Identification of genotype-phenotype associations in PhelanMcDermid Synd. using patient-sourced data

Paul Avillach, MD, PhD
Assistant Professor of Biomedical Informatics and Pediatrics
Harvard Medical School - Boston Children’s Hospital
Assistant Professor of Epidemiology
Harvard T.H. Chan School of Public Health
Toward Precision Medicine: Building a Knowledge Network for Biomedical Research and a New Taxonomy of Disease
Report from National academy of science, USA, 2011
Harvard IRB security levels:
Level 5 - Extremely sensitive information
Level 4 – Very sensitive information
Level 3 – Sensitive, or Confidential information
Level 2 - Benign information to be held confidentially
Level 1 - Non-confidential research information
Play with our API  http://bd2k-picsure.hms.harvard.edu
Phelan McDermid Syndrome

- Extremely rare genetic disease: \(~1100\) diagnosed patients worldwide today
- Autistic traits, intellectual deficiency, slight dysmorphic features
- Also called deletion 22q13 syndrome
- Caused by deletions of the terminus of chromosome 22
Heterogeneity of the genetic alterations

1 to 140 affected
Heterogeneity of the phenotypes

All organs can be affected:
- Neuro-developmental
- Facial dysmorphic features
- GERD (Gastro-Esophagal Reflux)
- Renal problems
- Lax joints
- Dysplastic toenails
- Congenital cardiac diseases
- ...

Sources:
globalgenes.org
autismspeaks.org
sfari.org
Deep phenotyping

Knowledge from Clinical Notes
## Box 1 | Natural language processing

<table>
<thead>
<tr>
<th>Step</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Boundary detection</td>
<td>Fx of obesity but no fx of coronary artery diseases.</td>
</tr>
<tr>
<td>Tokenization</td>
<td>Fx of obesity but no fx of coronary artery diseases.</td>
</tr>
<tr>
<td>Normalization</td>
<td>disease_</td>
</tr>
<tr>
<td>Part-of-speech tagging</td>
<td>NN IN NN CC DT NN IN JJ NN NNS</td>
</tr>
<tr>
<td>Shallow parsing</td>
<td>NP PP NP NN</td>
</tr>
<tr>
<td>Entity recognition</td>
<td>Obesity, Disease or disorder, UMLS ID: C0028754, Status: family history, Negated: no</td>
</tr>
<tr>
<td></td>
<td>Coronary artery disease, Disease or disorder, UMLS ID: C0010054, Status: family history, Negated: yes</td>
</tr>
<tr>
<td></td>
<td>Coronary artery, Anatomy, UMLS ID:C0205042</td>
</tr>
</tbody>
</table>

BEFORE validation

AFTER validation

Pop-up validation window
Sept 2016
590 patients enrolled (re consented)
390 patient Reported Outcomes
358 patients with Genetic data

https://pmsdn.hms.harvard.edu

Registration process with review for level 1
IRB for level 2

Need to be an Autism investigator
Objective:
Identify phenotypes linked with the deletion of other genes than SHANK3

Chr deletion size
Phenotypes
Results — Kidney malformations associated with other genes that SHANK3
Results — Gradient of gross motor delays showing a cumulative effect

Age of acquisition

- Early
- Late

Chromosomal coordinates

- Loss
- Mutation

- Walk unassisted
- Sit when placed
- Roll over back to stomach
- Crawl on hands and knees
- Hold head up on his/her own
- Climb stairs standing up without help
- Jump with both feet

P-values:
- P = 3.6e-07:
- P = 2.6e-06:
- P = 5.9e-05:
- P = 5.9e-05:
- P = 0.002:
- P = 0.04:
- P = 0.45:
• **Play with i2b2/tranSMART UI:**
  - https://grdr.hms.harvard.edu
  - https://nhanes.hms.harvard.edu
  - https://demo-ngs.hms.harvard.edu
  - https://pmsdn.hms.harvard.edu  *(ASD investigators only)*

• **Play with BD2K PIC-SURE RESTful API**
  - http://bd2k-picsure.hms.harvard.edu
  - http://exac.hms.harvard.edu
Creating an information commons for biomedical data centered on the patient.

PIC-SURE combines genetic, environmental, imaging, behavioral, and clinical data on individual patients from multiple sources into integrated sets.

www.pic-sure.org
PCORI
PMS_DN team

- Megan O’Boyle, PI
- Liz Horn, PhD, Co-PI, Network Director

PMSF Research director
- Geraldine Bliss, MSc

Project Manager
- Andria Cornell Mann

LGC Data Network Specialist
- Rebecca Davis

Family Engagement Specialist
- Jackie Malasky
Manager of Data Infrastructure
• Michael McDuffie, MSc

Developers
• Jeremy Easton-Marks
• Gabor Korodi
• Thomas DeSain
• Sean Finan
• Ranjay Kumar
• Alexander Nikitin
• Ken Hoflen

Graduate students
• Ombeline Dorval, MD
• Maxime Wack, MD
• Claire Hassen-Kodja, MD, MSc
• Emmanuelle Sylvestre, MD, MSc
• Yuri Ahuja, MD, PhD HMS candidate

Funding:

Project Manager
Cassandra Perry, MS, CGC

Research Associates - Postdocs
• Antoine Tran, MD, MSc
• Laurie Tran, MD, MSc
• Cartik Saravana, PhD
• Li Ly, PhD
• Joany Zachariasse, MD, MSc
• Antoine Neuraz, MD, MSc

Previous members
• Samuel Finlayson, MD, PhD HMS candidate
• Ephi Sachs, MD
• Pei Chen
• Sushma Hanawal

We are hiring now:
• Senior Software Developer *2
• Postdocs *5