Université de Montréal

The influence of smoking and occupational exposures on DNA methylation in the *AHRR* and *F2RL3* genes

Par

Michael Pham

Département de médecine sociale et préventive

École de Santé publique de l'Université de Montréal

Mémoire présenté en vue de l'obtention du grade de Maîtrise

en Épidémiologie

Juin 2022

Résumé

Objectif: Déterminer l'association entre le tabagisme et les expositions professionnelles, et les niveaux de méthylation dans les gènes *AHRR* et *F2RL3*, deux gènes impliqués dans le cancer du poumon.

Méthodes : CARTaGENE est la plus grande étude de cohorte prospective au Québec, Canada. Actuellement, une étude de cas-témoin nichée dans CARTaGENE examine l'association entre la méthylation des gènes *AHRR* et *F2RL3* et le risque de cancer du poumon (200 cas; 400 témoins). En utilisant les données de méthylation mesurées à partir de cette étude de cas-témoin nichée, les informations à propos du comportement tabagique et de l'emploi avec la plus longue durée des participants ont été obtenues à partir de questionnaires. Les informations concernant le statut tabagique, le nombre moyen de cigarettes fumées, la durée du tabagisme et le temps depuis la cessation (quand applicable) ont été paramétrées sous la forme d'un index cumulatif de tabagisme (continu). Les expositions professionnelles ont été estimées à partir de la matrice canadienne de l'exposition professionnelle. Dix-huit agents présents dans les milieux professionnels et également présents dans la fumée de tabac ont été retenus. Les ratios de méthylation de 40 sites CpG dans les gènes *AHRR* et *F2RL3* ont été mesurés avec le Sequenom Epityper. La moyenne des ratios de méthylation de tous les sites CpG a été calculée par gène et paramétrée comme une variable continue. Des modèles séparés de régression des moindres carrés ont été utilisés pour estimer les associations entre chacun des facteurs de risque et les niveaux de méthylation des gènes *AHRR* et *F2RL3* tout en ajustant pour des variables confondantes identifiées à l'aide de graphes acycliques dirigés.

Résultats : Le tabagisme est associé avec des niveaux moyens de méthylation plus faible dans chacun des gènes après ajustement pour les variables confondantes (*AHRR* : -0.014 par augmentation de l'écart-type de l'index cumulatif de tabagisme, 95% IC : -0.019, -0.010; *F2RL3* :

-0.019 par augmentation de l'écart-type de l'index cumulatif de tabagisme, 95% IC : -0.025, -0.012). Aucune association n'a été observée entre les expositions occupationnelles sélectionnées et les niveaux de méthylation dans ces deux gènes.

Conclusion : Nos observations indiquent que le tabagisme est associé avec une hypométhylation des gènes *AHRR* et *F2RL3*.

Mots-clés : Méthylation de l'ADN, épigénétiques, tabagisme, expositions occupationnelles, matrice d'exposition professionnelle.

Abstract

Objective: To determine the association between smoking and occupational exposures, and DNA methylation levels in the lung cancer-related genes, *AHRR* and *F2RL3*.

Methods: CARTaGENE is the largest ongoing prospective cohort study in Quebec, Canada. Currently, a nested case-control study in CARTaGENE is examining the association between *AHRR* and *F2RL3* gene methylation and lung cancer risk (200 cases; 400 controls). Using the methylation data measured from this nested case-control study, information on participants' smoking behavior and longest-held occupation were obtained from questionnaires. Information on smoking status and, where applicable, the average number of cigarettes smoked, duration of smoking, and time since cessation, was parameterized into a cumulative smoking index (CSI, continuous). Occupational exposures were estimated using the Canadian Job Exposure Matrix. Eighteen agents present in the occupational environment that are also found in cigarette smoke were of interest. In DNA isolated from blood samples collected at baseline, methylation ratios of 40 CpG sites in the *AHRR* and *F2RL3* genes were measured using the Sequenom Epityper. In each gene, average methylation levels across all CpG sites were calculated and parametrized as a continuous variable. Separate least squares regression models were used to estimate the associations between smoking and occupational exposures, and *AHRR* and *F2RL3* methylation levels while adjusting for potential confounders identified using directed acyclic graphs.

Results: In both genes, smoking was associated with lower average methylation levels after adjusting for confounding factors (*AHRR*: -0.014 per standard deviation increase in CSI, 95% CI: -0.019, -0.010; *F2RL3*: -0.019 per standard deviation increase in CSI, 95% CI: -0.025, -0.012). No association was found between the selected occupational exposures and average DNA methylation levels in the two genes.

Conclusion: Our findings support the hypothesis that tobacco smoking is associated with DNA hypomethylation of the *AHRR* and *F2RL3* genes.

Keywords: DNA methylation; epigenetics; smoking; occupational exposures; job exposure matrix

Table of Content

List of Tables

Chapter 2.

Chapter 4.

Chapter 5.

Chapter 6.

List of Figures

Chapter 5.

List of Acronyms

DNA: Deoxyribonucleic acid

CpG: Cytosine-phosphate-Guanine

EPIC: European Prospective Investigation into Cancer and Nutrition

CIHR: Canadian Institutes of Health Research

CANJEM: Canadian job-exposure-matrix

IARC: International Agency for Research on Cancer

CI: Confidence interval

CSI: Cumulative Smoking Index

PAHs: Polycyclic Aromatic Hydrocarbons

DAG: Directed Acyclic Graphs

ISCO: International Standard Classification of Occupations

CDC: Centre for Disease Control and Prevention

EWAS: Epigenome-wide association study

BMI: Body Mass Index

SES: Socioeconomic Status

List of Abbreviations

vs.: Versus

e.g.: For example

i.e.,: That is

et al.: and others

ACKNOWLEDGMENTS

Above all, I would like to thank my supervisors Dr. A. Koushik and Dr. V. Ho for their support and guidance throughout this research project. To Dr. V. Ho, thank you for providing me with an opportunity to be involved in this project, and for being patient and confident in my capacity to complete it. Your mentorship provided me with the tools to develop good research methodology, resilience, and self-confidence. To Dr. A. Koushik, thank you for providing me with advice throughout the year to improve my presentation and writing skills. Furthermore, I would like to acknowledge the faculty and staff in the Montreal University Public Health School's department of Social and Preventive Medicine for providing such a supportive and welcoming learning environment where I was able to grow as a student and as a person. Last but not least, to all my friends, thank you for believing in me unconditionally and with all your heart. Special thanks to Alice, Romain, Saeedeh, Laura and Najmeh for always offering me new perspectives on tackling incoming issues and helping me see the brighter side of things. Your ability to cheer me up has been unparalleled.

Finally, this project is funded by the Canadian Institute of Health Research. The investigators would like to thank the participants from the CARTaGENE study who made this project possible.

Chapter 1. INTRODUCTION

Globally, lung cancer remains a public health burden; in 2020, 2.21 million new cases of lung cancer were reported worldwide and 1.80 million people died from the disease (1). According to the Canadian Cancer Society, lung cancer is the second most commonly diagnosed cancer and the principal cause of death from cancer for both men and women in Canada (2). Smoking is known to be the main risk factor. But, other factors must play a role since lung cancer occurs among individuals who have never smoked and only a minority of smokers develops the disease (3). Besides smoking, the occupational environment is one of the most fruitful areas for research on lung cancer risk factors. According to Statistics Canada, around 60% of the Canadian population work (4), and many encounter hazards in the workplace (5). For example, studies have highlighted the associations between occupational exposures to asbestos, diesel exhaust, and other combustion products, and increased lung cancer risk (6, 7). However, despite knowledge of these associations, the mechanisms underlying the impact of smoking and many occupational exposures on lung cancer etiology remain poorly understood.

DNA methylation is a common biological process where a methyl group is added to cytosine bases of DNA (8). Most of the time, it occurs at cytosine residues that are adjacent to a guanine nucleotide, which then forms cytosine-phosphate-guanine (CpG) sites. DNA methylation is involved in several cellular processes. Global DNA methylation, which refers to the average methylation status that occurs across the whole genome, is distinguished from gene-specific DNA methylation which refers to the methylation status of specific genes (9). Evidence for the association between both aberrant global and gene-specific DNA methylation, and increased lung cancer risk has been accumulating (10, 11).

The aryl hydrocarbon receptor repressor (*AHRR*) and the coagulation factor II receptor-like 3 (*F2RL3*) genes encode proteins that are involved in many biochemical and physiological mechanisms including cell proliferation and differentiation or platelet activation (12, 13). There is emerging evidence that they are involved in the pathophysiology of several malignant tumors and are implicated in lung cancer etiology (14-16). Specifically, emerging evidence has suggested that the methylation in these two genes can serve as an intermediate biological marker underlying environmental exposures, such as smoking, and lung cancer risk (17, 18).

Given that certain workers are exposed to many of the same carcinogens present in cigarette smoke due to the nature of their occupation, this project proposes to examine the relationships between smoking and selected occupational exposures, and DNA methylation levels in the *AHRR* and *F2RL3* genes. This study will contribute to elucidating the impact of environmental factors on the epigenetic mechanism of gene-specific DNA methylation. This dissertation is composed of six main chapters. Chapter 2 presents an overview of lung cancer etiology and DNA methylation. Chapter 3 presents the study objectives and an overview of the study methodology. Next, Chapter 4 consists of the results, presented in the form of a manuscript that will be submitted to the journal entitled Mutation Research – Genetic Toxicology and Environmental Mutagenesis. Finally, Chapter 5 presents additional results and Chapter 6 presents a discussion of the key findings along with the strengths and limitations of this dissertation.

Chapter 2. LITERATURE REVIEW

2.1 Burden of lung cancer

Worldwide, cancer remains an important health issue. It is estimated that approximately 19.3 million new cancer cases arose in 2020 of which 11.4% were lung cancers. Furthermore, close to 10 million people died from cancer in 2020, and among them, 18% died from lung cancer (19). This makes lung cancer the second most diagnosed cancer and the one with the highest mortality rate among all cancers globally. The low survival rate can largely be explained by diagnosis at an advanced stage in the majority of cases (20) and the high relapse rate even among early-stage cancers (21). The lung cancer burden is predicted to double over the next decade in low- and middle-income countries with new lung cancer cases expected to reach 29 million by 2040 (22).

In Canada, lung cancer incidence and mortality rates have been declining in recent years. Overall, rates are converging between the sexes but remain higher among males compared to females. Specifically, the incidence is 20% higher in males than females while the mortality rate is about 30% higher in males in comparison with females (23). In 2020, it was estimated that lung cancer cases in both sexes combined accounted for 13% of new cancer cases. In parallel, lung cancer deaths accounted for 25% of all cancer deaths. The lung cancer survival rate is usually higher among females than males across all age groups, regardless of the province at diagnosis (23).

Lung carcinogenesis is a complex process that involves genetic mutations and epigenetic changes that modify cellular processes such as cell proliferation, differentiation, and metastasis (24). It is a disease with an estimated latency period of 20 years (25). There are different histological types of lung cancer. The most prevalent are adenocarcinoma (approximately 40% of all lung cancer cases) and squamous cell carcinoma (representing 25% of all lung cancer cases)

(26, 27). The considerable burden of lung cancer, despite the extensive knowledge accumulated to date, is the motivation for continued research, including this project.

2.2 Smoking and lung cancer

Globally, smoking is the main risk factor for lung cancer. Tobacco smoke is a mixture of about 7,000 chemicals. Among them, at least 70 are known for their carcinogenicity according to the Center for Disease Control and Prevention (28). A systematic review performed in 2012, using data from 287 studies, including both prospective cohort and case-control studies, found that the overall association of lung cancer with smoking was strong, evident for all lung cancer types, dose-related, and insensitive to covariate-adjustment with risk ratios equal to 5.50 (95% CI: 5.07-5.96), 8.43 (95% CI: 7.63-9.31) and 4.30 (95% CI: 3.93-4.71) for ever smokers, current smokers and former smokers as compared to never smokers, respectively (29). Approximately 75% of lung cancer cases could be attributable to tobacco smoking, globally. In Canada, the prevalence of tobacco smoking has decreased dramatically over the past 50 years, from 50% in 1965 to 15% in 2018 (30, 31). However, it remains the strongest risk factor for the disease and accounted for an estimated 72% of all lung cancer cases diagnosed in 2015 (32, 33).

2.3 Occupational exposures and lung cancer

Occupational exposures are defined by the Canadian Centre for Occupational Health and Safety as exposures to chemical, physical, or biological agents that occur as a result of one’s occupation. Some of these agents may be potentially harmful and many known carcinogens are present in the occupational environment (34). Globally, it is estimated that exposure to occupational carcinogens contributes to 102,000 deaths from lung cancer annually and nearly 969,000 disability-adjusted life years (35). Of the 28 definite lung carcinogens classified by the International Agency for Research on Cancer (IARC) up until 2018, 24 are primarily found in the occupational environment (36). In Canada, it was estimated that between 3.9% and 4.2% of all incident cases of cancer were caused by occupational exposures in 2011, and that lung cancer was the most prominent type of cancer caused by these exposures (37). Further, the same authors suggested that 15% of all lung cancer cases are attributable to workplace carcinogens (37).

Many of the harmful chemical agents present in tobacco smoke can also be found in the occupational environment and are of interest to this study. Such agents include polycyclic aromatic hydrocarbons (PAHs), benzo[a]pyrene, chromium VI, beryllium, cobalt, nickel, arsenic, cadmium, lead, formaldehyde, ethylene oxide, vinyl chloride, benzene, and aromatic amines. All these substances have been classified by IARC as Class 1 (carcinogenic to humans), 2A (probably carcinogenic to humans), or 2B (possibly carcinogenic to humans) carcinogens. Table 2.1 presents, for each substance, a brief definition, their IARC classification, and the occupations or industries in which they are predominantly found.

Overall, there is a vast literature on the associations between occupational exposure to these chemical agents and increased lung cancer risk. Selected references are indicated in Table 2.1.

Table 2.1 Profile of occupational chemical agents of interest (Part one)

Occupational agent	Definition[1]	IARC classification[2]	IARC Monograph (Volume, year)	Major industries exposed to occupational agent of interest in Canada[3]	References
PAHs from any source	PAHs are chemical compounds containing only carbon and hydrogen atoms, composed in three or more aromatic rings. They are found naturally in fossil fuels or can be formed by thermal decomposition of organic material containing hydrogen and carbon. PAHs can be separated into different categories depending on the source material which gives rise to PAH exposure.	1 to 2B	32, 1983; S7, 1987	Repair of motor vehicles and motorcycles, coal mining, manufacture of miscellaneous products of petroleum and coal, non-ferrous mining, and stone quarrying.	Singh, Kamal (38) Bruske-Hohlfeld, Mohner (39) Boffetta, Jourenkova (40)
PAHs from petroleum				Repair of motor vehicles and motorcycles, coal mining, non-ferrous ore mining, stone quarrying, and other passenger land transport	
PAHs from coal				Manufacture of miscellaneous products of petroleum and coal, manufacture of structural clay products, railway transport, manufacture of glass and glass products and manufacture of cement, lime and plaster	
PAHs from wood				Logging, forestry, manufacture of wooden and cane containers and small cane ware, sawmills, planning and other wood mills, and agriculture and livestock production	
PAHs from other sources				Manufacture of rubber products not elsewhere classified, tire and tube industries, manufacture of plastic products not elsewhere classified, manufacture of textiles not elsewhere classified and manufacture of structural clay products	
Benzo[a]pyrene	Benzo[a]pyrene is a PAH that comes from certain substances when they are not incompletely burned.	1		Manufacture of miscellaneous products of petroleum and coal, manufacture of structural clay products, repair of motor vehicles and motorcycles, iron and steel basic industries and non-ferrous metal basic industries	Nadon, Siemiatycki (41)
Formaldehyde	Formaldehyde is a colorless and flammable gas (at room temperature) with a strong smell. It is obtained by the oxidation of methyl alcohol.	1	*S7*, 1987; *62*, 1995b	Barber and beauty shops, manufacture of wearing apparel, restaurants, cafés and other eating and drinking places, fur dressing and dyeing industries and photographic studios	Kwak, Paek (42)
Vinyl chloride	Vinyl chloride is a colorless, flammable gas. It must be produced industrially for its use.	1	*19*, 1979; *S7*, 1987	Barber and beauty shops, manufacture of plastic products not elsewhere classified, manufacture of textiles not elsewhere classified, manufacture of fertilizers and pesticides and electrical repair shops	Mastrangelo, Fedeli (43)
Ethylene oxide	Ethylene oxide is a flammable and colorless gas with a sweet odor. It is mainly used as a chemical intermediate in ethylene glycol manufacture.	1	*60*, 1994; *S7*, 1987	Manufacture of synthetic resins, plastic materials and man-made fibers except glass, manufacture of basic industrial chemicals except fertilizers, medical, dental and other health services, manufacture of drugs and medicines and engineering, architectural and technical services	Mikoczy, Tinnerberg (44)

[1]: According to the National Cancer Institute (https://www.cancer.gov/)

[2]: Group 1: carcinogenic to humans; Group 2A: probably carcinogenic to humans; Group 2B: possibly carcinogenic to humans (45)

[3]: According to CANJEM (http://www.canjem.ca/)

Table 2.2. Profile of occupational chemical agents of interest (Part two)

Occupational agent	Definition[1]	IARC classification[2]	IARC Monograph (Volume, year)	Major industries exposed to occupational agent of interest in Canada[3]	References
Lead	Lead is a naturally occurring element found in Earth's crust in small quantities. Mostly used in industry (car batteries, pigments, cable sheathing, etc.)	2A	*23*, 1980; *S7*, 1987; *87*	Repair of motor vehicles and motorcycles, other passenger land transport, electrical repair shops, freight transport by road and urban, suburban and inter-urban highway passenger transport	t Mannetje, Bencko (46) Beveridge, Pintos (47) Wild, Bourgkard (48)
Chromium VI	Chromium is an odorless and tasteless metal found in Earth's crust. Naturally found in air, water, soil and food. Used widely in several industries (pigments, chrome plating, etc.).	1	*S7*, 1987; *49*, 1990	Tanneries and leather finishing, manufacture of engines and turbines, manufacture of paints, varnishes and lacquers, manufacture of jewelry and related articles and manufacture of cutlery, hand tools and general hardware	
Beryllium	Beryllium is a metal found in nature. Very good conductor of electricity and heat and non-magnetic. Used in aerospace components, transistors and nuclear reactors.	1	*S7*, 1987; *58*, 1993	Manufacture of fertilizers and pesticides, manufacture of wooden and cane containers and small cane ware, manufacture of miscellaneous products of petroleum and coal, tanneries and leather finishing and inland water transport	
Cobalt	As with Nickel, Cobalt can be found in Earth's crust. Used in numerous commercial and industrial applications (alloys, etc.).	1	*52*, 1991	Manufacture of metal and wood working machinery, manufacture of cutlery, hand tools and general hardware, manufacture of paints, varnishes and laquers, manufacture of fertilizers and pesticides and manufacture of special industrial machinery and equipment except metal and wood working machinery	
Nickel	Nickel is a silvery-white metallic element found in Earth's crust. It has many industrial uses. Mostly used in metal alloys.	1	*S7*, 1987; *49*, 1990	Manufacture of engines and turbines, manufacture of jewelry and related articles, manufacture of cutlery, hand tools and general hardware, manufacture of aircraft and manufacture of structural metal products	
Arsenic	Arsenic is a naturally occurring substance in air, water and soil. It can be released into the environment by agricultural and industrial processes (mining, etc.)	1	*84*, 2004	Non-ferrous metal basic industries, agriculture and livestock production, tanneries and leather finishing, supporting services to air transport and distilling, rectifying and blending spirits	
Cadmium	Metallic element that naturally occurs in air, water, soil, and food. By-product of zinc refining and used to make batteries, plastics or alloys.	1	*S7*, 1987; *58*, 1993	Authors, music composers and other independent artists not elsewhere classified, manufacture of paints, varnishes and laquers, electrical repair shops, printing, publishing and allied industries and manufacture of jewelry and related articles	
Benzene	Benzene is a colorless or light-yellow liquid at room temperature, which is derived from coal or petroleum. It is formed from both natural processes (volcanoes and forest fires) and human activities (crude oil or gasoline).	1	*29*, 1982; *S7*, 1987	Tyre and tube industries, manufacture of footwear, except vulcanized or moulded rubber or plastic footwear, manufacture of rubber products not elsewhere classified, manufacture of products of leather and leather substitutes and repair of footwear and other leather goods	Warden, Richardson (49)
Aromatic amines	Aromatic amines represent a category of chemical agents widely used as chemical intermediates. They consist of aromatic rings attached to an amine atom.	1 to 2B	*S7*, 1987; *77*, 2000 *57*, 1993 *4*, 1974; *S7*, 1987 *1*, 1972; *S7*, 1987	Photographic studios, barber and beauty shops, manufacture of rubber products not elsewhere classified, tanneries and leather finishing, fur dressing and dyeing industries	Tomioka, Saeki (50)

[1]: According to the National Cancer Institute (https://www.cancer.gov/)

[2]: Group 1: carcinogenic to humans; Group 2A: probably carcinogenic to humans; Group 2B: possibly carcinogenic to humans (45)

[3]: According to CANJEM (http://www.canjem.ca/)

2.4 The role of DNA methylation in carcinogenesis

The term “epigenetics” includes any biological process which alters gene activity without modifying the DNA sequence (51). Research in epigenetics investigates the influence of behaviors and environmental factors on the way genes work, and their influence on numerous disease etiologies. DNA methylation is a common epigenetic mechanism in which a methyl group is covalently transferred from an electrophilic methyl donor called *S*-adenosyl-*L*-homocysteine to the 5’ end of a cytosine which then becomes 5-methylcytosine (52). More than 98% of DNA methylation takes place in Cytosine-phosphate-Guanine (CpG) dinucleotide sites in somatic cells (53). It is one of the most studied epigenetic mechanisms as it is essential to mammalian development. In fact, DNA methylation plays a major role in multiple cellular processes in humans as it regulates gene expression, either by recruiting proteins involved in gene expression or by inhibiting the binding of transcription factors to DNA (54). During a person’s life, DNA methylation patterns in the genome change in an ongoing dynamic process. As a result, differentiated cells possess unique DNA methylation patterns that regulate tissue-specific gene transcription (8).

Studies have demonstrated that dysregulation of DNA methylation patterns can contribute to diseases such as cancers (55, 56). There are two main patterns of DNA methylation: global methylation of the genome and gene-specific methylation (57). Global DNA methylation refers to the average methylation status that occurs across the whole genome and gene-specific DNA methylation refers to the methylation status of specific genes (9). Generally, global DNA hypomethylation is thought to be a characteristic of cancer progression while gene-specific methylation changes constitute early events in the development of cancers and can be used as biomarkers of exposure and/or early effect to important carcinogens (58). Gene-specific methylation typically occurs in CpG-rich regions in the gene promoters and in DNA repeat sequences (59-61). When occurring in promoter regions, hypomethylation of CpG sites can lead to the activation of those genes and to their overexpression while hypermethylation usually leads to their silencing (62, 63). Both gene-specific DNA hypomethylation and hypermethylation patterns can be found in virtually all types of cancer. The deleterious aspects of gene-specific methylation (hypo- or hyper-) depend largely on the functionality of the specific genes under study. As a matter of fact, cancer-associated DNA hypomethylation is considered to be as prevalent as cancer-associated DNA hypermethylation (60). Both mechanisms have been associated with increased risks of lung cancer (64, 65).

2.5 Use of DNA methylation as an intermediate marker between smoking and occupational exposures, and lung cancer risk

Biomarkers are cellular, biochemical or molecular alterations that are measurable in biological media such as human tissues, cells, or fluids (66). In broader terms, they include all biological characteristics that can objectively be measured and evaluated as an indicator of biological processes, either normal or pathogenic (67). Today, the application of biomarkers in research, and in the diagnosis and management of cancer is well-known (68). For instance, biomarkers are used routinely to detect cancers early, to assign prognosis, and to better orient targeted anti-cancer molecular therapies (69, 70). Additionally, in research, biomarkers may be used to quantify exposure and intermediate events which is particularly useful in the study of diseases with a long latency period like lung cancer.

In the context of this thesis, the association between environmental exposures (i.e., smoking and occupational exposure to selected chemicals) and methylation levels of the *AHRR* and *F2RL3* genes was of interest. Specifically, *AHRR* and *F2RL3* methylation was conceptualized as a biomarker of intermediate effect (i.e., as a potential mechanism of action) underlying the environment-lung cancer relationship. To achieve accurate estimates of risk in molecular epidemiologic studies, the selection of a biomarker requires additional considerations, including the validity and reliability of the biomarker measurement method; the selection of the biologic matrix used to measure the biomarker; an understanding of the intra- and inter-individual variation of the biomarker and; finally, a critical appraisal of the current literature on the determinants of the biomarker, the relationship between the exposures of interest and the biomarker (Sections 2.7 and 2.8), and the association between the biomarker and the outcome of interest (Section 2.6) (71, 72).

The use of DNA methylation as a biomarker presents several advantages. First, it is a molecular biomarker that is chemically and biologically stable (73, 74). In comparison with other molecular biomarkers such as DNA mutations, methylation patterns are easier to detect as they are binary signals and can be amplified by methylation-specific polymerase chain reaction-based techniques (75). Furthermore, DNA methylation measurements can be compared with absolute reference points which makes it possible to accurately quantify levels (76). Many methods for the quantification of DNA methylation exist and they have similarly been shown to be reliable and thus, amenable to population health research (77). DNA methylation can be measured in small quantities of peripheral blood samples which has the advantage of being less invasive than tissue samples (78). Though, DNA methylation levels can vary in a relatively short period of time (within months) in response to environmental changes (or exposures) (79). It has already been shown that if environmental exposures persist, methylation patterns can remain stable through years (80). Consequently, DNA methylation can serve as a valuable biomarker of long-term exposure to environmental agents. The known determinants of DNA methylation include age, sex, body mass index (BMI), physical activity, tobacco smoking, diet, alcohol consumption, socioeconomic status (SES), and ethnicity (Table 2.3). Hence, studies using methylation as a biomarker of interest typically consider these as potential confounding factors.

Table 2.3. Summary of established determinants of DNA methylation

Determinant	Association with methylation	Reference
Age	Increasing age is associated with gene-specific DNA hypomethylation	Salameh, Bejaoui (81) Jung and Pfeifer (82)
Sex	Women have higher levels of global DNA methylation in comparison to men.	Hall, Volkov (83) Boks, Derks (84)
BMI	Obesity is associated with lower levels of global DNA methylation	Alegría-Torres, Baccarelli (85) Mendelson, Marioni (86) Reed, Suderman (87)
Physical activity	Engagement in physical activity is associated with higher levels of global DNA methylation	Alegría-Torres, Baccarelli (85) Zhang, Cardarelli (88) White, Sandler (89)
Tobacco smoking	Tobacco smoking is associated with lower levels of gene-specific DNA methylation	Fasanelli, Baglietto (18) Lee and Pausova (90) Tsaprouni, Yang (91) Dogan, Shields (92)
Diet	A diet poor in fruits and vegetables is associated with lower levels of global DNA methylation	Alegría-Torres, Baccarelli (85) Hibler, Huang (93)
Alcohol consumption	Alcohol consumption is associated with lower levels of global DNA methylation	Zakhari (94) Varela-Rey, Woodhoo (95)
SES	SES is inversely associated with gene-specific DNA methylation levels	McDade, Ryan (96) Stringhini, Polidoro (97) Needham, Smith (98)
Ethnicity	In comparison with Whites, different ethnic groups (Blacks, Hispanics and East-Asian) showed heterogeneous global and gene-specific DNA methylation levels.	Galanter, Gignoux (99) Park, Patel (100) Zhang, Cardarelli (101)

2.6 DNA methylation of the *AHRR* and *F2RL3* genes and lung cancer

It is recognized that the methylation patterns found in tumor cells are significantly altered in comparison to normal cells. In the context of lung cancer, studies have shown that genome-wide and also gene-specific methylation patterns differ in lung tumors compared to normal adjacent tissues (102). Given the increasing evidence supporting the importance of DNA methylation in the regulation of gene expression, and that aberrant methylation patterns may serve as early events in carcinogenesis (55, 56), DNA methylation of two genes, *AHRR* and *F2RL3*, has emerged as potential intermediate markers in lung cancer etiology.

Located on the human chromosome 5, the *AHRR* gene is a known tumor suppressor gene (103). AHRR represses the transcription activity of the aryl hydrocarbon receptor (AHR) which results in a down-regulation of genes regulated by AHR (104). AHR is a ligand-activated transcription factor implicated in a signaling cascade that mediates cell growth and differentiation as well as the detoxification of environmental contaminants. AHR is also involved in other pathological processes, cellular homeostasis (105), and tumor development (16). Carcinogenic chemicals such as the ones contained in tobacco smoke (e.g., PAHs, organic agents, and metals) trigger the AHR signaling pathway by directly binding to AHR (106, 107). This leads to several downstream effects that influence tumorigenesis, inflammation, cell proliferation (108), and AHR-dependent changes in gene expression (109, 110). Therefore, repression of AHR by *AHRR* overexpression can lead to carcinogenic processes.

Located on the human chromosome 19, the *F2RL3* gene encodes for the protease-activated receptor-4 (PAR-4) which is involved in cell signaling and in the pathophysiology of chronic inflammatory diseases such as coronary heart disease (111), prostate cancer (14), and lung cancer (112). Additionally, it is directly implicated in the physiological functions of pain, inflammation, thrombosis (113), and platelet activation (111). However, current knowledge does not clearly delineate its possible links to lung carcinogenic processes induced by exposure to tobacco smoke or environmental contaminants (114).

A nested case-control study composed of 143 cases of lung cancer and 457 age- and sex-matched healthy controls found that methylation levels of the *AHRR* and *F2RL3* genes in whole blood were associated with lung cancer risk. Specifically, odds ratios (OR) for lung cancer of 15.86 (95% confidence intervals (CI): 4.18-60.17) and 10.55 (95% CI: 3.44-32.31) were reported for *AHRR* and *F2RL3,* respectively, when comparing participants in the lowest quartile (representing hypomethylation and thus, higher expression of these genes) versus the highest quartile of DNA methylation (17). These findings have been replicated in an epigenome-wide association study (EWAS) of DNA isolated from pre-diagnostic blood samples from 132 case-control pairs in the

Norwegian Women and Cancer Study cohort that similarly reported significant associations between hypomethylation of CpG sites in the *AHRR* and *F2RL3* genes, and increased lung cancer odds after adjustment for smoking and blood cell composition (18). Consequently, strong emerging evidence supports the role of lower methylation in the *AHRR* and *F2RL3* genes in lung cancer etiology.

2.7 Smoking and methylation levels of the *AHRR* and *F2RL3* genes

Given the evidence supporting the associations between DNA methylation in the *AHRR* and *F2RL3* genes and lung cancer risk, epidemiologic studies have additionally examined aberrant methylation in these two genes in relation to smoking, the strongest risk factor for lung cancer.

An EWAS comparing current, former, and never smokers from participants of the population-based KORA F4 panel observed that DNA methylation levels of the *AHRR* and *F2RL3* genes, measured using genomic DNA from whole blood, were significantly lower in smokers compared to former and never smokers (115). Another study based on two EWAS nested in the European Prospective Investigation into Cancer and Nutrition cohort using peripheral blood DNA similarly observed that several CpG sites in both genes were hypomethylated in current smokers compared with former and non-smokers (114). These associations have been replicated by several other large case-control studies (18, 116), and prospective cohort studies (117, 118). All these studies estimated a percent difference in methylation among current smokers versus never smokers between -7% and -22%, and -8% and -18%, for *AHRR* and *F2RL3*, respectively.

Despite the consistency of these reported associations, certain methodological gaps remain. Indeed, all previous studies conceptualized smoking based only on status, therefore excluding several important dimensions of smoking including duration, intensity of smoking, and time since cessation (if applicable). More importantly, all these studies focused on the measurement of relatively few CpG sites in the promoter region of the *AHRR* and *F2RL3* genes. Specifically, for

AHRR, studies primarily focused on the CpG site cg05575921 and its surrounding region (i.e., ranging from two to 11 CpG sites within a 35-base pair (bp) distance). For *F2RL3*, studies mostly investigated the CpG site cg03636183 and one to three CpG sites within a 20-bp surrounding distance (18, 114-118). Increasingly, regional methylation patterns, as represented by the measurement of multiple CpG sites within the promoter region, are thought to better approximate gene expression levels (119-121). Thus, the measurement of additional CpG sites in the promoter regions of the *AHRR* and *F2RL3* genes may both reinforce the strong associations previously observed and shed light on additional regions of interest for future study.

2.8 Occupational exposures and methylation levels of the *AHRR* and *F2RL3* genes

Even though many chemical agents present in tobacco smoke can also be found in the occupational environment, very few studies have examined the association between occupational exposures and methylation levels of the *AHRR* and *F2RL3* genes.

A recent study compared *AHRR* and *F2RL3* methylation levels among 151 Swedish male chimney sweeps and creosote-exposed workers who were occupationally exposed to PAHs to 152 controls not occupationally exposed to PAHs. Lower *AHRR* methylation level was found among PAH-exposed workers (i.e., both chimney sweeps and creosote-exposed workers) as compared to controls; while, only creosote-exposed workers had lower *F2RL3* methylation than controls (122). However, the limits of this study included a modest sample size (303 participants) and the lack of consideration for concurrent exposure to other chemical hazards. Similar to the studies on smoking, this study quantified DNA methylation levels of three CpG sites within *AHRR* (cg05575921 and two sites within a 25-bp proximity) and two CpG sites within *F2RL3* (cg03636183 and one site within an 11-bp proximity). To the best of our knowledge, no further studies have investigated the association between occupational exposures to agents of interest in this study and DNA methylation of the *AHRR* and *F2RL3* genes.

2.9 Relevance of the study

The association between smoking and our selected occupational exposures, and increased lung cancer risks are well-accepted today. However, the mechanisms underlying these associations remain unclear. This project focused on DNA methylation which is a key epigenetic mechanism that has been extensively studied and implicated in lung cancer etiology. Specifically, the methylation levels of two genes, *AHRR* and *F2RL3*, involved in multiple physiological mechanisms including lung cancer, were of interest. Previous studies support the role of aberrant methylation levels in these two genes and increased lung cancer risks. In addition, there is emerging evidence that smoking and occupational exposures to certain agents can influence *AHRR* and *F2RL3* methylation patterns. This study improves upon the previous literature on the association between smoking and occupational exposures, and *AHRR* and *F2RL3* methylation, by measuring substantially more CpG sites located in the promoter regions of each gene. This will allow for better capture of regional methylation patterns in each gene which should arguably be more representative of gene expression. Furthermore, this study proposed to consider several dimensions of smoking that have been previously overlooked in the literature in the context of DNA methylation. Additionally, this study is among the first to examine the association between occupational exposures to multiple agents and gene-specific methylation patterns in the *AHRR* and *F2RL3* genes. Altogether, this study strives to provide a better understanding of the impact of common environmental exposures on lung cancer risk via epigenetic mechanisms.

Chapter 3. OBJECTIVES

The objectives of this thesis are as follows:

1. To investigate and quantify the association between smoking and DNA methylation levels in the *AHRR* and *F2RL3* genes.
2. To investigate and quantify the association between selected occupational exposures and DNA methylation levels in the *AHRR* and *F2RL3* genes.

It was hypothesized that exposure to chemical agents present in tobacco smoke and in certain occupations will be associated with hypomethylation in the *AHRR* and *F2RL3* genes.

Chapter 4. METHODS

4.1 Study design and population

We conducted a cross-sectional analysis of smoking history and occupational exposures on DNA methylation levels in the *AHRR* and *F2RL3* genes using data available from a nested case-control study. Specifically, an ongoing cumulative incidence case-control study was nested in CARTaGENE (co-directed by Dr. Vikki Ho and Dr. Anita Koushik) to investigate the relationship between DNA methylation in the *AHRR* and *F2RL3* genes, and lung cancer risk. Started in 2009, CARTaGENE is Quebec's largest ongoing prospective cohort study of 43,046 Quebec residents between 40 and 69 years of age living in Saguenay, Montreal, Quebec City, Trois-Rivières, Gatineau, and Sherbrooke. The goal of CARTaGENE is to investigate modifiable environmental and lifestyle factors and the genomic determinants of chronic diseases. Potential participants were identified via random selection of consenting individuals from provincial health insurance registries-FIPA files (fichier administrative des inscriptions des personnes assurées de la Régie de l'assurance maladie du Québec (RAMQ)). They were recruited in two phases: phase A in 2009 and phase B in 2012 (123).

Given the rich data available in CARTaGENE on smoking and occupational history, and that the DNA methylation measures were being quantified in the ongoing nested study, there was an opportunity to investigate the cross-sectional association between smoking and occupational exposures, and *AHRR* and *F2RL3* methylation levels. In this ongoing nested case-control study, the case group included all CARTaGENE participants with an incident diagnosis of lung cancer during the follow-up period (from baseline to 2016), and who donated a blood sample at baseline (N=200). Incident cases were identified via the linkage of CARTaGENE participants with the RAMQ and the Québec cancer registry. The control group included individuals who had not developed lung cancer by 2016, and who had DNA isolated from their blood sample donated at baseline (N=400). Controls were randomly selected from the CARTaGENE based on a ratio of 2:1, and were frequency-matched to cases based on age (5-year age groups), sex, and phase of blood sampling. The participants of this nested case-control study constituted the study population of this thesis.

4.2 Quantification of DNA methylation as the outcome of interest

Isolation of DNA from baseline blood samples was conducted at Biobanque Génome Québec (Chicoutimi) and stored at -80°C. Quantification of DNA methylation in the *AHRR* and *F2RL3* genes was conducted at CHU Sainte-Justine and Genome Quebec Integrated Centre for Pediatric Clinical Genomics. Bisulfite conversion treatment was performed on 1µg of DNA from each participant using the EZ DNA Methylation-Gold kit from ZymoResearch. The purpose of bisulfite conversion is to deaminate unmethylated cytosine to uracil. This process leads to a primary DNA sequence change that allows the differentiation of methylated cytosines to unmethylated ones. DNA methylation levels of the two genes were quantified using the Sequenom EpiTYPER® technology which uses base-specific cleavage and laser desorption/ionization-time of flight mass spectrometry (124-126).

Primers (i.e., short single-stranded DNA sequences used in polymerase chain reaction) were designed for the two genes of interest. The genomic region of interest for *AHRR* is located on the Genome Reference Consortium Human genome build 37 (GRCg37), or its equivalent Human Genome version 19 (hg19), more commonly referred to as GRCh37/hg19. The region spans 33599 base pairs from chr5:367471 to 401070 on the positive DNA strand. The region of interest for *F2RL3* spans 4946 base pairs on GRCh37/hg19 from chr19:16999071 to 17004017 on the DNA negative strand. For each gene, primers were chosen from the promoter region based on the findings of four prospective cohort studies (18). Specifically, they were chosen in proximity to CpG islands (i.e., regions of the genome that contain a large number of CpG dinucleotide repeats with a CG:GC ratio of more than 0.6) or to CpG island shores (i.e., 2-kb-regions that lie on both sides of a CpG island), to transcription factor binding sites, to DNAse hypersensitive sites, and to H3K27Ac marks suggestive of the presence of an active regulatory domain within each gene (i.e., epigenetic modifications that indicate acetylation of the lysine residue of the histone H3 protein, UCSC Genome Browser, http://genome.ucsc.edu/).

The data-cleaning strategies for the methylation measures were performed as described in Ho and *al.* (127). Regions of interest for *AHRR* and *F2RL3* were measured in six and one DNA fragments, respectively. For each fragment, 25 ng of bisulfite-converted DNA was utilized to quantify methylation ratios within CpG units (a unit consists of either an individual CpG site or aggregates of multiple CpG sites) located within each fragment. Methylation ratios (i.e., the percentage of methylated cytosines at a specific CpG site of a gene divided by the total number of copies of that CpG site in the sample) were calculated. For CpG units that consisted of multiple CpG sites, the methylation ratio of that CpG unit was assigned to each of the CpG sites within that unit. Seven 96-well plates containing participant samples were run per fragment. For quality control, two high-methylated human DNA controls manufactured by EpigenDx were included on each plate. Seventy-nine CpG sites within the promoter regions of the *AHRR* and *F2RL3* genes were measured: 72 CpG sites for the *AHRR* gene and seven CpG sites for the *F2RL3* gene. According to Sequenom, unmeasured sites, sites with high or low mass, sites with more than one overlapping silent peak, and duplicated sites were deemed unreliable for statistical analyses because methylation signals could not be assigned uniquely to one CpG unit or they could not be obtained. Based on these criteria, 30 CpG sites from the *AHRR* gene were excluded. Furthermore, three measured CpG sites from the *AHRR* gene with more than 25% missing methylation ratios among participants were also excluded to minimize the impact of missing data. Additionally, six measured CpG sites of the *AHRR* gene with methylation ratios that had a standard deviation inferior or equal to 0.02 (i.e., 20% methylation) were also excluded to ensure that only methylation ratios with meaningful differences were considered for analysis. All in all, out of the 79 CpG sites initially measured, only 40 were retained as informative for analysis in the study: 33 CpG sites from the *AHRR* gene and seven from the *F2RL3* gene. Next, 69 participants with more than 10% missing methylation ratios were excluded from the study in order to retain participants with only a small proportion of missing values imputed. In the end, 531 participants were retained in the study (179 cases and 352 controls). Figure 5.1 summarizes the data cleaning and exclusion processes based on the DNA methylation measures. All remaining missing values were assigned the mean methylation ratio for each CpG site. The reliability of the methylation measurements was assessed using two high-methylated quality control samples that were included on each plate. A coefficient of variation (CV) of 4.65% and 4.16% was estimated between plates and between fragments, respectively.

The main outcomes of interest were average methylation levels of the *AHRR* and *F2RL3* genes. Measures representing average methylation levels across all informative CpG sites (i.e., 33 *AHRR* CpG sites and seven *F2RL3* CpG sites) were calculated and parametrized as a continuous variable, with values between 0 and 1, for each gene separately.

4.3 Assessment of smoking

Smoking was one of the two main exposures of interest under investigation in this study. In CARTaGENE, questionnaires were used to collect information on a variety of factors including demographics, lifestyle behaviors, personal health information, and occupational history (Appendix I). Data collected on smoking history included current smoking status and when applicable (i.e., only for participants having smoked at least 100 cigarettes in their lifetime), age at initiation and cessation, and the average number of cigarettes smoked per week. To incorporate all these metrics related to smoking, a cumulative smoking index (CSI) developed by Hoffmann et al. (128) was derived for each participant:

$$\text{CSI} = (1\text{-}0.5^{dur/\tau})\,(0.5^{tsc/\tau})\,\ln\,(int + 1)$$

Where *dur* is the duration of smoking, *tsc* is the time since cessation, τ is the biological half-life of tobacco carcinogens (129) and *int* the average daily amount smoked in cigarettes. This index provides a reliable mathematical and continuous representation of the participants' smoking history and habits by including several aspects of smoking behavior into one parsimonious measure. In the main analyses, the CSI was parameterized as a standardized continuous variable. Of the 531 participants with methylation information retained in the study, 24 had missing information that prevented CSI calculation and were thus excluded from the main CSI-methylation analysis.

Additional analyses were conducted with participants categorized according to their smoking status at baseline to facilitate comparison with previous studies: never smokers (i.e., participants that had smoked less than 100 cigarettes in their lifetime), former smokers (i.e., participants that had stopped smoking at least 30 days before baseline), or current smokers (including participants that had stopped smoking in less than 30 days before baseline). Only one participant was excluded from the smoking status analysis due to missing information. There were 187 never-smokers, 237 former smokers, and 106 current smokers at baseline. A flowchart of the study participant exclusion schema based on CSI and smoking status is presented in Figure 5.2.

4.4 Assessment of occupational exposures

4.4.1 Occupational information

The second main exposure of interest was occupational exposures. At baseline, all CARTaGENE participants provided information on their longest-held job and current job including job name/title, job industry, and age at which the job started and ended through a self-administered questionnaire (Appendix II) that was mailed to them. Participants from phase A were re-contacted in 2011 and 2012 to complete a follow-up survey including information regarding their entire job history assessed through open-ended questions (Appendix III). In this thesis, information on the longest-held job was used to represent occupational exposures for all participants. In our study, study participants held their longest-held job, on average, for 23 years. All occupations were coded by an occupational hygienist according to the International Standard Classification of Occupations 1968 (ISCO-68) which is one of the main classification systems created by the International Labour Organization to categorize jobs into a set of groups according to the tasks and duties performed (130).

4.4.2 Methods to estimate retrospective occupational exposures

There are different approaches to retrospective occupational exposure assessment in population-based studies. Over the last 40 years, the most frequently used methods include self-assessment, expert exposure assessment, and job-exposure matrices, or JEMs (131). The expert-assessment approach is considered the gold-standard method, with high levels of reliability for retrospective exposure assessment (132). On top of that, it prevents errors involved with self-reported exposure (132, 133). However, expert assessment is not always feasible since it is a significantly more expensive undertaking (134). Consequently, many have advocated the use of JEMs for occupational exposure assessment. As a matter of fact, some studies suggest that JEMs are similar in reliability when compared with experts assessments (135, 136). Briefly, a JEM is a cross-tabulation of different levels of exposure to different agents for selected occupation titles with exposure information (137). JEMs can be constructed based on measurements, observations, experts assessments, self-assessment, or a mix of any of these approaches (138).

4.4.3 Use of CANJEM to estimate retrospective occupational exposures

4.4.3.1 Agent selection

To estimate occupational exposures among CARTaGENE participants, the Canadian Job Exposure Matrix was used (CANJEM). CANJEM provides information on the probability, frequency, and intensity of exposure to a list of 258 occupational exposures for a given job and time period. CANJEM was constructed using data from four case-control studies of various cancers conducted in the greater Montreal area from 1985 to 2004. From these studies, 31,673 jobs held from 1930 to 2005 by 8,912 subjects were evaluated by experts and occupational exposure to a list of over 258 agents was assigned by a team of experts according to the tasks, processes, and work environment (139). Essentially, CANJEM consists of three dimensions: the time period, the occupational/industrial classification, and the chemical agent of interest. Depending on the available information about study participants and the scope of the study, each of those dimensions can be specified accordingly to obtain occupational exposure estimates via CANJEM.

To select agents for this thesis, we prioritized CANJEM agents that were present in tobacco smoke and evaluated by IARC for carcinogenic potential (carcinogenic to humans, probably carcinogenic to humans, and possibly carcinogenic to humans). Overall, sixty-two chemical agents present in cigarette smoke have been classified as either carcinogenic (Group 1), probably carcinogenic (Group 2A), or possibly carcinogenic (Group 2B) to humans by IARC (140). Among these, eighteen were present in CANJEM's database: polycyclic aromatic hydrocarbons (PAHs) from any source, PAHs from petroleum, PAHs from wood, PAHs from other sources, PAHs from coal, benzo[a]pyrene, chromium VI, beryllium, cobalt, nickel, arsenic, cadmium, lead, formaldehyde, ethylene oxide, vinyl chloride, benzene, and aromatic amines. Table 3.1 presents the chemical agents retained according to their CANJEM chemical category.

Table 4.1 Chemical agents of interest in this study

CANJEM chemical category	Occupational agent
Organic liquids and vapors	PAHs from any source, PAHs from petroleum, PAHs from wood, PAHs from coal, PAHs from other sources, benzo[a]pyrene
Metallic compounds	Lead, chromium VI, beryllium, cobalt, nickel, arsenic, cadmium
Organic gases	Formaldehyde
Other organic gases	Vinyl chloride, ethylene oxide
Organic solvents	Benzene
Other organic products	Aromatic amines

4.4.3.2 Linkage of participants with CANJEM

Job classification systems like ISCO have different resolutions; ISCO codes can vary from a 2-digit resolution (broad) to a 5-digit resolution (specific). For instance, an ISCO code of 0-1 is attributed to “Physical scientists and related technicians” which is a broad category while an ISCO code of 0-11.20 is attributed to “Organic chemists” which is more specific. CANJEM extraction was performed with a time period ranging from 1950 to 2005 as it covers the years our study participants have worked. For all study participants, the job code pertaining to the longest-held job was linked to CANJEM using the highest resolution possible. In the case where a job code could not be linked with a 5-digit ISCO code resolution, subsequent linkage at a lower ISCO resolution was performed. Moreover, if participants had information on both their longest-held job and their lifetime occupational history (i.e., phase A participants), priority was given to the longest-held job code derived from the full job history as more details were provided for job coding. Fifty-four participants (10.2%) could not be linked to CANJEM either due to missing information on their longest-held job or because they had never worked. They were therefore excluded from our occupational exposure analyses. Overall, from the 531 participants with methylation information available, 89.8% were linked to CANJEM, yielding 477 participants in our occupational exposure-methylation analyses. Among them, 76.3% (n=364) were linked at a 5-digit resolution, 19.9% (n=95) at a 3-digit resolution and 3.8% (n=18) at a 2-digit resolution. Sixty-four % of the study participants that were linked to CANJEM had a full job history from which the longest-held job code was derived. The linkage of the participants to CANJEM is represented in Figure 5.2.

Once the linkage to the longest-held job of CARTaGENE participants was performed, CANJEM provided the probability, frequency, and intensity of exposure to each of the 18 chemical agents retained. In the context of this study, participants were considered "exposed" to a chemical agent in their longest-held job when their probability of exposure (i.e., percentage of jobs for which the agent was assessed to be present) to that agent was equal to or greater than 25%. Occupational exposures were parameterized in two ways. In our main analysis, occupational exposure to any of the 18 retained chemical agents was parameterized using a summary variable (3 categories): "Unexposed", "Exposed to one agent" and "Exposed to two or more agents". Additionally, individual exposure to the most prevalent agents (PAHs from all sources, PAHs from petroleum, lead, and formaldehyde) in our study population (prevalence of exposure ≥5%) was parametrized into two categories: "Unexposed" and "Exposed".

4.5 Additional Covariates

A comprehensive literature review on predictors of DNA methylation levels was additionally performed on the Ovid platform, using the MEDLINE and EMBASE databases. Based on multiple studies (refer to Table 2.3), we identified several predictors of DNA methylation including body mass index (BMI), physical activity, fruit and vegetable consumption, alcohol consumption, socioeconomic status (SES), and ethnicity (85, 96, 99).

Directed acyclic graphs (DAG) were constructed for each exposure separately (i.e., smoking and occupational exposures) to assess confounding and determine minimally sufficient adjustment sets for estimating the total effect of each main exposure on *AHRR* and *F2RL3* methylation. For both exposures, the minimally adjusted models included the frequency-matching variables (age, sex, phase of blood sampling). For smoking (Figure 5.3), the fully adjusted regression model included sex, SES, ethnicity as well as age, and phase of blood sampling (frequency-matching factors). For occupational exposures (Figure 5.4), the fully adjusted regression model included education level and the frequency-matching factors. In the occupational analysis, smoking was also added to the fully adjusted model as it is a strong risk factor for *AHRR* and *F2RL3* hypomethylation. Due to the information available, household annual income and education levels were used as surrogates of SES.

4.6 Statistical analysis

Least squares regression was used to quantify the relationship between smoking and occupational exposures, and DNA methylation in the *AHRR* and *F2RL3* genes. Regression coefficients and accompanying 95% confidence intervals were estimated. Analyses were conducted in the total population, then stratified by case-control status and by sex to determine if case-control status and/or sex had an influence on the results obtained.

In this thesis, several sensitivity analyses were carried out. Analyses considering methylation levels in individual CpG sites instead of the average methylation level for each gene were conducted in order to facilitate comparison with the literature. Next, occupational analysis using the participants' current employment was conducted to assess the robustness of the main analyses using the longest-held job information. Moreover, analyses excluding participants with more than 10% missing methylation measures for either the *AHRR* or *F2RL3* gene (i.e., yielding different sample sizes for each gene) were also performed to assess the influence of this exclusion criteria.

All statistical analyses were performed on R (version 1.3.109). The assumptions of the linearity between each exposure and the outcome, the homoscedasticity of the data as well as the normal distribution of errors in each model were graphically verified using R to ensure the validity of the linear regression models.

Chapter 5. MANUSCRIPT

This manuscript was written in accordance with the instructions for authors provided by the Mutation Research – Genetic Toxicology and Environmental Mutagenesis, a peer-reviewed journal.

Title:

The influence of smoking and occupational risk factors on DNA methylation in the *AHRR* and *F2RL3* genes

Authors:

Michael Pham[1,2], Alice Nguyen[1,2], Romain Pasquet[2], Laura Pelland-St-Pierre[1], Sherryl Taylor[3], Delphine Bosson-Rieutort[4], Jack Siemiatycki[1,2], Anita Koushik[1, 2], Vikki Ho[1, 2]

AFFILIATIONS

1. Department of Social and Preventive Medicine, University of Montreal, Montreal, Quebec, Canada
2. Health Innovation and Evaluation Hub, University of Montreal Hospital Research Centre (CRCHUM), Montreal, Quebec, Canada.
3. Department of Medical Genetics, Faculty of Medicine and Dentistry, University of Alberta, Edmonton, Alberta, Canada
4. Department of Health Management, Evaluation and Policy, University of Montreal, Montreal, Quebec, Canada

Correspondence to:
Dr. Vikki Ho
Université de Montréal Hospital Research Centre (CRCHUM),
850 Saint-Denis Street, 3rd Floor, S03-424
Montreal, Quebec H2X 0A9, Canada
E-mail: vikki.ho@umontreal.ca
Tel: 514-890-8000 ext. 31522
Fax: 514-412-7018

Abstract word count: 304
Manuscript word count: 4998 (without references and tables)
Tables: 10
Figures: 4

ABSTRACT

Objective: To determine the association of smoking and occupational exposures with DNA methylation levels in the lung cancer-related genes, *AHRR* and *F2RL3*.

Methods: Using data from a nested case-control study in CARTaGENE, the largest ongoing prospective cohort study of 43,046 people in Quebec, Canada, we examined the association between *AHRR* and *F2RL3* gene methylation and lung cancer risk in 200 cases and 400 controls. Information on participants' smoking behavior and longest-held occupation was obtained from questionnaires. Information on smoking status and, where applicable, the average number of cigarettes smoked, duration of smoking, and time since cessation, was parameterized into a cumulative smoking index (CSI, continuous). Occupational exposures were estimated using the Canadian Job Exposure Matrix. Eighteen agents present in the occupational environment that are also found in cigarette smoke were of interest. In DNA isolated from blood samples collected at baseline, methylation ratios of 40 CpG sites in the *AHRR* and *F2RL3* genes were measured using the Sequenom Epityper. In each gene, average methylation levels across all CpG sites were calculated and parametrized as a continuous variable. Separate least squares regression models were used to estimate the associations of smoking and occupational exposures with *AHRR* and *F2RL3* methylation levels while adjusting for potential confounders identified from directed acyclic graphs.

Results: In both genes, smoking was associated with lower average methylation levels after adjusting for confounding factors (*AHRR*: -0.014 per standard deviation increase in CSI, 95% CI: -0.019, -0.010; *F2RL3*: -0.019 per standard deviation increase in CSI, 95% CI: -0.025, -0.012). No association was found between occupational exposures and average DNA methylation levels in the two genes.

Conclusion: Our findings support the hypothesis that tobacco smoking is associated with DNA hypomethylation of the *AHRR* and *F2RL3* genes.

Keywords: DNA methylation; epigenetics; smoking; occupational exposures; job exposure matrix

HIGHLIGHTS

- AHRR and F2RL3 are respectively involved in cell proliferation and differentiation, and platelet activation. Lower methylation levels in two to eleven CpG sites of the *AHRR* and *F2RL3* genes have been associated with an increased risk of lung cancer.
- Smoking was associated with lower average DNA methylation levels of 33 and seven CpG sites in the *AHRR* and *F2RL3* genes, respectively.
- No significant association was found between the selected occupational exposures and methylation levels of the *AHRR* and *F2RL3* genes.

5.1 Introduction

Globally, lung cancer is the second most diagnosed cancer behind breast cancer and has the highest mortality rate among all cancers (1). Smoking is the main known risk factor, but other factors must play a role since lung cancer also occurs among individuals who have never smoked (2). Apart from smoking, several established lung carcinogens are widely found in the occupational environment. Around 60% of the North American population over 15 years of age work (3) and many encounter hazards in the workplace that have been linked to increased lung cancer risks (4-7). However, the mechanisms underlying the impact of smoking and these occupational hazards on lung cancer etiology remain poorly understood.

DNA methylation is an epigenetic mechanism that involves the transfer of a methyl group to cytosine residues in cytosine-phosphate-guanine (CpG) dinucleotide pairs (8). While it is involved in the regulation of normal cellular processes, evidence for the association between environmental exposures, aberrant methylation patterns, and increased lung cancer risk has been accumulating (9). It is possible to distinguish between global DNA methylation which refers to the average methylation status that occurs across the whole genome and gene-specific DNA methylation which refers to the analysis of the methylation status of specific genes (10). DNA hypomethylation refers to the unmethylated state of most CpG sites in a given sequence that is normally methylated (11).

Notably, DNA hypomethylation of the aryl hydrocarbon receptor repressor (*AHRR*) and the coagulation factor II receptor-like 3 genes (*F2RL3*) has been associated with lung cancer risk (12-16).

The *AHRR* and the *F2RL3* genes encode proteins that are involved in many biological mechanisms such as cell proliferation and differentiation, and platelet activation, respectively (17, 18).

Previous studies have also investigated the association between common environmental exposures and methylation of the *AHRR* and *F2RL3* genes. Given that smoking is an established risk factor for lung cancer, previous studies have reported that smokers and former smokers had lower methylation levels in the two genes when compared to never-smokers (14-16, 19-21). Another study has reported an association between occupational exposures to certain polycyclic aromatic hydrocarbons (PAHs) and hypomethylation of the *AHRR* and *F2RL3* genes (22). However, all these previous studies measured fewer than eleven CpG sites for the *AHRR* gene and fewer than three CpG sites for the *F2RL3* gene. It is posited that the measurement of substantially more CpG sites would allow for better capture of regional methylation patterns which arguably could be more representative of gene expression.

This study aimed at investigating the association between smoking and DNA methylation of the *AHRR* and *F2RL3* genes. Further, given that certain workers are exposed to many of the same carcinogens present in cigarette smoke due to the nature of their occupation, this study additionally examined selected occupational exposures in relation to *AHRR* and *F2RL3* methylation levels.

5.2 Material and Methods

5.2.1 Study population

A cumulative incidence nested case-control study investigating the association between DNA methylation in the *AHRR* and *F2RL3* genes, and lung cancer risk was nested within the CARTaGENE study. Started in 2009, CARTaGENE is Quebec's largest ongoing prospective cohort study of 43,046 Quebec residents between 40 and 69 years of age living in Saguenay, Montreal, Quebec City, Trois-Rivières, Gatineau and Sherbrooke. The goal of CARTaGENE is to investigate modifiable environmental and lifestyle factors, and the genomic determinants of chronic diseases. Participants consisted of a random selection of consenting individuals identified from provincial health insurance registries-FIPA files (fichier administrative des inscriptions des personnes assurées de la Régie de l'assurance maladie du Québec (RAMQ)). They were recruited in two phases: phase A in 2009 and phase B in 2012 (23).

In this nested case-control study, the case group included all CARTaGENE participants with an incident diagnosis of lung cancer (identified via linkage of participants with the RAMQ and the Québec cancer registry) during the follow-up period (from baseline to 2016), and who donated a blood sample at baseline (N=200). The control group included individuals who had not developed lung cancer by 2016, and who had DNA isolated from a blood sample donated at baseline (N=400). Controls were randomly selected from the CARTaGENE cohort in 2016 based on a ratio of 2:1 and were frequency-matched to cases based on age (5-year age groups), sex, and phase of blood sampling. This present study used the available information from this ongoing nested case-control study to examine the cross-sectional association between smoking and selected occupational exposures in relation to DNA methylation of the *AHRR* and *F2RL3* genes.

5.2.2 Quantification of DNA methylation

DNA isolation from peripheral blood samples was conducted at the Biobanque Genome Quebec. DNA methylation measurements of the *AHRR* and *F2RL3* genes were conducted at the CHU Sainte-Justine and Genome Quebec Integrated Centre for Pediatric Clinical Genomics. Bisulfite conversion treatment was performed on 1µg of DNA from each participant, in order to deaminate unmethylated cytosines to uracil, using the EZ DNA Methylation-Gold kit from ZymoResearch. DNA methylation levels of the two genes were quantified using the Sequenom EpiTYPER® technology which uses base-specific cleavage and laser desorption/ionization-time of flight mass spectrometry (21-23).

PCR primers were designed for the two genes of interest. Specifically, the genomic region of interest for *AHRR* spanned 33599 base pairs (GRCh37/hg19: chr5:367471-401070, positive strand) and was chosen from the promoter region based on the findings of Fasanelli and al. (15), in proximity to CpG islands or CpG island shores, transcription factor binding sites, DNAse hypersensitive sites, and H3K27Ac marks (UCSC Genome Browser, http://genome.ucsc.edu/) suggestive of the presence of an active regulatory domain within each gene. The region of interest for *F2RL3* was chosen similarly, spanned 4946 base pairs (GRCh37/hg19: chr19:16999071-17004017, negative strand), and included coverage of the promoter region and CpG island; additionally, CpG sites, located apart from the island approximately 1000 base pairs from the promoter region, were also targeted.

The regions of interest for *AHRR* and *F2RL3* were analyzed in six and one DNA fragments, respectively. For each of the fragments, 25 ng of bisulfite-converted DNA was used to quantify methylation ratios within CpG units (a unit consisting of either an individual CpG site or aggregates of multiple CpG sites) located within each fragment. A methylation ratio equals the percentage of methylated cytosines at a specific CpG site of a gene, divided by the total number of copies of that CpG site in the sample. For CpG units that consisted of multiple CpG sites, the methylation ratio of that CpG unit was assigned to each of the CpG sites within that unit. Seven 96-well plates containing participant samples were run per fragment and for quality control, two high-methylated human DNA controls manufactured by EpigenDx were included on each plate. Reliability was assessed using two high-methylated quality control samples included on each plate; a coefficient of variation (CV) of 4.65% and 4.16% was estimated between plates and between fragments, respectively.

5.2.3 Data cleaning for DNA methylation

Raw methylation data processing was conducted as described by Ho *et al* (24). First, unreliable methylation ratios for CpG sites with high or low mass, with more than one overlapping silent peak, and/or duplicated sites were excluded from statistical analyses. In total, 79 CpG sites within the promoter regions of the two genes were measured: 72 CpG sites for the *AHRR* gene and seven CpG sites for the *F2RL3* gene. As shown in Figure 5.1, 30 CpG sites with unreliable methylation ratios in the *AHRR* gene were excluded, followed by the exclusion of three additional CpG sites with more than 25% of participants with missing methylation ratios. CpG sites were subsequently defined as informative if they had methylation ratios with standard deviations >0.02 (i.e., 20% methylation); methylation ratios from six CpG sites of the *AHRR* gene with standard deviation inferior or equal to 0.02 were further excluded. This resulted in 33 and seven informative CpG sites for the *AHRR* and *F2RL3* genes, respectively. All remaining missing values were assigned the mean methylation ratio for each CpG site. To ensure that each participant had only a small proportion of missing values imputed, participants with >10% methylation data missing for the gene *AHRR* and *F2RL3* were excluded. Based on this cut-off, a total of 531 participants were retained in this study. Measures representing average methylation levels across all informative CpG sites were then calculated and parametrized as a continuous variable for each gene, separately.

5.2.4 Assessment of smoking

In CARTaGENE, questionnaires were used to collect baseline information on a variety of factors including demographics, lifestyle behaviors, personal health information, and occupational history. Smoking history included current smoking status and when applicable (i.e., only for participants having smoked at least 100 cigarettes in their lifetime), age at initiation and cessation (if applicable), and the average number of cigarettes smoked per week. Smoking was parameterized using the standardized cumulative smoking index (CSI) developed by Hoffmann *et al*., according to the following equation (25):

$$\mathrm{CSI} = (1\text{-}0.5\mathrm{dur}/\tau)\ (0.5\mathrm{tsc}/\tau)\ \ln\ (int + 1)$$

Where *dur* is the duration of smoking, *tsc* is the time since cessation, τ is the biological half-life of tobacco carcinogens (20) and *int* is the average daily amount smoked in cigarettes. This index provides a reliable mathematical and continuous representation of the participants' smoking history and habits by including several aspects of smoking behavior into one parsimonious measure. Of the 531 participants with methylation information retained in the study, 24 had missing information about their smoking history that prevented CSI calculation and were thus excluded from the main smoking-methylation analysis (Figure 5.2). Additional analyses were conducted with participants categorized according to their smoking status at baseline to facilitate comparison with previous studies: never smokers (i.e., participants that had smoked less than 100 cigarettes in their lifetime), former smokers (i.e., participants that had stopped smoking at least 30 days before baseline), or current smokers (including participants that had stopped smoking less than 30 days before baseline). Only one participant was excluded from the analysis based on smoking status (Figure 5.2).

5.2.5 Assessment of occupational exposures

All CARTAGENE participants provided information about their current and longest-held job including job name/title, industry (where applicable), and age at which the job started and ended. Further, phase A participants were re-contacted in 2011 and 2012 to complete a follow-up survey including information regarding their entire job history assessed through open-ended questions. All occupations were coded by an occupational hygienist according to the International Standard Classification of Occupations 1968 (ISCO-68). In this study, we used information on the longest-held job to estimate occupational exposures for all participants. Further sensitivity analyses considering participants' current occupation at baseline were performed as well.

To estimate occupational exposures, the Canadian Job Exposure Matrix (CANJEM) was used. CANJEM is a job exposure matrix that provides information on the probability, frequency, and intensity of exposure to a list of 258 occupational exposures for a given job code and time period (26). We selected agents present in tobacco smoke, categorized by the International Agency for Research on Cancer (IARC) as either carcinogenic (Group 1), probably carcinogenic (Group 2A), or possibly carcinogenic to humans (Group 2B) (27), and that are present in the CANJEM database. Agents of interest included: formaldehyde, ethylene oxide, vinyl chloride, benzene, chromium VI, beryllium, cobalt, nickel, arsenic, cadmium, lead (metallic compound), PAHs from any source, PAHs from petroleum, PAHs from wood, PAHs from other sources, PAHs from coal, benzo[a]pyrene and aromatic amines.

Extraction of exposure information from CANJEM was performed using the time period ranging from 1950 to 2005 providing coverage of the years the subjects worked. ISCO job codes vary in resolution from 2-digit codes (broadest) to 5-digit codes (most precise) (28). For all subjects, the job code pertaining to the longest-held job was linked to CANJEM using the highest possible resolution of the occupation code. Where a job code could not be linked at the 5-digit resolution, linkage at a lower resolution was performed. Among phase A participants who had information on both their longest-held job and their lifetime occupational history, priority was given to the longest-held job determined from the full job history as more details were provided for job coding. From the 531 participants with methylation information retained in the study, 54 participants could not be linked to CANJEM because of missing job information or an unlinkable job code, and therefore were excluded from the occupational exposures-methylation analysis. As illustrated in Figure 5.2, among the 477 participants that were linked to CANJEM, over 63% had a full job history from which the longest-held job code was derived. Overall, 76.3% of jobs were linked at a 5-digit resolution, 19.9% at a 3-digit resolution, and 3.8% at a 2-digit resolution.

Exposure to an occupational agent was determined based on the probability of exposure; defined by CANJEM as the proportion of jobs in a given cell that was considered exposed to the agent, ranging from 0% to 100%. We defined a participant as "exposed" to a given agent if the probability of exposure was equal to or greater than 25%.

In our main analysis, occupational exposure to any of the 18 retained occupational agents was parameterized into three categories as a summary variable: "Unexposed", "Exposed to one agent" and "Exposed to at least two agents." Additionally, exposure to the most prevalent agents in our study (prevalence of the exposure ≥5%) was parametrized into two categories: "Unexposed" and "Exposed". Prevalent agents examined included PAHs from any source, PAHs from petroleum, lead, and formaldehyde.

5.2.6 Statistical analysis

Least squares regression was used to assess the association between smoking and occupational exposures in relation to average DNA methylation levels in the *AHRR* and *F2RL3* genes. For the continuous representation of smoking (i.e., CSI), regression coefficients and associated 95% confidence intervals (95% CI) are interpretable as the change in DNA methylation levels per standard deviation increase in CSI. For categorical representations of exposure (i.e., smoking status, occupational exposures), differences in mean methylation level and their associated 95% CI were estimated. Minimally-adjusted models included frequency-matching factors (age, sex, and phase of blood sampling) while fully-adjusted models were determined using directed acyclic graphs (Figure 5.3 for smoking and Figure 5.4 for occupational exposures) and included all frequency-matching variables. The fully adjusted model for smoking included age, sex, phase of blood sampling, education level, annual household income, and ethnicity. The fully adjusted model for occupational exposures included age, sex, phase of blood sampling, education level, and smoking.

Sensitivity analyses considered associations stratified by case-control status and by sex for each analysis to investigate possible differences between groups. Also, analyses using the participants' current employment were conducted to assess the robustness of the main analyses using the longest-held job information.

Several assumptions had to be verified beforehand such as the linear relationship between each exposure and the outcome, the homoscedasticity of the data as well as the normal distribution of errors in each model. These assumptions were verified graphically using R version 1.3.1093 without additional packages.

5.3 Results

A summary of the selected characteristics of the study population is presented in Table 5.1. The mean age of our study population was 58.6 years at baseline and there was an equal representation of the population by sex. The majority of participants were recruited in phase A. Most of the participants reported an annual household income greater than $50,000 and held a university education. Also, most of them were White, former smokers, and unexposed to any of the eighteen occupational agents under study. Controls were more likely to have a university education, to have a higher annual household income, and to be exposed to at least two chemical agents in comparison to cases. Meanwhile, cases were more likely to be smokers at baseline and to be exposed to only one chemical agent in comparison to controls.

Table 5.2 presents the associations between smoking and average methylation levels in the *AHRR* and *F2RL3* genes. Specifically, minimally-adjusted models revealed that smoking was associated with lower average methylation levels in both genes (*AHRR*: -0.016 per standard deviation increase in CSI, 95% CI: -0.022, -0.009; *F2RL3*: -0.020 per standard deviation increase in CSI, 95% CI: -0.030, -0.010). Fully-adjusted results revealed the same, though slightly attenuated associations (*AHRR*: -0.014 per standard deviation increase in CSI, 95% CI: -0.019, -0.010; *F2RL3*: -0.019 per standard deviation increase, 95% CI: -0.025, -0.012). Furthermore, in comparison with never smokers in the fully-adjusted models, both current smokers (*AHRR*: -0.080, 95% CI: -0.089, -0.070; *F2RL3*: -0.102, 95% CI: -0.120, -0.087) and former smokers (*AHRR*: -0.017, 95% CI: -0.025, -0.010; *F2RL3*: -0.022, 95% CI: -0.034, -0.010) had a lower average methylation level in the *AHRR* and *F2RL3* genes.

Table 5.3 presents the associations between occupational exposures, represented as a summary variable, in relation to the average methylation levels of the two genes of interest. No association was observed in the total study population. Additional analyses considering the most prevalent agents in relation to methylation levels of the *AHRR* and *F2RL3* genes similarly revealed no associations (Table 5.4).

Stratified analyses by case-control status and by sex revealed no observable differences in the smoking-methylation associations (Supplementary Tables 5.1 and 5.2) nor the occupational exposure-methylation associations (Supplementary Tables 5.3 and 5.4). Further sensitivity analyses considering participants' current occupation at baseline instead of their longest-held occupation did not reveal any association between the selected occupational exposures and methylation levels of the *AHRR* and *F2RL3* genes (Supplementary Tables 5.5 and 5.6).

5.4 Discussion

Across all analyses, smoking was consistently associated with hypomethylation in the *AHRR* and *F2RL3* genes. The inverse associations observed in our study, conceptualizing smoking using the CSI or smoking status, are in line with the findings of six previous studies (14-16, 19-

21). Briefly, an epigenome-wide association study (EWAS) comparing DNA methylation levels of the *AHRR* and *F2RL3* genes measured in whole blood reported that DNA methylation levels of five CpG sites within the *AHRR* or *F2RL3* genes were significantly lower in current and former smokers as compared to never smokers (19). Another study, based on two EWAS nested in the European Prospective Investigation into Cancer and Nutrition cohort study, compared DNA methylation measured in lung tissue of smokers and non-smokers, and similarly observed in three CpG sites from the *AHRR* gene and one CpG site from the *F2RL3* gene that these sites were hypomethylated among smokers as compared to those in former and non-smokers (14). These inverse associations have all been replicated by four other studies that took a candidate gene approach: two case-control studies that quantified between one and two CpG sites for each gene (15, 16), and two prospective cohort studies that measured between two and eleven CpG sites for the *AHRR* gene and one to two CpG sites for the *F2RL3* gene (20, 21). All these studies estimated a percent difference in methylation among current smokers versus never smokers between -7% and -22%, and -8% and -18% for *AHRR* and *F2RL3*, respectively. Our results fall within these ranges for each gene and therefore corroborate the observations made in these studies.

In our study, there was no observable difference in the association between smoking and *AHRR* and *F2RL3* methylation levels by sex or by case-control status, similar to that observed in the two other studies that stratified by sex and by case-control status (15, 19).

Occupational exposure is an important area for research on lung cancer. Tobacco smoke is composed of multiple chemical agents; 68 have been classified by IARC according to their carcinogenic potential to humans (carcinogenic, probably carcinogenic, or possibly carcinogenic) (29). Among them, 18 were present in the CANJEM database such that exposure to these agents in the workplace could be examined in our study. Only one study has investigated occupational exposures in relation to DNA methylation in the *AHRR* and *F2RL3* genes. Specifically, this study compared *AHRR* and *F2RL3* methylation levels among 151 male workers who were occupationally exposed to PAHs (Swedish male chimney sweeps and creosote-exposed workers) and 152 controls (those who were not occupationally exposed to PAHs). Lower *AHRR* methylation was found among PAH-exposed workers (chimney sweeps and creosote-exposed workers) as compared to controls; while, only creosote-exposed workers had lower methylation of *F2RL3* than controls (22).

While looking at the literature for global DNA methylation (i.e., methylation levels across the entire genome), five studies have reported an association between occupational exposure to our selected occupational agents and aberrant global DNA methylation patterns (30-34). Occupational exposure to formaldehyde and benzene was associated with global DNA hypomethylation (32, 34) while occupational exposure to lead and vinyl chloride was associated with global DNA hypermethylation (30, 33). A review study identified patterns of both global DNA hypomethylation and hypermethylation associated with occupational exposure to cadmium, nickel, chromium, and arsenic (31). In our study, none of the selected occupational agents, which are also present in tobacco smoke, were found to be associated with DNA methylation in the *AHRR* and *F2RL3* genes when considering exposure in both the longest-held job and the current job. These results contrast with the observations reported by the previously mentioned studies.

The baseline collection of data in CARTaGENE, including both a wide range of risk factor information and pre-collected bio-samples, provided an opportunity to explore risk factors and their association with gene-specific DNA methylation levels in a short time frame and at a relatively low cost. To date, our study has quantified the most CpG sites in the *AHRR* and *F2RL3* genes in comparison with previous studies investigating the association between smoking and *AHRR* and *F2RL3* methylation levels. Being able to measure substantially more CpG sites in the promoter regions of each gene allowed us to better estimate regional DNA methylation patterns which arguably should better approximate gene expression levels. This is especially true with the *AHRR* gene, as 33 CpG sites were measured and retained as compared to only seven CpG sites retained for the *F2RL3* gene. The quantification of DNA methylation levels was performed using Sequenom EpiTYPER® technology which uses base-specific cleavage and laser desorption/ionization-time of flight mass spectrometry (35-37), a highly precise, accurate, and cost-effective method allowing DNA methylation measurements at a single-nucleotide resolution (38). The CV of 4.65% and 4.16%, estimated for between-plate and between-fragment, respectively, lends confidence in the reliability of our methylation measures.

A cumulative index was used to parametrize smoking in this study. Previous studies investigating the association between smoking and gene-specific DNA methylation in other genes have demonstrated that incorporating different dimensions of smoking history such as time since cessation, the intensity of smoking, or duration of smoking influence gene-specific DNA methylation levels in former and current smokers (19, 39, 40). In the context of *AHRR* and *F2RL3* methylation, this study is the first to use a cumulative index, encompassing all these aspects of smoking at the same time, which allowed us to better represent the participants' smoking history and behavior. Moreover, our study is the first to examine a variety of occupational exposures in relation to *AHRR* and *F2RL3* methylation.

One of the main limitations of this study rests in the limited statistical power we had to evaluate the association between selected occupational exposures and *AHRR* and *F2RL3* methylation. The majority of the study population was unexposed to any of the retained agents (79%). Similarly, occupational exposure was dichotomized based on the probability of exposure while intensity and duration of exposure were not considered which could have introduced the possibility of exposure misclassification that could also have prevented the detection of an effect. In fact, it is possible that the study participants were exposed at a lower level of intensity and frequency than those in previously reported studies which could explain the absence of an observed association in our study.

Furthermore, even though some studies suggest that JEMs are similar in reliability in comparison to expert exposure assessments (41, 42), a JEM cannot differentiate intra-group occupational exposure variations which could lead to non-differential misclassification of exposure and attenuate the observed associations (43). A study evaluating the validity of CANJEM in contrast with the expert assessment approach examined the impact of different approaches for exposure categorization using the probability of exposure. Probability of exposure thresholds between 25% and 50% were reported to be the most valid (44). The choice of an exposure threshold affects the balance between sensitivity and specificity with higher sensitivity, but lower specificity, associated with higher exposure thresholds, and vice versa. In this study, due to the low prevalence of occupational exposure to the selected occupational agents among our participants, analyses were performed using a probability of exposure threshold of 25%. Analogously, the use of a higher occupational exposure threshold in CANJEM would have decreased the number of exposed participants but increased the sensitivity in exchange for a lower specificity (45). Regardless of the threshold used, non-differential exposure misclassification in occupational exposure is likely in this study which would have biased our association towards the null (44).

In this study, gene-specific DNA methylation measurements were performed in DNA isolated from peripheral blood samples. Hence, the results obtained might not be generalizable to lung tissues since DNA methylation can vary between tissues and blood (46). Nevertheless, a prospective EWAS, published in 2013, reported results with similar effect sizes of smoking on *AHRR* and *F2RL3* methylation as measured in peripheral blood versus lung tissue. This lends confidence in the fact that measuring DNA methylation from peripheral blood would have a minimal influence on the estimated associations (14).

It is now commonly accepted that DNA methylation shows substantial variation across individual cell types (47). Given that there are several cell types in peripheral blood, blood cell composition may confound the underlying association. No reference data set was available for adjustment for blood cell composition in the study (48). But the similar smoking-methylation association observed in our study in relation to others that adjusted for blood cell composition (15) supports that our estimates were minimally impacted by blood cell composition.

Finally, potential confounding factors were identified through a literature review and using DAGs. SES is a complex variable that can be determined by several indicators such as income, financial debts and assets, poverty level, level of education, family size, access to quality healthcare facilities and social services among others. In this study, only information about household annual income and education level was available. Therefore, both factors were used to represent SES. However, we cannot exclude the possibility of residual confounding from the use of these surrogates that might not completely encapsulate the participants' SES. Additional limitations include the possibility of uncontrolled confounding due to factors such as chronic stress that we were not able to consider.

5.5 Conclusion

The findings of this study support an association between smoking and lower average methylation levels in both the *AHRR* and *F2RL3* genes while no association was observed between the selected occupational exposures and methylation levels in the two genes of interest.

Future studies with greater statistical power, targeting populations that have a higher burden of occupational exposure, and possibly incorporating frequency and intensity of exposure are needed to explore the influence of occupational exposures on *AHRR* and *F2RL3* methylation.

Acknowledgment

The authors would like to thank the CARTaGENE study participants who made this study possible. Vikki Ho holds a Sex and Sex Gender Chair in Cancer Research from the Canadian Institutes for Health Research. During the conduct of this work, she was supported by the Cancer Research Society, the Fonds de recherche du Québec – Santé and the Ministère de l'Économie, de la Science et de l'Innovation du Québec. Currently, Vikki Ho is the recipient of a Research Scholar J2 Award from the Fonds de recherche du Québec – Santé. Laura Pelland-St-Pierre holds a doctoral scholarship from the Institut de recherche Robert-Sauvé en santé et en sécurité du travail (IRSST).

Contributors

Michael Pham, Vikki Ho, Anita Koushik, Romain Pasquet, Laura Pelland-St-Pierre, and Jack Siemiatycki designed the study's analytical strategy. Sherryl Taylor oversaw the DNA methylation measures. Michael Pham conducted the analysis. Michael Pham and Vikki Ho interpreted the results of the analysis for this paper. Michael Pham drafted the manuscript under the supervision of Dr. Vikki Ho and Dr. Anita Koushik.

Funding

This study was supported and funded by the Canadian Institutes of Health Research (FRN#162502, 2019).

References

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA: A Cancer Journal for Clinicians. 2021;71(3):209-49.

2. Doll R, Peto R. The causes of cancer: quantitative estimates of avoidable risks of cancer in the United States today. J Natl Cancer Inst. 1981;66(6):1191-308.

3. Taux d'activité (% de la population âgée de 15 ans et plus) (estimation nationale) - North America | Data.

4. Spyratos D, Zarogoulidis P, Porpodis K, Tsakiridis K, Machairiotis N, Katsikogiannis N, et al. Occupational exposure and lung cancer. J Thorac Dis. 2013;5 Suppl 4:S440-5.

5. OCRC. Burden of Occupational Cancer in Ontario: Major Workplace Carcinogens and Prevention of Exposure. Occupational Cancer Research Centre, Cancer Care Ontario; 2019.

6. Labrèche F, Duguay P, Ostiguy C, Boucher A, Roberge B, Peters CE, et al. Estimating occupational exposure to carcinogens in Quebec. American Journal of Industrial Medicine. 2013;56(9):1040-50.

7. Gustavsson P, Jakobsson R, Nyberg F, Pershagen G, Jarup L, Scheele P. Occupational exposure and lung cancer risk: a population-based case-referent study in Sweden. Am J Epidemiol. 2000;152(1):32-40.

8. Moore LD, Le T, Fan G. DNA methylation and its basic function. Neuropsychopharmacology. 2013;38(1):23-38.

9. Tsou JA, Hagen JA, Carpenter CL, Laird-Offringa IA. DNA methylation analysis: a powerful new tool for lung cancer diagnosis. Oncogene. 2002;21(35):5450-61.

10. Mutize T, Mkandla Z, Nkambule BB. Global and gene-specific DNA methylation in adult type 2 diabetic individuals: a protocol for a systematic review. Systematic Reviews. 2018;7(1):46.

11. Peinado MA. Hypomethylation of DNA. In: Schwab M, editor. Encyclopedia of Cancer. Berlin, Heidelberg: Springer; 2011. p. 1791-2.

12. Zhang Y, Elgizouli M, Schöttker B, Holleczek B, Nieters A, Brenner H. Smoking-associated DNA methylation markers predict lung cancer incidence. Clinical Epigenetics. 2016;8(1):127.

13. Vaissiere T, Hung RJ, Zaridze D, Moukeria A, Cuenin C, Fasolo V, et al. Quantitative analysis of DNA methylation profiles in lung cancer identifies aberrant DNA methylation of specific genes and its association with gender and cancer risk factors. Cancer Res. 2009;69(1):243-52.

14. Shenker NS, Polidoro S, van Veldhoven K, Sacerdote C, Ricceri F, Birrell MA, et al. Epigenome-wide association study in the European Prospective Investigation into Cancer and Nutrition (EPIC-Turin) identifies novel genetic loci associated with smoking. Hum Mol Genet. 2013;22(5):843-51.

15. Fasanelli F, Baglietto L, Ponzi E, Guida F, Campanella G, Johansson M, et al. Hypomethylation of smoking-related genes is associated with future lung cancer in four prospective cohorts. Nat Commun. 2015;6:10192.

16. Baglietto L, Ponzi E, Haycock P, Hodge A, Bianca Assumma M, Jung CH, et al. DNA methylation changes measured in pre-diagnostic peripheral blood samples are associated with smoking and lung cancer risk. Int J Cancer. 2017;140(1):50-61.

17. Oikonomopoulou K, Hansen KK, Saifeddine M, Vergnolle N, Tea I, Diamandis EP, et al. Proteinase-mediated cell signalling: targeting proteinase-activated receptors (PARs) by kallikreins and more. Biol Chem. 2006;387(6):677-85.

18. Vogel CFA, Haarmann-Stemmann T. The aryl hydrocarbon receptor repressor - More than a simple feedback inhibitor of AhR signaling: Clues for its role in inflammation and cancer. Curr Opin Toxicol. 2017;2:109-19.

19. Zeilinger S, Kuhnel B, Klopp N, Baurecht H, Kleinschmidt A, Gieger C, et al. Tobacco smoking leads to extensive genome-wide changes in DNA methylation. PLoS One. 2013;8(5):e63812.

20. Elliott HR, Tillin T, McArdle WL, Ho K, Duggirala A, Frayling TM, et al. Differences in smoking associated DNA methylation patterns in South Asians and Europeans. Clin Epigenetics. 2014;6(1):4.

21. Dogan MV, Shields B, Cutrona C, Gao L, Gibbons FX, Simons R, et al. The effect of smoking on DNA methylation of peripheral blood mononuclear cells from African American women. BMC Genomics. 2014;15:151.

22. Alhamdow A, Lindh C, Hagberg J, Graff P, Westberg H, Krais AM, et al. DNA methylation of the cancer-related genes F2RL3 and AHRR is associated with occupational exposure to polycyclic aromatic hydrocarbons. Carcinogenesis. 2018;39(7):869-78.

23. Awadalla P, Boileau C, Payette Y, Idaghdour Y, Goulet JP, Knoppers B, et al. Cohort profile of the CARTaGENE study: Quebec's population-based biobank for public health and personalized genomics. Int J Epidemiol. 2013;42(5):1285-99.

24. Ho V, Ashbury JE, Taylor S, Vanner S, King WD. Quantification of gene-specific methylation of DNMT3B and MTHFR using sequenom EpiTYPER®. Data Brief. 2015;6:39-46.

25. Hoffmann K, Krause C, Seifert B. The German Environmental Survey 1990/92 (GerES II): primary predictors of blood cadmium levels in adults. Arch Environ Health. 2001;56(4):374-9.

26. Siemiatycki J, Lavoue J. Availability of a New Job-Exposure Matrix (CANJEM) for Epidemiologic and Occupational Medicine Purposes. J Occup Environ Med. 2018;60(7):e324-e8.

27. IARC. IARC Monographs on the evaluation of Carcinogenic Risks to Humans: International Agency for Research on Cancer; 1996 1996.

28. Organization IL. ISCO - International Standard Classification of Occupations: International Labour Organization; 2010 [Available from: https://www.ilo.org/public/english/bureau/stat/isco/.

29. IARC. Tobacco smoke and involuntary smoking. IARC Monogr Eval Carcinog Risks Hum. 2004;83:1-1438.

30. Weihrauch M, Markwarth A, Lehnert G, Wittekind C, Wrbitzky R, Tannapfel A. Abnormalities of the ARF-p53 pathway in primary angiosarcomas of the liver. Hum Pathol. 2002;33(9):884-92.

31. Salemi R, Marconi A, Di Salvatore V, Franco S, Rapisarda V, Libra M. Epigenetic alterations and occupational exposure to benzene, fibers, and heavy metals associated with tumor development (Review). Molecular Medicine Reports. 2017;15(5):3366-71.

32. Leso V, Macrini MC, Russo F, Iavicoli I. Formaldehyde Exposure and Epigenetic Effects: A Systematic Review. Applied Sciences. 2020;10(7):2319.

33. Kovatsi L, Georgiou E, Ioannou A, Haitoglou C, Tzimagiorgis G, Tsoukali H, et al. p16 promoter methylation in Pb2+-exposed individuals. Clinical Toxicology. 2010;48(2):124-8.

34. Bollati V, Baccarelli A, Hou L, Bonzini M, Fustinoni S, Cavallo D, et al. Changes in DNA methylation patterns in subjects exposed to low-dose benzene. Cancer Res. 2007;67(3):876-80.

35. Coolen MW, Statham AL, Gardiner-Garden M, Clark SJ. Genomic profiling of CpG methylation and allelic specificity using quantitative high-throughput mass spectrometry: critical evaluation and improvements. Nucleic Acids Res. 2007;35(18):e119.

36. Ehrich M, Nelson MR, Stanssens P, Zabeau M, Liloglou T, Xinarianos G, et al. Quantitative high-throughput analysis of DNA methylation patterns by base-specific cleavage and mass spectrometry. Proc Natl Acad Sci U S A. 2005;102(44):15785-90.

37. Ehrich M, Zoll S, Sur S, van den Boom D. A new method for accurate assessment of DNA quality after bisulfite treatment. Nucleic Acids Res. 2007;35(5):e29.

38. Suchiman HED, Slieker RC, Kremer D, Slagboom PE, Heijmans BT, Tobi EW. Design, measurement and processing of region-specific DNA methylation assays: the mass spectrometry-based method EpiTYPER. Front Genet. 2015;6:287.

39. Wilson R, Wahl S, Pfeiffer L, Ward-Caviness CK, Kunze S, Kretschmer A, et al. The dynamics of smoking-related disturbed methylation: a two time-point study of methylation change in smokers, non-smokers and former smokers. BMC Genomics. 2017;18(1):805.

40. Philibert R, Dogan M, Beach SRH, Mills JA, Long JD. AHRR methylation predicts smoking status and smoking intensity in both saliva and blood DNA. American Journal of Medical Genetics Part B: Neuropsychiatric Genetics. 2020;183(1):51-60.

41. Descatha A, Evanoff BA, Andersen JH, Baca M, Buckner-Petty S, Fadel M, et al. Comparison Between a Self-Reported Job Exposure Matrix (JEM CONSTANCES) to an Expertise-Based Job Exposure Matrix (MADE) for Biomechanical Exposures. Journal of Occupational and Environmental Medicine. 2019;61(9):e399.

42. Offermans NS, Vermeulen R, Burdorf A, Peters S, Goldbohm RA, Koeman T, et al. Comparison of expert and job-exposure matrix-based retrospective exposure assessment of occupational carcinogens in The Netherlands Cohort Study. Occup Environ Med. 2012;69(10):745-51.

43. Coughlin SS, Chiazze L. Job-exposure matrices in epidemiologic research and medical surveillance. Occup Med. 1990;5(3):633-46.

44. Pasquet R. Methodological considerations of the Canadian job-exposure matrix and the evaluation of the risk of brain cancer in relation to occupational exposure to metallic compounds. 2019.

45. Xu M, Ho V, Lavoue J, Richardson L, Siemiatycki J. Concordance of Occupational Exposure Assessment between the Canadian Job-Exposure Matrix (CANJEM) and Expert Assessment of Jobs Held by Women. Ann Work Expo Health. 2022:wxac008.

46. Lowe R, Slodkowicz G, Goldman N, Rakyan VK. The human blood DNA methylome displays a highly distinctive profile compared with other somatic tissues. Epigenetics. 2015;10(4):274-81.

47. Houseman EA, Kelsey KT, Wiencke JK, Marsit CJ. Cell-composition effects in the analysis of DNA methylation array data: a mathematical perspective. BMC Bioinformatics. 2015;16(1):95.

48. Houseman EA, Accomando WP, Koestler DC, Christensen BC, Marsit CJ, Nelson HH, et al. DNA methylation arrays as surrogate measures of cell mixture distribution. BMC Bioinformatics. 2012;13:86.

Table 5.1. Baseline characteristics of study population.

Selected Characteristics		**All**	**Cases**	**Controls**
		(N = 531)	(n=179)	(n=352)
Mean age at baseline (SD)		58.6 (±7.3)	58.6 (±7.4)	58.7 (±7.3)
		N (%)	N (%)	N (%)
Phase of blood sampling	Phase A	483 (91%)	162 (90.5%)	321 (91.2%)
	Phase B	48 (9.0%)	17 (9.5%)	31 (8.8%)
Sex	Male	260 (49.0%)	86 (48.0%	174 (49.4%)
	Female	271 (51.0%)	93 (52.0%)	178 (50.6%)
Ethnicity	White	502 (94.7%)	160 (89.4%)	342 (97.4%)
	Other	17 (3.2%)	10 (5.6%)	7 (2%)
	Missing	12 (2.1%)	9 (5%)	3 (0.6%)
Highest level of education	High school or lower	150 (28.4%)	66 (37.3%)	84 (23.9%)
	Technical school or college	158 (29.9%)	54 (30.5%)	104 (29.5%)
	University and above	221 (41.7%)	57 (32.2%)	164 (46.6%)
Annual household income	Below $25 000	58 (10.9%)	24 (13.4%)	34 (9.7%)
	$25 000 - $49 999	129 (24.3%)	52 (29.0%)	77 (21.9%)
	$50 000 - $74 999	140 (26.4%)	46 (25.7%)	94 (26.7%)
	$75 000 - $99 999	74 (13.9%)	24 (13.4%)	50 (14.2%)
	Above $100 000	102 (19.2%)	23 (12.8%)	79 (22.4%)
	Missing	28 (5.3%)	10 (5.6%)	18 (5.1%)
Smoking status	Never smokers	186 (35.2%)	42 (23.7%)	144 (40.9%)
	Former smokers	237 (44.8%)	75 (42.4%)	162 (46.0%)
	Current smokers	106 (20.0%)	60 (33.9%)	46 (13.1%)
Occupational exposure to the selected 18 agents in longest-held job[1]	Unexposed	377 (79.0%)	125 (80.1%)	252 (78.5%)
	Exposed to 1 agent	28 (5.9%)	15 (9.6%)	13 (4.0%)
	Exposed to 2+ agents	72 (15.1%)	16 (10.3%)	56 (17.4%)

[1] PAHs from any source, PAHs from petroleum, PAHs from wood, PAHs from other sources, PAHs from coal, benzo[a]pyrene, chromium VI, beryllium, cobalt, nickel, arsenic, cadmium, lead, formaldehyde, ethylene oxide, vinyl chloride, benzene and aromatic amines.

Table 5.2. Association between smoking and methylation levels in the *AHRR* and *F2RL3* genes.

Smoking status	N (%)	**Coefficient estimates (95% CI)**			
		AHRR		***F2RL3***	
		Minimally adjusted[1]	**Fully adjusted**[2]	**Minimally adjusted**[a]	**Fully adjusted**[b]
CSI (per s.d increase)	507 (100%)	-0.016 (-0.022, -0.009)	-0.014 (-0.019, -0.0010)	-0.020 (-0.030, -0.010)	-0.019 (-0.025, -0.012)
Never smokers	187 (35.3%)	Reference	Reference	Reference	Reference
Former smokers	237 (44.7%)	-0.018 (-0.026, -0.011)	-0.017 (-0.025, -0.010)	-0.024 (-0.036, -0.011)	-0.022 (-0.034, -0.010)
Current smokers	106 (20%)	-0.084 (-0.093, -0.075)	-0.080 (-0.089, -0.070)	-0.110 (-0.125, -0.095)	-0.102 (-0.120, -0.087)

[1] Adjusted for sex, age, and phase of blood sampling.

[2] Adjusted for sex, age, phase of blood sampling, ethnicity, education level and household annual income

Table 5.3. Association between occupational exposures to the selected 18 agents and methylation levels in the *AHRR* and *F2RL3* genes.

Exposure status[1]	**N (%)**	**Coefficient estimates (95% CI)**			
		AHRR		***F2RL3***	
		Minimally adjusted[2]	**Fully adjusted**[3]	**Minimally adjusted**	**Fully adjusted**
Unexposed	377 (79.0%)	Reference	Reference	Reference	Reference
Exposed to 1 agent	28 (5.9%)	-0.014 (-0.033, 0.005)	-0.003 (-0.022, 0.016)	-0.011 (-0.040, 0.017)	0.001 (-0.029, 0.030)
Exposed to at least 2 agents	72 (15.1%)	0.003 (-0.010, 0.016)	0.006 (-0.007, 0.018)	-0.003 (-0.022, 0.017)	0.002 (-0.018, 0.022)

[1] PAHs from any source, PAHs from petroleum, PAHs from wood, PAHs from other sources, PAHs from coal, benzo[a]pyrene, chromium VI, beryllium, cobalt, nickel, arsenic, cadmium, lead, formaldehyde, ethylene oxide, vinyl chloride, benzene and aromatic amines.

[2] Adjusted for sex, age and phase of blood sampling.

[3] Adjusted for sex, age, phase of blood sampling, smoking (standardized CSI) and education level

Table 5.4. Association between occupational exposure to the most prevalent agents and methylation levels in the *AHRR* and *F2RL3* genes in the fully adjusted model.

Exposure	**N (%)**	**Coefficient estimates**[1] **(95% CI)**	
		AHRR	***F2RL3***
PAHs from any source			
Unexposed	398 (83.4)	Reference	Reference
Exposed	79 (16.6%)	0.005 (-0.007, 0.018)	-0.000 (-0.019, 0.019)
PAHs from petroleum			
Unexposed	423 (88.7%)	Reference	Reference
Exposed	54 (11.3%)	-0.003 (-0.011, 0.017)	-0.007 (-0.029, 0.015)
Lead			
Unexposed	427 (89.5%)	Reference	Reference
Exposed	50 (10.5%)	0.008 (-0.007, 0.022)	-0.003 (-0.020, 0.026)
Formaldehyde			
Unexposed	452 (94.8%)	Reference	Reference
Exposed	25 (5.2%)	-0.011 (-0.031, 0.009)	-0.007 (-0.039, 0.023)

[1] Adjusted for sex, age, phase of blood sampling, smoking (standardized CSI) and education level

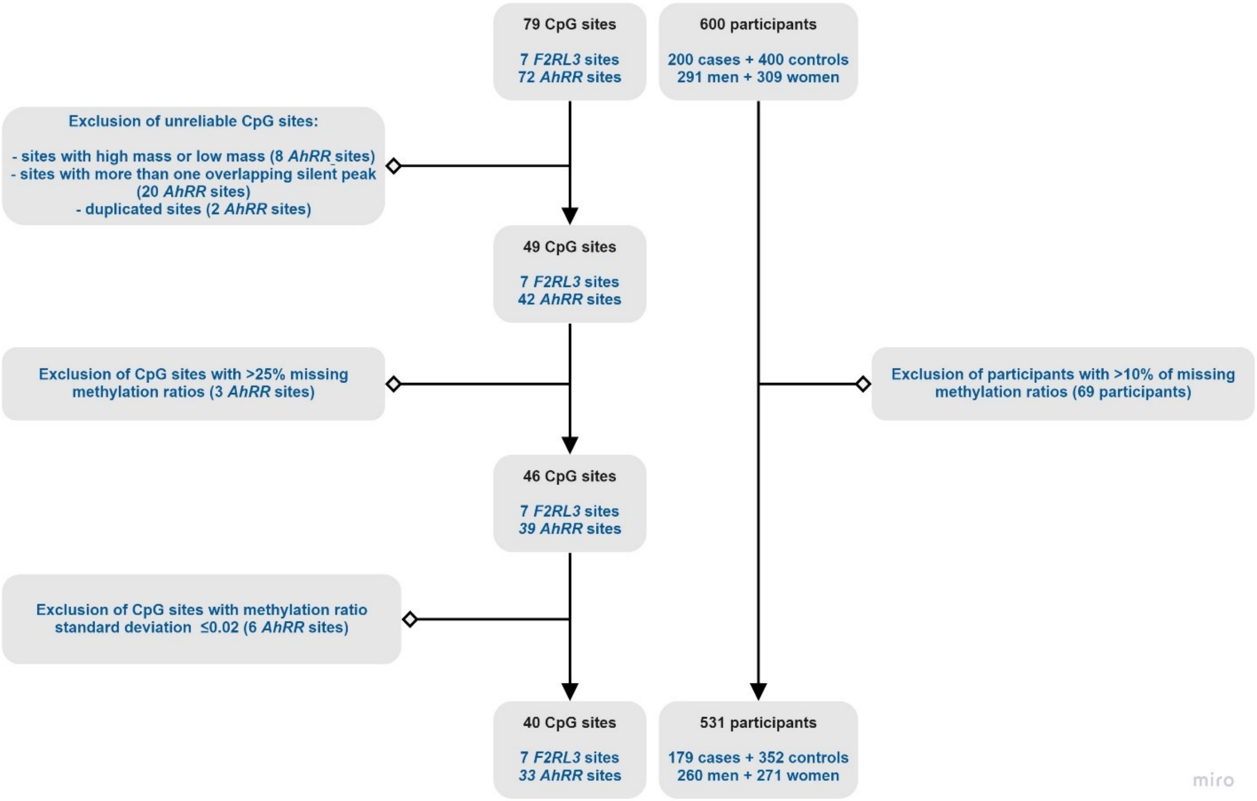

79 CpG sites
7 F2RL3 sites
72 AhRR sites
600 participants
200 cases + 400 controls
291 men + 309 women
Exclusion of unreliable CpG sites:
- sites with high mass or low mass (8 AhRR sites)
- sites with more than one overlapping silent peak (20 AhRR sites)
- duplicated sites (2 AhRR sites)
49 CpG sites
7 F2RL3 sites
42 AhRR sites
Exclusion of CpG sites with >25% missing methylation ratios (3 AhRR sites)
Exclusion of participants with >10% of missing methylation ratios (69 participants)
46 CpG sites
7 F2RL3 sites
39 AhRR sites
Exclusion of CpG sites with methylation ratio standard deviation ≤0.02 (6 AhRR sites)
40 CpG sites
7 F2RL3 sites
33 AhRR sites
531 participants
179 cases + 352 controls
260 men + 271 women
miro

Figure 5.1. Description of the participants and DNA methylation data cleaning.

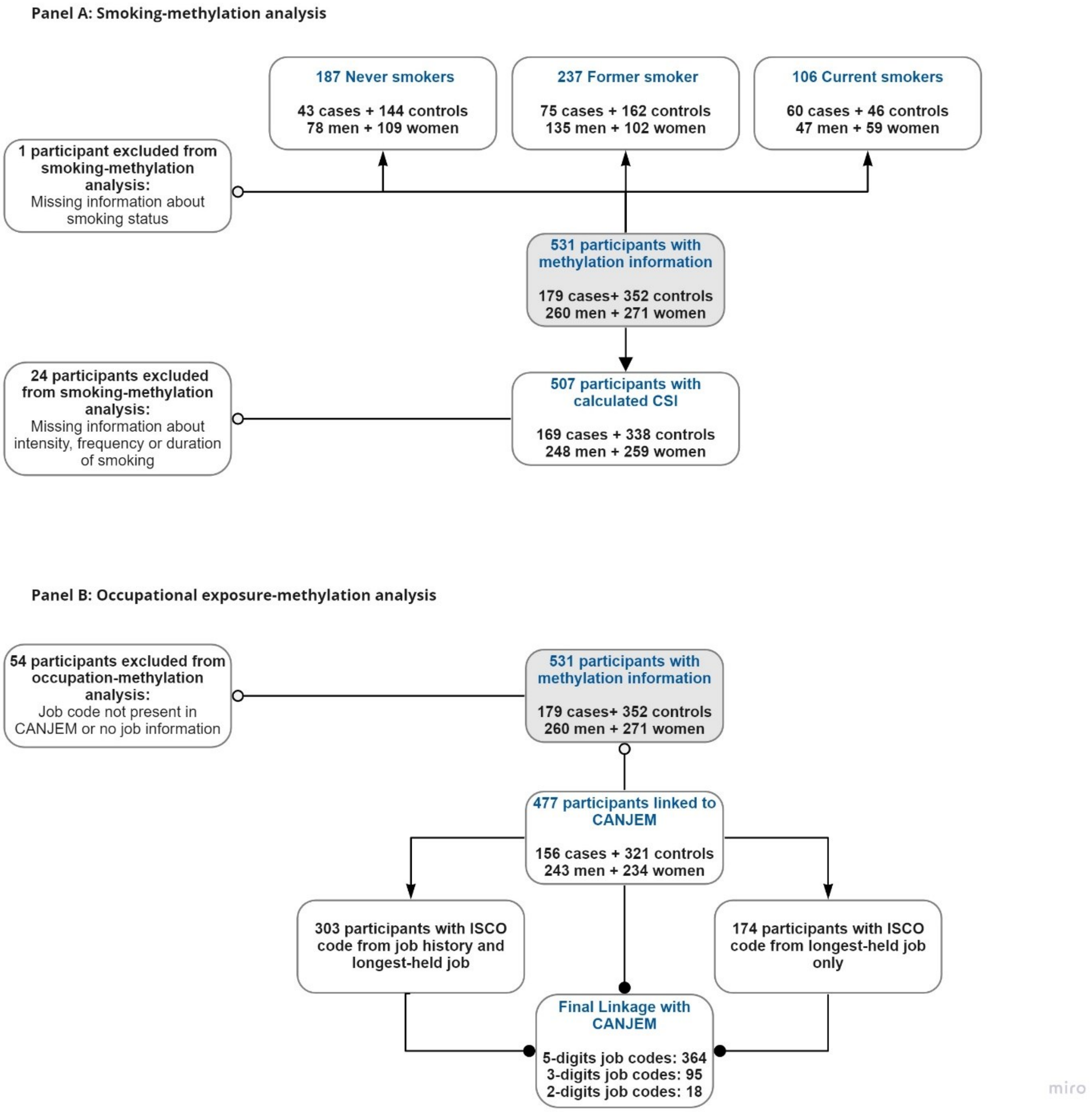

Panel A: Smoking-methylation analysis
187 Never smokers
43 cases + 144 controls
78 men + 109 women
237 Former smoker
75 cases + 162 controls
135 men + 102 women
106 Current smokers
60 cases + 46 controls
47 men + 59 women
1 participant excluded from smoking-methylation analysis:
Missing information about smoking status
531 participants with methylation information
179 cases+ 352 controls
260 men + 271 women
24 participants excluded from smoking-methylation analysis:
Missing information about intensity, frequency or duration of smoking
507 participants with calculated CSI
169 cases + 338 controls
248 men + 259 women
Panel B: Occupational exposure-methylation analysis
54 participants excluded from occupation-methylation analysis:
Job code not present in CANJEM or no job information
531 participants with methylation information
179 cases+ 352 controls
260 men + 271 women
477 participants linked to CANJEM
156 cases + 321 controls
243 men + 234 women
303 participants with ISCO code from job history and longest-held job
174 participants with ISCO code from longest-held job only
Final Linkage with CANJEM
5-digits job codes: 364
3-digits job codes: 95
2-digits job codes: 18
miro

Figure 5.2. Flowchart of the described study participant inclusion schema for the smoking-methylation analysis (Panel A) and the occupational exposure-methylation analysis (Panel B).

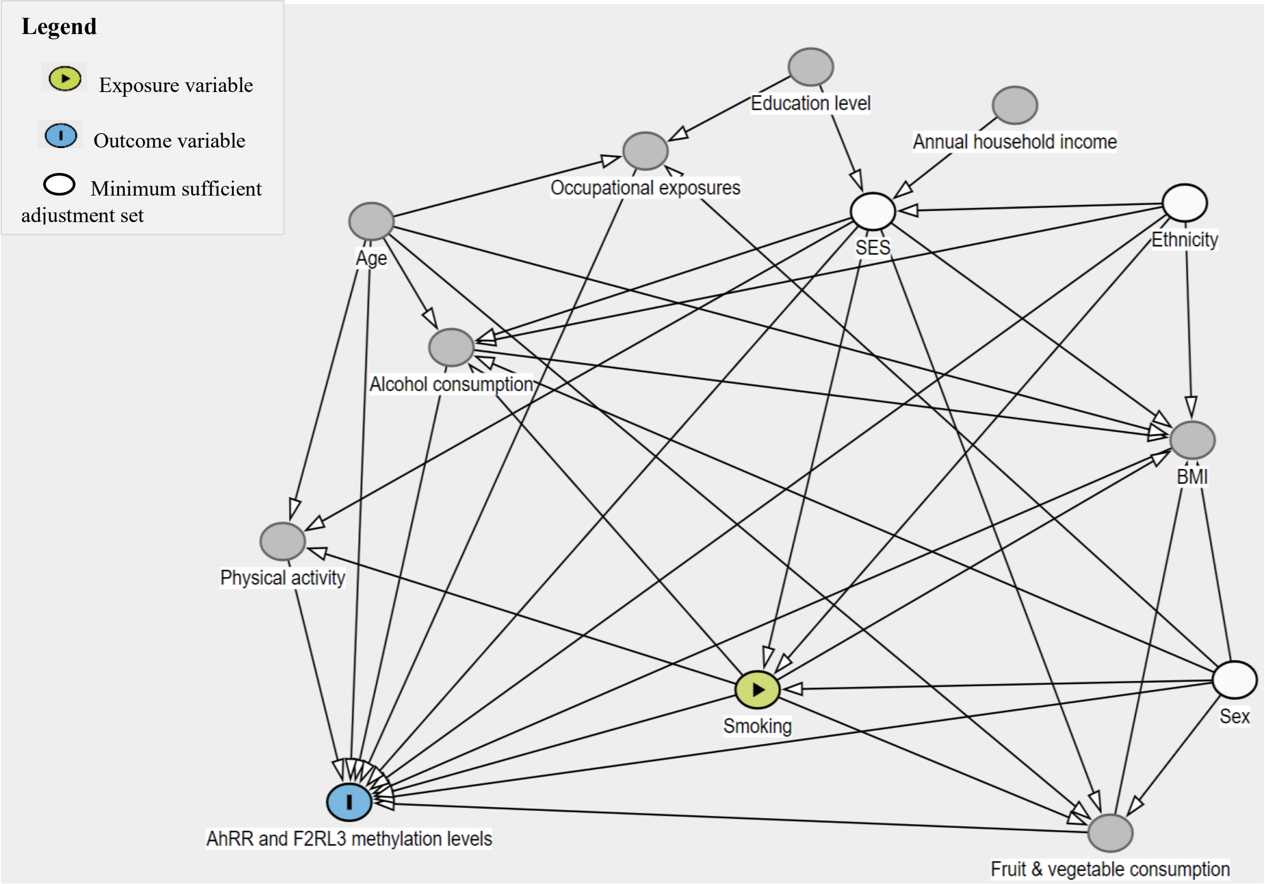


Figure 5.3. Directed acyclic graph of the association between smoking and methylation levels of the AHRR and F2RL3 genes.

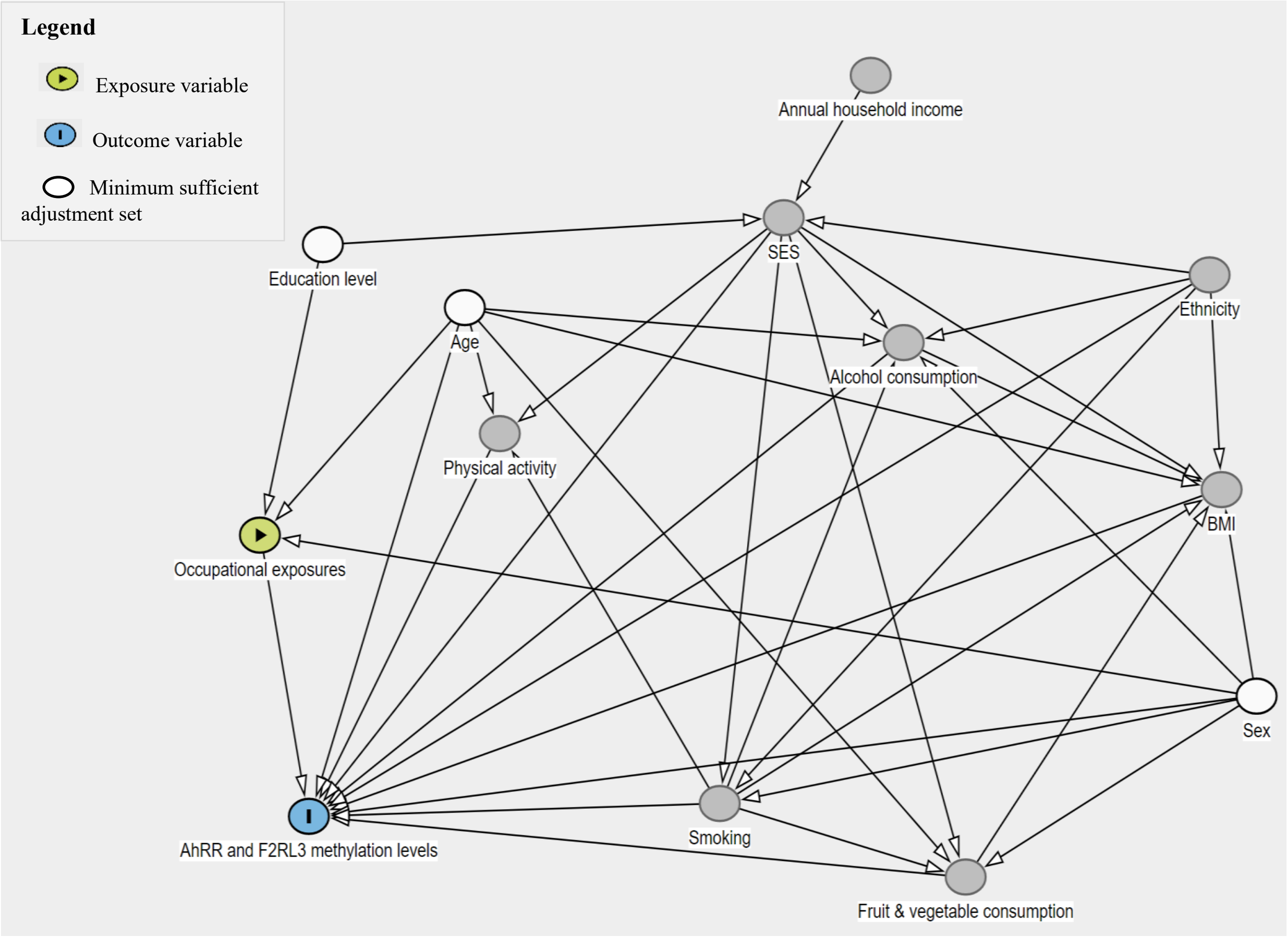


Figure 5.4. Directed acyclic graph of the association between selected occupational exposures and methylation levels of the AHRR and F2RL3 genes.

Article: Supplementary results

Supplementary Table 5.1. Association between smoking and methylation levels in the *AHRR* and *F2RL3* gene stratified by case-control status.

Smoking status	**N (%)**		**Coefficient estimates[1] (95% CI)**			
			AHRR		***F2RL3***	
	Case	**Control**	**Case**	**Control**	**Case**	**Control**
CSI (per s.d increase)	179 (100%)	352 (100%)	-0.015 (-0.023, -0.005)	-0.014 (-0.018, -0.010)	-0.020 (-0.034, -0.006)	-0.017 (-0.024, -0.010)
Never smokers	42 (23.7%)	144 (40.9%)	Reference	Reference	Reference	Reference
Former smokers	75 (42.4%)	162 (46%)	-0.024 (-0.043, -0.006)	-0.018 (-0.025, -0.010)	-0.039 (-0.067, -0.012)	-0.018 (-0.032, -0.004)
Current smokers	60 (33.9%)	46 (13.1%)	-0.088 (-0.107, -0.068)	-0.065 (-0.076, -0.054)	-0.118 (-0.146, -0.089)	-0.081 (-0.101, -0.061)

[1] Adjusted for sex, age, phase of blood sampling, ethnicity, education level, and household annual income

Supplementary Table 5.2. Association between smoking and methylation levels in the *AHRR* and *F2RL3* gene stratified by sex.

Smoking status	**N (%)**		**Coefficient estimates[1] (95% CI)**			
			AHRR		***F2RL3***	
	Men	**Women**	**Men**	**Women**	**Men**	**Women**
CSI (per s.d increase)	260 (100%)	271 (100%)	-0.014 (-0.020, -0.007)	-0.015 (-0.020, -0.009)	-0.019 (-0.029, -0.010)	-0.018 (-0.027, -0.009)
Never smokers	78 (30%)	108 (40.1%)	Reference	Reference	Reference	Reference
Former smokers	135 (51.9%)	102 (37.9%)	-0.012 (-0.024, 0.001)	-0.023 (-0.033, -0.013)	-0.013 (-0.031, 0.004)	-0.029 (-0.047, -0.012)
Current smokers	47 (18.1%)	59 (21.9%)	-0.087 (-0.102, -0.071)	-0.076 (-0.088, -0.063)	-0.104 (-0.128, -0.080)	-0.093 (-0.114, -0.072)

[1] Adjusted for age, phase of blood sampling, ethnicity, education level, and household annual income

Supplementary Table 5.3. Association between occupational exposure to the selected 18 agents and methylation levels in the *AHRR* and *F2RL3* genes stratified by case-control status.

Exposure status[1]	N (%)			Coefficient estimates[2] (95% CI)			
				AHRR		*F2RL3*	
	Total	**Case**	**Control**	**Case**	**Control**	**Case**	**Control**
Unexposed	377 (79.0%)	125 (80.1%)	252 (78.5%)	Reference	Reference	Reference	Reference
Exposed to 1 agent	28 (5.9%)	15 (9.6%)	13 (4.0%)	-0.007 (-0.044, 0.030)	0.001 (-0.019, 0.021)	0.006 (-0.047, 0.059)	-0.005 (-0.040, 0.031)
Exposed to at least 2 agents	72 (15.1%)	16 (10.3%)	56 (17.4%)	0.004 (-0.032, 0.040)	0.001 (-0.010, 0.012)	-0.000 (-0.052, 0.051)	-0.005 (-0.025, 0.014)

[1] PAHs from any source, PAHs from petroleum, PAHs from wood, PAHs from other sources, PAHs from coal, benzo[a]pyrene, chromium VI, beryllium, cobalt, nickel, arsenic, cadmium, lead, formaldehyde, ethylene oxide, vinyl chloride, benzene, and aromatic amines.

[2] Adjusted for sex, age, phase of blood sampling, ethnicity, education level, and household annual income

Supplementary Table 5.4. Association between occupational exposure to the selected 18 agents and methylation levels in the *AHRR* and *F2RL3* genes stratified by sex.

Exposure status[1]	N (%)			Coefficient estimates[2] (95% CI)			
				AHRR		*F2RL3*	
	Total	**Men**	**Women**	**Men**	**Women**	**Men**	**Women**
Unexposed	377 (79.0%)	125 (80.1%)	252 (78.5%)	Reference	Reference	Reference	Reference
Exposed to 1 agent	28 (5.9%)	15 (9.6%)	13 (4.0%)	0.001 (-0.026, 0.029)	-0.003 (-0.030, 0.023)	-0.007 (-0.049, 0.034)	0.017 (-0.027, 0.060)
Exposed to at least 2 agents	72 (15.1%)	16 (10.3%)	56 (17.4%)	0.006 (-0.010, 0.022)	0.006 (-0.019, 0.031)	-0.002 (-0.027, 0.022)	0.014 (-0.026, 0.055)

[1] PAHs from any source, PAHs from petroleum, PAHs from wood, PAHs from other sources, PAHs from coal, benzo[a]pyrene, chromium VI, beryllium, cobalt, nickel, arsenic, cadmium, lead, formaldehyde, ethylene oxide, vinyl chloride, benzene, and aromatic amines.

[2] Adjusted for age, phase of blood sampling, ethnicity, education level, and household annual income

Supplementary Table 5.5. Association between occupational exposures to the selected 18 agents in the current employment at baseline and methylation levels in the *AHRR* and *F2RL3* genes.

Exposure status[1]	**N (%)**	**Coefficient estimates (95% CI)**[2]	
		AHRR	***F2RL3***
Unexposed	209 (81.3%)	Reference	Reference
Exposed to 1 agent	14 (5.4%)	-0.004 (-0.032, 0.021)	-0.005 (-0.047, 0.037)
Exposed to at least 2 agents	34 (13.2%)	-0.002 (-0.020, 0.018)	-0.003 (-0.023, 0.017)

[1] PAHs from any source, PAHs from petroleum, PAHs from wood, PAHs from other sources, PAHs from coal, benzo[a]pyrene, chromium VI, beryllium, cobalt, nickel, arsenic, cadmium, lead, formaldehyde, ethylene oxide, vinyl chloride, benzene, and aromatic amines.

[2] Adjusted for sex, age, phase of blood sampling, ethnicity, education level, and household annual income

Supplementary Table 5.6. Association between occupational exposures status to the most prevalent agents from current employment at baseline and methylation levels in the *AHRR* and *F2RL3* genes.

Exposure status	**N (%)**	**Coefficient estimates**[1] **(95% CI)**	
		AHRR	***F2RL3***
PAHs from any source			
Unexposed	213 (82.9)	Reference	Reference
Exposed	44 (17.1%)	-0.002 (-0.019, 0.016)	-0.001 (-0.029, 0.027)
PAHs from petroleum			
Unexposed	229 (89.1%)	Reference	Reference
Exposed	28 (10.9%)	0.003 (-0.017, 0.022)	-0.010 (-0.023, 0.043)
Lead			
Unexposed	231 (89.9%)	Reference	Reference
Exposed	26 (10.1%)	-0.002 (-0.023, 0.019)	-0.024 (-0.069, 0.032)
Formaldehyde			
Unexposed	244 (94.9%)	Reference	Reference
Exposed	13 (5.1%)	-0.012 (-0.044, 0.020)	-0.022 (-0.073, 0.029)

[1] Adjusted for sex, age, phase of blood sampling, smoking (standardized CSI), and education level

Chapter 6. ADDITIONAL RESULTS

The primary objectives of this thesis were to investigate the associations between smoking and selected occupational exposures, and DNA methylation levels of the *AHRR* and *F2RL3* genes. In addition to the main analyses presented in Chapter 5, sensitivity analyses were conducted to evaluate the robustness of the study findings. This chapter presents additional findings corresponding to the estimation of the association between each risk factor and the DNA methylation levels of individual CpG sites, and finally the effects of excluding participants with missing methylation information for the gene *AHRR* or *F2RL3*.

For each gene, the main methylation analyses used the average methylation ratios across all informative CpG sites (33 *AHRR* CpG sites and seven *F2RL3* CpG sites). However, additional analyses using the methylation ratios of individual CpG sites for each gene were also performed in order to reveal whether certain individual CpG sites were more associated with smoking and occupational exposures. Moreover, this also provided the possibility to compare our observed individual results of hypomethylated CpG sites with results reported in the literature. Seventeen of 33 *AHRR* CpG sites and six of seven *F2RL3* CpG sites were found to be hypomethylated in relation to tobacco smoking (95% confidence intervals not including 0) (Table 6.1 and Table 6.2). Additionally, one *AHRR* CpG site (Chr5:369774) was hypomethylated in relation to occupational exposure to one of the 18 chemical agents retained. No other individual CpG site from any of the two genes was found to be hypomethylated in relation to occupational exposures (Table 6.3 and Table 6.4).

Table 6.1. Association between smoking (parametrized as standardized CSI) and methylation level in the individual CpG sites of the *AHRR* gene.

CpG site	**Coefficient estimates**[1]
Chr5:373249	**-0.015 (-0.024, -0.006)**
Chr5:373251	**-0.014 (-0.023, -0.006)**
Chr5:373300	**-0.028 (-0.039, -0.016)**
Chr5:373316	**-0.026 (-0.041, -0.009)**
Chr5:373379	**-0.040 (-0.052, -0.027)**
Chr5:373424	**-0.029 (-0.041, -0.017)**
Chr5:373473	**-0.039 (-0.053, -0.026)**
Chr5:373491	**-0.044 (-0.058, -0.031)**
Chr5:373495	**-0.044 (-0.058, -0.031)**
Chr5:373530	**-0.047 (-0.057, -0.037)**
Chr5:373610	**-0.011 (-0.020, -0.003)**
Chr5:368449	**-0.014 (-0.021, -0.007)**
Chr5:368447	**-0.014 (-0.021, -0.007)**
Chr5:368430	**-0.008 (-0.012, -0.003)**
Chr5:368278	-0.001 (-0.004, 0.002)
Chr5:368756	-0.002 (-0.005, 0.001)
Chr5:368805	**-0.010 (-0.016, -0.004)**
Chr5:368898	0.001 (-0.002, 0.003)
Chr5:368900	0.001 (-0.002, 0.003)
Chr5:392693	-0.007 (-0.013, 0.000)
Chr5:392704	0.002 (-0.001, 0.004)
Chr5:392940	-0.002 (-0.005, 0.002)
Chr5:392946	-0.002 (-0.006, 0.002)
Chr5:393073	**-0.006 (-0.011, -0.001)**
Chr5:393076	**-0.006 (-0.011, -0.001)**
Chr5:369774	-0.007 (-0.019, 0.004)
Chr5:369970	-0.003 (-0.011, 0.004)
Chr5:370021	-0.009 (-0.019, 0.001)
Chr5:377325	-0.000 (-0.004, 0.004)
Chr5:377359	0.002 (-0.002, 0.006)
Chr5:377361	0.003 (-0.001, 0.006)
Chr5:377438	0.003 (-0.000, 0.006)
Chr5:377453	0.002 (-0.002, 0.006)

[1] Adjusted for sex, age, phase of blood sampling, ethnicity, education level, and household annual income

Table 6.2. Association between smoking and methylation level in the individual CpG sites of the *F2RL3* gene.

CpG site	**Coefficient estimates**[1]
Chr19 :17000596	**-0.011 (-0.020, -0.001)**
Chr19 :17000585	**-0.017 (-0.024 -0.010)**
Chr19 :17000567	-0.008 (-0.017, 0.002)
Chr19 :17000552	**-0.023 (-0.030, -0.015)**
Chr19 :17000517	**-0.009 (-0.015, -0.003)**
Chr19 :17000476	**-0.022 (-0.030, -0.014)**
Chr19 :17000465	**-0.020 (-0.030, -0.012)**

[1] Adjusted for sex, age, phase of blood sampling ethnicity, education level, and household annual income

Table 6.3. Association between occupational exposure to the selected 18 occupational agents and methylation level in the individual CpG sites of the *AHRR* gene.

	Coefficient estimates (95% CI)[1,2]		
CpG sites	**Unexposed**	**Exposed to 1 agent**	**Exposed to at least 2 agents**
Chr5:373249	Reference	0.011 (-0.023, 0.049)	-0.002 (-0.027, 0.022)
Chr5:373251	Reference	0.013 (-0.022, 0.049)	-0.001 (-0.026, 0.023)
Chr5:373300	Reference	0.007 (-0.042, 0.055)	0.003 (-0.029, 0.036)
Chr5:373316	Reference	-0.019 (-0.083, 0.045)	-0.008 (-0.051, 0.036)
Chr5:373379	Reference	0.013 (-0.041, 0.068)	0.006 (-0.030, 0.043)
Chr5:373424	Reference	0.005 (-0.048, 0.058)	0.004 (-0.032, 0.040)
Chr5:373473	Reference	0.026 (-0.034, 0.085)	0.016 (-0.024, 0.056)
Chr5:373491	Reference	0.020 (-0.037, 0.077)	0.015 (-0.024, 0.054)
Chr5:373495	Reference	0.020 (-0.037, 0.077)	0.017 (-0.022, 0.054)
Chr5:373530	Reference	0.033 (-0.010, 0.077)	0.012 (-0.017, 0.041)
Chr5:373610	Reference	-0.007 (-0.046, 0.032)	-0.004 (-0.030, 0.023)
Chr5:368449	Reference	-0.015 (-0.046, 0.016)	0.003 (-0.018, 0.025)
Chr5:368447	Reference	-0.015 (-0.046, 0.016)	0.003 (-0.018, 0.025)
Chr5:368430	Reference	-0.013 (-0.032, 0.005)	0.001 (-0.012, 0.014)
Chr5:368278	Reference	-0.008 (-0.020, 0.005)	-0.003 (-0.0010, 0.005)
Chr5:368756	Reference	0.007 (-0.006, 0.021)	0.003 (-0.006, 0.012)
Chr5:368805	Reference	-0.022 (-0.048, 0.005)	0.002 (-0.016, 0.020)
Chr5:368898	Reference	-0.002 (-0.012, 0.009)	-0.001 (-0.008, 0.007)
Chr5:368900	Reference	-0.001 (-0.012, 0.010)	-0.001 (-0.008, 0.007)
Chr5:392693	Reference	-0.010 (-0.038, 0.018)	0.007 (-0.012, 0.026)
Chr5:392704	Reference	-0.004 (-0.012, 0.004)	-0.005 (-0.011, 0.000)
Chr5:392940	Reference	-0.012 (-0.029, 0.004)	-0.001 (-0.013, 0.010)
Chr5:392946	Reference	-0.012 (-0.029, 0.004)	-0.001 (-0.013, 0.010)
Chr5:393073	Reference	-0.004 (-0.026, 0.018)	0.007 (-0.006, 0.022)
Chr5:393076	Reference	-0.004 (-0.026, 0.018)	0.006 (-0.009, 0.021)
Chr5:369774	Reference	**-0.053 (-0.102,-0.005)**	0.005 (-0.028, 0.038)
Chr5:369970	Reference	-0.032 (-0.065, 0.001)	0.011 (-0.011, 0.034)
Chr5:370021	Reference	-0.017 (-0.059, 0.025)	0.010 (-0.019, 0.038)
Chr5:377325	Reference	0.005 (-0.011, 0.022)	0.002 (-0.009, 0.013)
Chr5:377359	Reference	0.008 (-0.007, 0.022)	-0.006 (-0.015, 0.004)
Chr5:377361	Reference	0.008 (-0.007, 0.022)	-0.008 (-0.018, 0.003)
Chr5:377438	Reference	0.005 (-0.009, 0.019)	-0.005 (-0.0116, 0.004)
Chr5:377453	Reference	-0.002 (-0.016, 0.012)	-0.007 (-0.017, 0.002)

[1] PAHs from any source, PAHs from petroleum, PAHs from wood, PAHs from other sources, PAHs from coal, benzo[a]pyrene, chromium VI, beryllium, cobalt, nickel, arsenic, cadmium, lead, formaldehyde, ethylene oxide, vinyl chloride, benzene, and aromatic amines.

[2] Adjusted for sex, age, phase of blood sampling, ethnicity, education level, and household annual income

Table 6.4. Association between occupational exposure to the selected 18 agents and methylation level in the individual CpG sites of the *F2RL3* gene.

	Coefficient estimates (95% CI)[1,2]		
CpG sites	**Unexposed**	**Exposed to 1 agent**	**Exposed to at least 2 agents**
Chr19:17000596	Reference	0.008 (-0.031, 0.049)	0.021 (-0.006, 0.047)
Chr19:17000585	Reference	0.000 (-0.033, 0.033)	-0.004 (-0.026, 0.018)
Chr19:17000567	Reference	-0.001 (-0.045, 0.044)	-0.019 (-0.050, 0.010)
Chr19:17000552	Reference	0.006 (-0.030, 0.042)	-0.004 (-0.028, 0.020)
Chr19:17000517	Reference	-0.009 (-0.035, 0.016)	-0.001 (-0.017, 0.016)
Chr19:17000476	Reference	0.006 (-0.030, 0.042)	-0.004 (-0.028, 0.020)
Chr19:17000465	Reference	0.001 (-0.040, 0.043)	-0.003 (-0.03, 0.025)

[1] PAHs from any source, PAHs from petroleum, PAHs from wood, PAHs from other sources, PAHs from coal, benzo[a]pyrene, chromium VI, beryllium, cobalt, nickel, arsenic, cadmium, lead, formaldehyde, ethylene oxide, vinyl chloride, benzene, and aromatic amines.

[2] Adjusted for sex, age, phase of blood sampling, ethnicity, education level, and household annual income

Finally, in the main methylation analyses, after the exclusion of uninformative CpG sites described earlier, several participants had missing methylation information. We prioritized having the same study sample for the *AHRR* and *F2RL3* analysis and thus only those participants with more than 10% methylation information missing for both the *AHRR* and *F2RL3* genes were excluded. To assess how this exclusion criterion could have affected the results obtained, sensitivity analyses excluding participants with more than 10% methylation information missing for the *AHRR* or *F2RL3* genes separately were performed in the total study population (Table 6.5 and Table 6.6). In the smoking-methylation analysis, this yielded 524 and 442 participants in the analyses with *AHRR* and F*2RL3*, respectively. In the occupation-methylation analysis, 458 and 422 participants were included in the analyses with *AHRR* and F*2RL3*, respectively. The results obtained did not reveal any observable association differences with the main analyses which indicates that this exclusion criterion did not noticeably influence the estimates. Indeed, very similar coefficient estimates and 95% confidence intervals were found.

Table 6.5. Association between smoking and methylation levels in the *AHRR* and *F2RL3* genes after exclusion of participants with more than 10% methylation information missing for the gene *AHRR* or *F2RL3*.

			Coefficient estimates (95% CI)			
	N (%)		***AHRR***		***F2RL3***	
	AHRR	***F2RL3***	**Minimally adjusted**[1]	**Fully adjusted**[2]	**Minimally adjusted**[a]	**Fully adjusted**[b]
CSI (per s.d increase)	524 (100%)	442 (100%)	-0.017 (-0.021, -0.012)	-0.014 (-0.018, -0.009)	-0.020 (-0.027, -0.014)	-0.018 (-0.025, -0.008)
Never smokers	185 (35.4%)	174 (39.5%)	Reference	Reference	Reference	Reference
Former smokers	231 (44.2%)	182 (41.4%)	-0.018 (-0.026, -0.011)	-0.018 (-0.025, -0.0010)	-0.015 (-0.023, -0.008)	-0.017 (-0.030, -0.002)
Current smokers	107 (20.5%)	84 (19.1%)	-0.084 (-0.093, -0.074)	-0.079 (-0.089, -0.066)	-0.078 (-0.087, -0.068)	-0.088 (-0.104, -0.070)

[1] Adjusted for sex, age, and phase of blood sampling.

[2] Adjusted for sex, age, phase of blood sampling, ethnicity, education level, and household annual income

Table 6.6 Association between occupational exposures to the 18 selected agents and methylation levels in the *AHRR* and *F2RL3* genes after exclusion of participants with more than 10% methylation information missing for the gene *AHRR* or *F2RL3*.

			Coefficient estimates (95% CI)			
	N (%)		***AHRR***		***F2RL3***	
Exposure status[1]	***AHRR***	***F2RL3***	**Minimally adjusted**[2]	**Fully adjusted**[3]	**Minimally adjusted**	**Fully adjusted**
Unexposed	380 (83%)	351 (83.2%)	Reference	Reference	Reference	Reference
Exposed to 1 agent	17 (3.7%)	19 (4.5%)	-0.012 (-0.030, 0.007)	-0.001 (-0.020, 0.018)	-0.013 (-0.041, 0.016)	0.000 (-0.030, 0.030)
Exposed to at least 2 agents	61 (13.3%)	52 (12.3%)	0.001 (-0.012, 0.015)	0.004 (-0.009, 0.016)	-0.001 (-0.019, 0.018)	0.004 (-0.016, 0.023)

[1] PAHs from any source, PAHs from petroleum, PAHs from wood, PAHs from other sources, PAHs from coal, benzo[a]pyrene, chromium VI, beryllium, cobalt, nickel, arsenic, cadmium, lead, formaldehyde, ethylene oxide, vinyl chloride, benzene, and aromatic amines.

[2] Adjusted for sex, age and phase of blood sampling

[3] Adjusted for sex, age, phase of blood sampling, smoking (standardized CSI), and education level

Chapter 7. DISCUSSION

The association between tobacco smoking and selected occupational exposures, and DNA methylation levels of the *AHRR* and *F2RL3* genes was investigated in this thesis. The availability of data from a case-control study nested in CARTaGENE provided a unique opportunity to perform a cross-sectional analysis of the impact of smoking and occupational exposures on DNA methylation, an essential epigenetic mechanism. The purpose of this chapter is to discuss what the results of this thesis bring to the relevant literature concerning environmental determinants of gene-specific DNA methylation in light of the strengths and limitations of the study.

7.1 Summary of key findings

This cross-sectional analysis used data available from a case-control study nested within the CARTaGENE study. The results of the main analyses revealed that tobacco smoking was associated with lower average methylation levels in the *AHRR* and *F2RL3* genes. Specifically, at baseline, current and former smokers had significantly lower methylation levels in the *AHRR* and *F2RL3* genes in comparison with never smokers. There was no observable difference in the association between cases and controls, and between men and women.

Moreover, no association between occupational exposure to the selected agents and average methylation levels of the *AHRR* and *F2RL3* genes was observed. Sensitivity analyses using participants' current employment at baseline, instead of their longest-held job, did not reveal any association. There was no observable difference between cases and controls, and between men and women.

Sensitivity analyses considering the methylation levels of individual CpG sites within the *AHRR* and *F2RL3* genes, instead of average methylation levels in the main analyses, revealed a significant association between tobacco smoking and lower methylation levels in 17 *AHRR* CpG sites out of 33 and six *F2RL3* CpG sites out of seven. Additional analyses performed after the exclusion of participants with more than 10% methylation information missing for the gene *AHRR* or *F2RL3* did not reveal any observable differences with the results of the main methylation analyses.

7.2 Comparison with the relevant literature

7.2.1 Smoking-methylation analysis

The results obtained in this thesis support the consistent association reported between tobacco smoking and hypomethylation of the *AHRR* and F*2RL3* genes. Previous studies ranged in sample size from 192 to 2272 individuals and used different media for methylation measurements. Specifically, a prospective EWAS comparing DNA methylation levels of the *AHRR* and *F2RL3* genes measured in whole blood between current, former, and never smokers reported that DNA methylation levels of five CpG sites within the *AHRR* and *F2RL3* genes were significantly lower in current compared to never smokers (115). Another prospective study based on two EWAS nested in the European Prospective Investigation into Cancer and Nutrition compared DNA methylation measured in lung tissue and similarly observed that in current smokers and former smokers compared to non-smokers, three CpG sites from the *AHRR* gene and one CpG site from the *F2RL3* gene were hypomethylated (114). These inverse associations have all been replicated by four other studies that took a candidate gene approach and two case-control studies that quantified between one and two CpG sites for each gene (18, 116), and two prospective cohort studies that measured between two and eleven CpG sites for the *AHRR* gene and one to two CpG sites for the *F2RL3* gene (117, 118). All these studies estimated a percent difference in methylation between current smokers and never smokers of between -7% and -22% and between -8% and -18% for *AHRR* and *F2RL3*, respectively. The results obtained in this study, using a more comprehensive representation and parameterization of smoking through a CSI, and measuring substantially more CpG sites in the *AHRR* and *F2RL3* genes, fall within these ranges for each gene. Therefore, they corroborate the observations made in the studies previously mentioned and indeed suggest an association between tobacco smoking and *AHRR* and *F2RL3* hypomethylation.

Despite the consistency of these reported associations, previous studies quantified methylation levels from a relatively few number of CpG sites. For *AHRR*, only cg05575921 and two to eleven CpG sites within a 35 bp distance were measured; for *F2RL3*, only cg03636183 and one to three CpG sites within a 20 bp distance were measured. Our study was able to measure substantially more CpG sites in the promoter regions of each gene (33 for the *AHRR* gene and seven for the *F2RL3* gene), and therefore provide a better estimate of regional methylation patterns which is arguably more representative of gene expression levels. Our smoking-methylation analyses identified 17 hypomethylated CpG sites in the *AHRR* gene; 14 of which are located within 150 bp of cg05575921. Six hypomethylated sites in the *F2RL3* gene were associated with smoking, all located within 50 bp of cg03636183. Aberrant methylation patterns in these individual sites can be explained by their direct proximity with cg05575921 (*AHRR*) and cg03636183 (*F2RL3*), respectively, two CpG sites considered markers of smoking behaviors (158). Thus, methylation measurements from these CpG sites reveal that regional hypomethylation patterns in the promoter region of each gene were associated with tobacco smoking.

In addition, previous studies did not consider several aspects of the participants' smoking history such as duration and intensity of smoking, and time since cessation (where applicable). This thesis built upon the previous work and parameterized smoking using a CSI which allowed for a continuous and more comprehensive representation of smoking behavior among the participants. In the context of *AHRR* and *F2RL3* methylation, this study is the first to incorporate all these different aspects of smoking to better represent the participants' smoking history and behavior. Our results concur with findings from previous studies investigating smoking and *AHRR* and *F2RL3* hypomethylation.

Our results, stratified by case-control status and by sex, did not reveal any difference in the association between smoking and the methylation levels of the genes *AHRR* and *F2RL3*. The results for the sex-stratified analysis are concordant with the only study that investigated the association between smoking and *AHRR* and *F2RL3* methylation by sex and which also reported no differences in association (115). Our results stratifying by case-control status were similar to those from a study employing a nested case-control design which reported similar smoking-related methylation levels in the *AHRR* and *F2RL3* in cases (i.e., participants with incident lung cancer during the follow-up period) and in healthy controls (i.e., participants without incident lung cancer during the follow-up period) (18).

7.2.2 Occupation-methylation analysis

No association was observed between occupational exposures to the selected 18 agents, that are commonly found in tobacco smoke, and methylation of the *AHRR* and *F2RL3* genes. Only one previous study investigated the relationship between occupational exposures and methylation levels of *AHRR* and *F2RL3*. Specifically, the study reported an association between occupational exposure to PAHs and lower methylation levels in the two genes (122).

Broadening, the literature review to encompass global DNA methylation revealed that multiple studies have examined occupational exposures to our selected agents in relation to global DNA methylation. These studies consistently reported an association between occupational exposures to these agents and aberrant global DNA methylation patterns (145-149). And even though the extrapolation of results from global DNA methylation to gene-specific methylation must be done with caution, as gene-specific methylation patterns do not always correlate with global methylation patterns, this body of literature is useful in order to better comprehend the association between occupational exposures and DNA methylation in the context of this thesis (159, 160). The results of the occupational exposure-methylation analyses found in our study do not support the findings reported by the previous studies.

All this being said, the result of this thesis should also be approached and interpreted considering possible bias that could have been introduced in the study.

7.3 Study validity: strengths and limitations

7.3.1 Selection bias

Selection bias arises from the procedures by which study participants are selected from the source population, or select themselves by agreeing to participate, and from factors that influence study participation (161, 162). Such a bias is introduced when the study population does not represent the source population in terms of exposure status and outcome, in this case, the CARTaGENE cohort (163). In the context of this cross-sectional study, available information from participants was obtained from an ongoing case-control study nested within CARTaGENE to elucidate the association between smoking and occupational exposures, and DNA methylation in the *AHRR* and *F2RL3* genes. Specifically, the nested case-control study was based on 200 incident cases of lung cancer that occurred during the follow-up period (from baseline to 2016) and who donated a blood sample, and 400 controls who had not developed lung cancer by the end of the follow-up period, and who had DNA isolated from a blood sample donated at baseline. Both cases and controls come from the same source population and share similar characteristics as controls were frequency-matched to cases by age, sex, and phase of blood sampling, and based on a ratio of two 2:1. However, though the cross-sectional association between environmental exposures and DNA methylation was assessed using information and blood samples collected at baseline (i.e., prior to lung cancer diagnosis), it was unknown whether the association found in the case group would be reflective of the exposure-outcome relationship in the larger source population. We tested this concern via stratified analysis by case-control status and the consistent associations observed in both case and control groups lend confidence that selection bias is not of concern in our study.

7.3.2 Information bias related to exposure assessment

Information bias occurs when the study variables are inaccurately measured or classified (164). For instance, it can happen when there is a systematic difference between groups of participants in the accuracy of the information collected or recalled (165). In the context of our study, it is very unlikely that participants could have been misreporting their smoking behavior and/or their occupational information according to their *AHRR* and *F2RL3* methylation levels as they did not know their methylation levels, thus limiting the potential for differential misclassification.

Nevertheless, collecting smoking information from self-assessed questionnaires can still lead to non-differential misclassification. In fact, errors can come from diverse reasons, including false reporting, inaccurate reporting, or memory failure among others (166). This being said, the results obtained, which are in agreement with findings from several previous studies using self-assessed questionnaires and/or interviews, suggest that the influence of such misclassification was minimal.

Besides, non-differential misclassification is still a potential limitation, particularly with respect to the assessment of occupational exposures. For instance, using exposure information estimated from the longest-held occupation rather than the participants' entire work history may have introduced exposure misclassification. However, since more than 61% of individuals in CARTaGENE population have held only one job, and even among those with more than one job the longest-held job still represented the majority of their working life (123), it is expected that the exposure misclassification associated with the use of the longest-held job to represent lifetime exposure is minimal.

CANJEM is a Canadian JEM that has been used to estimate exposure to selected occupational agents in this study, but there are some shortcomings to using a JEM for exposure assessment that deserve mention. Although it is a reliable tool to assess occupational exposures (132, 167), it cannot differentiate intra-group occupational exposure variations (137). Specifically, considering that a job may be composed of multiple tasks, each with its own specific exposure profile, assigning the same mean exposure to all workers sharing the same job title can result in exposure misclassification (168). In the context of this study, where people did not know their methylation status, the misclassification would have been non-differential and the direction of the bias would have been towards the null value, contributing to the absence of an association between occupational exposures to the selected agents and *AHRR* and *F2RL3* methylation.

A study evaluating the validity of CANJEM in contrast with the expert assessment approach examined the impact of different approaches for exposure categorization using the probability of exposure. Probability of exposure thresholds between 25% and 50% were reported to be the most valid (153). The choice of an exposure threshold affects the balance between sensitivity and specificity with higher sensitivity, but lower specificity, associated with higher exposure thresholds, and vice versa. In this study, due to the low prevalence of occupational exposure to our agents of interest (79% of participants were unexposed to any of the agents retained based on a probability of exposure threshold of 25%), analyses were performed using a probability of exposure threshold of 25%, and not 50%, to include a larger number of potentially exposed individuals. Analogously, the use of a higher occupational exposure threshold in CANJEM would have increased sensitivity and decreased the number of exposed participants (154). The use of a 25% versus 50% probability of exposure threshold might have contributed to non-differential exposure misclassification with a greater number of unexposed participants categorized as exposed and therefore biasing the association estimate towards the null value (153).

Finally, the occupational analyses were conducted based on categorical definitions of occupational exposure status (i.e., unexposed, exposed to one agent, exposed to two or more, and exposed/unexposed to the most prevalent agents) while intensity and duration of exposure were not considered which could have contributed to exposure misclassification. In fact, it is possible that the study participants were exposed at a lower level of intensity and frequency than those in previously reported studies which could explain the absence of an observed association in our study.

7.3.3 DNA methylation measurement

The quantification of DNA methylation levels of the genes *AHRR* and *F2RL3* were performed using the Sequenom EpiTYPER® technology which uses base-specific cleavage and laser desorption/ionization-time of flight mass spectrometry (124-126). This method is highly precise, accurate, and cost-effective allowing for DNA methylation measurements at a single-nucleotide resolution (150). As a matter of fact, a CV of 4.65% and 4.16% was estimated between plates and between fragments, respectively, based on the high methylated DNA quality controls. In the context of this study, those values suggest that the DNA methylation measurements performed with this technology are consistent and reliable.

7.3.4 Confounding

Confounding happens when the effects of the exposure of interest on a specific outcome are mixed with the effects of at least one additional factor, leading to bias (169). All in all, it occurs when the effects of two associated exposures have not been separated. It is a potential threat to the internal validity of an epidemiological study as it complicates the interpretation of the results by making it difficult to determine a clear causal association between the exposure and outcome. Directed acyclic graph, or DAG, is a common knowledge-based approach to select confounding factors to retain in the analyses. It is well-suited for situations where there is already an important body of literature concerning possible determinants of a given outcome (170). DNA methylation is an epigenetic mechanism that has been extensively studied. Potential confounding factors were identified through a literature review and DAGs were used to determine those to include in the regression models.

However, we cannot exclude the possibility of residual confounding in our analysis. For instance, SES is a complex variable that can be determined by several indicators such as income, financial debts and assets, poverty level, level of education, family size, access to quality health care facilities and social services among others. In this study, information about all these factors was unfortunately unavailable. Therefore, annual household income and education level were used instead to represent SES, but the use of these surrogates might not completely encapsulate the participants' SES which can lead to potential residual confounding.

Furthermore, it is now commonly accepted that DNA methylation shows substantial variation across individual cell types (156). Given that there are several cell types in peripheral blood, blood cell composition may confound the underlying association. As there was no reference data set available for comparison in the study, it was not possible to adjust the analyses for it (157). But the similar smoking-methylation association observed in our study in relation to others that adjusted for blood cell composition (18) supports that our estimates were minimally impacted by blood cell composition.

7.3.5 Temporality

The temporality of an association is a critical aspect of causality. Generally, one of the possible drawbacks of cross-sectional studies is their limitation in establishing temporal associations between exposures of interest and the outcome because both are examined at the same time. However, we conducted a cross-sectional analysis of data from a case-control study nested in a prospective cohort. It is evident that *AHRR* and *F2RL3* methylation did not influence smoking behaviors or occupational exposures. As a consequence, the temporality of the association between our exposures (i.e., smoking and occupational exposures to the selected agents) and the outcome of interest (i.e., DNA methylation of the *AHRR* and *F2RL3* genes) is clearly determined and should not be considered as a study limitation.

7.3.6 Design of the study

At the conception phase of this study, minimum detectable differences in DNA methylation levels effects were calculated for smoking and occupational exposures at 80% power and an α level of 0.05 (Appendix IV). Contrasting current versus never smokers, the minimal detectable effect size was estimated to be ±0.29; for occupational exposures, it was estimated to range from ±0.45 (PAHS from any source, 17.1% prevalence) to ±0.79 (Formaldehyde, 5.1% prevalence) when contrasting those exposed versus never exposed to individual agents. However, contrary to our initial prevalence estimates, the prevalence of occupational exposures to the retained chemical agents was lower than estimated and the exclusion of 54 participants (10.2%) due to the absence of a job code linkable to CANJEM further decreased the sample size in the occupation-methylation analysis; hence, limiting our statistical power to detect an association if it truly exists. Overall, with a probability of exposure threshold of 25%, 79% of the participants were unexposed to any of the 18 chemical agents retained. Future directions should consider exposure contrasts among industry-specific populations who have a distribution of exposure to these agents to inform on the possible association between occupational exposures to the selected agents and *AHRR* and *F2RL3* methylation.

7.3.7 Additional considerations

Exposure routes can influence the extent to which occupational agents affect the organism in terms of the duration and magnitude of DNA methylation (171, 172). Chemical agents are inhaled during smoking, but there are more exposure routes in the occupational environment according to the Canadian Centre for Occupational Health and Safety (ingestion, injection, and absorption through skin and eyes) (173). Thus, study participants were not necessarily occupationally exposed to the retained agents by inhalation which can contribute to the different results observed between the smoking- and the occupation-methylation analysis.

In addition, tobacco smoking exposes smokers to the 18 chemical agents retained at once which can result in a cumulative effect on DNA methylation levels. In contrast, only 15.1% of the study participants were exposed to two or more agents (Table 5.1). Therefore, in the occupation-methylation analysis, only a minority of participants were potentially exposed to a cumulative effect from the selected occupational agents on *AHRR* and *F2RL3* methylation.

The analyses performed using individual CpG sites (Table 6.1 to 6.4) did not account for multiple testing as they were only complementary analyses. On top of that, adjusting for multiple testing is not needed as the results of those additional analyses are consistent, and point toward the same observations established in the main analyses (i.e., hypomethylation of *AHRR* and *F2RL3* in relation to tobacco smoking, and no association between occupational exposure and *AHRR* and *F2RL3* methylation).

7.4 External validity

External validity is determined by the generalizability of the study results to persons or groups other than the original study population (174). The extent to which our results are applicable to other groups or populations is an important aspect of the study's external validity. CARTaGENE is Quebec's largest ongoing prospective cohort study including residents between the ages of 40 and 69 in metropolitan areas of Quebec (Montreal, Quebec, Sherbrooke, and Saguenay) which represent a total of 55.7% of the Quebec population. The participants of CARTaGENE were randomly recruited from a stratified sampling approach based on provincial official health insurance registries to be representative of the Quebec population (123). Our study participants share very similar socio-demographic characteristics with the CARTaGENE cohort. Hence, the findings of this study, using data from a case-control study nested in CARTaGENE, could be generalized to the Quebec population.

Ethnicity and SES, which is partly determined by the level of education, are factors influencing global and gene-specific methylation levels. In the context of this study, the association estimates (i.e., beta coefficients) were calculated from a study population that is mostly white (94.7%) and well-educated (university was the highest level of education for 41.7% of our study population). Overall, we can expect that smoking would result in similar levels of hypomethylation of the *AHRR* and *F2RL3* genes in other populations sharing similar socio-demographic characteristics. However, for populations with significantly more ethnic diversity and a different SES, even though hypomethylation of the two genes in question in relation to smoking is expected, the association estimates could be substantially different and not included in the 95% confidence intervals found in our analyses.

Pertaining to the occupation-methylation analysis, the generalizability of the results to other populations other than the study population should be done with caution. In fact, several limits have been discussed previously in the context of this thesis. Among them, the low prevalence of occupational exposure among the study participants was a major consideration. Consequently, the results obtained in this analysis might not be applicable to other populations. Indeed, a study specifically based on workers in the industries most exposed to these agents, instead of the general population, can be more suitable in order to estimate the association between occupational exposures to the selected agents and *AHRR* and *F2RL3* methylation.

7.5 Conclusion and future directions

This thesis took advantage of the unique opportunity offered by the ongoing CIHR-funded cumulative incidence case-control study nested in CARTaGENE (PIs: Vikki Ho and Anita

Koushik) and the availability of CANJEM to investigate the role of smoking and occupational exposures on the methylation levels of two lung cancer-related genes *AHRR* and *F2RL3*. The results obtained in this cross-sectional study suggest an association between smoking and lower average methylation levels in both genes, but they do not indicate any significant association between the selected occupational exposures and the methylation levels of the two genes of interest.

Future studies with higher statistical power, possibly based on workers involved in the industries most exposed to the selected agents, and possibly incorporating frequency and intensity of exposure, are needed to explore more in-depth the role of occupational exposures in the methylation of *AHRR* and *F2RL3*, and to substantiate or contradict our observations.

DNA methylation is an epigenetic mechanism involved in many cellular and biochemical processes. It plays a role in gene expression and can act as a biomarker for potential carcinogens. And, because gene-specific DNA methylation can constitute an early event in lung cancer development, it is important to comprehend how specific risk factors, such as smoking, can increase lung cancer risks through the modification of DNA methylation patterns in certain genes. Beyond that, since workers spend a large portion of their lives in the working environment where they are more exposed to certain harmful chemical agents on a regular basis than the general population, it is also crucial to understand whether occupational exposures can affect epigenetic mechanisms such as DNA methylation in order to refine public health efforts in reducing DNA methylation-related health issues.

ETHICAL CONSIDERATIONS

Ethics approval has been obtained from the Comité d'Ethique de la Recherche (CER) of the CHUM on the 27th February 2019. Access to CARTaGENE data has been granted.

REFERENCES

1. Organization WH. Cancer 2021 [Available from: https://www.who.int/news-room/fact-sheets/detail/cancer.

2. Society CC. Lung Cancer Statistics. 2019.

3. Doll R, Peto R. The causes of cancer: quantitative estimates of avoidable risks of cancer in the United States today. J Natl Cancer Inst. 1981;66(6):1191-308.

4. Canada S. Portrait of Canada's Labour Force. 2018.

5. Labrèche F, Duguay P, Ostiguy C, Boucher A, Roberge B, Peters CE, et al. Estimating occupational exposure to carcinogens in Quebec. American Journal of Industrial Medicine. 2013;56(9):1040-50.

6. Spyratos D, Zarogoulidis P, Porpodis K, Tsakiridis K, Machairiotis N, Katsikogiannis N, et al. Occupational exposure and lung cancer. J Thorac Dis. 2013;5 Suppl 4:S440-5.

7. Gustavsson P, Jakobsson R, Nyberg F, Pershagen G, Jarup L, Scheele P. Occupational exposure and lung cancer risk: a population-based case-referent study in Sweden. Am J Epidemiol. 2000;152(1):32-40.

8. Moore LD, Le T, Fan G. DNA methylation and its basic function. Neuropsychopharmacology. 2013;38(1):23-38.

9. Mutize T, Mkandla Z, Nkambule BB. Global and gene-specific DNA methylation in adult type 2 diabetic individuals: a protocol for a systematic review. Systematic Reviews. 2018;7(1):46.

10. Tsou JA, Hagen JA, Carpenter CL, Laird-Offringa IA. DNA methylation analysis: a powerful new tool for lung cancer diagnosis. Oncogene. 2002;21(35):5450-61.

11. Vaissiere T, Hung RJ, Zaridze D, Moukeria A, Cuenin C, Fasolo V, et al. Quantitative analysis of DNA methylation profiles in lung cancer identifies aberrant DNA methylation of specific genes and its association with gender and cancer risk factors. Cancer Res. 2009;69(1):243-52.

12. Oikonomopoulou K, Hansen KK, Saifeddine M, Vergnolle N, Tea I, Diamandis EP, et al. Proteinase-mediated cell signalling: targeting proteinase-activated receptors (PARs) by kallikreins and more. Biol Chem. 2006;387(6):677-85.

13. Vogel CFA, Haarmann-Stemmann T. The aryl hydrocarbon receptor repressor - More than a simple feedback inhibitor of AhR signaling: Clues for its role in inflammation and cancer. Curr Opin Toxicol. 2017;2:109-19.

14. Black PC, Mize GJ, Karlin P, Greenberg DL, Hawley SJ, True LD, et al. Overexpression of protease-activated receptors-1,-2, and-4 (PAR-1, -2, and -4) in prostate cancer. Prostate. 2007;67(7):743-56.

15. Tsay JJ, Tchou-Wong KM, Greenberg AK, Pass H, Rom WN. Aryl hydrocarbon receptor and lung cancer. Anticancer Res. 2013;33(4):1247-56.

16. Zudaire E, Cuesta N, Murty V, Woodson K, Adams L, Gonzalez N, et al. The aryl hydrocarbon receptor repressor is a putative tumor suppressor gene in multiple human cancers. J Clin Invest. 2008;118(2):640-50.

17. Zhang Y, Elgizouli M, Schöttker B, Holleczek B, Nieters A, Brenner H. Smoking-associated DNA methylation markers predict lung cancer incidence. Clinical Epigenetics. 2016;8(1):127.

18. Fasanelli F, Baglietto L, Ponzi E, Guida F, Campanella G, Johansson M, et al. Hypomethylation of smoking-related genes is associated with future lung cancer in four prospective cohorts. Nat Commun. 2015;6:10192.

19. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA: A Cancer Journal for Clinicians. 2021;71(3):209-49.

20. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2018. CA Cancer J Clin. 2018;68(1):7-30.

21. Duma N, Santana-Davila R, Molina JR. Non-Small Cell Lung Cancer: Epidemiology, Screening, Diagnosis, and Treatment. Mayo Clin Proc. 2019;94(8):1623-40.

22. Wild C. WE, Stewart B. WORLD CANCER REPORT: cancer research for cancer development. Place of publication not identified: IARC; 2020 2020.

23. Society CC. A 2020 special report on lung cancer. 2020.

24. Gomperts BN, Spira A, Massion PP, Walser TC, Wistuba II, Minna JD, et al. Evolving Concepts in Lung Carcinogenesis. Semin Respir Crit Care Med. 2011;32(1):32-43.

25. Alberg AJ, Brock MV, Ford JG, Samet JM, Spivack SD. Epidemiology of Lung Cancer: Diagnosis and Management of Lung Cancer, 3rd ed: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines. CHEST. 2013;143(5):e1S-e29S.

26. Zappa C, Mousa SA. Non-small cell lung cancer: current treatment and future advances. Transl Lung Cancer Res. 2016;5(3):288-300.

27. Travis WD, Brambilla E, Nicholson AG, Yatabe Y, Austin JHM, Beasley MB, et al. The 2015 World Health Organization Classification of Lung Tumors: Impact of Genetic, Clinical and Radiologic Advances Since the 2004 Classification. J Thorac Oncol. 2015;10(9):1243-60.

28. CDC. Tobacco and Cancer: Centers for Disease Control and Prevention; 2020 [Available from: https://www.cdc.gov/cancer/tobacco/index.htm.

29. Lee PN, Forey BA, Coombs KJ. Systematic review with meta-analysis of the epidemiological evidence in the 1900s relating smoking to lung cancer. BMC Cancer. 2012;12(1):385.

30. Kim HJ, Fay MP, Feuer EJ, Midthune DN. Permutation tests for joinpoint regression with applications to cancer rates. Stat Med. 2000;19(3):335-51.

31. Whittemore AS, McMillan A. Lung cancer mortality among U.S. uranium miners: a reappraisal. J Natl Cancer Inst. 1983;71(3):489-99.

32. Poirier AE, Ruan Y, Grevers X, Walter SD, Villeneuve PJ, Friedenreich CM, et al. Estimates of the current and future burden of cancer attributable to active and passive tobacco smoking in Canada. Prev Med. 2019;122:9-19.

33. Rushton L. The Global Burden of Occupational Disease. Curr Environ Health Rep. 2017;4(3):340-8.

34. CCHOS. Occupational Hygiene - Occupational Exposure Limits: Government of Canada; 2021 [Available from: https://www.ccohs.ca/oshanswers/hsprograms/occ_hygiene/occ_exposure_limits.html.

35. Driscoll T, Nelson DI, Steenland K, Leigh J, Concha-Barrientos M, Fingerhut M, et al. The global burden of disease due to occupational carcinogens. Am J Ind Med. 2005;48(6):419-31.

36. Agents Classified by the IARC Monographs, Volumes 1–130 – IARC Monographs on the Identification of Carcinogenic Hazards to Humans.

37. Labreche F, Kim J, Song C, Pahwa M, Ge CB, Arrandale VH, et al. The current burden of cancer attributable to occupational exposures in Canada. Prev Med. 2019;122:128-39.

38. Singh A, Kamal R, Ahamed I, Wagh M, Bihari V, Sathian B, et al. PAH exposure-associated lung cancer: an updated meta-analysis. Occup Med (Lond). 2018;68(4):255-61.

39. Bruske-Hohlfeld I, Mohner M, Pohlabeln H, Ahrens W, Bolm-Audorff U, Kreienbrock L, et al. Occupational lung cancer risk for men in Germany: results from a pooled case-control study. Am J Epidemiol. 2000;151(4):384-95.

40. Boffetta P, Jourenkova N, Gustavsson P. Cancer risk from occupational and environmental exposure to polycyclic aromatic hydrocarbons. Cancer Causes Control. 1997;8(3):444-72.

41. Nadon L, Siemiatycki J, Dewar R, Krewski D, Gérin M. Cancer risk due to occupational exposure to polycyclic aromatic hydrocarbons. American Journal of Industrial Medicine. 1995;28(3):303-24.

42. Kwak K, Paek D, Park J-T. Occupational exposure to formaldehyde and risk of lung cancer: A systematic review and meta-analysis. American Journal of Industrial Medicine. 2020;63(4):312-27.

43. Mastrangelo G, Fedeli U, Fadda E, Milan G, Turato A, Pavanello S. Lung cancer risk in workers exposed to poly(vinyl chloride) dust: a nested case-referent study. Occup Environ Med. 2003;60(6):423-8.

44. Mikoczy Z, Tinnerberg H, Björk J, Albin M. Cancer Incidence and Mortality in Swedish Sterilant Workers Exposed to Ethylene Oxide: Updated Cohort Study Findings 1972–2006. International Journal of Environmental Research and Public Health. 2011;8(6):2009-19.

45. CDC. Table 5.1, IARC evaluations of carcinogens in mainstream cigarette smoke. NCBI Center for Disease Control and Prevention. 2010.

46. t Mannetje A, Bencko V, Brennan P, Zaridze D, Szeszenia-Dabrowska N, Rudnai P, et al. Occupational exposure to metal compounds and lung cancer. Results from a multi-center case-control study in Central/Eastern Europe and UK. Cancer causes & control: CCC. 2011;22(12):1669-80.

47. Beveridge R, Pintos J, Parent M-É, Asselin J, Siemiatycki J. Lung cancer risk associated with occupational exposure to nickel, chromium VI, and cadmium in two population-based case–control studies in Montreal. American Journal of Industrial Medicine. 2010;53(5):476-85.

48. Wild P, Bourgkard E, Paris C. Lung Cancer and Exposure to Metals: The Epidemiological Evidence. In: Verma M, editor. Cancer Epidemiology: Modifiable Factors. Methods in Molecular Biology. Totowa, NJ: Humana Press; 2009. p. 139-67.

49. Warden H, Richardson H, Richardson L, Siemiatycki J, Ho V. Associations between occupational exposure to benzene, toluene and xylene and risk of lung cancer in Montréal. Occupational and Environmental Medicine. 2018;75(10):696-702.

50. Tomioka K, Saeki K, Obayashi K, Kurumatani N. Risk of Lung Cancer in Workers Exposed to Benzidine and/or Beta-Naphthylamine: A Systematic Review and Meta-Analysis. J Epidemiol. 2016;26(9):447-58.

51. Weinhold B. Epigenetics: The Science of Change. Environmental Health Perspectives. 2006;114(3):A160-A7.

52. Morgan HD, Santos F, Green K, Dean W, Reik W. Epigenetic reprogramming in mammals. Hum Mol Genet. 2005;14 Spec No 1:R47-58.

53. Lister R, Pelizzola M, Dowen RH, Hawkins RD, Hon G, Tonti-Filippini J, et al. Human DNA methylomes at base resolution show widespread epigenomic differences. Nature. 2009;462(7271):315-22.

54. Smith ZD, Meissner A. DNA methylation: roles in mammalian development. Nat Rev Genet. 2013;14(3):204-20.

55. Gopalakrishnan S, Van Emburgh BO, Robertson KD. DNA methylation in development and human disease. Mutat Res. 2008;647(1-2):30-8.

56. Robertson KD. DNA methylation and human disease. Nat Rev Genet. 2005;6(8):597-610.

57. Sincic N, Herceg Z. DNA methylation and cancer: ghosts and angels above the genes. Curr Opin Oncol. 2011;23(1):69-76.

58. Brzezianska E, Dutkowska A, Antczak A. The significance of epigenetic alterations in lung carcinogenesis. Mol Biol Rep. 2013;40(1):309-25.

59. Jones PA, Liang G. The Human Epigenome. In: Michels KB, editor. Epigenetic Epidemiology. Dordrecht: Springer Netherlands; 2012. p. 5-20.

60. Ehrlich M. DNA methylation in cancer: too much, but also too little. Oncogene. 2002;21(35):5400-13.

61. Wilson AS, Power BE, Molloy PL. DNA hypomethylation and human diseases. Biochim Biophys Acta. 2007;1775(1):138-62.

62. Sato N, Maitra A, Fukushima N, Heek NTv, Matsubayashi H, Iacobuzio-Donahue CA, et al. Frequent Hypomethylation of Multiple Genes Overexpressed in Pancreatic Ductal Adenocarcinoma. Cancer Research. 2003;63(14):4158-66.

63. Herman JG, Baylin SB. Promoter-region hypermethylation and gene silencing in human cancer. Curr Top Microbiol Immunol. 2000;249:35-54.

64. Rauscher GH, Kresovich JK, Poulin M, Yan L, Macias V, Mahmoud AM, et al. Exploring DNA methylation changes in promoter, intragenic, and intergenic regions as early and late events in breast cancer formation. BMC Cancer. 2015;15:816.

65. Helman E, Naxerova K, Kohane IS. DNA hypermethylation in lung cancer is targeted at differentiation-associated genes. Oncogene. 2012;31(9):1181-8.

66. Hulka BS, Wilcosky T. Biological markers in epidemiologic research. Arch Environ Health. 1988;43(2):83-9.

67. Perera FP, Weinstein IB. Molecular epidemiology: recent advances and future directions. Carcinogenesis. 2000;21(3):517-24.

68. Strimbu K, Tavel JA. What are biomarkers? Curr Opin HIV AIDS. 2010;5(6):463-6.

69. Villalobos P, Wistuba II. Lung Cancer Biomarkers. Hematol Oncol Clin North Am. 2017;31(1):13-29.

70. Scott A, Salgia R. Biomarkers in lung cancer: from early detection to novel therapeutics and decision making. Biomark Med. 2008;2(6):577-86.

71. Verma M, Patel P, Verma M. Biomarkers in Prostate Cancer Epidemiology. Cancers. 2011;3:3773-98.

72. Bennett MR, Devarajan P. Chapter 1 - Characteristics of an Ideal Biomarker of Kidney Diseases. In: Edelstein CL, editor. Biomarkers of Kidney Disease. San Diego: Academic Press; 2011. p. 1-24.

73. Sanchez-Cespedes M, Esteller M, Wu L, Nawroz-Danish H, Yoo GH, Koch WM, et al. Gene promoter hypermethylation in tumors and serum of head and neck cancer patients. Cancer Research. 2000;60(4):892-5.

74. Esteller M, Sanchez-Cespedes M, Rosell R, Sidransky D, Baylin SB, Herman JG. Detection of aberrant promoter hypermethylation of tumor suppressor genes in serum DNA from non-small cell lung cancer patients. Cancer Research. 1999;59(1):67-70.

75. Wani K, Aldape KD. PCR Techniques in Characterizing DNA Methylation. Methods in Molecular Biology (Clifton, NJ). 2016;1392:177-86.

76. Guo W, Zhu L, Yu M, Zhu R, Chen Q, Wang Q. A five-DNA methylation signature act as a novel prognostic biomarker in patients with ovarian serous cystadenocarcinoma. Clinical Epigenetics. 2018;10(1):142.

77. Šestáková Š, Šálek C, Remešová H. DNA Methylation Validation Methods: a Coherent Review with Practical Comparison. Biological Procedures Online. 2019;21(1):19.

78. Houseman EA, Kim S, Kelsey KT, Wiencke JK. DNA Methylation in Whole Blood: Uses and Challenges. Curr Envir Health Rpt. 2015;2(2):145-54.

79. Baccarelli A, Wright RO, Bollati V, Tarantini L, Litonjua AA, Suh HH, et al. Rapid DNA Methylation Changes after Exposure to Traffic Particles. Am J Respir Crit Care Med. 2009;179(7):572-8.

80. Palumbo D, Affinito O, Monticelli A, Cocozza S. DNA Methylation variability among individuals is related to CpGs cluster density and evolutionary signatures. BMC Genomics. 2018;19(1):229.

81. Salameh Y, Bejaoui Y, El Hajj N. DNA Methylation Biomarkers in Aging and Age-Related Diseases. Front Genet. 2020;11.

82. Jung M, Pfeifer GP. Aging and DNA methylation. BMC Biology. 2015;13(1):7.

83. Hall E, Volkov P, Dayeh T, Esguerra JLS, Salö S, Eliasson L, et al. Sex differences in the genome-wide DNA methylation pattern and impact on gene expression, microRNA levels and insulin secretion in human pancreatic islets. Genome Biol. 2014;15(12):522.

84. Boks MP, Derks EM, Weisenberger DJ, Strengman E, Janson E, Sommer IE, et al. The Relationship of DNA Methylation with Age, Gender and Genotype in Twins and Healthy Controls. PLOS ONE. 2009;4(8):e6767.

85. Alegría-Torres JA, Baccarelli A, Bollati V. Epigenetics and lifestyle. Epigenomics. 2011;3(3):267-77.

86. Mendelson MM, Marioni RE, Joehanes R, Liu C, Hedman ÅK, Aslibekyan S, et al. Association of Body Mass Index with DNA Methylation and Gene Expression in Blood Cells and Relations to Cardiometabolic Disease: A Mendelian Randomization Approach. PLOS Medicine. 2017;14(1):e1002215.

87. Reed ZE, Suderman MJ, Relton CL, Davis OSP, Hemani G. The association of DNA methylation with body mass index: distinguishing between predictors and biomarkers. Clinical Epigenetics. 2020;12(1):50.

88. Zhang FF, Cardarelli R, Carroll J, Zhang S, Fulda KG, Gonzalez K, et al. Physical activity and global genomic DNA methylation in a cancer-free population. Epigenetics. 2011;6(3):293-9.

89. White AJ, Sandler DP, Bolick SCE, Xu Z, Taylor JA, DeRoo LA. Recreational and household physical activity at different time points and DNA global methylation. European Journal of Cancer. 2013;49(9):2199-206.

90. Lee K, Pausova Z. Cigarette smoking and DNA methylation. Front Genet. 2013;4.

91. Tsaprouni LG, Yang T-P, Bell J, Dick KJ, Kanoni S, Nisbet J, et al. Cigarette smoking reduces DNA methylation levels at multiple genomic loci but the effect is partially reversible upon cessation. Epigenetics. 2014;9(10):1382-96.

92. Dogan MV, Shields B, Cutrona C, Gao L, Gibbons FX, Simons R, et al. The effect of smoking on DNA methylation of peripheral blood mononuclear cells from African American women. BMC Genomics. 2014;15(1):151.

93. Hibler E, Huang L, Andrade J, Spring B. Impact of a diet and activity health promotion intervention on regional patterns of DNA methylation. Clinical Epigenetics. 2019;11(1):133.

94. Zakhari S. Alcohol Metabolism and Epigenetics Changes. Alcohol Res. 2013;35(1):6-16.

95. Varela-Rey M, Woodhoo A, Martinez-Chantar M-L, Mato JM, Lu SC. Alcohol, DNA Methylation, and Cancer. Alcohol Res. 2013;35(1):25-35.

96. McDade TW, Ryan CP, Jones MJ, Hoke MK, Borja J, Miller GE, et al. Genome-wide analysis of DNA methylation in relation to socioeconomic status during development and early adulthood. American Journal of Physical Anthropology. 2019;169(1):3-11.

97. Stringhini S, Polidoro S, Sacerdote C, Kelly RS, van Veldhoven K, Agnoli C, et al. Life-course socioeconomic status and DNA methylation of genes regulating inflammation. International Journal of Epidemiology. 2015;44(4):1320-30.

98. Needham BL, Smith JA, Zhao W, Wang X, Mukherjee B, Kardia SLR, et al. Life course socioeconomic status and DNA methylation in genes related to stress reactivity and inflammation: The multi-ethnic study of atherosclerosis. Epigenetics. 2015;10(10):958-69.

99. Galanter JM, Gignoux CR, Oh SS, Torgerson D, Pino-Yanes M, Thakur N, et al. Differential methylation between ethnic sub-groups reflects the effect of genetic ancestry and environmental exposures. eLife.6.

100. Park SL, Patel YM, Loo LWM, Mullen DJ, Offringa IA, Maunakea A, et al. Association of internal smoking dose with blood DNA methylation in three racial/ethnic populations. Clinical Epigenetics. 2018;10(1):110.

101. Zhang FF, Cardarelli R, Carroll J, Fulda KG, Kaur M, Gonzalez K, et al. Significant differences in global genomic DNA methylation by gender and race/ethnicity in peripheral blood. Epigenetics. 2011;6(5):623-9.

102. Feinberg AP, Vogelstein B. Hypomethylation distinguishes genes of some human cancers from their normal counterparts. Nature. 1983;301(5895):89-92.

103. Haarmann-Stemmann T, Abel J. The arylhydrocarbon receptor repressor (AhRR): structure, expression, and function. Biol Chem. 2006;387(9):1195-9.

104. Cauchi S, Stucker I, Cenee S, Kremers P, Beaune P, Massaad-Massade L. Structure and polymorphisms of human aryl hydrocarbon receptor repressor (AhRR) gene in a French population: relationship with CYP1A1 inducibility and lung cancer. Pharmacogenetics. 2003;13(6):339-47.

105. Esser C. Biology and function of the aryl hydrocarbon receptor: report of an international and interdisciplinary conference. Arch Toxicol. 2012;86(8):1323-9.

106. Elbekai RH, El-Kadi AO. Modulation of aryl hydrocarbon receptor-regulated gene expression by arsenite, cadmium, and chromium. Toxicology. 2004;202(3):249-69.

107. Kasai A, Hiramatsu N, Hayakawa K, Yao J, Maeda S, Kitamura M. High levels of dioxin-like potential in cigarette smoke evidenced by in vitro and in vivo biosensing. Cancer Res. 2006;66(14):7143-50.

108. Tsay JJ, Tchou-Wong K-M, Greenberg AK, Pass H, Rom WN. Aryl Hydrocarbon Receptor and Lung Cancer. Anticancer research. 2013;33(4):1247-56.

109. Mimura J, Fujii-Kuriyama Y. Functional role of AhR in the expression of toxic effects by TCDD. Biochim Biophys Acta. 2003;1619(3):263-8.

110. Nebert DW, Dalton TP, Okey AB, Gonzalez FJ. Role of aryl hydrocarbon receptor-mediated induction of the CYP1 enzymes in environmental toxicity and cancer. J Biol Chem. 2004;279(23):23847-50.

111. Kahn ML, Nakanishi-Matsui M, Shapiro MJ, Ishihara H, Coughlin SR. Protease-activated receptors 1 and 4 mediate activation of human platelets by thrombin. J Clin Invest. 1999;103(6):879-87.

112. Han NA, Jin K, He K, Cao J, Teng L. Protease-activated receptors in cancer: A systematic review. Oncol Lett. 2011;2(4):599-608.

113. French SL, Hamilton JR. Protease-activated receptor 4: from structure to function and back again. Br J Pharmacol. 2016;173(20):2952-65.

114. Shenker NS, Polidoro S, van Veldhoven K, Sacerdote C, Ricceri F, Birrell MA, et al. Epigenome-wide association study in the European Prospective Investigation into Cancer and Nutrition (EPIC-Turin) identifies novel genetic loci associated with smoking. Hum Mol Genet. 2013;22(5):843-51.

115. Zeilinger S, Kuhnel B, Klopp N, Baurecht H, Kleinschmidt A, Gieger C, et al. Tobacco smoking leads to extensive genome-wide changes in DNA methylation. PLoS One. 2013;8(5):e63812.

116. Baglietto L, Ponzi E, Haycock P, Hodge A, Bianca Assumma M, Jung CH, et al. DNA methylation changes measured in pre-diagnostic peripheral blood samples are associated with smoking and lung cancer risk. Int J Cancer. 2017;140(1):50-61.

117. Dogan MV, Shields B, Cutrona C, Gao L, Gibbons FX, Simons R, et al. The effect of smoking on DNA methylation of peripheral blood mononuclear cells from African American women. BMC Genomics. 2014;15:151.

118. Elliott HR, Tillin T, McArdle WL, Ho K, Duggirala A, Frayling TM, et al. Differences in smoking associated DNA methylation patterns in South Asians and Europeans. Clin Epigenetics. 2014;6(1):4.

119. Yang M, Park JY. DNA Methylation in Promoter Region as Biomarkers in Prostate Cancer. Methods in molecular biology (Clifton, NJ). 2012;863:67-109.

120. Du X, Han L, Guo A-Y, Zhao Z. Features of Methylation and Gene Expression in the Promoter-Associated CpG Islands Using Human Methylome Data. Comp Funct Genomics. 2012;2012:598987.

121. Anastasiadi D, Esteve-Codina A, Piferrer F. Consistent inverse correlation between DNA methylation of the first intron and gene expression across tissues and species. Epigenetics & Chromatin. 2018;11(1):37.

122. Alhamdow A, Lindh C, Hagberg J, Graff P, Westberg H, Krais AM, et al. DNA methylation of the cancer-related genes F2RL3 and AHRR is associated with occupational exposure to polycyclic aromatic hydrocarbons. Carcinogenesis. 2018;39(7):869-78.

123. Awadalla P, Boileau C, Payette Y, Idaghdour Y, Goulet JP, Knoppers B, et al. Cohort profile of the CARTaGENE study: Quebec's population-based biobank for public health and personalized genomics. Int J Epidemiol. 2013;42(5):1285-99.

124. Coolen MW, Statham AL, Gardiner-Garden M, Clark SJ. Genomic profiling of CpG methylation and allelic specificity using quantitative high-throughput mass spectrometry: critical evaluation and improvements. Nucleic Acids Res. 2007;35(18):e119.

125. Ehrich M, Nelson MR, Stanssens P, Zabeau M, Liloglou T, Xinarianos G, et al. Quantitative high-throughput analysis of DNA methylation patterns by base-specific cleavage and mass spectrometry. Proc Natl Acad Sci U S A. 2005;102(44):15785-90.

126. Ehrich M, Zoll S, Sur S, van den Boom D. A new method for accurate assessment of DNA quality after bisulfite treatment. Nucleic Acids Res. 2007;35(5):e29.

127. Ho V, Ashbury JE, Taylor S, Vanner S, King WD. Quantification of gene-specific methylation of DNMT3B and MTHFR using sequenom EpiTYPER®. Data Brief. 2015;6:39-46.

128. Hoffmann K, Krause C, Seifert B. The German Environmental Survey 1990/92 (GerES II): primary predictors of blood cadmium levels in adults. Arch Environ Health. 2001;56(4):374-9.

129. Leffondre K, Abrahamowicz M, Xiao Y, Siemiatycki J. Modelling smoking history using a comprehensive smoking index: application to lung cancer. Stat Med. 2006;25(24):4132-46.

130. Organization IL. ISCO - International Standard Classification of Occupations: International Labour Organization; 2010 [Available from: https://www.ilo.org/public/english/bureau/stat/isco/.

131. Ge CB, Friesen MC, Kromhout H, Peters S, Rothman N, Lan Q, et al. Use and Reliability of Exposure Assessment Methods in Occupational Case–Control Studies in the General Population: Past, Present, and Future. Ann Work Expo Health. 2018;62(9):1047-63.

132. Siemiatycki J, Fritschi L, Nadon L, Gerin M. Reliability of an expert rating procedure for retrospective assessment of occupational exposures in community-based case-control studies. Am J Ind Med. 1997;31(3):280-6.

133. Teschke K, Olshan AF, Daniels JL, De Roos AJ, Parks CG, Schulz M, et al. Occupational exposure assessment in case-control studies: opportunities for improvement. Occup Environ Med. 2002;59(9):575-93; discussion 94.

134. Siemiatycki J, Dewar R, Richardson L. Costs and statistical power associated with five methods of collecting occupation exposure information for population-based case-control studies. Am J Epidemiol. 1989;130(6):1236-46.

135. Descatha A, Evanoff BA, Andersen JH, Baca M, Buckner-Petty S, Fadel M, et al. Comparison Between a Self-Reported Job Exposure Matrix (JEM CONSTANCES) to an Expertise-Based Job Exposure Matrix (MADE) for Biomechanical Exposures. Journal of Occupational and Environmental Medicine. 2019;61(9):e399.

136. Offermans NS, Vermeulen R, Burdorf A, Peters S, Goldbohm RA, Koeman T, et al. Comparison of expert and job-exposure matrix-based retrospective exposure assessment of occupational carcinogens in The Netherlands Cohort Study. Occup Environ Med. 2012;69(10):745-51.

137. Coughlin SS, Chiazze L. Job-exposure matrices in epidemiologic research and medical surveillance. Occup Med. 1990;5(3):633-46.

138. Petersen SB, Flachs EM, Prescott EIB, Tjønneland A, Osler M, Andersen I, et al. Job-exposure matrices addressing lifestyle to be applied in register-based occupational health studies. Occupational and Environmental Medicine. 2018;75(12):890-7.

139. Siemiatycki J, Lavoue J. Availability of a New Job-Exposure Matrix (CANJEM) for Epidemiologic and Occupational Medicine Purposes. J Occup Environ Med. 2018;60(7):e324-e8.

140. IARC. Tobacco smoke and involuntary smoking. IARC Monogr Eval Carcinog Risks Hum. 2004;83:1-1438.

141. Taux d’activité (% de la population âgée de 15 ans et plus) (estimation nationale) - North America | Data.

142. OCRC. Burden of Occupational Cancer in Ontario: Major Workplace Carcinogens and Prevention of Exposure. Occupational Cancer Research Centre, Cancer Care Ontario; 2019.

143. Peinado MA. Hypomethylation of DNA. In: Schwab M, editor. Encyclopedia of Cancer. Berlin, Heidelberg: Springer; 2011. p. 1791-2.

144. IARC. IARC Monographs on the evaluation of Carcinogenic Risks to Humans: International Agency for Research on Cancer; 1996 1996.

145. Weihrauch M, Markwarth A, Lehnert G, Wittekind C, Wrbitzky R, Tannapfel A. Abnormalities of the ARF-p53 pathway in primary angiosarcomas of the liver. Hum Pathol. 2002;33(9):884-92.

146. Salemi R, Marconi A, Di Salvatore V, Franco S, Rapisarda V, Libra M. Epigenetic alterations and occupational exposure to benzene, fibers, and heavy metals associated with tumor development (Review). Molecular Medicine Reports. 2017;15(5):3366-71.

147. Leso V, Macrini MC, Russo F, Iavicoli I. Formaldehyde Exposure and Epigenetic Effects: A Systematic Review. Applied Sciences. 2020;10(7):2319.

148. Kovatsi L, Georgiou E, Ioannou A, Haitoglou C, Tzimagiorgis G, Tsoukali H, et al. p16 promoter methylation in Pb2+-exposed individuals. Clinical Toxicology. 2010;48(2):124-8.

149. Bollati V, Baccarelli A, Hou L, Bonzini M, Fustinoni S, Cavallo D, et al. Changes in DNA methylation patterns in subjects exposed to low-dose benzene. Cancer Res. 2007;67(3):876-80.

150. Suchiman HED, Slieker RC, Kremer D, Slagboom PE, Heijmans BT, Tobi EW. Design, measurement and processing of region-specific DNA methylation assays: the mass spectrometry-based method EpiTYPER. Front Genet. 2015;6:287.

151. Wilson R, Wahl S, Pfeiffer L, Ward-Caviness CK, Kunze S, Kretschmer A, et al. The dynamics of smoking-related disturbed methylation: a two time-point study of methylation change in smokers, non-smokers and former smokers. BMC Genomics. 2017;18(1):805.

152. Philibert R, Dogan M, Beach SRH, Mills JA, Long JD. AHRR methylation predicts smoking status and smoking intensity in both saliva and blood DNA. American Journal of Medical Genetics Part B: Neuropsychiatric Genetics. 2020;183(1):51-60.

153. Pasquet R. Methodological considerations of the Canadian job-exposure matrix and the evaluation of the risk of brain cancer in relation to occupational exposure to metallic compounds. 2019.

154. Xu M, Ho V, Lavoue J, Richardson L, Siemiatycki J. Concordance of Occupational Exposure Assessment between the Canadian Job-Exposure Matrix (CANJEM) and Expert Assessment of Jobs Held by Women. Ann Work Expo Health. 2022:wxac008.

155. Lowe R, Slodkowicz G, Goldman N, Rakyan VK. The human blood DNA methylome displays a highly distinctive profile compared with other somatic tissues. Epigenetics. 2015;10(4):274-81.

156. Houseman EA, Kelsey KT, Wiencke JK, Marsit CJ. Cell-composition effects in the analysis of DNA methylation array data: a mathematical perspective. BMC Bioinformatics. 2015;16(1):95.

157. Houseman EA, Accomando WP, Koestler DC, Christensen BC, Marsit CJ, Nelson HH, et al. DNA methylation arrays as surrogate measures of cell mixture distribution. BMC Bioinformatics. 2012;13:86.

158. Tantoh DM, Wu M-C, Chuang C-C, Chen P-H, Tyan YS, Nfor ON, et al. AHRR cg05575921 methylation in relation to smoking and PM2.5 exposure among Taiwanese men and women. Clinical Epigenetics. 2020;12(1):117.

159. Chanda S, Dasgupta UB, Guhamazumder D, Gupta M, Chaudhuri U, Lahiri S, et al. DNA hypermethylation of promoter of gene p53 and p16 in arsenic-exposed people with and without malignancy. Toxicol Sci. 2006;89(2):431-7.

160. Reichard JF, Puga A. Effects of arsenic exposure on DNA methylation and epigenetic gene regulation. Epigenomics. 2010;2(1):87-104.

161. Porta M. A Dictionary of Epidemiology: Oxford University Press; 2008 2008/07/03/. 559 p.

162. Pearce N, Checkoway H, Kriebel D. Bias in occupational epidemiology studies. Occupational and Environmental Medicine. 2007;64(8):562-8.

163. Delgado-Rodríguez M, Llorca J. Bias. J Epidemiol Community Health. 2004;58(8):635-41.

164. Althubaiti A. Information bias in health research: definition, pitfalls, and adjustment methods. J Multidiscip Healthc. 2016;9:211-7.

165. Everson TM, Marsit CJ. Chapter 4 - Epidemiological concepts in environmental epigenetics. In: Fry RC, editor. Environmental Epigenetics in Toxicology and Public Health. Translational Epigenetics. 22: Academic Press; 2020. p. 89-105.

166. Coughlin SS. Recall bias in epidemiologic studies. J Clin Epidemiol. 1990;43(1):87-91.

167. Fritschi L, Nadon L, Benke G, Lakhani R, Latreille B, Parent ME, et al. Validation of expert assessment of occupational exposures. Am J Ind Med. 2003;43(5):519-22.

168. Benke G, Sim M, Fritschi L, Aldred G. Beyond the job exposure matrix (JEM): the task exposure matrix (TEM). Ann Occup Hyg. 2000;44(6):475-82.

169. Rothman KJ. Epidemiology: An Introduction: OUP USA; 2012 2012/06/21/. 281 p.

170. Evans D, Chaix B, Lobbedez T, Verger C, Flahault A. Combining directed acyclic graphs and the change-in-estimate procedure as a novel approach to adjustment-variable selection in epidemiology. BMC Medical Research Methodology. 2012;12(1):156.

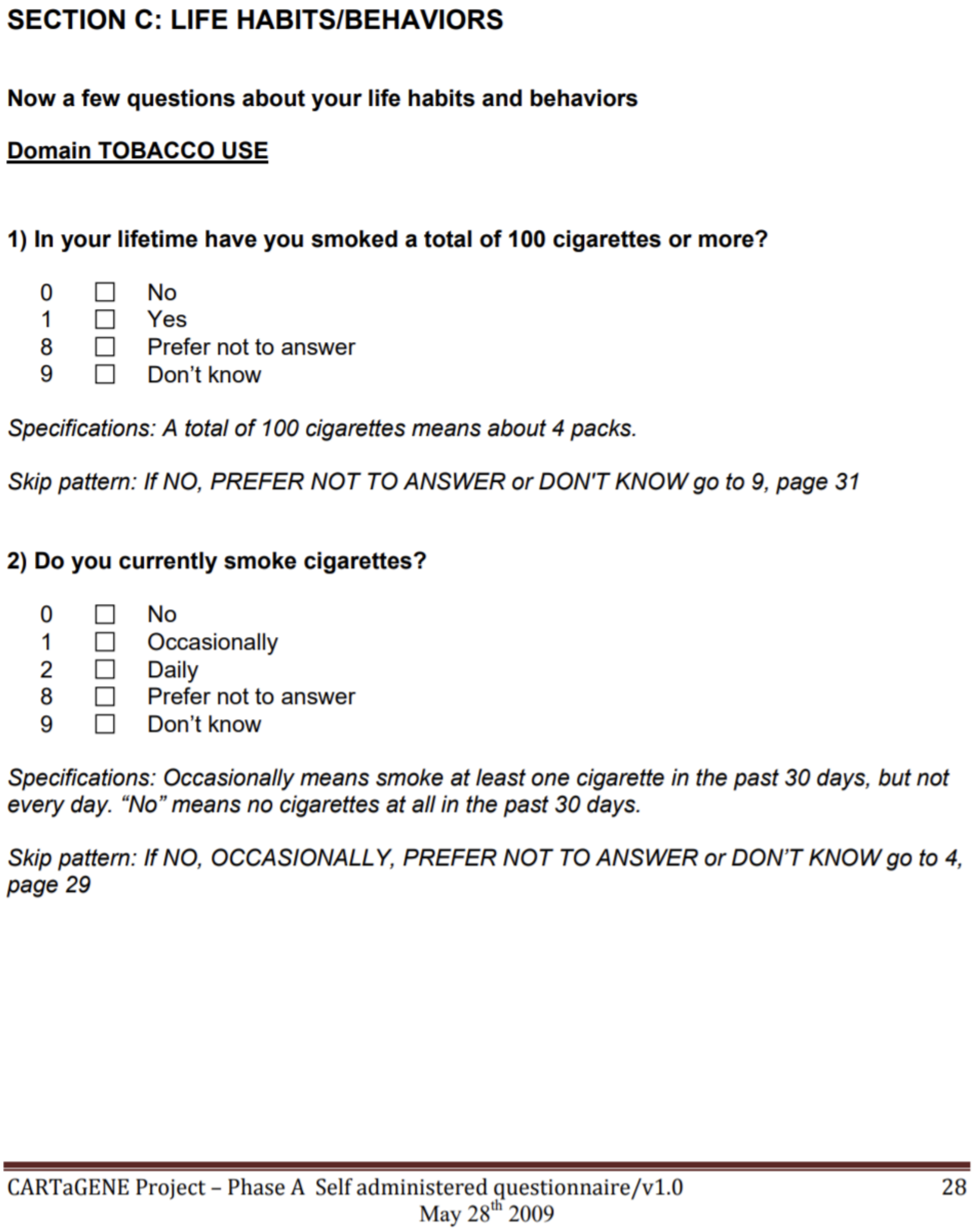
171. Ning ZH, Long S, Zhou YY, Peng ZY, Sun YN, Chen SW, et al. Effect of exposure routes on the relationships of lethal toxicity to rats from oral, intravenous, intraperitoneal and intramuscular routes. Regul Toxicol Pharmacol. 2015;73(2):613-9.

172. Lioy PJ. Assessing total human exposure to contaminants. A multidisciplinary approach. ACS Publications. 2002.

173. Government of Canada CCfOH, Safety. How Workplace Chemicals Enter the Body : OSH Answers. 2022.

174. Moher D, Schulz KF, Altman DG. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomised trials. Lancet. 2001;357(9263):1191-4.

Appendix I. Smoking history questionnaire



SECTION C: LIFE HABITS/BEHAVIORS

Now a few questions about your life habits and behaviors

Domain TOBACCO USE

1) In your lifetime have you smoked a total of 100 cigarettes or more?

0 □ No
1 □ Yes
8 □ Prefer not to answer
9 □ Don't know

Specifications: A total of 100 cigarettes means about 4 packs.

Skip pattern: If NO, PREFER NOT TO ANSWER or DON'T KNOW go to 9, page 31

2) Do you currently smoke cigarettes?

0 □ No
1 □ Occasionally
2 □ Daily
8 □ Prefer not to answer
9 □ Don't know

Specifications: Occasionally means smoke at least one cigarette in the past 30 days, but not every day. "No" means no cigarettes at all in the past 30 days.

Skip pattern: If NO, OCCASIONALLY, PREFER NOT TO ANSWER or DON'T KNOW go to 4, page 29

CARTaGENE Project – Phase A Self administered questionnaire/v1.0 28
May 28th 2009

3) On average how many cigarettes do you smoke per day or per week, including hand-rolled cigarettes?

Instructions: Use only one unit of time.

_____Cigarettes per day

OR

_____Cigarettes per week

888 □ Prefer not to answer
999 □ Don't know

Specification: A pack usually contains 25cigarettes

4) Have you ever smoked on most or all days?

0 □ No
1 □ Yes
8 □ Prefer not to answer
9 □ Don't know

Skip pattern: If NO, PREFER NOT TO ANSWER or DON'T KNOW go to 9, page 31.

5) During the period you smoked the most, either it is current or in the past, about how many cigarettes did you smoke?

____Cigarettes per day

OR

_____Cigarettes per week

888 □ Prefer not to answer
999 □ Don't know

Specification: A pack usually contains 25 cigarettes

6) For how long did this period last, in months or years?

Instructions: Use only one unit of time.

____Months

OR

____Years

88 □ Prefer not to answer
99 □ Don't know

7) What was your age when you first started smoking cigarettes on <u>most days</u>? Or in what year did you first start smoking cigarettes on <u>most days</u>?

____Age when started smoking on most days

OR

____Date when started smoking on most days (year)

8888 □ Prefer not to answer
9999 □ Don't know

Skip pattern: If the participant doesn't smoke cigarettes currently BUT has smoked cigarettes, go to 8, otherwise, go to 9, page 31.

8) What was your age when you last smoked cigarettes on <u>most days</u>? Or in what year did you last smoke cigarettes on <u>most days</u>?

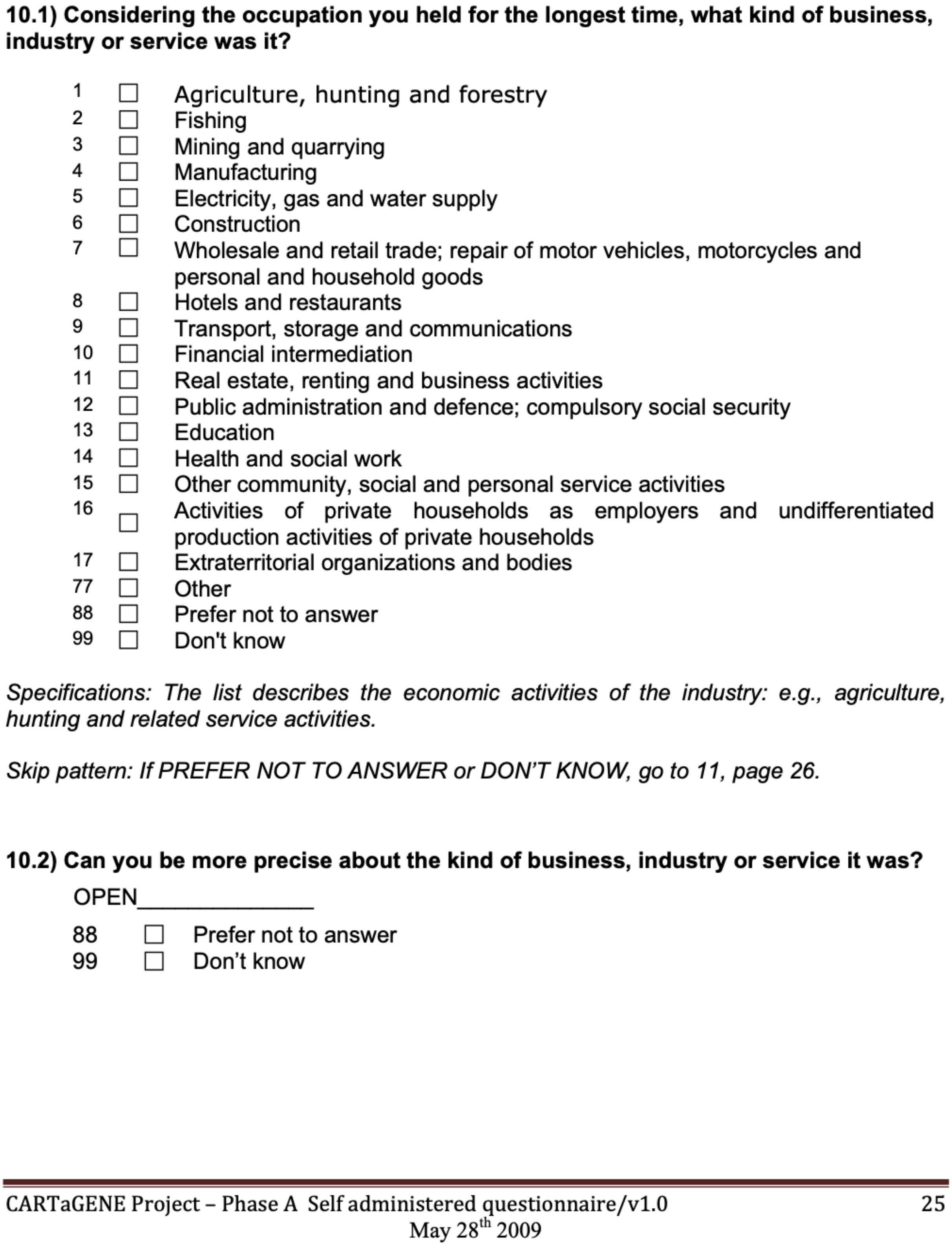
____Age when stopped smoking on most days

OR

____Date when stopped smoking on most days (year)

8888 □ Prefer not to answer
9999 □ Don't know

Appendix II. Longest-held job questionnaire



10.1) Considering the occupation you held for the longest time, what kind of business, industry or service was it?

1 □ Agriculture, hunting and forestry

2 □ Fishing

3 □ Mining and quarrying

4 □ Manufacturing

5 □ Electricity, gas and water supply

6 □ Construction

7 □ Wholesale and retail trade; repair of motor vehicles, motorcycles and personal and household goods

8 □ Hotels and restaurants

9 □ Transport, storage and communications

10 □ Financial intermediation

11 □ Real estate, renting and business activities

12 □ Public administration and defence; compulsory social security

13 □ Education

14 □ Health and social work

15 □ Other community, social and personal service activities

16 □ Activities of private households as employers and undifferentiated production activities of private households

17 □ Extraterritorial organizations and bodies

77 □ Other

88 □ Prefer not to answer

99 □ Don't know

Specifications: The list describes the economic activities of the industry: e.g., agriculture, hunting and related service activities.

Skip pattern: If PREFER NOT TO ANSWER or DON'T KNOW, go to 11, page 26.

10.2) Can you be more precise about the kind of business, industry or service it was?

OPEN______________

88 □ Prefer not to answer

99 □ Don't know

CARTaGENE Project – Phase A Self administered questionnaire/v1.0 25

May 28th 2009

11) What was the job title of the occupation that you have held for the longest time?

1 □ Legislators, senior-officials and managers
2 □ Professionals
3 □ Technicians and associate professionals
4 □ Clerks
5 □ Service workers and shop and market sales workers
6 □ Skilled agricultural and fishery workers
7 □ Craft and related workers
8 □ Plant and machine operators and assemblers
9 □ Elementary occupations
10 □ Armed forces
77 □ Other
88 □ Prefer not to answer
99 □ Don't know

12) What was your age when you started working there? Or, in what year did you start working there?

____Age when started working there

OR

____Date when started working there (year)

8888 □ Prefer not to answer
9999 □ Don’t know

13) Considering the occupation you held for the longest time, which of the following best describes your working schedule for this occupation?

1 □ Regular - daytime schedule or shift
2 □ Regular - evening shift
3 □ Regular - night shift
4 □ Rotating shift, changing periodically from days to evenings or to nights
5 □ Split shift, consisting of two or more distinct periods each day
6 □ Irregular schedule, or on call
7 □ Other
88 □ Prefer not to answer
99 □ Don't know

Specifications: A night shift is work during the early hours of the morning, after midnight. An evening shift is work during the evening ending at or before midnight.

14) What was your age when you stopped working there? Or, in what year did you stop working there?

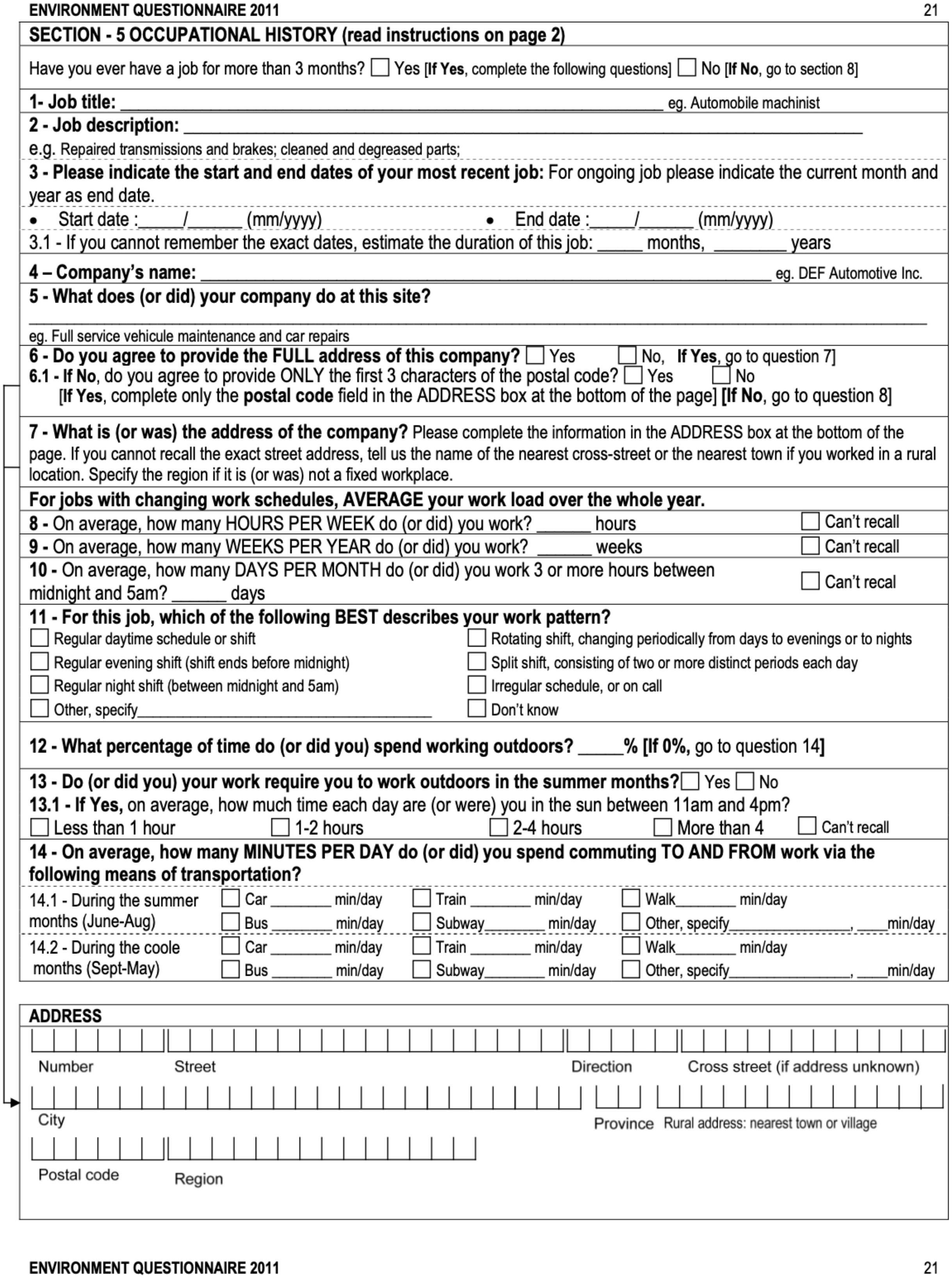
____Age when stopped working there

OR

____Date when stopped working there (year)

8888 □ Prefer not to answer
9999 □ Don't know

Appendix III. Job history questionnaire



ENVIRONMENT QUESTIONNAIRE 2011 21

SECTION - 5 OCCUPATIONAL HISTORY (read instructions on page 2)

Have you ever have a job for more than 3 months? □ Yes [**If Yes**, complete the following questions] □ No [**If No**, go to section 8]

1- Job title: ______________________________ eg. Automobile machinist

2 - Job description: ______________________________

e.g. Repaired transmissions and brakes; cleaned and degreased parts;

3 - Please indicate the start and end dates of your most recent job: For ongoing job please indicate the current month and year as end date.

- Start date :_____/______ (mm/yyyy)
- End date :_____/______ (mm/yyyy)

3.1 - If you cannot remember the exact dates, estimate the duration of this job: _____ months, ________ years

4 – Company's name: ______________________________ eg. DEF Automotive Inc.

5 - What does (or did) your company do at this site?

____

eg. Full service vehicule maintenance and car repairs

6 - Do you agree to provide the FULL address of this company? □ Yes □ No, **If Yes**, go to question 7]

6.1 - If No, do you agree to provide ONLY the first 3 characters of the postal code? □ Yes □ No

[**If Yes**, complete only the **postal code** field in the ADDRESS box at the bottom of the page] **[If No**, go to question 8]

7 - What is (or was) the address of the company? Please complete the information in the ADDRESS box at the bottom of the page. If you cannot recall the exact street address, tell us the name of the nearest cross-street or the nearest town if you worked in a rural location. Specify the region if it is (or was) not a fixed workplace.

For jobs with changing work schedules, AVERAGE your work load over the whole year.

8 - On average, how many HOURS PER WEEK do (or did) you work? ______ hours □ Can't recall

9 - On average, how many WEEKS PER YEAR do (or did) you work? ______ weeks □ Can't recall

10 - On average, how many DAYS PER MONTH do (or did) you work 3 or more hours between midnight and 5am? ______ days □ Can't recal

11 - For this job, which of the following BEST describes your work pattern?

□ Regular daytime schedule or shift
□ Regular evening shift (shift ends before midnight)
□ Regular night shift (between midnight and 5am)
□ Other, specify______________________________
□ Rotating shift, changing periodically from days to evenings or to nights
□ Split shift, consisting of two or more distinct periods each day
□ Irregular schedule, or on call
□ Don't know

12 - What percentage of time do (or did you) spend working outdoors? _____% [If 0%, go to question 14**]**

13 - Do (or did you) your work require you to work outdoors in the summer months? □ Yes □ No

13.1 - If Yes, on average, how much time each day are (or were) you in the sun between 11am and 4pm?

□ Less than 1 hour □ 1-2 hours □ 2-4 hours □ More than 4 □ Can't recall

14 - On average, how many MINUTES PER DAY do (or did) you spend commuting TO AND FROM work via the following means of transportation?

14.1 - During the summer months (June-Aug)	□ Car _______ min/day □ Bus _______ min/day	□ Train _______ min/day □ Subway_______ min/day	□ Walk_______ min/day □ Other, specify_______________, ____min/day
14.2 - During the coole months (Sept-May)	□ Car _______ min/day □ Bus _______ min/day	□ Train _______ min/day □ Subway_______ min/day	□ Walk_______ min/day □ Other, specify_______________, ____min/day

ADDRESS

Number Street Direction Cross street (if address unknown)

City Province Rural address: nearest town or village

Postal code Region

ENVIRONMENT QUESTIONNAIRE 2011 21

Appendix IV. ***A priori*** **estimation of the minimal detectable effects for the smoking-methylation and occupational-methylation analysis**

	Prevalence of exposure in the total study population (%)	Minimum detectable effect size (with a power of 80% and an alpha of 0.05)
Smoking[a]		
Former smokers	237(44.7%)	±0.24
Current smokers	106 (20%)	±0.29
Occupational exposure[b]		
Formaldehyde	13 (5.1%)	±0.79
Lead	26 (10.1%)	±0.57
PAHs from petroleum	28 (10.9%)	±0.55
PAHs from any source	44 (17.1%)	±0.45

[a] The total population for the smoking-methylation analysis consisted of 538 participants; the unexposed category was thus comprised of 195 participants.

[b] The total population for the occupational exposures-methylation analysis consisted of 257 participants; the unexposed category differed by agent and can be calculated by the difference of the total population and the estimated prevalence of exposure of each agent.