De Vocht, F. (2019). Residential exposure to radon and DNA methylation across the lifecourse: an exploratory study in the ALSPAC birth cohort. *Wellcome Open Research*, *4*, [3]. https://doi.org/10.12688/wellcomeopenres.14991.1

RESEARCH ARTICLE

Residential exposure to radon and DNA methylation across the lifecourse: an exploratory study in the ALSPAC birth cohort [version 1; referees: 1 approved]

Frank de Vocht [1], Matthew Suderman[1,2], Alberto Ruano-Ravina[3], Richard Thomas [1,4], Richard Wakeford[5], Caroline Relton[1,2], Kate Tilling[1,2], Andy Boyd [1,4]

[1]Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK
[2]MRC Integrative Epidemiology Unit, University of Bristol, Bristol, UK
[3]University of Santiago de Compostela, Santiago de Compostela, Spain
[4]Avon Longitudinal Study of Parents and Children, University of Bristol, Bristol, UK
[5]Centre for Occupational and Environmental Health, University of Manchester, Manchester, UK

v1 **First published:** 07 Jan 2019, **4**:3 (https://doi.org/10.12688/wellcomeopenres.14991.1)
Latest published: 07 Jan 2019, **4**:3 (https://doi.org/10.12688/wellcomeopenres.14991.1)

Abstract

Background: Radon (and its decay products) is a known human carcinogen and the leading cause of lung cancer in never-smokers and the second in ever-smokers. The carcinogenic mechanism from radiation is a combination of genetic and epigenetic processes, but compared to the genetic mechanisms, epigenetic processes remain understudied in humans. This study aimed to explore associations between residential radon exposure and DNA methylation in the general population.
Methods: Potential residential radon exposure for 75-metre area buffers was linked to genome-wide DNA methylation measured in peripheral blood from children and mothers of the Accessible Resource for Integrated Epigenomic Studies subsample of the ALSPAC birth cohort. Associations with DNA methylation were tested at over 450,000 CpG sites at ages 0, 7 and 17 years (children) and antenatally and during middle-age (mothers). Analyses were adjusted for potential residential and lifestyle confounding factors and were determined for participants with complete data (n = 786-980).
Results: Average potential exposure to radon was associated in an exposure-dependent manner with methylation at cg25422346 in mothers during pregnancy, with no associations at middle age. For children, radon potential exposure was associated in an exposure-dependent manner with methylation of cg16451995 at birth, cg01864468 at age 7, and cg04912984, cg16105117, cg23988964, cg04945076, cg08601898, cg16260355 and cg26056703 in adolescence.
Conclusions: Residential radon exposure was associated with DNA methylation in an exposure-dependent manner. Although residual confounding cannot be excluded, the identified associations may show biological mechanisms involved in early biological effects from radon exposure.

Open Peer Review

Referee Status: ✔

	Invited Referees 1
version 1 published 07 Jan 2019	✔ report

1 **Beata Peplonska**, Nofer Institute of Occupational Medicine, Poland

Any reports and responses or comments on the article can be found at the end of the article.

Keywords
ALSPAC, Radon, DNA methylation, epigenetics. ARIES, geo-spatial epidemiology

This article is included in the Avon Longitudinal Study of Parents and Children (ALSPAC) gateway.

Corresponding author: Frank de Vocht (frank.devocht@bristol.ac.uk)

Author roles: de Vocht F: Conceptualization, Formal Analysis, Methodology, Project Administration, Writing – Original Draft Preparation, Writing – Review & Editing; **Suderman M**: Data Curation, Formal Analysis, Investigation, Methodology, Resources, Software; **Ruano-Ravina A**: Conceptualization, Investigation, Methodology; **Thomas R**: Data Curation, Resources, Software, Visualization; **Wakeford R**: Conceptualization, Investigation; **Relton C**: Conceptualization, Funding Acquisition, Investigation, Methodology, Resources; **Tilling K**: Funding Acquisition, Investigation, Methodology; **Boyd A**: Conceptualization, Data Curation, Investigation, Methodology, Resources, Validation

Competing interests: No competing interests were disclosed.

Grant information: The Accessible Resource for Integrated Epigenomics Studies (ARIES) was funded by the BBSRC (BBI025751/1 and BB/I025263/1). Supplementary funding to generate DNA methylation data which are (or will be) included in ARIES has been obtained from the MRC, ESRC, NIH and other sources. ARIES is maintained under the auspices of the MRC Integrative Epidemiology Unit at the University of Bristol (MC_UU_12013/2, MC_UU_12013/8 and MC_UU_12013/9). This work was also supported by the BBSRC and ESRC (grant number ES/N000498/1), the UK National Institute for Health Research (NIHR) Health Protection Research Unit (HPRU) in Evaluation, NIHR Biomedical Research Centre (BRC) and NIHR School for Public Health Research (SPHR). The UK Medical Research Council and Wellcome Trust (102215/13/2) and the University of Bristol provide core support for ALSPAC. This publication is the work of the authors and FdV will serve as guarantors for the contents of this paper. RT and AB were supported by UK Natural Environment Research Council (R8/H12/83/NE/P01830/1) and the Wellcome Trust (WT086118/Z/08/Z). A comprehensive list of grants funding is available on the ALSPAC website (http://www.bristol.ac.uk/alspac/external/documents/grant-acknowledgements.pdf); No external funding was obtained for this study.

How to cite this article: de Vocht F, Suderman M, Ruano-Ravina A *et al.* **Residential exposure to radon and DNA methylation across the lifecourse: an exploratory study in the ALSPAC birth cohort [version 1; referees: 1 approved]** Wellcome Open Research 2019, **4**:3 (https://doi.org/10.12688/wellcomeopenres.14991.1)

First published: 07 Jan 2019, **4**:3 (https://doi.org/10.12688/wellcomeopenres.14991.1)

Introduction

Radon is a noble gas with no stable isotopes. Radon-222 (half-life ($t_{1/2}$) 3.82 days) and radon-220 ($t_{1/2}$ 55.2 seconds) are found in the environment as components of the radioactive decay chains of the naturally occurring, long-lived radionuclides uranium-238 and thorium-232, respectively, which are found to varying extents in all rocks and soil. Radon-222 is the product of the decay of long-lived radium-226 ($t_{1/2}$, 1600 years), and in most parts of the world, following their accumulation in enclosed spaces, inhalation of ^{222}Rn and its short-lived radioactive decay products is the largest contribution of human exposure to ionising radiation. Globally, inhalation of ^{222}Rn and its progeny is estimated to provide nearly half of the average annual effective dose (the radiation- and tissue-weighted whole-body absorbed dose) of 2.4 mSv from natural sources of ionising radiation[1]. However, the geographical variation of the effective dose from ^{222}Rn and its progeny is considerable, with a typical range of 0.2-10 mSv per annum. The contribution to the global average annual effective dose from the inhalation of ^{220}Rn and its progeny is much less at 0.1 mSv.

Radon-222 and some of its short-lived progeny deliver most of their radiation dose through short-range alpha-particle emission and, following inhalation, the radiation dose is received primarily by the bronchial epithelium. There is compelling epidemiological and experimental evidence that ^{222}Rn and its decay products (hereafter, "radon") cause lung cancer, with exposure-response associations approximately linear with no evidence of a threshold[2,3], and radon has been classified as a Group 1 carcinogen ("carcinogenic to humans") by the International Agency for Research on Cancer (IARC)[4].

Exposure to radon is considered the second leading cause of lung cancer after tobacco smoking, and the principal cause in never-smokers[5,6]. The fraction of the lung cancer burden attributable to indoor exposure to radon ranges from 3% to 14% across the world, and is estimated at 3.3%[7], or 1,100 lung cancer deaths annually, in the UK specifically[3]. Once inhaled, radon gas itself is mostly exhaled again, but a large proportion of the inhaled short-lived radon progeny deposits in the airways of the lungs with the alpha-particles emitted by ^{218}Po and ^{214}Po dominating the dose to the lung. In contrast, radon gas transported from the lung makes a larger contribution than its decay products to doses to organs/tissues other than the lung, particularly those with a comparatively high fat content (including the red bone marrow (RBM)). However, the evidence for radon causing cancers other than lung cancer is limited and relates to the fact that doses to other tissues from radon are relatively small. For example, the UK average annual equivalent dose (the radiation-weighted absorbed dose) to the RBM from radon is 80 µSv (children and adults) as compared to the RBM dose of 1430 µSv (5-year old) and 1070 µSv (adult) from all-natural sources[8]; this RBM dose from radon compares with that to the lung of 10,000 µSv.

Worldwide, the population-weighted geometric mean indoor level of radon is estimated to be 30 Bq m^{-3} [9], with a large geographical variation[3]. In England, the concentration in homes is about 20 Bq m^{-3} on average, but it ranges from 5 to 10,000 Bq m^{-3} and more in some radon-prone areas; for comparison, the average outdoor concentration is 4 Bq m^{-3} [2]. Variation between and within small geographical areas, as well as over time, can be the result of many factors including the abundance of ^{226}Ra in the ground, fissuring of rocks, permeability of the soil, openings in the foundations of buildings through which radon can enter, and the extent to which a particular structure retains radon, including ventilation[3,10]. In Great Britain, a strong correlation between domestic radon levels and socio-economic status (SES) has been observed, where lower SES residences have, on average, only two-thirds of the radon levels of those of the more affluent, which may be related to greater underpressure in warmer and better-sealed houses[11]. Because people spend a significant portion of their time indoors, homes are typically the primary source of indoor radon exposure[3], and within houses concentrations can also widely vary, with (in the USA) concentrations typically 50% higher in basements compared to the ground floor[12].

The World Health Organisation (WHO) and International Commission on Radiological Protection (ICRP) recommend radon reference levels for homes in the range of 100-300 Bq m^{-3} [13], with the ICRP reference level of 300 Bq m^{-3} having been incorporated as the upper limit for the reference level by the European Union[14]. The annual effective dose for a dwelling at 300 Bq m^{-3}, and given several assumptions, is estimated at about 14 millisievert (mSv)[15]. In the UK, Public Health England recommends that indoor radon levels should be below 200 Bq m^{-3} (averaged over the home; the Action Level), which corresponds to about 12 mSv annual effective dose[2], with 100 Bq m^{-3} being considered the Target Level for remediation work and for new buildings[2].

The multistage carcinogenic process is in all probability a mixture of genetic and epigenetic process. Ionizing radiation, in addition to producing mutations mainly by gene deletion and gross chromosomal damage, can also induce epigenetic effects[4]. Residential radon exposure has been associated with DNA-repair gene polymorphisms in adults[16] and children[17], with the latter study also reporting double-strand break repair gene polymorphisms. Epigenetics describe heritable chemical modifications of DNA and chromatin affecting gene expression, and include DNA methylation, histone modifications and microRNAs which can act in concert to regulate gene expression[18]. In addition, the 'bystander effect', in which cells that are not directly irradiated, but are in the neighbourhood of cells that have, also exhibit phenotypic features of genomic instability that is considered to be epigenetic in nature[4]. DNA methylation is the most stable and most readily quantifiable epigenetic marker and is sensitive to pre- and postnatal exogenous influences[19]. However, there is only limited data on effects of radon exposure on DNA methylation in humans, with some evidence from high exposed uranium miners in China[20].

This study aims to explore whether there is evidence of DNA methylation from residential radon exposure in the general population and assesses whether any methylation varies across the lifecourse.

Methods

Data

This study used data from the Avon Longitudinal Study of Parents and Children (ALSPAC)[21,22]. ALSPAC recruited 14,541 pregnant women with expected delivery dates between April 1991 and December 1992, which resulted in 14,062 live births of which 13,988 children were alive at 1 year of age. Details of all data searchable though are provided at the ALSPAC data dictionary.

A sub-sample of 1,018 ALSPAC mother–child pairs had DNA methylation measured using the Infinium HumanMethylation450 BeadChip (Illumina, Inc.)[23] as part of the Accessible Resource for Integrated Epigenomic Studies (ARIES) project[24]. For this study DNA methylation data generated from cord blood, venous blood samples at age 7 years and again at age 15 or 17, and additionally from the mothers during pregnancy and at middle age were used. All DNA methylation analyses were performed at the University of Bristol as part of the ARIES project and has been described in detail previously[24].

Ethical approval for this study was obtained from the ALSPAC Ethics and Law Committee and the Local Research Ethics Committees (Reference B2805).

DNA methylation

DNA methylation profiles for ALSPAC children were generated at birth from cord blood and in childhood from peripheral blood at ages 7 and 15–17 years using the Illumina Infinium HumanMethylation450 BeadChip as part of the Accessible Resource for Integrated Epigenomic Studies (ARIES)[25]. DNA was bisulphite-converted using the Zymo EZ DNA Methylation™ kit (Zymo, Irvine, CA). Infinium HumanMethylation450 BeadChips (Illumina, Inc.) and used to measure genome-wide DNA methylation levels at over 485,000 CpG sites. The arrays were scanned using an Illumina iScan, with initial quality review using GenomeStudio (version 2011.1). This assay detects methylation of cytosine at CpG islands using one probe to detect the methylated and one to detect the unmethylated loci. Single-base extension of the probes incorporated a labelled chain-terminating ddNTP, which was then stained with a fluorescence reagent. The ratio of fluorescent signals from the methylated site versus the unmethylated site determines the level of methylation at the locus.

Quality control and normalization of the profiles was performed using the meffil R package (version 1.1.0) as previously described[26]. The level of methylation is expressed as a percentage (β-value) ranging from 0 (no cytosine methylation) to 1 (complete cytosine methylation). Finally, to reduce influence of outliers in statistical models, normalized β-values were 90%-Winsorized.

Radon exposure

Potential residential radon exposure is available from the Health Protection Agency (HPA; now Public Health England) – British Geological Survey (BGS) 'radon potential dataset for Great Britain', and was obtained for the Avon area (which includes the original ALSPAC catchment area) from BGS after a data sharing agreement was agreed by BGS and the PI's Institute. Estimates of potential radon exposure were based on long-term radon measurements from 479,000 homes across Great Britain and provided with a spatial resolution of 75-metre buffers as the percentage of dwellings exceeding the 200 Bq m^{-3} Radon Action Level for 75-metre buffers in 6 classes: 1 (0-1%), 2 (1-3%), 3 (>3-5%), 4 (>5-10%), 5 (>10-30%) and 6 (>30-100%). More information is available at: http://www.bgs.ac.uk/radon/hpa-bgs.html. To assess measurement error, we also linked the ARIES dataset to estimates from the freely available radon 'indicative atlas'[27], which is based on the 'potential dataset', but provides the estimates in 1-km-side squares.

Residential histories of mothers and children were geocoded to postcode centroid level, and were linked to average potential radon exposure using ArcGIS software (version 10.6)[28] within the ALSPAC Data Safe Haven. This resulted in at least one address match for 986 mothers and 1001 young people (including two sets of twins). Once each residential address had a radon potential exposure class assigned, time spent at each address was calculated. This was merged with ARIES sample prevision dates, allowing time-weighted average potential radon exposures to be calculated up to the 'mothers at middle age', 'children at 7' and 'children at 15/17' sample extraction time points. For the cord and antenatal sample extractions, radon exposure potential of address at date of birth or closest address (temporally) to sample time point were assigned respectively.

These data were then linked to ALSPAC self-reported data selected to test for potential confounding (described below). After the linked exposure data were processed to minimise the risk of participant disclosure, the linked methylation-radon data were used for statistical analyses.

Statistical methods

For these analyses we only use participants with complete data. In the primary analyses average potential radon exposure was analysed as a continuous variable (range 1–6) to assess linear exposure-response associations. In addition, we also analysed associations based on binary exposure classifications (≤5% vs >5%).

Associations were tested using linear models using the limma R package (version 3.32.10)[29]. Associations were tested in (1) univariate analyses but with adjustment for the surrogate variables[30] to handle batch effects, sex differences, cell count heterogeneity and possible unknown confounders[31], and (2) additionally with adjustment for potential confounding factors maternal age at birth, maternal BMI, smoking during pregnancy, partner smoking during pregnancy, *AHRR* CpG site that detects own smoking nearly as accurately as self-report[32], mother alcohol intake in early pregnancy, equivalized income, parental occupation, and parental education, and (3) all factors of models 1 and 2 and additionally, for damp problems, central heating, boiler location, gas cooking, time windows open in the summer/winter day/night, and heavy traffic.

Associations at false discovery rate (FDR) less than 20% calculated using the *q* method[33] are reported. Where associations between MRB and methylation was positive, these sites were defined as "hypermethylated" and conversely when for inverse associations, these were defined as "hypomethylated".

Results

Participants and location

The results were based on 786 to 980 participants with complete information, depending on the time analyses. A graphical overview of the geographical study area and the distribution of potential radon exposure classes, as well as the distribution of addresses in each class, is shown graphically in Figure 1, and indicates that 79% of addresses are in areas with low probability (class 1 and 2) of exposure >200 Bq m^{-3}.

CpG sites

Results for CpG sites with FDR <0.20 are shown in Table 1. In mothers, average potential exposure to radon was only associated in an exposure-dependent manner with hypomethylation of cg25422346 during pregnancy ($p = 1.1x10^{-8}$, FDR = 0.005), with no associations observed at middle age. For the children, radon potential exposure was associated in an exposure-dependent manner with hypomethylation at cg16451995 at birth ($p = 3.2x10^{-7}$, FDR = 0.16) and with hypermethylation of cg01864468 at age 7 ($p = 1.1x10^{-8}$, FDR = 0.005). In adolescence (age 15–17) there was evidence of exposure-dependent methylation at several CpG sites. Cg04912984, cg16105117 and cg23988964 were hypermethylated with increased potential exposure, while cg04945076, cg08601898, cg16260355 and cg26056703 were hypomethylation proportionally to average potential radon exposure. The same CpG sites at the same time-points were identified when average potential exposure to >200 Bq m^{-3} was dichotomized into low (≤5%) and high (>5%) probability (Table 2), and similarly when using another cut-off (≤3% vs >3%); data not shown.

Regardless of exposure metric, there is little evidence of significant confounding with directions and sizes of associations similar for univariable and both multivariable models.

To assess the impact of measurement error, the same analyses were repeated but with exposure based on the 'indicative radon atlas' using 1-km^2 spatial resolution (Table 3). Results were comparable to those based on the 75-m buffers.

Discussion and conclusions

In this exploratory study we aimed to investigate associations between residential exposure to radon in the general population and DNA methylation. Associations were observed with increasing probability of average potential exposure of the residence over 200 Bq m^{-3} in children at birth, age 7 and during adolescence, with single CpG sites affected at birth (cg16451995) and age 7 (cg01864468) and seven sites affected at age 15–17 (cg04912984, cg04945076, cg08601898, cg16105117, cg16260355, cg23988964, cg26056703) after adjustment for important confounding factors. These also did not depend on the choice of cut-off used. For mothers an association with hypomethylation of cg25422386 was observed during pregnancy, but not at a later time point during middle age. To our knowledge, this is the first study identifying associations between radon exposure with methylation patterns in a general population.

Locations of the affected CpG sites of the children were on the *PMM2* gene (cordblood), associated with abnormalities in amniotic fluid and congenital disorders, upstream of *HCG14* (age 7), involved in the development of lung carcinoma, and *NDRG2* and *SGPL1*, associated with glioblastoma development and Alzheimer's disease and nephrotic syndrome, respectively, *LINC01197,* as well as upstream of *VANGL1*, associated with congenital disorders, and *FAM71A* genes (age 15–17), while for mothers cg25422346 is located upstream of *SMIM31*. These methylation patterns, describing both hyper- and hypomethylation associated with potential residential radon exposure, have not been reported elsewhere. In a candidate gene study of Chinese uranium miners, the authors reported increased methylation of promotor regions of $p16^{INK4a}$ and O^6-*MGMT* genes, as well as increased total methylation rate, depending on cumulative radon doses[20], and a study using BEAS-"B human lung cells exposed to 20,000 Bq m^{-3} radon for 30 minutes showing global hypomethylation and hypermethylation of candidate CpG-sites at *PTPRM* and *EDA2R* genes[34]. Similarly, these genes have also not come up in candidate gene studies of exposure to radon, in which gene-environment interactions with *p53*[35], *GSTM1* and *GSTT1*[36], *hOGG1* and *APE1*[37], *ADPRT*[38], *XPG*, *ADPRT* and *NBS1*[16], *LIG4*[39], and *NBS1* and *ATM1* have been reported. Possible explanations for the different genes for which hyper- or hypomethylation was associated with potential radon exposure in this study compared to other studies, may be that CpG sites identified in this study are involved in the 'bystander effect', rather than the result of direct irradiation, they may be a marker of earlier biological effects, it may be because methylation was measured in blood rather than in lung tissue, and of course residual confounding or chance findings can also not be completely excluded.

This study has several limitations. Most importantly, the exposure metric used in this study is a relatively weak one. It is not generally possible to accurately predict indoor radon concentrations for specific buildings without individual measurements[3]. Although people spend most of their time indoors at home, estimates are based on the modelled probability that a dwelling in the 75-m buffer that includes a person's home has a radon concentration exceeding 200 Bq m^{-3}. Because of high spatial and temporal variability[40,41] this will inevitably have led to considerable misclassification. Assuming measurement error in this case is non-differential, generally resulting in bias to the null, it is interesting that exposure-response associations were still observed in this study with a relatively small sample size. Furthermore, the possibility of misclassification of radon exposure should affect all participants in a similar way and is unlikely to bias associations with DNA methylation.

Although there was little evidence of significant confounding in these analyses, residual confounding explaining these findings cannot be excluded. For example, rurality is a known

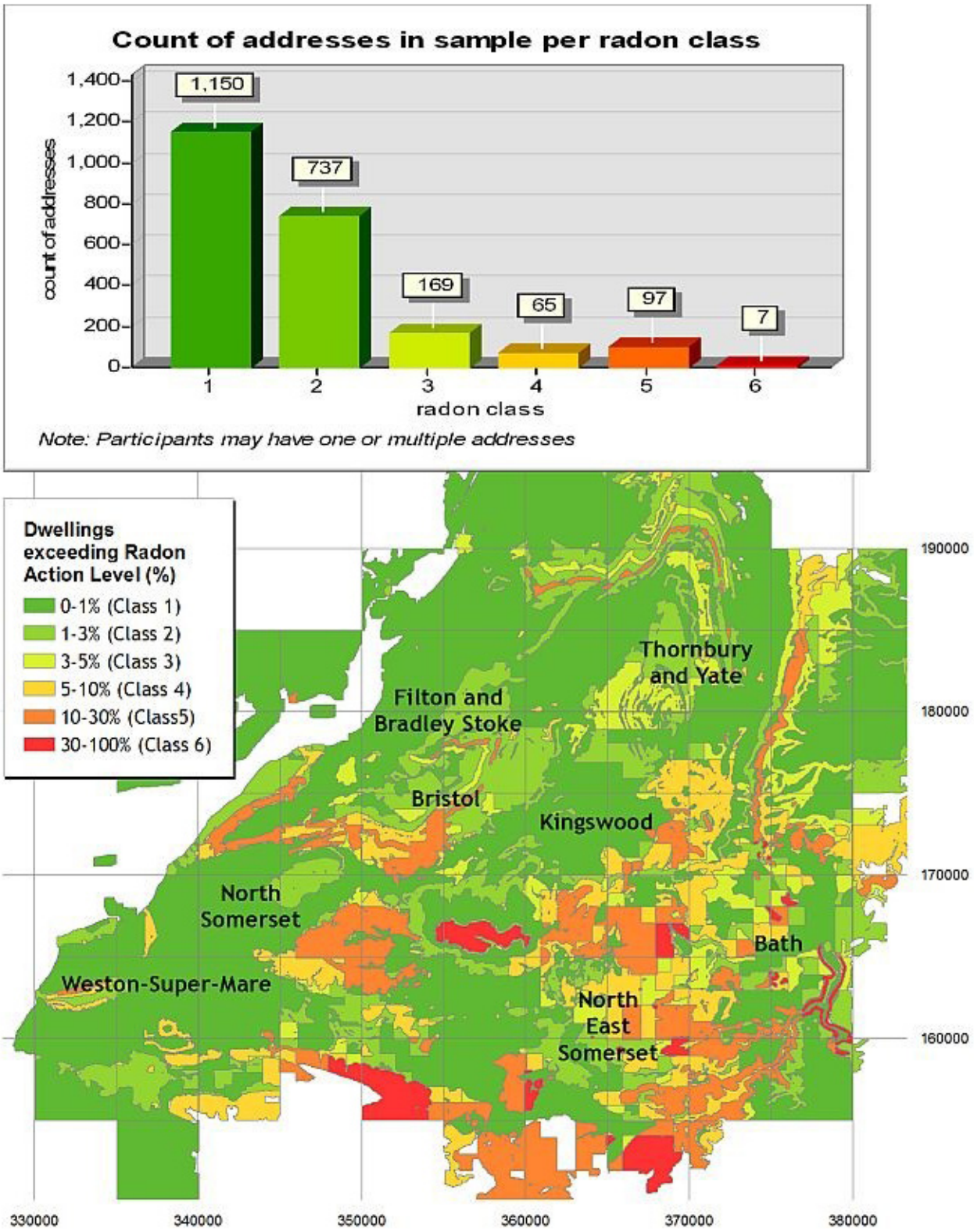

Figure 1. Geographical distribution of potential radon exposure classes, and number of addresses per class. Based upon the 'Radon Potential Dataset', reproduced with the permission of the British Geological Survey.

Table 1. Potential exposure to >200 Bq m^{-3} radon and methylation at specific CpG sites.

CpG	Chromosome	Gene	N (Model 1)	Model 1[1] Beta (se)	N (model 2)	Model 2[2] Beta (se)	N (model 3)	Model 3[3] Beta (se)	Min. FDR (adjusted)
Mothers during pregnancy									
cg25422346	4	upstream SMIM31	980	-0.002 (0.001)	712	-0.002 (0.001)	630	-0.003 (0.001)	0.005
Cordblood									
cg16451995	16	PMM2	912	-0.005 (0.001)	649	-0.005 (0.001)	576	-0.004 (0.001)	0.123
Age 7									
cg01864468	6	upstream HCG14	978	0.006 (0.002)	697	0.004 (0.002)	615	0.002 (0.002)	0.173
Age 15[4]									
cg04912984	1	upstream VANGL1	979	0.002 (0.001)	313	0.003 (0.001)	313	0.004 (0.001)	0.184
cg04945076	12		979	-0.007 (0.002)	313	-0.008 (0.003)	313	-0.008 (0.004)	0.184
cg08601898	8		979	-0.002 (0.001)	313	-0.004 (0.002)	313	-0.004 (0.002)	0.184
cg16105117	14	NDRG2	979	0.000 (0.000)	313	0.000 (0.000)	313	0.000 (0.000)	0.185
cg16260355	10	SGPL1	979	-0.001 (0.001)	313	0.000 (0.001)	239	-0.000 (0.001)	0.093
cg23988964	1	upstream FAM71A	979	0.004 (0.001)	313	0.001 (0.002)	239	0.002 (0.003)	0.185
cg26056703	15	LINC01197	979	-0.001 (0.001)	313	-0.000 (0.001)	239	-0.000 (0.001)	0.184
Mothers at middle age									
None									

[1]Adjusted for surrogate variable only. [2]Adjusted for surrogate variable, maternal age at birth, maternal BMI, smoking during pregnancy, partner smoking during pregnancy, mother alcohol intake in early pregnancy, equivalized income, parental occupation, parental education. [3]Adjusted for surrogate variable, maternal age at birth, maternal BMI, smoking during pregnancy, partner smoking during pregnancy, mother alcohol intake in early pregnancy, equivalized income, parental occupation, parental education, damp problems, central heating, boiler location, gas cooking, time windows open in the summer/winter day/night, heavy traffic. [4]At age 15/17, also adjusted for AHRR CpG site that detects own smoking.

confounding factor for studies on radon[41]. However, within ALSPAC and certainly within Avon there are few true 'rural' residential areas as the area quite heavily populated, so it is unlikely this will bias associations significantly. We also had no information on whether participants lived in houses or apartments (and in the latter case on which floor)[40] or whether houses had a basement[12], which will have added to further measurement error.

There are known limitations in quality of the ALSPAC residential address history data in terms of missingness and gaps; although in this study the impact of this will be limited as the postnatal ARIES sample dates are linked to direct contact with participants where address details would have been validated. However, to enable assignment of potential radon exposure to individuals over periods of unknown residence, remediation was carried out by (a) setting the address start date to child date of birth where first address start date fell after child date of birth, which is a reasonable assumption because often the address start date represents a data capture date as opposed to an actual move date, and (b) by rectifying all other temporal gaps by calculating a mean radon potential exposure class based on the radon potential at the preceding and succeeding addresses.

The current analysis lack directly measured blood-cell-type proportions, and we therefore included cell count heterogeneity using estimates obtained using surrogate variable analysis in the models[30]. This approach has been found to perform just as well or better[31] than the more commonly used method of Houseman *et al.*[42]. In this case, it probably performs better in DNA methylation profiles generated from childhood peripheral blood because DNA methylation references are available only for adult blood[43] and cord blood. Levels of methylation vary between tissue types and may relate differently to traits and exposures, which may limit inferences from this study. In the current study

Table 2. CpG site methylation and low (≤5%) vs high (>5%) probability of potential exposure to >200 Bq m^{-3} radon (FDR<0.20).

CpG	Chromosome	Gene	N (Model 1)	Model 1[1] Beta (se)	N (model 2)	Model 2[2] Beta (se)	N (model 3)	Model 3[3] Beta (se)	Min. FDR (adjusted)
Mothers during pregnancy									
cg25422346	4	upstream SMIM31	980	-0.012 (0.002)	712	-0.010 (0.002)	630	-0.011 (0.003)	0.005
Cordblood									
cg16451995	16	PMM2	912	-0.020 (0.004)	649	-0.018 (0.004)	576	-0.016 (0.005)	0.123
Age 7									
cg01864468	6	upstream HCG14	970	0.017 (0.006)	692	0.016 (0.008)	611	0.012 (0.009)	0.173
Age 15[4]									
cg04912984	1	upstream VANGL1	965	0.013 (0.003)	310	0.014 (0.005)	236	0.020 (0.005)	0.184
cg04945076	12		965	-0.037 (0.007)	310	-0.032 (0.012)	236	-0.029 (0.014)	0.184
cg08601898	8		965	-0.018 (0.004)	310	-0.031 (0.006)	236	-0.031 (0.007)	0.184
cg16105117	14	NDRG2	965	0.003 (0.001)	310	0.002 (0.001)	236	0.002 (0.001)	0.185
cg16260355	10	SGPL1	965	-0.007 (0.002)	310	-0.008 (0.004)	236	-0.009 (0.005)	0.093
cg23988964	1	upstream FAM71A	965	0.026 (0.006)	310	0.010 (0.008)	236	0.010 (0.009)	0.185
cg26056703	15	LINC01197	965	-0.014 (0.003)	310	-0.009 (0.005)	236	-0.014 (0.005)	0.184
Mothers at middle age									
None									

*Methylation for population with average potential radon exposure 5% or lower compared with population with probability >5%. [1]adjusted for surrogate variable only. [2]adjusted for surrogate variable, maternal age at birth, maternal BMI, smoking during pregnancy, partner smoking during pregnancy, mother alcohol intake in early pregnancy, equivalized income, parental occupation, parental education. [3]Adjusted for surrogate variable, maternal age at birth, maternal BMI, smoking during pregnancy, partner smoking during pregnancy, mother alcohol intake in early pregnancy, equivalized income, parental occupation, parental education, damp problems, central heating, boiler location, gas cooking, time windows open in the summer/winter day/night, heavy traffic. [4]At age 15/17 also adjusted for AHRR CpG site that detects own smoking.

we have methylation from blood samples, but it may have been beneficial had we been able to test associations in a more relevant cell type such as the lung.

And finally, this study had reduced statistical power due to the relatively limited sample size of the ARIES sub-sample, which was further diminished as a result of missing values. Alcohol consumption for adolescents could not be included as a potential confounding variable because this was only available for less than 200 teens. Because current approaches for multiple imputation are not feasible for genomic datasets including hundreds of thousands of measured variables (our study included variables corresponding to DNA methylation levels at over 480,000 CpG sites) we did not apply multiple imputation to increase sample size. In future, when feasible approaches have been developed, we plan to revisit these analyses.

The main strength of this study is the unique resource which allowed for the assessment of genome-wide methylation profiles at different time points linked to detailed phenotypic characterisation, which enabled assessment of the temporality of associations. In these analyses we used three cross-sectional models to compare methylation patterns at birth, age 7 and in adolescence, but with better characterization of the dynamic elements of the human methylome[44], longitudinal analyses will help to better elucidate persistent and reversible effects of (environmental) exposures as well as critical periods of effect[45]. Information on epigenetic signals across the life-course and radon exposure are of interest because they have the potential to describe early biological effects, and the estimated induction (lag) period of lung cancer to radon exposure is between 5 and 25 years[1].

In conclusion, this exploratory study is, to our knowledge, the first study to study association between genome-wide DNA methylation and (potential) residential exposure to radon. Despite the relatively weak exposure metric, differential methylation associated with increased potential residential radon exposure

Table 3. Indicative (1 km² spatial granularity) potential exposure to >200 Bq m⁻³ radon and CpG site methylation.

CpG	Chromosome	Gene	N (Model 1)	Model 1[1] Beta (se)	N (model 2)	Model 2[2] Beta (se)	N (model 3)	Model 3[3] Beta (se)	Min. FDR (adjusted)
Mothers during pregnancy									
cg25422346	4	upstream SMIM31	857	-0.001 (0.001)	615	-0.001 (0.001)	543	-0.002 (0.001)	0.005
Cordblood									
cg16451995	16	PMM2	786	-0.004 (0.001)	551	-0.004 (0.001)	485	-0.003 (0.001)	0.123
Age 7									
cg01864468	6	upstream HCG14	851	0.009 (0.002)	596	0.007 (0.002)	524	0.006 (0.002)	0.173
Age 15[4]									
cg04912984	1	upstream VANGL1	852	0.001 (0.001)	272	0.002 (0.001)	208	0.004 (0.002)	0.184
cg04945076	12			-0.002 (0.002)	272	-0.010 (0.004)	208	-0.008 (0.005)	0.184
cg08601898	8		852	-0.002 (0.001)	272	-0.002 (0.002)	208	-0.004 (0.002)	0.184
cg16105117	14	NDRG2	852	0.000 (0.000)	272	0.000 (0.000)	208	0.000 (0.000)	0.185
cg16260355	10	SGPL1	852	-0.002 (0.001)	272	-0.000 (0.001)	208	-0.001 (0.002)	0.093
cg23988964	1	upstream FAM71A	852	0.001 (0.002)	272	0.002 (0.003)	208	0.001 (0.003)	0.186
cg26056703	15	LINC01197	852	-0.001 (0.001)	272	-0.000 (0.002)	208	-0.002 (0.002)	0.184
Mothers at middle age									
None									

[1]Adjusted for surrogate variable only. [2]Adjusted for surrogate variable, maternal age at birth, maternal BMI, smoking during pregnancy, partner smoking during pregnancy, mother alcohol intake in early pregnancy, equivalized income, parental occupation, parental education. [3]Adjusted for surrogate variable, maternal age at birth, maternal BMI, smoking during pregnancy, partner smoking during pregnancy, mother alcohol intake in early pregnancy, equivalized income, parental occupation, parental education, damp problems, central heating, boiler location, gas cooking, time windows open in the summer/winter day/night, heavy traffic. [4]At age 15/17 also adjusted for AHRR CpG site that detects own smoking.

was observed prenatally in mothers, for children at birth, age 7, and especially at age 15–17, but not for the mothers in middle age. Future work in a larger population, with replication in an independent sample, and using a radon exposure estimation methodology, most notably personal exposure measures, can further elucidate these associations.

Data availability

The potential residential radon exposure was provided by the British Geological Survey (BGS) under license for the current study (Licence number 2017/017RAD ED British Geological Survey © NERC. All rights reserved) and can be requested from BGS (http://www.bgs.ac.uk/radon/hpa-bgs.html). Details on who will be granted access to the data, and whether there will be a charge for data access can be found online.

ALSPAC data access is through a system of managed open access. Full details of all available data can be accessed through a fully searchable data dictionary provided on the ALSPAC study website (http://www.bris.ac.uk/alspac/researchers/data-access/data-dictionary) and the steps below highlight how to apply for access to both the data included in this data note and all other ALSPAC data. The datasets presented in this data note are linked to ALSPAC project number B645; please quote this project number during your application. The ALSPAC variable codes highlighted in the dataset descriptions can be used to specify required variables.

1. Please read the ALSPAC access policy (PDF, 627kB) which describes the process of accessing the data and samples in detail and outlines the costs associated with doing so.
2. You may also find it useful to browse our fully searchable research proposals database, which lists all research projects that have been approved since April 2011.

3. Please submit your research proposal for consideration by the ALSPAC Executive Committee using the online process. You will receive a response within 10 working days to advise you whether your proposal has been approved.

If you have any questions about accessing data, please email alspac-data@bristol.ac.uk.

The ALSPAC data management plan describes in detail the policy regarding data sharing, which is through a system of managed open access.

Reporting guidelines

STROBE Guidelines for cohort studies have been used for this publication. DOI: https://doi.org/10.17605/OSF.IO/KGCHQ[46].

Grant information

The Accessible Resource for Integrated Epigenomics Studies (ARIES) was funded by the BBSRC (BBI025751/1 and BB/I025263/1). Supplementary funding to generate DNA methylation data which are (or will be) included in ARIES has been obtained from the MRC, ESRC, NIH and other sources. ARIES is maintained under the auspices of the MRC Integrative Epidemiology Unit at the University of Bristol (MC_UU_12013/2, MC_UU_12013/8 and MC_UU_12013/9). This work was also supported by the BBSRC and ESRC (grant number ES/N000498/1), the UK National Institute for Health Research (NIHR) Health Protection Research Unit (HPRU) in Evaluation, NIHR Biomedical Research Centre (BRC) and NIHR School for Public Health Research (SPHR). The UK Medical Research Council and Wellcome Trust (102215/13/2) and the University of Bristol provide core support for ALSPAC. This publication is the work of the authors and FdV will serve as guarantors for the contents of this paper. RT and AB were supported by UK Natural Environment Research Council (R8/H12/83/NE/P01830/1) and the Wellcome Trust (WT086118/Z/08/Z). A comprehensive list of grants funding is available on the ALSPAC website (http://www.bristol.ac.uk/alspac/external/documents/grant-acknowledgements.pdf); No external funding was obtained for this study.

Acknowledgements

We are extremely grateful to all the families who took part in this study, the midwives for their help in recruiting them, and the whole ALSPAC team, which includes interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, receptionists and nurses. The UK Medical Research Council and the Wellcome Trust (102215/2/13/2) and the University of Bristol provide core support for ALSPAC.

References

1. BEIR VI : **Health effects of exposure to radon.** Washington DC, UK: Committee on Health Risks of Exposure to Radon (BEIR VI); Board on Radiation Effects Research; Commission on Life Sciences; National Research Council, 1999.
Reference Source
2. HPA: **Limitation of Human Exposure to Radon. Advise from the Health Protection Agency.** Chilton, Didcot,UK: Health Protection Agency, Radiation CaEH; Report No.: RCE-15. 2010.
Reference Source
3. McColl N, Auvinen A, Kesminiene A, *et al.*: **European Code against Cancer 4th Edition: Ionising and non-ionising radiation and cancer.** *Cancer Epidemiol.* 2015; **39 Suppl 1**: S93–100.
PubMed Abstract | Publisher Full Text
4. IARC: **Radiation. Volume 100 D. A Review of Human Carcinogens.** Lyon: IARC, 2012.
Reference Source
5. Barros-Dios JM, Ruano-Ravina A, Pérez-Ríos M, *et al.*: **Residential radon exposure, histologic types, and lung cancer risk. A case-control study in Galicia, Spain.** *Cancer Epidemiol Biomarkers Prev.* 2012; **21**(6): 951–8.
PubMed Abstract | Publisher Full Text
6. Marsh JW, Laurier D, Tirmarche M: **Radon Dosimetry for Workers: Icrp's Approach.** *Radiat Prot Dosimetry.* 2017; **177**(4): 466–74.
PubMed Abstract | Publisher Full Text
7. WHO: **WHO Handbook on Indoor Radon: A Public Health Perspective.** World Health Organization. 2009.
Reference Source
8. Kendall GM, Fell TP: **Doses to the red bone marrow of young people and adults from radiation of natural origin.** *J Radiol Prot.* 2011; **31**(3): 329–35.
PubMed Abstract | Publisher Full Text
9. UNSCEAR: **Sources and Effects of Radiation.** UNSCEAR 2000 Report to the General Assembly, with Scientific Annexes. New York: United Nations, 2000.
Reference Source
10. Miles JC: **Temporal variation of radon levels in houses and implications for radon measurement strategies.** *Radiat Prot Dosimetry.* 2001; **93**(4): 369–76.
PubMed Abstract | Publisher Full Text
11. Kendall GM, Miles JC, Rees D, *et al.*: **Variation with socioeconomic status of indoor radon levels in Great Britain: The less affluent have less radon.** *J Environ Radioact.* 2016; **164**: 84–90.
PubMed Abstract | Publisher Full Text
12. Field RW, Krewski D, Lubin JH, *et al.*: **An overview of the North American residential radon and lung cancer case-control studies.** *J Toxicol Environ Health A.* 2006; **69**(7): 599–631.
PubMed Abstract | Publisher Full Text
13. Lecomte JF, Solomon S, Takala J, *et al.*: **ICRP Publication 126: Radiological Protection against Radon Exposure.** *Ann ICRP.* 2014; **43**(3): 5–73.
PubMed Abstract | Publisher Full Text
14. EU: **COUNCIL DIRECTIVE 2013/59/EURATOM.** European Union, Contract No.: L 13/1. 2014.
Reference Source
15. **Summary of ICRP Recommendations on Radon: Hearing before the INTERNATIONAL COMMISSION ON RADIOLOGICAL PROTECTION.** 2018.
Reference Source
16. Sinitsky MY, Larionov AV, Asanov MA, *et al.*: **Associations of DNA-repair gene polymorphisms with a genetic susceptibility to ionizing radiation in residents of areas with high radon (^{222}Rn) concentration.** *Int J Radiat Biol.* 2015; **91**(6): 486–94.
PubMed Abstract | Publisher Full Text
17. Larionov AV, Sinitsky MY, Druzhinin VG, *et al.*: **DNA excision repair and double-strand break repair gene polymorphisms and the level of chromosome aberration in children with long-term exposure to radon.** *Int J Radiat Biol.* 2016; **92**(8): 466–74.
PubMed Abstract | Publisher Full Text
18. Groom A, Elliott HR, Embleton ND, *et al.*: **Epigenetics and child health: basic principles.** *Arch Dis Child.* 2011; **96**(9): 863–9.
PubMed Abstract | Publisher Full Text
19. de Vocht F, Simpkin AJ, Richmond RC, *et al.*: **Assessment of Offspring DNA Methylation across the Lifecourse Associated with Prenatal Maternal Smoking Using Bayesian Mixture Modelling.** *Int J Environ Res Public Health.* 2015; **12**(11): 14461–76.
PubMed Abstract | Publisher Full Text | Free Full Text
20. Su S, Jin Y, Zhang W, *et al.*: **Aberrant promoter methylation of *p16*^{INK4a} and *O*^{6}-**

methylguanine-DNA methyltransferase genes in workers at a Chinese uranium mine. *J Occup Health.* 2006; **48**(4): 261–6.
PubMed Abstract | Publisher Full Text

21. Fraser A, Macdonald-Wallis C, Tilling K, *et al.*: **Cohort Profile: the Avon Longitudinal Study of Parents and Children: ALSPAC mothers cohort.** *Int J Epidemiol.* 2013; **42**(1): 97–110.
PubMed Abstract | Publisher Full Text | Free Full Text

22. Boyd A, Golding J, Macleod J, *et al.*: **Cohort Profile: the 'children of the 90s'--the index offspring of the Avon Longitudinal Study of Parents and Children.** *Int J Epidemiol.* 2013; **42**(1): 111–27.
PubMed Abstract | Publisher Full Text | Free Full Text

23. Dedeurwaerder S, Defrance M, Calonne E, *et al.*: **Evaluation of the Infinium Methylation 450K technology.** *Epigenomics.* 2011; **3**(6): 771–84.
PubMed Abstract | Publisher Full Text

24. Relton CL, Gaunt T, McArdle W, *et al.*: **Data Resource Profile: Accessible Resource for Integrated Epigenomic Studies (ARIES).** *Int J Epidemiol.* 2015; **44**(4): 1181–90.
PubMed Abstract | Publisher Full Text | Free Full Text

25. Relton CL, Gaunt T, McArdle W, *et al.*: **Data Resource Profile: Accessible Resource for Integrated Epigenomic Studies (ARIES).** *Int J Epidemiol.* 2015; **44**(4): 1181–90.
PubMed Abstract | Publisher Full Text | Free Full Text

26. Min JL, Hemani G, Davey Smith G, *et al.*: **Meffil: efficient normalisation and analysis of very large DNA methylation samples.** *bioRxiv.* 2017; 125963.
Publisher Full Text

27. Miles JCH, Appleton JD, Rees DM, *et al.*: **Indicative Atlas of Radon in England and Wales.** Chilton, Didcot, UK: Health Protection Agency, Centre for Radiation CaEH. Report No.: HPA-RPD-033. 2007.
Reference Source

28. ESRI: **ArcGIS Desktop: Release 10.** Redlands, CA: Environmental Systems Research Institute. 2011.
Reference Source

29. Smyth GK: **Linear models and empirical bayes methods for assessing differential expression in microarray experiments.** *Stat Appl Genet Mol Biol.* 2004; **3**: Article 3.
PubMed Abstract | Publisher Full Text

30. Leek JT, Storey JD: **Capturing heterogeneity in gene expression studies by surrogate variable analysis.** *PLoS Genet.* 2007; **3**(9): 1724–35.
PubMed Abstract | Publisher Full Text | Free Full Text

31. McGregor K, Bernatsky S, Colmegna I, *et al.*: **An evaluation of methods correcting for cell-type heterogeneity in DNA methylation studies.** *Genome Biol.* 2016; **17**: 84.
PubMed Abstract | Publisher Full Text | Free Full Text

32. Anderson AM, Philibert RA, Gibbons FX, *et al.*: **Accuracy and utility of an epigenetic biomarker for smoking in populations with varying rates of false self-report.** *Am J Med Genet B Neuropsychiatr Genet.* 2017; **174**(6): 641–50.
PubMed Abstract | Publisher Full Text | Free Full Text

33. Storey JD, Tibshirani R: **Statistical significance for genomewide studies.** *Proc Natl Acad Sci U S A.* 2003; **100**(16): 9440–5.
PubMed Abstract | Publisher Full Text | Free Full Text

34. Huang H, Ji Y, Zhang J, *et al.*: **Aberrant DNA methylation in radon and/or cigarette smoke-induced malignant transformation in BEAS-2B human lung cell line.** *J Toxicol Environ Health A.* 2017; **80**(23–24): 1321–30.
PubMed Abstract | Publisher Full Text

35. Taylor JA, Watson MA, Devereux TR, *et al.*: **p53 mutation hotspot in radon-associated lung cancer.** *Lancet.* 1994; **343**(8889): 86–7.
PubMed Abstract | Publisher Full Text

36. Ruano-Ravina A, Pereyra MF, Castro MT, *et al.*: **Genetic susceptibility, residential radon, and lung cancer in a radon prone area.** *J Thorac Oncol.* 2014; **9**(8): 1073–80.
PubMed Abstract | Publisher Full Text

37. Druzhinin VG, Volkov AN, Glushkov AN, *et al.*: **[Role of repair gene polymorphism in estimating the sensitivity of human genome to radon in concentrations exceeding maximum permissible level].** *Gig Sanit.* 2011; (5): 26–30.
PubMed Abstract

38. Minina VI, Druzhinin VG, Lunina AA, *et al.*: **Association of DNA repair gene polymorphism with chromosomal aberrations frequency in human lymphocytes.** *Russ J Genet Appl Res.* 2012; **2**(2): 171–6.
Publisher Full Text

39. Volkov AN, Druzhinin VG, Minina VI, *et al.*: **Research of the influence of the *LIG4* gene polymorphism on the chromosomal aberration level in human lymphocytes, with background and excessive exposure to a radon.** *Russ J Genet Appl Res.* 2014; **4**(5): 416–20.
Publisher Full Text

40. Denman AR, Groves-Kirkby NP, Groves-Kirkby CJ, *et al.*: **Health implications of radon distribution in living rooms and bedrooms in U.K. dwellings--a case study in Northamptonshire.** *Environ Int.* 2007; **33**(8): 999–1011.
PubMed Abstract | Publisher Full Text

41. Zahnd WE, Mueller-Luckey GS, Ratnapradipa K, *et al.*: **Predictors and Spatial Variation of Radon Testing in Illinois, 2005-2012.** *J Public Health Manag Pract.* 2018; **24**(2): e1–e9.
PubMed Abstract | Publisher Full Text

42. Houseman EA, Accomando WP, Koestler DC, *et al.*: **DNA methylation arrays as surrogate measures of cell mixture distribution.** *BMC Bioinformatics.* 2012; **13**(1): 86.
PubMed Abstract | Publisher Full Text | Free Full Text

43. Reinius LE, Acevedo N, Joerink M, *et al.*: **Differential DNA methylation in purified human blood cells: implications for cell lineage and studies on disease susceptibility.** *PLoS One.* 2012; **7**(7): e41361.
PubMed Abstract | Publisher Full Text | Free Full Text

44. Ziller MJ, Gu H, Müller F, *et al.*: **Charting a dynamic DNA methylation landscape of the human genome.** *Nature.* 2013; **500**(7463): 477–81.
PubMed Abstract | Publisher Full Text | Free Full Text

45. Mishra G, Nitsch D, Black S, *et al.*: **A structured approach to modelling the effects of binary exposure variables over the life course.** *Int J Epidemiol.* 2009; **38**(2): 528–37.
PubMed Abstract | Publisher Full Text | Free Full Text

46. De Vocht F: **Radon ALSPAC.** 2018.
http://www.doi.org/10.17605/OSF.IO/KGCHQ

Open Peer Review

Current Referee Status: ✔

Version 1

Referee Report 20 February 2019

https://doi.org/10.21956/wellcomeopenres.16353.r34801

✔ **Beata Peplonska**
Department of Environmental Epidemiology , Nofer Institute of Occupational Medicine, Lodz, Poland

The paper entitled "*Residential exposure to radon and DNA methylation across the lifecourse: an exploratory study in the ALSPAC birth cohort"* by Frank de Vocht et al. is a novel and stimulating contribution to studies on the effects of residential radon exposure. The study finds some significant associations between potential exposure to radon >200 Bq m3 and methylation pattern of cytosines within several CpGs in children and one site in mothers. Both hypo- and hypermethylations have been detected.

At the current stage of knowledge, the biological meaning of the detected epigenetic changes is not known, however, the work performed is still of value and the inferences observed may be clarified in the future research.

The authors discuss their findings providing description of the pathologies related to the mutations of the genes where the specific hypo- or hypermethylation has been found (lacking reference). Nevertheless, the epigenetic effects observed were neither located at the promoter sites of these genes nor enhancers, thus may not affect the genes activity, which in my opinion shall be stated explicitly.

The limits of the study, including lack of precise data on study participants exposure have been identified and described by the authors.
The paper is very clearly written and reads very well.

I have only few minor comments:
Introduction:
"Residential radon exposure has been associated with DNA repair gene polymorphisms in adults[16] and children[17], with the latter study also reporting double-strand break repair gene polymorphisms.". Please, check in the cited articles what associations have been reported, and correct the statement accordingly.
Methods:
2nd paragraph reports at which ages blood was collected from study subjects (children and mothers). The information for children is unnecessary repeated in the subsection describing "DNA methylation".

Is the work clearly and accurately presented and does it cite the current literature?
Yes

Is the study design appropriate and is the work technically sound?

Yes

Are sufficient details of methods and analysis provided to allow replication by others?
Yes

If applicable, is the statistical analysis and its interpretation appropriate?
Yes

Are all the source data underlying the results available to ensure full reproducibility?
Yes

Are the conclusions drawn adequately supported by the results?
Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: epidemiology

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.