Data Sharing, GeneMatcher and VariantMatcher

Nara Sobreira, MD, PhD
Johns Hopkins University
McKusick-Nathans Department of Genetic Medicine
Baylor-Hopkins Center for Mendelian Genomics
Data Sharing and Disease Gene Discovery

- Next-generation sequencing
  - Big data
    - Disease gene discovery
The number of genes associated with phenotypes listed in OMIM rose from ~2,500 (12.5%) to ~4,500 (~22%) over the last 11 years.
Next-generation sequencing

Big data

Disease gene discovery

Data Sharing
Data Sharing

- What data to share?
  - What data is needed for disease gene and variant discoveries?
  - What data can be shared to ensure patient confidentiality and privacy?
- How to share the data?
- Where to share the data?
- With whom to share data?
- How to ensure that the researchers that generated the data are recognized for their work?
- How to share the data in compliance with the IRB requirements?
Data Sharing Databases

- Restricted, NIH-based databases
  - dbGaP (The database of Genotypes and Phenotypes)
  - AnViL (The NHGRI Genomic Data Science Analysis, Visualization and Informatics Lab-space)
  - CAVATICA (Gabriella Miller Kids First)

- Public, variant-based databases
  - DECIPHER (copy number variation)
  - dbSNP
  - Exome Variant Server
  - 1000 Genomes Project
  - ExAC Browser - Harvard Medical School
  - gnomAD (The Genome Aggregation Database)
  - ClinVar
  - HGMD (Human Gene Mutation Database)
What is missing?

- Detailed phenotypic features
- Detailed family history
- Capability to reassess the individual with the variant of interest
How to address some of the problems?

- Share little but share the essential
  - Gene and/or variant of interest
  - The name and contact information of the individual responsible for the data
- Ensure simultaneous notification about the match
GeneMatcher

GeneMatcher is a freely accessible web site designed to enable connections between clinicians and researchers from around the world who share an interest in the same gene or genes.

Email: Required...
Password: Required...

Log in

Forgotten your password?
Create an account...

Baylor-Hopkins Center for Mendelian Genomics

VariantMatcher

Centers for Mendelian Genomics
GeneMatcher

- To enable connections between patients, their families, clinicians and researchers from around the world who share an interest in the same gene or genes
- De-identified data
- Automated matching
- Submitters choose to follow up at their discretion
- Also matching on phenotypic features, genomic location and OMIM disease number
Growth in number of genes and genes matched in GeneMatcher

As of 1 July 2021
- 13,707 genes from 11,729 submitters (93 countries)
- 8,671 (~63%) matched genes

https://genematcher.org/
Who are the GeneMatcher users?

- 7,980 users describe themselves as researchers
- 5,220 as health care providers
- 635 as patients
- Of the 13,707 genes in GeneMatcher, 323 were submitted by model organism researchers
  - 202 were from mouse
  - 55 from zebrafish
  - 38 from fruit fly
  - 14 from yeast
  - 14 were related to other non-human organisms
How has GeneMatcher contributed to disease gene discovery?

- GeneMatcher is cited in > 484 publications
- As of June 12th 2020, GeneMatcher was acknowledged in 325 publications and 302 genes were described in 249 publications
  - 157 were related to an autosomal dominant phenotype
  - 105 were related to an autosomal recessive phenotype
  - 24 were related to an autosomal dominant and autosomal recessive phenotype
  - 6 were related to a X-linked recessive or dominant phenotype
- 221 of these genes (~73%) were on GeneMatcher for more than a year before publication
- 231 genes (76.5%) are currently associated to a phenotype in OMIM
- 73 (~32%) had an OMIM phenotype before the gene was posted to GeneMatcher
- Among the genes entered >5 times in GeneMatcher and that had an OMIM phenotype as of 1 June 2020 (1,012), 445 genes were associated with an OMIM phenotype after being deposited in GeneMatcher
GeneMatcher and Matchmaker Exchange
Acknowledgments

- Nara Sobreira - JHU
- Renan Martin
- Elizabeth Wohler
- Eliete Rodrigues
- Corina Antonescu
- Kim Doheny
- Sean Griffith
- Alice Sanchez
- Laura Vail
- Jim Lupski - Baylor
- David Valle – JHU
- Ada Hamosh - JHU

Baylor-Hopkins Center for Mendelian Genomics