AWARD NUMBER: W81XWH-20-1-0709

TITLE: Integrating Environmental, Genomic, and Functional Data to Characterize Ir

PRINCIPAL INVESTIGATOR: Samuel Goldman, MD, MPH

CONTRACTING ORGANIZATION: Northern California Institute for Research (NCIRE

REPORT DATE: August 2023

TYPE OF REPORT: Annual

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

DISTRIBUTION STATEMENT: Approved for Public Release;
Distribution Unlimited

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
AUGUST 2023	Annual	1AUG2022 - 31JUL2023

4. TITLE AND SUBTITLE

Integrating Environmental, Genomic, and Functional Data to Characterize Individual Risk for Parkinson's Disease

5a. CONTRACT NUMBER
W81XWH-20-1-0709

5b. GRANT NUMBER
W81XWH-20-1-0709

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Samuel M. Goldman, MD, MPH

E-Mail: samuel.goldman@ucsf.edu

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

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

NORTHERN CALIFORNIA INSTITUTE FOR RESEARCH NCIRE
4150 CLEMENT ST 151 NC
SAN FRANCISCO CA 94121-1563

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

U.S. Army Medical Research and Development 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

14. ABSTRACT

The overall purpose of this work is to characterize which individuals may be susceptible to specific toxic insults, and the underlying mechanisms. We have three Aims: 1) use iPSC-derived neuronal cell lines to validate and extend our prior work that found an 11-fold increased risk of Parkinson's disease (PD) in persons exposed to the herbicide paraquat who lacked functional glutathione-S-transferase T1 (GSTT1), a key metabolic enzyme with anti-oxidant function; 2) perform whole genome sequencing of existing DNA specimens from a PD study with highly characterized pesticide exposures, using bioinformatic tools to identify likely functional gene-environment interactions; and 3) test our predictions using iPSC-derived neuronal cell lines that model these genetic variations, and exposing cellular lines to these same environmental agents. This is a partnering PI project with W81XWH-20-1-0710. The scope of project -0709 is to accomplish Aim 2, and is the focus of this report.

15. SUBJECT TERMS

Parkinson's disease; genetics; environment; pesticides; toxicants; gene-environment interaction

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			USAMRDC
U	U	U	UU	18	19b. TELEPHONE NUMBER *(include area code)*

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39.18

Table of Contents

I. INTRODUCTION

Parkinson's disease (PD) is a growing public health problem. As many as 1.5 million people in the U.S. currently suffer from PD, with an estimated 70,000 newly diagnosed cases per year. Veterans may be at increased risk, and more than 100,000 veterans with PD are currently treated in the Veterans Health Administration (VHA). Highly penetrant genetic causes of PD are uncommon. Numerous lines of research implicate a role for environmental causes, yet disease clusters are rarely identified. The vast majority of PD is therefore likely due to the deleterious effects of environmental factors on a background of genetic susceptibility, so-called gene-environment interaction (G*E). G*E is measured epidemiologically by the statistical deviation of the joint effects from independence. This statistical deviation is assumed to reflect an underlying biological interaction, but this assumption requires verification in valid biological models. The current project explores G*pesticide interaction in a highly exposed well-characterized cohort of professional pesticide applicators. Our aims are to 1) validate the epidemiologically observed G*E interaction in a cellular model; 2) identify novel, likely causal G*E interactions of particular relevance for service members; and 3) validate these novel interactions in cellular models.

II. KEYWORDS

Parkinson's disease; genetics; environment; pesticides; toxicants; gene-environment interaction

III. ACCOMPLISHMENTS

This is a collaborative (dual PI) project. The PI for project W81XWH-20-1-0709 is Samuel Goldman, MPH, MD; and the PI for project W81XWH-20-1-0710 is Raymond Swanson, MD. Project W81XWH-20-1-0710 has been completed, and a final report for that project is in preparation. Project W81XWH-20-1-0709 has received a no-cost extension through July 31, 2024, and is the focus of this Annual report. Only joint tasks and milestones, and those tasks and milestones specific to project -0709 are included in the current Annual report.

A. What were the major goals of the project?

Major Task 1: Implement procedures and methods for genetic data analysis and cell culture models [Swanson & Goldman]

Subtask 1.1: Establish Study Team Processes, regular meetings, train all study staff (months 1-3).
Subtask 1.2: Develop study databases, data security, quality assurance methods (months 3-5).
Subtask 1.3: Use the WTC11 iPSC line to model GSTT1 deficiency (months 6-12).

Major Task 2: Determine the consequences of GSTT1 loss of function for susceptibility to paraquat injury [Swanson]

Subtask 2.1: Differentiate at least 9 iPSC lines (3 each) with GSTT1 intact, down regulated, or deleted into dopaminergic (DA) neurons. (months 12-18)

Subtask 2.2: Treat the 9 lines of differentiated DA neurons with paraquat and other putative PD-related toxicants likely to be encountered in military settings (permethrin, 2,4-D, trichloroethylene) or left untreated. (months 15-20)
Subtask 2.3: In the lines of differentially expressing GSTT1 DA neurons, assess reactive oxygen species (ROS) generation, ROS sensitivity, a-synuclein levels and aggregation, neurite retraction and neurodegeneration. (months 18-24)

Milestone #1: *Development and characterization of GSTT1- deficient cell line (months 12-24)*

Major Task 3: Determine if GSTT1 overexpression can rescue neurodegeneration. (months 18-24) [Swanson]

Subtask 3.1: Using similar approaches as above, determine whether ectopic expression of GSTT1 slows or blocks toxicant-induced or synuclein-dependent neurodegeneration in the context of GSTT1 deletion and reverses ROS and aggregation phenotypes. This will be accomplished using transfection of the 9 previously generated cell lines
Subtask 3.2: Using similar approaches as above, in 9 lines determine whether treatment with antioxidants slows or blocks toxicant-induced or synuclein-dependent neurodegeneration in the context of GSTT1 deletion and reverses ROS and aggregation phenotypes.

Major Task 4: Whole-genome sequencing of 265 FAME study subject specimens [Goldman]

Subtask 4.1: prepare high quality aliquots of banked DNA from 265 study subjects to specified concentrations (months 4-6).
Subtask 4.2: DNA specimens from 265 subjects will be sequenced, and sequence data stored on multiple high capacity drives, and data uploaded to UCSF/Gladstone secure servers (months 7-12).

Milestone #2: *FAME sequencing completed (months 6-12)*

Major Task 5: Quantification of pesticide and other toxicant exposures within the FAME study database comprised of 498 study subjects in total (115 case and 383 controls) [Goldman]

Subtask 5.1: classify historical exposures for 498 FAME subjects, derive cumulative exposure estimates by duration and intensity for each agent (months 9-15)
Subtask 5.2: classify exposures for 498 FAME subjects according to mechanistic classes (e.g., mitochondrial Complex I inhibitors; redox-cycling/oxidative stressors; a-synuclein pro-aggregants) (months 12-16).
Subtask 5.3: classify serum toxicant exposures 498 FAME subjects (months 14-18).

Major Task 6: perform rare burden analysis and machine learning to identify genes likely to increase susceptibility to specific toxicants in FAME, pooling data from the 265 subjects sequenced in the current project with existing sequencing data for 70 subjects, for a total of 335 subjects [Goldman]

Subtask 6.1: Annotation of sequencing data for 335 FAME subjects (months 14-18)
Subtask 6.2 rare-burden analysis of variants for 335 FAME subjects (months 18-30)
Subtask 6.3 conduct exposed-only analyses of gene*environment interaction for 335 FAME subjects (months 20-30)
Subtask 6.4: use Machine Learning approaches for 335 FAME subjects to identify genes that may contribute to neurodegeneration in Parkinson's disease (months 24-30)

Milestone #3: *identification of combinations of variants likely to increase susceptibility to specific toxicants (months 24-30)*

Major Task 7: Validate novel G*E interactions relevant to PD pathogenesis [Swanson]

Subtask 7.1: lower the expression of the genes in human DA neurons and determine whether that confers sensitivity to specific toxicants, using the same paradigm and 9 cell lines as for Major Tasks 1 & 2 above (months 20-30)
Subtask 7.2: Determine the robustness of validated G*E interactions using additional cell lines. We will generate > 3 lines for each gene of interest. (months 24 -32)
Subtask 7.3: Using the lines above, demonstrate that specific genetic variants identified in FAME are sufficient to mediate G*E interaction. (months 32-36)

Milestone #4: *Manuscript on use of the identification and validation of specific genetic susceptibility to specific toxicants (months 32-36).*

B. What was accomplished under these goals?

Major Task 1: Implement procedures and methods for genetic data analysis and cell culture models [Swanson & Goldman]

Subtask 1.1: Establish Study Team Processes, regular meetings, train all study staff (months 1-3). [Swanson & Goldman]

Accomplishments: This Subtask is completed. We established regular meetings within the Goldman group, and between the Goldman, Swanson & Gladstone (subcontract) groups. These meetings have been held biweekly since August, 2020. All study staff are trained in study procedures and human subjects requirements of each of the associated institutions, with regular training updates as required.

Subtask 1.2: Develop study databases, data security and quality assurance methods (months 3-5). [Swanson & Goldman]

Accomplishments: This Subtask is completed. Study databases have been developed at Gladstone, and essential de-identified datasets are accessible to appropriate study staff. All databases are regularly backed up at Gladstone and stored on a secured server. Genomic sequencing data reside on redundant devices to ensure integrity and protect from any unlikely data loss. Intensive quality assurance procedures were implemented during DNA processing, prior to and during sequencing, and throughout

the annotation pipeline. The genomic data is stored at Gladstone is fully de- identified and maintained separately from the patient sample data at UCSF.

Major Task 4: Conduct whole-genome sequencing of 265 FAME study subject specimens [Goldman]

Subtask 4.1: prepare high quality aliquots of banked DNA from 265 study subjects to specified concentrations (months 4-6). [Goldman]

Accomplishments: This Subtask is completed. Banked DNA for 285 study subjects was aliquoted after assessment for quality and concentration, and prepped according to the sequencing contractor's specifications. Additional aliquots were prepped to account for instances of possible low DNA quality or other potential QA failures.

Subtask 4.2: DNA specimens from 265 subjects will be sequenced, and sequence data stored on multiple high capacity drives, and data uploaded to UCSF/Gladstone secure servers (months 7-12). [Goldman]

Accomplishments: This Subtask is completed. Genomic DNA was successfully sequenced for 270 subjects, 5 more than our target goal. Sequencing data quality was reviewed, and data uploaded to secure servers at UCSF/Gladstone.

Milestone #2: FAME sequencing completed (months 6-12)
Accomplishments: Milestone #2 was fully completed on schedule.

Major Task 5: Quantification of pesticide and other toxicant exposures within the FAME study database comprised of 498 study subjects in total (115 case and 383 controls) [Goldman]

Subtask 5.1: classify historical exposures for 498 FAME subjects, derive cumulative exposure estimates by duration and intensity for each agent (months 9-15)

Accomplishments: Cumulative exposure estimates are derived.

Subtask 5.2: classify exposures for 498 FAME subjects according to mechanistic classes (e.g., mitochondrial Complex I inhibitors; redox-cycling/oxidative stressors; a-synuclein pro-aggregants) (months 12-16).

Accomplishments: This Subtask is completed, though we continue to regularly search the scientific literature for additional insights regarding pesticide toxicologic mechanisms and potential mechanisms of specific relevance to PD etiology.

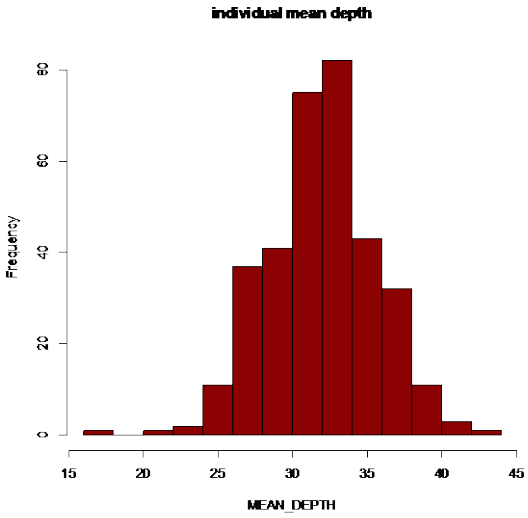
Subtask 5.3: classify serum toxicant exposures 498 FAME subjects (months 14-18).

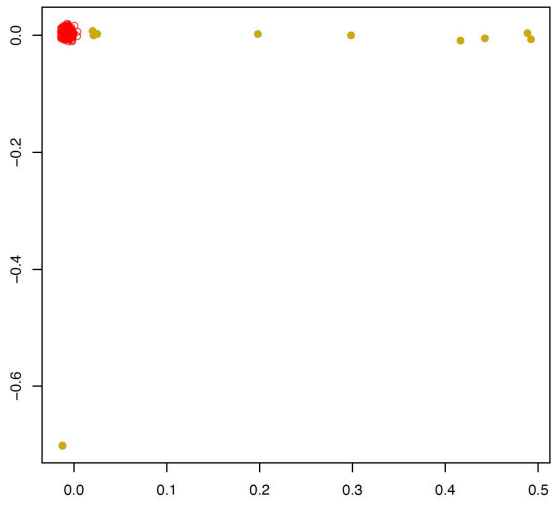
Accomplishments: We have derived lipid-adjusted values of previously measured persistent organic toxicants including polychlorinated biphenyl compounds and organochlorine pesticides. We used a beta-substitution method to impute values below the limits of detection.

Major Task 6: perform rare burden analysis and machine learning to identify genes likely to increase susceptibility to specific toxicants in FAME, pooling data from the 265 subjects sequenced in the current project with existing sequencing data for 70 subjects, for a total of 335 subjects [Goldman]

Subtask 6.1: Annotation of sequencing data for 335 FAME subjects (months 14-18)

Accomplishments: We have completed the QA and annotation of sequencing data for 340 FAME study subjects. Median sequencing depth per individual was 32 (see Figure below). Only a single individual with depth below 20 was excluded from analyses. Principal components analysis (see Figure below) identified a handful of (already self-identified) non-white individuals in this predominantly white European-descent study cohort, as well as a highly-related pair that necessitated the removal of one subject from analyses. Variants have been annotated as high impact (predicted loss of function, missense, n=1,233), low impact (n=182,133) or modifier (22,241,663). Rare and common variants (n=7,924,634) have been annotated as such.





Subtask 6.2: rare-burden analysis of variants for 335 FAME subjects (months 18-30)
Accomplishments: We have conducted rare-burden analyses (minor allele frequency < 1%). We have analyzed both rare high-impact variants, and pooled rare and (weighted) common high-impact variants (predicted loss of function, missense) using SKAT. The tables below summarize the most highly-associated genes among both candidate and agnostic genes.

All Genes	RANGE	NumVar	Pvalue
MIR548N	2:179246804-179541009	228	3.52 E-09
TTN	2:179610046-179672150;2:179390717-179672150;2:179390717-179672150;2:179390717-179672150;2:179390717-179672150	364	1.42 E-08
LAMA5	20:60884120-60942368	87	4.36 E-08
HSPG2	1:22148736-22263750	78	2.67 E-06
OBSCN	1:228395860-228548951;1:228395860-228566575	95	7.85 E-06
MUC16	19:8959519-9092018	150	8.44 E-06
MACF1	1:39796809-39952810;1:39547088-39952810	85	8.19 E-05

Subtask 6.3: Conduct exposed-only analyses of gene*environment interaction for 335 FAME subjects (months 20-30)

Accomplishments: We have conducted association analyses in pesticide exposed-only sub-cohorts for applicators of 1) rotenone, 2) paraquat, 3) permethrin, 4) benomyl, 5) dieldrin, 6) paraquat or diquat, 7) any pyrethroid, 8) dieldrin or aldrin, 9) any benzimidazole. Tests of interaction utilize SKAT-based burden tests for rare and (weighted) common variation across both a) apriori candidate genes involved in xenobiotic metabolism, transport, and glutathione metabolism, and PARK genes; and b) among all genes (i.e., agnostic analyses). The tables below include the statistically strongest interactions. These data will be presented at the August 2023 International Congress of Parkinson's Disease and Movement Disorders.

Exposed-Only Gene-Pesticide Interactions.

Pesticide Usage	Case n Exposed	Control n Exposed	Most Significant Genes*			
			Unselected		*A priori* Candidate	
			Gene	p-value	Gene	p-value
Rotenone	18	19	*FSCB* *SLC14A1* *SIGLEC12* *OR9Q1*	p=0.00018 p=0.00025 p=0.00035 p=0.00038	*BMP2* *VPS13C* *RDH12* *SLC1A2*	p=0.012 p=0.014 p=0.018 p=0.019
Paraquat	18	44	*SLC15A3*	p=0.00035	*BMP2* *ABCC8* *DHRS9* *POR* *GPX8*	p=0.0037 p=0.004 p=0.01 p=0.013 p=0.016
Permethrin	13	43	*CCDC110* *SIGLEC8* *EAPP* *GTF2B* *LRIT1* *BOC* *GTF2A1L*	p=1.9 x 10-5 p=2.2 x 10-5 p=3.1 x 10-5 p=4.5 x 10-5 p=6.3 x 10-5 p=7.6 x 10-5 p=9.7 x 10-5	*ABCC4^* *SLC1A1* *UGT2B4* *CYP2C16*	p=0.0008 p=0.003 p=0.01 p=0.01
Dieldrin	12	28	*ZNF471* *HRC* *NKX3-1* *SDR39U1*	p=3.1 x 10-5 p=0.0002 p=0.00025 p=0.00025	*ABCB10* *SLC7A11* *RDH5* *SLC25A22*	p=0.0029 p=0.0049 p=0.0049 p=0.005

			LSM11 *CTSG*	p=0.00033 p=0.00036	*SLC1A7*	p=0.01

*SKAT (Sequence Kernel Association Test) gene-burden tests of high impact variants
^significant interaction after Bonferroni correction

Frequencies of missense variants in exposed-only analyses.

Pesticide	Gene	Candidate Gene or Unselected	Variant Position (hg19)	Allele Frequency	
				Case	Control
Rotenone	*VPS13C (PARK23)*	Candidate: maintains mitochondrial transmembrane potential, regulates PINK1/ Parkin-mediated mitophagy	chr15:62219337	0.028	0
			chr15:62243187	0.028	0
			chr15:62254989	0.056	0
			chr15:62316035	0.056	0
			Any	0.17	0
	FSCB	Unselected	chr14:44975511	0.39	0.056
Paraquat	*VPS13C (PARK23)*	Candidate: maintains mitochondrial transmembrane potential, regulates PINK1/ Parkin-mediated mitophagy	chr15:62211600	0.028	0
			chr15:62219337	0.11	0.023
			chr15:62261612	0.028	0
			chr15:62283291	0.028	0
			Any	0.19	0.023
	SLC15A3	Unselected	chr11:60708672	0.028	0
			chr11:60709520	0.083	0
			chr11:60718792	0.17	0.034
			Any	0.28	0.034
Permethrin	*ABCC4*	Candidate: multidrug resistance protein; xenobiotic transporter	chr13:95735484	0.038	0
			chr13:95953517	0.11	0
			Any	0.15	0
	LRIT1	Unselected	chr10:85992390	0.19	0
Dieldrin	*ABCB10*	Candidate: mitochondrial membrane transporter	chr1:229654053	0.083	0
			chr1:229665958	0.042	0
			Any	0.13	0
	CTSG	Unselected	chr14:25043671	0.17	0

In addition, we have taken advantage of existing GWAS SNP data in this population to help inform next steps for in vitro modeling. Specifically, we performed imputation on approximately 750,000 SNPs typed on the Affymetrix UKBrainbank array. We have identified all SNP*pesticide-use interactions that are nominally statistically significant in either additive or binary models. In addition to hypothesis-free "agnostic" analyses, we have also explored pesticide interactions with genes and variants of *a priori* interest: 1) PARK genes, 2) PD GWAS-associated SNPs, and 3) genes involved in pesticide metabolism and transport. Selected preliminary results are reported in the tables below.

A priori Gene	Pesticide	dbSNPRSID	Location	OR_GE	OR_E_noG	OR_G_noE	OR_Inter-action	p-val
ABCB9	benzimidazole	rs116887147	Intron	15.62	0.96	1.20	13.55	0.031
ABCC1	pyrethroid	rs188352772	Intron	6.43	0.99	1.21	5.36	0.031
ABCC1	pyrethroid	rs79686715	Intron	12.62	1.30	1.05	9.22	0.040
ABCC1	benzimidazole	rs17287570	Intron	3.10	0.93	0.83	4.01	0.041
ABCC4	Dieldrin	rs4148493	Intron	3.58	1.11	0.81	3.99	0.015
ABCC4	rotenone	rs3742106	Utr3	6.50	1.91	0.95	3.56	0.031
ABCC4	benzimidazole	rs7324065	Intron	11.73	1.03	0.84	13.60	0.033
ABCC4	Blazer	rs78302301	Intron	11.49	0.97	0.87	13.64	0.036
ABCC4	Cyanides	rs9561784	Intron	5.03	0.86	1.11	5.29	0.038
ABCC4	rotenone	rs4148551	Utr3	6.70	2.10	1.02	3.13	0.050
ABCC6	permethrin	rs16967441	Intron	6.25	1.25	0.90	5.57	0.016
ABCC6	Ferbam	rs12598559	Intron	5.10	1.00	1.10	4.63	0.047
ABCC8	Dieldrin	rs4148618	Intron	4.70	1.21	0.90	4.32	0.020
GCLC	Blazer	rs16883912	Intron	4.36	0.81	0.88	6.17	0.016
GCLC	Benomyl	rs16883912	Intron	19.91	0.90	1.00	22.24	0.022
GCLC	pyrethroid		Intron	22.20	1.29	1.57	10.96	0.040
GLRX	rotenone	rs7700814	Intron	16.10	2.25	1.15	6.24	0.023
GLRX	rotenone	rs885303	Intron	6.48	1.77	0.85	4.28	0.025
GLRX	rotenone	rs74587548	Intron	35.48	2.48	1.25	11.49	0.035
GLRX	Ferbam	rs74587548	Intron	8.60	1.11	1.41	5.47	0.049
GSTM1	Aldrin	rs1056806	Synon	4.26	1.04	0.81	5.06	0.033
MGST1	Ferbam	rs73064116	Utr3	6.05	0.97	0.98	6.36	0.011
MGST1	benzimidazole	rs9332929	Intron	3.40	0.85	1.00	3.99	0.038
MGST1	Ferbam	rs10846355	Intron	4.14	0.99	1.14	3.67	0.039
NQO1	benzimidazole	rs1800566	**Nonsynon**	4.64	0.84	1.14	4.82	0.028
NXN	Ferbam	rs7223906	Intron	17.15	1.13	1.31	11.52	0.027
NXN	Blazer	rs17693812	Intron	2.86	0.91	0.92	3.43	0.048
SLC1A1	paraquat	rs10974575	Intron	7.01	1.67	0.86	4.90	0.023
SLC1A4	Benomyl	rs2268483	Intron	15.48	1.00	0.88	17.52	0.011
SLC1A4	Ferbam	rs2268483	Intron	5.26	1.04	0.85	5.94	0.032
SLC7A11	Dieldrin	rs4863768	Intron	3.18	1.02	0.85	3.67	0.022
SLC7A11	permethrin	rs72712336	Intron	7.91	1.30	1.51	4.04	0.037
TXNDC8	pyrethroid	rs7041938	**Nonsynon**	4.83	1.29	1.06	3.56	0.028
TXNRD1	Dieldrin	rs11111945	Intron	5.96	1.37	0.82	5.35	0.030
TXNRD2	Benomyl	rs7410379	Intron	5.83	0.83	0.93	7.52	0.026

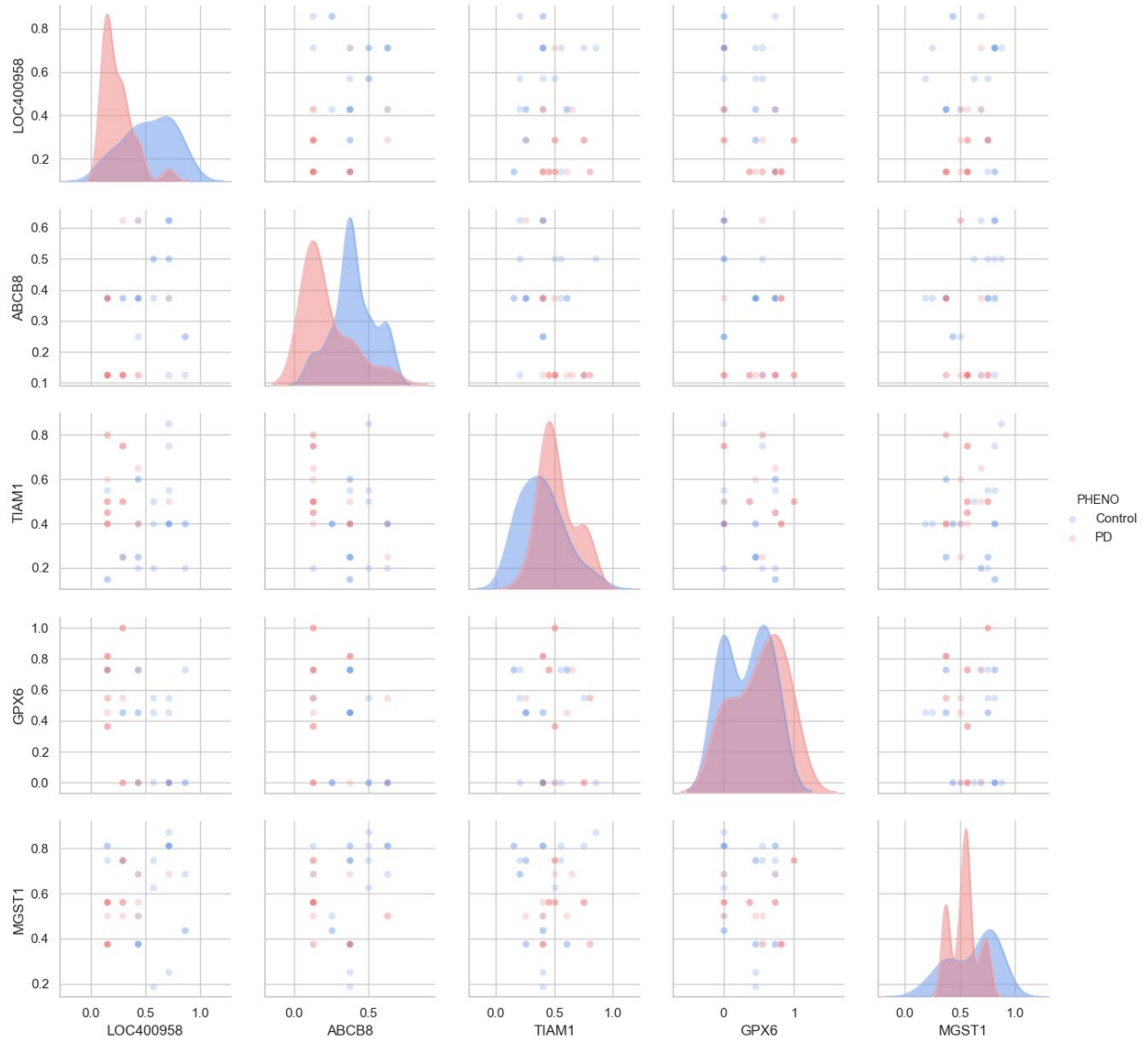
PARK Gene	Pesticide	dbSNPRSID	Location	OR_GE	OR_E_ noG	OR_G_ noE	OR_ Inter-action	p-val
PINK1	benzimidazole	rs148871409	**Nonsynon**	7.93	0.99	1.08	7.40	0.038
ATP13A2	fungicide	rs3738815	Synon	2.98	0.86	0.84	4.12	0.027
PARK7	Dieldrin	rs4908488	Intron	5.37	1.52	0.95	3.70	0.047
GIGYF2	permethrin		Intron	16.44	1.59	1.78	5.80	0.037
SNCA	rotenone	rs11097234	Intron	6.11	2.26	0.83	3.26	0.038
SNCA	rotenone	rs12502363	Intron	5.98	2.21	0.81	3.33	0.035
LRRK2	Blazer	rs7308720	**Nonsynon**	3.31	0.91	0.97	3.77	0.045
LRRK2	Blazer	rs10878372	Intron	2.60	0.83	1.03	3.06	0.043
LRRK2	rotenone	rs17466339	Intron	9.66	2.53	0.82	4.67	0.035
PARK2	Benomyl	rs9355360	Intron	4.83	0.89	1.05	5.12	0.049
PARK2	fungicide	rs58468575	Intron	2.68	0.92	0.85	3.39	0.047
PARK2	Cyanides	rs78224461	Intron	12.12	1.03	1.09	10.72	0.044
PARK2	Blazer	rs9458273	Intron	2.27	0.84	0.80	3.36	0.043
PARK2	Dieldrin	rs9456751	Intron	2.82	1.06	0.82	3.24	0.037
PARK2	Ferbam	rs3019443	Intron	3.17	0.86	0.87	4.22	0.032
PARK2	rotenone	rs9346879	Intron	4.84	1.35	0.88	4.07	0.030
PARK2	fungicide	rs11966738	Intron	2.83	0.97	0.86	3.36	0.030
PARK2	Benomyl	rs112913800	Intron	16.11	1.04	1.18	13.15	0.020

Subtask 6.4: use Machine Learning approaches for 335 FAME subjects to identify genes that may contribute to neurodegeneration in Parkinson's disease (months 24-30)

Accomplishments: We have implemented machine learning models to predict PD probability within pesticide-exposed subcohorts using genome-wide gene burden tests. Preliminary models based on GWAS data are encouraging, with ROC AUCs approaching 0.85 for some agents.

Permethrin, candidate, AUC = 0.80

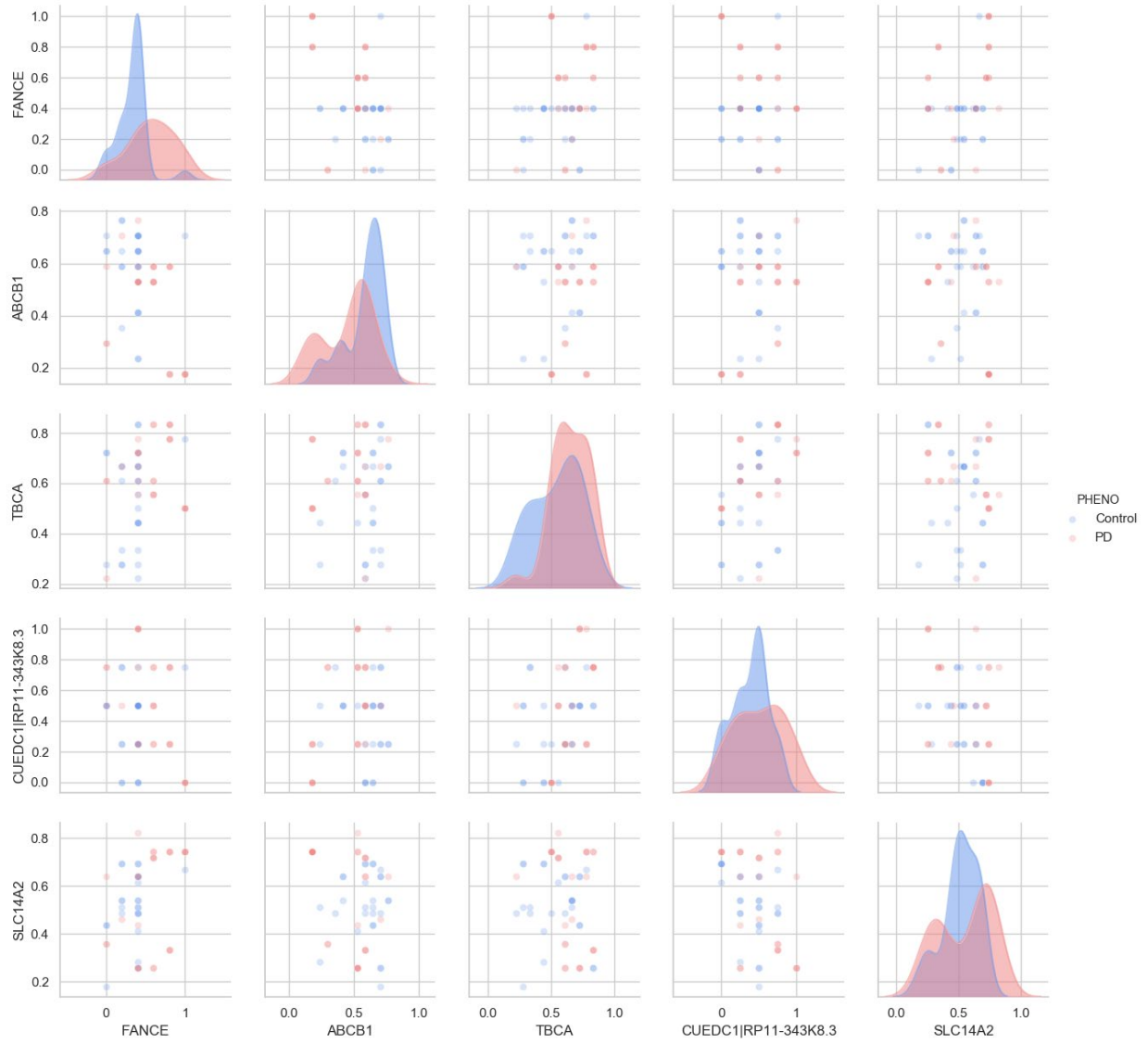
feature	importance	RVIS	RVIS_perc
LOC400958	0.075159	NA	NA
ABCB8	0.049438	0.295774947	71.64425572
TIAM1	0.040184	-0.913225292	9.843123378
GPX6	0.038189	1.771210783	96.79169615
MGST1	0.035953	-0.229483771	36.86010852
MYH9	0.033806	-1.993966113	1.745694739
RNMTL1	0.033553	0.20030965	67.3566879
RHOBTB3	0.032285	-0.201976964	38.98325077
CTNS	0.030495	-0.176292081	40.56381222
MED19	0.030222	-0.009020804	52.8544468



Rotenone, agnostic, AUC = 0.83

feature	importance	RVIS	RVIS_perc
FANCE	0.050225	-0.312207497	32.05944798
ABCB1	0.044598	0.099178678	60.75725407
TBCA	0.044326	-0.009020804	52.8544468
CUEDC1|RP11-343K8.3	0.04111	NA	NA
SLC14A2	0.038346	0.279193852	70.77730597
SEPT12	0.035527	-0.378346116	28.01368247
AC110781.3	0.03528	NA	NA
ABCA12	0.034959	-0.299763455	32.26586459
AC008734.1|MUC16	0.032689	NA	NA

RP11-242G20.1	0.031651	NA	NA



Milestone #3: identification of combinations of variants likely to increase susceptibility to specific toxicants

Accomplishments: As summarized above, we have made substantial progress toward achieving all elements of this Milestone.

C. What opportunities for training and professional development has the project provided?

Nothing to Report.

D. How were the results disseminated to communities of interest?

An abstract was presented at the Society for Neuroscience meeting in November 2022. An abstract has been accepted for presentation at the International Congress of Parkinson's Disease and Movement Disorders in August 2023.

E. What do you plan to do during the next reporting period to accomplish the goals?

During the final year of the project, we plan to complete work on the subtasks below.

Subtask 6.2 rare-burden analysis of variants for 335 FAME subjects (months 18-30) [Goldman]. We will continue to explore rare and common variation, including variation in non-coding regions likely to be eQTL or sQTL sites, or to otherwise influence gene expression.

Subtask 6.3: conduct exposed-only analyses of gene*environment interaction for 335 FAME subjects (months 20-30) [Goldman]. We will expand exposed-only analyses to identify meaningful gene*environment interactions that are likely to be causal by incorporating rare and common variation in eQTL and sQTLs, and by using WGS data to explore strong GWAS associations.

Subtask 6.4: use Machine Learning approaches for 335 FAME subjects to identify genes that may contribute to neurodegeneration in Parkinson's disease (months 24-30). We will continue work on this Subtask.

Milestone #4: Manuscript on use of the identification and validation of specific genetic susceptibility to specific toxicants. We will prepare manuscript(s), and will present additional findings at scientific meetings.

IV. IMPACT

A. What was the impact on the development of the principal discipline(s) of the project?

Nothing to Report

B. What was the impact on other disciplines?

Nothing to Report

C. What was the impact on technology transfer?

Nothing to Report

D. What was the impact on society beyond science and technology?

Nothing to Report

V. CHANGES/PROBLEMS

A. Changes in approach and reasons for change

Nothing to Report

B. Actual or anticipated problems or delays and actions or plans to resolve them

None

C. Changes that had a significant impact on expenditures

Nothing to Report

D. Significant changes in use or care of human subjects, vertebrate animals, biohazards, and/or select agents

Nothing to Report

E. Significant changes in use or care of human subjects

Nothing to Report

F. Significant changes in use or care of vertebrate animals

Nothing to Report

G. Significant changes in use of biohazards and/or select agents

Nothing to Report

VI. PRODUCTS

A. Publications, conference papers, and presentations

a. **Journal publications**

Nothing to Report

b. **Books or other non-periodical, one-time publications.**

Nothing to Report

c. **Other publications, conference papers, and presentations.**

Presentation at the 2023 International Congress on Parkinson's Disease and Movement Disorders: "Testing gene-environment interaction in Parkinson's disease (PD) using whole-genome sequencing of pesticide-exposed cohorts", Authors: Goldman, Kaye, Traglia, Swanson, Lima, Finkbeiner, Tanner

B. Website(s) or other Internet site(s)

Nothing to Report

C. Technologies or techniques

Nothing to Report

D. Inventions, patent applications, and/or licenses

Nothing to Report

E. Other Products

Nothing to Report

VII. PARTICIPANTS & OTHER COLLABORATING ORGANIZATIONS

A. What individuals have worked on the project?

Name	Role	Researcher Identifier	Nearest person-months worked	Contribution to Project	Funding support
Samuel Goldman, MD, MPH	PI	0000-0002-3039-9927	4	Project Oversight; FAME study exposure and genotyping lead	W81XWH-20-1-0709

B. Has there been a change in the active other support of the PD/PI(s) or senior/key personnel since the last reporting period?

Samuel Goldman:

Previously active projects that have closed:

Veterans Administration CSR&D
I01 CX002040-01 (PI: Goldman)
10/1/2019-9/30/2022
(VA 3/8 salary support)
Parkinson's Disease and Exposure to Chlorinated Solvents at Marine Base Camp Lejeune

Health Resources and Services Administration (HRSA)
D33HP31668 (PI: Blanc)
05/01/2018-04/30/2023
(0.6 calendar months)
Preventive Medicine Residencies

Newly active projects:

None.

C. What other organizations were involved as partners?

Gladstone Institute was a partner through partnering grant W81XWH-20-1-0710 (PI: Swanson)

VIII. SPECIAL REPORTING REQUIREMENTS

COLLABORATIVE AWARDS

This is a collaborative award: W81XWH-20-1-0709 (PI: Goldman), W81XWH-20-1-0710 (PI: Swanson). Award -0170 is completed, with final report in process. Award -0709, the subject of this progress report, has a no-cost extension through July 31, 2024.

QUAD CHARTS: The Quad Chart is submitted as a separate attachment.

IX. APPENDICES

None