



**CLUSTERING OF DISEASES
TO ASSIST IN DRUG
DISCOVERY FOR RARE
DISEASES**

TEAM 5B:

Stephan Villacencio
Abdulmelik Mohammed
Lina Fajardo Gomez
Kevin Nuno
Laura Borgerding

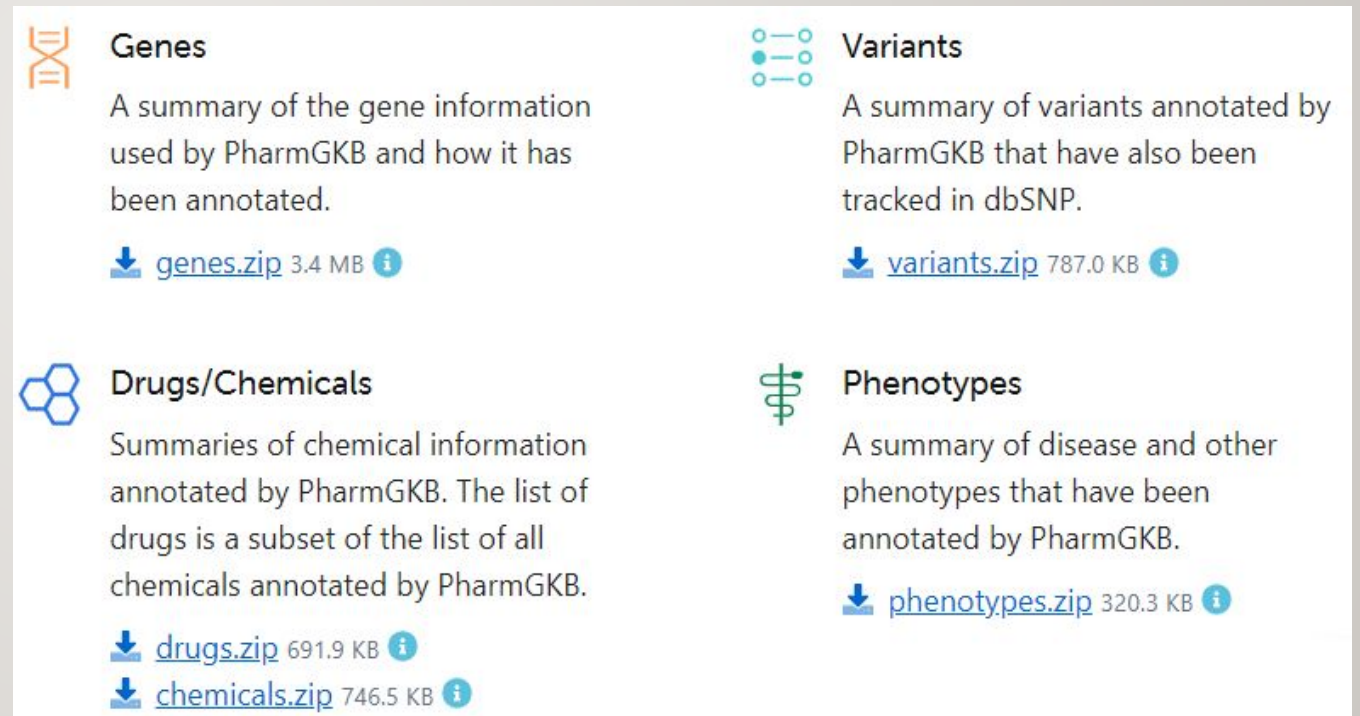
**FINAL PRESENTATION
MARCH 2ND 2022**

PROPOSAL: GENERAL AND RARE DISEASE COMBINED DATA SET AND CLUSTER ANALYSIS

- Are there associations between general and rare diseases?
- Can genomic trends in clusters suggest treatment?
- How is pharmacogenomic and pharmacological information utilized to analyze clusters?

Orphanet provides clinical and pharmacologic data regarding rare diseases and treatments

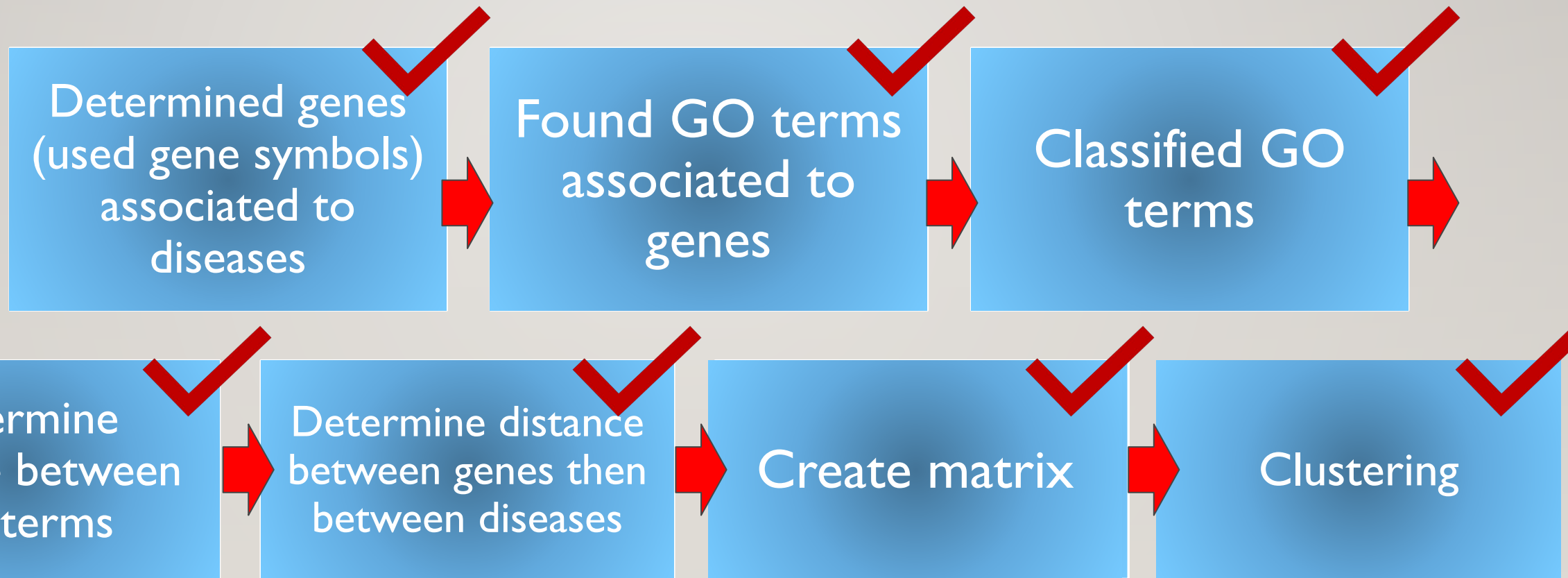
PharmGKB shares clinical data from participants of research studies



The screenshot displays four data categories available for download from PharmGKB, each with a description and a download link with file size and an information icon.

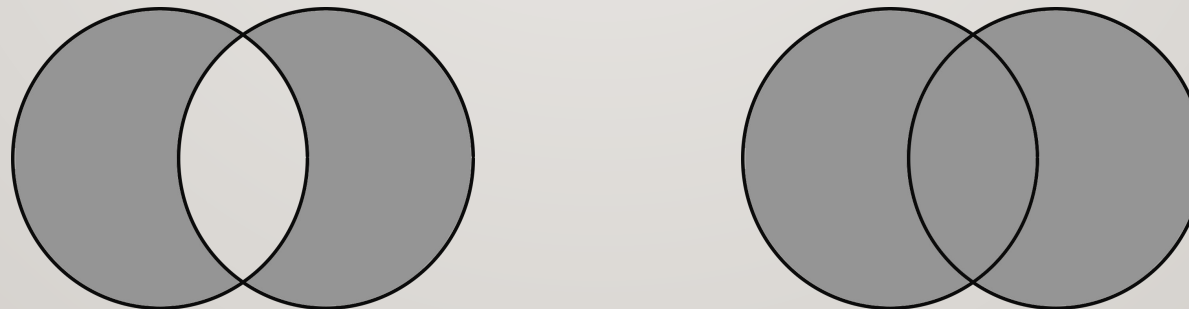
Category	Description	Download Link	File Size
Genes	A summary of the gene information used by PharmGKB and how it has been annotated.	genes.zip	3.4 MB
Variants	A summary of variants annotated by PharmGKB that have also been tracked in dbSNP.	variants.zip	787.0 KB
Drugs/Chemicals	Summaries of chemical information annotated by PharmGKB. The list of drugs is a subset of the list of all chemicals annotated by PharmGKB.	drugs.zip chemicals.zip	691.9 KB 746.5 KB
Phenotypes	A summary of disease and other phenotypes that have been annotated by PharmGKB.	phenotypes.zip	320.3 KB

WORK COMPLETED

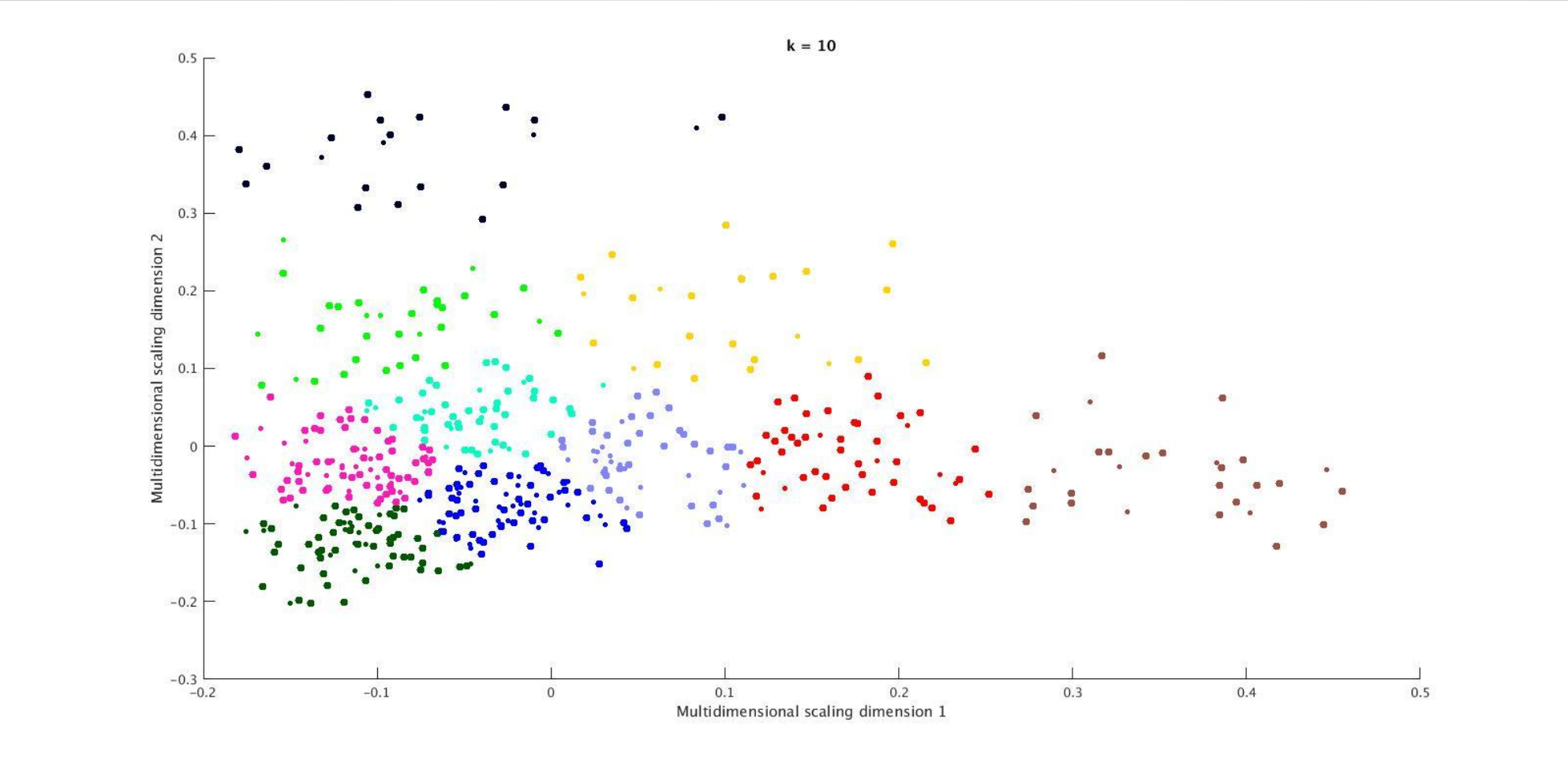


Distance between diseases gives a distance matrix

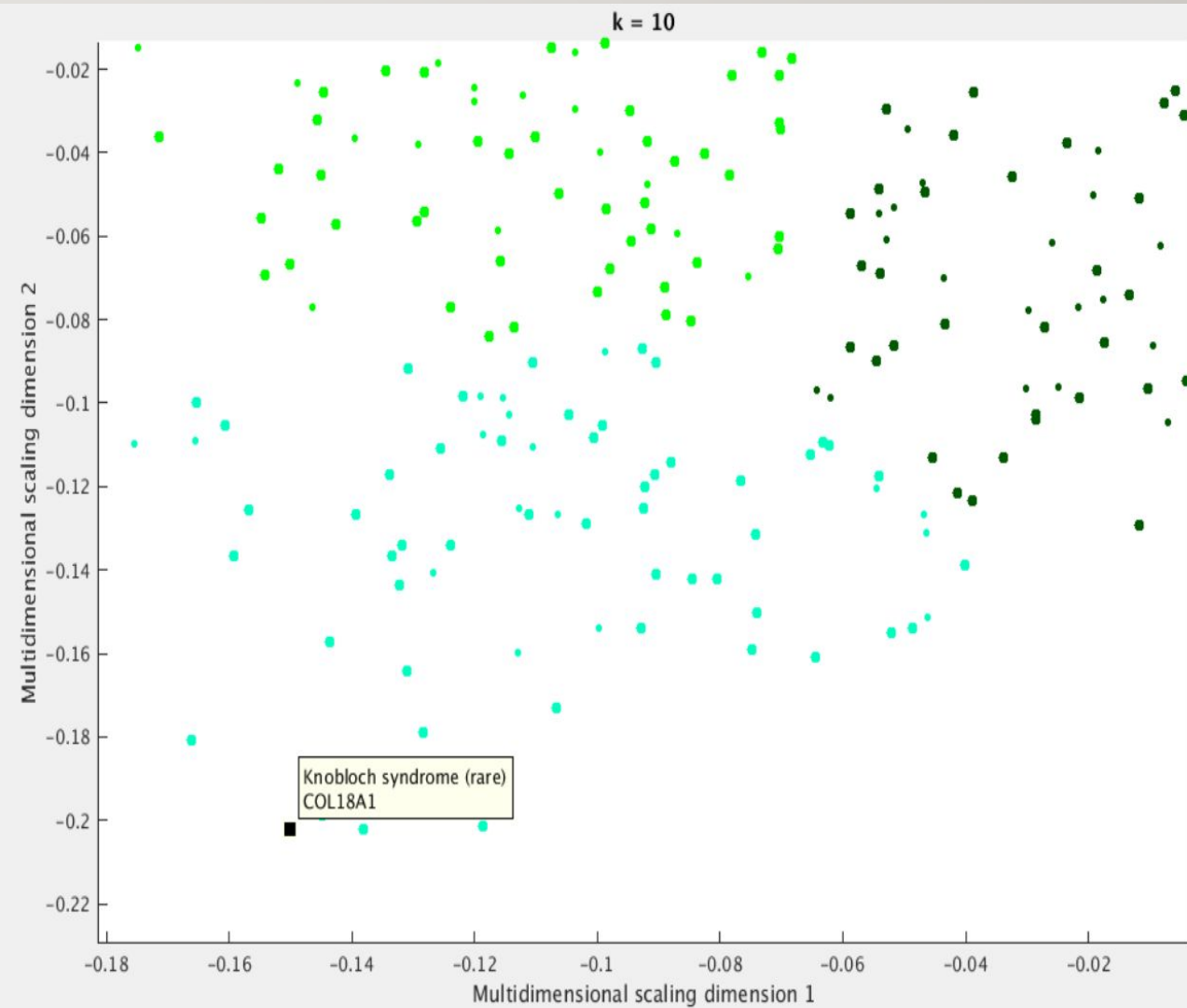
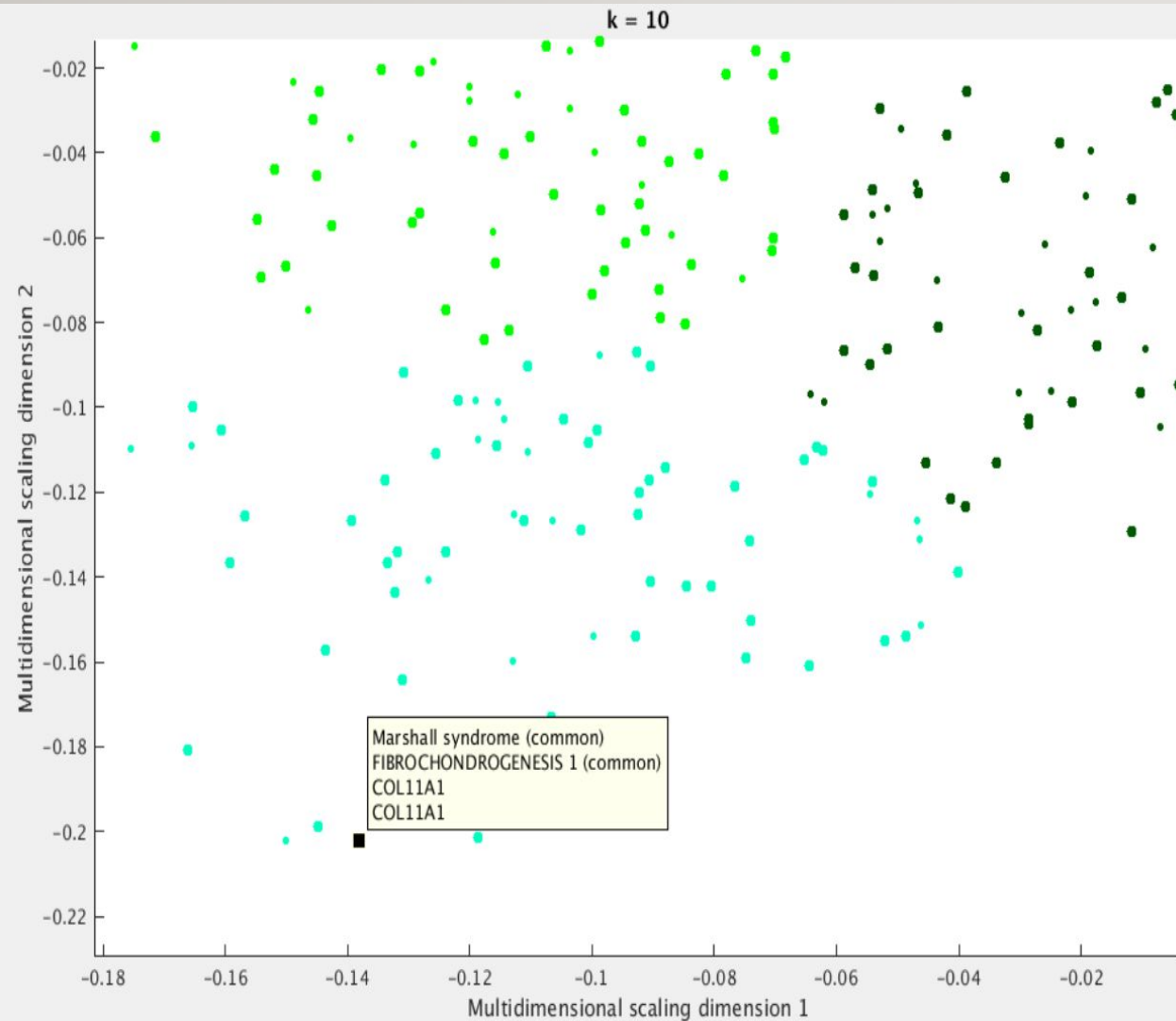
- Inferred from distance between genes that are associated with the diseases
- Dissimilarity score between genes computed from the intersection between the GO terms linked with the genes



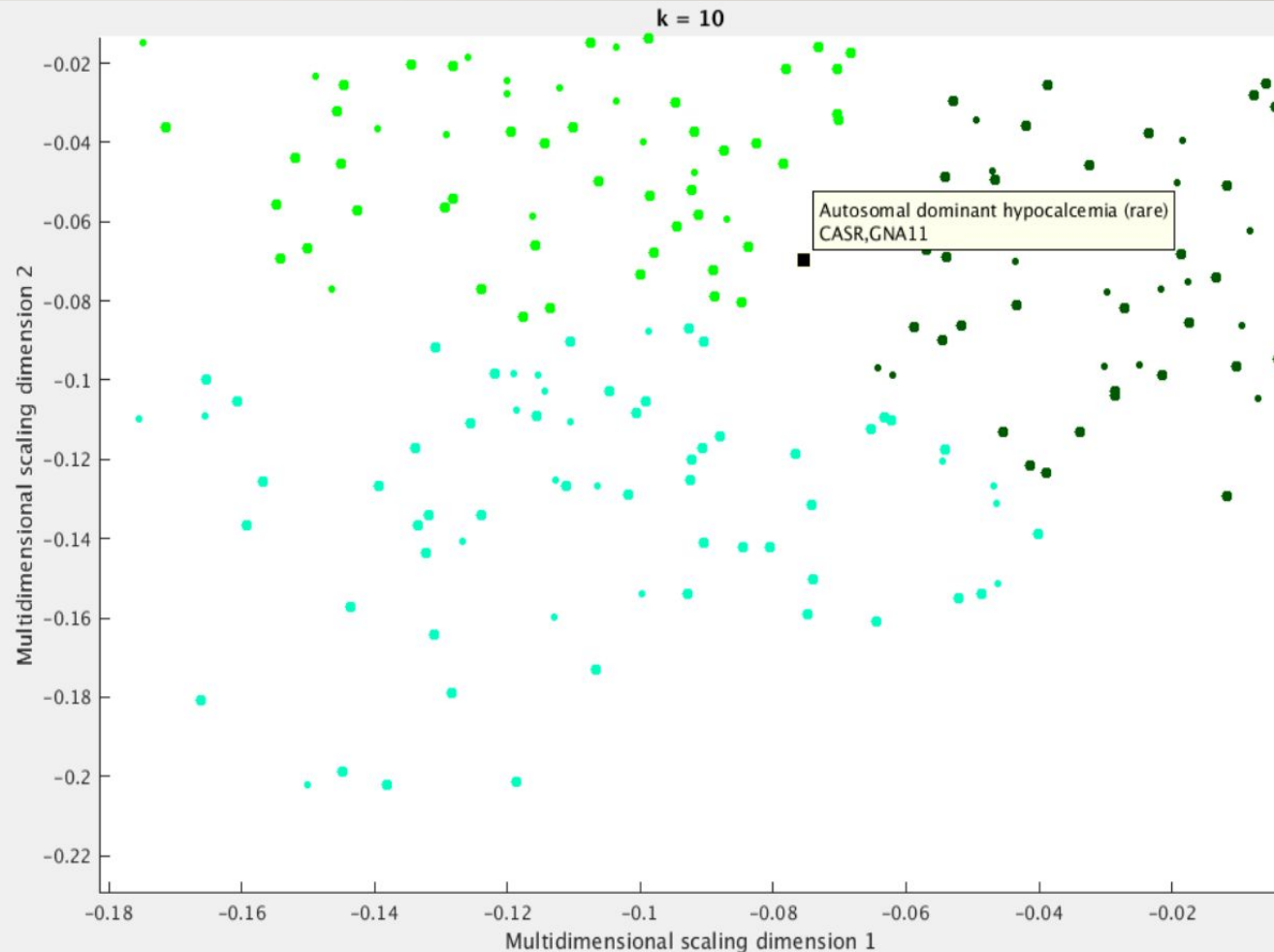
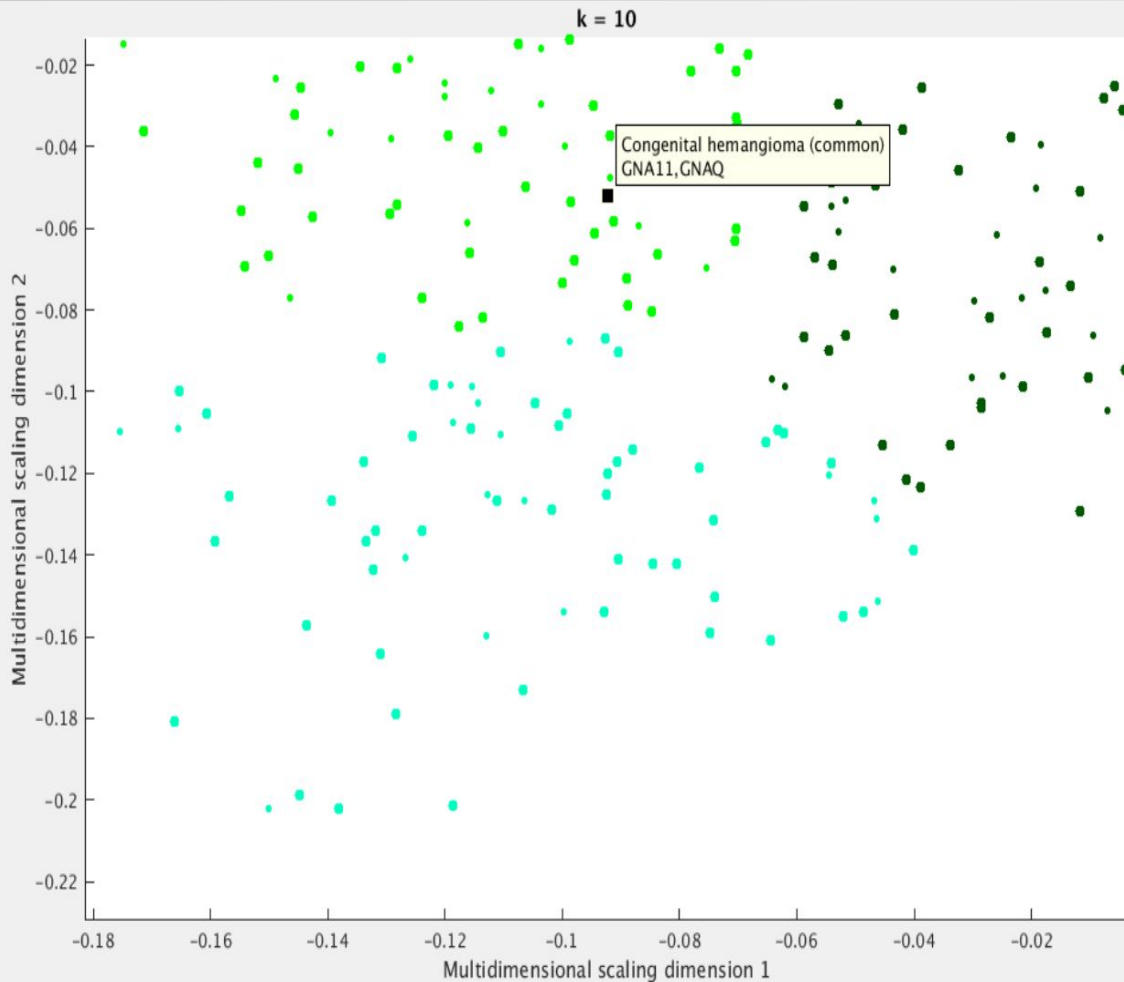
Multidimensional scaling of distance matrix followed by clustering (500 diseases)



EXAMPLE #1 OF RELATED DISEASES: MARSHALL SYNDROME VS KNOBLOCH SYNDROME - COL11A1/COL18A1



EXAMPLE #2 OF RELATED DISEASES: CONGENITAL HEMANGIOMA VS AUTOSOMAL DOMINANT HYPOCALCEMIA - GNA11



AFTER CLUSTERING: ANALYSIS (Ongoing)

What are the trends in each cluster?

- No organ system trends found, only gene related
- **Considering review of drugs associated to diseases**

How do rare diseases compare to the rest?

- Rare diseases are more likely to be linked to genetic onset
- Does provide some insight on related common diseases

Do different metrics yield new insights?

- Unable to use alternative metric for different clustering due to time constraints

LIMITATIONS



Distance metric

Smaller dataset

Limited data
available

MOVING FORWARD

Expand dataset

Add 1st line
treatment to
graph

Consider
alternate cluster
based on PK data

DATASETS USED

- [PharmGKB](#) (contains clinical data from research studies)
- [DisGeNET](#) (disease-gene association database)
- [Orphanet](#) (information about rare diseases and orphan drugs)