AWARD NUMBER: W81XWH-20-1-0849

TITLE: Plasma Cell-Free RNA as Noninvasive Biomarker for Parkinson's Disease

PRINCIPAL INVESTIGATOR: Laura Ibanez

CONTRACTING ORGANIZATION: Washington University, Saint Louis, MO

REPORT DATE: October 2022

TYPE OF REPORT: Annual

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

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. **REPORT DATE**	**2. REPORT TYPE**	**3. DATES COVERED**
October 2022	Annual	15Sep2021-14Sep2022

4. TITLE AND SUBTITLE

Plasma Cell-Free RNA as Noninvasive Biomarker for Parkinson's Disease

5a. CONTRACT NUMBER
W81XWH-20-1-0849

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Laura Ibanez

E-Mail:ibanezl@wustl.edu

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Washington University School of Medicine
660 S. Euclid Ave. St. Louis
63110, Missouri - US

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

Parkinson disease (PD) is the most common neurodegenerative disorder, after Alzheimer disease (AD). Many attempts have been made to find a good biomarker, including alpha-synuclein protein levels in the cerebrospinal fluid (CSF). Cell-free nucleic acids-based diagnostic tests have revolutionized prenatal screening. They have also been investigated in cancer and fetal development among other traits, including neurodegenerative diseases. We have successfully developed a preliminary predictive model for AD using cell-free plasma RNA sequencing (cfRNASeq) and machine learning techniques. We used an exploratory dataset (10 AD cases and 10 controls) to train a predictive model. We obtained an area under the ROC (AUC) of 0.84 in an independent replication dataset (10 independent AD cases and 10 controls). Moreover, this model provided similar accuracy (AUC=0.86) when tested in four preclinical AD. Using state-of-art deep neural network approaches, the accuracy increased up to 0.94. Overall, these results indicate that we can identify individuals that will progress to dementia. We think this technique can be applied to PD to generate disease-specific predictive model.

We ***hypothesize*** that there are detectable changes in the plasma free nucleic acid composition due to PD pathogenesis, even in early stages. We will use bioinformatics tools to construct a predictive model for PD, leveraging longitudinal plasma data that will allow the modeling of plasma cfRNA composition changes over the curse of the disease, thus maximizing the power of selecting informative transcripts to construct the predictive model.

We will firstly **accurately predict preclinical PD using cell-free nucleic species** in three steps: *A. Create prediction models for PD using cfRNA* generating cfRNASeq data from 200 plasma samples (50 PD individuals at 3 time-points – early pre-clinical (5-10 years before symptoms), pre-clinical (2-5 years before symptoms) and symptomatic (5-8 years after diagnostic) and 50 controls). We will use multiple analytical approaches including digital deconvolution and machine learning, feature selection and deep neural networks (similar to what we have used to generate our model for AD) to build a robust predictive model that includes the optimal number of transcripts. *B. Replication:* We will quantify the transcripts selected in A to be part of the predictive model using a more scalable and cost-effective technology such as Nanostring, Sequenom or custom transcript array to replicate the predictive model in an independent dataset (50 preclinical PD cases and 50 controls). We will also include subjects from African American and Latin ancestries (20 cases and 20 controls from each) and carriers of PD-causing mutations in *PARK1* and *PARK2* (n=20) to test the performance of the model in non-European ethnicities and in mutation carriers. *C. Specificity:* We will quantify the transcripts selected in A and replicated in B using the same scalable and cost-effective technology in 80 cases of other neurodegenerative diseases (AD, Lewy body dementia, progressive supranuclear palsy, amyotrophic lateral sclerosis and frontotemporal dementia) and additional 40 controls to test whether the predictive model is specific for PD or neurodegeneration. We expect the predictive model to be specific for PD; however, some overlap is expected due the commonalities of neurodegenerative diseases. We are currently generating longitudinal cfRNASeq data on AD individuals. With the data generated in the first aim, we will perform **integrative analyses of AD and PD** to describe biological differences and commonalities across the two most common neurodegenerative diseases. This will allow the description of biological mechanisms such as differences in genes or pathways across diseases, differences in the timeline of the disease for common genes/pathways and the improvement of the differential diagnosis.

If successful, this method could improve the cost-effectiveness of the currently available tools to diagnose and monitor PD, and provide a scalable blood-based early diagnostic screening tool.

Dr. Ibanez research interest is focused on using genetics to improve the management of individuals that suffer from neurodegenerative diseases, specially the early management by using high-throughput technologies and bioinformatics. Currently there is a great potential on using multi-omic approaches that integrate all levels of biological information. By using powerful bioinformatic tools we can combine genetic variance data with the RNA translation and the final protein production. This can be linked to diseases and be used to create predictive models that can easily be used in clinical settings to improve the management and quality of life of patients. In the last year and a half, Dr. Ibanez efforts have been focused on creating predictive models for Alzheimer Disease using different approaches. Even this model is on preliminary stages, she has proven that she can successfully create predictive models using cell-free RNA. This proposal will allow her to leverage the biology from the data generated, which could add some biological understanding to the biology of these neurodegenerative diseases.

This proposal has the potential of leading to a biomarker for early diagnosis and prognosis of PD. Moreover, the transcripts included in the prediction model will probably have biological relevance which, together with the integrative analyses with AD, might lead to potential drug targets in future studies.

15. SUBJECT TERMS

None listed.

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
					USAMRDC
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**	Unclassified	10	**19b. TELEPHONE NUMBER** *(include area code)*
Unclassified	Unclassified	Unclassified			

TABLE OF CONTENTS

Introduction

Parkinson's disease (PD) is the most common neurodegenerative disorder, after Alzheimer disease (AD). Many attempts have been made to find a good biomarker, including alpha-synuclein protein levels in the cerebrospinal fluid (CSF). Cell-free nucleic acids-based diagnostic tests have revolutionized prenatal screening. They have also been investigated in cancer and fetal development among other traits, including neurodegenerative diseases. We will use bioinformatics tools to construct a predictive model for PD, leveraging longitudinal plasma data that will allow the modeling of plasma cfRNA composition changes over the curse of the disease, thus maximizing the power of selecting informative transcripts to construct the predictive model.

Keywords

Parkinson's Disease, Biomarkers, cell-free RNA, Machine Learning

Accomplishments

What are the major goals of the project?

This project has four major goals:

1. Create a Predictive Model for Parkinson's Disease
2. Replicate the Predictive Model in an independent Dataset using a cost-effective Platform
3. Test the Specificity & Sensitivity of the Predictive Model
4. Conduct Integrative Analyses

What was accomplished under these goals?

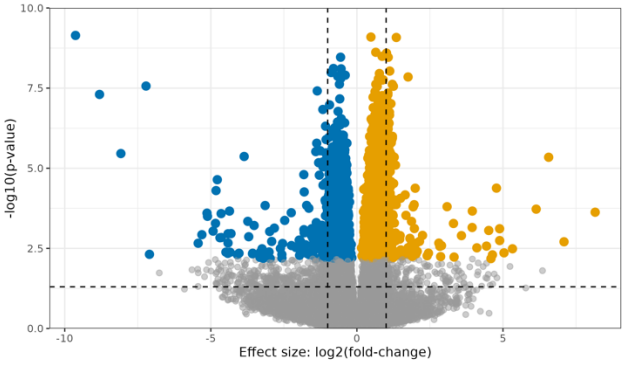
1) Major activities: We have successfully performed differential expression analyses on 189 plasma samples and generated preliminary predictive models. We have accessed 516 additional plasma samples to replicate the model. We have generated the libraries and we are currently finalizing the QC to sequence them.

2) Specific objectives: The goals for the second year of the award were finalize the differential expression analyses and generate a predictive model. We aimed to replicate the results in an independent dataset and finally, integrate all the data with that generated for AD.

3) Significant results or key outcomes:

<u>Differential Expression Analyses:</u> We have identified 1,558 down-regulated transcripts and 1,676 up-regulated that pass multiple test correction when comparing all PD cases to controls adjusting by biological (age at draw, sex and medication for PD (levodopa and dopamine agonist), medication for anxiety and depression, high blood pressure and dyslipidemia, all in the form of yes/no variables), and technical variables

Figure 1. Volcano plot showing the 3,782 transcripts differentially expressed in plasma of PD cases compared to Controls



(the first five surrogate variables for the transcriptomic data, calculated using the SVA package in R) (Figure 1).

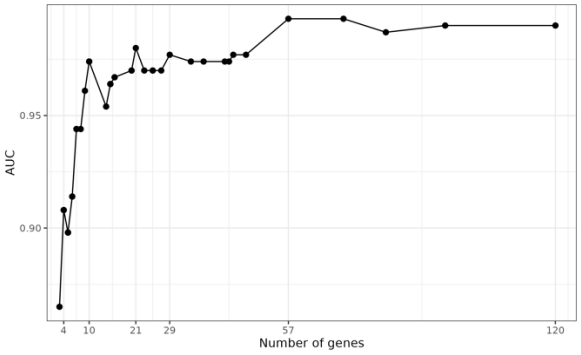
$$Status \sim Age + Sex + Batch_{Sequencing} + SVA_{1-5} + Medication$$

We then conducted pathway analyses with the top 20 differentially abundant transcripts (Table 1). Among the several significant pathways, we found Parkinson's disease ($p=2.30\times10^{-03}$), mitochondrion ($p=3.90\times10^{-04}$), and neurodegeneration ($p=2.90\times10^{-03}$). Interestingly, other neurodegenerative diseases were also identified such as amyotrophic lateral sclerosis, Huntington disease, and Alzheimer's disease. These results suggest that not all the changes that we are capturing are unique to PD but can be shared pathways with other neurodegenerative diseases.

Table 1. Top differentially abundant transcripts in plasma when comparing PD cases to Controls

Transcript	log2Fold Change	p-value	p adjusted
RIT1	0.938	7.13×10^{-11}	9.20×10^{-07}
SYTL4	1.349	8.31×10^{-10}	3.57×10^{-06}
PPP2R5A	0.480	8.06×10^{-10}	3.57×10^{-06}
SLC25A6	-0.559	3.44×10^{-09}	5.56×10^{-06}
DPYD	0.645	2.41×10^{-09}	5.56×10^{-06}
DPY19L1	1.072	3.45×10^{-09}	5.56×10^{-06}
CLIC4	0.862	3.22×10^{-09}	5.56×10^{-06}
MYCT1	1.014	2.56×10^{-09}	5.56×10^{-06}
PQBP1	-0.802	7.72×10^{-09}	1.02×10^{-05}
PTPN6	-0.540	7.91×10^{-09}	1.02×10^{-05}
AIG1	1.125	9.28×10^{-09}	1.09×10^{-05}
UBE2F	0.771	1.11×10^{-08}	1.10×10^{-05}
CXXC5	-0.881	1.03×10^{-08}	1.10×10^{-05}
NONO	-0.400	1.25×10^{-08}	1.14×10^{-05}
GABARAPL2	0.734	1.62×10^{-08}	1.14×10^{-05}
ERGIC3	-0.552	1.58×10^{-08}	1.14×10^{-05}
RPS27	-0.655	1.36×10^{-08}	1.14×10^{-05}
CALD1	0.899	1.73×10^{-08}	1.14×10^{-05}
TMSB4XP4	1.752	1.42×10^{-08}	1.14×10^{-05}
CD47	0.871	1.77×10^{-08}	1.14×10^{-05}

Figure 2. Performance of the predictive models in the testing dataset in increments of 10 genes per model.



Creation of the Predictive Model: After stringent quality control (described in the prior progress report), we have used the normalized counts for the 3,234 differentially expressed transcripts and Lasso regression to build a predictive model that includes the most informative transcripts. Briefly, we have used K=500 fold validation and randomly divided the sample 80% training and 20% testing for each fold. In each fold λ was optimized using cross-validation. In consequence, we obtained 500 different predictive models. Then we proceeded to select the best one. To do so, we calculated how many times each transcript was selected by lasso to be part of the final model and sort them. After randomly dividing the population in 80% testing and 20% training, we used ridge regression to calculate the performance of each group of genes in increments of ten to identify the optimal number of genes without sacrificing performance. We evaluated the area under the ROC curve (AUC) values for the testing in each model (Figure 2). We selected the final model based on the inflexion of the AUC evolution observed in the graph. The best models include 10 and 21 transcripts and have an area under the ROC curve of 0.97 (0.94-1.00) and 0.98 (0.95-1.00) respectively. We plan to replicate and potentially improve these predictive models in the replication dataset we are currently working on, and testing specificity.

Table 2. Summary demographics of the replication population

Group	N	Sex (Female %)	Age at draw (Mean age)	Age at Onset (Mean age)
Controls	277	56%	68.01	-
PD	239	39%	68.09	58.79

Independent Dataset Sample Selection: We have selected 239 PD cases and 277 controls from the Movement Disorders Clinic in Terrassa (Spain) via our collaborator Dr. Pastor (Table 2). We plan to use these samples to replicate the differential abundance and the predictive model results described above. The replication of the results, along with the investigation of PD Dementia and Lewy Body Dementia and generation of predictive models are part of the R00 obtained by the principal investigator using the data generated with this career development award, which has allowed for the inclusion of a larger sample size than that described in the original award.

Sample Processing: We have used the optimized protocols described in the previous progress report. We have extracted, ribodepleted, and generated the libraries for a total of 969 plasma samples. Of those 516 are PD cases and controls. We plan to submit the samples for sequencing in February, and target 40 millions of single end reads from each library. The Genome Technology Access Center (GTAC), the sequencing facility at Washington University will perform the final quality control for each library and generate the reads using a NovaSeq6000.

4) Other achievements. Include a discussion of stated goals not met.

Other Achievements: Via our collaboration, we have not only obtained plasma samples from PD individuals, but also a significant number of samples from individuals with Parkinson's Disease Dementia, and Dementia with Lewy Bodies to develop more specific predictive models. This is the main aim of the R00 obtained by Dr. Ibanez using the results from this career development award as preliminary data. Additionally, the analyses performed in AD are finalized and the paper was submitted for publication to Science Advances on December 27th. While we generate the replication dataset, we plan to prepare and submit a manuscript with the findings described in this report.

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

The present Early Investigator Research Award has facilitated the transition to independence of Dr. Ibanez. The data and analysis generated with the present award along with the generated in the K99 award was used a preliminary data for the successful transition to the R00 phase of her award. Similarly, the preliminary predictive models generated in the context of the project, were used as proof of concept for a project funded by the Michael J. Fox Foundation.

Dr. Ibanez has been invited as a speaker and participated in the following seminars and conferences (only reporting oral communication):

Alzheimer's Drug Discovery Foundation Investigators Meeting
28 November 2022 – San Francisco (US)
Plasma cell-free RNA signatures predict Alzheimer Disease

Invited lecture – Genetics and Genomics in Disease – DBBS Program (WUSTL)
17 October 2022 – Saint Louis (MO)
Genetic Influences on Early Neurological InStability after acute Ischemic Stroke

International Stroke Genetics Consortium – 27th Workshop
21-23 September 2022 - Bordeaux (France)
Genetic Influence on stroke severity measured by baseline NIHSS: Results from the GENISIS study

Knight-ADRC Tuesday Seminar Series – Washington University in Saint Louis
16 August 2022 – *Online*
Plasma cell-free RNA signatures predict Alzheimer Disease

Fundació ACE Seminar Series (Spain)
21 June 2022 – *Online*
Biomarker Tools for Alzheimer's Disease using High Throughput Data

iCMNar Seminar Series – VIB-UAntwerp Center for Molecular Neurology (Belgium)
31 May 2022 – *Online*
Biomarker Tools for Neurodegenerative Diseases using High Throughput Data

AD/PD Meeting 2021
15-20 March 2022 – Barcelona (Spain)
Circular RNA detection identifies circPSEN1 alterations in brain specific to autosomal dominant Alzheimer's Disease

NeuroGenomics and Transcriptomics Seminar Series – Washington University in Saint Louis
15 October 2021 – *Online*

Circular RNA detection Identifies circPSEN1 alterations in brain specific to Autosomal Dominant Alzheimer's Disease

Upcoming Invited Communications

AD/PD Meeting 2023

28 March-1 April 2023 – Gothenburg (Sweden)

Circular RNAs associated with idiopathic Parkinson's disease: case/control study results on the Parkinson's Progression Markers Initiative (PPMI)

How were the results disseminated to communities of interest?

The results described in this report will be reported in poster format in the upcoming AD/PD meeting (Gothenburg – March 28th to April 1st).

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

The main goals for the next reporting period are:

1. Finalize Aim 1: Finalize the data generation in the independent dataset to replicate the differential expression results and the predictive models.
2. Aim 2: We plan to investigate differences between AD and PD once all data is in hand via differential abundance. We will also apply unsupervised methods to understand the drivers of the differences (if not the disease).

Impact

What was the impact of the development of the principal discipline of the project?

Nothing to Report

What was the impact on other disciplines?

Nothing to Report

What was the impact on technology transfer?

Nothing to Report

What was the impact on society beyond science and technology?

Nothing to Report

Challenges / Problems

Changes in approach and reasons for change?

As stated in the previous progress report, reviewers and experts suggested to find an independent dataset to replicate our predictive models instead of finding an alternative technology for clinical translation. This has been the main focus of this reporting period. We have successfully secured a collaboration, obtained the samples, processed them, and we are working on the data generation.

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

Dr. Chen, postdoctoral research associate working on this project left the institution with very short notice in March 2022. We were not able to find a replacement until June 2022 when new postdoctoral

research associated joined Dr. Ibanez lab. Dr. Beric is not sourced to this project, but to its continuation via the R00, thus will be the one finalizing the analyses. Meanwhile, Dr. Ibanez has been the one processing the samples and performing the analyses needed.

Changes that had a significant impact on expenditures

Nothing to Report

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

Nothing to Report

Significant changes in use or care of human subjects

Nothing to Report

Significant changes in use or care of vertebrate animals.

Nothing to Report

Significant changes in use of biohazards and/or select agents

Nothing to Report

Products

Nothing to Report

Participants and other collaborating organizations

What individuals have worked on the project?

Name:	*Laura Ibanez*
Project Role:	*Principal Investigator*
Researcher Identifier (e.g. ORCID ID):	0000-0003-2381-7059
Nearest person month worked:	*3*
Contribution to Project:	*Dr. Ibanez has supervised all the steps for the correct development of the project. She performed half of the ribodepletion protocols, generated all the libraries and performed the quality control. She planned for the pooling and was in contact with the sequencing center during the sequencing project. She has performed the reported analyses and is training Dr. Beric.*
Funding Support:	*National Institute of Aging* *Alzheimer's Drug Discovery Foundation* *Michael J. Fox Foundation* *Bright Focus Foundation*

Name:	*Kristy Bergmann*
Project Role:	*Lab Technician*
Researcher Identifier (e.g. ORCID ID):	N/A
Nearest person month worked:	*3*
Contribution to Project:	*Ms. Bergmann has performed all the RNA extractions and quality control. She has also performed half of the the pooling and the adapter-primer cleaning. She is currently cleaning the pools prior to sequencing.*
Funding Support:	*National Institute of Aging* *Alzheimer's Drug Discovery Foundation*

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

Yes, the PI has transition to the independent phase of the NIH career development award, and secured two new grants:

- Successful transition from K99 to R00 phase of the Pathway Development Award from the NIH entitled "*Plasma cell-free RNA as non-invasive biomarker for Neurodegeneration*" in the total amount of for the period 02/01/2022 – 01/31/2025.

- Michael J. Fox Foundation Parkinson's Pathway Molecular Data Analysis Program entitled "*Resolution of Parkinson' associated loci and creation of predictive models using multi-omic data analyses"* in the amount of for the period 06/01/2022 – 11/30/2023.

- Internal Washington University in Saint Louis pilot award from the Knight-ADRC entitled "*Small RNA pathophysiology and multi-omic interactions in Alzheimer's Disease Brains*" in the amount of for the period 05/01/2022 – 04/30/2024.

What other organizations were involved as partners?

- **Organization Name:** Fundacio Docencia i Recerca MutuaTerrassa
  - **Location of Organization:** Spain
  - **Partner's contribution to the project:** In-kind support; the organization facilitated samples.

Special Reporting Requirements

N/A

Appendices