Phenomizer - User Guide

Sebastian Köhler, Peter Krawitz, Peter N. Robinson

Abstract

In this manual we will describe the features of the *Phenomizer* and guide the user through the workflow of using the Phenomizer for the differential diagnosis. In the first section we illustrate how to find and enter the clinical features of a patient and will explain the different methods for performing ontological searches with the Phenomizer. In the second section, we will show how to use the Phenomizer to provide suggestions for continuing the differential diagnostic process by identifying clinical features that can best differentiate among the top candidate differential diagnoses. Finally, we show in the last section how search results may be exported for documentation and/or further analysis.

Contents

1 Disclaimer

The *Phenomizer* is intended to be used by qualified and licensed physicians in order to provide assistance in reaching the correct diagnosis in patients with hereditary diseases and for use as a teaching aid. The *Phenomizer* does not make diagnoses. Rather, it produces a ranked list of possibilities that can be used by physicians as a part of the diagnostic workup. The *Phenomizer* does not contain information about all possible diagnoses or even all possible hereditary diseases. The *Phenomizer* should not be used to make medical decisions without the advice of a physician.

2 How to define the query

The medical terms accessible to the *Phenomizer* are based on the terminology of the Human Phenotype Ontology (HPO)[1]. The *Phenomizer* offers three different approaches to find the appropriate term for a phenotypic abnormality, indicated by the three tabs on the left hand side: *Feature*, *Disease* and *Ontology*.

2.1 Entering clinical data of a patient

Under the *Feature* tab you can search for the appropriate HPO terms by typing part of the name, synonym or HPO-identifier into the search box. This search box provides autocompletion functionality: Possible matches are shown below, while the query is typed in. After pressing the search button, features of the HPO, that best agree with the query, are listed in a table. For every feature displayed in the table you can open a *context menu* by right-clicking on the row. This will offer you further information about the specific HPO term (Figure 1). An HPO term can be added to the *Patient's features* list by double clicking, drag-and-drop or via the context menu. In our example we selected the phenotypic features *short stature*, *low posterior hair line*, and *downward slanting palpebral fissures*.
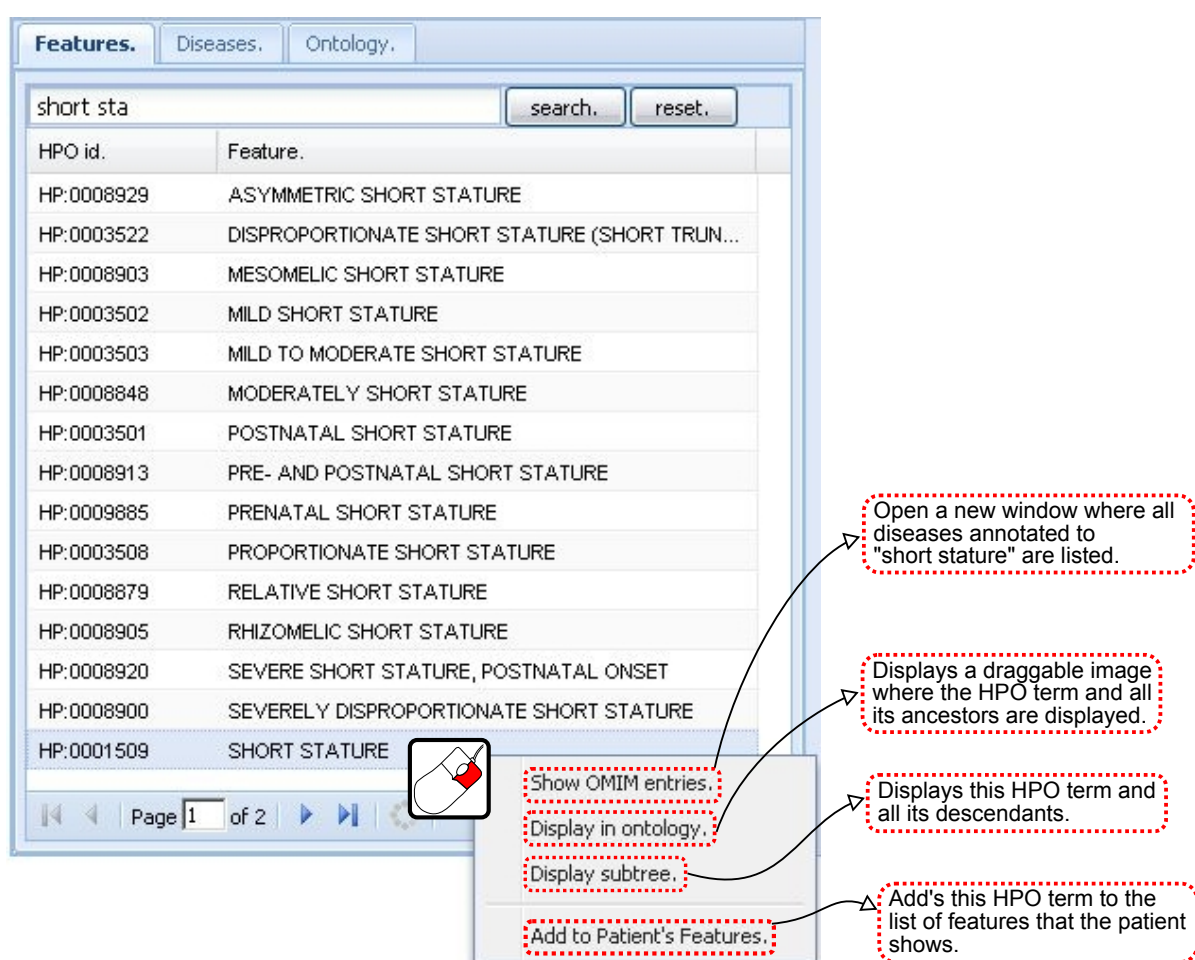

Figure 1: The context menu of an HPO-term allows you to add the feature to the list of *Patient's features*. You may also gather further information about the term, by listing all OMIM diseases that are annotated with this feature or by displaying the feature in the context of the structure of the Human Phenotype Ontology.

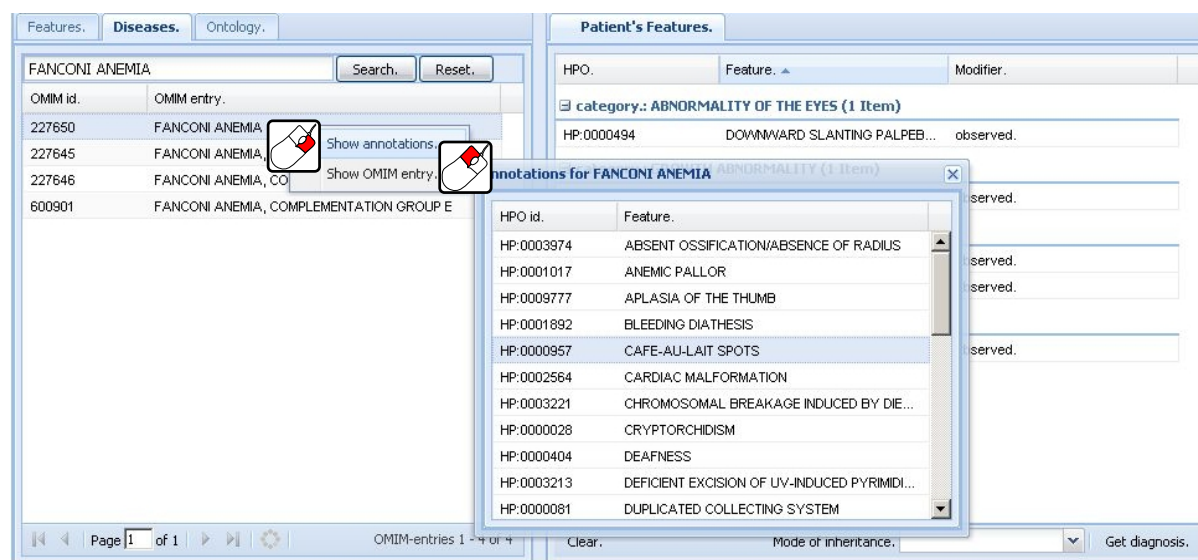[1]http://www.human-phenotype-ontology.org

Figure 2: You can use the search box under the *disease* tab to look for syndromes and display their annotated phenotypic features. Features that are relevant in your case can be added to the *Patient's features*.

If your patient shares some phenotypic abnormalities with a different genetic syndrome you are familiar with, the *Disease* tab offers a quick way to select appropriate features. Using this tab you can use the search box to search for syndromes by entering a part of their name or identifier. The *context menu* (right-click) shows a list of all the features used to describe the syndrome. You can also use the hyperlink to open a new window with the OMIM entry for this syndrome. Appropriate phenotypic features can now be added to the *Patient's features* by double clicking or drag and drop (Figure 2). In the example we have selected *Cafe au-lait spots*, a clinical feature in *Fanconi anemia* [MIM 227650].

You may also browse through the HPO under the *Ontology* tab to search for the most specific term that describes your patient's phenotypic feature. In an ontology child terms have the same or higher information content compared to their parents and will thus make a query more specific. Again, every feature has its own context menu and can be added to the list of the *Patient's features* as described above. In our example we chose *hypertelorism* as a precise description of the *abnormality of the eyes* found in our patient (Figure 3).

2.2   Patient's Features

The features that have been selected as above are now listed on the right-hand side of the Phenomizer window. In general, it is best to enter as many of the clinical features of a patient as possible in order to improve the specificity of the differential diagnoses proposed by the Phenomizer. However, it is also possible to add features one by one to explore the phenotypic space or to learn about the diseases that are characterized by sets of phenotypic features.

2.2.1   Remove features

The *Patient's features* are listed on the right hand side. To delete a feature from this list, right-click on the particular row and select 'remove' from the context-menu. To delete all features from the list you can use 'clear'-button below the table.

2.2.2   Refine query

You can choose for every feature whether it should be regarded as 'observed' or 'mandatory'. In the default setting ('observed') the similarities between the patient and particular syndromes are calculated as described in the paper. If a term is denoted as 'mandatory' all syndromes not having this particular feature (or a descendent thereof) will not be regarded as a possible differential diagnosis and are thus filtered out. You can also select the *mode of inheritance* for the syndrome observed in your patient if it is suspected because of the pedigree. Again, all syndromes that do not exhibit this mode of inheritance are not regarded as a possible differential diagnoses and are filtered out.
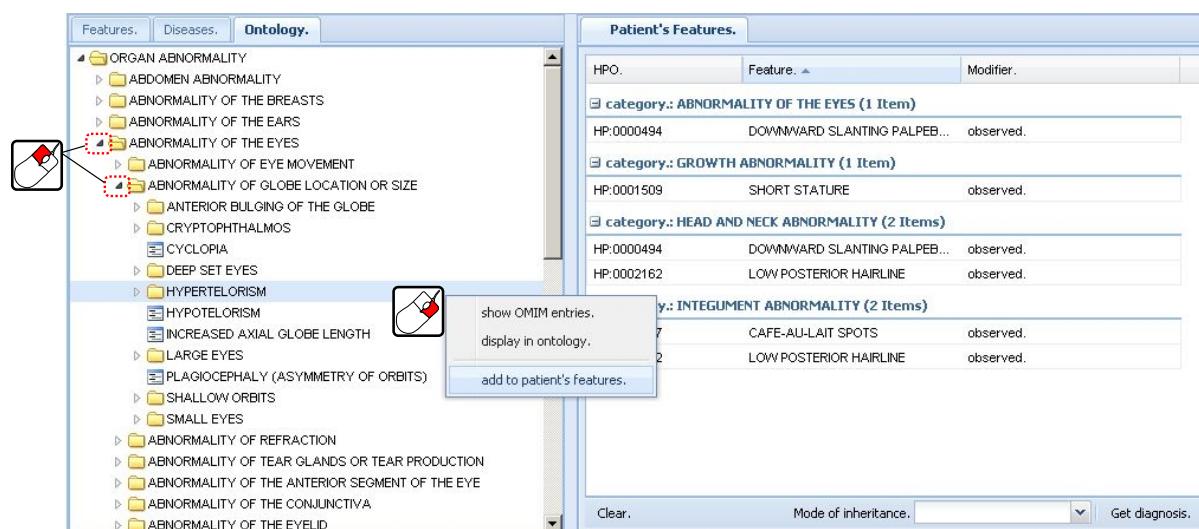
Figure 3: You can browse through the Human Phenotype Ontology in a top-down way to find the most specific term describing your patient's phenotypic feature.

3   Get differential diagnoses

3.1   Settings

Phenotypic similarity can be calculated in a one-sided (asymmetric) or two-sided (symmetric) fashion (see section 3.4). In general, we recommend starting with an asymmetric search, which is the default. However, using symmetric searches may better distinguish between diseases with few and many clinical features. This can be done using the settings menu on the upper-left.

You may also select a particular *multiple testing procedure* from the menu. The default setting is the Benjamini-Hochberg procedure, which is recommended for normal users.

3.2   The first differential diagnosis

Click the 'Get diagnosis' button to cause the *Phenomizer* to compute a list of differential diagnoses that are ranked by their P-value (Figure 4). A significant P-value does not mean that the diagnosis is confirmed. We suggest that it can be taken to mean that the diagnosis is at least plausible and that the physician should consider statistically significant differential diagnoses carefully. On the other hand, if no disease is found to have a significant P-value for the features entered by the physician, we would interpret this to mean that the combination of clinical features is not per se sufficient to make a diagnosis. It may mean that the disease which the patient has is not in the database being used by the program (which currently encompasses the mainly Mendelian diseases listed in OMIM). Physicians may then want to use the Improve differential diagnosis function to obtain a list of clinical features that may help with the further differential diagnostic process (see section 3.3).
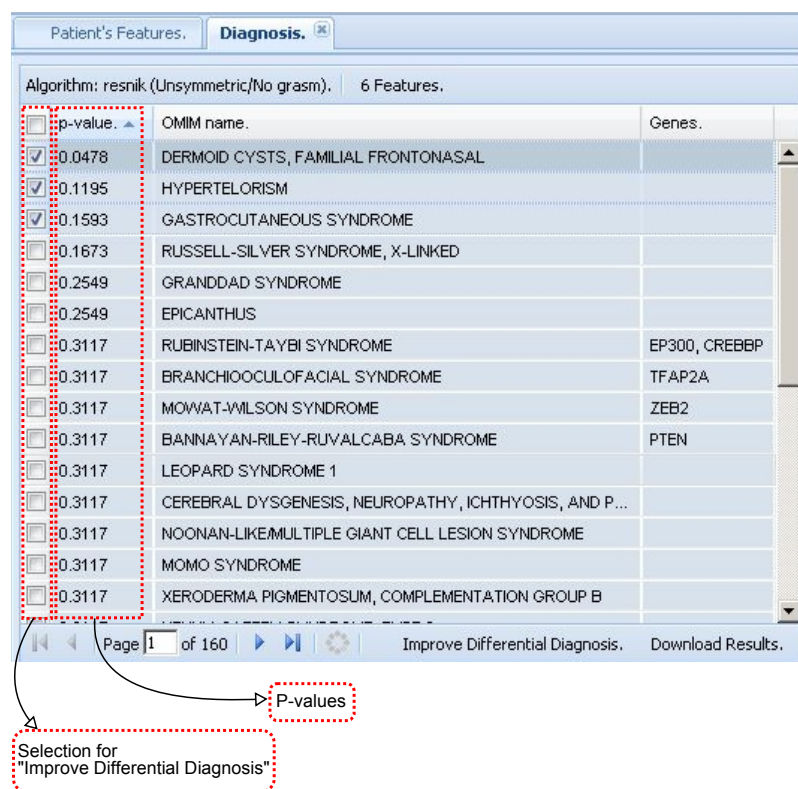

Figure 4: The Phenomizer ranks the diagnoses by their P-value.

For every row you can use the context-menu (right-click) and get further information about the suggested diagnosis, e.g. links to the genes known to cause this syndrome when mutated or you can get a visualization of the overlap between the query and a specific syndrome in the context of the HPO-structure, where the query terms are displayed in *blue*, the terms belonging to the disease in *yellow* and terms that are shared by query and disease in *red* (see Figure 5). Note that you you can drag the displayed image inside the popup-window.
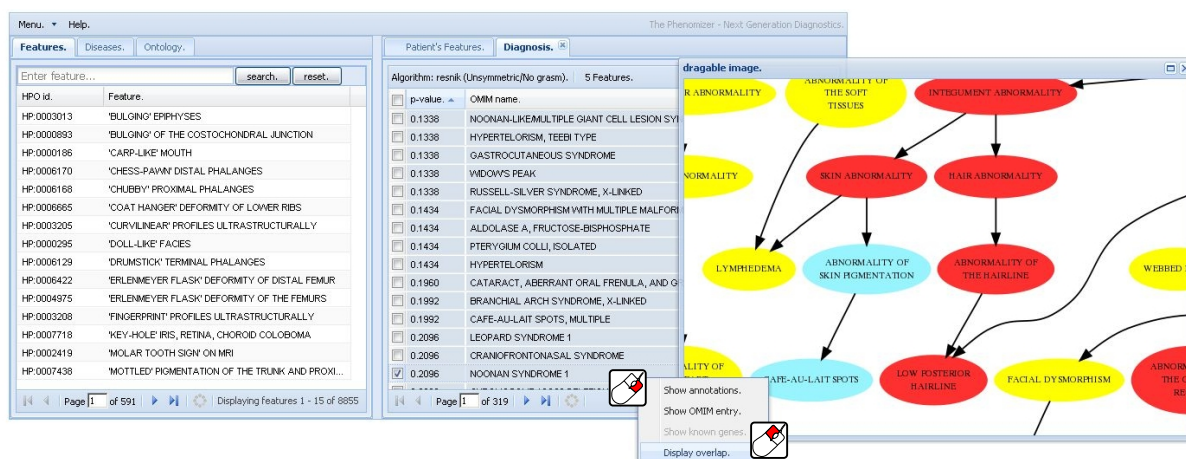
Figure 5: For each of the syndrome in the result list there exists a context menu from which allows you e.g. to display the overlap between the patients features and the syndrome's features. In this example, *Low posterior hairline* and the other red terms are shared by the disease and the query. *Lypmhedema* and the other yellow terms are used to annotate the disease but were not among the query terms. The blue term *Cafe-au-lait spots* was used in the query but is not used to annotate the disease.

3.3   Improve the differential diagnosis

The *Phenomizer* offers the possibility to improve your differential diagnoses. You can select several (at least two) of the listed syndromes by marking the corresponding checkbox. By clicking the 'Improve differential diagnosis' button the *Phenomizer* will generate a list of phenotypic features that are annotated only to a subset of the selected diseases (binary search) or that are specific for one syndrome (specific search).

3.3.1   *Binary* scoring of suggestive HPO-terms

The binary scoring method prioritizes terms that are used to annotate half of the selected differential diagnoses and are not present in the query.

The method starts by generating a list of all terms (*including ancestors*) of the selected syndromes. From this list all terms that already have been included in the query (patient's features) and their respective ancestors are removed. We thus have a set of phenotypic features that are only used to annotate the set of selected candidate-syndromes and have not already been used as query term. For each of these terms we calculate a score as the fraction of selected syndromes that show the particular feature compared to all selected syndromes. If multiple syndromes have the same score we favor terms that rather unspecific, meaning that we prioritize features that have a low information content. This feature can be used to identify general classes of phenotypic features that will best differentiate between about half of the marked differential diagnoses and may help to narrow down the differential diagnosis. Users can right-click on one of the proposal terms to obtain a submenu with more specific terms.

3.3.2   *Specific* scoring of suggestive HPO-terms

The specific scoring method prioritizes terms that are used to annotate a minimal number of the selected differential diagnoses and are not present in the query. This function can be more useful than the Binary search function if there is some reason to believe that one of the top differential diagnoses is the correct diagnosis and clinical features are being sought to further support the diagnosis.

The methods starts by taking all terms (*excluding ancestors*) of the selected syndromes. We then remove all terms that were implicitly listed in the query (i.e. the query terms and their descendents). For each of the terms we simply take the number of selected syndromes showing this feature. If multiple terms were used to annotate the same amount of selected syndromes we prioritize the features being most specific (highest information content).

3.4   Symmetric and Asymmetric Searches

Using the standard search method described in the main manuscript, only the terms of the diseases that best match the query are used for calculating the score. This may not be optimal for cases in which a patient is found

to have only a few clinical features but a disease is characterized by numerous signs and symptoms. In this case, a symmetric version of the score may be used.

$$\text{sim}(\mathcal{Q} \leftrightarrow \mathcal{D}) = \frac{1}{2} \text{avg} \left[\sum_{t_1 \in \mathcal{Q}} \max_{t_2 \in \mathcal{D}} IC(t_1, t_2) \right] + \frac{1}{2} \text{avg} \left[\sum_{t_2 \in \mathcal{D}} \max_{t_1 \in \mathcal{Q}} IC(t_1, t_2) \right]$$

This equation differs from the one presented in the main text by the fact that it also searches for the best match in the query terms for each of the terms annotating a disease. The Phenomizer automatically opens an option panel for switching to symmetric searches once 6 or more search terms have been entered.

4 Export search results

The results of the phenotypic analysis can be exported as a csv-file or PDF file. The csv file (comma-separated values) can be imported into spreadsheet programs such as OpenOffice of Microsoft Excel. The PDF file can be used to document the differential diagnostic process in the patient chart. Fields are provided to enter the patient's name and other information. A summary of the query terms and the ranked differential diagnoses is given. Finally, for every syndrome of the ranking a list of phenotypic features that are specific for this syndrome among all the diseases in the differential diagnosis is provided.