Title
Coffee Consumption, Smoking, and Genetic Factors of Parkinson's Disease: Gene-Environment Interaction and Genome-wide DNA Methylation Studies

Author
CHUANG, YU-HSUAN

Publication Date
2017

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Coffee Consumption, Smoking, and Genetic Factors of Parkinson's Disease: Gene-Environment Interaction and Genome-wide DNA Methylation Studies

A dissertation submitted in partial satisfaction of the requirements for the degree Doctor of Philosophy in Epidemiology

by

Yu-Hsuan Chuang

2017

ABSTRACT OF THE DISSERTATION

Coffee Consumption, Smoking, and Genetic Factors of Parkinson's Disease: Gene-Environment Interaction and Genome-wide DNA Methylation Studies

by

Yu-Hsuan Chuang

Doctor of Philosophy in Epidemiology

University of California, Los Angeles, 2017

Professor Beate R. Ritz, Chair

Parkinson's disease (PD) is a common neurodegenerative disorders with a complex and unclear etiology in which both environmental and genetic factors contribute to disease. Over the past decade, gene-environment interaction and DNA methylation association studies have been used to explore the biological mechanisms underlying PD in order to help develop new strategies for prevention and treatment.

Drinking caffeinated coffee has been reported to protect against PD. Caffeine is an adenosine A2A receptor (*ADORA2A*) antagonist that increases dopaminergic neurotransmission and Cytochrome P450 1A2 (*CYP1A2*) metabolizes caffeine. In a population-based case control study in Denmark (PASIDA), we estimated statistically significant interactions for *ADORA2A* rs5760423 and heavy vs. light coffee consumption in incident (OR interaction=0.66 [0.46-0.94], p=0.02) but not prevalent PD. In meta-analyses combining data from a large consortium study

(PEGASUS), PD associations with daily coffee consumption were strongest among carriers of variant alleles in both *ADORA2A* and *CYP1A2*.

There has also been growing interest in investigating whether inflammation-related genes interact with environmental factors such as smoking to influence PD risk. We replicate the previously reported inverse association human leukocyte antigen (HLA)-*DRB1* rs660895 and the risk of PD in the Danish PASIDA study. Moreover, both in the Danish study and in the pooled analysis combining two French studies (Terre and Partage), sub-multiplicative interactions were observed between rs660895 and smoking (OR interaction=1.54, p=0.001), such that the inverse association of rs660895-G with PD decreased among smokers.

Lastly, we investigated whether epigenetics play a role in health benefits of drinking coffee. We associate epigenome-wide DNA methylation levels to habitual coffee consumption from two studies with blood (Parkinson's Environment and Gene (PEG wave 1) and Women's Health Initiative), and one with saliva samples (PEG wave 2) using bi-weighted mid-correlation and meta-analysis. Adjusting for age, gender, and blood cell composition, one CpG (cg21566642 near *ALPPL2)* surpassed genome-wide significance ($p = 3.7x10^{-10}$) and from among ten additional CpGs significant at $p \leq 5.0x10^{-6}$. These CpGs are found in or near genes related to lipid metabolism and immune response. Interestingly, when we stratified by menopausal hormone therapy, methylation differences with coffee consumption were observed only in women who never used MHT. We did not replicate any of the associations found in blood in our saliva samples.

Overall, my dissertation provides epidemiologic and biological evidence supporting the hypotheses that caffeine plays a biological role in reducing PD risk and may have beneficial

effects in chronic disease or other neurodegenerative disorders. More importantly, neuroinflammation is an important contributor to the pathogenesis of PD.

The dissertation of Yu-Hsuan Chuang is approved.

Stefan Horvath

Marjan Javanbakht

Teresa E. Seeman

Beate R. Ritz, Committee Chair

University of California, Los Angeles

2017

TABLE OF CONTENTS

LIST OF TABLES AND FIGURES

ACKNOWLEDGEMENTS

I would like to thank a number of people who have assisted me with writing this dissertation and throughout my studies. I first thank my advisor, Dr. Beate Ritz, for the guiding, supporting, and training me to be an independent researcher. I would also like to thank my co-advisor, Dr. Stefan Horvath, for the excellent research opportunities and guidance throughout this process. I would like to thank Dr. Alexis Elbaz for his willingness to help and provide me best suggestions. I would like to thank Drs. Javanbakht and Seeman for serving on my dissertation committee and the helpful comments and feedback. I would also like to thank the study participants, staff, and research technicians, and collaborators who contributed to these researches. I am grateful for the financial support from UCLA Department of Epidemiology and the Burroughs Wellcome Fund training program. And finally, I would like to express my thanks to friends and family who provided emotional and spiritual support throughout the PhD program.

Chapter 2 was originally published in Neuroepidemiology published by Karger at: **Chuang YH**, Lill CM, Lee PC, Hansen J, Lassen CF, Bertram L, Greene N, Sinsheimer JS, Ritz B. Gene-Environment Interaction in Parkinson's Disease: Coffee, *ADORA2A*, and *CYP1A2*. *Neuroepidemiology* (2016).

Chapter 4 was originally published in European Journal of Human Genetics published by Nature at: **Chuang YH**, Quanch A, Absher D, Assimes T, Horvath S, Ritz B. Coffee consumption is associated with DNA methylation levels of human blood. *European Journal of Human Genetics* (2017).

VITA

Education:

2009 B.S., Medical Laboratory Science and Biotechnology, Taipei Medical University, Taiwan

2012 MPH, Epidemiology, Emory University, Atlanta, USA

Publications:

Chuang YH, Quach A, Absher D, Assimes T, Horvath S, Ritz B. Coffee consumption is associated with DNA methylation levels of human blood. Eur J Hum Genet. 2017 May;25(5):608-616.

Chuang YH, Lill CM, Lee PC, Hansen J, Lassen CF, Bertram L, Greene N, Sinsheimer JS, Ritz B. Gene-Environment Interaction in Parkinson's Disease: Coffee, *ADORA2A*, and *CYP1A2*. Neuroepidemiology. 2016;47(3-4):192-200.

Sun YV, Lazarus A, Smith JA, Chuang YH, Zhao W, Turner ST, Kardia SL. Gene-specific DNA methylation association with serum levels of C-reactive protein in African Americans.. PLoS One. 2013 Aug 19;8(8):e73480.

1. Background and Introduction

1.1 Dissertation Objectives

The aim of this dissertation is to investigate whether lifestyle factors i.e. coffee consumption and smoking, play biological roles in the etiology of PD. If the previously reported associations cannot be explained by bias, we would like to know how they interact with genetic factors and what possible mechanisms are in order to cause PD. We also aim to investigate the influence of coffee consumption on DNA methylation levels, which is believed to act as a potential mediator for caffeine's influence on health, including PD. Briefly, the first project examines the interaction of coffee consumption with the *ADORA2A* and *CYP1A2* polymorphisms in modifying PD risk, and whether this interaction is different between prevalent vs. incident PD cases. The second project examines the association of *HLA-DRB1* polymorphisms with PD, and whether smoking and the *HLA-DRB1* polymorphisms act together to influence the risk of PD. The third project investigates the association of coffee consumption and DNA methylation sites from a genome-wide screen in human blood, as well as in saliva.

1.2 Parkinson's disease background

Parkinson's disease is the second most common neurodegenerative disorder with a prevalence of 0.5-1% among people more than 65 years of age (Tanner et al. 1996). Each year 60,000 new cases are identified in the US, and by 2017, over 10 million people worldwide live with PD (Parkinson's Disease Foundation 2017). PD's four cardinal features include resting tremor, bradykinesia, rigidity, and postural instability, and they are attributed to the effects of a loss of dopaminergic neurons in substantia nigra pars compacta (Hoehn et al. 1967). However, PD is hard to accurately diagnose at the early stages for the following reasons: 1) the characteristic motor symptoms do not develop until 50-60% of the dopaminergic neurons are lost

and 80-85% of the dopamine is depleted in striatum; 2) other disorders such as postencephalitic, drug-induced, and arteriosclerotic parkinsonism share the symptoms and signs of PD. Currently, the most commonly used clinical criteria for PD diagnosis are the United Kingdom Brain Bank and Gelb criteria (Hughes et al. 1992; Gelb et al. 1999). People with the presence of at least two of the four cardinal features are considered as PD patients.

Epidemiologic studies have shown that age over 60, male gender, Hispanic/non-Hispanic white, and family history of PD are associated with a higher risk of PD (Van Den Eeden et al. 2003; Wirdefeldt et al. 2011). Although approximately 5-10% of PD is familial and appears to follow autosomal dominant inheritance, familial aggregation is stronger for early-onset cases than late-onset cases. Over 90% of PD patients are considered "idiopathic" without known genetic causes; it is likely that a combination of genetic and environmental factors together contribute to the disease. Lifestyle factors such as coffee consumption and smoking have been consistently associated with a decreased the risk of PD, with some study showing in women the protective effect of caffeine may be attenuated by estrogen use (Ascherio et al. 2004; Palacios et al. 2012). But some argue that the significant inverse association between coffee or smoking and PD can be explained by different systemic biases, such as reverse causation, competing risk, and confounding. For instance, PD patients may have changed coffee drinking behavior (e.g. tend to stop drinking coffee due to prodromal symptoms -- loss of smell or insomnia) or changed smoking habits (e.g. less prone to smoke or more likely to quit due to less sensitive nicotinic response) in the preclinical period of PD, thus a possibility of reverse causation cannot be ruled out. Furthermore, compared to non-smokers, smokers have a higher mortality of lung cancers, heart disease or other smoking-related diseases. We may observe less prevalent PD cases among smokers than non-smokers if smokers die earlier.

During the past decade, genome-wide association studies (GWAS) have been widely used to identify genes for idiopathic PD in a hypothesis-free manner. A recent meta-analysis of GWAS replicated associations with 24 previous identified single nucleotide polymorphisms (SNPs) and proposed 6 additional SNPs as risk variants for PD (Nalls et al. 2014). As expected, most SNPs contribute only weakly to population risk (OR<1.3) of PD, in spite of the fact that six SNPs further showed associations with gene expression or DNA methylation. Exploring gene-environment interaction and DNA methylation association may help understand the biologic mechanisms of PD and suggest helpful biomarker for PD prevention and treatment.

2. Gene-Environment Interaction in Parkinson's Disease: Coffee, *ADORA2A*, and *CYP1A2*

2.1 Introduction

More than 90% of Parkinson's disease (PD) is considered to be "idiopathic" – with genetic and environmental factors increasing risk of disease. A protective effect of coffee on PD has been postulated since many epidemiologic studies reported lower consumption of caffeinated coffee among PD patients (Nefzger et al. 1968; Hellenbrand et al. 1996). Ever vs. never drinkers have a 30% lower risk of PD and three additional cups of coffee per day lowered PD risk on average by 25-32% (Hernan et al. 2002). Caffeinated coffee is a very popular beverage in Northern European countries, especially Denmark, and we recently reported a 55% lower risk of PD among moderate coffee drinkers in Denmark (Kenborg et al. 2015). A landmark early prospective cohort study not only reported an inverse association for coffee, but also for caffeine from non-coffee sources, suggesting it might be the protective agent(Ross et al. 2000). This association was replicated in many prospective cohort and case-control studies and further strengthened by observations of exposure-response trends (Hernan et al. 2002; Saaksjarvi et al. 2008; Qi et al. 2014). Animal studies lent additional support to the idea that caffeine and its metabolites are neuro-protective (Xu et al. 2010). Yet, the biological mechanisms underlying neuroprotection derived from caffeine have yet to be established. Importantly, epidemiologic data – even from prospective studies - do not preclude reverse causality since those who later develop PD may stop drinking caffeinated coffee due to sleep disorders, anxiety, gastro-intestinal problems or simply a loss of smell that could make coffee drinking less enjoyable in the very long pre-motor stages of PD. Evidence for interactions between caffeinated coffee and genes that

metabolize caffeine or encode brain receptors targeted by caffeine, could help strengthen arguments that caffeine indeed plays a biological role in reducing PD risk.

Recently, a large consortium (PEGASUS) combined data from 1,325 PD cases and 1,735 controls and reported that PD risk was influenced by interactions between the single-nucleotide polymorphisms (SNP) rs5751876 and rs3032740 in *ADORA2A*, which encodes the adenosine A_{2A} receptor in dopamine neurons, and caffeinated coffee consumption (Popat et al. 2011); however, two much smaller studies did not find evidence for such interaction (Tan et al. 2006; Facheris et al. 2008). In addition, the PEGASUS study also observed stronger coffee-PD associations among carriers of the CC genotype of rs762551 in *CYP1A2* compared with CA or AA carriers (Popat et al. 2011), a gene that encodes the cytochrome P450, family 1, subfamily A, polypeptide 2, the main caffeine metabolizing enzyme.

Relying on data from a large population-based case control study of PD ("Parkinson's Disease in Denmark" [PASIDA]), here we re-examine interactions of coffee consumption with *ADORA2A* and *CYP1A2* polymorphisms and also assess whether reliance on prevalent versus incident PD cases influences results, a distinction that may have caused previous study results to disagree (Hill-Burns et al. 2011).

2.2 Materials and Methods

The PASIDA study was approved by the Institutional Review Boards of UCLA, the Danish Data Protection Agency, and the ethics committee of Copenhagen. Informed consent was obtained from all study participants

Study population

The PASIDA study enrolled idiopathic PD patients (ICD-8 342 and ICD-10 G20) treated at 10 neurological treatment centers and identified from the Danish National Hospital Register between 1996 and mid-2009 with subsequent validation of their diagnoses by medical record review. Population controls, free of PD when matched cases were diagnosed, were selected from the Danish Central Population Registry (individually matched on year of birth and sex). Detailed recruitment information was published previously (Kenborg et al. 2015). Of 3,700 recruited subjects, 1,575 (87%) PD cases and 1,607 (85%) controls provided DNA samples (saliva) for genotyping. We further excluded subjects who were diagnosed with dementia prior to interview, leaving 1,556 PD cases and 1,606 controls for analyses.

Exposure assessment and variable definition

Standardized telephone interviews were conducted between 2008 and 2010 to obtain participants' lifetime caffeinated coffee consumption history (drip- and instant-coffee) and information on other lifestyle factors. Due to the high prevalence (> 90%) of caffeinated coffee drinking in Denmark but little tea and caffeinated soda consumption during the study period, we omitted the latter caffeine sources. We collected lifetime amount and duration of caffeinated coffee-drinking, asking participants to report start and stop ages and the average number of cups the consumed per day. We consider an 'ever' coffee drinker someone who consumed at least one cup (6 oz) of coffee per week for a year. To obtain the amount of caffeine intake, we converted coffee cups per day into daily caffeine consumption (mg) using the U.S. Department of Agriculture criteria (Gebhardt et al. 2000). Only consumption before the index date contributed to our exposure measures, i.e. the date of first motor symptom recorded on the medical record, or the date of PD diagnosis for both cases and their matched controls.

Genotyping

DNA was extracted from saliva using standard protocols. Samples were genotyped on the QuantStudio™ 12K Flex Real-Time PCR System using multiplex Taqman allelic discrimination assays (Applied Biosystems) according to the manufacturer's protocol. Each 384-well plate included ~5% HapMap CEU samples genotyped in duplicates across plates to assess genotyping accuracy. To control for genotyping quality, we excluded samples with genotyping efficiency less than 80% and SNPs with low genotyping efficiency (<95%) and accuracy (<99.5%); all three SNPs (rs5760423, rs762551, rs2472304) in this study met these criteria.

Statistical Analysis

We tested for deviation from Hardy-Weinberg Equilibrium among controls using Pearson's chi-square test (all $p \geq 0.05$). We broke the matched pairs and conducted unconditional logistic regression analyses adjusting for sex, birth year, and onset/index age to estimate marginal associations between caffeinated coffee consumption and PD status as well as between the three *ADORA2A* or *CYP1A2* polymorphisms and PD status (additive genetic model), respectively. We broke the matched sets to avoid loss of entire pairs with only one subjects when conducting stratified analyses and to increase efficiency since many pairs shared the same matching variable values (Kenborg et al. 2015). However, we compared the overall results from conditional with the results from unconditional logistic regression adjusted for all of the matching variables and found them to be identical. Matching variables (i.e., year of birth, gender and onset/index age), potential confounders (i.e., any kind of tobacco smoking) and strong predictors of PD (i.e., family history) were included in all models. We treated coffee intake as a binary variable with light vs. heavy consumption (defined as 0 to <= median vs. > median cup-years (Hill-Burns et al. 2011)) and also as a continuous variable (number of cups per day). We

further created categories of caffeine intake in mg per day and years of coffee consumed using category definitions from our previous paper (Ahmed et al. 2014). The Wald test for trend was applied to categorized coffee variables testing for a linear relationship with PD. Information about ethnic diversity was not available but based on demographics of the Danish population provided by Denmark we are confident that the large majority were non-Hispanic Whites(2015).

We used multiplicative terms in logistic regression adjusted for confounders to assess whether the *ADORA2A* or *CYP1A2* polymorphisms modify caffeine-PD associations and the likelihood ratio chi-square tests was used to evaluate statistical significance . We also restricted all analyses to incident PD patients and their matched controls, i.e. those diagnosed close to their date of interview during 2006-2009, to assess whether survival or recall bias may have influenced results with prevalent patients. Analyses were conducted using SAS version 9.4 (SAS Institute, Inc., Cary, North Carolina).

Lastly, we conducted meta-analyses to aggregate results from PASIDA (incident cases only) and non-Hispanic Whites from PEGASUS (Popat et al. 2011), based on type of PD case (i.e. incident), control selection (i.e. population-based controls) and ethnicity (i.e. non-Hispanic Whites) using the metagen package in the R environment , which allows to fit fixed-effects and random- effects models (Friedrich 2013). In the meta-analysis, results of *ADORA2A* rs5760423 in PASIDA were equated with rs5751876 in PEGASUS because they are in high linkage disequilibrium (LD) (Byrne et al. 2012); also we combined our coffee category of 'heavy use' with 'ever' consumption in PEGASUS as well as and 'light' consumption in PASIDA with 'never' in PEGASUS since less than 10% of PASIDA participants reported having never consumed coffee.

2.3 Results

Our initial analysis included 3,162 Danish participants in the PASIDA study with high-quality genotyping data. The average age of PD onset or index age was 61 years for all participants (**Table 2-1**) and 64 years for incident PD patients and their-matched controls only. Sixty percent of participants were male and, compared with population controls, PD cases were more likely to have a positive family history of PD and smoke less. Ninety-four percent of PD cases and 97% of controls were ever coffee consumers.

Heavy coffee drinking in PASIDA is associated with a 25% lower risk of PD (OR=0.75[0.64-0.88]), and each additional cup of coffee consumed per day on average is associated with a 4% lower PD risk (OR=0.96[0.93-0.99]) (**Table 2-S1**); inverse coffee-PD associations are estimated for both prevalent and incident PD. Of note, the per-cup measure of daily coffee consumption was not associated with PD among incident cases. Odds ratio estimates adjusted solely for birth year, sex and onset/index age did not substantially differ from estimates further adjusted for ever smoking and PD family history. Finally, marginal associations of *ADORA2A* rs5760423 as well as of *CYP1A2* rs762551 and rs2472304 (in LD with rs762551: r^2=0.87, D'=0.99) with PD status (incident and prevalent) were null (**Table 2-S2**).

Interaction analyses based on all subjects, did not show statistically significantly varying effects of caffeine across genotypes of *ADORA2A* or *CYP1A2* polymorphisms, respectively (**Table 2-2**). However, there appeared to be a trend in coffee-PD effect estimates across *ADORA2A* rs5760423 genotypes: The OR for PD among heavy coffee drinkers, relative to light coffee drinkers, was 0.81 (0.62-1.05) for GG carriers compared with 0.68 (0.55-0.86) for GT and 0.54 (0.37-0.78) for TT carriers (OR interaction=0.85 [0.68-1.06], p for interaction =0.14). Further adjustment for smoking and PD family history did not change results (data not shown).

When we restricted our analyses to incident PD only, we observed a statistically significant interaction for the *ADORA2A* rs5760423 and heavy coffee drinking (OR interaction=0.66 (0.46-0.94), p for interaction=0.02): the OR for drinking coffee in GG carriers was 1.10 (0.72-1.68), 0.63 (0.44-0.92) for GT, and 0.58 (0.30-1.09) for TT carriers. When duration of caffeine intake was removed from the caffeine measure, the interaction of *ADORA2A* polymorphism and coffee was not statistically significant (p for interaction=0.28 in cup/day for all cases, and p=0.55 for incident cases, respectively). There was no evidence for association measure modification for *CYP1A2* rs762551 and rs2472304 (p for interaction=0.45 and 0.93 respectively). Similarly, restricting to incident cases only did not reveal statistically significant interactions for CYP1A2 SNPs. No interactions were found in prevalent case analyses. Results of interaction analyses using other coffee measures, daily intake of caffeine and years of coffee drinking are presented in **Table 2-S3**. We did not observe evidence for effect-measure modification with these measures and the SNPs we investigated.

Our meta-analytic results for coffee-PD associations did not differ much when we used random-effects versus fixed-effects models (**Table 2-3** and **Table 2-S4**). Based on random effects models, the *ADORA2A* gene polymorphisms and daily coffee consumption association for PD was strongest among rs5760423 TT carriers, OR=0.77 (0.51-1.16), compared with GT and GG carriers (OR=0.96 [0.90-1.02] and 0.97 [0.91-1.03], respectively). We saw similar patterns for *CYP1A2* polymorphisms in both rs762551 and rs24702304, i.e. the coffee-PD associations were strongest among homozygotes for the variant alleles (OR=0.86 [0.69-1.08] for rs762551 CC carriers and 0.84 [0.71-0.99] for rs24702304 GG carriers, **Figure 2-1**).

2.4 Discussion

We conducted analyses of gene-environment interactions in a Danish case-control study of PD. Our study follows up on the results from a previous consortium study (PEGASUS) with an equally large sample size. When we include both prevalent and incident PD cases in our analysis, we found no evidence for interactions with *ADORA2A/ CYP1A2* polymorphisms. However when we restricted analyses to incident cases only, we observed interactions between the *ADORA2A* polymorphism and coffee drinking. This difference in results suggests that survival or recall bias may affect studies that rely on or include prevalent PD cases. Moreover, when we combined our results for incident cases with the PEGASUS incident cases of European ancestry in a meta-analytical approach, both *ADORA2A* and *CYP1A2* polymorphisms modified coffee-PD associations, although the *CYP1A2* interaction was solely due to the influence of the PEGASUS study.

Our PASIDA findings are mostly consistent with those published by the PEGASUS consortium which previously reported ORs for PD risk of each additional cup of coffee consumed per day among coffee drinkers as 0.93 (0.84-1.03) and 0.92 (0.81-1.04) for CC or CT carriers of *ADORA2A* rs5751876, respectively, and 0.61 (0.46-0.81) for TT carriers (interaction p-value 0.01) in non-Hispanic Whites (Popat et al. 2011). Yet, two smaller studies did not find statistically significant interactions for *ADORA2A* polymorphisms and coffee in PD (Tan et al. 2006; Facheris et al. 2008). One study was conducted in a mixed-race population that did not find the expected inverse main effect for coffee consumption on PD, possibly because sibling controls were used (Facheris et al. 2008). Sibling controls are likely too similar to cases in terms of coffee consumption, making it hard to estimate effects of coffee consumption on PD risk. The second null result was reported for an Asian population with a low average coffee consumption

(2.9 in cases vs. 4.7 in controls (Tan et al. 2006) cup-years compared with 161.3 vs. 186.5 cup-years in PASIDA), such that the exposure levels and contrasts were likely insufficient.

Animal studies have shown that administration of caffeine or other adenosine A_{2A} receptor antagonists before dosing the animal with the neurotoxin 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine (MPTP) reduces loss of dopamine and dopaminergic neurons, suggesting that caffeine reduces PD risk by deactivating the A_{2A} receptors (Chen et al. 2001; Pierri et al. 2005). Also, the *ADORA2A* rs3032740 variant (in LD with rs5751876) has been shown to reduce protein expression (Alsene et al. 2003), and thus may result in reduced A_{2A} receptor function that together with further inhibition through coffee consumption may exert protective effects (Popat et al. 2011). We would thus expect the inverse coffee-PD association to be strongest in those with a TT genotype in *ADORA2A* rs5760423 (in LD with rs5751876). Adenosine A_{2A} receptors have also become the latest target for non-dopaminergic therapies in PD based on their interaction with dopamine D_2 receptors in striatopallidal neurons (Franco et al. 2000; Ferre et al. 2008).

Cytochrome P450 1A2 is the main caffeine-metabolizing enzyme that converts over 90% of caffeine in the liver to paraxanthine, and its activity depends on age, gender, smoking, and *CYP1A2* polymorphisms (Sachse et al. 1999; Backman et al. 2008; Gunes et al. 2008). Thus, we would expect neuroprotection due to caffeine to be stronger in slow metabolizers who carry the variant alleles we investigated. The PEGASUS consortium pooling incident case control studies (n=3,060), found evidence for coffee-*CYP1A2* interactions with inverse PD associations for coffee drinking being strongest in CC genotype carriers at rs762551 and GG genotype carriers at rs2472304 (Popat et al. 2011). The NeuroGenetic Research Consortium (n=2,389) did not find

interactions with *CYP1A2* but included prevalent cases and some studies used spousal controls (Hill-Burns et al. 2011).

Previously, concerns were raised that confounding by population structure in PEGASUS produced spurious results (Hill-Burns et al. 2011; Mellick et al. 2011) since allele frequencies for *CYP1A2* SNPs vary strongly across ethnicities and in PEGASUS the coffee- *CYP1A2* interactions did not reach statistical significance in non-Hispanic Whites alone (Popat et al. 2011). However, combining PEGASUS and PASIDA non-Hispanic Whites our meta-analysis produced a decreasing trend across *CYP1A2* rs762551 genotypes (lavender line, **Figure 1b**) and suggested an interaction with cups per day of consumption. Interestingly, ever (vs never) coffee consumption produced an inverted-U shape for the *CYP1A2* rs762551 polymorphisms, but ever (vs never) coffee consumption is a poor measure of average caffeine intake (**Table 3**).

Smoking is positively associated with coffee drinking and negatively with PD risk, which we have previously interpreted as a consequence of pre-motor prodromal PD (Hernan et al. 2002). Moreover, a study reported that the metabolic activity did not differ between AC or CC and AA carriers at rs762551in non-smokers suggesting that *CYP1A2* genotypes may influence enzyme activity only in smokers (Sachse et al. 1999). In our study, adjustment for smoking did not change the interaction estimates for either of the *CYP1A2* SNPs. Complicating the matter further, both caffeine and its CYP1A2 metabolite paraxanthine may non-selectively bind to adenosine receptor and act as a neuroprotector diminishing somewhat the potential importance of CYP1A2 enzyme activity (Xu et al. 2010). The average Danish study participants drank as much as 4 cups per day over 40 years implying that levels of caffeine and its metabolites might be chronically higher than in other populations consuming less coffee possibly rendering the contributions of the metabolizing enzyme less important.

Our study has several strengths. We have a large sample size with a homogenous ancestry; we selected population controls from Danish registers, assessed confounding (including smoking) extensively, and were able to distinguish between incident and prevalent cases. High coffee consumption in Denmark allowed us to assess dose-response relationships for coffee and PD with great statistical power and since we collected detailed information on lifetime coffee consumption we were able to define exposures in various ways.

Limitations are that very few (<10%) participants reported not drinking coffee such that CYP1A2 enzyme activity may not affect caffeine levels in the blood more than minimally making it hard to assess the influence of *CYP1A2* polymorphisms. PD prevalent cases tend to have more memory loss, therefore lifetime coffee consumption could be misreported or reflect changes in drinking habits after diagnosis such as due to sleep problems common in PD patients. Also, recall might also be impaired in all cases and controls as the population was on average 68 years of age at the time of interview, which might be causing non-differential exposure misclassification.

In conclusion, our study corroborates previous findings that interactions between *ADORA2A* rs5760423, *CYP1A2* rs762551 and rs2472304 variants and coffee consumption affect PD risk. However, since our study only found interaction between *ADORA2A* rs5760423 and coffee for a measure of 'total cup-years of coffee consumed' but not 'average number of coffee cups per day', a measure used in the previous study, we cannot exclude the possibility that reverse causation contributed to these results. The lack of a cup-per-day association may, however, also be explained by the generally very high coffee consumption levels among Danes, i.e. that few Danes consumed so little coffee that each additional cup would make a difference

(Popat et al. 2011) Therefore, additional data and studies are still needed in support of the hypothesis that a biological effect of caffeine protects against PD.

2.5 Tables and Figures

Table 2-1. Characteristics of study participants (n=3,162)

	All		Incident cases only		Prevalent cases only	
	Cases (n=1,556)	Controls (n=1,606)	Cases (n=554)	Controls (n=566)	Cases (n=1,002)	Controls (n=1,040)
Mean onset/index age (SD)[a]	61.1 (9.5)	61.3 (9.7)	64.3 (8.7)	64.3 (8.8)	59.7 (9.7)	60.1 (9.9)
Male (%)	932 (59.9)	975 (60.7)	335 (60.5)	357 (63.1)	597 (59.6)	618 (59.4)
Ever smoking (%)	775 (49.9)	1026 (64.2)	271 (49.2)	349 (62.2)	504 (50.4)	677 (65.3)
Packyears of cigarette smoking (SD)	7.0 (13.3)	11.3 (16.3)	8.0 (14.7)	11.0 (16.8)	6.5 (12.3)	11.5 (16.0)
Family history of PD (%)	211 (13.6)	89 (5.5)	74 (13.36)	28 (5.0)	137 (13.7)	61 (5.9)
Caffeinated coffee consumption						
Cup per day[b] (%)						
Never	85 (6.4)	43 (3.0)	25 (5.2)	18 (3.5)	60 (7.1)	25 (2.7)
1	111 (8.4)	79 (5.5)	36 (7.5)	23 (4.5)	75 (8.9)	56 (6.1)
2	244 (18.4)	243 (17.0)	77 (16.0)	71 (13.8)	167 (19.8)	172 (18.8)
>=3	884 (66.8)	1064 (74.5)	342 (71.3)	403 (78.3)	542 (64.2)	661 (72.3)
mean±SD	3.9 ± 3.1	4.4 ± 2.8	4.2 ± 2.9	4.5 ± 2.6	3.8 ± 3.3	4.4 ± 2.9
Cupyears[b] (%)						
Light (0,<=median)	806 (60.9)	746 (52.2)	253 (52.7)	236 (45.8)	553 (65.5)	510 (55.8)
Heavy (>median)	518 (39.1)	683 (47.8)	227 (47.3)	279 (54.2)	291 (34.5)	404 (44.2)
Caffeine mg/day (quartile)[b] (%)						
0%-25%	429 (32.4)	357 (24.9)	136 (28.3)	108 (21.0)	293 (34.7)	249 (27.2)
>25%-50%	430 (32.5)	445 (31.1)	154 (32.1)	171 (33.2)	276 (32.7)	274 (30.0)
>50-75%	204 (15.4)	269 (18.8)	73 (15.2)	95 (18.4)	131 (15.5)	174 (19.1)
>75%-100%	261 (19.7)	358 (25.1)	117 (24.4)	141 (27.4)	144 (17.1)	217 (23.7)
Years of coffee drinking[b] (%)						
0-37	519 (38.9)	467 (32.5)	155 (32.2)	123 (23.8)	364 (42.8)	344 (37.4)
>37-45	311 (23.4)	352 (24.5)	111 (23.1)	131 (25.4)	200 (23.5)	221 (24.0)
>45-53	301 (22.6)	354 (24.6)	124 (25.8)	142 (27.5)	177 (20.8)	212 (23.0)
>53	201 (15.1)	264 (18.4)	91 (18.9)	120 (23.3)	110 (12.9)	144 (15.6)

Polymorphisms						
ADORA2A rs5760423[c] (%)						
GG	534 (34.5)	533 (33.3)	201 (36.4)	193 (34.2)	333 (33.4)	340 (32.9)
GT	735 (47.5)	781 (48.9)	255 (46.2)	270 (47.8)	480 (48.2)	511 (49.5)
TT	279 (18.0)	284 (17.8)	96 (17.4)	102 (18.0)	183 (18.4)	182 (17.6)
CYP1A2 rs762551[c] (%)						
AA	874 (56.3)	873 (54.7)	312 (56.4)	322 (57.4)	562 (56.2)	551 (53.3)
CA	563 (36.3)	607 (38.1)	205 (37.1)	202 (36.0)	358 (35.8)	405 (39.2)
CC	116 (7.4)	115 (7.2)	36 (6.5)	37 (6.6)	80 (8.0)	78 (7.5)
CYP1A2 rs2472304[c] (%)						
AA	750 (48.4)	750 (46.8)	260 (47.1)	283 (50.3)	490 (49.1)	466 (44.9)
GA	630 (40.7)	681 (42.5)	234 (42.4)	229 (40.7)	396 (39.7)	452 (43.5)
GG	169 (10.9)	171 (10.7)	58 (10.5)	51 (9.0)	111 (11.2)	120 (11.6)

Age range at interview: cases 39-86 years old, controls 39-88 years old

Age at PD onset: 28-83 years old; age at diagnosis: 33-83 years old

[a]Age at the date of first motor symptoms recorded on medical record

[b]Missing information: smoking (n=13), packyear (n=416), coffee cup/day (n=409), cupyears (n=409), caffeine mg/day (n=409), years of coffee drinking (n=393); median value was determined based on controls (excluding non-drinkers (Hill-Burns et al. 2011)).

[c]Genotyping failures: rs5760423 (n=16), rs762551 (n=14), rs2472304 (n=12)

Table 2-2. Adjusted odds ratios (OR) and 95% confidence intervals (CI) for caffeinated coffee consumption and Parkinson's disease in PASIDA, by *ADORA2A* and *CYP1A2* genotypes

Caffeinated coffee	Homozygous Major				Heterozygous				Homozygous Minor				OR interaction	*p*-value
	Cases	Controls	cOR	95% CI	Cases	Controls	cOR	95% CI	Cases	Controls	cOR	95% CI		
Incident and prevalent cases: 1324 cases, 1429 controls [a]														
ADORA2A rs5760423														
Cups/day, mean ± SD [b]	4.1±2.7	4.5±2.8	0.94	0.90-0.99	4.2±2.9	4.6±2.8	0.95	0.91-0.99	4.4±4.1	4.6±2.6	0.98	0.92-1.04	1.02(0.98-1.06)	0.28
Light (0,<=median cupyears)	276	262	1	--	383	362	1	--	143	118	1	--	--	
Heavy (>median cupyears)	179	213	0.81	0.62-1.05	244	334	0.68	0.55-0.86	91	132	0.54	0.37-0.78	0.85(0.68-1.06)	0.14
CYP1A2 rs762551														
Cups/day, mean ± SD [b]	4.3±2.9	4.6± 2.8	0.96	0.93-1.00	4.2±3.5	4.6± 2.7	0.96	0.92-1.00	3.7±2.3	4.1± 2.7	0.91	0.81-1.03	0.99(0.94-1.03)	0.62
Light (0,<=median cupyears)	444	404	1	--	299	283	1	--	61	54	1	--	--	
Heavy (>median cupyears)	306	375	0.74	0.60-0.91	175	253	0.65	0.50-0.84	36	49	0.58	0.32-1.05	0.91(0.71-1.16)	0.45
CYP1A2 rs2472304														
Cups/day, mean ± SD [b]	4.3±2.9	4.6± 2.8	0.96	0.92-1.00	4.2±3.5	4.5± 2.7	0.96	0.93-1.00	3.9±2.4	4.3± 2.7	0.93	0.84-1.02	1.00(0.95-1.04)	0.85
Light (0,<=median cupyears)	382	339	1	--	333	317	1	--	85	87	1	--	--	
Heavy (>median cupyears)	266	328	0.71	0.57-0.89	197	287	0.66	0.52-0.84	54	67	0.77	0.47-1.25	1.01(0.80-1/27)	0.93
Incident cases only: 480 cases, 515 controls [a]														
ADORA2A rs5760423														
Cups/day, mean ± SD [b]	4.6±2.7	4.6±2.9	0.99	0.92-1.07	4.4±3.1	4.6±2.3	0.97	0.91-1.05	4.2±2.3	4.5±2.2	0.93	0.81-1.08	0.98(0.91-1.05)	0.55
Light (0,<=median cupyears)	83	91	1	--	127	110	1	--	42	35	1	--	--	
Heavy (>median cupyears)	94	90	1.10	0.72-1.68	95	131	0.63	0.44-0.92	37	57	0.58	0.30-1.09	0.66(0.46-0.94)	0.02

CYP1A2 rs762551														
Cups/day, mean ± SD [b]	4.6±2.9	4.5± 2.5	1.01	0.94-1.07	4.5±2.8	4.8± 2.6	0.95	0.88-1.04	3.3±1.8	3.7± 1.9	0.84	0.62-1.14	0.95(0.87-1.03)	0.20
Light (0,<=median cupyears)	144	132	1	--	91	84	1	--	17	19	1	--	--	
Heavy (>median cupyears)	133	156	0.79	0.57-1.11	80	104	0.72	0.47-1.10	14	15	0.73	0.25-2.12	1.03(0.69-1.54)	0.89
CYP1A2 rs2472304														
Cups/day, mean ± SD [b]	4.6±3.0	4.6± 2.6	1.00	0.94-1.07	4.4±2.8	4.7± 2.5	0.97	0.90-1.05	3.7±2.0	4.3± 2.5	0.84	0.69-1.03	0.95(0.88-1.03)	0.21
Light (0,<=median cupyears)	118	109	1	--	110	98	1	--	24	27	1	--	--	
Heavy (>median cupyears)	116	142	0.76	0.53-1.10	86	115	0.68	0.46-1.01	24	21	1.04	0.44-2.48	1.12(0.77-1.65)	0.54
Prevalent cases only: 844 cases, 914 controls [a]														
ADORA2A rs5760423														
Cups/day, mean ± SD [b]	3.8±2.6	4.4±2.7	0.90	0.85-0.97	4.0±2.7	4.6±3.0	0.94	0.89-0.99	4.5±4.9	4.6±2.8	0.99	0.93-1.05	1.05(1.00-1.09)	0.05
Light (0,<=median cupyears)	193	171	1	--	223	219	1	--	61	60	1	--	--	
Heavy (>median cupyears)	85	123	0.63	0.45-0.90	111	172	0.71	0.53-0.95	30	46	0.53	0.33-0.86	1.01(0.76-1.34)	0.94
CYP1A2 rs762551														
Cups/day, mean ± SD [b]	4.1±2.8	4.6± 3.0	0.94	0.89-0.98	4.1±3.8	4.4± 2.8	0.96	0.91-1.02	3.8±2.5	4.3± 3.0	0.93	0.81-1.06	1.01(0.96-1.07)	0.73
Light (0,<=median cupyears)	300	272	1	--	208	199	1	--	44	35	1	--	--	
Heavy (>median cupyears)	173	219	0.71	0.54-0.93	95	149	0.62	0.44-0.86	22	34	0.49	0.24-1.01	0.85(0.63-1.16)	0.31
CYP1A2 rs2472304														
Cups/day, mean ± SD [b]	4.1±2.8	4.6± 3.0	0.93	0.88-0.98	4.1±3.8	4.5± 2.8	0.96	0.92-1.01	4.0±2.5	4.3± 2.9	0.95	0.86-1.07	1.01(0.97-1.07)	0.48
Light (0,<=median cupyears)	264	230	1	--	223	219	1	--	61	60	1	--	--	
Heavy (>median cupyears)	150	186	0.68	0.51-0.91	111	172	0.65	0.48-0.89	30	46	0.63	0.35-1.15	0.94(0.70-1.25)	0.66

cOR = "crude" OR, i.e. adjusted only for the covariates year of birth, gender and onset/index age (continuous); *p* for interaction [coffee *genotype] based on chi-square test with *df*=1 (additive genetic model)

[a] Subjects with missing coffee information or SNP data were excluded.

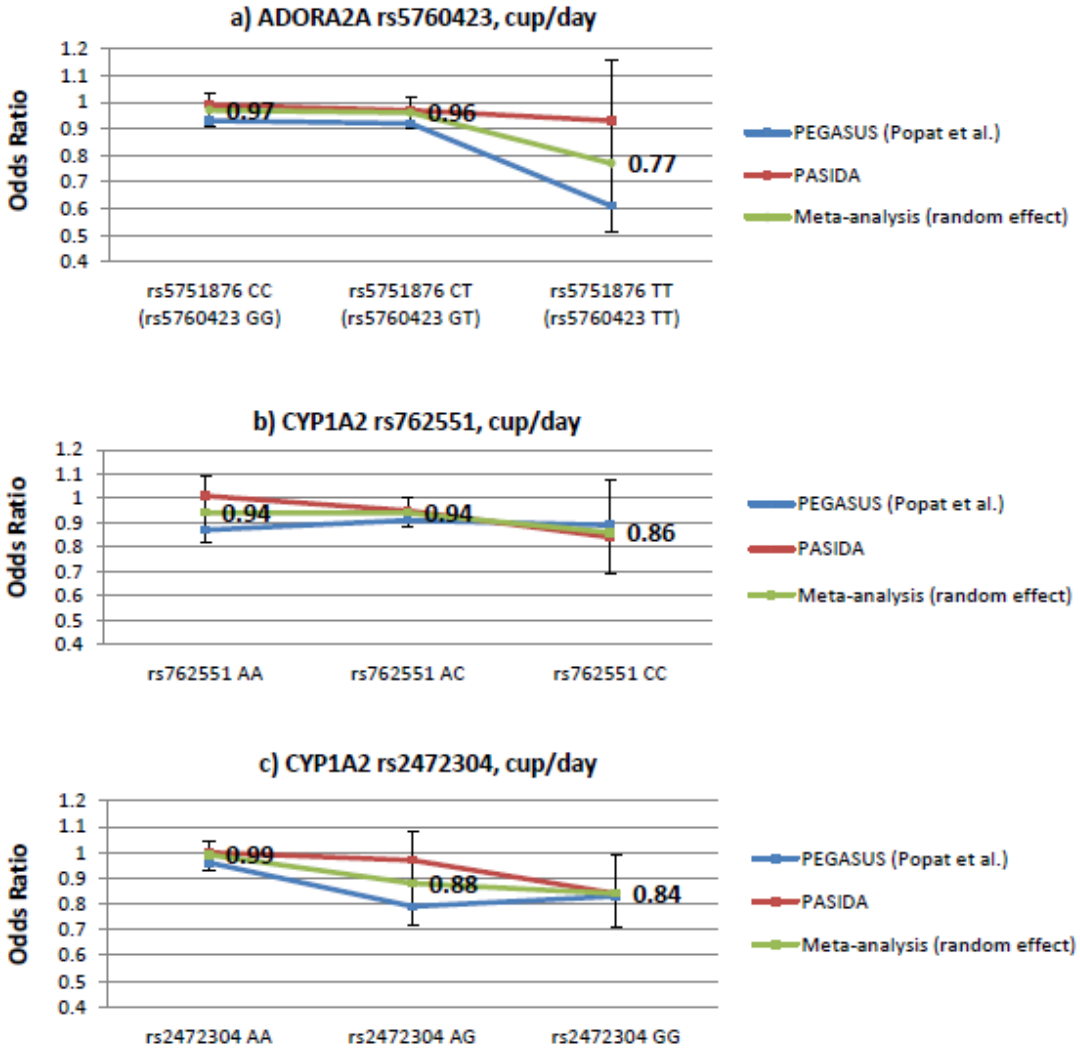
[b] Coffee non-drinkers were excluded (Popat et al. 2011).

Table 2-3. Adjusted odds ratios (OR) and 95% confidence intervals (CI) for the association between caffeinated coffee consumption and Parkinson's disease in the PEGASUS and PASIDA studies, by *ADORA2A* and *CYP1A2* genotypes: meta-analytic results using random-effect models

Caffeinated coffee	Homozygous Major		Heterozygous		Homozygous Minor	
	OR[a]	95% CI	OR[a]	95% CI	OR[a]	95% CI
ADORA2A rs5751876 in PEGASUS (LD with rs5760423 in PASIDA)						
Cups/day, mean ± SD	0.97	0.91-1.03	0.96	0.90-1.02	0.77	0.51-1.16
Ever vs Never*	0.85	0.53-1.36	0.66	0.51-0.84	0.65	0.43-0.98
CYP1A2 rs762551						
Cups/day, mean ± SD	0.94	0.82-1.09	0.94	0.88-1.00	0.86	0.69-1.08
Ever vs Never*	0.70	0.55-0.89	0.83	0.64-1.08	0.43	0.17-1.10
CYP1A2 rs2472304						
Cups/day, mean ± SD	0.99	0.93-1.04	0.88	0.72-1.08	0.84	0.71-0.99
Ever vs Never*	0.75	0.58-0.95	0.71	0.55-0.91	0.67	0.30-1.48

*Heavy vs light coffee consumption in the PASIDA study
OR[a]: Adjusted for age, sex and site in PEGASUS; adjusted for the covariates year of birth, gender and onset/index age (continuous) in PASIDA

Figure 2-1. Odds ratios for cups of coffee consumed per day and PD across *ADORA2A* / *CYP1A2* gene polymorphisms



2.6 Supplement

Supplementary Table 2-S1. Marginal odds ratios and 95% confidence intervals for coffee consumption and PD (incident cases only, as well as incident and prevalent cases)

	all [a]			2006-2009[a]		
	1556 cases/1606 controls	cOR (95% CI)	aOR (95% CI)	554 cases/566 controls	cOR (95% CI)	aOR (95% CI)
Cup per day[b]						
Continuous (mean)	3.9 / 4.4	0.94 (0.92-0.97)	0.96 (0.93-0.99)	4.2 / 4.5	0.97 (0.93-.1.01)	0.99 (0.94-1.04)
Cupyears						
Light (0,<=median)	806 / 746	1.00	1.00	253 / 236	1.00	1.00
Heavy (>median)	518 / 683	0.70 (0.60-0.82)	0.75 (0.64-0.88)	227 / 279	0.76 (0.59-0.98)	0.83 (0.64-1.08)
Caffeine mg/day (quartile)						
0%-25%	429 / 357	1.00	1.00	136 / 108	1.00	1.00
>25%-50%	430 / 445	0.83 (0.78-0.89)	0.87 (0.81-0.93)	154 / 171	0.88 (0.78-0.98)	0.92 (0.81-1.03)
>50-75%	204 / 269	0.70 (0.61-0.80)	0.75 (0.65-0.87)	73 / 95	0.77 (0.61-0.96)	0.84 (0.66-1.06)
>75%-100%	261 / 358	0.58 (0.47-0.71)	0.65 (0.53-0.81)	117 / 141	0.67 (0.48-0.94)	0.77 (0.54-1.10)
		p for trend=<.0001	0.0001		p for trend=0.02	0.15
Years of coffee drinking						
0-37	519 / 467	1.00	1.00	155 / 123	1.00	1.00
>37-45	311 / 352	0.83 (0.76-0.91)	0.85 (0.77-0.93)	111 / 131	0.77 (0.67-0.89)	0.78 (0.67-0.90)
>45-53	301 / 354	0.69 (0.58-0.83)	0.72 (0.60-0.86)	124 / 142	0.60 (0.45-0.80)	0.60 (0.45-0.81)
>53	201 / 264	0.58 (0.44-0.75)	0.61 (0.46-0.80)	91 / 120	0.46 (0.30-0.71)	0.47 (0.30-0.73)
		p for trend=<.0001	0.0004		p for trend=0.0005	0.001

cOR = "crude" OR, i.e. adjusted only for the covariates year of birth, gender and onset/index age (continuous)
aOR = adjusted OR, i.e. adjusted for year of birth, gender, onset/index age (continuous), ever smoking and family history of PD
[a]Missing coffee information: all (n=409), incident cases/controls (n=125); [b]Results remained the same after coffee non-drinkers were removed.

Supplementary Table 2-S2. Marginal odds ratios and 95% confidence intervals for ADORA2A / CYP1A2 gene polymorphisms and PD (additive genetic model; incident cases only, as well as incident and prevalent cases)

	all[a]			2006-2009[b]		
	1556 cases/1606 controls	cOR	aOR (95% CI)	554 cases/566 controls	cOR (95% CI)	aOR (95% CI)
ADORA2A rs5760423 (%)						
GG	34.5 / 34.4	1.00	1.00	36.4 / 34.2	1.00	1.00
GT	47.5 / 50.5	0.98 (0.89-1.08)	0.96 (0.87-1.06)	46.2 / 47.8	0.94 (0.79-1.11)	0.91 (0.77-1.08)
TT	18.0 / 18.3	0.96 (0.79-1.17)	0.92 (0.75-1.13)	17.4 / 18.1	0.88 (0.63-1.23)	0.83 (0.59-1.17)
CYP1A2 rs762551 (%)						
AA	56.5 / 56.4	1.00	1.00	56.4 / 57.4	1.00	1.00
CA	36.4 / 39.2	0.97 (0.87-1.08)	0.98 (0.87-1.10)	37.1 / 36.0	1.02 (0.84-1.24)	1.02 (0.84-1.24)
CC	7.5 / 7.4	0.94 (0.75-1.17)	0.96 (0.76-1.20)	6.5 / 6.6	1.04 (0.71-1.53)	1.04 (0.70-1.53)
CYP1A2 rs2472304 (%)						
AA	48.4 / 48.4	1.00	1.00	47.1 / 50.3	1.00	1.00
GA	40.7 / 44.0	0.97 (0.87-1.08)	0.98 (0.88-1.09)	42.4 / 40.7	1.11 (0.93-1.33)	1.13 (0.94-1.35)
GG	10.9 / 11.0	0.94 (0.76-1.16)	0.97 (0.78-1.20)	10.5 / 9.1	1.23 (0.86-1.77)	1.27 (0.88-1.83)

cOR = "crude" OR, i.e. adjusted only for the covariates year of birth, gender and onset/index age (continuous)

aOR = adjusted OR, i.e. adjusted for year of birth, gender, onset/index age (continuous), ever smoking and family history of PD

[a] Genotyping failed: rs5760423 (n=16), rs762551 (n=14), rs2472304 (n=12)

[b] Genotyping failed: rs5760423 (n=3), rs762551 (n=6), rs2472304 (n=5)

Supplementary Table 2-S3. Adjusted odds ratios (OR) and 95% confidence intervals (CI) for caffeinated coffee consumption and Parkinson's disease in PASIDA, by *ADORA2A* and *CYP1A2* genotypes

Caffeinated coffee	Homozygous Major				Heterozygous				Homozygous Minor				OR interaction	*p*-value
	Cases	Controls	cOR	95% CI	Cases	Controls	cOR	95% CI	Cases	Controls	cOR	95% CI		
Incident and prevalent cases: 1324 cases, 1429 controls [a]														
ADORA2A rs5760423														
Caffeine mg/day (quartile)														
0%-25%	157	128	1	--	201	167	1	--	71	60	1	--	--	--
>25%-50%	135	138	0.84	0.75-0.95	213	221	0.82	0.74-0.91	78	81	0.81	0.68-0.96	--	--
>50-75%	65	84	0.71	0.57-0.90	96	137	0.67	0.55-0.83	41	48	0.66	0.47-0.93	--	--
>75%-100%	98	125	0.60	0.43-0.85	117	171	0.55	0.41-0.75	44	61	0.53	0.32-0.89	1.01(0.92-1.11)	0.83
Years of coffee drinking														
0-37	185	152	1	--	245	230	1	--	86	83	1	--	--	--
>37-45	96	123	0.89	0.77-1.03	154	175	0.79	0.69-0.90	59	52	0.84	0.67-1.06	--	--
>45-53	96	116	0.79	0.59-1.07	146	166	0.62	0.48-0.80	57	68	0.71	0.45-1.12	--	--
>53	78	86	0.71	0.45-1.10	89	130	0.49	0.33-0.72	33	48	0.60	0.30-1.19	1.02(0.93-1.12)	0.69
CYP1A2 rs762551														
Caffeine mg/day (quartile)														
0%-25%	239	188	1	--	154	136	1	--	35	32	1	--	--	--
>25%-50%	232	254	0.86	0.78-0.94	162	154	0.80	0.72-0.90	35	32	0.77	0.59-1.01	--	--
>50-75%	120	136	0.73	0.61-0.88	69	116	0.65	0.51-0.82	14	15	0.60	0.35-1.02	--	--
>75%-100%	159	201	0.63	0.48-0.83	89	130	0.52	0.37-0.74	13	24	0.46	0.21-1.03	1.05(0.94-1.17)	0.40
Years of coffee drinking														
0-37	287	248	1	--	198	183	1	--	33	31	1	--	--	--
>37-45	188	205	0.82	0.72-0.92	97	119	0.85	0.74-0.98	25	27	0.87	0.60-1.26	--	--
>45-53	173	182	0.67	0.52-0.85	106	141	0.72	0.54-0.97	21	27	0.76	0.36-1.60	--	--
>53	105	148	0.54	0.38-0.78	77	97	0.62	0.40-0.95	19	18	0.66	0.22-2.02	0.96(0.87-1.08)	0.52
CYP1A2 rs2472304														
Caffeine mg/day (quartile)														
0%-25%	209	158	1	--	174	151	1	--	44	46	1	--	--	--
>25%-50%	198	215	0.85	0.77-0.93	178	182	0.81	0.73-0.91	52	47	0.82	0.66-1.02	--	--

>50-75%	97	114	0.72	0.59-0.87	82	130	0.66	0.53-0.83	22	25	0.67	0.44-1.04	--	--
>75%-100%	144	180	0.61	0.45-0.81	96	141	0.54	0.39-0.75	21	36	0.55	0.29-1.06	1.02(0.92-1.13)	0.76
Years of coffee drinking														
0-37	252	204	1	--	213	209	1	--	50	52	1	--	--	--
>37-45	165	179	0.76	0.66-0.86	108	137	0.91	0.80-1.04	38	35	0.87	0.65-1.17	--	--
>45-53	142	159	0.57	0.44-0.74	124	152	0.83	0.63-1.09	32	43	0.76	0.42-1.36	--	--
>53	91	129	0.43	0.29-0.64	89	109	0.76	0.50-1.14	21	25	0.66	0.27-1.59	0.94(0.85-1.04)	0.22
<u>Incident cases only</u>: 480 cases, 515 controls [a]														
ADORA2A rs5760423														
Caffeine mg/day (quartile)														
0%-25%	46	44	1	--	70	47	1	--	20	17	1	--	--	--
>25%-50%	54	57	0.99	0.82-1.19	70	79	0.79	0.66-0.94	29	34	0.85	0.63-1.14	--	--
>50-75%	26	28	0.97	0.67-1.40	35	51	0.62	0.44-0.88	12	16	0.72	0.40-1.30	--	--
>75%-100%	51	52	0.96	0.55-1.66	47	64	0.49	0.29-0.82	18	25	0.61	0.26-1.48	1.10(0.94-1.29)	0.24
Years of coffee drinking														
0-37	54	50	1	--	80	55	1	--	20	18	1	--	--	--
>37-45	40	48	0.90	0.72-1.13	57	64	0.65	0.52-0.81	14	19	0.95	0.65-1.39	--	--
>45-53	44	44	0.81	0.51-1.28	51	69	0.42	0.27-0.65	29	28	0.89	0.42-1.92	--	--
>53	39	39	0.73	0.37-1.45	35	54	0.27	0.14-0.52	16	27	0.85	0.27-2.66	1.12(0.96-1.32)	0.16
CYP1A2 rs762551														
Caffeine mg/day (quartile)														
0%-25%	76	60	1	--	47	39	1	--	13	9	1	--	--	--
>25%-50%	87	99	0.92	0.79-1.06	54	57	0.86	0.71-1.04	12	14	0.65	0.38-1.12	--	--
>50-75%	42	54	0.84	0.62-1.13	28	34	0.74	0.50-1.09	3	6	0.43	0.14-1.25	--	--
>75%-100%	72	75	0.77	0.49-1.20	42	58	0.64	0.36-1.14	3	5	0.28	0.06-1.40	1.10(0.92-1.33)	0.30
Years of coffee drinking														
0-37	91	66	1	--	55	47	1	--	8	7	1	--	--	--
>37-45	73	76	0.73	0.61-0.89	34	46	0.83	0.66-1.05	4	9	0.89	0.45-1.76	--	--
>45-53	65	75	0.54	0.37-0.79	47	52	0.69	0.43-1.11	12	13	0.80	0.21-3.11	--	--
>53	49	72	0.39	0.22-0.70	35	43	0.58	0.28-1.17	7	5	0.72	0.09-5.49	0.87(0.72-105)	0.15
CYP1A2 rs2472304														
Caffeine mg/day (quartile)														

0%-25%	63	48	1	--	55	47	1	--	18	12	1	--	--	--
>25%-50%	73	85	0.89	0.76-1.05	64	67	0.88	0.74-1.05	16	18	0.74	0.50-1.11	--	--
>50-75%	34	48	0.80	0.58-1.11	32	39	0.78	0.54-1.11	6	8	0.55	0.25-1.22	--	--
>75%-100%	64	70	0.72	0.44-1.16	45	60	0.68	0.40-1.17	8	10	0.41	0.13-1.36	1.05(0.88-1.25)	0.57
Years of coffee drinking														
0-37	79	54	1	--	62	55	1	--	13	13	1	--	--	--
>37-45	59	68	0.69	0.55-0.85	44	50	0.83	0.67-1.04	8	12	1.06	0.65-1.72	--	--
>45-53	52	65	0.47	0.31-0.72	52	61	0.69	0.45-1.07	19	16	1.12	0.43-2.96	--	--
>53	44	65	0.32	0.17-0.61	39	47	0.58	0.30-1.11	8	7	1.19	0.28-5.08	0.85(0.71-1.02)	0.07
Prevalent cases only: 844 cases, 914 controls [a]														
ADORA2A rs5760423														
Caffeine mg/day (quartile)														
0%-25%	111	84	1	--	131	120	1	--	51	43	1	--	--	--
>25%-50%	81	81	0.76	0.65-0.88	143	142	0.84	0.74-0.95	49	47	0.80	0.65-0.99	--	--
>50-75%	39	56	0.57	0.42-0.77	61	86	0.70	0.54-0.90	29	32	0.64	0.42-0.99	--	--
>75%-100%	47	73	0.43	0.27-0.68	70	107	0.59	0.40-0.85	26	36	0.52	0.27-0.98	0.95(0.84-1.08)	0.43
Years of coffee drinking														
0-37	131	102	1	--	165	175	1	--	66	65	1	--	--	--
>37-45	56	75	0.88	0.72-1.07	97	111	0.88	0.75-1.04	45	33	0.80	0.60-1.08	--	--
>45-53	52	72	0.77	0.52-1.14	95	97	0.78	0.56-1.08	28	40	0.65	0.36-1.17	--	--
>53	39	47	0.67	0.38-1.21	54	76	0.68	0.42-1.12	17	21	0.52	0.21-1.26	0.96(0.84-1.08)	0.49
CYP1A2 rs762551														
Caffeine mg/day (quartile)														
0%-25%	163	128	1	--	107	97	1	--	22	23	1	--	--	--
>25%-50%	145	155	0.82	0.73-0.92	108	97	0.78	0.67-0.90	23	18	0.83	0.61-1.13	--	--
>50-75%	78	82	0.67	0.53-0.85	41	82	0.60	0.45-0.81	11	9	0.68	0.37-1.27	--	--
>75%-100%	87	126	0.55	0.39-0.79	47	72	0.47	0.30-0.73	10	19	0.56	0.22-1.43	1.02(0.89-1.16)	0.82
Years of coffee drinking														
0-37	196	182	1	--	143	136	1	--	25	24	1	--	--	--
>37-45	115	129	0.88	0.75-1.02	63	73	0.86	0.72-1.04	21	18	0.90	0.56-1.42	--	--
>45-53	108	107	0.77	0.56-1.05	59	89	0.74	0.51-1.08	9	14	0.80	0.32-2.03	--	--
>53	56	76	0.67	0.42-1.07	42	54	0.64	0.37-1.12	12	13	0.72	0.18-2.89	1.02(0.89-1.17)	0.78

CYP1A2 rs2472304														
Caffeine mg/day (quartile)														
0%-25%	146	110	1	--	119	104	1	--	26	34	1	--	--	--
>25%-50%	125	130	0.82	0.72-0.93	114	115	0.78	0.68-0.90	36	29	0.86	0.66-1.12	--	--
>50-75%	63	66	0.67	0.52-0.86	50	91	0.61	0.46-0.80	16	17	0.74	0.43-1.24	--	--
>75%-100%	80	110	0.54	0.37-0.79	51	81	0.47	0.31-0.72	13	26	0.63	0.29-1.39	1.00(0.88-1.13)	0.98
Years of coffee drinking														
0-37	173	150	1	--	151	154	1	--	37	39	1	--	--	--
>37-45	106	111	0.80	0.68-0.95	64	87	0.97	0.81-1.15	30	23	0.78	0.54-1.14	--	--
>45-53	90	94	0.65	0.46-0.91	72	91	0.94	0.66-1.33	13	27	0.61	0.29-1.31	--	--
>53	47	64	0.52	0.31-0.87	50	62	0.91	0.54-1.53	13	18	0.48	0.15-1.50	1.00(0.88-1.13)	0.95

cOR = "crude" OR, i.e. adjusted only for the covariates year of birth, gender and onset/index age (continuous); *p* for interaction [coffee *genotype] based on chi-square test with *df*=1 (additive genetic model)

[a] Subjects with missing coffee information or SNP data were excluded.

Supplementary Table 2-S4. Adjusted odds ratios (OR) and 95% confidence intervals (CI) for the association between caffeinated coffee consumption and Parkinson's disease in the PEGASUS and PASIDA studies, by *ADORA2A* and *CYP1A2* genotypes: meta-analytic results using fixed effect models

Caffeinated coffee	Homozygous Major		Heterozygous		Homozygous Minor	
	OR[a]	95% CI	OR[a]	95% CI	OR[a]	95% CI
ADORA2A rs5751876 in PEGASUS (LD with rs5760423 in PASIDA)						
Cups/day, mean ± SD	0.97	0.91-1.03	0.96	0.90-1.02	0.85	0.75-0.97
Ever vs Never*	0.83	0.63-1.10	0.66	0.51-0.84	0.65	0.43-0.98
CYP1A2 rs762551						
Cups/day, mean ± SD	0.97	0.92-1.03	0.94	0.88-1.00	0.86	0.69-1.08
Ever vs Never*	0.69	0.55-0.87	0.83	0.64-1.08	0.42	0.21-0.85
CYP1A2 rs2472304						
Cups/day, mean ± SD	0.99	0.93-1.04	0.93	0.86-0.99	0.84	0.71-0.99
Ever vs Never*	0.75	0.58-0.95	0.71	0.55-0.91	0.64	0.37-1.10

*Heavy vs light coffee consumption in the PASIDA study
OR[a]: Adjusted for age, sex and site in PEGASUS; adjusted for the covariates year of birth, gender and onset/index age (continuous) in PASIDA

3. Pooled analysis of the interaction between *HLA-DRB1* and smoking in Parkinson's disease

3.1 Introduction

The human leukocyte antigen (HLA) system that encodes the major histocompatibility complex (MHC) proteins is a crucial part of the immune system. Class II MHC presents extracellular antigens to T lymphocytes. HLA-DR is a MHC class II cell surface receptor encoded by the human leukocyte antigen complex on chromosome 6p21.31 and is composed of an alpha and beta chain coded by the *HLA-DRA* and *HLA-DRB* genes, respectively. While the *HLA-DRA* locus is largely monomorphic and not prominent in HLA-disease association studies, the *HLA-DRB* locus has been reported to influence both autoimmune (e.g., rheumatoid arthritis,(Bowes et al. 2008; Bang et al. 2010) multiple sclerosis(Lincoln et al. 2005; Ockinger et al. 2016)) and neurodegenerative diseases (e.g., Alzheimer's disease,(Lambert et al. 2013) Parkinson's disease (PD)(Hamza et al. 2010; Saiki et al. 2010; Do et al. 2011; Ahmed et al. 2012; Sun et al. 2012)).

In our previous study, we examined 51 single nucleotide polymorphisms (SNPs) in the *HLA-DR* region in two French case-control datasets. We found that rs660895-G (in the *HLA-DRB1* region) showed the strongest association with PD (OR=0.70, 95% CI=0.57-0.87) and did not confirm the previously reported association with the *HLA-DRA* locus.(Hamza et al. 2010) In the same study, SNP-based imputation of HLA alleles showed an inverse association with the *HLA-DRB1*04* allele,(Ahmed et al. 2012) and this SNP was also the strongest HLA region hit in a genome-wide association study.(Do et al. 2011) This finding was corroborated in a British case-control study that reported a lower frequency of the *HLA-DRB1*04* allele in PD patients than in matched controls.(Saiki et al. 2010) A Chinese study also found a reduced risk of PD associated with the *HLA-DRB1*0406* allele.(Sun et al. 2012) Finally, based on SNP-based

imputation of HLA alleles, two GWAS showed an inverse association between *HLA*DRB1*0404* and PD.(Wissemann et al. 2013)

PD is likely to be caused by a combination of genetic and environmental factors, acting in concert with each other. While smoking increases the risk of many diseases, in particular through inflammatory mechanisms,(Walser et al. 2008; Baka et al. 2009) it has consistently been shown to be inversely associated with PD.(Ritz et al. 2007) It remains unknown, however, whether this robust inverse association is causal or not. Some have argued that it cannot be explained by survival bias or confounding,(Morens et al. 1995) but reverse causation cannot be discarded as an alternative explanation, due to increasing evidence of a long premotor period.(Ritz et al. 2014; Elbaz 2016)

The identification of gene-environment interactions represents a possible approach to better understand the biological mechanisms underlying the association between environmental exposures and disease and to improve causal inference. In our previous study we showed that the rs660895-PD association was restricted to never smokers (OR=0.63, 95% CI= 0.49-0.81) while there was no significant association in ever smokers (OR=0.93, 95% CI=0.61-1.37); however, the interaction test was not significant at the traditional α-level of 0.05 (p=0.11),(Ahmed et al. 2012) and larger studies are necessary to investigate this interaction.

Thus, the aims of this study were: (*i*) to investigate the interaction of rs660895 in the *HLA-DRB1* region and smoking in PD in a large population-based case-control study conducted in Denmark; (*ii*) to perform a pooled analysis of the smoking-rs660895 interaction based on individual data from the Danish and aforementioned French datasets.

3.2 Materials and Methods

Study population

The data used for analyses have been assembled in the Parkinson's disease in Denmark (PASIDA) study, and the French *Terre* and *Partage* studies.

The PASIDA study: enrolled idiopathic PD patients treated at 10 neurological treatment centers and identified from the Danish National Hospital Register between 1996 and 2009. Population controls were selected from the Danish Central Population Registry and matched on year of birth, sex, and being alive and free of PD at the time of case identification. Detailed recruitment information was published previously.(Kenborg et al. 2015) Briefly, from 2,762 eligible patients, 949 (34%) were excluded due to the following reasons: lack of a PD diagnosis on medical records (n=179), declined to participate (n=497), failure to confirm idiopathic PD according to accepted criteria in medical record review (n=273).(Wermuth et al. 2012) From 3,626 eligible controls, 1,887 (52%) completed an interview and questionnaire. Of these, DNA samples were available for 1,575 (87%) patients and 1,607 (85%) controls. Later, 20 subjects with dementia prior to interview and 20 subjects with missing information for either rs660895, ever smoking, or covariates were excluded. Overall, 1547 PD cases and 1595 controls were retained in the analysis.

Terre and *Partage* studies: Two population-based case-control studies were performed within a French health insurance system for persons working in agriculture and related occupations (*Mutualité Sociale Agricole*, MSA). PD patients from 62 metropolitan districts who received treatment between 1998 and 1999 were enrolled in the *Terre* study, whereas those from five districts treated between 2006 and 2007 were enrolled in *Partage*. These patients were examined by neurologists and PD was diagnosed using standard criteria.(Bower et al. 1999) We randomly selected eligible controls from among all MSA members who requested reimbursement for health expenses (*Terre*) or who were listed on electronic insurance records

(*Partage*). Controls were matched to cases on age, sex, and district of residency, and did not report cardinal signs of PD. Subjects with dementia or those who were bedridden were excluded. Participation rates were high and similar in cases and controls in both studies (*Terre*: cases, 83%, controls, 75%; *Partage*: cases, 82%, controls, 77%).(Elbaz et al. 2009; Ahmed et al. 2014) We further excluded subjects lacking DNA samples, of non-European ancestry or with missing data, leaving 202 PD and 529 controls (*Terre*) and 307 PD and 599 controls (*Partage*) for analysis.(Levecque et al. 2003; Galanaud et al. 2005)

The institutional review boards (IRB) of UCLA, the Danish Data Protection Agency, and the ethics committee of Copenhagen approved the PASIDA study, whereas the IRB of the ethnics committees of the Kremlin-Bicêtre and Pitié-Salpêtrière University Hospitals approved the *Terre* and *Partage* studies. All subjects provided written informed consent.

Data collection and smoking assessment

Participants provided information on demographics and lifestyle through telephone interviews or in writing (PASIDA) and in-person interviews (*Terre, Partage*). In PASIDA, we defined ever (habitual) smokers as persons who smoked at least one cigarette, pipe, or cigar/cheroot per week for at least 6 months. In the French studies, ever smokers included ex- and current smokers of cigarettes who had smoked at least one cigarette per week for at least 6 months; there were few pipe or cigar/cheroot smokers and all reported smoking cigarettes too.

For all studies, we collected lifetime smoking information, i.e., the amount and duration of smoking and changes of smoking behavior, and used this information to calculate total pack-years of cigarette smoking. Only smoking behavior that occurred prior to the date of the first motor symptom or PD diagnosis in cases and the same index date in matched controls were

considered. Information on coffee drinking was available in the PASIDA and *Partage* studies only.(Ahmed et al. 2014)

Genotyping

DNA was extracted from saliva (*PASIDA, Partage*) or blood (*Terre*) samples. PASIDA samples were genotyped for rs660895 using multiplex Taqman allelic discrimination assays (Applied Biosystems) according to the manufacturer's protocol. For *Terre* and *Partage* samples, rs660895 was included in a microarray that was processed by Integragen (Evry, France) using Illumina technology and Infinium iSelect custom genotyping.(Ahmed et al. 2012) All three datasets passed the criteria for quality assurance, i.e., genotyping efficiency >95% and accuracy >99.5%.

Statistical analysis

We assessed Hardy-Weinberg equilibrium in controls from each study using Pearson's chi-square test. For each study, we broke the matching and used unconditional logistic regression analysis adjusted for the matching factors (PASIDA: year of birth, sex, onset/index age (age at index date); *Terre, Partage*: index age, sex, and district of residency) to increase statistical efficiency. We first assessed the marginal associations of rs660895 and ever smoking with PD, and computed odds ratios (OR) and 95% confidence intervals (CI). We also considered family history of PD as a potential confounder. Similar to what has been previously reported in the literature, we analyzed rs660895 using co-dominant, additive, and dominant genetic models; recessive models were not considered due to the low frequency of homozygotes for the minor allele. We then examined the independent and joint effects of ever smoking and rs660895 in each study. In further analyses, we used pack-years of smoking defined as a three level variable (never smoker, ≤18, >18(Levecque et al. 2003)). The 18 pack-years cutoff was the median among not-

PD affected ever smokers in the *Terre* study, for PASIDA and *Partage*, the median pack-years among not-PD affected ever smokers was 15 and 12, respectively; we chose the cutoff of 18 to separate ligh/moderate from heaviest smokers. We also conducted a sensitivity analysis using 12 pack-years as the cut-off. Multiplicative interactions between ever smoking and rs660895 were examined by introducing a product term between these variables into the regression model. Interactions were tested on a multiplicative scale, as departure from multiplicativity is expected when at least one of the interacting factors is preventive.(Weinberg 1986)

Both for analyses of marginal associations and interactions, we pooled the estimates from the three studies using random-effects meta-analysis.(Bradburn et al. 1999) We tested for between-study heterogeneity with the chi square–based Q statistic (deemed significant for $p<.10$) and quantified its extent with I^2, which ranges from 0% to 100% and represents the proportion of between-study variability ascribed to heterogeneity rather than to chance;(Higgins et al. 2002; Higgins et al. 2003) I^2 values of 0%-24% suggest little heterogeneity, 25%-49% reflect moderate heterogeneity, 50%-74% reflect large heterogeneity, and >75% reflect very large heterogeneity.

Lastly, we assessed the association between rs660895 and smoking among cases and controls separately using unconditional logistic regression adjusted for onset/index age, sex, and matching factors in each study, and again pooled the three estimates using random-effects meta-analysis. Case-only studies have been shown to be statistically more powerful than case-control studies when assessing gene-environment interactions assuming that the gene-environment independence assumption holds, i.e., that rs660895 is not associated with smoking among controls.(Khoury et al. 1996)

In sensitivity analyses, we checked whether there was an interaction between rs660895 and coffee drinking and whether adjusting for coffee drinking modified our results for the rs660895-smoking interaction.

All analyses were performed using SAS v9.4 (SAS Institute, Cary, NC, USA) and STATA v14.1 (Stata Corporation, College Station, TX, USA).

Power calculation

Assuming the following parameters based on 1000 Genomes Project European ancestry samples and the literature (minor allele frequency, MAF: 28%; ever smoking: 70%; marginal genetic OR: 0.70;(Ahmed et al. 2012) marginal smoking OR: 0.64;(Noyce et al. 2012) n_{cases}=2,056, $n_{controls}$=2,723), our study had a statistical power of 97.3% to detect an interaction OR of 1.5 (as we previously reported***(Ahmed et al. 2012)***) at a 0.05 two-sided alpha level under an additive genetic model; power was 90.0% for an interaction OR of 1.4. For the case-only design, our study had 99.9% statistical power to detect an interaction OR of 1.5 and 98.9% for an interaction OR of 1.4. Power calculations were performed in Quanto v.1.2.4 (USC, Los Angeles, CA, USA.(Gauderman 2002)

3.3 Results

The three studies included a total of 2,056 cases (PASIDA, 1,547; *Terre*, 202; *Partage*, 307) and 2,723 controls (PASIDA, 1,595; *Terre*, 529; *Partage*, 599) for analysis (**Table 3-1**). There was no deviation from Hardy-Weinberg equilibrium for rs660895 in controls in any dataset (PASIDA, p=0.24; *Terre*, p=0.98; *Partage*, p=0.31). As expected, compared with control subjects, PD patients were more likely to be male, to report a positive family history of PD, and to have never smoked (**Table 3-1**). **Table S3-1** shows the cross tabulation of smoking and rs660895 in cases and controls from each study.

In each study and in the pooled analysis, ever-smoking was inversely associated with PD (pooled OR=0.56, 95%CI: 0.49-0.64, p<0.001; **Table 3-2**) with a dose-effect relation (p for trend <0.001) and without any heterogeneity across studies (I^2=0.0%). Similar to what we reported previously(Ahmed et al. 2012), we observed an inverse association between rs660895-G and PD in the Danish study (OR per one-minor allele=0.85, 95% CI=0.75-0.96, p=0.010). In the French studies, OR estimates adjusted solely for onset/index age and sex did not differ from estimates further adjusted for district of residency, thus only data from the former analysis are shown. The inverse association between rs660895-G and PD was significant in pooled analyses (OR per one-minor allele=0.79, 95% CI=0.68-0.93, p=0.003) with moderate heterogeneity (I^2=31.3%) and no heterogeneity under a dominant model (OR=0.79, 95% CI=0.70-0.90, p<0.001, I^2=0.0%).

Table S3-2 shows analyses of the interaction between smoking and rs660895 in each study and **Table 3-3** shows results of the pooled analysis. In PASIDA, the interaction between ever smoking and rs660895 was statistically significant, no matter which genetic model was considered. Under the dominant model, the inverse association between rs660895-G carriers and PD in never smokers (OR=0.66, 95% CI=0.53-0.82) was lost in ever smokers (OR=0.46/0.46=1.00, 95% CI=0.82-1.21) corresponding to an interaction OR of 1.52 (95% CI=1.13-2.04, p=0.005); the inverse association between smoking and PD seen in rs660895-G non-carriers (OR=0.46, 95% CI=0.38-0.56) was weaker in rs660895-G carriers (OR=0.46/0.66=0.70, 95% CI=0.54-0.86). Interactions ORs were of a similar size, but not statistically significant, in the French studies (*Terre*, OR=1.74, 95% CI=0.79-3.84, p=0.17; *Partage*, OR=1.52, 95% CI=0.76-3.06, p=0.24). In pooled analyses of the three studies (**Table 3-3**), there was no heterogeneity for any of the interaction estimates. Under a dominant model, rs660895-AG+GG was inversely associated with PD among never smokers (OR=0.64, 95%

CI=0.54-0.77) but not in ever smokers (OR=0.48/0.48=1.00, 95% CI=0.84-1.19), corresponding to an interaction OR of 1.54 (95% CI=1.19-2.00, p=0.001, I^2=0.0%); similarly, the inverse association between smoking and PD seen in rs660895-G non-carriers (OR=0.48, 95% CI=0.41-0.57) was weaker in rs660895-G carriers (OR=0.48/0.64=0.75, 95% CI=0.60-0.90). In analyses based on pack-years, interactions were significant for both light (OR=1.64, 95% CI=1.20-2.25, p=0.002) and heavy (OR=1.48, 95% CI=1.02-2.13, p=0.040) smoking; interaction ORs were somewhat stronger for light than heavy smoking but the difference was not significant (p=0.68). Results remained consistent when the cutoff of 12 pack-years was used (light smokers: OR interaction=1.71, 95% CI=1.22-2.41, p=0.002; heavy smokers: OR interaction=1.47, 95% CI=1.06-2.05, p=0.021).

Table 3-4 presents the results of case-only and control-only analyses. Case-only analyses showed a significant and positive association between ever smoking and rs660895-G in cases from PASIDA under all genetic models considered. This pattern was confirmed in pooled analyses without any heterogeneity across studies (I^2=0.0% for all genetic models). Among controls, rs660895-G tended to be less frequent among ever smokers than never smokers, but this difference was not significant in any of the studies and in pooled analyses.

Both in the *PASIDA* and *Partage* studies, smoking and coffee drinking were positively associated in controls. There was no significant interaction between rs660895 and coffee drinking (either as a binary variable or in cups per day) in both datasets and in pooled analyses; in addition, the rs660895-smoking interaction remained significant in PASIDA and in pooled analyses after adjusting for coffee consumption (data not shown).

3.4 Discussion

In this pooled analysis of individual data from three population-based case-control studies, we replicate an inverse association between rs660895-G and PD, and provide the first evidence of a statistically significant interaction between rs660895 and smoking for the risk of PD. The inverse association between rs660895-G and PD significantly decreased in smokers compared to non-smokers.

We previously reported significant associations of the *HLA-DRB1*04* allele and rs660895 with PD, and a strong correlation of rs660895-G with *HLA-DRB1*04* (Spearman's rho= 0.77, $p<0.0001$) in the French studies included here.(Ahmed et al. 2012) These results are in agreement with our findings for rs660895 in the Danish study. Additionally, a British study of 528 cases and ~3000 controls previously reported a lower *HLA-DRB1*04* allele frequency in PD patients compared with controls (17% vs. 20%, p=0.01). A subtype of *HLA-DRB1*04*, *HLA-DRB1*0406*, which is rare in European but common in Asian populations, was also significantly related to PD in a Han Chinese study of 567 cases and 746 controls (OR=0.12, $p=5x10^{-5}$).(Sun et al. 2012) An inverse association between *HLA*DRB1*0404* and PD was also reported in two GWAS using SNP-based imputation.(Wissemann et al. 2013) These findings support a role of *HLA-DRB1* in the etiology of PD across ethnicities. The *HLA-DRB1* gene encodes MHC class II beta chain proteins, which are widely expressed on antigen presenting cells such as B cells, dendritic cells, and macrophages, including brain microglia. Studies have reported microglial activation in the substantia nigra of PD patients and in animal models, and some suggested that modulation of microglia related oxidative stress and inflammation might be a treatment strategy for PD.(McGeer et al. 1988; McGeer et al. 2008) Observational studies showing an inverse association between NSAIDs use and PD also support a role for neuroinflammation in PD.(Gagne et al. 2010)

In addition, our findings suggest that the inverse association between *HLA*DRB1* and PD is present only in never smokers, and the smoking-PD inverse association was less pronounced in rs660895-G carriers. In analyses based on pack-years of smoking, although interaction ORs tended to be higher for light smokers than heavy smokers, the difference was small and not statistically significant. The finding that interaction effects on an endpoint decrease with exposure has previously been termed 'low exposure – gene effect' with several examples available in the cancer literature(Taioli et al. 1998); i.e. the interaction is relevant for the outcome at low doses while at high doses the interaction does not matter. Even larger studies would be needed to further investigate such dose dependent patterns for smoking and *HLA-DRB1* in PD.

Epidemiologic studies have consistently reported a significant inverse association between smoking and PD with a dose-effect relation.(Ritz et al. 2007) This has led some researchers to advocate the use of nicotine to prevent or treat PD.(Quik et al. 2012) There is however no consensus yet whether this association is causal or is due to bias or reverse causation, such that those who will develop PD have a diminished nicotinic response allowing them to quit more easily in the prodromal phase of PD; indeed, in PASIDA, the risk of developing PD was higher among former smokers who reported that quitting had been easy and who did not use nicotine substitutes.(Ritz et al. 2014) Gene-environment studies may represent a way to improve our understanding of the mechanisms underlying the puzzling association between smoking and PD. Our finding of a sub-multiplicative interaction between *HLA-DRB1* and smoking suggests that they are involved in common pathways, possibly related to neuroinflammation; further research is needed to better understand the mechanism underlying the *HLA-DRB1*-PD association and the interaction with smoking. In this context, it is of interest that smoking and HLA have been

previously reported to interact in other conditions, including multiple sclerosis (MS),(Simon et al. 2010; Ockinger et al. 2016) rheumatoid arthritis (RA),(Bang et al. 2010; Karlson et al. 2010; Bang et al. 2013) lymphoma,(Baecklund et al. 2017) and adult idiopathic inflammatory myopathies.(Chinoy et al. 2012) In MS patients, smoking influenced the proliferation of immune cells in the lung; the magnitude of the increase in alveolar macrophage concentration was associated with *HLA-DRB1* alleles.(Ockinger et al. 2016) In RA sufferers, smoking and *HLA-DRB1*04* interacted to influence their risk of cardiovascular disease;(Boechat Nde et al. 2012) the hypothesized mechanism was that both *HLA-DRB1*04* and smoking enhanced T-cell activation in the lung, which may lead to inflammatory processes in the endothelial cells of the vessel walls and subsequent development of cardiovascular disease. Regarding lymphoma, a recent study showed a positive association with *HLA-DRB1* shared epitope alleles (including *HLA-DRB1*04*) as well as a positive additive interaction between *HLA-DRB1* shared epitope alleles and smoking, such that there was no association with smoking in subjects without shared epitope alleles and a significant positive association in those who carried two shared epitope alleles.(Baecklund et al. 2017) Smoking induces the release of intracellular antigens and the authors hypothesized that shared epitope-containing *HLA-DRB1* variants may facilitate the presentation of these antigens to helper T cells and induce B-cell chronic activation which may increase the risk of abnormal proliferation.(Baecklund et al. 2017) Interestingly, a meta-analysis showed PD patients to have a 24% reduced risk of leukemia and lymphoma.(Bajaj et al. 2010) This finding is in agreement with associations in opposite directions for *HLA-DRB1* with PD and lymphoma; although it is unclear how smoking and *HLA-DRB1* may interact at the biological level, the observation of interactions in opposite directions in both diseases, one characterized by cell proliferation and the other by neuronal death, deserves further investigation. Lastly, smoking

and *HLA-DRB1* have also been reported to interact in adult idiopathic inflammatory myopathies.(Chinoy et al. 2012)

Our study has several strengths including its large size and study design (pooled analysis of individual data from three population-based studies) that provided us with sufficient statistical power to identify an interaction of the size previously reported(Ahmed et al. 2012) or even weaker. In case-only analyses, rs660895 was associated with smoking among cases and the assumption of gene-environment independence among controls was met. As studies were population-based, controls were representative of the underlying population from which the cases arose. In addition, in each study, participants were ethnically homogeneous and of European ancestry. The diagnosis of PD was clinically confirmed by neurologists in the two French studies, and by careful systematic review of complete medical records in PASIDA. Finally, the effect size of ever smoking was consistent with that reported in a previous meta-analysis of 67 studies (OR=0.64, 95% CI=0.60-0.69).(Noyce et al. 2012) However, there are also limitations. Although rs660895-G tags the *HLA-DRB1*04* allele, we cannot provide more detailed information about allele subtypes and haplotypes as additional SNPs were not typed. Besides, the recruitment of PD patients includes both incident and prevalent cases, but stratifying by disease duration produced the same results making survival bias unlikely (data not shown). Finally, residual confounding due to unmeasured factors cannot be ruled out; however, under gene-environment independence in the population, it has been shown that only unmeasured confounders that also interact with the gene under investigation may bias estimates of gene-environment interactions.(Vanderweele et al. 2013) In our analyses, coffee did not interact with rs660895-G for the risk of PD, and adjusting for coffee drinking had no impact on our findings. In our previous paper, we also reported that there was no interaction between *HLA-DRB1* and

professional pesticide exposure in the French dataset (p for interaction=0.88).(Ahmed et al. 2012)

In conclusion, our findings suggest that genetic variation in the *HLA-DRB1* locus has an effect on the risk of Parkinson's disease in interaction with smoking, and that the protective effect of the G allele of *HLA-DRB1* rs660895 in PD is present only in never smokers. Our study further emphasizes the importance of neuroinflammation in the development of PD.

3.5 Tables

Table 3-1. Characteristics of participants by study

	Denmark (PASIDA)		France (*Terre*)		France (*Partage*)	
Characteristics	**Cases**	**Controls**	**Cases**	**Controls**	**Cases**	**Controls**
N	1547	1595	202	529	307	599
Mean onset/index age (SD)	61.3 (9.6)	61.6 (9.7)	63.8 (7.0)	63.3 (7.8)	66.2 (7.6)	66.3 (7.7)
Mean age at diagnosis (SD)	62.3 (9.2)	-	64.8 (7.0)	-	66.8 (7.5)	-
Mean age at study (SD)	67.7 (8.4)	68.3 (8.7)	67.3 (6.7)	66.9 (7.4)	72.7 (6.6)	72.6 (6.7)
Sex, male (%)	926 (59.9)	970 (60.8)	116 (57.4)	311 (58.8)	177 (57.7)	356 (59.4)
European ancestry (%)	1547 (100.0)	1595 (100.0)	202 (100.0)	529 (100.0)	307 (100.0)	599 (100.0)
Family history of PD (%)	210 (13.6)	88 (5.5)	19 (9.4)	24 (4.5)	36 (11.7)	24 (4.0)
Smoking						
Never Smoker	773 (50.0)	572 (35.9)	150 (74.3)	349 (66.0)	233 (75.9)	402 (67.1)
Ever smoker	774 (50.0)	1023 (64.1)	52 (25.7)	180 (34.0)	74 (24.1)	197 (32.9)
Smoking intensity (%)						
Never Smoker	773 (56.1)	572 (42.1)	150 (74.6)	349 (66.0)	233 (75.9)	402 (67.1)
Light Smoker (≤ 18 pack-years) [a,b]	409 (29.6)	441 (32.4)	32 (15.9)	86 (16.3)	48 (15.6)	125 (20.9)
Heavy Smoker (> 18 pack-years) [a,b]	197 (14.3)	347 (25.5)	19 (9.5)	94 (17.7)	26 (8.5)	72 (12.0)
rs660895 (%)						
AA	974 (63.0)	929 (58.3)	146 (72.3)	354 (66.9)	231 (75.3)	399 (66.6)
AG	510 (33.0)	589 (36.9)	52 (25.7)	157 (29.7)	74 (24.0)	182 (30.4)
GG	63 (4.0)	77 (4.8)	4 (2.0)	18 (3.4)	2 (0.7)	18 (3.0)

[a] Number of subjects with missing data for pack-years: PASIDA n=403, *Terre* n=1.
[b] Median number of pack-years among ever smokers in controls in the *Terre* study (PASIDA: median 15 pack-years, *Partage*: median 12 pack-years).

Table 3-2. Associations between smoking, rs660895, and PD, in each study and overall

	Denmark (PASIDA)		France (*Terre*)		France (*Partage*)		Pooled analysis			
Characteristics	OR (95% CI)	p	OR (95% CI)	p	OR (95% CI)	p	OR (95% CI)	p	I^2 %	p-het
Smoking										
Never	1.00 (Ref.)		1.00 (Ref.)		1.00 (Ref.)		1.00 (Ref.)			
Ever	0.55 (0.47-0.63)	<0.001	0.62 (0.41-0.93)	0.022	0.60 (0.42-0.86)	0.005	0.56 (0.49-0.64)	<0.001	0.0	0.778
Smoking										
Never	1.00 (Ref.)		1.00 (Ref.)		1.00 (Ref.)		1.00 (Ref.)			
Light [a]	0.69 (0.57-0.82)	<0.001	0.78 (0.48-1.26)	0.306	0.62 (0.42-0.93)	0.021	0.68 (0.59-0.80)	<0.001	0.0	0.793
Heavy [a]	0.41 (0.33-0.50)	<0.001	0.44 (0.25-0.78)	0.005	0.57 (0.34-0.94)	0.029	0.43 (0.36-0.51)	<0.001	0.0	0.492
	p for trend=	<0.001	p for trend=	0.005	p for trend=	0.008	p for trend=	<0.001		
rs660895										
AA	1.00 (Ref.)		1.00 (Ref.)		1.00 (Ref.)		1.00 (Ref.)			
AG	0.83 (0.72-0.97)	0.016	0.83 (0.57-1.21)	0.331	0.71 (0.52-0.98)	0.035	0.81 (0.71-0.92)	0.001	0.0	0.675
GG	0.78 (0.55-1.10)	0.155	0.53 (0.17-1.59)	0.255	0.19 (0.04-0.84)	0.028	0.56 (0.28-1.11)	0.098	43.5	0.170
Additive model	0.85 (0.75-0.96)	0.010	0.80 (0.58-1.10)	0.162	0.65 (0.49-0.87)	0.003	0.79 (0.68-0.93)	0.003	31.3	0.233
Dominant model	0.83 (0.71-0.95)	0.009	0.80 (0.56-1.15)	0.222	0.66 (0.49-0.91)	0.010	0.79 (0.70-0.90)	<0.001	0.0	0.463

[a] Definition for smoking: non-smoker, light smoker (<=18 pack-years), and heavy smoker (>18 pack-years).
All models adjusted for index age (continuous), sex, family history of PD, and matching factor year of birth in the PASIDA study. p -het, p for heterogeneity.

Table 3-3. Pooled analysis of the joint effects of smoking and rs660895 for PD

		OR (95% CI)	p	I^2	p-het	Interaction OR (95% CI)	p	I^2	p-het
Smoking	**rs660895 (codominant)**								
Never	AA	1.00 (Ref.)							
Never	AG	0.65 (0.54-0.78)	<0.001	0.0%	0.989				
Never	GG	0.60 (0.37-0.98)	0.043	6.2%	0.344				
Ever	AA	0.48 (0.41-0.57)	<0.001	0.0%	0.767	1.00 (Ref.)			
Ever	AG	0.49 (0.41-0.60)	<0.001	0.0%	0.409	1.59 (1.22-2.07)	0.001	0.0%	0.811
Ever	GG	0.41 (0.25-0.66)*	<0.001	NE	NE	1.33 (0.66-2.68)*	0.430	NE	NE
Smoking	**rs660895 (additive)**								
Never	AA	1.00 (Ref.)							
Never	AG	0.69 (0.60-0.81)	<0.001	0.0%	0.597				
Never	GG	0.48 (0.36-0.65)	<0.001	0.0%	0.597				
Ever	AA	0.50 (0.42-0.58)	<0.001	0.0%	0.743				
Ever	AG	0.47 (0.40-0.56)	<0.001	0.0%	0.913	1.00 (Ref.)			
Ever	GG	0.46 (0.35-0.61)	<0.001	0.0%	0.927	1.37 (1.10-1.71)	0.005	0.0%	0.992
Smoking	**rs660895 (dominant)**								
Never	AA	1.00 (Ref.)							
Never	AG+GG	0.64 (0.54-0.77)	<0.001	0.0%	0.887				
Ever	AA	0.48 (0.41-0.57)	<0.001	0.0%	0.768	1.00 (Ref.)			
Ever	AG+GG	0.48 (0.40-0.57)	<0.001	0.0%	0.752	1.54 (1.19-2.00)	0.001	0.0%	0.952
Smoking	**rs660895 (dominant)**								
Never	AA	1.00 (Ref.)							
Never	AG+GG	0.65 (0.54-0.77)	<0.001	0.0%	0.884				
Light [a]	AA	0.57 (0.47-0.69)	<0.001	0.0%	0.885	1.00 (Ref.)			
Light	AG+GG	0.61 (0.48-0.77)	<0.001	0.0%	0.736	1.64 (1.20-2.25)	0.002	0.0%	0.962
Heavy [a]	AA	0.38 (0.30-0.48)	<0.001	0.0%	0.560	1.00 (Ref.)			
Heavy	AG+GG	0.35 (0.27-0.47)	<0.001	0.0%	0.475	1.48 (1.02-2.13)	0.040	0.0%	0.645

[a] Definition for smoking: non-smoker, light smoker (≤18 pack-years), and heavy smoker (>18 pack-years).
* Only Denmark study provided information for estimation.
p -het, p for heterogeneity; NE, not estimable.

Table 3-4. Association of smoking with rs660895 among a) PD cases-only and b) non-PD controls-only, in each study and overall.

	Denmark (PASIDA)		France (*Terre*)		France (*Partage*)		Pooled analysis			
rs660895	OR (95% CI)	p	OR (95% CI)	p	OR (95% CI)	p	OR (95% CI)	p	I^2 %	p-het
Case-only analysis										
AA	1.00 (Ref.)		1.00 (Ref.)		1.00 (Ref.)		1.00 (Ref.)			
AG	1.37 (1.10-1.71)	0.005	0.93 (0.43-2.01)	0.857	1.71 (0.87-3.37)	0.121	1.36 (1.11-1.67)	0.003	0.0%	0.505
GG	1.20 (0.71-2.03)	0.488	NE	NE	NE	NE	1.20 (0.71-2.03)[a]	0.487	NE	NE
Additive model	1.25 (1.05-1.50)	0.014	0.85 (0.41-1.76)	0.652	1.50 (0.79-2.87)	0.217	1.24 (1.05-1.47)	0.012	0.0%	0.494
Dominant model	1.35 (1.09-1.67)	0.005	0.89 (0.41-1.91)	0.761	1.63 (0.83-3.20)	0.153	1.34 (1.10-1.62)	0.004	0.0%	0.483
Control-only analysis										
AA	1.00 (Ref.)		1.00 (Ref.)		1.00 (Ref.)		1.00 (Ref.)			
AG	0.92 (0.74-1.15)	0.482	0.67 (0.42-1.07)	0.092	0.83 (0.54-1.29)	0.412	0.86 (0.72-1.04)	0.114	0.0%	0.468
GG	0.82 (0.50-1.34)	0.432	0.94 (0.31-2.85)	0.918	0.65 (0.18-2.36)	0.516	0.82 (0.54-1.25)	0.353	0.0%	0.913
Additive model	0.92 (0.77-1.10)	0.336	0.78 (0.53-1.13)	0.191	0.83 (0.57-1.21)	0.320	0.88 (0.76-1.02)	0.089	0.0%	0.698
Dominant model	0.91 (0.74-1.13)	0.392	0.70 (0.45-1.09)	0.112	0.82 (0.53-1.25)	0.352	0.86 (0.72-1.02)	0.087	0.0%	0.553

ORs (95% CI) estimate the association between rs660895 and ever smoking (dependent variable) in cases and controls. All models adjusted for index age (continuous), sex, and matching factor year of birth in the PASIDA study. p -het., p for heterogeneity; NE, not estimable. Please see Table S1 for the cross-tabulation of ever smoking and rs660895 in cases and controls.

[a] Only the Danish study contributed information.

3.6 Supplement

Table 3-S1. Joint distribution of smoking and rs660895 by study

		Denmark (PASIDA)		France (*Terre*)		France (*Partage*)	
		Cases	Controls	Cases	Controls	Cases	Controls
Smoking	**rs660895**	N	N	N	N	N	N
Never	AA	514	321	111	227	178	262
Never	AG	227	220	35	111	53	126
Never	GG	32	31	4	11	2	14
Ever	AA	460	608	37	127	53	137
Ever	AG	283	369	17	46	21	56
Ever	GG	31	46	0	7	0	4
Never	AA	514	321	111	227	178	262
Never	AG	227	220	35	111	53	126
Never	GG	32	31	4	11	2	14
Light [a]	AA	236	263	24	68	35	87
Light	AG	161	156	8	17	13	35
Light	GG	12	22	0	1	0	3
Heavy [a]	AA	117	197	10	59	18	50
Heavy	AG	65	136	9	29	8	21
Heavy	GG	15	14	0	6	0	1

[a] Definition for smoking: non-smoker, light smoker (≤18 pack-years), and heavy smoker (>18 pack-years)

Table 3-S2. Joint effects of smoking and rs660895 for PD by study

		Denmark (PASIDA)		France (*Terre*)		France (*Partage*)	
		OR (95% CI)	Interaction OR (95% CI)	OR (95% CI)	Interaction OR (95% CI)	OR (95% CI)	Interaction OR (95% CI)
Smoking	**rs660895 (codominant)**		*global P≤.0001*		*global P=0.13*		*global P=0.005*
Never	AA	1.00 (Ref.)		1.00 (Ref.)		1.00 (Ref.)	
Never	AG	0.66 (0.52-0.83)		0.66 (0.42-1.04)		0.64 (0.44-0.93)	
Never	GG	0.66 (0.40-1.11)		0.74 (0.23-2.39)		0.21 (0.05-0.95)	
Ever	AA	0.46 (0.38-0.56)	1.00 (Ref.)	0.52 (0.32-0.84)	1.00 (Ref.)	0.54 (0.36-0.81)	1.00 (Ref.)
Ever	AG	0.47 (0.38-0.58)	1.55 (1.14-2.10)	0.70 (0.37-1.32)	2.04 (0.91-4.58)	0.52 (0.30-0.92)	1.52 (0.75-3.08)
Ever	GG	0.41 (0.25-0.66)	1.33 (0.66-2.68)	NE	NE	NE	NE
Smoking	**rs660895 (additive)**		*P=0.01*		*P=0.41*		*P=0.29*
Never	AA	1.00 (Ref.)		1.00 (Ref.)		1.00 (Ref.)	
Never	AG	0.72 (0.60-0.87)		0.72 (0.49-1.05)		0.59 (0.42-0.83)	
Never	GG	0.52 (0.36-0.76)		0.52 (0.24-1.11)		0.35 (0.18-0.69)	
Ever	AA	0.48 (0.40-0.57)		0.55 (0.34-0.89)		0.55 (0.37-0.82)	
Ever	AG	0.47 (0.39-0.57)		0.53 (0.30-0.96)		0.46 (0.27-0.79)	
Ever	GG	0.46 (0.34-0.63)	1.36 (1.06-1.75)	0.52 (0.18-1.51)	1.34 (0.67-2.70)	0.39 (0.14-1.08)	1.42 (0.74-2.73)
Smoking	**rs660895 (dominant)**		*P=0.005*		*P=0.17*		*P=0.24*
Never	AA	1.00 (Ref.)		1.00 (Ref.)		1.00 (Ref.)	
Never	AG+GG	0.66 (0.53-0.82)		0.67 (0.43-1.03)		0.59 (0.41-0.86)	
Ever	AA	0.46 (0.38-0.56)	1.00 (Ref.)	0.52 (0.32-0.83)	1.00 (Ref.)	0.54 (0.36-0.81)	1.00 (Ref.)
Ever	AG+GG	0.46 (0.38-0.57)	1.52 (1.13-2.04)	0.60 (0.32-1.12)	1.74 (0.79-3.84)	0.49 (0.28-0.85)	1.52 (0.76-3.06)
Smoking	**rs660895 (dominant)**		P_{light}=0.005		P_{light}=0.29		P_{light}=0.35
Never	AA	1.00 (Ref.)	P_{heavy}=0.16	1.00 (Ref.)	P_{heavy}=0.12	1.00 (Ref.)	P_{heavy}=0.39
Never	AG+GG	0.66 (0.53-0.83)		0.67 (0.43-1.02)		0.59 (0.41-0.86)	
Light [a]	AA	0.56 (0.45-0.71)		0.65 (0.38-1.13)		0.56 (0.35-0.89)	
Light	AG+GG	0.61 (0.48-0.79)	1.66 (1.16-2.37)	0.77 (0.31-1.89)	1.77 (0.62-5.08)	0.50 (0.25-0.98)	1.49 (0.65-3.42)
Heavy [a]	AA	0.36 (0.28-0.48)		0.32 (0.15-0.67)		0.50 (0.28-0.92)	
Heavy	AG+GG	0.32 (0.24-0.44)	1.36 (0.89-2.07)	0.50 (0.23-1.11)	2.35 (0.79-6.95)	0.47 (0.20-1.12)	1.59 (0.56-4.55)

[a] Definition for smoking: non-smoker, light smoker (<=18 pack-years), and heavy smoker (>18 pack-years).
All models adjusted for index age (continuous), sex, family history of PD, and matching factor year of birth in the PASIDA study. P-values are for the interaction terms. NE: not estimable.

4. Coffee Consumption is Associated with DNA Methylation Levels of Human Blood

4.1 Introduction

Coffee is one of the most widely consumed beverages in the world and is believed to have potential health risks and benefits(Butt et al. 2011). Coffee consumption has been linked to a wide range of health outcomes including cardiovascular, metabolic, and neurocognitive function. Heavy coffee consumption induces cardiovascular responses and insomnia(Butt et al. 2011), but coffee consumption has also been associated with lower risk of type 2 diabetes, endometrial cancer(Giri et al. 2011), and neurodegenerative diseases such as Parkinson's disease (PD) and Alzheimer's disease (AD)(Qi et al. 2014). Caffeine is thought to prevent cognitive decline by inhibiting formation of beta-amyloid and by acting as an anti-inflammatory agent in AD(Chen et al. 2001; Arendash et al. 2006), whereas in PD, it is thought to reduce neuroinflammation and lipid-mediated oxidative stress(Farooqui et al. 2011; Ferrari et al. 2011). AD and PD are slowly progressive diseases with a long prodromal phase, making it difficult to rule out reverse causality such that at risk individuals may decrease coffee intake due to development of sleep problems or loss of smell(Wirdefeldt et al. 2011).

Genomic studies identified eight genetic loci that have an influence on habitual coffee consumption, including some near *CYP1A2* and *AHR,* encoding the caffeine metabolizing enzyme Cytochrome P450 1A2 and a CYP1A2 regulator *Aryl Hydrocarbon Receptor,* respectively,(Cornelis et al. 2011; Cornelis et al. 2015). DNA methylation (DNAm) might act as a potential epigenetic mediator for caffeine's influence on health(Petronis 2010). Mechanistic epigenetic studies of caffeine have mainly focused on animal models(Buscariollo et al. 2014; Ping et al. 2014; Wu et al. 2015). For example, maternal prenatal caffeine intake increased methylation of the steroidogenic factor-1 promoter in fetal adrenal tissue in mice(Ping et al. 2014) whereas caffeine elicited effects similar to acute exercise in rat skeletal muscle tissue and resulted in lower DNA methylation levels in

promoter regions of energy metabolism genes(Barres et al. 2012). Little is known whether epigenetic changes can be found in human due to their coffee consumption habits. Exploring whether coffee consumption affects DNA methylation can help identify epigenetic signatures and provide mechanistic insights for results from past epidemiological studies and possibly new insights into health risks or benefits of coffee consumption.

Here, for the first time we identified DNA methylation sites from a genome-wide screen that relate to habitual coffee consumption in humans. We conducted a meta-analysis of DNA methylation levels in blood samples from two different data sets: PD-free control subjects enrolled in the Parkinson's Environment and Genes (PEG first round, 2001-2007(Kang et al. 2005; Narayan et al. 2013)) study consisting of 215 non-Hispanic Caucasians, and women from the WHI consisting of 995 Caucasians, 431 Hispanics, and 674 African Americans. We also related coffee consumption to DNA methylation levels in saliva samples from 127 PD patients and 129 PD-free controls (age-, gender-, and ethnicity-matched) enrolled in the second round of the PEG study (2009-ongoing)(Costello et al. 2009; Narayan et al. 2015). Detailed information for each dataset can be found in **Table S1** and in **Methods**.

4.2 Materials and Methods

Description of PEG1 subjects

Study population: This dataset consists of 215 Caucasian population controls with complete information on coffee consumption and blood samples for DNA. The PEG1 study is a population-based case control study in central California (Fresno, Kern, or Tulare Counties) recruiting subjects from 2001 to 2007. To be eligible, participants had to be residents of one of three central California counties, had to have lived in California for at least 5 years, and to be at least 35 years of age(Kang et al. 2005). Population controls were identified from

Medicare lists and also using residential property tax assessor records. Potential controls were screened for eligibility by mail or telephone, and only 1 person per household was allowed to enroll(Costello et al. 2009; Narayan et al. 2015). The study was approved by the UCLA Institutional Review Board, and informed consent was obtained from all subjects.

Exposure assessment: Standardized interviews were conducted to obtain information on demographics, lifetime caffeinated beverage consumption, smoking, and MHT histories. In the interview, information on the frequency and amount of caffeinated beverage consumption at different periods of lifetime were collected: young adult <25 years, adult 25-44 years, middle-aged 45-64 years, and senior ≥65y years. We used this information to calculate weighted average daily coffee consumption. Only caffeinated coffee consumed during the past 12 months prior to the date of blood draw contributed to our exposure measures.

Description of PEG2 subjects

Study population: This dataset consists of data from 127 PD patients and 129 population controls with complete information on coffee consumption and saliva samples we extracted DNA from recruited for the PEG2 study which started in 2009 (ongoing). PD patients were identified using the California PD Registry for the three target counties in central California. Those who lived in the study area were eligible and were mailed invitations and those who agreed were examined by a UCLA movement disorder specialist who applied UK Brain Bank and Gelb diagnostic criteria(Hughes et al. 1992; Gelb et al. 1999). Controls selection was based on the same criteria as in the PEG1 study but only used tax assessor records to identify residents whom we recruited at the door step. Saliva samples selected from PEG2 participants were matched on age, gender, and race for cases and controls.

Exposure assessment: Methods for assessing exposure were identical to the methods employed in the PEG1 study.

Phenotype data and DNA methylation data of the PEG studies are available at GEO accession database GSE72775 (blood) and GSE78874 (saliva).

WHI subjects Description

Study population: This dataset consists of a sub-group of 2,100 women (995 Caucasians, 431 Hispanics, and 674 African Americans) with complete information on coffee consumption as well as genome-wide DNA methylation data from blood drawn at baseline. The WHI is a multi-center study launched in 1993 which enrolled post-menopausal women aged 50-79 years into either one or more randomized Clinical Trials (RCTs) or an observational study(Anonymous 1998). These women were originally selected from two WHI subcohorts for a nested genomic case-control study of coronary heart disease (CHD) with genome-wide genotype and cardiovascular disease related biomarker data.(Curb et al. 2003) Thus, fifty percent (n=1,053) of these WHI women were eventually diagnosed with CHD; however, disease status has no effect on DNA methylation level measured at baseline. The two cohorts are: 1) the WHI SNP Health Association Resource (SHARe) cohort which includes genotyping data from ~8500 African American and ~3500 Hispanic women through WHI core study M5-SHARe (www.whi.org/researchers/data/WHIStudies/StudySites/M5) as well as information on biomarker through WHI Core study W54-SHARe (...data/WHIStudies/StudySites/W54); 2) the two European Americans Hormonal Therapy (EA HT) trials selected for GWAS and biomarkers in core studies W58 (.../data/WHIStudies/StudySites/W58) and W63 (.../data/WHIStudies/StudySites/W63).

Exposure assessment: Information on demographics, smoking history, and MHT were obtained using a structured questionnaire at baseline. Food frequency questionnaires were used to collect information on daily coffee or tea (all types) consumption in the past 3 months

prior to baseline. Our exposure measures were directly taken from answers provided in response to the questionnaire.

DNA extraction and Genome-wide DNA methylation analysis

DNA methylation data were obtained from the Infinium HumanMethylation450 BeadChip (Illumina, San Diego, CA) using DNA samples extracted from peripheral blood cells and leukocytes in saliva. Methylation beta-values ranging from 0 (unmethylated) to 1 (fully methylated) were used for analysis.(Dunning et al. 2008)

Statistical analyses

The raw methylation data was preprocessed using the background normalization method from the Genome Studio software. To assess correlations between continuous coffee consumption (cup/day) and site-specific DNA methylation levels, biweight midcorrelation (bicor) was applied in a genome-wide screen. In the main correlation analysis using DNA methylation levels from blood, potential confounders such as age at blood draw, gender, and blood cell counts were adjusted for by regressing out the effects of these factors and retaining the residuals. Smoking status (ever vs. never) was further adjusted for in ancillary analyses. We used the Houseman algorithm in the *minfi* R package and epigenetic clock software for estimating blood cell counts(Houseman et al. 2012; Horvath 2013; Jaffe et al. 2014). All blood analyses were stratified by ethnicity, thus four subsets were generated: PEG1 Caucasians PD-free controls, WHI Caucasians, WHI Hispanics, and WHI African Americans. In order to obtain an overall p-value across the four subsets, we conducted a meta-analysis using Stouffer's method for combining Z-values (meta.Z) i.e. $\Sigma z_i/\text{sqrt}(4)$. The corresponding two-sided p-values (meta.pvalue) were calculated under the assumption of a normal distribution. These approaches were also applied to identify smoking-associated CpGs and CpGs influenced by both coffee and smoking. We then identified the top-ranked coffee-

associated CpGs by meta.pvalue, and applied functional enrichment analysis on 2,124 genes identified from the top 3,000 most significant coffee-associated CpGs (meta.pvalue threshold ~ $1.1x10^{-3}$) using the online bioinformatics tool – the Database for Annotation, Visualization and Integrated Discovery (DAVID v.6.7). We further conducted MHT-stratified meta-analysis for the top 11 coffee-associated CpGs using the WHI data in order to investigate the modifying effect MHT has on the coffee-DNAm association. In the analysis using DNA methylation levels from saliva in PEG2, potential confounders such as age at saliva collection, gender, and ethnicity were adjusted for as above. Analyses and scatter plots were created using the *WGCNA* package in R v.3.1.2, while Manhattan plots of epigenome-wide association study (EWAS) p-values were generated with the *qqman* package. QQ-plots of EWAS pvalues were also generated in R, and lambda i.e. median(chi^2)/0.454, were calculated to identify potential inflation.

4.3 Results and Discussion

Coffee consumption and DNA methylation levels in blood

In our EWAS study, we analyzed methylation levels of approximately 486k CpGs on the Illumina 450K array. Since many CpGs exhibit strong pairwise correlations, the Bonferroni-corrected significance threshold of $\alpha=0.05/500{,}000=1x10^{-7}$ was considered overly conservative; we used a modified threshold of $p<5x10^{-6}$ to evaluate genome-wide significance in our study. In the PEG1 and WHI datasets, adjusting for chronological age, gender, and imputed blood cell counts, we identified one CpG with genome-wide Bonferroni corrected significance: cg21566642 near the *ALPPL2* gene (meta.$p=3.7x10^{-10}$). Ten additional CpGs surpassed the significance threshold of $p<5.0x10^{-6}$ (**Table 4-1a** and **Figure 4-1a**) and are located in/near the genes *GPR132, BSCL2, MALRD1, GRK5, PSMD8, FSTL5, PTHLH,* etc. (**Table 4-1a**).

The top ranked CpGs appear to be linked to genes involved in lipid metabolism (Miranda et al. 2009; Vangaveti et al. 2010; Vergnes et al. 2013; Stringhini et al. 2015) and immune response [RefSeq, Jul 2008]. For instance, the protein encoded by *GPR132* is a receptor for oxidized free fatty acids and is a treatment target for diabetes because of its role in lipid metabolism and antioxidant activity(Vangaveti et al. 2010). In a mixed-race study of atherosclerosis, *GPR132* was found to be hypomethylated among low socioeconomic status (SES) individuals with increased inflammatory activity compared with high SES individuals(Stringhini et al. 2015). *BSCL2* encodes the transmembrane protein "seipin" residing in the endoplasmic reticulum. Variants in *BSCL2* cause congenital generalized lipodystrophy, characterized by the loss of adipose tissue and severe insulin resistance(Miranda et al. 2009). *MALRD1* encodes yet another lipid-related gene that has been shown to regulate bile acid and lipid levels in the enterohepatic system(Vergnes et al. 2013). Genes related to immune response include *GRK5* and *PSMD8*. The protein encoded by *GRK5* regulates polymorphonuclear leukocyte motility, while *PSMD8* encodes an immune-proteasome component related to major histocompatibility (MHC) class I antigen processing and presentation [provided by RefSeq, Jul 2008]. These coffee-associated CpGs were mostly located within 200 to 1,500 bps upstream of a transcription start site of a gene i.e. promoter region (Fisher's p=0.03, **Figure 4-1b**). Additional stratified analyses for the PEG and WHI samples, as well as comparisons of effect sizes between genders or ethnicities are provided in **Table 4-S2, 4-S3, and Table 4-1**.

Many studies have focused on associations between DNA methylation and smoking, and it is well known that a subgroup of coffee consumers is more likely to smoke. The highest correlation we observed between coffee intake and smoking in any of our cohorts was r=0.31 among PEG1 Caucasian controls ($p=3.4x10^{-6}$). Due to the common co-exposure to coffee and smoking, any adjustment for smoking is expected to affect associations between

DNA methylation and coffee consumption. Indeed, smoking adjustment reduced the statistical significance of cg21566642 (meta.p=5.4x10^{-4}, **Table 4-2 and Figure 4-2a**) to less than the genome-wide threshold. However, associations between the 11 top-ranked CpGs and coffee consumption were still preserved after smoking adjustment (meta.p<0.05/11=4.5x10^{-3}, **Table 4-2**). Further, we identified methylation differences for 135 CpGs associated with both coffee drinking and smoking (meta.p≤1.0x10^{-7}, **Table 4-S4**). After smoking adjustment, the most significant differentially methylated genes were *BSCL2* and GPR132, along with *CNTN4* and *ROBO3* which appear to be involved in axonal navigation. (meta.p ≤5x10^{-6}, **Table 4-1b**).

It is worth noting that genetic variants previously linked to coffee consumption in GWAS did not reach the significance threshold of p≤1.0x10^{-3} in our study, specifically *AHR*, *CYP1A1*, *CYP1A2*, *NRCAM*, and *ADORA2A*(Cornelis et al. 2011; Hamza et al. 2011; Amin et al. 2012; Byrne et al. 2012; Cornelis et al. 2015). However, our study corroborated the importance of the *STK11* gene (cg24145685: meta.Z=4.54, meta.p=5.7x10^{-6}; after smoking adjustment: meta.Z=3.99 and meta.p=6.6x10^{-5}), which encodes a member of the serine/threonine kinase family and interacts with another gene (*CAB39L*) identified in a previous GWAS focused on coffee consumption(Amin et al. 2012).

As mentioned above, previous studies reported reduced risks of developing PD and AD with habitual coffee consumption(Hernan et al. 2002; Barranco Quintana et al. 2007). To the best of our knowledge, our blood tissue data did not include AD or PD patients. Surprisingly, we found some CpGs located near genes linked with familial forms of PD associated with coffee consumption: *GBA* (meta.p=7.9x10^{-5}), *PARK2/Parkin* (meta.p=7.3x10^{-4}), and *PINK1* (meta.p=8.9x10^{-4}). Similarly, some GWAS-identified loci for AD were also associated with coffee intake: *PICALM* (meta.p=1.3x10^{-5}), *CLU* (meta.p=6x10^{-4}), and *EDC3* (meta.p=1.1x10^{-4})(Harold et al. 2009; Victor Junji Yamamoto 2015). The *PARK2* gene

encodes an ubiquitin protein ligase called Parkin that targets proteins for degradation in the proteasome. Pathways related to Parkin include oxidative stress, Class I MHC antigen processing and presentation and alpha-synuclein signaling(Andersen 2004). The *EDC3* gene is also of interest since it is located near *CYP1A1/CYP1A2* and the enzymes encoded by them may interact with coffee consumption in reducing AD risk.(Victor Junji Yamamoto 2015)

Using the 2,104 genes linked with the 3,000 most significant coffee-associated CpGs ($p \leq 1.1 \times 10^{-3}$ without smoking adjustment) in gene set enrichment analysis, we identified 2,901 CpGs (in/near 2,058 genes) that were hyper-methylated in habitual coffee drinkers, while only 3% (99 CpGs in/near 66 genes) were hypo-methylated. Results of the DAVID functional analysis showed that these coffee-associated genes are enriched in functional categories of transcription factor binding ($p=1.2 \times 10^{-6}$, **Table 4-3**) and protein kinase activity ($p=2.9 \times 10^{-5}$). The enriched biological terms remained statistically significant after correcting for multiple comparisons (Benjamini-adjusted $p < 0.05$).

It has previously been suggested that the potential protective action of coffee on PD in women may be abrogated by postmenopausal estrogen use (Ascherio et al. 2004; Palacios et al. 2012). Interestingly, when we stratified the participants from the WHI by MHT, we observed significant associations between coffee consumption and the 11 top CpG sites from **Table 4-1a** only in women who never used MHT and not in MHT users (**Table 4-4**).

Coffee consumption and DNA methylation levels in saliva

We also evaluated associations between coffee and genome-wide DNA methylation levels in saliva provided by 256 participants with and without PD enrolled in the PEG2 study. After adjustment for chronologic age, gender and ethnicity, no CpGs achieved genome-wide significance ($p \leq 10^{-7}$, **Figure 4-3a**). When examining the 11 most significant coffee-associated CpGs we previously identified in blood, none of the significant associations were preserved in saliva (**Table 4-2**). Moreover, we did not observe positive correlations between

meta-Z values for blood and Z values for saliva (**Figure 4-3b**). Further adjustment for PD status did not change these results (**Table 4-2**), suggesting that PD status did not affect DNA methylation levels in saliva for the coffee-related CpGs identified in blood. After adjusting for smoking, there appeared to be a significant correlation for coffee-related DNA methylation in blood and in saliva tissues, but in the opposite direction of what we expected (**Figure 4-4b and Table 4-S5**). This can be explained by the 'regression to the mean' effect, suggesting DNA methylation data from saliva did not replicate associations we found in blood.

Potential limitations

Our study has some potential limitations: First, we have a small amount of uncertainty regarding the reported ethnicity since we do not have genome-wide SNP data to compute the genetic principal components. However, ethnicity information in the studies we included have been carefully verified using 37 Ancestry Informative Markers (AIM). Moreover, we stratified by ethnicity to address the possibility of ethnic confounding. Second, stratification by gender in the PEG1 study removed the significance of some of the associations between coffee consumption and methylation loci. However, this might be due to the decreased sample size imparted by stratification. Third, similar to other DNA methylation studies, lambda values in this study were inflated; therefore, we should interpret p-values with caution. However, our findings in Caucasians were replicated in other ethnic groups giving some validation to the results. In addition, it remains debatable whether to present QQ-plots in DNA methylation studies, since CpGs are highly correlated and the distributional assumptions made in GWAS may not be met in EWAS (see **Figure 4-S1**).

4.4 Conclusions

In summary, in peripheral blood mononuclear cells we identified CpGs located near 11 genes that were associated with habitual coffee consumption based on the significance

threshold (meta.p≤$5.0x10^{-6}$) while adjusting for age, gender, and blood cell composition. Moreover, these correlations remained significant after further adjustment for smoking. Furthermore, many differentially methylated CpGs are located in/near genes reported to be associated with coffee-related chronic diseases or the common neurodegenerative diseases PD and AD for which coffee consumption has been suggested to be protective. Our results point to possible mechanisms through which coffee consumption may have beneficial effects and possibly may confer risk reduction. The measures of habitual coffee consumption we used in this study were based on recall over a short period (last 3 months in WHI or 12 months in PEG1); however, in PEG1 reported lifetime coffee consumption was highly correlated with coffee consumption reported for the past 12 months (cor=0.87), suggesting that coffee consumption is consistent across time. Moreover, this is a mixed-race and mixed-gender study, therefore the coffee associations with DNA methylation levels in blood appear to extend to both genders and different ethnic groups, even though in women, results also seemed to depend on MHT use. Lastly, our study suggests that while coffee affects DNA methylation levels in blood this does not seem to extend to saliva-derived tissue.

4.5 Tables and Figures

Table 4-1. The top-ranked CpG sites associated with coffee consumption in blood with/without smoking adjustment

List of CpGs associated with coffee consumption, (in/near) gene, chromosome, and CpG island location, gene region, (Stouffer's test) Z-value and p-value from meta-analysis, a robust correlation coefficient (known as biweight midcorrelation) and p-value for daily coffee consumption (in last 3 months in WHI or 12 months in PEG1) and DNA methylation levels within 4 population subsets.

									PEG1 Caucasian Ctrl only, adjusted for age, gender, cell counts (Subset1)		WHI Caucasian-only, adjusted for age, cell counts (Subset2)		WHI Hispanic-only, adjusted for age, cell counts (Subset3)		WHI African American-only, adjusted for age, cell counts (Subset4)	
	CpG	Gene	Chr.	Position (bp)[a]	Relation to UCSC CpG Island	UCSC RefGene Group[b]	meta. Z	meta. Pvalue	biCor	P-value	biCor	P-value	biCor	P-value	biCor	P-value
	(a) Adjusting for chronological age, gender and blood cell counts															
1	cg21566642	intergenic, near ALPPL2	2	233284661	Island		-6.26	3.73E-10	-0.29	3.60E-05	-0.14	1.34E-05	-0.07	1.28E-01	-0.10	1.33E-02
2	cg20333292	GPR132	14	105532012		TSS1500	5.05	4.33E-07	0.28	7.43E-05	0.05	8.67E-02	0.13	6.87E-03	0.07	9.10E-02
3	cg21163128	BSCL2	11	62477362	Island	TSS1500	5.00	5.70E-07	0.28	6.77E-05	0.10	2.40E-03	0.09	6.23E-02	0.04	2.80E-01
4	cg23303782	GRK5	10	120967744	S_Shore	Body	4.81	1.49E-06	0.29	3.55E-05	0.06	4.37E-02	0.09	6.39E-02	0.06	1.14E-01
5	cg26105150	FSTL5	4	163085403		TSS1500	4.78	1.77E-06	0.27	1.35E-04	0.05	8.98E-02	0.16	6.78E-04	0.02	5.45E-01
6	cg19723563	PTHLH	12	28123034	Island	TSS200	4.74	2.13E-06	0.17	2.14E-02	0.07	2.20E-02	0.13	5.22E-03	0.08	3.77E-02
7	cg17928869	intergenic, between EIF1 and KRT42P	17	39822542	S_Shore		4.68	2.86E-06	0.23	1.35E-03	0.08	1.17E-02	0.11	2.11E-02	0.05	1.93E-01
8	cg12866551	MALRD1	10	20019641			4.66	3.12E-06	0.23	1.48E-03	0.07	2.04E-02	0.14	2.75E-03	0.03	4.21E-01
9	cg15722372	PSMD8	19	38865020	N_Shore	TSS200	4.65	3.26E-06	0.20	4.55E-03	0.05	9.92E-02	0.10	4.67E-02	0.11	4.94E-03
10	cg19974428	TMEM130	7	98468047	Island	TSS1500	4.65	3.34E-06	0.22	2.29E-03	0.07	3.67E-02	0.11	2.64E-02	0.07	5.55E-02
11	cg15140902	FLJ22536	6	21667815	S_Shore	Body	4.60	4.14E-06	0.12	1.09E-01	0.10	1.52E-03	0.10	4.12E-02	0.09	1.72E-02
	(b) Further adjusting for smoking															
1	cg21163128	BSCL2	11	62477362	Island	TSS1500	4.95	7.36E-07	0.30	1.87E-05	0.09	7.19E-03	0.09	6.96E-02	0.04	2.82E-01

2	cg08119527	Intergenic, near PODXL	7	131340667			4.84	1.28E-06	0.30	2.81E-05	0.09	4.68E-03	0.09	6.27E-02	0.03	4.44E-01
3	cg26331135	CNTN4	3	2144068	S_Shelf	5'UTR	4.71	2.46E-06	0.31	1.31E-05	0.04	1.92E-01	0.13	7.96E-03	0.04	2.91E-01
4	cg20333292	GPR132	14	105532012		TSS1500	4.62	3.85E-06	0.24	7.22E-04	0.06	8.21E-02	0.12	1.36E-02	0.06	1.04E-01
5	cg08311403	ROBO3	11	124735215	Island	TSS200	4.57	4.98E-06	0.16	3.04E-02	0.07	3.16E-02	0.06	2.48E-01	0.14	2.67E-04

Abbreviations: ALPPL2 (Alkaline Phosphatase, Placental-Like 2), GPR132 (G Protein-Coupled Receptor 132), BSCL2 (Berardinelli-Seip Congenital Lipodystrophy 2 (Seipin)), GRK5 (G Protein-Coupled Receptor Kinase 5), FSTL5 (Follistatin-Like 5), PTHLH (Parathyroid Hormone-Like Hormone), EIF1 (Eukaryotic Translation Initiation Factor 1), KRT42P(Keratin 42 Pseudogene), MALRD1 (MAM And LDL Receptor Class A Domain Containing 1), PSMD8 (Proteasome (Prosome, Macropain) 26S Subunit, Non-ATPase, 8), TMEM130 (Transmembrane Protein 130), FLJ22536 (miscRNA), BSCL2 (Berardinelli-Seip Congenital Lipodystrophy 2 (Seipin)), PODXL (Podocalyxin-Like), CNTN4, GPR132 (G Protein-Coupled Receptor 132), ROBO3 (Roundabout guidance receptor 3).

[a]Location is based on NCBI genome build 37.

[b]TSS: transcription start site, TSS500: within 1500 bps of a TSS, TSS200: within 200 bps of a TSS, UTR: untranslated region.

Table 4-2. Smoking adjusted results and saliva results for the 11 top-ranked CpG sites in Table 4-1a

List of CpGs associated with coffee consumption, (in/near) gene, chromosome, and CpG island location, gene region, (Stouffer's test) Z-value and p-value from meta-analysis for the PEG1 and WHI studies, a robust correlation coefficient (known as biweight midcorrelation) Z-value and p-value for daily coffee consumption (in last 3 months in WHI or 12 months in PEG) and DNA methylation levels.

							Blood Samples		Saliva Samples					
							PEG1 and WHI adjusted for age, gender, cell counts, smoking (N=2,297)		PEG2 adjusted for age, gender and race (N=256)			PEG2 adjusted for age, gender, race and PD status (N=256)		
	CpG	Gene	Chr.	Position (bp)[a]	Relation to UCSC CpG Island	UCSC RefGene Group[b]	biCor	meta.Pvalue	biCor	Z-value	P-value	biCor	Z-value	P-value
1	cg21566642	intergenic, near ALPPL2	2	233284661	Island		-3.46	5.37E-04	-0.10	-1.65	9.87E-02	-0.09	-1.37	1.71E-01
2	cg20333292	GPR132	14	105532012		TSS1500	4.62	3.85E-06	0.06	0.90	3.71E-01	0.03	0.46	6.46E-01
3	cg21163128	BSCL2	11	62477362	Island	TSS1500	4.95	7.36E-07	-0.01	-0.11	9.14E-01	-0.03	-0.47	6.42E-01
4	cg23303782	GRK5	10	120967744	S_Shore	Body	4.27	1.92E-05	0.01	0.14	8.89E-01	0.04	0.58	5.62E-01
5	cg26105150	FSTL5	4	163085403		TSS1500	4.34	1.42E-05	0.02	0.33	7.44E-01	-0.01	-0.21	8.32E-01
6	cg19723563	PTHLH	12	28123034	Island	TSS200	4.22	2.43E-05	0.03	0.42	6.76E-01	0.00	-0.02	9.80E-01
7	cg17928869	intergenic, between EIF1 and KRT42P	17	39822542	S_Shore		4.06	4.94E-05	0.04	0.60	5.48E-01	0.02	0.29	7.73E-01
8	cg12866551	MALRD1	10	20019641			4.47	7.78E-06	-0.06	-0.94	3.48E-01	-0.06	-0.90	3.68E-01
9	cg15722372	PSMD8	19	38865020	N_Shore	TSS200	4.44	9.01E-06	-0.01	-0.10	9.17E-01	-0.02	-0.36	7.17E-01
10	cg19974428	TMEM130	7	98468047	Island	TSS1500	4.04	5.42E-05	-0.03	-0.49	6.23E-01	-0.05	-0.86	3.92E-01
11	cg15140902	FLJ22536	6	21667815	S_Shore	Body	4.04	5.39E-05	0.00	0.05	9.61E-01	-0.02	-0.25	8.03E-01

Abbreviations: ALPPL2 (Alkaline Phosphatase, Placental-Like 2), GPR132 (G Protein-Coupled Receptor 132), BSCL2 (Berardinelli-Seip Congenital Lipodystrophy 2 (Seipin)), GRK5 (G Protein-Coupled Receptor Kinase 5), FSTL5 (Follistatin-Like 5), PTHLH (Parathyroid Hormone-Like Hormone), EIF1 (Eukaryotic Translation Initiation Factor 1), KRT42P(Keratin 42 Pseudogene), MALRD1 (MAM And LDL Receptor Class A Domain Containing 1), PSMD8 (Proteasome (Prosome, Macropain) 26S Subunit, Non-ATPase, 8), TMEM130 (Transmembrane Protein 130), FLJ22536 (miscRNA).

[a]Location is based on NCBI genome build 37. [b]TSS: transcription start site, TSS500: within 1500 bps of a TSS, TSS200: within 200 bps of a TSS.

Table 4-3. Functional enrichment analysis for top 3,000 most coffee-associated CpG sites in 2,124 genes (meta.pvalue cutoff ~ $1.1x10^{-3}$)

List of functional categories that coffee-methylation related genes are enriched in, p-values, Bonferroni-corrected or Benjamini adjusted p-value, FDR, number of overlapping genes, and fold enrichment.

Rank	Category	Term	p-value	Bonferroni	Benjamini	FDR	Overlap Genes (n)	Fold Enrichment
1	GOTERM_MF_FAT	GO:0008134~transcription factor binding	1.24E-06	1.51E-03	1.51E-03	2.01E-03	92	1.65
2	GOTERM_MF_FAT	GO:0003713~transcription coactivator activity	2.06E-05	2.48E-02	3.59E-03	3.34E-02	45	1.94
3	GOTERM_MF_FAT	GO:0003712~transcription cofactor activity	3.06E-05	3.66E-02	4.13E-03	4.95E-02	66	1.67
4	GOTERM_MF_FAT	GO:0004672~protein kinase activity	2.88E-05	3.45E-02	4.37E-03	4.66E-02	99	1.50
5	GOTERM_MF_FAT	GO:0004674~protein serine/threonine kinase activity	6.33E-05	7.42E-02	6.99E-03	1.02E-01	74	1.58

Table 4-4. Stratified analysis of the 11 coffee-associated CpGs found in blood (in Table 4-1a) by MHT using WHI data only, adjusted for chronological age and blood cell counts

List of the 11 coffee-associated CpGs from Table 1a, (in/near) gene, chromosome, and CpG island location, (Stouffer's test) Z-value and p-value from meta-analysis using WHI data which consist of 3 subsets: Caucasians, Hispanics, and African Americans, stratified by MHT.

							WHI all ethnicities, adjusted for age and cell counts (N=2,100)		WHI all ethnicities MHT user, adjusted for age and cell counts (N=1,031)[c]		WHI all ethnicities MHT non-user, adjusted for age and cell counts (N=1,012)[c]	
	CpG	Gene	Chr.	Position(bp)[a]	Relation to UCSC CpG Island	UCSC RefGene Group[b]	meta.Z	meta.Pvalue	meta.Z	meta.Pvalue	meta.Z	meta.Pvalue
1	cg21566642	intergenic, near ALPPL2	2	233284661	Island		-3.47	5.23E-04	-2.15	3.19E-02	-4.48	7.43E-06
2	cg20333292	GPR132	14	105532012		TSS1500	4.49	7.23E-06	2.69	7.24E-03	4.09	4.38E-05
3	cg21163128	BSCL2	11	62477362	Island	TSS1500	4.34	1.41E-05	2.97	3.02E-03	3.51	4.44E-04
4	cg23303782	GRK5	10	120967744	S_Shore	Body	3.90	9.53E-05	2.22	2.65E-02	3.20	1.37E-03
5	cg26105150	FSTL5	4	163085403		TSS1500	4.43	9.57E-06	2.78	5.47E-03	3.24	1.18E-03
6	cg19723563	PTHLH	12	28123034	Island	TSS200	4.03	5.48E-05	2.27	2.34E-02	3.35	8.22E-04
7	cg17928869	intergenic, between EIF1 and KRT42P	17	39822542	S_Shore		3.82	1.33E-04	1.76	7.88E-02	3.81	1.37E-04
8	cg12866551	MALRD1	10	20019641			4.47	7.97E-06	2.84	4.47E-03	2.72	6.58E-03
9	cg15722372	PSMD8	19	38865020	N_Shore	TSS200	2.72	6.47E-03	1.60	1.11E-01	2.58	9.87E-03
10	cg19974428	TMEM130	7	98468047	Island	TSS1500	3.55	3.80E-04	2.01	4.41E-02	3.06	2.18E-03
11	cg15140902	FLJ22536	6	21667815	S_Shore	Body	4.39	1.15E-05	2.46	1.40E-02	4.61	3.99E-06

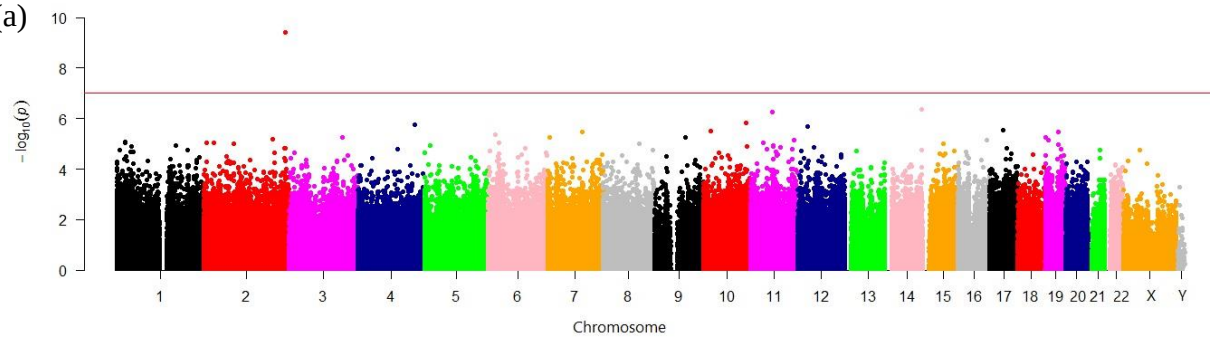
Abbreviations: ALPPL2 (Alkaline Phosphatase, Placental-Like 2), GPR132 (G Protein-Coupled Receptor 132), BSCL2 (Berardinelli-Seip Congenital Lipodystrophy 2 (Seipin)), GRK5 (G Protein-Coupled Receptor Kinase 5), FSTL5 (Follistatin-Like 5), PTHLH (Parathyroid Hormone-Like Hormone), EIF1 (Eukaryotic Translation Initiation Factor 1), KRT42P(Keratin 42 Pseudogene), MALRD1 (MAM And LDL Receptor Class A Domain Containing 1), PSMD8 (Proteasome (Prosome, Macropain) 26S Subunit, Non-ATPase, 8), TMEM130 (Transmembrane Protein 130), FLJ22536 (miscRNA).

[a]Location is based on NCBI genome build 37.

[b]TSS: transcription start site, TSS500: within 1500 bps of a TSS, TSS200: within 200 bps of a TSS.

[c]Missing MHT information: WHI Caucasian (N=35), Hispanic (N=11), African American (N=11)

(a)



(b)

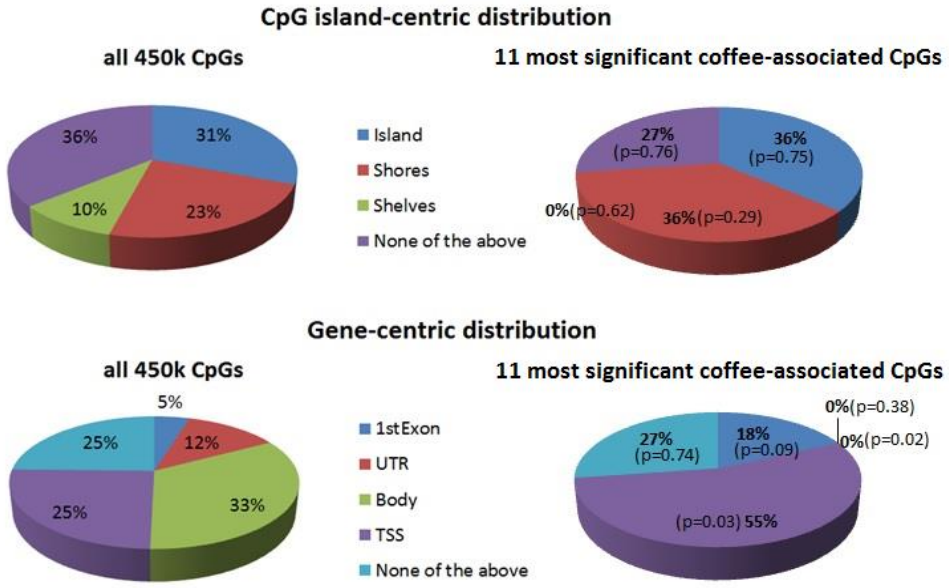
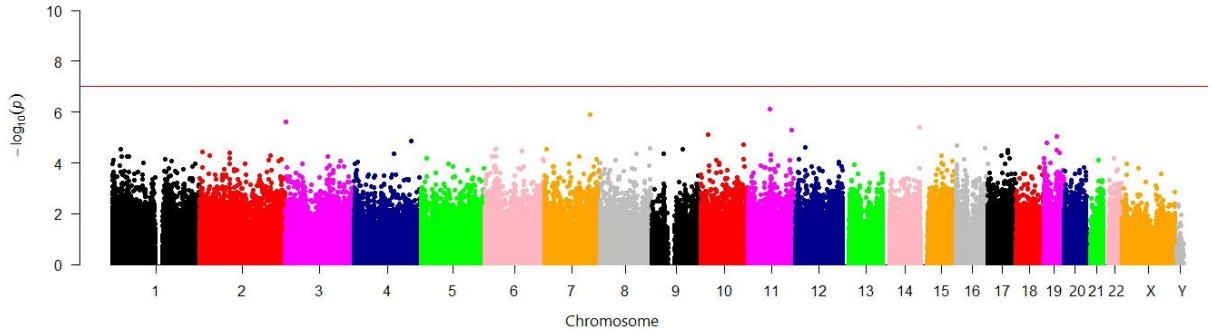


Figure 4-1. Blood DNA methylation levels associated with coffee consumption adjusted for age, gender and blood cell counts. (a) Manhattan plot of the meta-analysis methylation association p-values adjusted for chronological age, gender and blood cell counts. The red line indicates p-value threshold of 10^{-7}. One CpG on chromosome 2 passed this threshold. The x-axis corresponds to negative $\log_{10}$ transformed meta.p-value. The x-axis refers to chromosome number, and X and Y chromosomes. (b) Distributions of CpGs relative to CpG island and gene regions for all 450k CpGs on the microarray and the 11 most significant coffee-associated CpGs listed in **Table 4-1a**. P-values were obtained by Fisher's test for comparing proportions.

(a)



(b)

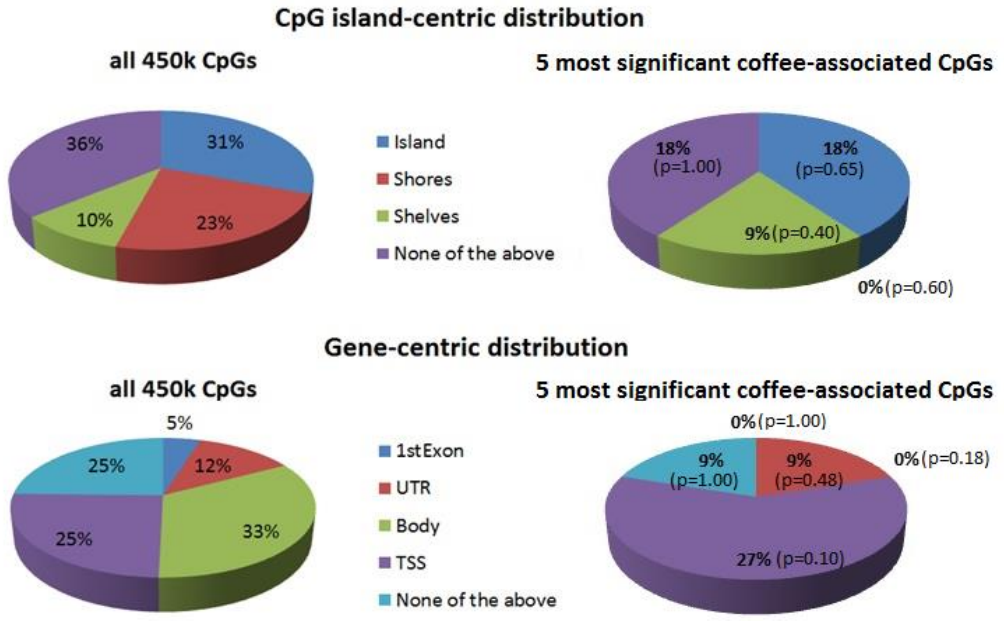
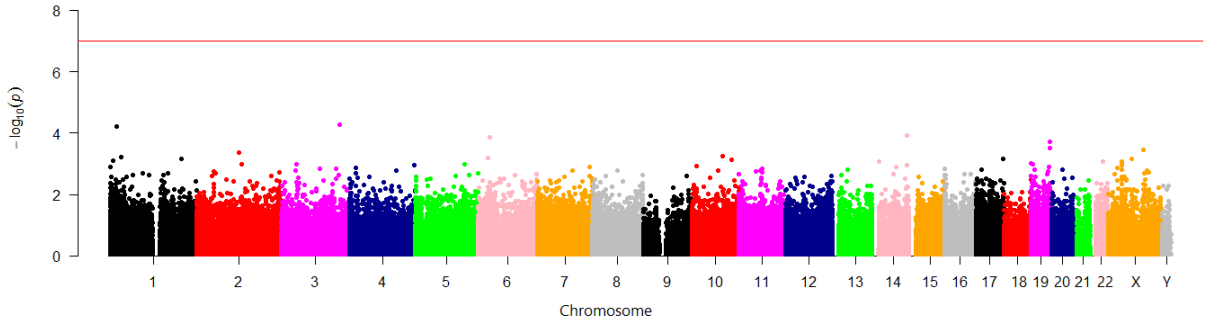


Figure 4-2. Blood DNA methylation levels associated with coffee consumption adjusted for age, gender, blood cell counts and smoking. (a) Manhattan plot of the meta-analysis methylation association p-values adjusted for chronological age, gender, blood cell counts, and smoking. Red line indicates p-value threshold of 10^{-7}- no CpG passed this threshold. X-axis corresponds to negative $\log_{10}$ transformed meta.p-value. Y-axis refers to chromosome number, X and Y chromosomes. (b) Distributions of CpGs relative to CpG island and gene regions for all 450k CpGs on the microarray and the 5 most significant coffee-associated CpGs listed in **Table 4-1b**. P-values were obtained by Fisher's test for comparing proportions.

(a)



(b)

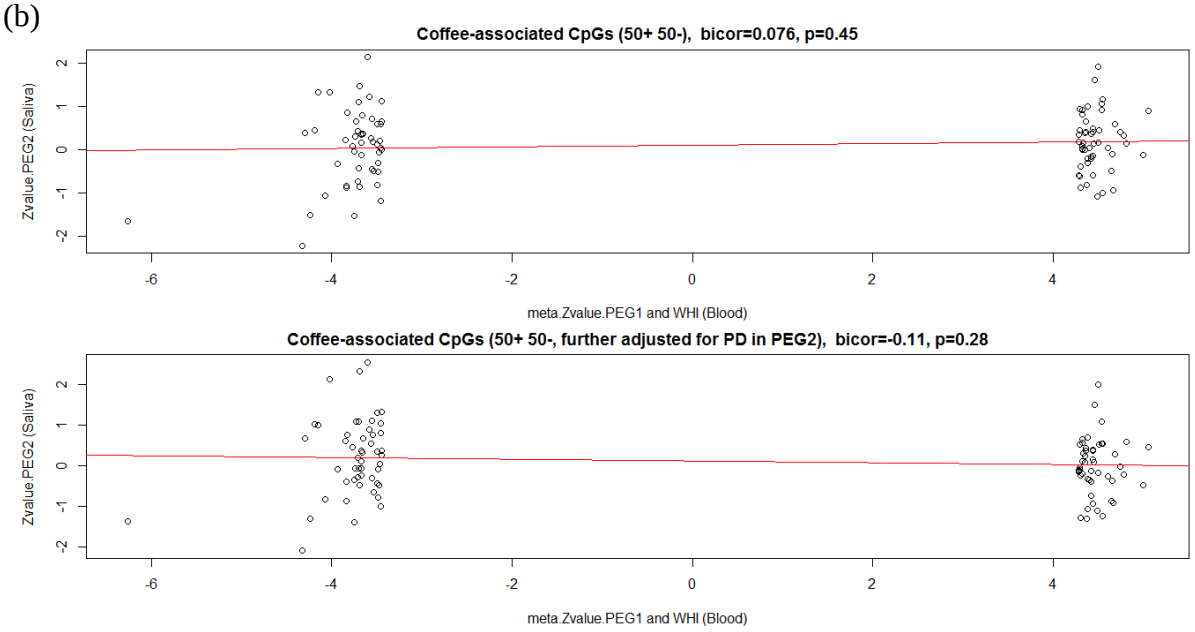
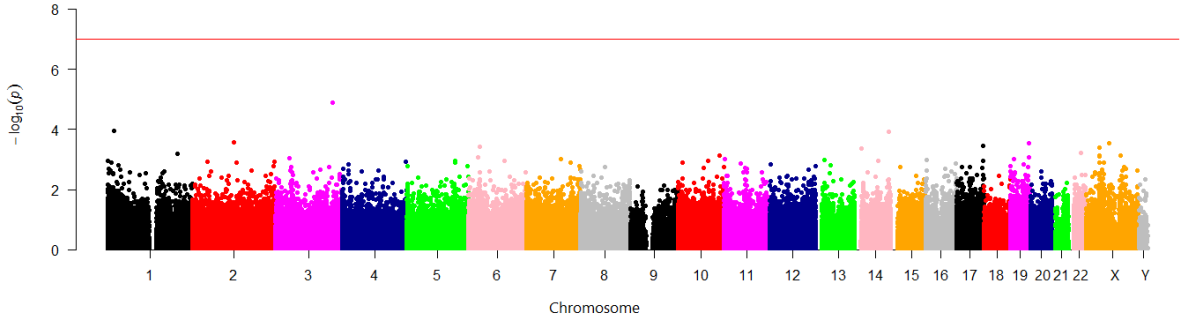


Figure 4-3. Saliva DNA methylation levels associated with coffee consumption adjusted for age, gender and ethnicity. (a) Manhattan plot of the methylation association p-values adjusted for chronological age, gender and ethnicity. Red lines indicate the p-value threshold of 10^{-7}-no CpGs passed this threshold. the x-axis corresponds to negative $\log_{10}$ transformed meta.p-value. the y-axis refers to chromosome number, and X and Y chromosomes. (b) Correlation between Z-values from biweight midcorrelations between DNA methylation and coffee consumption in blood and saliva for 50 most hypermethylated CpGs and 50 most hypomethylated CpGs in blood. The x-axis corresponds to the meta Z-values adjusted for age, gender and blood cell counts from PEG1 and WHI. The y-axis corresponds to the Z-values adjusted for age, gender and ethnicity from PEG2.

(a)



(b)

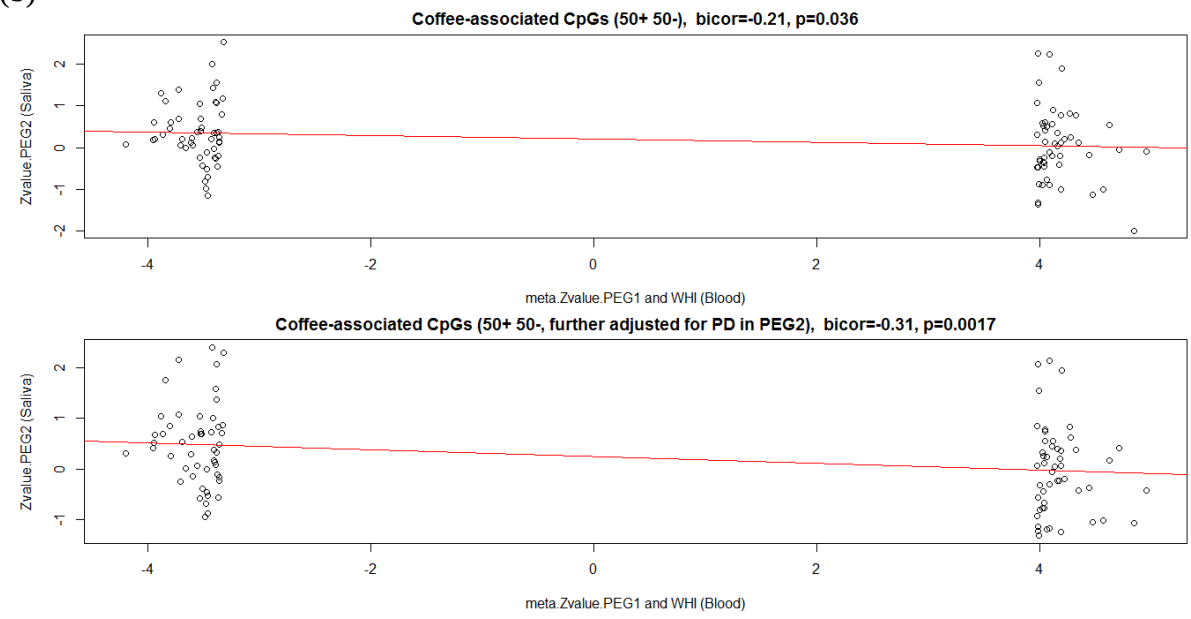


Figure 4-4. Saliva DNA methylation levels associated with coffee consumption adjusted for age, gender, ethnicity and smoking. (a)Manhattan plot of the methylation association p-values adjusted for chronological age, gender, ethnicity and smoking. Red line indicates p-value threshold of 10^{-7}-no CpG passed this threshold. X-axis corresponds to negative $\log_{10}$ transformed meta.p-value. Y-axis refers to chromosome number, X and Y chromosomes. (b) Correlation between Z-values from bi-weighted midcorrelations between DNA methylation and coffee consumption in blood and saliva for 50 most hypermethylated CpGs and 50 most hypomethylated CpGs in blood. X-axis corresponds to the meta Z-values adjusted for age, gender, blood cell counts and smoking from PEG1 and WHI. Y-axis corresponds to the Z-values adjusted for age, gender, ethnicity and smoking from PEG2.

4.6 Supplement

Supplementary Table 4-S1. Characteristics of subjects and exposures

The rows correspond to the datasets used in this article from three studies stratified by race and PD status. Columns report tissue types, disease status (N, %), mean age at blood draw (and range), number female (%), average of daily coffee/tea consumption (SD), smoking status (N, %), and information on MHT use in females (N, %).

	Blood samples (N=2,315)				Saliva samples (N=256)			
	PEG1 Ctrl (N=215)	WHI (N=2,100)			PEG2 PD patients (N=127)		PEG2 Ctrl (N=129)	
	Caucasian	Caucasian	Hispanic	African American	Caucasian	Hispanic	Caucasian	Hispanic
Ethnicity, N	215	995	431	674	81	46	84	45
Tissue types	Blood	Blood	Blood	Blood	Saliva	Saliva	Saliva	Saliva
PD status, N (%)	0 (0%)	.	.	.	81 (100%)	46 (100%)	0 (0%)	0 (0%)
CHD status, N (%)	.	500 (50%)	215 (50%)	335 (50%)	.	.	.	.
Age at blood draw, Mean (range)	68 (35-92)	67 (50-79)	61 (50-79)	62 (50-79)	69 (40-88)	69 (37-88)	69 (40-88)	67 (36-87)
Female, N (%)	101 (47%)	995 (100%)	431 (100%)	674 (100%)	37 (46%)	18 (39%)	38 (45%)	19 (42%)
Coffee or tea (all types) (cup/day), Mean (SD)	3.6 (3.1)	2.6 (1.9)	1.6 (1.5)	1.4 (1.5)	1.8 (1.4)	1.4 (1.2)	2.6 (3.1)	2.2 (2.1)
Ever coffee drinker, N (%)	187 (86%)	786 (79%)	328 (77%)	360 (54%)	64 (79%)	37 (80%)	74 (88%)	34 (76%)
Caffeinated coffee (cup/day), Mean (SD)	2.4 (2.7)	2.3 (1.6)	1.9 (1.3)	1.5 (1.2)	1.3 (1.1)	1.1 (1.2)	1.9 (1.9)	1.8 (1.8)
Decaffeinated coffee (cup/day), Mean (SD)	0.4 (0.9)	.	0.5 (0.8)	0.5 (0.8)	.	.	.	.
Ever Tea drinker, N (%)	169 (79%)	.	36 (21%)	49 (19%)	52 (68%)	20 (45%)	79 (94%)	33 (73%)
Caffeinated Tea (cup/day), Mean (SD)	1.0 (1.7)	.	1.6 (0.6)	1.6 (0.6)	0.8 (0.8)	0.5 (0.6)	0.9 (1.7)	0.6 (0.9)
Decaffeinated Tea (cup/day), Mean (SD)	0.3 (1.0)	.	.	.	.	.	.	.
Caffeinated soft drink (cup/day), Mean (SD)	0.6 (1.3)	.	.	.	.	.	.	.
Ever cigarette smoker, N (%)	129 (60%)	461 (46%)	168 (39%)	353 (53%)	29 (36%)	22 (48%)	47 (56%)	18 (40%)
MHT ever use in female, N (%)	72 (71%)	441 (46%)	260 (62%)	330 (50%)	14 (64%)	8 (67%)	18 (56%)	4 (22%)

NOTE: The percentage of ever coffee/tea drinkers was calculated after missing values were excluded. The mean and standard deviation (SD) of daily coffee/tea consumption is based on ever drinkers only. PD: Parkinson's Disease; CHD: Coronary Heart Disease; MHT: Menopausal Hormone Therapy. *Missing information (PEG1 Ctrl Caucasian/WHI Caucasian/WHI Hispanic/ WHI African American/PEG2 PD Caucasian/PEG2 PD Hispanic/PEG2 Ctrl Caucasian/PEG2 Ctrl Hispanic): tea (N=0/./256/418/4/2/0/0), smoking (N=0/10/5/3/0/0/0/0), MHT (N=0/35/11/11/15/6/6/1).

Supplementary Table 4-S2. Crude and adjusted associations of coffee and 11 top-ranked CpGs in blood by subsets among PEG1 and WHI subjects

Associations were evaluated with bi-weighted midcorrelations with or without adjustment for chronological age, gender, blood cell counts and smoking, stratified by race/ethnicity.

			PEG1 Caucasian Ctrl (N=215)							
			Unadjusted		Adjusted for age, gender		Adjusted for age, gender and cell counts		Adjusted for age, gender, cell counts and smoking	
	CpG	**Gene**	**biCor**	**P-value**	**biCor**	**P-value**	**biCor**	**P-value**	**biCor**	**P-value**
1	cg21566642	intergenic, near ALPPL2	-0.35	1.79E-07	-0.31	1.34E-05	-0.29	3.60E-05	-0.16	2.49E-02
2	cg20333292	GPR132	0.22	1.04E-03	0.25	3.50E-04	0.28	7.43E-05	0.24	7.22E-04
3	cg21163128	BSCL2	0.27	6.52E-05	0.28	9.38E-05	0.28	6.77E-05	0.30	1.87E-05
4	cg23303782	GRK5	0.25	2.16E-04	0.26	2.16E-04	0.29	3.55E-05	0.25	5.04E-04
5	cg26105150	FSTL5	0.23	6.48E-04	0.25	5.52E-04	0.27	1.35E-04	0.23	1.25E-03
6	cg19723563	PTHLH	0.16	1.55E-02	0.15	3.70E-02	0.17	2.14E-02	0.13	7.52E-02
7	cg17928869	intergenic, between EIF1 and KRT42P	0.18	7.33E-03	0.17	1.48E-02	0.23	1.35E-03	0.19	9.56E-03
8	cg12866551	MALRD1	0.12	8.06E-02	0.16	2.96E-02	0.23	1.48E-03	0.21	3.44E-03
9	cg15722372	PSMD8	0.18	9.28E-03	0.18	1.12E-02	0.20	4.55E-03	0.19	9.31E-03
10	cg19974428	TMEM130	0.19	4.49E-03	0.21	4.11E-03	0.22	2.29E-03	0.16	2.85E-02
11	cg15140902	FLJ22536	0.13	6.38E-02	0.13	6.98E-02	0.12	1.09E-01	0.10	1.83E-01

Supplementary Table 4-S2 (continued)

			WHI Caucasian (N=995)							
			Unadjusted		Adjusted for age		Adjusted for age and cell counts		Adjusted for age, cell counts and smoking	
	CpG	**Gene**	**biCor**	**P-value**	**biCor**	**P-value**	**biCor**	**P-value**	**biCor**	**P-value**
1	cg21566642	intergenic, near ALPPL2	-0.14	1.72E-05	-0.13	2.41E-05	-0.14	1.34E-05	-0.09	3.31E-03
2	cg20333292	GPR132	0.04	2.05E-01	0.04	1.95E-01	0.05	8.67E-02	0.06	8.21E-02
3	cg21163128	BSCL2	0.10	2.24E-03	0.09	3.49E-03	0.10	2.40E-03	0.09	7.19E-03
4	cg23303782	GRK5	0.05	1.11E-01	0.05	9.35E-02	0.06	4.37E-02	0.06	4.77E-02
5	cg26105150	FSTL5	0.03	3.24E-01	0.04	1.88E-01	0.05	8.98E-02	0.05	1.31E-01
6	cg19723563	PTHLH	0.06	6.86E-02	0.05	9.80E-02	0.07	2.20E-02	0.06	4.61E-02
7	cg17928869	intergenic, between EIF1 and KRT42P	0.05	1.12E-01	0.05	1.36E-01	0.08	1.17E-02	0.07	3.78E-02
8	cg12866551	MALRD1	0.06	5.83E-02	0.05	9.93E-02	0.07	2.04E-02	0.07	2.71E-02
9	cg15722372	PSMD8	0.04	1.77E-01	0.04	1.83E-01	0.05	9.92E-02	0.05	9.84E-02
10	cg19974428	TMEM130	0.05	1.21E-01	0.05	9.58E-02	0.07	3.67E-02	0.06	4.46E-02
11	cg15140902	FLJ22536	0.09	6.84E-03	0.09	3.67E-03	0.10	1.52E-03	0.08	8.09E-03

Supplementary Table 4-S2 (continued)

			WHI Hispanic (N=431)							
			Unadjusted		Adjusted for age		Adjusted for age and cell counts		Adjusted for age, cell counts and smoking	
	CpG	**Gene**	**biCor**	**P-value**	**biCor**	**P-value**	**biCor**	**P-value**	**biCor**	**P-value**
1	cg21566642	intergenic, near ALPPL2	-0.08	1.10E-01	-0.08	1.06E-01	-0.07	1.28E-01	-0.02	7.14E-01
2	cg20333292	GPR132	0.12	1.01E-02	0.12	1.32E-02	0.13	6.87E-03	0.12	1.36E-02
3	cg21163128	BSCL2	0.09	6.01E-02	0.09	6.48E-02	0.09	6.23E-02	0.09	6.96E-02
4	cg23303782	GRK5	0.07	1.39E-01	0.07	1.62E-01	0.09	6.39E-02	0.07	1.29E-01
5	cg26105150	FSTL5	0.11	1.71E-02	0.11	2.26E-02	0.16	6.78E-04	0.16	9.35E-04
6	cg19723563	PTHLH	0.14	3.62E-03	0.14	4.34E-03	0.13	5.22E-03	0.13	8.18E-03
7	cg17928869	intergenic, between EIF1 and KRT42P	0.10	3.77E-02	0.10	4.77E-02	0.11	2.11E-02	0.10	3.05E-02
8	cg12866551	MALRD1	0.15	1.45E-03	0.14	4.42E-03	0.14	2.75E-03	0.15	2.18E-03
9	cg15722372	PSMD8	0.08	1.09E-01	0.07	1.34E-01	0.10	4.67E-02	0.09	5.47E-02
10	cg19974428	TMEM130	0.07	1.23E-01	0.08	9.41E-02	0.11	2.64E-02	0.10	3.67E-02
11	cg15140902	FLJ22536	0.06	1.99E-01	0.07	1.42E-01	0.10	4.12E-02	0.09	6.58E-02

Supplementary Table 4-S2 (continued)

			WHI African American (N=674)							
			Unadjusted		Adjusted for age		Adjusted for age and cell counts		Adjusted for age, cell counts and smoking	
	CpG	**Gene**	**biCor**	**P-value**	**biCor**	**P-value**	**biCor**	**P-value**	**biCor**	**P-value**
1	cg21566642	intergenic, near ALPPL2	-0.09	1.73E-02	-0.09	1.47E-02	-0.10	1.33E-02	-0.05	1.72E-01
2	cg20333292	GPR132	0.06	1.13E-01	0.06	1.12E-01	0.07	9.10E-02	0.06	1.04E-01
3	cg21163128	BSCL2	0.03	4.08E-01	0.03	3.91E-01	0.04	2.80E-01	0.04	2.82E-01
4	cg23303782	GRK5	0.06	1.02E-01	0.06	9.94E-02	0.06	1.14E-01	0.06	1.22E-01
5	cg26105150	FSTL5	0.02	5.42E-01	0.02	6.06E-01	0.02	5.45E-01	0.02	5.45E-01
6	cg19723563	PTHLH	0.09	2.00E-02	0.09	2.16E-02	0.08	3.77E-02	0.08	4.43E-02
7	cg17928869	intergenic, between EIF1 and KRT42P	0.06	1.19E-01	0.06	1.17E-01	0.05	1.93E-01	0.05	2.05E-01
8	cg12866551	MALRD1	0.03	5.03E-01	0.03	4.27E-01	0.03	4.21E-01	0.03	4.72E-01
9	cg15722372	PSMD8	0.11	3.09E-03	0.11	3.19E-03	0.11	4.94E-03	0.10	7.21E-03
10	cg19974428	TMEM130	0.07	5.44E-02	0.07	6.55E-02	0.07	5.55E-02	0.07	7.63E-02
11	cg15140902	FLJ22536	0.09	2.47E-02	0.09	2.51E-02	0.09	1.72E-02	0.09	2.47E-02

Supplementary Table 4-S3. Associations of coffee and 11 top-ranked CpGs in blood (Table 1a, without smoking adjustment) by gender in the PEG study

Associations were evaluated with bi-weighted midcorrelations with adjustment for chronological age and blood cell counts, stratified by gender.

			PEG1 Caucasian Ctrl					
			All, adjusted for age, gender and cell counts (N=215)		Male, adjusted for age and cell counts (N=114)		Female, adjusted for age and cell counts (N=101)	
	CpG	**Gene**	**biCor**	**P-value**	**biCor**	**P-value**	**biCor**	**P-value**
1	cg21566642	intergenic, near ALPPL2	-0.29	3.60E-05	-0.29	1.51E-03	-0.26	9.52E-03
2	cg20333292	GPR132	0.28	7.43E-05	0.31	6.63E-04	-0.01	9.60E-01
3	cg21163128	BSCL2	0.28	6.77E-05	0.38	3.48E-05	0.15	1.30E-01
4	cg23303782	GRK5	0.29	3.55E-05	0.41	6.19E-06	-0.07	5.05E-01
5	cg26105150	FSTL5	0.27	1.35E-04	0.27	3.80E-03	0.18	7.45E-02
6	cg19723563	PTHLH	0.17	2.14E-02	0.23	1.20E-02	0.10	3.40E-01
7	cg17928869	intergenic, between EIF1 and KRT42P	0.23	1.35E-03	0.24	9.19E-03	0.19	5.11E-02
8	cg12866551	MALRD1	0.23	1.48E-03	0.26	4.45E-03	0.02	8.30E-01
9	cg15722372	PSMD8	0.20	4.55E-03	0.27	3.60E-03	-0.01	8.91E-01
10	cg19974428	TMEM130	0.22	2.29E-03	0.35	1.24E-04	-0.04	7.03E-01
11	cg15140902	FLJ22536	0.12	1.09E-01	0.12	1.93E-01	0.05	5.88E-01

Supplementary Table 4-S4. CpG sites influenced by both coffee consumption and smoking adjusted for chronological age, gender and blood cell counts

List of CpGs associated with both coffee consumption and smoking (p-value threshold ≤10^{-7}), (near) gene, chromosome, and CpG island they are located in, gene region, (Stouffer's test) Z-value and p-value from meta-analysis for coffee-DNAm and smoking-DNAm associations separately and combined.

							Coffee and Smoking		Coffee		Smoking	
	CpG	**Gene**	**Chr.**	**Position(bp)[a]**	**Relation to UCSC CpG Island**	**UCSC RefGene Group[b]**	**meta. Z**	**meta. Pvalue**	**meta. Z**	**meta. Pvalue**	**meta. Z**	**meta. Pvalue**
1	cg21566642	intergenic, near ALPPL2	2	233284661	Island		-19.27	<4.00E-16	-6.26	3.73E-10	-20.98	<4.00E-16
2	cg05951221		2	233284402	Island		-16.52	<4.00E-16	-4.33	1.50E-05	-19.03	<4.00E-16
3	cg05575921	AHRR	5	373378	N_Shore	Body	-16.79	<4.00E-16	-4.24	2.24E-05	-19.51	<4.00E-16
4	cg06644428		2	233284112	Island		-12.22	<4.00E-16	-4.07	4.63E-05	-13.21	<4.00E-16
5	cg12803068	MYO1G	7	45002919	S_Shore	Body	8.42	<4.00E-16	4.04	5.42E-05	7.87	3.55E-15
6	cg01940273		2	233284934	Island		-12.87	<4.00E-16	-3.46	5.42E-04	-14.74	<4.00E-16
7	cg23576855	AHRR	5	373299	N_Shore	Body	-8.33	<4.00E-16	-3.30	9.68E-04	-8.48	<4.00E-16
8	cg06126421		6	30720080			-13.51	<4.00E-16	-3.29	9.95E-04	-15.82	<4.00E-16
9	cg03636183	F2RL3	19	17000585	N_Shore	Body	-11.40	<4.00E-16	-3.26	1.13E-03	-12.87	<4.00E-16
10	cg17287155	AHRR	5	393347		Body	-9.65	<4.00E-16	-2.94	3.28E-03	-10.70	<4.00E-16
11	cg03604011	AHRR	5	400201		Body	8.84	<4.00E-16	2.86	4.18E-03	9.64	<4.00E-16
12	cg25189904	GNG12	1	68299493	S_Shore	TSS1500	-8.93	<4.00E-16	-1.99	4.63E-02	-10.64	<4.00E-16
13	cg13937905	RARG	12	53612551	N_Shore	Body	-8.14	4.44E-16	-3.39	6.96E-04	-8.12	4.44E-16
14	cg04180046	MYO1G	7	45002736	Island	Body	7.85	4.00E-15	3.84	1.24E-04	7.27	3.59E-13
15	cg09935388	GFI1	1	92947588	Island	Body	-7.85	4.22E-15	-3.12	1.79E-03	-7.97	1.55E-15
16	cg17924476	AHRR	5	323794	S_Shore	Body	7.38	1.61E-13	3.19	1.43E-03	7.25	4.28E-13
17	cg13193840		2	233285289	Island		-7.25	4.26E-13	-2.87	4.12E-03	-7.38	1.59E-13
18	cg14656441	NDUFS5	1	39500070		Body	7.23	4.68E-13	4.26	2.09E-05	5.98	2.29E-09
19	cg19572487	RARA	17	38476024	S_Shore	5'UTR	-7.12	1.08E-12	-1.09	2.76E-01	-8.98	<4.00E-16
20	cg03329539		2	233283329	N_Shore		-7.11	1.20E-12	-1.75	7.99E-02	-8.30	<4.00E-16
21	cg03991871	AHRR	5	368447	N_Shore	Body	-6.78	1.23E-11	-2.12	3.37E-02	-7.46	8.66E-14
22	cg14675361	LMO7	13	76334583		TSS1500	6.76	1.39E-11	3.91	9.37E-05	5.65	1.59E-08

23	cg23916896	AHRR	5	368804	N_Shore	Body	-6.73	1.68E-11	-1.53	1.27E-01	-7.99	1.33E-15
24	cg21322436	CNTP2	7	145812842	N_Shore	TSS1500	-6.72	1.87E-11	-2.68	7.40E-03	-6.82	9.14E-12
25	cg07967717	CNR2	1	24229682	S_Shore	5'UTR	6.68	2.43E-11	4.46	8.06E-06	4.98	6.38E-07
26	cg21806580		5	18746010			6.67	2.51E-11	3.76	1.67E-04	5.67	1.41E-08
27	cg13039251	PDZD2	5	32018601		Body	6.61	3.78E-11	2.74	6.11E-03	6.61	3.85E-11
28	cg02714303	LMO7	13	76334728		TSS200	6.53	6.75E-11	3.14	1.68E-03	6.09	1.15E-09
29	cg10581837	LMO7	13	76334707		TSS200	6.50	8.25E-11	3.39	6.98E-04	5.80	6.78E-09
30	cg13185177	GP5	3	194119885	S_Shore	5'UTR	6.48	9.44E-11	2.91	3.66E-03	6.25	4.05E-10
31	cg23288337	ECEL1P2	2	233250701	N_Shore	Body	6.44	1.21E-10	3.26	1.13E-03	5.85	4.93E-09
32	cg14179389	GFI1	1	92947961	Island	Body	-6.43	1.27E-10	-2.49	1.27E-02	-6.60	4.05E-11
33	cg23480021		3	22412746	N_Shore		6.41	1.50E-10	1.47	1.43E-01	7.59	3.13E-14
34	cg11660018	PRSS23	11	86510915	N_Shore	TSS1500	-6.31	2.77E-10	-0.85	3.94E-01	-8.07	6.66E-16
35	cg26242531	ZFYVE21	14	104190678	N_Shelf	Body	6.30	2.96E-10	3.34	8.26E-04	5.57	2.59E-08
36	cg00706683	ECEL1P2	2	233251030	N_Shore	Body	6.24	4.43E-10	2.58	9.97E-03	6.25	4.23E-10
37	cg03274391		3	22413232	N_Shore		6.23	4.65E-10	1.85	6.43E-02	6.96	3.37E-12
38	cg08305533	SFI1	22	32007224		Body	6.15	7.72E-10	3.98	6.90E-05	4.72	2.37E-06
39	cg14753356		6	30720108			-6.14	8.03E-10	-0.32	7.46E-01	-8.37	<4.00E-16
40	cg11492723		3	16577697			6.14	8.49E-10	4.24	2.27E-05	4.44	8.99E-06
41	cg11165912	COL25A1	4	110220063	N_Shelf	Body	6.13	8.84E-10	3.31	9.24E-04	5.36	8.55E-08
42	cg17033047	KC3	1	111214851	N_Shore	3'UTR	6.11	9.86E-10	3.14	1.72E-03	5.51	3.63E-08
43	cg11207515	CNTP2	7	146904205		Body	6.05	1.44E-09	1.30	1.93E-01	7.25	4.06E-13
44	cg19744173	FBLN7	2	112913178	N_Shelf	Body	6.02	1.70E-09	3.21	1.34E-03	5.31	1.08E-07
45	cg08940075	CNN3	1	95388680	N_Shelf	Body	6.00	1.95E-09	3.33	8.61E-04	5.16	2.53E-07
46	cg08606254	AHRR	5	323969	S_Shore	Body	6.00	2.01E-09	2.45	1.45E-02	6.04	1.58E-09
47	cg22947000	BCMO1	16	81272281		TSS200	5.99	2.13E-09	3.49	4.87E-04	4.98	6.36E-07
48	cg26601310	PRR5L	11	36397123	N_Shore	5'UTR	5.98	2.30E-09	4.45	8.78E-06	4.00	6.21E-05
49	cg18390495	DEFB132	20	239809		Body	5.97	2.39E-09	3.09	2.03E-03	5.36	8.55E-08
50	cg24227984	HDGF2	19	4474970	S_Shore	Body	5.97	2.40E-09	3.45	5.63E-04	4.99	5.97E-07
51	cg13008538	CARD11	7	3083741		TSS200	5.97	2.41E-09	3.20	1.37E-03	5.24	1.62E-07
52	cg00665106		1	201515370			5.97	2.43E-09	3.98	6.85E-05	4.46	8.34E-06
53	cg15693572		3	22412385	N_Shore		5.95	2.64E-09	1.82	6.92E-02	6.60	4.09E-11
54	cg21161138	AHRR	5	399360		Body	-5.92	3.15E-09	-0.95	3.40E-01	-7.42	1.14E-13

55	cg14027333	PRRT1	6	32116317	N_Shore	3'UTR	5.91	3.50E-09	2.53	1.15E-02	5.83	5.66E-09
56	cg18961281	CARD11	7	3083773		TSS200	5.90	3.69E-09	3.65	2.60E-04	4.69	2.76E-06
57	cg21771528		5	18745861			5.89	3.82E-09	2.98	2.93E-03	5.36	8.47E-08
58	cg07211476		18	53448189	S_Shore		5.86	4.64E-09	3.38	7.18E-04	4.90	9.40E-07
59	cg11557553	AHRR	5	404996		Body	5.85	4.97E-09	2.81	4.95E-03	5.46	4.75E-08
60	cg04951293		2	231688289			5.85	4.99E-09	4.33	1.48E-05	3.94	8.22E-05
61	cg02657160	CPOX	3	98311063	N_Shore	Body	-5.84	5.07E-09	-1.77	7.62E-02	-6.49	8.43E-11
62	cg25485805		3	13246862	S_Shore		5.84	5.18E-09	3.68	2.30E-04	4.58	4.71E-06
63	cg09570614		2	112124380			5.83	5.49E-09	3.52	4.31E-04	4.73	2.28E-06
64	cg25520249	CYP2S1	19	41700269	S_Shore	Body	5.82	5.80E-09	3.60	3.23E-04	4.64	3.51E-06
65	cg14977938	ZFYVE21	14	104190829	N_Shelf	Body	5.79	7.00E-09	3.11	1.84E-03	5.08	3.87E-07
66	cg20495738	CAC1C	12	2338399	N_Shore	Body	5.79	7.11E-09	3.04	2.36E-03	5.14	2.68E-07
67	cg00911794	HIC1	17	1962132	Island	3'UTR	5.78	7.31E-09	1.93	5.31E-02	6.25	4.24E-10
68	cg14120896		19	56086441	N_Shelf		5.77	7.92E-09	3.15	1.63E-03	5.01	5.43E-07
69	cg19713851	ALPP	2	233246594	S_Shore	3'UTR	-5.77	7.96E-09	-1.53	1.25E-01	-6.63	3.45E-11
70	cg01692968		9	108005349	N_Shore		-5.75	9.01E-09	-0.72	4.74E-01	-7.41	1.23E-13
71	cg13184736	GNG12	1	68299409	S_Shore	TSS1500	-5.75	9.14E-09	-1.15	2.49E-01	-6.97	3.12E-12
72	cg24145685	STK11	19	1227852	Island	3'UTR	5.74	9.53E-09	4.54	5.73E-06	3.58	3.44E-04
73	cg14817490	AHRR	5	392920		Body	-5.73	9.77E-09	-1.17	2.43E-01	-6.94	3.86E-12
74	cg09762515	CUX1	7	101556588	N_Shelf	Body	5.73	1.01E-08	3.50	4.64E-04	4.60	4.22E-06
75	cg23161492	ANPEP	15	90357202	N_Shore	5'UTR	-5.72	1.05E-08	-1.01	3.12E-01	-7.08	1.40E-12
76	cg08035323		2	9843525			5.72	1.08E-08	3.64	2.73E-04	4.45	8.77E-06
77	cg06635952	ANXA4	2	70025869		Body	5.70	1.22E-08	2.52	1.17E-02	5.53	3.12E-08
78	cg23621097	HIC1	17	1962236	Island	3'UTR	5.69	1.28E-08	2.75	5.96E-03	5.29	1.19E-07
79	cg25800753		10	94891725			5.67	1.45E-08	3.96	7.46E-05	4.05	5.04E-05
80	cg06885459	MCF2L	13	113689728	S_Shore	Body	5.66	1.48E-08	3.76	1.73E-04	4.25	2.10E-05
81	cg17272563	PRRT1	6	32116548	N_Shore	3'UTR	5.66	1.48E-08	2.82	4.75E-03	5.19	2.14E-07
82	cg05768005	C17orf59	17	8094486	S_Shore	TSS1500	5.66	1.49E-08	3.65	2.67E-04	4.36	1.28E-05
83	cg04961225		12	91332479			5.66	1.55E-08	3.65	2.58E-04	4.34	1.40E-05
84	cg05156137	RCAN1	21	35898975		5'UTR	5.64	1.69E-08	2.64	8.19E-03	5.33	9.60E-08
85	cg04912316	FAM100B	17	74266324		Body	-5.64	1.71E-08	-2.96	3.03E-03	-5.01	5.44E-07
86	cg04737759		6	155253937			5.61	1.98E-08	3.41	6.44E-04	4.53	6.01E-06

87	cg03008815	RPS6KL1	14	75372636		3'UTR	5.61	2.02E-08	3.58	3.40E-04	4.35	1.35E-05
88	cg19325791		17	46560683			5.58	2.41E-08	3.34	8.39E-04	4.55	5.32E-06
89	cg14397231		4	124940041			5.57	2.62E-08	2.42	1.54E-02	5.45	5.11E-08
90	cg25949550	CNTP2	7	145814306	S_Shore	Body	-5.55	2.78E-08	-0.09	9.25E-01	-7.76	8.44E-15
91	cg02612963	ZNF483	9	114287383	Island	TSS200	5.55	2.81E-08	3.48	5.10E-04	4.38	1.20E-05
92	cg24994127	RALGAPA2	20	20433328	Island	Body	5.54	2.98E-08	3.83	1.30E-04	4.01	6.03E-05
93	cg01352586		1	202995473	Island		5.53	3.25E-08	4.30	1.69E-05	3.51	4.41E-04
94	cg11902777	AHRR	5	368843	N_Shore	Body	-5.51	3.50E-08	-1.43	1.51E-01	-6.36	1.96E-10
95	cg09069072	TMEM51	1	15482753	S_Shore	5'UTR	-5.50	3.76E-08	-2.83	4.65E-03	-4.95	7.39E-07
96	cg27178191	SLC4A8	12	51819149	Island	Body	5.50	3.84E-08	2.99	2.79E-03	4.79	1.71E-06
97	cg18747408		10	130841097			5.50	3.84E-08	3.06	2.20E-03	4.71	2.44E-06
98	cg23771366	PRSS23	11	86510998	N_Shore	TSS1500	-5.50	3.89E-08	-0.41	6.82E-01	-7.36	1.82E-13
99	cg02152819		2	232258417	N_Shore		5.49	4.02E-08	2.82	4.84E-03	4.95	7.56E-07
100	cg25955180	PRRT1	6	32116538	N_Shore	3'UTR	5.48	4.23E-08	2.47	1.35E-02	5.28	1.28E-07
101	cg15037583	MACROD1	11	63767176	Island	Body	5.48	4.28E-08	3.49	4.87E-04	4.26	2.04E-05
102	cg19974428	TMEM130	7	98468047	Island	TSS1500	5.47	4.42E-08	4.65	3.34E-06	3.09	1.99E-03
103	cg07178945	FGF23	12	4488800		5'UTR	5.47	4.58E-08	2.11	3.49E-02	5.62	1.89E-08
104	cg13167816	FITM1	14	24601808	S_Shore	Body	5.46	4.64E-08	3.26	1.10E-03	4.46	8.01E-06
105	cg03489965	LOC390594	15	65368982	N_Shore	TSS200	5.45	4.94E-08	1.47	1.41E-01	6.24	4.34E-10
106	cg08553467	C3orf27	3	128292922		5'UTR	5.45	4.99E-08	3.77	1.63E-04	3.94	8.17E-05
107	cg00138101	HIC1	17	1961109	Island	Body	5.45	5.14E-08	3.53	4.16E-04	4.17	3.01E-05
108	cg25474399	MYOM1	18	3067624	S_Shore	Body	5.44	5.23E-08	2.58	9.83E-03	5.12	3.12E-07
109	cg12549300		13	113594776	N_Shelf		5.44	5.39E-08	3.15	1.62E-03	4.54	5.67E-06
110	cg24306779	MMP15	16	58060336	Island	1stExon	5.43	5.62E-08	2.88	3.95E-03	4.80	1.61E-06
111	cg06213060		3	16577726			5.43	5.66E-08	3.30	9.65E-04	4.38	1.20E-05
112	cg23973524	CRTC1	19	18873222	Island	Body	5.43	5.67E-08	3.68	2.32E-04	4.00	6.44E-05
113	cg04768713	ZNF701	19	53073334	Island	TSS200	5.41	6.30E-08	3.57	3.54E-04	4.08	4.53E-05
114	cg00336149	CAC1D	3	53700195		Body	5.41	6.30E-08	2.67	7.54E-03	4.98	6.39E-07
115	cg14422093		1	227009245			5.41	6.31E-08	3.33	8.65E-04	4.32	1.56E-05
116	cg01207684	ADCY9	16	4103167		Body	-5.41	6.36E-08	-2.54	1.09E-02	-5.10	3.32E-07
117	cg25197194	CCDC48	3	128758787		3'UTR	5.40	6.59E-08	2.05	4.01E-02	5.59	2.31E-08
118	cg08229199	GPR135	14	59932289	S_Shore	TSS1500	5.40	6.65E-08	3.38	7.12E-04	4.25	2.12E-05

119	cg20431135	MFAP4	17	19290762		TSS1500	5.39	7.17E-08	3.71	2.06E-04	3.91	9.38E-05
120	cg13015908		21	34481860			5.38	7.41E-08	4.12	3.80E-05	3.49	4.82E-04
121	cg17928869	intergenic, between EIF1 and KRT42P	17	39822542	S_Shore		5.38	7.48E-08	4.68	2.86E-06	2.93	3.42E-03
122	cg13380502	GPR3	1	27718221	N_Shore	TSS1500	5.38	7.66E-08	2.79	5.35E-03	4.82	1.46E-06
123	cg14823389		7	101398152			5.37	8.07E-08	3.17	1.54E-03	4.42	9.86E-06
124	cg00931843	TIAM2	6	155442993		5'UTR	5.36	8.10E-08	2.58	9.81E-03	5.00	5.59E-07
125	cg27067781	PRRT1	6	32116853	Island	3'UTR	5.36	8.17E-08	2.26	2.39E-02	5.33	1.01E-07
126	cg16331679	C6orf103	6	146920166	Island	1stExon	5.36	8.29E-08	3.25	1.16E-03	4.33	1.48E-05
127	cg19719391		4	26789915			5.36	8.40E-08	2.83	4.70E-03	4.75	2.03E-06
128	cg02627991		7	74032045			5.35	8.62E-08	3.99	6.67E-05	3.58	3.39E-04
129	cg23061027	PRRT1	6	32116207	N_Shore	3'UTR	5.35	8.69E-08	1.74	8.27E-02	5.83	5.41E-09
130	cg26390660		3	110344165			5.35	8.99E-08	2.66	7.71E-03	4.90	9.77E-07
131	cg05462761	PLEKHG5	1	6545157	Island	5'UTR	5.34	9.08E-08	3.44	5.83E-04	4.12	3.81E-05
132	cg16705546	IRF8	16	85936666	S_Shelf	Body	5.34	9.50E-08	3.24	1.20E-03	4.31	1.66E-05
133	cg14488466		9	93823465			5.33	9.69E-08	2.43	1.50E-02	5.11	3.23E-07
134	cg01899620	MCF2L	13	113689422	S_Shore	Body	5.33	9.73E-08	3.63	2.79E-04	3.91	9.37E-05
135	cg21878650	ADAMTS6	5	64558623		Body	5.33	9.97E-08	3.13	1.73E-03	4.40	1.08E-05

[a] Location is based on NCBI genome build 37.
[b]TSS: transcription start site, TSS500: within 1500 bps of a TSS, TSS200: within 200 bps of a TSS, UTR: untranslated region.

Supplementary Table 4-S5. Saliva results for the 5 top-ranked CpG sites in Table 1b and adjusted for chronological age, gender, race and smoking

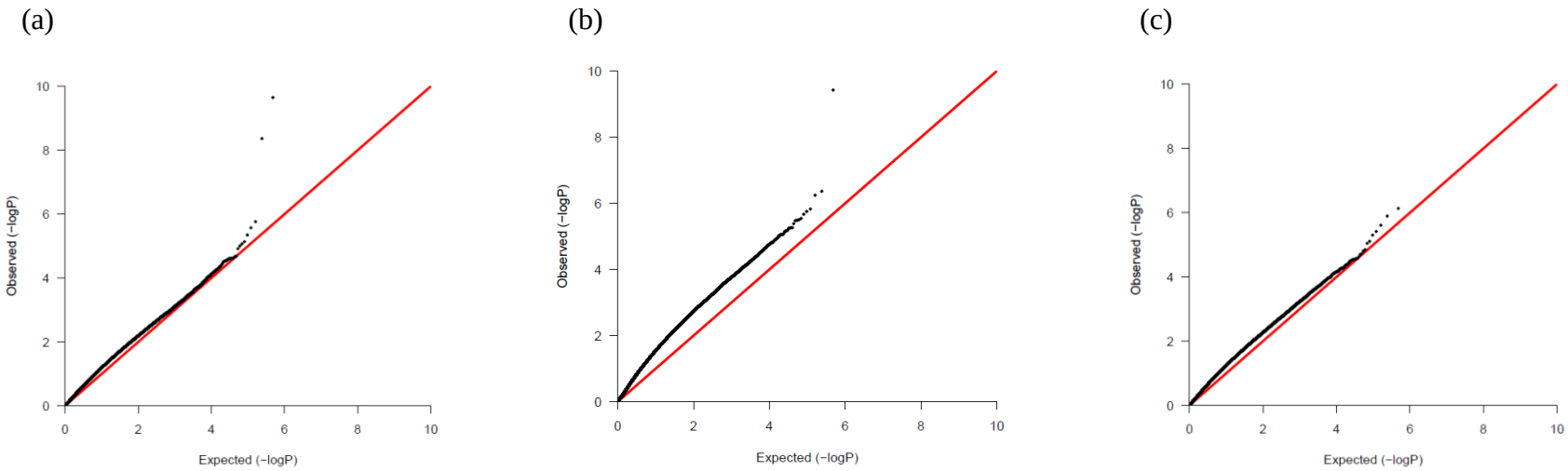
List of CpGs associated with coffee consumption, (near) gene, chromosome, and CpG island they are located in, gene region, a robust correlation coefficient (known as biweight midcorrelation) and p-value for daily coffee consumption (in last 12 months) and DNA methylation levels in the PEG2 study.

							PEG2 subjects, adjusted for age, gender, race, smoking			PEG2 subjects, adjusted for age, gender, race, smoking and PD status		
	CpG	**Gene**	**Chr.**	**Position (bp)[a]**	**Relation to UCSC CpG Island**	**UCSC RefGene Group[b]**	**biCor**	**Z-value**	**P-value**	**biCor**	**Z-value**	**P-value**
1	cg21163128	BSCL2	11	62477362	Island	TSS1500	-0.01	-0.09	9.26E-01	-0.03	-0.41	6.80E-01
2	cg08119527	Intergenic, near PODXL	7	131340667			-0.12	-2.00	4.65E-02	-0.07	-1.06	2.91E-01
3	cg26331135	CNTN4	3	2144068	S_Shelf	5'UTR	0.00	-0.05	9.57E-01	0.03	0.42	6.75E-01
4	cg20333292	GPR132	14	105532012		TSS1500	0.03	0.53	5.93E-01	0.01	0.16	8.70E-01
5	cg08311403	ROBO3	11	124735215	Island	TSS200	-0.06	-1.00	3.18E-01	-0.06	-1.01	3.12E-01

Abbreviations: BSCL2 (Berardinelli-Seip Congenital Lipodystrophy 2 (Seipin)), PODXL (Podocalyxin-Like), CNTN4, GPR132 (G Protein-Coupled Receptor 132), ROBO3 (Roundabout guidance receptor 3).

[a]Location is based on NCBI genome build 37.

[b]TSS: transcription start site, TSS500: within 1500 bps of a TSS, TSS200: within 200 bps of a TSS, UTR: untranslated region.



Supplementary Figure 4-S1. Q-Q plots of meta.Pvalue for the association between coffee consumption and each methylation site: a) adjust for age and gender (lambda=1.49); b) adjust for age, gender, and blood cell composition (lambda=2.24); c) adjust for age, gender, blood cell composition, and smoking (lambda=1.59).

5 Conclusion and Public Health Implications

This dissertation examines the combined effect of lifestyle (i.e. coffee consumption and smoking) and genetic factors on the risk of Parkinson's disease, as well as the association of coffee consumption with DNA methylation levels in both human blood and saliva tissues. In the first project, we found significant interactions between *ADORA2A*, *CYP1A2* variants, and coffee consumptions in reducing PD risk. The inverse association between coffee and PD is strongest among homozygous minor-allele carriers, which supports the hypothesis that caffeine has a neuroprotective effect against PD. Also, we suggest that survivor bias may affect results of studies that enroll prevalent PD cases. In the second project, we confirmed the inverse association of *HLA-DRB1* variants with PD, and found this association was only observed in non-smokers. We provide the first evidence for the involvement of both *HLA-DRB1* and smoking in the development of PD, possibly through neuroinflammation. In the third project, we found coffee affects DNA methylation levels in immune cells of the blood but not in saliva. And the coffee-associated methylation sites are linked to genes involved in lipid metabolism and immune response, suggesting possible mechanisms through which coffee consumption may be beneficial for health. We also provide support for the modified effect of postmenopausal estrogen use on coffee's effect.

Given the aging of the population worldwide, PD has a great impact not only on morbidity and mortality but on quality of life. Because the cause of idiopathic PD is unknown, proven ways to prevent the disease remain unclear. However, animal studies have showed that administration of caffeine or other adenosine A_{2A} receptor antagonists (e.g. istradefylline) before the administration of PD-causing neurotoxin MPTP, significantly inhibited the MPTP-induced loss of dopamine and dopaminergic neurons, suggesting that caffeine exposure in

early life could possibly reduce PD risk by inactivation of A_{2A} receptors (Chen et al. 2001; Pierri et al. 2005). In addition, adenosine A_{2A} receptor antagonist has been targeted for the non-dopaminergic treatment for PD in pharmacological studies, aiming to eliminate patients' side effect from dopamine targeted therapy. A double-blind randomized clinical trial was conducted to examine istradefylline and found to significantly decrease the wearing-off time without introducing dyskinesia compared to levodopa (LeWitt et al. 2008). The continuously exploration of molecular mechanisms underlying the development of PD may help make better prevention and treatment strategies and reduce the suffering.

References

(2015). "Indvandrere i Danmark 2015." from http://www.dst.dk/da/Statistik/Publikationer/VisPub?cid=20703.

Ahmed, I., P. C. Lee, et al. (2014). "Lack of replication of the GRIN2A-by-coffee interaction in Parkinson disease." PLoS Genet **10**(11): e1004788.

Ahmed, I., R. Tamouza, et al. (2012). "Association between Parkinson's disease and the HLA-DRB1 locus." Mov Disord **27**(9): 1104-1110.

Alsene, K., J. Deckert, et al. (2003). "Association between A2a receptor gene polymorphisms and caffeine-induced anxiety." Neuropsychopharmacology : official publication of the American College of Neuropsychopharmacology **28**(9): 1694-1702.

Amin, N., E. Byrne, et al. (2012). "Genome-wide association analysis of coffee drinking suggests association with CYP1A1/CYP1A2 and NRCAM." Mol Psychiatry **17**(11): 1116-1129.

Andersen, J. K. (2004). "Oxidative stress in neurodegeneration: cause or consequence?" Nat Med **10 Suppl**: S18-25.

Anonymous, A. (1998). "Design of the Women's Health Initiative clinical trial and observational study. The Women's Health Initiative Study Group." Control Clin Trials **19**(1): 61-109.

Arendash, G. W., W. Schleif, et al. (2006). "Caffeine protects Alzheimer's mice against cognitive impairment and reduces brain beta-amyloid production." Neuroscience **142**(4): 941-952.

Ascherio, A., M. G. Weisskopf, et al. (2004). "Coffee consumption, gender, and Parkinson's disease mortality in the cancer prevention study II cohort: the modifying effects of estrogen." Am J Epidemiol **160**(10): 977-984.

Backman, J. T., M. T. Schroder, et al. (2008). "Effects of gender and moderate smoking on the pharmacokinetics and effects of the CYP1A2 substrate tizanidine." Eur J Clin Pharmacol **64**(1): 17-24.

Baecklund, F., J. N. Foo, et al. (2017). "Possible Interaction Between Cigarette Smoking and HLA-DRB1 Variation in the Risk of Follicular Lymphoma." American journal of epidemiology **185**(8): 681-687.

Bajaj, A., J. A. Driver, et al. (2010). "Parkinson's disease and cancer risk: a systematic review and meta-analysis." Cancer causes & control : CCC **21**(5): 697-707.

Baka, Z., E. Buzas, et al. (2009). "Rheumatoid arthritis and smoking: putting the pieces together." Arthritis Res Ther **11**(4): 238.

Bang, S. Y., H. S. Lee, et al. (2013). "Interaction of HLA-DRB1*09:01 and *04:05 with smoking suggests distinctive mechanisms of rheumatoid arthritis susceptibility beyond the shared epitope." J Rheumatol **40**(7): 1054-1062.

Bang, S. Y., K. H. Lee, et al. (2010). "Smoking increases rheumatoid arthritis susceptibility in individuals carrying the HLA-DRB1 shared epitope, regardless of rheumatoid factor or anti-cyclic citrullinated peptide antibody status." Arthritis Rheum **62**(2): 369-377.

Barranco Quintana, J. L., M. F. Allam, et al. (2007). "Alzheimer's disease and coffee: a quantitative review." Neurol Res **29**(1): 91-95.

Barres, R., J. Yan, et al. (2012). "Acute exercise remodels promoter methylation in human skeletal muscle." Cell Metab **15**(3): 405-411.

Boechat Nde, O., M. M. Ogusku, et al. (2012). "Interaction between smoking and HLA-DRB1*04 gene is associated with a high cardiovascular risk in Brazilian Amazon patients with rheumatoid arthritis." PloS one **7**(8): e41588.

Bower, J. H., D. M. Maraganore, et al. (1999). "Incidence and distribution of parkinsonism in Olmsted County, Minnesota, 1976-1990." Neurology **52**(6): 1214-1220.

Bowes, J. and A. Barton (2008). "Recent advances in the genetics of RA susceptibility." Rheumatology **47**(4): 399-402.

Bradburn, M. J., J. J. Deeks, et al. (1999). "metan-an alternative meta-analysis command." Stata Technical Bulletin **8**(44).

Buscariollo, D. L., X. Fang, et al. (2014). "Embryonic caffeine exposure acts via A1 adenosine receptors to alter adult cardiac function and DNA methylation in mice." PLoS One **9**(1): e87547.

Butt, M. S. and M. T. Sultan (2011). "Coffee and its consumption: benefits and risks." Crit Rev Food Sci Nutr **51**(4): 363-373.

Byrne, E. M., J. Johnson, et al. (2012). "A genome-wide association study of caffeine-related sleep disturbance: confirmation of a role for a common variant in the adenosine receptor." Sleep **35**(7): 967-975.

Chen, J. F., K. Xu, et al. (2001). "Neuroprotection by caffeine and A(2A) adenosine receptor inactivation in a model of Parkinson's disease." J Neurosci **21**(10): RC143.

Chinoy, H., S. Adimulam, et al. (2012). "Interaction of HLA-DRB1*03 and smoking for the development of anti-Jo-1 antibodies in adult idiopathic inflammatory myopathies: a European-wide case study." Annals of the rheumatic diseases **71**(6): 961-965.

Cornelis, M. C., E. M. Byrne, et al. (2015). "Genome-wide meta-analysis identifies six novel loci associated with habitual coffee consumption." Mol Psychiatry **20**(5): 647-656.

Cornelis, M. C., K. L. Monda, et al. (2011). "Genome-wide meta-analysis identifies regions on 7p21 (AHR) and 15q24 (CYP1A2) as determinants of habitual caffeine consumption." PLoS Genet **7**(4): e1002033.

Costello, S., M. Cockburn, et al. (2009). "Parkinson's disease and residential exposure to maneb and paraquat from agricultural applications in the central valley of California." Am J Epidemiol **169**(8): 919-926.

Curb, J. D., A. McTiernan, et al. (2003). "Outcomes ascertainment and adjudication methods in the Women's Health Initiative." Ann Epidemiol **13**(9 Suppl): S122-128.

Do, C. B., J. Y. Tung, et al. (2011). "Web-based genome-wide association study identifies two novel loci and a substantial genetic component for Parkinson's disease." PLoS genetics **7**(6): e1002141.

Dunning, M., N. Barbosa-Morais, et al. (2008). "Statistical issues in the analysis of Illumina data." BMC Bioinformatics **9**(1): 85.

Elbaz, A. (2016). "Prodromal symptoms of Parkinson's disease: Implications for epidemiological studies of disease etiology." Revue neurologique **172**(8-9): 503-511.

Elbaz, A., J. Clavel, et al. (2009). "Professional exposure to pesticides and Parkinson disease." Annals of neurology **66**(4): 494-504.

Facheris, M. F., N. K. Schneider, et al. (2008). "Coffee, caffeine-related genes, and Parkinson's disease: a case-control study." Movement disorders : official journal of the Movement Disorder Society **23**(14): 2033-2040.

Farooqui, T. and A. A. Farooqui (2011). "Lipid-mediated oxidative stress and inflammation in the pathogenesis of Parkinson's disease." Parkinsons Dis **2011**: 247467.

Ferrari, C. C. and R. Tarelli (2011). "Parkinson's disease and systemic inflammation." Parkinsons Dis **2011**: 436813.

Ferre, S., C. Quiroz, et al. (2008). "An update on adenosine A2A-dopamine D2 receptor interactions: implications for the function of G protein-coupled receptors." Curr Pharm Des **14**(15): 1468-1474.

Franco, R., S. Ferre, et al. (2000). "Evidence for adenosine/dopamine receptor interactions: indications for heteromerization." Neuropsychopharmacology **23**(4 Suppl): S50-59.

Friedrich, T. (2013) "Generalised Inference in the Random Effects Meta Regression Model."

Gagne, J. J. and M. C. Power (2010). "Anti-inflammatory drugs and risk of Parkinson disease: a meta-analysis." Neurology **74**(12): 995-1002.

Galanaud, J. P., A. Elbaz, et al. (2005). "Cigarette smoking and Parkinson's disease: a case-control study in a population characterized by a high prevalence of pesticide exposure." Mov Disord **20**(2): 181-189.

Gauderman, W. J. (2002). "Sample size requirements for association studies of gene-gene interaction." American journal of epidemiology **155**(5): 478-484.

Gebhardt, S. and R. Thomas. (2000). "Nitritive Value of Foods."

Gelb, D. J., E. Oliver, et al. (1999). "Diagnostic criteria for Parkinson disease." Arch Neurol **56**(1): 33-39.

Giri, A., S. R. Sturgeon, et al. (2011). "Caffeinated coffee, decaffeinated coffee and endometrial cancer risk: a prospective cohort study among US postmenopausal women." Nutrients **3**(11): 937-950.

Gunes, A. and M. L. Dahl (2008). "Variation in CYP1A2 activity and its clinical implications: influence of environmental factors and genetic polymorphisms." Pharmacogenomics **9**(5): 625-637.

Hamza, T. H., H. Chen, et al. (2011). "Genome-wide gene-environment study identifies glutamate receptor gene GRIN2A as a Parkinson's disease modifier gene via interaction with coffee." PLoS Genet **7**(8): e1002237.

Hamza, T. H., C. P. Zabetian, et al. (2010). "Common genetic variation in the HLA region is associated with late-onset sporadic Parkinson's disease." Nature genetics **42**(9): 781-785.

Harold, D., R. Abraham, et al. (2009). "Genome-wide association study identifies variants at CLU and PICALM associated with Alzheimer's disease." Nat Genet **41**(10): 1088-1093.

Hellenbrand, W., A. Seidler, et al. (1996). "Diet and Parkinson's disease. I: A possible role for the past intake of specific foods and food groups. Results from a self-administered food-frequency questionnaire in a case-control study." Neurology **47**(3): 636-643.

Hernan, M. A., B. Takkouche, et al. (2002). "A meta-analysis of coffee drinking, cigarette smoking, and the risk of Parkinson's disease." Ann Neurol **52**(3): 276-284.

Higgins, J. P. and S. G. Thompson (2002). "Quantifying heterogeneity in a meta-analysis." Statistics in medicine **21**(11): 1539-1558.

Higgins, J. P., S. G. Thompson, et al. (2003). "Measuring inconsistency in meta-analyses." BMJ **327**(7414): 557-560.

Hill-Burns, E. M., T. H. Hamza, et al. (2011). "An attempt to replicate interaction between coffee and CYP1A2 gene in connection to Parkinson's disease." European journal of neurology : the official journal of the European Federation of Neurological Societies **18**(9): e107-108; author reply e109.

Hoehn, M. M. and M. D. Yahr (1967). "Parkinsonism: onset, progression and mortality." Neurology **17**(5): 427-442.

Horvath, S. (2013). "DNA methylation age of human tissues and cell types." Genome Biol **14**(10): R115.

Houseman, E. A., W. P. Accomando, et al. (2012). "DNA methylation arrays as surrogate measures of cell mixture distribution." BMC Bioinformatics **13**: 86.

Hughes, A. J., S. E. Daniel, et al. (1992). "Accuracy of clinical diagnosis of idiopathic Parkinson's disease: a clinico-pathological study of 100 cases." J Neurol Neurosurg Psychiatry **55**(3): 181-184.

Jaffe, A. E. and R. A. Irizarry (2014). "Accounting for cellular heterogeneity is critical in epigenome-wide association studies." Genome Biol **15**(2): R31.

Kang, G. A., J. M. Bronstein, et al. (2005). "Clinical characteristics in early Parkinson's disease in a central California population-based study." Mov Disord **20**(9): 1133-1142.

Karlson, E. W., S. C. Chang, et al. (2010). "Gene-environment interaction between HLA-DRB1 shared epitope and heavy cigarette smoking in predicting incident rheumatoid arthritis." Ann Rheum Dis **69**(1): 54-60.

Kenborg, L., C. F. Lassen, et al. (2015). "Lifestyle, family history, and risk of idiopathic Parkinson disease: a large Danish case-control study." Am J Epidemiol **181**(10): 808-816.

Khoury, M. J. and W. D. Flanders (1996). "Nontraditional epidemiologic approaches in the analysis of gene-environment interaction: case-control studies with no controls!" American journal of epidemiology **144**(3): 207-213.

Lambert, J. C., C. A. Ibrahim-Verbaas, et al. (2013). "Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease." Nature genetics **45**(12): 1452-1458.

Levecque, C., A. Elbaz, et al. (2003). "Association between Parkinson's disease and polymorphisms in the nNOS and iNOS genes in a community-based case-control study." Hum Mol Genet **12**(1): 79-86.

LeWitt, P. A., M. Guttman, et al. (2008). "Adenosine A2A receptor antagonist istradefylline (KW-6002) reduces "off" time in Parkinson's disease: a double-blind, randomized, multicenter clinical trial (6002-US-005)." Ann Neurol **63**(3): 295-302.

Lincoln, M. R., A. Montpetit, et al. (2005). "A predominant role for the HLA class II region in the association of the MHC region with multiple sclerosis." Nat Genet **37**(10): 1108-1112.

McGeer, P. L., S. Itagaki, et al. (1988). "Reactive microglia are positive for HLA-DR in the substantia nigra of Parkinson's and Alzheimer's disease brains." Neurology **38**(8): 1285-1291.

McGeer, P. L. and E. G. McGeer (2008). "Glial reactions in Parkinson's disease." Movement disorders : official journal of the Movement Disorder Society **23**(4): 474-483.

Mellick, G. D. and O. A. Ross (2011). "Caffeine and Parkinson's disease: are we getting our fix on risk-modifying gene-environment interactions?" Eur J Neurol **18**(5): 671-672.

Miranda, D. M., B. L. Wajchenberg, et al. (2009). "Novel mutations of the BSCL2 and AGPAT2 genes in 10 families with Berardinelli-Seip congenital generalized lipodystrophy syndrome." Clin Endocrinol (Oxf) **71**(4): 512-517.

Morens, D. M., A. Grandinetti, et al. (1995). "Cigarette smoking and protection from Parkinson's disease: false association or etiologic clue?" Neurology **45**(6): 1041-1051.

Nalls, M. A., N. Pankratz, et al. (2014). "Large-scale meta-analysis of genome-wide association data identifies six new risk loci for Parkinson's disease." Nat Genet **46**(9): 989-993.

Narayan, S., Z. Liew, et al. (2013). "Household organophosphorus pesticide use and Parkinson's disease." Int J Epidemiol **42**(5): 1476-1485.

Narayan, S., J. S. Sinsheimer, et al. (2015). "Genetic variability in ABCB1, occupational pesticide exposure, and Parkinson's disease." Environ Res **143**(Pt A): 98-106.

Nefzger, M. D., F. A. Quadfasel, et al. (1968). "A retrospective study of smoking in Parkinson's disease." Am J Epidemiol **88**(2): 149-158.

Noyce, A. J., J. P. Bestwick, et al. (2012). "Meta-analysis of early nonmotor features and risk factors for Parkinson disease." Annals of neurology **72**(6): 893-901.

Ockinger, J., M. Hagemann-Jensen, et al. (2016). "T-cell activation and HLA-regulated response to smoking in the deep airways of patients with multiple sclerosis." Clin Immunol **169**: 114-120.

Palacios, N., X. Gao, et al. (2012). "Caffeine and risk of Parkinson's disease in a large cohort of men and women." Mov Disord **27**(10): 1276-1282.

Parkinson's Disease Foundation, I. (2017). "Statistics on Parkinson's." 2017, from http://www.pdf.org/parkinson_statistics.

Petronis, A. (2010). "Epigenetics as a unifying principle in the aetiology of complex traits and diseases." Nature **465**(7299): 721-727.

Pierri, M., E. Vaudano, et al. (2005). "KW-6002 protects from MPTP induced dopaminergic toxicity in the mouse." Neuropharmacology **48**(4): 517-524.

Ping, J., J. F. Wang, et al. (2014). "Prenatal caffeine ingestion induces aberrant DNA methylation and histone acetylation of steroidogenic factor 1 and inhibits fetal adrenal steroidogenesis." Toxicology **321**: 53-61.

Popat, R. A., S. K. Van Den Eeden, et al. (2011). "Coffee, ADORA2A, and CYP1A2: the caffeine connection in Parkinson's disease." European journal of neurology : the official journal of the European Federation of Neurological Societies **18**(5): 756-765.

Qi, H. and S. Li (2014). "Dose-response meta-analysis on coffee, tea and caffeine consumption with risk of Parkinson's disease." Geriatr Gerontol Int **14**(2): 430-439.

Quik, M., X. A. Perez, et al. (2012). "Nicotine as a potential neuroprotective agent for Parkinson's disease." Movement disorders : official journal of the Movement Disorder Society **27**(8): 947-957.

Ritz, B., A. Ascherio, et al. (2007). "Pooled analysis of tobacco use and risk of Parkinson disease." Archives of neurology **64**(7): 990-997.

Ritz, B., P. C. Lee, et al. (2014). "Parkinson disease and smoking revisited: ease of quitting is an early sign of the disease." Neurology **83**(16): 1396-1402.

Ross, G. W., R. D. Abbott, et al. (2000). "Association of coffee and caffeine intake with the risk of Parkinson disease." JAMA **283**(20): 2674-2679.

Saaksjarvi, K., P. Knekt, et al. (2008). "Prospective study of coffee consumption and risk of Parkinson's disease." Eur J Clin Nutr **62**(7): 908-915.

Sachse, C., J. Brockmoller, et al. (1999). "Functional significance of a C-->A polymorphism in intron 1 of the cytochrome P450 CYP1A2 gene tested with caffeine." Br J Clin Pharmacol **47**(4): 445-449.

Saiki, M., A. Baker, et al. (2010). "Association of the human leucocyte antigen region with susceptibility to Parkinson's disease." J Neurol Neurosurg Psychiatry **81**(8): 890-891.

Simon, K. C., I. A. van der Mei, et al. (2010). "Combined effects of smoking, anti-EBNA antibodies, and HLA-DRB1*1501 on multiple sclerosis risk." Neurology **74**(17): 1365-1371.

Stringhini, S., S. Polidoro, et al. (2015). "Life-course socioeconomic status and DNA methylation of genes regulating inflammation." Int J Epidemiol **44**(4): 1320-1330.

Sun, C., L. Wei, et al. (2012). "HLA-DRB1 alleles are associated with the susceptibility to sporadic Parkinson's disease in Chinese Han population." PLoS One **7**(11): e48594.

Taioli, E., C. Zocchetti, et al. (1998). "Models of interaction between metabolic genes and environmental exposure in cancer susceptibility." Environmental health perspectives **106**(2): 67-70.

Tan, E. K., Z. Y. Lu, et al. (2006). "Exploring an interaction of adenosine A2A receptor variability with coffee and tea intake in Parkinson's disease." American journal of medical genetics. Part B, Neuropsychiatric genetics : the official publication of the International Society of Psychiatric Genetics **141B**(6): 634-636.

Tanner, C. M. and S. M. Goldman (1996). "Epidemiology of Parkinson's disease." Neurol Clin **14**(2): 317-335.

Van Den Eeden, S. K., C. M. Tanner, et al. (2003). "Incidence of Parkinson's disease: variation by age, gender, and race/ethnicity." Am J Epidemiol **157**(11): 1015-1022.

Vanderweele, T. J., Y. A. Ko, et al. (2013). "Environmental confounding in gene-environment interaction studies." American journal of epidemiology **178**(1): 144-152.

Vangaveti, V., V. Shashidhar, et al. (2010). "Free fatty acid receptors: emerging targets for treatment of diabetes and its complications." Ther Adv Endocrinol Metab **1**(4): 165-175.

Vergnes, L., J. M. Lee, et al. (2013). "Diet1 functions in the FGF15/19 enterohepatic signaling axis to modulate bile acid and lipid levels." Cell Metab **17**(6): 916-928.

Victor Junji Yamamoto, V. d. J. R. d. P., Orestes Vicente Forlenza, Bernardo dos Santos, Daniel Shikanai Kerr (2015). "Association study in Alzheimer's disease of single nucleotide polymorphisms implicated with coffee consumption." Arch. Clin. Psychiatry **42**(3).

Walser, T., X. Cui, et al. (2008). "Smoking and lung cancer: the role of inflammation." Proc Am Thorac Soc **5**(8): 811-815.

Weinberg, C. R. (1986). "Applicability of the simple independent action model to epidemiologic studies involving two factors and a dichotomous outcome." American journal of epidemiology **123**(1): 162-173.

Wermuth, L., C. F. Lassen, et al. (2012). "Validation of hospital register-based diagnosis of Parkinson's disease." Danish medical journal **59**(3): A4391.

Wirdefeldt, K., H. O. Adami, et al. (2011). "Epidemiology and etiology of Parkinson's disease: a review of the evidence." Eur J Epidemiol **26 Suppl 1**: S1-58.

Wissemann, W. T., E. M. Hill-Burns, et al. (2013). "Association of Parkinson disease with structural and regulatory variants in the HLA region." American journal of human genetics **93**(5): 984-993.

Wu, D. M., Z. He, et al. (2015). "Increased DNA methylation of scavenger receptor class B type I contributes to inhibitory effects of prenatal caffeine ingestion on cholesterol uptake and steroidogenesis in fetal adrenals." Toxicol Appl Pharmacol **285**(2): 89-97.

Xu, K., Y. H. Xu, et al. (2010). "Neuroprotection by Caffeine: Time Course and Role of Its Metabolites in the Mptp Model of Parkinson's Disease." Neuroscience **167**(2): 475-481.