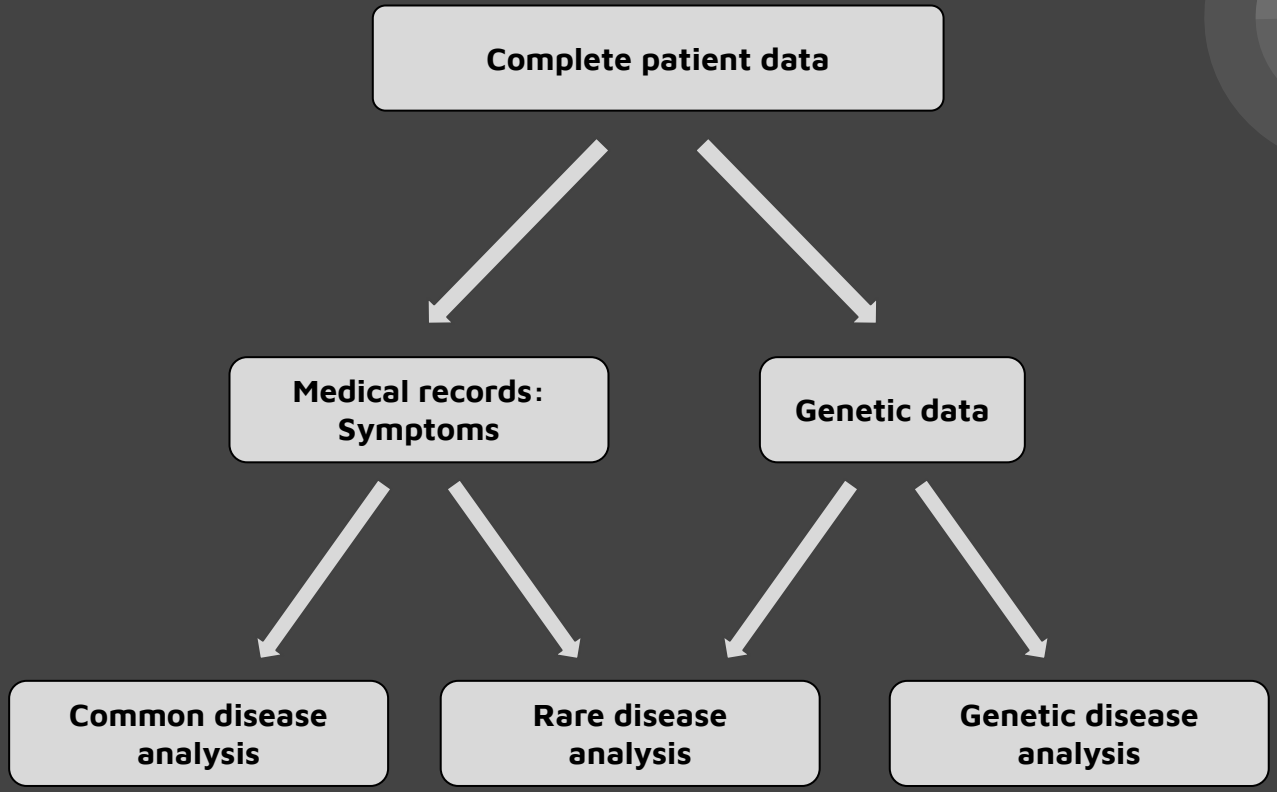
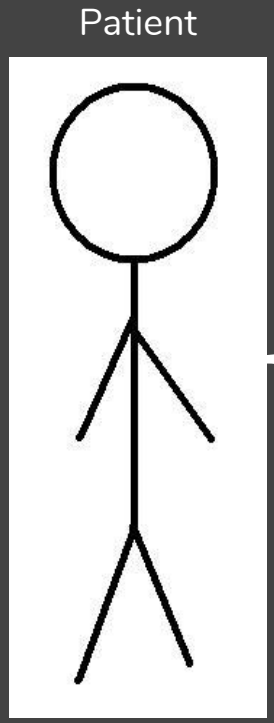
Automated diagnostics via disease-feature correlation

Toronto group for Undiagnosed-1:
Liam Haas-Neill, Duncan Kirby, Eugene Klyshko, Chris Nunn, Jeremy Rothschild & Matt Smart

Approach

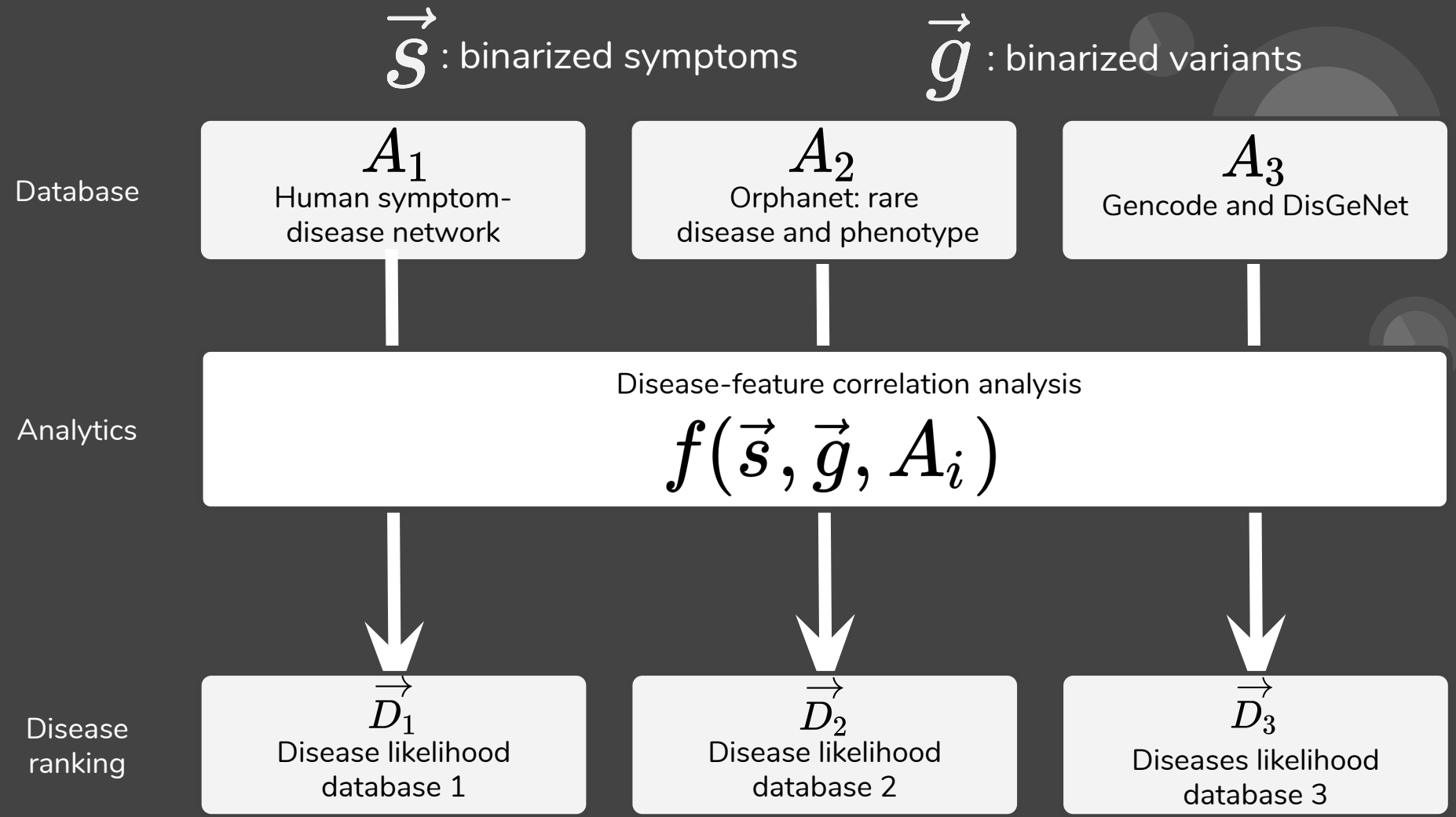

Complete patient data
Medical records: Symptoms
Genetic data
Common disease analysis
Rare disease analysis
Genetic disease analysis

Methodology



$\vec{s}$ binarized vector of symptoms

$\vec{g}$ binarized vector of genetic variants


$\vec{s}$: binarized symptoms
$\vec{g}$: binarized variants
Database
A_1
Human symptom-disease network
A_2
Orphanet: rare disease and phenotype
A_3
Gencode and DisGeNet
Analytics
Disease-feature correlation analysis
$f(\vec{s}, \vec{g}, A_i)$
Disease ranking
$\vec{D_1}$
Disease likelihood database 1
$\vec{D_2}$
Disease likelihood database 2
$\vec{D_3}$
Diseases likelihood database 3

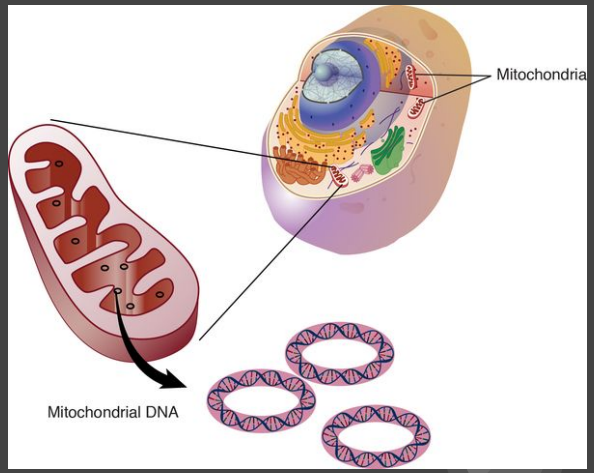
Diagnosis A - Early Parasitic infection

- Schistosomiasis mansoni came up #1 in our non-genetic analysis pipeline

- Early parasite infection (not necc. schistosomiasis) aligns well with
  - Sudden appearance of gastro symptoms during “trip” at age 2
  - Failure to thrive, very low percentile height and weight age 2 onwards
  - Immune system abnormalities; chronically high
    - Histamine
    - Interleukin-beta1
    - C-reactive protein
  - Intermittently elevated liver inflammation markers (ALT, AST) since young age
  - Primary symptoms revolve around gastrointestinal abnormalities esp. gastroparesis

- Remarks:
  - Patient does not currently have info on trip/vacation location at age 2
  - Normal eosinophil levels suggest absent or “non-active” (dormant) parasite infection
  - Scenario 1: he could have ongoing, mild (non-immunogenic?) parasitic infection -> identify pathogen, treat
  - Scenario 2: infection, followed by clearance, altered his immune and gastrointestinal development

Diagnosis B - MNGIE



- Mitochondrial Neurogastrointestinal Encephalopathy Disease (MNGIE)
- MNGIE came up #1 in our “rare-disease” non-genetic analysis
- High symptom alignment:
  - Neuro-gastrointestinal symptoms match patients gastroparesis and cachexia
  - Progressive pattern of disease with acceleration around age 20 fits the patient
- Case studies note heterogeneity in clinical and genetic profile
  - Homozygous variants in nuclear gene TYMP are common (regulates mtDNA synthesis)
- Remarks:
  - Recommend testing for key diagnostic criteria
    - Bloodwork : Increased plasma thymidine (>3 μmol/L) and deoxyuridine (>5 μmol/L)
    - Brain MRI: Asymptomatic leukoencephalopathy
    - Additional test: mtDNA copy number abnormalities
  - TYMP: patient heterozygous for significant insertions; homozygous for SNP (low confidence)
  - Patient is heterozygous for numerous VUS in other nuclear-encoded mitochondrial genes

Conclusion

We developed a pipeline to take binary input of symptoms/genetic variants and outputs a likelihood of diseases

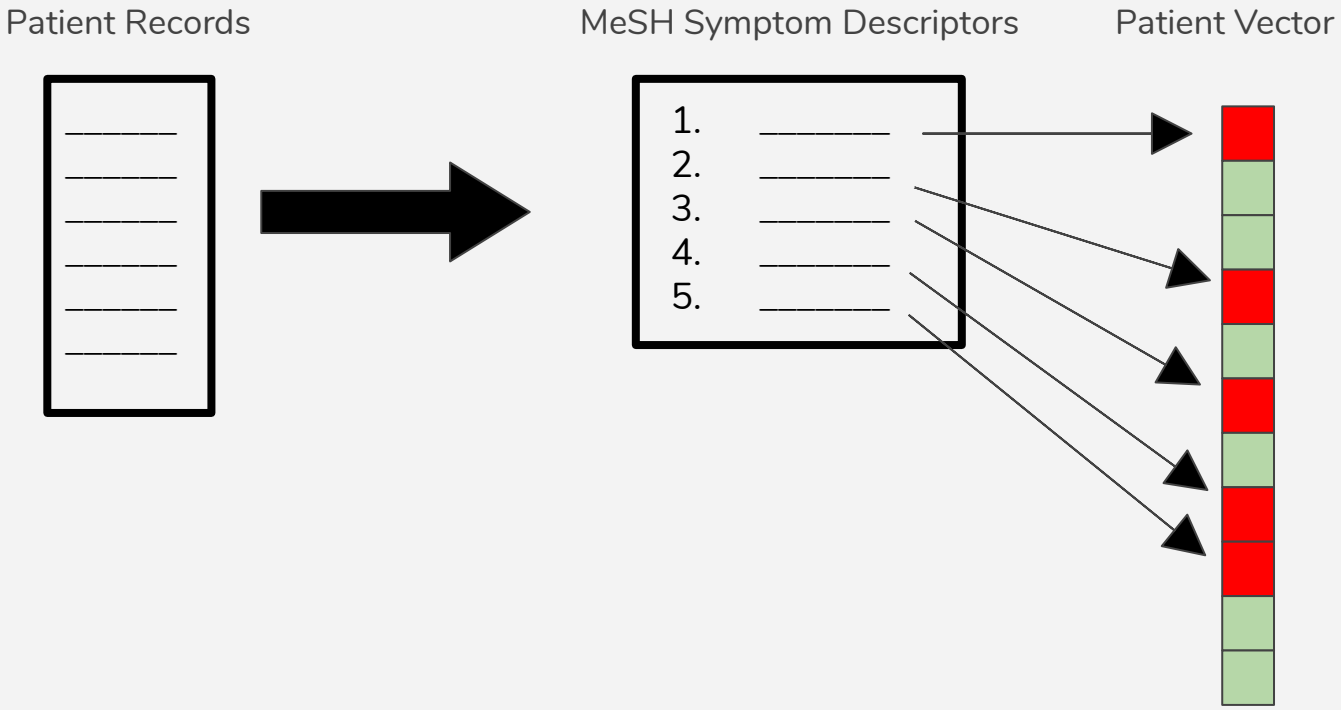
Diagnosis: **MNGIE** and/or **early parasitic infection**

Future work may include more rigorous analysis of the genetic variant to disease matrix

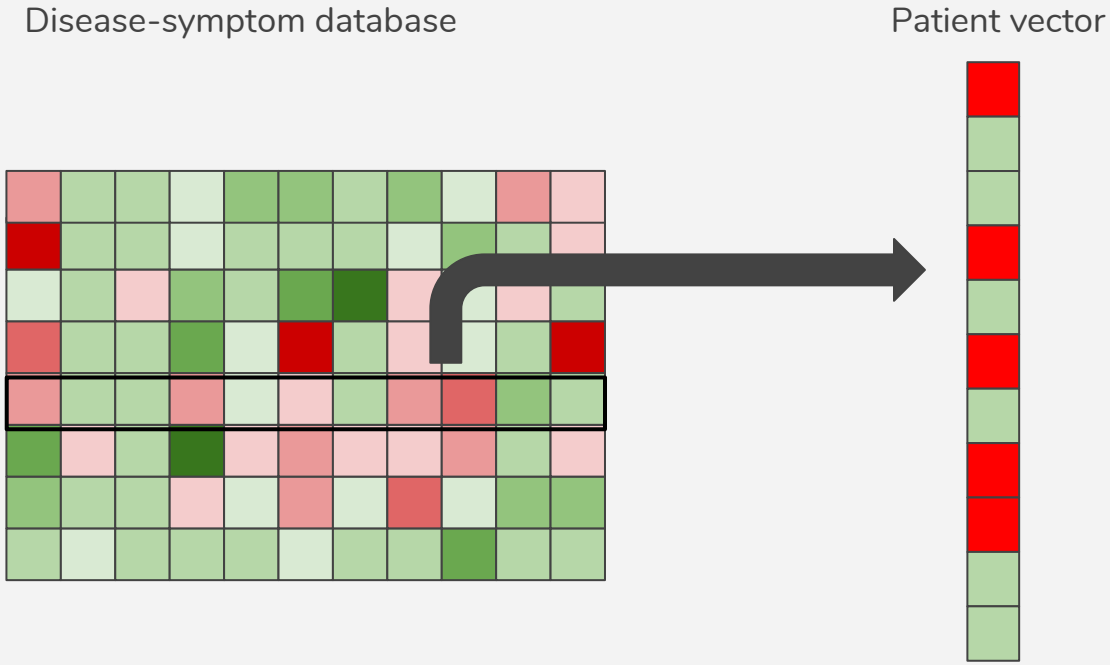
Questions? Email: rothschild@physics.utoronto.ca

SUPPLEMENTARY

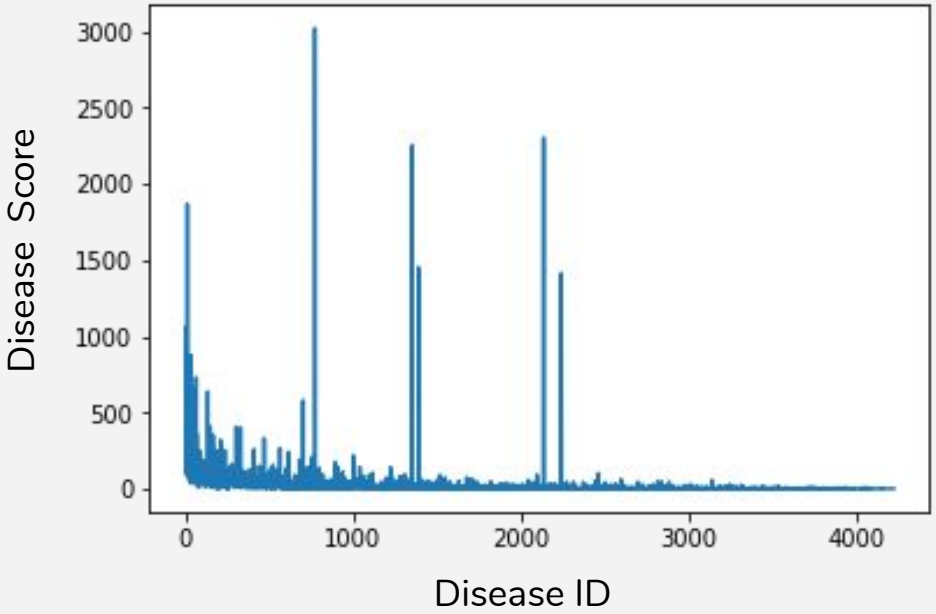
Methodologies



Methodologies



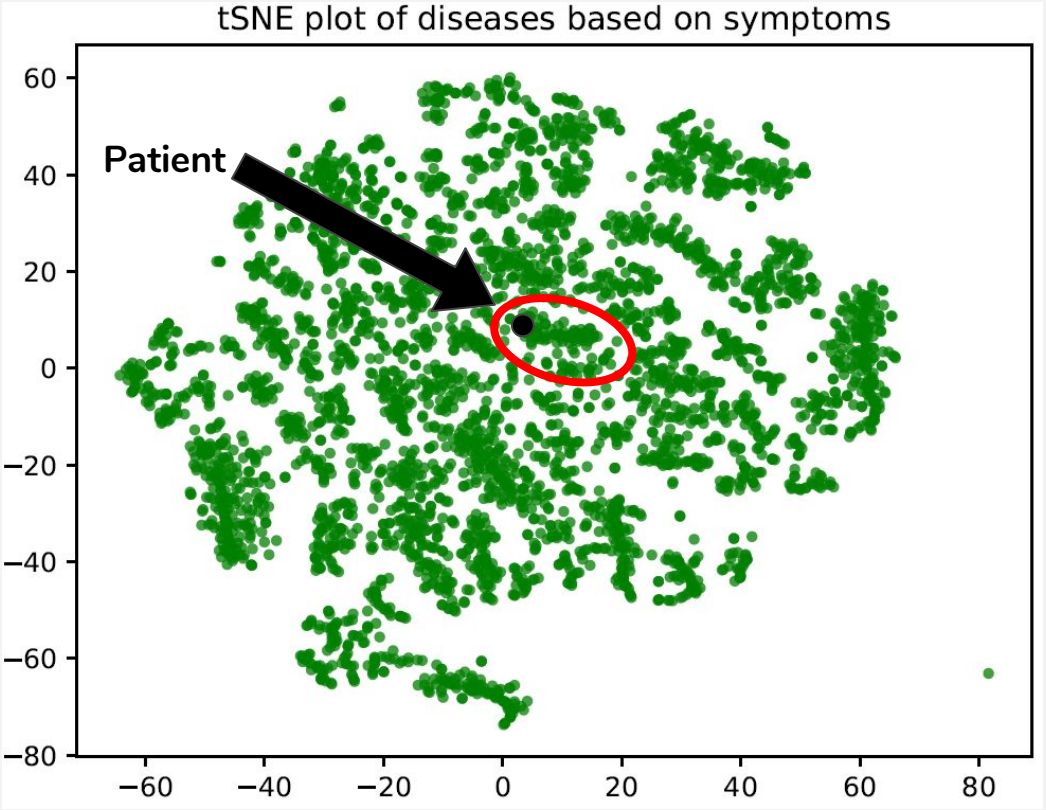
Methodologies



Ranked Diagnoses

1. _____
2. _____
3. _____
4. _____
5. _____
6. _____

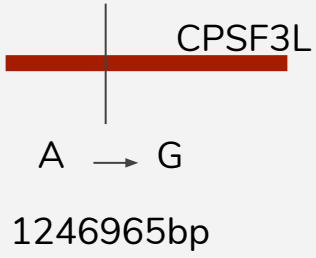
Clustering with tSNE



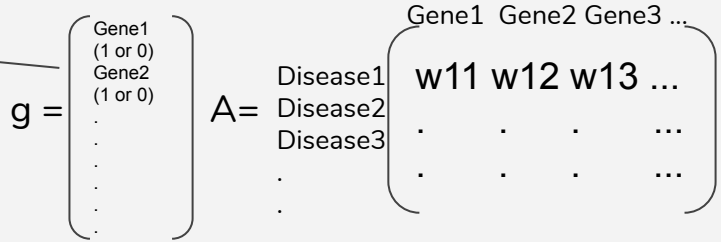
Genetic branch approach

All mutations from VCF files → (1) → List of unique genes containing at least one mutation → (2) → Ranked list of diseases associated with these genes

1: Compared mutation locations with canonical gene locations



2: Multiplied binary gene vector (g), with normalized gene-disease association database matrix (A).



Output disease list (d = g.A)

Genetic analysis results (list of top-10 candidates):

ALL genetic screens (excluding Exome)	ONLY cardio-neuro	ONLY carrier	ONLY pediatric
1 Diarrhea 5, With Tufting Enteropathy, Congenital 2 Borjeson-forssman-lehmann Syndrome 3 Duplication 15q11-q13 Syndrome 4 Ear Diseases 5 Charcot-marie-tooth Disease, X-linked, 1 6 Diastrophic Dysplasia 7 Achondrogenesis, Type Ib (Disorder) 8 Epiphyseal Dysplasia, Multiple, 4 9 Mental Retardation, Autosomal Recessive 18 10 Spinal Muscular Atrophy,	1 Kuru 2 Bardet-biedl Syndrome 13 3 Nephrotic Syndrome, Congenital, With Ocular Abnormalities And Congenital Myasthenic Syndrome 4 Familial Mesangial Sclerosis 5 Pterygium, Antecubital 6 Nephrotic Syndrome, Type 5, With Or Without Ocular Abnormalities 7 Hypomyelination With Brainstem And Spinal Cord Involvement And Leg Spasticity 8 Pierson Syndrome 9 Episodic Ataxia, Type 6 (Disorder) 10 Prion Diseases	1 Parkinson Disease 14, Autosomal Recessive 2 Amyloidosis, Cerebral, With Spongiform Encephalopathy 3 Peroxisome Biogenesis Disorder 12a (Zellweger) 4 Peroxisome Biogenesis Disorder, 5 Peroxisome Biogenesis Disorder, 6phosphoribosylpyrophosphate Synthetase Superactivity 7 Deafness, X-linked 1 (Disorder) 8 Charcot-marie-tooth Disease, 9 Ataxia, Fatal X-linked, 10 Familial Alzheimer-like Prion Disease	1 Ceroid Lipofuscinosis, Neuronal, 7 2 Hypomyelination With Brainstem And Spinal Cord Involvement And Leg Spasticity 3 Psychomotor Retardation, Epilepsy, And Craniofacial Dysmorphism 4 Neurodegeneration With Brain Iron Accumulation 2 (Disorder) 5 Brachydactyly syndactyly oligodactyly Syndrome 6 Cataract 20, Multiple Types 7 Mental Retardation, Autosomal Dominant 31 8 Microphthalmia, Isolated 3 9 Mental Retardation, Autosomal Recessive 41 10 Atrioventricular Septal Defect, Partial, With Heterotaxy Syndrome

Genetic branch: future work

Increase Pipeline Sophistication

(a) Implement more stringent mutation filtering (using PHRED quality, read quality, mapping quality etc.)

(b) Weight affected genes by the number of mutations observed.

(c) Compute variants across multiple reference genomes.

Exploration of Diagnosis B (MNGIE)

MNGIE associated diseases can result in a reduction in mtDNA copy number:

It would be useful to compare mtDNA read coverage to nuclear coverage while carefully handling coverage bias (i..e AT content bias in Illumina sequencing).

References:

Gene-Disease Associations

http://www.disgenet.org/downloads

Comprehensive gene annotation:

https://www.gencodegenes.org/human/

Human symptoms-disease network.

Zhou X, Menche J, Barabási AL, Sharma A (2014) Human symptoms-disease network. Nat Commun 5(May). doi:10.1038/ncomms5212.

Rare diseases database

https://www.orpha.net/consor/cgi-bin/index.php

Supplementary: TYMP Mutations

chr22	50527818	.	G	A	LowQual		1/1
chr22	50528362	.	C	T	LowQual		0/1
chr22	50528483	.	GCGGCGGTGACGGC	GCGGCGGTGACGGCGGTGACGGC			0/1
chr22	50528496	.	C	CGGTG			0/1
chr22	50528497	.	A	ACGGCG			0/1
chr22	50528569	.	A	T			0/1
chr22	50528623	.	C	A			0/1
chr22	50529134	.	T	C			0/1
chr22	50529146	.	C	T			0/1
chr22	50529148	.	C	G			0/1

X-ray structure of the thymidine phosphorylase from Salmonella typhimurium in complex with cytidine and sulphate