PhenoBlocks: Phenotype Comparison Visualizations

Michael Glueck, Peter Hamilton, Fanny Chevalier, Simon Breslav, Azam Khan, Daniel Wigdor, Michael Brudno

HAL Id: hal-01226032
https://hal.inria.fr/hal-01226032

Submitted on 14 Dec 2016

PhenoBlocks: Phenotype Comparison Visualizations

Michael Glueck, Peter Hamilton, Fanny Chevalier, Simon Breslav, Azam Khan, Daniel Wigdor, Michael Brudno

Abstract— The differential diagnosis of hereditary disorders is a challenging task for clinicians due to the heterogeneity of phenotypes that can be observed in patients. Existing clinical tools are often text-based and do not emphasize consistency, completeness, or granularity of phenotype reporting. This can impede clinical diagnosis and limit their utility to genetics researchers. Herein, we present PhenoBlocks, a novel visual analytics tool that supports the comparison of phenotypes between patients, or between a patient and the hallmark features of a disorder. An informal evaluation of PhenoBlocks with expert clinicians suggested that the visualization effectively guides the process of differential diagnosis and could reinforce the importance of complete, granular phenotypic reporting.

Index Terms—Clinical diagnosis, differential hierarchy comparison, ontology, genomics, phenomics, phenotype

---

INTRODUCTION

In medicine, the clinical differential diagnosis of hereditary disorders to determine the disease with which a patient is afflicted is a complex and iterative process. This practice involves significant data gathering efforts, including reviewing patient and family medical history, evaluating environmental exposures, performing physical examinations, diagnostic imaging, biochemical investigations, and psychological testing [21][32][51]. During differential diagnosis, a clinician characterizes a patient by a set of *phenotypes*, i.e., observable and measurable deviations from expected morphology, physiology, and behavior [4][51]. The observation of phenotypes is the single most important responsibility of a clinician [51], as phenotypes guide the development of a clinician's hypotheses. Such hypotheses are verified by additional investigations and analyses, or tracking the effectiveness of a course of treatment. The results of such tests provide additional patient phenotypes and enable clinicians to build a comprehensive characterization of a patient's phenotypes, one block at a time.

Prior investigations have reported that recorded phenotypic descriptions are often inconsistent, incomplete, or imprecise [51]. Existing differential diagnosis database tools, such as London Medical Databases [34] or POSSUM [44], allow clinicians to query potential disorders associated with observed phenotypes, but current search algorithms do not leverage the granularity of phenotype descriptions or semantic relationships between search terms [50]. Recent initiatives, such as the Human Phenotype Ontology (HPO) [30], promote the use of standardized terms for clinical reporting and support a framework for semantic matching. Such tools, however, are only accessible through text-based interfaces that do not reveal the hierarchical context of phenotypes and do little to encourage clinicians to improve the quality of their phenotypic reporting.

We address these shortcomings through *PhenoBlocks*, a novel visual analytics tool that supports clinical differential diagnosis. Using the HPO, a differential hierarchy comparison algorithm was developed to analyze phenotypes pairwise between patients and display the results using Sunburst radial hierarchy layouts [54]. This

visualization highlights the semantic context of the comparisons, uncovering hierarchical relationships that are not apparent in existing text-based interfaces (Fig. 1).

Using linked, differential views [17] organized as small-multiples, clinicians can visually compare the phenotypes of an undiagnosed query patient to a cohort of diagnosed reference patients who exhibited similar phenotypes (Fig. 5) and update the phenotype data in-place. Using phenotypes of existing diagnosed patients as baselines for comparison could encourage clinicians to improve the quality of their own phenotyping while also mitigating errors of probabilistic reasoning. Important phenotypes are made visually salient, allowing clinical investigations to be prioritized.

A user-centered approach informed the design of PhenoBlocks. First, clinicians were interviewed to understand existing workflows. A workflow analysis then identified key tasks and a preliminary evaluation elicited areas for improvement that could be aided by visual analytics. Addressing these improvements, we implemented PhenoBlocks and validated it via a use case scenario and feedback from clinicians. As a result, PhenoBlocks is a new visual analytics tool for clinical differential diagnosis to address the present and growing needs of clinicians and researchers alike. We contribute a differential hierarchy structure comparison algorithm, a visual encoding strategy, and a differential visualization to compare the structure of hierarchies.

- *M. Glueck is with Autodesk Research and the University of Toronto (e-mail: mglueck@dgp.toronto.edu)*
- *P. Hamilton is with the University of Toronto (e-mail: hamilton@dgp.toronto.edu)*
- *F. Chevalier is with INRIA (e-mail: fanny.chevalier@inria.fr)*
- *A. Khan and S. Breslav are with Autodesk Research (e-mail: azam.khan@autodesk.com; simon.breslav@autodesk.com)*
- *D. Wigdor is with the University of Toronto (e-mail: daniel@dgp.toronto.edu)*
- *M. Brudno is with the University of Toronto and the Hospital for Sick Children in Toronto (e-mail: brudno@cs.toronto.edu)*
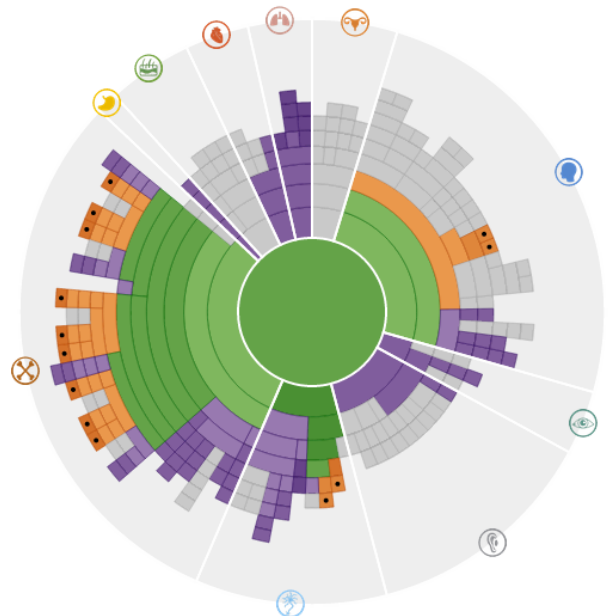
Fig. 1. An example of a single differential view in PhenoBlocks, comparing the phenotype hierarchies of an undiagnosed query patient to a diagnosed reference patient. During differential diagnosis, clinicians use shared (green) phenotypes to gauge confidence in their diagnostic hypothesis and missing (purple) phenotypes to identify candidates for subsequent analysis.

1 PHENOTYPES IN MEDICINE

Within the medical domain a *phenotype* is an observable and measureable deviation of some aspect of a patient from the expected manifestation within a population [4][51]. Thus, a phenotype describes an abnormality of a patient's morphology (i.e., structural features such as having a broad thumb or low-set ears), physiology (i.e., functional features such as cognitive impairment or seizures), or behavior (i.e., depression or impulsivity).

While genetic variations and inheritance could be the primary causes of phenotypes, not all phenotypes are present from birth. Onset may occur later in life (e.g., hearing loss) or could be due to environmental factors (e.g., cancers). Some phenotypes are also more likely to occur in males than females (e.g., red-green color-blindness), or differ between ethnic populations (e.g., epicanthal folds are more common in patients of Asian descent).

The study of phenotypes seeks a complete and detailed understanding of the spectrum of phenotypic abnormalities associated with diseases. Phenotypes are the primary data type used by clinicians when making a diagnosis [4]. The knowledge of phenotype frequency enables clinicians to differentiate a symptom of underlying syndromes from an isolated feature, resulting in a diagnosis and course of treatment [49]. A *rare* phenotype occurs in less than 5% of cases of a given syndrome, whereas a *hallmark* occurs in more than 90% [1].

1.1 Phenotype Granularity

Phenotype reports form part of a patient's Electronic Health Record (EHR). Since these entries are written using natural language, many notations, abbreviations, and synonymous terms have evolved [51] and are reported at different *granularities* of detail [35]. While comparing published results, insufficient granularity can introduce uncertainty of the actual observations of the clinician [49].

Research trends in phenomics, called *deep phenotyping*, rely on comprehensive patient phenotypes to infer the causality of gene variation [4]. Innovative computational approaches promise to deepen the understanding of the spectrum of phenotypic abnormalities by automating phenotype identification and quantification (i.e., analysis of medical imaging and 3D scans of morphology) [51]. Given current technologies, it is not practical to fully phenotype each patient. Thus, clinicians and researchers stand to benefit from advanced diagnosis tools that promote the use of consistent phenotype terminology and encourage complete and granular phenotyping in pursuit of next-generation personalized medicine.

1.2 Human Phenotype Ontology

The Human Phenotype Ontology (HPO) is an international project to address these shortcomings by standardizing terminology and describing the semantic relationships between phenotypes [30]. Each entity in the ontology describes a distinct phenotypic abnormality, and is arranged in a hierarchical structure representing subclass "is-a" relationships. As a phenotype may have more than one superclass, the HPO supports multiple inheritance. Thus, the HPO is a highly complex directed acyclic graph, following the *true-path rule*. The presence of a specific phenotype implies all ancestor phenotypes (i.e., Celiac Disease is a subclass of both Abnormality of the Intestine and Abnormality of the Immune System).

The HPO uses phenotype-disease co-occurrence annotations from existing medical databases, such as the Online Mendelian Inheritance in Man (OMIM) [40], to calculate an *information content* score. This score quantifies the diagnostic specificity of a given phenotype based on how frequently it occurs amongst all known disorders (i.e., phenotypes linked to fewer diseases have higher information content). Terms closer to the root are also less clinically specific and have less informational content than more granular terms farther from the root [50]. Since syndromes in OMIM are annotated to HPO terms, phenotypes can be used to describe all the signs, symptoms, and other manifestations that characterize a given disease. The HPO is an ongoing effort that currently includes over 10,000 terms and 50,000 annotations for over 5,000 known syndromes.

1.3 Hierarchical Phenotype Data

Phenotype data are often represented as a list of HPO terms, with each phenotype being *present* (i.e., the phenotype was investigated and found) or *absent* (i.e., the phenotype was investigated and not found) in the patient.

1.3.1 Phenotype Similarity

A list-based phenotype representation lends itself to calculating the cosine similarity between feature vectors, a technique used predominantly in existing clinical tools [50]. These searches match only exact terms and fail to account for the semantic similarity between phenotypes (e.g., the phenotypes Wide Nose, Long Nose, and Broad Nose are terms with high semantic similarity). When a patient's phenotypes are considered in context of the HPO, semantic similarity searches become possible. Phenomizer [29] and PhenoTips [15] demonstrated that semantic similarity scores outperform cosine similarity in both finding and ranking similar diseases.

1.3.2 Patient Phenotypes Comparisons

The calculation of phenotype similarity scores has benefited from the consideration of ontological relationships. However, explicitly representing phenotypes within the ontological hierarchy has not yet been applied to phenotype comparison visualizations. Representing phenotypes in a visual manner can support the discovery of novel insights that are not possible when list-based representations are used. The nearest common ancestor of two phenotypes, for example, can quickly be determined using a visual hierarchy.

In the present work, we introduce a visual analytics tool that supports the comparison of phenotypes between a new, undiagnosed *query* patient and a set of diagnosed *reference* patients. A database of diagnosed patients is searched and the most semantically similar patients to the query are used as references. Each visualization then depicts the differential hierarchy comparing the query patient and one reference patient (Fig. 1). The differential hierarchies are computed by converting patient phenotype data into hierarchical data and then comparing nodes (Fig. 2). This results in nine possible comparison states for each phenotype (Fig. 3). The differential views distinguish between *shared* (i.e., observed as present in both patients or absent in both patients), *divergent* (i.e., observed as present in one patient but absent in the other) and *missing* (i.e., observed in one patient, but neither present nor absent in the other) phenotypes.
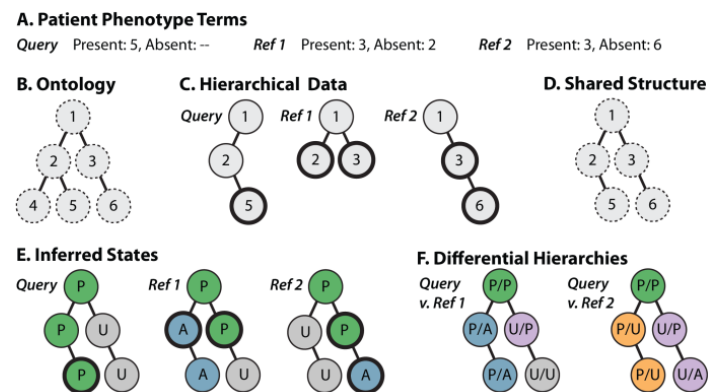

Fig. 2. Differential hierarchy computation. (A) For each patient, the list of phenotype terms are instantiated in the (B) HPO to create (C) hierarchical data. Bold strokes indicate input phenotype terms. (D) To facilitate visual comparisons across differential views, we derive a shared hierarchical structure as the union of query and all reference patient subgraphs. (E) Using the ontological structure, the states of relevant ancestor and descendant phenotypes are inferred, i.e., **P**resence in a child implies presence in the parent, **A**bsence in a parent implies absence in the children. Nodes in the shared hierarchy without data for a patient are marked as **U**nknown. (F) Finally, we compare nodes pairwise between the query and each reference to create the differential hierarchies displayed in the visualization.

Nine States		Reference			
		Present	*Absent*	*Unknown*	
Query	*Present*	P/P	P/A	P/U	**Shared in Both** / **Divergent in Both** / **Missing in Query**
	Absent	A/P	A/A	A/U	
	Unknown	U/P	U/A	U/U	

Shared in Both
Divergent in Both
Missing in Query
Missing in Reference
Unknown in Both

Fig. 3. Differential comparison of phenotypes yields nine states, which can be classified as shared, divergent, missing, and unknown.

2 CLINICAL DIFFERENTIAL DIAGNOSIS

To understand the clinical diagnosis workflow, interviews were conducted with two expert clinical diagnosticians, E1 and E2, who had 14 and 28 years of diagnostic experience, respectively. The clinicians have reviewed over 8,000 cases between them. The interviews introduced *hard handles*, described existing diagnosis tools, and provided a workflow overview. We generalize tasks and highlight improvements to the current workflow that informed PhenoBlocks.

2.1 Hard Handles

When patients with undiagnosed rare genetic disorders are referred to a specialist, the first step of differential diagnosis is to identify the most salient phenotypes of a patient, i.e., *hard handles*. Hard handles are highly indicative of specific disorders; they have high information content (rare among disorders) and occur with high frequency in specific disorders. Hard handles provide the strongest discriminative power to reduce the set of plausible diagnoses.

Clinicians rely on personal experience to identify hard handles: over time, they learn which phenotypes are rare or common, which combinations are expected or surprising, and so on. This is a necessary, but time consuming and error prone aspect of clinical diagnosis training. A clinician identifies hard handles by collecting data from sources, such as the patient's EHR or investigating family medical history, but primarily from an initial face-to-face consultation with the patient. During this consultation, a clinician focuses on externally observable phenotypes that present in the face, limbs, and extremities (e.g., Low-set Ears or Broad Thumbs). A clinician may visually identify dozens of phenotypes in a patient, but only a few qualify as hard handles. Factors, such as the patient's age, sex, and ethnicity are all taken into consideration when searching for hard handles, compounding the difficulty of diagnosis.

2.2 Existing Tools

Many hereditary disorders are rare and a clinician may have limited experience with specific cases [15]. There are currently over 8,000 named diseases and many thousands more yet to be discovered and classified [51]. Many tools help clinicians cross-reference phenotypes to associated syndromes and prioritize the order of investigation during differential diagnosis. The clinicians we interviewed use several tools, including textbooks, online databases, and research prototypes to support clinical differential diagnosis.

Smith's Recognizable Patterns of Human Malformation (Smith's) [24] is a popular reference textbook. The appendix lists syndromes associated with phenotypes and annotates the rate of co-occurrence as "frequent" or "infrequent".

The OMIM is a comprehensive online compendium of Mendelian (hereditary) disorders under development since the 1960s [40] and is the standard for naming and referencing disorders. Descriptions of diseases are regularly updated with the latest research results. Commonly associated phenotypes and genes are included, however, OMIM does not contain phenotype frequency information, does not integrate with the HPO, and only supports keyword search.

LMD is a differential diagnosis database system that identifies associated diseases based on lists of phenotypes and includes reference photos of example cases [34]. LMD and similar databases, such as POSSUM [44], are the *de facto* standard in clinical medicine. However, E1 noted that searches require exact phenotype terms be specified and searches with too many phenotype terms, or very granular phenotype terms, may not result in relevant matches. It is

often necessary to perform multiple searches using increasingly general phenotype terms, or even try different combinations of terms.

Recent research projects aim to improve the availability of phenotype frequency data to clinicians. Orphanet [45] catalogs published research results, updating base-rate frequencies between phenotypes and associated syndromes. This project remains an ongoing initiative due to the difficulty of finding sufficiently large cohorts of patients, particularly for rare disorders. Phenomizer [29] directly leverages the HPO and supports robust semantic searching, meaning the more complete and specific the phenotypes in the query, the better the results of search results. PhenoTips [15] provides a standardized and structured workflow to enter phenotype data using HPO terms. It also allows clinicians to share patient phenotypes and automatically suggests similar cases within the system. Unfortunately, such systems have not been broadly adopted in clinical settings.

2.3 Existing Workflow

In the literature, clinical diagnosis is characterized as an iterative process of hypothesis generation and testing [10][11][58]. A clinician typically follows five steps: (1) identify relevant clinical information about a patient; (2) interpret the meaning of the data; (3) generate hypotheses supporting a coherent explanation in light of the data; (4) test and refine the hypotheses through further data collection; and (5) establish a working diagnosis [10]. During our interviews, both clinicians described similar processes, but provided details specific to differential diagnosis of rare hereditary diseases.

Using identified hard handles, the clinician compiles a list of likely syndromes. The more hard handles unique to a syndrome, the more likely the syndrome is implicated by the observed phenotypes. Both E1 and E2 used several software tools during this process, including OMIM, LMD, Orphanet, and PhenoTips, as well as reference textbooks, such as *Smith's*. They noted that while each has advantages and disadvantages, each tool plays an important role in the process. Currently, no single tool provides all the needed information, so considerable effort is expended to manually compile, cross-reference, and compare data sources.

The identified disorders form an initial set of possible diagnoses, which must be iteratively evaluated and refined. Next, the clinicians search for phenotypic evidence that disconfirms any diagnosis. If a hard handle for a specific syndrome is absent in the patient, this provides some evidence against that diagnosis; the greater the number of absent hard handles, the stronger the disconfirming evidence. As not all phenotypes are externally visible, phenotypes requiring medical testing or imaging are then considered. Invasive testing is generally deferred until necessary due to cost and patient comfort.

By identifying the hard handles for each disorder, additional patient phenotypes can be prioritized for investigation. *"No one has time to work out the full phenotype for every patient. This is why experience is so necessary to make quick and accurate diagnoses. At a certain point, there's diminishing returns"* (E1). E1 explained that this process continues until she has reached a working diagnosis that explains the presence and absence of all phenotypes.

2.4 Tasks

Integrating the findings of the interviews and literature review, a set of high-level tasks informed the design of PhenoBlocks.

2.4.1 Collect Initial Data (T1)

Both clinicians identified that clinical differential diagnosis is a challenging undertaking that relies on the experience of a clinician, which is echoed in the literature [58]. More experienced clinicians may be able to make quicker differential diagnoses since they know that certain combinations of phenotypes, such as Developmental Delay and Broad Thumbs, are exemplary of a small number of syndromes. A less experienced clinician may only focus on the face, and note Developmental Delay and Long Eye Lashes. While these are still hard handles, they would return more results than the prior list, and thus more effort and time to cull at a greater cost and

inconvenience to the patient. Thus, clinicians must engage in a cost-benefit analysis, identifying a minimal set of hard handles that provide high enough discriminative power to select the most likely diagnosis.

2.4.2 Generate Hypotheses (T2)

Hypothesis generation is akin to a breadth-first search, where several competing hypotheses are considered concurrently with the ultimate goal of reducing the set through disconfirming evidence [3]. In this domain, the hypotheses are diagnoses, a *logical competitor set* [11] of plausible diseases with similar presentations but differing underlying causes. Effective diagnosticians use data-driven, forward reasoning. Backward reasoning increases the chance that contradictory evidence is considered anomalous and is often used by novices. The experience of a clinician is the primary factor determining the quality and diversity of hypotheses and typically continues until at least one potentially severe outcome is included [58]. On average, four hypotheses are considered at a given time by clinicians during hypothesis testing [10][58].

2.4.3 Seek Disconfirming Evidence (T3)

Failing to seek disconfirming evidence or incorrectly discounting observed disconfirming evidence are hallmarks of inefficient analysis methods [22]. Both forward reasoning and seeking disconfirming evidence require sufficient relevant knowledge of the problem domain. While generalized knowledge is sufficient to generate hypotheses, specialized knowledge is necessary for effective evaluation [3]. This is precisely a scenario where visual analytics tools can excel: combining the strengths of human and electronic processing into an interactive process that extracts knowledge from data [28].

An impediment to quick differential diagnosis lies in the identification of additional hard handles. In both initial phenotype collection and seeking disconfirming phenotypes, the lack of frequency information forces clinicians to rely on their own experience. Since all phenotypes occur with some frequency, it is possible that a patient with a specific disorder will not present with all hallmark features. Based on experience alone, this disorder may be prematurely discounted due to representational biases of the clinician and investigations of related phenotypes may not take place.

2.4.4 Evaluate and Refine Hypotheses (T4)

A successful working diagnosis integrates all available data into a single, cohesive explanation for a case [3] and relies on *abduction* [25]. These explanations disconfirm competing hypotheses in light of all available information [58]. The major cause of misdiagnosis is the failure to integrate clinical data properly, meaning that although all relevant data was collected and correctly interpreted, there was a failure to combine it into a coherent explanation [16]. Expert analysts require only minimal information to make an informed decision as they focus on fewer dominant factors that have the biggest influence, not more data. Thus, investigation towards these critical factors should be promoted [22]. These steps continue until all evidence can be integrated into a single explanation for the phenotypes of the patient.

2.5 Workflow Improvements

Conceptualizing phenotypes and their relationships is thus not a trivial task. Existing clinical tools provide text-based interfaces to phenotype data, requiring clinicians to consider many relationships abstractly in their minds. Visualization can alleviate this burden by explicitly representing phenotypes within the broader ontological context afforded by the HPO. After addressing the shortcomings of existing tools and the tasks elicited, a proof-of-concept visualization was designed and preliminary feedback was collected from clinicians.

2.5.1 Concurrent Investigation and Evaluation (G1)

Clinicians rarely access diagnostic databases concurrent with patient consultations, resulting in increased diagnosis durations because clinicians need to continually consult databases and then reevaluate the patient, i.e., *"If data can be entered at the time of seeing a patient, it could allow for quicker and more complete phenotyping"* (E1). This

parallelism of investigation and evaluation is described in Wheaton's multi-phasic model of intelligence analysis and has also been empirically observed [26]. A tighter coupling of phenotype observation and evaluation could help clinicians identify phenotypes to investigate more efficiently (T1).

2.5.2 Structured Externalizations (G2)

Clinicians often record investigations through *ad hoc*, hand-written notes. From a distributed cognition perspective, insight is an emergent property of interaction with externalized representations of information [33]. This mirrors Pirolli & Card's schemas [42] and Heuer's structured externalizations [22]. Interactive visualizations could thus benefit hypothesis generation by externalizing the thought process and help in the evaluation of competing hypotheses by offloading abstract comparisons to concrete visual representations.

Compared to novices, expert diagnosticians generate hypotheses that account for a wider range of potential medical outcomes [3] and with higher granularity due to their learned sensitivity to the expected base-rates of disorders [58]. Visualizations thus act as a record of the phenotypes investigated during a successfully diagnosed case. This could help novice clinicians learn by example and support expert clinicians when they have limited personal experience in a rare case (T4). In both cases, visualizations can suggest additional phenotypes to investigate (T3).

2.5.3 Availability of Frequency Information (G3)

Given the current workflow, clinicians spend significant time integrating information from multiple data sources. It is valuable to know whether a phenotype is hallmark and/or rare given a specific disorder (T1, T2, T3, T4). Clinicians use their knowledge of phenotype frequency to disconfirm diagnoses where these features are rare and highlight diagnoses where they are hallmarks. With such visualizations however, *"you no longer need to have first-hand experience if the system can provide you with a better picture through concrete examples of what potential combinations are"* (E1). Knowledge of phenotype frequency also enables a clinician to better prioritize the order of investigation; hallmark phenotypes should be investigated before lower frequency phenotypes.

2.5.4 Phenotype Completeness and Granularity (G4)

When provided with exhaustive lists of detailed phenotypes, existing diagnostic tools do not always return better results. This may inadvertently discourage complete and granular phenotyping. Presenting clinicians with examples of diagnosed cases provides a benchmark of the quantity and quality of phenotyping required for a given diagnosis. Moreover, exposure to an actual rare case of a diagnosed disorder could mitigate errors of probabilistic reasoning by concretely representing a case where some hallmark phenotypes are missing. Thus, visual analytics tools can provide better assistance to clinicians when they are shown examples of complete and granular data (T4). Since a major barrier to phenotyping is a lack of time, minimizing the cost of entering phenotype data into a system could encourage phenotype completeness and granularity and should be standard practice for clinicians.

3 RELATED WORK

PhenoBlocks builds upon a variety of prior visualization techniques from several domains: bioinformatics, healthcare, and ontology and hierarchy comparison.

3.1 Visualization in Bioinformatics

Interest in the domain of computational bioinformatics drove the application of algorithmic approaches to improve our understanding of the human genome [8][23]. Following this trend, many bioinformatics visualization tools support the exploration of genomic data, including genome browsers (e.g., Savant2 [13]), visualizations of genome sequence assemblies (e.g., ABySS-Explorer [39]), visualizations of genome comparison (e.g., Circos [31], MizBee [36]),

and tools for gene variant analysis (e.g., Variant View [12]). See Arndt [2] and Nielsen et al. [38] for reviews in this area. These tools do not typically focus on visualizing patient phenotypes.

The increasing availability of low-cost, high-throughput genome sequencing techniques has motivated research into disease genesis and evolution. New tools allow one to investigate the causality of gene variation on phenotypes (e.g., Arena3D [52], PheWAS [9], PheWAS-View [41], among others [5]) and rely on Manhattan plot visualizations of statistical analysis results or use machine learning to visualize complex associations between genes, proteins, and phenotypes. Such tools are integral to understanding how genes and phenotypes differ across disease cohorts. Our focus, however, is to visualize a specific patient's phenotype data against a cohort for healthcare purposes.

3.2 Visualization in Healthcare

A wealth of visualization tools have been developed to explore patient EHRs. Most efforts focus on patient medical histories in a time-oriented context, including monitoring a single patient's overall condition (e.g., LifeLines [43], VisuExplore [47]), analyzing a patient's response to a treatment (e.g., IPBC [7], CareCruiser [18]), and aggregating multiple patient records in a cohort as a baseline to compare a specific individual's evolving symptoms (e.g., Lifelines2 [57], LifeFlow [59]). See Rind et al. [46] and Shneiderman et al. [53] for reviews in the area. These tools are helpful decision-making aids for clinicians when determining a course of treatment for acute or chronic patient problems. To account for the range of problems that could present, these visualizations often focus on the complete EHR. In contrast, little attention has been devoted to specifically visualizing phenomic data in the context of diagnosing rare hereditary syndromes.

3.3 Ontology and Tree Comparison Visualization

An ontology formalizes domain knowledge by describing concepts and their interrelations, and often results in a hierarchy of multiple inheritance represented as a directed acyclic graph. Selecting an appropriate visualization technique can be challenging due to the complexity of the interrelations and the importance of representing leaf nodes. In their comparative survey, Katifori et al. [27] assert that hierarchical space-filling visualization techniques are good candidates for visualizing ontologies, with Treemaps, Icicle plots and Sunburst visualizations [54] as the most widespread examples. PhenoBlocks uses Sunburst visualizations to display the HPO as the radial layout equalizes the size of both ancestor and leaf nodes and has been successfully applied within other healthcare tools.

As the goal of PhenoBlocks is to facilitate comparison between multiple Sunburst visualizations, three categories of visual techniques are relevant to compare hierarchical visualizations, i.e., side-by-side views (e.g., TreeJuxtaposer [37], DaViewer [60], among others [6]), differential views (e.g., TreeVersity [19], among others [55][56]) and animation [48]. See Graham et al.'s survey [17] for a broader review. Consensus trees have been used to summarize the agreement of topological features across multiple trees [37][60]. The approach in PhenoBlocks differs in that not only are shared features computed, but we also explicitly account for dissimilar features and visually encode the source of the difference. These differential views provide specific pairwise comparisons between the phenotype hierarchies of two patients. Unlike prior work, where differential views highlight changes in node value within hierarchies [19][55][56], this work focuses on both topological and value differences between hierarchies.

4 PHENOBLOCKS TOOL DESCRIPTION

PhenoBlocks follows a clinician's workflow when diagnosing a new patient. This undiagnosed patient is the *query* patient. The clinician enters phenotypes for the query patient and PhenoBlocks searches its database for existing diagnosed patients with similar phenotypes. The results of this search are *reference* patients and are grouped by disease and ordered by semantic phenotype similarity. The number of diseases and number of patients per disease are configurable parameters in the system. The clinician then uses the visualizations to prioritize phenotype investigations with the query patient, and enters new phenotypes interactively until a diagnosis is reached.

4.1 Patient Phenotype Data

The prototype system uses a database of 47 patients across six genetic disorders, with between 3 and 11 patients per disorder. As part of the FORGE Canada Consortium [14], phenotype data for each patient was digitally transcribed from published research investigating cohorts of patients with diagnosed diseases.

- Floating Harbor Syndrome, SRCAP, 136140, PMID: 22265015
- MFDM, EFTUD2, 610536, PMID: 22305528
- MICCAP, STAMBP, 614261, PMID: 23542699
- Chudley-McCullough Syndrome, GPSM2, 604213, PMID: 22578326
- Hajdu-Cheney Syndrome, NOTCH2, 102500, PMID: 21681853
- Joubert, C5orf42, 614615, PMID: 22425360

The database consists of a list of present and absent HPO terms for each patient. Complete and granular phenotype data of diagnosed patients are becoming available through research prototypes (i.e., PhenoTips [15]). Alternatively, a clinician can enter custom data into the system to visualize comparisons of specific patients.

4.2 Sunburst Visualizations

To visualize the differential hierarchy of patient phenotypes (Fig. 2), we first convert the data into a tree by way of duplicating nodes with multiple inheritance. A typical differential hierarchy is attributed 10-20 phenotypes and the depth of the converted tree is around 10 levels deep, containing between 30-75 nodes.

PhenoBlocks displays the tree structure using a space-filling hierarchical visualization approach. The Sunburst layout was used as its radial shape makes the best use of display space and, unlike Treemaps, all nodes of the hierarchy are represented simultaneously. This is important for structure-based comparisons, such as identifying the nearest common ancestor of two phenotypes. In contrast to Icicle plots, leaf nodes are also afforded more space in the layout. Phenotypes around the periphery of the chart are critical to differential diagnosis as they are the most granular description of an abnormality.

Each Sunburst visualization represents a differential hierarchy between the query and a reference. Phenotypes are color-coded based on their state, with intensity communicating a *hard handle score*. This weighted score represents both the rareness of the phenotype among all disorders and the likelihood of occurring in the reference disorder, making potential hard handles visually salient. Phenotypes are grouped into high-level categories (e.g., Abnormality of the Head and Neck), defined by the children of the root HPO node. These categories visually divide the Sunburst into distinct regions. Working with clinicians, a set of icons were designed to representation each category in the visualization (Fig. 4).

In the Overview, the comparisons are displayed as small-multiples that are grouped by disorder and arranged into rows (Fig. 5A). Disorders are ordered by greatest similarity from top to bottom.

Fig. 4. Working with the clinicians, we developed a set of icons to represent the high-level categories of phenotypic abnormalities.
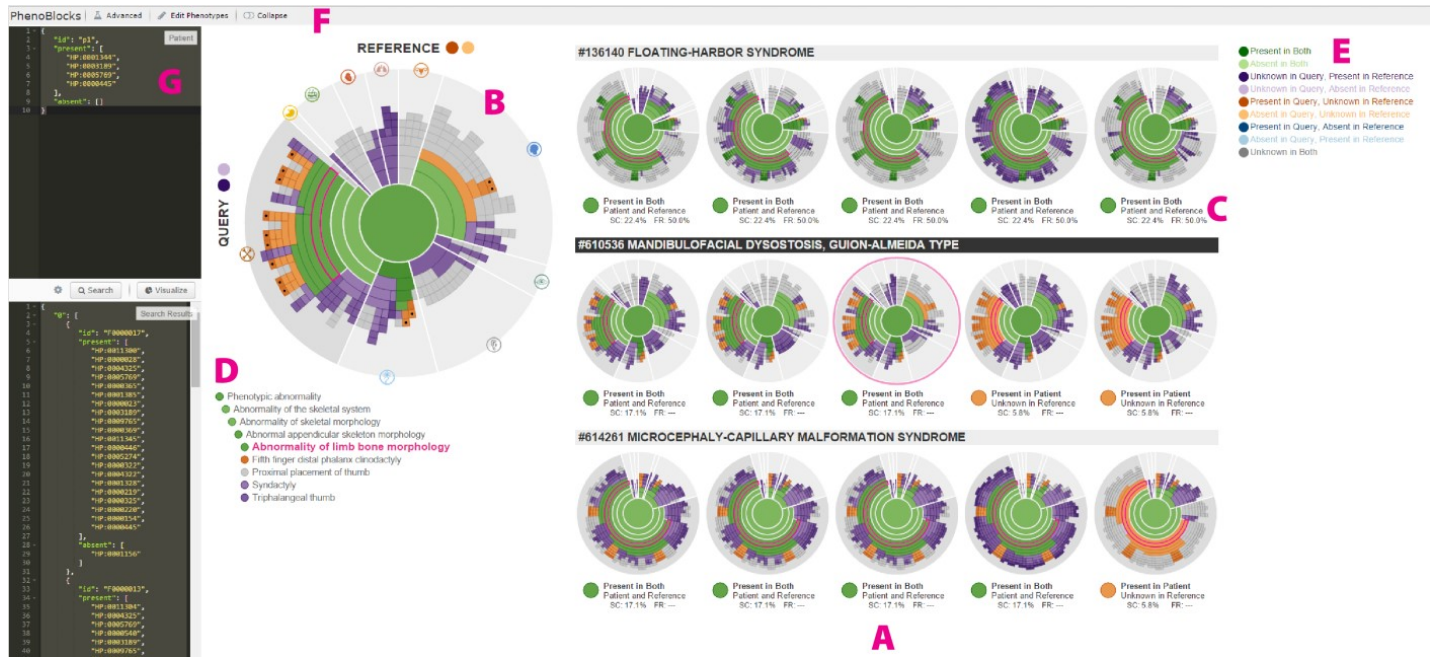
Fig. 5. PhenoBlocks user interface. (A) Overview of differential views using small-multiples. (B) Detail View of highlighted comparison in Overview. Hovering over a phenotype displays details under small multiples and indented list. (C) Details under small multiples include hard handle score and phenotype frequency. (D) Indented list of ancestors and descendants. (E) Interactive legend. (F) Toolbar. (G) Advanced panel.

Comparisons within disorders are ordered by greatest similarity from left to right. A Detail View shows a larger and more detailed version of a single comparison, selected in the Overview (Fig. 5B). The Detail View is the primary user interaction area. It supports details-on-demand through hover feedback, filtering of the hierarchy, and in-place modification of data. An interactive legend identifies the nine possible states for a phenotype (Fig. 5E).

4.3 Differential Views using Small-Multiples

To facilitate the pairwise comparison of the query patient with that of each of the reference patients, PhenoBlocks employs differential views using small-multiples in the Overview. Hence, the strengths of differential views are leveraged, as comparisons within a disease share the same structure, but with different visual encodings. When one hovers over a phenotype in the Detail View, the details of that phenotype within each small-multiple are displayed underneath, including the phenotype state and hard handle score (Fig. 5C).

These differential views allow for a concurrent breadth-first comparison of the query to several references across multiple disorders. Clinicians can quickly look over the range of potential disorders to guide their phenotypic investigations.

4.4 Details-on-Demand

The interface supports immediate feedback to the clinician. First, linked updates to the Overview reveal the phenotype state and hard handle score under each small-multiple comparison (Fig. 5C). Second, hovering over phenotypes in the Detail View displays an indented list of ancestors and descendants (Fig. 5D). Interactively exploring the comparisons may encourage a deeper analysis of what phenotypes could be important to their patients, including some less obvious phenotypes. Reporting on these phenotypes would improve the completeness and granularity of phenotype reporting (G4).

4.5 Interactive Filtering

The visualizations can also be interactively filtered. First, interactive legends can be used to filter phenotypes of a certain state (Fig. 5E). Second, the first level of phenotypes define general categories, such as Abnormality of the Skeletal System. These categories can be collapsed or expanded by clicking on their icons to filter unwanted information or reduce visual clutter (i.e., if a clinician wants to focus on a particular category of abnormalities). Third, the Collapse Button in the Toolbar reduces the visual complexity of the hierarchy by removing redundant intermediary phenotypes (i.e., phenotypes with only a single child) (Fig. 5F). In all cases, smooth animated transitions maximize change awareness.

4.6 Adding and Editing Query Phenotypes

Several mechanisms allow clinicians to add and edit the phenotypes of the query interactively. First, the Edit Phenotypes button in the Toolbar displays a modal window to edit the present and absent phenotype terms (Fig. 5F). Tokenized, auto-completing dropdown lists allow one to search for specific phenotypes using HPO names or IDs. Second, clicking on a phenotype in the Detail View supports in-place data editing, modifying the state of the phenotype from unknown to present or absent. Third, an Advanced Panel can be displayed and used to edit the underlying data in JSON format directly (Fig. 5G). In all cases, the visualization immediately updates (G1). By supporting multiple methods of quickly editing a patient's phenotypic data, we hope to encourage more complete phenotyping (G4).

4.7 Example-Based Comparisons and Frequency

At the core of PhenoBlocks are comparison visualizations of query to similar references in multiple disorders. Unlike existing clinical tools, these comparisons are displayed visually and in the context of the relevant subgraph of the HPO (G2). These visualizations act as an aid to externalize the thought process of the clinician and can be used as a visual record of the diagnosis process.

Since the frequency of phenotypes within disorders is paramount to diagnosis, we take the approach of displaying example-based comparisons to existing diagnosed cases (G3). In addition to the hard handle scores visualized in the comparisons, these concrete references provide exemplars of the hard handles considered by other clinicians in their diagnoses. References can also highlight outlier cases that are similar to the query, which can alert a clinician not to discount a diagnosis prematurely. As probabilistic reasoning can be difficult even for experts [22], explicitly presenting an example of a diagnosed case where a hallmark phenotype is absent can be helpful.

Seeing examples of complete and granular diagnosed patient phenotypes may encourage clinicians to provide a similar level of reporting completeness and granularity (G4).

4.8 Visual Encoding

The visual encoding scheme was designed to highlight potential hard handles and identify how the phenotypes of patients differ. Finding an optimal visual encoding strategy for the dataset was difficult as there were nine ordinal states to differentiate and the hard handle ratio score. Although one could use the divergent-divergent color scheme proposed by Harrower & Brewer [20], the colors were too disparate, making interpretation tedious, and requiring frequent legend consultations. What was needed was to confer a semantic meaning to groups of colors. We wanted to group based on the states of phenotypes into shared, divergent, missing in query, missing in reference, and unknown (Fig. 3). What complicated matters was that we also wanted to differentiate whether a present or an absent phenotype was involved in the comparison. We could not find an existing color palette that met all three of these needs.

Absent phenotypes do not occur as frequently as present phenotypes in the data, likely because it is more difficult to confirm the absence rather than presence of a feature. This was confirmed during our conversations with clinicians. To differentiate between the key phenotype states, represent a ratio value, and confer presence and absence, a novel palette based on Brewer's two-pair categorical scale was developed (Fig. 6). The darker color variant was used for present phenotype comparisons, and the lighter variant for absent phenotypes. A fully saturated absent phenotype would never be as salient as a fully saturated present phenotype, but since absent phenotypes are less prevalent and they tend to be more granular, they stand out in contrast as islands within larger areas of darker, present comparisons. This provides aesthetically pleasing results while not hindering the interpretation of the visualization.

Shared Phenotypes (Greens): Shared phenotypes between the query and reference are the primary indicator of overlapping traits. These shared phenotypes can be present in both, or absent in both. Patients may not always share the same granularity of phenotypic abnormality description; however, by investigating the nearest common ancestor, a decision can be made regarding the similarity of the two traits. These are represented by shades of green with purple and orange descendants.

Potential Hard Handles (Purples): Phenotypes that are missing in the query, but present in the reference are good indicators of potential hard handles. These are represented as shades of purple.

Divergent Phenotypes (Blues): In some cases, there is a clear difference between the phenotype of the query and the reference; the phenotype is present in one and absent in the other.

Missing Reference Phenotypes (Oranges): Sometimes, a query will present with a phenotype that is unknown in the reference. These suggest areas where a difference may exist, but there is a lack of supporting data. If these phenotypes have high hard handle scores, they may indicate a mismatch of the query patient to that disorder, or provide a starting point to seek additional information from the clinician of the reference patient.

Input Data: The actual phenotypes entered into the system are represented using a black dot. These provide landmarks indicating where the clinician has already investigated.
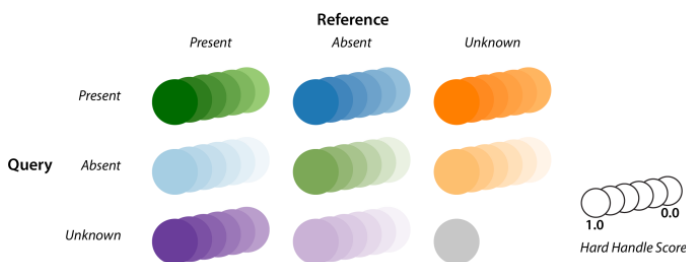

Fig. 6. In PhenoBlocks, the color communicates the state of a phenotype comparison. The intensity of the color indicates the strength of the hard handle score.

4.9 Comparison Algorithms

All the scores represented in the PhenoBlocks visualization are based on a phenotype's information content [29]:

$$IC(n) = -\log\left(\frac{\#\ of\ disorders\ annotated\ with\ n}{\#\ of\ disorders}\right)$$

Information content is proportional to the rarity of HPO term, *n*. As information content correlates with diagnostic value, it is non-decreasing when traversing from parents to children.

Similarity Score: The similarity score ranks similar reference patients in the database based on the phenotypes of a given query patient. We calculate similarity using the PhenoTips algorithm [15], but account for absent terms (i.e., sharing absent phenotypes is also a similarity). The similarity score algorithm is interchangeable with any other method with corresponding inputs and outputs.

Importance Score: The similarity score for a phenotype is based on combining the similarity scores of its descendants. While useful to calculate an overall similarity score, it is insufficient to identify hard handles, where attention should be drawn to specific intermediary phenotypes and not necessarily their ancestors. To account for this, an importance score based on phenotype state was developed (Fig. 3).

Similar in Both: When a phenotype is common in two patients, we can use the information content score of that phenotype, directly. The most important shared phenotypes are always the most specific.

Missing in a Patient: When a phenotype appears in only one patient, the importance score is calculated as the information content of that phenotype reduced by the highest information content among all shared ancestors of that phenotype. This results in a lower importance score for phenotypes that share a closer nearest common ancestor, while increasing importance score for phenotypes where the nearest common ancestor is farther, to highlight larger differences.

Absent Phenotypes: Absent phenotypes complicate the calculation of the importance score. Unlike present phenotypes, absence is inherited by descendants, not ancestors, implying that more general absent phenotypes (e.g., Abnormality of the Skeletal System) have greater discriminatory power than more specific phenotypes (e.g., Abnormality of the Finger). Thus, in states involving absent phenotypes, the importance score is calculated as the complement of the information content of that phenotype.

Frequency Score: Using frequency annotations from Orphanet [45], we calculate a frequency score based on the reference patient's disorder. If no frequency information exists for a phenotype, it is assigned the highest frequency among its annotated descendants. If no annotated descendants exist, it is assigned the average frequency of all annotations. If no frequency information exists, the computed average frequency of all phenotypes (0.51) is used.

Hard Handle Score: The hard handle score is the average of the importance and frequency scores for a given phenotype.

5 CLINICAL USE CASE SCENARIO

PhenoBlocks was designed to be used by clinicians during all phases of differential diagnosis. In this section, we detail how concrete tasks can be completed using PhenoBlocks. While the dataset does not currently include a full range of disorders, we illustrate use through a simplified, hypothetical example.

5.1 Collecting Phenotypes and Generating Hypotheses

Consider an undiagnosed patient who is referred to a specialist for consultation. Immediately, two striking features stand out: they are unable to speak and have short stature. These terms are entered into PhenoBlocks and these phenotypes are visualized, along with similar reference patients from different disorders. The search results are limited to three similar disorders, with five references each (Fig. 7).

The three disorders are listed, ordered by similarity score: Floating-Harbor (FLHS), Microcephaly-Capillary Malformation (MCMS), and Chudley-McCullough (CMCS). Looking at the small-

multiples, the darker green FLHS comparisons indicate more important phenotypes are shared between the query and references (Fig. 7A). First, we investigate the highest ranked comparison.

Immediately, we notice the deep green of Short Stature, which signifies it is a likely hard handle (Fig. 7B). Hovering over this phenotype displays details under each small-multiple (Fig. 7EF). We notice that Short Stature is present in all FLHS references and that it is a hallmark (FR 90%) of the disorder (Fig. 7E). The details of both MCMS and CMCS indicate with orange that this phenotype is missing in the reference (Fig. 7F). In MCMS frequency information indicates this phenotype is rare (FR 5%), so this is not a hard handle for MCMS. All CMCS patients are missing Growth Abnormalities, indicating this may not be an important feature, but too early to discount completely.

Next, we turn to Absent Speech. While this phenotype is missing in the reference we note the direct ancestor, Delayed Speech and Language Development, has a deep green color (Fig. 7C). This indicates semantically similar terms are used in the reference patient, and suggests related phenotypes to investigate. When comparing across disorders, we note Delayed Speech and Language Development is shared by all FLHS references but only by one CMCS reference. This is additional evidence against CMCS.

In summary, there appears to be moderate evidence for FLHS, but further investigation is necessary. In this first step, we collected initial phenotypes (T1), started generating hypotheses (T2), and demonstrated the value of frequency information (G3). The use of visualization has helped us understand the similarities and differences between disorders (G2).

5.2 Identifying Disconfirming Hard Handles

Now that we have a plausible diagnosis (i.e., FLHS), we want to identify potential hard handles that can help differentiate the disorders. For this, we look for deep purple areas in the comparisons. There are many, so we focus on specific categories that are different in size between disorders. For example, Abnormality of the Skeletal System is large in both FLHS and MCMS, but small in CMCS; Abnormality of the Head and Neck has many features for FLHS, and not the others.

Looking at the skeletal abnormalities of FLHS and MCMS, across the small multiples, most references (4/5) share Abnormality of the Fingers. We can quickly view all descendants of a phenotype with the indented list view (Fig. 7GH). In FLHS, these abnormalities are Broad Fingertip, Broad Thumb, and Fifth Finger (Fig. 7G). Most MCMS references (4/5) share Short Distal Phalanx of Finger (Fig. 7H). Investigating head and neck abnormalities, we see that in FLHS references these are limited to the face, while in CMCS references they are limited to skull size. This information is quickly parsed using visual encodings; traditional methods would involve cross-referencing textbooks and databases.

These observations lead us to look more closely at the hands and face of our patient to try and identify more hard handles. We find that our patient has a triangular face, with a long, wide nose, and also confirm presence of Fifth Finger Clinodactyly. However, we find no broad thumbs or fingertips. After further investigation, we generalize the absence of Abnormality of the Phalanxes of Fingers and Abnormality of Skull Size. Broader absences are more discriminative.

By investigating where phenotypes of the query patient differ from the references, we are able to prioritize an investigation of disconfirming hard handles (T3). This process can be done interactively during a patient consultation (G1). By comparing to diagnosed references, granular present phenotypes, such as Fifth Finger Clinodactyly, and general absent phenotypes such as no Abnormality of Skull Size (G4) can be indicated.

5.3 Hypothesis Evaluation

These new phenotypes are added directly into the visualization by clicking on the respective phenotypes and selecting the new state, thus receiving immediate feedback as the visualization is updated (Fig. 8). We see many more shared phenotypes across FLHS references, and note that the green color reaches the periphery, indicating specific shared features (Fig. 8A). This contrasts MCMS references where shared phenotypes are more central. The presence of orange highlights differences between the patient and the MCMS references (Fig. 8B). By exploring the details of the updated visualization, hypotheses can be further evaluated (T4).
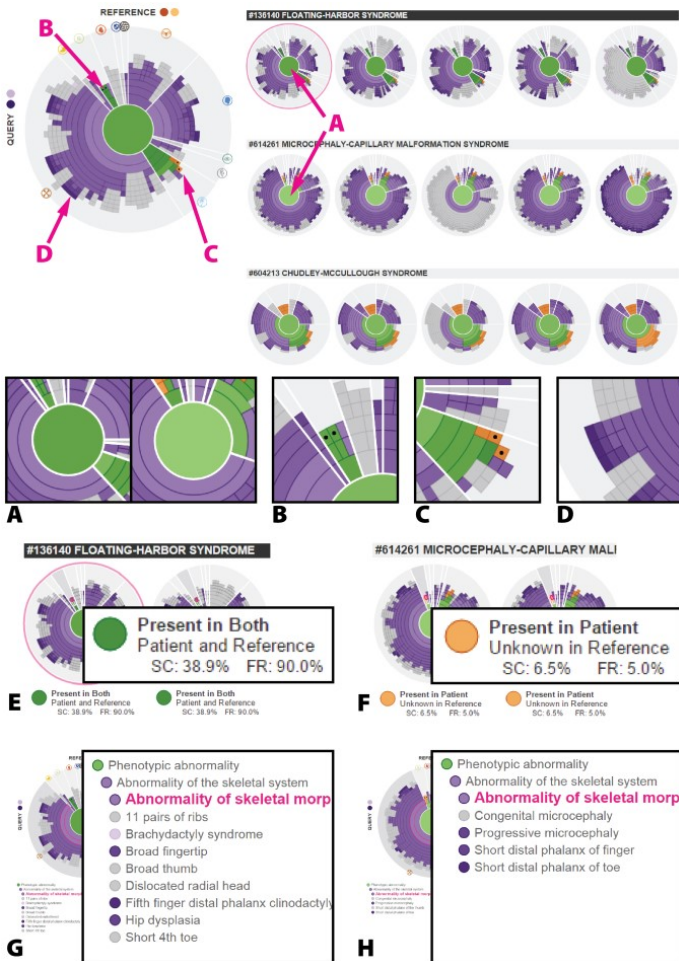

Fig. 7. Initial view. (A) Deeper green implies more important shared features. (B) Short Stature is a shared phenotype in FLHS. (C) Absent Speech is missing (orange), but its direct ancestor is shared (green). (D) Possible hard handles appear in deep purple. (E, F) Hovering over Short Stature allows comparison of phenotype details between differential views. (G, H) Hovering over Abnormality of Skeletal Morphology yields an indented list of all children, supporting quick overview of granular phenotypes across all differential views.
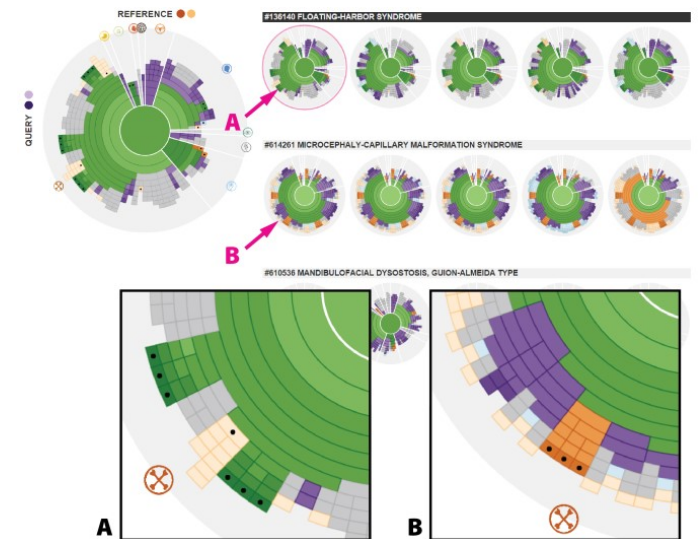

Fig. 8. Updated view. (A) Green phenotypes reach the periphery in FLHS, indicating specific features are shared, in contrast to (B) MCMS where only central general features are shared and granular phenotypes are purple or orange.

6 Informal Evaluation

PhenoBlocks was presented to the clinicians consulted in follow-up interviews. First, we explained the visual encoding strategy and the calculation of the hard handle score. Second, we described the interface and interactions. Finally, we walked through the usage scenario described above to elicit general feedback about the tool and to reflect on its use in their existing workflow.

Overall, the clinicians were excited by the ability to visually compare patient phenotypes, *"it is amazing to be guided by what the biggest differences are [... for example,] one that has a specific eye phenotype and the other two don't. The clear next step is to input the phenotype for the patient's eye. This helps me focus on the diagnosis, not searching databases [...] I think that's the main difference between this and any of the other tools that I've seen"* (E1).

Realistic Usage Scenario: The use case scenario focused on using PhenoBlocks from the very beginning of the differential diagnosis workflow, entering phenotypes one at a time. We were cautioned this is unlikely, as clinicians are trained to quickly filter through possible disorders while collecting initial phenotypes. However, the power of the tool lies in investigating details, *"[PhenoBlocks] comes in to be helpful as soon as you [...] enter [all the phenotypes]. Then you start to think 'ok, what do I need to look specifically for, that I haven't necessarily checked for?'"* (E1).

One clinician commented that ideally PhenoBlocks could be used in conjunction with PhenoTips to create a "diagnostic hub" that would "*take over the functionality of the LMD database from the point of view of inputting the phenotypes*" (E1). LMD would still be useful as a reference tool for clinical descriptions and reference photographs.

Importance and Frequency: The clinicians noted the hard handle score was the most important part of this tool for both hypothesis generation and evaluation. Because it is a "*composite of information content and OMIM frequencies [it provides] a way to visually interpret both scores simultaneously*" (E2), which is currently not possible. E1 stated, "*if my patient had a phenotype that was really uncommon in a particular syndrome or was absent in a syndrome I was looking at and I thought the phenotype was really important [...] then that would lead me away from that particular syndrome.*"

This score also provides guidance in prioritizing the next tests to run, "*I don't want to go x-ray a [patient] for short phalanges if that's only present in 2% of [cases] because that isn't going to help me. But, if 100% of [patients] with this condition have an abnormal thumb on x-ray, that's the type of thing that would make you push for that particular investigation*" (E1). The clinicians agreed that seeing an outlier case was more powerful than knowing the frequency alone, "*you could have a patient with [a disorder], but without the hallmark features. If you're presented with an example of [that disorder] that lacks these features [...] it helps to ground the probability in a concrete example. So, it's really helpful to see that in a graphical representation*" (E1).

Absent phenotypes: With respect to the underrepresentation of absent phenotypes in the dataset, E2 commented, "*in my own experience, absences are notes of an explicit evidence trail that a hypothesis was tested.*" Even though they are more discriminative, general absences are seldom reported because the time they take to investigate does not align well with hypothesis-driven testing.

Other Application Areas: One clinician commented that this style of visualization could be applied to general clinical diagnosis scenarios, "*I still think this is a really, really interesting way of thinking about diagnosing. Not just for genetics but for any area of medical specialty [...] If somebody comes in with chest pain, what you really want are those conditions where chest pain is typically the presenting feature*" (E1).

While using PhenoBlocks during our first interview, one clinician quickly noticed, none of the reference patients of Floating-Harbor Syndrome exhibited one of the hallmark features, which was highly atypical. Upon further investigation, we found some phenotypes were omitted from the data set during transcription. PhenoBlocks could thus also be helpful in validating new data against existing examples.

7 Future Work and Conclusions

PhenoBlocks was well received by the clinicians we interviewed and they saw that it could improve their existing diagnosis workflow. We are interested in integrating PhenoBlocks with PhenoTips, as suggested in our feedback. We also want to extend the features of PhenoBlocks to also account for additional patient features such as age, sex, and ethnicity. These would be additional inputs that could personalize the importance score for a given patient, e.g., an adult-onset phenotype could be given less weight when the query patient is a child. Lastly, we are interested in a longitudinal study with clinicians, to investigate how their workflows are affected when given access to the tool in a real-world usage scenario.

PhenoBlocks is a novel visual analytics tool to aid clinicians during differential diagnosis of rare genetic disorders. A user-centered approach was used and the informal evaluation suggested that visually comparing new patients to diagnosed cases can lead to insights not currently supported by existing text-only tools. These visual analyses can also help clinicians to quickly generate and evaluate hypotheses, as well as prioritize clinical investigations with patients. Potential also exists to use PhenoBlocks in a genetics research to visualize phenotypic variation between patients of a gene variant cohort. The visualization design and visual encoding used by PhenoBlocks could also be applied more generally to non-genetic clinical diagnosis.

Open-Source Release

PhenoBlocks will be released as a publically accessible tool and the source code will be made available on GitHub:
www.phenoblocks.org
www.github.com/mgdgp/phenoblocks

References

[1] Annotation Guide, Human Phenotype Ontology. Accessed March 2015. http://www.human-phenotype-ontology.org/contao/index.php/annotation-guide.html

[2] Arndt, T. (2008). Visual software tools for bioinformatics. *JVLC*, *19*(2), 291-301.

[3] Arocha, J. F., Patel, V. L., & Patel, Y. C. (1993). Hypothesis generation and the coordination of theory and evidence in novice diagnostic reasoning. *Medical Decision Making*, 13(3), 198-211.

[4] Baynam, G., Walters, M., Claes, P.,... & Goldblatt, J. (2014). Phenotyping: Targeting genotype's rich cousin for diagnosis. *Journal of paediatrics and child health*.

[5] Binder, H., Hopp, L., Lembcke, K. & Wirth, H. (2014) Chapter 15: Personalized disease phenotypes from massive OMICs data. *Big Data Analytics in Bioinformatics and Healthcare*. Med. Info. Sci. Reference.

[6] Bremm, S., von Landesberger, T., Heß, M., Schreck, T., Weil, P., & Hamacherk, K. (2011). Interactive visual comparison of multiple trees. In *Proc. VAST 2011 IEEE*. 31-40.

[7] Chittaro, L., Combi, C., & Trapasso, G. (2003). Data mining on temporal data: a visual approach and its clinical application to hemodialysis. *JVLC*, *14*(6), 591-620.

[8] Clare, A. (2005). Chapter 6: Integration of Genomic and Phenotypic Data. *Data analysis and visualization in genomics and proteomics*. John Wiley.

[9] Denny, J. C., Ritchie, M. D., Basford., ... & Crawford, D. C. (2010). PheWAS: demonstrating the feasibility of a phenome-wide scan to discover gene–disease associations. *Bioinformatics*, *26*(9), 1205-1210.

[10] Elstein, A. S., Shulman, L. S., & Sprafka, S. A. (1978). *Medical problem solving: An analysis of clinical reasoning*. Harvard University Press.

[11] Feltovich, P. J., Johnson, P. E., Moller, J. H., & Swanson, D. B. (1984). LCS: The role and development of medical knowledge in diagnostic expertise. *Readings in medical artificial intelligence*, 275-319.

[12] Ferstay, J. A., Nielsen, C. B., & Munzner, T. (2013). Variant view: Visualizing sequence variants in their context. *TVCG*, *19*(12), 2546-2555.

[13] Fiume, M., Smith, E. J., Brook, A.,… & Brudno, M. (2012). Savant Genome Browser 2: visualization and analysis for population-scale genomics. *Nucleic acids research*, *40*(1), 615-621.

[14] FORGE Canada Consortium. Accessed March 2015. http://www.genomebc.ca/research-programs/projects/health/finding-of-rare-disease-genes-in-canada-forge-canada/

[15] Girdea, M., Dumitriu, S., Fiume, M.,... & Brudno, M. (2013). PhenoTips: patient phenotyping software for clinical and research use. *Human mutation*, 34(8), 1057-1065.

[16] Groves, M., O'Rourke, P., & Alexander, H. (2003). The clinical reasoning characteristics of diagnostic experts. *Medical teacher*, 25(3), 308-313.

[17] Graham, M., & Kennedy, J. (2010). A survey of multiple tree visualisation. *Information Visualization*, *9*(4), 235-252.

[18] Gschwandtner, T., Aigner, W., Kaiser, K., Miksch, S., & Seyfang, A. (2011). CareCruiser: exploring and visualizing plans, events, and effects interactively. In *Proc. PacificVis,* 43-50.

[19] Guerra-Gomez, J., Pack, M. L., Plaisant, C., & Shneiderman, B. (2013). Visualizing change over time using dynamic hierarchies: TreeVersity2 and the StemView. *TVCG*, *19*(12), 2566-2575.

[20] Harrower, M., & Brewer, C. A. (2003). ColorBrewer. org: an online tool for selecting colour schemes for maps. *The Cartographic Journal*, 40(1), 27-37.

[21] Hennekam, R., & Biesecker, L. G. (2012). Next-generation sequencing demands next-generation phenotyping. *Human mutation*, 33(5), 884-886.

[22] Heuer, R. J. (1999). *Psychology of intelligence analysis*. Lulu. com.

[23] Houle, D., Govindaraju, D. R., & Omholt, S. (2010). Phenomics: the next challenge. *Nature Reviews Genetics*, *11*(12), 855-866.

[24] Jones, K. L., Jones, M. C., & Del Campo, M. (2013). *Smith's recognizable patterns of human malformation*. Elsevier Health Sciences.

[25] Josephson, J. R., & Josephson, S. G. (1996). *Abductive inference: Computation, philosophy, technology*. Cambridge University Press.

[26] Kang, Y. A., & Stasko, J. (2011). Characterizing the intelligence analysis process: Informing visual analytics design through a longitudinal field study. In *Proc VAST*, 21-30.

[27] Katifori, A., Halatsis, C., Lepouras, G., Vassilakis, C., & Giannopoulou, E. (2007). Ontology visualization methods—a survey. *CSUR*, 39(4), 10.

[28] Keim, D., Andrienko, G., Fekete, J. D., Görg, C., Kohlhammer, J., & Melançon, G. (2008). *Visual analytics: Definition, process, and challenges*, 154-175. Springer.

[29] Köhler, S., Schulz, M. H., Krawitz, P.,... & Robinson, P. N. (2009). Clinical diagnostics in human genetics with semantic similarity searches in ontologies. *Am J Human Genetics*, 85(4), 457-464.

[30] Köhler, S., Doelken, S. C., Mungall,, ... & Robinson, P. N. (2013). The Human Phenotype Ontology project. *Nucleic acids research*.

[31] Krzywinski, M., Schein, J., Birol, I., ... & Marra, M. A. (2009). Circos: an information aesthetic for comparative genomics. *Genome research*, 19(9), 1639-1645.

[32] Lanktree, M. B., Hassell, R. G., Lahiry, P., & Hegele, R. A. (2010). Phenomics: expanding the role of clinical evaluation in genomic studies. *Journal of Investigative Medicine*, 58(5), 700-706.

[33] Liu, Z., Nersessian, N. J., & Stasko, J. T. (2008). Distributed cognition as a theoretical framework for information visualization. *TVCG,* 14(6),1173-1180.

[34] London Medical Databases. Accessed March 2015. http://www.lmdatabases.com

[35] Lussier, Y. A., & Liu, Y. (2007). Computational approaches to phenotyping: high-throughput phenomics. In *Proc. Am Thoracic Society*, 4(1), 18-25.

[36] Meyer, M., Munzner, T., & Pfister, H. (2009). MizBee: a multiscale synteny browser. *TVCG*, 15(6), 897-904.

[37] Munzner, T., Guimbretière, F., Tasiran, S., Zhang, L., & Zhou, Y. (2003). TreeJuxtaposer: scalable tree comparison using Focus+ Context with guaranteed visibility. *ACM TOG*, 22(3), 453-462.

[38] Nielsen, C. B., Cantor, M., Dubchak, I.,...& Wang, T. (2010). Visualizing genomes: techniques and challenges. *Nature methods*, *7*, S5-S15.

[39] Nielsen, C. B., Jackman, S. D., Birol, I., & Jones, S. J. (2009). ABySS-Explorer: visualizing genome sequence assemblies. *TVCG*, *15*(6), 881-888.

[40] Online Mendelian Inheritance in Man (OMIM). Accessed March 2015. http://www.omim.org/

[41] Pendergrass, S. A., Dudek, S. M., Crawford, D. C., & Ritchie, M. D. (2012). Visually integrating and exploring high throughput Phenome-Wide Association Study results using PheWAS-View. *BioData mining*, *5*(5).

[42] Pirolli, P., & Card, S. (2005). The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis. In *Proc Intl Conf on Intelligence Analysis,* 5, 2-4.

[43] Plaisant, C., Mushlin, R., Snyder, A., Li, J., Heller, D., & Shneiderman, B. (1998). LifeLines: using visualization to enhance navigation and analysis of patient records. In *Proc. AMIA,* 76.

[44] Pictures of Standard Syndromes and Undiagnosed Malformations (POSSUM). Accessed March 2015. http://www.possum.net.au/

[45] Rath, A., Olry, A., Dhombres,… , & Ayme, S. (2012). Representation of rare diseases in health information systems: the Orphanet approach to serve a wide range of end users. *Human mutation*, 33(5), 803-808.

[46] Rind, A., Wang, T. D., Aigner, W.,… & Shneiderman, B. (2011). Interactive information visualization to explore and query electronic health records. *Foundations and Trends in Human-Computer Interaction*, *5*(3), 207-298.

[47] Rind, A., Aigner, W., Miksch, S., Wiltner, S., Pohl, M., Turic, T., & Drexler, F. (2011). *Visual exploration of time-oriented patient data for chronic diseases: Design study and evaluation*, 301-320. Springer.

[48] Robertson, G., Cameron, K., Czerwinski, M., & Robbins, D. (2002). Animated visualization of multiple intersecting hierarchies. *Information Visualization*, 1(1), 50-65.

[49] Robinson, P. N., Köhler, S., Bauer, S., Seelow, D., Horn, D., & Mundlos, S. (2008). The Human Phenotype Ontology: a tool for annotating and analyzing human hereditary disease. *Am J Hum Genetics*, 83(5), 610-615.

[50] Robinson, P. N., & Mundlos, S. (2010). The human phenotype ontology. *Clinical genetics*, 77(6), 525-534.

[51] Robinson, Peter N. (2012). Deep phenotyping for precision medicine. *Human mutation*, 33(5), 777-780.

[52] Secrier, M., Pavlopoulos, G. A., Aerts, J. & Schneider, R. (2012). Arena3D: visualizing time-driven phenotypic differences in biological systems. *Bioinformatics* 13(45).

[53] Shneiderman, B., Plaisant, C., & Hesse, B. W. (2013). Improving healthcare with interactive visualization. *Computer*, (5), 58-66.

[54] Stasko, J., & Zhang, E. (2000). Focus+ context display and navigation techniques for enhancing radial, space-filling hierarchy visualizations. In *Proc. Infovis,* 57-65.

[55] Tu, Y. & Shen, H. W. (2007) Visualizing changes of hierarchical data using treemaps. *TVCG,* 13, 1286-1293.

[56] van Ham, F. (2003) Using multilevel call matrices in large software projects. In *Proc. Infovis,* 227-232.

[57] Wang, T. D., Plaisant, C., Quinn, A. J., Stanchak, R., Murphy, S., & Shneiderman, B. (2008). Aligning temporal data by sentinel events: discovering patterns in electronic health records. In *Proc. CHI,* 457-466.

[58] Weber, E. U., Böckenholt, U., Hilton, D. J., & Wallace, B. (1993). Determinants of diagnostic hypothesis generation: effects of information, base rates, and experience. *J of Exp Psych*, 19(5), 1151.

[59] Wongsuphasawat, K., Guerra Gómez, J. A., Plaisant, C., Wang, T. D., Taieb-Maimon, M., & Shneiderman, B. (2011). LifeFlow: visualizing an overview of event sequences. In *Proc CHI*, 1747-1756.

[60] Zhao, J., Chevalier, F., Collins, C., & Balakrishnan, R. (2012). Facilitating discourse analysis with interactive visualization. *TVCG*, *18*(12), 2639-2648.