



HAL
open science

Recommendations for genetic variation data capture in developing countries to ensure a comprehensive worldwide data collection

George P. Patrinos, Jumana Al Aama, Aida I Al Aqeel, Fahd Al-Mulla, Joseph Borg, Andrew Devereux, Alex E. Felice, Finlay Macrae, Makia J Marafie, Michael B Petersen, et al.

► **To cite this version:**

George P. Patrinos, Jumana Al Aama, Aida I Al Aqeel, Fahd Al-Mulla, Joseph Borg, et al.. Recommendations for genetic variation data capture in developing countries to ensure a comprehensive worldwide data collection. *Human Mutation*, 2010, 32 (1), pp.2. 10.1002/humu.21397 . hal-00599480

HAL Id: hal-00599480

<https://hal.science/hal-00599480>

Submitted on 10 Jun 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



**Recommendations for genetic variation data capture in
developing countries to ensure a comprehensive worldwide
data collection**

Journal:	<i>Human Mutation</i>
Manuscript ID:	humu-2010-0429.R1
Wiley - Manuscript type:	Special Article
Date Submitted by the Author:	15-Oct-2010
Complete List of Authors:	<p>Patrinou, George; University of Patras, School of Health Sciences, Dept of Pharmacy; Erasmus MC, Bioinformatics Al Aama, Jumana; King Abdulaziz University, Department of Genetic Medicine, Faculty of Medicine and Princess Al-Jawhara Center of Excellence in Research of Hereditary Disorders AL Aqeel, Aida; Riyadh Military Hospital and Stem Cell therapy program, King Faisal Specialist Hospital and Research Center, Department of Pediatrics Al-Mulla, Fahd; Kuwait University, Molecular Pathology Unit, Borg, Joseph; University of Malta, Department of Physiology and Biochemistry, Laboratory of Molecular Genetics Devereux, Andrew; NGRl Manchester, St Mary's Hospital, Department of Medical Genetics Felice, Alex; University of Malta, Department of Physiology and Biochemistry, Laboratory of Molecular Genetics Macrae, Finlay; University of Melbourne, The Royal Melbourne Hospital,, Department of Colorectal Medicine and Genetics Marafie, Makia; Kuwait Medical Genetics Centre, Maternity Hospital, Sabah Medical Area, Clinical Genetics, Petersen, Michael; Institute of Child Health, Department of Genetics Qi, Ming; Center for Genetic and Genomic Medicine, The First Affiliated Hospital of Zhejiang University School of Medicine, James Watson Institute of Genomic Sciences Ramesar, Raj; University of Cape Town Medical school, Department of Human genetics Zlotogora, Joel; Ministry of Health, Community genetics Cotton, Richard; Genomic Disorders Research Centre, Howard Florey Institute</p>
Key Words:	Developing countries, National/ethnic mutation databases, Populations, Genetic variation

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

SCHOLARONE™
Manuscripts

For Peer Review

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Recommendations for genetic variation data capture in developing countries to ensure a comprehensive worldwide data collection

George P. Patrinos^{1,*}, Jumana Al Aama², Aida Al Aqeel³, Fahd Al-Mulla⁴, Joseph Borg⁵,
Andrew Devereux⁶, Alex E. Felice⁷, Finlay Macrae⁸, Makia J. Marafie⁹, Michael B. Petersen¹⁰,
Ming Qi¹¹, Rajkumar S. Ramesar¹², Joel Zlotogora¹³, and Richard G. Cotton¹⁴

¹ University of Patras, School of Health Sciences, Department of Pharmacy, Patras, Greece;

² Department of Genetic Medicine, Faculty of Medicine and Princess Al-Jawhara Center of Excellence in Research of Hereditary Disorders, King Abdulaziz University, Jeddah, Saudi Arabia;

³ Department of Pediatrics, Riyadh Military Hospital and Stem Cell therapy program, King Faisal Specialist Hospital and Research Center, Riyadh, Saudi Arabia;

⁴ Molecular Pathology Unit, Kuwait University, Kuwait;

⁵ University of Malta, Faculty of Health Sciences, Department of Applied Biomedical Sciences, Msida, Malta;

⁶ NGRL Manchester, Department of Medical Genetics, St Mary's Hospital, Manchester, United Kingdom;

⁷ University of Malta, Department of Physiology and Biochemistry, Laboratory of Molecular Genetics, Msida, Malta;

⁸ Department of Colorectal Medicine and Genetics and University of Melbourne, The Royal Melbourne Hospital, Australia;

1
2
3⁹ Clinical Genetics, Kuwait Medical Genetics Centre, Maternity Hospital, Sabah Medical Area,
4
5
6 Kuwait;

7
8¹⁰ Institute of Child Health, Department of Genetics, Athens, Greece;

9
10¹¹ Center for Genetic and Genomic Medicine, The First Affiliated Hospital of Zhejiang
11
12 University School of Medicine and James Watson Institute of Genomic Sciences, P. R. China;

13¹² MRC Human Genetics Research Unit, Division of Human Genetics, Institute for Infectious
14
15 Diseases and Molecular Medicine, Faculty of Health Sciences, University of Cape Town, South
16
17 Africa;

18
19¹³ Department of Community Genetics, Public Health Services, Ministry of Health, Jerusalem,
20
21 Israel;

22
23¹⁴ Genomic Disorders Research Centre, Howard Florey Institute, Melbourne, Australia & Faculty
24
25 of Medicine, Dentistry and Health Sciences, The University of Melbourne, Australia
26
27
28
29
30
31
32
33
34
35

36 Running head: Recommendations for genetic variation data capture in developing countries
37
38
39
40

41 * Corresponding author at:

42
43 Department of Pharmacy, School of Health Sciences, University of Patras,

44
45 University Campus, Rion, GR-265 04, Patras,

46
47
48 Greece

49
50 Telephone/Fax: +30-2610-969.834,

51
52 E-mail: gpatrinos@upatras.gr
53
54
55
56
57
58
59
60

Abstract

Developing countries have significantly contributed to the elucidation of the genetic basis of both common and rare disorders, providing an invaluable resource of cases due to large family sizes, consanguinity and potential founder effects. Moreover, the recognised depth of genomic variation in indigenous African populations, reflecting the ancient origins of humanity on the African continent, and the effect of selection pressures on the genome, will be valuable in understanding the range of both pathological and non-pathological variations. The involvement of these populations in accurately documenting the extant genetic heterogeneity is more than essential. Developing nations are regarded as key contributors to the Human Variome Project (HVP; <http://www.humanvariomeproject.org>), a major effort to systematically collect mutations that contribute to or cause human disease and create a cyber infrastructure to tie databases together. However, biomedical research has not been the primary focus in these countries even though such activities are likely to produce economic and health benefits for all. Here, we propose several recommendations and guidelines to facilitate participation of developing countries in genetic variation data documentation, ensuring an accurate and comprehensive worldwide data collection. We also summarize a few well-coordinated genetic data collection initiatives that would serve as paradigms for similar projects.

Key words: Developing countries, national/ethnic mutation databases, populations, genetic variation

Introduction

The recent elucidation of the genome sequence of several individuals, such as Caucasians (Jim Watson [Wheeler et al., 2008] and Craig Venter [Levy et al., 2007]), West African [Bentley et al., 2008], and Han Chinese [Wang et al., 2008], suggests that a large number of SNPs and other sequence variation exists in the human population. Estimates from African genetic diversity and the Pan Asian SNP initiative suggest that approximately 80–90% of human genomic variation resides in the world's developing countries. It is therefore obvious that there is a huge need to comprehensively document the extant genetic variation in various ethnic groups and populations around the globe to ensure maximal data coverage.

Since 2002, there have been several initiatives to thoroughly document the genetic heterogeneity in various populations worldwide in National/Ethnic Mutation databases (NEMDBs), a large number of which reside in developing countries. Despite the fact that data content is rather homogeneous, facilitated by the existence of an off-the-shelf database management system [van Baal et al., 2010], these databases have been developed without appropriate guidelines and recommendations to ensure not only a thorough description of common and rare genetic diseases in these populations but also maximal visibility and usefulness of these resources in the respective countries while maintaining the individual local ethical and cultural values in them. Furthermore, developing countries are often a valuable resource of clinical samples that derive from consanguineous families with very unusual phenotypes. These genotype and phenotype data are worth being recorded in NEMDBs not only to make them readily available to interested parties but also to develop and implement genetic epidemiological approaches for estimating the relative frequency of rare autosomal recessive disorders and an estimate of their prevalence in these populations.

1
2
3 Here, we propose a number of guidelines and recommendations that would assist in
4 genetic variation data capture in developing countries to ensure a comprehensive worldwide data
5 collection, one of the aims and goals of the Human Variome Project [HVP; Cotton et al 2007;
6 Kaput et al., 2009; www.humanvariomeproject.org]. These guidelines are largely the result of
7 discussions among the participants of the developing countries working group during the
8 International HVP Planning and Implementation meetings in May 2008 and 2010, respectively,
9 coming from a wide range of developed and developing countries and extended, based on
10 ongoing data documentation efforts in these countries.
11
12
13
14
15
16
17
18
19
20
21
22
23
24

25 **Background**

26
27
28
29 Any formal attempt to identify the extent of genetic variation must include geographical
30 regions that have not been included in previous haplotype mapping projects, e.g., the HapMap
31 project, particularly since the latter project included racial and not population groups. Although
32 the Population Reference Sample (POPRES) will address some of these missing populations
33 [Nelson et al., 2008], this effort is again designed as a mapping project that is not expected to
34 focus on functional polymorphisms or mutations. Also, the Arab region has long played a key
35 role in discovering new genes and it is really awkward why this region has not been so actively
36 involved in today's genomics research and genome-wide association studies.
37
38
39
40
41
42
43
44
45
46
47
48

49 Hence, one of the main activities of the HVP is the analysis and subsequent inclusion of
50 genetic variation data from clinical samples originating from diverse populations or ethnic
51 groups [Kaput et al., 2009]. The distinct advantage of some ethnic populations is the opportunity
52 to study genetic diseases due to: (a) large family size, (b) consanguinity, and (c) potential
53 founder effects [Saadallah and Rashed, 2007]. In fact, several disorders have been described for
54 the first time in certain populations residing in developing countries e.g Sanjad-Sakatti
55
56
57
58
59
60

1
2
3 syndrome, Teebi-Shaloot syndrome, Teebi Hypertelorism and others. Based on the
4
5 abovementioned advantages, developing nations are regarded as major contributors to the HVP
6
7 even though biomedical research has not been the focus of these countries, despite such activities
8
9 being likely to produce economic and health benefits for all [Singer and Daar, 2001; Kaput et al.,
10
11 2009].
12
13

14
15 The NEMDBs are one of the main components of the HVP that could accommodate the
16
17 extant genetic heterogeneity in the developing countries and, for that matter, all countries. In
18
19 particular, the NEMDBs are mutation depositories needing continuous updating, recording
20
21 extensive information over the described genetic heterogeneity of an ethnic group or population
22
23 [reviewed in Patrinos, 2006]. NEMDBs can help optimize national molecular diagnostic
24
25 services, by stratifying the corresponding mutation spectra in various inherited diseases, and
26
27 assist in interpreting diagnostic test results in countries with heterogeneous populations,
28
29 particularly in minority ethnic groups. NEMDBs can also contribute toward the elucidation of
30
31 the origin and migration of populations, acting as platforms for comparative genomic studies that
32
33 can reciprocally provide insights into, e.g., the demographic history of human populations,
34
35 patterns of their migration and admixture, gene/mutation flow, *etc* [Patrinos, 2006]. Such
36
37 comprehensive databases would include information on genotype and reasonable phenotype, in a
38
39 level of detail that should be agreed by the community, which would also enable those interested
40
41 in these diseases to investigate the possible influence of geographical environment or genetic
42
43 modifiers on expression of the disease/phenotype. Examples include the genetics of obesity,
44
45 hypertension or type II diabetes. Relevant genotype/phenotype collection projects, such as
46
47 GEN2PHEN (<http://www.gen2phen.org>) or P3G (<http://www.p3g.org>) could serve as models to
48
49 establish the necessary requirements for the depth of the recorded genotype and phenotype
50
51 information.
52
53
54
55
56
57
58
59
60

Recommendations

The recommendations and guidelines that follow (summarized in Box 1) are intended to facilitate genetic variation data capture in developing countries, either by suggesting specific measures aiming to enhance awareness among clinicians, bioscientists, and the general public about the range of commonest genetic disorders suffered by certain populations and/or ethnic groups, or by sharing the experience from ongoing projects to document genetic heterogeneity in different countries.

1. Develop procedures to include clinicians and researchers in developing nations, mostly by creating networks among those of common interest.

This would be particularly important for clinicians/researchers working in under-resourced institutes and research environments. Where construction of a NEMDB is decided, selected representatives from participating laboratories, mainly established scientists and principal investigators involved in human genetics should take the lead in this effort. These individuals would form the advisory committee of the NEMDB with one of them assigned as national coordinator with a defined (e.g. yearly) term. These members will then assign eligible members from their groups that would serve as data curators responsible for data entry/modification.

Such efforts would flourish if placed under the umbrella of specific scientific initiatives or consortia (see also below), such as the Human Genome Variation Society (HGVS; www.hgvs.org), a Society which generates recommendations and guidelines towards human genome variation capture, the HVP and ORPHANET (<http://www.orpha.net>), a database of information on rare diseases and orphan drugs, with presence in 37 mostly European and Arab countries, that aims to contribute to the improvement of the diagnosis, care and treatment of

1
2
3 patients with rare diseases. This approach is already adopted in the Israeli and Greek genetics
4 communities, mostly to share genetic laboratory information, but also to accelerate the data
5 collection process. Partnering with local human genetic societies is also liable to have the same
6 result.
7
8
9
10
11

12 Moreover, in order to increase awareness among interested parties, the resulting
13 NEMDBs should be made known to their potential users using various communication channels.
14 These do not necessarily include international scientific journals to which local institutes in these
15 countries might not have access, but local scientific journals, news bulletins, scientific events
16 (symposia, workshops), one-to-one communication, and so on. Such an approach was followed
17 by the Israeli NEMDB, by producing a review article in the Israeli Medical Association Journal,
18 a leading peer-reviewed scientific medical journal in Israel [Zlotogora et al., 2009].
19
20
21
22
23
24
25
26
27
28
29
30
31

32 **2. Create international networks from closely related populations to enhance** 33 **communication, interaction, and initiate research networks between them.** 34

35 In this case, it would be mutually beneficial for closely related countries such as in Latin
36 America, the Middle East or regions in Africa, to initiate new or strengthen existing
37 collaborative ties towards comprehensive genetic variation data capture. Similarly, specific
38 resources would be committed in different countries so that there will be no redundancy in their
39 research efforts. In this regard, the Catalogue for Transmission Genetics in Arabs (CTGA;
40 Tadmouri et al., 2006; see also below) initiated from the Centre for Arab Genomic Studies
41 (<http://www.cags.org.ae>) in Dubai, United Arab Emirates, is a pilot project to document genetic
42 disorders in Arab populations from the published literature. The current version of CTGA
43 database contains nearly 760 full-text records for nearly 1,290 Mendelian disorders and related
44 genes, including extensive data from the United Arab Emirates, Bahrain, Oman, and Qatar.
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

3. Foster communication, interaction, and research networks between developing and developed countries.

This would be particularly beneficial in both data collection from the literature or generation of new genetic variation data and would include:

- (a) Education of healthcare providers, the public, and government officials from researchers and health care providers from developed countries, regarding data collection, data generation and interpretation, and so on. To this end, the Pan-Arab Human Genetics conferences, with international participation from developed countries, could very well be an example of such interactions. The African Society for Human Genetics is also spearheading efforts to create a database of genetic variation in indigenous African populations (Raj Ramesar; personal communication; Al Aama and co-workers, submitted)
- (b) Assistance in data handling and storage in large databases and servers in developed countries, particularly where massive storage units (e.g. GRIDs) are needed to accommodate large datasets. Such possibilities are currently offered by LSDB and NEMDB software developers (LOVD [Fokkema et al., 2005] and ETHNOS [van Baal et al., 2010], respectively) in the Netherlands and Greece, respectively, to either download for use or to store in their local servers, data from interested parties and curators that do not have either the knowledge or the resources needed, although ultimately the aim would be to provide the infrastructure locally.
- (c) Consultancy in the design and coordination of large-scale genome projects (see also below) if such need arises, including patients recruitment, experimental design, software/database development (see also Point 5),
- (d) Outsourcing of high-throughput experiments, such as microarray or next-generation sequencing experiments, to laboratories in the developed countries, where such

1
2
3 infrastructure and expertise is more likely to be available, but also provide expert external
4 training to those countries where the facilities are available but outstrip the local expertise.
5
6

- 7
8 (e) To work interactively, leading from all of the above, to train postgraduate students and other
9 researchers towards capacity development in developing countries.
10
11

12
13
14
15 **4. Ensure that the database management systems that are being used or developed can be**
16 **utilized by those in a limited resource environment.**
17
18

19
20 Nowadays, LSDBs and NEMDBs, which capture the core information in any one of these
21 domains of information, are developed by experts and curated with invaluable skill and
22 experience. Often, developing countries suffer from a limited resource research environment;
23 hence this reality should be carefully considered when database management systems are
24 developed for this purpose.
25
26
27
28
29
30

31
32 Therefore, the database management system should be designed and built such that the
33 resulting database would:
34

- 35
36 (a) Record the specific variants of inherited disorders found in a population or ethnic group in a
37 comprehensive and user-friendly manner, i.e., by providing consistent and simplified
38 information accompanied by numerous links to other databases and related resources
39 [Patrinos, 2006],
40
41
42
43
44
45 (b) Provide the end user with an interface for easy data query and retrieval of information
46 pertaining to the mutation spectrum of each disorder together with the corresponding
47 mutation frequencies and references,
48
49
50
51
52 (c) Allow permanent data storage and secure data entry and modification. In this case, software
53 pipelines should also allow automated genotype data transfer between the clinic and/or
54 laboratory to the database, and ideally from the genotyping equipment (sequencer, scanner,
55 *etc*) directly to the database,
56
57
58
59
60

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

(d) Operate under strict quality assurance guidelines, such as intellectual property rights, the professional code of ethics (<http://ihealthcoalition.org/ethics/ehcode.html>) or specific ethical guidelines [Povey et al., 2010], laws on computing and liberties (e.g., www.ama-assn.org/ama/pub/category/1905.html), electronic data protection, specific guidelines and any law or regulation applicable in addition to the local ethical issues that may be specific to a given country.

Such databases should be able to operate remotely, should the required support not be available locally (see also Point 3)

5. Provide support to developing countries to build capacity and to fully participate in the collection, analysis and sharing of genetic variation information.

Comprehensive genetic variation data capture usually entails highly specialized researchers or diagnostic laboratory staff and state-of-the-art infrastructure, particularly where generation of new genomic data comes to play. Developing countries usually suffer from brain drain and lack of highly sophisticated equipment to perform cutting-edge research. To this end, developed countries have the obligation to alleviate this burden and to assist developing nations by every means possible [Kaput et al., 2009]. It should be noted that developing countries have a valuable resource of clinical samples, often with very unusual phenotypes, particularly when consanguineous families are involved since in this case, the elucidation of the genetic basis of the underlying phenotype becomes much easier. Developing countries should therefore:

(a) Educate healthcare providers and government officials to demonstrate the need to initiate documentation of the genetic variation in their countries and to underline the benefits of cooperation in biomedical research [Bhan et al., 2007],

- 1
2
3
4 (b) Increase awareness in the general population about the need to comprehensively record the
5 underlying genetic heterogeneity in their country and informing them objectively about the
6 benefits and risks of such projects,
7
8
9
- 10 (c) Ascertain all parties involved that the research outcome will not be misused. Certain
11 populations or ethnic groups in developing countries usually mistrust research involving
12 genetic analyses or fear that results can stigmatize them or worse
13 (<http://www.eubios.info/ASIAE/BIAE201.htm>). For example, Malay-Muslims, Chinese, and
14 Indians in Singapore expressed anxiety about breach of confidentiality, the misuse of their
15 DNA for cloning, and possibilities of being diagnosed with disease [Wong et al., 2004].
16 Community-based research collaborations may provide fora for addressing cultural and
17 ethical concerns of biomedical research [Povey et al., 2010]. Analyzing the extent of human
18 genome variation creates an ideal opportunity for the developed and the developing nations
19 of the world to forge meaningful partnerships and to work together in an unprecedented
20 way, initially to identify variation causing disease, and then to simplify mutation detection in
21 specific ethnic groups, and then to understand how genetic variation contributes to human
22 phenotypic diversity. By ensuring that all nations and ethnic groups have an equal and fair
23 opportunity to share data and technology, evidence-based information will be provided such
24 that all populations can benefit from a global society health network.
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
- 46 (d) Emphasize in every stage of the collaborative project that external input is needed from the
47 design and coordination of NEMDB development and curation to database hosting and
48 technical support, patient recruitment, and experimental design to performing high-
49 throughput experiments in the case of large-scale genome projects (see also Point 3). To this
50 end, funding agencies (e.g. UKIERI; UK-India Education and Research Initiative) could
51 assist by funding innovative projects so that developing countries become independent and
52 self-sustained.
53
54
55
56
57
58
59
60

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

(e) Develop partnerships for the equitable sharing of knowledge and technology between developed and developing countries. These partnerships should be ideally formed under the umbrella of international organizations, such as the HGVS or the Human Genome Organization (HUGO; <http://www.hugo-international.org>) that would also organize inter-institutional specialized educational programmes and workshops. The biennial mutation detection workshop would be a paradigm for this initiative and may sometimes be held in developing countries.

6. Develop a framework to facilitate interactions between the coordinating centre and national, regional and international agencies.

Researchers and their academic institutions that participate in national data collection projects and/or NEMDBs development should jointly develop infrastructures that would ensure smooth communication between them and national and international agencies. These include: (a) national governments and their Departments of Health, Science and Technology, Development, Education and so on, local human genetic societies, (b) regional organizations, such as the NEPAD (New Programme for Africa's Development) and (c) international organizations, such as the European Commission (EC), United Nations (UN), United Nations Children's Fund (UNICEF), World's Health Organization (WHO), Organization for Economic Cooperation and Development (OECD), *etc.*

This in turn would: (a) create new collaboration ties with neighbouring countries interested in initiating similar projects, (b) increase the chances of receiving funding for these activities (see also below), (c) increase visibility of the progress to the international scientific community.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

7. Ensure that all ethical, legal, religious, and social issues are thoroughly considered when NEMDBs and/or data capture projects are launched in developing countries.

The goal of establishing LSDBs and NEMDBs in developing countries is the sharing of genetic variation information for the benefit of the respective populations. This includes the protection of privacy, which in this context is the right of the individual and members of their family to be protected against intrusion into their personal information and further intrusions ensuing from access to this, by publication of information. Of course, there is no such issue when summary level data are reported for a population, as it is the case in some NEMDBs. Therefore, there should be a balance between the public's interest in the value of the shared information and its interest in the strict protection of privacy. This balance would be viewed differently in different cultures and ethnicities [Al-Aqeel, 2007].

For many of these NEMDBs, it seems reasonable that an independent group of qualified scientists oversees these resources and related activities not only at their initiation but on an ongoing basis. Where needed, the main purpose of the particular NEMDB should be clarified, recognising that this may change with time. This will allow evaluation of the exact requirements for data deposition in the NEMDBs, and whether compliance with the remaining guidelines will be possible with the database content freely available to the public or partly restricted to registered persons, namely curators, physicians, and so on. This measure was implemented in the Israeli NEMDB, where not all information is made publicly available [Zlogotora et al., 2007]), since access to specific cases is password-protected [Zlotogora et al., 2009]. A similar measure would entail anonymizing genetic data, either by permanently removing or coding the personal identifiers [Povey et al., 2010]. Of interest, the newest version of the LOVD software [Fokkema et al., 2005] has the option to store data that are not public but that can be queried. The result of a query hitting non-public data is a notification that there is such information in the database but that the curator needs to be contacted to get more information.

1
2
3 The depth of the ethnic and geographic origin data should be determined in advance,
4 particularly in the case of small villages where carriers of rare variants could be identified even if
5 personal information were not provided. In this case, data donors should be well informed and
6 provided with an explanation as to why their data documentation is useful and how their privacy
7 is secured. These individuals should consent to the inclusion of their genetic information in the
8 database and reserve the right to withdraw their data from the database whenever they desire.
9 This is particularly important for persons unable to consent because of disability or young age.
10 Standardized consent forms need to be translated to different local languages and/or dialects.
11
12
13
14
15
16
17
18
19
20
21

22 Finally, these databases would also benefit from some independent/external ethical
23 review, perhaps from an international group (HGVS or HVP) who could advise them
24 accordingly.
25
26
27
28
29
30
31

32 **Nation-wide genomic projects that would serve as paradigm for genetic** 33 **variation data collection** 34 35 36 37 38 39

40 By definition, a nation-wide genomic project is a national multicentre initiative to
41 determine the underlying genetic heterogeneity of a population with the goal to ascertain the
42 genetic variation leading to common or rare inherited diseases. Such initiatives consist of three
43 parts: (a) documenting the genetic heterogeneity that has already been reported in the published
44 literature, (b) generating new genetic data by screening a large proportion of the population using
45 high throughput genotyping approaches, and (c) accumulating data causing inherited disease
46 from hospitals, laboratories and other healthcare facilities. In every case, except perhaps (c), a
47 large national genetic database/portal is being developed to accommodate the sheer amount of
48 data gathered. To date, there are few nation-wide genomic projects in various developing
49 countries that are currently in progress. These projects (summarized in Table 1) implement some
50
51
52
53
54
55
56
57
58
59
60

1
2
3 of the guidelines described previously. Also, there are several other NEMDB projects, all of
4 which more or less follow these guidelines, that have already been reviewed [Patrinos, 2006; van
5 Baal et al., 2007] and will, therefore, not be further discussed since they lie outside the scope of
6 the present article.
7
8
9
10
11

12 The Estonian Genome Project (EGP; <http://www.geenivaramu.ee/index.php?lang=eng>) is
13 a research venture of the University of Tartu initiated in 2001 [Metspalu, 2002] and aims to
14 create a database of health, genealogical and genetic data representing approximately 10% of
15 Estonia's population. The database will make it possible for researchers both in Estonia and
16 outside to look for links between genes, environmental factors and common diseases (such as
17 cancer, diabetes, depression, cardio-vascular diseases, *etc*). Currently, EGP gene bank contains
18 data contributed by over 20,000 donors with an expected increase to 100,000 within the next
19 three or four years.
20
21
22
23
24
25
26
27
28
29
30
31

32 The Indian Genome Variation database (IGVdb; <http://www.igvdb.res.in>) is a consortium
33 activity established in 2003 to document the inherent genetic variability of the Indian
34 subpopulations as a first step towards identifying genetic biomarkers to understand susceptibility
35 for any genetic disease and population migration, pharmacogenomic markers to understand
36 variable drug response. The Indian Genome Variation (IGV) consortium aims to provide data on
37 validated SNPs and repeats, both novel and reported, along with gene duplications, in over 1,000
38 genes (selected on the basis of their relevance as functional and positional candidates in many
39 common diseases including genes relevant to pharmacogenomics), in 15,000 individuals drawn
40 from Indian subpopulations [Indian Genome Variation Consortium, 2005].
41
42
43
44
45
46
47
48
49
50
51
52

53 Kuwait established within the Faculty of Medicine, Department of Pathology a Molecular
54 Genetics Diagnostic Service Division (<http://www.al-mulla.org>), integrated within a research
55 laboratory, which is focused on delivering state-of-the-art diagnostic service for the Kuwaiti
56 population. These services, which focus on autozygosity mapping in families with
57
58
59
60

1
2
3 consanguineous marriages, provide the research laboratory with a unique opportunity to identify
4 novel genes in autosomal recessive disorders. Moreover, various unique cases suffering from
5 cancer are currently being analysed using the Affymetrix Array 6.0 microarray technology to
6 identify copy number changes and homozygosity stretches. The data generated from these
7 studies are deposited in CAGS database (see above), and LOVD, while the Kuwaiti NEMDB has
8 been recently established databases to encourage data storage for genetic diseases in different
9 subpopulations in Kuwait.

10
11
12
13
14
15
16
17
18
19
20 The United Kingdom has developed a central database to collect all UK genome variation
21 data, and their effects, from participating diagnostic laboratories. The diagnostic mutation
22 database (DMuDB; <http://ngri.org.uk/Manchester/dmudb.html>) provides a route for sharing
23 mutation data within and between diagnostic laboratories in the United Kingdom to support
24 genetic testing services for patients. Currently DMuDB has over 4,884 records for 33 genes,
25 namely Breast Cancer (*BRCA1*, *BRCA2*), Lowe Syndrome (*OCRL*), Sotos Syndrome (*NSD1*),
26 HNPCC (*MLH1*, *MSH2*), Cystic Fibrosis (*CFTR*), X-Linked Retinitis Pigmentosa (*XLRP*),
27 Neurofibromatosis, Types I and II (*NF1*, *NF2*), Alstrom Syndrome (*ALMS1*), CADASIL
28 (*NOTCH3*), Familial adenomatous polyposis (*FAP*) and Muscular Dystrophy (*DMD*), containing
29 over 10,985 individual variants.

30
31
32
33
34
35
36
37
38
39
40
41
42
43
44 The Hellenic Genome Variation database aims to document the genetic heterogeneity in
45 the Hellenic population. The first version of the Hellenic NEMDB was created in 2004 [Patrinos
46 et al., 2005]. In 2007, research institutes involved in hemoglobinopathies research initiated a
47 pilot multicenter project to determine the mutation spectrum of the thalassemia syndromes in
48 various geographical regions in the country with the goal to document the findings in the
49 Hellenic NEMDB. These data showed that although Greece is a rather homogeneous country,
50 there are big qualitative (type of mutations) and quantitative (allelic frequencies) differences in
51 various parts of the country, namely central Greece [Samara et al., 2007] and South-Western
52
53
54
55
56
57
58
59
60

1
2
3 Greece [Papachatzopoulou et al., 2010]. These data compared to findings from North-Western
4
5 Greece [Georgiou et al., 2003] and the rest of the country [Patrinos et al., 2005] reveal valuable
6
7 information for region-specific population carrier screening and for efficient provision of
8
9 prenatal diagnosis of β -thalassemia in Greece.
10
11

12
13 Finally, there are four international consortia with activities related to the documentation
14
15 of causative mutations and allelic frequencies in developing countries:
16
17

- 18
19 (a) The International Society for Gastrointestinal Hereditary Tumours (InSiGHT;
20
21 <http://www.insight-group.org>) is an international multidisciplinary, scientific organisation
22
23 with a mission "...to improve the quality of care of patients and their families with any
24
25 condition resulting in hereditary gastrointestinal tumours." The InSiGHT consortium
26
27 maintains a LOVD-based database for genetic variations for the *EPCAM*, *MLH1*, *MLH3*,
28
29 *MSH2*, *MSH6*, *MUTYH*, *PMS1*, and *PMS2* genes leading to gastrointestinal tumours and
30
31 encourages the establishment of national repositories for patients suffering from
32
33 gastrointestinal tumours, particularly in developing countries.
34
35
36
37
38 (b) The International Thalassemia Network (ITHANET; <http://www.ithanet.eu>) is a European
39
40 Commission-funded project that comprises a network of all major European research
41
42 institutions active in the field of molecular and clinical research of thalassemia and related
43
44 hemoglobinopathies [Lederer et al., 2009]. Among its activities, ITHANET hosts a database
45
46 (ITHANETbase) that includes mutation frequencies in various populations at-risk for
47
48 thalassemia, particularly from developing countries. In many instances though, this
49
50 repository is largely redundant to HbVar database of hemoglobin variants and thalassemia
51
52 mutations [<http://globin/bx.psu.edu/hbvar>; Patrinos et al., 2004].
53
54
55
56
57 (c) The MEDGENET project (<http://medgenet.tredueuno.it>) is a European Consortium, funded
58
59 by the European Commission, aiming to expand the human expertise in clinical genetics in
60

1
2
3 Mediterranean Partner Countries (MPC) through the transfer of knowledge and technology
4
5 between the two rims of the Mediterranean which share a common burden of genetic
6
7 diseases. Among the actions foreseen by MEDGENET is the building of Medical Genetics
8
9 databases for the various partner countries. Several NEMDBs are built from scratch, such as
10
11 the Tunisian and Egyptian NEMDBs or expanded from the existing NEMDBs, such as the
12
13 Israeli [Zlotogora et al., 2007], Cypriot [Kleanthous et al., 2006], and Lebanese NEMDBs
14
15 [Megarbane et al., 2007], using the ETHNOS database management system for NEMDB
16
17 development and curation [van Baal et al., 2010].
18
19
20
21

- 22
23 (d) The Pharmacogenomics for Every Nation Initiative (PGENI; www.pgeni.org) that aims to
24
25 determine the frequencies of the commonest pharmacogenomic markers in various
26
27 populations, particularly from developing countries worldwide.
28
29

30
31 Finally, there are a few nation-wide DNA biobanks, such as the National Laboratory for
32
33 the Genetics of Israeli Populations, representing a national repository for human cell-lines and
34
35 DNA from the unique and large ethnic variation of the Israeli populations (Tel Aviv University;
36
37 available at <http://www.tau.ac.il/medicine/NLGIP>) or the Biomaterial Banking Medical Research
38
39 Initiative (BBMRI) in Malta, which of course lie at the borderline of the nation-wide genomic
40
41 projects described above.
42
43
44
45
46

47 **Funding and incentives**

48
49
50
51

52
53 There are currently two main obstacles that hinder further efforts to expand the NEMDB
54
55 field, namely funding and providing incentives to data submitters.
56

57
58 Funding for collecting genetic variation data has traditionally been difficult due to the
59
60 extreme fragmentation of the field [Patrinos and Brookes, 2005]. Therefore, it would be expected
that efforts to receive funding for genetic variation data capture are more likely to flourish if

1
2
3 treated as concerted efforts. GEN2PHEN (<http://www.gen2phen.org>) represents such initiative
4
5 that have already received funding from the EC and aims to unify human and model organism
6
7 genetic variation databases towards increasingly holistic views into Genotype-To-Phenotype
8
9 data, and to link this system into other biomedical knowledge sources via genome browser
10
11 functionality. Linking population and ethnic-specific information into LSDBs and diagnostic
12
13 databases, creating NEMDBs and developing genomic databases, similar to the Indian Genome
14
15 Variation database are key activities for the main GEN2PHEN work-packages. Obviously, other
16
17 strategically minded funding agencies (e.g. NIH), charities, and foundations would also launch
18
19 similar calls. Recently the Australian Node of the HVP has been funded to pilot software for data
20
21 flow from hospitals/clinics via country specific databases to LSDBs and NCBI/EBI/UCSC
22
23 browsers (unpublished). Involving universal organizations such as the WHO and UNESCO in
24
25 the fund-raising project will clearly have a strong impact.
26
27
28
29
30

31
32 On the other hand, the development of comprehensive analyses of human genetic
33
34 variation often reaches a barrier in the form of the reward system for clinicians and academic
35
36 researchers. First of all, database entries could be made a mandatory quality control standard for
37
38 clinical laboratories and clinicians. Notably, according to a new initiative of the Israeli
39
40 government, genetic laboratories are mandated to report the details of the tests they are
41
42 performing in the Israeli NEMDB in order to obtain accreditation from the Israeli Ministry of
43
44 Health [van Baal et al., 2010].
45
46
47

48
49 For researchers, a publication or Web-based system establishing microattribution and
50
51 community annotation of mutations and cited data will enable measurable contributions to the
52
53 scientific knowledge base [Anonymous, 2008; Hoffmann, 2008]. Similarly, database-journals
54
55 would also serve this task by providing a forum for publishing gene variation data that would be
56
57 eventually deposited in the PubMed literature database. One such journal, the open access
58
59 Human Genomics and Proteomics (<http://www.sage-hindawi.com/journals/hgp>), which is
60

1
2
3 affiliated with FINDbase [<http://www.findbase.org>; van Baal et al., 2007], will focus on studies
4
5 characterizing causative mutation and/or biomarker frequency spectra. Accepted contributions
6
7 including datasets will be linked in FINDbase and deposited in PubMed [Patrinis and Petricoin,
8
9 2009]. Also, the journal *Human Genetics* has been accepting mutation data and this is linked to
10
11 HGMD (<http://www.hgmd.org>).
12
13

14
15 Journals, tenure, and promotion committees, and funding agencies would be encouraged
16
17 to cite these contributions to and citations of LSDB and international national databases [Patrinis
18
19 and Brookes, 2005]. The rapid progress in genetic research has led to a declining interest of
20
21 scientific clinical genetic journals in case reports with single or few novel mutations. This may
22
23 deter clinicians from reporting these new mutations, which collectively become important, unless
24
25 another venue becomes visible.
26
27
28
29
30

31 32 **Concluding remarks**

33
34
35
36
37 As stated above, documenting population/ethnic-specific data in developing countries are
38
39 among the primary objectives of the HVP that appreciates the genomic sovereignty/equality for
40
41 all countries to be involved and acknowledges the value of “human capital” within all
42
43 populations. Real and tangible benefits of the HVP to improve health will be generated for
44
45 participating populations. Hence, the voluntary participation of the greatest number of countries
46
47 would ensure a more general applicability, and it is hoped that many countries will eventually
48
49 decide to participate.
50
51
52
53
54

55 56 **Acknowledgements**

57
58
59
60

1
2
3 We thank the members of the developing countries working group at the HVP planning
4 and implementation meetings for lively and helpful discussions. Part of this work has received
5 funding from the European Commission [ITHANET (FP6-026539), EuroGenTest (FP6-512148),
6 MEDGENET (FP6-31968) and GEN2PHEN (FP7-200754) projects], and the Kuwait
7 Foundation for the advancement of Sciences (KFAS) grant number 2006-1302-07.
8
9
10
11
12
13
14
15
16
17

18 **References**

- 19
20
21
22
23 Al Aqeel AI. 2007. Islamic ethical framework for research into and prevention of genetic
24 diseases. *Nat Genet* 39:1293-1298.
25
26
27 Anonymous. 2008. Human variome microattribution reviews. *Nat Genet* 40:1.
28
29
30 Bentley DR, Balasubramanian S, Swerdlow HP, Smith GP, Milton J, Brown CG, Hall KP, Evers
31 DJ, Barnes CL, Bignell HR, Boutell JM, Bryant J, et al. 2008. Accurate whole human
32 genome sequencing using reversible terminator chemistry. *Nature* 456:53–59.
33
34
35 Bhan A, Singh JA, Upshur RE, Singer PA, Daar AS. 2007. Grand challenges in global health:
36 engaging civil society organizations in biomedical research in developing countries. *PLoS*
37 *Med* 4:e272.
38
39
40
41
42
43 Cotton RG, Al Aqeel AI, Al-Mulla F, Carrera P, Claustres M, Ekong R, Hyland VJ, Marafie MJ,
44 Paalman MH, Patrinos GP, Qi M, Ramesar RS, Scott RJ, Sijmons RH, Sobrido MJ, Vihinen
45 M. (2009). Capturing All Disease Causing Mutations For Clinical And Research Use:
46 Towards An Effortless System For The Human Variome Project. *Genet Med*, 11:843-849.
47
48
49
50
51
52 Cotton RG, Appelbe W, Auerbach AD, Becker K, Bodmer W, Boone DJ, Boulyjenkov V,
53 Brahmachari S, Brody L, Brookes A, Brown AF, Byers P, Cantu JM, Cassiman JJ, Claustres
54 M, Concannon P, Cotton RG, den Dunnen JT, Flicek P, Gibbs R, Hall J, Hasler J, Katz M,
55 Kwok PY, Laradi S, Lindblom A, Maglott D, Marsh S, Masimirembwa CM, Minoshima S,
56
57
58
59
60

- 1
2
3 de Ramirez HZ, Scriver CR, Sherry S, Shimizu N, Shimizu N, Stein L, Tadmouri GO,
4
5 Taylor G, Watson M. 2007. Recommendations of the 2006 Human Variome Project
6
7 meeting. *Nat Genet* 39:433–436.
8
9
- 10 Fokkema IF, den Dunnen JT, Taschner PE. 2005. LOVD: easy creation of a locus-specific
11
12 sequence variation database using an "LSDB-in-a-box" approach. *Hum Mutat* 26:63-68.
13
14
- 15 Georgiou I, Makis A, Chaidos A, Bouba I, Hatzi E, Kranas V, Zilidis C, Bourantas KL. 2003.
16
17 Distribution and frequency of beta-thalassemia mutations in northwestern and central
18
19 Greece. *Eur J Haematol* 70:75-78.
20
21
- 22 Hinds DA, Stuve LL, Nilsen GB, Halperin E, Eskin E, Ballinger DG, Frazer KA, Cox DR. 2005.
23
24 Whole-genome patterns of common DNA variation in three human populations. *Science*
25
26 307:1072–1079.
27
28
- 29 Hoffman R. 2008. A wiki for the life sciences where authorship matters. *Nat Genet* 40:1047–
30
31 1051.
32
33
- 34 Indian Genome Variation Consortium. 2005. The Indian Genome Variation database (IGVdb): a
35
36 project overview. *Hum Genet* 118:1-11.
37
38
- 39 Kaput J, Cotton RG, Hardman L, Watson M, Al Aqeel AI, Al-Aama JY, Al-Mulla F, Alonso S,
40
41 Aretz S, Auerbach AD, Bapat B, Bernstein IT, Bhak J, Bleoo SL, Blöcker H, Brenner SE,
42
43 Burn J, Bustamante M, Calzone R, Cambon-Thomsen A, Cargill M, Carrera P, Cavedon L,
44
45 Cho YS, Chung YJ, Claustres M, Cutting G, Dagleish R, den Dunnen JT, Díaz C,
46
47 Dobrowolski S, Dos Santos MR, Ekong R, Flanagan SB, Flicek P, Furukawa Y, Genuardi
48
49 M, Ghang H, Golubenko MV, Greenblatt MS, Hamosh A, Hancock JM, Hardison R,
50
51 Harrison TM, Hoffmann R, Horaitis R, Howard HJ, Barash CI, Izagirre N, Jung J, Kojima
52
53 T, Laradi S, Lee YS, Lee JY, Gil-da-Silva-Lopes VL, Macrae FA, Maglott D, Marafie MJ,
54
55 Marsh SG, Matsubara Y, Messiaen LM, Möslin G, Netea MG, Norton ML, Oefner PJ,
56
57 Oetting WS, O'Leary JC, de Ramirez AM, Paalman MH, Parboosingh J, Patrinos GP,
58
59
60

- 1
2
3 Perozzi G, Phillips IR, Povey S, Prasad S, Qi M, Quin DJ, Ramesar RS, Richards CS,
4
5 Savige J, Scheible DG, Scott RJ, Seminara D, Shephard EA, Sijmons RH, Smith TD,
6
7 Sobrido MJ, Tanaka T, Tavigian SV, Taylor GR, Teague J, Töpel T, Ullman-Cullere M,
8
9 Utsunomiya J, van Kranen HJ, Vihinen M, Webb E, Weber TK, Yeager M, Yeom YI, Yim
10
11 SH, Yoo HS. 2009. Planning the Human Variome Project. The Spain report. *Hum Mutat*
12
13 30:496-510.
14
15
16
17 Kleanthous M, Patsalis PC, Drousiotou A, Motazacker M, Christodoulou K, Cariolou M, Baysal
18
19 E, Khrizi K, Pourfarzad F, Moghimi B, van Baal S, Deltas CC, Najmabadi S, Patrinos GP.
20
21 2006. The Cypriot and Iranian National Mutation databases. *Hum Mutat* 27:598-599.
22
23
24 Lederer CW, Basak AN, Aydinok Y, Christou S, El-Beshlawy A, Eleftheriou A, Fattoum S,
25
26 Felice AE, Fibach E, Galanello R, Gambari R, Gavrilu L, Giordano PC, Grosveld F,
27
28 Hassapopoulou H, Hladka E, Kanavakis E, Locatelli F, Old J, Patrinos GP, Romeo G, Taher
29
30 A, Traeger-Synodinos J, Vassiliou P, Villegas A, Voskaridou E, Wajcman H, Zafeiropoulos
31
32 A, Kleanthous M. 2009. An electronic infrastructure for research and treatment of the
33
34 thalassemias and other hemoglobinopathies: the Euro-mediterranean ITHANET project.
35
36 *Hemoglobin* 33:163-176.
37
38
39
40 Levy S, Sutton G, Ng PC, Feuk L, Halpern AL, Walenz BP, Axelrod N, Huang J, Kirkness EF,
41
42 Denisov G, Lin Y, MacDonald JR, Pang AW, Shago M, Stockwell TB, Tsiamouri A, Bafna
43
44 V, Bansal V, Kravitz SA, Busam DA, Beeson KY, McIntosh TC, Remington KA, Abril JF,
45
46 Gill J, Berman J, Rogers YH, Frazier ME, Scherer SW, Strausberg FL, Venter JC. 2007. The
47
48 diploid genome sequence of an individual human. *PLoS Biol* 5:e254.
49
50
51
52 Mégarbané A, Chouery E, van Baal S, Patrinos GP. 2006. The Lebanese National Mutation
53
54 frequency databases. *Eur J Hum Genet* 14(Suppl 1):365.
55
56
57 Metspalu A. 2002. Estonian Genome Project--before the take-off and take-off. *Bioinformatics* 18
58
59 Suppl 2:S152.
60

- 1
2
3 Nelson MR, Bryc K, King KS, Indap A, Boyko AR, Novembre J, Briley LP, Maruyama Y,
4
5 Waterworth DM, Waeber G, Vollenweider P, Oksenberg JR, Hauser SL, Sitrnadel HA,
6
7 Kooner JS, Chambers JC, Jones B, Mooser V, Bustamanta CD, Roses AD, Burns DK, Ehm
8
9 MG, Lia EH. 2008. The population reference sample, POPRES: a resource for population,
10
11 disease, and pharmacological genetics research. *Am J Hum Genet* 83:347–358.
12
13
14
15 Papachatzopoulou A, Vantarakis A, Kourakli A, Stavrou E, Frangou E, Patrinos GP,
16
17 Athanassiadou A. 2010. Documentation of the incidence of beta-thalassemia and *HBB*
18
19 mutation frequencies in South-Western Greece. *Hemoglobin* 34:333-342.
20
21
22 Patrinos GP. 2006. National/Ethnic mutation databases: Recording populations' genography.
23
24 *Hum Mutat* 27:879-887.
25
26
27 Patrinos GP, Brookes AJ. 2005. DNA, diseases and databases: Disastrously deficient. *Trends*
28
29 *Genet*, 21:333-338.
30
31
32 Patrinos G, Petricoin E. 2009. A new scientific journal linked to a genetic database: towards a
33
34 novel publication modality. *Hum Genomics Proteomics* 1:e597478.
35
36
37 Patrinos GP, Giardine B, Riemer C, Miller W, Chui DH, Anagnou NP, Wajcman H, Hardison
38
39 RC. 2004. Improvements in the HbVar database of human hemoglobin variants and
40
41 thalassemia mutations for population and sequence variation studies. *Nucleic Acids Res*
42
43 32:D537-D541.
44
45
46 Patrinos GP, van Baal S, Petersen MB, Papadakis MN. 2005. The Hellenic National Mutation
47
48 database: A prototype database for inherited disorders in the Hellenic population. *Hum*
49
50 *Mutat*, 25:327-333.
51
52
53 Povey S, Al Aqeel AI, Barash CI, Cambon-Thomsen A, Dalglish R, den Dunnen JT, Firth HV,
54
55 Greenblatt M, Parker M, Patrinos GP, Savige J, Sobrido MJ, Cotton RG. 2010. Practical
56
57 Guidelines addressing ethical issues pertaining to the curation of human Locus specific
58
59 variation databases (LSDBs). *Hum Mutat* (in press).
60

- 1
2
3 Saadallah AA, Rashed MS. 2007. Newborn screening: experiences in the Middle East and North
4
5 Africa. *J Inherit Metab Dis* 30:482–489.
6
7
8 Samara M, Chiotoglou I, Kalamaras A, Likousi S, Chassanidis C, Vagenas A, Vagenas C,
9
10 Eftichiadis E, Vamvakopoulos N, Patrinos GP, Kollia P. 2007. Large-scale population
11
12 genetic analysis for hemoglobinopathies reveals different mutation spectra in central Greece
13
14 compared to the rest of the country. *Am J Hematol* 82:634-636.
15
16
17 Singer PA, Daar AS. 2001. Harnessing genomics and biotechnology to improve global health
18
19 equity. *Science* 294:87–89.
20
21
22 Tadmouri GO, Al Ali MT, Al-Haj Ali S, Al Khaja N. 2006. CTGA: the database for genetic
23
24 disorders in Arab populations. *Nucleic Acids Res.* 34:D602-D606.
25
26
27 van Baal S, Kaimakis P, Phommavinh M, Koumbi D, Cuppens H, Riccardino F, Macek Jr M,
28
29 Scriver CR, Patrinos GP. 2007. FINDbase: a relational database recording frequencies of
30
31 genetic defects leading to inherited disorders worldwide. *Nucleic Acids Res* 35:D690–D695.
32
33
34 van Baal S, Zlotogora J, Lagoumintzis G, Gkantouna V, Tzimas I, Poulas K, Tsakalidis A,
35
36 Romeo G, Patrinos GP. 2010. ETHNOS: A versatile electronic tool for the development and
37
38 curation of National Genetic databases. *Hum Genomics* 4:361-368.
39
40
41 Wang J, Wang W, Li R, Li Y, Tian G, Goodman L, Fan W, Zhang J, Li J, Guo Y, Feng B, Li H,
42
43 Lu Y, Feng X, Liang H, Du Z, Li D, Zhao Y, Hu Y, Yang Z, Zheng H, Hellmann I, Inouye
44
45 M, Pool J, Yi X, Zhao J, Duan J, Zhou Y, Qin J, Ma L, Li G, Yang Z, Zhang G, Liu D, Lu
46
47 Z, Li N, Guo G, Zhang J, Ye J, Fang L, Hao Q, Chen Q, Liang Y, Su Y, San A, Ping C,
48
49 Yang S, Chen F, Li L, Zhou K, Zheng H, Ren Y, Yang L, Gao Y, Yang G, Li Z, Feng X,
50
51 Kristiansen K, Wong GK, Nielsen R, Durbin R, Bolund L, Zhang X, Li S, Yang M, Wang J.
52
53 2008. The diploid genome sequence of an Asian individual. *Nature* 456:60–65.
54
55
56
57
58 Wheeler DA, Srinivasan M, Egholm M, Shen Y, Chen L, McGuire A, He W, Chen YJ,
59
60 Makhijani V, Roth GT, Gomes X, Tartaro K, Niazi F, Turcotte CL, Irzyk GP, Lupski JR,

- 1
2
3 Chinault C, Song XZ, Liu Y, Yuan Y, Nazareth L, Qin X, Muzny DM, Marquies M,
4
5 Weinstock GM, Gibbs RA, Rothberg GM. 2008. The complete genome of an individual by
6
7 massively parallel DNA sequencing. *Nature* 452:872–876.
8
9
10 Wong ML, Chia KS, Wee S, Chia SE, Lee J, Koh WP, Shen HM, Thumboo J, Sofjan D. 2004.
11
12 Concerns over participation in genetic research among Malay-Muslims, Chinese and Indians
13
14 in Singapore: a focus group study. *Community Genet* 7:44–54.
15
16
17 Zlotogora J, van Baal S, Patrinos GP. 2007. Documentation of inherited disorders and mutation
18
19 frequencies in the different religious communities in Israel in the Israeli National Genetic
20
21 Database. *Hum Mutat* 28:944-949.
22
23
24 Zlotogora J, van Baal S, Patrinos GP. 2009. The Israeli National Genetic Database. *IMAJ* 6:373-
25
26
27 375.
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Box 1: Guidelines and recommendations for data capture in developing countries

1. Develop procedures to include clinicians and researchers in developing nations, mostly by creating networks among those of common interest.
2. Create international networks from closely related populations to enhance communication, interaction, and initiate research networks between them.
3. Foster communication, interaction, and research networks between developing and developed countries.
4. Ensure that the database management systems that are being used or developed can be utilized by those in a limited resource environment.
5. Provide support to developing countries to build capacity and to fully participate in the collection, analysis and sharing of genetic variation information.
6. Develop a framework to facilitate interactions between the coordinating centre and national, regional and international agencies.
7. Ensure that all ethical, legal, religious, and social issues are thoroughly considered when NEMDBs and/or data capture projects are launched in developing countries.

Table 1. Nation-wide genome and related projects (N/A: Not available).

Project	URL	Aims and scope	References
Nation-wide genome projects			
The Estonian Genome Project	http://www.geenivaramu.ee/index.php?lang=eng	To create a database of health, genealogical and genome data representing approximately 10% of Estonia's population	Metspalu, 2002
The MEDGENET Project	http://medgenet.tredueuno.it	To expand the human expertise in clinical genetics in Mediterranean Partner Countries (MPC) that share a common burden of genetic diseases.	van Baal et al., 2010
The Indian Genome Variation database	http://www.igvdb.res.in	To create a DNA variation database of the people of India and make it available to researchers for understanding human biology with respect to disease predisposition, adverse drug reaction, population migration	Indian Genome Variation Consortium, 2005
The Catalogue of Transmission Genetics in Arabs	http://www.cags.org.ae/ctga_search.html	To provide the user with clinical and molecular genetic descriptions of genetic disorders, epidemiological details of reports of the disorders in the Arab population and clinical case details wherever available	Tadmouri et al., 2006
Other related projects			
Kuwaiti Molecular Genetics Diagnostic Service	http://www.al-mulla.org	To provide autozygosity mapping services, Information on Arab genome and molecular karyotyping of cancers	N/A
UK diagnostic mutation database	http://ngri.org.uk/Manchester/dmudb.html	To provide a route for sharing mutation data within and between diagnostic laboratories in the United Kingdom to support genetic testing services for patients.	N/A
Hellenic Genome Variation database	http://www.goldenhelix.org/hellenic	To create a compendium for genetic diseases and to encourage genetic studies to elucidate the genetic basis of inherited diseases in Greece	Patrinos et al., 2005
The International Society for Gastrointestinal Hereditary Tumours (InSiGHT)	http://www.insight-group.org	To improve the quality of care of patients and their families with any condition resulting in hereditary gastrointestinal tumours	N/A
International Thalassemia Network (ITHANET)	http://www.ithanet.eu	To strengthen research and treatment of hemoglobinopathies	Lederer et al., 2009
Pharmacogenomics for Every Nation Initiative	http://www.pgeni.org	To determine the frequencies of the commonest pharmacogenomic markers in various populations, particularly from developing countries worldwide	N/A

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Recommendations for genetic variation data capture in developing countries to ensure a comprehensive worldwide data collection

George P. Patrinos ^{1*}, Jumana Al Aama ², Aida Al Aqeel ³, Fahd Al-Mulla ⁴, Joseph Borg ⁵,

Andrew Devereux ⁶, Alex E. Felice ⁷, Finlay Macrae ⁸, Makia J. Marafie ⁹, Michael B. Petersen

¹⁰ Ming Qi ¹¹, Rajkumar S. Ramesar ¹², Joel Zlotogora ¹³, and Richard G. Cotton ¹⁴

Deleted: 5

Deleted: 7

Deleted: 8

Deleted: 9

Deleted: 0

Deleted: 1

Deleted: 2

Deleted: 3

¹ University of Patras, School of Health Sciences, Department of Pharmacy, Patras, Greece;

² Department of Genetic Medicine, Faculty of Medicine and Princess Al-Jawhara Center of Excellence in Research of Hereditary Disorders, King Abdulaziz University, Jeddah, Saudi Arabia;

³ Department of Pediatrics, Riyadh Military Hospital and Stem Cell therapy program, King Faisal Specialist Hospital and Research Center, Riyadh, Saudi Arabia;

⁴ Molecular Pathology Unit, Kuwait University, Kuwait;

⁵ University of Malta, Faculty of Health Sciences, Department of Applied Biomedical Sciences, Msida, Malta;

Deleted: University of Malta, Department of Physiology and Biochemistry, Laboratory of Molecular Genetics, Msida, Malta;

⁶ NGRL Manchester, Department of Medical Genetics, St Mary's Hospital, Manchester, United Kingdom;

⁷ University of Malta, Department of Physiology and Biochemistry, Laboratory of Molecular Genetics, Msida, Malta;

Deleted: 7

⁸ Department of Colorectal Medicine and Genetics and University of Melbourne, The Royal Melbourne Hospital, Australia;

⁹ Clinical Genetics, Kuwait Medical Genetics Centre, Maternity Hospital, Sabah Medical Area,
Kuwait;

Deleted: 8

¹⁰ Institute of Child Health, Department of Genetics, Athens, Greece;

Deleted: 9

¹¹ Center for Genetic and Genomic Medicine, The First Affiliated Hospital of Zhejiang

Deleted: 0

University School of Medicine and James Watson Institute of Genomic Sciences, P. R. China;

¹² MRC Human Genetics Research Unit, Division of Human Genetics, Institute for Infectious

Deleted: 1

Diseases and Molecular Medicine, Faculty of Health Sciences, University of Cape Town, South

Africa;

¹³ Department of Community Genetics, Public Health Services, Ministry of Health, Jerusalem,

Deleted: 2

Israel;

¹⁴ Genomic Disorders Research Centre, Howard Florey Institute, Melbourne, Australia & Faculty

Deleted: 3

of Medicine, Dentistry and Health Sciences, The University of Melbourne, Australia

Running head: Recommendations for genetic variation data capture in developing countries

* Corresponding author at:

Department of Pharmacy, School of Health Sciences, University of Patras,

University Campus, Rion, GR-265 04, Patras,

Greece

Telephone/Fax: +30-2610-969.834,

E-mail: gpatrinos@upatras.gr

Abstract

Developing countries have significantly contributed to the elucidation of the genetic basis of both common and rare disorders, providing an invaluable resource of cases due to large family sizes, consanguinity and potential founder effects. Moreover, the recognised depth of genomic variation in indigenous African populations, reflecting the ancient origins of humanity on the African continent, and the effect of selection pressures on the genome, will be valuable in understanding the range of both pathological and non-pathological variations. The involvement of these populations in accurately documenting the extant genetic heterogeneity is more than essential. Developing nations are regarded as key contributors to the Human Variome Project (HVP; <http://www.humanvariomeproject.org>), a major effort to systematically collect mutations that contribute to or cause human disease and create a cyber infrastructure to tie databases together. However, biomedical research has not been the primary focus in these countries even though such activities are likely to produce economic and health benefits for all. Here, we propose several recommendations and guidelines to facilitate participation of developing countries in genetic variation data documentation, ensuring an accurate and comprehensive worldwide data collection. We also summarize a few well-coordinated genetic data collection initiatives that would serve as paradigms for similar projects.

Key words: Developing countries, national/ethnic mutation databases, populations, genetic variation

Introduction

The recent elucidation of the genome sequence of several individuals, such as Caucasians (Jim Watson [Wheeler et al., 2008] and Craig Venter [Levy et al., 2007]), West African [Bentley et al., 2008], and Han Chinese [Wang et al., 2008], suggests that a large number of SNPs and other sequence variation exists in the human population. Estimates from African genetic diversity and the Pan Asian SNP initiative suggest that approximately 80–90% of human genomic variation resides in the world's developing countries. It is therefore obvious that there is a huge need to comprehensively document the extant genetic variation in various ethnic groups and populations around the globe to ensure maximal data coverage.

Since 2002, there have been several initiatives to thoroughly document the genetic heterogeneity in various populations worldwide in National/Ethnic Mutation databases (NEMDBs), a large number of which reside in developing countries. Despite the fact that data content is rather homogeneous, facilitated by the existence of an off-the-shelf database management system [van Baal et al., 2010], these databases have been developed without appropriate guidelines and recommendations to ensure not only a thorough description of common and rare genetic diseases in these populations but also maximal visibility and usefulness of these resources in the respective countries while maintaining the individual local ethical and cultural values in them. Furthermore, developing countries are often a valuable resource of clinical samples that derive from consanguineous families with very unusual phenotypes. These genotype and phenotype data are worth being recorded in NEMDBs not only to make them readily available to interested parties but also to develop and implement genetic epidemiological approaches for estimating the relative frequency of rare autosomal recessive disorders and an estimate of their prevalence in these populations.

1
2 Here, we propose a number of guidelines and recommendations that would assist in
3 genetic variation data capture in developing countries to ensure a comprehensive worldwide data
4 collection, one of the aims and goals of the Human Variome Project [HVP; Cotton et al 2007;
5 Kaput et al., 2009; www.humanvariomeproject.org]. These guidelines are largely the result of
6 discussions among the participants of the developing countries working group during the
7 International HVP Planning and Implementation meetings in May 2008 and 2010, respectively,
8 coming from a wide range of developed and developing countries and extended, based on
9 ongoing data documentation efforts in these countries.
10
11
12
13
14
15
16
17

Deleted: respective

18 19 20 **Background**

21
22
23 Any formal attempt to identify the extent of genetic variation must include geographical
24 regions that have not been included in previous haplotype mapping projects, e.g., the HapMap
25 project, particularly since the latter project included racial and not population groups. Although
26 the Population Reference Sample (POPRES) will address some of these missing populations
27 [Nelson et al., 2008], this effort is again designed as a mapping project that is not expected to
28 focus on functional polymorphisms or mutations. Also, the Arab region has long played a key
29 role in discovering new genes and it is really awkward why this region has not been so actively
30 involved in today's genomics research and genome-wide association studies.
31
32
33
34
35
36
37
38

Deleted: i.

39 Hence, one of the main activities of the HVP is the analysis and subsequent inclusion of
40 genetic variation data from clinical samples originating from diverse populations or ethnic groups
41 [Kaput et al., 2009]. The distinct advantage of some ethnic populations is the opportunity to study
42 genetic diseases due to: (a) large family size, (b) consanguinity, and (c) potential founder effects
43 [Saadallah and Rashed, 2007]. In fact, several disorders have been described for the first time in
44
45
46
47
48
49

1
2 certain populations residing in developing countries e.g Sanjad-Sakatti syndrome, Teebi-Shaloot
3 syndrome, Teebi Hypertelorism and others. Based on the abovementioned advantages,
4
5 developing nations are regarded as major contributors to the HVP even though biomedical
6
7 research has not been the focus of these countries, despite such activities being likely to produce
8
9 economic and health benefits for all [Singer and Daar, 2001; Kaput et al., 2009].

11 The NEMDBs are one of the main components of the HVP that could accommodate the
12
13 extant genetic heterogeneity in the developing countries and, for that matter, all countries. In
14
15 particular, the NEMDBs are mutation depositories needing continuous updating, recording
16
17 extensive information over the described genetic heterogeneity of an ethnic group or population
18
19 [reviewed in Patrinos, 2006]. NEMDBs can help optimize national molecular diagnostic services,
20
21 by stratifying the corresponding mutation spectra in various inherited diseases, and assist in
22
23 interpreting diagnostic test results in countries with heterogeneous populations, particularly in
24
25 minority ethnic groups. NEMDBs can also contribute toward the elucidation of the origin and
26
27 migration of populations, acting as platforms for comparative genomic studies that can
28
29 reciprocally provide insights into, e.g., the demographic history of human populations, patterns of
30
31 their migration and admixture, gene/mutation flow, *etc* [Patrinos, 2006]. Such comprehensive
32
33 databases would include information on genotype and reasonable phenotype, in a level of detail
34
35 that should be agreed by the community, which would also enable those interested in these
36
37 diseases to investigate the possible influence of geographical environment or genetic modifiers
38
39 on expression of the disease/phenotype. Examples include the genetics of obesity, hypertension
40
41 or type II diabetes. Relevant genotype/phenotype collection projects, such as GEN2PHEN
42
43 (<http://www.gen2phen.org>) or P3G (<http://www.p3g.org>) could serve as models to establish the
44
45 necessary requirements for the depth of the recorded genotype and phenotype information.

Deleted: detail of

Formatted: Indent: First line: 0 pt

Recommendations

Formatted: Font: 14 pt, Bold

Formatted: Font: 14 pt

Formatted: Font: 14 pt, Bold

Formatted: Font: 14 pt, Bold

The recommendations and guidelines that follow ([summarized in Box 1](#)) are intended to facilitate genetic variation data capture in developing countries, either by suggesting specific measures aiming to enhance awareness among clinicians, bioscientists, and the general public about the range of commonest genetic disorders suffered by certain populations and/or ethnic groups, or by sharing the experience from ongoing projects to document genetic heterogeneity in different countries.

1. Develop procedures to include clinicians and researchers in developing nations, mostly by creating networks among those of common interest.

This would be particularly important for clinicians/researchers working in under-resourced institutes and research environments. Where construction of a NEMDB is decided, selected representatives from participating laboratories, mainly established scientists and principal investigators involved in human genetics should take the lead in this effort. These individuals would form the advisory committee of the NEMDB with one of them assigned as national coordinator with a defined (e.g. yearly) term. These members will then assign eligible members from their groups that would serve as data curators responsible for data entry/modification.

Such efforts would flourish if placed under the umbrella of specific scientific initiatives or consortia (see also below), such as the Human Genome Variation Society (HGVS; www.hgvs.org), a Society which generates recommendations and guidelines towards human genome variation capture, the HVP and ORPHANET (<http://www.orpha.net>), a database of information on rare diseases and orphan drugs, with presence in 37 mostly European and Arab

1 countries, that aims to contribute to the improvement of the diagnosis, care and treatment of
2 patients with rare diseases. This approach is already adopted in the Israeli and Greek genetics
3 communities, mostly to share genetic laboratory information, but also to accelerate the data
4 collection process. Partnering with local human genetic societies is also liable to have the same
5 result.
6
7
8
9

10
11 Moreover, in order to increase awareness among interested parties, the resulting
12 NEMDBs should be made known to their potential users using various communication channels.
13 These do not necessarily include international scientific journals to which local institutes in these
14 countries might not have access, but local scientific journals, news bulletins, scientific events
15 (symposia, workshops), one-to-one communication, and so on. Such an approach was followed
16 by the Israeli NEMDB, by producing a review article in the Israeli Medical Association Journal, a
17 leading peer-reviewed scientific medical journal in Israel [Zlotogora et al., 2009].
18
19
20
21
22
23
24

25
26
27 **2. Create international networks from closely related populations to enhance**
28 **communication, interaction, and initiate research networks between them.**
29

30
31 In this case, it would be mutually beneficial for closely related countries such as in Latin
32 America, the Middle East or regions in Africa, to initiate new or strengthen existing collaborative
33 ties towards comprehensive genetic variation data capture. Similarly, specific resources would be
34 committed in different countries so that there will be no redundancy in their research efforts. In
35 this regard, the Catalogue for Transmission Genetics in Arabs (CTGA; Tadmouri et al., 2006; see
36 also below) initiated from the Centre for Arab Genomic Studies (<http://www.cags.org.ae>) in
37 Dubai, United Arab Emirates, is a pilot project to document genetic disorders in Arab populations
38 from the published literature. The current version of CTGA database contains nearly 760 full-text
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Deleted: the

1 records for nearly 1,290 Mendelian disorders and related genes, including extensive data from the
2
3 United Arab Emirates, Bahrain, Oman, and Qatar.
4
5
6

7 **3. Foster communication, interaction, and research networks between developing and**
8 **developed countries.**
9

10 This would be particularly beneficial in both data collection from the literature or
11 generation of new genetic variation data and would include:
12
13

- 14 (a) Education of healthcare providers, the public, and government officials from researchers and
15 health care providers from developed countries, regarding data collection, data generation and
16 interpretation, and so on. To this end, the Pan-Arab Human Genetics conferences, with
17 international participation from developed countries, could very well be an example of such
18 interactions. The African Society for Human Genetics is also spearheading efforts to create a
19 database of genetic variation in indigenous African populations (Raj Ramesar; personal
20 communication; Al Aama and co-workers, submitted)
21
22 (b) Assistance in data handling and storage in large databases and servers in developed countries,
23 particularly where massive storage units (e.g. GRIDs) are needed to accommodate large
24 datasets. Such possibilities are currently offered by LSDB and NEMDB software developers
25 (LOVD [Fokkema et al., 2005] and ETHNOS [van Baal et al., 2010], respectively) in the
26 Netherlands and Greece, respectively, to either download for use or to store in their local
27 servers, data from interested parties and curators that do not have either the knowledge or the
28 resources needed, although ultimately the aim would be to provide the infrastructure locally.
29
30 (c) Consultancy in the design and coordination of large-scale genome projects (see also below) if
31 such need arises, including patients recruitment, experimental design, software/database
32 development (see also Point 5),
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49

- 1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
- (d) Outsourcing of high-throughput experiments, such as microarray or next-generation sequencing experiments, to laboratories in the developed countries, where such infrastructure and expertise is more likely to be available, but also provide expert external training to those countries where the facilities are available but outstrip the local expertise.
 - (e) To work interactively, leading from all of the above, to train postgraduate students and other researchers towards capacity development in developing countries.

4. Ensure that the database management systems that are being used or developed can be utilized by those in a limited resource environment.

Deleted: s

Nowadays, LSDBs and NEMDBs, which capture the core information in any one of these domains of information, are developed by experts and curated with invaluable skill and experience. Often, developing countries suffer from a limited resource research environment; hence this reality should be carefully considered when database management systems are developed for this purpose.

Therefore, the database management system should be designed and built such that the resulting database would:

- (a) Record the specific variants of inherited disorders found in a population or ethnic group in a comprehensive and user-friendly manner, i.e., by providing consistent and simplified information accompanied by numerous links to other databases and related resources [Patrinos, 2006],
- (b) Provide the end user with an interface for easy data query and retrieval of information pertaining to the mutation spectrum of each disorder together with the corresponding mutation frequencies and references,

- 1
2 (c) Allow permanent data storage and secure data entry and modification. In this case, software
3 pipelines should also allow automated genotype data transfer between the clinic and/or
4 laboratory to the database, and ideally from the genotyping equipment (sequencer, scanner,
5 *etc*) directly to the database,
6
7
8
9
10 (d) Operate under strict quality assurance guidelines, such as intellectual property rights, the
11 professional code of ethics (<http://ihealthcoalition.org/ethics/ehcode.html>) or specific ethical
12 guidelines [Povey et al., 2010], laws on computing and liberties (e.g., [www.ama-](http://www.ama-assn.org/ama/pub/category/1905.html)
13 [assn.org/ama/pub/category/1905.html](http://www.ama-assn.org/ama/pub/category/1905.html)), electronic data protection, specific guidelines and any
14 law or regulation applicable in addition to the local ethical issues that may be specific to a
15 given country.
16
17
18
19
20

21 Such databases should be able to operate remotely, should the required support not be
22 available locally (see also Point 3)
23
24
25
26

27
28 **5. Provide support to developing countries to build capacity and to fully participate in the**
29 **collection, analysis and sharing of genetic variation information.**
30

31 Comprehensive genetic variation data capture usually entails highly specialized
32 researchers or diagnostic laboratory staff and state-of-the-art infrastructure, particularly where
33 generation of new genomic data comes to play. Developing countries usually suffer from brain
34 drain and lack of highly sophisticated equipment to perform cutting-edge research. To this end,
35 developed countries have the obligation to alleviate this burden and to assist developing nations
36 by every means possible [Kaput et al., 2009]. It should be noted that developing countries have a
37 valuable resource of clinical samples, often with very unusual phenotypes, particularly when
38 consanguineous families are involved since in this case, the elucidation of the genetic basis of the
39 underlying phenotype becomes much easier. Developing countries should therefore:
40
41
42
43
44
45
46
47
48
49
50

- 1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
- (a) Educate healthcare providers and government officials to demonstrate the need to initiate documentation of the genetic variation in their countries and to underline the benefits of cooperation in biomedical research [Bhan et al., 2007],
 - (b) Increase awareness in the general population about the need to comprehensively record the underlying genetic heterogeneity in their country and informing them objectively about the benefits and risks of such projects,
 - (c) Ascertain all parties involved that the research outcome will not be misused. Certain populations or ethnic groups in developing countries usually mistrust research involving genetic analyses or fear that results can stigmatize them or worse (<http://www.eubios.info/ASIAE/BIAE201.htm>). For example, Malay-Muslims, Chinese, and Indians in Singapore expressed anxiety about breach of confidentiality, the misuse of their DNA for cloning, and possibilities of being diagnosed with disease [Wong et al., 2004]. Community-based research collaborations may provide fora for addressing cultural and ethical concerns of biomedical research [Povey et al., 2010]. Analyzing the extent of human genome variation creates an ideal opportunity for the developed and the developing nations of the world to forge meaningful partnerships and to work together in an unprecedented way, initially to identify variation causing disease, and then to simplify mutation detection in specific ethnic groups, and then to understand how genetic variation contributes to human phenotypic diversity. By ensuring that all nations and ethnic groups have an equal and fair opportunity to share data and technology, evidence-based information will be provided such that all populations can benefit from a global society health network.
 - (d) Emphasize in every stage of the collaborative project that external input is needed from the design and coordination of NEMDB development and curation to database hosting and technical support, patient recruitment, and experimental design to performing high-

1
2 throughput experiments in the case of large-scale genome projects (see also Point 3). To this
3
4 end, funding agencies (e.g. UKIERI; UK-India Education and Research Initiative) could
5
6 assist by funding innovative projects so that developing countries become independent and
7
8 self-sustained.

- 9
10 (e) Develop partnerships for the equitable sharing of knowledge and technology between
11
12 developed and developing countries. These partnerships should be ideally formed under the
13
14 umbrella of international organizations, such as the HGVS or the Human Genome
15
16 Organization (HUGO; <http://www.hugo-international.org>) that would also organize inter-
17
18 institutional specialized educational programmes and workshops. The biennial mutation
19
20 detection workshop would be a paradigm for this initiative and may sometimes be held in
21
22 developing countries.

Deleted: -annual

23
24
25 **6. Develop a framework to facilitate interactions between the coordinating centre and**
26
27 **national, regional and international agencies.**
28

29
30 Researchers and their academic institutions that participate in national data collection
31
32 projects and/or NEMDBs development should jointly develop infrastructures that would ensure
33
34 smooth communication between them and national and international agencies. These include: (a)
35
36 national governments and their Departments of Health, Science and Technology, Development,
37
38 Education and so on, local human genetic societies, (b) regional organizations, such as the
39
40 NEPAD (New Programme for Africa's Development) and (c) international organizations, such as
41
42 the European Commission (EC), United Nations (UN), United Nations Children's Fund
43
44 (UNICEF), World's Health Organization (WHO), Organization for Economic Cooperation and
45
46 Development (OECD), *etc.*

1
2 This in turn would: (a) create new collaboration ties with neighbouring countries
3 interested in initiating similar projects, (b) increase the chances of receiving funding for these
4 activities (see also below), (c) increase visibility of the progress to the international scientific
5 community.
6
7
8
9

10
11 **7. Ensure that all ethical, legal, religious, and social issues are thoroughly considered when**
12 **NEMDBs and/or data capture projects are launched in developing countries.**
13
14

15 The goal of establishing LSDBs and NEMDBs in developing countries is the sharing of
16 genetic variation information for the benefit of the respective populations. This includes the
17 protection of privacy, which in this context is the right of the individual and members of their
18 family to be protected against intrusion into their personal information and further intrusions
19 ensuing from access to this, by publication of information. Of course, there is no such issue when
20 summary level data are reported for a population, as it is the case in some NEMDBs. Therefore,
21 there should be a balance between the public's interest in the value of the shared information and
22 its interest in the strict protection of privacy. This balance would be viewed differently in
23 different cultures and ethnicities [Al-Aqeel, 2007].
24
25
26
27
28
29
30
31
32

33 For many of these NEMDBs, it seems reasonable that an independent group of qualified
34 scientists oversees these resources and related activities not only at their initiation but on an
35 ongoing basis. Where needed, the main purpose of the particular NEMDB should be clarified,
36 recognising that this may change with time. This will allow evaluation of the exact requirements
37 for data deposition in the NEMDBs, and whether compliance with the remaining guidelines will
38 be possible with the database content freely available to the public or partly restricted to
39 registered persons, namely curators, physicians, and so on. This measure was implemented in the
40 Israeli NEMDB, where not all information is made publicly available [Zlogotora et al., 2007],
41
42
43
44
45
46
47
48
49
50

1 since access to specific cases is password-protected [Zlotogora et al., 2009]. A similar measure
2
3 would entail anonymizing genetic data, either by permanently removing or coding the personal
4
5 identifiers [Povey et al., 2010]. Of interest, the newest version of the LOVD software [Fokkema
6
7 et al., 2005] has the option to store data that are not public but that can be queried. The result of a
8
9 query hitting non-public data is a notification that there is such information in the database but
10
11 that the curator needs to be contacted to get more information.
12

13
14 The depth of the ethnic and geographic origin data should be determined in advance,
15
16 particularly in the case of small villages where carriers of rare variants could be identified even if
17
18 personal information were not provided. In this case, data donors should be well informed and
19
20 provided with an explanation as to why their data documentation is useful and how their privacy
21
22 is secured. These individuals should consent to the inclusion of their genetic information in the
23
24 database and reserve the right to withdraw their data from the database whenever they desire.
25
26 This is particularly important for persons unable to consent because of disability or young age.
27
28 Standardized consent forms need to be translated to different local languages and/or dialects.
29

30
31 Finally, these databases would also benefit from some independent/external ethical
32
33 review, perhaps from an international group (HGVS or HVP) who could advise them
34
35 accordingly.
36

37 **Nation-wide genomic projects that would serve as paradigm for genetic** 38 39 **variation data collection** 40 41

42
43
44 By definition, a nation-wide genomic project is a national multicentre initiative to
45
46 determine the underlying genetic heterogeneity of a population with the goal to ascertain the
47
48 genetic variation leading to common or rare inherited diseases. Such initiatives consist of three
49

1 parts: (a) documenting the genetic heterogeneity that has already been reported in the published
2 literature, (b) generating new genetic data by screening a large proportion of the population using
3 high throughput genotyping approaches, and (c) accumulating data causing inherited disease
4 from hospitals, laboratories and other healthcare facilities. In every case, except perhaps (c), a
5 large national genetic database/portal is being developed to accommodate the sheer amount of
6 data gathered. To date, there are few nation-wide genomic projects in various developing
7 countries that are currently in progress. These projects (summarized in Table 1) implement some
8 of the guidelines described previously. Also, there are several other NEMDB projects, all of
9 which more or less follow these guidelines, that have already been reviewed [Patrinos, 2006; van
10 Baal et al., 2007] and will, therefore, not be further discussed since they lie outside the scope of
11 the present article.

12
13
14
15
16
17
18
19
20
21
22
23
24 The Estonian Genome Project (EGP; <http://www.geenivaramu.ee/index.php?lang=eng>) is
25 a research venture of the University of Tartu initiated in 2001 [Metspalu, 2002] and aims to
26 create a database of health, genealogical and genetic data representing approximately 10% of
27 Estonia's population. The database will make it possible for researchers both in Estonia and
28 outside to look for links between genes, environmental factors and common diseases (such as
29 cancer, diabetes, depression, cardio-vascular diseases, *etc*). Currently, EGP gene bank contains
30 data contributed by over 20,000 donors with an expected increase to 100,000 within the next
31 three or four years.

32
33
34
35
36
37
38
39
40 The Indian Genome Variation database (IGVdb; <http://www.igvdb.res.in>) is a consortium
41 activity established in 2003 to document the inherent genetic variability of the Indian
42 subpopulations as a first step towards identifying genetic biomarkers to understand susceptibility
43 for any genetic disease and population migration, pharmacogenomic markers to understand
44 variable drug response. The Indian Genome Variation (IGV) consortium aims to provide data on
45
46
47
48
49

1 validated SNPs and repeats, both novel and reported, along with gene duplications, in over 1,000
2 genes (selected on the basis of their relevance as functional and positional candidates in many
3 common diseases including genes relevant to pharmacogenomics), in 15,000 individuals drawn
4 from Indian subpopulations [Indian Genome Variation Consortium, 2005].
5
6
7
8

9
10 Kuwait established within the Faculty of Medicine, Department of Pathology a Molecular
11 Genetics Diagnostic Service Division (<http://www.al-mulla.org>), integrated within a research
12 laboratory, which is focused on delivering state-of-the-art diagnostic service for the Kuwaiti
13 population. These services, which focus on autozygosity mapping in families with
14 consanguineous marriages, provide the research laboratory with a unique opportunity to identify
15 novel genes in autosomal recessive disorders. Moreover, various unique cases suffering from
16 cancer are currently being analysed using the Affymetrix Array 6.0 microarray technology to
17 identify copy number changes and homozygosity stretches. The data generated from these studies
18 are deposited in CAGS database (see above), and LOVD, while the Kuwaiti NEMDB has been
19 recently established databases to encourage data storage for genetic diseases in different
20 subpopulations in Kuwait.
21
22
23
24
25
26
27
28
29
30

31 The United Kingdom has developed a central database to collect all UK genome variation
32 data, and their effects, from participating diagnostic laboratories. The diagnostic mutation
33 database (DMuDB; <http://ngri.org.uk/Manchester/dmudb.html>) provides a route for sharing
34 mutation data within and between diagnostic laboratories in the United Kingdom to support
35 genetic testing services for patients. Currently DMuDB has over 4,884 records for 33 genes,
36 namely Breast Cancer (*BRCA1*, *BRCA2*), Lowe Syndrome (*OCRL*), Sotos Syndrome (*NSD1*),
37 HNPCC (*MLH1*, *MSH2*), Cystic Fibrosis (*CFTR*), X-Linked Retinitis Pigmentosa (*XLRP*),
38 Neurofibromatosis, Types I and II (*NF1*, *NF2*), Alstrom Syndrome (*ALMS1*), CADASIL
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Deleted: the

Deleted: some time ago

Deleted: diagnostic and its effect from

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

(*NOTCH3*), Familial adenomatous polyposis (*FAP*) and Muscular Dystrophy (*DMD*), containing over 10,985 individual variants.

The Hellenic Genome Variation database aims to document the genetic heterogeneity in the Hellenic population. The first version of the Hellenic NEMDB was created in 2004 [Patrinos et al., 2005]. In 2007, research institutes involved in hemoglobinopathies research initiated a pilot multicenter project to determine the mutation spectrum of the thalassemia syndromes in various geographical regions in the country with the goal to document the findings in the Hellenic NEMDB. These data showed that although Greece is a rather homogeneous country, there are big qualitative (type of mutations) and quantitative (allelic frequencies) differences in various parts of the country, namely central Greece [Samara et al., 2007] and South-Western Greece [Papachatzopoulou et al., 2010]. These data compared to findings from North-Western Greece [Georgiou et al., 2003] and the rest of the country [Patrinos et al., 2005] reveal valuable information for region-specific population carrier screening and for efficient provision of prenatal diagnosis of β -thalassemia in Greece.

Finally, there are four international consortia with activities related to the documentation of causative mutations and allelic frequencies in developing countries:

- (a) The International Society for Gastrointestinal Hereditary Tumours (InSiGHT; <http://www.insight-group.org>) is an international multidisciplinary, scientific organisation with a mission "...to improve the quality of care of patients and their families with any condition resulting in hereditary gastrointestinal tumours." The InSiGHT consortium maintains a LOVD-based database for genetic variations for the *EPCAM*, *MLH1*, *MLH3*, *MSH2*, *MSH6*, *MUTYH*, *PMS1*, and *PMS2* genes leading to gastrointestinal tumours and encourages the establishment of national repositories for patients suffering from gastrointestinal tumours, particularly in developing countries.

- 1
2 (b) The International Thalassemia Network (ITHANET; <http://www.ithanet.eu>) is a European
3
4 Commission-funded project that comprises a network of all major European research
5
6 institutions active in the field of molecular and clinical research of thalassemia and related
7
8 hemoglobinopathies [Lederer et al., 2009]. Among its activities, ITHANET hosts a database
9
10 (ITHANETbase) that includes mutation frequencies in various populations at-risk for
11
12 thalassemia, particularly from developing countries. In many instances though, this repository
13
14 is largely redundant to HbVar database of hemoglobin variants and thalassemia mutations
15
16 [<http://globin.bx.psu.edu/hbvar>; Patrinos et al., 2004].
17
- 18 (c) The MEDGENET project (<http://medgenet.tredueuno.it>) is a European Consortium, funded
19
20 by the European Commission, aiming to expand the human expertise in clinical genetics in
21
22 Mediterranean Partner Countries (MPC) through the transfer of knowledge and technology
23
24 between the two rims of the Mediterranean which share a common burden of genetic
25
26 diseases. Among the actions foreseen by MEDGENET is the building of Medical Genetics
27
28 databases for the various partner countries. Several NEMDBs are built from scratch, such as
29
30 the Tunisian and Egyptian NEMDBs or expanded from the existing NEMDBs, such as the
31
32 Israeli [Zlotogora et al., 2007], Cypriot [Kleanthous et al., 2006], and Lebanese NEMDBs
33
34 [Megarbane et al., 2007], using the ETHNOS database management system for NEMDB
35
36 development and curation [van Baal et al., 2010].
37
- 38 (d) The Pharmacogenomics for Every Nation Initiative (PGENI; www.pgeni.org) that aims to
39
40 determine the frequencies of the commonest pharmacogenomic markers in various
41
42 populations, particularly from developing countries worldwide.
43

44
45 Finally, there are a few nation-wide DNA biobanks, such as the National Laboratory for
46
47 the Genetics of Israeli Populations, representing a national repository for human cell-lines and
48
49
50
51
52
53
54
55
56
57
58
59
60

1 DNA from the unique and large ethnic variation of the Israeli populations (Tel Aviv University;
2 available at <http://www.tau.ac.il/medicine/NLGIP>) or the Biomaterial Banking Medical Research
3 Initiative (BBMRI) in Malta, which of course lie at the borderline of the nation-wide genomic
4 projects described above.
5
6
7
8
9

10 **Funding and incentives**

11
12
13
14
15
16 There are currently two main obstacles that hinder further efforts to expand the NEMDB
17 field, namely funding and providing incentives to data submitters,
18
19

20 Funding for collecting genetic variation data has traditionally been difficult due to the
21 extreme fragmentation of the field [Patrinos and Brookes, 2005]. Therefore, it would be expected
22 that efforts to receive funding for genetic variation data capture are more likely to flourish if
23 treated as concerted efforts. GEN2PHEN (<http://www.gen2phen.org>) represents such initiative
24 that have already received funding from the EC and aims to unify human and model organism
25 genetic variation databases towards increasingly holistic views into Genotype-To-Phenotype
26 data, and to link this system into other biomedical knowledge sources via genome browser
27 functionality. Linking population and ethnic-specific information into LSDBs and diagnostic
28 databases, creating NEMDBs and developing genomic databases, similar to the Indian Genome
29 Variation database are key activities for the main GEN2PHEN work-packages. Obviously, other
30 strategically minded funding agencies (e.g. NIH), charities, and foundations would also launch
31 similar calls. Recently the Australian Node of the HVP has been funded to pilot software for data
32 flow from hospitals/clinics via country specific databases to LSDBs and NCBI/EBI/UCSC
33 browsers (unpublished). Involving universal organizations such as the WHO and UNESCO in the
34 fund-raising project will clearly have a strong impact.
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49

1
2 On the other hand, the development of comprehensive analyses of human genetic
3 variation often reaches a barrier in the form of the reward system for clinicians and academic
4 researchers. First of all, database entries could be made a mandatory quality control standard for
5 clinical laboratories and clinicians. Notably, according to a new initiative of the Israeli
6 government, genetic laboratories are mandated to report the details of the tests they are
7 performing in the Israeli NEMDB in order to obtain accreditation from the Israeli Ministry of
8 Health [van Baal et al., 2010].
9

10 For researchers, a publication or Web-based system establishing microattribution and
11 community annotation of mutations and cited data will enable measurable contributions to the
12 scientific knowledge base [Anonymous, 2008; Hoffmann, 2008]. Similarly, database-journals
13 would also serve this task by providing a forum for publishing gene variation data that would be
14 eventually deposited in the PubMed literature database. One such journal, the open access Human
15 Genomics and Proteomics (<http://www.sage-hindawi.com/journals/hgp>), which is affiliated with
16 FINDbase [<http://www.findbase.org>; van Baal et al., 2007], will focus on studies characterizing
17 causative mutation and/or biomarker frequency spectra. Accepted contributions including
18 datasets will be linked in FINDbase and deposited in PubMed [Patrinos and Petricoin, 2009].
19 Also, the journal *Human Genetics* has been accepting mutation data and this is linked to HGMD
20 (<http://www.hgmd.org>).
21

22 Journals, tenure, and promotion committees, and funding agencies would be encouraged
23 to cite these contributions to and citations of LSDB and international national databases [Patrinos
24 and Brookes, 2005]. The rapid progress in genetic research has led to a declining interest of
25 scientific clinical genetic journals in case reports with single or few novel mutations. This may
26 deter clinicians from reporting these new mutations, which collectively become important, unless
27 another venue becomes visible.
28

Concluding remarks

As stated above, documenting population/ethnic-specific data in developing countries are among the primary objectives of the HVP that appreciates the genomic sovereignty/equality for all countries to be involved and acknowledges the value of “human capital” within all populations. Real and tangible benefits of the HVP to improve health will be generated for participating populations. Hence, the voluntary participation of the greatest number of countries would ensure a more general applicability, and it is hoped that many countries will eventually decide to participate.

Acknowledgements

We thank the members of the developing countries working group at the HVP planning and implementation meetings for lively and helpful discussions. Part of this work has received funding from the European Commission [ITHANET (FP6-026539), EuroGenTest (FP6-512148), MEDGENET (FP6-31968) and GEN2PHEN (FP7-200754) projects], and the Kuwait Foundation for the advancement of Sciences (KFAS) grant number 2006-1302-07.

References

Al Aqeel AI. 2007. Islamic ethical framework for research into and prevention of genetic diseases. *Nat Genet* 39:1293-1298.

- 1
2 Anonymous. 2008. Human variome microattribution reviews. *Nat Genet* 40:1.
- 3
4 Bentley DR, Balasubramanian S, Swerdlow HP, Smith GP, Milton J, Brown CG, Hall KP, Evers
5
6 DJ, Barnes CL, Bignell HR, Boutell JM, Bryant J, et al. 2008. Accurate whole human
7
8 genome sequencing using reversible terminator chemistry. *Nature* 456:53–59.
- 9
10 Bhan A, Singh JA, Upshur RE, Singer PA, Daar AS. 2007. Grand challenges in global health:
11
12 engaging civil society organizations in biomedical research in developing countries. *PLoS*
13
14 *Med* 4:e272.
- 15
16 Cotton RG, Al Aqeel AI, Al-Mulla F, Carrera P, Claustres M, Ekong R, Hyland VJ, Marafie MJ,
17
18 Paalman MH, Patrinos GP, Qi M, Ramesar RS, Scott RJ, Sijmons RH, Sobrido MJ, Vihinen
19
20 M. (2009). Capturing All Disease Causing Mutations For Clinical And Research Use:
21
22 Towards An Effortless System For The Human Variome Project. *Genet Med*, 11:843-849.
- 23
24 Cotton RG, Appelbe W, Auerbach AD, Becker K, Bodmer W, Boone DJ, Boulyjenkov V,
25
26 Brahmachari S, Brody L, Brookes A, Brown AF, Byers P, Cantu JM, Cassiman JJ, Claustres
27
28 M, Concannon P, Cotton RG, den Dunnen JT, Flicek P, Gibbs R, Hall J, Hasler J, Katz M,
29
30 Kwok PY, Laradi S, Lindblom A, Maglott D, Marsh S, Masimirembwa CM, Minoshima S,
31
32 de Ramirez HZ, Scriver CR, Sherry S, Shimizu N, Shimizu N, Stein L, Tadmouri GO, Taylor
33
34 G, Watson M. 2007. Recommendations of the 2006 Human Variome Project meeting. *Nat*
35
36 *Genet* 39:433–436.
- 37
38 Fokkema IF, den Dunnen JT, Taschner PE. 2005. LOVD: easy creation of a locus-specific
39
40 sequence variation database using an "LSDB-in-a-box" approach. *Hum Mutat* 26:63-68.
- 41
42 Georgiou I, Makis A, Chaidos A, Bouba I, Hatzi E, Kranas V, Zilidis C, Bourantas KL. 2003.
43
44 Distribution and frequency of beta-thalassemia mutations in northwestern and central Greece.
45
46 *Eur J Haematol* 70:75-78.
- 47
48
49
50
51
52
53
54
55
56
57
58
59
60

- 1
2 Hinds DA, Stuve LL, Nilsen GB, Halperin E, Eskin E, Ballinger DG, Frazer KA, Cox DR. 2005.
3
4 Whole-genome patterns of common DNA variation in three human populations. *Science*
5
6 307:1072–1079.
7
8 Hoffman R. 2008. A wiki for the life sciences where authorship matters. *Nat Genet* 40:1047–
9
10 1051.
11
12 Indian Genome Variation Consortium. 2005. The Indian Genome Variation database (IGVdb): a
13
14 project overview. *Hum Genet* 118:1-11.
15
16 Kaput J, Cotton RG, Hardman L, Watson M, Al Aqeel AI, Al-Aama JY, Al-Mulla F, Alonso S,
17
18 Aretz S, Auerbach AD, Bapat B, Bernstein IT, Bhak J, Bleoo SL, Blöcker H, Brenner SE,
19
20 Burn J, Bustamante M, Calzone R, Cambon-Thomsen A, Cargill M, Carrera P, Cavedon L,
21
22 Cho YS, Chung YJ, Claustres M, Cutting G, Dagleish R, den Dunnen JT, Díaz C,
23
24 Dobrowolski S, Dos Santos MR, Ekong R, Flanagan SB, Flicek P, Furukawa Y, Genuardi M,
25
26 Ghang H, Golubenko MV, Greenblatt MS, Hamosh A, Hancock JM, Hardison R, Harrison
27
28 TM, Hoffmann R, Horaitis R, Howard HJ, Barash CI, Izagirre N, Jung J, Kojima T, Laradi S,
29
30 Lee YS, Lee JY, Gil-da-Silva-Lopes VL, Macrae FA, Maglott D, Marafie MJ, Marsh SG,
31
32 Matsubara Y, Messiaen LM, Möslin G, Netea MG, Norton ML, Oefner PJ, Oetting WS,
33
34 O'Leary JC, de Ramirez AM, Paalman MH, Parboosingh J, Patrinos GP, Perozzi G, Phillips
35
36 IR, Povey S, Prasad S, Qi M, Quin DJ, Ramesar RS, Richards CS, Savige J, Scheible DG,
37
38 Scott RJ, Seminara D, Shephard EA, Sijmons RH, Smith TD, Sobrido MJ, Tanaka T,
39
40 Tavtigian SV, Taylor GR, Teague J, Töpel T, Ullman-Cullere M, Utsunomiya J, van Kranen
41
42 HJ, Vihinen M, Webb E, Weber TK, Yeager M, Yeom YI, Yim SH, Yoo HS. 2009. Planning
43
44 the Human Variome Project. The Spain report. *Hum Mutat* 30:496-510.
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

- 1
2 Kleanthous M, Patsalis PC, Drousiotou A, Motazacker M, Christodoulou K, Cariolou M, Baysal
3 E, Khrizi K, Pourfarzad F, Moghimi B, van Baal S, Deltas CC, Najmabadi S, Patrinos GP.
4
5 2006. The Cypriot and Iranian National Mutation databases. *Hum Mutat* 27:598-599.
6
7 Lederer CW, Basak AN, Aydinok Y, Christou S, El-Beshlawy A, Eleftheriou A, Fattoum S,
8
9 Felice AE, Fibach E, Galanello R, Gambari R, Gavrilu L, Giordano PC, Grosveld F,
10
11 Hassapoulou H, Hladka E, Kanavakis E, Locatelli F, Old J, Patrinos GP, Romeo G, Taher
12
13 A, Traeger-Synodinos J, Vassiliou P, Villegas A, Voskaridou E, Wajcman H, Zafeiropoulos
14
15 A, Kleanthous M. 2009. An electronic infrastructure for research and treatment of the
16
17 thalassemias and other hemoglobinopathies: the Euro-mediterranean ITHANET project.
18
19 *Hemoglobin* 33:163-176.
20
21 Levy S, Sutton G, Ng PC, Feuk L, Halpern AL, Walenz BP, Axelrod N, Huang J, Kirkness EF,
22
23 Denisov G, Lin Y, MacDonald JR, Pang AW, Shago M, Stockwell TB, Tsiamouri A, Bafna
24
25 V, Bansal V, Kravitz SA, Busam DA, Beeson KY, McIntosh TC, Remington KA, Abril JF,
26
27 Gill J, Berman J, Rogers YH, Frazier ME, Scherer SW, Strausberg FL, Venter JC. 2007. The
28
29 diploid genome sequence of an individual human. *PLoS Biol* 5:e254.
30
31 M egarban e A, Chouery E, van Baal S, Patrinos GP. 2006. The Lebanese National Mutation
32
33 frequency databases. *Eur J Hum Genet* 14(Suppl 1):365.
34
35 Metspalu A. 2002. Estonian Genome Project--before the take-off and take-off. *Bioinformatics* 18
36
37 Suppl 2:S152.
38
39 Nelson MR, Bryc K, King KS, Indap A, Boyko AR, Novembre J, Briley LP, Maruyama Y,
40
41 Waterworth DM, Waeber G, Vollenweider P, Oksenberg JR, Hauser SL, Sitrnadel HA,
42
43 Kooner JS, Chambers JC, Jones B, Mooser V, Bustamanta CD, Roses AD, Burns DK, Ehm
44
45 MG, Lia EH. 2008. The population reference sample, POPRES: a resource for population,
46
47 disease, and pharmacological genetics research. *Am J Hum Genet* 83:347-358.
48
49
50
51
52
53
54
55
56
57
58
59
60

- 1
2 Papachatzopoulou A, Vantarakis A, Kourakli A, Stavrou E, Frangou E, Patrinos GP,
3
4 Athanassiadou A. 2010. Documentation of the incidence of beta-thalassemia and *HBB*
5
6 mutation frequencies in South-Western Greece. *Hemoglobin* 34:333-342.
- 7
8 Patrinos GP. 2006. National/Ethnic mutation databases: Recording populations' genography.
9
10 *Hum Mutat* 27:879-887.
- 11
12 Patrinos GP, Brookes AJ. 2005. DNA, diseases and databases: Disastrously deficient. *Trends*
13
14 *Genet*, 21:333-338.
- 15
16 Patrinos G, Petricoin E. 2009. A new scientific journal linked to a genetic database: towards a
17
18 novel publication modality. *Hum Genomics Proteomics* 1:e597478.
- 19
20 Patrinos GP, Giardine B, Riemer C, Miller W, Chui DH, Anagnou NP, Wajcman H, Hardison
21
22 RC. 2004. Improvements in the HbVar database of human hemoglobin variants and
23
24 thalassemia mutations for population and sequence variation studies. *Nucleic Acids Res*
25
26 32:D537-D541.
- 27
28 Patrinos GP, van Baal S, Petersen MB, Papadakis MN. 2005. The Hellenic National Mutation
29
30 database: A prototype database for inherited disorders in the Hellenic population. *Hum Mutat*,
31
32 25:327-333.
- 33
34 Povey S, Al Aqeel AI, Barash CI, Cambon-Thomsen A, Dagleish R, den Dunnen JT, Firth HV,
35
36 Greenblatt M, Parker M, Patrinos GP, Savige J, Sobrido MJ, Cotton RG. 2010. Practical
37
38 Guidelines addressing ethical issues pertaining to the curation of human Locus specific
39
40 variation databases (LSDBs). *Hum Mutat* (in press).
- 41
42 Saadallah AA, Rashed MS. 2007. Newborn screening: experiences in the Middle East and North
43
44 Africa. *J Inherit Metab Dis* 30:482-489.
- 45
46 Samara M, Chiotoglou I, Kalamaras A, Likousi S, Chassanidis C, Vagenas A, Vagenas C,
47
48 Eftichiadis E, Vamvakopoulos N, Patrinos GP, Kollia P. 2007. Large-scale population genetic
49
50

- 1
2 analysis for hemoglobinopathies reveals different mutation spectra in central Greece
3
4 compared to the rest of the country. *Am J Hematol* 82:634-636.
- 5
6 Singer PA, Daar AS. 2001. Harnessing genomics and biotechnology to improve global health
7
8 equity. *Science* 294:87–89.
- 9
10 Tadmouri GO, Al Ali MT, Al-Haj Ali S, Al Khaja N. 2006. CTGA: the database for genetic
11
12 disorders in Arab populations. *Nucleic Acids Res.* 34:D602-D606.
- 13
14 van Baal S, Kaimakis P, Phommarinh M, Koumbi D, Cuppens H, Riccardino F, Macek Jr M,
15
16 Scriver CR, Patrinos GP. 2007. FINDbase: a relational database recording frequencies of
17
18 genetic defects leading to inherited disorders worldwide. *Nucleic Acids Res* 35:D690–D695.
- 19
20 van Baal S, Zlotogora J, Lagoumintzis G, Gkantouna V, Tzimas I, Poulas K, Tsakalidis A,
21
22 Romeo G, Patrinos GP. 2010. ETHNOS: A versatile electronic tool for the development and
23
24 curation of National Genetic databases. *Hum Genomics* 4:361-368.
- 25
26 Wang J, Wang W, Li R, Li Y, Tian G, Goodman L, FanW, Zhang J, Li J, Guo Y, Feng B, Li H,
27
28 Lu Y, Feng X, Liang H, Du Z, Li D, Zhao Y, Hu Y, Yang Z, Zheng H, Hellmann I, Inouye
29
30 M, Pool J, Yi X, Zhao J, Duan J, Zhou Y, Qin J, Ma L, Li G, Yang Z, Zhang G, Liu D, Lu Z,
31
32 Li N, Guo G, Zhang J, Ye J, Fang L, Hao Q, Chen Q, Liang Y, Su Y, San A, Ping C, Yang S,
33
34 Chen F, Li L, Zhou K, Zheng H, Ren Y, Yang L, Gao Y, Yang G, Li Z, Feng X, Kristiansen
35
36 K, Wong GK, Nielsen R, Durbin R, Bolund L, Zhang X, Li S, Yang M, Wang J. 2008. The
37
38 diploid genome sequence of an Asian individual. *Nature* 456:60–65.
- 39
40 Wheeler DA, Srinivasan M, Egholm M, Shen Y, Chen L, McGuire A, He W, Chen YJ,
41
42 Makhijani V, Roth GT, Gomes X, Tartaro K, Niazi F, Turcotte CL, Irzyk GP, Lupski JR,
43
44 Chinault C, Song XZ, Liu Y, Yuan Y, Nazareth L, Qin X, Muzny DM, Marqulies M,
45
46 Weinstock GM, Gibbs RA, Rothberg GM. 2008. The complete genome of an individual by
47
48 massively parallel DNA sequencing. *Nature* 452:872–876.
- 49
50
51
52
53
54
55
56
57
58
59
60

- 1
2 Wong ML, Chia KS, Wee S, Chia SE, Lee J, Koh WP, Shen HM, Thumboo J, Sofjan D. 2004.
3
4 Concerns over participation in genetic research among Malay-Muslims, Chinese and Indians
5
6 in Singapore: a focus group study. *Community Genet* 7:44-54.
7
8 Zlotogora J, van Baal S, Patrinos GP. 2007. Documentation of inherited disorders and mutation
9
10 frequencies in the different religious communities in Israel in the Israeli National Genetic
11
12 Database. *Hum Mutat* 28:944-949.
13
14 Zlotogora J, van Baal S, Patrinos GP. 2009. The Israeli National Genetic Database. *IMAJ* 6:373-
15
16 375.
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Box 1: Guidelines and recommendations for data capture in developing countries

1. Develop procedures to include clinicians and researchers in developing nations, mostly by creating networks among those of common interest.
2. Create international networks from closely related populations to enhance communication, interaction, and initiate research networks between them.
3. Foster communication, interaction, and research networks between developing and developed countries.
4. Ensure that the database management systems that are being used or developed can be utilized by those in a limited resource environment.
5. Provide support to developing countries to build capacity and to fully participate in the collection, analysis and sharing of genetic variation information.
6. Develop a framework to facilitate interactions between the coordinating centre and national, regional and international agencies.
7. Ensure that all ethical, legal, religious, and social issues are thoroughly considered when NEMDBs and/or data capture projects are launched in developing countries.

Deleted: s

Deleted: Table 1. Nation-wide genome and related projects (N/A: Not available).
Project

For Peer Review

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47

Page 30: [1] Deleted

ADMIN

10/15/2010 11:15:00 AM

Table 1. Nation-wide genome and related projects (N/A: Not available).

Project	URL	Aims and scope
Nation-wide genome projects		
The Estonian Genome Project	http://www.geenivaramu.ee/index.php?lang=eng	To create a database of health, genealogical genome data representing approximately 100,000 Estonia's population
The MEDGENET Project	http://medgenet.tredueuno.it	To expand the human expertise in clinical genetics in the Mediterranean Partner Countries (MPC) that share a common burden of genetic diseases.
The Indian Genome Variation database	http://www.igvdb.res.in	To create a DNA variation database of the population of India and make it available to researchers for understanding human biology with respect to genetic predisposition, adverse drug reaction, population migration
The Catalogue of Transmission Genetics in Arabs	http://www.cags.org.ae/ctga_search.html	To provide the user with clinical and molecular descriptions of genetic disorders, epidemiological details of reports of the disorders in the Arab world and clinical case details wherever available
Other related projects		
Kuwaiti Molecular Genetics Diagnostic Service	http://www.al-mulla.org	To provide autozygosity mapping services, on Arab genome and molecular karyotyping
UK diagnostic mutation database	http://ngri.org.uk/Manchester/dmudb.html	To provide a route for sharing mutation data between diagnostic laboratories in the United Kingdom to support genetic testing services for patients
Hellenic Genome Variation database	http://www.goldenhelix.org/hellenic	To create a compendium for genetic diseases in Greece, encourage genetic studies to elucidate the genetic basis of inherited diseases in Greece
The International Society for Gastrointestinal Hereditary Tumours (InSiGHT)	http://www.insight-group.org	To improve the quality of care of patients and their families with any condition resulting in hereditary gastrointestinal tumours
International Thalassemia Network (ITHANET)	http://www.ithanet.eu	To strengthen research and treatment of hemoglobinopathies
Pharmacogenomics for Every Nation Initiative	http://www.pgeni.org	To determine the frequencies of the common pharmacogenomic markers in various populations, particularly from developing countries worldwide

1
2
3 **We thank both referees and the communicating and managing editors for their**
4 **very positive and constructive criticism. Please find below our reply to the**
5 **comments in boldface.**
6
7

8 Referee: 1

9 Comments to the Author

10 This paper represents an interesting extension of previous work carried out by the
11 authors and is very much in keeping with the federation of network of LSDBs and
12 CGDBs recently proposed by Patrinos et al. (Locus-specific database domain and data
13 content analysis: Evolution and content maturation towards clinical use. Human
14 Mutation, Vol. 31, Issue 10, 2010).

15
16 However in my opinion one important aspect is completely missing among the
17 recommendations for genetic variation data capture in developing countries. Although
18 the authors acknowledge “that developing countries have a valuable resource of
19 clinical samples, often with very unusual phenotypes, particularly when
20 consanguineous families are involved” they do not stress enough among their
21 recommendations the importance of collecting information about the possible
22 consanguinity of parents of each individual patient and of recording the causative
23 mutation(s) the patient carries, when this information is available.

24
25 As Dr. Patrinos knows very well, today it is possible to develop an approach for
26 estimating the relative frequency of autosomal recessive disorders using mutation data
27 originated by the molecular genetic centers which have diagnosed the affected
28 offspring of consanguineous couples. The rationale of this new approach is based on
29 the possibility that the child born of consanguineous parents carries the same mutation
30 in double copy (true homozygosity) or alternatively carries two different mutations in
31 the same gene (compound heterozygosity). In the latter case the two mutations must
32 have been inherited through two different ancestors of the consanguineous parents (in
33 such case the two mutated alleles are not “identical by descent”). In a very recent
34 paper, Leo ten Kate et al. showed a positive correlation between compound
35 heterozygosity and gene frequency (L.P. Ten Kate et al.: Autosomal recessive disease
36 in children of consanguineous parents: inferences from the proportion of compound
37 heterozygous. J Community Genet. Jan 2010).

38
39 The interesting aspect of this new genetic epidemiological approach is that it makes
40 use of data already existing in molecular diagnostic laboratories. I believe it would be
41 very valuable to find this information also in mutation databases. The practical benefit
42 for developing countries will be that the genetic epidemiological approach just
43 mentioned will yield not only a ranking order of the frequencies of rare monogenic
44 disorders in each of them but also an estimate of prevalence.

45
46 I strongly encourage the authors to add this recommendation to their paper and to
47 promote a comprehensive worldwide data collection also of this information.
48
49
50

51
52 **We thank Referee 1 for his comments. Based also on the Communicating editor’s**
53 **comments, we have chosen to add a couple of sentences in the Introduction to**
54 **emphasize the Referee 1 comments on consanguinity.**
55

56 Referee: 2

57 Comments to the Author

58 This manuscript is a laudable and timely effort to provide guidelines on the collection
59 of human genetic variation data in the developing world, and to offer advice on how
60 these data may be gathered in a fashion and format which maximizes their usefulness

1
2
3 to the local (national) and international communities. This paper will be of interest to
4 readers of Human Mutation and I recommend it for publication, pending one change
5 and a few minor points detailed below.
6
7

8 Major point:

9
10 p6 129: "...reasonable detail of phenotype..." is a rather vague description. Whilst the
11 authors need not explain in detail what "reasonable detail" means, they should
12 perhaps rephrase with something like "...appropriate detail..." or "...level of detail
13 agreed by the community...". Could some relevant references to other phenotype
14 collection projects be included e.g. P3G, GEN2PHEN, Biobanks, etc.?
15
16

17 **We have rephrased this sentence and add some references from relevant**
18 **projects.**

19
20 Minor corrections:

21
22 Abstract, last sentence: insert "a" before "few" - "We also summarize a few well-
23 coordinated genetic data..."
24
25

26
27 p4 158: suggest naming the "...respective working group..." (respective to what?).
28

29 p5 119/20: "...mapping projects, i.e., the HapMap..." - should this be "e.g." rather than
30 "i.e."?
31

32
33 p8 117: Remove "The" preceding "Latin America"
34

35 p9 150: Replace "...the databases management systems..." with "...the database
36 management systems..."
37

38
39 p17 112/13: Replace "The United Kingdom has developed the central database some
40 time ago to collect all UK genome variation data from diagnostic and its effect from
41 participating laboratories." with "The United Kingdom has developed a central
42 database to collect all UK genome variation data, and their effects, from participating
43 diagnostic laboratories."
44

45
46 p19 127/28: Replace "...there are few nation-wide DNA biobanks..." with "...there are
47 a few nation-wide DNA biobanks..."
48

49 p19 144/45/ and 51/52: "incentivation" is a little (if ever) used portmanteau word and
50 should be avoided. Perhaps replace with "motivation", "encouragement", or
51 "incentive".
52

53
54 **We thank Referee 2 for thoroughly going through the text and providing the**
55 **places that corrections are required. We have fully corrected all minor points**
56 **indicated.**
57

58
59 EDITORIAL BOARD'S COMMENTS

60 This is a very useful manuscript which summarizes a small set of important
recommendations to be considered regarding the collection of genetic variation data

1
2
3 in developing countries. This paper should be published with top priority, pending
4 some revision.
5

6
7 The authors are asked to attend to the points raised by Reviewer 2, all of which are
8 minor.
9

10 **We have fully addressed Referee's 2 comments**

11
12 The comments made by Reviewer 1 require some consideration, and some
13 input/response/discussion from the authors would be welcome. Reviewer 1's
14 comments seem perfectly reasonable, but care must be taken not to unbalance the
15 manuscript. The authors might like to consider these options (or suggest another):
16
17

- 18 1) add a sentence or two, emphasizing Reviewer 1's point, in the Introduction where
19 consanguinity is mentioned;
- 20 2) make an additional recommendation along the lines of "the minimum set of data
21 recommended by HUGO-MDI/HGVS/HVP be collected, but that more data should be
22 gathered if possible", then provide examples of additional data, e.g. consanguinity,
23 and why it is important?
24
25

26 **We have chosen to add a couple of sentences in the Introduction to emphasize** 27 **the Referee 1 comments on consanguinity.**

28
29 Please make edits to the manuscript using "track changes" if possible, or by
30 highlighting added text.
31
32

33
34 **Changes are made using the "track changes" option and two file versions are**
35 **provided, one with the changes indicated and another one with the changes**
36 **accepted, the latter file to be forwarded for typesetting.**
37

38 **MANAGING EDITOR COMMENTS:**

39
40 Please respond to these points under "Response to Managing Editor", otherwise the
41 final decision could be delayed.
42
43

- 44 1) Paper organization:
45 a-Please ensure that the subheadings are presented in the proper hierarchy to convey
46 your meaning. For example, While Introduction is a top-level heading and
47 Background is subordinate to Intro, the numbered recommendations actually appear
48 as subordinate to Background. A new top-level heading called "Recommendations"
49 might be necessary before the paragraph starting "The recommendations and
50 guidelines..."
51
52

53 **The new top-level heading has been added.**

54
55 b-Please cite "Box 1" in the text.
56
57

58 **Box 1 has been cited in the text.**

- 59 2) On resubmission:
60

1
2
3 a-Please double check the author names and affiliations carefully. These are often a
4 source of typographical errors.
5
6

7 **Names were double-checked.**

8
9 b-Title Page, Abstract (180-200 words max), Key Words, Main Text, and References
10 should be combined into one file for the manuscript and submitted as a *.doc file.
11
12

13 **Title Page, Abstract (180-200 words max), Key Words, Main Text, and**
14 **References were combined and submitted as one main manuscript (.doc format).**
15

16 c-Tables must be submitted individually as separate *.doc files (with their titles and
17 legends included). Please use the MS Word table format if possible. Excel (*.xls) files
18 should not be submitted.
19

20 --Do not use custom paper sizes - only use Letter or A4.
21

22 **Table 1 has been saved and submitted separately to the main text**
23

24 3) Please check to see that the references follow the journal's standard format and are
25 cited properly. See our online Author Instructions
26 [http://onlinelibrary.wiley.com/journal/10.1002/%28ISSN%291098-](http://onlinelibrary.wiley.com/journal/10.1002/%28ISSN%291098-1004/homepage/ForAuthors.html)
27 [1004/homepage/ForAuthors.html](http://onlinelibrary.wiley.com/journal/10.1002/%28ISSN%291098-1004/homepage/ForAuthors.html)
28
29

30 **Reference format is according to the journal's style.**
31

32
33 4) Human Mutation abides by the NIH Mandate. If your work was funded by the
34 NIH, be sure to include a grant number in an "Acknowledgments" section of the
35 manuscript right before the References. Visit this site for more information:
36 <http://www.wiley.com/go/nihmandate>
37
38

39 **Not applicable**
40

41 **5) IMPORTANT INFORMATION REGARDING PREPRINTS**

42 a-Human Mutation is now publishing online preprints of accepted manuscripts prior
43 to typesetting and page proof corrections. It is therefore crucial that you revise your
44 manuscript carefully so that errors (typographical and grammatical) are corrected
45 BEFORE the final accepted manuscript is posted online. The accepted preprint
46 version will remain online until the corrected proofs are received and the typeset
47 manuscript is finalized. At that time, the preprint version will be replaced with the
48 final, typeset version online, in Early View.
49
50

51
52 b-It is essential that you submit a copyright transfer agreement (CTA) upon
53 submission of your revised manuscript. This will avoid delay in publication of your
54 article upon acceptance. If possible, the CTA must be signed by the corresponding
55 author and should be signed by all contributing authors if practical. All authors must
56 be made aware of the CTA and the rights it conveys to them.
57

58 The CTA can be found here:

59 www.wiley.com/go/ctaus

60 and must be filled out completely, including the article title and manuscript number.
Please fax it to this number: (201) 748-6091.

1
2
3
4
5 **The CTA has been filled in and faxed to the above number.**
6

7 6) N.B.: Referee #1 offered to forgo anonymity and discuss the issue raised with you
8 if necessary. However, the editorial board placed considerably less emphasis on
9 Reviewer 1's comments than Reviewer 2's and suggested a means of dealing with that.
10 It is up to you whether you wish to communicate with Referee 1 and I'd be happy to
11 put you in contact if you wish. Please write me as needed.
12

13 **We feel that there is no need to communicate with Referee 1 since the issue was**
14 **addressed based on the Communicating editor's suggestions.**
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For Peer Review