

ASSOCIATION STUDIES ARTICLE

# Occupational exposure to gases/fumes and mineral dust affect DNA methylation levels of genes regulating expression

Diana A. van der Plaat<sup>1,2,†</sup>, Judith M. Vonk<sup>1,2,‡,#</sup>, Natalie Terzikhan<sup>3,4,#</sup>, Kim de Jong<sup>1,2,¶</sup>, Maaike de Vries<sup>1,2,||</sup>, Sacha La Bastide-van Gemert<sup>1,††</sup>, Cleo C. van Diemen<sup>6,‡‡</sup>, Lies Lahousse<sup>3,4,5,§</sup>, Guy G. Brusselle<sup>3,4,5</sup>, Ivana Nedeljkovic<sup>4,§§</sup>, Najaf Amin<sup>4,¶¶</sup>, BIOS Consortium<sup>¶¶</sup>, Hans Kromhout<sup>7,|||</sup>, Roel C. H. Vermeulen<sup>7</sup>, Dirkje S. Postma<sup>8,2</sup>, Cornelia M. van Duijn<sup>4</sup> and H. Marika Boezen<sup>1,2,\*</sup>

<sup>†</sup>Diana A. van der Plaat, <http://orcid.org/0000-0002-0361-357X>

<sup>‡</sup>Judith M. Vonk, <https://orcid.org/0000-0001-7531-4547>

<sup>¶</sup>Kim de Jong, <https://orcid.org/0000-0002-3948-1046>

<sup>||</sup>Maaike de Vries, <https://orcid.org/0000-0001-7210-8174>

<sup>††</sup>Sacha La Bastide-van Gemert, <https://orcid.org/0000-0003-0666-6690>

<sup>‡‡</sup>Cleo C. van Diemen, <https://orcid.org/0000-0001-9283-2207>

<sup>§</sup>Lies Lahousse, <https://orcid.org/0000-0002-3494-4363>

<sup>§§</sup>Ivana Nedeljkovic, <http://orcid.org/0000-0002-0370-1473>

<sup>¶¶</sup>Najaf Amin, <https://orcid.org/0000-0002-8944-1771>

<sup>|||</sup>Hans Kromhout, <https://orcid.org/0000-0002-4233-1890>

<sup>#</sup>These authors should be regarded as joint second authors.

<sup>¶¶</sup>Group author: BIOS Consortium (Biobank-based Integrative Omics Study)

*Management Team:* Bastiaan T. Heijmans (chair)<sup>1</sup>, Peter A.C. 't Hoen<sup>2</sup>, Joyce van Meurs<sup>3</sup>, Aaron Isaacs<sup>4</sup>, Rick Jansen<sup>5</sup>, Lude Franke<sup>6</sup>.

*Cohort collection:* Dorret I. Boomsma<sup>7</sup>, René Pool<sup>7</sup>, Jenny van Dongen<sup>7</sup>, Jouke J. Hottenga<sup>7</sup> (Netherlands Twin Register); Marleen MJ van Greevenbroek<sup>8</sup>, Coen D.A. Stehouwer<sup>8</sup>, Carla J.H. van der Kallen<sup>8</sup>, Casper G. Schalkwijk<sup>8</sup> (Cohort study on Diabetes and Atherosclerosis Maastricht); Cisca Wijmenga<sup>6</sup>, Lude Franke<sup>6</sup>, Sasha Zernakova<sup>6</sup>, Etti F. Tigchelaar<sup>6</sup> (LifeLines Deep); P. Eline Slagboom<sup>1</sup>, Marian Beekman<sup>1</sup>, Joris Deelen<sup>1</sup>, Diana van Heemst<sup>9</sup> (Leiden Longevity Study); Jan H. Veldink<sup>10</sup>, Leonard H. van den Berg<sup>10</sup> (Prospective ALS Study Netherlands); Cornelia M. van Duijn<sup>4</sup>, Bert A. Hofman<sup>11</sup>, Aaron Isaacs<sup>4</sup>, André G. Uitterlinden<sup>3</sup> (Rotterdam Study). *Data Generation:* Joyce van Meurs (Chair)<sup>3</sup>, P. Mila Jhamai<sup>3</sup>, Michael Verbiest<sup>3</sup>, H. Eka D. Suchiman<sup>1</sup>, Marijn Verkerk<sup>3</sup>, Ruud van der Breggen<sup>1</sup>, Jeroen van Rooij<sup>3</sup>, Nico Lakenberg<sup>1</sup>. *Data management and computational infrastructure:* Hailiang Mei (Chair)<sup>12</sup>, Maarten van Iterson<sup>1</sup>, Michiel van Galen<sup>2</sup>, Jan Bot<sup>13</sup>, Dasha V. Zernakova<sup>6</sup>, Rick Jansen<sup>5</sup>, Peter van 't Hof<sup>12</sup>, Patrick Deelen<sup>6</sup>, Irene Nooren<sup>13</sup>, Peter A.C. 't Hoen<sup>2</sup>, Bastiaan T. Heijmans<sup>1</sup>, Matthijs Moed<sup>1</sup>.

*Data Analysis Group:* Lude Franke (Co-Chair)<sup>6</sup>, Martijn Vermaat<sup>2</sup>, Dasha V. Zernakova<sup>6</sup>, René Luijk<sup>1</sup>, Marc Jan Bonder<sup>6</sup>, Maarten van Iterson<sup>1</sup>, Patrick Deelen<sup>6</sup>, Freerk van Dijk<sup>14</sup>, Michiel van Galen<sup>2</sup>, Wibowo Arindrarto<sup>12</sup>, Szymon M. Kielbasa<sup>15</sup>, Morris A. Swertz<sup>14</sup>, Erik. W van Zwet<sup>15</sup>, Rick Jansen<sup>5</sup>, Peter-Bram 't Hoen (Co-Chair)<sup>2</sup>, Bastiaan T. Heijmans (Co-Chair)<sup>1</sup>.

*Management Team:* Bastiaan T. Heijmans (chair)<sup>1</sup>, Peter A.C. 't Hoen<sup>2</sup>, Joyce van Meurs<sup>3</sup>, Aaron Isaacs<sup>4</sup>, Rick Jansen<sup>5</sup>, Lude Franke<sup>6</sup>.

*Cohort collection:* Dorret I. Boomsma<sup>7</sup>, René Pool<sup>7</sup>, Jenny van Dongen<sup>7</sup>, Jouke J. Hottenga<sup>7</sup> (Netherlands Twin Register); Marleen MJ van Greevenbroek<sup>8</sup>, Coen D.A. Stehouwer<sup>8</sup>, Carla J.H. van der Kallen<sup>8</sup>, Casper G. Schalkwijk<sup>8</sup> (Cohort study on Diabetes and Atherosclerosis Maastricht); Cisca Wijmenga<sup>6</sup>, Lude Franke<sup>6</sup>, Sasha Zernakova<sup>6</sup>, Etti F. Tigchelaar<sup>6</sup> (LifeLines Deep); P. Eline Slagboom<sup>1</sup>, Marian Beekman<sup>1</sup>, Joris Deelen<sup>1</sup>, Diana van Heemst<sup>9</sup> (Leiden Longevity Study); Jan H. Veldink<sup>10</sup>, Leonard H. van den Berg<sup>10</sup> (Prospective ALS Study Netherlands); Cornelia M. van Duijn<sup>4</sup>, Bert A. Hofman<sup>11</sup>, Aaron Isaacs<sup>4</sup>, André G. Uitterlinden<sup>3</sup> (Rotterdam Study). *Data Generation:* Joyce van Meurs (Chair)<sup>3</sup>, P. Mila Jhamai<sup>3</sup>, Michael Verbiest<sup>3</sup>, H. Eka D. Suchiman<sup>1</sup>, Marijn Verkerk<sup>3</sup>, Ruud van der Breggen<sup>1</sup>, Jeroen van Rooij<sup>3</sup>, Nico Lakenberg<sup>1</sup>.

*Data management and computational infrastructure:* Hailiang Mei (Chair)<sup>12</sup>, Maarten van Iterson<sup>1</sup>, Michiel van Galen<sup>2</sup>, Jan Bot<sup>13</sup>, Dasha V. Zernakova<sup>6</sup>, Rick Jansen<sup>5</sup>, Peter van 't Hof<sup>12</sup>, Patrick Deelen<sup>6</sup>, Irene Nooren<sup>13</sup>, Peter A.C. 't Hoen<sup>2</sup>, Bastiaan T. Heijmans<sup>1</sup>, Matthijs Moed<sup>1</sup>.

*Data Analysis Group:* Lude Franke (Co-Chair)<sup>6</sup>, Martijn Vermaat<sup>2</sup>, Dasha V. Zernakova<sup>6</sup>, René Luijk<sup>1</sup>, Marc Jan Bonder<sup>6</sup>, Maarten van Iterson<sup>1</sup>, Patrick Deelen<sup>6</sup>, Freerk van Dijk<sup>14</sup>, Michiel van Galen<sup>2</sup>, Wibowo Arindrarto<sup>12</sup>, Szymon M. Kielbasa<sup>15</sup>, Morris A. Swertz<sup>14</sup>, Erik. W van Zwet<sup>15</sup>, Rick Jansen<sup>5</sup>, Peter-Bram 't Hoen (Co-Chair)<sup>2</sup>, Bastiaan T. Heijmans (Co-Chair)<sup>1</sup>.

Received: November 1, 2018. Revised: March 13, 2019. Accepted: March 23, 2019

© The Author(s) 2019. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

1. Molecular Epidemiology Section, Department of Medical Statistics and Bioinformatics, Leiden University Medical Center, Leiden, The Netherlands. 2. Department of Human Genetics, Leiden University Medical Center, Leiden, The Netherlands. 3. Department of Internal Medicine, ErasmusMC, Rotterdam, The Netherlands. 4. Department of Genetic Epidemiology, ErasmusMC, Rotterdam, The Netherlands. 5. Department of Psychiatry, VU University Medical Center, Neuroscience Campus Amsterdam, Amsterdam, The Netherlands. 6. Department of Genetics, University of Groningen, University Medical Centre Groningen, Groningen, The Netherlands. 7. Department of Biological Psychology, VU University Amsterdam, Neuroscience Campus Amsterdam, Amsterdam, The Netherlands. 8. Department of Internal Medicine and School for Cardiovascular Diseases (CARIM), Maastricht University Medical Center, Maastricht, The Netherlands. 9. Department of Gerontology and Geriatrics, Leiden University Medical Center, Leiden, The Netherlands. 10. Department of Neurology, Brain Center Rudolf Magnus, University Medical Center Utrecht, Utrecht, The Netherlands. 11. Department of Epidemiology, ErasmusMC, Rotterdam, The Netherlands. 12. Sequence Analysis Support Core, Leiden University Medical Center, Leiden, The Netherlands. 13. SURFsara, Amsterdam, the Netherlands. 14. Genomics Coordination Center, University Medical Center Groningen, University of Groningen, Groningen, the Netherlands. 15. Medical Statistics Section, Department of Medical Statistics and Bioinformatics, Leiden University Medical Center, Leiden, The Netherlands

<sup>1</sup>Department of Epidemiology and <sup>2</sup>Groningen Research Institute for Asthma and COPD (GRIAC), University Medical Center Groningen, University of Groningen, Groningen, The Netherlands, <sup>3</sup>Department of Respiratory Medicine, Ghent University Hospital, Ghent, Belgium, <sup>4</sup>Department of Epidemiology and <sup>5</sup>Department of Respiratory Medicine, Erasmus Medical Center, Rotterdam, the Netherlands, <sup>6</sup>Department of Genetics, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands, <sup>7</sup>Division Environmental Epidemiology, Institute for Risk Assessment Sciences, Utrecht University, Utrecht, The Netherlands and <sup>8</sup>Department of Pulmonary Diseases, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands

\*To whom correspondence should be addressed at: Department of Epidemiology, University Medical Center Groningen, University of Groningen, Hanzeplein 1, 9700 RB Groningen, The Netherlands. Tel: +33 503610739; Email: h.m.boezen@umcg.nl

## Abstract

Many workers are daily exposed to occupational agents like gases/fumes, mineral dust or biological dust, which could induce adverse health effects. Epigenetic mechanisms, such as DNA methylation, have been suggested to play a role. We therefore aimed to identify differentially methylated regions (DMRs) upon occupational exposures in never-smokers and investigated if these DMRs associated with gene expression levels. To determine the effects of occupational exposures independent of smoking, 903 never-smokers of the LifeLines cohort study were included. We performed three genome-wide methylation analyses (Illumina 450 K), one per occupational exposure being gases/fumes, mineral dust and biological dust, using robust linear regression adjusted for appropriate confounders. DMRs were identified using comb-p in Python. Results were validated in the Rotterdam Study (233 never-smokers) and methylation-expression associations were assessed using Biobank-based Integrative Omics Study data ( $n = 2802$ ). Of the total 21 significant DMRs, 14 DMRs were associated with gases/fumes and 7 with mineral dust. Three of these DMRs were associated with both exposures (*RPLP1* and *LINC02169* ( $2\times$ )) and 11 DMRs were located within transcript start sites of gene expression regulating genes. We replicated two DMRs with gases/fumes (*VTRNA2-1* and *GNAS*) and one with mineral dust (*CCDC144NL*). In addition, nine gases/fumes DMRs and six mineral dust DMRs significantly associated with gene expression levels. Our data suggest that occupational exposures may induce differential methylation of gene expression regulating genes and thereby may induce adverse health effects. Given the millions of workers that are exposed daily to occupational exposures, further studies on this epigenetic mechanism and health outcomes are warranted.

## Introduction

Daily, millions of workers worldwide are exposed to chemical agents, fumes and (in)organic dusts (1). The leading occupational causes of death in 2000 were unintentional injuries (41%), chronic obstructive pulmonary disease (COPD, 40%) and lung cancer (13%) (1). This is not remarkable, since the skin and the lungs are most directly exposed to occupational pollutants, which could be prevented by implementing protective measures. Studies focusing on specific occupations, like pig farmers, miners, construction and textile workers, found associations between job-specific exposures and a faster annual decline in lung function ( $FEV_1$ ) (2–4). In addition, we have previously shown that exposure to gases/fumes, mineral and biological dust is associated with small and large airways obstruction (5,6).

Even though occupational exposures are common, it is still largely unknown how these exposures are involved in (lung) disease development. Epigenetic mechanisms such as DNA methylation have been suggested to play a role, and researchers have

therefore advocated the importance of epigenetic studies into environmental exposures and lung health (7). Environmental exposures, like occupational exposures, induce changes in DNA methylation levels, which can affect gene expression, possibly aiding in disease development (8). DNA methylation is the addition of a methyl group to the DNA without altering its sequence. This usually occurs at sites where a cytosine base is adjacent to a guanine base (CpG) and can have a regulatory function on gene expression (9). Several small studies showed suggestive evidence that specific compounds found in occupational exposures, like cadmium, lead and mercury, affect DNA methylation (8,10–12).

To date, no large hypothesis-free genome-wide DNA methylation studies assessing the association between occupational exposures and DNA methylation levels have been performed. We therefore aimed to identify differentially methylated CpG sites (CpGs) and differentially methylated regions (DMRs) associated with occupational exposure to gases/fumes, mineral dust and biological dust, and to assess the effects of these regions on

gene expression levels. To determine the effects of occupational exposures independent of smoking exposure, the analyses were restricted to never-smokers.

## Results

### Population characteristics

Our identification cohort comprised 903 never-smokers of the LifeLines cohort study with genome-wide DNA methylation data (Illumina 450 K) and complete data on all covariates (13). The validation cohort comprised 233 never-smokers of the Rotterdam Study and the characteristics of both cohorts are presented in Table 1, with additional characteristics of LifeLines subjects shown in Supplementary Material, Table S1 (14).

Three genome-wide methylation analyses were performed in never-smokers, one per occupational exposure being gases/fumes, mineral dust and biological dust, and consequently DMRs were identified. We present the results of our analyses per occupational exposure, and see Figure 1 for an overview of the performed analyses and corresponding results. The results of all analyses can be found in the supplementary Excel file including all supplementary tables and the Manhattan plots are shown in Supplementary Material, Figure S1.

### Gases/fumes

**Genome-wide methylation analysis.** In the genome-wide methylation analysis in never-smokers of the identification cohort, two single CpGs were epigenome-wide significantly associated with gases/fumes exposure [false discovery rate (FDR) < 0.05] (Supplementary Material, Table S2). These CpGs are annotated to ribosomal protein L37a (RPL37A) and Grid2-interacting protein (GRID2IP).

**Identification of DMRs.** Thirteen DMRs were significantly associated with exposure to gases/fumes (Supplementary Material, Table S3). The three most significant DMRs are annotated to long intergenic non-protein coding RNA 2169 (LINC02169), ribosomal protein lateral stalk subunit P1 (RPLP1) and leptin (LEP). The genome-wide significant CpG annotated to RPL37A was not located within an identified DMR.

**Validation of the DMRs.** In the validation analysis, two DMRs contained a significantly replicated CpG and exposure to gases/fumes was associated with lower methylation levels at

**Table 1.** Characteristics of the never-smokers included in the LifeLines cohort study (discovery cohort) and the Rotterdam Study (validation cohort)

	LifeLines	Rotterdam Study
N with no missing data	903	233
Males, N (%)	508 (57)	100 (43)
Age (years), median (min–max)	46 (18–80)	57 (47–89)
Occupational exposure, N	<b>No/low/high</b>	<b>No/low/high</b>
Gases/fumes	637/150/116	177/51/5
Mineral dust	673/105/125	210/20/3
Biological dust	720/69/114	N/A

SD, standard deviation; N/A, not applicable. See Supplementary Material, Table S1 for the characteristics of the LifeLines cohort separately per exposure level. NB: The LifeLines sample is a selected population, not a sample from the general population.

these CpGs in both cohorts (Tables 2 and S4). These two DMRs are annotated to Vault RNA 2–1 (VTRNA2–1, a.k.a. MIR886) and guanine nucleotide-binding protein alpha stimulating activity (GNAS) (Fig. 2A and B).

**Gene expression analysis.** We found that CpGs within 9 out of 14 DMRs were significantly associated with differential gene expression, the direction of effect was predominantly negative. Table 3 presents the significant methylation–expression associations of CpGs within replicated DMRs. For the results of all DMRs, see Supplementary Material, Table S5. The replicated DMRs annotated to GNAS were associated with lower expression of NPEPL1.

### Mineral dust

**Genome-wide methylation analysis and identification of DMRs.** No CpGs were genome-wide significantly associated with mineral dust exposure in our identification cohort (FDR < 0.05), but seven DMRs were (Supplementary Material, Table S3). The three most significant hits are annotated to RPLP1, LINC02169 and major histocompatibility Complex class I E (HLA-E), and the first two mentioned DMRs were also associated with exposure to gases/fumes.

**Validation of the DMRs.** The DMR annotated to coiled-coil domain containing 144 family, N-terminal like (CCDC144NL) contained a significantly replicated CpG and the association between mineral dust exposure and methylation levels was negative in both cohorts (Table 2, Supplementary Material, Table S4 and Fig. 2C).

**Gene expression analysis.** In total, CpGs within six out of seven DMRs were significantly associated with differential gene expression and the direction of effect was predominantly negative (Supplementary Material, Table S6). The replicated DMR annotated to CCDC144NL was associated with lower expression of abhydrolase domain containing 17A pseudogene 6 (ABHD17AP6), dehydrogenase/reductase 7B (DHR57B) and galectin 9B (LGALS9B) (Table 3).

### Biological dust

No single CpGs or DMRs were genome-wide significantly associated with biological dust exposure in never-smokers of the identification cohort (FDR < 0.05). Therefore, no validation of results or methylation–expression analyses was performed.

## Discussion

This is the first genome-wide DNA methylation study assessing the association between occupational exposures and DNA methylation. Since it is well known that smoking is associated with extensive changes in DNA methylation levels, we restricted our analyses to never-smokers (15). In these never-smokers, occupational exposure to gases/fumes and to mineral dust was associated with 14 and 7 DMRs, respectively. Three of these DMRs were associated with both gases/fumes and mineral dust (one DMR in RPLP1 and two DMRs in LINC02169). We were able to replicate the result of two DMRs associated with gases/fumes, and one DMR was associated with mineral dust in the Rotterdam Study. These three DMRs were annotated to VTRNA2-1, GNAS

Table 2. Results of the CpGs within the replicated DMRs associated with occupational exposures in never-smokers

DMR	Chr	Start	End	N	Annotation			P	LifeLines			Rotterdam Study		
					Gene	Feature	Island		Beta	SE	P	Beta	SE	P
Gases/fumes														
GN4	chr5	135416331	135416579	7	VTRNA2-1	TSS200	Island	$2.68 \times 10^{-5}$	-2.84	0.83	$5.96 \times 10^{-4}$	-5.68	3.49	0.103
									-1.31	0.56	$1.90 \times 10^{-2}$	-3.25	2.47	0.188
									-1.88	0.59	$1.50 \times 10^{-3}$	-5.81	2.67	0.029
									-1.96	0.63	$1.80 \times 10^{-3}$	-2.04	2.89	0.479
									-1.10	0.44	$1.18 \times 10^{-2}$	-2.33	2.33	0.318
									-1.59	0.55	$3.75 \times 10^{-3}$	-6.23	3.32	0.061
									-2.21	0.84	$8.84 \times 10^{-3}$	-6.87	4.50	0.127
GN13	chr20	57427713	57427880	6	GNAS; GNAS-ASI	Intron; TSS1500	Island	$2.44 \times 10^{-2}$	-1.31	0.52	$1.25 \times 10^{-2}$	-2.45	1.69	0.146
									-1.20	0.42	$4.24 \times 10^{-3}$	-2.52	1.61	0.119
									-1.39	0.46	$2.53 \times 10^{-3}$	-3.15	2.24	0.160
									-1.23	0.45	$6.27 \times 10^{-3}$	-5.2	1.88	0.783
									-1.06	0.54	$4.69 \times 10^{-2}$	-4.54	2.31	0.049
									-1.32	0.59	$2.54 \times 10^{-2}$	-1.81	2.75	0.510
Mineral dust														
MN7	chr17	20799408	20799694	6	CCDC144NL; RP11-344E13.3	TSS200; 5'UTR	Island	$1.13 \times 10^{-3}$	-1.06	0.54	$5.03 \times 10^{-2}$	-6.07	2.30	0.008
									-3.36	1.36	$1.37 \times 10^{-2}$	-9.67	6.06	0.111
									-3.05	0.93	$1.05 \times 10^{-3}$	-8.33	7.66	0.277
									-2.81	0.92	$2.14 \times 10^{-3}$	-10.25	5.65	0.070
									-3.20	1.01	$1.48 \times 10^{-3}$	-14.71	7.64	0.054
									-2.05	0.85	$1.62 \times 10^{-2}$	-2.90	3.94	0.461

Chr, Chromosome; CpG, DNA-methylation site; SE, standard error.

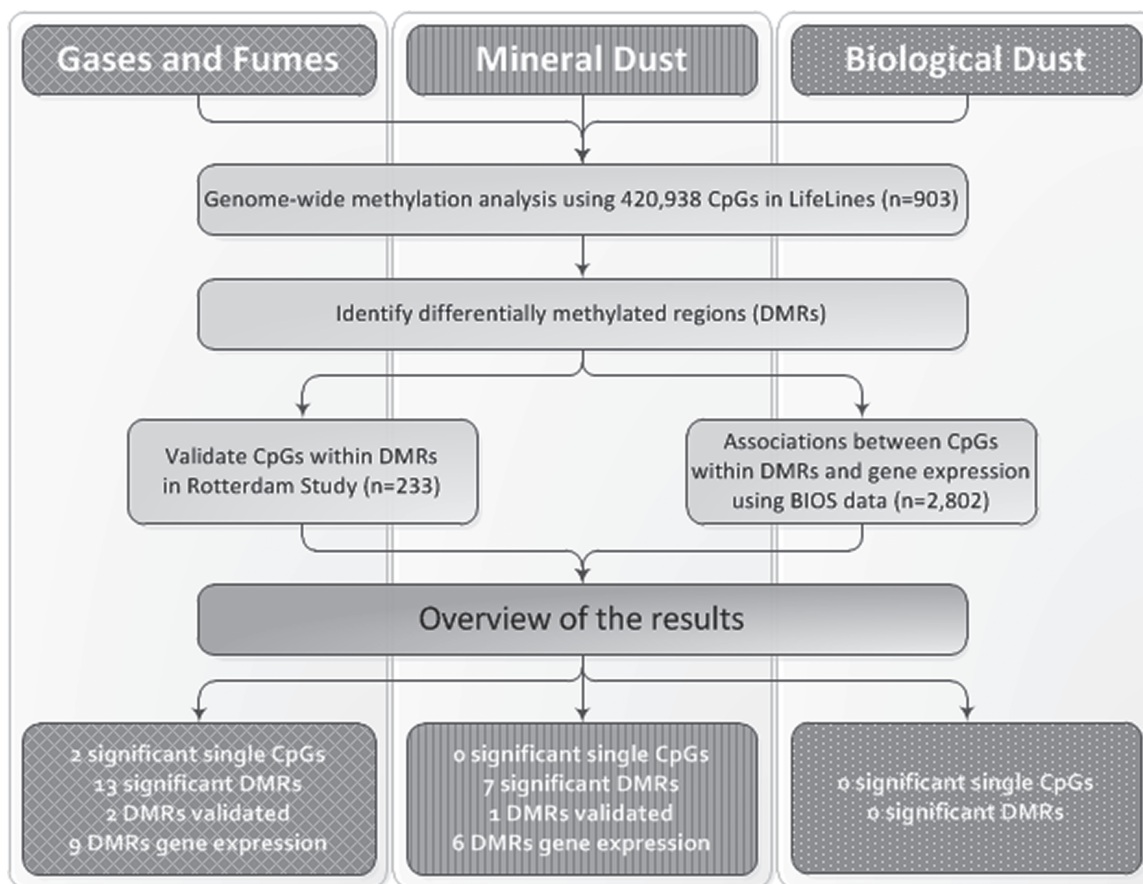


Figure 1. Overview of the performed analyses and results per occupational exposure. All analyses were performed for the three exposures in never-smokers.

Table 3. Results of replicated DMRs in never-smokers who were associated with gene expression levels for genes located within 1 MB of the CpG (n = 2802)

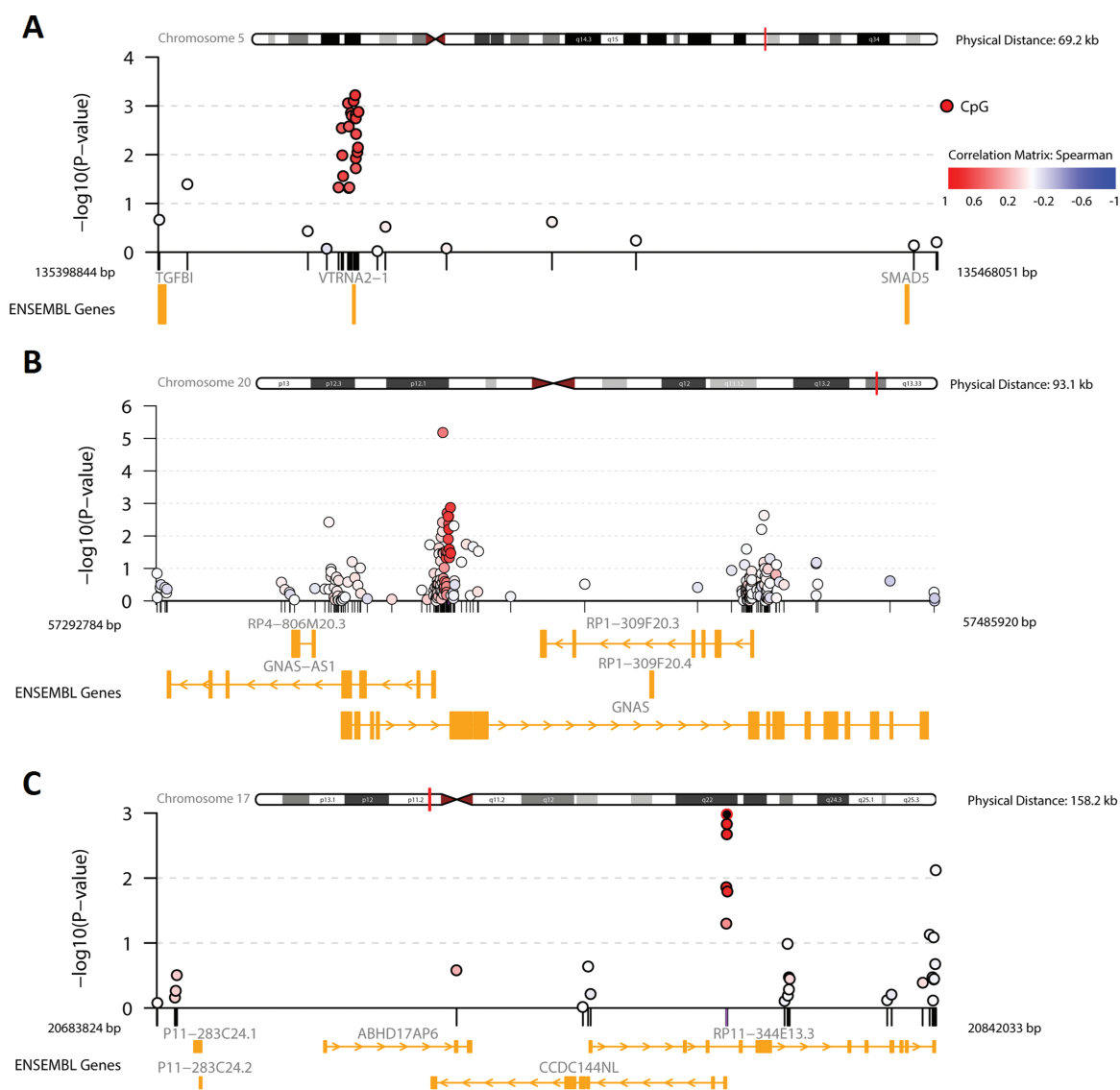
DMR	CpG	Annotated gene	Ensembl_ID	Gene	B	SE	P adjusted
<i>Gases/fumes</i>							
GN13	cg04257105	GNAS	ENSG00000254419	NPEPL1	-0.428	0.134	2.19*10 <sup>-2</sup>
	cg17414107	GNAS	ENSG00000254419	NPEPL1	-0.388	0.125	2.81*10 <sup>-2</sup>
	cg19589727	GNAS	ENSG00000254419	NPEPL1	-0.580	0.188	3.03*10 <sup>-2</sup>
	cg20528838	GNAS	ENSG00000254419	NPEPL1	-0.472	0.156	3.68*10 <sup>-2</sup>
<i>Mineral dust</i>							
MN7	cg06809326	CCDC144NL	ENSG00000226981	ABHD17AP6	-1.132	0.251	9.44*10 <sup>-5</sup>
			ENSG00000109016	DHRS7B	0.129	0.044	2.53*10 <sup>-2</sup>
	cg08288433	CCDC144NL	ENSG00000226981	ABHD17AP6	-0.992	0.231	2.67*10 <sup>-4</sup>
	cg14560110	CCDC144NL	ENSG00000226981	ABHD17AP6	-0.977	0.187	2.57*10 <sup>-6</sup>
			ENSG00000170298	LGALS9B	0.554	0.168	7.13*10 <sup>-3</sup>
	cg21980100	CCDC144NL	ENSG00000226981	ABHD17AP6	-1.327	0.304	1.92*10 <sup>-4</sup>
			ENSG00000109016	DHRS7B	0.145	0.053	4.72*10 <sup>-2</sup>
cg22570042	CCDC144NL	ENSG00000226981	ABHD17AP6	-1.032	0.216	2.57*10 <sup>-5</sup>	

CpG, DNA-methylation site; B, beta; SE, standard error; P adjusted, FDR correct meta-analysis p-value based on genes with available data located within the 1 MB window of the CpG.

and CCDC144NL. CpGs within the DMRs annotated to GNAS and CCDC144NL were significantly associated with lower expression levels of NPEPL1 and ABHD17AP6, respectively. Moreover, 14 out of 21 DMRs were associated with gene expression levels and 11 DMRs were located within the transcript start sites (TSS) of a gene. Together, our data suggest that occupational exposures may induce differential DNA methylation at specific genomic

locations and this may be a mechanism through which occupational exposures affect health.

Interestingly, the majority of identified DMRs were located within the TSS of a gene; 55.2% and 48.5% of the CpGs within the DMRs associated with gases/fumes and mineral dust, respectively, were located in the TSS, compared to 25.7% of all included CpGs in the study. The three replicated DMRs were also located in



**Figure 2.** Regional association plots (R package comet) for the three replicated DMRs in never-smokers. (A) DMR annotated to *VTRNA2-1*, (B) DMR annotated to *GNAS*, and (C) DMR annotated to *CCDC144NL*. x-axis, megabase (Mb) position on the chromosome; y-axis, negative log<sub>10</sub> of the P-values; dots, CpG sites; and see inset legend for the correlation explanation between CpGs.

the TSS, of which two were also associated with gene expression levels (*GNAS* and *CCDC144NL*). The general idea of the function of DNA methylation at these TSSs is that it blocks the initiation of transcription and thereby lowers gene expression (9). In the current study, we observed that occupational exposure is associated with lower DNA methylation levels which in turn are associated with higher gene expression levels for most DMRs associated with gene expression levels. This observation thus corroborates our knowledge of the function of DNA methylation at TSSs. Moreover, several of the DMRs associated with gene expression were not associated with the annotated gene. This is partly due to the fact that for 11 of our identified DMRs no gene expression data was available for the annotated gene, including the replicated DMR annotated to *VTRNA2-1*. For others, CpGs within a DMR were nominally associated with expression levels of the annotated gene but did not survive the multiple testing correction (e.g. the replicated DMR annotated to *GNAS*).

Another intriguing observation is that several DMRs that we identified are annotated to or associated with the expres-

sion of genes with unknown function, RNA genes or pseudo-genes, like *CCDC144NL*, *ABHD17AP6*, *NPEPL1*, *RP11-373 N24.2* and *LINC02169*. It is therefore challenging to understand the relation between these genes and occupational exposures. Long non-coding RNAs (lncRNAs) are known to play a role in gene expression regulation during development, cell differentiation, genomic imprinting and sex chromosomal dosage compensation (16). The gene *ZSCAN26* is a zinc finger (transcription factor) and may therefore also be involved in gene expression regulation (17). In addition, multiple microRNAs and lncRNAs were shown to be key regulators of gene expression in lung diseases such as asthma and COPD (18). These might even be biomarkers or therapeutic targets, but more research into the function of these genes is warranted. For your interest, results of Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways and Gene Ontology (GO) term enrichment analyses are included in [Supplementary Material, Tables S7 and S8](#). Overall, gene expression changes could drive DNA methylation changes due to cellular differentiation as a response to occupational exposures.

However, it is more likely that occupational exposures may affect regulation of gene expression by changing DNA methylation levels of particular genes that regulate the expression of other genes.

Interestingly, the three DMRs annotated to *RPLP1* and *LINC02169* (2×) were identified in both the gases/fumes and mineral dust analyses. In addition, CpGs annotated to *VTRNA2-1* were also associated with occupational exposure to pesticides in our previous study (13). *RPLP1* is a ribosomal protein regulating translation and *VTRNA2-1* is indirectly also related to the innate immune response, since it was shown to inhibit protein kinase R (*EIF2AK2*) (19,20). This could indicate that different types of occupational exposures affect similar pathways; alternatively it could result from multiple occupational exposures in specific jobs. For example, construction workers can be exposed to mineral dust and gases/fumes at the same time and crop farmers distribute pesticides over their fields using fuelled machines (gases/fumes exposure). Notably, eight subjects of our cohort were highly exposed to all three occupational exposures and the exposures are moderately to strongly correlated (correlation between gases/fumes and mineral dust = 0.85, between gases/fumes and biological dust = 0.66 and between mineral dust and biological dust = 0.56; [Supplementary Material, Table S9](#)). Since we used broad categories of occupational exposures, it was not possible to investigate specific exposure molecules. Occupational exposure levels were also estimated based on current or last held job, and duration of exposure was not taken into account. It is likely that some subjects classified as non-exposed have changed from an 'exposed' to a 'non-exposed' job, because they experienced adverse effects from the exposures. Therefore, we may have underestimated the effect of occupational exposures on DNA methylation. However, in our cohort on average 72% of the subjects currently exposed to gases/fumes, mineral or biological dust had this job for >5 years and thus had been exposed for a substantial time period in the same job.

Another restriction of our study is the use of blood DNA methylation levels. DNA methylation is cell and tissue specific, and the main route of occupational exposure is via inhalation or skin absorption. However, we have validated a number of CpGs associated with cigarette smoke exposure in lung tissue that were originally identified in whole blood (21). Thus using whole blood could be an efficient way to identify differential DNA methylation upon exposures as an accessible proxy for changes in lung tissue. For your interest, the associations between the lung function measurements  $FEV_1$ , FVC,  $FEV_1/FVC$  and  $FEF_{25-75}$  and CpGs within DMRs are provided in Table S10. Furthermore, using a job exposure matrix does not allow to assess specific chemical compounds present in occupational exposures, nor the effect of lifetime exposure. Different types of jobs classified into the same exposure category might contain different chemical compounds as well. Therefore, our results reflect the effect of current or recent occupational exposure on DNA methylation.

In conclusion, our data suggest that occupational exposures may induce differential methylation of genes that regulate gene expression and therefore occupational exposures may induce adverse health effects via this methylation. Several of our identified differentially methylated regions upon occupational exposure to gases/fumes and mineral dust were associated with gene expression levels. Some regions were even associated with two types of occupational exposure. Given the millions of workers that are exposed daily to occupational exposures, further studies on this epigenetic mechanism and health outcomes are warranted. For example, since 40% of the occupational cause

of death is due to COPD, especially in developing countries without proper precautions, further studies on this epigenetic mechanism could aid in reducing the global burden of COPD (1).

## Materials and Methods

### Population and measurements

From the LifeLines Cohort Study, 1656 unrelated subjects were selected for DNA methylation assessment (13). Subject selection was based on creating relatively equal-sized groups based on age, smoking, occupational exposures and spirometry. In the current study only never-smokers were included in order to determine the effects of occupational exposures independent of smoking exposure. No, low and high occupational exposures to gases/fumes, mineral dust and biological dust were estimated using the ALOHA+ job exposure matrix, based on current or last held job (6,22). See [Figure 1](#) and [supplementary methods](#) for an overview and more detailed information on the methods.

### Genome-wide methylation data and analysis

Illumina Infinium Human Methylation 450 K arrays were used to obtain genome-wide DNA methylation data from whole blood. Samples were processed using the Illumina protocol. Quality control (QC) using *Minfi* and normalization using *DASEN* (*watermelon*) were performed in R (23,24). QC steps included the removal of samples with >1% of all probes having a detection  $P > 0.01$ , and samples with incorrect sex or SNP prediction. We removed single probes with a detection  $P > 0.01$ , sex chromosome probes, cross-reactive probes (25), probes measuring SNPs and probes where the CpG itself or the single base extension (SBE) site is a SNP. After QC, we had complete data for 420 938 CpG probes in 903 never-smoking subjects.

We performed three genome-wide methylation analyses, one per occupational exposure being gases/fumes, mineral dust and biological dust. We assessed associations between DNA methylation levels (beta-values ranging from 0 to 1) and the three occupational exposures separately using robust linear regression in R [*MASS* package]. Models included low and high exposure dummy-variables (no exposure as reference), and were adjusted for sex, age, technical variances and differential blood counts (eosinophilic, neutrophilic and basophilic granulocytes, lymphocytes and monocytes, all obtained using standard laboratory techniques). Single CpGs with a FDR adjusted  $P < 0.05$  for the high-exposure dummy variable were considered genome-wide significant.

### Identification of DMRs

To identify DMRs, *comb-p* in Python was used (14). *Comb-p* identifies regions of enrichment by combining adjacent  $P$ -values into FDR adjusted regional  $P$ -values using auto-correlation and sliding windows. As input we used  $P$ -values of the high-exposure dummy variable with the following settings: seed = 0.01 and distance = 300. Each CpG within a DMR with a Šidák-corrected  $P < 0.05$  was further investigated (26).

### Validation of DMRs

DMRs identified in LifeLines were validated in the baseline assessment of the third Rotterdam Study cohort (RS-III-1,  $n = 722$ ) (27). Blood DNA methylation levels were measured using Illumina 450 K arrays and processed similar to LifeLines

as described elsewhere (28). All CpGs within the DMRs were validated in RS-III-1 and the statistical models (robust linear regression) were similar to LifeLines. Single CpGs and CpGs within the DMRs with a nominal validation  $P < 0.05$  and same direction of effect in both cohorts were considered significantly replicated.

### Association between CpGs within DMRs and gene expression

To assess whether the CpGs within the DMRs were associated with gene expression levels, we used data from four population-based cohorts within Biobank-based Integrative Omics Studies (BIOS), from the Biobanking and Biomolecular Resources Research Infrastructure for The Netherlands (BBMRI-NL) (29). In total, 2802 subjects were included in the analyses (independent samples of LifeLines,  $n=727$ ; Rotterdam Study III-2,  $n=589$ ; Netherlands Twin Registry,  $n=900$ ; and Leiden Longevity Study,  $n=586$ ) (30–33). In each cohort, probesets (read counts from RNA sequencing) within 1 Mb around the CpG were assessed and the linear regression was adjusted for sex, smoking, age and technical variances. Effect estimates of the cohorts were meta-analyzed. CpGs with a meta-analysis FDR-corrected  $P$ -value below 0.05 were considered significant.

### Supplementary Material

Supplementary Material is available at HMG online.

### Acknowledgements

The authors wish to acknowledge the services of the LifeLines Cohort Study and the Rotterdam Study, the contributing research centers delivering data to LifeLines and the participating general practitioners in the Rotterdam Study and all the study participants. The authors also thank Mr Michael Verbiest, Ms Mila Jhamai, Ms Sarah Higgins, Mr Marijn Verkerk and Lisette Stolk, PhD, for their help in creating the methylation database.

The LifeLines Biobank initiative has been made possible by funds from FES (Fonds Economische Structuurversterking), SNN (Samenwerkingsverband Noord Nederland) and REP (Ruimtelijke Economisch Programma). The Rotterdam Study is funded by Erasmus Medical Center and Erasmus University, Rotterdam, Netherlands Organization for the Health Research and Development (ZonMw), the Research Institute for Diseases in the Elderly (RIDE), the Ministry of Education, Culture and Science, the Ministry for Health, Welfare and Sports, the European Commission (DG XII), and the Municipality of Rotterdam. The generation and management of the Illumina 450 K methylation array data (EWAS data) for the Rotterdam Study was executed by the Human Genotyping Facility of the Genetic Laboratory of the Department of Internal Medicine, Erasmus MC, the Netherlands. The EWAS data was funded by the Genetic Laboratory of the Department of Internal Medicine, Erasmus MC, and by the Netherlands Organization for Scientific Research (NWO; project number 184021007) and made available as a Rainbow Project (RP3; BIOS) of the Biobanking and Biomolecular Research Infrastructure Netherlands (BBMRI-NL). The Biobank-Based Integrative Omics Studies (BIOS) Consortium is funded by BBMRI-NL, a research infrastructure financed by the Dutch government (NWO 184.021.007).

**Conflict of Interest statement.** L.L. reports personal fees from Boehringer Ingelheim GmbH, non-financial support from Novartis, grants from AstraZeneca, grants and non-financial support from European Respiratory Society and grants and non-financial support from Belgian Respiratory Society, outside the submitted work. DSP reports: The University of Groningen has received money for D.S.P. regarding a grant for research from Astra Zeneca, Chiesi, Genentec, GSK and Roche. Fees for consultancies were given to the University of Groningen by Astra Zeneca, Boehringer Ingelheim, Chiesi, GSK, Takeda and TEVA. All other authors declare they have no actual or potential competing financial interest.

### Funding

Lung Foundation Netherlands (4.1.13.007, 4.113.007 to D.A.v.d.P. and N.A.); Research Foundation—Flanders (postdoctoral fellowship to L.L.).

### References

1. Concha-Barrientos, M., Nelson, D.I., Driscoll, T., Steenland, N.K., Punnett, L., Fingerhut, M., Prüss-Üstün, A., Leigh, J., Tak, S.W. and Corvalan, C. (2010) Chapter 21: Selected occupational risk factors. In *Comparative Quantification Of Health Risks*. World Health Organization, pp. 1651–1801.
2. Iversen, M. and Dahl, R. (2000) Working in swine-confinement buildings causes an accelerated decline in FEV1: a 7-yr follow-up of Danish farmers. *Eur. Respir. J.*, **16**, 404–408.
3. Bakke, B., Ulvestad, B., Stewart, P. and Eduard, W. (2004) Cumulative exposure to dust and gases as determinants of lung function decline in tunnel construction workers. *Occup. Environ. Med.*, **61**, 262–269.
4. Wang, X., Zhang, H.X., Sun, B.X., Dai, H.L., Hang, J.Q., Eisen, E., Su, L. and Christiani, D.C. (2008) Cross-shift airway responses and long-term decline in FEV1 in cotton textile workers. *Am. J. Respir. Crit. Care Med.*, **177**, 316–320.
5. de Jong, K., Boezen, H.M., Kromhout, H., Vermeulen, R., Vonk, J.M., Postma, D.S. and Study, L.C. (2014) Occupational exposure to vapors, gases, dusts, and fumes is associated with small airways obstruction. *Am. J. Respir. Crit. Care Med.*, **189**, 487–490.
6. de Jong, K., Boezen, H.M., Kromhout, H., Vermeulen, R., Postma, D.S., Vonk, J.M. and Study, L.C. (2014) Pesticides and other occupational exposures are associated with airway obstruction: the LifeLines cohort study. *Occup. Environ. Med.*, **71**, 88–96.
7. Melén, E., Barouki, R., Barry, M., Boezen, H.M., Hoffmann, B., Krauss-Etschmann, S., Koppelman, G.H. and Forsberg, B. (2018) Promoting respiratory public health through epigenetics research: an ERS environment health committee workshop report. *Eur. Respir. J.*, **51**.
8. Ruiz-Hernandez, A., Kuo, C.C., Rentero-Garrido, P., Tang, W.Y., Redon, J., Ordovas, J.M., Navas-Acien, A. and Tellez-Plaza, M. (2015) Environmental chemicals and DNA methylation in adults: a systematic review of the epidemiologic evidence. *Clin. Epigenetics*, **7**, eCollection 2015.
9. Jones, P.A. (2012) Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Natl. Rev.*, **13**, 484–492.
10. Hossain, M.B., Vahter, M., Concha, G. and Broberg, K. (2012) Low-level environmental cadmium exposure is associated with DNA hypomethylation in Argentinean women. *Environ. Health Perspect.*, **120**, 879–884.



11. Li, C., Yang, X., Xu, M., Zhang, J. and Sun, N. (2013) Epigenetic marker (LINE-1 promoter) methylation level was associated with occupational lead exposure. *Clin. Toxicol. (Phila.)*, **51**, 225–229.
12. Goodrich, J.M., Basu, N., Franzblau, A. and Dolinoy, D.C. (2013) Mercury biomarkers and DNA methylation among Michigan dental professionals. *Environ. Mol. Mutagen.*, **54**, 195–203.
13. van der Plaat, D.A., de Jong, K., de Vries, M., van Diemen, C.C., Nedeljković, I., Amin, N., Kromhout, H., Biobank-based Integrative Omics Study Consortium, Vermeulen, R., Postma, D.S. et al. (2018) Occupational exposure to pesticides is associated with differential DNA methylation. *Occup. Environ. Med.*, **75**, 427–435.
14. Pedersen, B.S., Schwartz, D.A., Yang, I.V. and Kechris, K.J. (2012) Comb-p: software for combining, analyzing, grouping and correcting spatially correlated P-values. *Bioinformatics*, **28**, 2986–2988.
15. Zeilinger, S., Kuhnel, B., Klopp, N., Baurecht, H., Kleinschmidt, A., Gieger, C., Weidinger, S., Lattka, E., Adamski, J., Peters, A. et al. (2013) Tobacco smoking leads to extensive genome-wide changes in DNA methylation. *PLoS One*, **8**, e63812.
16. Fatica, A. and Bozzoni, I. (2014) Long non-coding RNAs: new players in cell differentiation and development. *Natl. Rev.*, **15**, 7–21.
17. Klug, A. (2010) The discovery of zinc fingers and their applications in gene regulation and genome manipulation. *Annu. Rev. Biochem.*, **79**, 213–231.
18. Booton, R. and Lindsay, M.A. (2014) Emerging role of MicroRNAs and long noncoding RNAs in respiratory disease. *Chest*, **146**, 193–204.
19. Martinez-Azorin, F., Remacha, M. and Ballesta, J.P. (2008) Functional characterization of ribosomal P1/P2 proteins in human cells. *Biochem. J.*, **413**, 527–534.
20. Kunkeaw, N., Jeon, S.H., Lee, K., Johnson, B.H., Tanasanvimon, S., Javle, M., Pairojkul, C., Chamgramol, Y., Wongfieng, W., Gong, B. et al. (2013) Cell death/proliferation roles for nc886, a non-coding RNA, in the protein kinase R pathway in cholangiocarcinoma. *Oncogene*, **32**, 3722–3731.
21. de Vries, M., van der Plaat, D.A., Nedeljkovic, I., Verkaik-Schakel, R.N., Kooistra, W., Amin, N., van Duijn, C.M., Brandsma, C.-A., van Diemen, C.C., Vonk, J.M. et al. (2018) From blood to lung tissue: effect of cigarette smoke on DNA methylation and lung function. *Respir. Res.*, **19**, 212.
22. Matheson, M.C., Benke, G., Raven, J., Sim, M.R., Kromhout, H., Vermeulen, R., Johns, D.P., Walters, E.H. and Abramson, M.J. (2005) Biological dust exposure in the workplace is a risk factor for chronic obstructive pulmonary disease. *Thorax*, **60**, 645–651.
23. Aryee, M.J., Jaffe, A.E., Corrada-Bravo, H., Ladd-Acosta, C., Feinberg, A.P., Hansen, K.D. and Irizarry, R.A. (2014) Minfi: a flexible and comprehensive bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics*, **30**, 1363–1369.
24. Pidsley, R., Y Wong, C.C., Volta, M., Lunnon, K., Mill, J. and Schalkwyk, L.C. (2013) A data-driven approach to preprocessing illumina 450K methylation array data. *BMC Genomics*, **14**, 293.
25. Chen, Y.A., Lemire, M., Choufani, S., Butcher, D.T., Grafodatskaya, D., Zanke, B.W., Gallinger, S., Hudson, T.J. and Weksberg, R. (2013) Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray. *Epigenetics*, **8**, 203–209.
26. Sidak, Z. (1967) Rectangular confidence regions for the means of multivariate Normal distributions. *J. Am. Stat. Assoc.*, **62**, 626–633.
27. Hofman, A., Brusselle, G.G., Darwish Murad, S., van Duijn, C.M., Franco, O.H., Goedegebure, A., Ikram, M.A., Klaver, C.C., Nijsten, T.E., Peeters, R.P. et al. (2015) The Rotterdam Study: 2016 objectives and design update. *Eur. J. Epidemiol.*, **30**, 661–708.
28. Ligthart, S., Steenaard, R.V., Peters, M.J., van Meurs, J.B., Sijbrands, E.J., Uitterlinden, A.G., Bonder, M.J., Consortium, B., Hofman, A., Franco, O.H. et al. (2016) Tobacco smoking is associated with DNA methylation of diabetes susceptibility genes. *Diabetologia*, **59**, 998–1006.
29. Biobanking and Biomolecular Resources Research Infrastructure for The Netherlands (BBMRI-NL) (2017) Biobank-based integrative Omics studies (BIOS). 2017. Available: <https://www.bbMRI.nl/acquisition-use-analyze/bios>.
30. Tigchelaar, E.F., Zhernakova, A., Dekens, J.A., Hermes, G., Baranska, A., Mujagic, Z., Swertz, M.A., Munoz, A.M., Deelen, P., Cenit, M.C.C. et al. (2015) Cohort profile: LifeLines DEEP, a prospective, general population cohort study in the northern Netherlands: study design and baseline characteristics. *BMJ Open*, **5**, e006772–e002014-006772.
31. Hofman, A., van Duijn, C.M., Franco, O.H., Ikram, M.A., Janssen, H.L., Klaver, C.C., Kuipers, E.J., Nijsten, T.E., Stricker, B.H., Tiemeier, H. et al. (2011) The Rotterdam Study: 2012 objectives and design update. *Eur. J. Epidemiol.*, **26**, 657–686.
32. Willemsen, G., Vink, J.M., Abdellaoui, A., den Braber, A., van Beek, J.H., Draisma, H.H., van Dongen, J., van't Ent, D., Geels, L.M., van Lien, R. et al. (2013) The adult Netherlands twin register: twenty-five years of survey and biological data collection. *Twin Res. Hum. Genet.*, **16**, 271–281.
33. Westendorp, R.G., van Heemst, D., Rozing, M.P., Frolich, M., Mooijaart, S.P., Blauw, G.J., Beekman, M., Heijmans, B.T., de Craen, A.J., Slagboom, P.E. et al. (2009) Nonagenarian siblings and their offspring display lower risk of mortality and morbidity than sporadic nonagenarians: the Leiden longevity Study. *J. Am. Geriatr. Soc.*, **57**, 1634–1637.