Research Article

For reprint orders, please contact: reprints@futuremedicine.com

Pharmacogenomics

Minimum Information required for a DMET Experiment reporting

Aim: To provide pharmacogenomics reporting guidelines, the information and tools required for reporting to public omic databases. **Material & methods:** For effective DMET data interpretation, sharing, interoperability, reproducibility and reporting, we propose the Minimum Information required for a DMET Experiment (MIDE) reporting. **Results:** MIDE provides reporting guidelines and describes the information required for reporting, data storage and data sharing in the form of XML. **Conclusion:** The MIDE guidelines will benefit the scientific community with pharmacogenomics experiments, including reporting pharmacogenomics data from other technology platforms, with the tools that will ease and automate the generation of such reports using the standardized MIDE XML schema, facilitating the sharing, dissemination, reanalysis of datasets through accessible and transparent pharmacogenomics data reporting.

First draft submitted: 27 January 2016; Accepted for publication: 15 April 2016; Published online: 22 August 2016

Keywords: bioinformatics • DMET • minimum information requirement guidelines • personalized genomics • personalized medicine • pharmacogenomics • standardization

Judit Kumuthini*, Mamana Mbiyavanga, Emile Chimusa, Jyotishman Pathak, Panu Somervuo, Ron HN Van Schaik, Vita Dolzan, Clint Mizzi, Kusha Kalideen, Raj S Ramesar, Milan Macek, George P Patrinos & Alessio Squassina

*Author for correspondence: Centre for Proteomic & Genomic Research, Cape Town, South Africa jkumuthini@gmail.com

For full author affiliations list, please see the back page

Pharmacogenomics investigates the role of genetic variation in individual drug responses [1,2]. Once ingested, pharmaceutical agents undergo both a pharmacokinetic (the action the body takes on the drug, i.e., drug metabolism) and pharmacodynamic processes (the action the drug has on the body) [1,3]. Variants in genes involved in pharmacokinetic (QC: Quality control, toxicity of the drug) and pharmacodynamic (action of the drug) result in differential rates of processing at every step [1,4–7].

A number of pharmacogenes are highly polymorphic and display a nonuniform distribution of functionally relevant alleles between individuals and populations [7–13]. For example, the well-characterized multidrug resistance gene ATP-binding cassette, sub-family B (MDR/TAP), member 1 (*ABCB1),* has shown significantly different allele frequencies for certain SNPs in populations across the world [9–14]. Recently, additional genetic variants have been identified as contributors to variations in dosage of medications between individuals [15].

The US FDA has recognized the value of genotyping known genetic variants and now recommends genetic testing prior to prescription of certain pharmaceuticals [16,17]. Pharmacogenomic research is vital and the knowledge of pharmacogenomic markers associated with drug response broadens on a daily basis. As a result, the FDA welcomes any voluntary genomic data submissions generated from pharmacogenetic research. This research stands to benefit clinical trial procedures as well as drug development and dosage guidelines [16].

Conventional genotyping methods (restriction fragment length polymorphism, allele-specific PCR, real-time PCR-based methods) are time consuming and generate limited data. In contrast, SNP microarrays generate a greater amount of data within a more com-

Future Medicine part of fsg

pact time frame. High-throughput (e.g., 1–2.5 million) SNP chips are available, but do not provide answers to specific phenotypic questions. Consequently, the use of specialized, that is, custom-made microarrays is becoming increasingly widespread [18]. An alternative to specialized arrays is targeted resequencing, utilizing next-generation sequencing platforms. Using whole genome sequencing, pharmacogenomics profiles were created and studied for one individual [19], or for multiple individuals [20]. Although next-generation sequencing can reveal novel, undiscovered variants, array-based genotyping is, however, currently the most feasible and affordable solution for screening hundreds to thousands of samples are array based. Pharmacogenomics studies are expanding exponentially, both in terms of number and amount of data generated.

The microarray known as drug metabolizing enzymes and transporters (DMET plus) comprises 1936 variants across 231 genes, including five copy number loci involved in drug metabolism and transport. To facilitate future research and diagnostic applications, Affymetrix® (CA, USA) developed a microarray focused on probing known genetic variations in absorption, distribution, metabolism, excretion, toxicity in genes [21]. This array is designed to identify genetic variants in patients presenting different drug responses and risks for adverse drug reactions. Furthermore, the DMET array is considered a small to midscale pathway analysis genotyping platform [22] that screens and identifies both common and rare variants (SNPs, copy number variants, insertions, deletions and trialleles) [21]. The data generated from DMET provide valuable information that can be useful in pharmacogenomics interpretation for accurately predicted drug prescriptions and optimal doses for an individual. DMET technology has steadily advanced leading to increased demands to develop new bioinformatics software, analysis tools, algorithms, web applications and specific statistical techniques. With the advent of personalized medicine, it is evident that a large number of pilot studies will be conducted globally in the foreseeable future [23–25]. There are several issues concerning the management and effective use of information (or raw and meta data), generated from these types of studies in biomedical research and personalized medicine [26]. On the other hand, the number of pharmacogenomics studies have increased, and generated data in large volume, diverse in content as well beset with gaps and ambiguities in the description and characterization of diseases, drugs being studied, experimental design, analytical protocol, techniques and methods. Cohesive informatics methods and standards in each of these entities are critical for enabling collaboration between researchers, data sharing, unambiguous representation of such data as such active international community efforts contributes toward standardization in reporting for example, Minimum Information for Biological and Biomedical Investigations (MIBBI), OBO foundry and biosharing. In addition, these entities are enablers in correcting interpretation of biomedical data and semantic (meaningful) search; and help the integration of data; and ultimately for ensuring data quality, reliability and reproducibility, for, for example, discrepancies or lack of information on algorithms implemented or QC threshold used will yield different results for the same study. Here, we assess and address the issues concerning meta data capturing for the data analysis and reporting tools for DMET experiments. Similarly to pharmacogenomics data produced from other technology platforms, our assessment focuses on the relevant information required for reporting to public databases as well as capturing meta data pertaining to DMET experiments as a part of the routine protocol, and provides guidelines that would facilitate effective DMET data interpretation, sharing, reporting and use across the scientific community in the future.

Materials & methods

Common elements of MIDE

The common elements proposed for MIDE were derived from consolidating other Minimum Information Guidelines and Standards published in MIBBI as to which elements and ontology where appropriate to use. However, these were not available for most of the elements or some elements have had multiple fields or inconsistency in the ontology. This was addressed by consulting with researchers from Pharmacogenomic for Every Nation's Initiative, Technology platform, Affymetrix and the DMET service labs to consolidate what is essential and optional information to capture and how to phrase it such that validation, interoperability and reproducibility of DMET experiments be increased and seamless.

DMET analysis

DMET data analysis workflow (Figure 1) is similar to any other array-based genotyping experiment. After DNA sample processing and hybridization, the DMET array is scanned and probes intensities transformed into numeric values. Results of the scanning are stored in an image/intensity (*.CEL), sample attribute (*.ARR) and genotyping (*.CHP) files, which provide the starting point for computer-based data analysis at primary, secondary and tertiary levels. In the next sections and in Table 1 we will discuss and provide details for the different levels of analyzes and

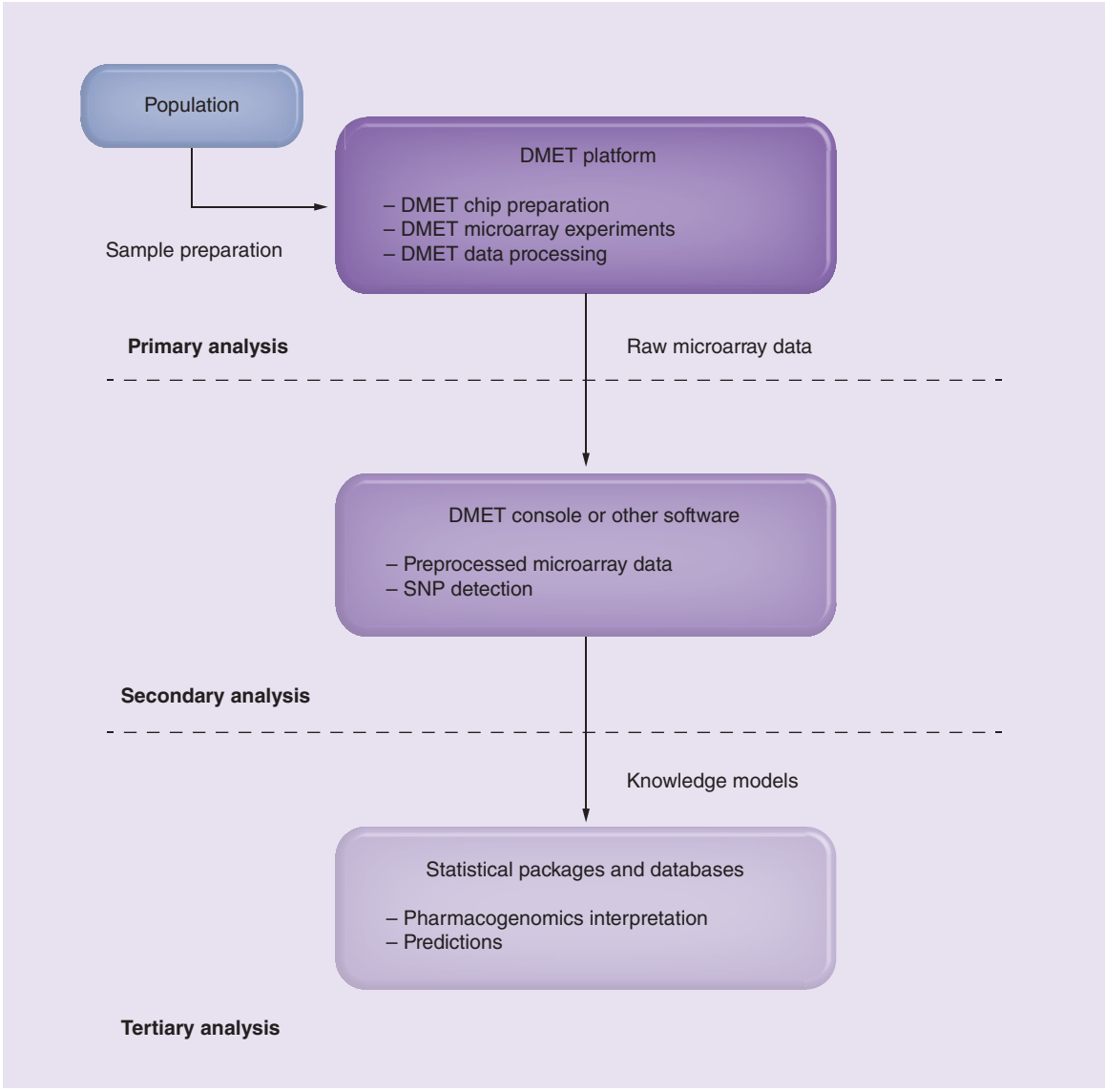

Figure 1. DMET analysis workflow. Data analysis of a DMET experiment is divided into primary (preparation of raw microarray data), secondary (through DMET Console) and tertiary (interpretation of the results) steps.

demonstrate the value MIDE adds to typical DMET experiment by capturing relevant details that is, meta data. The details of these analyzes are described below.

Primary analysis: normalizing raw data

Affymetrix provides the primary work package for raw DMET data analysis at no additional cost. DMET Console software normalizes raw data and provides tools that convert allelic variants into the standard star nomenclature format [28]. Alternatively, data processing can be done in command line using Affymetrix Power Tools or other propriety software for example, DMET Miner [29]. The DMET Console graphical user interface does, however, have the added benefit of providing tools for QC that enable visual investigation of individual markers via cluster plots. The current limitation of DMET Console is that it does not provide tools for statistical analysis of an experiment. This therefore increases the likelihood of researchers directing their efforts toward developing the necessary tools for advanced processing of data in the workflow. One such example is DMET-Analyzer [27,29], although the statistical analysis is currently limited to Fisher's Exact test in a case/control study and a calculating test for Hardy–Weinberg equilibrium. The DMET Console also supports some tertiary analysis, including phenotype prediction and genotype conversion into haplotype level star allele results.

Files required for DMET Console analysis

Sample-related data are stored as intensity (*.CEL) or genotyping (*.CHP), after genotypes calling and as sample information (*.ARR) files. Prior to data ana-

Table 1. Minimum Information required for a DMET Experiment checklist for authors, reviewers, data managers, data curators, analysis software developers and journal editors.

Study specific information	Importance	Description
Aim of the study	O	Aim(s) and ID of the study
Link to publication(s)	R	(PMID or DOI) or electronic record
Leading institution or source	O	Name and address of the research institute where the study PI comes from and if necessary the actual experimental lab is based. Main PI's name/researcher ID (or other unique identifier)
Date(s)	O	Date study begun and completed
Sample collection, storage, shipping	O	How sample was collected, stored and shipped
Sample processing, labeling, hybridization	O	Sample processing details
Scanning	R	Instrument, software, parameters
Probe signal normalization	R	Software, parameters including reference data versions
Protocol(s)	O	Name/ID
Marker level summarization, software, parameters	R	Technical analysis workflow used: name of software including versions and parameters used
Genotyping method, software, parameters	R	Technical analysis workflow used: name of software including versions and parameters used
Tertiary data analysis	O	Technical analysis workflow used: name of software including versions, parameters used and database versions and URL of the date of access
Sample name	R	Specimen annotation
Biomaterial	O	Species or cell line name
Covariates	R	Clinical observations: ethnic group, gender, age, weight, height, region, survival (yes/no), other; disease state, treatment or normal of each sample
Quality control steps taken	R	Type of QC checks carried out for the study
Replicates	R	Identify, if any, replicates are technical/biological by differentiating sample IDs
Cell line used	O	Tissue part, source provider, distributer, company, catalogue no., conditions of storage, contact details of laboratory
Cell culture conditions	O	Conditions and characteristics of cell culture
Genotype call rate	R	For each sample
Experimental aim	R	Description of experimental aim associated with a study (ID), for example, identify association between SNPs and genotype variations to a disease(s), use of genetics markers to predict response to medicine
Summary of results	R	What was the outcome of the experiment/study?
Experimental design	R	For example, compound treatment design, dose response design, stimulus or stress design, injury design and other
Data files	O	Table showing sample/raw data file/processed data file associations containing following information. Naming convention or nomenclature (e.g., rs, *, HGVS) chosen for the study
File name	O	Full names of files (raw data and analyzed data) and locations
File format	R	Type of formats it is available in, for example, ARR, CEL, CHP, among others.
Explanation of missing data	O	Why file(s) or sample(s) were missing. Missing sample data which needs to be added before the software will analyze the data

D: Desirable Information, should be submitted if available; E: Essential information to be submitted with the paper; O: Optional Information, should be submitted if available; PI: Principal investigator; QC: Quality control; R: Required information to be submitted with the paper.

lysis, reference files need to be downloaded from the Affymetrix website [30] and installed onto DMET Console. These files are DMETPlus Genotyping files for primary analysis and the DMETPlus Marker Annotation and Allele Translation files for translating genotypes into star alleles. Additionally, when using DMET Console version 1.3 (or higher), a Metabolizer bin file is required if the workflow includes phenotype prediction. Any user modification of the Metabolizer Bin file should be included in the information submitted to a data repository to ensure result reproducibility.

DMET Console versions & compatibility

DMET Console has been updated four-times since its initial release, with the current version (1.3) available for download from the Affymetrix website, similarly to DMET miner [29]. The releases introduced additional genotyping methods, with the current release supporting all the previous genotyping methods. Version 1.3 introduced the ability to do phenotype prediction and reporting. There are also different versions of the algorithms used in genotype calling depending on which version of software is used. In an effort to ensure reproducibility, the genotyping method used should be provided when submitting data.

Secondary analysis: genotype calling using the DMET Console

Secondary analysis consists of making genotype calls for each marker on the array using DMET Console, Affymetrix Power Tool or similar software. Genotyping with DMET Console can be performed using one of two methods. The default is fixed boundary analysis, while the alternate is dynamic boundary analysis.

Fixed boundary analysis compares signals generated by the sample to predefined clustering models in order to make the genotype call. In contrast, dynamic boundary analysis adapts the cluster models according to the data being analyzed in the current data set [21]. Only fixed boundary analysis gives directly comparable results between samples analyzed at different research sites. At present, both algorithms have two different versions available.

Tertiary analysis: submission of data to pharmacogenomic-specific databases

Tertiary analysis requires the use of additional work packages that support DMET data incorporation with databases such as dbSNP [25], the pharmacogenomics knowledge base (PharmGKB, [31]) [32,33] or other pharmacogenomics interpretation tools [33,34]. PharmGKB is currently the most prominent one out of the very few databases that are available to date specifically for pharmacogenomic data and allows users to submit data in various formats (e.g., genotype data, phenotype data, information about pathways). In addition, the database allows the submission of SNP array data, specifically from Affymetrix and Illumina platforms. An important requirement of PharmGKB is that only raw data can be submitted [34,35].

Results

Proposed MIDE & its rationale

The quintessential information about DMET data concerns the experiment itself. This should include but not limited to keeping with good laboratory practice, experimental design, extract/sample preparation, labeling, hybridization procedures, measurement specifications, QC information and data extraction information, which, if reported in concise way in one place, will increase their accessibility and interpretability. Since the Affymetrix DMET kit comes with a rigorous protocol, any deviation from the protocol must be captured and noted in any publication and when submitting the data to specific databases. Standardizing how this information is captured and published in a comparable and consistent manner is crucial for researchers to understand the experiment and subsequently interpret the data generated.

The adoption of common data reporting standards for DMET generated data will, in particular, allow the aggregation of data generated from different platforms, as well as the integration of secondary data (for example, Q-PCR data, clinical and epidemiological data, drug interaction data, among others) in tertiary analysis platforms. With a need for such a standardized manner of reporting DMET-generated data, we propose a guideline to the MIDE.

MIDE is based on a checklist (Table 1) for DMET data extracted form the guidelines for microarrays as provided in [36,37]. We also endeavored to adhere to two criteria used by several other guidelines such as the minimum information about a peptide array experiment authors [26], namely sufficiency and practicability. Sufficiency states that the minimum information requirements are constructed in such a way that the reader is able to critically evaluate experiment findings, as well as able interpret and reproduce the experiments. Practicability refers to the incorporation of precise and limited information, which are of significant importance to the experiment results. MIDE relies on other important prerequisites to support the scientific community in publishing high-quality experimental data with respect to reproducibility, compatibility and reusability and interoperability.

Reproducibility

The guidelines proposed in MIDE shall enable DMET experiment findings to be more easily reproduced and verified as the information required for reproducing are captured in a systematic fashion for example, which algorithm was used, without it different genotype results will be obtained. It will also ease the interpretation of findings by peers, as a clear idea of experimental procedures will more effectively orient a journal editor or a reviewer and will also allow meta-analysis of data produced by different laboratories.

Comparability & reusability

Findings between different studies can be compared effectively if data reporting standards are in place. Furthermore, if data are extracted in a concise and

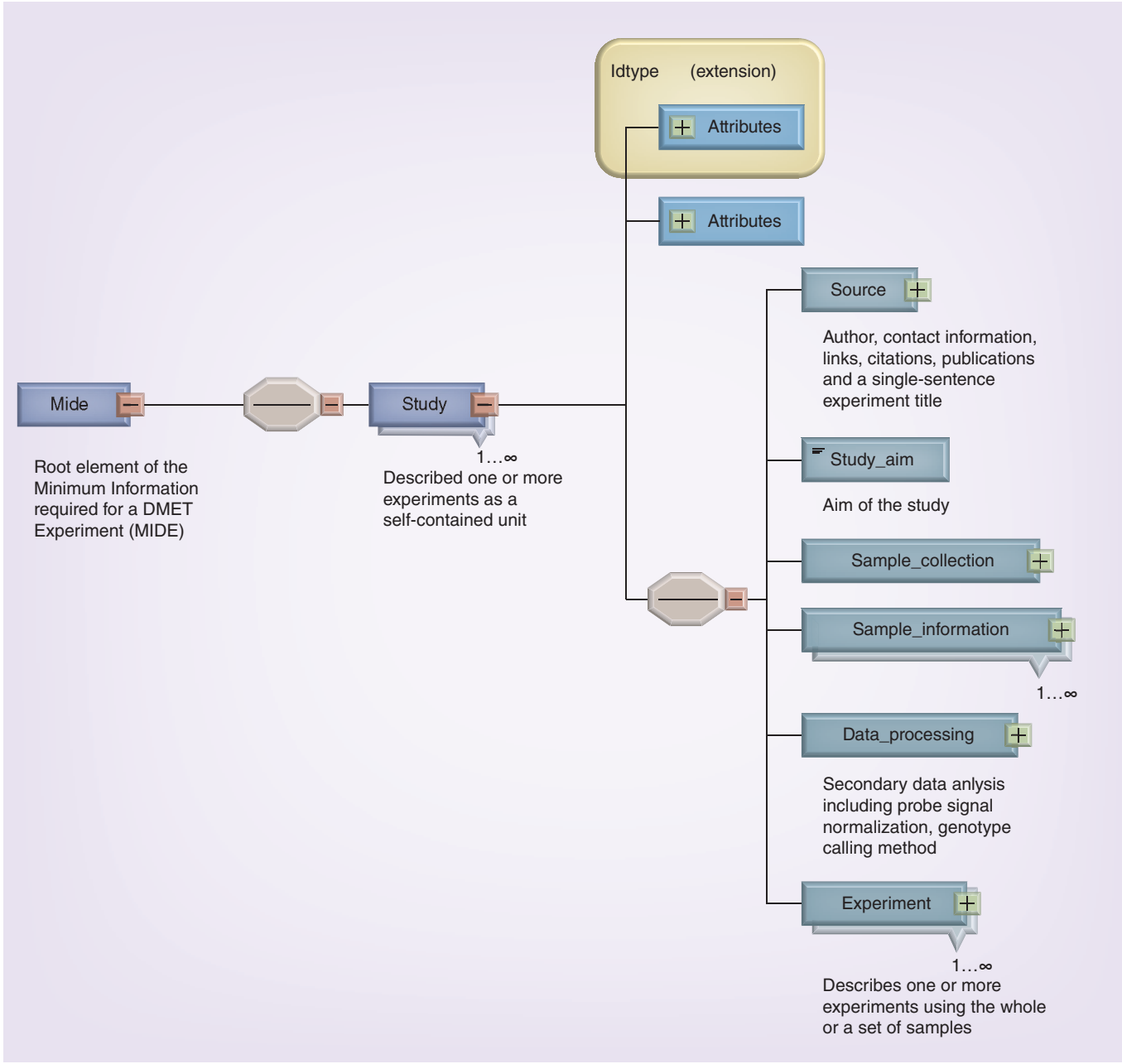

Figure 2. Graphical representation of minimum information required for a DMET experiment XML 1.0 format. This figure shows the minimum required for DMET experiment described by elements which are grouped by source, study aim, sample_collection_and_process, sample_information, data_processing and experiment. Some elements have been collapsed (+) for ease of visualization. The full MIDE XML 1.0 schema is accessible from the MIBBI project page [38].
MIDE: Minimum information required for DMET experiment.

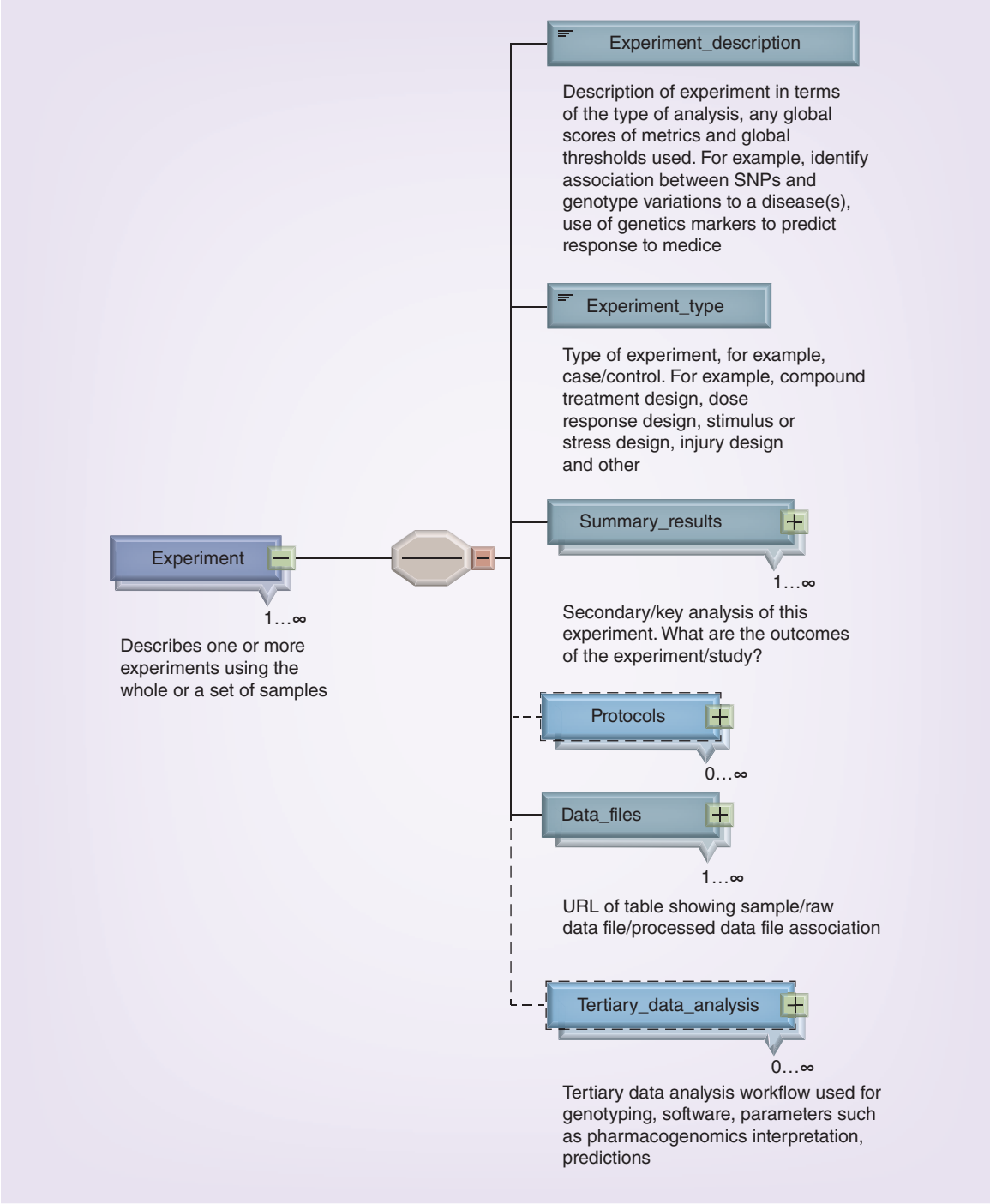

Figure 3. A graphical representation of the experiment element of minimum information required for a DMET experiment XML 1.0. The experiment is formed by a number of attributes, which hold information on the description of the experiment, the type of experiment, protocols used, data files and tertiary data analysis.

correct manner, they can be used in subsequent experiments. We feel that quality of data supersedes quantity and using a concise method of data extraction from DMET arrays can greatly increase the experimenter's ability to sort biologically meaningful information from background noise and experimental error.

Interoperability

Ultimately, the field of pharmacogenomics will evolve to integrate data from various omics and technology platforms. Implementing standards for submission of data for each of the various platforms (microarrays, SNP arrays, proteomics and DMET), will aid the development of pipelines able to consolidate the dif-

Table 2. An example of DMET study completed using the Minimum Information required for a DMET Experiment checklist. The checklist illustrates the usefulness of Minimum Information required for a DMET Experiment and the information pertaining to study, experiment and subject levels.

Information	
Study-specific Information	
Aim of the study†	To profile different SNPs between the Xhosa, Caucasian and the mix-ancestry population groups
Link to publication(s)‡	All the files containing the results reported from the analysis are located on the "/home/minah/DMET_Data/Mixrace and Xhosa analysis" directory. This contains: – 2013-01-24_145318_translations – a directory for the translation reports – Differences of allele frequencies – Genotype correlation Bonferroni – Genotype results Xhosa, Caucasian and Mixed Ancestry_GT.txt
Leading institution or source†	(J Kumuthini), (Centre for Proteomic and Genomic Research [CPGR], Upper Level, St Peter's Mall, Cnr Anzio and Main Road, Observatory, 7925).
Date(s)†	(Started: 10/10/2012) (Completed: 14/08/2013)
Sample collection, storage, shipping†	DNA samples were submitted by collaborator and with any information about collection, storage and shipping
Sample processing, labeling, hybridization†	Samples were processed in accordance with the Affymetrix DMET Plus protocol. The samples were taken through the following procedures: mPCR, annealing of mPCR products to genomic DNA, gap fill through amplification followed by a PCR cleanup step. The cleaned up PCR products were fragmented and labelled and hybridized to DMET arrays for 18 h. Following hybridization, the arrays were washed and stained using the GeneChip Fluidics Station 450 and the arrays scanned using the GeneChip® Scanner 3000 7G
Scanning‡	Instrument: GeneChip® Scanner 3000 7G Software: AGCC
Probe signal normalization‡	DMET Console V1.0
Protocol(s)†	Affymetrix DMET plus array protocol
Marker level summarization, software, parameters‡	(DMET Console V1.0), (fixed genotype boundary 2), (maximum confidence score threshold 0.001 and minimum prior observations)
Genotyping method, software, parameters‡	(DMET Console V1.0), (use of the standardized star allele nomenclature under DMET-Analyzer tool)
Tertiary data analysis†	DMET Console_1_3_64 bit was installed and used to carry out genotype analysis and allele translation on the Xhosa and mixed-ancestry DMET data. The output (probe set IDs and alleles belonging to different samples in a .txt file format) was submitted to the DMET Analyzer1.0.1 to identify the significantly discriminative SNPs (Guzzi *et al.*) [27] Annotation Analysis using dbSNP and PharmGKB
Sample/subject-specific information	
Sample name(s)‡	See attached file for the mixed sample data (see Table 3 in Supplementary Material 1)
Biomaterial†	Homo sapiens
Covariates‡	Ethnicity: mixed, Caucasian, Xhosa groups were observed and given in BatchEdit_Mixed_24_MA.xls

†Desirable or optional Information, should be submitted if available.
‡Essential or Required information to be submitted with the article.
QC: Quality control; mPCR: Multiplex PCR; NA: Not applicable.

Table 2. An example of DMET study completed using the Minimum Information required for a DMET Experiment checklist. The checklist illustrates the usefulness of Minimum Information required for a DMET Experiment and the information pertaining to study, experiment and subject levels (cont.).

Information	
Sample/subject-specific information (cont.)	
Quality control steps taken‡	RNA concentration and quality – nanodrop spectrophotometer
	DNA integrity – gel electrophoresis
	QC to confirm amplification – Gel electrophoresis
	QC to confirm successful fragmentation – Gel electrophoresis
	QC call rate for genotyping (In bound) = 99%
Replicates‡	No technical replicates recorded. All are biological replicates
Cell line used†	NA
Cell culture conditions†	NA
Genotype call rate‡	98% average rate across all samples
Experiment-specific information	
Experimental aim‡	Establish simple analysis pipeline to achieve the main objective of the study
Summary of results‡	Preliminary SNP profiles were generated between the Xhosa and the mix-ancestry population group for all the samples with 98% call rate. There is a huge discrepancy between the observed SNPs in the two population groups. A total of 13 SNPs with a p-value ≤ 0.05 indicated a significant differential distribution between the two population groups
Experimental design‡	Evaluation of presence and absent of ADME variants in different ethnic groups
Data files†	See attached file for the mixed sample data (BatchEdit_Mixed_24_MA.xls)
File name†	See attached file for the mixed sample data (See Table 3 in Supplementary Material)
File format‡	ARR, CEL, CHP
Explanation of missing data†	No missing data observed

†Desirable or optional Information, should be submitted if available.
‡Essential or Required information to be submitted with the article.
QC: Quality control; mPCR: Multiplex PCR; NA: Not applicable.

ferent platforms and the standards by which each abides.

Implementation & development of MIDE

In the interest of coherent and coordinated development of such a guideline, the project is registered on the MIBBI portal [38]. The MIBBI project is a collaboration between leaders in the biological and biomedical fields, which acts as a meeting point for the coordination of minimum information guidelines for technological and field applications in life sciences [35]. It is also integrated with the standards setting and sharing of biological data reporting called Biosharing [39]. The guidelines for reporting and the XML format described in this document, and in future API and related documents, will be made available on [40]. Freely accessible, MIBBI provides, through a web-based platform, checklist for projects such as MIDE and other resources such as controlled vocabularies (CVs), tools, databases and complementary data formats, while ensuring that efforts on new checklists are not redundant or repeated. It is important to iterate that this reporting guideline, at its current stage, represents a staring point.

For ease of usage and to enable automated generation of MIDE reports using tools that would be developed by the DMET scientific community, the MIDE guidelines provide an XML schema MIDE version 1.0 in line with the requirements of MIBBI [41]. The MIDE data model reporting is based on a simple XML schema that represents how the data should be formatted, which basically provides the ability to define cardinality of XML elements and define type to the XML element attributes. Due to nonavailability of well defined CVs for DMET experiments, which could ensure the use of the same terminology to described an element by the data producer, this version of the MIDE schema does not integrate any CV or an ontology. Figure 2 describes version 1.0 of the MIDE schema pointing out key entries required for a DMET experiment. Extended description of elements of the MIDE

schema is provided in the documentation of the MIDE guidelines on MIBBI [42].

The root element of the MIDE XML 1.0 is the MIDE. It can contain one or multiple study elements. MIDE XML 1.0 revolves around a study, which is the core element describing one or more DMET experiments with all related meta data including sample and data processing. Thus, multiple study elements can easily be merged into a single MIDE element. The study element consists of several elements that contain the minimum information required for reporting an experiment on Affymetrix DMET described in this manuscript. It includes a source element that describes the origin of the data in the study, along with the study description and aim. The sample_collection_and_processing element describes how the study sample was collected, stored and processed. The element sample_information contains essential information for each sample, including the ID, sample name, sample type, cell line, genotype call rate, covariates, essential QC matrices. The next element, experiment contains all DMET experiments for the study (Figure 3). To illustrate the utility of the MIDE guidelines, an example using the South African decent DMET data generated at the Centre for Proteomic and Genomic Research [J Kumuthini, Unpublished Data], which study and data are presented in detail in Supplementary Material 1, is provided to illustrate the use of MIDE guidelines and highlights the need for information not usually included in the 'Methods' section of published experiments (Table 2).

Discussion

Though the goal of the MIDE guidelines is far from dictating how designs and analyzes of DMET experiments should be implemented, it will aid the extraction of useful information about a patient's pharmacogenomic profile and support effective clinical management, as well as facilitate collaboration between institutions. This may also serve as a critical "checklist for authors, reviewers, data managers, data curators, analysis software developers and journal editors. The MIDE format should have all required fields and sections in one document type and should allow updates as need arises from the community. Similar to other experiments requiring a minimum degree of information, the MIDE guidelines provide that all essential information should be readily available electronically with all published data and articles as a minimum. This version-controlled document with support material should be made available in a centralized repository, MIBBI. In order to store and make all data files and sample information related to DMET experiments available in a coherent way, we have provided an appropriate XML template for DMET data. This can be found via the MIBBI portal [40]. We have illustrated the discussed work-flow of DMET (see 'Methods' section) and the proposed MIDE guidelines using an unpublished data of three populations.

Several benefits will arise from the widespread recognition and acceptance of MIDE. To reap the full benefits of applying MIDE, as seen for other minimum information framework projects requiring minimum information, it requires implementation at a higher level. This can be achieved by implementing MIDE compliancy as a requirement for publication of research articles (at journal level), data submission to propriety and public database repositories (at project and framework level), funding and grant proposals (from funders level) and possibly encouragement from open source projects and data repositories.

While we hope that MIDE adds significant value to data generators, data consumers and end users, we are mindful of emerging technologies, some of which may supersede the DMET system. Nonetheless, an adherence to the recommendations provided in this article will ensure a move toward interoperability and sustainability of information generating systems. Also, laboratories performing pharmacogenetic testing should be accredited in order to assure quality of services – either in Europe according to ISO 15189:2013 [43] – [44] and in USA for example, CPA / CLIA, among others – this will add to the 'quality' of these recommendations.

Present MIDE challenges & limitations in practice

Though many researchers may share the current need of standardized guidelines for reporting on DMET experiments, the emergence of a complete consensus for such guidelines by the research community is highly improbable. This may be due to arguments including the practicability of their development and implementation, as well as issues around the generation of such reports for all the technology platforms, which might be time consuming and tedious. However, because of the important nature of DMET experiments that need to be reproducible, a set of essential information that can enable exact repetition of experiments under the same conditions not only by the original experimentalists but also by other researchers, standards such as the proposed MIDE are of high relevance. The current version of MIDE XML schema does not include CV for all the elements because of the unavailability of such resources, which may, for now, result into some inconsistencies in the reporting of DMET experiments through MIDE guidelines. The use of such CV would aid in the sorting, as well as the storage of MIDE reports, thus enabling automated methods of extracting and analyzing data from MIDE reports. With the growing need

of standards across the scientific community, especially for DMET experiments, the availability of such CV or ontology will enable the extention of MIDE XML schema and take advantage of such resources. Numerous standard guidelines have been developed in recent year and widely accepted by the scientific community, including the minimum information about microarray experiment and many other standards listed on the MIBBI portal. The acceptance of MIDE guidelines by the scientific community at their current stage may need time, until time-effort-resources saving tools to automate the generation of MIDE reports are developed. These tools will enable processing of experimental data and extraction of meta-data for automated generation of MIDE reports as well as enough documentation and user support. This will ultimately be shaped be the community of DMET researchers.

Conclusion

No data-reporting standard exists for DMET or drug toxicology array data reporting. We introduced and have developed a DMET raw and meta-data deporting and guidelines through data sharing, MIBBI foundry requirements and principals. The MIDE guidelines, which will benefit the scientific community from ongoing and future DMET experiments, with tools that will ease and automate the generation of such reports using the Standardized MIDE XML schema, will facilitate the sharing, dissemination, reanalysis of datasets through accessible and transparent DMET experiment reporting. A MIDE API is also under development, which will facilitate the validation and generation of MIDE reports by providing the user with easy-to-use and powerful interface to store, retrieve and export MIDE documents into different formats.

Supplementary data

To view the supplementary data that accompany this paper, please visit the journal website at: www.futuremedicine.com/doi/full/10.2217/pgs-2016-0015

Acknowledgement

We thank members of the Centre for Proteomic and Genomic Research, Cape Town, South Africa and the interns, who contributed toward drafting this article and everyone who provided expert opinion and helped review the manuscript.

Financial & competing interests disclosure

The CPGR is supported through institutional funding from TIA (Technology Innovation Agency) and the DST (Department of Science and Technology). This work is partially funded by NIH Common Fund Award/NHGRI Grant Number U41HG006941 through H3AbioNet project. M Macek was supported by NF-CZ11-PDP-3-003-2014, 00064203 and COST LD14073. The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

No writing assistance was utilized in the production of this manuscript.

Executive summary

Aim

- It is increasingly popular within the life sciences subdomains to establish data reporting guidelines.
- Many articles been published for minimum information required to report data including microarray, proteomics data and RT-PCR among others.
- Biosharing and Minimum Information for Biological and Biomedical Investigations are 'gold' standard in monitoring existing standards and guidelines available in life sciences.

Results

- We have introduces minimum information required for a DMET experiment, Minimum Information required for DMET Experiment guideline.
- The reporting template also available in XML format to increase the user friendliness of data capturing and the data governance.

Conclusion

- Reporting scientific raw data with required metadata to be able to be shared, stored and managed to increase the efficiency of the data analysis process, interoperability.
- These efforts are also proven to increase the value of the data to the end user and ease of use.

References

Papers of special note have been highlighted as: • of interest

1 Thomas FJ, Mcleod HL, Watters JW. Pharmacogenomics: the influence of genomic variation on drug response. *Curr. Top. Med. Chem.* 4(13), 1399–1409 (2004).

• **Highlights essential integration of both pharmacogenomic and clinical data can provide powerful predictive tools for stratifying responders from nonresponders and identifying patients at increased risk for toxicity.**

2 Watson RG, Mcleod HL. Pharmacogenomic contribution to drug response. *Cancer J.* 17(2), 80–88 (2011).

3 Pirmohamed M. Drug metabolism. *Medicine* 36(7), 355–359 (2008).

4 Wang L, Mcleod HL, Weinshilboum RM. Genomics and drug response. *N. Engl. J. Med.* 364(12), 1144–1153 (2011).

5 Weinshilboum RM, Wang L. Pharmacogenetics and pharmacogenomics: development, science, and translation. *Annu. Rev. Genomics Hum. Genet.* 7, 223–245 (2006).

6 Pinto N, Dolan ME. Clinically relevant genetic variations in drug metabolizing enzymes. *Curr. Drug Metab.* 12(5), 487–497 (2011).

7 Daly AK. Pharmacogenetics and human genetic polymorphisms. *Biochem. J.* 429(3), 435–449 (2010).

8 Johnson JA. Ethnic differences in cardiovascular drug response: potential contribution of pharmacogenetics. *Circulation* 118(13), 1383–1393 (2008).

9 Engen RM, Marsh S, Van Booven DJ, Mcleod HL. Ethnic differences in pharmacogenetically relevant genes. *Curr. Drug Targets* 7(12), 1641–1648 (2006).

10 Nebert DW, Menon AG. Pharmacogenomics, ethnicity, and susceptibility genes. *Pharmacogenomics J.* 1(1), 19–22 (2001).

11 Scott SA, Khasawneh R, Peter I, Kornreich R, Desnick RJ. Combined *CYP2C9*, *VKORC1* and *CYP4F2* frequencies among racial and ethnic groups. *Pharmacogenomics* 11(6), 781–791 (2010).

12 Urban TJ. Race, ethnicity, ancestry, and pharmacogenetics. *Mt Sinai. J. Med.* 77(2), 133–139 (2010).

13 Yen-Revollo JL, Van Booven DJ, Peters EJ *et al.* Influence of ethnicity on pharmacogenetic variation in the Ghanaian population. *Pharmacogenomics J.* 9(6), 373–379 (2009).

14 Wolf SJ, Bachtiar M, Wang J, Sim TS, Chong SS, Lee CG. An update on *ABCB1* pharmacogenetics: insights from a 3D model into the location and evolutionary conservation of residues corresponding to SNPs associated with drug pharmacokinetics. *Pharmacogenomics J.* 11(5), 315–325 (2011).

15 Santos PC, Soares RA, Santos DB *et al. CYP2C19* and *ABCB1* gene polymorphisms are differently distributed according to ethnicity in the Brazilian general population. *BMC Med. Genet.* 12, 13 (2011).

16 Laurence J. Getting personal: the promises and pitfalls of personalized medicine. *Transl. Res.* 154(6), 269–271 (2009).

17 Hamburg MA, Collins FS. The path to personalized medicine. *N. Engl. J. Med.* 363(4), 301–304 (2010).

18 Peters EJ, Mcleod HL. Ability of whole-genome SNP arrays to capture 'must have' pharmacogenomic variants. *Pharmacogenomics* 9(11), 1573–1577 (2008).

19 Salleh MZ, Teh LK, Lee LS *et al.* Systematic pharmacogenomics analysis of a Malay whole genome: proof of concept for personalized medicine. *PLoS ONE* 8(8), e71554 (2013).

20 Mizzi C, Peters B, Mitropoulou C *et al.* Personalized pharmacogenomics profiling using whole-genome sequencing. *Pharmacogenomics* 15(9), 1223–1234 (2014).

21 Affymetrix. White Paper DMET™ plus genotyping and copy number methods. (2011).

22 Sissung TM, English BC, Venzon D, Figg WD, Deeken JF. Clinical pharmacology and pharmacogenetics in a genomics era: the DMET platform. *Pharmacogenomics* 11(1), 89–103 (2010).

- **Great review with DMET design, DMET platform applications in clinical pharmacology and pharmacogenetics studies.**

23 Oetjens MT, Denny JC, Ritchie MD *et al.* Assessment of a pharmacogenomic marker panel in a polypharmacy population identified from electronic medical records. *Pharmacogenomics* 14(7), 735–744 (2013).

24 Yang L, Price ET, Chang CW *et al.* Gene expression variability in human hepatic drug metabolizing enzymes and transporters. *PLoS ONE* 8(4), e60368 (2013).

25 Day IN. dbSNP in the detail and copy number complexities. *Hum. Mutat.* 31(1), 2–4 (2010).

26 York WS, Agravat S, Aoki-Kinoshita KF *et al.* MIRAGE: the minimum information required for a glycomics experiment. *Glycobiology* 24(5), 402–406 (2014).

27 Guzzi PH, Agapito G, Di Martino MT *et al.* DMET-analyzer: automatic analysis of Affymetrix DMET data. *BMC Bioinformatics* 13, 258 (2012).

28 Affymetrix. White Paper DMET Plus allele translation reports: summary of comprehensive drug disposition genotyping into commonly recognized allele names. (2012).

29 Agapito G, Guzzi PH, Cannataro M. DMET-Miner: efficient discovery of association rules from pharmacogenomic data. *J. Biomed. Inform.* 56, 273–283 (2015).

30 Affymetrix. www.affymetrix.com

31 PharmGKB. www.pharmgkb.org

32 Thorn CF, Klein TE, Altman RB. PharmGKB: the Pharmacogenomics Knowledge Base. *Methods Mol. Biol.* 1015, 311–320 (2013).

- **Drug metabolisms and related data embedded in an interactive tool integrated with pharmacogenomics knowledge base, PharmGKB, for researchers investigating relation between genetic variation and drug response.**

33 Potamias G, Lakiotaki K, Katsila T *et al.* Deciphering next-generation pharmacogenomics: an information technology perspective. *Open Biol.* 4(7), 140071, (2014).

- **Pharmacogenomics information systems and its challenges for future of personalized medicine treatments. This includes drug information linked genotype-to-phenotype pharmacogenomics data as well as to support biomedical researchers in the identification of pharmacogenomics-related gene variants.**

34 Klein TE, Chang JT, Cho MK *et al.* Integrating genotype and phenotype information: an overview of the PharmGKB project. *Pharmacogenomics J.* 1(3), 167–170 (2001).

35 Kettner C, Field D, Sansone SA *et al.* Meeting report from the second "minimum information for biological and biomedical investigations" (MIBBI) workshop. *Stand. Genomic Sci.* 3(3), 259–266 (2010).

36 Botha G, Kumuthini J. Minimum Information About a Peptide Array Experiment (MIAPepAE). *EMBnet. J.* 18(1), 14 (2012).

- **A data standard/reporting guidlines for peptide array experiment to increase the value of peptide array data reported.**

37 Quinn TA, Granite S, Allessie MA *et al.* Minimum Information about a Cardiac Electrophysiology Experiment (MICEE): standardised reporting for model reproducibility, interoperability, and data sharing. *Prog. Biophys. Mol. Biol.* 107(1), 4–10 (2011).

- **Another example of standardized reporting for model to enhance reproducibility, interoperability and data sharing in life science research toward good practice.**

38 Mibbi.
http://mibbi.sourceforge.net/projects/MIDE.shtml

39 BioSharing.
https://biosharing.org/

40 BioSharing.
https://biosharing.org/bsg-000628

41 BioWeb.
http://bioweb.cpgr.org.za/mide/mide.xsd

42 BioSharing.
https://biosharing.org/bsg-000281

43 EuroGentest.
www.eurogentest.org

44 EuroGentest.
www.eurogentest.org/index.php?id=139

Affiliations

Judit Kumuthini
[1]Centre for Proteomic & Genomic Research, Cape Town, South Africa

Mamana Mbiyavanga
[1]Centre for Proteomic & Genomic Research, Cape Town, South Africa

Emile Chimusa
[1]Centre for Proteomic & Genomic Research, Cape Town, South Africa
and
Computational Biology Group, Institute for Infectious Diseases & Molecular Medicine, University of Cape Town, South Africa

Jyotishman Pathak
Division of Biomedical Statistics & Informatics, Department of Health Sciences Research, Mayo Clinic, 200 First Street SW, Rochester, MN 55905, USA

Panu Somervuo
Institute of Biotechnology, University of Helsinki, Helsinki, Finland

Ron HN Van Schaik
Department of Clinical Chemistry, Erasmus University Medical Center Rotterdam, Room Na-415, Wytemaweg 80, 3015CN Rotterdam, The Netherlands

Vita Dolzan
Pharmacogenetics Laboratory, Institute of Biochemistry, Faculty of Medicine, University of Ljubljana, Vrazov trg 2, SI-1000 Ljubljana, Slovenia

Clint Mizzi
Department of Bioinformatics, Faculty of Medicine & Health Sciences, Erasmus University Medical Center, Rotterdam, The Netherlands
and
Department of Physiology & Biochemistry, Faculty of Medicine and Surgery, University of Malta, Malta

Kusha Kalideen
UCT/SA MRC Human Genetics Research Unit, Division of Human Genetics, Institute for Infectious Diseases & Molecular Medicine, Division of Human Genetics, University of Cape Town, South Africa

Raj S Ramesar
UCT/SA MRC Human Genetics Research Unit, Division of Human Genetics, Institute for Infectious Diseases & Molecular Medicine, Division of Human Genetics, University of Cape Town, South Africa

Milan Macek
Department of Biology & Medical Genetics, Charles University Prague & 2nd Faculty of Medicine, Prague, Czechia

George P Patrinos
Department of Bioinformatics, Faculty of Medicine & Health Sciences, Erasmus University Medical Center, Rotterdam, The Netherlands
and
Department of Pharmacy, School of Health Sciences, University of Patras, Patras, Greece

Alessio Squassina
Laboratory of Pharmacogenomics, Section of Neuroscience & Clinical Pharmacology, Department of Biomedical Sciences, University of Cagliari, sp 8 Sestu-Monserrato, Km 0.700, 09042 Cagliari, Italy