PhenoTips: Patient Phenotyping Software for Clinical and Research Use

Marta Girdea, Sergiu Dumitriu, Marc Fiume, Sarah Bowdin, Kym M. Boycott, Sébastien Chénier, David Chitayat, Peter N. Ray, Michael Brudno, Joyce So, and Peter N. Robinson

Department of Computer Science, University of Toronto, Ontario Canada; Centre for Computational Medicine, Hospital for Sick Children, Toronto, Ontario Canada; Department of Paediatrics, University of Toronto, Ontario Canada, Division of Clinical and Metabolic Genetics, Hospital for Sick Children, Toronto, Ontario Canada; Children’s Hospital of Eastern Ontario Research Institute, University of Ottawa, Ottawa, Ontario Canada; Molecular Genetics Laboratory, Hospital for Sick Children, Toronto, Ontario Canada; The Prenatal Diagnosis and Medical Genetics Program, Department of Obstetrics and Gynecology, Mount Sinai Hospital, Toronto, Ontario Canada; Mount Sinai Hospital, University of Toronto, Toronto, Ontario Canada; University Health Network, University of Toronto, Toronto, Ontario Canada; Genetics and Genome Biology Program, Hospital for Sick Children, Toronto, Ontario Canada; Centre for Addiction and Mental Health, Toronto, Ontario Canada

Communicated by Peter N. Robinson

Received 18 January 2013; accepted revised manuscript 18 April 2013. Published online 1 May 2013 in Wiley Online Library (www.wiley.com/humanmutation). DOI: 10.1002/humu.22347

ABSTRACT: We have developed PhenoTips: open source software for collecting and analyzing phenotypic information for patients with genetic disorders. Our software combines an easy-to-use interface, compatible with any device that runs a Web browser, with a standardized database back end. The PhenoTips’ user interface closely mirrors clinician workflows so as to facilitate the recording of observations made during the patient encounter. Collected data include demographics, medical history, family history, physical and laboratory measurements, physical findings, and additional notes. Phenotypic information is represented using the Human Phenotype Ontology; however, the complexity of the ontology is hidden behind a user interface, which combines simple selection of common phenotypes with error-tolerant, predictive search of the entire ontology. PhenoTips supports accurate diagnosis by analyzing the entered data, then suggesting additional clinical investigations and providing Online Mendelian Inheritance in Man (OMIM) links to likely disorders. By collecting, classifying, and analyzing phenotypic information during the patient encounter, PhenoTips allows for streamlining of clinic workflow, efficient data entry, improved diagnosis, standardization of collected patient phenotypes, and sharing of anonymized patient phenotype data for the study of rare disorders. Our source code and a demo version of PhenoTips are available at http://phenotyps.org.

Hum Mutat 34:1057–1065, 2013. © 2013 Wiley Periodicals, Inc.

KEYWORDS: phenotype; clinical genetics; ontologies; data standardization

Introduction

Perhaps more than any other field, the availability of low-cost genomic sequencing has revolutionized the practice of Clinical Genetics. Whole-exome sequencing has allowed for the discovery of the genetic cause for over 100 Mendelian disorders [Rabbani et al., 2012], opening up new avenues to diagnostics and treatment. One of the most successful methods for identifying the causes of Mendelian conditions is the comparison of the genomes of unrelated individuals. However, for many rare disorders, identifying multiple unrelated patients with similar manifestations is a non-trivial task. Most such patients will be seen by different clinicians at different hospitals, and the clinician working with one family will typically be unaware of other cases. The key to the identification of such cases is the sharing of patient data among hospitals. For this to be effective, patient phenotypes must be recorded using a standardized vocabulary, enabling automated comparison of patients’ data across institutions. Unfortunately, depending on the hospital, only a small fraction of patient information (e.g., height, weight, blood biochemistry) is stored in standard formats, whereas most phenotypic descriptions are typically recorded as free-form notes. This is especially true within a Genetics Clinic, where a doctor often deals with disorders of multiple systems, and relies on observation of physical findings and patient dysmorphisms, as much as on precise measurements and laboratory tests.

The need for a standardized phenotyping database emerged during our work with the Molecular Diagnostic Laboratory (MDL) at the Hospital for Sick Children (Toronto). Our primary goal was developing computational methodologies for correlating phenotype and genotype information; however, the phenotyping information provided for each patient was not suited for automated analysis. This information was recorded as free-form clinician notes, often with typos, inconsistent abbreviations, or multiple wordings for the same phenotype. Several such examples extracted from this database are given in Table 1. Furthermore, there was commonly a lack of precision in the reported phenotypes. For example, although “Developmental delay” was reported for many of the patients, there was no information on whether the delay was in gross motor skills, fine motor skills, language, or was global. In a parallel study, our colleagues were reviewing the full medical records of MDL patients to refine the phenotypes; however, they had no software to record their results using a standardized vocabulary.
Table 1. Examples of Different Texts Designating the Same Phenotype, Found in the Dataset of Patient Reports Provided by the MDL at the Hospital for Sick Children

| Actual phenotype                      | Text variations                  |
|---------------------------------------|----------------------------------|
| Behavioral problems                   | behavioral problem               |
|                                       | behavioural problems             |
|                                       | behaviour problem                |
| Congenital heart defect               | congenital heart                 |
| Developmental delay                   | dd                               |
| Learning problems                     | learning delay                   |
| Mental retardation                    | mental retard                    |
| Lists of phenotypes with inconsistent separators |
|                                       | dd, cong. malfor, behav. pro.    |
|                                       | dd, mental retardation           |
|                                       | dd, delayed puberty              |
|                                       | dd, df mr                        |

The phenotypic descriptions extracted from these patient reports often contain typos, inconsistent abbreviations, or multiple wordings for the same phenotype.

The recording of detailed and standardized phenotypes for patients displaying a broad variety of clinical features requires an ontology, which has a rich vocabulary, with clear semantic relationships between the terms, to allow for the identification of similar (yet not identical) findings. Although the London Dysmorphology Database presented one of the first efforts to organize phenotypes typically seen by a Clinical Geneticist, currently the Human Phenotype Ontology (HPO) [Robinson and Mundlos, 2010] is the most complete vocabulary available for recording such patient data. However, the broad use of complex ontologies such as HPO is hindered by their size and complexity. The HPO has almost 10,000 terms, and only few of them are relevant for a specific patient. Although software to search for HPO terms exists, for example, PhenExplorer [Robinson and Mundlos, 2010], these typically do not allow for the selection and persistent storage of ontology terms for specific patients. The PhenoloDB application [Hamosh et al., 2013], developed in parallel and independently of our work, does support patient data collection, but is specifically designed for research studies rather than clinical use. Hence, it does not provide features that match the clinical examination workflow, such as measurement evaluation or diagnosis assistance.

In this report, we present PhenoTips, a deep phenotyping tool and database, specifically designed for phenotyping patients with genetic disorders. Because the best way to ensure accurate and detailed patient data is to record the phenotypes during the patient encounter and clinical examination, PhenoTips closely follows the clinician’s workflow. In addition to an easy-to-use interface for collecting phenotypic descriptions using a standardized vocabulary (HPO), our software provides a series of features that help reduce the clinician’s workload during the clinical examination, and facilitates the safe sharing of de-identified data among medical institutions. In addition to phenotypic data, PhenoTips supports entering demographic information, medical history (including prior laboratory results), family history, standard measurements, relevant images depicting manifestations of the patient’s disorder, genetic tests and their results, as well as additional notes for each of these categories. The software automatically plots growth curves for a variety of measurements, selects phenotypes reflecting abnormal measurements, instantly finds Online Mendelian Inheritance in Man (OMIM) diseases that most closely match the phenotypic description, and suggests additional clinical features whose evaluation could contribute to a more accurate diagnosis. Observations can be recorded directly during the patient encounter, and the interface is compatible with any device that runs a modern Web browser (including mobile devices such as tablets). PhenoTips is already used both in research studies and in the clinic, including the phenotyping of patients for the FORGE (Finding Of Rare disease Genes) Canada project (http://care4rare.ca/), and is freely available (open source) at http://phenotips.org.

**PhenoTips Features**

PhenoTips provides an easy-to-use Web interface and standardized database back-end for collecting clinical findings observed in patients with possible genetic disorders. The main goals of this software are (1) to efficiently capture patient data in standard formats, facilitate effortless exchange of anonymized data, and enable automated searches in annotated gene and disease databases; and (2) to provide advanced functionalities and a friendly user interface that help reduce the clinician’s workload, permitting seamless use of this application within the clinician’s routine.

**Building and Managing a Database of Phenotypic Profiles**

PhenoTips helps building a database of phenotypic profiles of patients with genetic disorders. The information recorded for each patient can be subsequently modified, and all versions of the examination notes remain available for comparison. The database can be fully or selectively exported in a comma-separated values file, compatible with Microsoft Excel, for external processing. Although it could be used as an extension to a medical records system, it is intended as a knowledge base of phenotypic descriptions where multiple patients with similar indications can be identified. In addition to simple browsing of the existing patient records, the software allows users to filter the data by various criteria, including phenotype or disorder name.

The software supports several different user roles and levels of access, which can be enabled or disabled depending on the purpose of each PhenoTips installation. Installations intended for clinical use or for private research projects will allow access to a specific group of users, and authenticated users can have multiple roles: viewer (i.e., can browse the data without the ability to modify it), contributor (can contribute data and modify their own contributions), data administrator (full access to the entire data, together with the ability to modify or remove it), and server administrator (full access to the entire data, with the ability to modify other users’ permissions).
Matching the Clinical Examination Routine

To better address the needs of the professionals collecting the information, PhenoTips supports the collection of most types of patient data typically acquired during a comprehensive examination in a Genetics clinic. The user interface contains multiple sections corresponding to the steps of the examination (see Fig. 1). Although some sections allow the clinician to record free-form notes, or upload of external files (e.g., test reports or medical imaging) into the database, others enable for the recording of detailed patient data in standardized formats, and reduce clinician workload for manual and time consuming steps. A typical PhenoTips Web form intended for a clinical examination report contains the following sections:

- **Demographic information**: At a minimum, the patient’s date of birth (which is limited to only month and year if the subject is to remain anonymous) and the patient’s gender should be included. For clinical use, we allow for the storage of name, exact date of birth, and other identifiable data such as health card number. These are not displayed for research studies.
- **Family history**: Here, the user lists known disorders identified in family members and uploading pedigrees. A pedigree drawing tool that will be fully integrated with PhenoTips is in development.
- **Medical history**: Health issues known prior to the clinical examination are entered here. PhenoTips also allows the user to upload PDF files with previous medical records and test results.
- **Measurements**: The software supports a comprehensive set of entries, from general measurements (i.e., weight, height, and head circumference) to more specific anthropometric data (e.g., interpupillary distance and ear length). Multiple rounds of measurements taken at different examination dates can be recorded.
- **Clinical symptoms and physical findings**: The clinician can select standardized phenotypes, which are mapped to a controlled vocabulary. This is discussed in detail in section Collection of Standardized Phenotypic Data.
- **Diagnosis**: Here, the software presents in real time the results of automated searches in OMIM of disorders that best match the current phenotype selection.
- **Genetic testing**: This is a summary of disease-specific genetic tests performed on this patient, variants found to be related to the observed manifestations, as well as comments on their results.

In the two subsequent sections, we discuss PhenoTips’ ability to collect standardized patient data directly during the patient examination (section Collection of Standardized Phenotypic Data),
and several features meant to simplify and clinical workflows and reduce the time required for common manual tasks (section *Simplifying Clinical Workflows*).

**Collection of Standardized Phenotypic Data**

Among the many existing controlled vocabularies providing phenotypic descriptions, we chose to standardize phenotypic data using the HPO. The HPO provides a well-maintained, comprehensive hierarchy of approximately 10,000 standard terms, most of them with detailed definitions and alternative names. The HPO is used in a variety of additional resources, such as phenotype annotations [Kohler et al., 2009; Robinson et al., 2008] of OMIM [Hamosh et al., 2005] disorders and implicated genes.

We facilitate the adoption of HPO by clinicians for patient phenotypic descriptions by providing a user interface, which can ensure easy access to the terms relevant for each specific patient. The interface combines a list of predefined options corresponding to the most commonly encountered phenotypes and a search box for any phenotypic description not mentioned in the predefined list.

**Predefined phenotype options**

The predefined phenotype list is easily configurable, allowing the user to capture the most relevant findings for a specific type of examination. Each of the predefined terms with subterms within the HPO can be expanded to allow for the selection of the more specific phenotypes. An example is given in Figure 2. The phenotypes are organized in subsections or categories (also customizable). Each subsection includes its own phenotype search box, enabling a phenotype search narrowed down to that category of phenotypes (e.g., search only for cardiac abnormalities).

**Selecting any standard HPO term**

The phenotype search box offers the ability to search and browse the ontology to compose the phenotypic description of a patient, somewhat similar to PhenExplorer [Robinson and Mundlos, 2010]. In our case, however, the data associated with a patient is persistent, and the search is tolerant to spelling errors. The phenotype search in PhenoTips is handled by Solr (http://lucene.apache.org/solr/), a powerful search platform, which can handle complex queries and has extensive support for spelling errors, abbreviations, and phonetic searches. The entire HPO (including semantic relations between its terms) is indexed in Solr, allowing PhenoTips to instantly provide suggestions of standard terms matching the query (Fig. 3). Synonyms annotated within HPO for the same phenotype (e.g., “Mental retardation” and “Intellectual disability”) will both point the user to the same standard term (in this case, HP:0001249). This overcomes inconsistent terminology issues that often come up when using free-form notes. The user can also browse the ontology for terms closely related to the ones suggested. Once a term is selected, its unique identifier in HPO is recorded as part of the patient’s phenotypic description (e.g., HP:0000369 for “low-set ears”). However, the user only sees and interacts with human-readable text (the name of this phenotype, its definition, etc.), which ensures a familiar terminology for medical specialists and hides the technical complexities of the underlying ontology.

**Capturing the relevance of absent phenotypes**

One of the issues initially identified by our users was that a list of checkboxes does not enable the marking of a phenotype as “investigated, but not found.” Although the majority of possible phenotypes that are not present in a patient are irrelevant, some patient descriptions can be considered incomplete without mentioning disorder manifestations, which are expected, yet not observed, and which can be relevant for diagnostics. Consequently, instead of using simple checkboxes, we offer the possibility to mark a phenotype as either present (Y), absent (N), or unexplored (NA, selected by default) in each of the phenotype selection modes (Figs. 2 and 3).

**Handling phenotypic descriptions not present in HPO**

Although HPO is a substantial collection of phenotypic descriptions, it still has insufficient details in some categories (e.g., oncology and prenatal markers). Our software tries to enforce the use of standard terms for the patient’s phenotypic profile, but at the same time, it must avoid limiting the user to only phenotypes present in HPO. Within the phenotype search box, after presenting suggestions from HPO that best match the user’s input, PhenoTips also allows the user to select and save the free-form text actually entered as the phenotypic description. From the user’s perspective, this feature offers increased flexibility, and enables building a complete phenotypic profile. From a data coordination perspective, it forces the user to consider standard alternatives before selecting a nonstandard term, and allows for the monitoring of phenotype entries that are not mapped to HPO for further standardization and expansion of the ontology (as explained in section *Improving the HPO Based on PhenoTips Usage*).

**Advanced information about each phenotypic description**

Often, the standard HPO term alone does not provide sufficient detail, and the clinician may want to supplement it with information regarding the age of onset of the observed abnormality, its pace of progression, as well as additional comments and test reports or medical images supporting the presence of the phenotype. As shown in Figure 2, these optional data can be specified together with each of the selected phenotypic descriptions. Although the primary purpose of this information is to support a more comprehensive clinical examination summary, the age of onset and the pace of progression are expressed in standardized terms as defined by the HPO, also making them useful for automated diagnosis inference methods.

**Simplifying Clinical Workflows**

Although the core reason for developing PhenoTips was to facilitate the efficient and accurate collection of patient phenotype data, we introduced additional functionality to maximize the utility of PhenoTips and encourage clinicians to use the software by increasing the accuracy and speed of tasks that clinicians typically perform manually. These were identified through numerous interviews with clinicians as time-consuming steps in their workflows.

**Automated plotting and interpretation of measurement values**

A clinical examination often involves taking certain quantitative measurements to determine whether the patient’s
Figure 2. Example of predefined phenotypes list in the PhenoTips form, with easily configurable phenotype categories, each consisting of a group of options mapped to standard HPO terms, and a summary view of the current phenotype selection. The user can also associate to each selected phenotypic description additional data such as the onset of the abnormality, the pace of progression, medical imaging or test reports, which support the presence of that phenotype, and free-form comments on the manifestation.
Figure 3. Standardized phenotype search in PhenoTips. The quick search box enables text searches with tolerance for spelling errors. It provides suggestions of terms from HPO that best match the entered keywords, and permits to broaden the search with other terms related to the given suggestions. This feature allows the user to browse the entire ontology of human phenotypes by navigating according to their semantic relations.

physical development falls within normal parameters. To make this evaluation, clinicians must compute percentiles or standard deviations from mean and update growth curves for each measurement over several exams, a time-consuming process that lends itself to automation.

In PhenoTips, we developed functionality to do these computations automatically for a number of measurements currently supported by our form: weight, height, head circumference, sitting height, arm span, inner and outer canthal distance, palpebral fissure length, interpupillary distance, philtrum length, ear length, hand, palm, and foot lengths (Fig. 4). Adding other measurements is straightforward. Using the patient’s age at the time of measurement, which is obtained from the measurement date and the date of birth, the values entered for these measurements are interpreted based on World Health Organization and Center for Disease Control and Prevention standard measurements, as well as other published sources [Blais et al., 1956; Chouke, 1929; Farkas, 1981; Feingold and Bossert, 1974; Laestadius et al., 1969; Tanner and Whitehouse, 1978; Thomas et al., 1987; Zankl et al., 2002]. We also instantly plot growth curves, percentiles, and Z-scores (number of standard deviations from the mean) for all of the measurements entered at different ages, replacing a time-consuming manual step, and enabling the clinician to immediately observe if certain measurements are slightly or significantly abnormal. For all the measurements, extreme percentile values (top or bottom percentile or Z-score) trigger the automatic selection of corresponding phenotypes (e.g., microcephaly for a head circumference in the first percentile, or hypertelorism for interpupillary distance 2 standard deviations above the mean). This ensures consistency in the specification of these phenotypes, improving data accuracy.

Automated phenotype suggestions

Although a clinical genetics assessment typically involves a full patient examination, specific symptoms or phenotypes that are critical for a diagnosis may still be missed by the clinician. To reduce such occurrences, PhenoTips automatically identifies additional phenotypic features that are associated with those previously selected, and would further improve the depth of the phenotypic description. With every change in the phenotypic description of a patient, all of the remaining unselected phenotypic features are ranked by a relevance score, computed so that features that are specific (occur in fewer OMIM disorders) and likely (best correlate with already selected phenotypes) rank highly, as explained in the Supp. Methods. The top ranking phenotypic description based on this score, together with the most likely disorder(s), is presented to the clinician as an additional list of checkboxes, thus allowing the clinician to carefully consider this additional information in the assessment.

Automated diagnostics assistance

During our interviews with clinicians, they identified the search for the set of observed phenotypes in OMIM as one of the most time-consuming steps in their analysis. To reduce the time spent on this task, we implemented within PhenoTips an algorithm to identify
The Measurements section of the PhenoTips Web form permits to record standard measures taken during clinical exams: using the patient’s gender and the age at the time of measurement (which is computed based on the recorded date of birth), these parameters are automatically translated into quantiles. Growth curves are plotted instantly based on entered values. Abnormal values are automatically reported and the corresponding phenotypes are selected (e.g., microcephaly for an abnormally small-head circumference), though the clinician may choose to modify these later.

**Figure 4.** The Measurements section of the PhenoTips Web form permits to record standard measures taken during clinical exams: using the patient’s gender and the age at the time of measurement (which is computed based on the recorded date of birth), these parameters are automatically translated into quantiles. Growth curves are plotted instantly based on entered values. Abnormal values are automatically reported and the corresponding phenotypes are selected (e.g., microcephaly for an abnormally small-head circumference), though the clinician may choose to modify these later.
OMIM disorders that best match the observed phenotypic features. We report to the clinician a ranked list of 20 disorders based on a similarity score with respect to the selected features, allowing them to explore candidate disorders and finally select one or more of these as the diagnosis where appropriate. We emphasize here that the goal of our algorithm is not to “diagnose” the patient, but rather to give to the clinician a list of potential disorders for consideration.

Our algorithm is based on information retrieval concepts, which have been previously proposed for clinical diagnostics using semantic similarity searches in HPO [Kohler et al., 2009]; however, our approach differs in several ways:

- it accounts for disorder frequencies in the general population according to Orphanet [Rath et al., 2012];
- it supports negative phenotypes, that is, relevant symptoms that were not observed in the patient;
- it can handle free-form text inputs entered by the clinician, in addition to the standard HPO term identifiers, when searching for matches.

This functionality relies on the availability of a knowledge base consisting of OMIM disorders together with their frequencies in the general population, their clinical features, as well as features known not to occur in this particular disorder (hereafter called negative phenotypic features). The frequencies are obtained from Orphanet based on the available mapping between Orphanet and OMIM disorder identifiers [Rath et al., 2012]. The disorder-specific clinical features and negative phenotypic features are retrieved from the mapping between OMIM and HPO developed by the HPO team [Robinson et al., 2008].

Every change in the patient’s phenotypic description triggers a search for the matching diagnoses. Each disorder is assigned a relevance score [Manning et al., 2008], which ensures higher relevance scores for disorders that: (1) have a higher frequency in the general population; (2) match more terms/keywords from the query; (3) are annotated with very specific symptoms/keywords (which do not occur often in the annotations of other disorders); and (4) do not have many more symptoms in addition to those from the query. The diagnosis prediction method in PhenoTips relies on the Solr engine, similar to the phenotype keyword search in HPO discussed in section Collection of Standardized Phenotypic Data; we built upon Solr’s search functionality and calibrated it for this specific task. Additional details on the ranking algorithm are given in the Supp. Methods.

The diagnosis suggestion method used in PhenoTips is loosely coupled with the rest of the software, and can be easily replaced by another method with the same inputs (HPO terms and free-form text phenotypes) and outputs (OMIM diagnosis suggestions). In earlier versions of PhenoTips, we implemented the diagnosis suggestion method of Kohler et al. [2009], which remains included in our source code as an alternative. One can also implement other approaches, for example, the Bayesian ontology querying method proposed in Bauer et al. [2012].

PhenoTips Deployments and Use Cases

Current Deployments of PhenoTips

Research deployments

PhenoTips is currently used in three research projects as a secure system for collecting anonymized patient information, including a project by the MDL at SickKids to collect detailed patient phenotype data for individuals who are undergoing whole-genome microarrays, the FORGE Canada project (http://care4rare.ca/) to identify genes responsible for a wide spectrum of rare pediatric-onset disorders, and a project by clinicians at Mount Sinai Hospital and the Centre for Addiction and Mental Health (CAMH) for gathering dysmorphism data and measurements for adults with neuropsychiatric disorders and developmental or other systemic comorbidities. Although in the first two projects the data is entered after the patient encounter, in the third one the clinician enters the data directly into PhenoTips while examining the patient for the study. The Web form used for collecting the data has been tailored to each project or user/group of users within the project, including the degree of detail for the demographic information, the list of predefined options for most common phenotypes, types of standard measurements, and whether dysmorphism photos or additional laboratory reports are being stored within the system.

Within these research projects, each PhenoTips installation is expected to lead to the accumulation of a rich knowledge base of deeply phenotyped patients with specific disorders, which can eventually improve the understanding of disorder manifestations and serve as a resource of data against which one can compare and evaluate new, undiagnosed patients. Among the three research-based PhenoTips installations, approximately 250 cases have been phenotyped to date.

Clinical deployment

In addition to our three research-based installations, where PhenoTips is being used to record patient data after the patient encounter has concluded, it is also used at the Hospital for Sick Children and within NIH Undiagnosed Diseases Program for collecting complete demographic and phenotypic data for patients seen by a Clinical Geneticist during their patient encounter. In this context, the ability of the software to automate calculation of growth charts and suggest clinical diagnoses based on the current phenotype selection reduces the clinician’s effort and eliminates the need for redundant data entry.

Improving the HPO Based on PhenoTips Usage

As data are being collected in different institutions dealing with multiple categories of patients, our users may encounter phenotypes that are not currently present in the HPO. During our demonstrations and discussions with medical professionals of different specialties, our audience has occasionally noticed that HPO does not fully cover some categories, such as prenatal or oncology-related phenotypes. To address this shortcoming, our Web forms allow users to enter free-form text describing the observed abnormality whenever they are unable to find a standard term matching their observations. These nonstandard entries will be collected and regularly reported to the HPO team. Once their inclusion in the ontology is validated, the free-text entries stored in our databases will be remapped to the newly added standard terms. With these steps, we avoid constraining our users to a predefined set of terms, and at the same time, we gather feedback that can improve HPO over time and benefit many other users of this ontology.

Conclusions and Future Directions

Here, we present PhenoTips, an easy-to-use Web interface and standardized database back end for collecting clinical symptoms and physical findings observed in patients with genetic disorders.
Our development process was guided by interviews with Clinical Geneticsists and a number of other specialists from major hospitals across Canada (Hospital for Sick Children, Toronto; Mount Sinai Hospital, Toronto; Children’s Hospital for Eastern Ontario, Ottawa; Centre for Addiction and Mental Health, Toronto; University Health Networks, Toronto), as well as the FORGE Canada Consortium steering committee. PhenoTips is designed to be used in real time during the clinical examination, and provides an interface for quickly recording observed phenotypes in a standard vocabulary. In addition to allowing the collection of patient data in standard formats, it provides advanced functionalities that help reduce the clinician’s workload, permitting seamless use of this application within the clinician’s routine. The software is freely available (http://phenotips.org/), and it can be downloaded and used without restrictions either for research studies or in clinical settings.

Comparing the genomes of unrelated individuals is one of the most successful approaches to identify novel genes causing Mendelian disorders. The key to the identification of such individuals, especially when dealing with very rare disorders, is the sharing of accurate patient data among hospitals. PhenoTips can significantly simplify such exchange by allowing the collection of data in a standard, machine-readable format. Although the direct sharing of patient records is impossible due to privacy concerns, PhenoTips can create de-identified patient records by exporting only select, nonidentifiable fields. Thus, data from multiple installations can be unified within a central repository, allowing for the identification of other patients or families with specific disorders.

For patients with rare disorders, effective de-identification is con-

founded by the patients’ uniqueness; the small number of individuals with a specific diagnosis makes it feasible to reconstruct the identity of a given patient based on casual knowledge of the family, or, for example, dates of clinic visits. Consequently, the sharing must occur with strict limitations on access to the data, and with a sharing model that enables the discovery of patients with specific indications without the ability to directly view full patient information, such as the Cafe Variome technology for genomic data (http://www.cafevariome.org/).

Finally, our limited knowledge of the phenotypic effects of indi-

vidual variants precludes the clinical use of all but a small fraction of the information contained in an individual’s exome or genome. By facilitating multi-institutional efforts to comprehensively catalogue correlations between individual phenotypes and genomic variants, PhenoTips can help address this central challenge in genomic medicine.

Acknowledgments

The development of PhenoTips was possible thanks to the valuable ideas and feedback on our early work from many Clinical Genetics professionals affiliated with the Hospital for Sick Children (SickKids) Toronto, Children’s Hospital of Eastern Ontario (CHEO) Ottawa, the Mount Sinai Hospital Toronto, and the Centre for Addiction and Mental Health (CAMH) Toronto, as well as from the members of the FORGE Canada (Finding Of Rare disease GEnes in Canada) steering committee: Kym Boycott (leader; University of Ottawa), Jan Friedman (co-lead; University of British Columbia), Jacques Michaud (co-lead; Université de Montréal), Francois Bernier (University of Calgary), Michael Brudno (University of Toronto), Bridget Fernandez (Memorial University), Bartha Knoppers (McGill University), Mark Samuels (Université de Montréal), Steve Scherer (University of Toronto).

References

Bauer S, Kohler S, Schulz MH, Robinson PN. 2012. Bayesian ontology querying for accurate and noise-tolerant semantic searches. Bioinformatics 28:2502–

2508.

Blais MM, Green WT, Anderson M. 1956. Lengths of the growing foot. J Bone Joint Surg Am 38:998–1000.

Chouke KS. 1929. The epicantlus or Mongolian fold in Caucasian children. Am J Phys Anthropol 13:255–279.

Farkas, LG. 1981. Anthropometry of the head and face in medicine. New York: Elsevier.

Feingold M, Bossert WH. 1974. Normal values for selected physical parameters: an aid to syndrome delineation. Birth Defects Org Artic Ser 10:1–16.

Hames SH, Scott AF, Amberger JS, Bocchini CA, McKusick VA. 2005. Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. Nucleic Acids Res 33:D314–D317.

Hames SH, Sobirea N, Hoover-Fong J, Sutton VR, Boehm C, Schiettecatte F, Valle D. 2013. PhenosDR: a new web-based tool for the collection, storage, and analysis of phenotypic features. Hum Mutat 34(4):566–571.

Kohler S, Schulz MH, Krawitz P, Bauer S, Dolken S, Ott CE, Mundlos C, Horn D, Mundlos S, Robinson PN. 2009. Clinical diagnostics in human genetics with semantic similarity searches in ontologies. Am J Hum Genet 85:457–

464.

Laestadius ND, Aase JM, Smith DW. 1969. Normal inner canthal and outer orbital dimensions. J Pediatr 74:465–468.

Manning CD, Raghavan P, Schütze H. 2008. Introduction to information retrieval. Cambridge: Cambridge University Press.

Rabbani B, Mahdieh N, Hosomichi K, Nakaoka H, Inoue I. 2012. Next-generation sequencing: impact of exome sequencing in characterizing Mendelian disorders. J Hum Genet 57:621–632.

Rath A, Olry A, Dhombres F, Brandt MM, Urbero B, Ayme S. 2012. Representation of rare diseases in health information systems: the Orphanet approach to serve a wide range of end users. Hum Mutat 33:803–808.

Robinson PN, Kohler 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 Genet 83:610–615.

Robinson PN, Mundlos S. 2010. The Human Phenotype Ontology. Clin Genet 77:525–

534.

Tanner J, Whitehouse R. 1978. Standards for sitting height and subischial leg length from birth to maturity. British children. Ware, Hertfordshire, UK: Castlemead Publications.

Thomas TT, Gaitantzis YA, Frias JL. 1987. Palpebral fissure length from 29 weeks gestation to 14 years. J Pediatr 111:267–268.

Zankl A, Eberle L, Molinari L, Schinzel A. 2002. Growth charts for nose length, nasal protrusion, and philtrum length from birth to 97 years. Am J Med Genet 111:388–

391.