Project I: User Case

- Develop *AD Knowledge Base*, a computational system that enables users to mine genomics data of Alzheimer's Disease (AD).
- Raw Data
  - Gene interaction (entrez id)
  - Expression profile (entrez id)
  - Entrez id uniprot mapping (entrez id -> uniprot id)
  - Uniprot KB and Schema (uniprot id)
  - Patient information

Project I: Requirement

- The knowledge base should at least answer the following questions in a quick response time:
  - Given a gene, find all of its n-order interacting genes
  - Given a gene, find mean and std of gene expression values for AD/MCI/NCI, respectively
  - Given a gene, find all other information associated with this gene.
  - Given a patient id, find all patient information (age, gender, education etc.)

Project I: Requirement

- A Python command-line client interface for database creation and query
- Use at least three types of SQL or NoSQL stores (MySQL, Key-value, Document, Column Family, or Graph)

Project I: Requirement

- Document (no hand-writing, in print!)
  - Design diagram
  - All queries
  - Potential improvements (e.g. how to speed up query)
- All source codes
- A README file detailed how to install and run the codes
- Two-person team
- Sent through blackboard
- All documents and codes are in a fold named after Student1LastName_Student2LastName
- Due: 11:59pm, April 6
- Project review: April 9

Project I: Rubric

- Database design: 30%
- Query functionality: 30%
- Client interface: 20%
- Documentation (including REAME) and presentation: 20%
- Bonus point: 20%
  - A Django graphic interface