PhenoTips & PhenomeCentral: Deep Phenotyping and Data Sharing for Rare Disorders

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

IRDiRC Conference, 8 November 2014

Deep Phenotyping

- Describe the features of an individual
  - To enable a diagnosis (especially of a rare disease)
  - To distinguish between similar disorders
  - To enable genotype-phenotype correlations

Slide courtesy of Ada Hamosh, JHU

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 anamoly
- congenital anomly
- congential anomaly
- congentital anomaly
- cong. m.
- cong. Mal
- cong. malfor
- congenital malform
- congenital m.
- multiple congenital anomalies
- multiple congenital abormalities
- multiple congenital abnormalities

Examples of lists:
* dd. cong. malfor. behav. pro.
* dd. mental retardation
* df< delayed puberty
* df&lt
* dd df mr
* mental retar.short stature

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 anamoly
- congenital anomly
- congential anomaly
- congentital anomaly
- cong. m.
- cong. Mal
- cong. malfor
- congenital malform
- congenital m.
- multiple congenital anomalies
- multiple congenital abormalities
- multiple congenital abnormalities

Phenotypic description (Clinical symptoms)

Behavior, Cognition and Development
- ☐ Global development delay
- ☐ Fine motor delay ☐ Gross motor delay
- ☐ Language delay
- ☐ Learning disability
- ☐ Mental retardation
  - ☐ Mild
  - ☐ Moderate
  - ☐ Severe
- ☐ Attention deficit hyperactivity disorder
- ☐ Autism
- ☐ Pervasive developmental delay
- ☐ Psychiatric disorders (Specify below)
- ☐ Other: _____

Neurological
- ☐ Hypotonia
- ☐ Seizures
- ☐ Ataxia
- ☐ Dystonia
- ☐ Chorea

Cardiac
- ☐ ASD
- ☐ VSD
- ☐ AV canal defect
- ☐ Coarctation of aorta
- ☐ Tetralogy of fallot
- ☐ Other: _____

Craniofacial
- ☐ Craniosynostosis
- ☐ Cleft lip ☐ Cleft palate
- ☐ Microretrognathia ☐ Retrognathia
- ☐ Facial dysmorphism (Specify below)
- ☐ Other: _____

Eye Defects
- ☐ Blindness
- ☐ Coloboma
- ☐ Epicanthus
- ☐ Eyelid abnormality (Specify bellow)
- ☐ Other: _____

Problems with the status quo

- Phenotypic descriptions that are very evocative for humans, unreadable to a computer:
  - myopathic electromyography
  - still walking 25 years after onset

- The following descriptions mean the same to you: “generalized amyotrophy”, “generalized muscle atrophy”, “muscular atrophy, generalized”

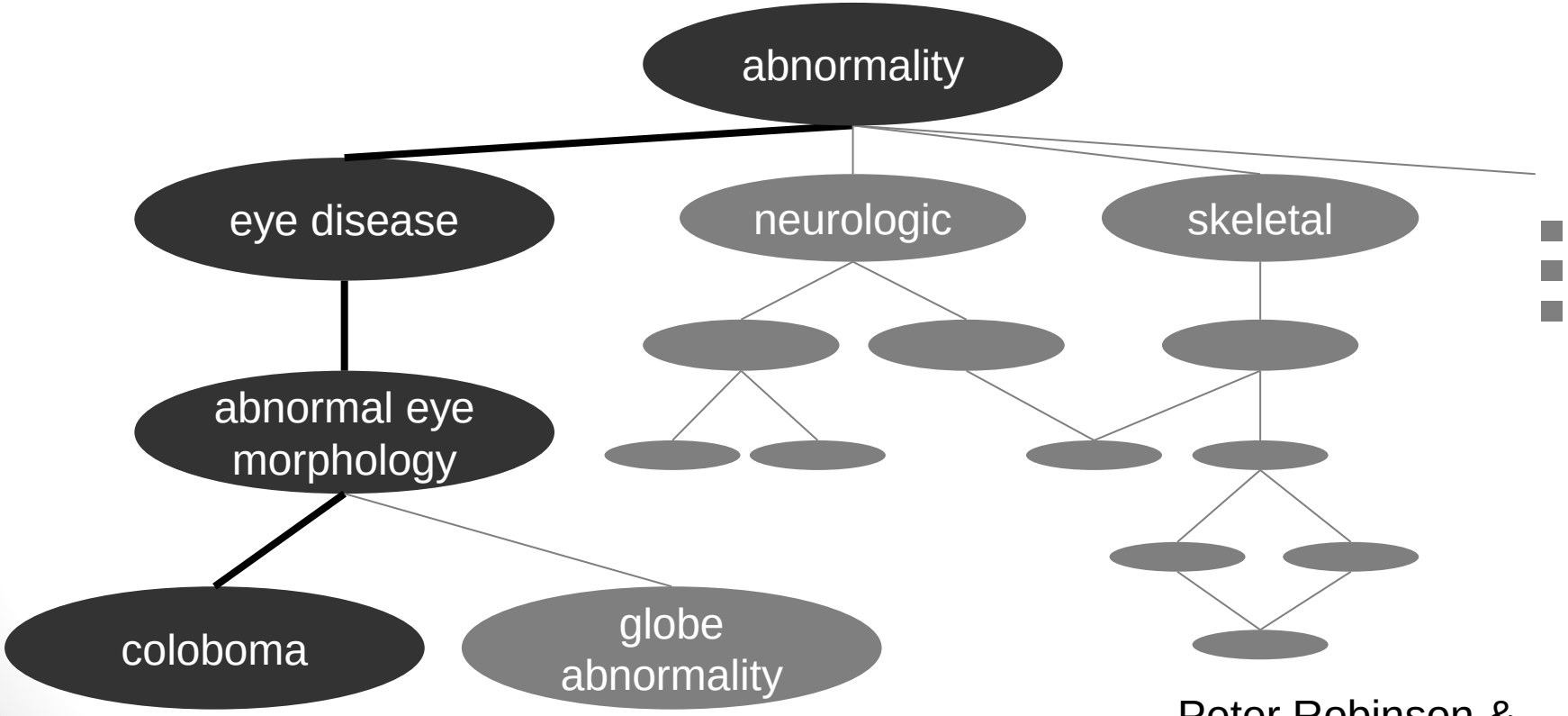
- It is impossible to define distances between phenotypes

- Databases don’t talk to one another about phenotypes

Next-generation **phenotyping**

Human Phenotype Ontology (HPO):

- 10,700+ terms
- 100,000+ links to 5,000+ OMIM/ORDO Disorders



Peter Robinson &
Monarch Consortium

phenotips.org

PhenoTips:
Deep Phenotyping Platform
for clinical and research use

Using ontologies patient-side

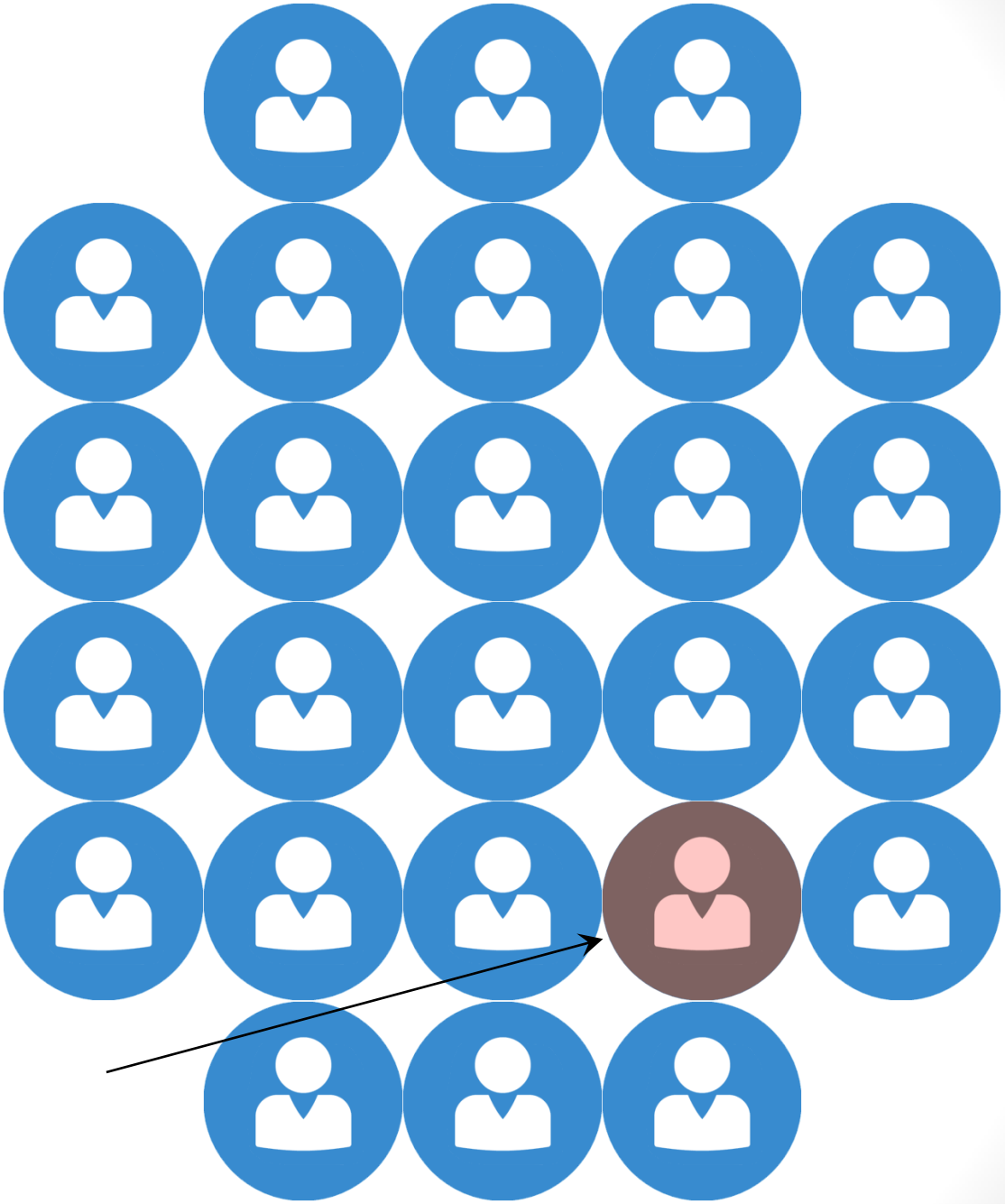
- Ontologies are large (HPO has > 10,000 terms) and difficult to use
- Re-mapping data to an ontology post-visit is time consuming and prone to error
- Best time to phenotype using ontologies is during the patient visit
- Goals of PhenoTips
  - Make deep phenotyping simple
  - Make it “faster than paper”

Demo

Advantages of ontologies & PhenoTips

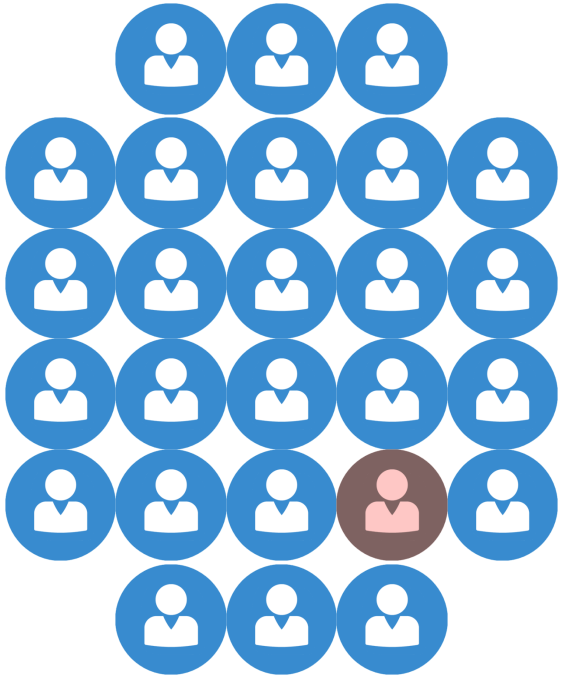
- Integration of data between studies, identification of unrelated patients
- Better and more thorough automated genome analysis and variant prioritization
- Phenotype patients at the bedside – more accurate, with less redundant data entry
- Training for next generation:
  - Diagnosis assistance
  - Identify previously seen similar patients
  - Decisions based on prior outcomes

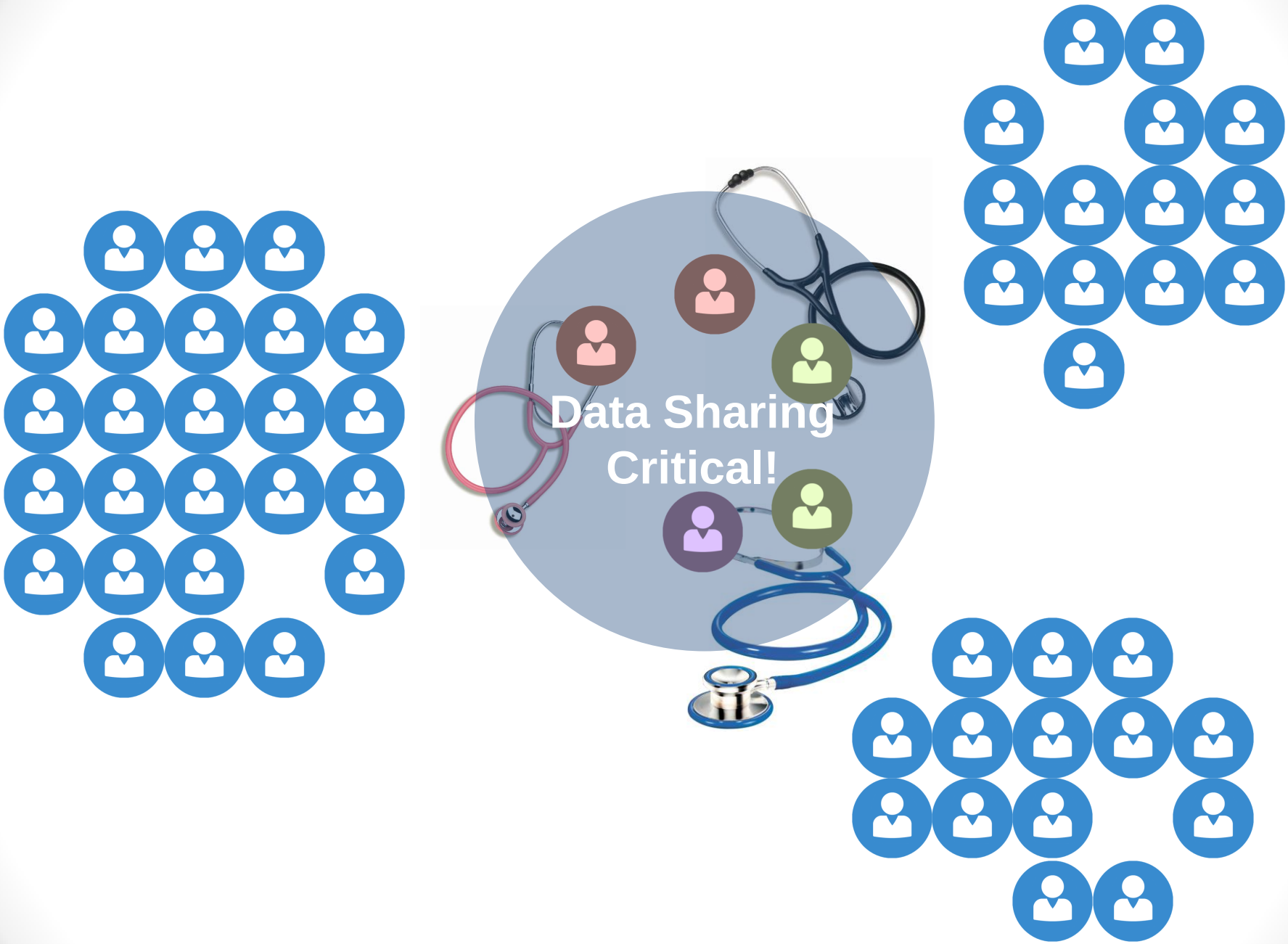
PhenomeCentral: an integrated portal for sharing phenotype and genotype data for rare genetic disorders



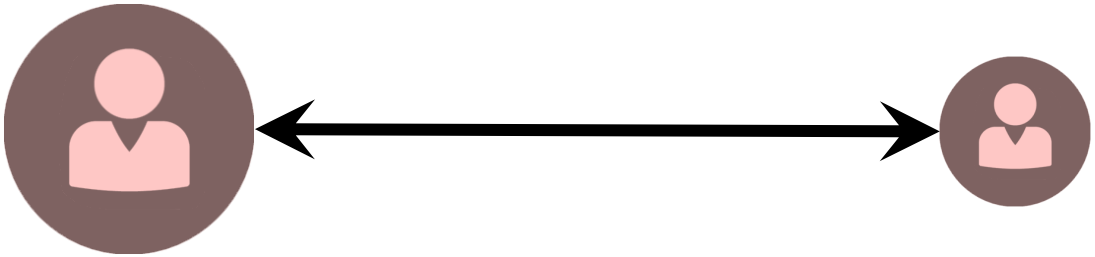
rare genetic disease

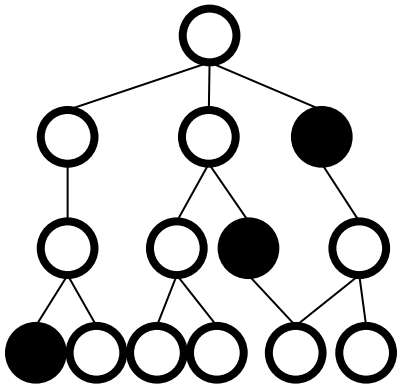
- might not recognize known disease
- insufficient sample size for novel gene

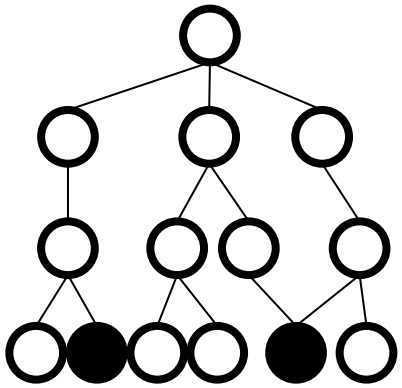



Data Sharing
Critical!

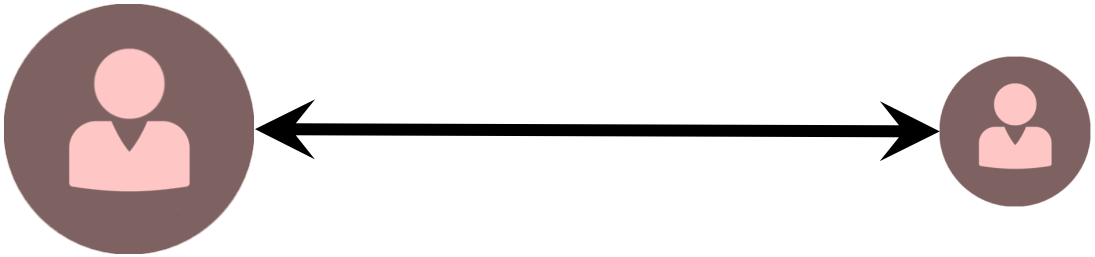
Finding similar patients

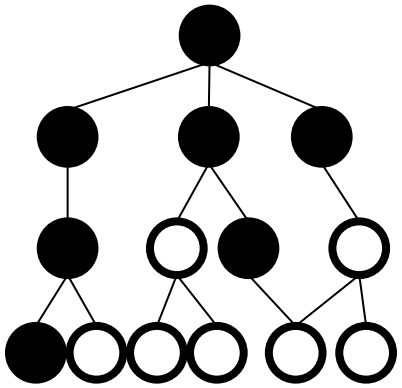


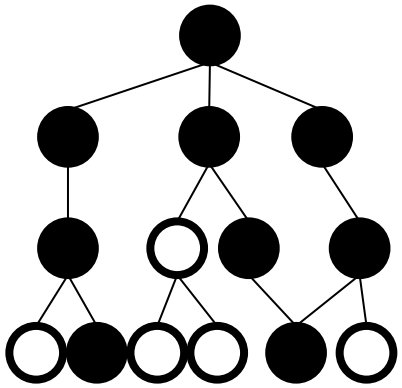




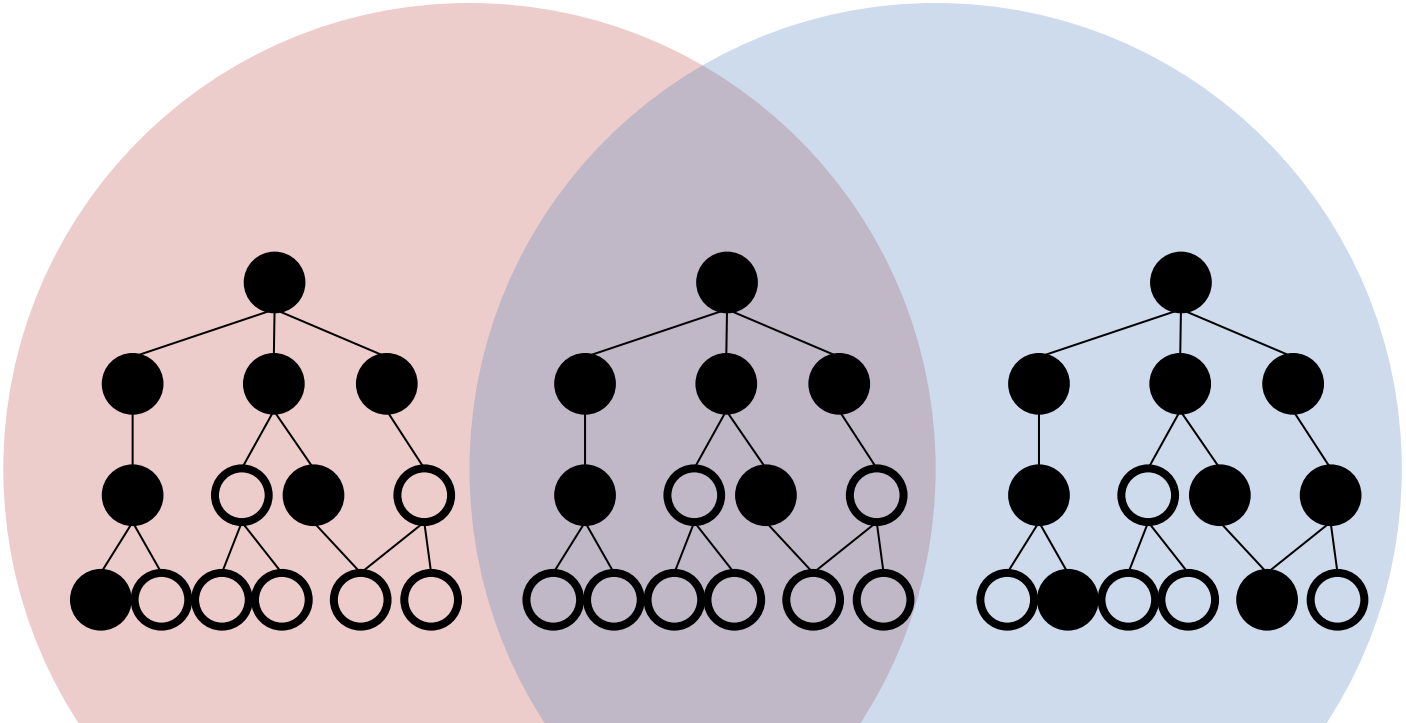
Finding similar patients



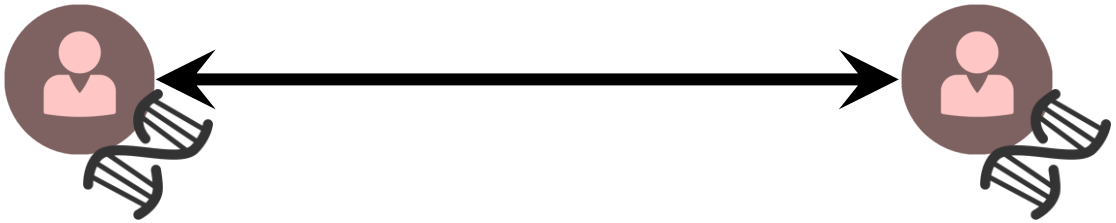




Finding similar patients

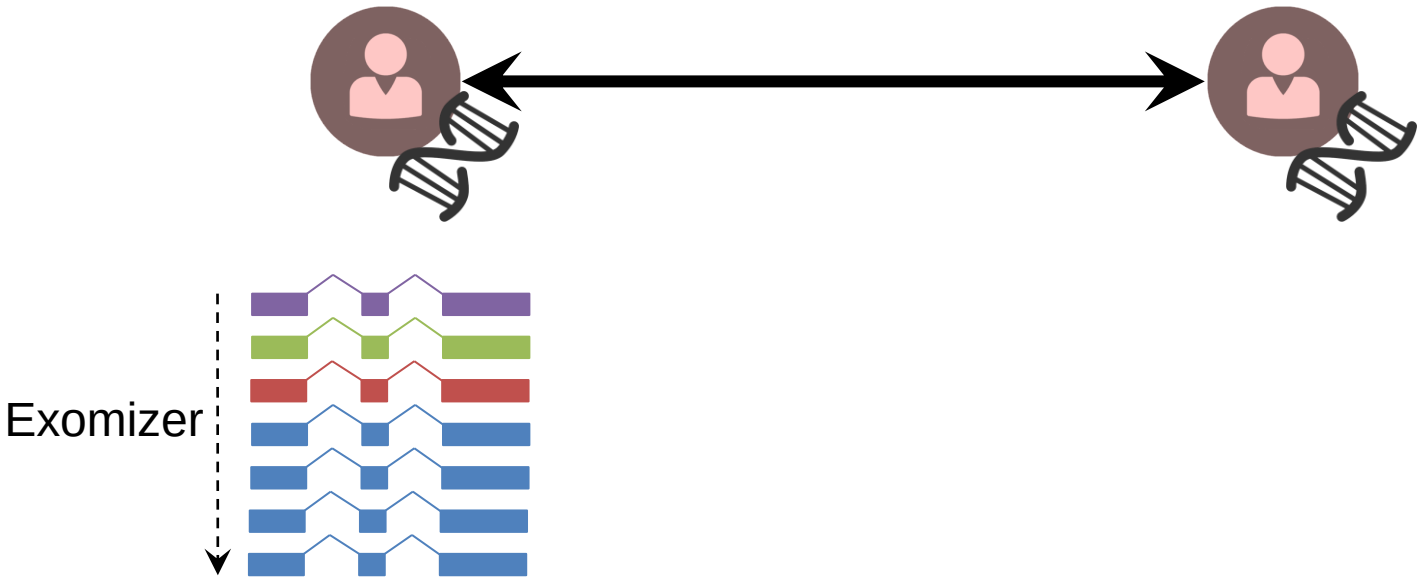


Incorporating gene data



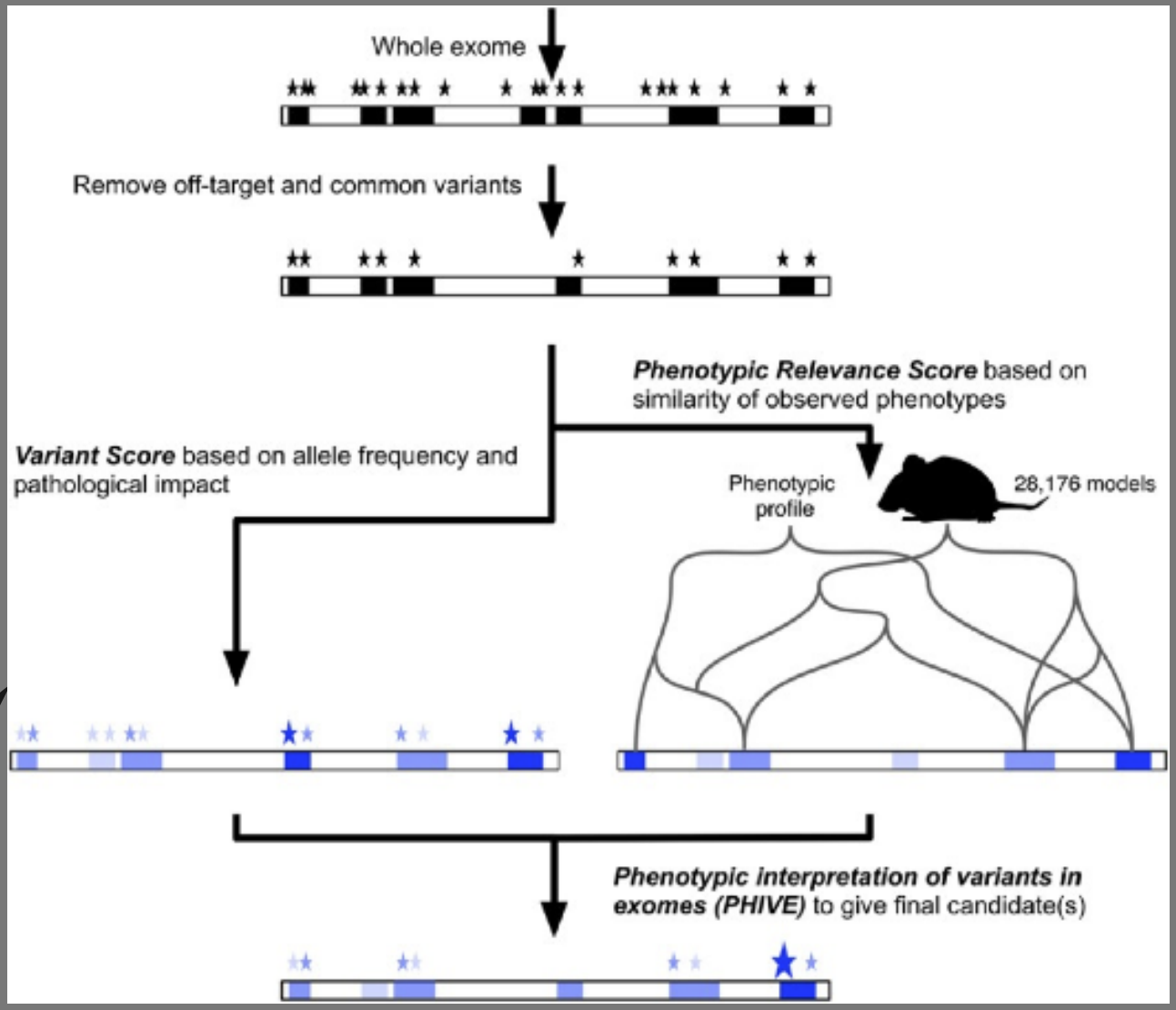
Incorporating gene data

Incorporating gene data



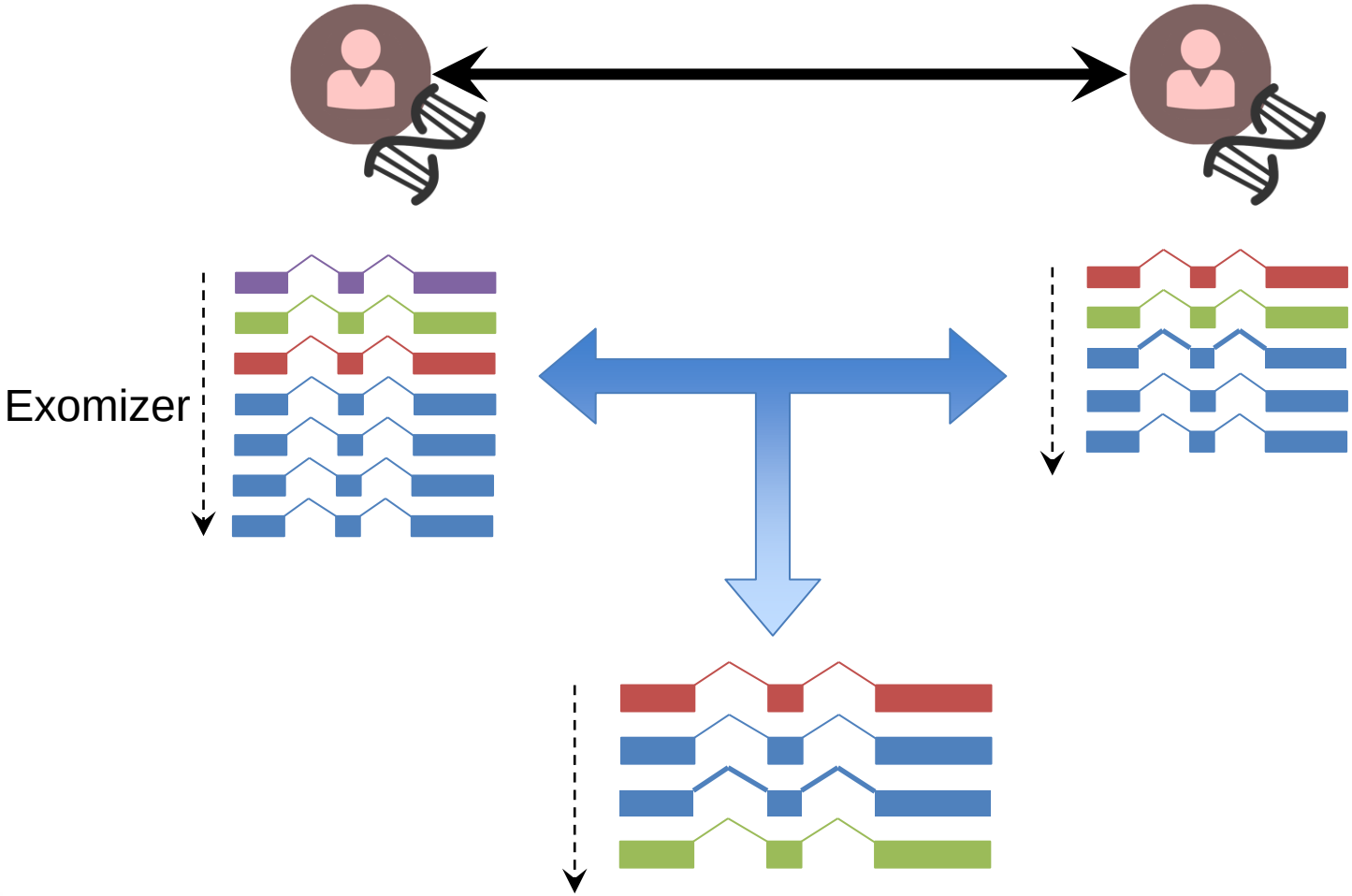
Incorporating gene data

Exomizer



(Robinson et al., 2014)

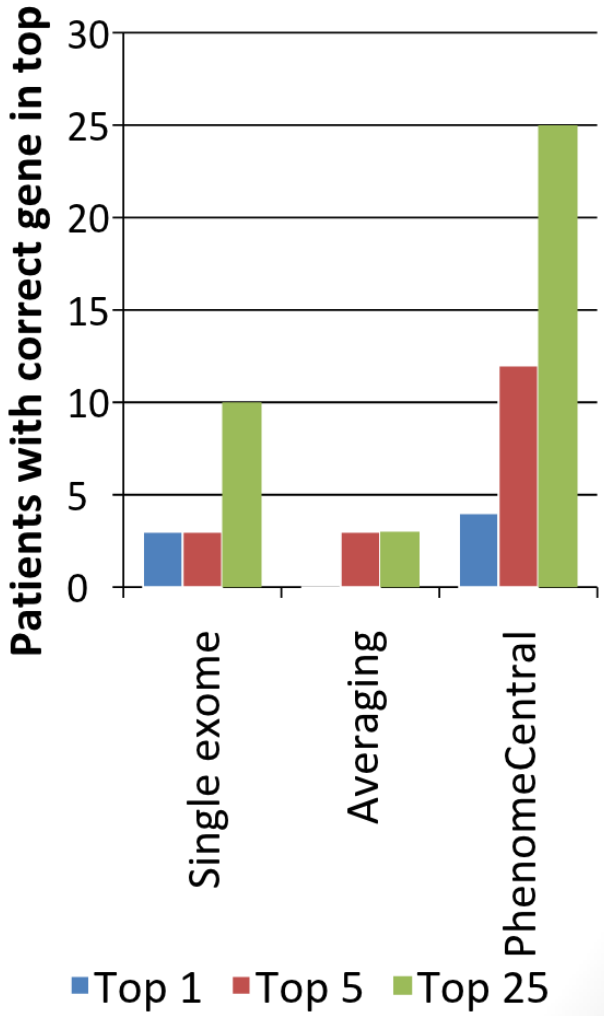
Incorporating gene data



Validation

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

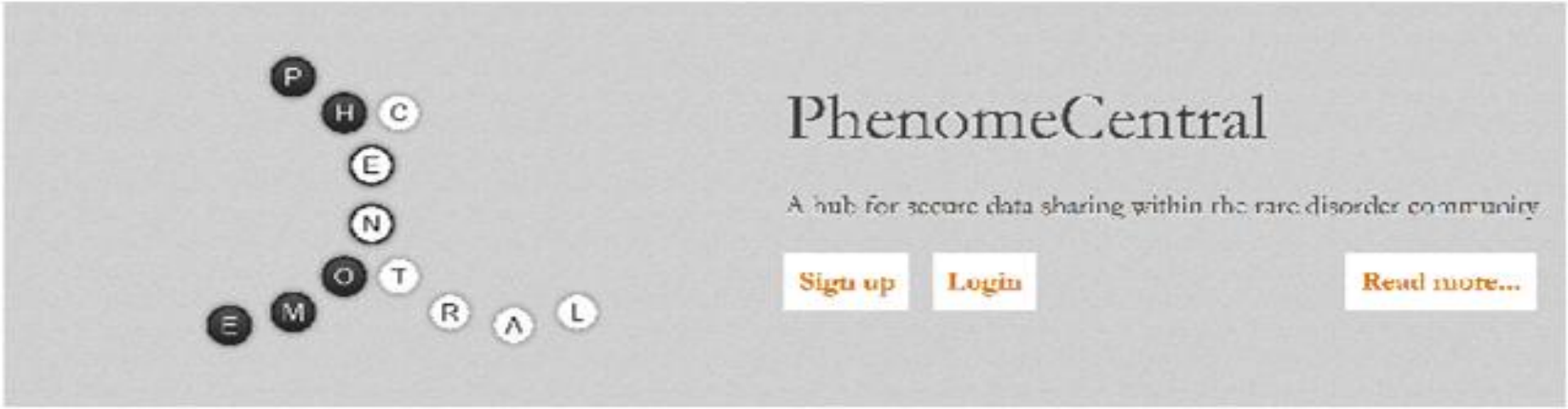
- Consistently Finds Similar Patients
  - **73%** top phenotype matches are in the same cohort
- Prioritizes solved or lead genes
  - **4x** known/lead genes ranked in top 5 compared to single exome



to encourage data sharing we built a

user-friendly, privacy-aware portal

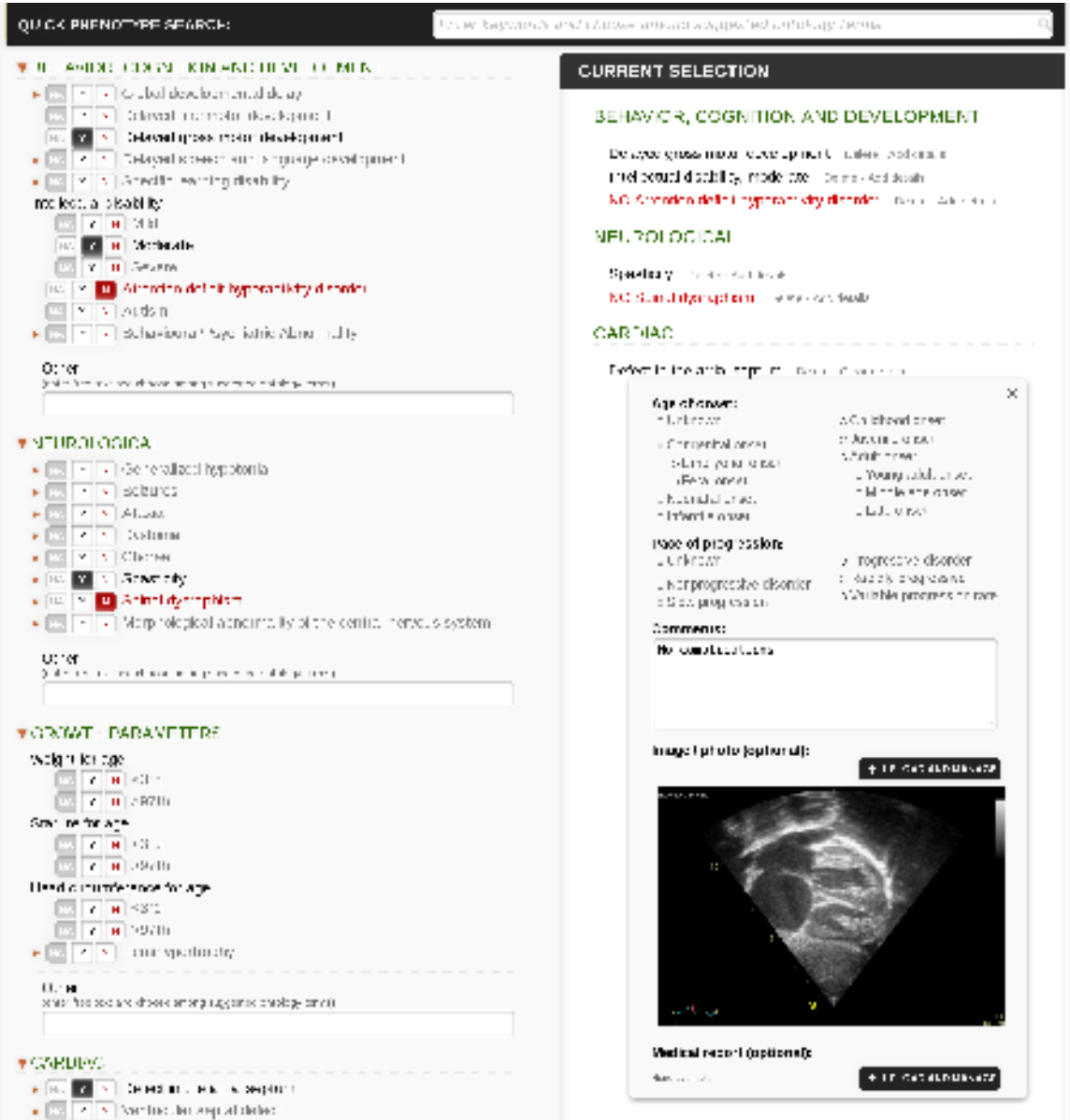
for discovering patients similar to yours



- 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

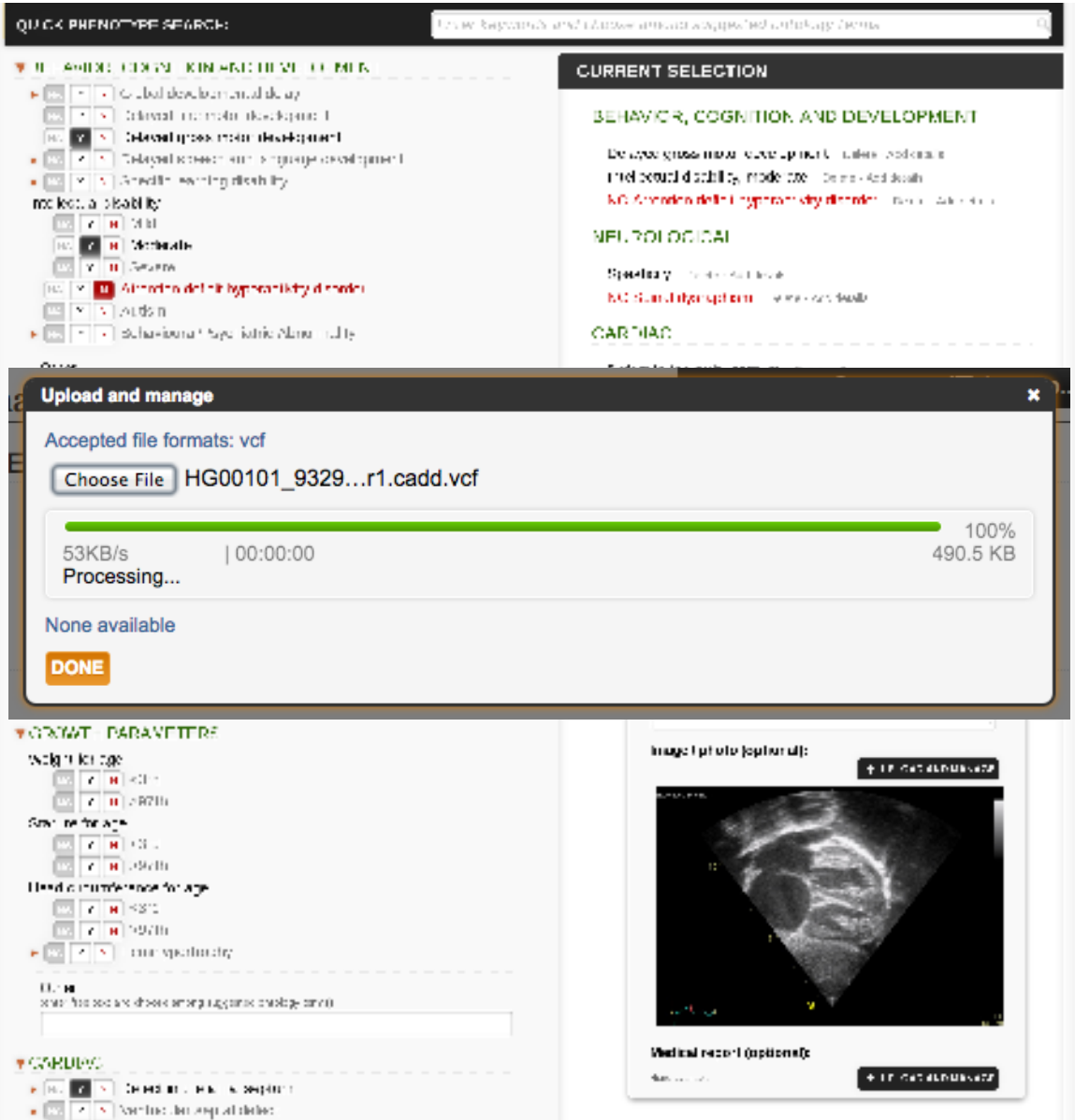
phenomecentral.org

Step 1: submit your patient



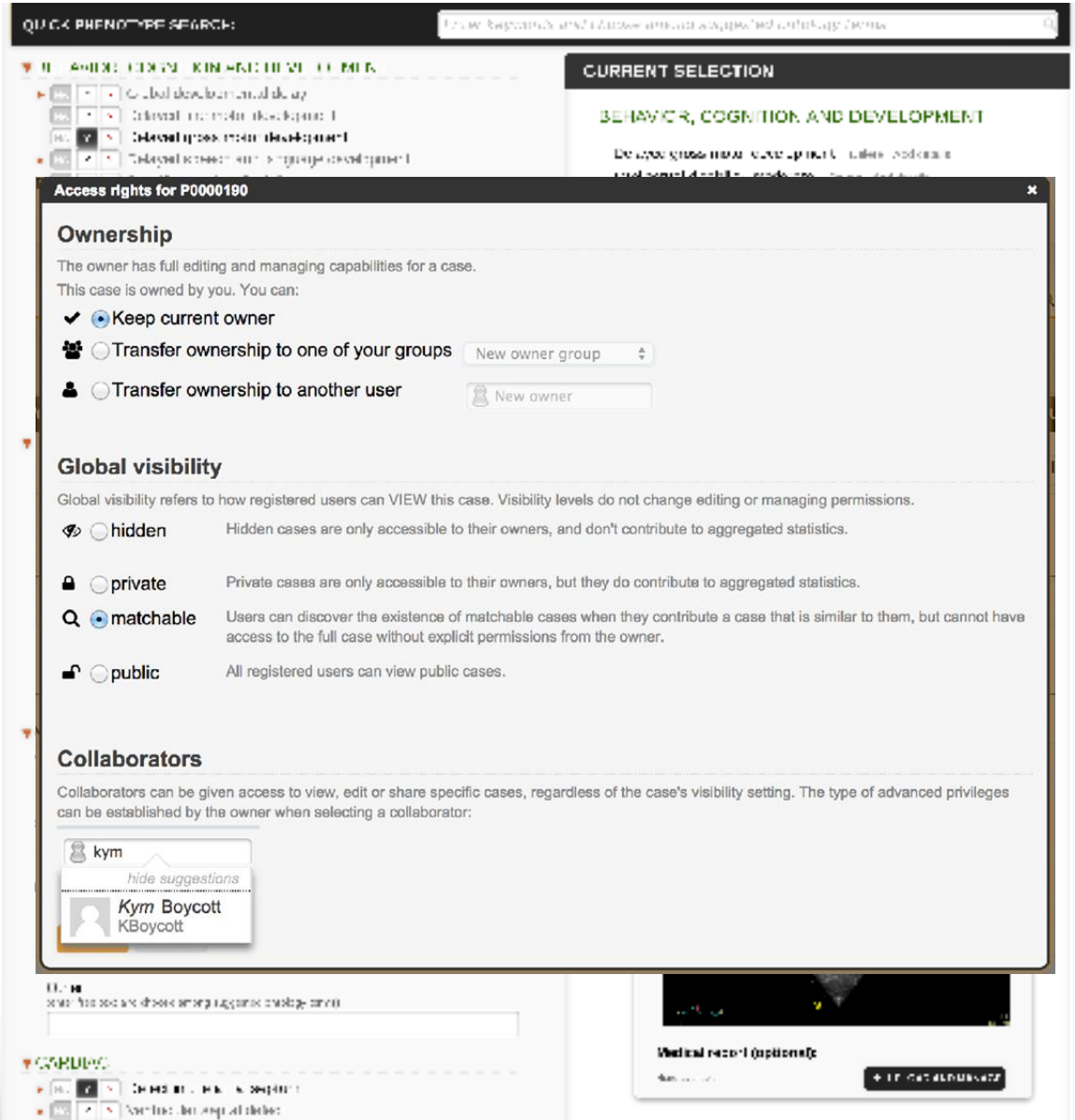
- Built-in PhenoTips interface
- Export anonymized records from other instances
- Add a VCF file and/or gene list
- Set permission and add collaborators

Step 1: submit your patient



- Built-in PhenoTips interface
- Export anonymized records from other instances
- Add a VCF file and/or gene list
- Set permission and add collaborators

Step 1: submit your patient



- Built-in PhenoTips interface
- Export anonymized records from other instances
- Add a VCF file and/or gene list
- Set permission and add collaborators

Step 2: see patients similar to yours

F0000010 Reported by **Marta Girdea (admin)** on 2013/09/29 18:10 · Last modified by **Marta Girdea** on 2013/09/30 14:00

This case is owned by Care4Rare, it is public and it is shared with 1 collaborator.

Patient information

Identifier:	KB_174_FHS1-1
Sex:	Female

Clinical symptoms and physical findings

CRANIOFACIAL
- Low hanging columella
- Thin upper lip vermilion
- Short philtrum
- Triangular face
- Wide nose
- Prominent nasal tip
- Narrow nasal bridge
- Long nose
- NO Wide mouth

EAR DEFECTS
- Low-set ears
- Recurrent otitis media

MUSCULOSKELETAL
- Broad fingertip
- Brachydactyly syndrome
- Broad thumb

GENITOURINARY
- Nephrocalcinosis
- Hydronephrosis

BEHAVIOR, COGNITION AND DEVELOPMENT
- Moderate expressive language delay

Diagnosis

OMIM disorder:	#136140 FLOATING-HARBOR SYNDROME; FLHS

Similar cases available in the database

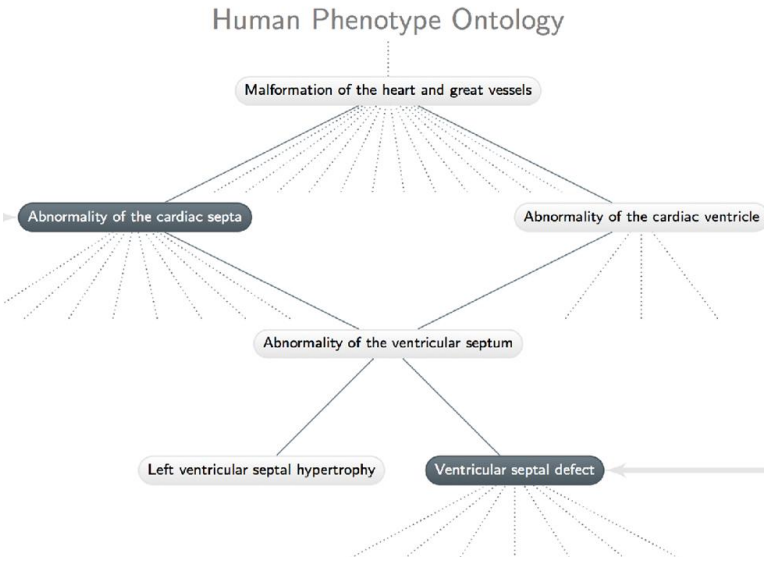
Showing 10 similar cases REFRESH

Case ID	Diagnosis	Relevance	Details	
F0000021	#136140 FLOATING-HARBOR SYNDROME; FLHS	■■■□	Matches found for 14 out of 17 features.	Show matches...
F0000019	#136140 FLOATING-HARBOR SYNDROME; FLHS	■■■□	Matches found for 14 out of 17 features.	Show matches...
F0000012	#136140 FLOATING-HARBOR SYNDROME; FLHS	■■■□	Matches found for 14 out of 17 features.	Show matches...
F0000009	#136140 FLOATING-HARBOR SYNDROME; FLHS	■■■□	Matches found for 14 out of 17 features.	Show matches...
F0000011	#136140 FLOATING-HARBOR SYNDROME; FLHS	■■■□	Matches found for 14 out of 17 features.	Show matches...
F0000020	#136140 FLOATING-HARBOR SYNDROME; FLHS	■■■□	Matches found for 14 out of 17 features.	Show matches...
F0000014	#136140 FLOATING-HARBOR SYNDROME; FLHS	■■■□	Matches found for 13 out of 17 features.	Show matches...
F0000017	#136140 FLOATING-HARBOR SYNDROME; FLHS	■■■□	Matches found for 13 out of 17 features.	Show matches...
F0000016	#136140 FLOATING-HARBOR SYNDROME; FLHS	■■■□	Matches found for 11 out of 17 features.	Show matches...
F0000015	#136140 FLOATING-HARBOR SYNDROME; FLHS	■■■□	Matches found for 14 out of 17 features.	Show matches...

Abnormality of the cardiac septa
- Infantile onset
- Slow progression

↔

Ventricular septal defect
- Congenital onset
- Nonprogressive disorder



Step 2: see patients similar to yours

F0000010

Reported by **Marta Girdea (admin)** on 2013/09/29 18:10 · Last modified by **Marta Girdea** on 2013/09/30 14:00

This case is owned by Care4Rare, it is public and it is shared with 1 collaborator.

Patient information

Identifier:	KB_174_FHS1-1
Sex:	Female

Clinical symptoms and physical findings

CRANIOFACIAL
- Low hanging
- Thin upper li
- Short philtrun
- Triangular fa
- Wide nose
- Prominent na
- Narrow nasal
- Long nose
- NO Wide mou

EAR DEFECTS
- Low-set ears
- Recurrent oti

MUSCULOSKELI
- Broad fingerti
- Brachydactyl
- Broad thumb

GENITOURINAR
- Nephrocalcin
- Hydronephro

BEHAVIOR, COG
- Moderate exp

Diagnosis

OMIM disorder:

Similar cas

Showing 10 simila

Case ID	Di
F0000021	#1 SY
F0000019	#1 SY
F0000012	#1 SY
F0000009	#1 SY
F0000011	#1 SY
F0000020	#1 SY
F0000014	#1 SY
F0000017	#1 SY
F0000016	#136140 FLOATING-HARBOR SYNDROME; FLHS — Matches found for 11 out of 17 features. Show matches...
F0000015	#136140 FLOATING-HARBOR SYNDROME; FLHS — Matches found for 14 out of 17 features. Show matches...

PHENOTYPIC FEATURES BREAKDOWN

ABNORMALITY OF THE METOPIC SUTURE — 75%

The current patient (P0000081) presented with:
- Prominent metopic ridge

The matched patient presented with:
- *1 undisclosed feature*

ABNORMALITY OF PERIAURICULAR REGION — 66%

The current patient (P0000081) presented with:
- Preauricular skin tag

The matched patient presented with:
- *1 undisclosed feature*

ABNORMALITY OF THE PINNA — 62%

The current patient (P0000081) presented with:
- Microtia

The matched patient presented with:
- *1 undisclosed feature*

CLEFT PALATE — 47%

The current patient (P0000081) presented with:
- Submucous cleft hard palate
- Median cleft palate

The matched patient presented with:
- *1 undisclosed feature*

ABNORMALITY OF THE SKELETAL SYSTEM — 13%

The current patient (P0000081) presented with:
- Micrognathia
- Delayed skeletal maturation
- Microcephaly

The matched patient presented with:
- *2 undisclosed features*

UNMATCHED

The current patient (P0000081) presented with:
- Hypoplastic toenails
- Delayed CNS myelination
- Defect in the atrial septum
- Downslanted palpebral fissures
- Seizures
- Global developmental delay
- Malar flattening
- Stenosis of the external auditory canal
- Conductive hearing impairment
- Choanal atresia
- Cleft eyelid

The matched patient presented with:
- -

GENE MATCHING BREAKDOWN

EDN1 — HIDE VARIANTS...

Estimated relevance for the observed phenotype in the current patient (P0000081): 95%

Estimated relevance for the observed phenotype in the matched patient: 95%

VARIANT	ESTIMATED HARMFULNESS
chr6: 12290906 - 12290915 GCCAAGGAGC → - (FS_DELETION)	95%

VARIANT	ESTIMATED HARMFULNESS
Undisclosed position	95%

STAG2 — SHOW VARIANTS...

Estimated relevance for the observed phenotype in the current patient (P0000081): 77%

Estimated relevance for the observed phenotype in the matched patient: 77%

ABCA1 — SHOW VARIANTS...

Estimated relevance for the observed phenotype in the current patient (P0000081): 73%

Estimated relevance for the observed phenotype in the matched patient: 50%

ANKRD20A4 — SHOW VARIANTS...

Estimated relevance for the observed phenotype in the current patient (P0000081): 34%

Estimated relevance for the observed phenotype in the matched patient: 34%

ANKRD20A3 — SHOW VARIANTS...

Estimated relevance for the observed phenotype in the current patient (P0000081): 13%

Estimated relevance for the observed phenotype in the matched patient: 13%

Step 3: contact the other submitter

F0000024 | PhenomeCentral

Taking Screenshots in Mac OS X...

phenomecentral.org/data/F0000024

take screenshot mac

EAR DEFECTS

Deafness

- Sensorineural

BEHAVIOR, COGNITION AN

Delayed fine motor develop

Delayed gross motor devel

Intellectual disability

- Mild

Attention deficit hyperactivit

NEUROLOGICAL

Generalized hypotonia

Absent Achilles reflex

Reduced tendon reflexes

Sensory neuropathy

Autonomic dysregulation

OTHER

Neonatal hypotonia

Similar cases availabl

Showing 10 similar cases

Case ID	Diagnosis
P0000144	
Undisclosed identifier	Undisclosed diagnosis
Undisclosed identifier	Undisclosed diagnosis
Undisclosed identifier	Undisclosed diagnosis
Undisclosed identifier	Undisclosed diagnosis
Undisclosed identifier	Undisclosed diagnosis
Undisclosed identifier	Undisclosed diagnosis
Undisclosed identifier	Undisclosed diagnosis
Undisclosed identifier	Undisclosed diagnosis
Undisclosed identifier	Undisclosed diagnosis

REFRESH

Show matches...

Contact a non-public case owner

1 Configure your message

SUBJECT

Interested in one of your non-public cases

Information about you:

- [x] DISCLOSE YOUR NAME
- [x] DISCLOSE YOUR EMAIL
- [x] DISCLOSE YOUR MEMBERSHIP TO PHENOMECENTRAL GROUPS

Information about your case (F0000024):

- [x] INCLUDE DIAGNOSIS INFORMATION
- [x] INCLUDE A PHENOTYPE SUMMARY

Your requests:

- [x] REQUEST MUTUAL VIEW ACCESS TO THE TWO SIMILAR CASES

If the recipient accepts, they gain view access to your case and you gain view access to theirs.

- [ ] REQUEST CONTACT INFORMATION

OTHER INFORMATION TO INCLUDE IN YOUR MESSAGE

2 Preview your message

This is the message the other user will receive:

SUBJECT

[PhenomeCentral] Interested in one of your non-public cases

MESSAGE

Hello *<undisclosed recipient name>*,

A PhenomeCentral user is interested in one of your non-public cases: *<undisclosed case identifier>*. Please see their message below.

PhenomeCentral has identified significant similarities between one of your cases and one of mine.

My patient is undiagnosed presents the following phenotypic features:

- Absent Achilles reflex
- Anosmia
- Attention deficit hyperactivity disorder
- Autonomic dysregulation
- Decreased corneal reflex
- Delayed fine motor development
- Delayed gross motor development
- Generalized hypotonia
- Intellectual disability, mild
- Neonatal hypotonia
- Reduced tendon reflexes
- Sensorineural hearing impairment
- Sensory neuropathy

I would like to grant you the rights to view my case and to obtain view access to your case, and to learn your contact information in order to further discuss these abnormalities with you.

Regards,
Marta Girdea
marta@phenotips.org

To accept view privileges from this user **and** to grant them view access to *<undisclosed case identifier>*, follow this link: *<undisclosed URL>*.

Best wishes,
The PhenomeCentral team

SEND CANCEL

PhenomeCentral

phenomecentral.org

PHENOTYPIC FEATURES BREAKDOWN

FEEDING DIFFICULTIES IN INFANCY 55%

The current patient (P0000089) presented with: Gastrostomy tube feeding in infancy

The matched patient (P0000079) presented with: Gastrostomy tube feeding in infancy

ABNORMALITY OF THE THUMB 47%

The current patient (P0000089) presented with: Triphalangeal thumb

The matched patient (P0000079) presented with: Proximal placement of thumb

GLOBAL DEVELOPMENTAL DELAY 46%

The current patient (P0000089) presented with: Global developmental delay

The matched patient (P0000079) presented with: Global developmental delay

MICROGNATHIA 45%

The current patient (P0000089) presented with: Micrognathia

The matched patient (P0000079) presented with: Micrognathia

MICROCEPHALY 44%

The current patient (P0000089) presented with: Microcephaly

The matched patient (P0000079) presented with: Microcephaly

ABNORMALITY OF THE OUTER EAR 35%

The current patient (P0000089) presented with: Low-set ears

The matched patient (P0000079) presented with:
Microtia
Preauricular skin tag
Atresia of the external auditory canal

ABNORMALITY OF THE EYE 22%

The current patient (P0000089) presented with: Strabismus

The matched patient (P0000079) presented with: Myopia

UNMATCHED

The current patient (P0000089) presented with:
Syndactyly
Subglottic stenosis
Vocal cord paralysis
Patent ductus arteriosus

The matched patient (P0000079) presented with:
Aplasia/Hypoplasia of the middle ear
Malar flattening
Conductive hearing impairment

GENE MATCHING BREAKDOWN

EFTUD2 HIDE VARIANTS...

Estimated relevance for the observed phenotype in the current patient (P0000089): 44%

VARIANT	ESTIMATED HARMFULNESS
chr17: 42931653 - 42931653 - → G (FS_INSERTION)	95%

Estimated relevance for the observed phenotype in the matched patient (P0000079): 44%

VARIANT	ESTIMATED HARMFULNESS
chr17: 42929131 - 42929131 G → A (STOPGAIN)	95%

SBK1 SHOW VARIANTS...

Estimated relevance for the observed phenotype in the current patient (P0000089): 6%

Estimated relevance for the observed phenotype in the matched patient (P0000079): 6%

MFDM patient (EFTUD2 mutation) matched to a known one despite atypical presentation

includes data from:

CARE forRARE

RD Connect

NIH Undiagnosed Diseases Program

PhenomeCentral

phenomecentral.org

PHENOTYPIC FEATURES BREAKDOWN

PROXIMAL MUSCLE WEAKNESS 59%

The current patient (P0000131) presented with: Proximal muscle weakness

The matched patient (P0000371) presented with: Proximal muscle weakness

MYOPATHY 55%

The current patient (P0000131) presented with: Myopathy

The matched patient (P0000371) presented with: EMG: myopathic abnormalities

THROMBOCYTOPENIA 55%

The current patient (P0000131) presented with: Thrombocytopenia

The matched patient (P0000371) presented with: Thrombocytopenia

ABNORMALITY OF THE MUSCULATURE 27%

The current patient (P0000131) presented with: Exercise-induced muscle fatigue

The matched patient (P0000371) presented with: Frequent falls

ABNORMALITY OF THE NERVOUS SYSTEM 15%

The current patient (P0000131) presented with:
- Areflexia
- Delayed gross motor development
- Waddling gait

The matched patient (P0000371) presented with:
- Abnormality of peripheral nerve conduction

UNMATCHED

The current patient (P0000131) presented with:
- Petechiae
- Bruising susceptibility
- Impaired platelet aggregation
- Limited hip extension
- Epistaxis
- Decreased hip abduction

The matched patient (P0000371) presented with:
- Hepatic steatosis
- Elevated serum creatine phosphokinase

GENE MATCHING BREAKDOWN

STIM1 HIDE VARIANTS...

Estimated relevance for the observed phenotype in the current patient (P0000131): 1%

VARIANT	ESTIMATED HARMFULNESS
chr11: 4045175 - 4045175 A → T (NONSYNONYMOUS)	80%

Estimated relevance for the observed phenotype in the matched patient (P0000371): 1%

VARIANT	ESTIMATED HARMFULNESS
chr11: 4045175 - 4045175 A → T (NONSYNONYMOUS)	80%

Two similar patients with STIM1 mutations matched despite inconsistent terminology

includes data from:

CARE forRARE

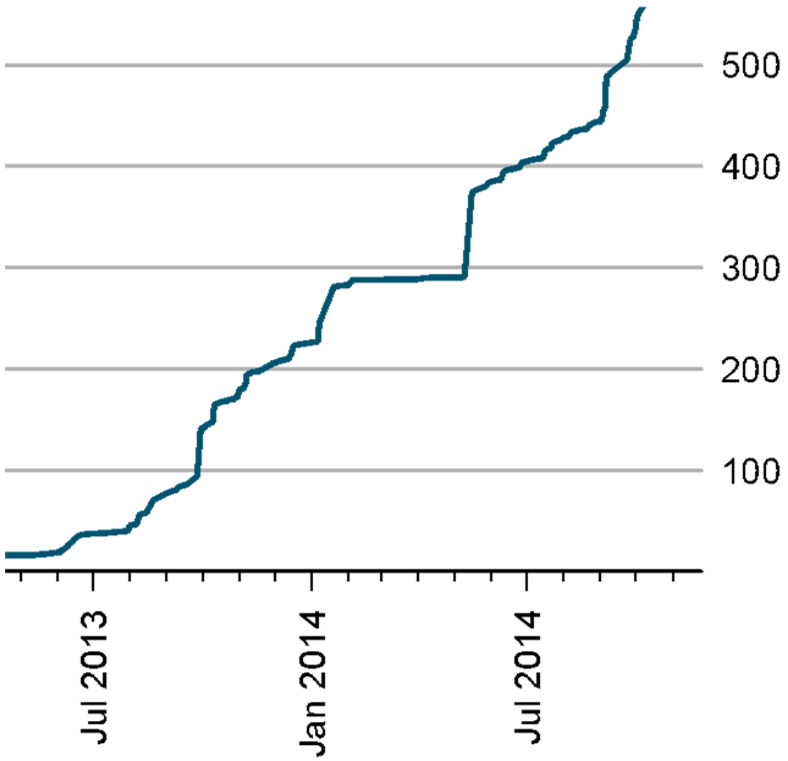
RD Connect

NIH Undiagnosed Diseases Program

PhenomeCentral

phenomecentral.org

- **557** cases
  - deeply phenotyped
  - most with exome data
  - most undiagnosed
- **272** users



includes data from:

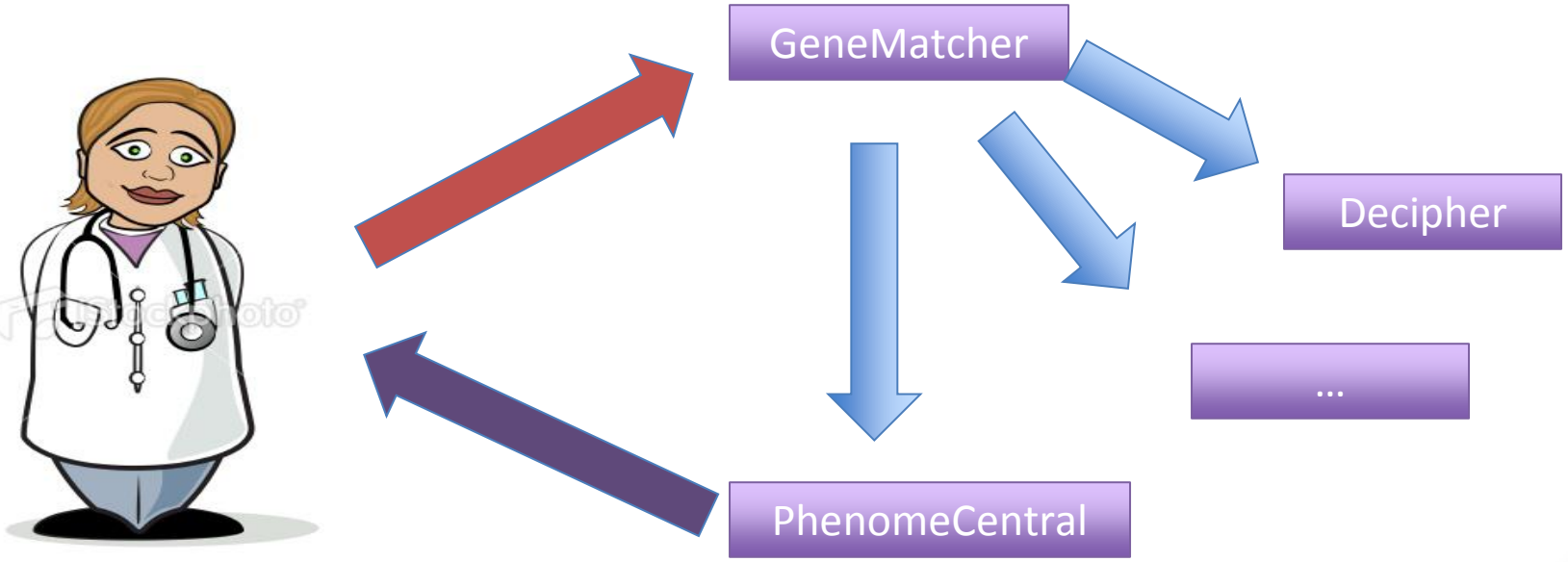
CARE forRARE

RD Connect

NIH Undiagnosed Diseases Program

So Many Matchers, So Little Time

- Rare disorder data is collected at many databases (GeneMatcher, Decipher, GEM.app, PhenomeCentral)
- Developed a query "API" for a patient profile to be entered once, searched everywhere



Acknowledgments

PhenoTips/ PhenomeCentral Development Team:

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

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, CIHR, NSERC

MatchMaker Exchange API

Send a description of a patient

- Each site decides how to process the query
- Result can be sent back to requesting site or directly to the user
- Adopted by the GA4GH

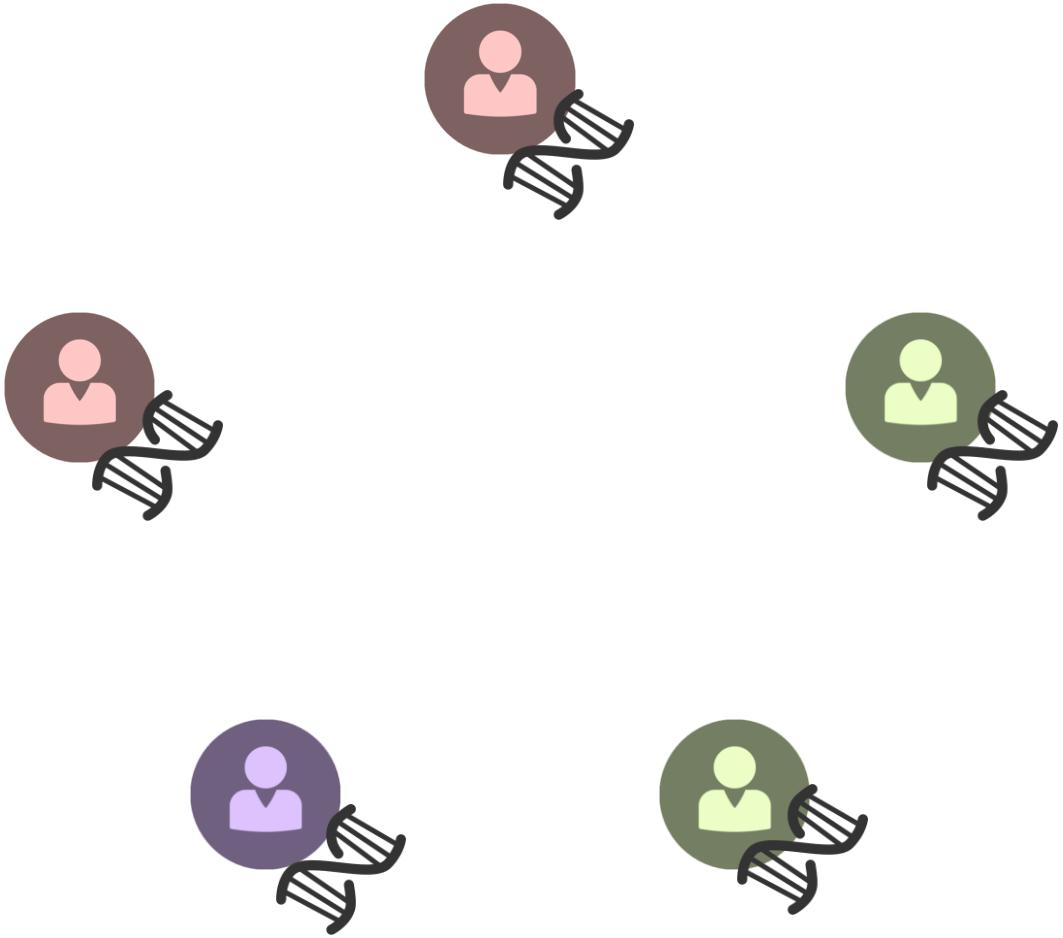
Questions/Challenges:

- How best to run queries
- Allow more precise queries?

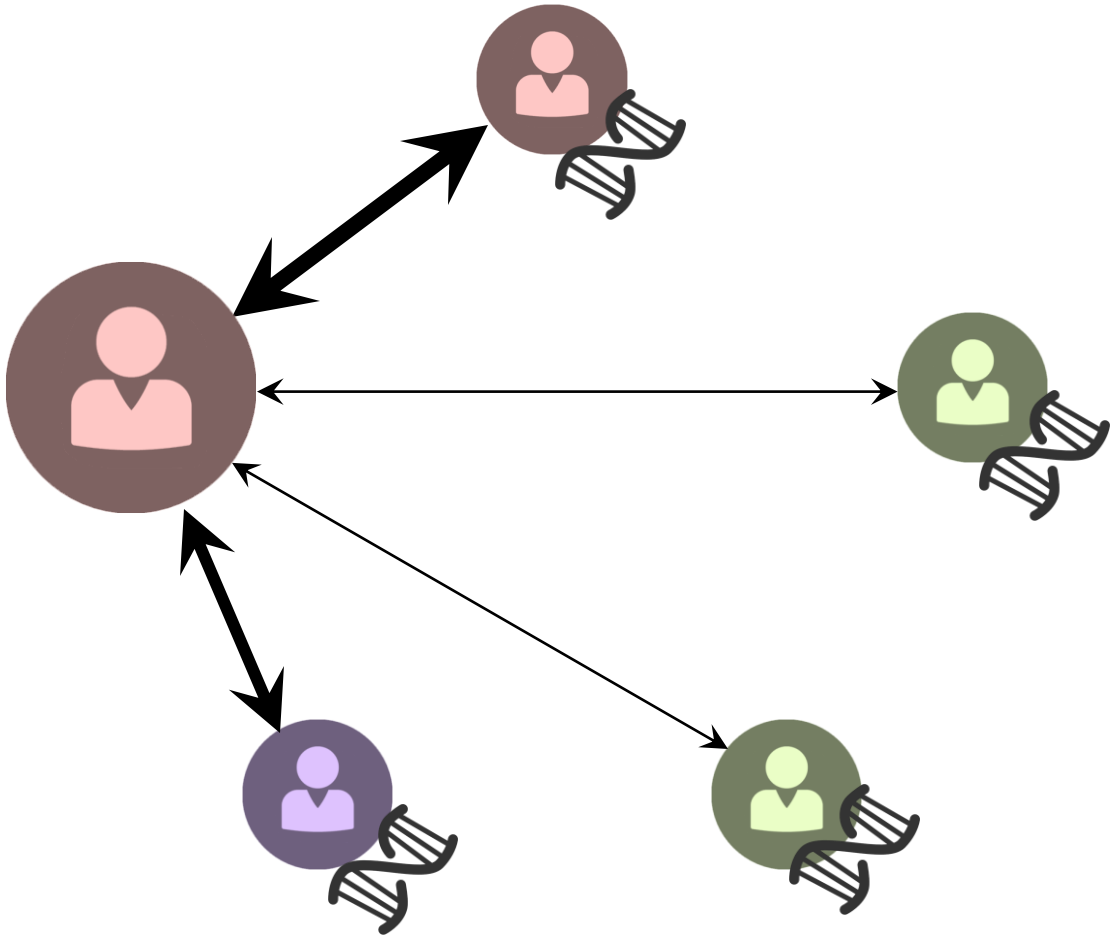
```
"gender" : "M"|"F",
"age_of_onset" : <number of years>|<number of months>|<age code>,
"mode_of_inheritance" : <inheritance code>,

"disorders" : [
  "MIM:######",
  "ORPHA#####",
  …
],
"features" : [
  {
    "id" : <ICHPT code>,
    "observed" : "yes"|"no"|"unknown",
    "age_of_onset" : "…"
  },
  …
],
"genes" : [
  {
    "gene" : <gene name>|<ensembl gene ID>|<entrez gene ID>,
    "chr" : "1"|"2"|…|"X"|"Y",
    "start_position" : <number>,
    "end_position" : <number>,
    "ref" : "-"|"A"|"ACG"|…,
    "alt" : "-"|"A"|"ACG"|…,
    "zygosity" : <number>,
    "type" : <mutation type>,
    "assembly" : "hg##"
  },
```

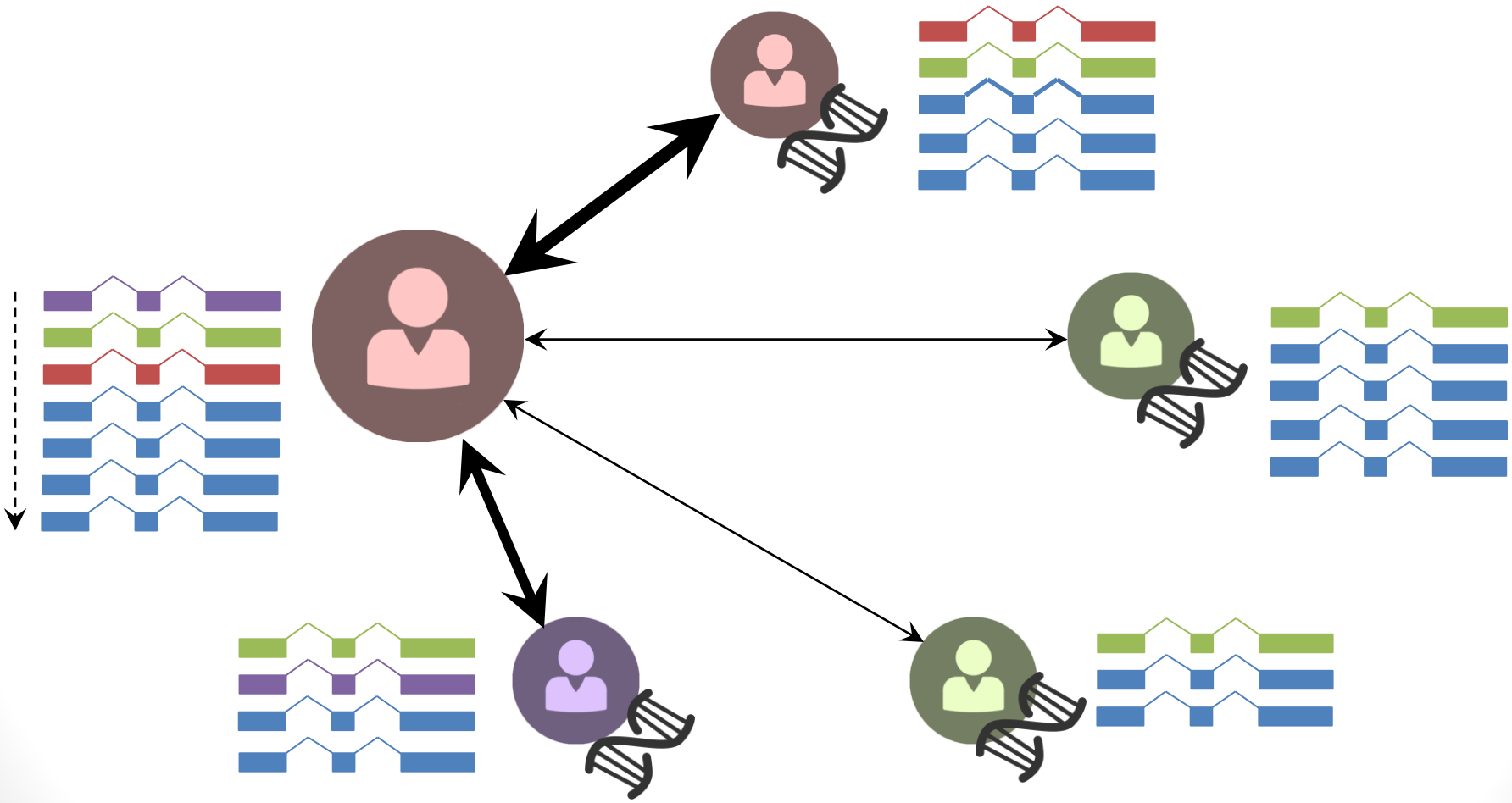
Identifying candidate genes



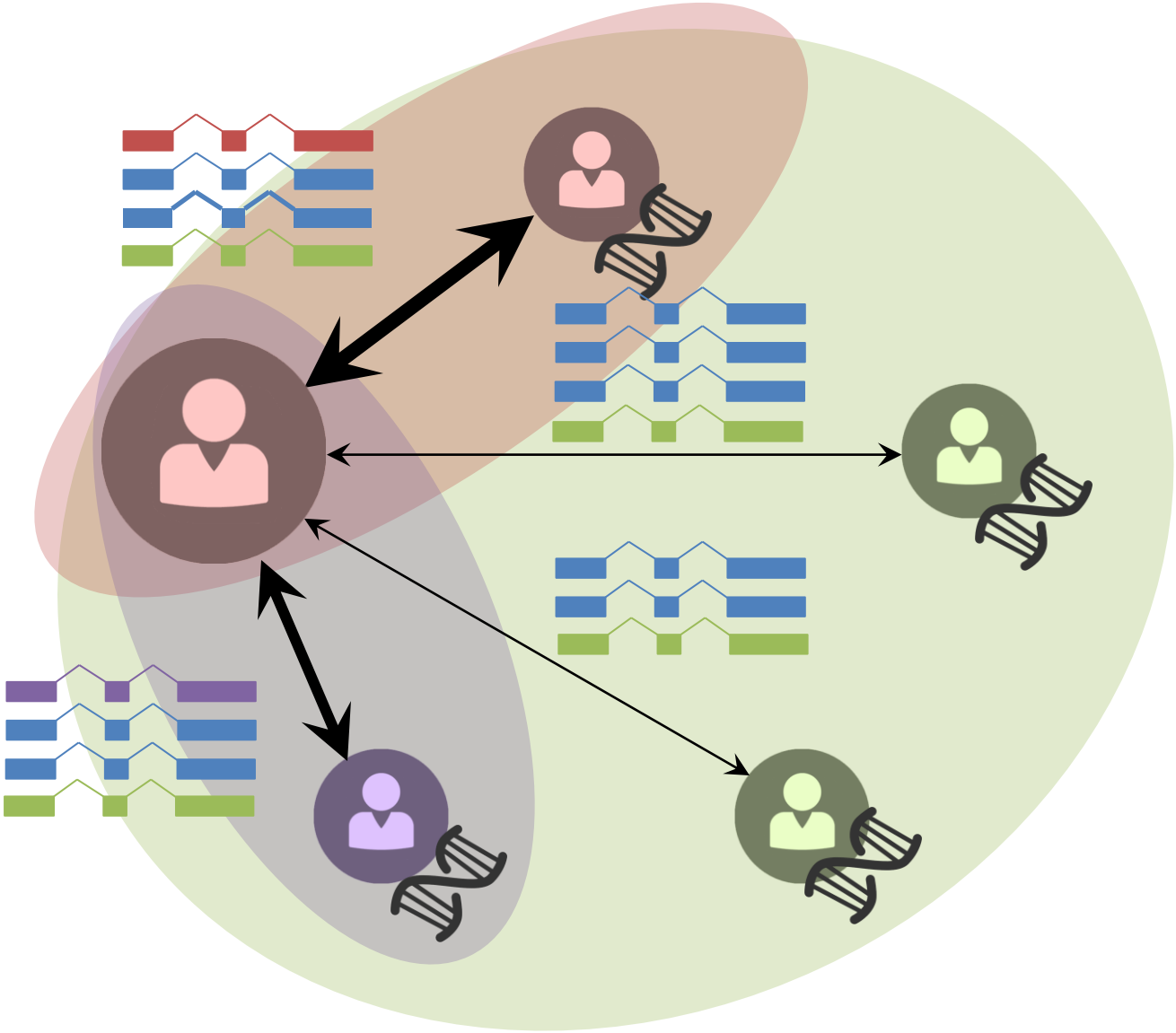
Identifying candidate genes



Identifying candidate genes



Identifying candidate genes



Finding similar patients

p_{term} from ORDO corpus

$IC(term) = \log(1/p_{term})$

$LS(term) \approx IC(term) - \max_{parents} IC(parent)$