Exploring Existing Disease-to-Clinical Feature Annotations for Case-Level Phenotypic Descriptions in Genomic Databases

J. Krier[1], A. Henrie[2], A. Roberts[3], H. Rehm1,[4], D. Maglott[5], M. Benson[5], W. Rubinstein[5], K. Eilback[2], D. Miller[3]

1) Brigham and Women's Hospital, Harvard Medical School; 2) Department of Biomedical Informatics, University of Utah School of Medicine; 3) Boston Children's Hospital, Harvard Medical School; 4) Laboratory for Molecular Medicine, Partners Healthcare; 5) National Center for Biotechnology Information, National Institutes of Health

Background

- Phenotypic data is underrepresented in human genomic databases.
- The ClinGen Project's Phenotype Working Group aims to to optimize the collection and display of clinical phenotype data to enhance the value of variant data in the ClinVar database.
- A primary challenge for including phenotypic data in genomic databases is identifying standard concepts and terms to be annotated for cases of a given disorder
- Existing disease-to-phenotypic feature annotation sets, such as the OMIM "Clinical Synopsis" terms mapped to Human Phenotype Ontology (HPO) terms, could be used to support consistent phenotypic term collection.
- However, such syndrome-to-phenotypic term associations were not designed for collecting case-level phenotypic data and their utility for this function has not been explored.

Approach

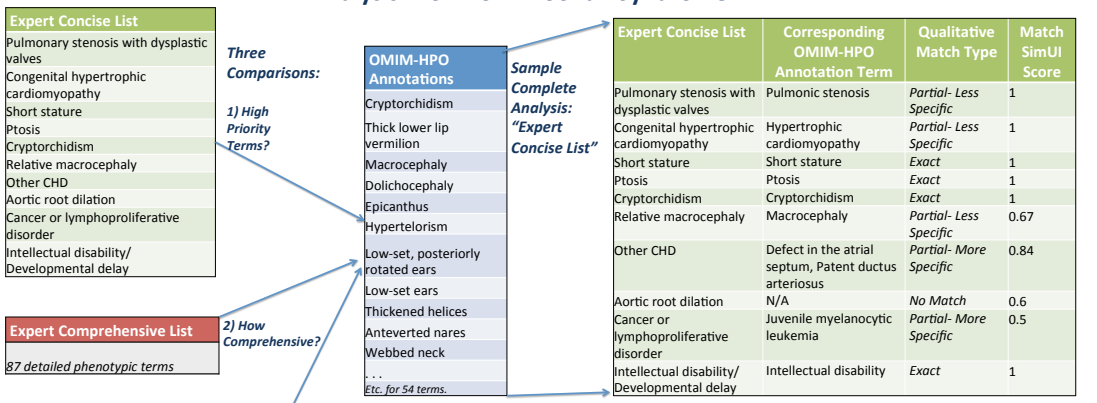
- We compared "Clinical Synopsis" terms from OMIM to expert derived phenotypic term lists for two Mendelian diseases: Noonan Syndrome and Hypertrophic Cardiomyopathy.
- We used OMIM-HPO annotations as reported by MedGen (http://www.ncbi.nih.gov/medgen/).
- We compared the existing OMIM-HPO annotations to:
  (1) A "concise" expert-generated high-yield phenotypic term list.
  (2) A more comprehensive expert-generated phenotypic term list.
  (3) Phenotypic term lists derived from clinical laboratory requisition forms.
- We used both qualitative/conceptual and quantitative semantic similarity (SimUI) term comparisons.*
- SimUI quantifies the similarity in meanings of two phenotypic terms or lists of terms by utilizing knowledge about meaning and relatedness represented in the graphical structure of the HPO.**

*Gentleman R, "Visualizing and Distances Using GO", 2005

** Kohler et al: "The Human Phenotype Ontology Project: ", 2014

Acknowledgments. This research was supported by NIH Grant U41 HG006834-01A1, U01 HG006500, and the Intramural Research Program of the NIH, National Library of Medicine

Analysis Workflow: Noonan Syndrome



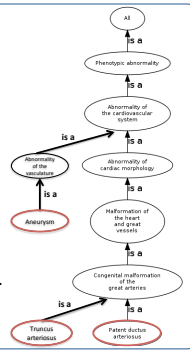
Expert Concise List	Corresponding OMIM-HPO Annotation Term	Qualitative Match Type	Match SimUI Score
Pulmonary stenosis with dysplastic valves	Pulmonic stenosis	*Partial- Less Specific*	1
Congenital hypertrophic cardiomyopathy	Hypertrophic cardiomyopathy	*Partial- Less Specific*	1
Short stature	Short stature	*Exact*	1
Ptosis	Ptosis	*Exact*	1
Cryptorchidism	Cryptorchidism	*Exact*	1
Relative macrocephaly	Macrocephaly	*Partial- Less Specific*	0.67
Other CHD	Defect in the atrial septum, Patent ductus arteriosus	*Partial- More Specific*	0.84
Aortic root dilation	N/A	*No Match*	0.6
Cancer or lymphoproliferative disorder	Juvenile myelanocytic leukemia	*Partial- More Specific*	0.5
Intellectual disability/ Developmental delay	Intellectual disability	*Exact*	1

SimUI

- For each term in the query list, an algorithm identifies the most similar term in the other list.
- Each term is defined by itself and all of its ancestors (more general terms) on the ontology.
- SimUI quantifies similarity (range 0-1) between two terms by determining the proportion of ancestors in common.

Example: "Patent ductus arteriosus" compared to "Truncus arteriosus" and "Aneurysm"

- "Truncus arteriosus" has a higher proportion of ancestors in common with "Patent ductus arteriosus" than does "Aneurysm".
- "Truncus arteriosus-Patent ductus arteriosus" therefore has a higher SimUI score than "Aneurysm-Patent ductus arteriosus".



Results

- OMIM-HPO annotations included exact or partial matches for 9/10 terms from the Noonan syndrome "Expert concise list" and 10/12 terms from the Familial Hypertrophic Cardiomyopathy "Expert concise list".
- Based on the SimUI measure, OMIM-HPO annotations included 5 exact, 3 close, and 2 distant matches for the Noonan syndrome "Expert concise list" and 6 exact, 5 close, and 1 distant matches for the Familial Hypertrophic Cardiomyopathy "Expert concise list".
- OMIM-HPO annotations included exact matches for 22/82 and partial matches for 20/82 terms from the "Expert detailed list" per clinical qualitative analysis. 22 exact, 28 close, and 32 distant or absent matches were identified for the OMIM-HPO annotations were "Expert detailed list" of 82 Noonan syndrome clinical features.

Conclusions

- Existing disease-to-phenotypic feature annotations provide good coverage of key clinical features identified by disease experts for Noonan Syndrome and Familial Hypertrophic Cardiomyopathy, but are not comprehensive.
- An expert-led curative process to develop syndrome-specific high-yield phenotypic term lists could reduce the variability of clinical phenotype data collected by clinical molecular laboratories and lead to improved phenotypic data content in genomic databases such as ClinVar.