



The local environment and germline genetic variation predict cancer risk in the UK Biobank prospective cohort[☆]

Alessio Felici ^a, Giulia Peduzzi ^a, Francesca Giorgolo ^b, Andrea Spinelli ^b, Marco Calderisi ^b, Anna Monreale ^c, Riccardo Farinella ^a, Roberto Pellungrini ^d, Federico Canzian ^e, Daniele Campa ^{a,*}

^a Department of Biology, University of Pisa, Pisa, Italy

^b Kode Solutions s.r.l., Pisa, Italy

^c Department of Computer Science, University of Pisa, Pisa, Italy

^d Classe di Scienze, Scuola Normale Superiore, Pisa, Italy

^e Genomic Epidemiology Group, German Cancer Research Center (DKFZ), Heidelberg, Germany

ARTICLE INFO

Keywords:

Local environment
Distance to coast
Cancer
UK Biobank
Gene-environment interactions
Mediation

ABSTRACT

Background: There is a growing body of evidence on the effect of the local environment exposure on cancer susceptibility. Nonetheless, several of the associations remain controversial. Moreover, our understanding of the possible interaction between the local environment and the genetic variability is still very limited.

Objective: The aim of this study was to clarify the role of the local environment and its possible interplay with genetics on common cancers development.

Methods: Using the UK Biobank (UKBB) prospective cohort, we selected 12 local environment exposures: nitrogen oxides, nitrogen dioxides, particulate matter (10 and 2.5 µm), noise pollution, urban traffic, living distance from the coast, percentage of greenspace, natural environment, water, and domestic garden within 1000 m from the residential coordinates of each participant. All these exposures were tested for association with 17 different types of cancer for a total of 53,270 cases and 302,645 controls. Additionally, a polygenic score (PGS) was computed for each cancer, to test possible gene-environment interactions. Finally, mediation analyses were carried out.

Results: Thirty-six statistically significant associations considering multiple testing ($p < 2.19 \times 10^{-4}$) were observed. Among the novel associations we observed that individuals living farther from the coast had a higher risk of developing prostate cancer ($OR = 1.13$, $CI95\% = 1.06-1.20$, $P = 1.98 \times 10^{-4}$). This association was partially mediated by physical activity (indirect effect (IE) = -8.48×10^{-7}) and the time spent outdoor (IE = 9.07×10^{-6}). All PGSs showed statistically significant associations. Finally, genome-environment interaction analysis showed that local environment and genetic variability affect cancer risk independently.

Discussion: Living close to the coast and air pollution were associated with a decreased risk of prostate cancer and skin melanoma, respectively. These findings from the UKBB support the role of the local environment on cancer development, which is independent from genetics and may be mediated by several lifestyle factors.

1. Introduction

Over the last years there has been a growing body of evidence linking the local environment (*i.e.*, all the exposures that characterize the built

environment in which an individual lives) and its associated pollution and lifestyle with cancer susceptibility (Fazeli Dehkordi et al., 2022). To date, numerous studies aimed at clarifying these associations, most of which are related to the damaging effects of air pollution on different types of cancer (Pourvakhshoori et al., 2020). Nevertheless, the findings

* The present research was funded with intramural funding of the University of Pisa and German Cancer Research Center (DKFZ). Data used for this research were gathered from the UK Biobank Resource under application No. 66591. Bona fide researchers can access the UK Biobank dataset by registering and applying at <http://ukbiobank.ac.uk/register-apply/>. All participants have provided their informed consent to participate. Studies based on UK Biobank were performed in accordance with the Declaration of Helsinki. Ethical approval and consent were granted by the North West-Haydock NRES multicentre ethics committee, REF: 16NW/ 0274.

* Corresponding author. Department of Biology, University of Pisa, Via Derna 1, 56126, Pisa, Italy.

E-mail address: daniele.campa@unipi.it (D. Campa).

Abbreviations

CI	Confidence Interval
CLL	Chronic Lymphocytic Leukemia
DE	Direct Effect
DLBCL	Diffuse Large B-Cell Lymphoma
GxE	Gene-Environment interactions
ICD-10	International Classification of Diseases, 10th revision
ICD-11	International Classification of Diseases, 11th revision
IE	Indirect Effect
LD	Linkage Disequilibrium
Lden	Day-evening-night noise Level
NOx	Nitrogen Oxides
NO2	Nitrogen Dioxides
OR	Odds Ratio
PDAC	Pancreatic Ductal Adenocarcinoma
PGS	Polygenic Score
PM2.5	Particulate matter with diameter <2.5 µm
PM10	Particulate matter with diameter <10 µm
SNP	Single Nucleotide Polymorphism
TE	Total Effect
UKBB	UK Biobank

reported for most of these environmental variables are controversial (Fazeli Dehkordi et al., 2022; Zare Sakhvati et al., 2022).

The effect of several environmental variables on cancer risk, such as noise pollution or the effect of living close to the coast, has been poorly investigated, with most of the studies focusing only on few cancer types (Claeson et al., 2012; Haraldsdottir et al., 2017; Korgavkar et al., 2014; Reynolds et al., 2004, 2005). Similarly, studies analyzing the effect of urban or natural visible water bodies are almost entirely focused on blood malignancies, especially in children (Sánchez et al., 2022).

Throughout the last years, the development of large-scale projects such as UK Biobank (UKBB) has played a key role in the understanding of the possible effects that the environment exerts on human health. Nevertheless, only a small number of studies are focused on cancer (Huang et al., 2021; Wang et al., 2022).

Another challenge in studying local environment exposures is represented by understanding the possible gene-environment interactions for which our knowledge is currently very limited except for the effect of air pollution on lung cancer (Huang et al., 2021; Wang et al., 2022; Yu et al., 2018).

Considering all these premises, we aimed at studying whether and to which extent the different exposures that characterize the local environment affect the development of seventeen of the most frequent types of cancer. Moreover, using polygenic scores, we also aimed at studying the presence of gene-environment interactions. All analyses were conducted in the UKBB participants, for a total of 355,915 individuals.

2. Methods

2.1. The UK Biobank cohort

This case-control study used data from the UKBB prospective cohort. A detailed description of UKBB has already been provided (Sudlow et al., 2015). In brief, more than 500,000 subjects, aged between 37 and 73 years, have been recruited between 2006 and 2010 in the United Kingdom. During the assessment procedures data from each participant were collected by touchscreen self-completed questionnaires, blood sampling, physical measurement, imaging, and genotyping. Before data collection, each participant provided informed written consent. The UKBB study has received the approval from the North-West Multi-centre Research Ethics Committee (MREC). Environmental and genetic data of

UKBB participants were obtained from UKBB (project ID 66591).

2.2. Outcome ascertainment

Starting from a population of 502,420 individuals, only participants with a European ethnicity were selected using UKBB field 21000 (*i.e.*, white (code 1), British (code 1001), Irish (code 1002), and with any other white background (code 1003)), resulting in 472,622 individuals.

Cancer cases were selected according to cancer registry data. Specifically, three different UKBB fields were used: type of cancer – ICD10 (40006), histology of cancer tumor (40011), and behavior of cancer tumor (40012). To obtain better anatomically differentiated outcomes, the ICD codes taken from UKBB category 40006 were converted from version 10 to version 11 using the ICD10/ICD11 mapping by WHO (<https://icd.who.int/browse11/l-m/en>).

The ICD11 codes were then combined with UKBB histology codes to obtain a unique code that determines the specific histological tumor subtype. Only primary malignant tumors and microinvasive tumors were selected. Seventeen cancers (bladder cancer, breast cancer, chronic lymphocytic leukemia (CLL), colorectal cancer, cancer of corpus uteri, diffuse large B-cell lymphoma (DLBCL), glioma, kidney cancer, lung cancer, melanoma, multiple myeloma, oesophageal cancer, ovarian cancer, pancreatic ductal adenocarcinoma (PDAC), prostate cancer, stomach cancer, and thyroid cancer) with more than 500 cases each were included in this study. Breast cancer was analyzed only in females because we considered that male breast cancer is a distinct disease (Gucalp et al., 2019). Controls were selected among all those subjects who did not report a code that identifies a diagnosis of cancer within 3 different categories: cancer code, self-reported (20001), type of cancer – ICD10 (40006), and diagnoses – ICD10 (41270). A total of 302,645 controls were included. The total number of cases and controls is reported in Table 1 alongside age and sex distribution.

2.3. Environmental exposure measurements and outcomes of interest

In this study, 31 exposures that characterize the local environment were selected to be tested for their association with the risk of developing the 17 cancers described above. A correlation matrix using the Pearson correlation with a threshold of $r \leq 0.8$ identified 12 independent variables that were analyzed. The list of all exposures and the correlation matrix is reported in Fig. 1. The variables considered were: concentration in the air of nitrogen oxides (NOx), nitrogen dioxides (NO₂), particulate matter with a diameter of 10 µm (PM₁₀) and 2.5 µm (PM_{2.5}), traffic intensity (the average total number of vehicles on the nearest major road per 24 h), noise pollution (measured as the day-evening-night noise level indicator, Lden), each participant's living distance to the nearest major road (expressed as the inverse of distance, 1/meters), living distance to the coast and percentage of greenspace, natural environment, water, and domestic garden in a 1000 m radius centered on the residential coordinates of each participant. Detailed information on the measurement of each variable is given as supplementary methods. NOx showed a strong correlation with PM_{2.5} ($r = 0.85$), while greenspace was correlated with natural environment ($r = 0.97$) but were all maintained in the analysis to compare our results with previous studies since in the literature these variables are usually used individually.

2.4. Polygenic score (PGS) selection

All genetic data were downloaded from UKBB, using PLINK v2.0 with the method "extract" (Chang et al., 2015) and processed with a self-developed pipeline in Python v3.8.10. The code is reported as supplementary methods. The UKBB participants were genotyped on two different arrays, the Applied Biosystems UK BiLEVE Axiom Array by Affymetrix (49,950 individuals), and the Applied Biosystems UK Biobank Axiom Array (438,427 individuals). The genotyping was followed

Table 1
Study subjects and number of SNPs used in each PGS.

Cancer	Cases			SNPs used in PGS ^a	
	Females	Males	Total		
Breast	16,199 (100%)	NA	16,199 (100%)	58.44 [9.3]	257
Prostate	NA	12,838 (100%)	12,838 (100%)	66.34 [6.3]	106
Colorectal	2156 (40.4%)	3177 (59.6%)	5333 (100%)	63.96 [8.2]	71
Melanoma	1675 (44.9%)	1379 (45.1%)	3054 (100%)	60.9 [10.4]	21
Lung	1303 (49.1%)	1352 (50.9%)	2655 (100%)	67.76 [6.8]	21
Corpus uteri	2068 (100%)	NA	2068 (100%)	62.03 [7.8]	17
Kidney	627 (36.9%)	1073 (63.1%)	1700 (100%)	64.06 [8.4]	17
Bladder	334 (22.1%)	1182 (77.9%)	1516 (100%)	64.79 [9.0]	13
Ovary	1396 (100%)	NA	1396 (100%)	59.96 [10.6]	27
DLBCL	449 (44.8%)	552 (55.2%)	1001 (100%)	64.63 [8.9]	8
Multiple myeloma	386 (42.1%)	531 (57.9%)	917 (100%)	65.90 [8.1]	21
CLL	318 (37.4%)	532 (62.6%)	850 (100%)	64.65 [7.8]	31
Pancreas	390 (47.8%)	426 (52.2%)	816 (100%)	67.54 [6.9]	34
Oesophagous	128 (15.8%)	681 (84.2%)	809 (100%)	66.95 [7.0]	14
Glioma	305 (41.4%)	431 (58.6%)	736 (100%)	63.01 [10.6]	28
Thyroid	532 (74.2%)	185 (25.8%)	717 (100%)	56.83 [11.5]	9
Stomach	199 (29.9%)	466 (70.1%)	665 (100%)	66.26 [8.1]	3
Total	28,465 (53.4%)	24,805 (46.6%)	53,270 (100%)	63.76	—

A total number of 302,645 controls (mean age = 55.63 years) was used. Among these subjects, a total of 161,049 females (53.2%, mean age = 55.8 years) and 141,596 males (46.8%, mean age = 55.39 years) were included. Only men were used as controls for prostate cancer analyses and only women for breast, corpus uteri and ovarian cancers.

^a Number of single nucleotide polymorphisms (SNP) used to build the PGS for each cancer type.

by general quality controls, such as Principal Component Analysis (PCA), missing rate, heterozygosity, sex mismatches, and relatedness analysis. Finally, imputation of genetic data was carried out by combining data from different reference panels, specifically the Haplotype Reference Consortium (HRC) and the merged UK10K and 1000 Genomes Phase 3. A detailed description of each procedure has been previously reported (Bycroft et al., 2018).

PGSs for each cancer outcome were selected from PGS catalog (Lambert et al., 2021). Due to the differential computation of each PGS, we selected all genetic scores for every cancer included in the study based on the development method (e.g., LD clumping or pruning) and parameters (e.g., MAF ≥ 0.01, $r^2 \leq 0.1$, $P \leq 5 \times 10^{-8}$) used in the original studies to compute the score. For each selected PGS, linkage disequilibrium (LD) pruning was performed via the LDlink's SNPclip tool (Machiela and Chanock, 2015) using the non-Finnish Europeans of the 1000 Genomes Project retaining only one SNP for markers in high LD ($r^2 > 0.8$). Additionally, only SNPs with a minor frequency allele > 0.05 were included in the scores. The number of SNPs used in each PGS is given in Table 1.

Each PGS was computed as the unweighted sum of the number of risk alleles:

$$PGS_i = \sum_{j=1}^n X_{ij}$$

where n is the number of SNPs used to build the score and X_{ij} is the number of risk alleles (0, 1 or 2) for the participant i at the SNP j .

Based on the distribution of the PGS among the controls, each PGS was then categorized in tertiles, therefore identifying low, intermediate, and high genetic risk categories.

2.5. Statistical analyses

Odds ratios (OR) and 95% confidence intervals (95% CI) for the associations between each environmental exposure and each cancer and between each PGS with its relative cancer were estimated with multi-variable logistic regression models. Each model was adjusted for age and, when necessary, sex. Every environmental variable was analyzed as a continuous exposure and divided in quintiles. PGSs were analyzed as tertiles and as a continuous variable. For environmental exposures and PGS the reference group was quintile 1 (Q1) and tertile 1 (T1), respectively. Moreover, the association between each environmental exposure and female-specific cancers (breast, ovarian and corpus uteri) and all cancers taken together were also investigated. Participants with missing data in at least one of the selected exposures were excluded from the analysis. A comparison between participants with full and missing data is provided as Supplementary Table S1.

Mediation analyses were performed to explain observed significant associations between environmental exposures and each cancer outcome. These analyses were performed by fitting three regression models, as proposed by Baron and Kenny (Baron and Kenny, 1986). Two logistic and one linear model were fitted when the mediator was a continuous variable, while three logistic models were fitted when the mediator was a binary variable.

$$model1 = Y \sim A + Cs$$

$$model2 = Y \sim A + M + Cs$$

$$model3 = M \sim A + Cs$$

Model 1 is the logistic model with the outcome (Y) regressed on the exposure (A) and the covariates (Cs); model 2 is the logistic model with the outcome regressed on the exposure, the covariates, and the mediator (M); and model 3 is the linear or logistic model with the mediator regressed on the exposure and the covariates.

When two mediators were used (i.e., multi-mediator models), an additional regression model was included so that four regression models were fitted.

$$model4 = M2 \sim M1 + Cs$$

In this fourth model, the second mediator (M2) is regressed on the first mediator (M1) and the covariates. Each model was corrected with age and, when necessary, sex. A mediation effect was considered to be significant when the effects of the exposure on the mediator, the mediator on the outcome, and the exposure on the outcome were all significant under the statistical threshold of $P < 0.05$. Significant results were further investigated with a second approach. Specifically, indirect (IE), direct (DE) and total effects (TE), and proportion of mediation were computed using the "mediation" package (v4.5.0) in Rstudio (Tingley et al., 2014). For single mediator analyses, a nonparametric bootstrap with 2000 iterations was used to estimate bias-corrected and accelerated 95% CIs (DiCiccio and Efron, 1996) and P-values. For multiple mediator analyses, percentile 95% CIs were computed using the "multimed" function, which is based on the implementation of the Imai and Yamamoto method (Imai and Yamamoto, 2013).

Detailed information on the mediators used is provided as supplementary methods.

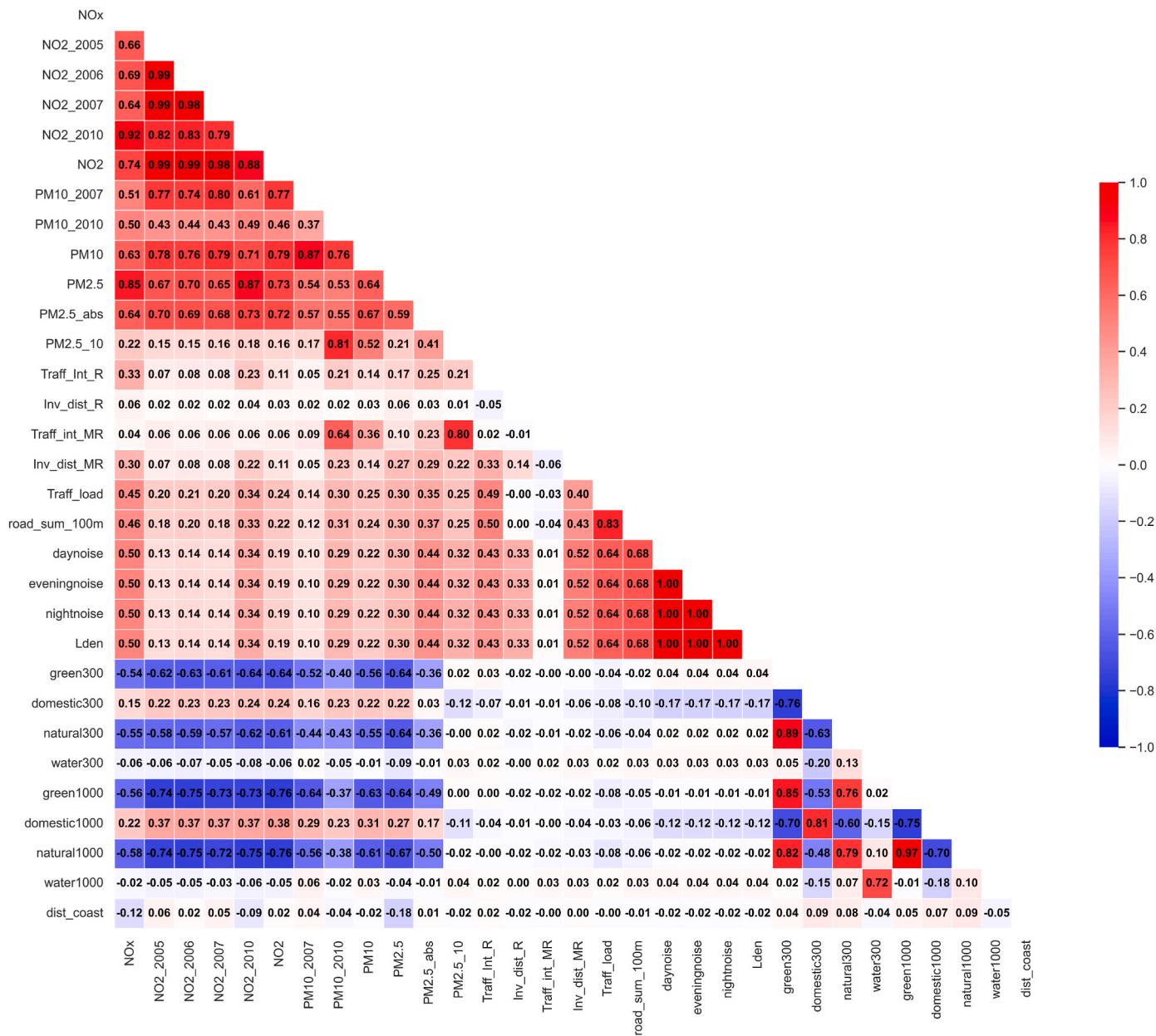


Fig. 1. Heatmap for the correlation among the selected local environment exposures.

The local environment exposures included in this study were first selected according to their correlation coefficient. The threshold used was $r \leq 0.8$. Positive and negative correlation are identified with gradually stronger red and blue colorations, respectively. Even if their correlation was higher than the threshold, NOx and PM2.5 ($r = 0.85$), and greenspace and natural environment ($r = 0.97$) were retained to study their main effect taken individually, as done in the literature.

In the gene-environment (GxE) interaction analysis NOx, NO₂, PM₁₀, PM_{2.5}, and Lden were categorized as low and high exposures groups. The cut off points used were the median for NO₂ Huang et al. (2021); WHO 2005 air quality guidelines for NO₂, PM₁₀, and PM_{2.5}. For Lden, a 55 dB cut off was used according to the Environmental Noise Directive of the European Community. For all other variables quintiles were used.

Multivariable logistic regression models corrected for age and sex (when necessary) were fitted including the interaction term:

$$\text{logit}(P(Y=1)) = \beta_0 + \beta_1 \text{Age} + \beta_2 \text{Sex} + \beta_3 E + \beta_4 \text{PGS} + \beta_5 \text{ExPGS}$$

where E is the environmental exposure categorized as above and PGS is the unweighted PGS divided in tertiles. All multivariable logistic regression models were carried out using Rstudio, version 4.2.2.

Regression results were also tested for significance under Bonferroni's correction. The threshold was calculated dividing 0.05 by the

number of test performed for each outcome ($0.05/(19$ (i.e., the 17 individual cancers, plus the pan-cancer and the female cancer categories) $\times 12)$) and resulted in $P = 2.19 \times 10^{-4}$. For PGS analysis the statistical threshold was $P = 0.05/17 = 2.9 \times 10^{-3}$. For the GxE interaction a threshold of 2-sided $P = 0.05/(17 \times 13) = 2.26 \times 10^{-4}$ was used.

3. Results

3.1. Associations between environmental variables and cancer

One hundred and eighteen associations were observed considering a conventional $P = 0.05$ threshold, while 36 were statistically significant also correcting for multiple testing ($P < 2.19 \times 10^{-4}$). Among the latter, high concentration of NOx, NO₂, PM₁₀, and PM_{2.5} were associated with increased lung cancer risk, with ORs ranging from 1.75 to 1.90, and p-

values ranging from 4.42×10^{-17} to 2.15×10^{-22} . The associations were statistically significant when analyzing individuals belonging to the highest quintile vs the lowest quintile and analyzing the variables as continuous (Table 2).

When comparing the highest quintile with the lowest one, high concentrations of NOx were associated with oesophageal cancer risk ($OR = 1.54$, 95% CI = 1.23–1.94, $P = 1.75 \times 10^{-4}$). Also, PM_{2.5} resulted strongly associated with oesophageal cancer risk in the continuous analysis: $OR = 1.15$, 95% CI = 1.08–1.23, $P = 2.76 \times 10^{-5}$ per 1 $\mu\text{g}/\text{m}^3$ increase (Table 2).

Additionally, high concentrations of NOx, NO₂, PM₁₀, and PM_{2.5} were associated with a decreased risk of developing melanoma, specifically NOx: $OR = 0.74$, 95% CI = 0.66–0.84, $P = 7.54 \times 10^{-7}$; NO₂: $OR = 0.66$, 95% CI = 0.59–0.74, $P = 5.22 \times 10^{-12}$; PM₁₀: $OR = 0.76$, 95% CI = 0.67–0.85, $P = 2.00 \times 10^{-6}$; PM_{2.5}: $OR = 0.71$, 95% CI = 0.63–0.80, $P = 3.46 \times 10^{-8}$ (Table 2).

Living near major roads was associated with lower risk of developing melanoma when comparing individuals in the highest quintile with individuals in the lowest quintile: $OR = 0.80$, 95% CI = 0.72–0.89, $P = 9.27 \times 10^{-5}$, and analyzing the variable as continuous: $OR = 0.93$, 95% CI = 0.90–0.97, $P = 8.39 \times 10^{-5}$ per 1% increase (Table 2).

When comparing the highest with the lowest quintile, high percentage of natural environment near residential coordinates showed a strong protective effect for lung cancer risk ($OR = 0.58$, 95% CI = 0.51–0.66, $P = 2.59 \times 10^{-16}$), while a high percentage of greenspace resulted in increased risk of melanoma ($OR = 1.27$, 95% CI = 1.12–1.43, $P = 1.32 \times 10^{-4}$) (Table 2). The living distance to the coast was associated with increased prostate cancer risk when comparing the highest vs lowest quintile ($OR = 1.13$, 95% CI = 1.06–1.21, $P = 1.98 \times 10^{-4}$) and considering the variable as continuous ($OR = 1.02$, 95% CI = 1.01–1.03, $P = 7.01 \times 10^{-7}$ per 10 Km increase).

Moreover, when comparing the highest quintile with the lowest one, we found a significant association between the living distance to coast and female-specific cancers risk ($OR = 1.10$, 95% CI = 1.05–1.15, $P = 1.70 \times 10^{-4}$). Finally, in the analysis that considered all cancer types together, we found a statistically significant association for both NO₂ ($OR = 1.26$, 95% CI = 1.12–1.41, $P = 9.95 \times 10^{-5}$ per 100 $\mu\text{g}/\text{m}^3$) and PM₁₀ ($OR = 2.91$, 95% CI = 1.72–4.94, $P = 7.30 \times 10^{-5}$ per 100 $\mu\text{g}/\text{m}^3$), and when comparing the highest quintile with the lowest one only for PM₁₀ ($OR = 1.06$, 95% CI = 1.03–1.10, $P = 1.95 \times 10^{-4}$). A statistically significant association was also observed between the living distance to the coast and all cancer risk in both continuous ($OR = 1.01$, 95% CI = 1.00–1.02, $P = 3.62 \times 10^{-11}$ per 10 Km increase) and quintile analysis ($OR = 1.11$, 95% CI = 1.07–1.14, $P = 4.55 \times 10^{-10}$).

The results that were significant after correction for multiple testing are shown in Table 2. All results are shown in Fig. 2 and in Supplementary Table S2. Significant results for female-specific cancer and all cancer analyses are shown in Supplementary Fig. S1.

3.2. Mediation analysis

Mediation analysis was carried out to clarify the association between the living distance from the coast and prostate cancer. Physical activity (walking, moderate, or vigorous physical activity), oily and non-oily fish intake, and time spent outdoor were selected as plausible mediators, as described by White and colleagues (White et al., 2020). In the models where each mediator is regressed on the exposure, the distance to coast significantly predicted moderate physical activity ($\beta = 0.0051$, SE = 0.0023, $P = 0.024$) and the time spent outdoor ($\beta = -0.0023$, SE = 0.0002, $P = 2.72 \times 10^{-25}$). When analyzing the mediator effects, moderate physical activity and the time spent outdoor displayed significant but small effects on prostate cancer risk ($\beta = -0.0029$, SE = 0.0006, $P = 6.04 \times 10^{-7}$ and $\beta = -0.067$, SE = 0.0064, $P = 1.27 \times 10^{-25}$, respectively). Finally, significant results were observed when analyzing the effect of the living distance from the coast on prostate cancer risk when adjusting for each mediator ($\beta = 0.0024$, SE = 0.0004,

$P = 2.30 \times 10^{-8}$ for moderate physical activity; $\beta = 0.0023$, SE = 0.0004, $P = 1.70 \times 10^{-7}$ for time spent outdoor). In the analysis of the indirect effects, we found significant but very small effects exerted by both moderate physical activity (IE = -8.48×10^{-7} , 95% CI = -1.82×10^{-7} – -1.87×10^{-6} , $P = 0.017$) and time spent outdoor (IE = 9.07×10^{-6} , 95% CI = 6.92×10^{-6} – 1.16×10^{-5} , $P < 2 \times 10^{-16}$) on the association between the living distance to coast and prostate cancer. We found no mediation effect for both oily and non-oily fish intake.

In multi-mediator models, the effect of the time spent outdoor on prostate cancer risk was significantly associated with increased levels of vitamin D ($\beta = 0.002$, SE = 0.0006, $p = 4.27 \times 10^{-4}$) and increased vigorous physical activity ($\beta = 0.002$, SE = 0.0006, $p = 3.92 \times 10^{-4}$). However, for both vitamin D levels and vigorous physical activity and using the time spent outdoor as the main mediator, indirect effects were not significant (95% CI includes 0). Mediation results with the Baron and Kenny method are reported in Fig. 3. TEs, DEs and IEs, and proportion of mediation (with their respective 95% CIs and P-values) estimated for each mediator with the “mediation” package are reported in Supplementary Table S3. For melanoma, we did not observe any mediation effect.

3.3. Associations between PGSs and cancer

All PGS showed statistically significant associations when analyzing the tertiles (highest vs lowest, $p < 0.05$). The largest effect was observed for CLL ($OR = 4.82$, 95% CI = 3.82–6.07, $P = 1.93 \times 10^{-40}$), the smallest for stomach cancer ($OR = 1.36$, 95% CI = 1.12–1.65, $P = 1.53 \times 10^{-3}$). The most significant association was observed for breast cancer ($OR = 2.53$, 95% CI = 2.42–2.64, $P < 10^{-250}$). All PGS showed a statistically significant association when analyzed as a continuous variable. The results of the association for each PGSs are reported in Supplementary Table S4.

3.4. Gene – environment (GxE) interactions

In GxE analysis, no results below the Bonferroni threshold of $p = 2.26 \times 10^{-4}$ were found. However, an interesting but non-significant result was found for colorectal cancer. Specifically, when compared with individuals in the lowest exposure group (*i.e.*, quintile Q1) and with the lowest genetic risk (*i.e.*, PGS tertile T1), participants in the highest exposure group for the distance to coast (*i.e.*, quintile Q5) and with the highest genetic risk (*i.e.*, PGS tertile T3) displayed almost three times the risk of developing colorectal cancer ($OR = 2.63$, CI 95% = 1.45–4.78, $P = 1.55 \times 10^{-3}$). The results of all GxE analysis are reported in Supplementary Table S5.

4. Discussion

In the last decade several studies suggested that local environment may affect the risk of developing cancer through several pollutants (Fazeli Dehkordi et al., 2022). However, the associations reported are limited to a small number of cancers and the understanding of their possible interaction with the genetic background is still incomplete. In this study we analyzed 12 variables related to the local environment in relation to the risk of developing 17 cancers. We have confirmed some previously reported associations and identified new ones.

We observed that living close to the coast is associated with a decreased risk of developing prostate cancer ($p = 7.01 \times 10^{-7}$). To clarify this association, we used mediation analysis considering lifestyle factors that could be different in coastal areas compared to more central ones. For example, the diet may differ, and physical activity and the time spent outdoor are reported to be higher in coastal areas (White et al., 2020). We found significant but very small mediation effects for physical activity ($P = 0.017$) and the time spent outdoor ($P < 2 \times 10^{-16}$). Considering the IE scales (IE = -8.48×10^{-7} for moderate physical activity, and IE = 9.07×10^{-6} for the time spent outdoor), these two

Table 2Statistically significant results under the Bonferroni corrected p-value ($p = 2.19 \times 10^{-4}$).

Local environment exposure	Oesophagus	P-value	Lung	P-value	Melanoma	P-value
Nitrogen Oxides (NOx)	NA		1.011 (1.01–1.01)	1.55×10^{-21}	0.993 (0.990–0.995)	7.83×10^{-8}
Nitrogen Oxides (NOx) - Q2	NA		NA		NA	
Nitrogen Oxides (NOx) - Q3	NA		NA		NA	
Nitrogen Oxides (NOx) - Q4	NA		1.35 (1.18–1.54)	1.33×10^{-5}	0.80 (0.71–0.89)	8.78×10^{-5}
Nitrogen Oxides (NOx) - Q5	1.54 (1.23–1.94)	1.75×10^{-4}	1.897 (1.67–2.16)	2.15×10^{-22}	0.74 (0.66–0.84)	7.54×10^{-7}
Nitrogen dioxide (NO2)	NA		1.023 (1.02–1.03)	4.30×10^{-24}	0.983 (0.979–0.987)	8.03×10^{-14}
Nitrogen dioxide (NO2) - Q2	NA		NA		NA	
Nitrogen dioxide (NO2) - Q3	NA		NA		NA	
Nitrogen dioxide (NO2) - Q4	NA		1.468 (1.29–1.67)	1.37×10^{-8}	0.80 (0.72–0.89)	6.45×10^{-5}
Nitrogen dioxide (NO2) - Q5	NA		1.8 (1.58–2.05)	7.63×10^{-19}	0.66 (0.59–0.74)	5.22×10^{-12}
Particulate matter PM2.5	1.15 (1.08–1.23)	2.76×10^{-5}	1.22 (1.17–1.26)	8.84×10^{-25}	0.90 (0.86–0.93)	1.10×10^{-8}
Particulate matter PM2.5 - Q2	NA		NA		NA	
Particulate matter PM2.5 - Q3	NA		NA		NA	
Particulate matter PM2.5 - Q4	NA		1.42 (1.24–1.63)	5.19×10^{-7}	NA	
Particulate matter PM2.5 - Q5	NA		1.85 (1.62–2.11)	1.53×10^{-19}	0.71 (0.63–0.80)	3.46×10^{-8}
Particulate matter PM10	NA		1.10 (1.08–1.12)	3.40×10^{-19}	0.94 (0.92–0.96)	2.00×10^{-10}
Particulate matter PM10 - Q2	NA		NA		NA	
Particulate matter PM10 - Q3	NA		NA		NA	
Particulate matter PM10 - Q4	NA		1.44 (1.26–1.65)	9.52×10^{-8}	NA	
Particulate matter PM10 - Q5	NA		1.75 (1.54–2.00)	4.42×10^{-17}	0.76 (0.67–0.85)	2.00×10^{-6}
Natural environment ^a	NA		0.993 (0.991–0.994)	6.74×10^{-18}	1.04 (1.02–1.05)^b	4.82×10^{-8}
Natural environment - Q2	NA		NA		NA	
Natural environment - Q3	NA		0.77 (0.68–0.87)	2.82×10^{-5}	NA	
Natural environment - Q4	NA		0.68 (0.60–0.77)	2.21×10^{-9}	1.28 (1.13–1.44)	5.62×10^{-5}
Natural environment - Q5	NA		0.58 (0.51–0.66)	2.59×10^{-16}	1.37 (1.21–1.54)	2.18×10^{-7}
Greenspace ^a	NA		0.992 (0.99–0.994)	7.70×10^{-14}	1.04 (1.02–1.05)^b	9.69×10^{-6}
Greenspace - Q2	NA				NA	
Greenspace - Q3	NA				NA	
Greenspace - Q4	NA		0.75 (0.65–0.85)	2.02×10^{-5}	NA	
Greenspace - Q5	NA		0.63 (0.55–0.73)	1.08×10^{-10}	1.27 (1.12–1.43)	1.32×10^{-4}
Inverse distance ^c	NA		NA		0.93 (0.90–0.97)	8.39×10^{-5}
Inverse distance - Q2	NA		NA		NA	
Inverse distance - Q3	NA		NA		0.81 (0.73–0.91)	2.24×10^{-4}
Inverse distance - Q4	NA		NA		0.81 (0.73–0.91)	2.16×10^{-4}
Inverse distance - Q5	NA		NA		0.80 (0.72–0.89)	9.27×10^{-5}
Distance to coast ^d	NA		NA		NA	
Distance to coast - Q2	NA		NA		NA	
Distance to coast - Q3	NA		NA		0.78 (0.69–0.88)	6.95×10^{-5}
Distance to coast - Q4	NA		NA		NA	
Distance to coast - Q5	NA		NA		NA	
Local environment exposure	Prostate	P-value	Female-specific	P-value	All cancer	P-value
Nitrogen Oxides (NOx)	NA		NA		NA	
Nitrogen Oxides (NOx) - Q2	NA		NA		NA	
Nitrogen Oxides (NOx) - Q3	NA		NA		NA	
Nitrogen Oxides (NOx) - Q4	NA		NA		NA	
Nitrogen Oxides (NOx) - Q5	NA		NA		NA	
Nitrogen dioxide (NO2)	NA		NA		1.26 (1.12–1.41)^f	9.95×10^{-5}
Nitrogen dioxide (NO2) - Q2	NA		NA		NA	
Nitrogen dioxide (NO2) - Q3	NA		NA		NA	
Nitrogen dioxide (NO2) - Q4	NA		NA		NA	
Nitrogen dioxide (NO2) - Q5	NA		NA		NA	
Particulate matter PM2.5	NA		NA		2.91 (1.74–4.94)^f	7.30×10^{-5}
Particulate matter PM2.5 - Q2	NA		NA		NA	
Particulate matter PM2.5 - Q3	NA		NA		NA	
Particulate matter PM2.5 - Q4	NA		NA		NA	
Particulate matter PM2.5 - Q5	NA		NA		1.06 (1.03–1.10)	1.95×10^{-4}
Particulate matter PM10	NA		NA		NA	
Particulate matter PM10 - Q2	NA		NA		NA	
Particulate matter PM10 - Q3	NA		NA		NA	
Particulate matter PM10 - Q4	NA		NA		NA	
Particulate matter PM10 - Q5	NA		NA		NA	
Natural environment ^a	NA		NA		NA	
Natural environment - Q2	NA		NA		NA	
Natural environment - Q3	NA		NA		NA	
Natural environment - Q4	NA		NA		NA	
Natural environment - Q5	NA		NA		NA	
Greenspace ^a	NA		NA		NA	

(continued on next page)

Table 2 (continued)

Local environment exposure	Prostate	P-value	Female-specific	P-value	All cancer	P-value
Greenspace - Q2	NA		NA		NA	
Greenspace - Q3	NA		NA		NA	
Greenspace - Q4	NA		NA		NA	
Greenspace - Q5	NA		NA		NA	
Inverse distance ^c	NA		NA		NA	
Inverse distance - Q2	NA		NA		NA	
Inverse distance - Q3	NA		NA		NA	
Inverse distance - Q4	NA		NA		NA	
Inverse distance - Q5	NA		NA		NA	
Distance to coast ^d	1.02 (1.01–1.03) ^e	7.01 × 10 ⁻⁷	NA		1.01 (1.00–1.02) ^e	3.62 × 10 ⁻¹¹
Distance to coast - Q2	NA		NA		NA	
Distance to coast - Q3	1.14 (1.07–1.21)	1.03 × 10 ⁻⁴	1.10 (1.05–1.16)	6.34 × 10 ⁻⁵	1.08 (1.05–1.11)	2.48 × 10 ⁻⁶
Distance to coast - Q4	1.19 (1.11–1.27)	2.60 × 10 ⁻⁷	NA		1.12 (1.08–1.15)	1.02 × 10 ⁻¹¹
Distance to coast - Q5	1.13 (1.06–1.20)	1.98 × 10 ⁻⁴	1.10 (1.05–1.5)	1.70 × 10 ⁻⁴	1.11 (1.07–1.14)	4.55 × 10 ⁻¹⁰

NA is reported when the result is not significant under the Bonferroni corrected p-value threshold.

^a Natural environment and greenspace are referred to percentage levels of greenspace and natural environment in a 1000 m radius centered on the residential coordinates of each participant.

^b Continuous results for greenspace and natural environment are referred for a 10% increase.

^c Logarithm of the inverse distance (1/m) of each participant residential co-ordinates to the nearest major road.

^d Living distance to the coast.

^e Continuous results for distance to coast are reported for 10 Km increase.

^f Continuous results for NO₂, PM₁₀ and all cancer are reported for a 100 µg/m³ increase.

effects could be both negligible in the association between the living distance to the coast and prostate cancer risk. Moreover, in multi-mediator models, we found that the indirect effect of the time spent outdoor on prostate cancer risk was not significant when vitamin D levels and physical activity were individually included in the models. Therefore, we think that the protective effect of living near the coast on prostate cancer risk is mediated by a complex interplay of several lifestyle factors that reflect differences in lifestyle of individuals who live near the coast and individuals living far from it. This is the first evidence for an association between living near the coast and prostate cancer, although a lower incidence was also observed in a study carried out in the United States using spatial patterns and autocorrelation (Mather et al., 2006). Besides prostate cancer, we also found an increased risk of breast cancer associated with living far from the coast in both continuous and quintile analyses. The living distance to the coast was also associated with an increased risk of female-specific cancers. Regarding breast cancer, our result is in line with recent findings (Haraldsdottir et al., 2017), where a reduced risk for breast cancer was observed for women who lived their puberty (*i.e.*, up to the age of 20 years old or more) in coastal villages when compared with women who lived their puberty in the capital city of Reykjavik. However, in the present study, the significance level of our result ($P = 1.21 \times 10^{-3}$) does not overcome the Bonferroni threshold ($P < 2.19 \times 10^{-4}$).

Another novel finding is that high level of NOx, NO₂, PM₁₀, and PM_{2.5} showed an inverse association with melanoma risk ($P = 7.54 \times 10^{-7}$, $P = 5.22 \times 10^{-12}$, $P = 2.00 \times 10^{-6}$, and $P = 3.46 \times 10^{-8}$, respectively) that could be explained by a reduced sun exposure that characterizes cities, where air pollution is higher. We also found that living close to major roads was associated with lower melanoma risk, supporting the idea that the association with air pollution and melanoma risk is just a proxy of a lower sun exposure. We tested this hypothesis in mediation analysis, using the time spent outdoor as a proxy for sun exposure. We found no mediation effects. This null result may be explained by the fact that the variable reporting the time spent outdoor is computed in UKBB as the self-reported average hours spent outside, not considering neither the period of the day nor the direct sun exposure.

In our study, high levels of NOx were positively associated with oesophageal cancer risk in the quintile analysis ($P = 1.75 \times 10^{-4}$), while PM_{2.5} was found to be associated in the continuous analysis ($P = 2.76 \times 10^{-5}$). The effect of NOx on oesophageal cancer risk has never been reported, while the effect of PM_{2.5} was investigated by several studies conducted in China and in the United States (Coleman et al., 2020; Li

et al., 2021, 2022). The authors report an increased risk for the disease for individuals with more exposure to PM_{2.5}, in agreement with our results. Our study is the first to be carried out in a European population.

NOx, NO₂, PM₁₀, and PM_{2.5} were also associated with increased lung cancer risk as strongly supported by previously published epidemiological evidence (Giabattini et al., 2021).

Moreover, even if not significant under the Bonferroni-corrected P-value threshold, we found significant associations between nitrogen air pollutants (NOx and NO₂) and corpus uteri and ovarian cancer risk. In addition, PM₁₀ pollution resulted associated with an increased risk of corpus uteri cancer in both continuous and quintile analyses, while PM_{2.5} pollution was associated with ovarian cancer only in the continuous analysis. Regarding the association between NOx, NO₂ and corpus uteri cancer risk, our results are discordant with literature findings (Raaschou-Nielsen et al., 2011), while an association between PM₁₀ and corpus uteri cancer risk has never been reported before. Regarding ovarian cancer, our results are discordant with recent findings (Coleman et al., 2020).

Additionally, we also observed that living close to greenspaces increases the risk of melanoma and decreases the risk of lung cancer, reflecting the fact that individuals living close to greenspaces are exposed to more sunlight (Astell-Burt et al., 2014) and reduced air pollution (J. G. Su et al., 2011). Our results are conflictual with recent findings (Cao et al., 2023).

Finally, we found a significant association between NO₂ and PM₁₀ and all risk of cancer. Nonetheless, our result is discordant from what is reported in literature. Indeed, different authors found that PM₁₀ was not associated with the overall risk of cancer (Radespiel-Tröger et al., 2018; Shin et al., 2022). As for PM₁₀, literature results about the association between all cancer risk and NO₂ are inconclusive (Al-Ahmadi and Al-Zahrani, 2013; S.-Y. Su et al., 2019).

We observed strong effects of all PGSS on risk of the respective cancers, with relatively large ORs for individuals with the highest count of deleterious alleles compared to individuals with the least (average OR = 2.14) and very robust associations (p-values ranging from $P = 1.53 \times 10^{-3}$ to $P < 10^{-250}$). These results are particularly interesting since in the gene - (local) environment interaction analysis we show that the two are independent from each other and therefore could be combined in multifactorial risk scores to stratify high risk individuals as it has been done for several cancer types (Galeotti et al., 2021; Lee et al., 2022; Torres et al., 2019). Several studies have attempted to identify GxE interaction in cancer, however, with very few exceptions none has

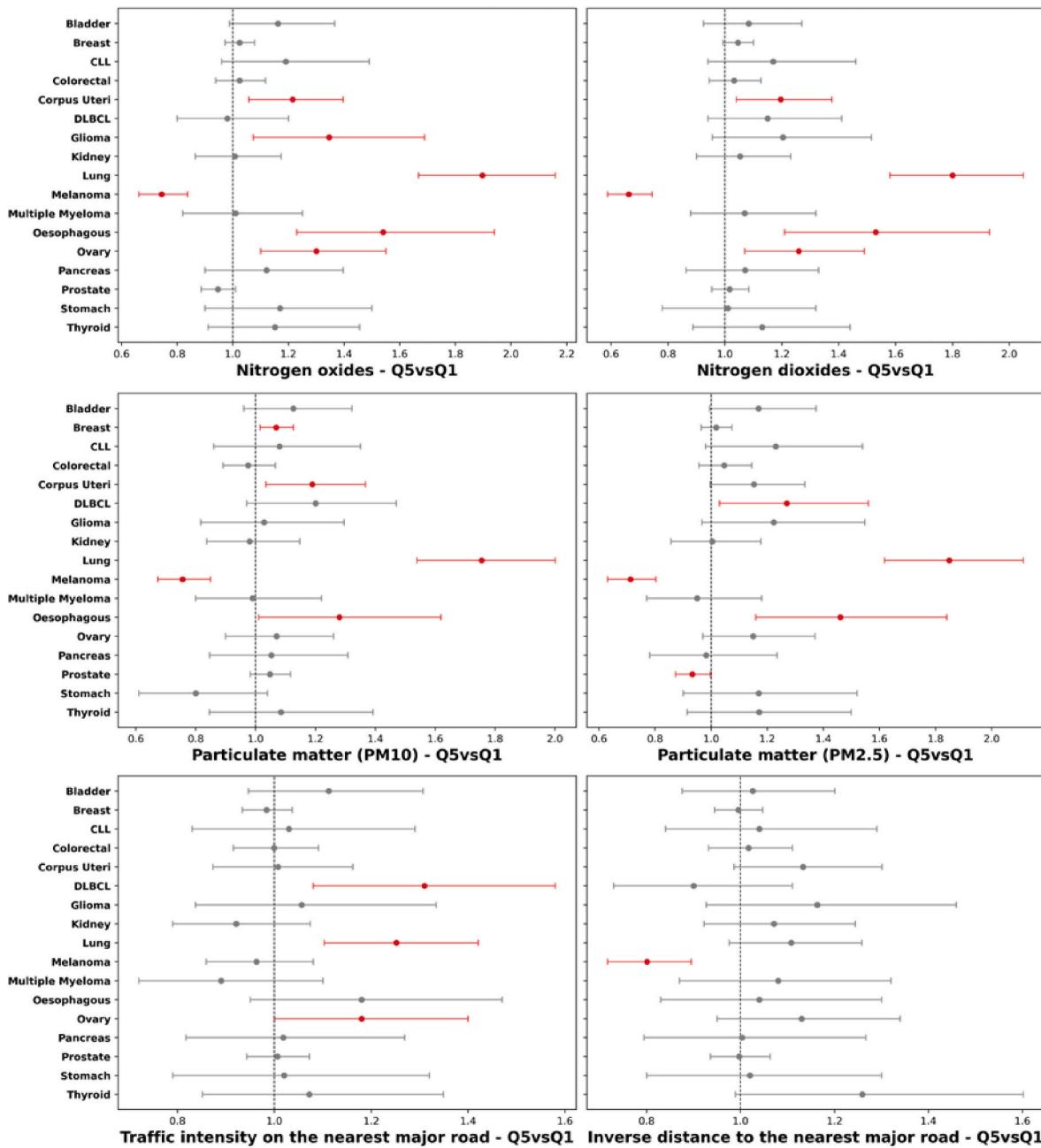


Fig. 2. Forest plots for the associations between every selected local environment exposure and each cancer type.

been found yet (Choi et al., 2021; Huang et al., 2021; Kapoor et al., 2020). We applied a novel approach using the SNPs grouped in a score instead of analyzing the individual interactions with the environment. These results indicate that if an interaction between genetic variability and environmental exposure exists, it must be identified with other approaches, such as genome-wide interactions studies (GWIS), that however need sample sizes much larger than UKBB (Burns et al., 2023; Gref et al., 2017).

Our study has several strengths, such as the comprehensive analysis of the environmental exposure, the integration of genetic data in the form of PGSs, and the homogeneity of the measures across the cohort. We are aware of possible limitations. In UKBB there are missing data on the variations of the concentrations of air pollutants throughout the years. With a longitudinal study it would have been possible to consider the effect of the duration of the exposures, and the possible changes in

the environment that could happen in the lifetime of each participant. Another possible limitation may be given by the ethnicity and the geographic position, since our study was limited to persons of European ancestry living in the UK, and therefore our findings cannot be generalized to other ethnicities and/or geographic location.

5. Conclusions

In conclusion, our results show that the local environment exerts an effect on risk of many cancer types. We confirmed several known associations and identified for the first time the effect of living close to the coast on prostate cancer risk, which could be mediated by physical activity and the time spent outdoor.

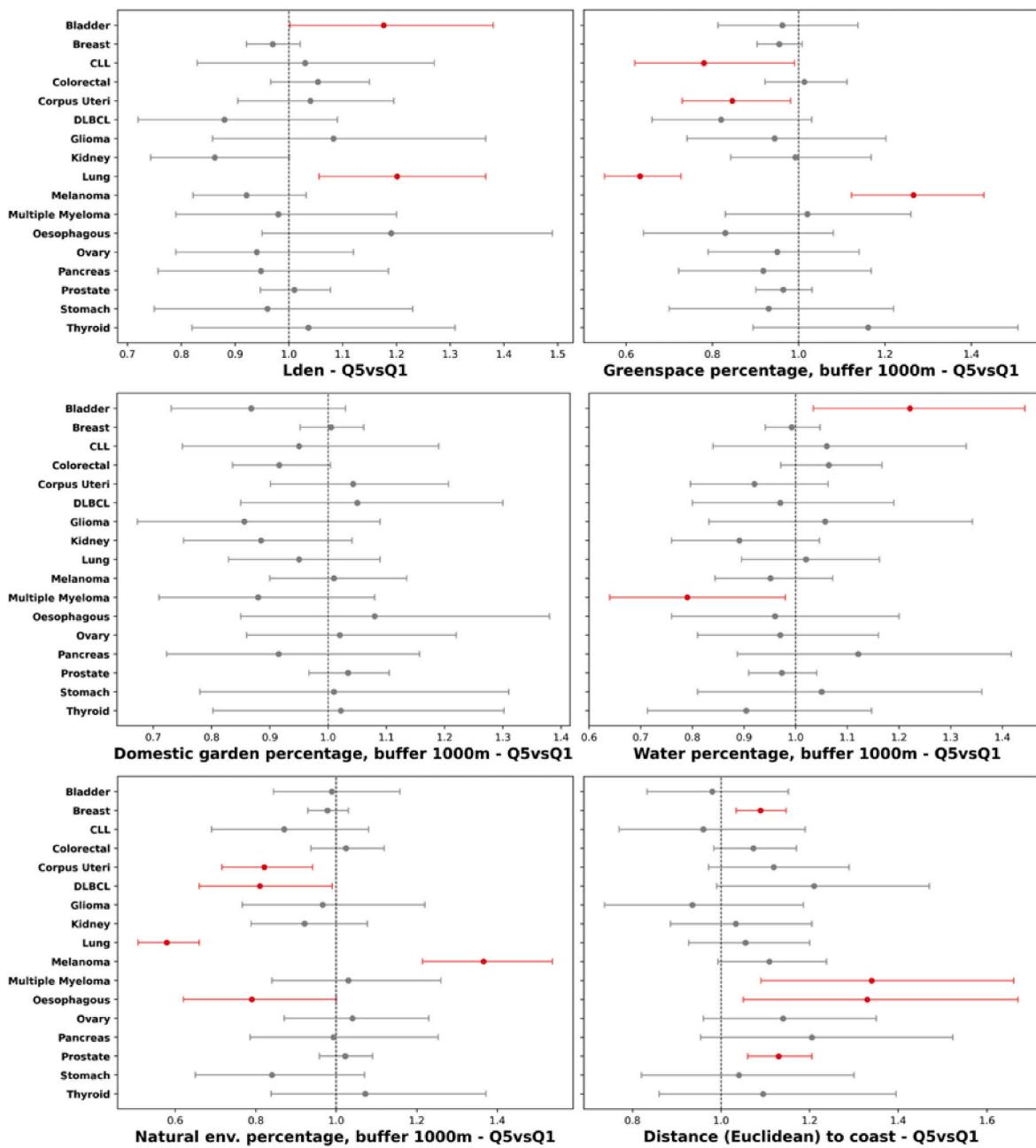


Fig. 2. (continued).

CRediT author statement

Alessio Felici: Conceptualization, Methodology, Formal analysis, Investigation, Software, Validation, Data Curation, Writing – Original Draft, Writing – Review & Editing, Visualization. **Giulia Peduzzi:** Methodology, Resources, Software, Validation, Formal analysis, Data Curation, Writing – Original Draft, Writing – Review & Editing, Visualization. **Francesca Giorgolo:** Software, Validation, Writing – Original Draft. **Andrea Spinelli:** Software, Validation, Writing – Original Draft. **Marco Calderisi:** Software, Resources, Writing – Original Draft, Supervision, Project administration. **Anna Monreale:** Software, Writing – Original Draft. **Riccardo Farinella:** Validation, Writing – Original Draft, Visualization. **Roberto Pellungrini:** Software, Validation, Writing – Original Draft. **Federico Canzian:** Conceptualization, Validation, Resources, Writing – Original Draft, Writing – Review & Editing, Supervision, Project administration, Funding acquisition. **Daniele Campa:**

Conceptualization, Methodology, Validation, Resources, Writing – Original Draft, Writing – Review & Editing, Supervision, Project administration, Funding acquisition.

Funding

The present research was funded with intramural funding of the University of Pisa and German Cancer Research Center (DKFZ). This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

This research is supported by the Ministry of University and Research (MUR) as part of the PON 2014–2020 “Research and Innovation” resources – Green/Innovation Action – DM MUR 1061/2022.

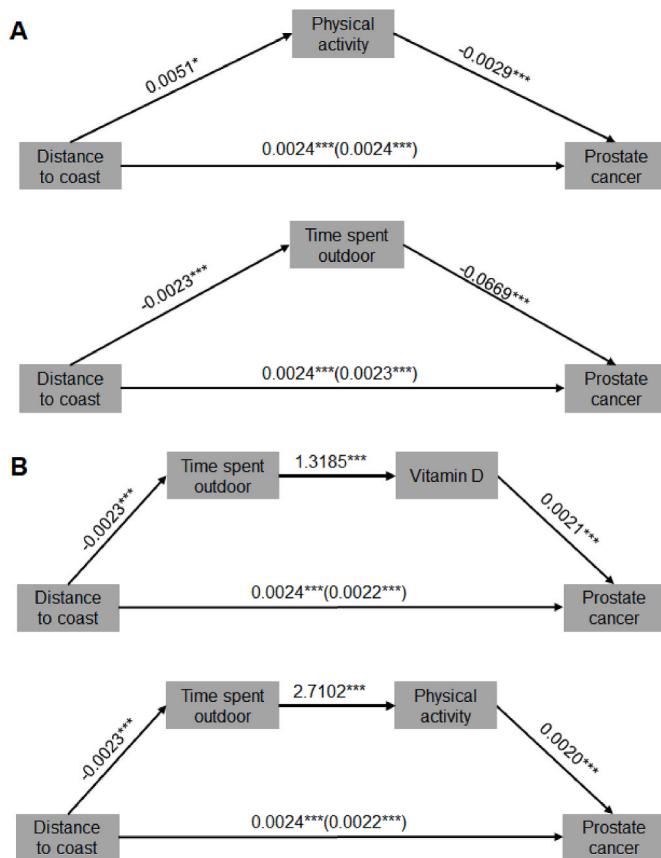


Fig. 3. Mediation results for the living distance to coast and prostate cancer. Panel A and panel B report the mediation results (betas) for the living distance to coast and prostate in cancer in single mediator (A) and multi-mediator (B) analyses with the Baron and Kenny method. Only statistically significant results ($p < 0.05$) are reported. In panel A, sun exposure and physical activity resulted in a weak but significant mediation effect. In panel B, sun exposure effect resulted partially mediated by increased levels of vitamin D and an increase physical activity. For the exposure-outcome path (i.e., distance to coast – prostate cancer), the value within the brackets represents the effect of the exposure when the mediator is present, while the value outside the brackets represents the total effect of the exposure when the mediator is not included. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Ethics statement

All participants provided written informed consent before their enrollment in the UK Biobank cohort. The UKBB study has received the approval from the North-West Multi-centre Research Ethics Committee (MREC).

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors do not have permission to share data.

Acknowledgements

Data used for this research were gathered from the UK Biobank Resource (www.ukbiobank.ac.uk) under application No. 66591. The authors want to thank the UK Biobank participants for their

contribution.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.envres.2023.117562>.

References

- Al-Ahmadi, K., Al-Zahrani, A., 2013. NO(2) and cancer incidence in Saudi Arabia. *Int. J. Environ. Res. Publ. Health* 10 (11), 5844–5862. <https://doi.org/10.3390/ijerph10115844>.
- Astell-Burt, T., Feng, X., Kolt, G.S., 2014. Neighbourhood green space and the odds of having skin cancer: multilevel evidence of survey data from 267072 Australians. *J. Epidemiol. Community Health* 68 (4), 370–374. <https://doi.org/10.1136/jech-2013-203043>.
- Baron, R.M., Kenny, D.A., 1986. The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *J. Pers. Soc. Psychol.* 51 (6), 1173–1182. <https://doi.org/10.1037/0022-3514.51.6.1173>.
- Burns, A.C., Phillips, A.J.K., Rutter, M.K., Saxena, R., Cain, S.W., Lane, J.M., 2023. Genome-wide gene by environment study of time spent in daylight and chronotype identifies emerging genetic architecture underlying light sensitivity. *Sleep* 46 (3), 1–10. <https://doi.org/10.1093/SLEEP/ZSAC287>.
- Bycroft, C., Freeman, C., Petkova, D., Band, G., Elliott, L.T., Sharp, K., Motyer, A., Vukcevic, D., Delaneau, O., O'Connell, J., Cortes, A., Welsh, S., Young, A., Effingham, M., McVean, G., Leslie, S., Allen, N., Donnelly, P., Marchini, J., 2018. The UK Biobank resource with deep phenotyping and genomic data. *Nature* 562 (7726), 203–209. <https://doi.org/10.1038/s41586-018-0579-z>.
- Cao, Z., Xu, C., Li, S., Wang, Y., Yang, H., 2023. Residential greenspace and risk of cancer: a prospective cohort study from the UK Biobank. *Sci. Total Environ.* 871, 162145 <https://doi.org/10.1016/J.SCITOTENV.2023.162145>.
- Chang, C.C., Chow, C.C., Tellier, L.C.A.M., Vattikuti, S., Purcell, S.M., Lee, J.J., 2015. Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience* 4 (1). <https://doi.org/10.1186/S13742-015-0047-8>.
- Choi, J., Jia, G., Wen, W., Shu, X.O., Zheng, W., 2021. Healthy lifestyles, genetic modifiers, and colorectal cancer risk: a prospective cohort study in the UK Biobank. *Am. J. Clin. Nutr.* 113 (4), 810–820. <https://doi.org/10.1093/AJCN/NQAA404>.
- Ciabattini, M., Rizzello, E., Lucaroni, F., Palombi, L., Boffetta, P., 2021. Systematic review and meta-analysis of recent high-quality studies on exposure to particulate matter and risk of lung cancer. *Environ. Res.* 196 <https://doi.org/10.1016/j.envres.2020.110440>.
- Claeson, M., Andersson, E.M., Wallin, M., Wastensson, G., Wennberg, A.-M., Paoli, J., Gonzalez, H., 2012. Incidence of cutaneous melanoma in Western Sweden, 1970–2007. *Melanoma Res.* 22 (5), 392–398. <https://doi.org/10.1097/CMR.0b013e32835861d6>.
- Coleman, N.C., Burnett, R.T., Ezzati, M., Marshall, J.D., Robinson, A.L., Pope, C.A., 2020. Fine particulate matter exposure and cancer incidence: analysis of SEER cancer registry data from 1992–2016. *Environ. Health Perspect.* 128 (10), 107004 <https://doi.org/10.1289/EHP7246>.
- DiCiccio, T.J., Efron, B., 1996. Bootstrap confidence intervals. *Stat. Sci.* 11 (3) <https://doi.org/10.1214/ss/1032280214>.
- Fazeli Dehkordi, Z.S., Khatami, S.M., Ranjbar, E., 2022. The associations between urban form and major non-communicable diseases: a systematic review. *J. Urban Health* : Bull. N. Y. Acad. Med.
- Galeotti, A.A., Gentiluomo, M., Rizzato, C., Obabee, O., Neoptolemos, J.P., Pasquali, C., Nentwich, M., Cavestro, G.M., Pezzilli, R., Greenhalf, W., Hollecze, B., Schroeder, C., Schötzker, B., Ivanauskas, A., Ginocchi, L., Key, T.J., Hegyi, P., Archibugi, L., Darvasi, E., et al., 2021. Polygenic and multifactorial scores for pancreatic ductal adenocarcinoma risk prediction. *J. Med. Genet.* 58 (6), 369–377. <https://doi.org/10.1136/JMEDGENET-2020-106961>.
- Gref, A., Merid, S.K., Gruzieva, O., Ballereau, S., Becker, A., Bellander, T., Bergström, A., Bossé, Y., Bottai, M., Chan-Yeung, M., Fuertes, E., Ierodiakonou, D., Jiang, R., Joly, S., Jones, M., Kober, M.S., Korek, M., Kozyrskyj, A.L., Kumar, A., et al., 2017. Genome-wide interaction analysis of air pollution exposure and childhood asthma with functional follow-up. *Am. J. Respir. Crit. Care Med.* 195 (10), 1373–1383. https://doi.org/10.1164/RCCM.201605-1026OC/SUPPL_FILE/DISCLOSURES.PDF.
- Gucalp, A., Traina, T.A., Eisner, J.R., Parker, J.S., Selitsky, S.R., Park, B.H., Elias, A.D., Baskin-Bey, E.S., Cardoso, F., 2019. Male breast cancer: disease distinct from female breast cancer. *Breast Cancer Res. Treat.* 173 (1), 37–48. <https://doi.org/10.1007/s10549-018-4921-9>.
- Haraldsdottir, A., Steingrimsdottir, L., Valdimarsdottir, U.A., Aspelund, T., Tryggvadottir, L., Harris, T.B., Launer, L.J., Mucci, L.A., Giovannucci, E.L., Adami, H.O., Gudnason, V., Torfadottir, J.E., 2017. Early life residence, fish consumption, and risk of breast cancer. *Cancer Epidemiol. Biomark. Prev.* 26 (3), 346–354. <https://doi.org/10.1158/1055-9965.EPI-16-0473-T>.
- Huang, Y., Zhu, M., Ji, M., Fan, J., Xie, J., Wei, X., Jiang, X., Xu, J., Chen, L., Yin, R., Wang, Y., Dai, J., Jin, G., Xu, L., Hu, Z., Ma, H., Shen, H., 2021. Air pollution, genetic factors, and the risk of lung cancer: a prospective study in the UK Biobank. *Am. J. Respir. Crit. Care Med.* 204 (7), 817–825. <https://doi.org/10.1164/RCCM.202011-4063OC>.
- Imai, K., Yamamoto, T., 2013. Identification and sensitivity analysis for multiple causal mechanisms: revisiting evidence from framing experiments. *Polit. Anal.* 21 (2), 141–171. <https://doi.org/10.1093/pan/mps040>.

- Kapoor, P.M., Lindström, S., Behrens, S., Wang, X., Michailidou, K., Bolla, M.K., Wang, Q., Dennis, J., Dunning, A.M., Pharoah, P.D.P., Schmidt, M.K., Kraft, P., García-Closas, M., Easton, D.F., Milne, R.L., Chang-Claude, J., 2020. Assessment of interactions between 205 breast cancer susceptibility loci and 13 established risk factors in relation to breast cancer risk, in the Breast Cancer Association Consortium. *Int. J. Epidemiol.* 49 (1), 216. <https://doi.org/10.1093/IJE/DYZ193>.
- Korgavkar, K., Lee, K.C., Weinstock, M.A., 2014. Higher melanoma incidence in coastal versus inland counties in California. *Melanoma Res.* 24 (3), 280–285. <https://doi.org/10.1097/CMR.0000000000000067>.
- Lambert, S.A., Gil, L., Jupp, S., Ritchie, S.C., Xu, Y., Buniello, A., McMahon, A., Abraham, G., Chapman, M., Parkinson, H., Danesh, J., MacArthur, J.A.L., Inouye, M., 2021. The Polygenic Score Catalog as an open database for reproducibility and systematic evaluation. *Nat. Genet.* 53 (4), 420–425. <https://doi.org/10.1038/s41588-021-00783-5>.
- Lee, A., Yang, X., Tyrer, J., Gentry-Maharaj, A., Ryan, A., Mavaddat, N., Cunningham, A. P., Carver, T., Archer, S., Leslie, G., Kalsi, J., Gaba, F., Manchanda, R., Gayther, S., Ramus, S.J., Walter, F.M., Tischkowitz, M., Jacobs, I., Menon, U., et al., 2022. Original research: comprehensive epithelial tubo-ovarian cancer risk prediction model incorporating genetic and epidemiological risk factors. *J. Med. Genet.* 59 (7), 632. <https://doi.org/10.1136/JMEDGENET-2021-107904>.
- Li, P., Guo, X., Jing, J., Hu, W., Wei, W.Q., Qi, X., Zhuang, G., 2022. The lag effect of exposure to PM2.5 on esophageal cancer in urban-rural areas across China. *Environ. Sci. Pollut. Control Ser.* 29 (3), 4390–4400. <https://doi.org/10.1007/S11356-021-15942-8/FIGURES/3>.
- Li, P., Jing, J., Guo, W., Guo, X., Hu, W., Qi, X., Wei, W.Q., Zhuang, G., 2021. The associations of air pollution and socioeconomic factors with esophageal cancer in China based on a spatiotemporal analysis. *Environ. Res.* 196, 110415 <https://doi.org/10.1016/J.ENVRES.2020.110415>.
- Machiela, M.J., Chanock, S.J., 2015. LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. *Bioinformatics* 31 (21), 3555–3557. <https://doi.org/10.1093/BIOINFORMATICS/BTV402>.
- Mather, F.J., Chen, V.W., Morgan, L.H., Correa, C.N., Shaffer, J.G., Srivastav, S.K., Rice, J.C., Blount, G., Swalm, C.M., Wu, X., Scribner, R.A., 2006. Hierarchical modeling and other spatial analyses in prostate cancer incidence data. *Am. J. Prev. Med.* 30 (2) <https://doi.org/10.1016/j.amepre.2005.09.012>.
- Pourvakhshoori, N., Khankeh, H.R., Stueck, M., Farrokhi, M., 2020. The association between air pollution and cancers: controversial evidence of a systematic review. *Environ. Sci. Pollut. Control Ser.* 27 (31), 38491–38500. <https://doi.org/10.1007/S11356-020-10377-Z/TABLES/5>.
- Raaschou-Nielsen, O., Andersen, Z.J., Hvidberg, M., Jensen, S.S., Ketzel, M., Sørensen, M., Hansen, J., Loft, S., Overvad, K., Tjønneland, A., 2011. Air pollution from traffic and cancer incidence: a Danish cohort study. *Environ. Health* 10 (1), 67. <https://doi.org/10.1186/1476-069X-10-67>.
- Radespiel-Tröger, M., Geiss, K., Twardella, D., Maier, W., Meyer, M., 2018. Cancer incidence in urban, rural, and densely populated districts close to core cities in Bavaria, Germany. *Int. Arch. Occup. Environ. Health* 91 (2), 155–174. <https://doi.org/10.1007/s00420-017-1266-3>.
- Reynolds, P., Hurley, S.E., Quach, A.T., Rosen, H., Von Behren, J., Hertz, A., Smith, D., 2005. Regional variations in breast cancer incidence among California women, 1988–1997. *Cancer Causes Control* 16 (2), 139–150. <https://doi.org/10.1007/S10552-004-2616-5/METRICS>.
- Reynolds, P., Hurley, S., Goldberg, D.E., Anton-Culver, H., Bernstein, L., Deapen, D., Horn-Ross, P.L., Peel, D., Pinder, R., Ross, R.K., West, D., Wright, W.E., Ziogas, A., 2004. Regional variations in breast cancer among California teachers. *Epidemiology* 15 (6), 746–754. <https://doi.org/10.1097/01.EDE.0000134863.45834.50>.
- Sánchez, C.O., García-Pérez, J., Gómez-Barroso, D., Domínguez-Castillo, A., Romaguera, E.P., Cañete, A., Ortega-García, J.A., Ramis, R., 2022. Exploring blue spaces' effects on childhood leukaemia incidence: a population-based case-control study in Spain. *Int. J. Environ. Res. Publ. Health* 19 (9). <https://doi.org/10.3390/IJERPH19095232>.
- Shin, M., Kim, O.-J., Yang, S., Choe, S.-A., Kim, S.-Y., 2022. Different mortality risks of long-term exposure to particulate matter across different cancer sites. *Int. J. Environ. Res. Publ. Health* 19 (6). <https://doi.org/10.3390/ijerph19063180>.
- Sudlow, C., Gallacher, J., Allen, N., Beral, V., Burton, P., Danesh, J., Downey, P., Elliott, P., Green, J., Landray, M., Liu, B., Matthews, P., Ong, G., Pell, J., Silman, A., Young, A., Sprosen, T., Peakman, T., Collins, R., 2015. UK Biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med.* 12 (3) <https://doi.org/10.1371/journal.pmed.1001779>.
- Tingley, D., Yamamoto, T., Hirose, K., Keele, L., Imai, K., 2014. mediation: R package for causal mediation analysis. *J. Stat. Software* 59 (5). <https://doi.org/10.18637/jss.v059.i05>.
- Torres, D., Lorenzo Bermejo, J., García Mesa, K., Gilbert, M., Briceño, I., Pohl-Zeidler, S., González Silos, R., Boekstegers, F., Plass, C., Hamann, U., 2019. Interaction between genetic ancestry and common breast cancer susceptibility variants in Colombian women. *Int. J. Cancer* 144 (9), 2181–2191. <https://doi.org/10.1002/IJC.32023>.
- Wang, X., Qian, Z., Min, J., Zhang, Z., Cai, M., Chen, L., Wu, Y., Li, H., Liu, E., McMillin, S. E., Lin, H., 2022. Population attributable fraction of lung cancer due to genetic variants, modifiable risk factors, and their interactions: a nationwide prospective cohort study. *Chemosphere* 301. <https://doi.org/10.1016/J.CHEMOSPHERE.2022.134773>.
- White, M.P., Elliott, L.R., Gascon, M., Roberts, B., Fleming, L.E., 2020. Blue space, health and well-being: a narrative overview and synthesis of potential benefits. *Environ. Res.* 191 <https://doi.org/10.1016/J.ENVRES.2020.111016>.
- Yu, P., Kusuma, J.D., Suarez, M.A.R., Shiao, S.Y.P.K., 2018. Lung cancer susceptibility from GSTM1 deletion and air pollution with smoking status: a meta-prediction of worldwide populations. *Oncotarget* 9 (57), 31120. <https://doi.org/10.18632/ONCOTARGET.25693>.
- Zare Sakhlidi, M.J., Yang, J., Mehrparvar, A.H., Dzhambov, A.M., Ebrahimi, A.A., Dadvand, P., Jacquemin, B., 2022. Exposure to greenspace and cancer incidence, prevalence, and mortality: a systematic review and meta-analyses. *Sci. Total Environ.* 838, 156180 <https://doi.org/10.1016/J.SCITOTENV.2022.156180>.