

# PhenoStacks:

## Cross-Sectional Cohort Phenotype Comparison Visualizations

Michael Glueck, Alina Gvozdkik, Fanny Chevalier, Azam Khan, Michael Brudno, Daniel Wigdor

IEEE TRANSACTIONS ON VISUALIZATION AND COMPUTER GRAPHICS VOL. 23, NO. 1, JANUARY 2017

Hooman Shariati

1

Or better yet:  
Why you must use the task  
abstracts that we learned in  
class

2

## What's a phenotype?

- A characteristic of an individual compared to the rest of its population
- Can be anything
  - Morphological, biochemical, behavioural
- Ex: the way a particular bird builds its nest
  - I.e. how it looks and functions

3

## Disney Male Phenotypes:



Source: Disney

4

## Problem

- Phenotype datasets don't match in:
  - Structure
  - Coverage
  - Granularity

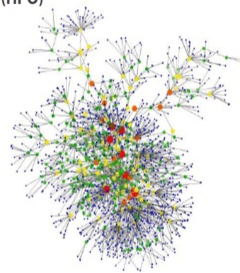
5

## Human Phenotype Ontology (HPO)

**Standardized Terminology**  
11,000+ phenotype terms

**Hierarchical Structure**  
Abstraction of relationships  
Linked synonyms  
Multiple inheritance

**Supports Computation**  
Information content

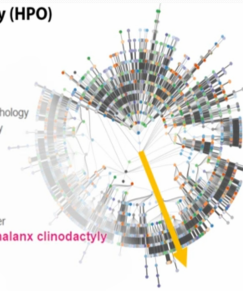


www.human-phenotype-ontology.org  
Köhler et al. (2014). Nucleic Acids Research, 42, 966-974.

6

## Human Phenotype Ontology (HPO)

- Phenotypic abnormality
- Abnormality of the skeletal system
- Abnormality of skeletal morphology
- Abnormal appendicular skeleton morphology
- Abnormality of limb bone morphology
- Abnormality of the upper limb
- Abnormality of the hand
- Abnormality of finger
- Deviation of finger
- Deviation of the 5th finger
- Clinodactyly of the 5th finger
- Fifth finger distal phalanx clinodactyly



www.human-phenotype-ontology.org

7

## First attempt at visualizing HPO

- By the same authors (Jan 2016)
  - interviews with 2 clinicians identified 4 major domain tasks
- No Task Abstraction
- Chose to: support the comparison of phenotypes between a new, undiagnosed query patient and a set of diagnosed reference patients
- Result: Algorithm for comparing an individual to a cohort

8

## Demo:

9

## What: Data

(C1, P1, H1, present)  
(C1, P1, H2, absent)  
(C1, P1, H3, present)  
(C1, P2, H2, present)  
(C1, P2, H3, absent)  
...

**Phenotype Data**  
(Cohort, Patient, HPO, Observ.)

Source: Paper



**HPO Topology**

10

## What: Derived

- Union of all recorded phenotypes is a cohort graph



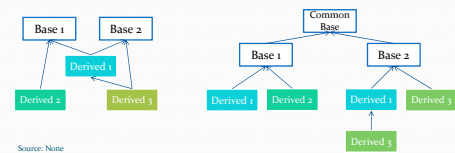
**Cohort C1**  
(P1, present)  
(P1, absent)  
(P2, present)

**Cohort Subgraph**

Source: Paper

11

## What wasn't shown in their video:



Source: None

12

## What wasn't said in their paper:

- "feedback from researchers indicated duplication of phenotypes was extremely confusing because it gave the false impression that there were more clusters."
- Oops

13

## Second attempt at visualizing HPO

- Interviewed 6 experts
- Identified 9 domain tasks
- Synthesized them to common visualization tasks using Brehmer & Muntzner's multi-level typology
- Result: algorithm for comparing multiple cohorts
- Basically added a second view to their previous tool

14

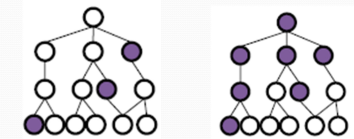
## Why: Tasks

- **Explore** → **Summarize**
  - emergent patterns (W1), disease prevalence (W5), disease characterization (W6), compare patterns across disease subtypes (B1)
- **Locate** → **Identify**
  - outlier patients/phenotypes (W2), subcohort discovery (W3), inform clinical practice (B3)
- **Browse** → **Compare**
  - audit data quality (W4), validate data quality (B2)

15

## What: Derived

- Infer some of the HPO labels that weren't observed



Source: Author Presentation

16



## How: Scale

- Patients or Patient Cohorts: One dozen
- Phenotypes: About a 100
- HPO Graph Nodes: several thousands