

LETTER

Establishment of a biobank and pharmacogenetics database of African populations

European Journal of Human Genetics (2008) 16, 780–783;
doi:10.1038/ejhg.2008.49; published online 2 April 2008

Pharmacogenetics and genomics research has experienced great advances over the past decade as witnessed by the completion of the human genome in 2003 (www.genome.gov/HGP). The field has been driven by the belief that understanding the human genome, that of pathogens, and interindividual genetic variability would result in radical advances in medicine. Anticipated measurable end points include increased targets for which drug discovery campaigns could be initiated, increased understanding of human susceptibility to disease and variability in drug response hence development of diagnostic tools to realize individualized treatment where drugs would be given to patients in whom they are predicted to work and at doses predicted to be safe.¹ Toward the realization of this biomedical paradigm, Biobanking and Pharmacogenetics Databasing have become well established in developed countries (www.biobanks.se, www.icelandbio.com, www.ukbiobank.ac.uk). Little has, however, been done in developing countries.² Starting with a workshop organized by the African Institute of Biomedical Science and Technology (www.aibst.com) in 2003 on Pharmacogenetics of Drug Metabolism in Nairobi, Kenya, a number of African scientists initiated a consortium for the biobanking and pharmacogenetics databasing of African populations. We here report the results of the first phase of this initiative that has seen research groups from five different African countries with collaborative support from leading experts in Europe and America establish a biobank of blood and DNA from nine ethnic groups from across the African continent. The biobank of anonymous samples has been used to establish baseline frequency distribution of SNPs of genes important in drug metabolism, hence the initiation of a pharmacogenetics database (<http://www.aibst.com/biobank.html>).

Ethical approval for the study was obtained from each of the countries from which samples were collected. Blood

samples were collected from 50 to 100 adult volunteers from ethnic groups in Nigeria, Kenya, Tanzania, Zimbabwe and South Africa (Figure 1). Portions of each blood sample were used to prepare DNA, blot on filter paper or store at -80°C . The biobank consists of 1488 DNA samples from nine ethnic groups (Yoruba, Hausa, Ibo, Luo, Kikuyu, Maasai, Shona, San and Venda). The utility of the biobank was illustrated by studying the frequency distribution of some polymorphisms of known phenotype characteristics (www.imm.ki.se/CYPalleles, <http://louisville.edu/medschool/pharmacology/Human.NAT2.pdf>) of six genes (Table 1) important in drug metabolism (CYP2B6, 2C19, 2D6, GSTM, GSTT and NAT-2). Genotyping was carried out using established PCR–RFLP methods.^{4–10} A database cataloging the samples and the genotype results was designed using Microsoft Access and Visual Basics software packages (<http://www.aibst.com/biobank.html>) and access is currently limited to authorized individuals involved in the research projects.

The stratification of populations based on allele frequency data was evaluated using principal component analysis using the program SIMCA P+ (www.umetrics.com).

The frequency distribution of the polymorphisms (Table 1) analyzed stratified the major populations, Caucasian (European and North American), Oriental (Asian) and African into distinct clusters (Supplementary Figure 2). This is in agreement with current evolutionary understanding of the stratification of these major populations based on other genetic markers.^{11,12} We further analyzed the data for possible stratification of the African ethnic groups alone. No distinct differentiation was observed. Studies using other genetic markers have demonstrated that there is great genetic diversity among African populations compared to Caucasian or Oriental populations.¹² These preliminary findings from our studies could indicate that the number and/or type of genes and/or SNPs analyzed do not carry enough resolution power to capture the genetic diversity of African populations reported in other studies.

The current biobank represents a significant contribution to efforts to jump-start pharmacogenetic and genomics research in African populations in that it contains samples from the most diverse representation of African populations to date from north-west Africa, the Hausa, Yoruba and Ibo, from central-eastern Africa, the Kikuyu, Luo and Maasai and from southern Africa, the Shona, Venda and the San. These regional groups represent some of the major known ethnicities based on linguistic classifications generally used to identify the populations (<http://www.ethnologue.com>). This work complements biobanking projects on samples from West Africa

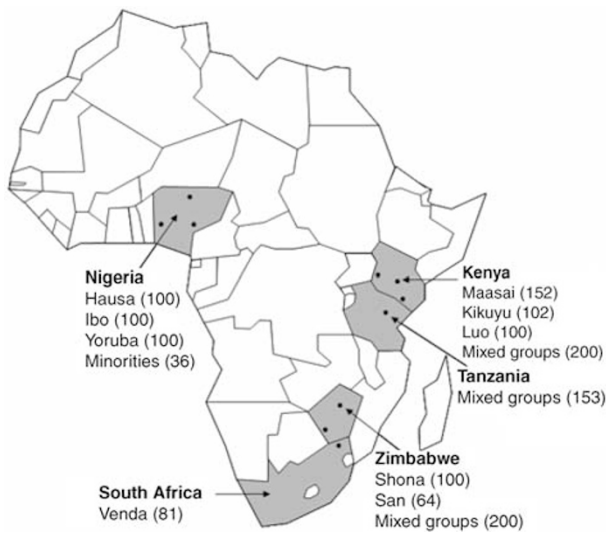


Figure 1 Samples currently in the biobank. The number of available DNA samples from each group is shown. Black dots show location of the ethnic groups where the samples were collected. The Hausa are found mostly in the northern part of Nigeria, the Ibo in the east and the Yoruba on the west. The Maasai of Kenya are located mostly toward the border of Tanzania, the Luo in the west and the Kikuyu in south. The Venda are located at the northeast border of South Africa and Zimbabwe. The Shona are the major ethnic group in Zimbabwe and the San reside on the border of Zimbabwe and Botswana.

(Nigeria, Cameroon and Gambia)^{13–15} and African Americans.¹⁶ It also adds to the increasing international biomedical resources for genetics research (<http://www.p3gconsortium.org>, www.pharmgkb.org, <http://hgvsbase.cgb.ki.se>, www.hapmap.org). In the execution of these studies, we opted for the anonymization of the samples as a consensus document is being drafted on how to handle ethical issues in an area where most African ethics review boards do not have clear guidelines and recommendations. However, there is a limitation of not being able to trace back to the individual should any interesting genetic findings be found in a particular sample. In sample collection, challenges were faced when assigning ethnicity that can be complex due to interethnic marriages, therefore in this study ethnicity was assigned based on the submission that parents and grandparents of the volunteers were of the same self-identified ethnic group.

Extrapolations of possible clinical implications of the baseline frequency distribution of some of the alleles based on established phenotypic characteristics could guide doctors in drug prescription decision-making and/or provide explanations of ethnic-specific adverse effects in metropolitan medical practice. The high frequency of individuals who are homozygous for the CYP2B6*6 allele (18–25%) in African populations is predictive of reduced capacity to metabolize and dispose efavirenz compared to

Table 1 Allele frequencies in the African populations in this study and other ethnicities or populations

Population	CYP2C19			CYP2D6					NAT2				GST		CYP2B6		
	*2	*3	*2/2	*3	*4	*5	*10	*9	*17	*29	*5	*6	*7	*14		M1 del/del	T1 del/del
Oriental	30	10	2	0	1	6	51	0	0	0	5	25	13	0	55	65	18
Chinese	37	8	1	0	1	6	51	0	0	0	6	31	16	0	58	53	21
Japanese	35	11	1	0	1	3	43	0	0	0	2	19	10	0	44	44	16
Koreans	21	12	0	0	2	6	51	0	0	0	3	19	11	0	53	60	15
Caucasian	15	0	5	2	25	5	2	2	0	0	49	27	2	0	50	15	21
Swedes	17	0	1	3	23	5	1	0	0	0	51	28	2	0	51	20	20
Germans	18	0	2	2	20	2	2	0	0	0	46	27	4	0	51	21	21
American	14	0	2					0			45	28	2	0	54	15	
Mixed African	16	1	2	<1 ^a	2	4	6	0	30	15	34	20	5	13	30		40
African American	25	0	1	<1 ^a	7	6	4	1	15	5	30	22	2	9	28	24	47
Tanzanian	18	<1 ^a	3	0	2	4	4	0	18	20	34	21	3	13	33	25	39
Shona	13	<1 ^b	2	0	2	4	6	0	34	17	31	21	6	14	24	26	38
Venda	21	0		0	3	5	12	0	24	6	39	22	5	11	23	20	36
Ghanaian			2	0	7	6	3	0	28						39		49
Ethiopians	14	2	15	0	4	3	9	0	9								
Kikuyu	16	0		0	1			0	33	14	58	24			28	25	34
Luo	18	0		0	4		6	0	23	16	34	22	3	14	29	22	37
Maasai	11	<1 ^a		0	8		5	0	18	8	42	27	4	9	16	40	35
Igbo	29	0		0	8		10	0	14	20	28	29	4	11	23	36	38
Yoruba	10	0		0	3		7	0	22	10	33	27	3	8	31	35	42
Hausa	12	0		0	2		13	0	18	10	27	33	3	3	37	42	42
San	12				9			0	22	2	20	8			45		40

Abbreviation: glutathione-S-transferase.

The numbers in bold are frequencies obtained in this study, the nonbold are literature data compilations.³ Fifteen SNPs/polymorphisms were analyzed and compared to literature data. The CYP2D6 polymorphisms (*5 and *2/2) were not analyzed in this study, but were included in the stratification analysis. The exact description of the CYP alleles can be found on www.imm.ki.se/CYPalleles.

^aThe frequency of the allele is rare (>1%), a few individuals have been described but constitute a rare occurrence.

Where frequency is 0, so far the allele has not been found in the individuals studied in the respective population.

Caucasians where lower genotype frequencies of 5–10% have been reported.^{17,18} This is in agreement with clinical observations in which Africans have been reported to have significantly higher plasma concentrations of the drug compared to Caucasians when given at the standard 600 mg per day doses.¹⁹ Ongoing studies in our laboratory in HIV/AIDS patients taking the anti-HIV drug, efavirenz, indicate the need to lower the dose of this drug in people of African origin homozygous for the *CYP2B6*6* allele (Nyakutira *et al*, unpublished). This could go a long way in reducing adverse effects hence increase treatment compliance. This could also reduce the costs for individuals requiring lower doses and thus partly contributes to the diagnostic costs of identifying such individuals. For the African specific *CYP2D6*17* variant²⁰ similar clinical effects could be explored in the use of antipsychotic drugs^{21,22} in African populations based on the observed high frequency of the variant (14–34%) across the major African populations reported in this study (Table 1). Though the current pharmacogenetics database carries no direct phenotype information linked to carriers of the variants, clinical extrapolations from well-established *in vivo* effects of these polymorphisms give an important baseline that can be the basis for population-specific clinical trial design for optimal use of some medicines.

The use of the biobank, expatriation of samples and intellectual property issues can be both sensitive and controversial. While the steering committee of the consortium is working on a guideline document in consultation with ethics review boards of relevant countries, for samples collected in this study, a working position has been taken to the effect that no samples will be permanently be expatriated from African Institute of Biomedical Science and Technology (AiBST, the project's technical headquarters) or national laboratories and that, should samples be temporarily be expatriated for specialized analysis, approval must be sort from the steering committee. To encourage North–South and South–South collaboration and maximal utilization of the biobank, scientists are welcome to visit the AiBST laboratories to carry out studies on the samples on research that will have been approved by the steering committee. This is hoped to encourage both scientific and technology transfer involving visits by international scientists interested in genetic studies in African populations as they will have to come and carry out their research work at AiBST or laboratories of consortium members hence share their skills and possibly invest in equipment.

The future for the Biobank and Pharmacogenetics Databasing project of African populations will aim to collect samples driven by hypotheses based on phenotypes of interest to health-care challenges in Africa such as malaria, TB, HIV/AIDS infections and side effects to anti-infectives. Such projects with nonanonymous sampling will require rigorous ethical considerations that are being

explored by the steering committee but are necessary as they yield the most informative genotype–phenotype results. Toward increasing the number of samples in the biobank, there will be a move from blood sampling to buccal swabs, which is faster, poses less risk for infection to collectors and easy to transport. This effort will hopefully spur pharmacogenetics and genomics research capacity strengthening at African institutions through investment in molecular biology platforms in the form of DNA automatic sequencers, real-time PCR thermocyclers, bioinformatics tools and the training of biomedical scientists.

Acknowledgements

Alice Matimba was a recipient of the Flemish International Council scholarship. We acknowledge the technical assistance of Chancemore Matonhodze. The extensive technical and scientific assistance provided by Dr Andrea Gaedigk is greatly acknowledged. We also acknowledge the advisory role played by Professor Leif Bertilsson throughout this study.

Alice Matimba¹, Margaret N Oluka², Benjamin U Ebeshi³, Jane Sayi⁴, Oluseye O Bolaji³, Anastasia N Guantai² and Collen M Masimirembwa*¹

¹African Institute of Biomedical Science and Technology, Harare, Zimbabwe;

²Department of Pharmacology and Pharmacognosy, School of Pharmacy, AiBST–UON Lab, University of Nairobi, Nairobi, Kenya;

³Department of Pharmaceutical Chemistry, Obafemi Awolowo University, Ile-Ife, Nigeria;

⁴Department of Pharmaceutical Chemistry, Muhimbili University, Dar-es-Salam, Tanzania

*Correspondence: Dr CM Masimirembwa, African Institute of Biomedical Science and Technology, P O Box 2294, Harare, Zimbabwe.

Tel: +263 4 710563-4; Fax: +263 4 710562;

E-mail: collen.masimirembwa@aibst.com or collenmasimirembwa@yahoo.com

References

- 1 Evans WE: Pharmacogenomics: marshalling the human genome to individualise drug therapy. *Gut* 2003; **52** (Suppl 2): ii10–ii18.
- 2 Sgaier SK, Jha P, Mony P *et al*: Public health. Biobanks in developing countries: needs and feasibility. *Science* 2007; **318**: 1074–1075.
- 3 Akillu E, Dandara C, Bertilsson L, Masimirembwa C: Pharmacogenetics of cytochrome P450s in African populations: clinical and molecular evolutionary implications. In: Suarez-Kurtz G (ed.): *Pharmacogenomics in Admixed Populations*. Texas, USA: Landes Bioscience, 2007.
- 4 Abe M, Suzuki T, Deguchi T: An improved method for genotyping of *N*-acetyltransferase polymorphism by polymerase chain reaction. *Jpn J Hum Genet* 1993; **38**: 163–168.
- 5 Bell DA, Taylor JA, Butler MA *et al*: Genotype/phenotype discordance for human arylamine *N*-acetyltransferase (NAT2)

- reveals a new slow-acetylator allele common in African-Americans. *Carcinogenesis* 1993; **14**: 1689–1692.
- 6 Brockmoller J, Gross D, Kerb R, Drakoulis N, Roots I: Correlation between trans-stilbene oxide-glutathione conjugation activity and the deletion mutation in the glutathione S-transferase class mu gene detected by polymerase chain reaction. *Biochem Pharmacol* 1992; **43**: 647–650.
 - 7 de Morais SM, Wilkinson GR, Blaisdell J, Nakamura K, Meyer UA, Goldstein JA: The major genetic defect responsible for the polymorphism of S-mephenytoin metabolism in humans. *J Biol Chem* 1994; **269**: 15419–15422.
 - 8 Gaedigk A, Gotschall RR, Forbes NS, Simon SD, Kearns GL, Leeder JS: Optimization of cytochrome P4502D6 (CYP2D6) phenotype assignment using a genotyping algorithm based on allele frequency data. *Pharmacogenetics* 1999; **9**: 669–682.
 - 9 Pemble S, Schroeder KR, Spencer SR *et al*: Human glutathione S-transferase theta (GSTT1): cDNA cloning and the characterization of a genetic polymorphism. *Biochem J* 1994; **300** (Part 1): 271–276.
 - 10 Rotger M, Colombo S, Furrer H *et al*: Influence of CYP2B6 polymorphism on plasma and intracellular concentrations and toxicity of efavirenz and nevirapine in HIV-infected patients. *Pharmacogenet Genomics* 2005; **15**: 1–5.
 - 11 Cavalli-Sforza LL: The Human Genome Diversity Project: past, present and future. *Nat Rev Genet* 2005; **6**: 333–340.
 - 12 Rosenberg NA, Pritchard JK, Weber JL *et al*: Genetic structure of human populations. *Science* 2002; **298**: 2381–2385.
 - 13 Agyemang C, Bhopal R, Brujinzeels M: Negro, Black, Black African, African Caribbean, African American or what? Labelling African origin populations in the health arena in the 21st century. *J Epidemiol Community Health* 2005; **59**: 1014–1018.
 - 14 Ely B, Wilson JL, Jackson F, Jackson BA: African-American mitochondrial DNAs often match mtDNAs found in multiple African ethnic groups. *BMC Biol* 2006; **4**: 34.
 - 15 Sirugo G, Schim vdL, Sam O *et al*: A national DNA bank in the Gambia, West Africa, and genomic research in developing countries. *Nat Genet* 2004; **36**: 785–786.
 - 16 Kaiser J: Genomic medicine. African-American population biobank proposed. *Science* 2003; **300**: 1485.
 - 17 Klein K, Lang T, Saussele T *et al*: Genetic variability of CYP2B6 in populations of African and Asian origin: allele frequencies, novel functional variants, and possible implications for anti-HIV therapy with efavirenz. *Pharmacogenet Genomics* 2005; **15**: 861–873.
 - 18 Nolan D, Phillips E, Mallal S: Efavirenz and CYP2B6 polymorphism: implications for drug toxicity and resistance. *Clin Infect Dis* 2006; **42**: 408–410.
 - 19 Rodriguez-Novoa S, Barreiro P, Jimenez-Nacher I, Soriano V: Overview of the pharmacogenetics of HIV therapy. *Pharmacogenomics J* 2006; **6**: 234–245.
 - 20 Masimirembwa C, Persson I, Bertilsson L, Hasler J, Ingelman-Sundberg M: A novel mutant variant of the CYP2D6 gene (CYP2D6*17) common in a black African population: association with diminished debrisoquine hydroxylase activity. *Br J Clin Pharmacol* 1996; **42**: 713–719.
 - 21 Kirchheiner J, Brosen K, Dahl ML *et al*: CYP2D6 and CYP2C19 genotype-based dose recommendations for antidepressants: a first step towards subpopulation-specific dosages. *Acta Psychiatr Scand* 2001; **104**: 173–192.
 - 22 Masimirembwa CM, Hasler JA: Genetic polymorphism of drug metabolising enzymes in African populations: implications for the use of neuroleptics and antidepressants. *Brain Res Bull* 1997; **44**: 561–571.

Supplementary Information accompanies the paper on European Journal of Human Genetics website (<http://www.nature.com/ejhg>)

Maternal genotype effects can alias case genotype effects in case-control studies

European Journal of Human Genetics (2008) **16**, 783–785; doi:10.1038/ejhg.2008.74; published online 9 April 2008;

With the increasing popularity of case-control association studies in human genetics, it is worth recalling that other genetic mechanisms may masquerade as case genotype effects. In particular, in a case-control study, any maternal genotype effects are aliased with case genotype effects. The maternal genotype partially determines the uterine environment, leading to the possibility of detrimental effects in the developing fetus. Such maternal genotype effects have been implicated in developmental disorders, such as spina bifida and autism.^{1–4} Maternal-fetal interactions are a separate phenomenon not discussed here.⁵

When an allele contributes to susceptibility only in the mother, the offspring will be enriched for that allele simply by Mendelian inheritance. At a locus with two alleles A and a, with frequencies p and $q = 1 - p$, respectively, let r_1^m be the relative risk of disease given a single copy of A in the mother and r_2^m the relative risk of disease given two copies of A in the mother. Writing D_c for 'disorder present in the child,' M for the number of copies of the A allele in the mother, and C for the number of copies of the A allele in the child, we have

$$\begin{aligned} P(D_c|C = j) &= \sum_i P(D_c|M = i, C = j)P(M = i|C = j) \\ &= \sum_i P(D_c|M = i)P(M = i|C = j) \\ &= b \sum_i r_i^m P(M = i|C = j) \end{aligned}$$

where b is the prevalence of the disorder among offspring with mothers with genotype aa and $r_0^m = 1$. The conditional frequencies, under Hardy-Weinberg equilibrium, of the mothers' genotypes can be easily calculated⁶ but are given in Table 1 for easy reference.

By way of example, suppose $p = 0.1$, $r_2^m = 2$, and $r_1^m = 1.5$. For an aa child, by Table 1, the probability of an Aa mother