

Coverage rate of ADME genes from commercial sequencing arrays

Nabil Zaid, PhD^{a,b}, Youness Limami, PhD^c, Nezha Senhaji, PhD^d, Nadia Errafiy, PhD^e, Loubna Khalki, PhD^e, Youssef Bakri, PhD^f, Younes Zaid, PhD^{c,*}, Saaid Amzazi, PhD^{a,b}

Abstract

Pharmacogenomics offers remarkable potential for the rapid translation of discoveries into changes in clinical practice. In the present work, we are interested in evaluating the ability of commercially available genome-wide association sequencing chips to cover genes that have high pharmacogenomics potential.

We used a set of 2794 variations within 369 absorption, distribution, metabolism, and elimination (ADME) genes of interest, as previously defined in collaboration with the Pharma ADME consortium. We have compared the Illumina TrueSeq and both Agilent SureSelect and HaloPlex sequencing technologies. We have developed Python scripts to evaluate the coverage for each of these products. In particular, we considered a specific list of 155 allelic variants in 34 genes which present high pharmacogenomics potential. Both the theoretical and practical coverage was assessed.

Given the need to have a good coverage to establish confidently the functionality of an enzyme, the observed rates are unlikely to provide sufficient evidence for pharmacogenomics studies. We assessed the coverage using enrichment technology for exome sequencing using the Illumina TruSeq exome, Agilent SureSelectXT1 V4 and V5, and Haloplex exome, which offer a coverage of 96.12%, 91.61%, and 88.38%, respectively.

Although pharmacogenomic advances had been limited in the past due in part to the lack of coverage of commercial genotyping chips, it is anticipated that future studies that make use of new sequencing technologies should offer a greater potential for discovery.

Abbreviations: ADME = absorption, distribution, metabolism, and elimination, ADRs = adverse drug reactions, FDA = food and drug administration, GWA = genome-wide association, Hg = human genome, LD = linkage disequilibrium, SNPs = single nucleotide polymorphisms.

Keywords: ADME variants, coverage, enrichment, genotyping, pharmacogenomics

1. Introduction

Despite major recent progress in the knowledge of the human genome, interindividual genetic polymorphism, the metabolism of the drugs, and of the molecular techniques of biology, therapeutic individualization is not a current practice.^[1,2] Even with promising results, conclusive data are often missed and both effectiveness and tolerance of statistical data are still taken into consideration before suggesting treatment to a patient. The standardization of both protocols and doses, that are useful to a

period of development of new therapeutic indications, shows now its limits in term of effectiveness and tolerance.^[3]

Pharmacogenetics focuses on the study of metabolic differences between individuals as well as their absorption, distribution, metabolism, and elimination of drugs.^[4,5] The study of phenotypic expressions allowed genetic analysis, particularly interindividual differences in the enzymatic equipment and their effects on the metabolism of the organism.^[6-8] Pharmacogenomics, with the development of molecular genetics techniques, applies to the gene itself and not only its expression.^[9-11] It gathers pharmacogenetics and renews it by identifying different variations in the genome which are responsible for a modification in responses of the body.^[9] Pharmacogenomics, therefore, concerns the study of all the genes involved in drug response and refers to the use of genomics in the search for new therapeutic targets.^[12] Bioinformatics is an essential tool in pharmacogenomics.^[13,14] Thanks to its capacity to collect and treat important quantities of data, it allows the use of new methods in various fields, in particular, health.^[6,15]

The safety and efficacy of medication are of great clinical concern and several teams worldwide are investing concerted efforts toward the identification of the genetic causes of variable drug responses in the hopes of offering genetically-determined personalized therapy.^[14] The effective concentration of a drug at its effective site is a key determinant of the safety and efficacy of the drug. It depends on the absorption, distribution, metabolism, and elimination (ADME) mechanisms of the drug.^[15] Genetic variations in ADME genes can lead to large differences in drug exposure between individuals. Unlike other factors, genetic variations remain stable throughout a person's lifetime and

Editor: Nikhil Jain.

The authors have no conflict of interest to declare.

^a Faculty of Sciences, Mohammed V University, Rabat, ^b Biochemistry and Immunology Laboratory, Rabat, ^c Research Center, Abulcasis University of Health Sciences, Rabat, ^d Faculty of Medicine and Pharmacy of Casablanca, Laboratory of Genetics and Molecular pathologies, Hassan II University, ^e Mohammed VI University of Health Sciences (UM6SS), Casablanca, ^f Faculty of Sciences, Laboratory of Biology of Human Pathology, Center of Genomics of Human Pathologies, Mohammed V University, Rabat, Morocco.

* Correspondence: Younes Zaid, Abulcasis University of Health Sciences, Rabat, Morocco (e-mail: younes_zaid@yahoo.ca).

Copyright © 2019 the Author(s). Published by Wolters Kluwer Health, Inc. This is an open access article distributed under the terms of the Creative Commons Attribution-Non Commercial License 4.0 (CCBY-NC), where it is permissible to download, share, remix, transform, and buildup the work provided it is properly cited. The work cannot be used commercially without permission from the journal.

Medicine (2019) 98:3(e13975)

Received: 5 July 2018 / Received in final form: 20 November 2018 / Accepted: 10 December 2018

<http://dx.doi.org/10.1097/MD.0000000000013975>

provide a valuable means to predict drug response and prevent adverse drug reactions.^[16,17] Pharmacogenetics has achieved impressive progress toward the personalization of pharmaceutical treatment, with over 100 drugs in the list of the U.S. Food and Drug Administration (FDA) with genetic testing recommendations, ensuring drug safety and efficacy.^[18]

The interaction of these many genes and pathways are very complex, and current commercial platforms do not allow good coverage of the ADME variants.^[19,20]

Single nucleotide polymorphisms (SNPs) are the most frequent type of polymorphism in the human genome. They can provide a huge number of useful genetic markers for many genetic analyses (eg, phylogenetic analysis, ultra-dense genetic mapping, and genotype/phenotype association studies) and important applications (eg, cultivar identification and marker-assisted selection), which are simplified as these markers are most often biallelic.^[21]

This analysis aims to calculate the coverage rate of ADME genes for each of the variants and amplicons lists from different technologies. To this end, the genomic position of each genetic variation present in one of the ADME gene was used, according to the most recent build of the human genome (Hg19). Using Python scripts, the coverage rate of each of the genetic variation of interest was computed as an intersection between Interest list and ADME list divided over ADME list.

Coverage of the markers of interest that can be achieved with linkage disequilibrium (LD) with neighboring markers was also considered. To refine the work and make it more significant, theoretical results according to the targeted coverage of the different technologies and the practical results were compared. The result of this analysis allows for the identification of the technology which has the best possible coverage of the ADME genes.

2. Materials and methods

2.1. Generating data

Different technologies of interest were compared regarding their coverage of different ADME lists (Table 1). Variants of interest were extracted from ADME genes (Table 2) that were determined to be associated with drug metabolism.^[4,5,22] A total of 155 variations extracted from 34 genes that were considered to be of highest importance, the ADME Core.

In terms of enrichment platforms, we have:

1. the TrueSeq list, includes 20,794 genes that have 201,121 exons, spanning different databases: the NCBI CCDS (97.2%), NCBI RefSeq (96.4%), NCBI RefSeq including the non-coding DNA (88.3%), EncodeGenome of UCSC Genome Bioinformatics (93.2%), and microRNA (77.6%). The TrueSeq list is the result of the TrueSeq Technology product by Illumina, which represents the latest advances in Illumina targeted sequencing by capture hybridization.^[23]

- 2.

Table 1

ADME and genomic platforms used in this study.

Platform	Company	Number of genes
TrueSeq	Illumina	~21,000 genes
SureSelect V4	Agilent	~20,000 genes
SureSelect V5	Agilent	~21,000 genes
HaloPlex	Agilent	~21,000 genes

ADME = absorption, distribution, metabolism, and elimination.

Table 2

Most important ADME genes.

<i>ABCB1</i>	<i>CYP2B6</i>	<i>CYP3A4</i>	<i>NAT1</i>	<i>SLC01B1</i>	<i>UGT2B15</i>
<i>ABCC2</i>	<i>CYP2C19</i>	<i>CYP3A5</i>	<i>NAT2</i>	<i>SLC01B3</i>	<i>UGT2B17</i>
<i>ABCG2</i>	<i>CYP2C8</i>	<i>DPYD</i>	<i>SLC15A2</i>	<i>SLC02B1</i>	<i>UGT2B7</i>
<i>CYP1A1</i>	<i>CYP2C9</i>	<i>GSTM1</i>	<i>SLC22A1</i>	<i>SULT1A1</i>	<i>VKORC1</i>
<i>CYP1A2</i>	<i>CYP2D6</i>	<i>GSTP1</i>	<i>SLC22A2</i>	<i>TPMT</i>	
<i>CYP2A6</i>	<i>CYP2E1</i>	<i>GSTT1</i>	<i>SLC22A6</i>	<i>UGT1A1</i>	

ADME = absorption, distribution, metabolism, and elimination.

The SureSelect, a list of covered polymorphisms from the “Agilent Technologies” product for capture hybridization, which has the following website: <http://www.genomics.agilent.com/>. This list includes 554,751 amplicons likely to cover the variants of each one of the ADME lists.

3. The HaloPlex, technology provides outstanding performance, streamlined workflow, and low sample input requirements for next-generation sequencing of human exomes. The HaloPlex Exome has been optimized to provide comprehensive coverage of the coding regions of the human genome.

2.2. Steps

Before starting the calculation of the coverage of our interest lists, the data were processed to ensure organization in the appropriate format, complementation by finding the position of each ADME gene, purification by eliminating duplicates of each list, and standardization by unifying the build on hg19.

2.3. Selection of the database

The various databases offer essentially the same information. It was, therefore, necessary to select only 1 database to avoid redundancy. The 1000 Genomes Project is the first project to sequence the genomes of a large number of people, to provide a comprehensive resource on human genetic variation. Also, this database allows finding most genetic variants that have frequencies of at least 1% in the populations studied.^[24] Since the 1000 genomes database allows the extraction of these data in BED format files, which is an ideal format for data mining using PLINK, we chose this database as a study tool.

3. Results

3.1. Theoretical coverage of enrichment platforms

For enrichment platforms, we relied on probes contained in these lists. The coverage rates of the ADME genes by enrichment platforms are recorded in Table 3.

The TrueSeq list can theoretically cover 137 out of 155 ADME variants, the coverage rate is 88%. The coverage of the SureSelect

Table 3

Summary of coverage rates of our ADME lists of interest by enrichment platforms.

	Reference	Core (155)
TrueSeq	Target regions	88.38% (137)
SureSelect V4	Baits only	60.64% (94)
Sure Select V5	Baits only	70.96% (110)
HaloPlex	Baits only	92.26% (143)

ADME = absorption, distribution, metabolism, and elimination.

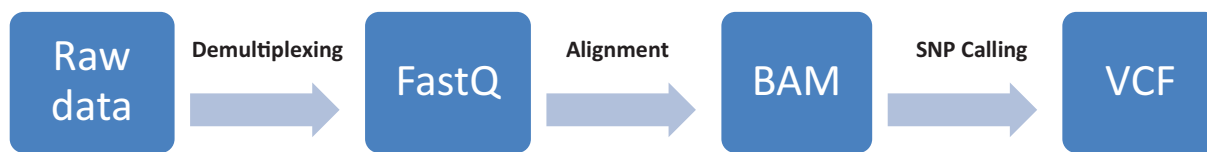


Figure 1. Steps of calculating the practical coverage.

Table 4
Practical coverage of ADME core variants by sequencing platforms according to the depth.

	5x	10x	15x	20x	25x
TrueSeq	147	146	146	142	142
SureSelect	147	136	126	111	107
HaloPlex	132	124	116	112	104

ADME = absorption, distribution, metabolism, and elimination.

is around 60% (version 4), and 70% (version 5). The Haloplex can theoretically cover 143 of the 155 ADME variants Core, which presents a rate of 92%. As we can infer, from enrichment platforms previously described, the Haloplex list allows the best coverage of the ADME variants. But this coverage remains theoretical.

3.2. Practical results

To refine the work and make it more meaningful, we calculated the coverage of the 155 ADME core variants that was achieved from the 3 sequencing platforms (TruSeq, SureSelect, and HaloPlex) by using read datasets made available in our lab according to a bioinformatics pipeline (Fig. 1).

We demultiplexed the sequencing raw data of different samples which allowed us to get the results as a “FASTQ” files. Then, we acceded to the pipeline using these files. The pipeline consists of aligning the sequences of the “FastQ” files according to a

reference genome.^[25] The result of this protocol is a “.BAM” file from which we calculated the depth (number of tags) of the practical coverage in 3 platforms, base by base, using the following command:

```
samtools mpileup FILE.bam
```

The last step of the pipeline consists in variant calling. In our study, we were limited to the calculation of the coverage from the alignment (“.BAM” file) because of the number of samples.

3.3. Comparison of the 3 sequencing platforms

The practical coverage (Table 4) of TrueSeq (made at 130x) list is between 95% (147 of 155) 5x, and 92% (142 of 155) 25x. SureSelect V5 has a practical coverage between 95% (147 of 155) at 5x, and 69% (107 of 155) at 25x. The SureSelect platform, unlike TruSeq, is made at 49x, which explains the low coverage at depths greater than 30x. The practical coverage of HaloPlex list (made at 42x) is between 85% (132 of 155) at 5x, and 67% (104 of 155) at 25x.

To make the comparison more significant between these 3 platforms, we normalized TruSeq and Sureselect at 42x. To this end, we developed a python script to eliminate randomly some reads from “bam files” of these platforms, to make them at 42x. The diagram (Fig. 2) shows the coverage of different platforms normalized at 42x, previously described in terms of depth.

As the diagram shows, by normalizing the data, Sureselect and TruSeq allow a good coverage (96% and 95%,

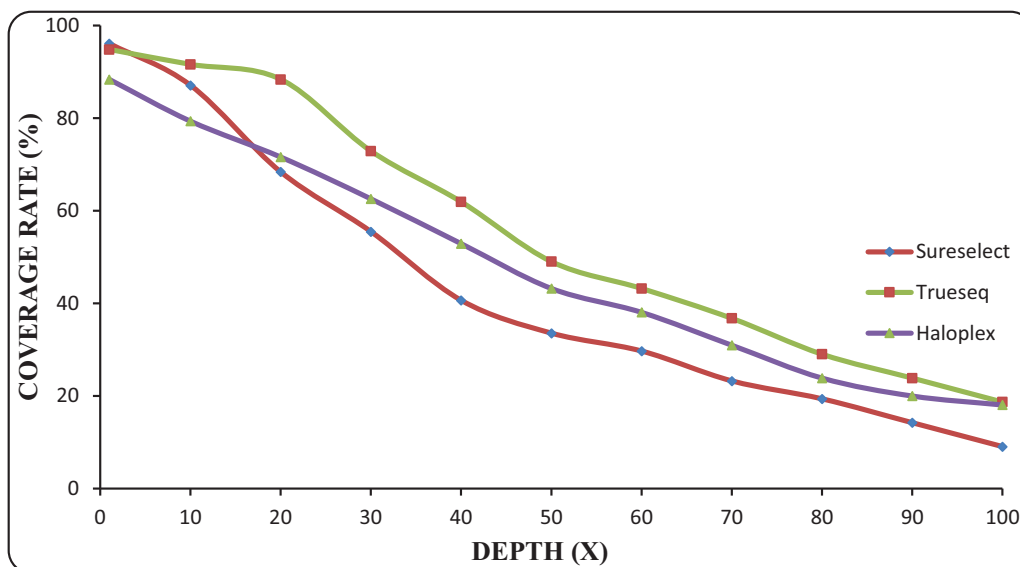


Figure 2. Comparison of the coverage for 3 sequencing platforms normalized at 42x.

Table 5
Pharmacogenomic biomarkers in drug labeling.

Drug	Therapeutic area	HUGO symbol
Dexlansoprazole (2)	Gastroenterology	<i>CYP1A2</i>
Carisoprodol	Rheumatology	<i>CYP2C19</i>
Citalopram (1)	Psychiatry	<i>CYP2C19</i>
Clobazam	Neurology	<i>CYP2C19</i>
Clopidogrel	Cardiology	<i>CYP2C19</i>
Dexlansoprazole (1)	Gastroenterology	<i>CYP2C19</i>
Diazepam	Psychiatry	<i>CYP2C19</i>
Drospirenone and Ethinyl Estradiol	Neurology	<i>CYP2C19</i>
Esomeprazole	Gastroenterology	<i>CYP2C19</i>
Lansoprazole	Gastroenterology	<i>CYP2C19</i>
Omeprazole	Gastroenterology	<i>CYP2C19</i>
Pantoprazole	Gastroenterology	<i>CYP2C19</i>
Prasugrel	Cardiology	<i>CYP2C19</i>
Rabeprazole	Gastroenterology	<i>CYP2C19</i>
Ticagrelor	Cardiology	<i>CYP2C19</i>
Celecoxib	Rheumatology	<i>CYP2C9</i>
Flurbiprofen	Rheumatology	<i>CYP2C9</i>
Warfarin (1)	Cardiology or Hematology	<i>CYP2C9</i>
Amitriptyline	Psychiatry	<i>CYP2D6</i>
Aripiprazole	Psychiatry	<i>CYP2D6</i>
Atomoxetine	Psychiatry	<i>CYP2D6</i>
Carvedilol	Cardiology	<i>CYP2D6</i>
Cevimeline	Dermatology	<i>CYP2D6</i>
Citalopram (2)	Psychiatry	<i>CYP2D6</i>
Clomipramine	Psychiatry	<i>CYP2D6</i>
Clozapine	Psychiatry	<i>CYP2D6</i>
Codeine	Anesthesiology	<i>CYP2D6</i>
Desipramine	Psychiatry	<i>CYP2D6</i>
Dextromethorphan and Quinidine	Neurology	<i>CYP2D6</i>
Doxepin	Psychiatry	<i>CYP2D6</i>
Fluoxetine	Psychiatry	<i>CYP2D6</i>
Fluvoxamine	Psychiatry	<i>CYP2D6</i>
Galantamine	Neurology	<i>CYP2D6</i>
Iloperidone	Psychiatry	<i>CYP2D6</i>
Imipramine	Psychiatry	<i>CYP2D6</i>
Metoprolol	Cardiology	<i>CYP2D6</i>
Modafinil	Psychiatry	<i>CYP2D6</i>
Nefazodone	Psychiatry	<i>CYP2D6</i>
Nortriptyline	Psychiatry	<i>CYP2D6</i>
Paroxetine	Psychiatry	<i>CYP2D6</i>
Perphenazine	Psychiatry	<i>CYP2D6</i>
Pimozide	Psychiatry	<i>CYP2D6</i>
Propafenone	Cardiology	<i>CYP2D6</i>
Propranolol	Cardiology	<i>CYP2D6</i>
Protriptyline	Psychiatry	<i>CYP2D6</i>
Quinidine	Cardiology	<i>CYP2D6</i>
Risperidone	Psychiatry	<i>CYP2D6</i>
Tetrabenazine	Neurology	<i>CYP2D6</i>
Thioridazine	Psychiatry	<i>CYP2D6</i>
Tolterodine	Urology	<i>CYP2D6</i>
Tramadol	Analgesic	<i>CYP2D6</i>
Trimipramine	Psychiatry	<i>CYP2D6</i>
Venlafaxine	Psychiatry	<i>CYP2D6</i>
Fluorouracil (1)	Dermatology	<i>DPYD</i>
Isosorbide and Hydralazine	Cardiology	<i>NAT1-2</i>
Azathioprine	Rheumatology	<i>TPMT</i>
Indacaterol	Pulmonary	<i>UGT1A1</i>
Warfarin (2)	Cardiology or Hematology	<i>VKORC1</i>

respectively) at a depth of 1x. However, we cannot make a “SNP calling” with 1x. At depths from 10x to 20x, Sureselect coverage falls to 68% while TruSeq still allows a good coverage (88%). At depths greater than 20x, it is the TruSeq

that allows the best coverage, even if it does not exceed the HaloPlex coverage rate significantly.

4. Discussion

The aim of this work is to evaluate the coverage of clinically relevant ADME genes and specific variants of interest in those genes that can be obtained from commercial sequencing platforms so that this can guide platform selection in preparation of a study design.

According to previous studies, the highest coverage of genotyping arrays was 67% (105 of 155 variants), taking into account a LD threshold of 0.8. Among the genotyping platforms assessed previously by Gamazon,^[3] no platform showed a good coverage of ADME genes (after accounting for LD) sufficiently to conduct pharmacogenomic studies. This is particularly important as several metabolizer phenotypes can only be predicted from diplotype status.

According to the FDA, the most important pharmacogenes are *CYP2D6* (rs765776661, rs267608302) and *CYP2C19* (rs12248560, rs17884712, rs140278421, rs138142612, rs4917623, rs192154563) (Table 5). From the 87 FDA approved drugs (excluding biomarkers of cancerous tissues or virus), these 2 genes are linked to 49 drugs, which presents 56% of all FDA approved drugs.^[26] In term of coverage, Omni 5.0 chip contains 8 more pharmacogenes, namely: *CYP1A1*, *CYP2C9*, *DPYD*, *NAT1*, *NAT2*, *TPMT*, *UGT1A1*, and *VKORC1*, in addition to *CYP2D6*, and *CYP2C19*, which allows a coverage of 58 drugs as shown in Table 5.

The lack of coverage is due to the choice of a set of homologous genes. However, these genes are the most important in personalized medicine and pharmacogenomics, hence the importance of these results to those interested in the evaluation of genomic data.

Previous studies have shown the limits of genome-wide methods for pharmacogenomic testing. Gamazon et al.^[3] focused on a set of the most important genes in pharmacogenomics and personalized medicine, using only genotyping platforms. Their findings demonstrated that even by taking into consideration the SNPs which are in LD, the coverage rate of these genes by genotyping platforms is sub-optimal. In this study, we assessed also exome sequencing platforms. The HaloPlex enrichment allows the best coverage of the ADME variants. However, this coverage remains theoretical. To further evaluate the coverage that can be obtained using the sequencing technologies, we used practical sequencing results from 3 platforms, namely: TruSeq, HaloPlex, and Sureselect. According to the results obtained, we can conclude that among the 4 commercial platforms tested, TruSeq offers a good coverage of our ADME variants of interest.

Author contributions

Conceptualization: Nabil Zaid, Nezha Senhaji, Loubna Khalki, Younes Zaid, and Saaid Amzazi.

Data curation: Nabil Zaid, Youness Limami, Loubna Khalki, Younes Zaid, and Saaid Amzazi.

Formal analysis: Nabil Zaid, Nezha Senhaji, and Saaid Amzazi.

Investigation: Nabil Zaid, Youness Limami, and Saaid Amzazi.

Methodology: Nabil Zaid, Youness Limami, Nezha Senhaji, Nadia Errafiy, Youssef Bakri, Younes Zaid, and Saaid Amzazi.

Resources: Nadia Errafiy.

Software: Nabil Zaid, Youness Limami, and Nadia Errafiy.

Supervision: Youness Limami, Younes Zaid, and Saaid Amzazi.

Validation: Nabil Zaid, Youness Limami, Youssef Bakri, Younes Zaid, and Saaid Amzazi.

Visualization: Nabil Zaid, Nadia Errafiy, and Saaid Amzazi.

Writing – original draft: Nabil Zaid, Nezha Senhaji, Loubna Khalki, Youssef Bakri, and Younes Zaid.

Writing – review & editing: Younes Zaid and Saaid Amzazi.

References

- [1] Engels FK, Loos WJ, van der Bol JM, et al. Therapeutic drug monitoring for the individualization of docetaxel dosing: a randomized pharmacokinetic study. *Clin Cancer Res* 2011;17:353–62.
- [2] Tanaka N, Kurose T, Seino Y. Therapeutic inertia or individualization? Delay in clinical management of type 2 diabetes mellitus. *Curr Med Res Opin* 2016;32:1477–8.
- [3] Gamazon ER, Skol AD, Perera MA. The limits of genome-wide methods for pharmacogenomic testing. *Pharmacogenet Genom* 2012;22:261–72.
- [4] Grossman I. ADME pharmacogenetics: current practices and future outlook. *Expert Opin Drug Metab Toxicol* 2009;5:449–62.
- [5] Lubomirov R, di Iulio J, Fayet A, et al. ADME pharmacogenetics: investigation of the pharmacokinetics of the antiretroviral agent lopinavir coformulated with ritonavir. *Pharmacogenet Genom* 2010;20:217–30.
- [6] Bogenpohl JW, Mignogna KM, Smith ML, et al. Integrative analysis of genetic, genomic, and phenotypic data for ethanol behaviors: a network-based pipeline for identifying mechanisms and potential drug targets. *Methods Mol Biol* 2017;1488:531–49.
- [7] Wu D, Zhang H, Wang H, et al. Phenotypic and genetic analysis of a boy with partial trisomy of 1q. *Zhonghua Yi Xue Yi Chuan Xue Za Zhi* 2017;34:398–401.
- [8] Zhong F, Zhong F, Lan F, et al. Phenotypic and genetic analysis of an inv dup(15) case with a BP3:BP3 rearrangement. *Zhonghua Yi Xue Yi Chuan Xue Za Zhi* 2017;34:402–5.
- [9] Auman JT, McLeod HL. Cancer pharmacogenomics: DNA genotyping and gene expression profiling to identify molecular determinants of chemosensitivity. *Drug Metab Rev* 2008;40:303–15.
- [10] Garcia-Campelo R, Alonso-Curbera G, Anton Aparicio LM, Rosell R. Pharmacogenomics in lung cancer: an analysis of DNA repair gene expression in patients treated with platinum-based chemotherapy. *Expert Opin Pharmacother* 2005;6:2015–26.
- [11] Kassack MU. Pharmacogenomics and gene expression analysis. *Functional genome research for individual application to patients. Med Monatsschr Pharm* 2003;26:165–6.
- [12] Lubomirov R, Csajka C, Telenti A. ADME pathway approach for pharmacogenetic studies of anti-HIV therapy. *Pharmacogenomics* 2007;8:623–33.
- [13] Wang JF, Wei DQ. Role of structural bioinformatics and traditional Chinese medicine databases in pharmacogenomics. *Pharmacogenomics* 2009;10:1213–5.
- [14] Yan Q. The integration of personalized and systems medicine: bioinformatics support for pharmacogenomics and drug discovery. *Methods Mol Biol* 2008;448:1–9.
- [15] Thorn CF, Klein TE, Altman RB. Pharmacogenomics and bioinformatics: PharmGKB. *Pharmacogenomics* 2010;11:501–5.
- [16] Bertsche T, Pfaff J, Schiller P, et al. Prevention of adverse drug reactions in intensive care patients by personal intervention based on an electronic clinical decision support system. *Intensive Care Med* 2010;36:665–72.
- [17] Burkhardt H, Wehling M, Gladisch R. Prevention of adverse drug reactions in older patients. *Z Gerontol Geriatr* 2007;40:241–54.
- [18] Fabbri C, Porcelli S, Serretti A. From pharmacogenetics to pharmacogenomics: the way toward the personalization of antidepressant treatment. *Can J Psychiatry* 2014;59:62–75.
- [19] Krejsa CM, Horvath D, Rogalski SL, et al. Predicting ADME properties and side effects: the BioPrint approach. *Curr Opin Drug Discov Devel* 2003;6:470–80.
- [20] Smith DA. Evolution of ADME science: where else can modeling and simulation contribute? *Mol Pharm* 2013;10:1162–70.
- [21] Zhang L, Li WH. Human SNPs reveal no evidence of frequent positive selection. *Mol Biol Evol* 2005;22:2504–7.
- [22] Brown AM, Renaud Y, Ross C, et al. Development of a broad-based ADME panel for use in pharmacogenomic studies. *Pharmacogenomics* 2014;15:1185–95.
- [23] Guo Y, Long J, He J, et al. Exome sequencing generates high quality data in non-target regions. *BMC Genomics* 2012;13:194.
- [24] Genomes Project C, Abecasis GR, Altshuler D, et al. A map of human genome variation from population-scale sequencing. *Nature* 2010;467:1061–73.
- [25] Bonfield JK, Mahoney MV. Compression of FASTQ and SAM format sequencing data. *PLoS One* 2013;8:e59190.
- [26] Jain KK. Synthetic biology and personalized medicine. *Med Princ Pract* 2013;22:209–19.