AD_____

Award Number: W81XWH-04-1-0404

TITLE: Large Scale Single Nucleotide Polymorphism Study of PD Susceptibility

PRINCIPAL INVESTIGATOR: J. William Langston, M.D.

CONTRACTING ORGANIZATION: Parkinson's Institute
Sunnyvale, CA 94089-1605

REPORT DATE: March 2006

TYPE OF REPORT: Final

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
01-03-2006	Final	1 Mar 2004 – 28 Feb 2006

4. TITLE AND SUBTITLE

Large Scale Single Nucleotide Polymorphism Study of PD Susceptibility

5a. CONTRACT NUMBER

5b. GRANT NUMBER
W81XWH-04-1-0404

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

J. William Langston, M.D.

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Parkinson's Institute
Sunnyvale, CA 94089-1605

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES
Original contains colored plates: ALL DTIC reproductions will be in black and white.

14. ABSTRACT

SEE ATTACHED PAGE

15. SUBJECT TERMS
Parkinson's disease genetics; SNPs; Molecular Inversion Probe Assay; haplotype analysis; genotyping

16. SECURITY CLASSIFICATION OF:				18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
					USAMRMC
a. REPORT	b. ABSTRACT	c. THIS PAGE			19b. TELEPHONE NUMBER *(include area code)*
U	U	U	UU	16	

ABSTRACT

Although genetic contributions of Parkinson's disease (PD) have gained support from the recent identification of eight genetic loci in the familial PD, the results of intensive investigations of polymorphisms in dozens of genes related to sporadic, late onset, typical PD have not shown consistent results. Recent rapid progress in the investigation of single nucleotide polymorphisms (SNPs) has provided a new tool for this area of research. Millions of SNPs have been identified and compiled in several public accessible databases. A highly multiplexed genotyping technology called Molecular Inversion Probe Assay has recently been developed. This technique is capable of genotyping over 2000 SNPs in a single tube and is currently the most inexpensive platform for genotyping of SNPs with high call rate and high accuracy. Another major finding in recent genomic studies is that haplotype linkage disequilibrium is composed of blocks of sequence with average size of 7.8 kb and could be used in association studies. In this proposal, we plan to perform a large-scale association study by using the high throughput Molecular Inversion Probe Assay in PD to (1) investigate the association between classical, sporadic PD and 2386 SNPs in 23 genes implicated in the pathogenesis of PD; (2) construct haplotypes based on the SNP genotyping results to identify haplotypes associated with PD. This proposal is the first large-scale SNP association study in PD. The results of this study may lead to the further understanding of the pathogenesis of the disease and possibly new therapeutic approaches. In addition, experiences derived from this study may be applied in other complex disorders for the identification of susceptibility genes, as well as in genome-wide SNP association studies.

Table of Contents

	Page No.
Cover……	**1**
SF 298……..……	**2**
Table of Contents……..……	**3**
Introduction…….……....	**4**
Body…….	**4**
Key Research Accomplishments…….……	**9**
Reportable Outcomes……	**10**
Conclusions…….	**10**
References……	**10**
Appendices……	**13**

FINAL REVISED REPORT

INTRODUCTION

There is an increasing consensus in the scientific community that Parkinson's disease is likely to result from a combination of environmental and genetic factors. In regard to the latter, a great deal of research has focused on finding functionally significant polymorphisms that predispose to the development of Parkinson's disease (i.e., genetically determined risk factors).

Many research groups have sought to identify PD susceptibility genes by carrying out population-based case-control association studies of candidate genes, identified on the basis of their involvement in a biologically relevant pathway that might be expected to lead to the degenerative process that underlies PD. However, no single genetic susceptibility factor has been shown to consistently increase the risk for PD (Mellick et al 1999, Nicholl et al 1999, Grevle et al 2000, Xu *et al* 2002, Zheng et al 2003, Le et al 2003, Marx et al 2003). These inconsistencies have prompted increasing concern about how future association studies should be performed (Nicholl et al. 1999, Markopoulou et al 1999). One problem is that most studies have investigated only one or a few polymorphisms in a particular candidate gene, with the assumption that the polymorphism would have an impact on the function/expression of the gene, at transcriptional, translational or post-translational level, or that the polymorphism is in linkage disequilibrium with another nearby critical variant.

This research project was designed to identify genetic susceptibility factor(s) for PD in a large group of PD cases and controls using two different approaches, haplotype tagging and whole genome screening. The first approach investigates the halpotypes that may be associated within 23 genes implicated in PD. The second approach investigates 20,000 non-synonymous cSNPs across the entire human genome. These two approaches allow us to map and find variations that could be associated with disease by focusing on those SNPs that are likely to be biologically important and potentially disease causing.

BODY

I) IRB: This project has been approved by our IRB. Only non-Hispanic whites are included in the study solely for the purpose of reducing the genetic variation in the study subjects, which is critical to the study design.

II) II) Sample collections: 200 PD patients and 200 controls were recruited for this study. The criteria for the patient inclusion were as follows: (1) at least three of the four following cardinal features of PD on neurological examination: resting tremor, rigidity, bradykinesia, and postural instability; (2) a clear-cut response to levodopa; (3) an absence of clinical features suggesting atypical parkinsonism; (4) age of onset over 50 years; (5) a disease duration of 10 years or less (this criteria has been highly recommended by our epidemiologists to avoid survivor bias); (6) non-Hispanic white (to reduce the genetic complexity as noted above); (7) no familial history. The criteria for control subjects included the following: Non-Hispanic white without a history of known neurodegenerative disorders, matched in age, gender with the patient group.

DNA quality was checked for each sample with a spectrophotometer, and detailed information on ethnic origin, disease history, clinical presentation, drug response and family history was collected for each case. Confidentially has been strictly maintained throughout the study.

III) DNA purification: Genomic DNA from patient and control samples was extracted and quantified by spectrophotometer; DNA quality verified by both Gel electrophoresis and spectrophotometer to check for DNA integrity and any possible contamination of DNA samples by RNA and/or protein. We required a 260/280 ratio between 1.7 to 2 for all extracted DNA and no degradation by electrophoresis gel, confirming the high quality of DNA.

IV) Whole genome screening: The ability to genotype of thousands of cSNPs across the genome and to perform association analysis between cases and controls enables a discovery-based approach for determining the underpinnings of complex human genetic disorders, including Parkinson's disease. In this study we used Affymetrix MegAllele™ Genotyping Human 20K cSNP kit, which utilizes Molecular Inversion Probe (MIP) Technology, to identify the possible cSNPs associated with PD. The DNA collected from our 200 PD patients and 200 controls was separately amplified, labeled, and loaded to 20K cSNP chip as recommended by Affymetrix of Santa Clara, CA (http://www.affymetrix.com/support/technical/manuals.affx). Briefly, 20K oligonucleotide probes with recognition sequences at each terminus were hybridized with 4µg of genomic DNA. The oligonucleotide probe forms a circular structure, with the ends of the probe abutting. This leaves a single base gap at the location of a SNP. This gapped-duplex was labeled in four separate reactions, each with a single dNTP species present, in which successful polymerization/ligation provides allelic differentiation. The probes were subsequently released from the genomic DNA and those that have been covalently circularized in the correct allele/nucleotide reaction combinations are amplified using a "universal" PCR primer pair. Each amplified probe contains a unique tag sequence that is complementary to a sequence on the universal tag array. Tags have been selected to have a similar Tm and base composition and to be maximally orthogonal in sequence complementarity. Amplicons were fluorescently labeled and the tag sequences released from the genome homology regions using a restriction endonuclease treatment. The tags were then detected using a complementary 20,000 tag array chip.

IV-1) Data analyzing: Genotypes were called by Targeted Genotyping Analysis software (GTGS) www.affymertix.com. In total eight million genotypes were generated. To investigate the possible association of each cSNPs with Parkinson's disease statistical R-package for exact Fisher test was used.

IV-2) cSNPs associated with Parkinson's disease: This study produced several novel and potentially exciting findings. Because MegAllele™ Genotyping Human 20K cSNP panel includes non-synonymous SNPs, this study enabled us to screen SNPs that code for functional changes across the genome. The list of the cSNPs that are associated with Parkinson's disease, with p-value$<10^{-6}$, is provided in Table-1. In this table ID number for each SNP, gene names, and p-values are provided. Interestingly, three cSNPs with ID numbers rs11007567, rs11007612,

rs11600 are located on chromosome 10 very close to each other. As indicated in Table-1 the chromosomal positions for rs11007567, rs11007612, rs11600 are 29750643, 29810557, and 29750946 respectively. rs11007567, rs11007612 are both located on the same gene, Loc387647, a predicted gene. Loc387647 gene is similar to RAB18 which is a member of RAS oncogene family. Because this data indicates that the region from 29750643 to 29810557 on chromosome 10 might be associated with PD we are planning to genotype all SNPs located within that region. On chromosome 10, two other cSNPs, rs10510114 (p-value = 10^{-6}) and rs11168070 (p-value = 10^{-6}), are located at positions 124510729 and 14856454 respectively, but distant from rs11007567, rs11007612, rs11600.

Two other cSNPs that were found to be strongly associated with the Parkinson's disease are rs16891982 and rs26722 and are both located on the same gene MATP (membrane-associated transporter protein) on chromosome 5. Mutations in MATP was previously found to be the cause of oculocutaneous albinism type 4 (oca4) [mim:606574]. oca4 is an autosomal recessive disorder of pigmentation characterized by reduced biosynthesis of melanin in the skin, hair and eyes and it leads to reduced visual acuity [mim:606574].

In regard to future studies beyond the duration of this proposal, pending success in obtaining additional funding, we will first confirm these results by using a different technology. If the results are confirmed, we will then repeat these studies in a different population to see if they are generalizable.

V) Analysis of haplotypes within selected candidate genes implicated in Parkinson's disease.

The aim of this study was to analyze haplotypes in candidate genes that may contribute to the development of typical sporadic PD. These candidate genes included;

- Enzymes involved in the metabolism of environmental toxins: The candidate enzymes in this category include cytochrome-P450 enzymes (CYP2D6, CYP2E1, CYP1A1), glutathione S-transferases (GSTM1, GSTT1, GSTP1, GSTZ1), paraoxonase 1 (PON1), N-acetyltransferases 2 (NAT2).
- Enzymes involved in the metabolism of dopamine: The candidate enzymes in this category include monoamine oxidases (MAOA, MAOB), catechol-O-methyltransferase (COMT), dopamine transporter (DAT), vesicular monoamine transporter-2 (VMAT2) and dopamine receptors (DRD1, DRD2, DRD3, DRD4, DRD5).
- Genes involved in the familial parkinsonism: The candidate enzymes in this category include *alpha-synuclein*, *parkin, UCHL1,* and *tau* genes.

V-1) Primer design, PCR and genotyping: The sequence and SNPs within above mentioned-genes were retrieved using NCBI (http://www.ncbi.nlm.nih.gov/Genbank/GenbankSearch.html) and hapmap (http://www.hapmap.org/) databases. For haplotype tagging, SNPs were selected by halpoview software. Forward and reverse primers for selected SNPs were designed using primers-3 program. One primer of each pair has a 5' biotin-triethylene glycol label necessary for post-PCR processing. Pyrosequencing template was made by PCR amplification (15 ng)

of genomic DNA in a 50-µL reaction in the presence of 5 pmol of each forward and reverse primer, 200 µM each deoxynucleotide triphosphate, 50 mM KCl, 10 mM Tris (pH 8.3), and 0.01 µg/µL gelatin. Template PCRs were set up on a MultiPROBE II HT liquid handler (PerkinElmer, Inc.). Thermocycling conditions for all PCRs were 35 cycles of 30 s at 95 °C, 30 s, and 30 s at 72 °C. A final 72 °C hold for 10 min ended the reaction. Each PCR plate contained one negative control (no template), and one positive control per genotype. PCR products were size-controlled by gel electrophoresis.

Ten microliters of each PCR product were used to determine the genotypes for each individual sample using PSQ (pyrosequencing). From the PCR-generated template, the desired strand to be sequenced was separated from its complement, and an appropriate sequencing primer was annealed to it according to the manufacturer's instructions. Briefly, 25 µL of PCR-generated template was incubated with 90 µg of streptavidin-coated Dynabead M-280 magnetic beads (Dynal AS) and a volume-to-volume quantity of 2x BW-buffer II (10 mmol/L Tris-HCl, 2 mol/L NaCl, 1 mmol/L EDTA, 1 mL/L Tween 20, pH 7.6). This mixture was incubated at 65 °C for 15 min, with shaking, to bind the template to the beads. A 96-pin magnetic tool (Pyrosequencing AB) was used to transfer up to 96 samples at a time to solutions as follows. The beads with bound template were first transferred to 50 µL of denaturing solution (0.5 mol/L NaOH), then to 100 µL of 1x annealing buffer (20 mmol/L Tris-acetate, 5 mmol/L magnesium acetate, pH 7.6), and finally into a solution of 1x Annealing buffer containing 5 pmol of the appropriate sequencing primer. Lastly, this mixture was heated to 95 °C for 2 min and then cooled to and incubated at room temperature for at least 5 min to bind the sequencing primer to the template. PSQ reactions were performed on all post-PCR pools in three replicas, and on all five PCR replicas of all pre-PCR pools. The PCR products were prepared for PSQ using a PSQ96 Sample Prep tool and PSQ reactions were performed using the PSQt96 SNP reagent kit, both steps according to manufacturer's instructions (www.pyrosequencing.com).

V-2) Data analysis: The program PSQt96 evaluation AQ software was used to obtain the ratio of one allele peak height to the sum of height of both allele peaks. To allow the conversion of this peak height ratio into allele frequency, a standard curve based on the individual samples was made. The ratio of one allele peak height to the sum of height of both allele peaks was plotted against the sample frequency of that allele, e.g., in a single sample the frequency of T in a C>T SNP would be 0, 0.5, and 1 for a CC, C/T, and TT genotype, respectively.

V-3) Haplotype analysis: one of the focus of this study is on finding haplotypes, which may contribute to the development of typical sporadic PD. Selected candidate genes in our studies are among enzymes involved in the metabolism of environmental toxins, enzymes involved in the metabolism of dopamine, and genes involved in the familial parkinsonism. The list of candidate genes and their functionalities are provided in Table-2. The information about exact genomic regions analyzed in our study for haplotype association study with PD is summarized in Table-3. In the same table (Table-3), the results of haplotype analysis are also provided.

In regard to finding haplotypes within candidate genes, first, allele frequencies within selected genes were estimated, and conformity to Hardy-Weinberg equilibrium was tested using the $\chi 2$

test with 1 df. Wright's FST statistic was used to estimate the proportion of variation attributable to differences in case and control populations. Then, linkage disequilibrium (LD) between pairs of polymorphic sites was measured using the statistics D' (Lewontin, 1964) and r2 (Hill & Robertson, 1968). Haplotype diversity and frequency for multiple loci were estimated by the expectation-maximization (EM) method using the Alrequin program (Schneider et al. 2000).

In total we identified 3 haplotypes associated with typical sporadic PD in our study (Table-3). No other haplotype within selected candidate genes was associated with typical sporadic PD.

The first haplotype associated with typical sporadic PD is within GSTM1 gene on chromosome 1p13.3 at genomic region 110042530 to 110042975. GTSM1 is a member of the glutathione S-transferases (GST) family, which are enzymes responsible for the metabolism of a broad range of xenobiotics and carcinogens. GTSM1 enzyme catalyzes the reaction of glutathione with a wide variety of organic compounds to form thioethers, a reaction that is sometimes a first step in a detoxification process leading to mercapturic acid formation. Given the role of glutathione S-transferases (GSTs) in the conjugation of
electrophiles and protection against reactive oxygen species, the frequency of single or several polymorphisms within genes encoding the GSTs were examined in several association studies of PD. In total seven studies investigated the role of the GST polymorphisms (GSTM1, GSTT1, GSTP1, and GSTZ1) in the pathogenesis of typical sporadic PD. Three found that polymorphisms of the GSTM1 loci resulting from complete gene depletions increased susceptibility to PD (Stroombergen et al. 1999, De Palma et al. 1998, Ahmadi et al. 2000). One of these studies further examined the presence of null genotypes of GSTM1, GSTT1, and two polymorphisms of mEPHX in subjects with PD, and found raised levels of homozygosity of the histidine (H) 113 isoform of mEPHX (Ahmadi et al. 2000). The remaining four studies did not confirm these results, and did not find any association between GST polymorphisms and the risk of PD. The variability of these findings precluded a definitive conclusion regarding the association of GST polymorphism with typical sporadic PD risk. To our knowledge our study is the first study that investigates and reports an association between a haplotype within GSTM1 gene and typical sporadic Parkinson's disease. In the future studies we're planning to confirm our obtained results in a larger population.

In our study, the second haplotype that was found associated with typical sporadic PD is a halpotype located in the promoter region of alpha-synuclein gene. Alpha-synuclein gene belongs to a family of structurally related proteins that are prominently expressed in the central nervous system. Aggregated alpha-synuclein proteins form brain lesions that are hallmarks of Parkinson's disease. Up to present, three mutations (Ala53Thr, Ala30Pro, Glu46Lys), duplication, and triplication of alpha-synuclein have been identified as causative mutations in familial PD. It has been also reported that SNCA promoter variability may contribute to susceptibility to typical sporadic PD. Pals et al. (2004) found overrepresentation of minimum promoter haplotypes spanning approximately 15.3 kb, in Belgian PD patients. The haplotypes encompassed the Rep1 promoter region but did not rely on Rep1 genotypes. Alleles at NACP-Rep1, the polymorphic microsatellite repeat located approximately 10 kb upstream of the SNCA gene, which acts as a negative modulator of SNCA transcription, were also found to be associated with differing risks of sporadic Parkinson disease (Chiba-Falek

and Nussbaum 2001 and Chiba-Falek et al. 2003). Our study appears to confirm these results as we found overrepresentation of minimum promoter haplotype spanning approximately 15,338bp within the *SNCA* promoter at genomic region 90978070 to 90978910 on chromosome 4q21.

In our study, the third haplotype that was found associated with typical sporadic PD is located within PARK2 gene sequence, which is exactly located at genomic region 162784307 to 162784829 on chromosome 6q25.2-q27. PARK2 gene is involved in protein degradation as a ubiquitin-protein ligase and is mutant in autosomal recessive juvenile parkinsonism. Recently, mutations in the LRRK2 gene, especially the mutation Gly2019Ser, were found to be present in 1.6% of sporadic (Gilks etal. 2005) and 2.8–6.6% of familial PD (Kachergus et al. 2005, Di Fonzo et al. 2005, Nichols et al. 2005). Twin studies seem to contradict hereditary theories for the most common late-onset forms of PD (Duvoisin et al. 1981, Tanner et al. 1999). In contrast, previous studies on familial aggregation suggested a three- to fourfold increased risk in the families of PD patients (Payami et al. 1994, Autere et al. 2000, Kurz et al. 2003). A population-based study with genealogic data of the Icelandic population underlines the genetic component for the late-onset form as for the early-onset form of PD (Sveinbjornsdottir et al. 2000). Recently, strong evidence was reported for a possible role of Parkin gene variations in the late-onset form of PD (age of onset >45 years). Parkin mutations appear to contribute to the common late-onset form and mutations, especially heterozygous mutations, which may play a role as susceptibility alleles for sporadic PD (Oliveira et al. 2003, Lucking et al. 2003, Foroud et al. 2003).

Overall, the results obtained from the current study, support that genetic susceptibility may be important in typical sporadic PD. Furthermore, for the first time we have identified an association between the haplotype located at genomic region 162784307 to 162784829 on chromosome 6q25.2-q27 and typical sporadic PD.

KEY RESEARCH ACCOMPLISHMENTS

- IRB approval was obtained for the proposal
- 200 PD patients and 200 control subjects enrolled in the study, and blood samples have been collected.
- Genomic DNA from all of patients and control samples was been extracted.
- DNA for all collected samples was quantified by spectrophotometer and DNA quality was verified by both Gel electrophoresis and Spectrophotometer.
- We genotyped selected SNPs within 23 candidate genes to find possible haplotype (s) associated with the disease.
- *In total, 3 haplotypes were found to be associated with typical sporadic PD in study.*
- Whole genome screening was performed using 20,000 potentially informative non-synonymous SNPs.
- *Over 20 SNPs were found that showed a high correlation with PD.*

REPORTABLE OUTCOMES

Virtually all of the data outlined under section *IV-2* are reportable. We are in process of writing two manuscripts, which will present these results outline in the previous sections of this final report. The haplotying results are also of interest, and will be reported, even if the studies fail to show associations, as this data would be important for the scientific community.

CONCLUSIONS

This whole genome study using non-synonymous SNPs to search for genetic factors in PD has proved to be exceedingly rewarding. As noted above, we found 22 SNPs that showed a highly significant association with PD. The whole genome screening performed in this study is only the second ever whole genome screening in PD, and the first ever using non-synonymous SNPs, which are much more likely to have functional consequences, and therefore be related to disease causation. Furthermore, using hapoltype analysis of candidate genes we identified three haplotypes associated with typical sporadic PD. While these findings need to be replicated using different techniques, and in different PD populations, they appear to be very promising in terms of furthering our efforts to understand the genetic contributions to PD.

REFERENCES

Ahmadi A, Fredrikson M, Jerregard H, Akerback A, Fall PA, Rannug A et al. GSTM1 and mEPHX polymorphisms in Parkinson's disease and age of onset. *Biochem. Biophys. Res. Commun.* 269: 676–680, 2000.

Autere JM, Moilanen JS, Myllyla VV, Majamaa K. Familial aggregation of Parkinson's disease in a Finnish population. *J. Neurol. Neurosurg Psychiatry* 69:107–9, 2000.

Chiba-Falek O, Nussbaum RL. Effect of allelic variation at the NACP-Rep1 repeat upstream of the alpha-synuclein gene (SNCA) on transcription in a cell culture luciferase reporter system. *Hum. Molec. Genet.* 10: 3101-3109, 2001.

Chiba-Falek O, Touchman JW, Nussbaum RL. Functional analysis of intra-allelic variation at NACP-Rep1 in the alpha-synuclein gene. *Hum. Genet.* 113: 426-431, 2003.

De Palma G, Mozzoni P, Mutti A, Calzetti S, Negrotti A. Case–control study of interactions between genetic and environmental factors in Parkinson's disease. *Lancet* 352: 1986–1987, 1998.

Di Fonzo A, Rohe CF, Ferreira J et al. A frequent LRRK2 gene mutation associated with autosomal dominant Parkinson's disease. *Lancet* 365: 412–5, 2005.

Duvoisin RC, Eldridge R, Williams A, Nutt J, Calne D. Twin study of Parkinson disease. *Neurology* 31: 77–80, 1981.

Foroud T, Uniacke SK, Liu L et al. Heterozygosity for a mutation in the parkin gene leads to later onset Parkinson disease. *Neurology* 60: 796–801, 2003.

Gilks WP, Abou-Sleiman PM, Gandhi S et al. A common LRRK2 mutation in idiopathic Parkinson's disease. *Lancet* 365: 415–6, 2005.

Grevle L, Guzey C, Hadidi H, Brennersted R, Idle JR, Aasly J. Allelic association between the DRD2 TaqI A polymorphism and Parkinson's disease. *Mov Disord* 15: 1070–1074, 2000.

Hartl DL, Clark AG. Principles of population genetics, 3rd editions. Sinauer Associates, Sunderland. 1997

Hill WG, Robertson A. Linkage disequilibrium in finite populations. *Theor. Appl. Genet* 38: 226–231, 1968.

Kachergus J, Mata IF, Hulihan M et al. Identification of a novel LRRK2 mutation linked to autosomal dominant parkinsonism: evidence of a common founder across European populations. *Am. J. Hum. Genet.* 76: 672–80, 2005.

Kurz M, Alves G, Aarsland D, Larsen JP. Familial Parkinson's disease: a community-based study. *Eur. J. Neurol.* 10: 159–63, 2003.

Le WD, Xu P, Jankovic J, Jiang H, Appel SH, Smith RG et al. Mutations in NR4A2 associated with familial Parkinson disease. *Nat Genet* 33: 85–89, 2003.

Lewontin RC. Interaction of selection and linkage. I. General considerations; heterotic models. *Genetics* 49: 49–67, 1964.

Lucking CB, Chesneau V, Lohmann E et al. Coding polymorphisms in the parkin gene and susceptibility to Parkinson disease. *Arch. Neurol.* 60: 1253–6,2003.

Markopoulou K Langston JW. Candidate genes and Parkinson's disease. *Neurology* 53: 1382-1383, 1999

Marx FP, Holzmann C, Strauss KM, Li L, Eberhardt O, Gerhardt E, Cookson MR et al. Identification and functional characterization of a novel R621C mutation in the synphilin-1 gene in Parkinson's disease. *Hum. Molec. Genet.* 12: 1223-1231, 2003.

Mellick GD, Buchanan DD, McCann SJ, James KM, Johnson AG, Davis DR et al. Variations in the monoamine oxidase B (MAOB) gene are associated with Parkinson's disease. *Mov. Disord.* 14: 219–224, 1999.

Nicholl DJ, Bennett P, Hiller L, et al. A study of five candidate genes in Parkinson's disease and related neurodegenerative disorders. European Study Group on Atypical Parkinsonism. *Neurology* 53: 1415-1421, 1999.

Nichols WC, Pankratz N, Hernandez D et al. Genetic screening for a single common LRRK2 mutation in familial Parkinson's disease. *Lancet* 365: 410–2, 2005.

Oliveira SA, Scott WK, Martin ER et al. Parkin mutations and susceptibility alleles in late-onset Parkinson's disease. *Ann. Neurol.* 53: 624–9, 2003.

Pals P, Lincoln S, Manning J, Heckman M, Skipper L, Hulihan M, Van den Broeck M, De Pooter T, Cras P, Crook J, Van Broeckhoven C, Farrer MJ. Alpha-synuclein promoter confers susceptibility to Parkinson's disease. *Ann. Neurol.* 56: 591-595, 2004.

Payami H, Larsen K, Bernard S, Nutt J. Increased risk of Parkinson's disease in parents and siblings of patients. *Ann. Neurol.* 36: 659–61,1994.

Schneider S, Roessli D, Excoffier L. (2000) Arlequin: A software for population genetics data analysis . Ver 2.000. Genetics and Biometry Lab, Dept of Anthropology, University of Geneva.

Stroombergen MCMJ, Waring RH. Determination of glutathione S-transferase m and y polymorphisms in neurological disease. *Hum. Exp. Toxicol.* 18: 141–145, 1999.

Sveinbjornsdottir S, Hicks AA, Jonsson T et al. Familial aggregation of Parkinson's disease in Iceland. *N. Engl. J. Med.* 343: 1765–70, 2000.

Tanner CM, Ottman R, Goldman SM et al. Parkinson disease in twins: an etiologic study. *JAMA* 281:341–6, 1999.

Warner TT, Schapira AH. Genetic and environmental factors in the cause of Parkinson's disease. *Ann. Neurol.* 53: 16–23,2003.

Xu PY, Liang R, Jankovic J, Hunter C, Zeng YX, Ashizawa T et al. Association of homozygous 7048G7049 variant in the intron six of Nurr1 gene with Parkinson's disease. *Neurology* 58: 881–884, 2002.

Zheng K, Heydari B, Simon DK. A common NURR1 polymorphism associated with Parkinson disease and diffuse Lewy body disease. *Arch Neurol* 60: 722–725, 2003.

List of personnel receiving pay from the research effort

Dr. J. William Langston, MD., Principal Investigator, and Julie Doostzadeh, Ph.D., Staff Scientist, received salary support on this research study.

APPENDICES

TABLE-1. cSNPs associated with Parkinson's disease. Assay number, ID numbers and p-values, chromosomal locations, exact positions of each cSNPs, and gene information for cSNPs associated with Parkinson's disease are provided in this table.

Assay ID	P-Value	SNP ID	Chromosome	Position (bp)	Gene
312845	1.58E-10	rs16891982	5	33,987,450	MATP
315658	1.06E-06	rs11007567	10	29,750,643	LOC387647
178250	1.16E-06	rs11007612	10	29,810,557	SVIL
289111	1.20E-06	rs11600	10	29,750,946	LOC387647
319061	2.28E-06	rs12505221	4	141,740,726	LOC152586
290479	3.82E-06	rs10510114	10	124,510,729	FLJ46361
180801	5.86E-06	rs4915691	1	65,579,540	DNAJC6
288088	7.23E-06	rs153478	5	150,619,632	GM2A
174464	1.03E-05	rs3818717	17	17,647,830	RAI1
184695	1.78E-05	rs25695	9	71,579,400	TMEM2
179382	2.15E-05	rs1320308	5	76,207,008	S100Z
315115	2.44E-05	rs5952606	X	44,263,631	LOC392451
185735	2.73E-05	rs26722	5	33,999,627	MATP
179194	2.90E-05	rs1438307	2	136,332,898	ENSG00000176119.1
305080	2.95E-05	rs2280523	3	161,955,155	LOC389172
179491	3.10E-05	rs11259332	10	14,856,454	C10orf45
178143	4.21E-05	rs11168070	5	148,186,120	ENSG00000169256.1
323875	4.25E-05	rs9455190	6	71,374,681	LOC442230
182495	4.61E-05	rs1799999	7	113,112,385	PPP1R3A
181217	5.95E-05	rs6022903	20	52,045,956	BCAS1
180700	6.92E-05	rs7309681	12	110,800,594	ENSG00000178555.6
326919	7.18E-05	rs12231744	12	110,939,775	ENSG00000111300

TABLE-2: The list of candidate genes and their functionalities.

Function	Gene	Name
Enzymes involved in the metabolism of environmental toxins	**CYP2D6**	Cytochrome P450 2D6
	CYP2E1	Cytochrome P450 2E1
	CYP1A1	Cytochrome P450 1A1
	GSTM1	Glutathione S-Transferase M1
	GSTP1	Glutathione S-Transferase P1
	GSTT1	Glutathione S-Transferase T1
	GSTZ1	Glutathione S-Transferase Z1
	PON1	Paraoxonase 1
	NAT2	N-acetyltransferases
Enzymes involved in the metabolism of dopamine	MAOA	Monoamine Oxidase A
	MAOB	Monoamine Oxidase B
	COMT	Catechol-O-methyltransferase
	DAT	Dopamine Transporter
	VMAT2	Vesicular Monoamine Transporter-2
	DRD1	Dopamine Receptor D1
	DRD2	Dopamine Receptor D2
	DRD3	Dopamine Receptor D3
	DRD4	Dopamine Receptor D4
	DRD5	Dopamine Receptor D5
Genes involved in familial PD	SNCA	Alpha Synuclein
	UCHL1	Ubiquitin Carboxyl-Terminal Esterase L1
	PINK1	PTEN-Induced Putative Kinase1
	PARK7	DJ1
	PARK2	Parkin

TABLE-3: The information about chromosomal locations, genomic regions covered in our study for haplotype analysis, and associated haplotypes with PD is provided in this table.

Gene Name	Chromosome	Genomic Region Covered in Haplotype Analysis	Associated Haplotype with PD
CYP2D6	22	40848326 - 40871229	No
CYP2E1	10	135193728 -135202090	No
CYP1A1	15	72788283 -72806502	No
GSTM1	1	110024530 -110042975	110042530 - 110042975
GSTP1	11	67108832 -67110155	No
GSTT1	22	22676550 -22732321	No
GSTZ1	14	76860625 -76866394	No
PON1	7	94765613 -94790628	No
NAT2	8	18296349 -18302274	No
MAOA	X	43411887 -43477666	No
MAOB	X	43512943 -43611338	No
COMT	11	18310109 -18335692	No
DAT	5	1447522 -1495521	No
VMAT2	13	118992657 -119022265	No
DRD1	5	174800505 -174802802	No
DRD2	11	112786283 -112846601	No
DRD3	3	115340891 - 115374239	No
DRD4	11	626399 -631191	No
DRD5	4	9388678 -9397033	No
SNCA & Promoter region of SNCA	4	90865909 -90995188	90978070 - 90978910
UCHL1	4	40959306 -40963950	No
PINK1	1	20835414 --20844635	No
PARK7	1	7947179 -7963939	No
PARK2	6	161687527 -163069487	162784307 – 162784829