The BDDS Platform for Big Biomedical Data Management and Analysis: Discovering the role of Amyloid Deposition in Neurodegenerative Diseases

Ravi K Madduri, Michael D'Arcy, Kyle Chard, Alexis Rodriguez, Judy Pa, Naveen Ashish, Ben Heavner, Gustavo Glusman, Ivo Dinov, John Van Horn, Eric Deutsch, Nathan Price, Leroy Hood, Carl Kesselman, Ian Foster, Joseph Ames, Arthur Toga

University of Chicago, University of Southern California, Institute for Systems Biology, University of Michigan

ark/88120/r83w2c

BIG DATA FOR DISCOVERY SCIENCE

Abstract

Objective: Investigate if there are commonalities in patterns of ***amyloid deposition*** in individuals with Alzheimer's Disease or Parkinson's Disease that identify those individuals with or at risk for cognitive dysfunction

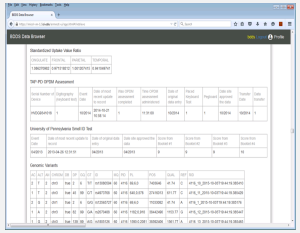
Approach: We created a ***platform*** for integration of multi-omic data, facilitate data discovery, cohort creation, enable rapid, scalable and reproducible analysis and finally publish results

Approach

We employed a systematic, reproducible approach that leveraged existing capabilities for understanding commonalities in amyloid deposition that included the following stages and challenges

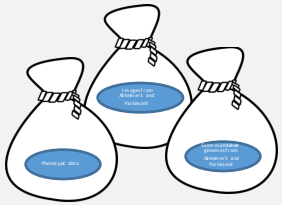
Data Ingest	• Multi-omic data • Multiple distributed repositories • Data usage agreements • Data cleanup and integration
Data Query and Exchange	• Easy to find data • Enable creation of patient cohorts • Send data for analysis
Data Analysis	• Multiple analysis platforms • Ease of use • Scale
Data Publication and Integration	• Enable discovery • End-to-end data view

The BDDS Platform



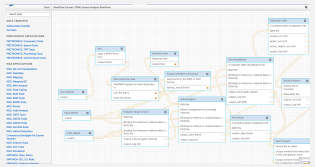
1. Data Ingest - ERMRest

- Phenotypic, Genotypic and Imaging datasets
- Multiple repositories
- Data Usage Agreements
- Automated data clean up and ingest



2. Data Query and Exchange - BagIt

- ERMRest provides high-level API for query
- Adopted the BagIt specification to create databags
- Enhanced BagIt to support big biomedical data

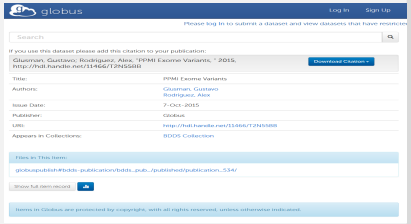


3. Data Analysis – Pipeline and Globus Genomics

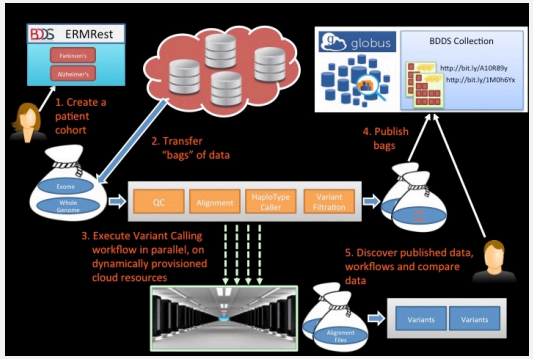
- LONI Pipeline runs the analysis on the amyloid PET and MRI images and outputs amyloid index values (SUVR: standard uptake value ratio) for each brain region, for each subject using computational resources at USC
- Genomics analysis using Globus Genomics on the Amazon cloud
- Leverages elastic provisioner that optimizes performance and costs

4. Data Publication and Integration – Globus Publication and ERMRest

- Results from analysis are integrated with ERMRest
- Results are published in Globus Publication service with appropriate metadata



Analyzing role of Amyloid Burden



Discussion & Conclusions

- We created a powerful platform by integrating several existing services and capabilities
- Platform development resulted in adopting and extending BagIt format for biomedical big data
- Future work includes applying the platform for analyzing Alzheimer's data
- Reference implementation of the NIH Commons

Bibliography

1. The BagIt File Packaging Format (V0.97) available at : http://tools.ietf.org/html/draft-kunze-bagit-11
2. Minids: http://minid.bd2k.org