

# Drug-SNPing: an integrated drug-based, protein interaction-based tagSNP-based pharmacogenomics platform for SNP genotyping

Cheng-Hong Yang<sup>1</sup>, Yu-Huei Cheng<sup>2,\*</sup>, Li-Yeh Chuang<sup>3,\*</sup> and Hsueh-Wei Chang<sup>4,5,6,\*</sup>

<sup>1</sup>Department of Electronic Engineering, National Kaohsiung University of Applied Sciences, Kaohsiung, <sup>2</sup>Department of Digital Content Design and Management, Toko University, Chiayi, <sup>3</sup>Department of Chemical Engineering & Institute of Biotechnology and Chemical Engineering, I-Shou University, Kaohsiung, <sup>4</sup>Department of Biomedical Science and Environmental Biology, <sup>5</sup>Graduate Institute of Natural Products and <sup>6</sup>Cancer Center, Kaohsiung Medical University Hospital, Kaohsiung Medical University, Kaohsiung, Taiwan

Associate Editor: Martin Bishop

## ABSTRACT

**Summary:** Many drug or single nucleotide polymorphism (SNP)-related resources and tools have been developed, but connecting and integrating them is still a challenge. Here, we describe a user-friendly web-based software package, named Drug-SNPing, which provides a platform for the integration of drug information (DrugBank and PharmGKB), protein–protein interactions (STRING), tagSNP selection (HapMap) and genotyping information (dbSNP, REBASE and SNP500Cancer). DrugBank-based inputs include the following: (i) common name of the drug, (ii) synonym or drug brand name, (iii) gene name (HUGO) and (iv) keywords. PharmGKB-based inputs include the following: (i) gene name (HUGO), (ii) drug name and (iii) disease-related keywords. The output provides drug-related information, metabolizing enzymes and drug targets, as well as protein–protein interaction data. Importantly, tagSNPs of the selected genes are retrieved for genotyping analyses. All drug-based and protein–protein interaction-based SNP genotyping information are provided with PCR-RFLP (PCR-restriction enzyme length polymorphism) and TaqMan probes. Thus, users can enter any drug keywords/brand names to obtain immediate information that is highly relevant to genotyping for pharmacogenomics research.

**Availability and implementation:** Drug-SNPing and its user manual are freely available at <http://bio.kuas.edu.tw/drug-snping/>.

**Contact:** chuang@isu.edu.tw; yuhuei.cheng@gmail.com; changhw@kmu.edu.tw

Received on November 10, 2012; revised on December 20, 2012; accepted on January 19, 2013

## 1 INTRODUCTION

Pharmacogenomics is a type of personalized medicine that studies drug therapy in terms of efficacy or adverse events with respect to each patient's genetic variation such as single nucleotide polymorphisms (SNPs) (Daly, 2010). SNP profiling in terms of significantly informative SNPs is an important factor when carrying out drug selection, assessing dosages and deciding on a therapeutic approach. Associating drug responses with a patient's SNP genotype for genes related to the drug metabolism

and targeting is straight forward if the relevant pharmacogenomics information is readily available.

DrugBank (Knox *et al.*, 2011; Wishart, 2008; Wishart *et al.*, 2008) and PharmGKB (Gong *et al.*, 2008; Owen *et al.*, 2008; Sangkuhl *et al.*, 2008; Thorn *et al.*, 2010) are two notable databases for pharmacogenomics-related information. Although many SNP technologies associated with drug discovery have been widely discussed (Beckstead *et al.*, 2008; Chuang *et al.*, 2008; Shen *et al.*, 2009; Voisey and Morris, 2008), experimental information that allows widespread SNP genotyping is unavailable in DrugBank and PharmGKB.

Recently, many SNP-related tools, such as SNP500Cancer (Packer *et al.*, 2006), SNP-RFLPing (Chang *et al.*, 2006, 2010), Seq4SNPs (Field *et al.*, 2009), SNP ID-info (Yang *et al.*, 2008), PineSAP (Wegrzyn *et al.*, 2009), Seq-SNPing (Chang *et al.*, 2009a), MapNext (Bao *et al.*, 2009) and CandiSNPer (Schmitt *et al.*, 2010) have been developed. However, these lack the ability to allow communication between SNP genotyping information and drug targeting/metabolism/transporter/carrier. Moreover, SNP–SNP and protein–protein interactions are not considered in these SNP-related tools. In contrast, a powerful mining tool for protein–protein interaction named STRING (Jensen *et al.*, 2009) has been developed, but no SNP-related information is included. Therefore, it is still difficult to integrate cheminformatics with SNP interactivity and genotyping for pharmacogenomics purposes (Chang *et al.*, 2012).

To overcome these problems, we developed an integrated pharmacogenomics-based and protein interaction-based platform for SNP genotyping information. TagSNPs retrieved from HapMap (Deloukas and Bentley, 2004; Thorisson *et al.*, 2005), which can provide narrowed down informative SNP genotypes, have been integrated in the proposed Drug-SNPing system. This study presents a novel and user-friendly web-based pharmacogenomics tool that allows the study of drug-related SNP interactions and related genotyping.

## 2 MATERIALS AND METHODS

### 2.1 Implementation

The flow chart for the six modules used in Drug-SNPing is shown in Figure 1: (i) input module; (ii) drug info query module; (iii) tagSNPs analysis module; (iv) SNP-RFLP analysis module; (v) STRING

\*To whom correspondence should be addressed.

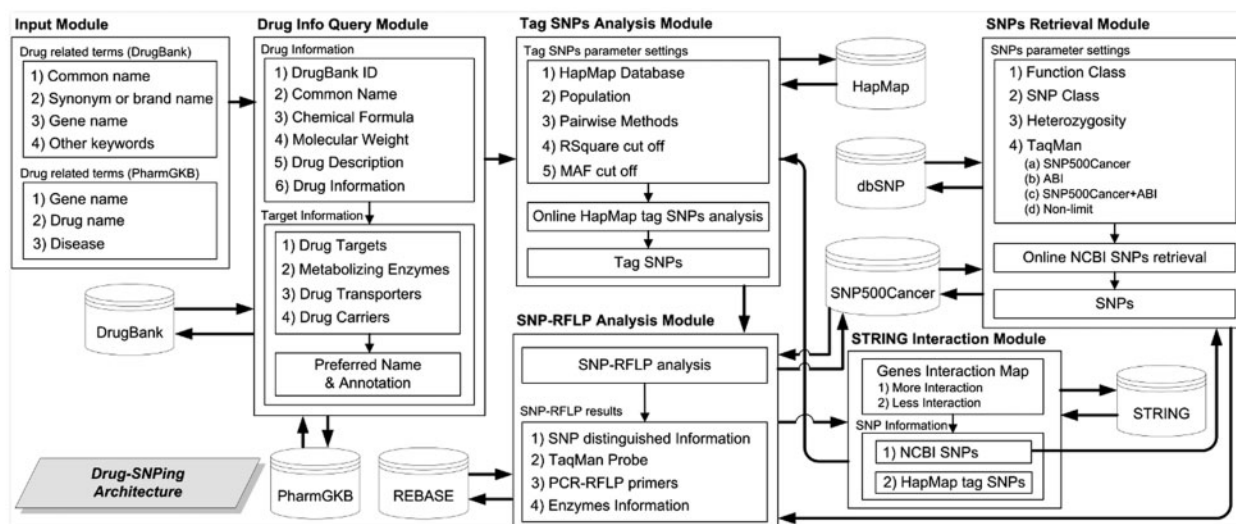


Fig. 1. System structure and flow chart for Drug-SNPing

interaction module; and (vi) SNP retrieval module. These modules are described in detail below.

(i) The input module accepts drug-related terms, including common name, synonym or brand name, gene name and other keywords based on DrugBank, as well as gene name, drug name and disease based on PharmGKB as shown in Figure 2. (ii) The drug info query module provides drug-related information (drug-orientated and gene-centric inputs) from DrugBank (Wishart, 2008; Wishart *et al.*, 2008), especially in terms of drug targets, metabolizing enzymes, transporters and carriers. (iii) In the tagSNPs analysis module, a user-selected gene is entered to retrieve the appropriate tagSNPs from HapMap (Thorisson *et al.*, 2005). The retrieval tagSNPs are fed into the SNP-RFLP analysis module for further analysis. (iv) In the SNP-RFLP analysis module, SNP distinguished information from dbSNP in NCBI (Sherry *et al.*, 2001), TaqMan probes from SNP500Cancer (Packer *et al.*, 2006) and from dbSNP in NCBI (ABI), if available, together with PCR-RFLP primers from Prim-SNPing (Chang *et al.*, 2009b), and restriction enzyme information from REBASE (Roberts *et al.*, 2010) are integrated. (v) The STRING interaction module provides visualization of the protein interaction map retrieved online from STRING (Jensen *et al.*, 2009) for the selected genes. The information from NCBI dbSNPs (Sherry *et al.*, 2001) and/or HapMap tagSNPs is available as part of the STRING map reconstructed in our system. (vi) The SNP retrieval module provides the SNP information. This information is integrated into the SNP-RFLP analysis module.

## 2.2 Database description

Some data have been previously downloaded from other databases, and some data are retrieved online. For example, DrugBank (Knox *et al.*, 2011; Wishart, 2008; Wishart *et al.*, 2008) (<http://www.drugbank.ca/downloads>) and PharmGKB (Gong *et al.*, 2008; Owen *et al.*, 2008; Sangkuhl *et al.*, 2008; Thorn *et al.*, 2010) (<http://www.pharmgkb.org/>) for input and its drug info modules, REBASE (Roberts *et al.*, 2010) (<http://rebase.neb.com/rebase/rebase.html>) for restriction enzyme mining to be used in RFLP analysis and SNP500Cancer (Packer *et al.*, 2006) (<http://variantgps.nci.nih.gov/cgfsnp/pages/snp500.do>) for the TaqMan probe information are downloaded and have been built into local databases. Some bulky databases, such as HapMap (Thorisson *et al.*, 2005) for tagSNPs information, dbSNP (Sherry *et al.*, 2001) in NCBI for information on basic SNP functions and TaqMan probes and STRING (<http://string.embl.de/>) (Jensen *et al.*, 2009) for protein

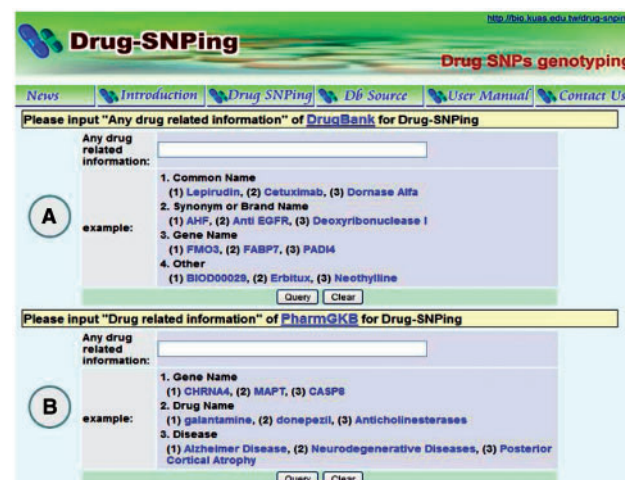


Fig. 2. Overview of Drug-SNPing web interface. Inputs for drugs are based on (A) DrugBank and (B) PharmGKB

interaction data, are accessed online, and then analysed and presented by our web platform.

## 2.3 Availability and update frequency

The Drug-SNPing website is freely accessible at <http://bio.kuas.edu.tw/drug-snping/>. Online example inputs demonstrating each of the main functions of Drug-SNPing are described in the user manual, which can be downloaded from [http://bio.kuas.edu.tw/drug-snping/user\\_manual.jsp](http://bio.kuas.edu.tw/drug-snping/user_manual.jsp). The downloaded databases are updated dynamically and built into a local database. The databases retrieved online are trimestral updated by checking the retrieval formats to provide stable online extraction.

## 2.4 Test examples

An example of synonym name such as AHF (Antithrombotic Factor), which is the coagulation factor VIII precursor or the procoagulant

component, was used to demonstrate the function of the Drug-SNPing in the result section.

### 3 RESULTS

As shown in Figure 2, input items such as common name, synonym or brand name, gene name (HUGO official gene symbol) and other keywords are line fed through the web interface for analysis in terms of drug-based SNP genotyping information. The outputs for the query results in the pipeline and for the SNP genotyping information from Drug-SNPing are shown in Figures 3 and 4, respectively.

#### 3.1 Query results in the Drug-SNPing pipeline

For the example of AHF (synonym or brand name) as an input, the query drug information is provided in Figure 3; it includes DrugBank ID with hyperlink (Fig. 3A-a1), common name, chemical formula, molecular weight, drug description and DrugBank information (drug targets, metabolizing enzymes, transporters and carriers). The DrugBank information (Fig. 3A-a2) provides a hyperlink that allows the presentation of the information on: (b1) drug targets (Coagulation factor X, Coagulation factor IX, von Willebrand factor, Phytanoyl-CoA dioxygenase peroxisomal, Asialoglycoprotein receptor 2, 78 kDa glucose-regulated protein and others), (b2) metabolizing enzymes (Prothrombin and Vitamin K-dependent protein C), (b3) drug transporters (not available) and (b4) drug carriers (not available) for the input (e.g. AHF) as shown in Figures 3B-b1, -b2, -b3 and -b4, respectively. Subsequently, a candidate of interest (e.g. Coagulation factor X) can be selected by clicking (as indicated by the arrow in Fig. 3B). Other examples, such as HUGO gene name input, are shown in the user manual.

During the pre-stage of tagSNP retrieval, some parameters in HapMap, such as HapMap database version, population, pairwise methods, R Square cut-off and minimal allele frequency (MAF) cut-off (Fig. 3C-c1), are adjustable before clicking an 'preferred name' (e.g. F10 as shown in Fig. 3C-c2).

At the retrieval stage for protein interaction and tagSNP information, the interactive proteins and their tagSNPs, which are listed on the left side and center of Figure 3D, are retrieved online from STRING (Jensen *et al.*, 2009) and HapMap (Thorisson *et al.*, 2005), respectively. All these gene names and tagSNPs can be hyperlinked to Entrez Gene and dbSNP in NCBI, respectively. The PharmGKB info button provides a hyperlink to access all kinds of pharmacogenomics-related information stored in PharmGKB. The genotype information for tagSNP (obtained via SNP-RFLP button) is provided as described later in Figure 4. The 'STRING interaction' box at the bottom of Figure 3D (as indicated by the arrow) provides further STRING interactions.

As shown in Figure 3E, proteins that interact to a greater or lesser extent in relation to the selected genes associated with the drug metabolism can be retrieved online from STRING (Jensen *et al.*, 2009). Moreover, STRING-based visualization (Fig. 3E-e3) is implemented with a hyperlink to our developed SNP genotyping information system (as described later in Fig. 4) rather than the original STRING function for protein annotation. This SNP genotype information is based on the settings for

NCBI SNPs and HapMap tagSNPs (Fig. 3E-e1 and -e2, respectively). Many SNP sub-functions in dbSNP of NCBI are analysed in Drug-SNPing, e.g. functional class, SNP class, heterozygosity and TaqMan resources [SNP500Cancer (Packer *et al.*, 2006) and ABI in NCBI dbSNP] for SNP genotyping information.

#### 3.2 Outputs of SNP genotyping information

In the example, rs2740171 (Fig. 4A) was selected to show the SNP genotyping information for the SNP fasta sequence (as shown in Fig. 4A-a1) and for the TaqMan probe from dbSNP in NCBI (Applied Biosystems; ABI) (Fig. 4A-a2). A hyperlink is available to enter the TaqMan probe for ABI as shown in Figure 4A-a2. Natural primers and RFLP restriction enzymes for traditional PCR-RFLP analysis are provided in Figures 4A-a3 and -a4, respectively. In Figure 4A-a3, the locations of the natural primers are visualized in the SNP-containing sequences. In Figure 4A-a4, information in relation to RFLP restriction enzymes for both the sense and antisense strands are available via REBASE (Roberts *et al.*, 2010).

In another example, rs1051740 (Fig. 4B) was selected to show the SNP genotyping information for mismatch (mutagenic) PCR-RFLP. The TaqMan probe information, including sequences for the paired TaqMan probe and the forward/reverse primers, was retrieved online from SNP500Cancer (Packer *et al.*, 2006) (Fig. 4B-b1). The enzyme information for PCR-RFLP using mutagenic primers is also available (Fig. 4A-b2).

For both natural and mutagenic primers (Fig. 4A-a3 and -b2), much SNP genotyping information, such as primer sequences, positions, primer lengths, GC%, T<sub>m</sub> values, T<sub>m</sub> value differences and product lengths after PCR-RFLP analyses, are also available from Drug-SNPing.

#### 3.3 Example scenarios

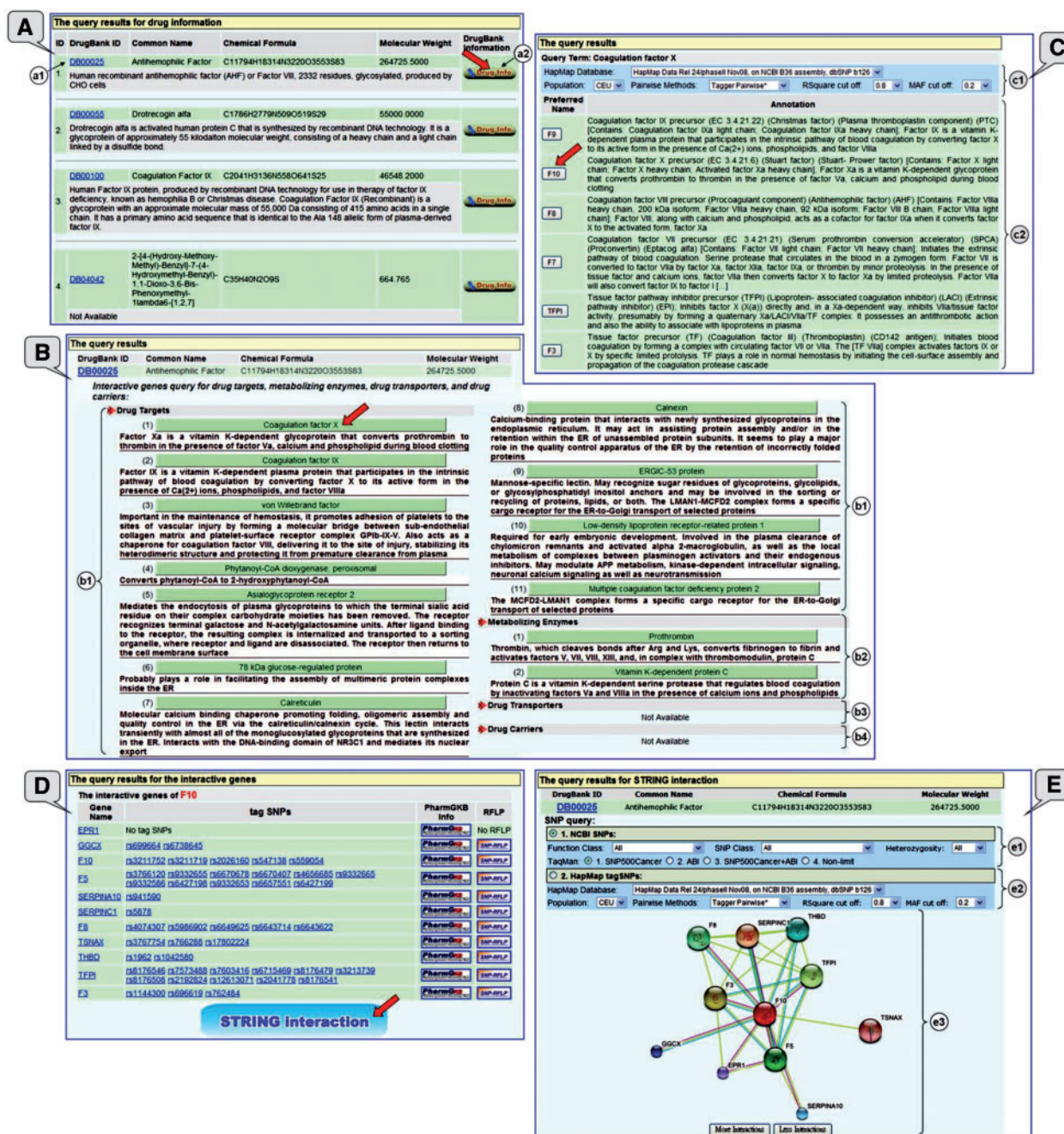
Seventeen example scenarios with animations were provided online ([http://bio.kuas.edu.tw/drug-snping/user\\_manual.jsp](http://bio.kuas.edu.tw/drug-snping/user_manual.jsp)), including: (i) The query results for drug information; (ii) The query results for the drug targets, metabolizing enzymes, drug transporters and drug carriers; (iii) The query results for the interactive genes; (iv) The query results for STRING interaction; (v) The query results for STRING interaction with more interactions; (vi) The query results for STRING interaction with less interactions; (vii) NCBI SNP selection for SNP genotyping information; (viii) HapMap tagSNPs for SNP genotyping information; (ix) Example of result for SNP fasta information; (x) Example of result for TaqMan probe information; (xi) Example of result for Natural primer information; (xii) Example of result for Enzyme Information; (xiii) Example of result for Mutagenic primer information; (xiv) TagSNPs genotyping information; (xv) One tagSNP genotyping information; (xvi) PharmGKB information; and (xvii) PharmGKB STRING interaction.

### 4 DISCUSSION

SNP-based association studies are common indicators used in personalized medicine, especially for pharmacogenomics.

Although many pharmacogenomics resources (Burgarella *et al.*, 2005; Burgoon *et al.*, 2006; Freimuth *et al.*, 2005;





**Fig. 3.** Representative query results in the pipeline of Drug-SNPing. **(A)** Drug information for input. (a1) Hyperlinking to the DrugBank. (a2) DrugBank information to the Drug-SNPing. **(B)** The query results for (b1) drug targets, (b2) metabolizing enzymes, (b3) drug transporters and (b4) drug carriers. **(C)** Condition setting for the online analysis of STRING-based protein interaction mining and HapMap-based tagSNP retrieval. (c1) TagSNP and (c2) preferred name selections for analyses. **(D)** The query results for the interactive genes and their corresponding tagSNPs and genotyping information. **(E)** The query results for more or less STRING interactions in an interactive way. (e1) SNP query setting for dbSNP in NCBI, (e2) HapMap tagSNPs and (e3) Interactive protein interaction for SNPs or tagSNP selection are included. The results of the arrows in Figures 3A, 3B, 3C and 3D are shown in Figures 3B, 3C, 3D and 3E, respectively

Hayes *et al.*, 2005; Hug *et al.*, 2003; Mattingly *et al.*, 2003, 2006; Salter, 2005; Sangkuhl *et al.*, 2008; Sun *et al.*, 2002; Tong *et al.*, 2003; Wishart, 2007; Wishart *et al.*, 2008; Xirasagar *et al.*, 2006; Zheng *et al.*, 2007) and chemoinformatics databases (Chen *et al.*, 2005; Masciocchi *et al.*, 2009; Miteva *et al.*, 2006; Neugebauer *et al.*, 2007; Timmers *et al.*, 2008; Wang *et al.*, 2007) have been

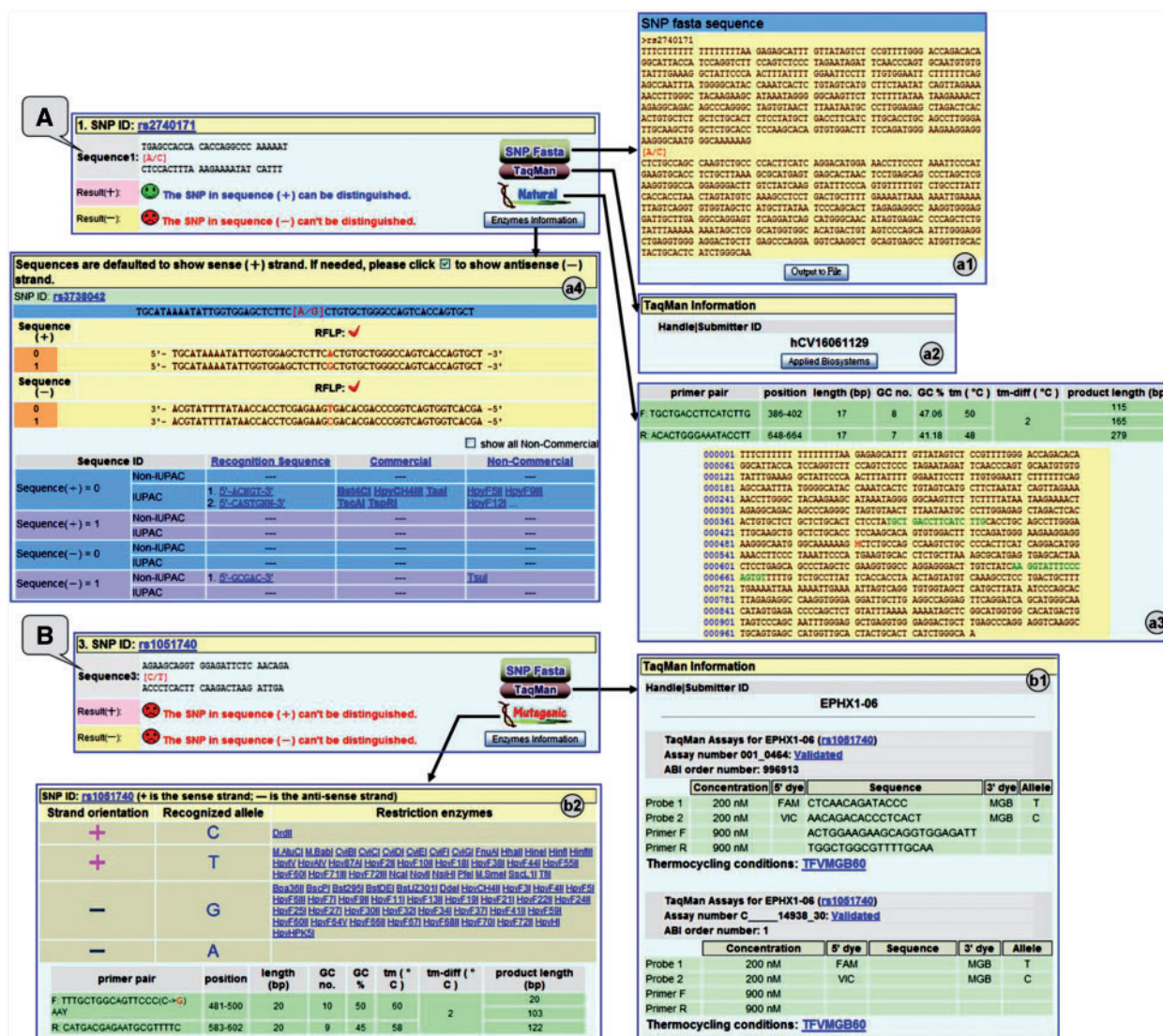


Fig. 4. Representative outputs for SNP genotyping information in Drug-SNPing. (A) Natural primers for rs2740171. (B) Mutagenic primers for rs1051740. SNP genotyping information such as (a1): SNP fasta sequence, (a2/b1): TaqMan probe from SNP500Cancer/NCBI, (a3/b2): natural/mutagenic RFLP primers and (a4/b2): restriction enzymes for natural/mutagenic RFLP primers are shown

(Royo and Galan, 2009) and TaqMan (Shen *et al.*, 2009). However, the materials for SNP array analysis and the machine for pyrosequencing are more expensive than the TaqMan method. Therefore, we selected the TaqMan probe information for implementation in the Drug-SNPing. Currently, two major resources for SNP-related TaqMan probe information are available, namely SNP500Cancer and dbSNP in NCBI (ABI), but neither can provide TaqMan probe information for all SNPs. Furthermore, TaqMan probes are sometimes unsuitable for genotyping of some SNPs (Demirci *et al.*, 2007; Kocsis *et al.*, 2008; Yoshiya *et al.*, 2008). To compensate for such problems, we also implemented a RFLP genotyping module in Drug-SNPing. When implementing RFLP genotyping, some SNPs do not have natural primers (regular PCR-RFLP primers). As an alternative, we introduced mutagenic primers whereby the module

computationally introduces mismatched nucleotides into the forward (F) primer sequence to artificially generate a new restriction enzyme site for SNP genotyping. The lack of natural primers can be overcome by the use of mutagenic primers, as can be seen for rs1051740 in Figure 4B. Other examples are presented in the user manual.

Although DrugBank (Wishart *et al.*, 2008) contains a versatile search function for many inputs, its gene name input for 'drug target' is limited to the full name rather than the HUGO official gene symbol. For example, FABP7 (fatty acid-binding protein, brain), PADI4 (protein-arginine deiminase type 4) and PTGS2 (prostaglandin G/H synthase 2) are listed in the 'drug cards' of DB00132, DB00155 and DB00154 in DrugBank, respectively. But for gene name (HUGO) inputs, such as FABP7, PADI4 and PTGS2, the appropriate drug records related to



these genes can not be found, e.g. <http://www.drugbank.ca/search/search?query=FABP7>, <http://www.drugbank.ca/search/search?query=PADI4> and <http://www.drugbank.ca/search/search?query=PTGS2>. In contrast, Drug-SNPing allows input based on the HUGO gene name, which improves the gene-centric search for drugs with input drug targets. This is demonstrated in the user manual for Drug-SNPing. In Drug-SNPing, we also integrated PharmGKB (Gong *et al.*, 2008; Owen *et al.*, 2008; Sangkuhl *et al.*, 2008; Thorn *et al.*, 2010) for annotation of important genes related to drug responses and pathways as a complement to DrugBank. The contributions in Drug-SNPing, PharmGKB and DrugBank are shown in Table 1.

The importance of SNP-SNP interactions in association studies is increasing, but most studies only focus their data analysis on limited and known SNPs (Lin *et al.*, 2008, 2009; Yang *et al.*, 2009, 2012; Yen *et al.*, 2008; Zheng *et al.*, 2008). To fill this gap, we incorporated a widely used protein-protein interaction tool, STRING (Jensen *et al.*, 2009), in Drug-SNPing. We perform SNP-SNP interaction analysis based on the protein-protein interaction data from STRING via online retrieval. In the current version of Drug-SNPing, however, many other protein-protein interaction tools are not yet included, such as GWIDD (Kundrotas *et al.*, 2010) for genome-wide protein docking database, STITCH 2 (Kuhn *et al.*, 2010) for an interaction network database for small molecules and proteins, Gene Interaction Miner (Ikin *et al.*, 2010) for data mining contextual information for protein-protein interaction analysis, Path (Zamar *et al.*, 2009) to facilitate pathway-based genetic association analysis and Polymorphism Interaction Analysis (PIA) (Mechanic *et al.*, 2008) as a method for investigating complex gene-gene interactions, and protein-protein interaction databases (Lehne and Schlitt, 2009). Similarly, more drug-related chemoinformatics resources are also not yet included in the current version of Drug-SNPing, such as DCDB (Liu *et al.*, 2010) as a drug combination database, drug-binding databases (Timmers *et al.*, 2008), SuperSite (Andre Bauer *et al.*, 2008) as a dictionary of metabolite and drug-binding sites in proteins and the drug adverse reaction target database (DART) (Ji *et al.*, 2003) for proteins related to adverse drug reactions. In future, we intend

to integrate the available ftp data released in these tools in order to add more value to the Drug-SNPing product.

The performance of the Drug-SNPing was assessed by at least 50 different inputs (see the section of 'Assessment of Drug-SNPing' at [http://bio.kuas.edu.tw/drug-snping/user\\_manual.jsp](http://bio.kuas.edu.tw/drug-snping/user_manual.jsp)), and all of them were functional to get the output by Drug-SNPing. Moreover, the operating time for integration in Drug-SNPing can save time compared with non-integration.

## 5 CONCLUSION

In this article, we developed a novel integrated web-based interface, Drug-SNPing, that provides user-friendly connections between the tools DrugBank for chemoinformatics, pharmGKB for pharmacogenomics, STRING for protein-protein interaction, genotyping information for TaqMan probes and PCR-RFLP. Moreover, gene-centric inputs can also be used as a means of finding corresponding drugs for drug targeting and metabolism. This tool thus provides all the drug-orientated and gene-centric inputs which are needed to mine all possible metabolic enzymes and targets for these drugs, as well as bench mark information that allows SNP genotyping to be performed. This platform will be helpful for the development of pharmacogenomics as part of personalized medicine.

**Funding:** National Science Council in Taiwan (grants NSC101-2622-E-151-027-CC3, NSC100-2221-E-214-071, NSC98-2622-E-151-001-CC2 and NSC101-2221-E-464-001, and the funds of the Department of Health, Executive Yuan, R.O.C. (TAIWAN) (DOH102-TD-C-111-002); the Kaohsiung Medical University Research Foundation (KMUER001); the NSYSU-KMU JOINT RESEARCH PROJECT (#NSYSUKMU 102-034; and the I-Shou University plan (ISU100-02-05).

**Conflict of Interest:** none declared.

## REFERENCES

- Andre Bauer, R. *et al.* (2008) SuperSite: dictionary of metabolite and drug binding sites in proteins. *Nucleic Acids Res.*, **37**, D195–D200.
- Bao, H. *et al.* (2009) MapNext: a software tool for spliced and unspliced alignments and SNP detection of short sequence reads. *BMC Genomics*, **10** (Suppl. 3), S13.
- Beckstead, W.A. *et al.* (2008) SNP2RFLP: a computational tool to facilitate genetic mapping using benchtop analysis of SNPs. *Mamm. Genome*, **19**, 687–690.
- Burgarella, S. *et al.* (2005) MicroGen: a MIAME compliant web system for microarray experiment information and workflow management. *BMC Bioinformatics*, **6** (Suppl. 4), S6.
- Burgoon, L.D. *et al.* (2006) dbZach: a MIAME-compliant toxicogenomic supportive relational database. *Toxicol. Sci.*, **90**, 558–568.
- Chang, H.W. *et al.* (2006) SNP-RFLPing: restriction enzyme mining for SNPs in genomes. *BMC Genomics*, **7**, 30.
- Chang, H.W. *et al.* (2009a) Seq-SNPing: multiple-alignment tool for SNP discovery, SNP ID identification, and RFLP genotyping. *OMICS*, **13**, 253–260.
- Chang, H.W. *et al.* (2009b) Prim-SNPing: a primer designer for cost-effective SNP genotyping. *Biotechniques*, **46**, 421–431.
- Chang, H.W. *et al.* (2010) SNP-RFLPing 2: an updated and integrated PCR-RFLP database tool for SNP genotyping. *BMC Bioinformatics*, **11**, 173.
- Chang, H.W. *et al.* (2012) The importance of integrating SNP and cheminformatics resources to pharmacogenomics. *Curr. Drug Metab.*, **13**, 991–999.
- Chen, X. and Sullivan, P.F. (2003) Single nucleotide polymorphism genotyping: biochemistry, protocol, cost and throughput. *Pharmacogenomics J.*, **3**, 77–96.
- Chen, J. *et al.* (2005) ChemDB: a public database of small molecules and related cheminformatics resources. *Bioinformatics*, **21**, 4133–4139.

**Table 1.** The contributions in Drug-SNPing, PharmGKB and DrugBank

Function	Drug-SNPing	PharmGKB	DrugBank
System			
Disease information	✓	✓	✗
TaqMan assays information	✓	✗	✗
PCR-RFLP primers	✓	✗	✗
Protein-protein interaction	✓	✗	✗
Clinical interpretations	✗	✓	✗
Pathways	✗	✓	✓
Rich resources and database download	✗	✓	✓
Drug target information	✓	✗	✓
Chemical formula query	✓	✗	✓

**Note:** Symbol '✓' means that the function is provided, and symbol '✗' means that the function is not provided.

- Chuang, L.Y. et al. (2008) Restriction enzyme mining for SNPs in genomes. *Anticancer Res.*, **28**, 2001–2007.
- Daly, A.K. (2010) Genome-wide association studies in pharmacogenomics. *Nat. Rev.*, **11**, 241–246.
- Deloukas, P. and Bentley, D. (2004) The HapMap project and its application to genetic studies of drug response. *Pharmacogenomics J.*, **4**, 88–90.
- Demirci, F.Y. et al. (2007) Association study of Toll-like receptor 5 (TLR5) and Toll-like receptor 9 (TLR9) polymorphisms in systemic lupus erythematosus. *J. Rheumatol.*, **34**, 1708–1711.
- Field, H.I. et al. (2009) Seq4SNPs: new software for retrieval of multiple, accurately annotated DNA sequences, ready formatted for SNP assay design. *BMC Bioinformatics*, **10**, 180.
- Freimuth, R.R. et al. (2005) PolyMAPr: programs for polymorphism database mining, annotation, and functional analysis. *Hum. Mutat.*, **25**, 110–117.
- Gong, L. et al. (2008) PharmGKB: an integrated resource of pharmacogenomic data and knowledge. *Curr. Protoc. Bioinformatics*, **Chapter 14**, Unit 14.17.
- Hayes, K.R. et al. (2005) EDGE: a centralized resource for the comparison, analysis, and distribution of toxicogenomic information. *Mol. Pharmacol.*, **67**, 1360–1368.
- Hug, H. et al. (2003) ADRIS—the adverse drug reactions information scheme. *Pharmacogenetics*, **13**, 767–772.
- Ikin, A. et al. (2010) The Gene Interaction Miner: a new tool for data mining contextual information for protein-protein interaction analysis. *Bioinformatics*, **26**, 283–284.
- Jensen, L.J. et al. (2009) STRING 8—a global view on proteins and their functional interactions in 630 organisms. *Nucleic Acids Res.*, **37**, D412–D416.
- Ji, Z.L. et al. (2003) Drug Adverse Reaction Target Database (DART): proteins related to adverse drug reactions. *Drug Saf.*, **26**, 685–690.
- Knox, C. et al. (2011) DrugBank 3.0: a comprehensive resource for ‘Omics’ research on drugs. *Nucleic Acids Res.*, **39**, D1035–D1041.
- Kocsis, A.K. et al. (2008) Association of beta-defensin 1 single nucleotide polymorphisms with Crohn’s disease. *Scand. J. Gastroenterol.*, **43**, 299–307.
- Kuhn, M. et al. (2010) STITCH 2: an interaction network database for small molecules and proteins. *Nucleic Acids Res.*, **38**, D552–D556.
- Kundrotas, P.J. et al. (2010) GWIDD: genome-wide protein docking database. *Nucleic Acids Res.*, **38**, D513–D517.
- Lehne, B. and Schlitt, T. (2009) Protein-protein interaction databases: keeping up with growing interactomes. *Hum. Genomics*, **3**, 291–297.
- Lin, G.T. et al. (2008) SNP combinations in chromosome-wide genes are associated with bone mineral density in Taiwanese women. *Chin. J. Physiol.*, **91**, 1–10.
- Lin, G.T. et al. (2009) Combinational polymorphisms of seven CXCL12-related genes are protective against breast cancer in Taiwan. *OMICS*, **13**, 165–172.
- Liu, Y. et al. (2010) DCDB: drug combination database. *Bioinformatics*, **26**, 587–588.
- Masciocchi, J. et al. (2009) MMsINC: a large-scale chemoinformatics database. *Nucleic Acids Res.*, **37**, D284–D290.
- Mattingly, C.J. et al. (2003) The Comparative Toxicogenomics Database (CTD). *Environ. Health Perspect.*, **111**, 793–795.
- Mattingly, C.J. et al. (2006) The Comparative Toxicogenomics Database (CTD): a resource for comparative toxicological studies. *J. Exp. Zool. A Comp. Exp. Biol.*, **305**, 689–692.
- Mechanic, L.E. et al. (2008) Polymorphism Interaction Analysis (PIA): a method for investigating complex gene-gene interactions. *BMC Bioinformatics*, **9**, 146.
- Miteva, M.A. et al. (2006) FAF-Drugs: free ADME/tox filtering of compound collections. *Nucleic Acids Res.*, **34**, W738–W744.
- Neugebauer, A. et al. (2007) Prediction of protein-protein interaction inhibitors by chemoinformatics and machine learning methods. *J. Med. Chem.*, **50**, 4665–4668.
- Owen, R.P. et al. (2008) PharmGKB and the International Warfarin Pharmacogenetics Consortium: the changing role for pharmacogenomic databases and single-drug pharmacogenetics. *Hum. Mutat.*, **29**, 456–460.
- Packer, B.R. et al. (2006) SNP500Cancer: a public resource for sequence validation, assay development, and frequency analysis for genetic variation in candidate genes. *Nucleic Acids Res.*, **34**, D617–D621.
- Peters, E.J. and McLeod, H.L. (2008) Ability of whole-genome SNP arrays to capture ‘must have’ pharmacogenomic variants. *Pharmacogenomics*, **9**, 1573–1577.
- Roberts, R.J. et al. (2010) REBASE—a database for DNA restriction and modification: enzymes, genes and genomes. *Nucleic Acids Res.*, **38**, D234–D236.
- Royo, J.L. and Galan, J.J. (2009) Pyrosequencing for SNP genotyping. *Methods Mol. Biol.*, **578**, 123–133.
- Salter, A.H. (2005) Large-scale databases in toxicogenomics. *Pharmacogenomics*, **6**, 749–754.
- Sanguhl, K. et al. (2008) PharmGKB: understanding the effects of individual genetic variants. *Drug Metab. Rev.*, **40**, 539–551.
- Schmitt, A.O. et al. (2010) CandiSNPer: a web tool for the identification of candidate SNPs for causal variants. *Bioinformatics*, **26**, 969–970.
- Shen, G.Q. et al. (2009) The TaqMan method for SNP genotyping. *Methods Mol. Biol.*, **578**, 293–306.
- Sherry, S.T. et al. (2001) dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.*, **29**, 308–311.
- Sun, L.Z. et al. (2002) ADME-AP: a database of ADME associated proteins. *Bioinformatics*, **18**, 1699–1700.
- Thorisson, G.A. et al. (2005) The International HapMap Project Web site. *Genome Res.*, **15**, 1592–1593.
- Thorn, C.F. et al. (2010) Pharmacogenomics and bioinformatics: PharmGKB. *Pharmacogenomics*, **11**, 501–505.
- Timmers, L.F. et al. (2008) Drug-binding databases. *Curr. Drug Targets*, **9**, 1092–1099.
- Tong, W. et al. (2003) ArrayTrack—supporting toxicogenomic research at the U.S. Food and Drug Administration National Center for Toxicological Research. *Environ. Health Perspect.*, **111**, 1819–1826.
- Voisey, J. and Morris, C.P. (2008) SNP technologies for drug discovery: a current review. *Curr. Drug Discov. Technol.*, **5**, 230–235.
- Wang, H. et al. (2007) Chemical data mining of the NCI human tumor cell line database. *J. Chem. Inf. Model.*, **47**, 2063–2076.
- Wegrzyn, J.L. et al. (2009) PineSAP—sequence alignment and SNP identification pipeline. *Bioinformatics*, **25**, 2609–2610.
- Wishart, D.S. (2007) In silico drug exploration and discovery using DrugBank. *Curr. Protoc. Bioinformatics*, **Chapter 14**, Unit 14.14.
- Wishart, D.S. (2008) DrugBank and its relevance to pharmacogenomics. *Pharmacogenomics*, **9**, 1155–1162.
- Wishart, D.S. et al. (2008) DrugBank: a knowledgebase for drugs, drug actions and drug targets. *Nucleic Acids Res.*, **36**, D901–D906.
- Xirasagar, S. et al. (2006) Chemical effects in biological systems (CEBS) object model for toxicology data, SysTox-OM: design and application. *Bioinformatics*, **22**, 874–882.
- Yang, C.H. et al. (2008) SNP ID-info: SNP ID searching and visualization platform. *OMICS*, **12**, 217–226.
- Yang, C.H. et al. (2009) Novel generating protective single nucleotide polymorphism barcode for breast cancer using particle swarm optimization. *Cancer Epidemiol.*, **33**, 147–154.
- Yang, C.H. et al. (2012) Single nucleotide polymorphism barcoding to evaluate oral cancer risk using odds ratio-based genetic algorithms. *Kaohsiung J. Med. Sci.*, **28**, 362–368.
- Yen, C.Y. et al. (2008) Combinational polymorphisms of four DNA repair genes XRCC1, XRCC2, XRCC3, and XRCC4 and their association with oral cancer in Taiwan. *J. Oral Pathol. Med.*, **37**, 271–277.
- Yoshiya, G. et al. (2008) Influence of cancer-related gene polymorphisms on clinicopathological features in colorectal cancer. *J. Gastroenterol. Hepatol.*, **23**, 948–953.
- Zamar, D. et al. (2009) Path: a tool to facilitate pathway-based genetic association analysis. *Bioinformatics*, **25**, 2444–2446.
- Zheng, C.J. et al. (2007) PharmGED: pharmacogenetic effect database. *Nucleic Acids Res.*, **35**, D794–D799.
- Zheng, S.L. et al. (2008) Cumulative association of five genetic variants with prostate cancer. *N Engl. J. Med.*, **358**, 910–919.