A susceptibility locus for papillary thyroid carcinoma on chromosome 8q24. *Cancer Res*. 2009;69(2):625–631.
18. Crowther-Swanepoel D, Broderick P, Di Bernardo MC, et al. Common variants at 2q37.3, 8q24.21, 15q21.3 and 16q24.1 influence chronic lymphocytic leukemia risk. *Nat Genet*. 2010; 42(2):132–136.
19. Cook MB, Graubard BI, Quraishi SM, et al. Genetic variants in the 8q24 locus and risk of testicular germ cell tumors. *Hum Genet*. 2008;123(4):409–418.
20. Shete S, Hosking FJ, Robertson LB, et al. Genome-wide association study identifies five susceptibility loci for glioma. *Nat Genet*. 2009;41(8):899–904.
21. Sladek R, Rocheleau G, Rung J, et al. A genome-wide association study identifies novel risk loci for type 2 diabetes. *Nature*. 2007;445(7130):881–885.
22. Huyghe JR, Van Laer L, Hendrickx JJ, et al. Genome-wide SNP-based linkage scan identifies a locus on 8q24 for an age-related hearing impairment trait. *Am J Hum Genet*. 2008; 83(3):401–407.
23. Holmans PA, Riley B, Pulver AE, et al. Genomewide linkage scan of schizophrenia in a large multicenter pedigree sample using single nucleotide polymorphisms. *Mol Psychiatry.* 2009; 14(8):786–795.
24. Zöllner S, Su G, Stewart WC, et al. Bayesian EM algorithm for scoring polymorphic deletions from SNP data and application to a common CNV on 8q24. *Genet Epidemiol*. 2009;33(4): 357–368.
25. Zandi PP, Zöllner S, Avramopoulos D, et al. Family-based SNP association study on 8q24 in bipolar disorder. *Am J Med Genet B Neuropsychiatr Genet*. 2008;147B(5):612–618.
26. Birnbaum S, Ludwig KU, Reutter H, et al. Key susceptibility locus for nonsyndromic cleft lip with or without cleft palate on chromosome 8q24. *Nat Genet*. 2009;41(4):473–477.
27. Grant SF, Wang K, Zhang H, et al. A genome-wide association study identifies a locus for nonsyndromic cleft lip with or without cleft palate on 8q24. *J Pediatr.* 2009;155(6): 909–913.
28. Bertrand P, Bastard C, Maingonnat C, et al. Mapping of MYC breakpoints in 8q24 rearrangements involving non-immunoglobulin partners in B-cell lymphomas. *Leukemia*. 2007;21(3):515–523.
29. Pomerantz MM, Ahmadiyeh N, Jia L, et al. The 8q24 cancer risk variant rs6983267 shows long-range interaction with MYC in colorectal cancer. *Nat Genet*. 2009;41(8):882–884.
30. Prokunina-Olsson L, Hall JL. No effect of cancer-associated SNP rs6983267 in the 8q24 region on co-expression of MYC and TCF7L2 in normal colon tissue. *Mol Cancer.* 2009;8(1): 96. (doi:10.1186/1476-4598-8-96).
31. Huppi K, Volfovsky N, Runfola T, et al. The identification of microRNAs in a genomically unstable region of human chromosome 8q24. *Mol Cancer Res*. 2008;6(2):212–221.
32. Lu Y, Ryan SL, Elliott DJ, et al. Amplification and over-expression of Hsa-miR-30b, Hsa-miR-30d and KHDRBS3 at 8q24.22-q24.23 in medulloblastoma. *PLoS One*. 2009; 4(7):e6159.
33. Pomerantz MM, Beckwith CA, Regan MM, et al. Evaluation of the 8q24 prostate cancer risk locus and MYC expression. *Cancer Res*. 2009;69(13):5568–5574.

34. Gonzalez CA. The European Prospective Investigation Into Cancer and Nutrition (EPIC). *Public Health Nutr.* 2006; 9(1A):124–126.
35. Palli D, Berrino F, Vineis P, et al. A molecular epidemiology project on diet and cancer: the EPIC-Italy Prospective Study. Design and baseline characteristics of participants. *Tumori*. 2003;89(6):586–593.
36. Barrett JC, Fry B, Maller J, et al. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics*. 2005;21(2):263–265.
37. Stephens M, Smith NJ, Donnelly P. A new statistical method for haplotype reconstruction from population data. *Am J Hum Genet*. 2001;68(4):978–989.
38. Stephens M, Donnelly P. A comparison of Bayesian methods for haplotype reconstruction from population genotype data. *Am J Hum Genet*. 2003;73(5):1162–1169.
39. Wacholder S, Chanock S, Garcia-Closas M, et al. Assessing the probability that a positive report is false: an approach for molecular epidemiology studies. *J Natl Cancer Inst*. 2004; 96(6):434–442.
40. Sotelo J, Esposito D, Duhagon MA, et al. Long-range enhancers on 8q24 regulate c-Myc. *Proc Natl Acad Sci U S A*. 2010;107(7):3001–3005.
41. Wright JB, Brown SJ, Cole MD. Upregulation of c-MYC in *cis* through a large chromatin loop linked to a cancer risk-associated single-nucleotide polymorphism in colorectal cancer cells. *Mol Cell Biol*. 2010;30(6):1411–1420.
42. Harismendy O, Frazer KA. Elucidating the role of 8q24 in colorectal cancer. *Nat Genet*. 2009;41(8):868–869.
43. Tuupanen S, Turunen M, Lehtonen R, et al. The common colorectal cancer predisposition SNP rs6983267 at chromosome 8q24 confers potential to enhanced Wnt signaling. *Nat Genet*. 2009;41(8):885–890.
44. Cicek MS, Slager SL, Achenbach SJ, et al. Functional and clinical significance of variants localized to 8q24 in colon cancer. *Cancer Epidemiol Biomarkers Prev*. 2009;18(9):2492–2500.
45. Jia L, Landan G, Pomerantz M, et al. Functional enhancers at the gene-poor 8q24 cancer-linked locus. *PLoS Genet*. 2009; 5(8):e1000597. (doi:10.1371/journal.pgen.1000597).
46. Kindich R, Florl AR, Kamradt J, et al. Relationship of NKX3.1 and MYC gene copy number ratio and DNA hypomethylation to prostate carcinoma stage. *Eur Urol*. 2006;49(1):169–175.
47. Ohtsuki T, Nishitani K, Hatamochi A, et al. Analysis of methylation in the c-MYC gene in five human myeloma cell lines. *Br J Haematol*. 1991;77(2):172–179.
48. Beuten J, Gelfond JA, Martinez-Fierro ML, et al. Association of chromosome 8q variants with prostate cancer risk in Caucasian and Hispanic men. *Carcinogenesis*. 2009;30(8):1372–1379.
49. Ghoussaini M, Song H, Koessler T, et al. Multiple loci with different cancer specificities within the 8q24 gene desert. UK Genetic Prostate Cancer Study Collaborators/British Association of Urological Surgeons' Section of Oncology; UK Protect Study Collaborators. *J Natl Cancer Inst*. 2008;100(13): 962–966.
50. Yeager M, Xiao N, Hayes RB, et al. Comprehensive resequence analysis of a 136 kb region of human chromosome 8q24 associated with prostate and colon cancers. *Hum Genet*. 2008;124(2):161–170.