






SOFTWARE TOOL ARTICLE

Pharmosome: an integrative and collective database for exploration and analysis of single nucleotide polymorphisms associated with disease [version 1; peer review: awaiting peer review]

Peter T. Habib ^{1,2}, Alsamman M. Alsamman ³, Sameh E. Hassanein ^{1,4}, Kerolos M. Yousef¹, Aladdin Hamwiah²

¹Department of Bioinformatics and Functional Genomics, College of Biotechnology, Misr University for Science and Technology, Giza, Egypt

²Department of Biodiversity and Crop Improvement Program, International Center for Agricultural Research in the Dry Areas, Cairo, Egypt

³Department of Genome Mapping, Molecular Genetics and Genome Mapping Laboratory, Agricultural Genetic Engineering Research Institute, Giza, Egypt

⁴Department of Bioinformatics & Computer Networks, AGERI, Agricultural Research Center, Giza, Egypt

v1 First published: 10 Jan 2020, 9:14 (<https://doi.org/10.12688/f1000research.21773.1>)

Latest published: 10 Jan 2020, 9:14 (<https://doi.org/10.12688/f1000research.21773.1>)

Abstract

Current single nucleotide polymorphism (SNP) databases are limited to a narrow set of SNPs, which has led to a lack of interactivity between different databases, limited tools to analyze and manipulate the already existing data, and complexity in the graphical user interface. Here we introduce Pharmosome, a web-based, user-friendly and collective database for more than 30,000 human disease-related SNPs, with dynamic pipelines to explore SNPs associated with disease development, drug response and the pathways shared between different genes related to these SNPs. Pharmosome implements several tools to design primers to detect SNPs in large genomes and facilitates analysis of different SNPs to determine relationships between them by aligning sequences, constructing phylogenetic trees, and providing consensus sequences illustrating the connections between SNPs. Pharmosome was written in the Python programming language using the Django web framework in combination with HTML, CSS, and JavaScript to receive user inputs, and process and export the sorted result to the interface. Pharmosome is available from: <https://pharmosome.herokuapp.com/>.

Keywords

SNP, Disease, Python, Bioinformatics

Open Peer Review

Reviewer Status AWAITING PEER REVIEW

Any reports and responses or comments on the article can be found at the end of the article.

Corresponding author: Peter T. Habib (p.habib911@gmail.com)

Author roles: **Habib PT:** Conceptualization, Investigation, Software, Validation, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Alsamman AM:** Validation, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Hassanein SE:** Conceptualization, Project Administration, Supervision, Writing – Original Draft Preparation, Writing – Review & Editing; **Yousef KM:** Validation, Writing – Review & Editing; **Hamwiah A:** Funding Acquisition, Supervision

Competing interests: No competing interests were disclosed.

Grant information: The author(s) declared that no grants were involved in supporting this work.

Copyright: © 2020 Habib PT *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Habib PT, Alsamman AM, Hassanein SE *et al.* **Pharmosome: an integrative and collective database for exploration and analysis of single nucleotide polymorphisms associated with disease [version 1; peer review: awaiting peer review]**

F1000Research 2020, 9:14 (<https://doi.org/10.12688/f1000research.21773.1>)

First published: 10 Jan 2020, 9:14 (<https://doi.org/10.12688/f1000research.21773.1>)

Introduction

With the extending and deciphering of genomic data produced by sequencing technologies and the Human Genome Project¹, much information has been discovered, such as exons, introns, domains, coding sequence, and non-coding sequences. Mutated signals have attracted tremendous attention because of their crucial impact in altering gene expression, in particular, single nucleotide polymorphisms (SNPs)², which also act as gene molecular-markers for an associated trait. Additionally, the existence of a specific SNP can indicate precisely what disease is foreseen and the possible drugs to treat it, which is considered the ultimate goal of pharmacogenomics.

Pharmacogenetics, or its inclusive version pharmacogenomics (since it covers proteomic, genomic, epigenomic and transcriptional effects on disease and drug response), has produced a vast body of research since 1997³, due to its key importance in personalized medicine through investigating how far genetic variations (e.g. SNPs) are involved in disease development and determining drug targets. Thereby the safety and efficacy of an individualized drug therapy can be improved. Since the relationships between molecular data pertaining to patients and their disease phenotype are complex and difficult to determine manually, scientists have begun to develop and enrich the bioinformatics knowledge base with more sophisticated and accurate molecular tools to detect genetic variations for example, SNPector is a recent tool developed by the authors to detect SNP effect in drug response and disease development⁴. This will allow interpretation of how these tiny variations may cause direct errors, e.g. X-SNP is a common SNP type that gives rise to premature termination codons that halt gene expression⁵. Some SNP types involved in alteration of the protein production process cause disease⁶, while other regulatory SNPs⁷ may disrupt pathways causing cascade errors that lead to collapse of a whole pathway thereby causing disease development.

Elucidations about SNPs play a central role in providing recommendations for practicing physicians. In addition, a wide range of research fields have arisen out of pharmacogenomics, such as vaccinomics, which study aberrant immune responses to vaccines based on genetic makeup. For example, a specific SNP in the *TLR3* gene was found to be responsible for the reduction of humoral immune responses and cell-mediated immunity to the measles vaccine⁸. Nutrigenomics is the science of gene–nutrient interactions, which involves research methods

and clinical implementation to detect and treat nutrient-related diseases. One of the best known examples of nutrigenomics is lactase persistence, in which the gene encoding lactase is expressed past weaning. Lactase persistence in Europeans is caused by a polymorphism called “C-13910T” in the lactase phlorizin hydrolase gene promoter⁹, and the lack of C-1390T in many adults can lead to severe gastrointestinal discomfort and diarrhea resulting from ingesting milk due to the inability to metabolize lactose¹⁰.

Here we introduce Pharmosome, a web-based, user-friendly and collective database for more than 30,000 human disease-related SNPs, with dynamic pipelines to explore SNPs associated with disease development, drug response and the pathways shared between different genes related to these SNPs. Pharmosome implements several tools to design primers to detect SNPs in large genomes and facilitates analysis of different SNPs to determine relationships between them by aligning sequences, constructing phylogenetic trees, and providing consensus sequences illustrating the connections between SNPs.

Methods

Implementation

We collected SNP-related data (e.g. SNP ID, annotation, pathway and phenotypes) from PharmGKB¹¹, NCBI^{12,13}, Ensemble¹⁴, DiseaseEnhancer¹⁵, GeneCards¹⁶ and Reactome¹⁷ in tab-delimited format in order to link between different available information. The collected data were categorized into four main sub-databases: SNP, Gene, Chemical and Disease. The Python 3+ programming language was utilized to read, select, filter and sort the data and to link the Python scripts with HTML and JavaScript codes.

Data collection. About 50% of the data was downloaded from PharmGKB, which is considered the most common database for SNP annotation. The data comprises the associated phenotype, the clinical perspectives and, considering storage space limitations, the remaining data are imported on-demand to use later by a set of Python functions we built to get access to the Application Programming Interface (API) of different databases. These other databases include GeneCards, DiseaseEnhancer, Ensembl and Reactome; thereby a user can return specific data using preset IDs and then export this information to the HTML interface. The number of data entries collected by Pharmosome is shown in Table 1.

Table 1. Number of data entries collected from each database used by Pharmosome.

Type of data	NCBI	Ensemble	Reactome	PharmGKB	DiseaseEnhancer	GeneCards	KEGG	DrugBank
Gene	18,975	0	0	27,000	1,059	13,500	20,109	0
Transcript	0	381,060	0	0	0	0	0	0
Pathway	0	0	2,256	0	0	0	0	0
Disease	0	0	0	3,546	1,059	0	2,287	0
Chemical	0	0	0	3,393	0	0	0	11,926
SNP	440,000,000	0	0	10,780	0	0	0	0
Annotation	10,845	0	2,256	3,393	0	0	20,109	0

Data fetching and exporting. We constructed a Python module containing 12 functions. These functions connect in different ways to import the data from the databases above either from tab-delimited files or using APIs. User-requested information is extracted and exported to the Pharmosome web interface. The Django web framework was used to build the Pharmosome web interface with HTML programming language. We used Django to build functions that receive the user input from the Pharmosome interface, process requests using a Python script, and finely export the result to the interface.

```
return render(request, 'WebInterface.html', context={})
```

Operation

Google Chrome browser is recommended to use the Pharmosome web interface (<https://pharmosome.herokuapp.com/>), but other internet browsers can also be used.

Functions of Pharmosome

SNP sub-database. In the SNP sub-database, users can enter the ID of a SNP (e.g. rs141033578) and receive output data about its related gene, chromosomal location, gene bands, summary of the normal function of the gene, the gene part responsible for enhancing the disease occurrence, pathway of defective gene (if available), gene transcripts and different splicing variants. Pharmosome also provides information that can be used to retrieve data of recent studies (specifically, SNP reference nucleotide and alternative nucleotide data and an explanation of how the SNP contributes to disease development and drug response). For example, an SNP in *GSTP1* was associated with overall survival in 107 patients with metastatic colorectal cancer who received 5-FU/oxaliplatin combination chemotherapy that caused the replacement of isoleucine with valine at amino acid position 105 of the protein, which is known to substantially diminish enzyme activity¹⁸.

Disease sub-database. The Disease sub-database allows users to search for disease data collected from **KEGG Disease** and **PharmGKB** databases by entering the disease name. The search result is a list of genes responsible for the disease, the ID of the SNP occurring in this gene, a description of clinical annotations related to this gene and the gene-specific chemical used in the treatment. The sub-database can be used, for example, to identify the SNP, gene and chemical related to coronary artery disease caused by a high frequency of a particular polymorphism in the *PLA2* gene. This gene encodes glycoprotein IIIa, which is associated with a high prevalence of premature myocardial infarction^{19,20}.

Chemical sub-database. The Chemical sub-database has data from four sources: **KEGG**, **PharmGKB**, **DrugBank** and **ChemSpider**. After users enter a chemical name, the output is a list of the chemical name, trend and generic names, structure, description and pharmacodynamics. For example, by inputting bortezomib (a proteasome inhibitor), users receive output data relating to the clinical success of bortezomib, which established the ubiquitin (Ub)+proteasome system as a key therapeutic target in multiple myeloma^{21,22}.

Gene sub-database. The Gene sub-database links between different sources (NCBI GenBank, Ensemble, Reactome and DiseaseEnhancer). In particular, the DiseaseEnhancer dataset represents a new approach that determines the gene part that is responsible for enhancing occurrence of disease. NCBI provides information about gene name, location, a summary of gene function and chromosomal location. Reactome displays the pathway in which genes are involved and gives an overview description. The Gene sub-database can be used to identify the pathway, splicing variants, disease enhancing region, genomic and proteomic expression profile or even to get general information. An example of this would be looking at the *SLCO1B1* gene, for which various groups tested the hypothesis of whether polymorphisms in *SLCO1B1* affect pharmacokinetics and the effects of drugs in humans²³.

SNP collector. The SNP collector is a tool within Pharmosome that is designed to find all SNPs present on the gene, related to disease, associated with the chemical compound. Users can choose between different options to collect SNPs and their clinical annotations and chemicals related to each SNP. This tool can be used to find SNPs or other information. For example, it could be used to detect the bond work which occurs on location 118 of the mu opioid gene, which was 3 fold more vigorous than the wildtype in its interaction with b-endorphin. Other regulatory-SNPs in this region of the gene can be linked to other phenotypes²⁴.

Pick primer. The detection of SNP existence in a DNA sample taken from patients depends on designing an appropriate primer. Primers should be compatible with the flanking sequences of SNP. As the presence of an SNP in the genome may result in disease and affect the choice of drug, there is a need to detect the presence of SNP e.g. for early disease diagnosis. The pick primer tool within Pharmosome has the important function of designing primers to detect an SNP in the genome by retrieving the SNP sequence record from the NCBI database, locating the SNP position and designing primers 50 bases before, after and within the SNP sequence.

SNP phylogeny. As discussed in previous sections, there is always some relationship between different SNPs due to the complicated interaction network between different genes. In order to determine how these SNPs are related to each other, the SNP phylogeny tool constructs a phylogenetic tree that illustrates the relationships between different SNPs²⁵ by downloading SNP and flanking sequences and commencing multiple sequence alignment to determine how far each sequence is related to others. This function could be used clarify connections in studies, such as that of Thompson *et al.* that showed an association of 43 SNPs in 16 genes with the response drug of atorvastatin²⁶.

Workflow

Figure 1 shows the flow of information to meet the needs of users.

Use case

Pharmosome deploys seven sub-databases and tools. Our approach during the building of Pharmosome, is to achieve the easiest

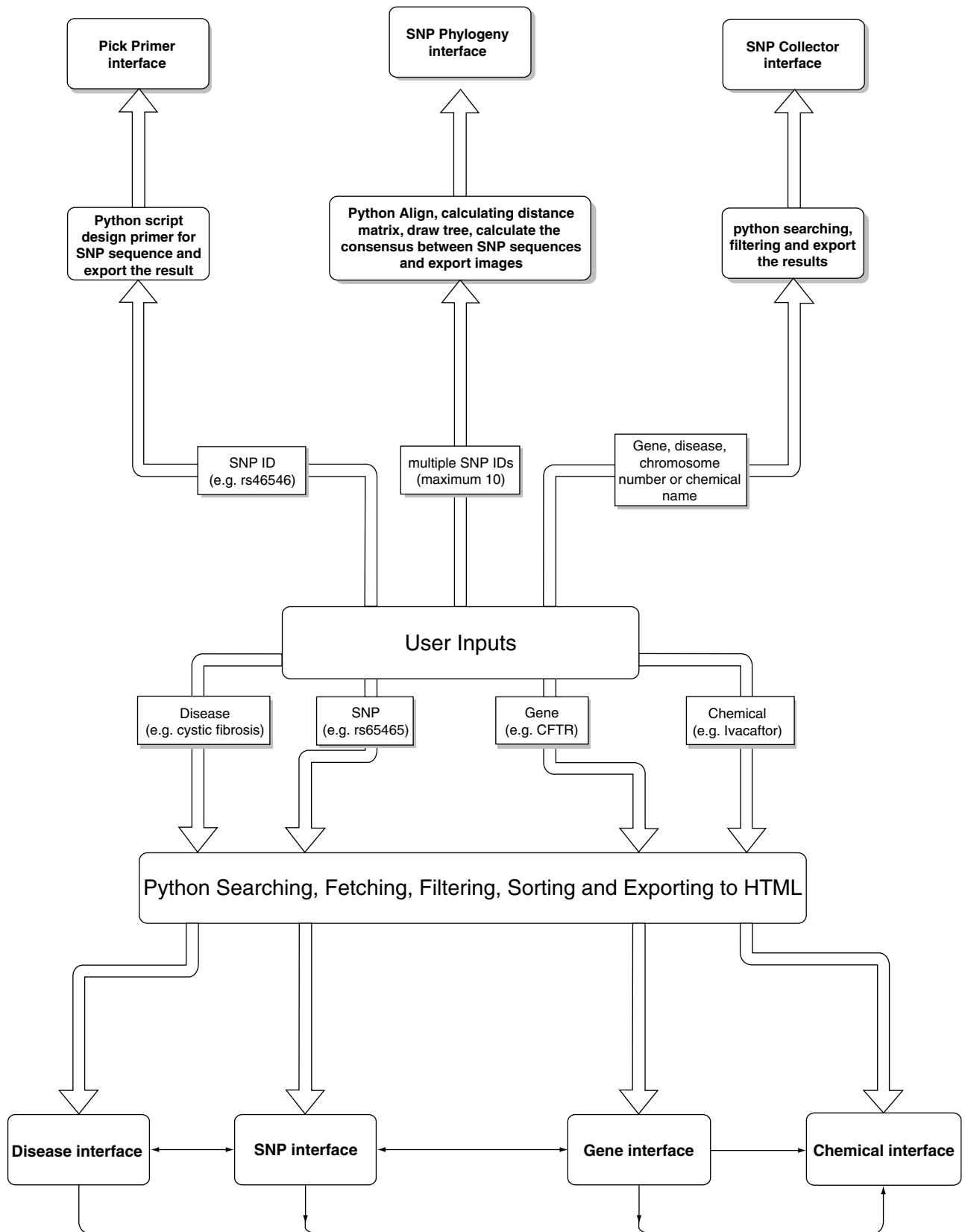


Figure 1. Pharmosome workflow illustrating the processing in the background of the database and the relationships and links between different interface webpages.

usage. We designed each tool and sub-database to receive the user input with minimum required parameters (as shown in Table 2). Users enter the target input they require and the

Pharmosome interface will automatically redirect to another page that shows the user the output results. Figure 2–Figure 5 show output on the Pharmosome web interface.

Table 2. Use cases for Pharmosome detailing the input required in the web interface and a summary of the output of Pharmosome.

Function	Input required	Input example	Output description
SNP sub-database	SNP ID	rs75527207	SNP related disease and drug response annotation
Disease sub-database	Disease Name	Heart Failure	Gene involved in disease, chemical used in treatment, and SNP causing the disease (Figure 1)
Chemical sub-database	Chemical Name	Ivacaftor	Description of target disease (Figure 2)
Gene sub-database	Gene Symbol	CFTR	Gene annotation (Figure 3)
SNP collector	Gene, Chromosome, Disease, or Chemical	CFTR, 11, Heart Failure, or Ivacaftor	List of SNPs (Figure 4)
Pick primer	SNP ID	rs75527207	Forward and reverse primer
SNP phylogeny	SNP List	rs113993960 rs2853741 rs1045642 rs3857532 rs10817464 rs16969968	Illustration shows the phylogenetic tree of the input SNPs

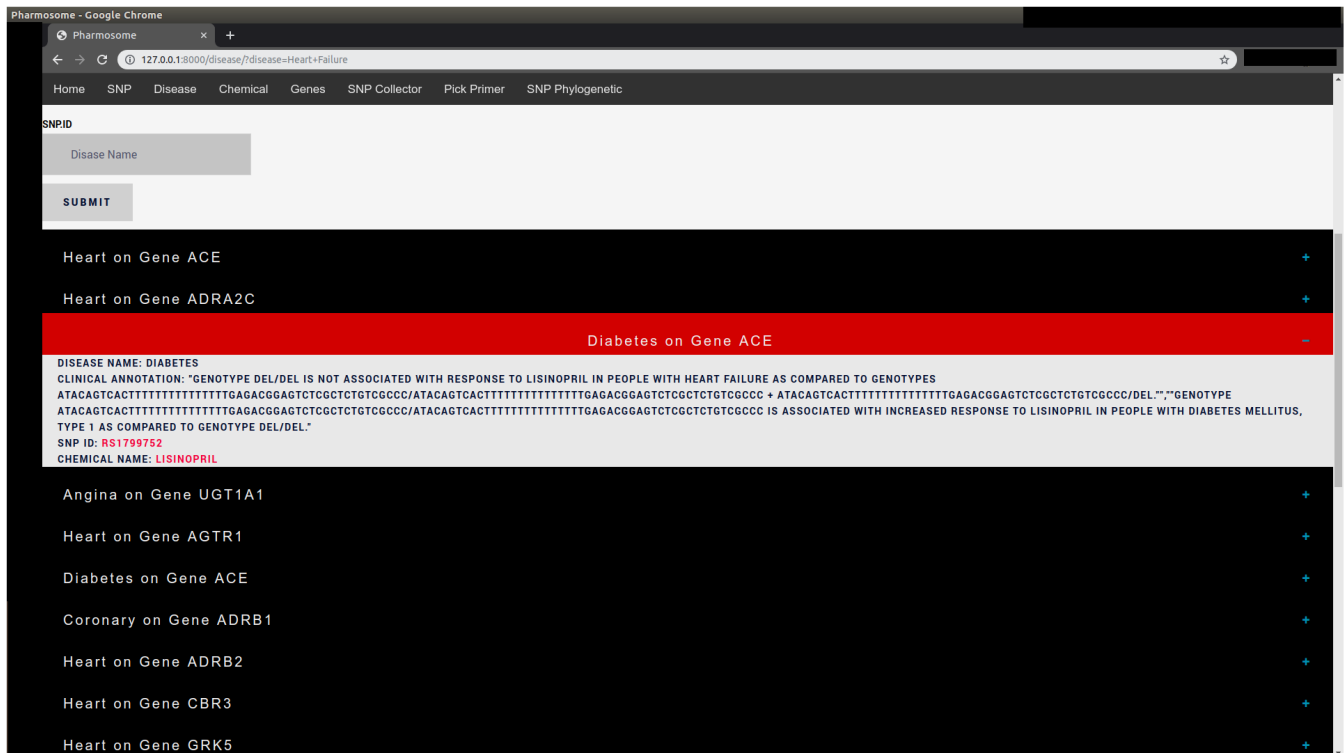


Figure 2. Disease sub-database output on Pharmosome web interface after input of “Heart failure”. The results consist of list of drop-down menus. Each menu describes disease annotation and the associated gene to the disease.

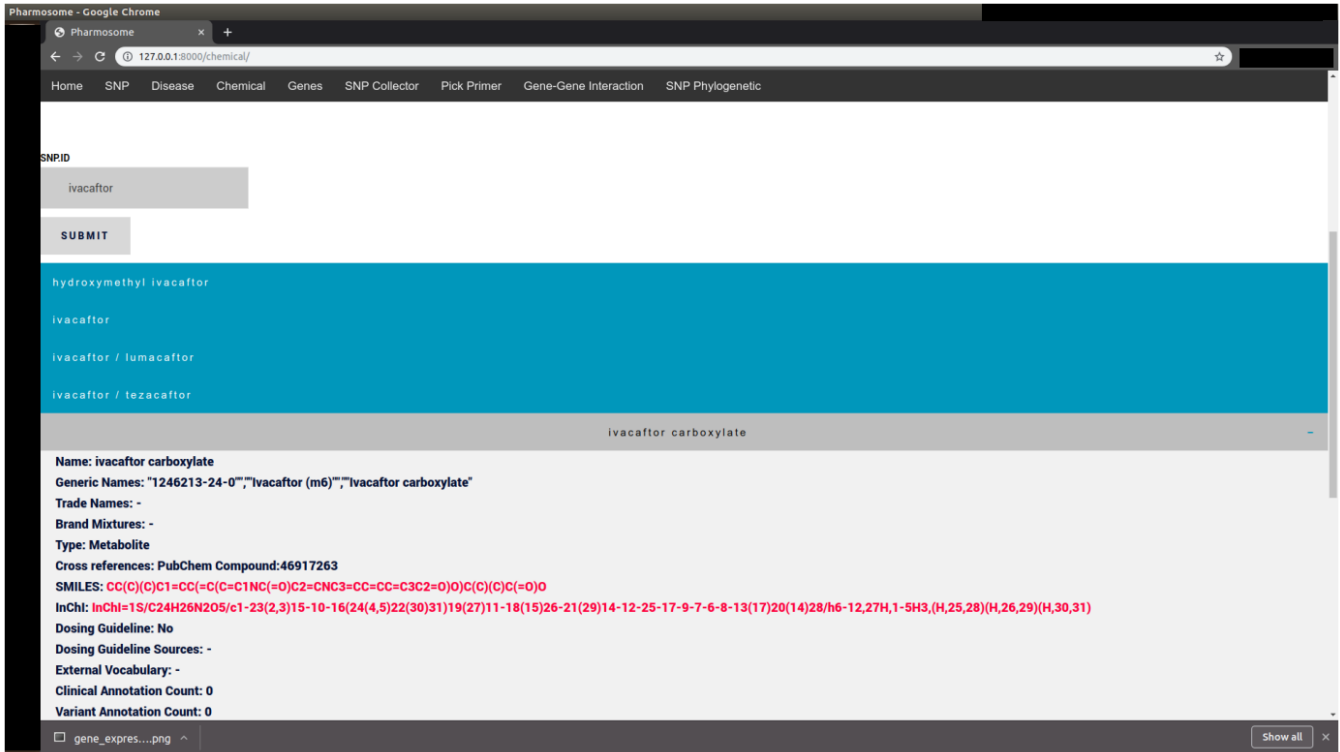


Figure 3. Chemical sub-database output on Pharmosome web interface after input of “ivacaftor”. The results consists of a list of drop-down menus. Each menu describes Drug/Chemical annotations.

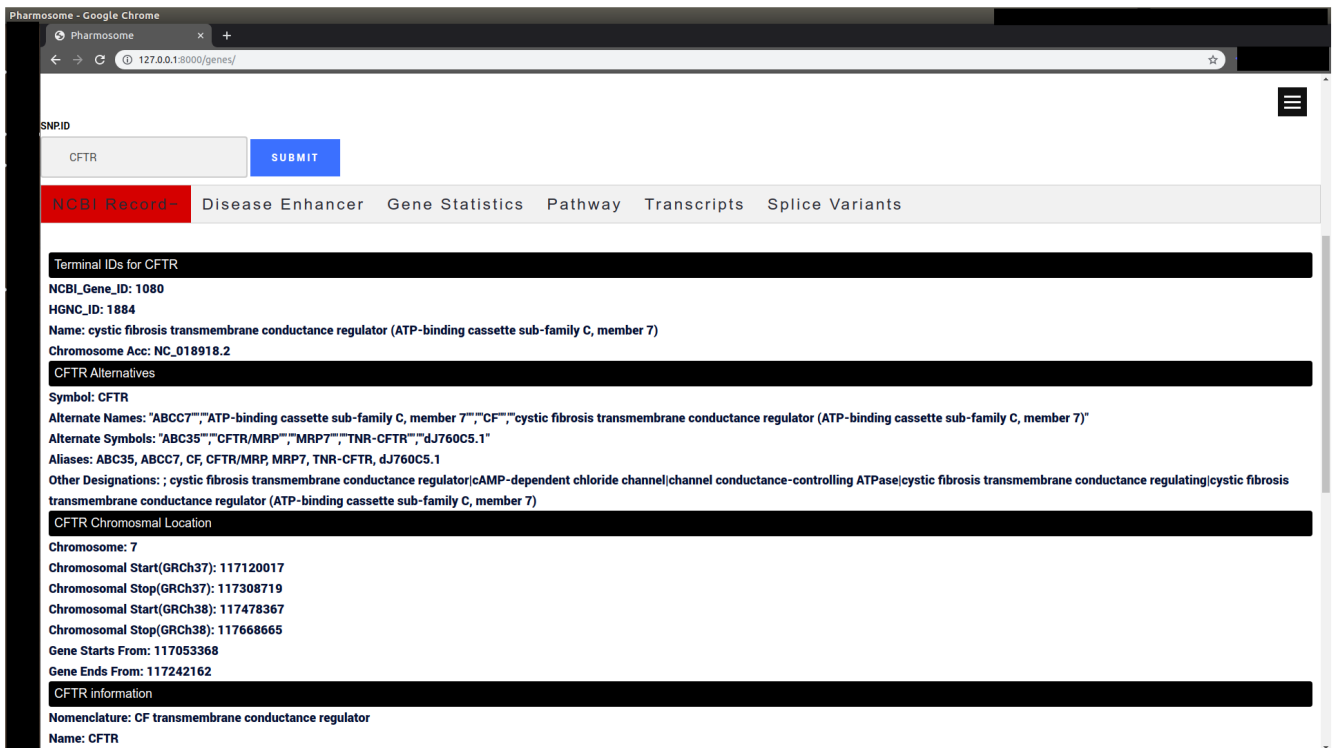


Figure 4. Gene sub-database output on Pharmosome web interface after input of “CFTR”. The results consist of a navigation bar, and each button expands to a different annotation.

First: Choose Input Type
 Gene Chromosome Chemical Disease

Second: Enter The Data:
 SNPID
 CFTR
 SUBMIT

Third: The Results
 Found 8 SNP Related To CFTR

SNP ID
rs113993960
rs113993960
rs113993960
rs75527207
rs113993960
rs113993960
rs75527207
rs75527207

Figure 5. SNP collector output on Pharmsome web interface after input of “CFTR”. The results consists of a list of SNPs located on this gene.

Summary

In this study, we introduce Pharmsome, an integrative and collective database for exploring and analysing human SNPs and the associated disease and drug response. Our tool deploys various functions to determine the relationships between different SNPs, construct the consensus sequence between different SNPs and to determine the pathways shared between different genes. Pharmsome also includes sub-databases to simplify, link and display data about gene functions, pathways, transcriptomes of genes, different splicing variants, clinical annotation, chemical structures and annotations of chemicals involved in the disease. The returned data are informative, user-friendly and easy to navigate. Pharmsome was written in

Python 3.5, HTML and CSS with the implementation of Django (Python library) to design links between Python scripts and other languages.

Software availability

Pharmsome web interface: <https://pharmsome.herokuapp.com/>

Source code available from: https://github.com/peterhabib/Pharmsome_Web

Archived source code as at time of publication: <http://doi.org/10.5281/zenodo.3583191>²⁷.

License: MIT

References

- Lander ES, Linton LM, Birren B, *et al.*: **Initial sequencing and analysis of the human genome.** *Nature.* 2001; **409**(6822): 860–921.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Vignal A, Milan D, SanCristobal M, *et al.*: **A review on SNP and other types of molecular markers and their use in animal genetics.** *Genet Sel Evol.* 2002; **34**(3): 275–305.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Gurwitz D, Pirmohamed M: **Pharmacogenomics: the importance of accurate phenotypes.** *Pharmacogenomics.* 2010; **11**(4): 469–70.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Habib PT, Alsamman AM, Hassanein SE, *et al.*: **SNPector: SNP inspection tool for diagnosing gene pathogenicity and drug response in a naked sequence [version 1; peer review: awaiting peer review].** *F1000Res.* 2019; **8**: 2133.
[Publisher Full Text](#)
- Savas S, Tuzmen S, Ozelik H: **Human SNPs resulting in premature stop codons and protein truncation.** *Hum Genomics.* 2006; **2**(5): 274–86.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Bond J, Scott S, Hampshire DJ, *et al.*: **Protein-truncating mutations in ASPM cause variable reduction in brain size.** *Am J Hum Genet.* 2003; **73**(5): 1170–7.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- De Gobbi M, Viprakasit V, Hughes JR, *et al.*: **A regulatory SNP causes a human genetic disease by creating a new transcriptional promoter.** *Science.* 2006;

- 312(5777): 1215–7.
[PubMed Abstract](#) | [Publisher Full Text](#)
8. Dhiman N, Poland GA, Cunningham JM, *et al.*: **Variations in measles vaccine-specific humoral immunity by polymorphisms in *SLAM* and *CD46* measles virus receptors.** *J Allergy Clin Immunol.* 2007; **120**(3): 666–72.
[PubMed Abstract](#) | [Publisher Full Text](#)
 9. Enattah NS, Sahi T, Savilahti E, *et al.*: **Identification of a variant associated with adult-type hypolactasia.** *Nat Genet.* 2002; **30**(2): 233–7.
[PubMed Abstract](#) | [Publisher Full Text](#)
 10. Kaput J, Rodriguez RL: **Nutritional genomics: the next frontier in the postgenomic era.** *Physiol Genomics.* 2004; **16**(2): 166–77.
[PubMed Abstract](#) | [Publisher Full Text](#)
 11. Hewett M, Oliver DE, Rubin DL, *et al.*: **PharmGKB: the Pharmacogenetics Knowledge Base.** *Nucleic Acids Res.* 2002; **30**(1): 163–5.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 12. **Gene.** Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information; 2004; Accessed 2019-02-16.
[Reference Source](#)
 13. **dbSNP.** Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information; 1998; Accessed 2019-02-16.
[Reference Source](#)
 14. Kersey PJ, Allen JE, Allot A, *et al.*: **Ensembl Genomes 2018: an integrated omics infrastructure for non-vertebrate species.** *Nucleic Acids Res.* 2018; **46**(D1): D802–8.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 15. Zhang G, Shi J, Zhu S, *et al.*: **DiseaseEnhancer: a resource of human disease-associated enhancer catalog.** *Nucleic Acids Res.* 2018; **46**(D1): D78–84.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 16. Safran M, Dalah I, Alexander J, *et al.*: **GeneCards Version 3: the human gene integrator.** *Database (Oxford).* 2010; **2010**: baq020.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 17. Fabregat A, Sidiropoulos K, Viteri G, *et al.*: **Reactome diagram viewer: data structures and strategies to boost performance.** *Bioinformatics.* 2018; **34**(7): 1208–14.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 18. Stoehlmacher J, Park DJ, Zhang W, *et al.*: **Association between glutathione S-transferase P1, T1, and M1 genetic polymorphism and survival of patients with metastatic colorectal cancer.** *J Natl Cancer Inst.* 2002; **94**(12): 936–42.
[PubMed Abstract](#) | [Publisher Full Text](#)
 19. Weiss EJ, Bray PF, Tayback M, *et al.*: **A polymorphism of a platelet glycoprotein receptor as an inherited risk factor for coronary thrombosis.** *N Engl J Med.* 1996; **334**(17): 1090–4.
[PubMed Abstract](#) | [Publisher Full Text](#)
 20. Walter DH, Hink U, Asahara T, *et al.*: **The *in vivo* bioactivity of vascular endothelial growth factor/vascular permeability factor is independent of N-linked glycosylation.** *Lab Invest.* 1996; **74**(2): 546–56.
[PubMed Abstract](#)
 21. Richardson PG, Mitsiades CS, Hideshima T, *et al.*: **Novel biological therapies for the treatment of multiple myeloma.** *Best Pract Res Clin Haematol.* 2005; **18**(4): 619–34.
[PubMed Abstract](#) | [Publisher Full Text](#)
 22. Richardson PG, Schlossman RL, Alsina M, *et al.*: **PANORAMA 2: panobinostat in combination with bortezomib and dexamethasone in patients with relapsed and bortezomib-refractory myeloma.** *Blood.* 2013; **122**(14): 2331–7.
[PubMed Abstract](#) | [Publisher Full Text](#)
 23. König J, Seithel A, Gradhand U, *et al.*: **Pharmacogenomics of human OATP transporters.** *Naunyn Schmiedebergs Arch Pharmacol.* 2006; **372**(6): 432–43.
[PubMed Abstract](#) | [Publisher Full Text](#)
 24. Bond C, LaForge KS, Tian M, *et al.*: **Single-nucleotide polymorphism in the human mu opioid receptor gene alters beta-endorphin binding and activity: possible implications for opiate addiction.** *Proc Natl Acad Sci U S A.* 1998; **95**(16): 9608–13.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 25. Habib PT, Alsamman AM, Hamwiah A: **BioAnalyzer: Bioinformatic software of routinely used tools for analysis of genomic data.** *Biotechnology.* 2019; **10**: 33–41.
[Publisher Full Text](#)
 26. Thompson JF, Man M, Johnson KJ, *et al.*: **An association study of 43 SNPs in 16 candidate genes with atorvastatin response.** *Pharmacogenomics J.* 2005; **5**(6): 352–8.
[PubMed Abstract](#) | [Publisher Full Text](#)
 27. Peter: **Pharmosome: An Integrative and Collective Database for Explorations and Analysis of SNP Associated with Disease.** *Zenodo.* 2019.
<http://www.doi.org/10.5281/zenodo.3583191>

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias
- You can publish traditional articles, null/negative results, case reports, data notes and more
- The peer review process is transparent and collaborative
- Your article is indexed in PubMed after passing peer review
- Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com

F1000Research