Pharmacogenetics in the Era of Next Generation Sequencing

Bottillo I[1]*, Morrone A[2] and Grammatico P[1]

[1]U.O.C. Medical Genetics Laboratory, "Sapienza – University of Rome", Molecular Medicine Department, S. Camillo-Forlanini Hospital, Rome, Italy
[2]S. Camillo-Forlanini Hospital, Healthcare Administration, Rome, Italy

Pharmacogenetics is the study of the effects of each individual genome on drug response. Information about genetic testing is now part of drug label as for abacavir, warfarin, clopidogrel, irinotecan, maraviroc, cetuximab, imatinib [1,2]. At present a wide range of drugs are employed for treating pathological conditions and chronic disorders, like cancer, diabetes, heart disease, etc. Ideally a drug is delivered so that it achieves a blood concentration that is within the therapeutic range: a too low drug concentration is ineffective, while one that is too high can be toxic. None of the drugs known to date proved to be 100% effective in every patient: the same dose of a given medicine may cause no response in some individuals, may produce the right therapeutic response in others, and may give rise to adverse responses in another part of them. These changes in the response may result from clinical differences between individuals (for example age or the presence of other pathological conditions), from diagnostic uncertainty, and from environmental factors. Besides, genetics has been proven to modulate in a significant way inter-patients variation in drug reaction. Genes are sequences of DNA that contain information determining the amino acid sequence of proteins, which participate in every cellular process, including drug response. Polymorphisms are DNA variations that occur at a frequency of 1% or greater, while DNA mutations occur less frequently than 1%: this definition is merely statistical and it does not reflect the functional role of both kind of genetic variants. In both cases in fact, if the DNA carries a sequence variation (a mutation or a polymorphism), the proteins may not be expressed properly, or not function correctly, or not reach the appropriate site of action. Regarding pharmacogenetics, individual genetic variations in enzymes of drug metabolizing enzymes, transporters, receptors, ion channels, target enzymes and signal transduction pathways, may lead to inadequate therapeutic responses. For example the genetic information of genes *CYP2C9* (encoding a member of the cytochrome P450 superfamily of enzymes, which catalyze many reactions involved in drug metabolism and synthesis of cholesterol, steroids and other lipids) and *VKORC1* (encoding the enzyme which activates vitamin K) can assist in selection of the starting dose of warfarin [3] an anticoagulant which acts by inhibiting vitaminK-dependent coagulation factors. In particular, warfarin interferes with VKORC1enzyme, and DNA variants in *CYP2C9* are related to differences in the drug's clearance. The *VKORC1* G1639A polymorphism is associated with lower dose requirements and increased bleeding risk, as well as inferior initial warfarin doses have been associated with the CYP2C9*2 and CYP2C9*3 variants [4-6]. Algorithms for genetics-based dosing have been developed and are available online (Online calculator for initiation of warfarin dosing based on pharmacogenetic algorithms, www.WarfarinDosing.org) [7].

The history of pharmacogenetics starts in 1957 when Arno Motulsky proposed that "*...hereditary gene-controlled enzymatic factors determine why, with identical exposure, certain individuals become "sick" whereas others are not affected*" [8] (Figure 1). Two years later Vogel coined the term "pharmacogenetics" to describe the relationship between genetic factors and response to medications [9] (Figure 1). The periods in and around the 1950s and the year 2000, capped by the discovery of the DNA double helix and the maturation of the human genome initiative respectively, were periods of remarkable achievement

in the quest for knowledge of heredity. Moreover biochemistry and molecular genetics researches were focused on studying both enzymes involved in drugs' metabolism and their corresponding DNA coding sequences. Afterwards the introduction of sequencing technologies simplified the identification of genetic DNA sequence variants that can control enzymatic activity. Registers of these variants are now available online (http://www.cypalleles.ki.se/) [10]. However the *loci* relevant for the pharmacokinetics and especially the pharmacodynamics of most drugs are still unknown. Discovery of most types of genetic variation depends on DNA sequencing. Through Sanger method, also now known as "first generation sequencing" [11] (Figure 1), earned Frederick Sanger the Nobel Prize in 1980, it has been possible to complete the sequence of the human genome [12] (Figure 1). Sanger sequencing has dominated the past few decades [13], even facing limitations in the experiments' throughput (number of samples and number of genes that can be analyzed in a single experiment) and costs. As that the Human Genome Project lasted 13 years [14], the need to conduct large-scale sequencing projects for the rapid incorporation of genetic information into clinical care has led to the development of a variety of "next-generation sequencing" (NGS) technologies (Figure 1). On comparing with "one experiment-one fragment" Sanger method, the massively parallel techniques permit to analyze millions of DNA molecules in a single experimental run. Perhaps more important than the sequencing throughput provided by this technology and its relative low cost compared with traditional sequencing methods, is the type of data it generates. Instead of long reads generated from a PCR-amplified and Sanger-sequenced sample, NGS methods provide much shorter reads, but millions of them. The short reads can be counted and quantified, allowing accurate copy number assessment of each genomic region. At present the whole-genome sequence can be done in a few weeks. Another NGS approach is sequencing of target genomic regions (both candidate genes and regulatory non-coding regions) by an experimental option that is more effective both for cost and time. A targeted but genome-wide approach is the exome (the protein coding sequences for all genes, comprising approximately 1–2% of the genome) sequencing [15].
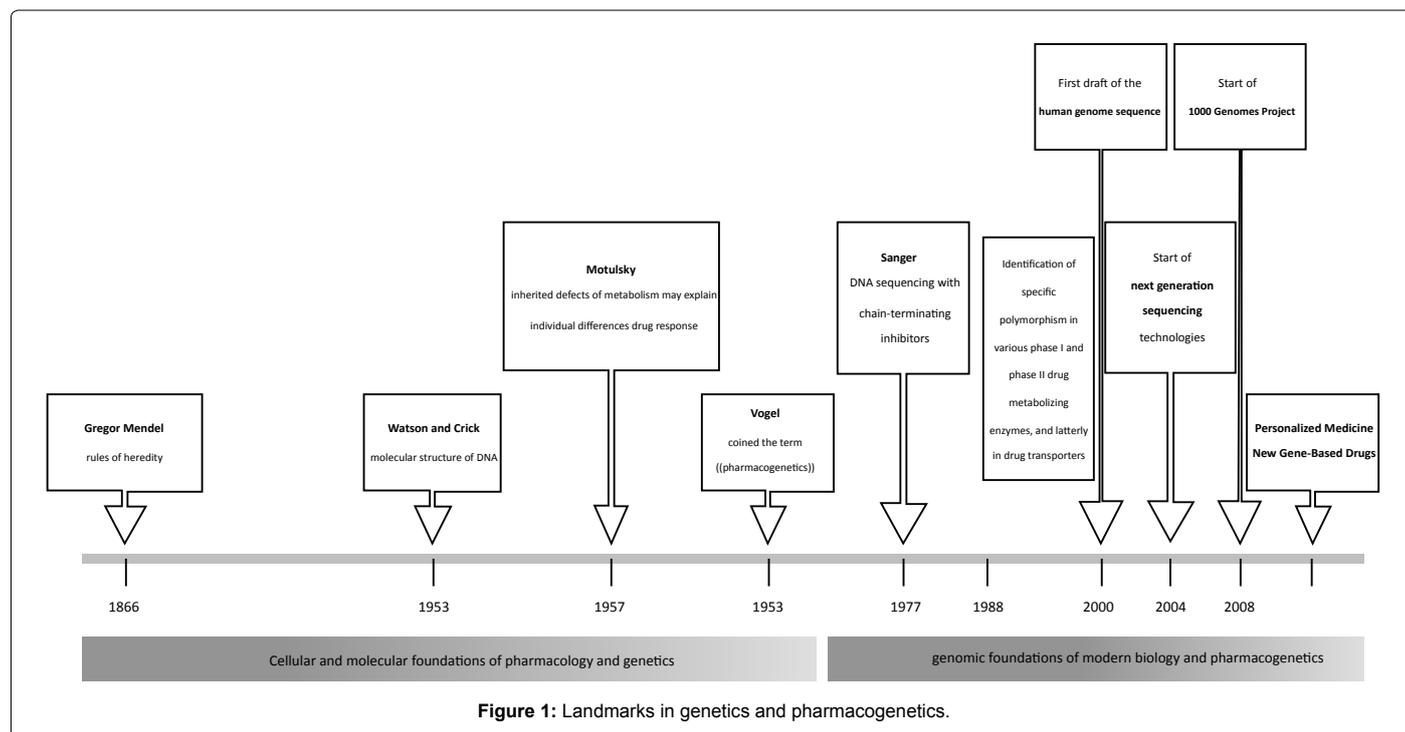
At present clinical application of NGS revealed a great amount of genetic differences among individuals including DNA variations in the coding sequence and/or regulatory regions of genes encoding drug metabolizing enzymes, transporters, receptors, or drug target genes

Figure 1: Landmarks in genetics and pharmacogenetics.

[16]. It has been estimated that the sequencing of about 1000 individuals from the general population will allow the identification of 99% of common and uncommon human genomic variants [17]. To this end a funded project (the 1000 Genomes Project) was launched in 2008 using NGS technologies to obtain complete genome sequences for this many subjects (Figure 1). This project is now underway (www.1000genomes.org) and its specific objectives are to identify most of human genome single nucleotide polymorphisms (SNPs) and structural variations (such as copy-number variants, insertions, deletions and inversions) in individuals from the general population [18]. Moreover a pharmacogene database enhanced by the 1000 Genomes Project was built for the community to immediately evaluate and utilize the sequencing data, every time they are released (http://genemed1.bsd.uchicago.edu/pharmacodb/thougen/main.php) [19]. Particularly, this database can be used to access SNP genotypic calls of 39 pharmacogenetic candidate genes, maintained by the Very Important Pharmacogenes (VIP) project of the Pharmacogenetics Knowledge Base (PharmGKB, www.PharmGKB.org). TheVIP project is an initiative to provide annotated information regarding genes, variants, haplotypes and splice variants of particular relevance for pharmacogenetics.

In the next generation sequencing era, given the means to comprehensively catalog human genetic variation, the challenge for pharmacogenetics is aimed to discover new genomic variants modifying inter-individual drugs' response and to apply this new knowledge to the clinical practice. This will allow the prescription of the safest and most effective drug for every patient. Thanks to new technologies, genome wide approaches to discover variants important in drug response and/or toxicity have emerged, shifting the studies of variants within one or more candidate genes (pharmacogenetics), towards evaluating the entire genome for associations with pharmacologic phenotypes (pharmacogenomics). Besides, one of the major impacts of whole genome methods is that it has become possible to identify DNA variants outside the genes already known to be involved in drugs' metabolism, disposition and/or response. Therefore new genetic *loci*

can emerge as important elements in the biology of a given drug. At present identifying and validating genetic *loci* involved in drugs' response and potential therapeutic drug targets, are studies that can be approached by the use of a single technological NGS platform, applying different experimental methods including: (i) whole genome DNA sequencing for identificationof new coding and non-coding DNA variations modulating drugs' metabolism; (ii) RNA sequencing to perform gene expression studies of "affected versus normal" tissues [20]; (iii) chromatin immunoprecipitation sequencing (NGS-ChIP-Seq) to identify DNA-interacting proteins [21].

The future of pharmacogenomics and the dream of providing the right drug, in the right dosage, to the right patient, is dependent on the cross fertilization of ideas and research; that is, the coordination of efforts in both the laboratories and clinics to translate bench science findings into the clinic. Pharmacogenomic studies will increase our knowledge of genetic markers predictive of drug response and toxicity, and most importantly, will improve efficacy and safety of current medications. The analysis of the data produced from exome and genome sequencing is massive. Even more challenging is the interpretation of these data. Improvements in bioinformatics tools will be needed to assess the functional significance of rare variants identified in genes of potential pharmacogenetic impact. Major advances will be needed to transfer the genetic information to the clinical practice and it is too early to know whether use of these data will routinely improve patient outcomes, leading to substantial healthcare cost-savings: if treatment will be guided by "pharmacogenomic profiles", therapeutic failure and drug toxicity would be greatly reduced or avoided in genetically susceptible persons, as already happens for warfarin: identification of the genetic variants in *CYP2C9* and *VKORC1* increases the need for more frequent PT-INR (prothrombin time-international normalized ratio) monitoring only in at-risk patient.

References

1. Ginsburg GS, Willard HF (2009) Genomic and personalized medicine: foundations and applications. Transl Res 154: 277-287.

2. Bakhouche H, Slanař O (2012) Pharmacogenetics in clinical practice. Prague Med Rep 113: 251-261.

3. Rieder MJ, Reiner AP, Gage BF, Nickerson DA, Eby CS, et al. (2005) Effect of VKORC1 haplotypes on transcriptional regulation and warfarin dose. N Engl J Med 352: 2285-2293.

4. Lindh JD, Lundgren S, Holm L, Alfredsson L, Rane A (2005) Several-fold increase in risk of overanticoagulation by CYP2C9 mutations. Clin Pharmacol Ther 78: 540-550.

5. Sanderson S, Emery J, Higgins J (2005) CYP2C9 gene variants, drug dose, and bleeding risk in warfarin-treated patients: a HuGEnet systematic review and meta-analysis. Genet Med 7: 97-104.

6. Wadelius M, Chen LY, Downes K, Ghori J, Hunt S, et al. (2005) Common VKORC1 and GGCX polymorphisms associated with warfarin dose. Pharmacogenomics J 5: 262-270.

7. Klein TE, Chang JT, Cho MK, Easton KL, Fergerson R, et al. (2001) Integrating genotype and phenotype information: an overview of the PharmGKB project. Pharmacogenetics Research Network and Knowledge Base. Pharmacogenomics J 1: 167-170.

8. MOTULSKY AG (1957) Drug reactions enzymes, and biochemical genetics. J Am Med Assoc 165: 835-837.

9. Vogel F (1959) Moderne Probleme der Humangenetik. Ergeb Inn Med Kinderheilkd 12:52-125.

10. Sim SC, Ingelman-Sundberg M (2006) The human cytochrome P450 Allele Nomenclature Committee Web site: submission criteria, procedures, and objectives. Methods Mol Biol 320: 183-191.

11. Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors 1977. Proc Natl Acad Sci USA 74: 5463-5467.

12. Wheeler DL, Barrett T, Benson DA, Bryant SH, Canese K, et al. (2008) Database resources of the National Center for Biotechnology Information. Nucleic Acids Res 36: D13-D21.

13. Metzker ML (2010) Sequencing technologies - the next generation. Nat Rev Genet 11: 31-46.

14. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, et al. (2001) The sequence of the human genome. Science 291:1304-1351.

15. Ng SB, Turner EH, Robertson PD, Flygare SD, Bigham AW, et al. (2009) Targeted capture and massively parallel sequencing of 12 human exomes. Nature 461: 272-276.

16. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, et al. (2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet 43: 491-498.

17. Ionita-Laza I, Lange C, M Laird N (2009) Estimating the number of unseen variants in the human genome. Proc Natl Acad Sci U S A 106: 5008-5013.

18. 1000 Genomes Project Consortium, Abecasis GR, Altshuler D, Auton A, Brooks LD, et al. (2010) A map of human genome variation from population-scale sequencing. Nature 467: 1061-1073.

19. Gamazon ER, Zhang W, Huang RS, Dolan ME, Cox NJ (2009) A pharmacogene database enhanced by the 1000 Genomes Project. Pharmacogenet Genomics 19: 829-832.

20. Shendure J (2008) The beginning of the end for microarrays? Nat Methods 5: 585-587.

21. Nutiu R, Friedman RC, Luo S, Khrebtukova I, Silva D, et al. (2011) Direct measurement of DNA affinity landscapes on a high-throughput sequencing instrument. Nat Biotechnol 29: 659-664.