Whole-genome resequencing in pharmacogenomics: moving away from past disparities to globally representative applications

Africa suffers from a high burden of disease; nonetheless, it has been one of the most under-represented continents with regard to genomic research. It can be argued that this disproportionate research is related to the fact that the genome architecture of African individuals is poorly suited to SNP-based genome-wide association studies, given existing genotyping platforms. However, this argument is no longer plausible with the arrival of next-generation sequencing technologies, which allow for the analysis of entire genomes. Using pharmacogenes to critically examine the merit of next-generation sequencing technologies in pharmacogenomics, we found a substantial amount of novel/uncharacterized variation, which was predicted to alter protein function. This variation was predominantly observed in African individuals, emphasizing the benefit of next-generation sequencing technologies specifically for these individuals. We also observed an improvement in the reliability of sequencing technologies in a relatively short time. Therefore, as sequencing technologies develop and decrease in cost, the ability to reliably detect variation will improve and these technologies will begin to replace other less comprehensive genotyping assays.

KEYWORDS: Africa ■ genomics ■ next-generation sequencing ■ pharmacogenes ■ pharmacogenomics

Britt I Drögemöller[1], Galen EB Wright[1], Dana JH Niehaus[2], Robin A Emsley[2] & Louise Warnich*[1]

[1]Department of Genetics, Stellenbosch University, Private Bag XI, Matieland 7602, South Africa
[2]Department of Psychiatry, Stikland Hospital, Stellenbosch University, South Africa
**Author for correspondence:*
Tel.: +27 218 085 888
Fax: +27 218 085 833
lw@sun.ac.za

Inequalities in genomic research

There has been a large disparity in genomic research, with approximately 75% of genome-wide association studies (GWAS) having been performed in individuals of European descent [1], while the remaining GWAS provide a poor representation of African individuals. When consulting the existing database of GWAS literature (last updated 29 July 2011), out of a total of 5854 possible 'GWAS hits', 465 were related to the search term 'African' [101]. However, from these results, only six studies referred to specific populations residing within the continent (i.e., west Africans, Nigerians, Malawians, Gambians and Ghanaians [2–7]), while all other studies referred to African–Americans or African-descent individuals. As can be seen in Figure 1, the 'GWAS hits' that have been obtained from African populations represent a very small portion of the continent and only fall in areas of the Niger–Kordofanian language family distribution [8,9]. Since Africa is host to 30.5% of the world's 6909 living languages [102], it is not realistic to refer to Africans as one homogenous population group, nor are the results obtained from the GWAS performed on the highly admixed African–American population a good representation of the genetic variation present in the populations residing within Africa [10].

The under-representation of studies examining populations residing within Africa seems ironic when considering that out of the 28 countries worldwide that have a death rate higher than 1850 per 100,000, 26 of these countries are situated in Africa [103], making Africa one of the continents most likely to benefit from the translation of GWAS results. Furthermore, due to the shortage of genomic research in Africa, those GWAS that have been performed include the more frequently studied tuberculosis, HIV and malaria [3,6,7], while other disorders, such as complex psychiatric disorders, are absent. The lack of psychiatric GWAS in Africa is further highlighted by the absence of these studies in the psychiatric GWAS consortium, whose goal is to conduct meta-analyses utilizing GWAS data for psychiatric disorders [104]. This absence is a result of the better genome-wide coverage and greater amounts of data available for European descent individuals; however, it is the long-term goal of the psychiatric GWAS consortium to include African descent individuals [104]. This emphasizes a need for additional studies within Africa that move beyond the more frequently examined epidemics to cover a wider variety of diseases and disorders. It should also be noted that one of the predominant explanations for the lack of genomic research in Africa can be attributed to a lack of funding, resources and infrastructure within the continent. This in turn results in a lack of expertise and subsequently a shortage of comprehensive genetic/genomic studies pertaining

future medicine part of fsg

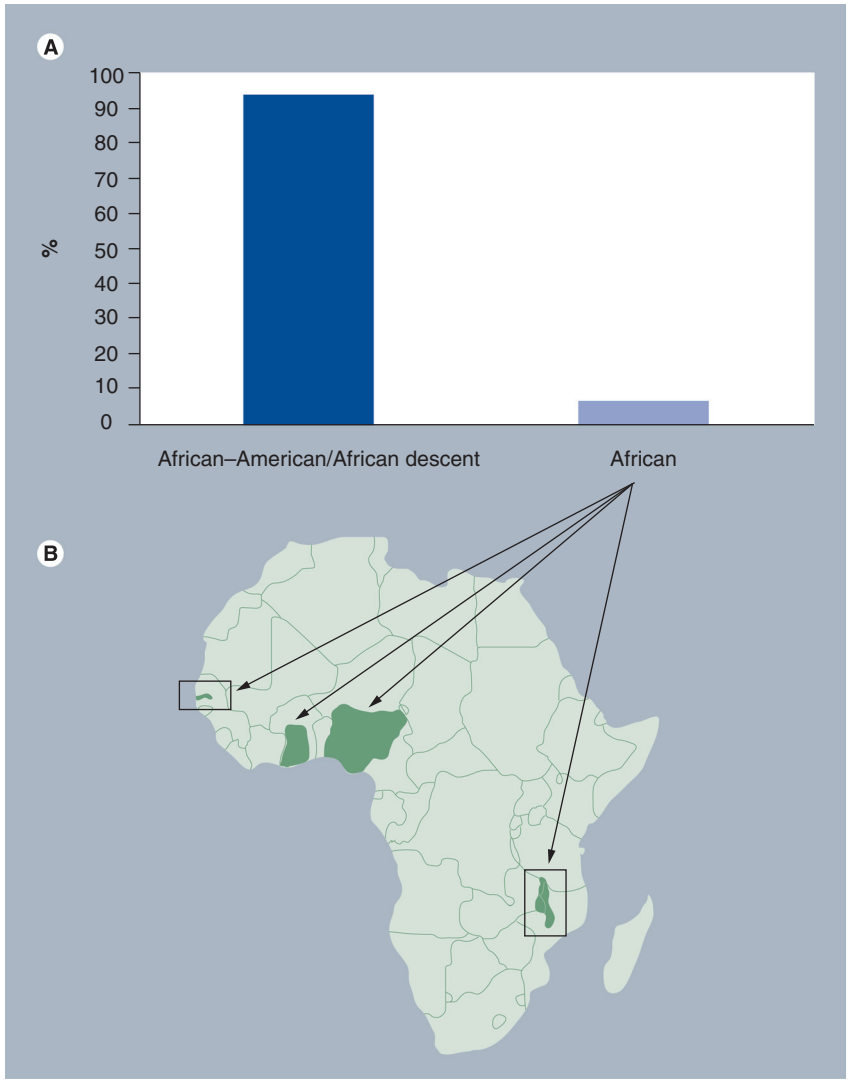

Figure 1. Disparities in genome-wide association study hits. (A) Percentage African 'genome-wide association studies hits', which represent approximately 8% of the genome-wide association studies performed to date, stratified by those studies using specific populations within Africa and those using populations defined as African–American or African descent [101]. **(B)** The countries in Africa from which the genome-wide association studies cohorts were gathered. This image highlights the fact that those few studies that have used populations from within Africa represent only a very small portion of the continent.

to African individuals [11]. Without access to large cohorts of well-characterized African individuals, current technologies have been unable to incorporate the genomic architecture of these populations into genome-wide assays.

Fortunately the realization that we need to address these research disparities has come at the right time. Although the GWAS of the past have focused on European-descent populations, it can be argued that these types of GWAS, utilizing SNP genotyping, are well suited to the genomic architecture of European descent individuals, while the low linkage disequilibrium (LD) present in African individuals is uniquely suited for fine-mapping [1,12]. Therefore by utilizing individuals of African ancestry in resequencing studies, we may be able to more easily elucidate causal variants, owing to the lower LD in these individuals. By identifying causal variants as opposed to those that are merely in LD with these variants, we may be able to increase the amount of evidence-based guidelines to aid in the application of genomic results into the clinical setting [13]. Furthermore, Africa provides the ultimate setting for the translation of genomic results, due to its high disease burden and ample disease cohorts [14,15], with reference to both communicable [16–19] and noncommunicable diseases [20,21]. However, before Africa enters the genomic research arena, it is important that researchers focus on utilizing techniques that will benefit from the complex architecture of African genomes. Bearing this in mind, there is evidence to suggest that SNP-based GWAS employing commercial microarrays may be replaced, in many areas, with the use of 'next-generation' sequencing technologies [22]. Therefore, Africa may be entering the field of genomics at the perfect time, coinciding with the explosion in sequencing technologies.

Next-generation sequencing technologies

This year, as we commemorate the decade that has passed since the human genome sequence was published, a retrospective look at sequencing reveals one of the most rapidly advancing technologies to date. Although Sanger sequencing, published in 1977 [23], earned Frederick Sanger the Nobel Prize in 1980, the sequencing of the human genome using this technique took 13 years to complete [24]. Today, with the use of next-generation sequencing technologies, the whole-genome sequence can be obtained in a few weeks and the associated cost has dropped approximately 1 million-fold [25], making the field of genomics one of the leaders in science and technology at present. Even so, the potential of high-throughput sequencing remains largely unharnessed and there is much room for improvement. By smoothing out the associated flaws and implementing whole-genome sequencing on a regular basis, it is hoped that the criticism associated with the lack of translatable results obtained from the Human Genome Project can be addressed and genomic applications in the clinical setting will become routine.

Exciting as this technology may be, it is important that the biases of past genomic research are not repeated and that all populations are equally represented. Even though European ancestry

individuals have been over-represented in GWAS to date, the lack of clinically applicable results obtained from these studies have fortunately prevented a large increase in global health disparities as a result of this research bias. However, if this uneven research continues into whole-genome sequencing studies, which are more likely to identify causal variants, the knowledge gap and health disparities between developing and developed countries are likely to increase, to the detriment of developing countries [11]. Unfortunately, even in the short time that whole-genome resequencing technologies have been available, research disparities are already emerging. These disparities refer both to sequencing-related equipment and resequencing data. While a substantial amount of both Caucasian and Asian genomes are being sequenced, the resequencing of African genomes is lagging behind [26]. Due to the fact that all humans originate from Africa and African populations have largely avoided the bottleneck effects experienced by non-Africans [27,28], genomic characterization of these populations should provide the most comprehensive catalogue of human variation. Therefore, if anything, these individuals should be over-represented rather than under-represented in genomic research. As mentioned in the previous section, the lack of African GWAS data can be justified by the difficulty in completely capturing the variation in African genomes using SNP-based genotyping techniques designed for European populations, due to high levels of genetic variation and low levels of LD [1]. However, whole-genome sequencing technologies allow for the almost complete capture of high levels of genomic variation, and low LD may even improve the ability to identify causal variants. Unfortunately, although the sequencing technologies for genomic studies are available and affordable, and Africa provides the required disease cohorts for these studies [14,15], to our knowledge, no published African disease genomes are currently available. In addition, while the highly diverse genome architecture of African individuals appears to be most well suited to benefit from sequence analysis, only nine personal African genomes/exomes, excluding those generated by the 1000 Genomes Project, have been published to date [29–31].

These issues are, however, being addressed and the 1000 Genomes Project is in the process of resequencing 100 individuals from each of the five selected African populations from Nigeria, Kenya, Gambia and Malawi, as well as an additional population from either Sierra Leone or Nigeria [105]. These data should play a role in adding to the knowledge of not only the African genome, but also the human variome. This is important, as examination of the available resequencing data has demonstrated that African genomes consistently exhibit more variation, both novel and known, than non-Africans [29–32]. Unfortunately, with the exception of Kenya and Sierra Leone, the populations that are included in the 1000 Genomes Project originate from the same areas in Africa where the 'GWAS hits' of the past have been obtained (Figure 1) [2–7]. Therefore, once again the majority of Africa remains unaccounted for. However, every sequenced genome will add to our understanding of human variation and will create a stepping stone for future resequencing projects in more African populations. Through the careful implementation of whole-genome sequencing in Africa, taking into consideration the possibility of identifying causal variants contributing to diseases, it is hoped that genomic findings can be translated into the clinical context to improve health. These results are likely to have the biggest impact in Africa where improvement of healthcare is urgently required.

Pharmacogenomics: a front runner for translation of genomic results into the clinical setting?

Although the Human Genome Project was expected to play a role in understanding disease, the most tangible results were in fact related to personalized medicine and the effective treatment of disease after diagnosis [33]. More specifically, pharmacogenetics/genomics was both predicted and has been proven to show promise for clinical applications [34,35]. This is demonstrated by the translation of results from pharmacogenetic studies into the clinical setting [36,37], in combination with the number of US FDA-approved drugs with pharmacogenomic information on their labels [106]. Furthermore, in some cases the cost–effectiveness has been calculated with regard to the utilization of specific pharmacogenetic tests to improve drug efficacy and decrease adverse drug reactions (ADRs) [38–44]. Although, these studies have shown that the pharmacogenetic tests are in some cases very expensive, which is a particular hindrance for implementation in developing countries, they have also shown that pharmacogenetic tests can be cost effective if the appropriate considerations are made and the allele frequencies of the specific variants are high enough in the population being tested [39–43]. Results such as these, providing evidence for cost–benefit outcomes, are important as it has been widely reported that ADRs and treatment failure are responsible for

large economic and healthcare burdens worldwide [45,46,107]. In addition, these results emphasize a need to characterize and understand the variation present in the genes influencing drug response and to determine the role they play in treatment outcomes in different populations.

Unfortunately, studies providing examples of the cost–effectiveness of pharmacogenomics within Africa are missing, which may in part be attributed to a lack of data pertaining to the specific variants affecting treatment outcome. This is of serious consequence as Africa is suspected to be burdened with a high rate of nonoptimal treatment regimens. One of the reasons that Africa is burdened by ADRs and poor drug efficacy is the high prevalence of HIV/AIDS, which requires lifelong treatment with antiretrovirals [108]. Not only are patients more likely to experience ADRs due to the presence of the disease, but the use of concomitant drugs increases the likelihood of drug–drug interactions, which subsequently results in nonoptimal treatment outcomes [45,47]. Furthermore, diseases and the treatment thereof are aggravated by poverty and a lack of resources, and while 10% of the global burden of disease can be attributed to diseases occurring in poor countries, unfortunately only 1% of new drugs are developed for the treatment of these diseases [48]. Therefore, not only are the relevant drugs not developed for Africa, but those drugs that are developed are predominantly designed using European descent individuals as a reference, which is unlikely to be an optimal fit for the highly diverse genomes of African individuals. Therefore, research to determine the presence and effect of variation in pharmacogenes (i.e., pharmacogenetic/genomically relevant genes) needs to be performed in Africa to assess the types and dosages of medication required for optimal treatment.

Validation for the requirement of these studies is provided by previous studies, which have resequenced pharmacogenes in Africans and consistently discovered novel variation that is likely to affect treatment outcome [49–55]. As most African pharmacogenetic studies have been performed focusing on candidate genes [56], there is much that future pharmacogenomic studies in Africa stand to gain from the implementation of whole-genome sequencing.

Critical evaluation of the variation present in pharmacogenes

Before large and expensive projects are designed and initiated, proof of the value of these studies is required. The most cost-effective way of doing this is to review data that is already publicly available. Therefore, as this article is focused predominantly on the application of pharmacogenomic data, we critically examined and compared the variation present in the pharmacogenes of African and non-African individuals. In order to obtain a comprehensive overview of this variation, we utilized publically available high-throughput data derived from genome-wide studies utilizing either large-scale SNP genotyping or whole genome/exome resequencing.

To evaluate the value of whole genome/exome resequencing data, we utilized the Galaxy library and corresponding tutorial to compare the variation observed in the pharmacogenes of 13 ethnically diverse individuals, which were broadly divided into seven Africans and six non-Africans (Supplementary Table 1; www.future-medicine.com/doi/suppl/10.2217/pgs.11.119) [109]. More specifically, we examined the top ten pharmacogenes (*ABCB1*, *CYP1A2*, *CYP2B6*, *CYP2C9*, *CYP2C19*, *CYP2D6*, *CYP3A4*, *TPMT*, *UGT1A1* and *VKORC1*) as described by Thorn *et al.* [57]. Although these pharmacogenes have all been extensively studied, using the data obtained from the Galaxy library for the 13 individuals, we found that half of these genes contained novel/uncharacterized nonsynonymous SNPs that were predicted to be 'damaging' by the sorting intolerant from tolerant (SIFT) algorithm (Table 1) [58]. In the context of this article we defined uncharacterized variants as those variants, which although predicted to alter the respective protein products, have not been described in either the allele nomenclature websites [110,111] or on PharmGKB [112]. The presence of these variants emphasizes the variability of the genes and suggests that more comprehensive genotyping assays may be required in future studies. More specifically, not only was there consistently a higher level of variation present in the pharmacogenes of African individuals when compared with non-Africans, but approximately three-quarters of the novel/uncharacterized variation was detected in African individuals (Table 1). These data provide a strong argument for the resequencing of pharmacogenes, specifically in African descent individuals. The data also provides a potential explanation for the higher frequency of genotype–phenotype discordance observed in individuals with African ancestry, which may be attributed to uncharacterized variation present in pharmacogenes [59].

It is, however, equally important to consider variation that has been characterized

and proven to be of value to pharmacogenetic applications. Therefore, we examined variation, known to affect drug response, in the highest ranked genes with perceived importance for gene–drug interactions, as described by the Clinical Pharmacogenetics Implementation Consortium (Table 2) [60]. We also included *COMT* and *TMPT*–cisplatin to this list of drug–gene interactions, as it has recently been reported to be highly significant for pharmacogenetic applications [61]. To evaluate and compare the presence and frequency of this variation in Africans (Yoruba individuals from Nigeria [YRI]) and non-Africans (northern and western European ancestry individuals from Utah [CEU]), we utilized data from the International HapMap project [113]. Unfortunately, consultation of the HapMap data revealed that genotype information was absent for almost half of the variants listed in Table 2, emphasizing a need for population-based studies focusing specifically on pharmacogenetic variation.

Even with the limited data, clear differences in allele frequencies between the two HapMap populations were observed (Figure 2A). In general, the variants appeared to be more representative of the CEU population, highlighting the disproportionate research that has been performed with regard to the characterization of pharmacogenetic variation. Although not all of the relevant pharmacogenetic variants were detected in the African-descent individuals, these variants have for the most part been well characterized, and are thus likely to be accounted for by already available pharmacogenetic genotyping assays. Of further interest, although many variants were not detected in African-descent individuals, some variants were present at much higher frequencies in these individuals and may therefore have even more value for pharmacogenetic applications in Africa (e.g., *COMT* rs9332377 and *TMPT* rs12201199, both associated with cisplatin-induced hearing loss [61]). The differences in allele frequencies observed between the YRI and CEU can be extended to the other nine HapMap populations, as has been well demonstrated by Adeyemo and Rotimi, who showed that the more removed the HapMap populations were from one another, the more the allele frequencies differed [62]. Together these results serve to support the notion that multiple populations need to be studied in order for us to gain a comprehensive understanding of human variation and its effect on pharmacogenetic traits.

More specifically, even those studies that have taken African variation into account have predominantly used the YRI as a reference for African individuals. This is not always an accurate representation, as demonstrated by Rotimi and Jorde with regard to the differences in allele frequencies observed for the *HLA-B*5701* abacavir hypersensitivity syndrome-associated variant [63], and further highlighted in Figure 2B for *VKORC1* -1639 G>A, which is included in three of the five FDA-approved *in vitro* diagnostic tests for warfarin dosing [55]. Both these variants show that although the allele frequencies were extremely low in the YRI population, the same is not true for other African populations. Therefore, if a particular variant is not of relevance to one African population, it should not be assumed that it will not be relevant for pharmacogenetic applications in another African population. This once again highlights that the handful of African populations that have been included in GWAS and resequencing projects are unlikely to be representative of the entire continent.

In summary, these data demonstrate that available pharmacogenetic information has greater relevance for non-African populations, which may be attributed to the greater amount of research that has been performed, in combination with the lower levels of genetic variation and greater homogeneity of these populations, when compared with Africans. By contrast, not only is there a large proportion of uncharacterized variation present in African genomes, but the value of characterized variation differs between populations. These data all support the resequencing of many, diverse African genomes. Although we have demonstrated that resequencing applications have the greatest value for African populations, as sequencing becomes cheaper and more readily available, the replacement of other less comprehensive genotyping assays with this technology would be beneficial to all population groups.

Table 1. Novel/uncharacterized nonsynonymous SNPs predicted to be damaging, from resequencing data of ten pharmacogenes in 13 ethnically diverse individuals.

Gene	Amino acid changes	
	Africans	***Non-Africans***
CYP1A2		T324I
CYP2C9	L287M	
CYP2C19	L17H	
CYP2D6		R365H
CYP3A4	I118F	
	Q200H	
	L401P	

Table 2. List of important previously characterized variants affecting treatment response.

Gene	Drug	Variant	ADR/treatment outcome
CYP2D6	Tamoxifen	**3* (rs35742686)†	Unfavorable treatment outcome
		**4* (rs3892097)†	
		**5* (gene deletion)†	
		**6* (rs5030655)†	
		**7* (rs5030867)†	
CYP2C19	Clopidogrel	**2* (rs4244285)	Decrease in platelet responsiveness
		**3* (rs4986893)†	Decrease in platelet responsiveness
		**17* (rs12248560)	Improved response
VKORC1	Warfarin	-1639 G>A (rs9923231)	Increased risk of bleeding
		1173 C>T (rs9934438)	Increased risk of bleeding
CYP2C9	Warfarin	**2* (rs1799853)	Increased risk of bleeding
		**3* (rs1057910)	Increased risk of bleeding
HLA-B	Abacavir	**5701* (rs2395029)	Hypersensitivity syndrome
TPMT	Mercaptopurine	**2* (rs1800462)†	Toxicity
		**3* (rs1142345)	Toxicity
		**4* (rs1800584)†	Toxicity
TPMT	Cisplatin	rs12201199	Hearing loss
COMT	Cisplatin	rs9332377	Hearing loss

†*No genotype information available on HapMap.*
ADR: Adverse drug reaction.

Hurdles associated with sequencing technologies

Before large-scale resequencing projects are initiated, it is important that we are realistic about the potential hurdles associated with these new technologies. Although the time and costs associated with the generation of sequence data are steadily decreasing, the production of vast amounts of data offers a whole new range of constraints. A few years ago, the generation of sequence data was considered a rate-limiting step, now the analyses of these data create a variety of bottlenecks of their own. With regard to the cost of sequencing, it is often cheaper to resequence the data than to store it [114]. Thus, it is crucial that the technologies and techniques to store and analyze the generated data are developed at a rate equal to, or faster than, the rate at which sequencing technologies are developing. These additional costs and considerations may prevent the application of resequencing technologies in certain contexts and alternative genotyping strategies may need to be considered based on the resources available to the specific research unit. These genotyping assays may include commercially available pharmacogenetic tests, such as the AmpliChip® CYP450 Test [115] from Roche or the DMET™ chip from Affymetrix [116]. It should, however, be noted that these assays have predominantly been designed according to populations of European descent and may therefore not account for African-specific alleles [52,53]. Although customized assays that are designed according to African populations will address these issues, they will most likely not account for rare variants, which are more common in African individuals [52,53]. Sequencing of candidate genes will provide data for all variants and will eliminate the bottlenecks caused by the vast amounts of data generated by whole-genome sequencing; however, this strategy will also eliminate the unbiased results that can be obtained by examining the entire genome. Therefore, it is essential to consider both the research question and the resources available before implementing a genotyping assay.

On a technical level, next-generation sequencing technologies are far from perfect. However, the ability to detect variation has already improved dramatically since next-generation sequencing was first introduced. Utilizing the resequencing data for the 13 individuals described in Supplementary Table 1, we divided the individuals into eight individuals sequenced prior to 2010 and five individuals sequenced in 2010. Examination of the three *CYP2C19* alleles described in Table 2 showed that the ability to detect these variants differed according to the date that the sequencing data was published (Supplementary Table 1 & Figure 3). This clearly demonstrates the improvement, not only in cost, but also in quality, of next-generation

sequencing technologies (including techniques for the targeted capture of exomes) in a short period of time, the advancement of which is likely to continue. As technologies continue to improve in accuracy and decrease in cost, the likelihood of genome sequencing becoming the genotyping assay of choice in the future will increase.

However, there is still room for improvement, which is demonstrated by examining genotype data for *CYP2D6*, which is arguably one of the most important pharmacogenes [60]. The complex nature of the gene has resulted in both SNP-based and next-generation sequencing technologies having difficulty capturing *CYP2D6* gene variation. This may in part be attributed to the high levels of sequence similarity observed between *CYP* genes and their corresponding pseudogenes [64], and is demonstrated by the lack of data pertaining to *CYP2D6* in the examined databases. In the HapMap database the allele frequencies for only two SNPs are available and neither of these SNPs are of relevance to pharmacogenetic applications [113]. Although genotype data for more SNPs is available in the 'Very Important Pharmacogenes and 1000 Genomes database' [65], this genotype data is far less than that obtained for other less variable genes (e.g., there are 82 *CYP2D6* alleles reported to date, which is more than double the number of alleles reported for the second most variable *CYP* gene present in the Very Important Pharmacogenes genes and 1000 Genomes database).

Furthermore, the complexity of *CYP2D6* makes accurate allele prediction difficult. To demonstrate this, we reviewed the *CYP2D6* alleles reported for Craig Venter and James Watson by Ng *et al.*, to evaluate the potential for error in allele classification [66]. In addition, we included *CYP1A2*, *CYP2C9*, *CYP2C19* and *CYP3A4* in this evaluation of allele classification, as these genes fell into our previous criteria for the top ten pharmacogenes. Out of the ten predicted allele combinations for these five genes, we disagreed with four of the allele combinations reported by Ng and colleagues [66] (Table 3). For *CYP2C9* the discordance is of little consequence as both allele combinations result in extensive metabolizer phenotypes. With regard to *CYP2C19*, however, we predicted that Craig Venter would be a CYP2C19 poor metabolizer (confirmed by the National Center for Biotechnology Information dbSNP database), while Ng and colleagues predicted that he would be an extensive metabolizer [66]. In addition, when looking at James Watson's genome, the presence of *CYP2C19*27*, which was reported on the CYP allele nomenclature website post the Ng *et al.* publication [66,110], serves as a reminder that new data are continually being discovered and should be added to databases on a regular basis. The *CYP2D6* data is, however, the most convincing with regard to the ease with which errors can occur in the allele classification process. Ng and colleagues predicted that James Watson was homozygous for *CYP2D6*10*, which correlates with the resequencing data, such that he was homozygous for the P34S and S486T SNPs, which form the decreased function *CYP2D6*10* allele [66]. However, when examining the data more extensively, we also noticed that there was no genotype information available for the 1846 G>A splicing-defect variant, which along with the P34S and S486T SNPs, form the *CYP2D6*4* null allele – the most frequent nonfunctional allele in Caucasians [59]. When we consulted National

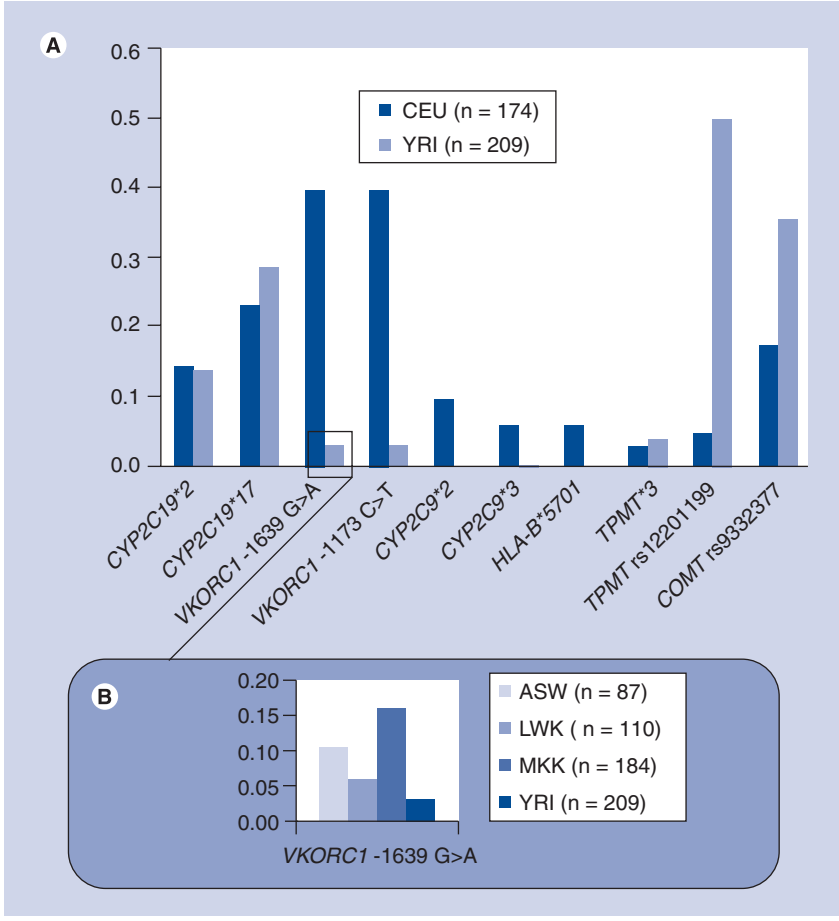

Figure 2. Allele frequency differences between populations for important pharmacogenetic variants. (A) The differences in allele frequencies observed between the CEU and YRI individuals for known pharmacogene variants that have been genotyped by the HapMap project. **(B)** The differences in allele frequencies observed between African ancestry individuals for *VKORC1* -1639 G>A.
ASW: African ancestry in southwest USA; CEU: Utah residents with northern and western European ancestry from the Centre d'Etude du Polymorphisme Humain collection; LWK: Luhya in Webuye, Kenya; MKK: Maasai in Kinyawa, Kenya; YRI: Yoruban in Ibadan, Nigeria.
All data were obtained from the HapMap Genome Browser release #28 [113].

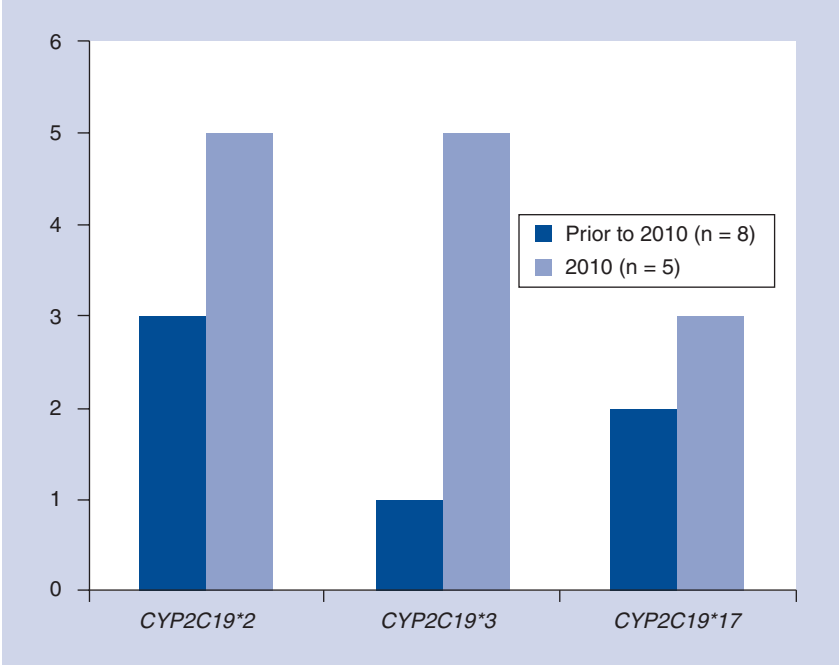

Figure 3. A comparison of the number of individuals for which genotype information was available for the examined SNPs, stratified according to the date that the sequencing data was published. Although the sequencing data published prior to 2010 examines more individuals (eight individuals vs five individuals), genotype data was available for fewer individuals for each of the SNPs. With regards to *CYP2C19*17*, this SNP occurs beyond the regions captured for exome sequencing, thus explaining the inability to detect this variant in the two individuals for whom only exome sequences were published in 2010.

Center for Biotechnology Information's Entrez SNP database [117], we found that James Watson was indeed homozygous for the 1846G>A splicing defect and was thus a poor metabolizer rather than an intermediate metabolizer. Another consideration with regard to *CYP2D6*, is that any individual genotyped as homozygous for an allele, as both Craig Venter and James Watson were, may instead be hemizygous for that allele due to the presence of the *CYP2D6*5* gene-deletion allele. Thus, for the time being, it appears that accurate *CYP2D6* allele prediction will require specially designed genotyping assays which will cover the spectrum of variation of this gene, accounting for the gene's homologous nature and including gene deletions and duplications, as well as hybrid genes. Furthermore, with regard to the assignment of pharmacogene alleles, it appears that allele-naming software, to eliminate potential human error, may be required. It is however important to remember that human interpretation is essential in order to critically examine the data and interpret the findings.

Conclusion

Genomic technologies have developed at a rapid pace. Unfortunately, although the technologies provide opportunities to aid in the improvement of human health, not all populations have benefited equally from these technologies. The vast majority of genomic studies have been performed in European descent populations, followed by Asian populations. Even though the genetically diverse and ancient African populations are most representative of all populations, they have been consistently under-represented in genomic research. Although it is generally accepted that Africa should be included in future studies, even with the advent of next-generation sequencing technologies that are uniquely suited to the complex genomes of Africans, Africa remains neglected. Next-generation sequencing provides us with the much needed opportunity to catalogue human variation on a scale that would have been unimaginable a few years ago. Therefore we should take advantage of these opportunities and include all human populations in global genomic research endeavors.

It is however essential that we remain realistic with regard to what can be achieved both with reference to the technical challenges associated with current sequencing technologies and the translation of results. Sequencing technologies are still developing and there remains much room for improvement. This paper has shown that resequencing technologies are not yet capable of detecting all variation in all gene regions. The complex nature of certain genes, such as *CYP2D6*, are problematic for present-day technologies, which have not yet completely overcome obstacles such as high sequence similarity. In addition, the almost overwhelming amounts of data allows ample opportunity for the incorporation of errors, both human and mechanical. Therefore, it is important that we develop methods to analyze these data in an efficient and reliable manner that can be applied to the clinical setting.

Furthermore, if sequencing projects are to be implemented within Africa, with its lack of resources and infrastructure, there are several improvements that need to be made. These improvements include the development of appropriate facilities, collaborations and educational programs, all of which require a large amount of financial support. It is for this reason that the Human, Hereditary and Health in Africa initiative has been proposed. This initiative, focusing on genomic research in Africa, aims to build resources and infrastructure within the continent, and in so doing contribute to the education and training of African scientists [107]. However, additional funding from organizations such as the government and private sectors is required. In order to convince these organizations of the value of the proposed

research, proof of the clinical utility of genomic studies should be provided. Although, it may be argued that the translation of results into the clinical context has been slow, there have been a few good examples with regard to pharmacogenomics, which is highlighted by the fact that this field has been included as one of the suggested research avenues for the Human, Hereditary and Health in Africa initiative [107]. Thus, pharmacogenomics may be a good starting point for the utilization of whole-genome resequencing for implementation in the clinic.

The question remains whether, at present, personalized medicine can be applied to every individual, all around the world. The short answer to this is 'no'. At present, this is not feasible, especially in developing countries. First, there is not enough evidence to convince medical practitioners of the value of personalized medicine. Second, there is not enough information pertaining to the variation present in poorly characterized populations to confidently prioritize variants in pharmacogenes for genetic tests, the costs of which remain too expensive for routine use. Furthermore, sequencing technology still needs to develop, both with regard to quality and cost, to the point where it can be used in the clinical context. Sequencing technology is, however, in its infancy and if we look at the variation that has already been detected in the genomes of only five resequenced southern African individuals [31], it is clear that there remains much to be discovered from the analyses of African genomes.

Future perspective

In the future, whole-genome sequencing is likely to be applied more regularly in research. In 5–10 years some of the main hurdles associated with present-day sequencing will have been overcome. As the technology improves, the quality will improve and with this, the reliability of variant detection. With regard to analysis of the data, optimized pipelines to suit the needs of the specific research project will be available, and computational training will be incorporated into biological degrees. Institutes making use of multidisciplinary collaboration will become commonplace and large research networks, allowing for the creation of large cohorts, will be the norm. Furthermore, as the quality and reliability of next-generation sequencing technologies improve and the costs decrease, these assays may replace the traditional SNP-based genotyping assays and will be of more value for detecting uncharacterized and rare variation, which has particular relevance for the highly diverse genomes of African individuals.

Table 3. Allele combination comparisons utilizing data from our analyses to compare to that of Ng *et al*.

Gene	Craig Venter		James Watson	
	Ng* et al. *alleles	***Our alleles***	***Ng* et al. *alleles***	***Our alleles***
CYP1A2	**1F/*1F*	**1F/*1F*	**1F/*1F*	**1F/*1F*
CYP2C9	**1A/*1B*	**1A/*1B*	****1A/*1A***	****1A/*1C***
CYP2C19	****1B/*1B***	****2/*2***	****1B/*1B***	****1B/*27***[†]
CYP2D6	**1A/*1A*	**1A/*1A*	****10/*10***	****4/*4***
CYP3A4	**1A/*1A*	**1A/*1A*	**1A/*1B*[†]	**1A/*1B*[†]

Alleles in bold are the combinations that do not correlate.
[†]The genotype given by the National Center for Biotechnology Information SNP database does not correlate with the given genotype, highlighting inconsistencies between different databases.
Data taken from [66].

However, although sequencing technologies may become more commonplace, it is important to remember that their use in research will depend largely on the context to which they are applied. Even though the associated sequencing costs are likely to continue to decrease, the costs and resources required to store the vast amounts of data may still limit the application of these technologies. Therefore before applying sequencing technologies, it will be necessary to carefully consider whether the large amounts of data that will be obtained are necessary, or reliable enough, to answer the research question of interest.

With regard to the sequencing data that is generated in the future, there are three main areas that we feel will be positively affected: databases, the understanding of the genome and complex disorders, and the incorporation of genomic data into the clinical setting.

An increase in data will result in an increase in what is known regarding variation and as this data is added to the growing databases, it will become easier to correlate this variation to a specific phenotype or disease. More comprehensive databases will allow for more comprehensive studies. At present we only understand approximately 1% of the human genome's functions. By sequencing the whole genome and identifying variation that affects specific functions, we can learn more about noncoding DNA and the role it plays. The increase in understanding with regard to the genome and the complex interactions that exist within this system, in combination with the ability to detect all variation, including rare and novel variation, will give us insight into complex disorders. Lastly, as whole-genome sequencing becomes more reliable and easier to implement, we will reach a point where the whole-genome sequence of an individual will become part of the medically relevant information of a patient and will be used to guide, among other things, treatment regimens.

Executive summary

Inequalities in genomic research

- Approximately 75% of genome-wide association studies (GWAS) have been performed in European-descent individuals.
- Of the approximately 8% of GWAS performed in African-descent individuals, only a very small portion have been performed in specific populations from Africa.
- The handful of African GWAS that have been performed have focused on the African epidemics, such as tuberculosis, HIV and malaria, thereby neglecting other disorders.

Next-generation sequencing technologies

- Sequencing is one of the most rapidly advancing technologies.
- Even in the short time that whole-genome sequencing has been available, Africans have been under-represented.
- African genomes are likely to provide the most comprehensive catalogue of human variation.
- The 1000 Genomes Project is aiding in the representation of African populations; however, more studies are required.

Pharmacogenomics: a front runner for translation of genomic results into the clinical setting?

- Pharmacogenomics has provided the most clinically applicable results.
- Africa has a high rate of adverse drug reactions; therefore, pharmacogenomic applications are urgently required.
- Detection of variation in pharmacogenes is important for the successful application of pharmacogenetics.

Critical evaluation of the variation present in pharmacogenes

- There is a substantial amount of novel/uncharacterized variation, predicted to alter the protein product of pharmacogenes, specifically in Africans.
- Well-characterized pharmacogenetic variation is more applicable to European-descent individuals.
- The frequencies and relevance of well-characterized pharmacogenetic variation differ between African populations.

Hurdles associated with sequencing technologies

- The downstream analysis of sequencing data requires optimization.
- The quality of sequencing data is rapidly improving.
- Variation in complex gene regions is not easily detected.
- Large amounts of data allow for the easy incorporation of errors; therefore, comprehensive pipelines for comprehensive analysis need to be developed to eliminate these errors.

Conclusion

- As sequencing technologies improve, they are likely to become the genotyping assay of choice for many studies.

Financial & competing interests disclosure

The authors would like to acknowledge the South African National Research Foundation and Stellenbosch University for financial assistance. 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.

References

Papers of special note have been highlighted as:
- of interest
- of considerable interest

1 Rosenberg NA, Huang L, Jewett EM, Szpiech ZA, Jankovic I, Boehnke M. Genome-wide association studies in diverse populations. *Nat. Rev. Genet.* 11(5), 356–366 (2010).

- **Reviews the past genome-wide association studies and associated challenges.**

2 Adeyemo A, Gerry N, Chen G *et al.* A genome-wide association study of hypertension and blood pressure in African Americans. *PLoS Genet.* 5(7), e1000564 (2009).

3 Jallow M, Teo YY, Small KS *et al.* Genome-wide and fine-resolution association analysis of malaria in West Africa. *Nat. Genet.* 41(6), 657–665 (2009).

4 Xiong DH, Liu XG, Guo YF *et al.* Genome-wide association and follow-up replication studies identified *ADAMTS18* and *TGFBR3* as bone mass candidate genes in different ethnic groups. *Am. J. Hum. Genet.* 84(3), 388–398 (2009).

5 Kang SJ, Chiang CW, Palmer CD *et al.* Genome-wide association of anthropometric traits in African- and African-derived populations. *Hum. Mol. Genet.* 19(13), 2725–2738 (2010).

6 Thye T, Vannberg FO, Wong SH *et al.* Genome-wide association analyses identifies a susceptibility locus for tuberculosis on chromosome 18q11.2. *Nat. Genet.* 42(9), 739–741 (2010).

7 Petrovski S, Fellay J, Shianna KV *et al.* Common human genetic variants and HIV-1 susceptibility. A genome-wide survey in a homogeneous African population. *AIDS* 25(4), 513–518 (2011).

8 Sirugo G, Hennig BJ, Adeyemo AA *et al.* Genetic studies of African populations. An overview on disease susceptibility and response to vaccines and therapeutics. *Hum. Genet.* 123(6), 557–598 (2008).

9 Tishkoff SA, Reed FA, Friedlaender FR *et al.* The genetic structure and history of Africans and African Americans. *Science* 324, 1035–1044 (2009).

10 Bryc K, Auton A, Nelson MR *et al.* Genome-wide patterns of population structure and admixture in West Africans and African Americans. *Proc. Natl Acad. Sci. USA* 107(2), 786–791 (2010).

11 Need AC, Goldstein DB. Next generation disparities in human genomics. concerns and remedies. *Trends Genet.* 25(11), 489–494 (2009).

12 Musunuru K, Strong A, Frank-Kamenetsky M *et al.* From noncoding variant to phenotype via SORT1 at the 1p13 cholesterol locus. *Nature* 466(7307), 714–719 (2010).

13 Khoury MJ, Gwinn M, Yoon PW, Dowling N, Moore CA, Bradley L. The continuum of translation research in genomic medicine. How can we accelerate the appropriate integration of human genome discoveries into health care and disease prevention? *Genet. Med.* 9(10), 665–674 (2007).

14 Lopez AD, Mathers CD, Ezzati M, Jamison DT, Murray CJ. Global and regional burden of disease and risk factors, 2001. systematic analysis of population health data. *Lancet* 367(9524), 1747–1757 (2006).

15 Ezeh AC, Izugbara CO, Kabiru CW *et al.* Building capacity for public and population health research in Africa. the consortium for advanced research training in Africa (CARTA) model. *Glob. Health Action* 3, doi:10.3402/gha.v3i0.5693 (2010).

16 Dye C, Harries AD, Maher D *et al.* Chapter 13: Tuberculosis. In: *Disease and mortality in sub-Saharan Africa (2nd Edition).* Jamison DT, Feachem RG, Makgoba MW *et al.* (Eds). The World Bank, Washington, DC, USA (2006).

17 Mboup S, Musonda R, Mhalu F, Essex M. Chapter 17: HIV/AIDS. In: *Disease and mortality in sub-Saharan Africa. (2nd Edition).* Jamison DT, Feachem RG, Makgoba MW *et al.* (Eds). The World Bank, Washington, DC, USA (2006).

18 Snow RW, Omumbo JA. Chapter 14: Malaria. In: *Disease and Mortality in sub-Saharan Africa. (2nd Edition).* Jamison DT, Feachem RG, Makgoba MW *et al.* (Eds). The World Bank, Washington, DC, USA (2006).

19 Kebede S, Duales S, Yokouide A, Alemu W. Trends of major disease outbreaks in the African region, 2003–2007. *East Afr. J. Public Health* 7(1), 20–29 (2010).

20 Fiander AN. The prevention of cervical cancer in Africa. *Womens Health (London).* 7(1), 121–132 (2011).

21 Maher D, Sekajugo J. Research on health transition in Africa. time for action. *Health Res. Policy Syst.* 9(1), 5 (2011).

22 Zhang J, Chiodini R, Badr A, Zhang G. The impact of next generation sequencing on genomics. *J. Genet. Genomics* 38(3), 95–109 (2011).

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

24 Wheeler DA, Srinivasan M, Egholm M *et al.* The complete genome of an individual by massively parallel DNA sequencing. *Nature* 452, 872–876 (2008).

25 Nair P. QnAs with Eric S Lander. *Proc. Natl Acad. Sci. USA* 108(28), 11319 (2011).

26 Human genome: genomes by the thousand. *Nature* 467(7319), 1026–1027 (2010).

27 Voight BF, Adams AM, Frisse LA, Qian Y, Hudson RR, Di Rienzo A. Interrogating multiple aspects of variation in a full resequencing data set to infer human population size changes. *Proc. Natl Acad. Sci. USA* 102(51), 18508–18513 (2005).

28 Campbell MC, Tishkoff SA. African genetic diversity. Implications for human demographic history, modern human origins, and complex disease mapping. *Annu. Rev. Genomics Hum. Genet.* 9, 403–433 (2008).

29 Bentley DR, Balasubramanian S, Swerdlow HP *et al.* Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* 456(7218), 53–59 (2008).

30 Kidd JM, Cooper GM, Donahue WF *et al.* Mapping and sequencing of structural variation from eight human genomes. *Nature* 453(7191), 56–64 (2008).

31 Schuster SC, Miller W, Ratan A *et al.* Complete Khoisan and Bantu genomes from southern Africa. *Nature* 463(7283), 943–947 (2010).

▪ **A reminder of the amount of variation that is likely to be detected through the resequencing of African individuals.**

32 1000 Genomes Project Consortium. A map of human genome variation from population-scale sequencing. *Nature* 467(7319), 1061–1073 (2010).

▪▪ **Describes the value of the resequencing data obtained from the 1000 Genomes Project.**

33 Caplan A. Genome-sequencing anniversary. What will drive genomics over the next 10 years? *Science* 331(6018), 691 (2011).

34 Ginsburg GS, Konstance RP, Allsbrook JS, Schulman KA. Implications of pharmacogenomics for drug development and clinical practice. *Arch. Intern. Med.* 165(20), 2331–1336 (2005).

35 Deverka PA, Vernon J, McLeod HL. Economic opportunities and challenges for pharmacogenomics. *Annu. Rev. Pharmacol. Toxicol.* 50, 423–437 (2010).

36 Swen JJ, Wilting I, de Goede AL *et al.* Pharmacogenetics: from bench to byte. *Clin. Pharmacol. Ther.* 83(5), 781–787 (2008).

37 Swen JJ, Nijenhuis M, de Boer A *et al.* Pharmacogenetics. From bench to byte – an update of guidelines. *Clin. Pharmacol. Ther.* 89(5), 662–673 (2011).

38 Costa-Scharplatz M, van Asselt AD, Bachmann LM, Kessels AG, Severens JL. Cost–effectiveness of pharmacogenetic testing to predict treatment response to angiotensin-converting enzyme inhibitor. *Pharmacogenet. Genomics* 17(5), 359–368 (2007).

39 Schackman BR, Scott CA, Walensky RP, Losina E, Freedberg KA, Sax PE. The cost–effectiveness of *HLA-B*5701* genetic screening to guide initial antiretroviral therapy for HIV. *AIDS* 22(15), 2025–2033 (2008).

40 Eckman MH, Rosand J, Greenberg SM, Gage BF. Cost–effectiveness of using pharmacogenetic information in warfarin dosing for patients with nonvalvular atrial fibrillation. *Ann. Intern. Med.* 150(2), 73–83 (2009).

41 Gold HT, Hall MJ, Blinder V, Schackman BR. Cost effectiveness of pharmacogenetic testing for uridine diphosphate glucuronosyltransferase 1A1 before irinotecan administration for metastatic colorectal cancer. *Cancer* 115(17), 3858–3867 (2009).

42 Leey JA, McCabe S, Koch JA, Miles TP. Cost–effectiveness of genotype-guided warfarin therapy for anticoagulation in elderly patients with atrial fibrillation. *Am. J. Geriatr. Pharmacother.* 7(4), 197–203 (2009).

43 Hagaman JT, Kinder BW, Eckman MH. Thiopurine *S*-methyltranferase testing in idiopathic pulmonary fibrosis. A pharmacogenetic cost–effectiveness analysis. *Lung* 188(2), 125–132 (2010).

44 Vegter S, Jansen E, Postma MJ, Boersma C. Economic evaluations of pharmacogenetic and genomic screening programs. Update of the literature. *Drug Dev. Res.* 71, 492–501 (2010).

45 Mehta U, Durrheim DN, Blockman M, Kredo T, Gounden R, Barnes KI. Adverse drug reactions in adult medical inpatients in a south African hospital serving a community with a high HIV/AIDS prevalence. prospective observational study. *Br. J. Clin. Pharmacol.* 65(3), 396–406 (2008).

46 Wester K, Jönsson AK, Spigset O, Druid H, Hägg S. Incidence of fatal adverse drug reactions. a population based study. *Br. J. Clin. Pharmacol.* 65(4), 573–579 (2008).

47 Jones AE, Brown KC, Werner RE *et al.* Variability in drug metabolizing enzyme activity in HIV-infected patients. *Eur. J. Clin. Pharmacol.* 66(5), 475–485 (2010).

48 Chirac P, Torreele E. Global framework on essential health R&D. *Lancet* 367(9522), 1560–1561 (2006).

49 Dandara C, Masimirembwa CM, Magimba A *et al.* Arylamine *N*-acetyltransferase (NAT2) genotypes in Africans. The identification of a new allele with nucleotide changes 481C>T and 590G>A. *Pharmacogenetics* 13(1), 55–58 (2003).

50 Wang J, Sönnerborg A, Rane A *et al.* Identification of a novel specific *CYP2B6* allele in Africans causing impaired metabolism of the HIV drug efavirenz. *Pharmacogenet. Genomics* 16(3), 191–198 (2006).

51 Gaedigk A, Coetsee C. The *CYP2D6* gene locus in south African Coloureds. Unique allele distributions, novel alleles and gene arrangements. *Eur. J. Clin. Pharmacol.* 64(5), 465–475 (2008).

52 Matimba A, Del-Favero J, van Broeckhoven C, Masimirembwa C. Novel variants of major drug-metabolising enzyme genes in diverse African populations and their predicted functional effects. *Hum. Genomics* 3(2), 169–190 (2009).

53 Drögemöller BI, Wright GE, Niehaus DJ *et al.* Characterization of the genetic profile of *CYP2C19* in two south African populations. *Pharmacogenomics* 11(8), 1095–1103 (2010).

54 Wright GE, Niehaus DJ, Drögemöller BI, Koen L, Gaedigk A, Warnich L. Elucidation of *CYP2D6* genetic diversity in a unique African population. implications for the future application of pharmacogenetics in the Xhosa population. *Ann. Hum. Genet*. 74(4), 340–350 (2010).

55 Mitchell C, Gregerson N, Krause A. Novel *CYP2C9* and *VKORC1* gene variants associated with warfarin dosage variability in the south African black population. *Pharmacogenomics* 12(7), 953–963 (2011).

56 Warnich L, Drögemöller BI, Pepper MS, Dandara C, Wright GEB. Pharmacogenomic research in south Africa. Lessons learned and future opportunities in the rainbow nation. *Curr. Pharmacogenomics Person. Med.* 9(3), 191–207 (2011).

57 Thorn CF, Klein TE, Altman RB. Pharmacogenomics and bioinformatics. PharmGKB. *Pharmacogenomics* 11(4), 501–505 (2010).

58 Ng PC, Henikoff S. SIFT. Predicting amino acid changes that affect protein function. *Nucleic Acids Res*. 31(13), 3812–3814 (2003).

59 Gaedigk A, Simon SD, Pearce RE, Bradford LD, Kennedy MJ, Leeder JS. The CYP2D6 activity score. translating genotype information into a qualitative measure of phenotype. *Clin. Pharmacol. Ther*. 83(2), 234–242 (2008).

60 Relling MV, Klein TE. CPIC. Clinical Pharmacogenetics Implementation Consortium of the Pharmacogenomics Research Network. *Clin. Pharmacol. Ther*. 89(3), 464–467 (2011).

61 Ross CJ, Katzov-Eckert H, Dubé MP *et al.* Genetic variants in *TPMT* and *COMT* are associated with hearing loss in children receiving cisplatin chemotherapy. *Nat. Genet.* 41(12), 1345–1349 (2009).

62 Adeyemo A, Rotimi C. Genetic variants associated with complex human diseases show wide variation across multiple populations. *Public Health Genomics* 13(2), 72–79 (2010).

▪ **Represents the value of studying multiple different populations in order to obtain a better understanding of human disease.**

63 Rotimi CN, Jorde LB. Ancestry and disease in the age of genomic medicine. *N. Engl. J. Med*. 363(16), 1551–1558 (2010).

▪▪ **Reviews human variation and the corresponding considerations.**

64 Meijerman I, Sanderson LM, Smits PH, Beijnen JH, Schellens JH. Pharmacogenetic screening of the gene deletion and duplications of *CYP2D6*. *Drug Metab. Rev*. 39(1), 45–60 (2007).

65 Gamazon ER, Zhang W, Huang RS, Dolan ME, Cox NJ. A pharmacogene database enhanced by the 1000 Genomes Project. *Pharmacogenet. Genomics* 19(10), 829–832 (2009).

66 Ng PC, Zhao Q, Levy S, Strausberg RL, Venter JC. Individual genomes instead of race for personalized medicine. *Clin. Pharmacol. Ther*. 84(3), 306–309 (2008).

▪▪ **Perspective on individualized genomes and their value for personalized medicine.**

Websites

101 HuGE Navigator (version 2.0). www.hugenavigator.net/HuGENavigator/gWAHitStartPage.do

102 Ethnologue. www.ethnologue.com/ethno_docs/distribution.asp?by=area

103 WHO. Age standardized death rates (per 100,000 population) (2004). http://gamapserver.who.int/gho/interactive_charts/mbd/as_death_rates/atlas.html

104 Psychiatric GWAS consortium. https://pgc.unc.edu/index.php

105 1000 Genomes Project. www.1000genomes.org/about#ProjectSamples

106 US FDA. Table of Pharmacogenomic Biomarkers in Drug Labels. www.fda.gov/Drugs/ScienceResearch/ResearchAreas/Pharmacogenetics/ucm083378.htm

107 Abdelhak S, Adebamowo C, Adeyemo A *et al.* Harnessing genomic technologies toward improving health in Africa. Opportunities and challenges. http://h3africa.org/h3africa_whitepaper.pdf'

108 International HIV/AIDS alliance. www.hacccambodia.org/store_files/book_library/AntiretraviralARVtreatmentfactsheetARVtreatment.pdf

109 Galaxy data libraries. http://main.g2.bx.psu.edu/library

▪ **Easy to use tool to analyze publicly available sequencing data.**

110 Human cytochrome P450 (CYP) allele nomenclature committee. www.cypalleles.ki.se

111 *UGT1A1* SNPs table. www.pharmacogenomics.pha.ulaval.ca/webdav/site/pharmacogenomics/shared/Nomenclature/UGT1A/UGT1A1.htm'

112 PharmGKB. www.pharmgkb.org

113 International HapMap project. http://hapmap.ncbi.nlm.nih.gov/cgi-perl/gbrowse/hapmap28_B36

114 What to do with all that data? www.genomeweb.com/blog/what-do-all-data

115 Roche AmpliChip® CYP450 Test. www.roche.com/products/product-details.htm?type=product&id=17

116 Affymetrix DMET Plus Premier Pack. www.affymetrix.com/browse/products.jsp?productId=131412&navMode=34000&navAction=jump&aId=productsNav#1_1

117 NCBI Entrez SNP. www.ncbi.nlm.nih.gov/snp