

The Human Phenotype Ontology in 2017

Köhler et al. Nucleic Acid Research, 2016

doi: [10.1093/nar/gkw1039](https://doi.org/10.1093/nar/gkw1039)

Wei-Hung Weng

December 14, 2016

BMI705 Precision Medicine — Integrating Clinical and Genomic Data

<https://ckbjimmy.github.io/705hpo.pdf>



Deep Phenotyping

- Precise and comprehensive analysis of phenotypic abnormalities in which the individual components of the phenotype are observed and described
- Signs, symptoms, laboratory findings, imaging studies
- Ultimate goal - Calculating the similarity between diseases, symptoms, and patients (Robinson 2015)

Human Phenotype Ontology

Home About Downloads Tools Documentation Users History
FAQ License Citation Contact

New HPO Release (September)
September 5, 2016
HPO release - September 2016

New HPO Release (July)
July 4, 2016
HPO release - July 2016



Human Phenotype Ontology

- Integrating phenotype information across fields and databases
- Standard for phenotypic abnormalities / variation
 - Orphanet, NIH UDP, UDN, DECIPHER...
- Three component
 - Phenotype vocabulary (5717 classes, 11813 terms)
 - Disease–phenotype annotations (123724 for rare disease, 132620 for common disease)
 - Algorithms for phenotype driven genomic discovery and diagnostics

REPORT

The Human Phenotype Ontology: A Tool for Annotating and Analyzing Human Hereditary Disease

Peter N. Robinson,^{1,2,*} Sebastian Köhler,^{1,2} Sebastian Bauer,¹ Dominik Seelow,^{1,3} Denise Horn,¹ and Stefan Mundlos^{1,2,4}

There are many thousands of hereditary diseases in humans, each of which has a specific combination of phenotypic features, but computational analysis of phenotypic data has been hampered by lack of adequate computational data structures. Therefore, we have developed a Human Phenotype Ontology (HPO) with over 8000 terms representing individual phenotypic anomalies and have annotated all clinical entries in Online Mendelian Inheritance in Man with the terms of the HPO. We show that the HPO is able to capture phenotypic similarities between diseases in a useful and highly significant fashion.

How HPO Looks Like?

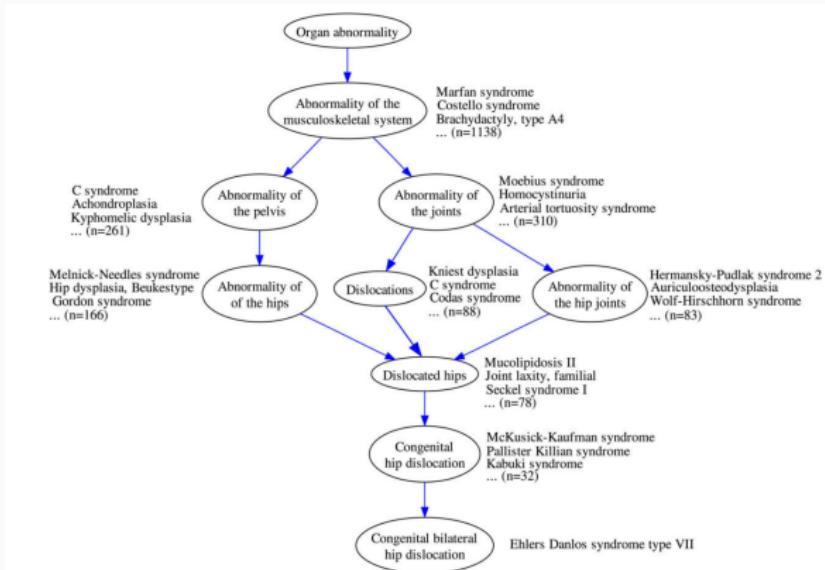
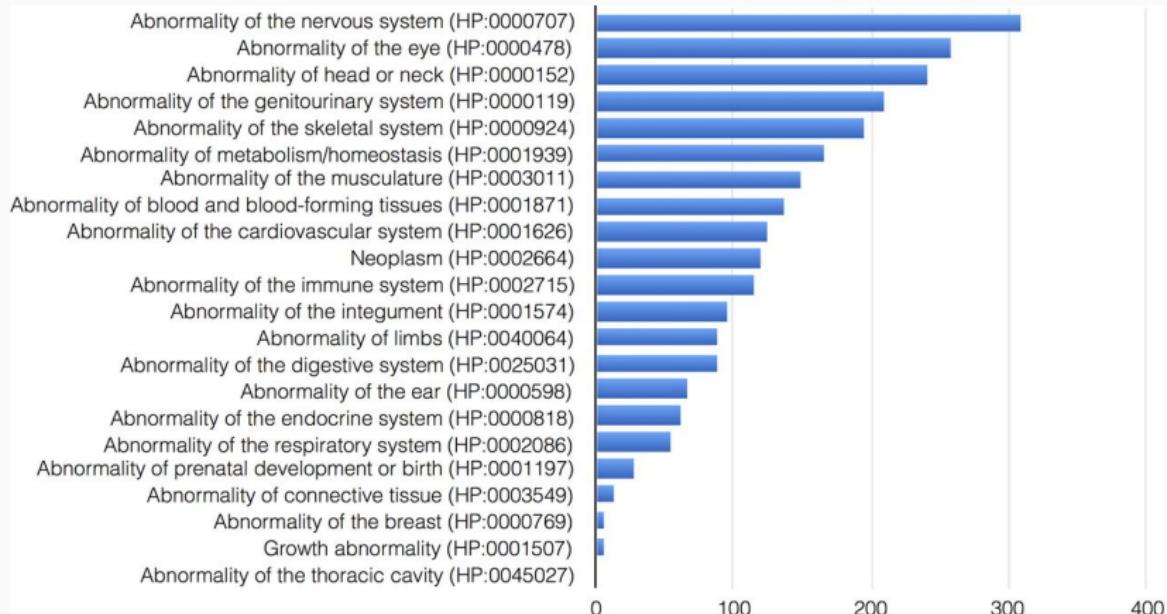


Figure 1. The Human Phenotype Ontology

The HPO term *Bilateral congenital hip dislocation* and all paths to the root that emanate from this term are shown. Some of the annotated disease entries from OMIM, as well as the total number of annotated diseases, are shown next to the terms. Note that because of the true-path rule, a disease that is directly annotated to a specific term is also implicitly annotated to all ancestors of that term. For instance, Ehlers Danlos syndrome type VII is directly annotated to *Bilateral congenital hip dislocation* and is thereby implicitly annotated to *Abnormality of the hips*, *Dislocations*, and the other terms shown in the figure.

Distribution of HPO Classes



Terms, Annotations and Ontology Integration

- Subontologies (different categories)
 - Phenotypic abnormality
 - Mode of inheritance: allows disease models to be defined according to Mendelian or non-Mendelian inheritance modes
 - Mortality/Aging: allows the age of death typically associated with a disease or observed in a specific individual to be annotated
 - Clinical modifier: provide terms to characterize and specify the phenotypic abnormalities defined in the Phenotypic abnormality subontology

id	name	is_obsolete	is_root	subontology	comment	acc	term1_id	term2_id	relationship_type
2	Abnormality of body height	✓	✓	✗	none	HR-J000002	14	100645	is_a
3	Multicystic kidney dysplasia	0	0	○	HULL	HP:0000003	15	8691	is_a
5	Mode of inheritance	0	1		HULL	HP:0000005	15	12619	is_a
6	Autosomal dominant inheritance	0	0		HULL	HP:0000006	20	10992	is_a
7	Autosomal recessive inheritance	0	0		HULL	HP:0000007	22	8775	is_a
8	Abnormality of female internal genitalia	0	0	○	HULL	HP:0000008	22	9714	is_a
9	Functional abnormality of the bladder	0	0	○	HULL	HP:0000009	22	12872	is_a
10	Recurrent urinary tract infections	0	0	○	HULL	HP:0000010	25	26	is_a
11	Neurogenic bladder	0	0	○	ENTRE	HP:0000011			

Annotations

- Main application: rare disorders, personalized medicine → phenotypic-driven analysis with text mining!
- Large corpus of disease-HPO annotation using OMIM, Orphanet, DECIPHER
- PubMed corpus → infer 132620 annotations for 3145 common diseases (validated, with precision of 67%)
- Adopted by GWAS analysis, and maybe PheWAS in the future
- Algorithms for cross-species searching
- **Monarch Initiative** has annotation sufficiency meter (rating system) → helping annotators to generate annotation profile

term_id	external_object_disease_id	evidence_code	is_negative	frequency_modifi...	annotation_id	annotated_date	annotated_by
1233	273460	IEA	0		411688	2009-02-17	HPO
1376	273460	TAS	0	hallmark	411689	2016-01-13	orphanet
2987	273460	IEA	0		411690	2009-02-17	HPO
2997	273460	TAS	0	hallmark	411691	2016-01-13	orphanet
3016	273460	TAS	0		411692	2013-03-30	HPO
3042	273460	TAS	0	hallmark	411693	2016-01-13	orphanet

Integration

- Phenotype before: billing and QC
- Phenotype future: diagnosis and cross-species comparisons
- UMLS covered 54% HPO concepts in 2014
- UMLS integrates HPO in 2015AB
- Good for cross reference and mapping!

↓ Screenshot from UMLS 2016AA

CUI	LAT	TS	LUI	STT	SUI	ISPREF	AUI	SAUI	SCUI	SDUI	SAB	TTY	CODE	STR
Go forward to the next source location (%F10)					0006	Y	A24665785	NA	NA	HP:0000034	HPO	ET	HP:0000034	Hydrocele
2 C3489396	ENG	P	L6506454	VCL	S14971697	Y	A24665788	NA	NA	HP:0000044	HPO	ET	HP:0000044	Isolated hypogonadotropic hyp
3 C0022679	ENG	S	L0185587	PF	S0250196	Y	A24665806	NA	NA	HP:0000107	HPO	ET	HP:0000107	Cystic kidney disease
4 C4020894	ENG	P	L6501211	PF	S7468873	Y	A24665820	NA	NA	HP:0000176	HPO	ET	HP:0000176	Submucous clefting
5 C0158646	ENG	P	L0185531	VO	S14968914	Y	A24665824	NA	NA	HP:0000202	HPO	ET	HP:0000202	Cleft lip; cleft palate
6 C4020890	ENG	P	L12073483	PF	S14972657	Y	A24665892	NA	NA	HP:0000456	HPO	ET	HP:0000456	Notched nasal tip
7 C0015397	ENG	S	L0015397	VO	S14970420	Y	A24665896	NA	NA	HP:0000478	HPO	ET	HP:0000478	Eye disease
8 C1842584	ENG	P	L6496640	PF	S7523137	Y	A24665929	NA	NA	HP:0000570	HPO	ET	HP:0000570	Impaired saccades
9 C4020886	ENG	P	L12072181	PF	S14969579	Y	A24665950	NA	NA	HP:0000657	HPO	ET	HP:0000657	Defective or absent horizontal v
10 C0004941	ENG	P	L0004941	VC	S7397463	Y	A24665968	NA	NA	HP:0000708	HPO	ET	HP:0000708	Behavioral symptoms
11 C0013491	ENG	P	L0013491	VO	S0036188	Y	A24666036	NA	NA	HP:0000978	HPO	ET	HP:0000978	Echymoses

Clinical Utility

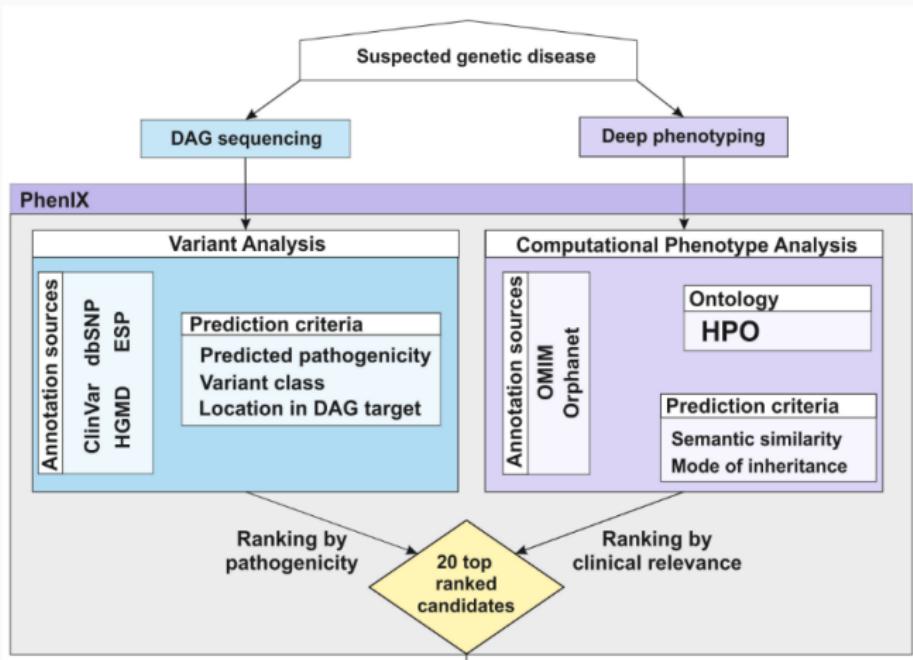
- Assert the pathogenic variants: identifying at least two unrelated cases
- How HPO can help? All rare disease project can use HPO to match phenotype and genotype data! (Table 1)
- BDO+HPO for genetic skeletal disorders, DECIPHER DDD study for developmental disorders, NIH UDP, NIH UDN, EU RD-Connect, EURenOmics, NeurOmics, RES project for rare epilepsy, 1000 GP...
- Integrating automated annotation of HPO terms to 3D facial images

Clinical Utility

- Tools
 - PhenoTips: pheno-genotype data collection tools (can be incorporated into EMR to characterize patient using HPO terms!)
 - Patient Archive: patient data management, HPO-centric patient phenotyping in clinical notes
 - Phenopolis: Platform for functional studies for relevant genetic diseases
 - ... Table 2

Translational Research

- Diagnosis for rare disease patient is difficult (< 50% diagnostic rate)
- PhenIX tries to solve this problem (HPO terms + VCF file)
- Using HPO to generate individualized phenotype-driven gene panels for diagnosis
 - Identify predicted pathogenic mutation in exome
 - Rank relevant genes according to phenotypic relevance by Phenomizer (semantic similarity)
 - (Basic idea of ontological search: do not need exact match, but semantically similar diseases score)



PhenIX (Zemojtel 2014)

Gene Identification Research

- Gene filtering using HPO terms
- Phenotype Similarity Regression (SimReg)
 - Algorithm for identifying composite phenotypes associated with rare variation in specific genes
- Matchmaker exchange (MME) platform for rare disease gene discovery, link to federated network of phenotype-genotype database (30000 cases, 6 databases)

Collect Patient Phenotypes

- Clinical free text
- Ask authors to provide selected HPO terms, HPO-coded phenotype data (to PhenomeCentral) for research paper
- Patient phenotyping: HPO class labels for layperson (plain language terms) to provide increased access to HPO!
 - e.g. dyschromatopsia vs. color-blind
 - Now HPO has 6K layperson synonyms
 - Data interoperability across clinicians and patients

Assessment by NIHR

Table 3. NIHR-RD-TRC assessment scale

Stage	Description	Example
Foundation	The basis of characterizing the disease in HPO needs to be developed	HPO is good for describing dysmorphologies especially across species: how do you model and use dyslexia?
Formulation	The theory is defined but key details need to be defined and handled in the ontology computations	HPO models biology, where diseases are caused by environmental factors, e.g. cancers — how can an environment ontology be included?
Refinement	The key data sets and definitions for the disease are identified and available but require ‘translation’	Theme based registry systems hold collections of data in other coding systems (registry-specific or ICD) — how can these be mapped onto HPO?
Maturity	The HPO framework is in place and productive results are being obtained, the HPO term set continues to evolve	The HPO basics are in place and a set of Phenotypes in place — do we need more terms or do existing terms need modification?

Table 4. NIHR-RD-TRC assessment of HPO maturity

Theme	Foundation	Formulation	Refinement	Maturity
Cancer	✓✓✓✓✓	✓✓		
Cardiovascular	✓✓✓✓✓✓	✓✓✓	✓✓	
Central Nervous System	✓✓✓			
Eye Diseases	✓✓✓✓✓✓	✓✓✓✓✓✓	✓✓✓✓✓✓	✓✓
Gastrointestinal	✓✓✓✓✓	✓✓✓		
Immunological Disorders	✓✓✓✓✓✓	✓✓✓	✓✓	
Paediatric (cross-cutting)	✓✓✓✓✓✓	✓✓✓	✓✓✓	✓
Metabolic & Endocrine Diseases	✓✓✓✓✓✓	✓✓		
Musculoskeletal Disorders	✓✓✓✓✓✓	✓✓✓✓✓✓	✓✓✓	
Muscle & Nerve Diseases	✓✓✓✓✓✓	✓✓✓	✓	
Non-malignant Haematology	✓✓✓✓✓✓	✓✓✓✓✓✓	✓✓✓✓✓✓	✓✓✓
Renal	✓✓✓✓✓✓	✓✓✓✓✓✓	✓✓✓	✓
Respiratory Diseases	✓✓✓	✓		
Skin Diseases	✓✓✓✓✓✓	✓✓✓✓✓✓	✓✓✓	

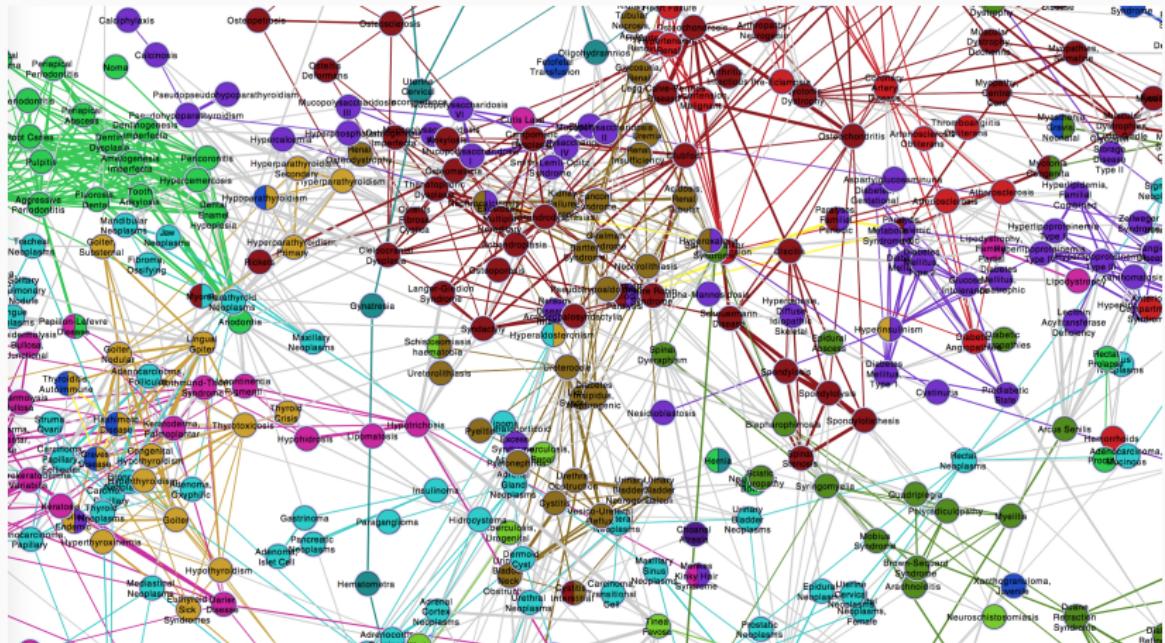
Future Direction

- International work
- Rare disease specific informatics ecosystem + Rare disease ontology
- Extension to non-Mendelian inheritant diseases, cancer, ...
- Now we have Monarch for cross-species comparison

Some Thoughts

- Phenotypes for behavior, metabolism, craniofacial
- Consider time course, multimorbidity, medication, treatment, side effect, ...
- Molecular taxonomy
- Probably a good starting point for "new taxonomy" of precision medicine
- Deep learning is not enough (features are difficult to interpret), still requires good knowledge representation

Phenotype Network of Common Diseases



Groza, AJHG 2015