



FINDbase

Genome Variation Allele Frequencies Worldwide

www.findbase.org

**THE WORLDWIDE DATABASE FOR FREQUENCIES
OF CAUSATIVE MUTATIONS AND
PHARMACOGENOMIC MARKERS**

User Guide

Introduction

Frequency of INherited Disorders database (<http://www.findbase.org>) documents frequencies of pathogenic mutations leading to inherited disorders and pharmacogenomic markers worldwide.

Data content is freely available to the public and there are no registration requirements for data querying. The querying interface is based on a recently launched program by Microsoft, namely the PivotViewer [<http://www.getpivot.com>] based on Microsoft Silverlight technology [<http://www.silverlight.net>], which offers useful tools for querying large datasets in multiple ways. The whole application provides an elegant, web-based multimedia interface for population-based variation data collection and retrieval. Database records include the population, the ethnic group and/or the geographic region, the gene name and its variation parameters, the rare allele frequencies, accompanied by links to the respective Online Mendelian Inheritance in Man (OMIM) and the HGMD (pathogenic mutations) or PharmGKB (pharmacogenomic markers) entries. All entries are recorded against their unique PubMed and ResearcherIDs.

The whole system architecture is based on a three-tier client-server model, namely the client application, the application server and the database server. The first contains the presentation logic only. As a result, less resources are needed for the client workstation and no client modification is needed if database location changes. Changes to business logic are automatically enforced by the server and possible changes require only new application server software to be installed. The n-tier architecture is a robust model and flexible enough to aggregate multiple information sources and integrate modular developments. All the database records presented above are a collection that combines large groups of similar items. Conceptually, a collection behaves as any other web content. There is a set of files on a server, and a local client that has the capability to display them. The files are traditionally HTML and image files. In the collection case, the files are CXML and Deep Zoom-formatted (DZC) images [<http://msdn.microsoft.com/en-us/library/cc645077%28VS.95%29.aspx>]. When the user browses the collection from a web page, the PivotViewer uses the Silverlight Control to display the files.

Querying FINDbase

The user can select to zoom in from massive amounts of data to particular gene-specific, variation-specific, or population-specific details. FINDbase PivotViewer web application is

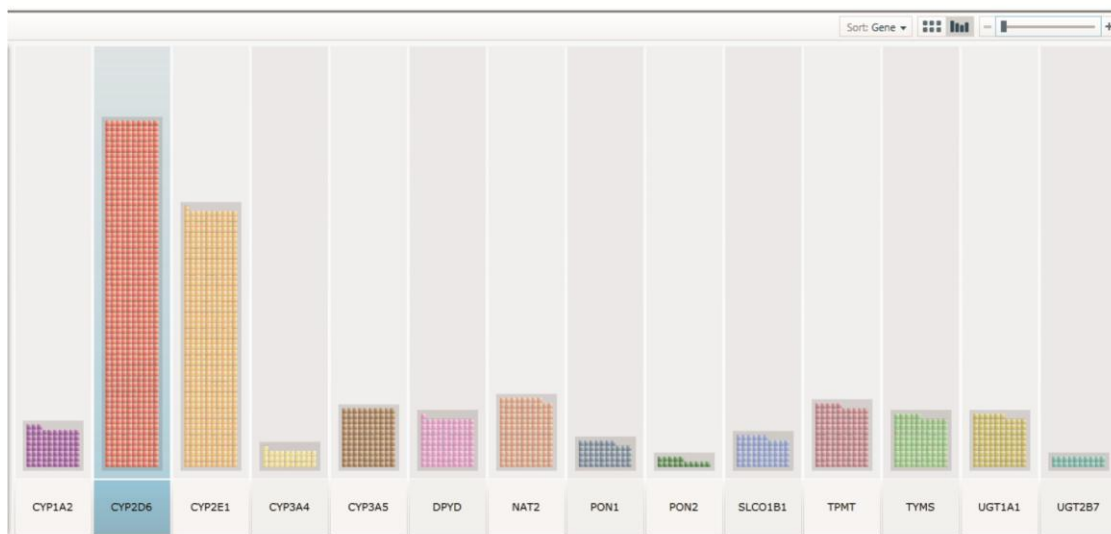
designed to provide a much more natural way for humans to digest large palettes of information without losing their way — an idea that anyone who has analyzed giant spreadsheets would welcome. It supports dynamical data visualization, sorting, organization and categorization. FINDbase PivotViewer web application combines large groups of similar items, so that the user can begin viewing the relationships between individual pieces of information in a new way. By visualizing hidden patterns, FINDbase PivotViewer web application enables users to discover new insights while interacting with thousands of data at once. FINDbase provides users a convenient environment to smoothly and quickly arrange collections according to common characteristics that can be selected from the data query menu and then zoom in for a closer look, by either filtering the collection to get a subset of information or clicking on a particular item. A display item in the form of a card (**Exhibit 1**) is provided for each marker, along with a sidebar textbox with in-depth data concerning the particular marker and population. Hyperlinks for each gene name to OMIM database and PharmGKB, offer to the user the possibility of easily accessing additional information.



Exhibit 1. Display items in the form of cards are provided for each pathogenic mutation (upper panel) or pharmacogenomic marker (lower panel), along with a sidebar textbox with in-depth data concerning the particular mutation or marker and population.

In particular, FINDbase enables the user to: i) Make sense from a sheer amount of data (i.e. population-based variation data), ii) Visualize and sort, organize and categorize data dynamically, iii) Discover trends across all items, using different views. For example, a user can perform a query for each of the 14 pharmacogenetically-relevant genes by selecting the field “Gene” (**Exhibit 2A**). Additional filtering allows the user to see specific allelic variants and to further zoom in a particular population. A useful query would also lead to data arrangement based on the frequency of the rare allele (**Exhibit 2B**).

A



B



Exhibit 2. Data content sorted using the columns option (See top-right corner of the screen), according to Gene name (**A**) or Rare allele frequency (**B**). The user can select a specific column (e.g. 2nd column in panel A showing variants of the *CYP2D6* gene highlighted) to retrieve more specific information.

Apart from these simple queries, the querying interface also allows the user to formulate compound queries. Such query would be the identification of rare pharmacogenomic markers (frequencies 0-10%) in the Japanese population, sorted out by gene name. The query output includes 51 rare alleles for 11 genes (7 alleles for *CYP1A2*, 12 alleles for *CYP2D6*, 3 alleles for *CYP2E1*, 4 alleles for *CYP3A4*, 1 allele for *CYP3A5*, 13 alleles for *DPYD*, 2 alleles for *NAT2*, 1 allele for *PON1*, 5 alleles for *TPMT*, 2 alleles for *TYMS* and 1 allele for *UGT2B7*; **Exhibit 3**).



Exhibit 3. A. Query formulation to retrieve the rare alleles in the Japanese population with an allelic frequency of 10.07% or lower (indicated with red arrows). **B.** The query returns 51 alleles in equal number of display items, sorted out in columns arranged by Gene name (red arrow). The total number of alleles is also shown in the query box along with the rare alleles with frequencies below 10,07% in all populations recorded in FINDbase-PGx. This query indicates that contrary to the Japanese and Italian populations, only 2 alleles in the Jamaican population have a frequency of less that 10.07%.

Another example of a variant that prompts critical usage of the anticancer drug 5'-fluorouracil (5'FU), is the *DPYD*-*2A allele (also known as *IVS14+1G>A*); this variant has been identified so far only in Caucasians, rarely in Asians, but not in Africans (**Exhibit 4**).

The user can use a plethora of querying options. For more information, the user is prompted to contact us using the contact information available.



Exhibit 4. Query formulation to retrieve the pharmacogenomic markers for the drugs 5'FU and debrisoquine (red arrow). This query returns 63 alleles in 2 genes, 33 for debrisoquine (for *CYP2D6* gene) and 30 for 5'FU (*DPYD* gene). The query output is sorted out by population (red arrow), using again the columns option.

In case you encounter any problems or want to report possible bugs please contact Manolis Viennas (viennas AT gmail DOT com).

If you are interested in implementing the ETHNOS software to establish a new National or Ethnic Genetic database please contact the principal investigator. Detailed contact information is provided in the contact page.