FINDbase: worldwide database for clinically relevant genomic variation allele frequencies

http://www.findbase.org

Branka Zukic[1], Biljana Stankovic[1], Nikola Kotur[1], Vladimir Gasic[1], Kristel Klaassen[1], Ksenija Vukicevic[2], George Patrinos[3] and Sonja Pavlovic[1]

[1] Laboratory for Molecular Biomedicine, Institute of Molecular Genetics and Genetic Engineering, University of Belgrade, Serbia

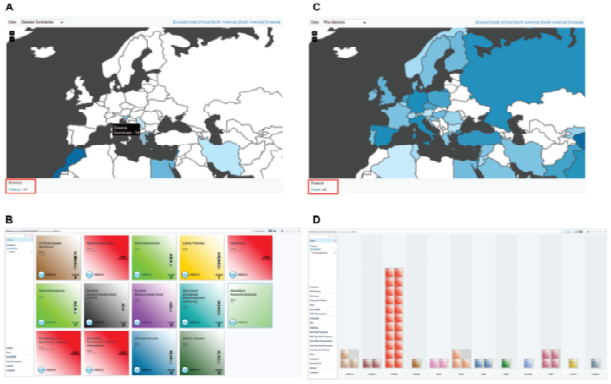
[2] Department of Physiology, Faculty of Medical Sciences, University of Kragujevac, Serbia

[3] Department of Pharmacy, School of Health Sciences, University of Patras, University Campus Rion, Greece

FINDbase (Frequency of INherited Disorders database), is a comprehensive online data repository that in a structured manner records the prevalence of clinically relevant genomic variants in various populations worldwide, such as pathogenic variants leading mostly to monogenic disorders and pharmacogenomics biomarkers, all in **well-distinct data modules**. The incidence of rare genetic diseases in various populations is recorded also.

The data came from previously published reports as well as from unpublished information contributed from individual researchers prior of publication. The contributor's unique ResearcherID follows the **microattribution approach**, allowing unambiguous identification of curated data. FINDbase is freely available and no registration is needed for data querying.

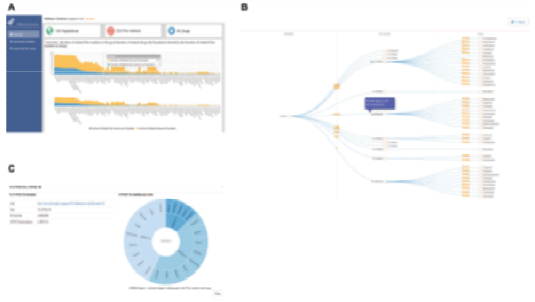
This work was supported by Ministry of Education, Science and Technological Development, Republic of Serbia (Grant No. III41004).



Overview of the new FINDbase map display visualization tool. (A) Overview of the documented genetic disease summaries in European countries (part of North Africa is also shown). By hovering the cursor over Greece, a text box appears providing an overview of the number of disease summaries documented in FINDbase for the Hellenic population. By clicking on the country, a link appears in the bottom down part of the map (indicated in a red rectangle). Clicking on the link opens a new browser page, where all the different genetic disease summaries in the (B) Hellenic population appear, where the user can navigate for further information. (C) Overview of the documented PGx markers in European countries (part of North Africa is also shown). Again, by clicking on a specific country (e.g. Serbia), a link appears in the bottom down part of the map (indicated in a red rectangle), which opens a new browser page, where all the different PGx markers in the Serbian population appear, (D) grouped per gene. Please note that the intensity of the blue color is indicative of the number of data records that are documented in FINDbase for every country per data module, respectively.

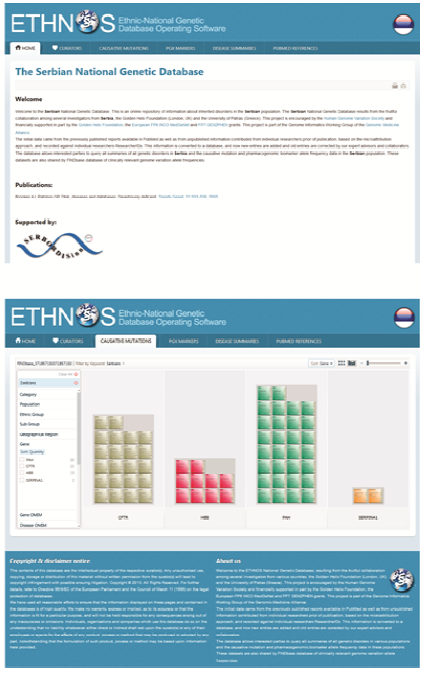
Pharmacogenomic Markers module documents the frequencies of pharmacogenetically relevant SNVs in various populations worldwide. Database records include the population, the ethnic group and/or the geographic region, the gene name and its variation parameters, the rare allele frequencies, linked to the respective OMIM and the PharmGKB entries. This module includes data for 144 pharmacogenomic markers across 14 genes, representing 87,000 individuals from 150 populations. Interrelating the PGx data module with DruGeVar database provide users with the best of both resources.

Causative Mutations module documents the frequencies of causative mutations leading to inherited disorders in various populations worldwide. Database records include the population, the ethnic group and/or the geographic region, the gene name and its variation parameters, the rare allele frequencies, linked to the respective OMIM and the HGMD entries. This module includes data for more than 3,800 disease-causing mutations across 26 genes, representing over 100,000 individuals from 92 populations.



Integration of DruGeVar with the PGx markers data module of FINDbase. (A) Bar chart showing the number of the related PGx biomarkers to drugs and the number of related drugs by population sorted by the number of related PGx biomarkers to drugs. The subchart is used for selecting specific region of the bar chart. (B) Collapsible tree diagram showing the relation between population, PGx biomarkers and drugs. The number between population and PGx biomarker defines the rare allele frequency and the label between PGx number and drug the effect on the toxicity or efficacy of the biomarker to the related drug. Clicking on the nodes of PGx biomarkers or drugs additional information for the selected element is appearing in a pop up window. (C) Pop up window showing additional information for the selected PGx biomarker. Left: Details for the selected PGx marker. Right: Sunburst visualization, showing the related additional PGx markers and drugs of the gene of the selected PGx biomarker. Clicking on the elements of the PGx biomarkers, the visualization is reconstructed focusing on the children nodes (related drugs) of the selected node, while clicking on the elements of the drugs, the visualization is reconstructed focusing on the 'children' of the parent node.

National/Ethnic Genetic Databases (NEGDB) modules records information over the described genetic heterogeneity of an ethnic groups/populations. FINDbase comprises 90 individual ETHNOS-based NEGDBs in all 5 continents.



Development of the Serbian NEGDB, catalyzed by the upgraded version of the ETHNOS software. (A) Home page of the Serbian NEGDB, where all available information for pathogenic variants, PGx markers and genetic disease summaries is provided in different tabs. (B) Overview of the pathogenic variants in the *CFTR*, *HBB*, *PAH* and *SERPINA1* genes, leading mostly to monogenic disorders in the Serbian population.