Original Paper

Public Health Genomics

Public Health Genomics 2017;20:142–147
DOI: 10.1159/000478254

Received: May 31, 2017
Accepted: June 9, 2017
Published online: July 14, 2017

Application of the DruGeVar Database in Cancer Genomics and Pharmacogenomics

Konstantinos Sarris[a] Angeliki Komianou[b] George P. Patrinos[a, c]
Theodora Katsila[a]

[a]Department of Pharmacy, School of Health Sciences, University of Patras, Patras, Greece; [b]The Institute of Cancer Research, London, UK; [c]Department of Pathology, College of Medicine and Health Sciences, United Arab Emirates University, Al Ain, UAE

Keywords
Databases · DruGeVar · Pharmacogenomics · POWER-BI · Visualization interface

Abstract
In the post-genomic era, there is an increasing and urgent need for managing and visualizing big data. Data complexity and size will turn information growth into knowledge growth only if presented in a comprehensive and user-friendly way. In such a context, the information technology community collaborates in a multidisciplinary manner with other scientific fields searching for and/or developing tools and services for data management and visualization. We have previously developed DruGeVar, a comprehensive database that triangulates drugs with genes and pharmacogenomic biomarkers to serve clinical pharmacogenomics. To empower its functionality, we explored and implemented new visualization tools, such as POWER-BI (Microsoft), which allows for interactive visualization. Herein, we describe the synergy of POWER-BI and DruGeVar, focusing on cancer genomics data in light of translational pharmacogenomics.

Introduction

The post-genomic era dictates us to exploit big data and to develop and optimize information tools and databases [1]. To turn such information growth into knowledge growth we need to achieve data interpretation. For this, data should be handled in such a way to ensure data quality, robustness, and reproducibility via user-friendly interfaces. Indeed, there are no ideal means, just good ones for users with certain goals and needs.

Data visualization is a communication form used already by several multidisciplinary subfields, such as science and/or business [2, 3] and thus, holds the promise of interoperability. For this, if we aim for the most readable and at the same time user-friendly and comprehensive means of handling data, particularly in the context of a database, we need to empower data visualization. The latter strongly depends on the information that needs to be handled. Additionally, one shall distribute a user-friendly interface from the database to the users towards efficient data exploitation. The management and use of information alongside with the virtual devices in mind tend to be united. Today, the "Internet of things" is an umbrella key-

E-Mail karger@karger.com
www.karger.com/phg

Theodora Katsila
Department of Pharmacy, School of Health Sciences
University of Patras, University Campus, Rion
GR–265 04 Patras (Greece)
E-Mail thkatsila@upatras.gr

Table 1. Current visualization tools

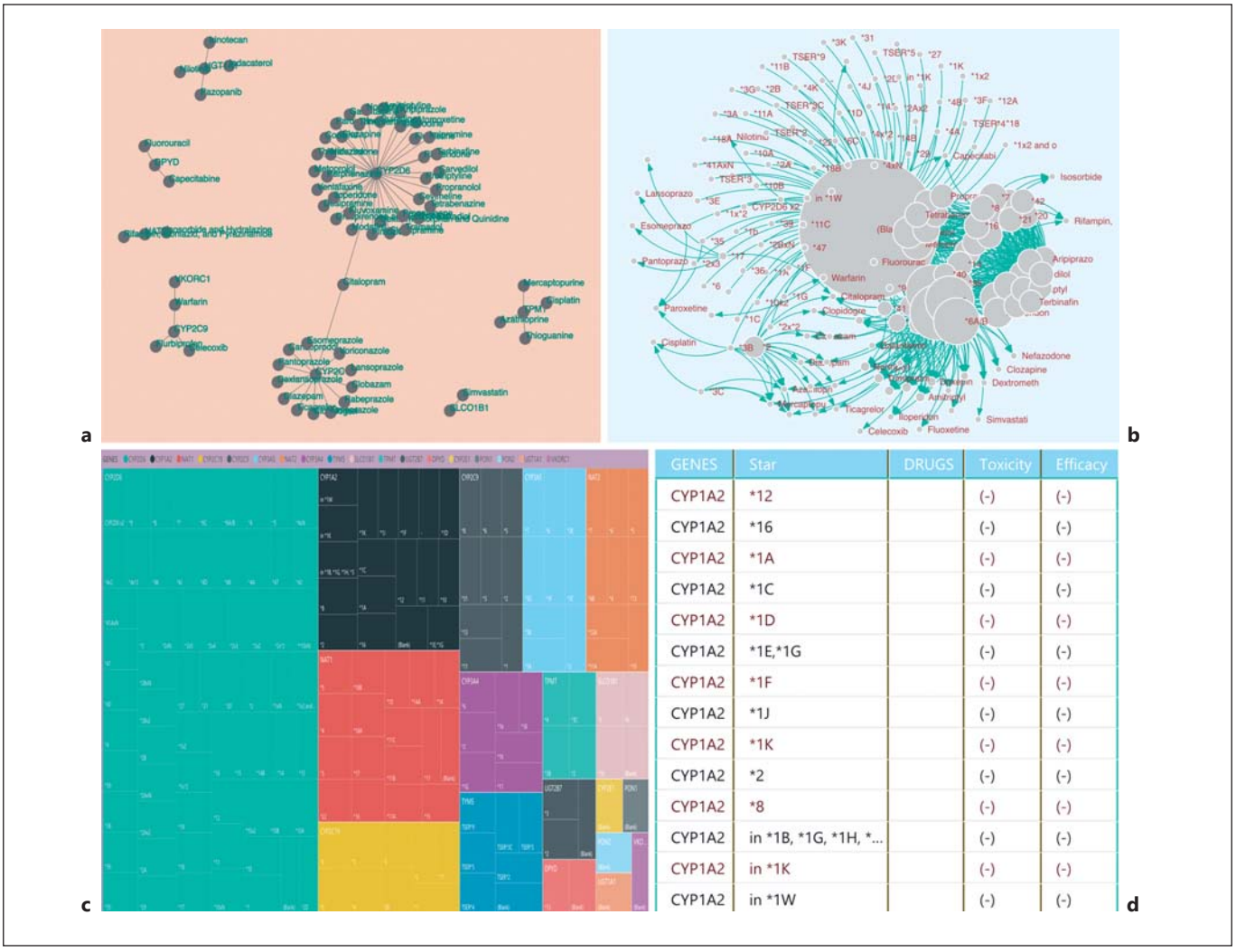
Data visualization tools (URL)	Interactivity	Advantages	Disadvantages
Matplotlib (http://matplotlib.org)	Yes, but not effective	Easy creation of graphs, using a few lines of code	Low interactivity (big databases)
DrasticData (http://www.drasticdata.nl)	Yes	Several technologies are supported	Low flexibility (big databases) Treemap (main use)
D3.js, SVG (https://d3js.org)	Yes	A large range of graphs	Complexity
PhiloGL (http://www.senchalabs.org/philogl)	Yes	3-D graphs Clear and well-documented code	Abstract logic, mainly used for games/interactive webpages
HighCharts (http://www.highcharts.com)	Yes, but not effective	Open source for developers Easy changes Flexible for all browsers	Not efficient interactivity A small range of graphs (bar chart, pies)

word that describes the extension of the Web and the Internet into the physical realm, as spatially distributed devices with embedded identification, sensing, and/or actuation capabilities are widely deployed [4]. Thus, even though at its infancy, the Internet of things shares the vision of a future in which digital and physical entities can be linked, on the basis of appropriate information and communication technologies, enabling new services and applications [5]. This presents unmet advances for decision-making.

Focusing on data visualization, the concept is not new. People have always been using pictures to interpret and communicate data (from maps and graphs in the 17th century to pie charts in the early 1800s) [6, 7]. Mapping Napoleon's invasion of Russia is one of the most cited examples of statistical graphics, as C. Minard depicted the path of Napoleon's retreat from Moscow and the size of the army and linked that information to temperature and time scales [8, 9]. Notwithstanding, it is technology itself that truly revolutionized data visualization as we are thinking of today. Since computers allowed data processing at lightning-fast speeds, data visualization is currently considered as the rapidly evolving blend of science and art that already affects the corporate landscape [10].

Currently, several efforts of visualizing data along with a minimum comparison among them are ongoing (summarized in Table 1). No doubt, every technology exhibits advantages and disadvantages. Depending on the meaning and, of course, the size, use, needs, and aims of a database in question, the most suitable technology should be chosen. This becomes a tricky situation that often confuses code developers for making the right choice. DruGeVar (http://www.genomicmedicinealliance.org/working-groups/pharmacogenomics/80-genomic-medicine-alliance/137-drugevar-database) [11] is a comprehensive database that triangulates drugs-genes-pharmacogenomic (PGx) biomarkers with the aim of serving clinical pharmacogenomics. DruGeVar has been developed to be either a database plug-in module or a standalone resource [12]. This application allows querying large datasets in a very dynamic and user-friendly manner and in multiple ways. Database records include the drug, the gene name, and its genomic variation parameters (in the official dbSNP [http://www.ncbi.nlm.nih.gov/projects/SNP], Human Genome Variation Society [http://www.hgvs.org], and star allele nomenclatures), linked to the Online Mendelian Inheritance in Man (OMIM; http://www.omim.org) and PharmGKB entries [11].

Considering the fact that DruGeVar consists of a large number of data that further complicates data interactivity and data management, herein, we aimed to enhance the battery of the database's data visualization tools that would not only maximize its user-friendliness but also allow building interfaces with other resources. Here, we describe the synergy of POWER-BI and DruGeVar to achieve (i) easy access to data and (ii) a flexible structure that provides effortlessly the designing pattern of the requested graph. In light of translational pharmacogenomics, we will focus on cancer genomics data and describe the benefits, after integrating POWER-BI into DruGeVar.



GENES	Star	DRUGS	Toxicity	Efficacy
CYP1A2	*12		(-)	(-)
CYP1A2	*16		(-)	(-)
CYP1A2	*1A		(-)	(-)
CYP1A2	*1C		(-)	(-)
CYP1A2	*1D		(-)	(-)
CYP1A2	*1E,*1G		(-)	(-)
CYP1A2	*1F		(-)	(-)
CYP1A2	*1J		(-)	(-)
CYP1A2	*1K		(-)	(-)
CYP1A2	*2		(-)	(-)
CYP1A2	*8		(-)	(-)
CYP1A2	in *1B, *1G, *1H, *...		(-)	(-)
CYP1A2	in *1K		(-)	(-)
CYP1A2	in *1W		(-)	(-)

Fig. 1. Visualization of interactive drug-gene-pharmacogenomic (PGx) biomarkers outcomes. **a** Spider graph depicting drug-gene relations in a fully interactive way that allows users to select for a gene or a drug and collect all available information (the search option is activated) via the immediate dynamic change of all charts (**b–d**). **b** Drug-PGx markers network visualization of outcomes (the search option is inactivated), again summarizing all available information in a fully interactive way. **c** Treemap, describing genes-PGx markers relations (the search option for "genes" is activated) in a fully interactive way. **d** Excel-like tabular graph for genes-star alleles-drugs relations linked to efficacy/toxicity information. (+) = "yes"; (–) = "no".

Methods

POWER-BI Data Integration

To achieve the emergence of specific data fields, DruGeVar data were integrated into Power-BI. Power-BI consists of software services, applications, and connectors, which synergize to turn unrelated data sources into coherent, visually immersive, and interactive insights. Irrespective of the data format, Power-BI facilitates (i) an easy connection to data sources, (ii) data visualization and/or discovery, and (iii) data sharing (https://powerbi.microsoft.com/en-us/). Data input may refer to a simple Excel spreadsheet or cloud-based and on-premises hybrid data warehouses.

Create and Exploit Graphs

We have subsequently focused on DruGeVar cancer genomics data and started to build our graphs to exploit the benefits of POWER-BI at the highest level. First, we aimed to empower data visualization (drugs) towards an easy search functionality. For this, drugs were incorporated as a panel to achieve improved search functionality. We made the search engine as handy as possible to distribute the details of each drug upon consideration and then, we created additional graphs for data analysis. We provided the required fields comprehensively and grouped for better understanding. Following a two-level division, POWER-BI enabled the analysis of fields of data as a group of two objects. Next, all field outcomes could be compared, as the searching machinery adjusts to

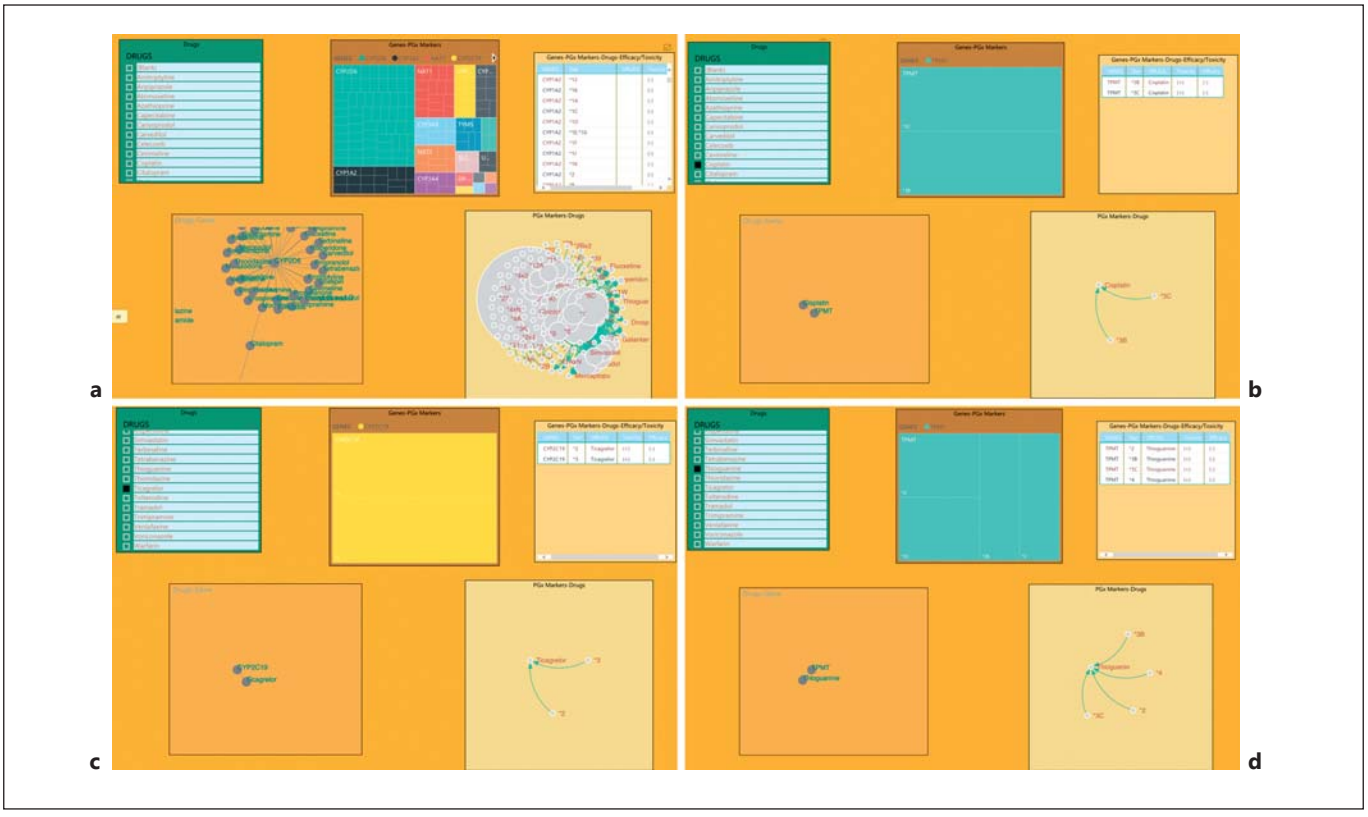


Fig. 2. POWER-BI integration into DruGeVar focusing on oncology drugs. **a** A user interface devoid of any query. **b** When a drug of interest is selected (in this case, cisplatin), the user can visualize interactive graphs and all related available information. The graphs respond immediately and present the relations, according to the selected drug. **c** When a gene of interest is selected (in this case, *CYP2C19*), the user can visualize interactive graphs and all related available information (in this case, gene-pharmacogenomic (PGx) markers-drugs-efficacy/toxicity), also linked to a drug of interest (in this case, ticagrelor). **d** Advanced relation of "genes" with "PGx markers" and "drugs", including toxicity and efficacy outcomes, can be visualized for the drug of interest (in this case, thioguanine). Note the PGx markers-drug relationships depending on star alleles.

every result point automatically. Such groups of information that relate to fields of data are depicted in Figure 1: drugs and genes, drugs and PGx markers, genes and PGx markers, genes/PGx markers/drugs and efficacy and/or toxicity (https://app.powerbi.com/view?r=eyJrIjoiODA0NTc2Y2QtM2FjMi00NDU2LWJjZTYtOTk0NDQwODU3ZGY5IiwidCI6ImQyYTYwMTAyLWFlNGUtNGM0MC05N2U2LTMzM2M4YzEzZTgzMiIsImMiOjh9).

Results and Discussion

Interactive Outcomes

The automatically generated interactivity of POWER-BI offers much greater flexibility to the database users. All relevant DruGeVar information was divided into multiple graphs, allowing for better distribution, immediate evaluation of information, and ease of access. Figure 2a depicts an indicative illustration of the resulted graphs prior to a user query, whereas Figure 2b depicts a scenario during which search functionality for a specific object (drug) occurs, following a user query. Notably, immediate response was ensured even though the length of data rows was quite big.

Another important functionality that characterizes this updated version of DruGeVar, following the integration of POWER-BI, refers to the flexible way of data querying, both when drugs (format similar to Excel for improving search performance) alone or in combined queries of genes-PGx markers and drugs-genes were explored. In this present database version, the user can query the data using one or more filters, being able to accomplish complex searches, if needed. Indicatively, if one desires to run a search about one specific gene, the graph "genes-PGx markers" can be used, followed by a click on that

particular gene, resulting in all graphs reacting due to this search. Figure 2c depicts such a query for "*TPMT*" (gene), showing that the results in all available graphs respond immediately to the information that the "gene" information provides.

Interactive Oncology Drug Outcomes

To ease decision-making and implement PGx information in the clinic, we envisaged collaborative systems to orchestrate data [13–15] and facilitate multi-omics data integration to account for interindividual variability [16–18]. Focusing on cancer genomics, we tested the synergy of POWER-BI and DruGeVar on oncology drugs sharing pharmacogenomic recommendations to maximize drug efficacy and minimize toxicity. Given the increasingly recognized importance of determining and administering the biologically effective drug, we could explain how clinical pharmacodynamic biomarkers specific to the agent mechanism of action could be used for the development of pharmacodynamics-guided biologically effective dosage regimens (PD-BEDR) to maximize the efficacy and minimize the toxicity of targeted therapies [19].

First, we implemented the graph of drugs at the application, independently, in a format similar to Excel, for the sake of simplifying the overall searching procedure. Using "mercaptopurine," "cisplatin," "thioguanine," and "fluorouracil" as case drugs, we selected for those directly (by clicking on to the drugs in the first graph of the outcome illustrated as shown below and getting more in-depth information) and mined all relevant data in order of appearance. Figure 2 illustrates such functions for "cisplatin" and "thioguanine." For drug/gene/PGx markers toxicity/efficacy, the following outcomes were obtained (Fig. 2):

- Mercaptopurine – *TPMT* – *4, *3C, *3B, *2 – toxicity mode: active/efficacy mode: not active
- Cisplatin – *TPMT* – *3B, *3C – toxicity mode: active/efficacy mode: not active
- Thioguanine – *TMPT* –*3B, *4, *2, *3C – toxicity mode: active/efficacy mode: not active
- Fluorouracil – *DPYD* – *13 – toxicity mode: active/efficacy mode: not active

Conclusions and Future Perspectives

POWER-BI, upon its integration into DruGeVar, expands the meaning of efficient data querying and presents a new ability of interactivity. Herein, we succeeded to manage the information of DruGeVar into multiple visualization for better performance, while the overall structure of the graphs was able to invoke every single piece of data from DruGeVar, when requested. Additionally, we improved the reaction time and graphics outcome by separating the data objects into single graphs, instead of adding all information into a single graph, and we provided concurrency of all related fields. Furthermore, we added the auto-update data attribute, according to which, when the DruGeVar administrator appends more information, the graphs will respond automatically and will generate the added objects.

Today, the ultimate goal when dealing with big data applications is to improve performance coupled to management reliability. Both such aspects should always be updated to ensure compatibility to contemporary technologies. The next generation of POWER-BI graphs in DruGeVar will be real-time visualizations. The latter constitute a new level of data entrance and distribute the set of information in the most realistic form. The optimization of speed response is another important issue that must be handled further. Due to multiple objects and the pieces of data related with these objects, the delay often is an obvious issue.

Acknowledgements

This work was encouraged and coordinated by the Genomic Medicine Alliance Cancer Genomics and Genome Informatics Working Groups and partly funded by research grants from the European Commission (FP7-305444; RD-Connect and H2020-668353; U-PGx) and the Greek General Secretariat of Research and Technology (ELIXIR-GR) to G.P.P.

Disclosure Statement

The authors declare no conflict of interests.

References

1 Chen CP, Zhang C-Y: Data-intensive applications, challenges, techniques and technologies: a survey on Big Data. Inf Sci 2014;275:314–347.
2 Keim D, Andrienko G, Fekete J-D, Görg C, Kohlhammer J, Melançon G: Visual analytics: definition, process, and challenges; in Kerren A, Stasko J, Fekete J-D, North C (eds): Information Visualization – Human-Centered Issues and Perspectives. Berlin, Springer, 2008, pp 154–175.

3 Chen H, Chiang RH, Storey VC: Business intelligence and analytics: from big data to big impact. MIS Quarterly 2012;36:1165–1188.
4 Sarma AC, Girão J: Identities in the future internet of things. Wireless Pers Commun 2009; 49:353–363.
5 Miorandi D, Sicari S, De Pellegrini F, Chlamtac I: Internet of things: vision, applications and research challenges. Ad Hoc Netw 2012; 10:1497–1516.
6 Manovich L: What is visualisation? Vis Stud 2011;26:36–49.
7 Chen C: Information visualization. Wiley Interdiscip Rev Comput Stat 2010;2:387–403.
8 Friendly M, Denis DJ: Milestones in the History of Thematic Cartography, Statistical Graphics, and Data Visualization. http://www.datavis.ca/milestones, 2001.
9 Friendly M: Visions and re-visions of Charles Joseph Minard. J Educ Behav Stat 2002;27: 31–51.
10 Cairo A: The Functional Art: An Introduction to Information Graphics and Visualization. Berkeley, New Riders, 2012.
11 Dalabira E, Viennas E, Daki E, Komianou A, Bartsakoulia M, Poulas K, Katsila T, Tzimas G, Patrinos GP: DruGeVar: an online resource triangulating drugs with genes and genomic biomarkers for clinical pharmacogenomics. Public Health Genomics 2014;17: 265–271.
12 Viennas E, Komianou A, Mizzi C, Stojiljkovic M, Mitropoulou C, Muilu J, Vihinen M, Grypioti P, Papadaki S, Pavlidis C: Expanded national database collection and data coverage in the FINDbase worldwide database for clinically relevant genomic variation allele frequencies. Nucleic Acids Res 2017;45:D846–D853.
13 Makedon F, Karkaletsis V, Maglogiannis I: Overview: Computational analysis and decision support systems in oncology. Oncol Rep 2006;15:971–974.
14 Karacapilidis N, Koukouras D: A web-based system for supporting collaboration towards resolving oncology issues. Oncol Rep 2006; 15:1101–1107.
15 Tsiliki G, Karacapilidis N, Christodoulou S, Tzagarakis M: Collaborative mining and interpretation of large-scale data for biomedical research insights. PLoS One 2014;9:e108600.
16 Katsila T, Konstantinou E, Lavda I, Malakis H, Papantoni I, Skondra L, Patrinos GP: Pharmacometabolomics-aided pharmacogenomics in autoimmune disease. EBioMedicine 2016;5:40–45.
17 Balasopoulou A, Patrinos GP, Katsila T: Pharmacometabolomics informs viromics toward precision medicine. Front Pharmacol 2016;7: 411.
18 Katsila T, Patrinos GP: Whole genome sequencing in pharmacogenomics. Front Pharmacol 2015;6:61.
19 Manolio TA, Fowler DM, Starita LM, Haendel MA, MacArthur DG, Biesecker LG, Worthey E, Chisholm RL, Green ED, Jacob HJ: Bedside back to bench: building bridges between basic and clinical genomic research. Cell 2017;169: 6–12.