PhenomeCentral: an integrated portal for sharing patient phenotype and genotype data for rare genetic disorders

Michael Brudno
Centre for Computational Medicine
Hospital for Sick Children & University of Toronto

European Conference on Rare Diseases
RareConnect Workshop
Rare diseases and the long tail

- **6.2%** total prevalence of rare diseases in Orphanet
- **4.8%** heart disease prevalence in Canada in 2007 (PHAC)

Rare: < 1 / 2,000 (in EU)
Finding the cause of a genetic disease...

1. GENE:185delA

2. GENE:185delA

3.
Finding additional families can be difficult...

rare diseases are rare.
rare genetic disease
› might not recognize known disease

› insufficient sample size for novel gene
Data Sharing Critical!
Deep Phenotyping

› Describe the features of an individual, rather than a disease

– To enable a diagnosis (especially of a rare disease)

– To distinguish between similar disorders

– To enable genotype-phenotype correlations
### Previous State of Clinical Phenotyping

- Two Alternatives: free text or checkboxes

**Dysmorphic features**
- df
- dysmorphic
- dysmorphic faces
- dysmorphic features

**Congenital malformation/anomaly:**
- congenital anomaly
- congenital malformation
- congenital anomaly
- congenital anomaly
- congenital anomaly
- congenital anomaly
- congenital anomaly
- congenital anomaly
- congenital anomaly
- congenital anomaly
- congenital anomaly

**Examples of lists:**
- dd. cong. malfor. behav. pro.
- dd. mental retardation
- df< delayed puberty
- df & li
- mental retard short stature

### Phenotypic description (Clinical symptoms)

<table>
<thead>
<tr>
<th>Behavior, Cognition and Development</th>
</tr>
</thead>
<tbody>
<tr>
<td>Global development delay</td>
</tr>
<tr>
<td>Fine motor delay</td>
</tr>
<tr>
<td>Gross motor delay</td>
</tr>
<tr>
<td>Language delay</td>
</tr>
<tr>
<td>Learning disability</td>
</tr>
<tr>
<td>Mental retardation</td>
</tr>
<tr>
<td>Moderate</td>
</tr>
<tr>
<td>Severe</td>
</tr>
<tr>
<td>Attention deficit hyperactivity disorder</td>
</tr>
<tr>
<td>Autism</td>
</tr>
<tr>
<td>Pervasive developmental delay</td>
</tr>
<tr>
<td>Psychiatric disorders (Specify below)</td>
</tr>
<tr>
<td>Other: ____________________________</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Cardiac</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASD</td>
</tr>
<tr>
<td>VSD</td>
</tr>
<tr>
<td>AV canal defect</td>
</tr>
<tr>
<td>Coarctation of aorta</td>
</tr>
<tr>
<td>Tetralogy of fallot</td>
</tr>
<tr>
<td>Other: ____________________________</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Craniofacial</th>
</tr>
</thead>
<tbody>
<tr>
<td>Craniostenosis</td>
</tr>
<tr>
<td>Cleft lip</td>
</tr>
<tr>
<td>Cleft palate</td>
</tr>
<tr>
<td>Microtremognathia</td>
</tr>
<tr>
<td>Retrognathia</td>
</tr>
<tr>
<td>Facial dysmorphism (Specify below)</td>
</tr>
<tr>
<td>Other: ____________________________</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Neurological</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hypotonia</td>
</tr>
<tr>
<td>Seizures</td>
</tr>
<tr>
<td>Ataxia</td>
</tr>
<tr>
<td>Dystonia</td>
</tr>
<tr>
<td>Chorea</td>
</tr>
<tr>
<td>Other: ____________________________</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Eye Defects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blindness</td>
</tr>
<tr>
<td>Coloboma</td>
</tr>
<tr>
<td>Epicanthus</td>
</tr>
<tr>
<td>Eyelid abnormality (Specify below)</td>
</tr>
<tr>
<td>Other: ____________________________</td>
</tr>
</tbody>
</table>
Problems with the status quo

- Phenotypic descriptions that are very evocative for humans, unreadable to a computer:
  - “first words at 5 years”
  - “has trouble spelling”
  - “recognizes only close relatives”

- Multiple terms have the same meaning:
  - “generalized amyotrophy”, “generalized muscle atrophy”
  - “muscular atrophy, generalized”

- It is difficult to define distances between phenotypes

- Cannot do computation with phenotypes!
Next-generation **phenotyping**

Human Phenotype Ontology (HPO):

- 11,000+ terms
- 100,000+ links to 5,000+ OMIM/Orphanet Disorders

- eye disease
- abnormal eye morphology
- coloboma
- globe abnormality
- neurologic
- skeletal

Peter Robinson & Monarch Consortium
Graduate Student Phenotyping

- [HP:0000708] Behavioural/Psychiatric Abnormality
  - Abnormal emotion/affect behavior
    - Abnormal aggressive, impulsive or violent behavior
  - Abnormal fear/anxiety-related behavior
    - Agoraphobia
    - Anxiety
    - Emotional lability
    - Mood changes
    - Mood swings
    - Anhedonia
    - Apathy
  - Autism spectrum disorder
  - Conspicuously happy disposition
  - Depression
  - Echolalia
  - Inappropriate behavior
    - Inflexible adherence to routines or rituals
  - Disinhibition
  - Inappropriate laughter
  - Inappropriate sexual behavior
  - Irritability
  - Lack of insight
  - Lack of motivation
  - Lack of spontaneous play
  - Low frustration tolerance
  - Mutism
  - Oppositional defiant disorder
  - Overfriendliness

- [HP:0100025] Overfriendliness
- [HP:0002193] Pseudobulbar behavioral symptoms
- Restlessness
- Restrictive behavior
- Shyness
Finding similar patients
Finding similar patients
Finding similar patients
Finding similar patients

\[ p_{\text{term}} \text{ from OMIM corpus} \]

\[ \text{IC}(\text{term}) = \log\left(\frac{1}{p_{\text{term}}}\right) \]

\[ \text{LS}(\text{term}) \approx \text{IC}(\text{term}) - \max_{\text{parents}} \text{IC}(\text{parent}) \]
Incorporating gene data

Exomizer
Incorporating gene data

Exomizer

Variation Score based on allele frequency and pathological impact

Whole exome

Remove off-target and common variants

Phenotypic Relevance Score based on similarity of observed phenotypes

Phenotypic interpretation of variants in exomes (PHIVE) to give final candidate(s)

(Robinson et al., 2014)
Incorporating gene data
Validation

491 rare disease patients, 171 in cohorts, 78 with known/lead genes

› Consistently Finds Similar Patients
  ▪ 73% top phenotype matches are in the same cohort

› Prioritizes solved or lead genes
  ▪ 3x known/lead genes ranked in top 5 compared to single exome
to encourage data sharing we built a user-friendly, privacy-aware portal for discovering patients similar to yours
PhenomeCentral is a Matchmaker
  – Find out about other similar patients
  – Easily connect with other clinicians

Each Patient Record can be:
  – *Public* – Visible to all registered users
  – *Private* – Only visible to specified users/consortia
  – *Matchable* – Private visibility, but existence can be "discovered" by users who submit similar patients

phenomcentral.org
Step 1: submit your patient phenomecentral.org

- Select positive and negative HPO terms
- Predictive, error-tolerant search of HPO
- Add a VCF file and/or gene list
- Set permission and add collaborators
Step 2: see patients similar to yours
Step 3: contact the other submitter
MFDM patient (EFTUD2 mutation) matched to a known one despite atypical presentation

includes data from:

[Logos of CARE for RARE, RD Connect, and NIH Undiagnosed Diseases Program]
Two similar patients with STIM1 mutations matched despite inconsistent terminology.
PhenomeCentral

>2000 cases

includes data from:

>500 users

phenomecentral.org
PhenomeCentral.org

Development Team:
Marta Gîrdea, Orion Buske, Sergiu Dumitriu, Bailey Gallinger, Heather Trang, Jonathan Zung, Andriy Misyura, Anton Kats,

Consortia:
- CARE for RARE Canada
  Kym Boycott, Taila Hartley, Sarah Sawyer, Chandree Beauleiu
- NIH-UDP
  Neal Boerkoel, William Gahl, David Adams, William Bone
- Care for Rare Australia
  Tracy Dudding
- RD-Connect
  Hanns Lochmuller, Rachel Thompson

HPO, Exomiser, Monarch:
Peter Robinson, Melissa Haendel, Damian Smedley, Sebastian Kohler, Nicole Washington

Funding:
Genome Canada (CARE for RARE), CIHR, NSERC
The state of affairs as of the initial meeting in October 2013, ASHG.

Multiple disconnected projects
Multiple disconnected projects

- Gene and Phenotype (HPO)
- Variants
- LOVD
- Variome
- Disease and Variants
- Genome Connect
- Model Organisms
- Monarch
- Variant and Disease
- GEM.app
- Disease and VCFs
- Gene and Phenotype (HPO)
- Gene Yenta
- Phenome Central
- GEM.app
- PEER
- DECIPHER
- Matchmaker
- Exchange
- Gene Matcher
- undiag.
- Diseases
- Program
- GEM.app
- ClinGen
- Monarch
- Genome Connect
- PEER
- Multiple disconnected projects
- Disease and VCFs
- Gene and Phenotype (HPO)
- Gene and Phenotype (HPO)
- Gene and Phenotype (HPO)
Making a Match

Linking multiple databases is not without its challenges...
The Real Story

Exchanging Data

Gene Matcher

DECIPHER

Phenome Central

API v 1.0

Courtesy Ben Hutton, DECIPHER
API v1 Data Fields

ID (Mandatory) - The internal identifier (obfuscated or not) that can be used by the originating system to reference the patient data.

Label (Optional) - A name/identifier assigned by the user which can be used to reference the patient in a recognizable manner (in an email for example); it should not contain any personally identifiable information.

Query type (Optional)
  - Accepted values: once or periodic

Submitter (Mandatory) - Consists of contact information of the person that submitted the search

Gender (Optional)

Age of onset (Optional)

Mode of inheritance (Optional)

Disorders (Optional) - A list of OMIM (MIM:######) or OrphaNet (ORPHA####) identifiers, can be empty

Features – It is mandatory to have at least clinical features or gene(s ); having both is preferred
  - HPO terms for clinical features
  - gene name(s)
Currently Connected MME Services

Matchmaker Exchange

API v1.0

Gene Matcher

US

PEER

RD-Connect

Monarch

Canada

Phenome Central

GEM.app

Undiag. Diseases Network

DECIIPHER

Gene Yenta

LOVD

Café Variome

UK
MME Results in PC

Remote server: Decipher Production Server

Anyone who proposes to publish material which uses data obtained from the DECIPHER database agrees to:

- Acknowledge the DECIPHER Consortium; and
- Contact the coordinator of the centre that entered the data on any individual who they wish to include in their report and offer appropriate agreed recognition of their contribution, which may include co-authorship if the magnitude of the contribution warrants it to at least one representative from the project/participating centre (possibly the member who submitted the patient data). This can be achieved by emailing a request to decipher@sanger.ac.uk

Access to bulk data may be obtained from DECIPHER and is subject to a Data Access Agreement, in which the user certifies that no attempt to identify individual patients will be undertaken. The same restrictions apply to the public data displayed on this website: no one is authorized to attempt to identify patients by any means.

The DECIPHER consortium provides these data in good faith as a research tool, but without verifying the accuracy, clinical validity or utility of the data. The DECIPHER consortium, makes no warranty, express or implied, nor assumes any legal liability or responsibility for any purpose for which the data are used.

Showing 5 similar cases

<table>
<thead>
<tr>
<th>Remote ID</th>
<th>Case ID</th>
<th>Potential diagnoses</th>
<th>Contact</th>
<th>Local relevance</th>
<th>Remote relevance</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>263271</td>
<td></td>
<td>Undiagnosed</td>
<td>DECIPHER</td>
<td>★★★★☆</td>
<td>★★★★★★</td>
<td>SHOW PHENOTYPE AND GENOTYPE SIMILARITY...</td>
</tr>
</tbody>
</table>

PHENOTYPIC FEATURES BREAKDOWN

- **ABNORMALITY OF THE PHILTRUM**
  - The current patient (F0000010) presented with: Short philtrum
  - The matched patient (263271) presented with: Long philtrum

- **ABNORMALITY OF THE THUMB**
  - The current patient (F0000010) presented with: Broad thumb
  - The matched patient (263271) presented with: Short thumb

- **LOW-SET EARS**
  - The current patient (F0000010) presented with:
  - The matched patient (263271) presented with:
Matchmaker Exchange is a Demonstration Project for the GA4GH

- Success highly dependent on large international effort
- Critical need for standards
- Activity spans multiple workgroups
  1. Data (data format and interfaces)
  2. Regulatory and Ethics (patient consent)
  3. Security (patient privacy and user authentication)
  4. Clinical (phenotyping and matching algorithms)
Website Guided Use

Matchmaker Exchange
Genomic discovery through the exchange of phenotypic & genotypic profiles

www.matchmakerexchange.org
Guiding Use by the Community

These tables help define the database location, content and approaches for each MME service to help guide the user in choosing a MME service for case deposition.

<table>
<thead>
<tr>
<th>Matchmaker Exchange Site</th>
<th>Server Location</th>
<th>Phenotypes</th>
<th>Genotype</th>
<th>Candidates</th>
</tr>
</thead>
<tbody>
<tr>
<td>PhenomeCentral</td>
<td>Canada</td>
<td>√</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>GeneMatcher</td>
<td>USA</td>
<td>√</td>
<td>√</td>
<td>√</td>
</tr>
<tr>
<td>DECIPHER</td>
<td>UK</td>
<td>√</td>
<td></td>
<td>√</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Matchmaker Exchange Site</th>
<th>Parameters Used for Matching</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Gene</td>
</tr>
<tr>
<td>PhenomeCentral</td>
<td>√</td>
</tr>
<tr>
<td>GeneMatcher</td>
<td>√</td>
</tr>
<tr>
<td>DECIPHER</td>
<td>√</td>
</tr>
</tbody>
</table>
Acknowledgements

S Balasubramanian
Mike Bamshad
Sergio Beltran Agullo
Jonathan Berg
Kym Boycott
Anthony Brookes
Catherine Brownstein
Michael Brudno
Han Brunner
Orion Buske
Deanna Church
Raymond Dalgliesh
Victor de la Torre
Andrew Devereau
Johan den Dunnen
Sergiu Dumitriu
Helen Firth
Paul Flicek
Jan Friedman
Richard Gibbs
Marta Girdea
Robert Green
Ada Hamosh
Melissa Haendel
Ingrid Holm
Matt Hurles
Ben Hutton
Ekta Khurana
Sebastian Kohler
Joel Krier
Owen Lancaster
Melissa Landrum
Farrah Ladha
Paul Lasko
Rick Lifton
Daniel MacArthur
Alex MacKenzie
Danielle Azzariti
Aleksander Milosavljevic
Chris Mungall
Debbie Nickerson
Woong-Yan Park
Justin Paschall
Anthony Philippakis
Heidi Rehm
Peter Robinson
Gary Saunders
Francois Schiettecatte
Rolf Sijmons
Nara Sobreira
Jawahar Swaminathan
Morris Swertz
Peter Taschner
Sharon Terry
Rachel Thompson
Stephan Zuchner
MME: Patient facing efforts

Several teams are working to bring patients directly to MME

• GenomeConnect (ClinVar Consortium)
• PEER (Genetic Alliance)
• MyGene2 (U Washington)
• RareConnect (with our team)
MME: Patient facing challenges

- Different relationship with user
- Different goals
- Data accuracy is unproven
- Unreasonable expectations (high chance of false positives)
  - Should an intermediary look over results?
- Clinicians/Researchers may not want to use same system as patients
What should a patient-facing portal do?

- Connect with other families
- Find advocacy organizations
- Search for advice and information
- Discover specialists, clinical trials
Early prototype

Hi! I'm a PhD student at the University of Toronto and the Hospital for Sick Children. I focus on building software that helps doctors identify cohorts of similar rare disease patients. Now, we're trying to take what we've learned and the methods we've developed and apply them to help families more directly.

Signs and symptoms:

- Curly hair
- Blepharitis
- Conspicuously happy disposition

CORD holding rally in Vancouver to get Health Ministers to adopt a national plan for drugs for rare diseases. Jan 19, leaving the Fairmont Pacific Rim at 1:30pm.

It's a happy day. We've been hard at work and are happy to introduce two new features in response to the many comments we've received from patients, families, and advocates so far:
- a scientific contribution section (with 1 challenge so far)
- this discussion board
What should a Patient-Facing Site Do?

• Connect you to other patients?
• Forum/Discussion functionality?
• Connect you to your doctor(s)?
• Allow you to participate in research?
• Allow you to be contacted regarding relevant trials?
RareConnect/SickKids Collaboration

Development Team:
Orion Buske, Marta Girdea, Veronika Koltunova, Faye Feng

RareConnect Team:
Denis Costello, Laura Amorini, Robert Pleticha, Yann Le Cam