An Association Analysis of Gene-Environment Interactions in the PON Region in Late-Onset Sporadic Parkinson's Disease

Nirupama Shridhar

A thesis

submitted in partial fulfillment

of the requirement for the degree of

Masters in Public Health Genetics

University of Washington

2012

Committee:

Karen Edwards

Wylie Burke

Program Authorized to Offer Degree

Public Health Genetics

University of Washington

Abstract

An Association Analysis of Gene-Environment Interactions in the PON Region in Late-Onset Sporadic Parkinson's Disease

Nirupama Shridhar

Chair of Supervisory Committee:

Dr. Karen Edwards

Institute for Public Health Genetics (IPHG)

Objective: The PON region (7q21.3-7q22), comprising the three PON genes, PON1, PON2, and PON3, has been assessed for its role as a potential gene for risk susceptibility in Parkinson's Disease (PD). The gene products, the three different paraoxonase enzymes, hydrolyze organophosphates. Three environmental factors, namely, cigarette smoking, coffee, and non-steroidal anti-inflammatory drugs (NSAIDs) have shown to be inversely associated with risk of developing PD. No prior studies have evaluated Gene-Environment Interactions (GxE) between genes in the PON region and these established risk factors with PD. Further, this genetic epidemiologic project also raises questions about return of research results, particularly if evidence supports evidence for interactions between genetic variability and potentially modifiable environmental factors, such as cigarette smoking, coffee consumption and NSAID use.

Methods: Case-Control Association Study was conducted using data from the NeuroGenetics Research Consortium (NGRC) which includes 2000 PD cases and 1986 unrelated controls (n= 3986). The dataset, includes genotype information for the PON locus (n= 467) and three the environmental factors (cigarette smoking, coffee, and NSAIDs use).

Results: No interactions were statistically significant once the Bonferroni adjustment for multiple testing was made. However, of the top hit SNPs from each of the models (smoking, coffee and NSAIDs), two SNPs rs705379 (OR=1.58, 95% CI: 1.10, 2.11) and rs75071114 (OR = 0.61, 95% CI: 0.23, 1.58), are cSNPs of uncertain significance on PON2 and PON1 respectively. Conclusion: This is the first assessment of interactions between variants in the PON region and cigarette smoking, coffee, and NSAIDs. While the results of this study did not provide support for statistically significant interactions, it does raise additional issues about the need for dialog by the various stakeholders to formulate policy around offering return of individual research results for findings that are of uncertain clinical significance, and possessing more personal utility than clinical.

ACKNOWLEDGEMENTS

The author wishes to express sincere appreciation to the NeuroGenetics Research Consortium, Dr. Cyrus Zabetian, and Dr. Ignacio Mata for all their support and help through this project.

The author also wishes to acknowledge the support of Jia Yin (Angel) Wan, whose extensive knowledge and immense patience is much appreciated by the author.

The author wishes to acknowledge Dr. Karen Edwards, who has been a constant source of encouragement, support and whose vision has been the foundation on which this thesis rests.

Finally, the author would like to acknowledge that none of this would have been possible without the unconditional support provided by the author's family and friends.

DEDICATION

To my Dog, Pouchy.

Table of Contents

LIST OF FIGURES

LIST OF TABLES

Overview

The Brain is the Achilles heel of an ageing society. Loss of neuronal function is often regarded as a consequence of ageing. There are important public health implications for society in age-associated cognitive and motor declines. According to the World Health Organization, the proportion of people 60 and older is the fastest growing age group globally (World Health Organization, 2012). While this is a success story for public health policies that have helped reduce mortality from childhood and early-adulthood diseases, an older population poses challenges in terms of maximizing health outcomes and maintaining functional capacity.

This thesis is composed of two parts: Part A, a Genetic Epidemiology study of gene-environment interactions in the PON locus and their associations with the risk of developing Parkinson's Disease (PD). The hypothesis underlying Part A is that interactions between PON1 and the three environmental variables (smoking, coffee and NSAIDs) are associated with risk of developing PD. Part B, is an exploration of the issues surrounding the Return of Individual Research Results to individuals participating in genetic research studies of Parkinson's Disease.

Part A: Introduction

Parkinson's Disease

One such challenge is Parkinson's Disease (PD). Eponymously named for Dr. James Parkinson, a British doctor who first described the disorder in his book almost 200 years ago, it is a chronic and progressive neurodegenerative disorder of uncertain etiology and a prevalence of 1 % over the age of 65 (Jankovich, 2008).

Cardinal features of PD include tremors (of limbs and face), bradykinesia (slow movement), postural instability (impaired coordination and balance) and rigidity (of limbs and body) (Parkinson's Disease Foundation, 2012). As the disease advances, cognitive and behavioral problems also arise. The clinical features of the disease, both motor and non-motor, can be attributed to loss of neuronal function. It is canonically characterized by loss and impairment of vital neurons in the mid-brain or mesencephalon, in an area called the substantia nigra. This area is involved in the production of dopamine. Dopamine is a chemical signaling molecule transmitting signals for controlling smooth physical movement and loss of dopamine is tied to motor impairment. When the dopamine producing neurons die, communication between the brain and muscle is weakened and eventually the brain is unable to control muscle movement.

α-synuclein, a protein of largely unknown function and normally seen in the brain as an unstructured soluble protein, polymerizes to form insoluble fibrils (β-sheets) called Lewy bodies when protein concentrations are high or in a mutant form. In most cases of PD, Lewy bodies are seen in the dead and impaired neurons. Lewy bodies are structurally altered neurofilaments

thought to be similar to aggregosomes (proteinaceous inclusion bodies comprised of several different proteins including α-synuclein and ubiquitin), that form in a cell. Aggregosomes form in response to cellular stress, which causes impaired clearance of protein by the lysosomal system and the ubiquitin-proteasomal system. Whether Lewy bodies are causal to the disease or are associated with a protective cellular mechanism is as yet unknown. Cases without Lewy bodies are usually referred to as Parkinsonism, instead of PD.

Genetic research on PD is a recent entrant to research on PD. Although a genetic component to the disease was attributed only in the late 80's, there has been prolific research on the genetic mechanisms of the disease over the past two decades. There are rare Mendelian forms inherited in families and the more common, non-Mendelian, sporadic forms of PD that still cluster in families but without clear Mendelian segregation. Research has been successful in identifying specific loci and genes associated with Mendelian forms of PD and has provided a deeper insight into our understanding of both familial and non-familial forms of PD (Hardy, 2006). However, assigning causal roles to genetic determinants in non-Mendelian forms of PD have been scarce as is information needed to improve prevention, detection and treatment of PD (Lesage, 2009).

PD Etiology

PD is traditionally classified into two forms: Familial PD, which accounts for about 15% of all PD cases and Sporadic PD, of unknown etiologic causes. Although the pathoetiology for both forms are thought to be different, both forms have been shown to include a genetic component. For example, causal mutations for the familial form of PD have been identified in the PARK loci and include genes such as Parkin, PINK1, LRRK2, DJ1, ATP13A2, and SNCA genes (Shin, 2009). Mutations in these genes represent only a small proportion of all known PD cases. Most research on sporadic PD indicate that the etiology is a combination of genetic and

environmental factors, with some of the strongest evidence coming from twin studies that imply an environmental etiology or as yet unknown gene-environment interactions (Environmental Molecular Epidemiology Research (EMERG) Group, 2008). While it does appear unlikely that sporadic PD is caused by the deficiency of a single gene product and inherited in a Mendelian manner, about 25% of all patients do report having a relative with PD. Sporadic PD has been hypothesized to be caused by associated genes and susceptibility variants in those genes that confer greater risk when combined with other environmental factors. Idiopathic or sporadic PD is the second most common neurodegenerative disorder, second only to Alzheimer's Disease.

Age is an important risk factor in the development of PD. 96% of all cases are late-onset disease beginning after 50 years. Early-onset is mostly associated with familial inherited PD. Gender is another risk factor with men being one and a half times more likely to develop PD than women. PD has a prevalence of approximately 1 million cases in the United States and about 60,000 incident cases are diagnosed each year (National Institutes for Health, 2007). In spite of focused research, the pathoetiology of the disease is still elusive and relative contributions of genetic versus environmental factors are still in debate. However, it is known that by the time the disease manifests through the symptoms, more than 50% of dopamine neurons and 75% of striatal dopamine neurons are dead or impaired, suggesting an etiology that is well upstream of the symptomatic disease manifestation timeline (Collier, 2002) .

Environmental risk factors known to be associated with PD

Epidemiologic studies have evaluated many environmental risk factors for PD, factors that are associated with elevated risk such as exposure to pesticides and rural living (Barbeau, 1987). Studies have also identified factors that are associated inversely with risk such as coffee, smoking and non-steroidal anti-inflammatory drugs (NSAIDs). Coffee, Smoking and NSAIDs

have been found, over the course of many decades and in multiple studies, to reduce the risk of developing PD (Checkoway, 2002) (Samii, 2009). Whether this evidence represents true biologic protection is not entirely known but research that focuses on examining whether the correlations reflect true neuroprotective effects or pre-existing differences in the brains of PD patients that preclude them from consuming either continues. Coffee and tobacco have been shown to be protective in Drosophila models of PD (Trinh, 2010). Research has shown that caffeine-free coffee and nicotine-free tobacco confer the same neuroprotective benefits as caffeinated coffee and tobacco with nicotine in flies. Using two knockout fly models: one with the homozygous loss of function mutation in PARKIN and the other, a transgenic β-amyloid strain over-expressing alpha-synuclein protein in dopamine neurons, researchers were able to provide evidence for the neuroprotection conferred by using food supplemented with coffee and tobacco.

Coffee. Several scientific studies have looked at pharmacologic and health effects of coffee with respect to a wide variety of conditions. Some of them have been contradictory as to any specific health benefits or risks. In order to find consistent results, studies have to account for the fact that coffee is prepared in multiple ways that either increase or decrease the potency of the molecules in the beverage. Coffee is a compound of many molecules including diterpenes such as cafestol and kahweol, acids like caffeic acid and chlorogenic acid, polyphenols, and caffeine (an alkaloid molecule and a known stimulant of the central nervous system).

In a prospective longitudinal study conducted at the Honolulu heart program, researchers found that age-adjusted incidence of PD declined in relation to coffee intake, with progressively lower risk as coffee consumption increased, from 10.4 per 10,000 person-years in men who drank no coffee to 1.9 per 10,000 person-years in men who drank at least 28 oz/d (P<.001 for trend). Consumption of increasing amounts of coffee was also associated with lower risk of PD in men who were never, past, and current smokers at baseline (P=.049, P=.22, and P=.02, respectively, for trend) (Webster, 2000).

Similar results were found in a much larger prospective cohort of about 135,000 people comprising the Health Professionals Follow-Up Study (HPFS) and Nurses' Health Study (NHS). An inverse association was found with consumption of coffee, but not non-caffeinated coffee. Among men, after adjustment for age and smoking, the relative risk of Parkinson's disease was 0.42 (95% CI: 0.23–0.78; *p* for trend < 0.001) (Ascherio, 2001).

A systematic review of 8 case-control and 5 cohort studies summarized the epidemiologic evidence between the risk of PD and coffee-drinking. Compared with non–coffee drinkers, relative risk of Parkinson's disease was 0.69 (95% CI, 0.59–0.80) for coffee drinkers. The relative risk per three additional cups of coffee per day was 0.75 (95% CI, 0.64–0.86) in case–control studies and 0.68 (95% CI, 0.46–1.00) in cohort studies (Hernán, 2002).

In a Genome-Wide Association and Interaction Study (GWAIS) that was undertaken to identify genes that influenced the inverse association of coffee with the risk of developing PD, the most significant signal came from rs4998386 and nearby SNPs, housed in the GRIN2A gene, coding for a glutamate-receptor subunit and involved in regulating excitatory neurotransmission in the brain. In stratified GWAS, the GRIN2A signal was present in heavy coffee-drinkers (OR=0.43; P=6×10(-7)) but not in light coffee-drinkers (Hamza, 2011).

Other theories as to the mechanisms of action leading to the protection hypothesize that caffeine acts by blocking the Adenosine A_{2A} receptor in the brain which then acts as a barrier to neuronal damage (Schwarzschild, 2002). Recent evidence from knockout fly models implicate NRF2 antioxidative pathways in the neuroprotective effect conferred by coffee on the risk of developing PD (Trinh, 2010). There is considerable ongoing research in developing neuroprotective therapies using the adenosine antagonists and glutamate antagonist compounds. There are several potential biomarkers that may be identified as well from research of this nature.

Smoking (cigarettes). Cigarettes are a tobacco product. Although a cigarette has many chemicals and additives in both the tobacco and the filter, one of the key ingredients in tobacco is the stimulant and psycho-active chemical, nicotine. Nicotine is an alkaloid and constitutes approximately 0.6 - 3.0 % of the tobacco weight. It is capable of almost instantly crossing the blood-brain barrier when inhaled (Le Houezec, 2003).

One of the environmental risk factors shown to have an inverse association with risk of developing PD is smoking (Checkoway, 2002). In a case -control study nested within a prospective cohort study conducted on residents of a retirement community in Laguna Hills, researchers observed that having been a smoker in the past reduces the odds for developing PD by as much as 50% with even greater reduction in risk for continued smokers. The multivariate odds ratios (95% confidence intervals) were 0.42 (0.22–0.80) for current cigarette smokers of 1+ pack/day, 0.62 (0.48–0.80) , a 60 % reduction in risk relative to never smokers (Paganini-Hill, 2001). In a meta-analysis that looked at 44 case-control and 4 cohort studies, the risk for PD in smokers compared with never smokers was 0.59 (95% CI, 0.54–0.63) for ever smokers, 0.80 (95% CI, 0.69–0.93) for past smokers, and 0.39 (95% CI, 0.32–0.47) for current smokers. The relative risk per 10 additional pack-years was 0.84 (95% CI, 0.81–0.88) in case–control studies and 0.78 (95% CI, 0.73–0.84) in cohort studies (Hernán, 2002).

One interpretation of the results is that there are inherent biases that skew the results towards inverse risk associations. For instance, it could be hypothesized that the occurrence of a slowly progressing, long-standing disease may have an impact on addictive behaviors such as smoking. It may also be viewed that people who develop PD are predisposed in some way to not tolerate nicotine and therefore don't smoke. Data from the World War II Twins Cohort showed that twins (monozygotic and dizygotic) discordant for PD had significant differences in their pattern of smoking. In 33 discordant MZ pairs and 39 discordant DZ pairs in which at least one twin had smoked, the twins without PD smoked more than their brothers smoked (32.5 vs.

22.7 pack-years, $p = 0.026$). This was more marked in the MZ pairs (37.1 vs. 25.3 pack-years, $p = 0.077$) than in the DZ pairs (28.6 vs. 20.5 pack-years, $p = 0.17$). Since monozygotic twins are identical in their genomic makeup and may be seen as also having greater environmental similarities, differences are usually attributed to environmental influences (Tanner, 2002).

Rat models show reduced proneurotoxins in the brains of rats exposed to cigarette smoke solution. These are neurotoxin precursors that are either endogenous or exogenous. It is thought that by reducing the levels of rat brain proneurotoxins, 1,2,3,4-Tetrahydroisoquinoline (TIQ) and 1,2,3,4-tetrahydro-beta-carboline (THbetaC), smoking prevents neurodegeneration of the striatal dopaminergic neurons (Soto-Otero, 2001). In rat models, the capability of nicotine to prevent striatal dopamine loss triggered by lesions in the substantia nigra is dependent on the frequency and concentration of nicotine as well as the extent of the lesion (Costa G. A.-C., 2001). This is another promising area for drug development and treatment of PD.

NSAIDs. Non-Steroidal Anti-Inflammatory Drugs (NSAIDs) are non-narcotic drugs that are anti-inflammatory, analgesic and anti-pyretic. The three common OTC NSAIDs are Ibuprofen, Aspirin and Naproxen, all of which are phenolic acids by way of chemical composition.

NSAIDs show a similar inverse-risk relationship with PD. The role of inflammation and oxidative stress have been discussed as potential triggers to neurodegeneration (Trinh, 2010). The involvement of inflammatory pathways is also supported by PD models looking at neuroglia interaction and inflammatory processes (Esposito, 2007). The reduced risk observed in this group of NSAIDs users is consistent with a possible neuroinflammatory mechanism in the pathoetiology of PD.

In a large Danish Study using pharmacy registries, no evidence was found for the inverse risk-association between NSAIDs and PD among those diagnosed with PD and being NSAID users shortly before PD diagnosis. There was no difference in the risk with aspirin use (OR = 0.97; 95% CI 0.82, 1.14) or non-aspirin NSAID use (OR = 0.97; 95% CI 0.86, 1.09), regardless of intensity of use. Additionally, ibuprofen use also showed no associations (Manthripragada, 2011). However, it must be noted from the research that data on NSAIDs use was tracked only from 1995 and all PD diagnoses in the study was made between 2001 and 2006. The study captured chronology of possible NSAIDs use ranging between 5-11 years before diagnosis. Inflammation is a long-term process and the NSAIDs use looked at in this dataset may not have captured upstream inflammatory processes set into motion many years before PD became symptomatic.

A meta-analyses of 11 observational studies on NSAIDs use and risk of PD, found that NSAIDs, as a class of drugs, do not modify the risk of PD. The only evidence in this study that was protective was ibuprofen use. The pooled risk ratio of PD with NSAID use was 0.95 (95% CI 0.80, 1.12). The pooled risk ratio of PD with high-dose or long-duration NSAID use was 0.91 (95% CI 0.78, 1.05). The pooled risk ratio of PD for aspirin (acetylsalicylic acid) users was 1.08 (95% CI 0.93, 1.26). The pooled risk ratio of PD among ibuprofen users was 0.76 (95% CI 0.65, 0.89). The pooled risk ratio of PD in men using NSAIDs was 0.79 (95% CI 0.69, 0.92), and in women using NSAIDs, it was 0.72 (95% CI 0.45, 1.15) (Samii A. E., 2009).

The association between NSAIDs use and PD is weaker than the other two environmental factors, coffee and smoking. This may have more to do with the classes of drugs that fall under the umbrella of NSAIDs. They are distinct in their operational parameters and the biologic pathways they act on. Yet they are all NSAIDs for information-gathering purposes. Additionally, recall bias may have a greater role to play in this environmental variable. Coffee and smoking are lifestyle choices that make recollection of quantification easier. Whereas

recollecting the dose, duration and type of NSAIDs accurately is a challenging task, especially when the chronology that is required in the recollection is so extensive.

Joint effect of all three environmental factors and Risk for PD. Each of the environmental factors have been looked at separately in order to assess their effect on the risk for PD. A study that looked at joint effects, for individual, two-way and three-way combinations of these factors found significant reductions in risk. For the individual factors, the reduction in risk was 20% - 30%. For the two-way combination of factors, the reduction in risk was 37-49%. The three-way combination of factors showed 62 % reduction in risk, with the greatest reduction coming from those who were stratified at the higher end of consumption for all three environmental factors - an 87% reduction in risk (Powers, 2007). The results from this study tie in to the dose-response relationship model that was observed in assessing each of the individual factors to their PD risk.

In addition to the established risk factors described above, there are also known genetic effects and likely GxE. There is research being conducted on the entire range of possible susceptibility genes involved in Parkinson's Disease including UCHL1, GBA (Maraganore, 2004), (Almeida, 2012). The PON region, PON1 in specific, has also been evaluated as a possible susceptibility gene in PD

PON Genes

Environmental insults are thought to play a role in PD etiology .The PON region has been evaluated as a potential candidate gene in PD on account of its role in metabolizing environmental organophosphates This is an area that is still being evaluated and studied, more so because there are contrasting reports in assessing the role of the PON variants in PD.

The PON cluster of genes is a group of three genes, PON 1, PON2 and PON3, located on the q arm of chromosome 7, at 7q 21.3 - 22.1 and spanning ~ 136 kb. PON1 is the most centromeric and PON2, the least. PON is a serum enzyme and the gene cluster codes for the expression of three distinct forms of the enzyme Paraoxonase, an aryl esterase, involved in the hydrolysis of organophosphates (Sergio L. Primo-Parmo., 1996) . Increasing evidence from research in the last decade attribute the gene products from this cluster to also be involved in guarding against cellular damage and oxidative stress caused by toxic organophosphates (Liang, 2003). PON 1 and PON3 are similar in activity but vary in substrate specificity. PON2 is a ubiquitously expressed anti-oxidant protein that hydrolyses lactones and also acts as bacterial quorum-sensing mediators with a potential role in defending against pathogenic bacterial infections (NCBI, 2012) , (Costa, 2004).

Some of the functional SNPs have been evaluated for association with PD. PON1 polymorphism codon 192 (SNP rs662) was found to be associated with a higher risk of developing PD (Kondo, 1998). Another study in a Finnish population, found no risk associated between the functional polymorphisms in PON1 and sporadic PD (Clarimon, 2004) . A recent meta-analysis of 12 studies found no evidence associating the functional polymorphisms L55M (rs854560) and Q192R (rs662) with PD (Ying-Li, 2012).

PON locus polymorphisms have also been found associated in studies looking at Alzheimer's Disease. The evidence from the study supported the link between variants in PON contributing to AD risk (Erlich, 2006).There is no evidence yet suggestive of a causal link between PON and Alzheimer's Disease. But since Alzheimer's Disease and PD are both neurodegenerative diseases and present common features such as dementia and the insoluble amyloidal protein plaques in the neurons, any evidence of an association in the PON region in Alzheimer's may indicate as yet unknown role in the etiology of PD. Polymorphisms in the PON gene family are also associated with Clopidogrel efficacy. Clopidogrel is an anti-platelet agent

used in preventing strokes and heart attacks. Clinical efficacy of the drug is tied into Paraoxonase 1 which bioactivates the drug. The polymorphism Q192R was found associated with the rate of formation of the active metabolite (Bouman, 2010), (Pare, 2012).It is possible that functional polymorphisms in the PON family that alter the activity of Paraoxonase might contribute to altered effects of environmental factors that act on neurodegenerative pathways.

GxE as a Model for Parkinson's Disease Research

Sporadic PD is a multi-factorial disorder, caused by the combinatory effects of multiple genes, environmental factors and impacted by lifestyle choices. While PD clusters in families, it is difficult to assess risk based on just inheritance. The nature of multi-factorial diseases makes it unlikely that all causal mechanisms are discernible through research conducted only on genetics. Hence, the importance of Gene-Environment Interaction studies (GxE) . In a GxE model, the risk of developing PD is modulated by a cumulative and interactive effect of genetics and exposures. They are tools to help discover undetected risk relationships that might exist between genetic predisposition and non-genetic environmental factors that will not be detected in a Genome Wide Association Studies (GWAS). A study that looked at the effects of 4 candidate susceptibility genes: synuclein(*SNCA*) promoter polymorphism REP1, microtubule-associated protein tau (*MAPT*) H1/H2 haplotypes, apolipoprotein E (*APOE*) _2/_3/_4 polymorphism, ubiquitin carboxyl-terminal esterase L1 (*UCHL1*) S18Y variant along with two exposures (cigarette smoking and caffeinated coffee) on PD, found strong evidence for GxE. Two novel interactions were detected: *APOE* with coffee ($P = 0.005$), and REP1 with smoking ($P = 0.021$). While the individual interactive effects of genes and exposures were modest, each yielding OR < 1.6, the effects were cumulative, with some combinations reaching OR = 12.6 (95% CI: 5.9–26.8) (McCulloch, 2008).

Rationale for the GxE on PON. Each of the environmental factors included in this project has consistently been shown to be associated with a reduced risk of developing PD, and are thought to interact with other genes. The PON locus, by virtue of its function in metabolizing organophosphates is a Parkinson's susceptibility locus and has been researched as a potential candidate gene for PD. There is scant data from human studies on GxE in the PON region influencing susceptibility to PD. The PON cluster of genes has been shown to be strongly associated with metabolism of pesticides, some of which are known to cause Parkinsonism-like symptoms. PON, as the main enzyme involved in the metabolism of organophosphates has also been researched as a potential candidate gene for PD. Past research on the three environmental exposures (in PD) have not yielded results that are clinically noteworthy yet. The GxE model is designed to detect novel interactions that will not rise to significance in a GWAS study on the same.

It is plausible to theorize that a gene product having organophosphates as its substrate might be linked to the three organic environmental exposures (although none of them are organophosphates), directly or indirectly via biotransformative processes in the body to the interactions seen between the environmental covariates and inverse risk for developing PD.

One notion on risk modeling in PD is that cumulative risk factors having modest individual effects play an important role in the etiology of the disease. PON has not been evaluated for possible interactions with the three known environmental factors associated with reduced risk in PD, however all have been implicated in PD. The focus of this analysis is to evaluate interactions between PD and the three environmental exposures; smoking, caffeine and NSAIDs.

Materials and Methods

Subjects and Data Collection in the NGRC dataset

The NeuroGenetics Research Consortium (NGRC) is a multi-center study of genetic and environmental risk factors in PD. It began in Oregon as a genetic study of PD in 1996 and grew to encompass collaborators in Washington (2002), New York (2004), and Georgia (2005), all housed in various academic, federal and state settings. The Consortium is currently comprised of eight movement disorder clinics in the four states. The genetic component of the study has been funded by the NIH since 1998. In 2004, the NGRC was initiated into the Michael J Fox Foundation as a Global Genetic Consortium funded by the Foundation. Created in 2004, the study's epidemiology arm introduced the environmental exposure questionnaire to the study (dbGaP, 2012).

The NGRC dataset aims to study the gene-environment effects on PD . The study was approved by the Institutional Review Boards of the various Institutions that form a part of the Consortium. The NGRC study has a large number of samples in the dataset and it has been the source of data for dozens of publications thus far.

The Recruitment Process. All subjects were recruited at one of the eight NGRC-affiliated movement disorder clinics from among the four states. NGRC includes eight movement disorder clinics in four states, led by Drs. Stewart Factor (Emory University), John Nutt (Oregon Health & Sciences University), Cyrus Zabetian (University of Washington and Puget Sound Veterans Medical Center), Eric Molho (Albany Medical College), and Donald Higgins (Albany Veterans Medical Center). NGRC's molecular and statistical genetics laboratories are at New York State Department of Health (Haydeh Payami) and Puget Sound

Veterans Medical Center (Cyrus Zabetian) (dbGaP, 2012). The study collected tissue samples from cases (n=2013) and controls (n=1995). Approximately 85 % of the cases and 85 % of the controls who were invited to volunteer in the study consented to be a part of the process (McCulloch, 2008). Patients were enrolled sequentially, as and when they agreed to participate in the study. There was no preference in enrollment based on age of onset or family history of disease.

The subjects were all Caucasian and self-identified as white-Americans or white-Europeans. This was done in order to minimize confounding caused by population substructure in genetic association studies. In addition, subjects missing data on any of three covariates (site of recruitment, age (at blood draw) or sex), patients with onset before age 21 and patients with age of blood draw before age 20 were also excluded.

In order to qualify as a control in the study, a participant had to be willing to sign the informed consent forms to participate in research and donate tissue for DNA extraction. They also needed to be genetically unrelated to the cases and to each other. Controls were spouses and community volunteers genetically unrelated to the patients. The controls did not fill out a standardized family history questionnaire and only those recruited after 2004 filled out the Environmental Exposure Questionnaire (EEQ). The controls also had to be free of neurodegenerative disease by self-report or exam. Excluded diseases include Alzheimer's, Bipolar Disorder, Multiple Sclerosis, Amyotrophic Lateral Sclerosis, Ataxia, Dystonia, Parkinson's, Autism, Dementia, Epilepsy, Stroke and Schizophrenia.

In order to qualify as a case in the study, a participant had to be willing to sign the informed consent forms to participate in research and donate tissue for DNA extraction. It was also required that cases be completely genetically unrelated to each other and to the controls being recruited. All cases filled out a standardized family history questionnaire and those

recruited after 2004 needed to also fill out the Environmental Exposure Questionnaire (EEQ). Finally, in order to be considered a case, a subject needed to have had a confirmed diagnosis of PD by a neurologist, as diagnosed using the U.K Brain Bank Diagnostic Criteria (dbGaP, 2012).

The Diagnostic Process. Lewy bodies in the brain are the pathological hallmark of PD. However, there are no tests currently available to confirm the presence of the biomarker (Lewy body) in the Substantia nigra, short of an autopsy. This makes diagnosis of PD rather challenging. The U.K Parkinson's Disease Society Brain Bank Clinical Diagnostic Criteria is the gold standard for diagnosis of PD in the absence of any biomarker-based tools (Gibb, 1988). The criteria follow a three-step process.

The First step is to evaluate the symptoms as being Parkinsonian . The presence of bradykinesia (slowness of voluntary movement) along with at least one of three of the following symptoms: muscular rigidity, resting tremor and postural instability (not caused by loss of cerebellar, vestibular, visual dysfunction), is collectively used to make that diagnosis. The first step establishes the patient under the Parkinsonian umbrella of diseases.

The Second step is an exclusion step, to exclude those diagnoses that have a non-PD etiology. This includes assessing the symptoms for other causative sources such as exposure to the neurotoxin precursor MPTP (1-methyl-4-phenyl-1,2,3,6,tetrahydropyridine), history of stroke, head injury or encephalitis, dementia, cerebral tumors, Babinski's sign (signaling the presence of disease or lesions in the spinal cord and brain)and a negative response to Levadopa treatment (if malabsorbtion has been excluded). Traditionally, those with a familial form of PD are excluded under this criteria on grounds that their PD is not truly idiopathic. The NGRC dataset modified this criteria so as not to exclude subjects with a positive family history of PD. The second step serves to eliminate all those patients whose symptoms have other

etiologic sources.

The Third step, then classifies the remaining patients as having Parkinson's Disease. The inclusion criteria requires any three of the following: unilateral onset, rest tremors, progressive disorder, persistent asymmetry affecting the side of onset most response (70-100%) to levodopa, severe levodopa-induced chorea, levodopa response for 5 years or more, and a clinical course of these symptoms charted for over 10 years.

The stringency of the diagnostic criteria ensures that a diagnosis of PD is validated through multiple steps. Additionally, in the NGRC dataset, the mean difference in the dataset between age at diagnosis and age at blood draw (8.36 years) adds another layer to the accuracy of the diagnosis. In the unlikely event of a misdiagnoses, given the difference in mean years between the two variables, it is likely that the misdiagnosis surfaced well before any participation in the study.

The Subset of Data for this Analysis. For the purpose of this analysis, I am using a subset of SNPs, both genotyped and imputed (n=510) focused on the PON cluster of genes. All the subjects who participated in the NGRC study are represented in the controls (n=1986) and cases (n=2000). On account of the modified U.K Brain Bank Clinical Diagnostic Criteria, the data does include a few individuals with mutations known to confer causal risk in PD. e.g. The G2019S mutation in LRRK2 (n=1 homozygote AA and n=24 heterozygote GA) and includes a few cases (n=25) diagnosed at a younger age of between 25-35. However, these are very small numbers and unlikely to have much of an impact on the data. . Since controls did not fill any family history questionnaire, the analysis does not take into account any susceptibility the controls may have based on family-history of PD.

Genotyping and Molecular Analysis

Unamplified DNA was collected from whole blood at concentrations ≥ 50 ng/µl via standard methods of peripheral blood extraction. Genotyping was done using Illumina Human Onmi1 v1-0B SNP chip. The chip was customized for optimal tag SNP content and genome coverage. The chip provided genotyped information on 1,051,295 SNPs evenly spaced on the genome for a genome-wide scan of observed genotypes. Additionally, the dataset also has PD-associated gene mutations genotyped to exacting specifications, in order to distinguish between haplotypes (MAPT SNP rs1800547), repeat lengths (SNCA REP1), isoforms (APOE) and alternate alleles (UCHL1 and LRRK2) (McCulloch, 2008).

For the purpose of this analysis, SNPs localized in the PON1, 2 and 3 region (7q21.3-7q22) were both genotyped and imputed (n=510) to focus on the PON cluster of genes. A total of 292 SNPs were imputed and 218 SNPs were genotyped.

The coverage of the genotyped SNPs ranged from GRCh37/hg19 7-94912404 (genotyped to location 7-94750340 from earlier build NCBI36/hg18) to GRCh37/hg19 7-95082010 (genotyped to location 7-94919946 from build NCBI36/hg18), spanning a segment of 169.670 kb. The average coverage of the PON gene cluster by the genotyped SNPs was 778bp/SNP. This value, by no means denotes equidistant coverage 778 bp apart. It just approximates a value based on total chromosomal segment length and number of SNPs representing that segment.

In order to allow better coverage of the region and to detect any variability not secured by the genotyped and observed tag SNPs, additional SNPs were imputed in the PON reference panel using Markov Chain Monte Carlo (MCMC) algorithm in the Impute2 program (Impute2,

2012). It is a structurally sound method that reconstructs the genotype distribution given a set of referent haplotypes and the observed, genotyped data. It is limited by the fact that it can only approximate target distributions and is less valid than an observed genotype.

All SNPs in the region 94750kn to 94920kb on Chromosome 7 (build 36.3) containing PON1, 2, and 3 SNPs were then imputed using HapMap3 as well as the 1000 Genomes European panel and the software program Impute2.Genotypes were called if the maximum genotype probability was greater than 80%; otherwise genotype for the individual was set to missing. A total of 510 SNPs were imputed and summarized in the genotype summary file but 43 SNPs were dropped from the dataset because the missing rate was >0.05.The final dataset contains a total of 467 SNPs for 3986 individuals.

The coverage of the imputed and genotyped SNPs ranged from chromosomal basepair location 7-94912311 in GRCh37/hg19 (genotyped to location 7-94750247 from earlier build NCBI36/hg18) to chromosomal basepair location 7-95082010 (genotyped to location 7-94919946 from build NCBI36/hg18), spanning a segment of 169.699 kb. The average coverage of the PON gene cluster by the imputed and genotyped SNPs altered to 363.39 bp/SNP.

In order to ensure quality control of the imputation process, additional exclusionary criteria were followed. The imputed SNPs were evaluated for imputation certainty based on the value of the information metric in order to ensure that the imputed value was a product of the algorithm and not an artifact of the imputation process. The Information metric is calculated based on the ratio of empirical and observed variance in expected allele counts. Values closer to 1 indicate better quality of imputation. Imputation certainty is dependent on the minor allele frequency (MAF). Imputed SNPs were evaluated by their information metric, in which values between 0.3-0.5 were flagged as those SNPs with moderate imputation certainty and lower acceptable imputation quality.

Environmental Exposure Information

Environmental exposure information was first collected in 2004, with the advent of the epidemiologic arm of the study. Collection of genetic samples, however, began earlier in 1996. As a result, exposure assessment data is limited to those subjects enrolled after 2004, especially in Oregon (Powers, 2007). Overall, there is exposure information for about 80% of the cases and controls in the dataset. The Environmental Exposure Questionnaire (EEQ) was a standardized self-administered questionnaire. Apart from age at environmental data collection, the EEQ collected information on five categories of environmental variables and are summarized in Table 1.

Tea and soda consumption were collected in a single question each, as an ever/never answer category. They were both collected in reference to their caffeine content. The questionnaire specifically refers to only caffeinated soda and caffeinated tea as part of the data gathering process.

The other three key environmental variables: Smoking, Coffee and NSAIDs had multiple questions gathering information on a variety of factors on each of those variables. Information was collected on the primary ever/never to each of these variables, consumption load, status (on current smoking) and threshold values denoting low or high loads. NSAIDs was further subdivided into OTC and prescription categories. In view of the fact that NSAIDs is an umbrella term and includes many medications that are hard to recollect going back in time as the EEQ did, a list of OTC and prescription NSAIDs was presented to the subjects during EEQ, as well as a reminder that Aspirin and Acetaminophen were not classified as NSAIDs. Start and stop ages were obtained on the other two environmental variables to get an idea of the duration of exposure, but NSAIDs use is as-needed based on health requirements and sporadic during a

person's lifetime. Therefore, no start and stop ages were collected for both categories of NSAIDs use (Powers, 2007).

The Subset of Environmental Exposure Variables for this Analysis. The primary exposures of interest are Smoking, Coffee and NSAIDs. From among the many options the data presented, the analysis used only the ever/never option as it was the most complete and consistent data across all over variables. In other words, the variables collected answered the question: did the subject ever consume coffee, did the subject ever smoke and did the subject ever used NSAIDs. Further the NSAIDs were limited to only OTC.

Statistical Tools and Methods

The data was used to evaluate potential interactions between each of the three environmental exposures and the SNPs in PON region with respect to the risk they conferred on PD susceptibility.

We first evaluated Hardy-Weinberg Equilibrium (HWE) in the control group using PLINK for all SNPs, both genotyped and imputed. After adjustment for multiple comparisons using a Bonferroni correction, none of the SNPs were found to deviate significantly from HWE.

STATA version 12 (STATA, 2012) was the software used for most of the data analysis. and R (R Development Core Team, 2012) was used for generating Manhattan plots. The two datasets (PON reference panel information and covariate information) were combined after genotyped and imputed SNPs data were processed for quality control checks on SNPs for missingness, duplicates, minor allele frequency (MAF) and Hardy-Weinberg equilibrium (HWE) and imputation quality.

Logistic regression was the method of choice used in this analysis. It was well-suited to evaluate potential GxE interactions and risk of PD and it also allowed for adjustment of potentially confounding factors such as age at blood draw, and sex. Specifically, for each environmental exposure, the model included age at blood draw, sex, an additive SNP effect, the binary environmental factor, and an interaction term (environmental factor*SNP). The adjustments allow for minimizing the underlying variability in the data as well as adjust imbalanced baseline variables known to be related to the outcome. Models looking at two and three exposure combinations were beyond the scope of this analysis

Because we evaluated 467 SNPs for each environmental exposure, Bonferroni Correction was used to adjust for multiple comparisons generated by the association analysis (α =.05, n= 467 SNPs).Threshold for significance or α/n = .0001 or 1×10^{-4} for each environmental exposure.

Since the genotyped data was initially obtained on NCBI Build 36.3, all chromosomal locations map to that build. NCBI Build 37.3, released in October 2011, is the newer version and annotates to different chromosomal locations as compared to Build 36.3. In order to standardize the genomic location of the SNPs, all SNPs that were significant were converted to Build 37.3 and all genes and locations mapped to the SNPs use the newer Build using UCSC's Genome Browser (UCSC). The old build was used to map the Manhattan plot in order to maintain integrity of this analysis with previous analyses performed prior to my thesis.

Results

The descriptive and demographic characteristics (Table 1) reveal that the mean age of blood draw is greater in controls than in cases. This was intentional, in order to eliminate the possibility that some of the controls cold be pre-symptomatic cases if their age at blood draw was lower than that of the cases. The proportion of subjects in the ever/never category for each of the environmental variables is similar between cases and controls for the most part. The median values for the load (e.g. in the case of smoking this would be pack years) between cases and controls are similar in smoking , slightly different for NSAIDs OTC, and quite different in coffee.

Table 1: Descriptive Demographics

Descriptive Demographics			
COVARIATES	TOTAL (N=3986)	CASES (N=2000)	CONTROLS (N=1986)
age blood draw (mean, sd)	68.78, 12.59	67.26 , 10.67	70.32 , 14.09
age blood draw (p25, median, p75)	60, 69, 78	60, 68.5, 75	60, 71, 83
sex (% male)	53.06%	67.30%	38.72%
smoking (% ever)	46.00%	45.80%	46.21%
smoking load (p25, median, p75)	0, 0, 7.5	0, 0, 9	0, 0, 5
coffee (% ever)	86.64%	86.55%	86.77%
coffee load (p25, median, p75) **	22.5, 60,108	21, 57, 93	30, 67.5, 125.45
tea (% ever)	63.19%	59.86%	68.47%
soda (% ever)	79.51%	79.62%	79.33%
NSAID OTC (% ever)	64.46%	62.16%	66.83%
NSAID OTC load (p25, median, p75)* & **	0.75, 2.7, 5.25	0.75, 3, 6	0.75, 2.25, 4.5
NSAID Rx (% ever)	34.11%	33.29%	35.41%
NSAID Rx load (p25, median, p75)* & **	0.25, 1.25, 4	0.25, 1.2, 4	0.27, 1.35, 4

* times/day for lifetime use
** load calculation only calculated for those that have ever taken

The Locus Information for 7q21-22 (Table 2) maps out the PON region that we focus the investigation on. It is a relatively small region and the PON genes are clustered together. There is also a pseudogene spanning 1604 bp between PON1 and PON3.

Table2: Locus Information for 7q21-22 containing the PON region (Build 37.3)

Gene	PPP1R9A	PON1	Pseudogene	PON3	PON2	ASB4
Size	388,779	26216	1604	36504	30211	54333
From (bp)	94536949	94927669	94979476	94989184	95034174	95115213
To (bp)	94925727	94953884	94981079	95025687	95064384	95169543
Strand		Minus	None	Minus	Minus	

The results from the logistic regression models showed no significant evidence for interaction for any of the three environmental factors (coffee, smoking and NSAIDs) in the PON region in PD, after adjusting for multiple comparisons. Shown below are the results for each of the environmental exposures presented separately, as well as a summary of all top hits (Table 6), . Manhattan plots were also generated for each environmental exposure to better visualize the gene-environment interactions associated with PD in the PON region. The plots depict the $-\log_{10}$(p-value) along y-axis and the basepair location (Build 36) on the x-axis. The low interactions did not meet the significance threshold of $-\log_{10}$(p-value) >4.

Coffee

Coffee*PON interaction was the most significant of the three environmental factors that were a part of this analysis. In Table 3, the PON variant that displayed the strongest statistical association in the coffee association with PD was a non-coding polymorphism rs17876205, an intergenic SNP between PON3 and PON2 (OR = 7.78, 95% CI 1.67 - 36.21, *P* = 0.008). While the effect size, as denoted by the odds ratio was very high, the p-value did not rise to the pre-determined level of significance (p=.0001, the equivalent of α=.05 after multiple comparisons).

Three other SNPs that were among the top hit SNPs in the coffee interaction model were all intronic SNPs in PON2. Rs2237585 (OR = 1.58, 95% CI 1.12-2.23, *P* = 0.008),

rs2375005 (OR = 1.57, 95% CI 1.1-2.22, *P* = 0.009), and rs9640633 (OR = 1.57, 95% CI 1.11-2.22, *P* = 0.009).

Table 3: Top Hits for Interaction Model between coffee (ever/never caffeinated coffee) and PON SNPs associated with PD: Adjusted for age at blood draw and sex

SNP	N	OR	OR_LOW95	OR_UP95	Z	PVAL	GENE	POSITION (BUILD 37)	LOCATION
rs2237585_C* coffee_use	2620	1.58	1.12	2.23	2.63	0.008	PON2	Chr 7, 95049818	Intron
rs17876205_G * coffee_use	2581	7.78	1.67	36.21	2.61	0.008	PON3 & PON2	Chr 7, 95033104	Intergenic
rs2375005_A* coffee_use	2615	1.57	1.12	2.22	2.60	0.009	PON2	Chr 7, 95036876	Intron
rs9640633_G* coffee_use	2620	1.57	1.11	2.22	2.60	0.009	PON2	Chr 7, 95036897	Intron
rs987539_T* coffee_use	2620	1.57	1.11	2.22	2.60	0.009	PON2	Chr 7, 95036992	Intron
rs705379_G* coffee_use	2564	1.52	1.10	2.11	2.57	0.010	PON1	Chr 7, 94953895	CSNP

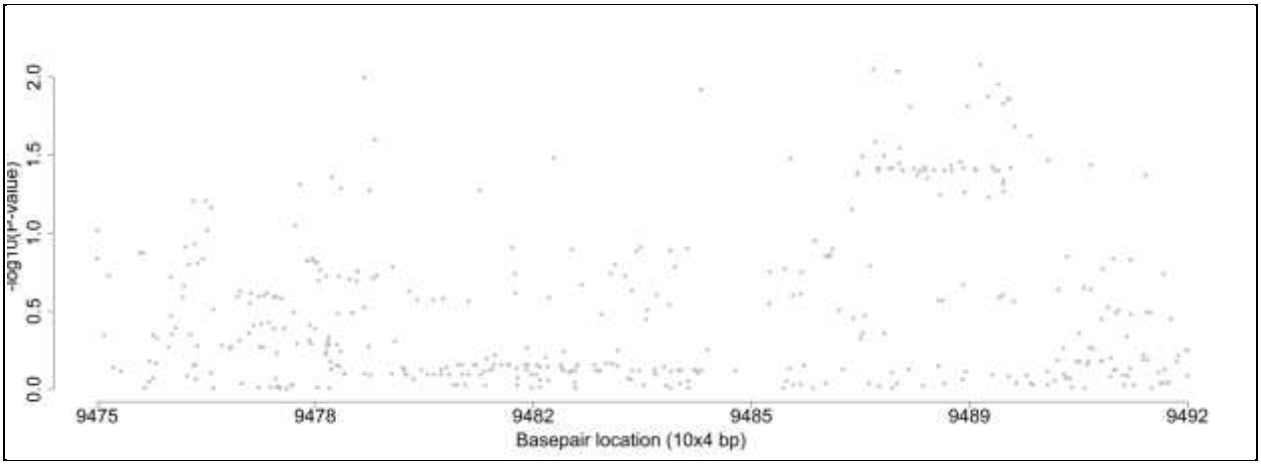

Figure 1: Coffee*SNP Interaction

As seen in the plot, none of the PON region SNP *coffee interactions reach the threshold of significance at 1 x $10^{-4,}$ even though the coffee model has the highest values of all three environmental covariates modeled. There are a few SNPs that rise to 10^{-2}. These formed the top hit SNPs for the coffee model. The PON variant in Table 4 that showed the strongest association in the smoking association with PD was rs75071114, a SNP in the coding region of

the region of the gene (cSNP), on PON2 (OR = 0.61, 95% CI = 0.23-1.58, *P* = 0.309). This SNP was imputed for the PON reference panel and at the time of imputation, in NCBI Build 36.3 was not yet designated a SNP and was a variable (v419) that mapped to chromosomal location -94903143. When the values were converted to Build 37.1, v419 mapped to 7-95065207, using 1000 Genomes as the reference source for the minor allele frequency in the SNP. In this case too, the clinical significance of the cSNP is as yet uncertain.

Smoking

The PON region SNP* smoking interactions, at their highest, range from -log10 (p-values) = 1.2 - 1.4. Again, this does not reach close to the level of significance as determined by the Bonferroni correction.

Table 4: Top Hits for Interaction Model between smoking (ever/never 100 cigarettes) and PON SNPs associated with PD: Adjusted for age at blood draw and sex

SNP	N	OR	OR_LOW95	OR_UP95	Z	PVAL	GENE	POSITION(BUILD 37)	LOCATION
rs75071114_A*smk_use	3058	0.61	0.23	1.58	1.01	0.309	PON2	Chr 7, 95065207	CSNP
rs62469566_A* smk_use	3091	0.65	0.23	1.78	0.82	0.407	PON2	Chr 7, 95037208	Intron
rs62467349_T* smk_use	3074	0.7	0.49	0.98	2.03	0.042	PON1	Chr7, 94946795	Intron
rs10240398_T* smk_use	3097	1.08	0.87	1.34	0.71	0.474	PON2 and ASB4	Chr 7, 95082010	Intergenic
rs2299258_A* smk_use	3049	0.71	0.50	0.99	1.99	0.046	PON1	Chr7, 94942917	Intron
rs7794650_A* smk_use	2957	0.30	0.09	0.98	1.98	0.047	PON2	Chr7, 95061804	Intron

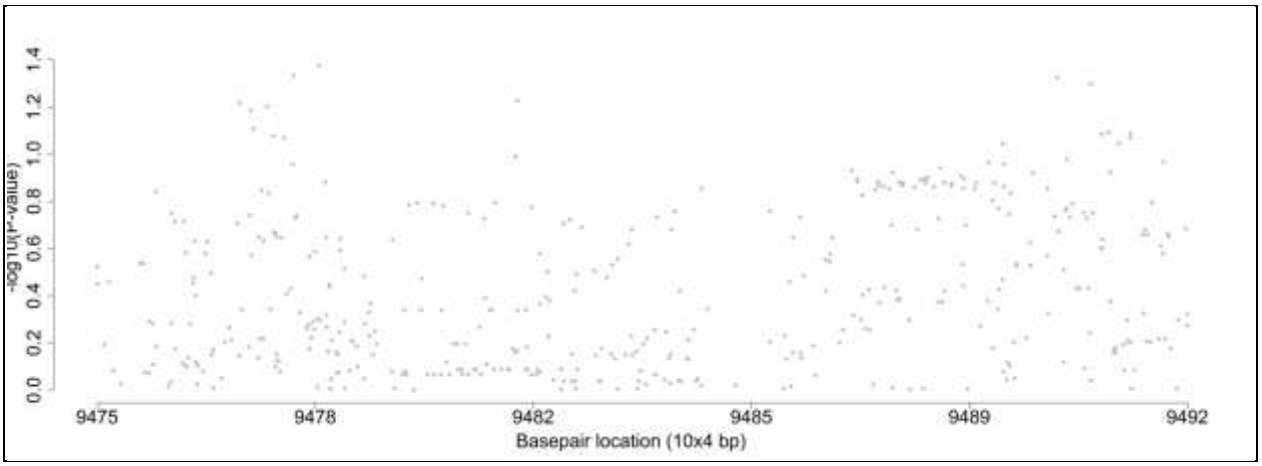

Figure 2: Smoking *SNP Interaction

NSAIDS

The PON variant in Table 5 showing the strongest association in the NSAIDs association with PD was rs43037, an intergenic SNP located between PON2 and ASB4 (OR = 0.75, 95% CI = 0.58-0.97 , *P* = 0.027). While the Gene View on the SNP in dbSNP refers to the SNP as being on PON2, the chromosomal location of PON2 as given by dbSNP for Build 37.1 are 7-95,034,174 - 7-95,064,384. This SNP, rs43037, is located outside the boundary of the gene at 7-95067006.

Table 5: Top Hits for Interaction Model between NSAIDs_OTC (ever/never NSAIDs OTC) and PON SNPs associated with PD: Adjusted for age at blood draw and sex

SNP	N	OR	OR_LOW95	OR_UP95	Z	PVAL	GENE	POSITION (BUILD 37)	LOCATION
rs43037_C*nsaids_otc_use	2553	0.75	0.58	0.97	2.20	0.027	PON2&ASB4	Chr7, 95067006	Intergenic
rs112227853_G*nsaids_otc_use	2521	0.49	0.25	0.95	2.09	0.036	PON2&ASB4	Chr 7, 95070043	Intergenic
rs6962107_C*nsaids_otc_use	2548	0.20	0.04	1.00	1.95	0.050	PON1&PON3	Chr7, 94977753	Intergenic
rs62469566_A*nsaids_otc_use	2542	0.64	0.22	1.87	0.81	0.417	PON2	Chr7, 95037208	Intron
rs17883991_G*nsaids_otc_use	2510	0.17	0.02	1.10	1.85	0.063	PON1&PON3	Chr7, 94981188	Intergenic

rs17876184_T* nsaids_otc_use	2550	0.37	0.13	1.06	1.84	0.065	PON2	Chr7, 95064026	Intron

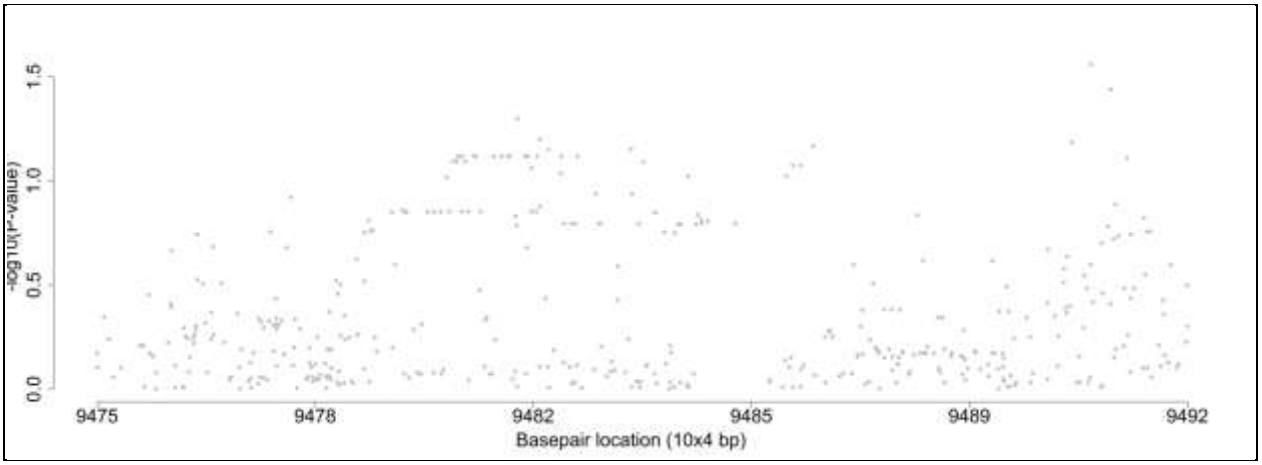

Figure 3: NSAIDs*SNP Interaction

The plot for the PON*NSAIDs OTC interaction is very similar to smoking with top hit values

< -log10 (p-values) =1.5.

Table 6: Imputation Values on Top Hit SNPs from All Environmental Exposures modeled for Interaction

SNP	Location (Build 37.3)	Location (Build 36.3)	Exposure	SNP Status	HWE P-Values	Info	Certainty
rs62469566_a	95037208	94875144	All three	Imputed	1	0.787	0.992
rs17876184_t	95064026	94901962	All three	Genotyped	0.4323	1	0.999
rs75071114_a (v419)	95065207	94903143	Smoking	Imputed	0.2435	0.741	0.987
rs62467349_t	94946795	94784731	Smoking	Imputed	0.7433	0.966	0.991
rs10240398_t	95082010	94919946	Smoking	Genotyped	0.321	1	0.998
rs2299258_a	94942917	94780853	Smoking	Imputed	0.7489	0.961	0.989
rs7794650_a	95061804	94899740	Smoking	Imputed	1	0.632	0.977
rs2237585_c	95049818	94887754	Coffee	Genotyped	0.682	1	1
rs17876205_g	95033104	94871040	Coffee	Imputed	0.2624	0.759	0.988
rs2375005_a	95036876	94874812	Coffee	Imputed	0.8917	0.99	0.995
rs9640633_g	95036897	94874833	Coffee	Genotyped	1	1	1
rs987539_t	95036992	94874928	Coffee	Genotyped	1	1	1
rs705379_g	94953895	94791831	Coffee	Imputed	0.009189	0.959	0.976
rs43037_c	95067006	94904942	NSAIDS	Genotyped	0.3819	1	1
rs112227853_g (v436)	95070043	94907979	NSAIDS	Imputed	1	0.73	0.996
rs6962107_c	94977753	94815689	NSAIDS	Imputed	1	0.892	0.998
rs17883991_g	94981188	94819124	NSAIDS	Imputed	1	0.527	0.986

In Table 6, about a third of the top hit SNPs were genotyped from the GWAS panel. All the imputed SNPs from among the top hit models have high certainty of imputation as indicated by information metrics greater than 0.5. Two SNPs were allotted variable names 7-94903143 (v419) and 7-94907979 (v436) when data was first imputed using Build 36.3. In subsequent Builds, they have been identified and established as SNPs in dbSNP named rs75071114_a and rs112227853_g respectively.

From the results in this dataset, the PON region may be thought to have weak interactions at best with the environmental factors known to influence PD. However, none of the interactions were statistically significant, once the Bonferroni adjustment for multiple testing was made.

Discussion

This analysis provides, to our knowledge, the first evaluation of gene-environment interactions between variants in the PON region and the three known environmental risk factors for PD. The PON region may have weak interactions with environmental factors known to be inversely-associated with PD risk, While none of the SNPs reached the threshold for significance after accounting for multiple comparisons, there were a few SNPs with relatively large effects.

Although we did not detect any statistically significant evidence for interactions after our conservative adjustment for multiple comparisons, it is interesting to note that two of the top-hit SNPs in the environmental interaction models are located on cSNPs (housed within coding

regions of a gene) of uncertain significance, i.e., a variant whose association with disease risk is as yet unknown. The location of the SNPs makes it plausible to hypothesize that any risk susceptibility conferred by these SNPs may be linked to the gene product. One explanation for the cSNPs being relevant to the interaction model is that they may be involved in alternate spliced variants. It has been estimated that 35-59% of human genes are alternately spliced and produce altered transcripts. That number is thought to be even higher for genes with multiple-exons (Lander, 2001). One possible next step is creating a comprehensive catalog of cSNPs that have been identified in the NGRC dataset. A comprehensive list of the cSNPs and the final protein product may help identify the mechanisms by which cSNPs influence susceptibility to PD.

Whether housed within coding regions of a gene or non-coding intergenic/intronic regions, all the top hit SNPs in this analysis are variants of uncertain significance or VUS. These represent a variation in the genetic sequence whose association with disease risk is unknown. cSNPs, by virtue of representing coding regions are more likely to be associated with an altered gene product. They also represent avenues for further exploration by way of RNA expression profiling and proteomics to determine how they impact the protein and what effect that might have on PD etiology. However, the majority of the SNPs from the top hit SNPs in the analysis are in non-coding regions of the genome, and our current knowledge of how they may impact disease etiology is limited.

Emerging research is altering the framework of functionality on non-coding SNPs (located on intronic and intergenic regions, i.e., outside the coding area of genes within the genome). These were at one point thought to be non-informative genomic markers and not considered to be of much value in targeted gene research. These segments are now known to house miRNA's, a type of RNA that are post-transcriptional regulators of gene expression, usually involved in a gene-silencing mechanism by binding to complementary sequences in the

3'UTR of target mRNA's. The identification of regulatory elements originating in these regions ties in to the notion that non-coding SNPs may also be valuable in their contribution to disease susceptibility mechanisms in PD mediated by environmental exposures.

All the top hit SNPs in the coffee*SNP association with PD have point estimates (ORs) higher than 1.5 with rs17876205_G (an intergenic SNP between PON3 and PON2) with OR = 7.78 (1.67, 36.21), being associated with much greater risk of PD in coffee*SNP interactions in the PON region. However, the width of the interval suggests a level of uncertainty in the measurement. Rs705379_G with OR = 1.52 (1.10 - 2.11) is a cSNP of uncertain significance on PON1.

Among the top hit SNPs in the smoking*SNP association with PD, 5 out of 6 are associated with reduced risk of developing PD. Rs75071114_A*, a cSNP of uncertain significance on PON2 is associated with a PD risk reduction of 39%in the interaction model. Another SNP, rs7794650, in the intronic region of PON2 is associated with a 70% reduction in risk of developing PD in interaction with smoking with OR = 0.30 (0.09- 0.98).

In the NSAIDs*SNP association with PD, there was considerable reduction in risk. Rs17883991_G, an intergenic SNP between PON1 and PON3 with an OR = 0.17 (0.02 - 1.10) is associated with an 83 % reduction in risk of developing PD in interaction with NSAIDs

Imputing SNPs allows non-genotyped SNPs to be included in the analysis and can provide additional information and power, even when average imputation accuracy is poor (Yongtao, 2008) . Imputation uses information about the degree of linkage disequilibrium (LD) between the observed and unobserved SNPs, allowing the genotypes of unobserved SNPs to be inferred (Biernacka, 2009). Further, the degree of accuracy can also be evaluated, from the genotypes of the observed SNPs. Imputing SNPs for this analysis has more than doubled the coverage for the reference panel in the PON cluster of genes as compared to the genotyped

SNPs for the region from the NGRC GWAS panel. The result is improved coverage, but only for SNPs with a MAF >5%, thus rare variants are not imputed. Although coverage is improved in the region, the trade-off is the need to adjust for additional comparisons (219 actually genotyped vs. 467 total comparisons with imputed markers). The trade-off between increased coverage (of common variants) and power is one that needs to be considered.

With most genetic association studies and GxE studies in particular, there are a number of well documented limitations. Power is the most important limitation of them all. The ability to detect interaction effect hinges on the ability of case-control association studies to detect differences in allelic frequencies between cases and controls in the markers being analyzed. The likelihood that a study will detect an effect depends on the sample size, allele frequency and effect size. Despite the fact that the NGRC dataset, with its well-characterized cases and controls, is one of the largest of its kind in PD and is well-powered for GWAS, it may not be as well-powered for a GxE interaction study. Adequate power is of concern in most studies evaluating gene-environment interactions, as it has been shown that it may take between 5,000 - 50,000 participants to detect interactions in a GxE, depending on the frequency of the minor allele (Luan, 2001).

Misclassification of environmental variables is also of concern in GxE studies and can also reduce power. Environmental exposure questionnaires date people's retrospective reports of exposure. Retrospective recall includes aspects of revisionist bias, normal memory lapse, biases by knowledge of disease outcome and in cases of neurodegenerative diseases, biases stemming from cognitive decline.

In the NGRC dataset, the method of data-gathering for the EEQ is subject to some of the same concerns that are common to all GxE. Assuming misclassification in the NGRC is non-differential in that both cases and controls are subject to the same challenges posed by

retrospective recall, the effect would be to push the point estimates towards the null. It is also possible that there are differences in the recall stemming from cognitive decline, between cases and controls. In the case of non-differential misclassification, the ability to predict the impact on the point estimate is more difficult, as it could go either direction: towards the null or way from the null.

In terms of assessing the environmental exposures, recall bias presents more of a problem in NSAIDs than in coffee or smoking for both cases and controls. Coffee and smoking are lifestyle choices and the consumption of both follow somewhat consistent patterns in individuals and might be easier to recall accurately. NSAIDs however, is an exposure that is used as required, given during a time of duress and encompasses many drugs classified under a single umbrella category. There could be strong contextual-effects in participant recollection of the NSAIDs use causing differences in recall between cases and controls leading to flaws in measurement exposures and non-differential misclassification and bias the odd ratio towards null. NSAIDs is not seeing odds towards null in the association with PON in this analysis, but it may show more extreme an effect in the absence of the measurement error.

The limitations of the environmental-data gathering on coffee consumption revolves around the composition of the brewed beverage. Different types of bean roast and different methods of preparing coffee lead to distinctly altered chemical composition of the coffee itself. Metal filters allow oils to go through into the percolated liquid. Paper filters block out the diterpenes in the coffee. Steeping the coffee grinds as is done in a French press alters the polyphenols. Heating it to varying temperatures alters composition by breaking down some of the alkaloids. No specific compound in coffee has been associated with the inverse-risk for PD and yet, the default assumption hinges around caffeine. The questions in the EEQ are tailored only to caffeinated coffee consumption and not the method of preparation. In spite of not accounting for these differences, if the observed results from the coffee interaction is still in the

range of 8.3×10^{-3}, it can be hoped that once these differences are accounted for coffee should rise to a level of significance in interactions with PON looking at PD.

The limitations of the data-gathering on smoking as a variable is two-pronged. First off, the EEQ does not gather any information for other source of smoking and compounds in cigarette smoke such as cigars, beedis, cheroots, chewing tobacco, second-hand inhalation and nicotine patches. It also does not distinguish between filtered and unfiltered cigarettes. Also based on the rat model of nicotine and dopamine lesions, the frequency of smoking and the smoking interval are thought to play a role in disease development. That information is not a part of the questionnaire.

However, the NGRC dataset does have fairly comprehensive exposure information on all the variables and is well-positioned to evaluate the questions asked. That each of the variables are complex compounds comprised of dozens of smaller compounds, and we are yet to understand which of these is responsible for the inverse association seen on PD risk, does in no way detract from the capability of the dataset to explore the question of interactions of the three environmental factors in the PON region influencing PD risk.

Discrete classification of variables is appropriate for exploratory studies, but is limited in its ability as compared to continuous data, that can gather a range of responses on a continuous scale and is more accurate at detecting differences within the scale.

For this analysis, environmental factors were defined as ever/never use: 'ever' referring to use or exposure to one of the three environmental factors used in the analysis and 'never' to a lack of use or exposure. When compared to a continuous load variable that accepts values on a scale, this method may result in misclassification or lower power or both. However, the trade off was that the ever/never variables had the largest number of subjects with responses in NGRC and was, therefore, more complete.

The lack of significant findings could be that the effect of the environmental factors interacting with the PON region in PD is somewhat independent of all the parameters that were considered for this analysis or more likely that even in this large study, we do not have enough power to detect interactions. It is possible that power could be increased by using quantitative measures of our exposures, however, this information was not available for all subjects. . Including them may also provide different patterns of interaction than what is observed in this data. There are some suggestive interactions which may not be strong enough and observing them may require more power.

The stringent criteria used by the NGRC to diagnose PD greatly reduces misclassification of PD in the dataset. Most large disease studies in neurodegenerative diseases often rely on self report (rather than independent professional verification) to categorize cases. Self-report can lead to misclassification, especially within diseases sharing phenotypic similarities. The NGRC's process of standardizing the procedure creates an intentional rigor and validates that those participants selected as cases of PD are true cases. This ensures that the case group is likely to be more homogeneous and not include other related disorders.

The NGRC dataset is also one of very few datasets to contain environmental exposure information on the subjects. This offers the option of gene-environment interaction studies on a large database with considerable power for interaction studies.

In summary, PD is a disease with a complex etiology likely resulting from the interplay of genetic and environmental factors. Although significant evidence for GxE were not detected in this study, identifying gene-environment interactions can provide information on susceptibility loci and the environmental triggers that act on those loci and that maybe involved in modulating risk of PD. It must also be noted that while there was no evidence for significant interactions

with individual exposures in the data, it is possible that by considering combinations of all exposures, the effects may have been stronger.

In conclusion, targeting the multiple factors in PD, including detecting environmental interactions, is the direction research needs to focus on to identify what modulates the development and appearance of the PD phenotype. This will greatly aid the development of screening, diagnostic, therapeutic, and preventive treatment options.

Part B: Introduction

Rapid improvement in DNA sequencing technologies has made large scale genotyping possible in clinical studies. However, in order to detect associations between specific variants, particularly rare variants, and disease, large numbers of subjects are needed. The response to this problem has led to a proliferation of genomic biobanks. It has led to the emergence of genotype-driven research and large multi-site genomic consortia to meet sample size requirements. The trend has also raised questions about how and whether to return individual research results, requiring a new look at the current policy around these issues. This section will explore issues related to the return of genetic research results to individuals participating in genetic studies of Parkinson's Disease, including an assessment of the various stakeholders whose views are likely to shape the policy landscape on return of individual research results in PD.

In the United States, the Common Rule for the Protection of Human Research Subjects has no policy, guideline or regulation involving 'Return of Research Results' back to participants, except to state that information that may affect a participants decision to remain enrolled in a study needs to be given back to them. When this policy is viewed through the framework of biomedical ethics, specifically the principle of respect for persons, it poses the question of whether research has an obligation to return individual results back to research participants. There are many viewpoints around the issue, but no guidance clarifying a policy action (Murphy, 2008).

PD is a neurodegenerative disorder in which there is a loss of function in the dopaminergic neurons of the substantia nigra in the brain. The loss of function is associated with motor symptoms including tremors, bradykinesia and postural instability. Some non-motor

symptoms, including decreased colonic transit time and gastric motility have been found to precede the motor symptoms and other non-motor symptoms such as cognitive impairment and dementia manifest after the appearance of motor symptoms.

PD was historically thought to be largely environmental, and a genetic component was found only in the mid-90's. As a result, progress in understanding genetic influences in PD is in its infancy when compared to some other diseases. However, other challenges in PD add to the importance of the research results, including the fact that the pathoetiology of the disease is unknown, there are no accurate biomarkers for diagnosis or gauging progression of the disease. There is no known cure, treatment options to alleviate symptoms are minimal, and once the disease sets in, the prognosis almost certainly includes a slow degenerative process with a diminishing quality of life (National Institutes of Health, 2012).

Genetic research on PD aims to delineate the genetic components and their interactions with environmental factors in causal pathways. Genetic targets enable the study of mechanisms underlying the onset, progression and characteristic hallmarks of the disease and predict risk. Mutations in six genes have been causally linked to early-onset PD (appearing in those under 40, unlike the more common late-onset version that appears in those over 65). Early-onset PD also has a more significant genetic component that late-onset PD (Bekris, 2010). More genes are linked to increased risk. This finding raises important issues concerning screening for PD around these six genes. It also raises interesting questions on how we may approach incorporating the on-going research agenda in PD into translational benefit for high-risk groups. Like any multi-factorial disease, the risk variants are low in penetrance and confer marginally higher risk. NextGen sequencing is expected to generate a lot more data around the exome and alternate splice sites and produce the transformation that makes the case for return of research results.

Genetic research in individuals with Parkinson's Disease (PD) represents an example of the challenges posed by the lack of definitional clarity in the return of research results. PD is a disease that poses a degree of uncertainty in disease characterization, has no cure, and offers minimal treatment options. Returning research results is unlikely to be on account of clinical utility. Clinical utility of genetic information refers to the use of that information to improve health outcomes. Since genetic research on PD has only marginal clinical utility, any return of results in PD assumes value only when viewed through the lens of personal utility, i.e., genetic information benefitting a person's life even in the absence of a health benefit from that information.

Offering individual return of genetic research results in PD may enhance a patient's personal utility by helping them plan their future and also provides information on the nature of potential risk to offspring. A policy standard that addresses the offer to return of aggregates and individual research results has the potential to maximize benefit to research participants.

Context of PD Research

Goals of genetic research in PD

Research has many goals including understanding disease biology, identifying biomarkers diagnosing disease, and disease progression as well as using biological data points during the course of progression as targets for treatment. Among the goals of genetic research on PD pertaining to public health would be a genetic test-based screen that identifies at-risk individuals accurately and tests that identify biomarkers of exposure that increase risk of

developing PD. The evidence we have from current research does not have the capability to accomplish that goal. Much more research is required to provide the evidence base to address that need. The more immediate steps in identification of risk to individuals are centered around research participants and policies addressing how best and if we should return evidence from research to the individual study participants. There are many questions that need to be answered through a stakeholder analysis in order for us to have a comprehensive research model incorporating all the requisite elements.

Key Stakeholders

Key stakeholders influence and shape the research process. There are many stakeholders who have a vested interest in policy development around return of research results for individuals participating in genetic studies of PD , including but not limited to enrolled and prospective participants (both patients and families of patients), researchers, Institutional Review Boards (IRB's), advocacy groups, health insurance companies, long-term care insurance companies, clinicians, pharmaceutical companies, testing laboratories, clinicians, and funders of private/public research.

With respect to return of individual research results, I will be looking at a subset of the stakeholder group that is potentially the most impacted by any policy formulation around the topic. The stakeholder interests intersect in some areas and are distinct in others. They are defined by whose interests they seek to preserve and why.

Participants. The participant group in genetic research on PD are mostly people with PD and their family members. There has been no research conducted so far on this group of participants as to their preferences on whether or not they want research results returned to

them. PD patients, once diagnosed, have no therapeutic treatment that offers a cure, only symptom management options. Additionally, the uncertainty involved in the disease etiology and progression are factors that might propel patients to participate in the research, especially so if individual results providing clinically or personally valuable information were to be returned. It may be viewed as a tool to maximize benefit to themselves, their families and the public (Beskow L. N., 2011).

The genetic research on PD so far has not yielded any clinical utility. Clinical utility is based on actionability of the data and clinical significance, which is a three step process involving statistical significance, effect size and the usefulness of the result. Even with the explicit recognition that the results have marginal clinical utility, the results may be meaningful to participants from a personal standpoint. It is important to keep in mind that return of results should not be a blanket return for all participants. It should be offered as an option to participants in order to be respectful and mindful of participant preferences, and give them the choice of refusal should they not want to receive results.

Researchers. PD researchers are focused on understanding disease biology in PD and untangling biologic mechanisms that lead to the etiology of the disease. Time-based and financial constraints are tied in to the progress of the research, especially in the current research landscape. While as a group, they are in agreement with the principle of respect for persons, they do not have consent from the participants to enable them to return results. They are also concerned that participants may attribute more meaning to the results than scientific evidence would warrant about the inherent value of the result to the participant (Bollinger, 2012).

Advocacy Groups. Advocacy groups are an integral part of the stakeholders in that they represent the collective voices of most patients. One of their goals is to establish a dialog between the research community and the patients to help identify what areas of research in PD constitute the most immediate need and prioritize the direction of research. PD has multiple advocacy groups (National Parkinson Foundation, Parkinson's Disease Foundation, Michael J. Fox Foundation for Parkinson's Research, and The Parkinson Alliance, to name a few), each with a strong backing of the patient community, and nuanced research agendas. The research they endorse varies according to their research focus. This system enables parallel research in diverse areas and is thought to benefit the patient community overall. Discussion around return of individual research results is likely to be viewed in a positive light by these groups as it empowers their objective of establishing the dialog between researchers and patients. It must be noted that while advocacy groups claim represent most of the patient community, there are likely some perspectives left unrepresented.

Research Funders. Funding for PD is not homogeneous in origin. There are three main sources of research funding.

Private research funding by Direct-to-consumer companies like 23 and Me are promoting research with the dual aim of furthering research and providing returning results and information in accordance to their business model. Additionally, private research also funds much of the pharmaceutical efforts in PD. Although return of genetic results in PD may have pharmacogenomic implications, there is much research yet to happen in that field before return of results are relevant to the pharmaceutical industry.

Funding from advocacy groups somewhat straddles the boundary between private and

public funding, in that funds raised publicly are applied towards specifically directed research. They are more likely to want aggregate results, in keeping with their policy of being an umbrella organization representing the varied needs of patients, families, clinicians and researchers, but depending on the nature of the research and the result in question, may require individual results.

A substantive portion of PD research funding is governmental in origin. The National Institutes of Health (NIH) allocated 151 million dollars in 2011 towards PD research with almost two thirds of that funding coming from the National Institute of Neurological Disorders and Stroke (NINDS). Protecting the research commons is one of the major goals of public research funding. Resources are finite and the agencies allocating funds are concerned about the value of the result in terms of actionability to the participants. They are also concerned about diminishing the quality of research by layering non-research activities on top of the research framework.

Summary of Main Findings

Background

PD is the second most common neurodegenerative disorder affecting about 1% of the population in the United States. It is more common in men than in women and seen more in Caucasian ancestries. About 20 % of PD patients report a family history. In the remaining 80 %

or so, it is of idiopathic origin, and therefore thought to be multi-factorial, polygenic with significant gene-environment interactions (Bekris, 2010) .

Lack of Standardization

Currently, there is not much emphasis being placed on the implications of predictive information garnered from PD genetic studies with respect to return of research results to research participants. It is reflective of the lack of direction in the parent bodies that guide research. The National Institutes of Health (NIH) 'Policy for Sharing Data Obtained in NIH Supported or Conducted Genome-Wide Associations Studies', stated that they expect the return of individual results to research participants to be a rare occurrence (The National Institutes of Health, 2007). No policies have been formulated around decision-making in the return of test results landscape and yet, an important consideration for participants in the Personal Human Genome Project at Harvard was the public availability of their data (Personalgenomes.org, 2011). The NIH are aware of the unmet need in this area and have recently funded projects to gather data around it.

Rationale for Return of Genomic Research

Sequencing the human genome set the stage for large-scale genomic data-generation and paved the way for accelerated progress in bioinformatics and sequencing technologies. It is now possible to obtain a level of data that was unheard of even early last decade. While the genotype itself is unlikely to change, the meaning attributed to each genotype is constantly evolving. Similar to other genomic research, genetic research on PD is constantly generating new meaning to existing data which need to be analyzed, synthesized and disseminated for it to be translated to clinical and public health benefit. Part of the dissemination process involves setting up a process to offer participants the return of validated results from research they participated in as a means of maintaining respect for persons. Offering individual return of

results enables the creation an informed cohort and maximizes respect for participant preference.

However, any conversation around this shifting model of research, once it accounts for the underlying questions of validity of the technology and test results, quickly gets into the mechanisms of the return of results model, with the primary questions being

- What constitutes a result
- Which of the results obtained will be returned
- Who is to return the results
- What model of return of results is optimal for maximum uptake (individual session with genetic counselor, multiple sessions, online data tool, etc.)
- Who controls the data
- For how long does the obligation to return data exist
- Does it apply equally to secondary and de-identified data as it does to primary
- Are there other research contexts that need to be considered
- Is there a need for a new body to provide regulatory oversight to the process

Return of Research Results from Gene-Environment Interaction Studies - Any rationale that may be attributed for return of genetic research results becomes more robust when the research is one of gene-environment interactions. It may be argued that there is a more direct path to actionability and utility of results in gene-environment interaction studies as a consequence of research that enables the characterization of environmental risk. Environmental risk, unlike genetic risk can be modulated through policy and personal health behaviors. It can be acted upon through ecological models of social determinants of health. Research on genetics, while extremely valuable, rarely identifies actionable causal variants or

even causal variants. Most genetic research only point to susceptibility loci and variants of uncertain significance. There is a dearth of translational research, given the wide spectrum of genetics and genomics research being conducted. Gene-environment interaction studies offer a different pathway into translational research. Some environmental interactions are more readily modifiable than others. Those that are easily modified could be targeted for interventions to break the interaction of environmental risk factors leading into disease phenotypes or boost interactions of environmental factors conferring benefits to health.

The Discussion around Return of Genomic Research Results

Information on stakeholders, their interests, and their motivations guide policy makers into adopting policies that not only incorporate the needs of the stakeholders but also in formulating realistic policies around return of individual genomic research results that are sustainable in the long run.

There are many pros and cons discussed, each representative of established thought processes, but there is little consensus on many issues. Return of research results may be aggregate (for the entire participant group) or individual. They may be deliberate (research that was planned and executed to answer a specific question) or incidental (a per-chance discovery in a patient unrelated to the study question.

The Pros. Offer of return of research results is a method of endorsing public opinion and support. It will increase enrolment in research and public interest in research activities (Fernandez, 2006) . Return of research results is in keeping with the ethical principle of respect for persons. Individuals have the right to so choose to have results returned to them if they want it (Partridge, 2004). Respect for research participant outweighs the burden of disclosure to the researcher and research community (David I. Shalowitz, 2005). While there is a duty to inform the

participant of all risks discovered, the bar that defines the risk is set rather high and entirely within the threshold of clinical utility (Greely, 2007). Additionally, return of research results to participants is soundly grounded in the Kantian Categorical Imperative that participants are ends in themselves. Participants are research partners and should not treated as means to an end.

The Cons. The genotyping will need to be redone, in accordance with the Clinical Laboratories and Improvement Amendment (CLIA), in a CLIA-certified lab in order for results to be returned and the costs associated with the process are staggering (Bookman, 2006). Retraining personnel to deliver results could delay progress of genomic research and hinder researchers by creating new versions of old dilemmas (Klitzman, 2006). Genetic information is different from clinical information in that it may be applicable to family members as well. Once the genie is out of the bottle, it can't go back in. This raises interesting questions on who should be consented (Knoppers, 2007). Re-contacting participants raises privacy implications that need to be dealt with. If return of results is made a duty, then breach of duty claims have the potential to alter the ability to conduct research (Meltzer, 2006) . Finally, most results so far are only variants of uncertain significance and in practice, it is often not possible to confidently establish the significance of a variant for the purpose of timely clinical decision making (Murray, 2011). It may on the other hand, lead to psychological harms by offering results that are not meaningful. This has the potential to damage the dialog between researchers and participants unless the results are offered in a manner that is conducive for individual participants to select the nature and type of results they choose to receive from their participation in genomic research.

Recommendations by Various Groups on the Return of Genetic Research Results

The National Bioethics Advisory Commission in its August 1999 report, recommended the return of genomic research results only if there was an immediately available course of action.

The National Heart, Lung, and Blood Institute (NHLBI) convened a working group in 2004, with experts weighing in on genetics, clinical research, and the ethical, social, and legal implications on if, when and how genetic information should be reported back to study participants. The group concluded that genetic test results should be reported to study participants when the associated risk for the disease is significant; the disease has important health implications such as premature death or substantial morbidity or has significant reproductive implications; and proven therapeutic or preventive interventions are available. The group also recommended the need for uniform guidelines in this regard (Bookman, 2006).

Updates in 2009 to the 2004 NHLBI working group noted the change in the landscape that had taken place in the 5 years since the original recommendations and produced five recommendations on returning genetic study results to research participants. The recommendations pertain to the criteria that must be met by the results (health implication, actionability, analytic validity of test and consent from patient), the time-bound obligation to return results (limited to the duration of the funding), the establishment of an independent review committee to offer guidance, that the investigator can return results even if all criteria are not met so long as it is grounded in sound science, has the approval of the IRB and the consent of the participant, and finally, in identifiable communities, investigators should engage the community and return of results may be either aggregate or individual depending on the preference of the community (Fabsitz, 2010).

An Example of Parkinson's Disease Consortia and Return of Research Results

The NeuroGenetics Research Consortium (NGRC) began as a genetic study of PD in Oregon in the 1990's and has since grown to include collaborations at several academic, federal and state institutions. The genetic arm of the study has been funded by NIH since 1998 (R01 NS36960). In 2004, the consortium was formalized as a Michael J Fox Foundation Funded Global Genetic Consortium, and an epidemiologic arm was implemented. NGRC includes eight movement disorder clinics in four states: Oregon, Washington, Georgia and New York. 2000 patients and 2000 controls were recruited at the disorder clinic. They were all white, cases were diagnosed with PD, and they all consented to give tissue sample for DNA extraction and filled out an environmental questionnaire on exposure assessment in order to conduct GxE studies on the data. The consent, in keeping with the standard consent used in PD research does not seek permission for return of research results from the study (dbGAP, 2011). If the magnitude of any deliberate or incidental finding from this study warrant the return of the research result, the participants will have to be recontacted and reconsented for such a return option. Any secondary research being conducted on this data have no return options whatsoever as their data is de-identified.

Policy options

Maintaining the Status Quo

Genomics research conducted through large scale genetic epidemiology studies does not usually return research results to participants. Deliberate inaction is definitely one policy action in the return of genomic results to research participants. It may be argued that the scant

evidence from current genomic research in accurate risk prediction is reason enough to shelve the policy action and guidance for a later time when we have more concrete evidence and clinically actionable research. Most genomic results, at this point, are risk variants whose allelic frequencies are not known across different populations, that contribute to marginally increased risk, are found to be of low penetrance, and are in intergenic or non-coding regions of the genome. Genomics is sometimes unable to accurately characterize even those variants that are involved in an altered expression of a cellular function, simply because the entire pathway is unclear or its association with causal elements in the disease etiology and progression in unknown (Murray, 2011). The pros of this action revolve around not defining policy when the evidence-base is uncertain. The cons of this action are slightly nuanced in that the lack of any guidance might send a message that return of research results to participants is not likely to be considered anytime soon. The lack of clarity could lead to repetitive work and confusion in assessing if there is a need to return results. It could also lead to a waning of interest in genomics research which might be reflected in a funding shift.

Return of Genomic Research Results Based on Significance of Result

Most advocates for return of genomic research results to participants are grounded in the ACCE Model of analytic validity, clinical validity, clinical utility and ELSI to best gauge if the result is worth returning. The recommendations of various groups, commissions and steering committees of experts specifically set up to address the issue tie in to this framework of thought (Fabsitz, 2010). Medical actionability of the result is the gold standard to which research results are held up. If the results are determinative of a disease, and there are treatments that cure the disease, alleviate symptoms, slow the progression of the disease or enhance quality of life, only then are the research results considered actionable enough to return back to participants. However, in the absence of research regulations on the subject, even when all the criteria are met, the decision to pursue return of results rests solely within the study and its researchers and

possibly the IRB that oversees the study. This was observed in the eMERGE Networks study for four of the bio-repositories (Fullerton, 2012). The pros of this policy are the evidence-based evaluation on the need to return results is consistent with all other policies in this realm. The cons of this policy approach is that the participants have no choice in assessing what may be significant or valuable to them.

Context-Based Return of Results

The research context can be described as the physical setting of the research: where was it conducted, who participated and how. Based on this, the conduct of the research can either be primary or secondary with respect to data, clinical research or community research or some other type of research based on the setting. Even within a clinical setting, the context of research may be sub-classified by the depth of relationship: if it involves the clinicians, it is more personal and based on the doctor-patient relationship versus a researcher obtaining a sample blood draw and filling out a questionnaire of sorts in a single interaction with the participant. In order for the debate around return of genomic research result to be meaningful, it must incorporate research context into any proposed model.

Other than a fundamental duty to rescue when in a position to be able to do so when risk or harm to the research community is minimal, researchers obligations are modified by three main factors (Beskow L. B., 2010):

1) Degree of Vulnerability: Participants are at the lower end of the power differential in the researcher-participant relationship and are vulnerable to discretionary actions by the researchers that may affect participant well-being. This differential is amplified when the participant is also a patient, which alters their status to patient-subjects and not just human subjects.

2) Depth of the Relationship: Researchers have a stronger moral responsibility to

engage with the participants when the relationship is deeper. Research in a clinical setting involving the patients' physicians have more depth than ones with only research interaction. Researchers working on secondary data have no relationship with the patients/participants and their obligations are at best, weak.

3) Degree of Dependence: Refers to participation in research because there are no other options available to the participant. The main reference here is still clinical utility in that the degree of dependence answers the question 'what difference the return of results will have participant health?'

The Need to Consider Participant Motivation

There is a growing body of evidence that points to participant interests in wanting results returned and some evidence that is contrary as well. There is a lot of debate on defining the threshold for determining importance, particularly around the medical actionability of data. Clinical actionability is considered to the criterion for return of results from genetic studies. It may be argued that clinical utility is a narrow interpretation of utility, and given the context, personal utility plays an important role. Clinical utility incorporates personal utility and medical actionability. Personal utility refers to the benefit of information to the participant, in the face of a lack of evidence to anything actionable. Currently, there is an unmet need to consider participant motivations and preferences, in the context of genomics research in PD.

It is reasonable to assume that the motivations of participants to take part in genomics research can range from entirely altruistic to using their participation as a barter right to obtain information. The key driver deciding where research participants fall in that spectrum bookended by altruism and barter, is contingent on the research in question and the personal utility of that information. There is however, no research available currently on identifying the range of research motivations while considering research context. It may be hypothesized that a

participant in genomics research for asthma that is non-fatal, alters quality of life marginally, is well-controlled by medicine and behavioral changes, will have less personal utility from the genomic information than someone with PD. Alternately, if we consider cancer genomics research, the disease is fatal in many instances, alters quality of life significantly and the genomics research is of great clinical utility (in putting an end to the diagnostic odyssey and helping select the right treatment option) but may be considered not as valuable for personal utility. Also, it is not always of value to the family, unless the genotype in question is intrinsic to the patient and not a mutation found only in the cancer tissue. The context of PD research may be thought to combine the gravity of cancer genomics but with complex, multi-factorial pathways and a surprising lack of knowledge on most aspects of the disease. The two combined, do not lend themselves to the successes that cancer genomics has seen in translational medicine and disease characterization.

PD Participant Motivation - PD for all reasons discussed earlier (unknown etiology, lack of diagnostic markers to detect disease and track progression, lack of a cure, lack of therapeutics that alleviate symptoms, multi-factorial disease, environmental exposures play a large role, degenerative process, loss of quality of life) falls into a participant context that has more personal utility than anything else from the results from genomic research. PD research participants are seeking to find meaning from research. The meaning is not self-evident. An important question to be resolved by better understanding of patient perspectives is the recognition of the many situations in which personal utility bridges that gap and the notion of personal utility being valuable even when the clinical utility is non-existent. In the absence of traditionally viewed meaning from research, it is reasonable to argue that PD falls into a context of research where even variants of uncertain significance (VUS) are meaningful and empowering to the patients as they represent an active process of information-gathering about the disease when there is little else that can be actively done to manage the disease. It may

also be extrapolated that VUS or risk variants associated with environmental triggers may alter behavioral patterns in the patient and the family. For instance, if genomics research in PD shows a variant of uncertain significance or a risk-allele conferring a slightly elevated risk to PD in the Aryl hydrocarbon receptor (AhR), a gene that serves as a transcription factor and the main clearance pathway for organochlorides in the cytosol, it may induce a behavioral change to avoid pesticide exposure, which is known to be associated with PD. This ties in to the concept that even VUS and risk variants of low risk have the capability to induce a behavioral change. This will not qualify as actionable under the clinical utility mode, but it does qualify as actionable in the personal utility model.

This context would be applicable to other diseases in the neurogenetics research field like Alzheimer's Disease and Dementia. Returning genomic research results in the absence of any evidence will not reduce the burden of disease but it may reduce the psychological burden and is therefore, the ethical path to follow in this situation.

However, there is a singular lack of any study regarding participant motivation and preferences based on context in these areas. Reasonable person standards do not necessarily provide any evidence, except a logical, conjecture-based hypothesis. More research is definitely needed in analyzing participant motivations and preferences with respect to the degree of dependence housed entirely within personal utility to participants. Research in that area will shed light on participant motivations and expectations, and help provide clarity in policy formulation.

Conclusions

In conclusion, there is no one size fits all answer to the question of return of genomic research results to participants. The context of research needs to be defined first before any consensus can emerge on the issue. Further deliberation is needed to analyze if participant preferences vary by the disease in question and if so, further delineating of the types of context is needed in order to unravel the research participant mindset and how it determines the framework of research.

Return of research results to participants blurs the line between researcher and clinician and may contribute to a misconception on the meaning and value of results. Traditionally health law and policy have kept that distinction dichotomous, because the duties are different for the two sets of people. Questions have also been raised on whether return of results will alter the framework of research and if it will make research more tentative, require tedious paperwork unrelated to the research itself and impede the progress of research.

While there is deliberation around making the offer of return of research results obligatory and a part of the research continuum, it is important to note that in none of the deliberations around the topic is it being viewed as a duty. This is in line with the current thought process of research as a standalone hermeneutic process that benefits the society at large when it progresses. Enforcing a duty to research will stifle research, innovation, and creativity.

The next step in addressing the issue ought to be delineating the various contexts of research and identify what each contextual framework obligates the researcher to. Once that is

agreed upon, policies need to be made for each context. This is an immediate and unmet need. Genetic studies are increasingly requiring large consortia to detect evidence. It would be a monumental endeavor if the studies were not designed appropriately and subsequent policies required reconsenting to disseminate vital information.

Summary

Return of genomic research results is an area that needs policy development. There are increasingly more and more genomic studies, comprising thousands of individuals and there are no processes in place to return results back to the participants. While there are many pros and cons to the situation, including but not limited to returning accurate results, the actionability of those results and the alteration it may cause to the way we conduct research, current discussions in the field point to the fact that we do not have the evidence to conclude on how best to proceed without conducting further research in the field. We require more information on the context of research that defines participant motivation and preferences. We also need a better understanding of how the context influences personal utility of the research results, in the absence of any known clinical utility or medical actionability. PD research is subject to the same uncertainties and while it may be appropriate to make reasonable assumptions for the sake of hypothesis generation, definitive research is needed to provide the evidence that will help make policies in this area.

BIBLIOGRAPHY

National Institutes of Health. (2012, March 7). *Parkinson's Disease Research Web Overview*. Retrieved March 12, 2012, from National Institute of Neurological Disorders and Stroke: http://www.ninds.nih.gov/research/parkinsonsweb/index.htm

Almeida, M. d. (2012). Glucocerebrosidase Involvement in Parkinson Disease and Other Synucleinopathies. *Frontiers in Neurology* , 2012; 3: 65. doi: 10.3389/fneur.2012.00065.

Ascherio, A. Z. (2001). Prospective study of caffeine consumption and risk of Parkinson's disease in men and women. *Annals of Neurology* , DOI: 10.1002/ana.1052.

Barbeau, A. R. (1987). Ecogenetics of Parkinson's disease: prevalence and environmental aspects in rural areas. *The Canadian journal of neurological sciences. Le journal canadien des sciences neurologiques,* , Feb; 14(1): 36-41.

Bekris, L. M. (2010). The Genetics of Parkinson Disease. *J Geriatr Psychiatry Neurol.* , December; 23(4): 228–242.

Beskow, L. B. (2010). Offering Individual Genetic Research Results: Context Matters. *Sci Transl Med* , doi: 10.1126/scitranslmed.3000952.

Beskow, L. N. (2011). Research participants' perspectives on genotype-driven research recruitment. *J Empir Res Hum Res Ethics.* , Dec;6(4):3-20.

Biernacka, J. T. (2009). Assessment of genotype imputation methods. *BMC Proceedings* , 3(Suppl 7):S5 www.biomedcentral.com/1753-6561/3/S7/S5.

Bollinger, J. S. (2012). Public preferences regarding the return of individual genetic research results: findings from a. *Genetics in Medicine* , doi:10.1038/gim.2011.66.

Bookman, E. L. (2006). Reporting Genetic Results in Research Studies: Summary and Recommendations of an NHLBI Working Group. *Am J Med Genet A.* , May 15; 140(10): 1033–1040.

Bouman, J. H.-G.-t. (2010). Paraoxonase-1 is a major determinant of clopidogrel efficacy. *Nature Medicine* , 17,110–116(2011) doi:10.1038/nm.2281.

Checkoway, H. P.-W. (2002). Parkinson's Disease Risks Associated with Cigarette Smoking, Alcohol Consumption, and Caffeine Intake . *American Journal of Epidemiology* , 155(8): 732-738 doi:10.1093/aje/155.8.732 .

Clarimon, J. E. (2004). Paraoxonase 1 (PON1) gene polymorphisms and Parkinson's disease in a Finnish population. *Neuroscience Letters* , Volume 367, Issue 2, 2 September 2004, Pages 168–170 http://dx.doi.org.offcampus.lib.washington.edu/10.1016/j.neulet.2004.05.108.

Collier, K.-S. M. (2002). Etiology of Parkinson's disease: Genetics and environment revisited. *PNAS* , doi: 10.1073/pnas.242594999 .

Costa, G. A.-C. (2001). Nicotine prevents striatal dopamine loss produced by 6-hydroxydopamine lesion in the substantia nigra. *Brain Research* , Jan 12;888(2):336-342.

Costa, L. G.-M. (2004). Paraoxonase 2 (PON2) polymorphisms and Parkinson's disease. *Neurosciences Research Communications* , Neurosci. Res. Comm., 34: 130–135. doi: 10.1002/nrc.20008.

David I. Shalowitz, A. (2005). Disclosing Individual Results of Clinical Research. *JAMA* , 294(6):737-740. doi: .

dbGaP. (2012, March 12). *Genome-Wide Association Study of Parkinson Disease: Genes and Environment dbGaP Study Accession: phs000196.v1.p1* . Retrieved April 10th, 2012, from http://www.ncbi.nlm.nih.gov: http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000196.v1.p1

dbGAP. (2011, July 15th). *Genome-Wide Association Study of Parkinson Disease: Genes and Environment dbGaP Study Accession: phs000196.v1.p1*. Retrieved March 13, 2012, from www.ncbi.nlm.nih.gov: http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000196.v1.p1&phv=75991&phd=1642&pha=&pht=875&phvf=&phdf=&phaf=&phtf=&dssp=1&consent=&temp=1

Environmental Molecular Epidemiology Research (EMERG) Group. (2008). *Gene-Metal Interactions and Parkinson's Disease.* Ann Arbor: University of Michigan.

Erlich, M. P. (2006). Polymorphisms in the PON gene cluster are associated with Alzheimer disease. *Human Molecular Genetics* , Vol. 15, No. 1 77–85 doi:10.1093/hmg/ddi428.

Esposito, E. D. (2007). Non-steroidal anti-inflammatory drugs in Parkinson's disease. *Experimental Neurology* , 205 (2007), pp. 295–312.

Fabsitz, R. M. (2010). Ethical and Practical Guidelines for Reporting Genetic Research Results to Study Participants : Updated Guidelines From a National Heart, Lung, and Blood Institute Working Group. *Circulation Cardiovascular Genetics: Journal of the American Heart Association* , DOI: 10.1161/CIRCGENETICS.110.958827.

Fernandez, C. W. (2006). Obligations in offering to disclose genetic research results. *The American journal of bioethics* , The American journal of bioethics.

Fullerton, S. W. (2012). Return of individual research results from genome-wide association studies: experience of the Electronic Medical Records and Genomics (eMERGE) Network. *Genetics in Medicine* , doi:10.1038/gim.2012.15.

Gibb, W. L. (1988). The relevance of the Lewy body to the pathogenesis of idiopathic Parkinson's disease. *Journal of Neurology, Neurosurgery, and Psychiatry* , 51:745-752.

Greely, H. T. (2007). The Uneasy Ethical and Legal Underpinnings of Large-Scale Genomic Biobanks. *Annual Review of Genomics and Human Genetics* , Vol. 8: 343-364.

Hamza, T. C.-B. (2011). Genome-wide gene-environment study identifies glutamate receptor gene GRIN2A as a Parkinson's disease modifier gene via interaction with coffee. *PLoS Genetics* , Aug;7(8):e1002237. Epub 2011 Aug 18.

Hardy, J. C.-H. (2006). Genetics of Parkinson's disease and parkinsonism. *Annals of Neurology* , Volume 60, Issue 4, pages 389–398, October 2006 DOI: 10.1002/ana.21022.

Hernán, A. M.-I.-O. (2002). A meta-analysis of coffee drinking, cigarette smoking, and the risk of Parkinson's disease. *Annals of Neurology* , Volume 52, Issue 3, pages 276–284, September 2002 DOI: 10.1002/ana.10277.

Impute2. (2012, April 19). Retrieved June 5, 2012, from Mathgen.stats.ox.ac.uk: http://mathgen.stats.ox.ac.uk/impute/impute_v2.html

Jankovich, J. (2008). Parkinson’s disease: clinical features and diagnosis. *Journal of Neurology, Neurosurgery & Psychiatry with Practical Neurology* , 79:368-376 doi:10.1136/jnnp.2007.131045.

Klitzman, R. (2006). Questions, complexities, and limitations in disclosing individual genetic results. *The American Journal of Bioethics:* , Nov-Dec;6(6):34, 6.

Knoppers, B. (2007). *Genomics and Public Health: Legal and Socio-Ethial Perspectives.* Martinus Nijhoff Publishers.

Kondo, I. Y. (1998). Genetic polymorphism of paraoxonase 1 (PON1) and susceptibility to Parkinson's disease. *Brain Research* , Volume 806, Issue 2, 28 September 1998, Pages 271–273 http://dx.doi.org.offcampus.lib.washington.edu/10.1016/S0006-8993(98)00586-1.

Lander, E. (2001). Initial Sequencing and Analysis of the Human Genome. *Nature* , (0028-0836), 409(6822), p.860 DOI:10.1038/35057062.

Le Houezec, J. (2003). Role of nicotine pharmacokinetics in nicotine addiction and nicotine replacement therapy: a review. *International Journal of Tuberculosis and Lung Disease* , Sep;7(9):811-9. 12971663.

Lesage, S. B. (2009). Parkinson's disease: from monogenic forms to genetic susceptibility factors. *Human Molecular Genetics* , 18 (R1): R48-R59. doi: 10.1093/hmg/ddp012.

Liang, H.-L. L.-P.-C. (2003). Paraoxonase gene polymorphisms, oxidative stress, and diseases. *J Mol Med* , 81:766–779 DOI 10.1007/s00109-003-0481-4.

Luan, J. W. (2001). Sample size determination for studies of gene-environment interaction. *International Journal of Epidemiology* , 30:1035-1040.

Mackness, M. M. (2002). Paraoxonase and coronary heart disease. *Atherosclerosis Supplements* , Volume 3, Issue 4, December 2002, Pages 49–55 http://dx.doi.org.offcampus.lib.washington.edu/10.1016/S1567-5688(02)00046-6.

Manthripragada, A. S. (2011). Non-Steroidal Anti-Inflammatory Drug Use and the Risk of Parkinson's Disease. *Neuroepidemiology* , v36 n3 (2011): 155-161.

Maraganore, D. A. (1999). Autopsy patterns for Parkinson's disease and related disorders in Olmsted County, Minnesota. *Neurology October 1,* , vol. 53 no. 6 1342.

Maraganore, D. L.-H. (2004). UCHL1 is a Parkinson's disease susceptibility gene. *Annals of Neurology* , Apr;55(4):512-21.

McCulloch, C. C. (2008). Exploring gene-environment interactions in Parkinson's disease. *Human Genetics* , 23:257–265 DOI 10.1007/s00439-008-0466-z.

Meltzer, L. (2006). Undesirable implications of disclosing individual genetic results to research participants. *The American Journal of Bioethics* , Nov-Dec;6(6):28, 30.

Murphy, J. S. (2008). Public Expectations for Return of Results from Large-cohort Genetic Research. *Am J Bioeth* , 8(11): 36–43.

Murray, M. C. (2011). Follow-up of carriers of BRCA1 and BRCA2 variants of unknown significance: Variant reclassification and surgical decisions. *Genetics IN Medicine* , DOI: 10.1097/GIM.0b013e318226fc15.

National Institutes for Health. (2007, July). *Genetics Home Reference: Parkinson's Disease*. Retrieved March 1, 2012, from http://ghr.nlm.nih.gov: http://ghr.nlm.nih.gov/condition/parkinson-disease

NCBI. (2012, May 11). *PON2 Paraoxonse 2 [Homo sapiens]*. Retrieved May 14, 2012, from http://www.ncbi.nlm.nih.gov: http://www.ncbi.nlm.nih.gov/sites/entrez?Db=gene&Cmd=ShowDetailView&TermToSearch=5445

Paganini-Hill, A. (2001). Risk Factors for Parkinson's Disease: The Leisure World Cohort Study. *Neuroepidemiology* , 20:118-124 (DOI: 10.1159/000054770).

Pare, G. R. (2012). Effect of PON1 Q192R Genetic Polymorphism on Clopidogrel Efficacy and Cardiovascular Events in the Clopidogrel in the Unstable Angina to Prevent Recurrent Events Trial and the Atrial Fibrillation Clopidogrel Trial With Irbesartan for Prevention of Vascula. *Circulation: Cardiovascular Genetics* , Issue: Volume 5(2), April 2012, p 250–256.

Parkinson's Disease Foundation. (2012). *Understanding Parkinson's*. Retrieved March 1, 2012, from http://www.pdf.org: http://www.pdf.org/en/understanding_pd

Partridge, A. H.-W. (2004). Oncology physician and nurse practices and attitudes regarding offering clinical trial results to study participants. *Natl Cancer Inst* , 96:629–632.

Personalgenomes.org. (2011, April 22). *Important Considerations*. Retrieved March 10, 2012, from http://www.personalgenomes.org: http://www.personalgenomes.org/considerations.html

Powers, K. K. (2007). Combined Effects of Smoking, Coffee, and NSAIDs on Parkinson's disease risk. *Movement Disorders* , Vol. 23, No. 1, 2008, pp. 88-95.

R Development Core Team. (2012). *R: A language and Environment for Statistical Computing*. Retrieved June 5th, 2012, from R-project: www.R-project.org

Samii, A. E. (2009). NSAID Use and the Risk of Parkinson's Disease: Systematic Review and Meta-Analysis of Observational Studies. *Drugs & Ageing* , Volume 26(9), 1 September 2009, pp 769-779 DOI: 10.2165/11316780-000000000-00000.

Samii, A. E. (2009). NSAID Use and the Risk of Parkinson's Disease: Systematic Review and Meta-Analysis of Observational Studies. *Drugs & Ageing* , Volume 26(9), 1 September 2009, pp 769-779 DOI: 10.2165/11316780-000000000-00000.

Schwarzschild, M. C. (2002). Caffeinated clues and the promise of adenosine A(2A) antagonists in PD. *Neurology* , 58:1154–60 .

Sergio L. Primo-Parmo., R. C. (1996). The Human Serum Paraoxonase/Arylesterase Gene (PON1) Is One Member of a Multigene Family. *Genomics* , Volume 33, Issue 3, 1 May 1996, Pages 498–507 http://dx.doi.org.offcampus.lib.washington.edu/10.1006/geno.1996.0225, .

Shin, J. D. (2009). SnapShot: Pathogenesis of Parkinson's Disease. *Cell* , Volume 139, Issue 2, 16 October 2009, Pages 440.e1–440.e2 DOI 10.1016/j.cell.2009.09.026.

Soto-Otero, R. M.-A.-S.-L.-R. (2001). Reduction of rat brain levels of the endogenous dopaminergic proneurotoxins 1,2,3,4-tetrahydroisoquinoline and 1,2,3,4-tetrahydro-beta-carboline by cigarette smoke. *Neuroscience Letters* , Feb 9;298(3):187-90.

STATA. (2012). *Stata release 11*. Retrieved June 5th, 2012, from Stata.com: http://www.stata.com/stata11/

Tanner, C. G. (2002). Smoking and Parkinson's disease in twins. *Neurology* , Feb 26;58(4):581-8.

The National Institutes of Health. (2007, August 28). *Policy for Sharing Data Obtained in NIH Supported or Conducted Genome-Wide Association Studies (GWAS)*. Retrieved June 18, 2012, from grants.nih.gov: http://grants.nih.gov/grants/guide/notice-files/NOT-OD-07-088.html

Trinh, K. A. (2010). Decaffeinated Coffee and Nicotine-Free Tobacco Provide Neuroprotection in Drosophila Models of Parkinson's Disease through an NRF2-Dependent Mechanism. *Journal of Neuroscience* , 30(16): 5525-5532; doi: 10.1523/JNEUROSCI.4777-09.2010 .

UCSC. (n.d.). *Human (Homo sapiens) Genome Browser Gateway*. Retrieved June 5th, 2012, from Genome.ucsc.edu: http://genome.ucsc.edu/cgi-bin/hgGateway

Webster, G. A. (2000). Association of Coffee and Caffeine Intake With the Risk of Parkinson Disease. *JAMA* , 283(20):2674-2679. doi:10-1001/pubs.JAMA-ISSN-0098-7484-283-20-joc91293.

World Health Organization. (2012). *Ageing and Life Course - Good health adds life to years: Global brief for World Health Day 2012.* WHO.

Ying-Li, L. J.-W.-M.-N.-W.-B. (2012). Paraoxonase 1 polymorphisms L55M and Q192R were not risk factors for Parkinson's disease: A HuGE review and meta-analysis. *Gene* , Volume 501, Issue 2, 15 June 2012, Pages 188–192 http://dx.doi.org.offcampus.lib.washington.edu/10.1016/j.gene.2012.03.067.

Yongtao, G. S. (2008). Practical Issues in Imputation-Based Association Mapping. *PLoS Genetics* , doi:10.1371/journal.pgen.1000279 .

Yuan, Z. Y. (2011). Association between PON1 activity and coronary heart disease risk: A meta-analysis based on 43 studies. *Molecular Genetics an Metabolism* , Volume 105, Issue 1, January 2012, Pages 141–148 http://dx.doi.org.offcampus.lib.washington.edu/10.1016/j.ymgme.2011.09.018,.