Project PPMI — A Report

Christian Bracher
Mijail Gomez

August 14, 2014

THE MICHAEL J. FOX FOUNDATION
FOR PARKINSON'S RESEARCH

BAYES IMPACT

AYASDI Cure
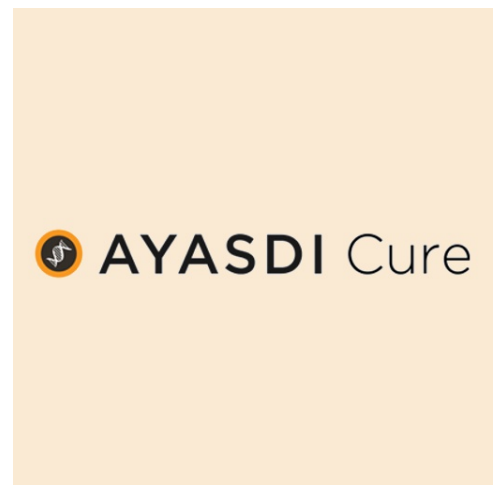
Acknowledgments

For their assistance with this project, we would like to thank:

- Eric Liu, Paul Duan, and Andrew Jiang (Bayes Impact)

- Ken Kubota and the Michael J. Fox Foundation

- The Parkinson's Progression Markers Initiative Team

- Devi Ramanan (Ayasdi)

- Vesela Gateva and Uri Laserson (mentors)

- Ryan Orban and Jonathan Dinu (Zipfian Academy)

Overview

- Parkinson's Disease — A Challenge

- Outline of the PPMI Study

- Evaluating Markers of Parkinson's Disease

- The Olfactory – REM Sleep – Motor Symptom Triad

- Leveraging Social Networks for Parkinson's Research

- A Machine Learning Web Application for PPMI

- Future Investigations – A Few Suggestions

Parkinson's Disease — A Challenge

PD presents multiple challenges:

- Progressive, fatal disease; no cure

- Medications only treat symptoms

- No definitive diagnosis without autopsy

- Causes of PD are mostly unknown

- Possibility of distinct PD subtypes

- Early detection of PD

Lewy Body
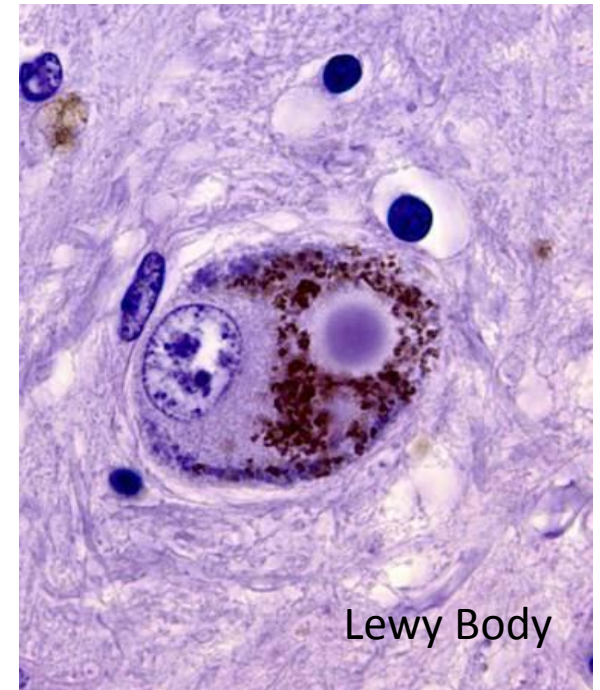
Parkinson's Progression Marker Initiative

PPMI is designed to yield answers to these challenges:

- $60 M four year longitudinal, clinical and observational study

- Massive horizontal data set:
  Biomarker, psychological, neurological, imaging, genetics data

- Several cohorts with altogether ~ 1,000 subjects

Use machine learning, advanced statistical techniques to evaluate data, find patterns

Evaluating Markers of Parkinson's Disease

Physiological changes that have been linked to PD:

- Brain function (dopamine response)

- Lipid metabolism

- Alteration in protein levels in cerebrospinal fluid (CSF)

- Genetic variants, gene expression

- Neurological abnormalities:
  Olfactory, sleep, spatiovisual capability, ...

- Psychological dysfunction (depression, ...)

SPECT Brain Imaging (DaTSCAN)

- "Gold Standard" in PD diagnosis

- Expensive ($5,000)

- Requires specialized equipment:

  – Radioactive marker

  – SPECT imaging

- Not widely available

HEALTHY   PARKINSON'S

SPECT Brain Imaging (DaTSCAN)

- But DaTSCAN is far from perfect:

  - Healthy subjects pass DaTSCAN, but

  - 15 % of PD subjects are misclassified

- DaTSCAN unveils a new subject cohort:

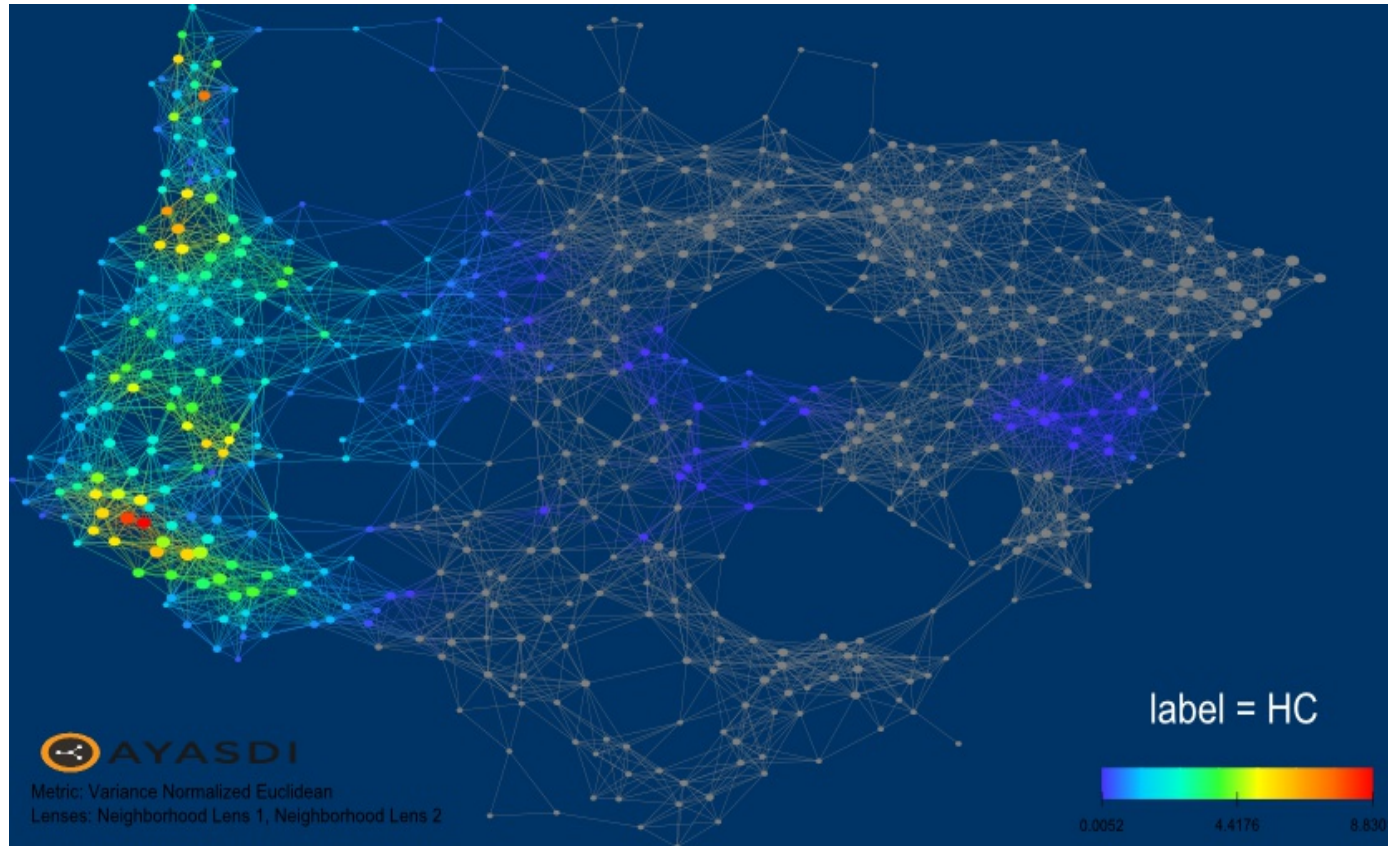 SWEDD — subjects without evidence of dopaminergic dysfunction

- Identification of SWEDD patients?

SWEDD

How Well Does DaTSCAN Perform?
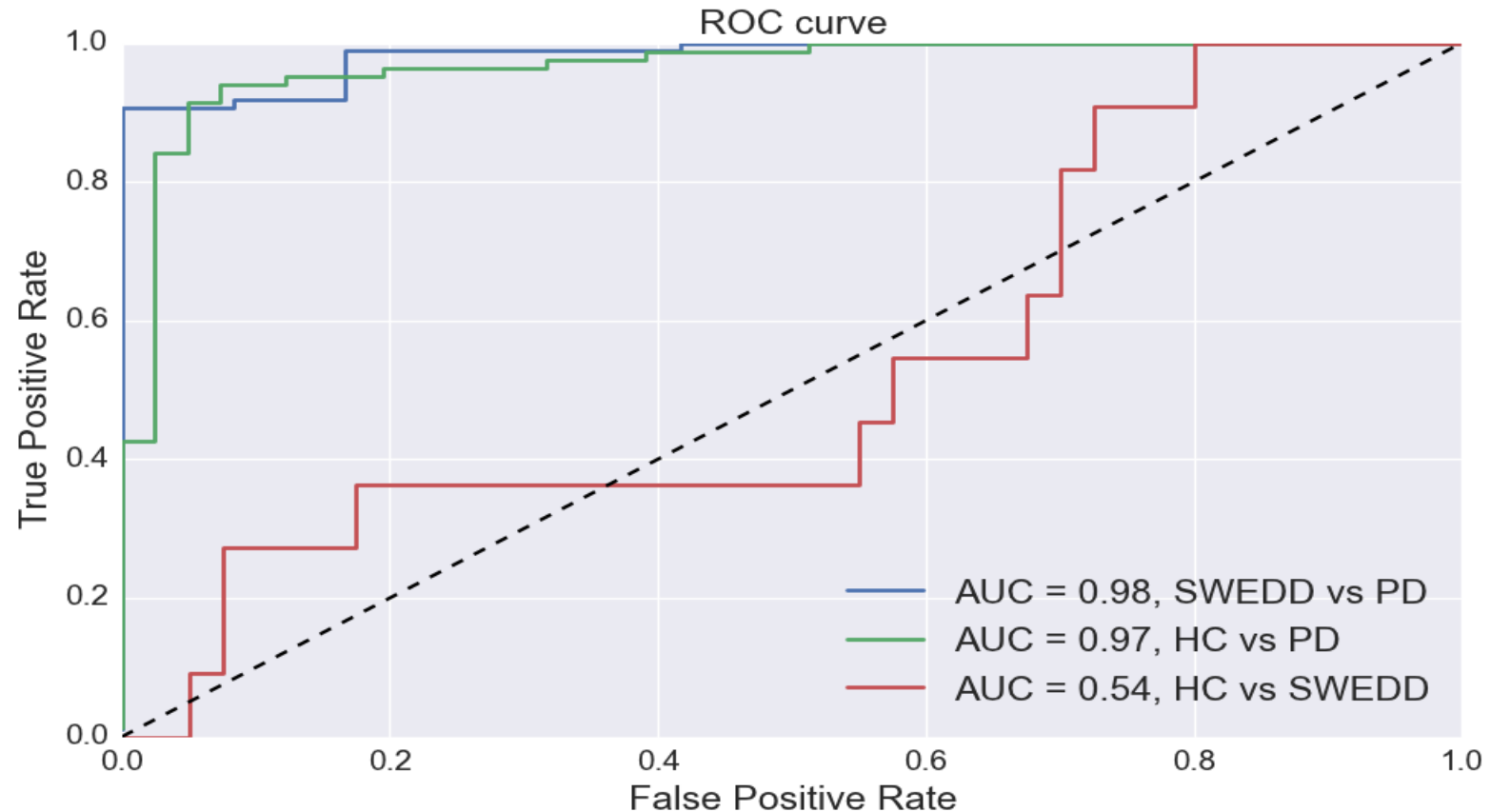
Supervised learning using DaTSCAN fails to separate Healthy Control (HC) and SWEDD subjects
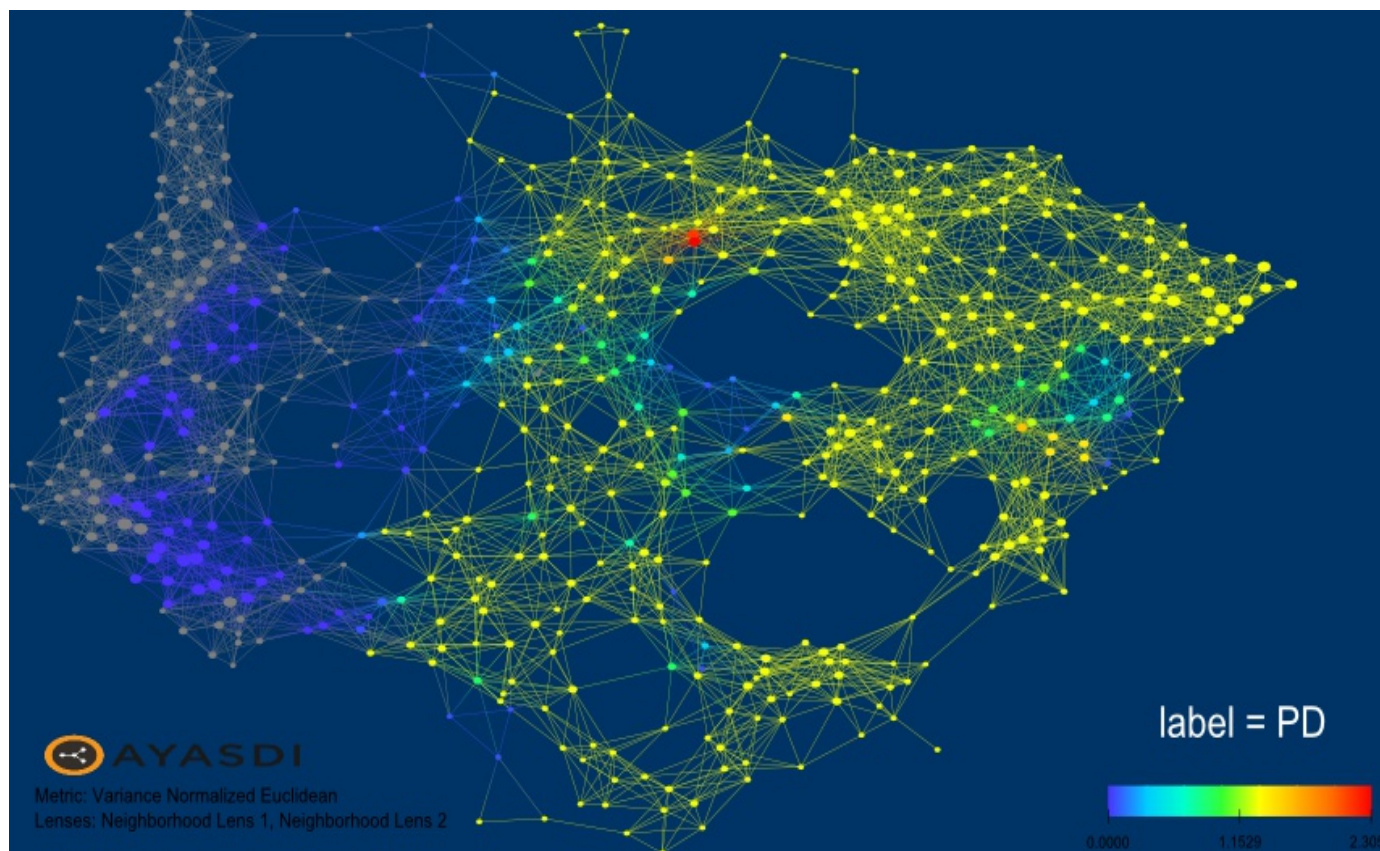
DaTSCAN Clustering Analysis (Ayasdi)

Graph-based Topological Cluster Analysis (proprietary algorithm) — healthy subjects

DaTSCAN Clustering Analysis (Ayasdi)
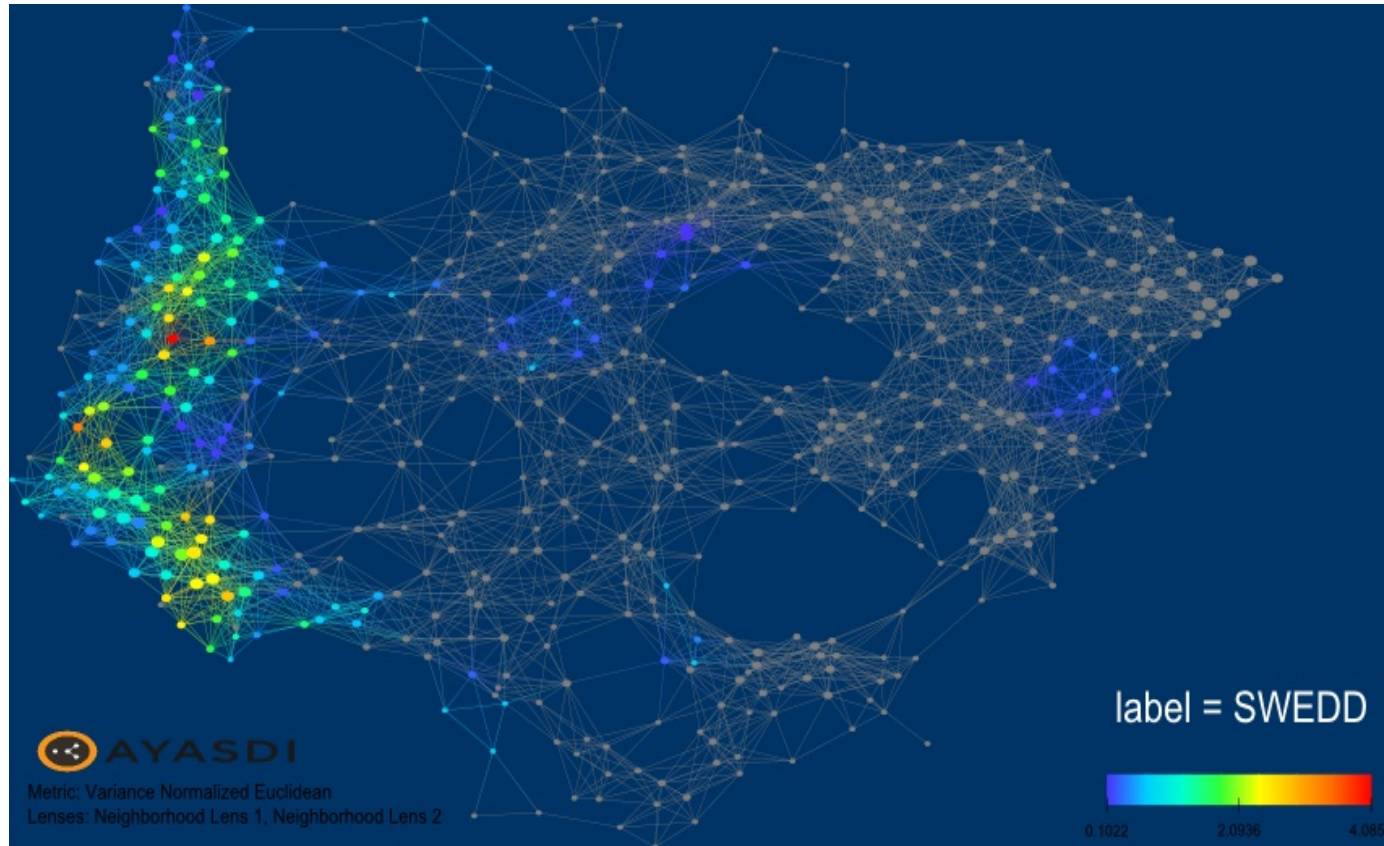
Graph-based Topological Cluster Analysis (proprietary algorithm) — SWEDD population
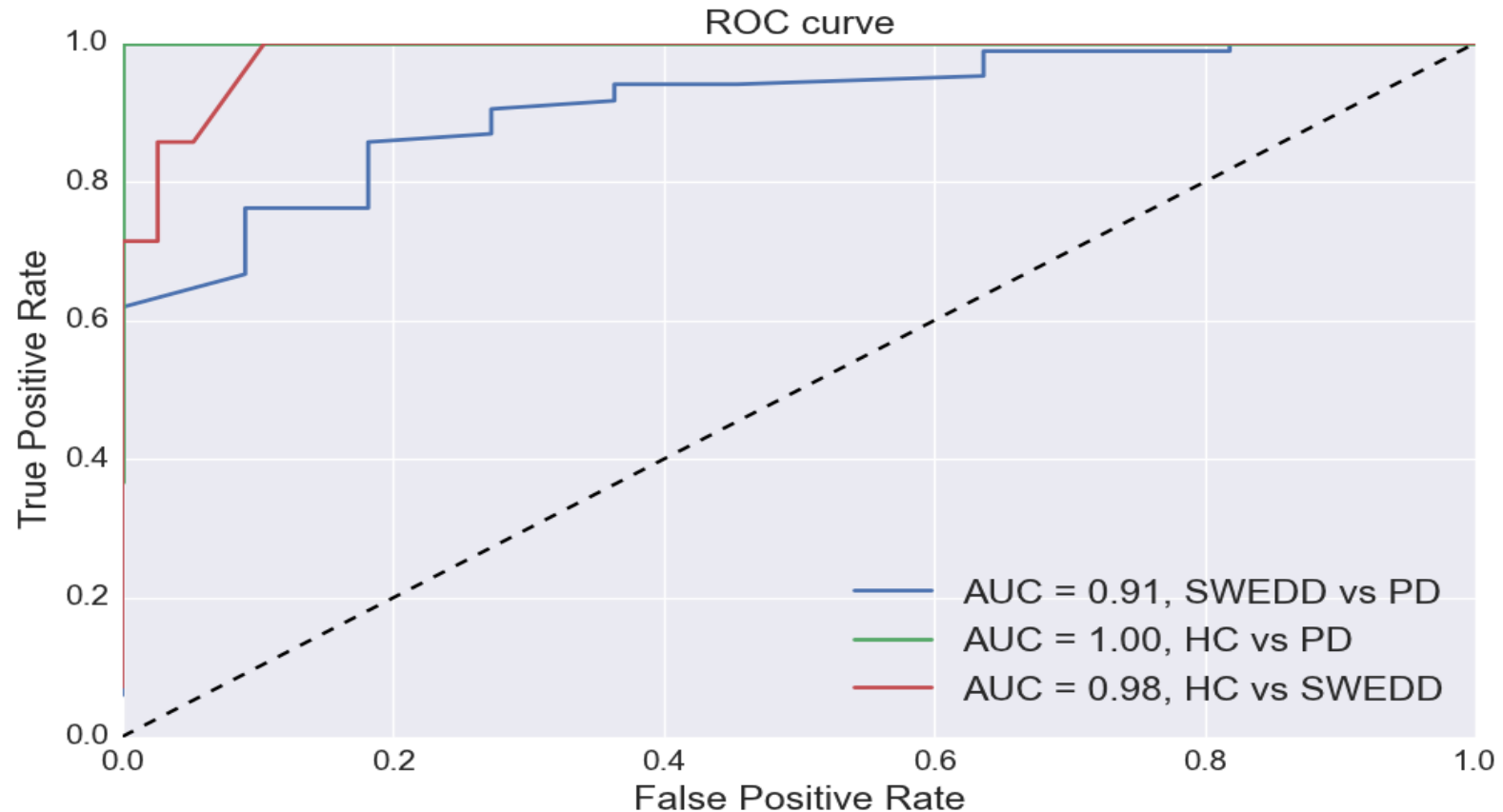
DaTSCAN Clustering Analysis (Ayasdi)

Graph-based Topological Cluster Analysis (proprietary algorithm) — PD patients

Identifying PD Using Clinical Assessments

Are there other markers of Parkinson's Disease?

- Earlier studies indicate that PD patients deviate from healthy controls in many clinical parameters

- Using test of significance, we found three areas that were good predictors for PD and SWEDD:

  - Olfactory (smelling) ability (UPSIT)

  - REM sleep behavior (acting out dreams)

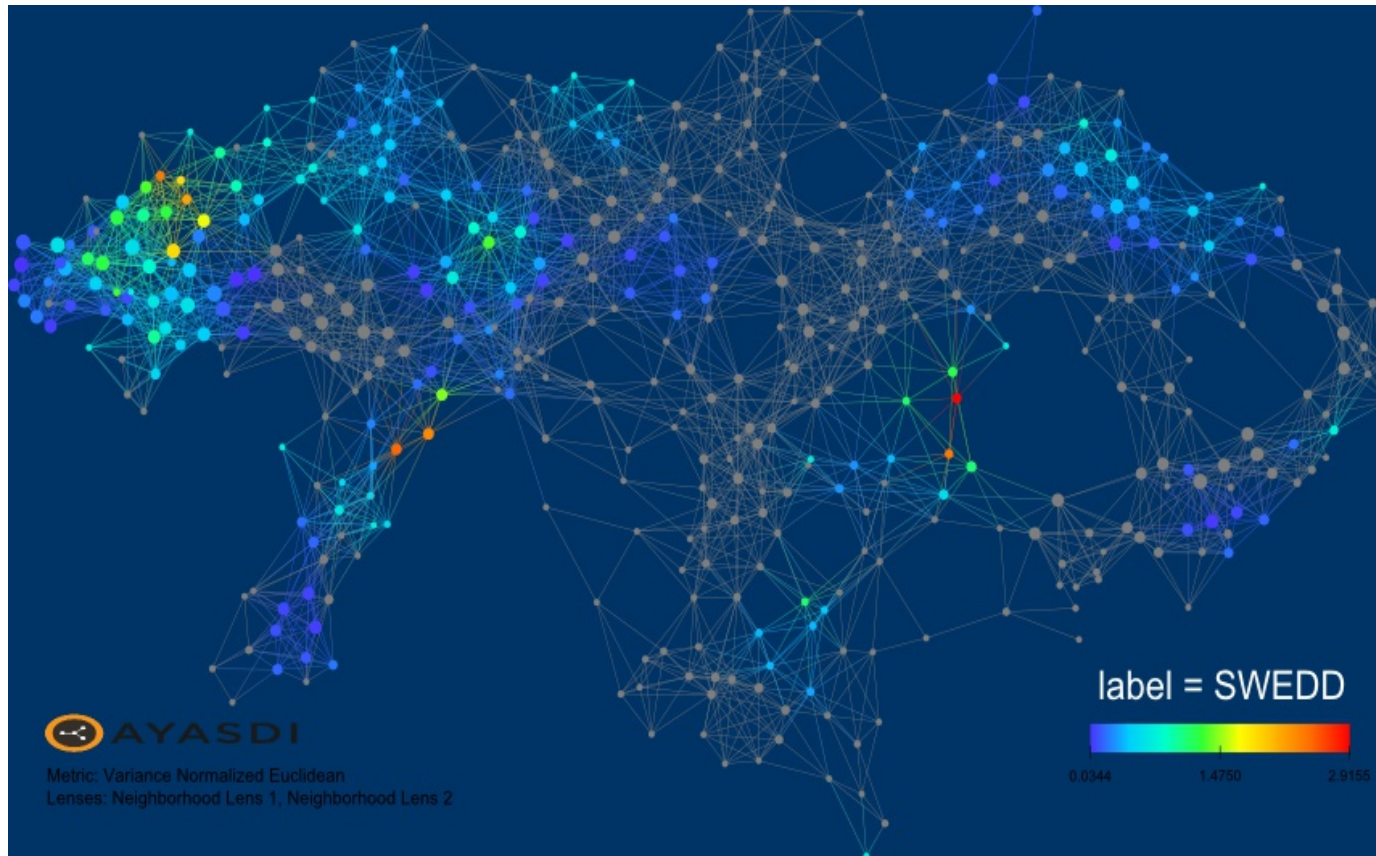 - Self-evaluation of motor difficulties (UPDRS Part II)

Classification Using Smell, REM, and Motor

Smell, REM, and motor assessments together outperform DaTSCAN!!!

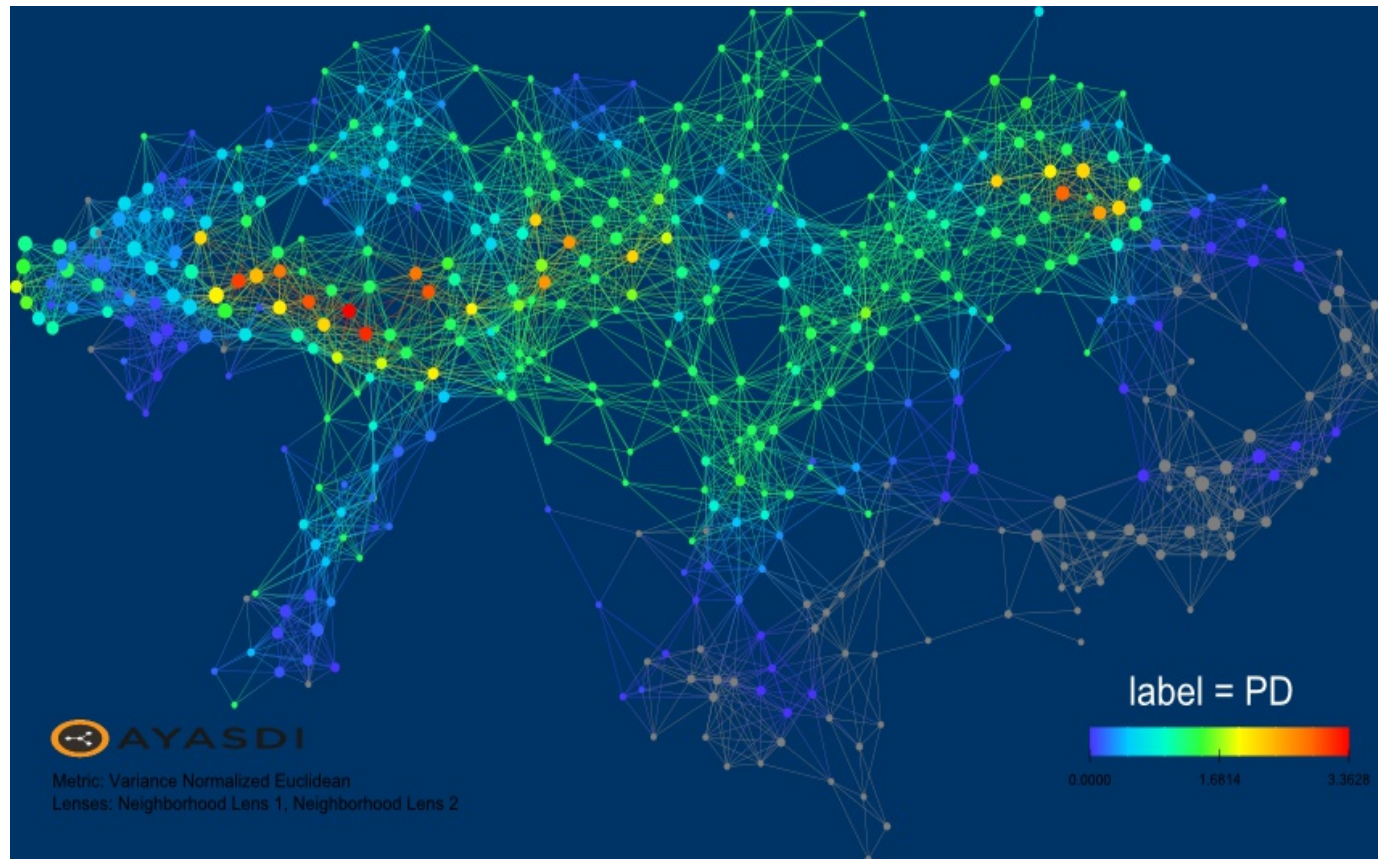# Clustering Analysis — Smell, REM, Motor
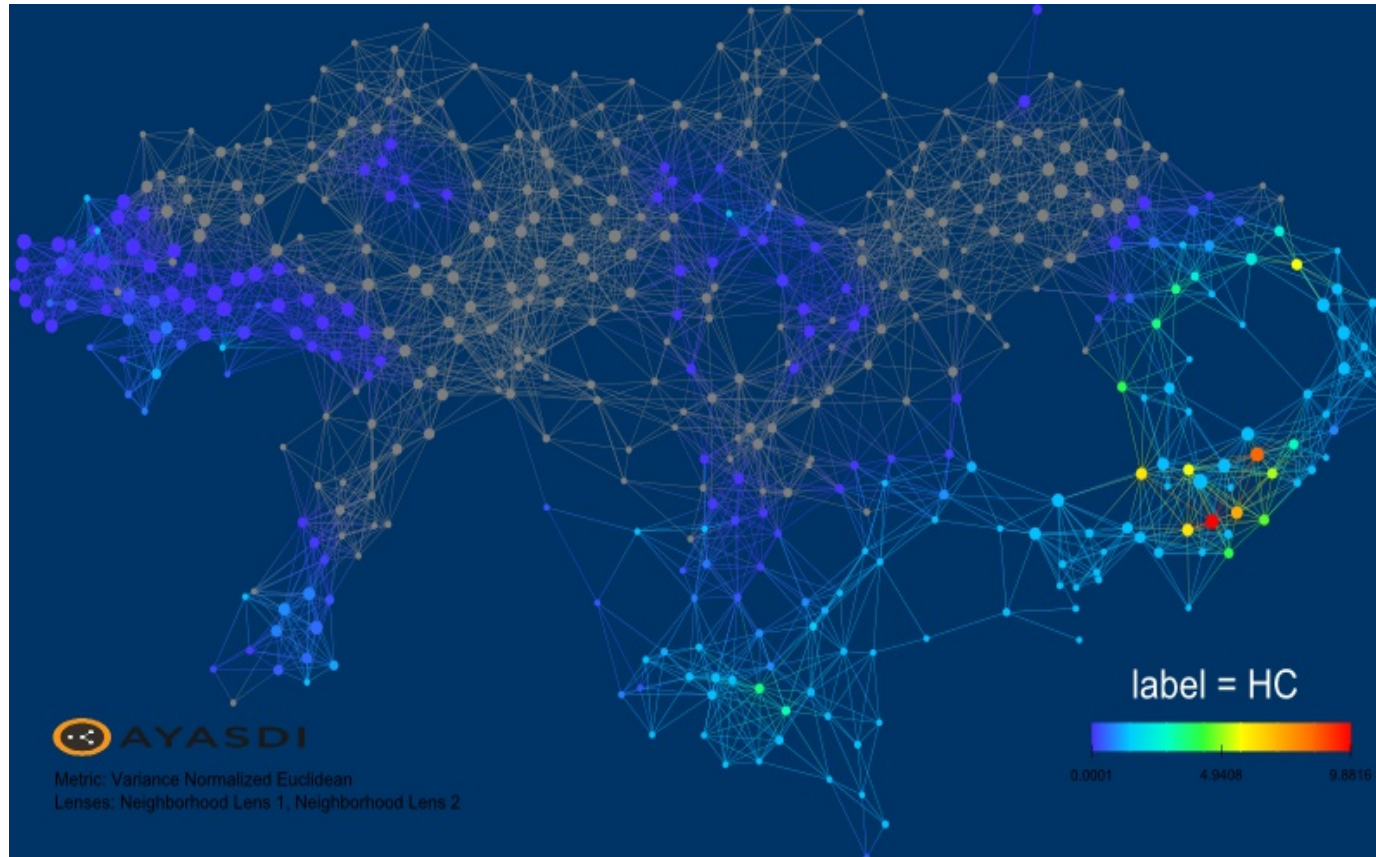
Graph-based Topological Cluster Analysis (proprietary algorithm) — healthy subjects
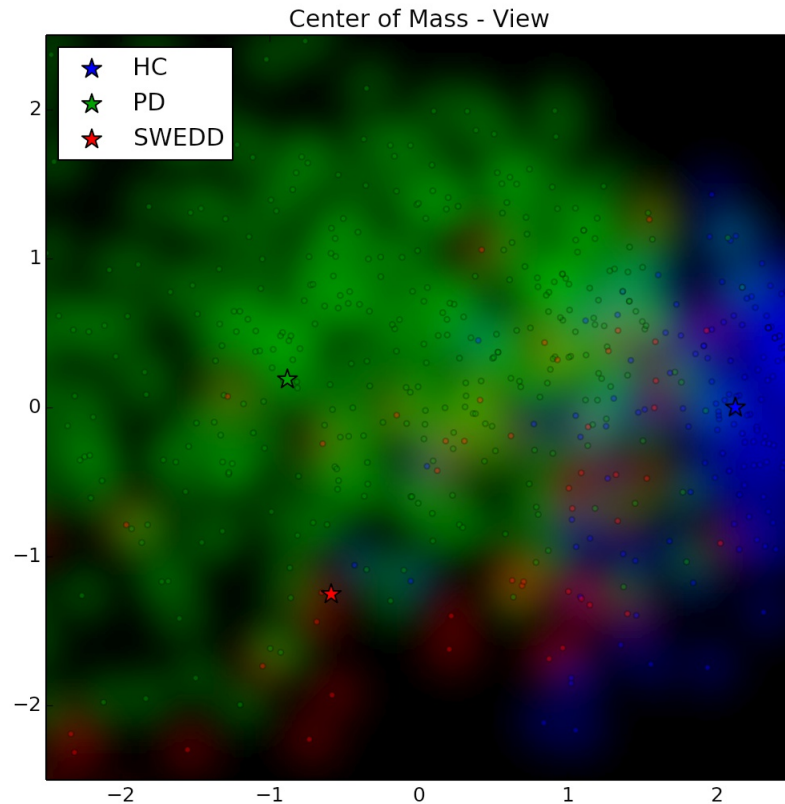
Clustering Analysis — Smell, REM, Motor

Graph-based Topological Cluster Analysis (proprietary algorithm) — SWEDD subjects

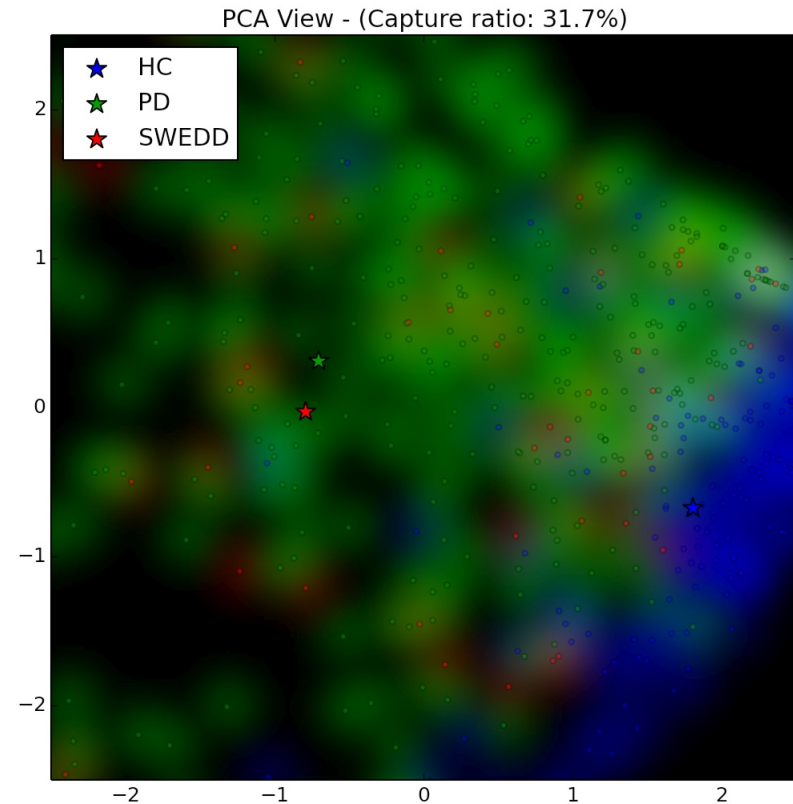# Clustering Analysis — Smell, REM, Motor

Graph-based Topological Cluster Analysis (proprietary algorithm) — PD patients

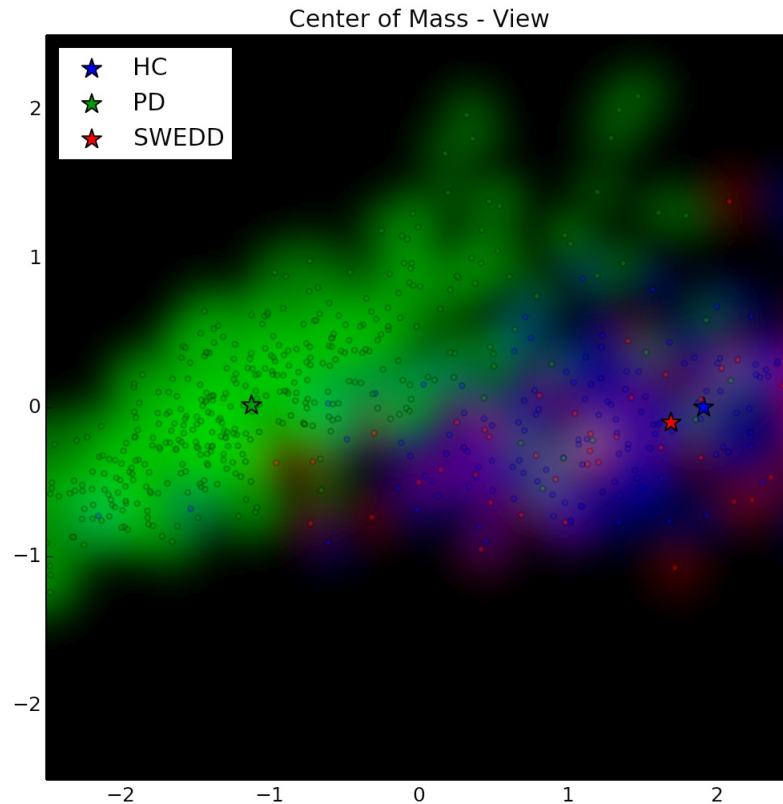# Smell – REM – Motor: Feature Reduction

Center of Mass - View

Canonical Discriminant Analysis

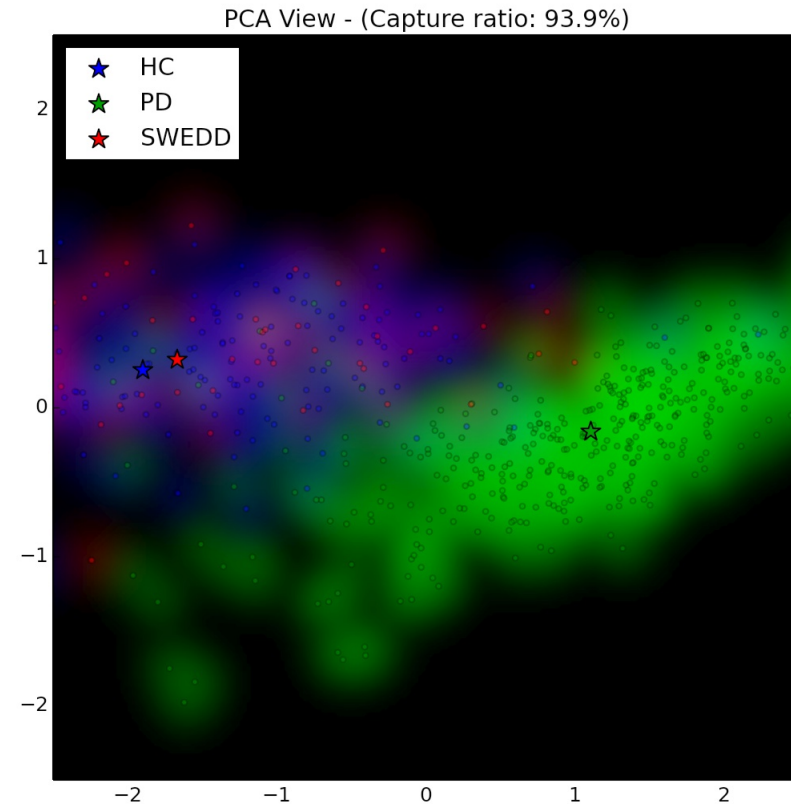PCA View - (Capture ratio: 31.7%)

Principal Component Analysis

Smell – REM – Motor classification relies on multidimensional patterns in feature space!

DaTSCAN: Feature Reduction

Canonical Discriminant Analysis

Principal Component Analysis

DaTSCAN classification depends on a single direction in feature space

Assessments That Lack Resolving Power

PPMI captures many other markers that have been deemed important in PD. How do they perform?

- Markers that appear altogether independent of disease status:

    - Example: Lipid metabolism

- Markers showing statistically significant differences between cohorts on average, but variation within groups is sufficiently large to lack discriminatory power:

    - Example: PD-related proteins in cerebrospinal fluid (CSF)

    - Example: RNA transcription rates of PD-related proteins

Blood Lipids vs. PD Status

PCA View - (Capture ratio: 78.3%)

Principal Component Analysis

ROC Curve (Random Forest)

AUC HC vs. PD: 0.55
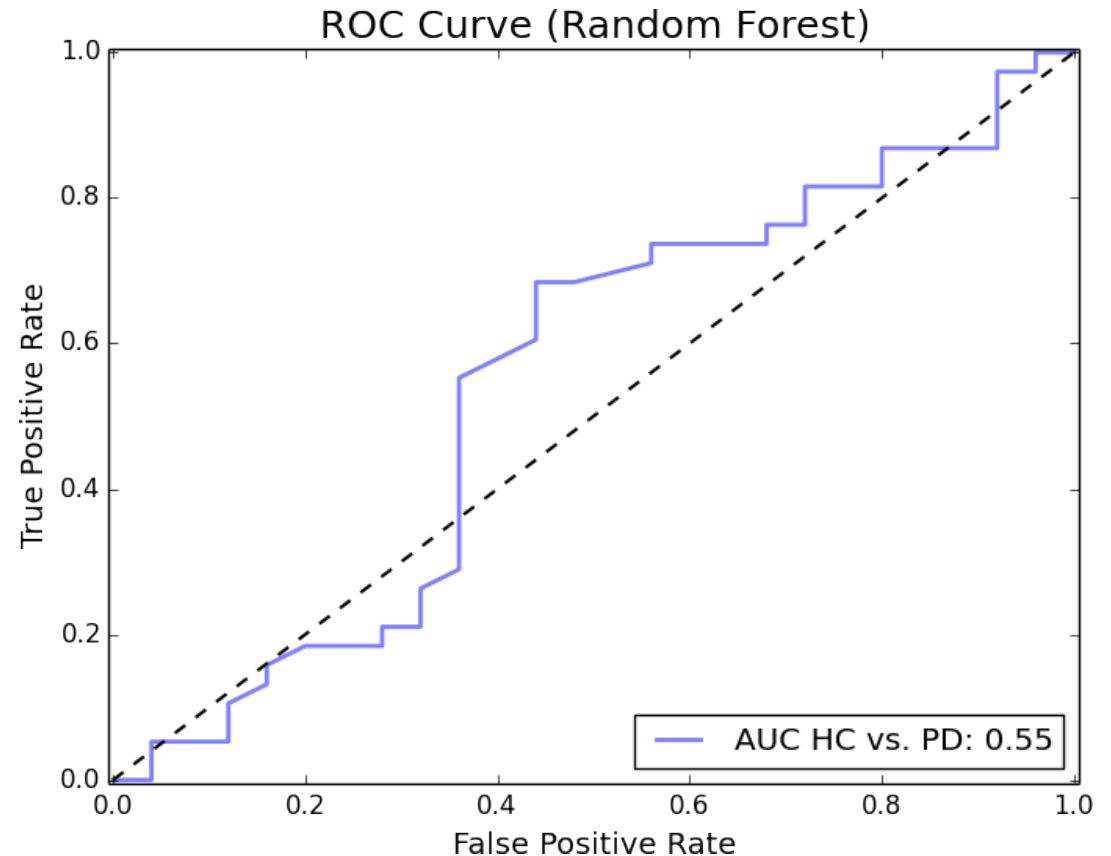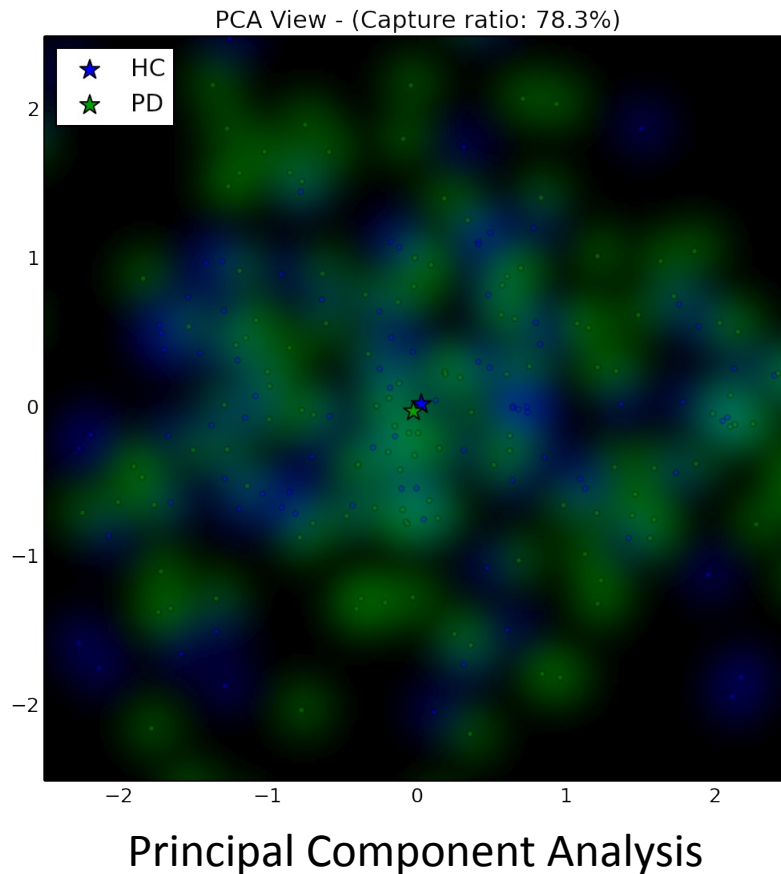
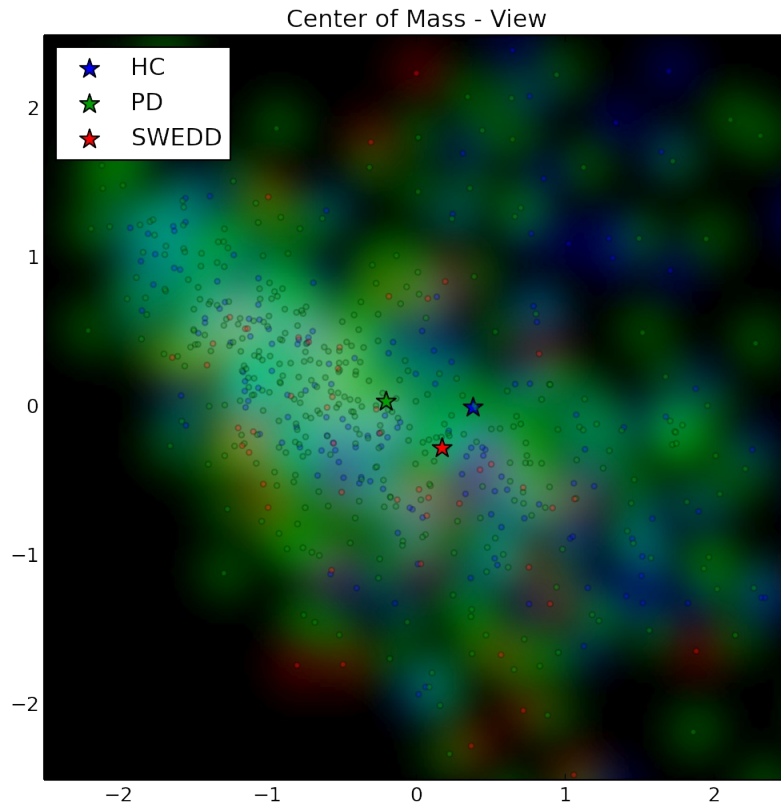No correlation between blood lipids and PD cohort found. (*Note*: SWEDD data missing.)

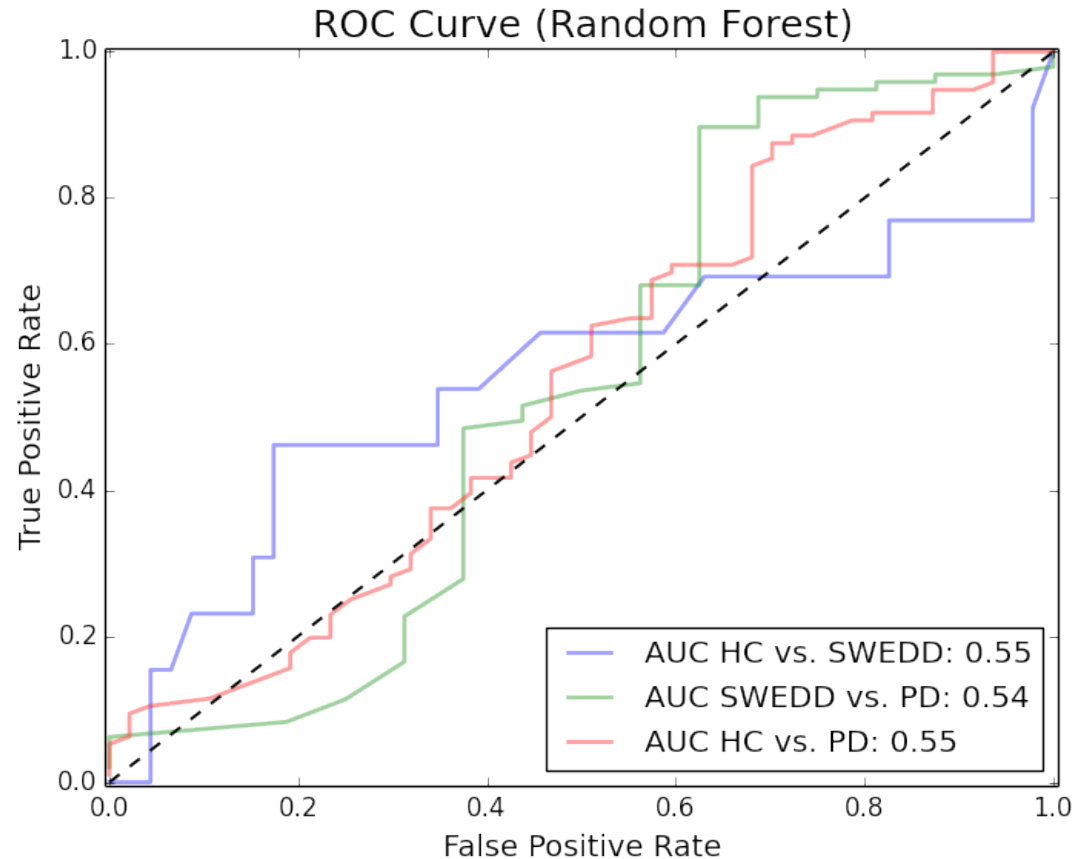# CSF Profiles: α-Synuclein and Amyloid-β$_{42}$
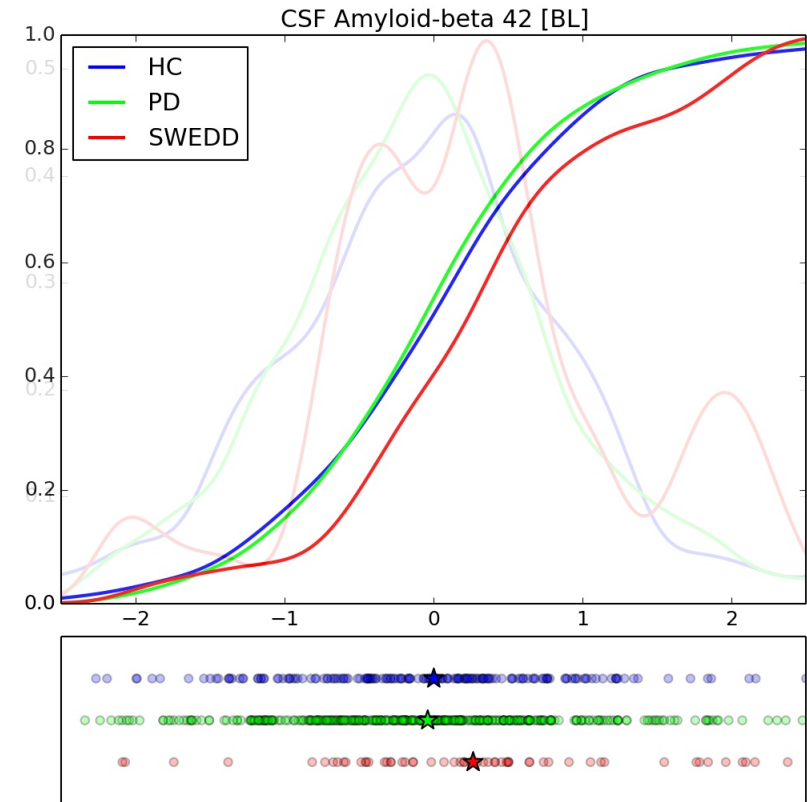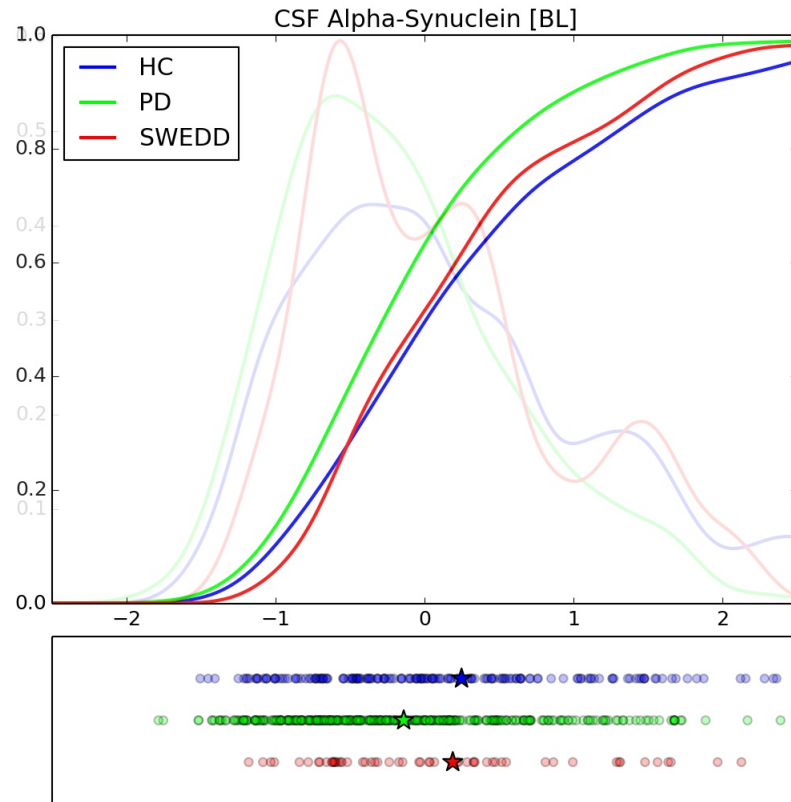
Cumulative probability (bold) and probability density (light) distributions

Significant decrease of α-syn in PD cohort, increase of Amyloid-β$_{42}$ in SWEDD *on average*
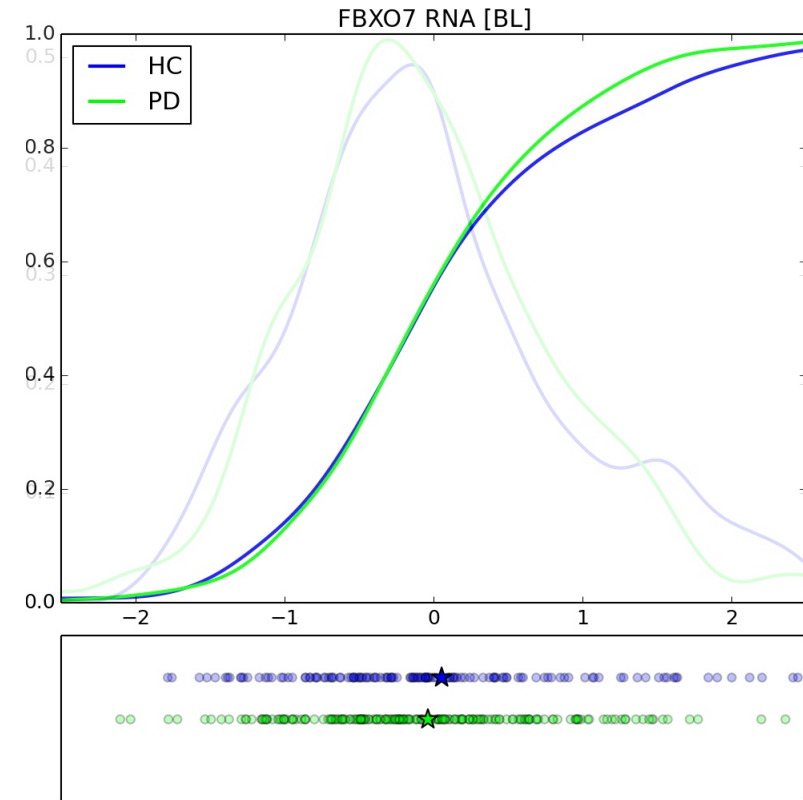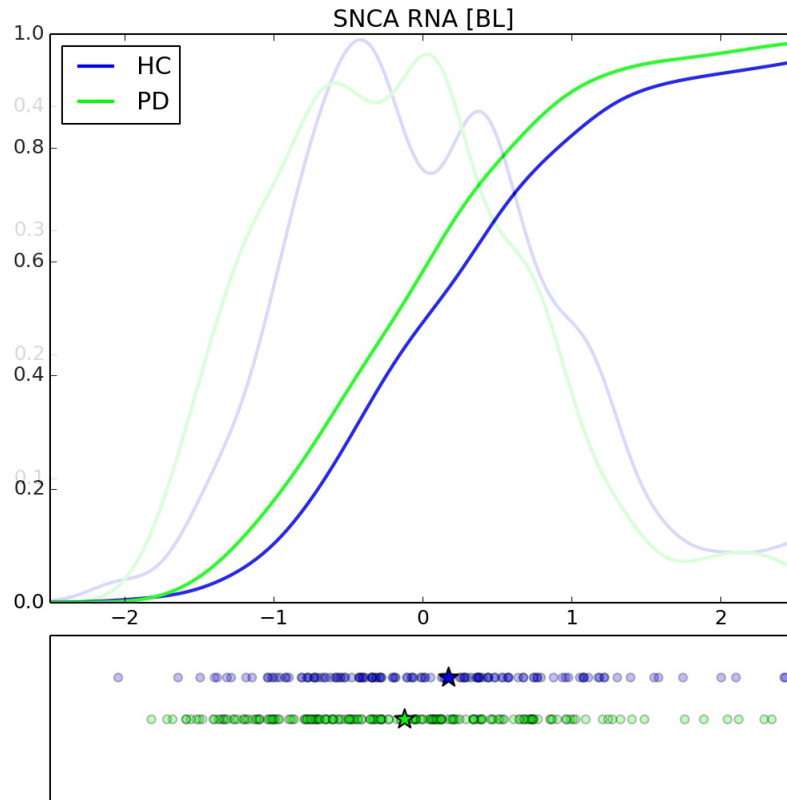
Analysis: CSF Proteins

Center of Mass - View

Canonical Discriminant Analysis

ROC Curve (Random Forest)

AUC HC vs. SWEDD: 0.55
AUC SWEDD vs. PD: 0.54
AUC HC vs. PD: 0.55

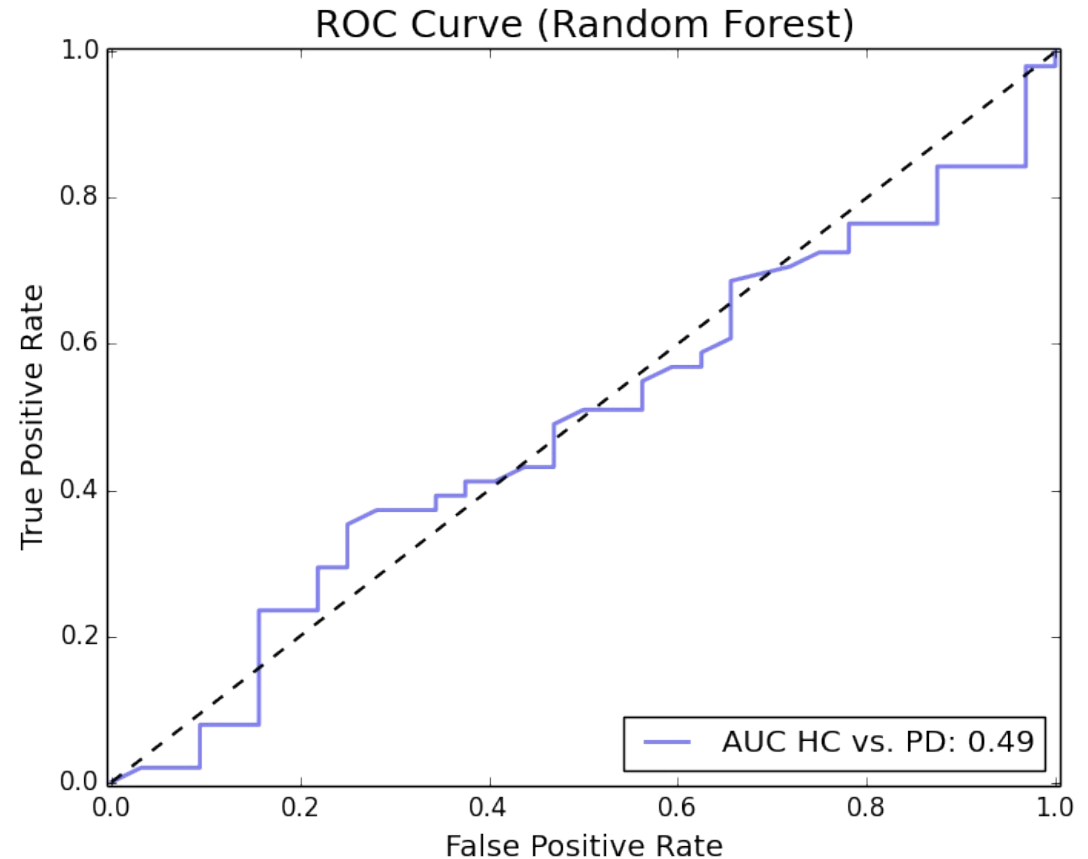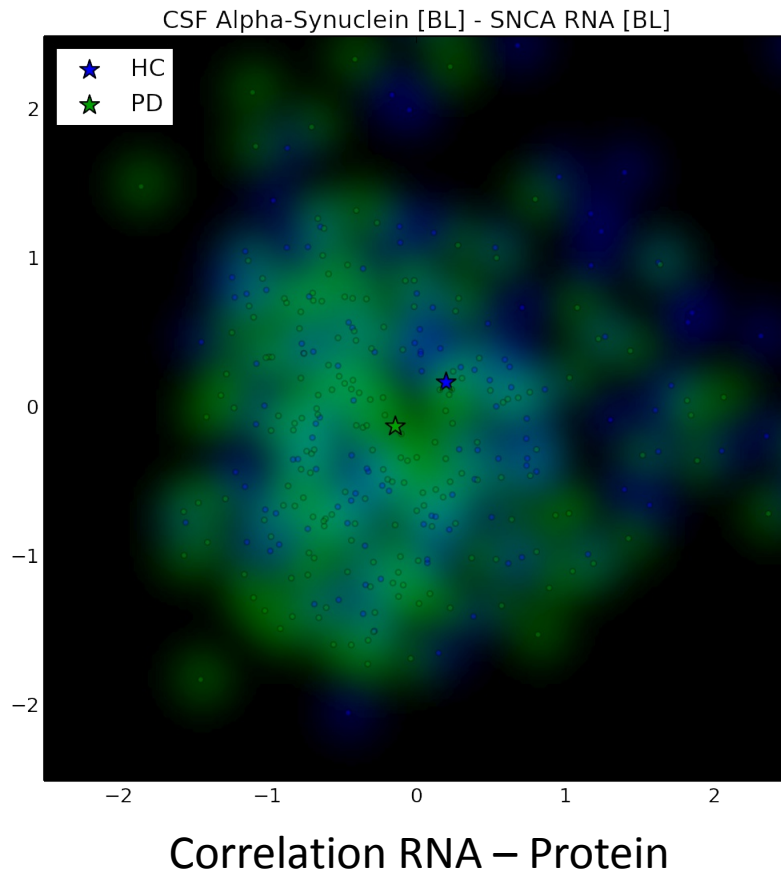Cohorts strongly overlap, weak predictive power of CSF protein data

RNA Counts SNCA (all types), FBXO7 (all types)

Cumulative probability (bold) and probability density (light) distributions

Transcription of *SNCA* RNA suppressed in PD cohort, but not of *FBXO7* (no SWEDD data)

Are SNCA RNA, α-Syn Protein Correlated?

CSF Alpha-Synuclein [BL] - SNCA RNA [BL]

★ HC
★ PD

Correlation RNA – Protein

ROC Curve (Random Forest)

True Positive Rate / False Positive Rate

AUC HC vs. PD: 0.49

Transcription of *SNCA* RNA and CSF concentration of α-Syn protein are suppressed in PD patients, but transcription and protein level display little correlation

# Leveraging the Web for Parkinson's Research

Smell-REM-Motor classifier contains self-reported data:

- Use social media to raise PD awareness

- Gather large-scale population data from online community

- Outreach via websites, apps, gamification?

Smell-REM-Motor testing requires no specialist knowledge:

- Define "best practices" in PD diagnosis and differential testing

- Provide tools for family practitioners to identify likely PD sufferers

Online Self-Assessment of Motor Skills
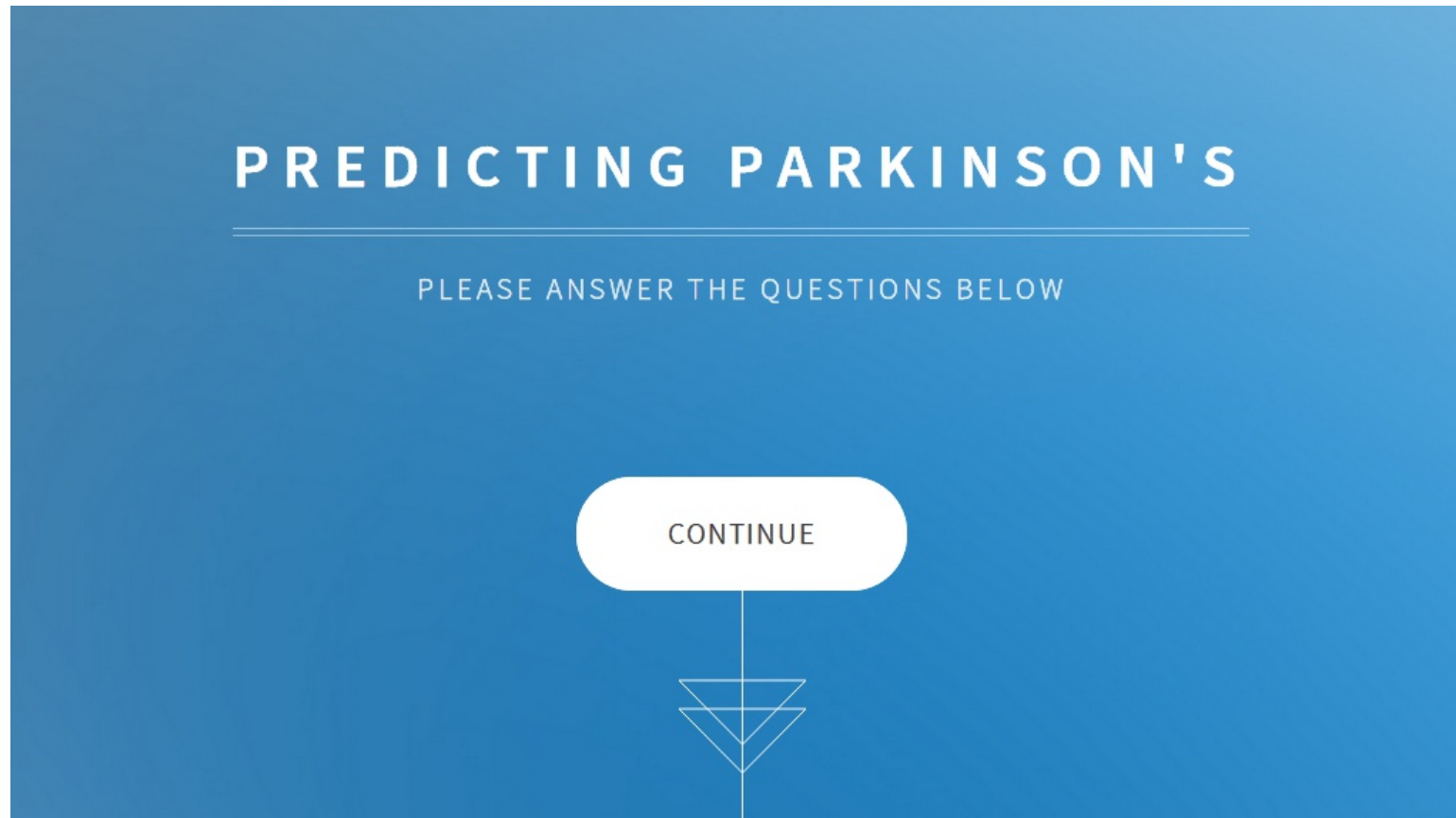
"Predicting Parkinson's" — Screenshot of landing page. Online at: www.mgomezm.com

# A Machine Learning Application for PPMI

Background:

- Machine learning is a promising tool for analyzing medical data, but acceptance is low:

  - Few biologists and physicians are familiar with advanced statistics

  - Few data scientists have a background in the life sciences

- Idea: Provide a utility that makes data mining techniques accessible to PPMI research community

A Machine Learning Application for PPMI

Specifications:

- Utility should be easy to operate
  - Intuitive approach: "Select and click"
  - Reduce numerical information, rely on graphical display instead

- PPMI is an international study
  - World-wide access: Employ website as user interface

- Medical data is privileged, sensitive information:
  - Keep PPMI data, machine learning core on central server
  - Transmit only aggregate data to address privacy concerns

Machine Learning Utility Framework
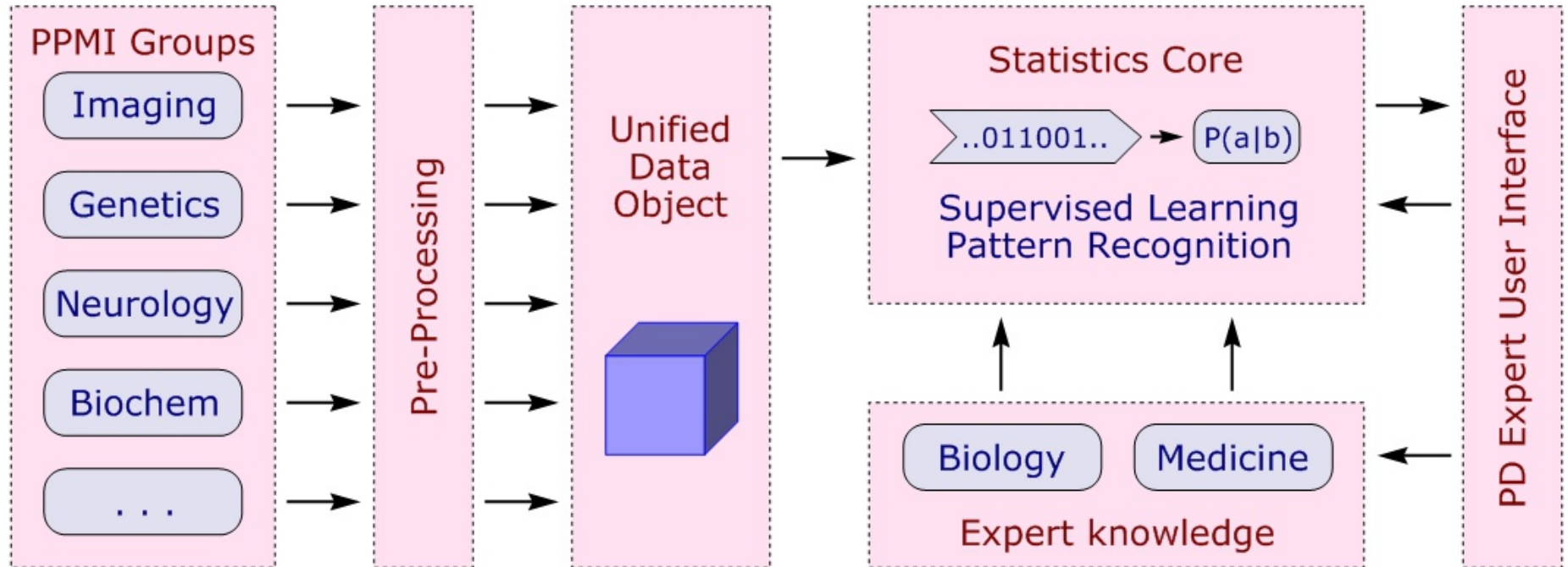
PPMI Data

ETL Stage:
Load & Assemble Data

Statistics & ML Core:
Data Selection & Graphics
Communication with UI

WWW

PD-LEARN User Interface

PD-LEARN
Statistical Analysis with PPMI Data

THE MICHAEL J. FOX FOUNDATION FOR PARKINSON'S RESEARCH

BAYES IMPACT

Step #1: Select Data Sets

These data sets are available for analysis. Mark the sets you'd like to include, then click the **Continue** button:

	SC	BL	V01	V02	V03	V04	V05	V06	V07	V08	V09	V10	V11	V12
UPSIT Olfactory Score	0	655	0	0	0	0	0	0	0	0	0	0	0	0
DaTSCAN Left Putamen	649	0	0	0	0	323	0	130	0	0	0	2	0	0
DaTSCAN Right Putamen	649	0	0	0	0	323	0	130	0	0	0	2	0	0
DaTSCAN Left Caudate	649	0	0	0	0	323	0	130	0	0	0	2	0	0
DaTSCAN Right Caudate	649	0	0	0	0	323	0	130	0	0	0	2	0	0
CSF Amyloid-beta 42	0	635	0	263	0	300	0	0	0	0	0	0	0	0
CSF Alpha-Synuclein	0	635	0	262	0	300	0	0	0	0	0	0	0	0
CSF Total Tau-Protein	0	629	0	262	0	299	0	0	0	0	0	0	0	0
CSF Phosphorylated Tau-Protein	0	633	0	263	0	299	0	0	0	0	0	0	0	0

Tip　　　Continue

Summary: Assessing Disease Markers

- PPMI is a large-scale, multi-feature study of Parkinson's Disease.

- We investigated variables for their value in diagnosing PD.

- Combining olfactory ability, REM sleep abnormalities, and perceived motor disability, we found a superior predictor for PD.

- Unlike DaTSCAN, our test reliably identifies SWEDD patients.

- We applied machine learning methods to PPMI data to test markers suggested in the PD literature. We found that the amount of disease-related proteins in the CSF, and translation of genes implied in Parkinson's disease are not good predictors for PD.

Summary: Using the Internet for PD

- The olfactory – REM sleep – motor method is simple, noninvasive, and affordable, and suitable for non-clinical settings.

- We built a WebApp, based on a self-assessment of motor difficulties, to identify people at risk of Parkinson's.

- Advanced statistical methods and machine learning are underutilized in medical research.

- We are developing a web-based utility that extracts PPMI data, and gives non-experts access to ML tools through a graphical interface.

Further Investigations

- Using machine learning on longitudinal data in the PPMI study: How do PD patients' symptoms progress over time?

- PPMI has a plethora of genetic information that should be studied.

- We did not have full access to the rich PPMI brain imaging data. Further analysis may improve on the classification of PD.

- Use Topological Analysis and unsupervised learning to identify clusters in the data linked to subtypes of PD (like SWEDD).

- Use Natural Language Processing (NLP) on patients diaries/journals and doctors' notes to obtain additional features.

- Use social media (Twitter, ...) for PD awareness and early detection.

Contact Information

- Christian Bracher

  cbracher69@gmail.com

  http://www.christianbracher.net/

  http://www.linkedin.com/in/christianbracher/

  https://github.com/cbracher69/PD-Learn/

- Mijail Gomez

  mijail.gomez.munoz@gmail.com

  https://www.linkedin.com/in/mijailgomez

  https://github.com/mgomezm/Predicting-Parkinsons