

Towards an Ecology of Collective Innovation: Human Variome Project (HVP), Rare Disease Consortium for Autosomal Loci (RaDiCAL) and Data-Enabled Life Sciences Alliance (DELSA)

Vural Ozdemir^{1,*}, David S. Rosenblatt^{2,3,*}, Louise Warnich^{4,*}, Sanjeeva Srivastava^{5,*}, Ghazi O. Tadmouri⁶, Ramy K. Aziz⁷, Panga Jaipal Reddy⁵, Aresha Manamperi⁸, Edward S. Dove¹, Yann Joly¹, Ma'n H. Zawati¹, Candan Hizel^{9,10}, Yasemin Yazan⁹, Leela John¹¹, Emmanuelle Vaast¹², Adam S. Ptolemy^{13,*}, Samer A. Faraj^{12,*}, Eugene Kolker^{14,15,*} and Richard G.H. Cotton^{16,17,*}

¹Centre of Genomics and Policy, Department of Human Genetics, Faculty of Medicine, McGill University, Montreal, QC, Canada; ²Department of Human Genetics, Faculty of Medicine, McGill University, Montreal, QC, Canada; ³Rare Disease Consortium for Autosomal Loci (RaDiCAL), McGill University, Montreal, QC, Canada; ⁴Department of Genetics, Stellenbosch University, Stellenbosch, South Africa; ⁵Wadhvani Research Center for Biosciences and Bioengineering, Department of Biosciences and Bioengineering, Indian Institute of Technology Bombay, Powai, Mumbai, India; ⁶Centre for Arab Genomic Studies, Dubai, United Arab Emirates; ⁷Department of Microbiology and Immunology, Faculty of Pharmacy, Cairo University, Cairo, Egypt; ⁸Molecular Medicine Unit, Faculty of Medicine, University of Kelaniya, Ragama, Sri Lanka; ⁹Faculty of Pharmacy, Anadolu University and the Personalized Genomics Healthcare Center, Pharmvation-BİGEM, Eskişehir, Turkey; ¹⁰C2H-Vichy Genomics, Vichy, France, and Kaanmed, Üsküdar, Istanbul, Turkey; ¹¹Consultant in Clinical Pharmacy and Global Health Diagnostics, 2324 West First Avenue, Vancouver BC, Canada; ¹²Desautels Faculty of Management, McGill University, Montreal, QC, Canada; ¹³Department of Pathology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA; ¹⁴Predictive Analytics Seattle Children's Hospital, Bioinformatics and High-Throughput Analysis Laboratory, Seattle Children's Research Institute, Departments of Pediatrics and Medical Education & Biomedical Informatics, Seattle, WA, USA; ¹⁵The Data Enabled Life Sciences Alliance International (DELSA), Seattle, WA, USA; ¹⁶The Human Variome Project International (HVP), Melbourne, VIC, Australia; ¹⁷Genomic Disorders Research Centre, Florey Neuroscience Institutes, and Faculty of Medicine, Dentistry and Health Sciences, University of Melbourne, Melbourne, VIC, Australia

Keywords: Anticipatory ethics, Data-Enabled Life Sciences Alliance, DELSA, genomics without borders, global ethics and international development in pharmacogenomics, Human Variome Project, RaDiCAL, Rare Disease Consortium for Autosomal Loci.

“Where the common disease communities bring opportunities of scale, the heritable disease community brings insights from detail and experience. Both are concerned with personalized medicine and should work together.”

-- Burton *et al.* 2011 [1]

*Address correspondence to the CPPM Editors and the International Consortia Chief Scientific Officers: R.G.H. Cotton at the *Human Variome Project International* (HVP), University of Melbourne, Melbourne, VIC, Australia, Email: cotton@unimelb.edu.au; D.S. Rosenblatt at the *Rare Disease Consortium for Autosomal Loci* (RaDiCAL), McGill University, Montreal, QC, Canada, Email: david.rosenblatt@mcgill.ca; E. Kolker at the *Data Enabled Life Sciences Alliance International* (DELSA), Seattle Children's Research Institute, Seattle, WA, USA, E-mail: eugene.kolker@seattlechildrens.org; L. Warnich at Stellenbosch University, Matieland, South Africa, E-mail: lw@sun.ac.za; S. Srivastava at the Indian Institute of Technology Bombay, Powai, Mumbai, India, Email: sanjeeva@iitb.ac.in; A.S. Ptolemy at Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA, E-mail: aptolemy@partners.org; S.A. Faraj at the Desautels Faculty of Management, McGill University, Montreal, QC, Canada, Email: samer.faraj@mcgill.ca; and V. Ozdemir at the Centre of Genomics and Policy, Department of Human Genetics, McGill University, Montreal, QC, Canada, Email: vural.ozdemir@mcgill.ca

1. CHANGING MINDSETS AND THINKING BEYOND BORDERS

1.1. Rise of Collective Innovation

The Millennium Summit in 2000 established the Millennium Development Goals (MDGs), which were agreed upon by 193 countries and 23 international organizations to combat extreme poverty and other pressing global priorities for human development. In the December 2011 issue of CPPM, Borda-Rodriguez and Huzair present an analysis of the close ties and synergies among the MDGs, pharmacogenomics and personalized medicine [2]. Notably, MDGs promote the creation of *collective innovation*, a concept with both substantive and instrumental pertinence for the personalized medicine R&D that is currently

undergoing rapid globalization. The ethos for collective innovation in global health is also embodied in the Paris Declaration on Aid Effectiveness, endorsed in 2005 by more than 100 signatories, including donor and developing country governments, regional development banks and international aid agencies [3].

Launched in 2008, CPPM is a peer-reviewed integrated forum for pharmacogenomics and personalized medicine as seen through the lens of global public health [4], with particular attention to emerging applications in developing countries [5-7]. If the field of personalized medicine is to equitably expand to low and middle income countries (LMICs) to address neglected tropical diseases, tuberculosis, HIV/AIDS, malaria [5] and the current global epidemic of noncommunicable diseases (NCDs) [8], significant changes in mindsets are necessary [7, 9, 10].

The traditional Westphalian model of independent sovereign nation states does not adequately recognize the complex interdependencies in global governance for health [9] or the need to think about genomics beyond national borders [7, 11-13]. Moreover, the pharmacogenomics community has been “on pause” in incorporating critical social sciences and humanities research [14], while other emerging postgenomics data-intensive fields such as nanotechnology have been appropriately on “fast forward” [15, 16]. Creating healthy and sustainable modern societies in the 21st century requires “integrated knowledge ecosystems” with contributions from all sectors of science and society:

Discussing the state of health at the beginning of the 21st century is not dissimilar to discussing the state of the environment. Both are in crisis and run counter to the notion of sustainable wellbeing, both focus around the ways of life that have developed in our societies and both indicate that significant changes are required at the level of policy and of society...Both our extensive knowledge about what creates health, as well as the exponentially rising rates of chronic disease, obesity and mental health problems, indicate that we need to shift course and apply a radically new mindset to health...Some warning voices state that the present generation of children – born at the turn of the century – might be the first to have a lower life expectancy than their parents. After the extraordinary gains in health and longevity over the last 100 years, such a development would indeed represent a phenomenal failure...Clearly if health is created by factors outside the remit of the traditional health sector then the creation of a healthy society needs commitment and action from all sectors of government, new partnerships for health, and citizens’ participation [17].

The above insights describe precisely what is needed to transform the “one-scientist-one-project paradigm” that has hitherto prevailed throughout the 20th century towards crossdisciplinary collaborative science in the 21st century [7, 18-20]. Collective innovation offers the promise of novel diagnostics for complex diseases and single gene disorders, as well as customized drug therapies, vaccines, cell-based therapies and nutrition [7]. Diagnostics are now being developed by innovators well beyond the industrialized

countries of the North [5-7, 21]. In particular, the Asia-Pacific region is rapidly emerging, owing to substantial investments in data-intensive life sciences research [6, 7, 22]. A recent study on the globalization of science found, using quantitative indicators, that the current growth rate for state-level research funding in China greatly exceeds that in the US and the European Union (EU) [23, 24].

The *Human Variome Project* (HVP), an incorporated and globally inclusive charitable body headquartered at the University of Melbourne, Australia, is a large-scale international consortium for collective innovation on the critical path to novel diagnostics [11, 25]. The HVP offers the promise to vastly impact and inform the continuum from disease prevention and diagnosis to personalized rational therapeutics. As a global collective initiative in partnerships with individual country genetic societies, the HVP will collect and curate all human genetic variations effecting human disease. The endorsement in 2011 of the HVP by the United Nations Educational, Scientific and Cultural Organization (UNESCO) [25], together with the recent pledge by the Chinese Government of US\$300 million to establish a new institute to directly contribute to the efforts of the HVP, brought about a renewed enthusiasm for novel diagnostics as well as formidable expectations for global public health and personalized healthcare.

Diseases with a strong heritable component (often referred to as Mendelian or single gene disorders) provide many potential avenues for novel diagnostics on the path to global personalized medicine [1, 26]. Indeed, the Online Mendelian Inheritance in Man (OMIM) database indicates that there are 1638 autosomal phenotypes for which the molecular basis is unknown [13]. Historically, such disorders have been neglected in discussions on personalized medicine, perhaps because of their rarity and the difficulty in identifying the underlying genes and mutations [1, 13]. The *Rare Disease Consortium for Autosomal Loci* (RaDiCAL) led by McGill University in Canada is a recent effort to address precisely this knowledge gap [13] on putative autosomal recessive diseases. The RaDiCAL consortium takes a whole exome (or whole genome) approach and gives priority to autosomal recessive conditions so as to yield far fewer false positive hits when searching for either a homozygous deleterious mutation or two deleterious mutations in a single gene.

Postgenomics personalized medicine is experiencing a “data deluge” from emerging high-throughput technologies and a myriad of sensors and exponential growth in electronic records. Data-driven, data-intensive science (referred to herein as “data-enabled science”) was originally proposed by Jim Gray and colleagues as the Fourth Paradigm of scientific inquiry [22]. It builds on and greatly expands the three precedent paradigms – theory, experimentation and simulation (modeling) [22]. Data-enabled science is defined by the National Science Foundation (NSF) as a new research paradigm that firmly depends on a data commons [27] or “infrastructure science” and is inseparable from the classic “discovery science” [22, 28]. The data-enabled science community recognized that shared problems such as the current transition of bioinformatics to clouds demand

collective solutions and moved to establish the *Data Enabled Life Sciences Alliance* (DELSA) with the vision of sustainable and shared access to data, knowledge and tools for 21st century postgenomics life sciences [29].

As we begin to turn towards collective innovation in postgenomics R&D, this editorial analysis provides an overview of the HVP, RaDiCAL and DELSA that share a global ethos for collective innovation in data-enabled sciences. In the spirit of collective research infrastructure capacity building for postgenomics novel diagnostics, we present a survey of the insights and expectations from the leading experts in global regions where disease burdens stand to benefit from the HVP, RaDiCAL and DELSA. Developing countries, in particular, have a significant role to play in 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. Research in any one country can also inform the genomics healthcare for individuals and family groups in diaspora in other countries. We also wish to note the generation of a new initiative, the International Rare Disease Research Consortium (IRDiRC) [30], whose main thrust is to develop therapeutics and diagnostics for rare disorders with a smaller input into global registries.

We conclude the editorial by underscoring that the HVP, RaDiCAL and DELSA attest to the ever increasing diversity in the “ecology of collective innovation”, which is essential for discovery of truly novel diagnostics that can benefit global personalized medicine and public health.

2. THE HUMAN VARIOME PROJECT (HVP)

The global community needs total annotation of the human genome and it could still be decades before this is complete. The Human Variome Project [25] was initiated to collect all variations effecting human disease. The project was launched at a World Health Organization (WHO)-hosted meeting in Melbourne, Australia in 2006 by representatives from the WHO, the UNESCO, European Commission, Organisation for Economic Co-operation and Development (OECD), March of Dimes, major databases, 30 countries, and world experts [31]. Recommendations from this inaugural meeting were subsequently published [32]. Most recently, the Human Variome Project has been incorporated as a charitable body and the third Human Variome Project Meeting at Paris UNESCO Headquarters [33] elected an interim scientific advisory committee and ratified a roadmap [25].

The aim of the Human Variome Project is to facilitate the collection of all mutations effecting human disease in all genes and from all countries, to link this to clinical and phenotypic data, and to make the data freely and openly available for use in research and patient care worldwide. The roadmap specifies a two-pronged approach to this collection. The first approach is via gene/locus/disease specific databases (LSDBs). A number of LSDBs have been in existence for many years, providing expert oversight and curation of data associated with single genes or diseases. Existing databases must be encouraged to grow and expand, while new databases need to be created to fill the gaps in genome coverage. The second approach is via country-

specific nodes, which are relatively novel and which are needed for proper genetic healthcare, funding and strategy in each country [34]. In the past, LSDBs usually focused on the first instance of a particular mutation, but there is a need to collect each instance of a mutation for various reasons [34].

Considerable progress has been made already by the HVP. This includes: (a) a pilot data collection strategy and databasing by the inherited colon cancer community (InSiGHT) [35] with other areas of biology/disease progressing; (b) the launch of the Human Variome Project Australian Node (www.hvpaustralia.org.au) on April 1, 2011 as a pilot to collect all data within a country and to provide software and strategy to those who need it; (c) 12 countries so far that have agreed to set up country nodes; (d) China agreeing to fund a substantial amount of the project (see above); (e) the Project being granted the status of ‘Non Governmental Organization (NGO) in Operational Relations with UNESCO’; and (f) the holding of international forums and lectures in 16 countries since 2006.

A major meeting focusing on new international collaborations will be held in Beijing on December 8-12, 2011 and the fourth HVP meeting will be held at UNESCO Headquarters in Paris on June 11-15, 2012 [25]. A global call has been made to encourage all to join the HVP [25].

3. RARE DISEASE CONSORTIUM FOR AUTOSOMAL LOCI (RaDiCAL)

RaDiCAL is an attempt by researchers, initially based in the Department of Human Genetics at McGill University in Montreal, Quebec, Canada, to fast track gene discovery for autosomal diseases using approaches such as exome and whole genome sequencing [13]. The thought behind RaDiCAL is that since many in the scientific and medical communities are driven by personal career and academic motives, there is a need to stimulate a new approach that will result in rapid discoveries that will immediately benefit patients. The goal of RaDiCAL is to find as many genes as possible as quickly as possible for the purpose of better diagnosis and genetic counseling. The hope is to work with clinical colleagues around the world to collect a single well-characterized proband for each autosomal recessive disease for which the gene is not known. Importantly, the limiting factor in this approach is the clinical diagnosis and not the molecular biology. All gene assignments would have to be independently validated by looking at additional patients with the same clinical condition. The plan is to publish or post candidate genes for each disease quickly and without functional validation. In addition, there would be no attempt to look at known disease genes in the samples from patients that are not candidates for the diseases in question. How these candidates will be made accessible to the community at large remains unclear. Either a journal or society will champion this approach or it will be necessary to create and maintain a centralized database that would simply post candidate genes. The arrival of data-enabled sciences and the increasing recognition of “data” and “data citation” as legitimate units of microattribution (*e.g.*, as a complement to classic journal article citations) for scientific credit and academic promotion, could catalyze and transform upstream discovery initiatives such as RaDiCAL.

4. DATA-ENABLED LIFE SCIENCES ALLIANCE (DELSA)

The data-enabled science paradigm is a departure from traditional hypothesis-driven life science research. Because it opens up new avenues of scientific exploration that do not rely on prior assumptions on disease pathophysiology or drug action, the paradigm thus paves the way for innovative solutions to current problems in life sciences, medicine and global health. In this era of data-enabled science, life scientists produce diverse, multifaceted data at an unprecedented and accelerated pace that is too often unmatched by the availability of adequate cyber-infrastructure capacity and real-life policies to govern its use. In September 2010, the data-enabled community explored issues and challenges associated with these new avenues [36] and in May 2011, the community established the DELSA with the vision of advancing 21st century life sciences by moving from “one scientist-one project” to “collective innovation”. Hosted by the Seattle Children's Research Institute [29], co-founders are geographically distributed with research interests spanning an extensive array of fields such as ecology, biodiversity, bioinformatics, pharmacogenomics, pharmacoproteomics and sustainability sciences. DELSA also includes experts in computer sciences, super-computing, data and cyberinfrastructure.

The data deluge the life sciences community currently faces is part of the Fourth Paradigm of scientific inquiry [22]. This brings about a need for organizational and cultural changes in addition to those changes/advances usually associated with a technical realm. These changes will come about through a supporting ecosystem that includes: research funding agencies and foundations in developed countries and LMICs; governments; academia; industry; and publics. All must work together to build capacity for data-enabled life sciences infrastructure and discovery. In particular, cloud-based and distributed computing technologies are some of the salient focus areas for DELSA. Cloud computing offers the possibility of “on demand” flexible computing that is of great relevance for global personalized medicine, including in rural and resource-limited settings in LMICs. With cloud computing, 1000 computers used for one hour costs the same as one computer used for 1000 hours [29].

5. FROM CONSORTIA TO COLLECTIVE INNOVATION

Both frequent contacts between data/knowledge generators (*e.g.*, scientists, technology designers) and end-users (*e.g.*, publics, policy-makers) and creating mechanisms between supply and demand of knowledge and technology are necessary to move towards responsible innovation in democratic societies [9, 10]. Hence, we present below expert insights on the current situation in global regions, including developing countries, in relation to collective innovation in data-enabled life sciences, and what we may anticipate ahead.

5.1. South Africa

South Africa is home to many distinct populations, displaying abundant genetic diversity but also homogeneity

due to founder effects [5]. While these populations, as those from other sub-Saharan countries, have been largely underrepresented in genomics and postgenomics research, international initiatives such as HVP, RaDiCAL, DELSA, as well as the Human Heredity and Health in Africa (H3Africa) [37] can enable such research in Africa. South Africa is already linked to the HVP through the InSiGHT group, who will pilot country-specific collections for four colon cancer genes [38]. Furthermore, the country has a number of established clinical genetic units that could participate in RaDiCAL, as first and second generation sequencing facilities are now readily available locally for the characterization or validation of candidate genes by whole exome or targeted re-sequencing. While the establishment of biorepositories and bioinformatic expertise and networks are featuring prominently on the agendas of H3Africa and the Southern African Human Genome Project [39], the training and retraining of clinical geneticists and genetic counselors should also be a high priority, as they are key role players in basic as well as translational research. The commitment of local and national government to contribute to long-term sustainable funding for research and translational services will thus have a significant impact on the long-term success and impact of postgenomic research in South Africa [5, 40]. Future challenges will be linking the genetic variations to more common and complex human diseases and traits, and understanding the interaction of these variations with the different and unique environmental conditions in Africa.

In the immediate term, the benefits of studies such as HVP and RaDiCAL will include the more thorough characterization of southern African diversity through the analyses of whole exome and genome data of population samples, and linking some of the genetic variants to human Mendelian disorders. Efforts by DELSA to develop cyberinfrastructure and high-throughput analysis capacity, including the emerging cloud technologies, can potentially extend the reach of genomics medicine to rural communities in the African continent.

5.2. United Arab Emirates and Other Arab States

In the United Arab Emirates, and more generally in other Arab countries, the highly demanding data intensive field of collective genomics is mainly available in selected research centers that are distributed over a vast geographical area including North Africa, the Eastern Mediterranean, and the Gulf States. Some of the Arab countries, such as Tunisia, Morocco and Lebanon, have integrated adequately into international collaborative networks, but this engagement primarily centers on the implementation of advanced high-throughput omics molecular technologies [41]. That is, postgenomics research in many Eastern Mediterranean and Gulf States seems to be fuelled by collaborations at a downstream clinical innovation level essentially targeting endogenous country-specific public health needs, despite the wealth of technologies that are available for *upstream* discovery-oriented postgenomics research in the region [12].

The capacity for collective innovation in genomics in the United Arab Emirates and Arab countries can be developed by:

- coordinated transfer of know-how to local researchers as well as addressing the serious imbalance in genomics R&D funding by local governments;
- accelerating development aid in genomics and personalized medicine, but with careful consideration of the priorities of the recipient countries, and not only of the donors (see also the editorial discussion in this issue of CPPM by Borda-Rodriguez and Huzair) [2]; and
- comprehension of the importance of genomic applications for global public health in LMICs (see the previous CPPM interview article by Muin J. Khoury) [4].

At the time of the publication of this editorial, nearly 979 conditions are known to cause various forms of inherited disorders in Arab populations [42]. Of these, a large majority, 718 disorders, are associated with well-documented gene pathologies in other global populations. Yet, population genomic analyses of almost half of these disorders have not been carried out in the region and this represents a serious loss of a *shallow infonome* (a term previously coined by a co-author of this editorial, Dr. G.O. Tadmouri, referring to the total information gained by study of disease-causing variation) that may be remedied rapidly by appropriate use of genomics tools. A promising trend, however, is that 71 disorders have been clinically and genetically characterized over the last 12 years for the first time in Arab families. Additionally, there are 261 inherited disorders that are almost exclusively described in Arab families and the available knowledge is only confined to clinical studies with no efforts to study their genetic etiologies. Deciphering the molecular pathologies of these disorders, including many monogenic Mendelian diseases, will definitely result in unearthing the *deep infonome* (i.e., the information gained by study of the total variome content of a genome). The knowledge of the deep infonome will be even more attractive when combined with large-scale studies aiming to map non-disease variome patterns in the highly charged gene pool in Arab populations. The consortia for collective innovation discussed in this editorial, the HVP, RaDiCAL and DELSA, are therefore in direct synergy with the extant needs of genomics medicine and the resources presently available in the United Arab Emirates and other Arab countries.

5.3. Egypt

A recent analysis of the scientific publication data from the Web of Science over the past 30 years (1980 – 2009) found that Asia's share of world scientific output grew by 155%, while the Middle East, as a region, displayed a growth four times faster than at the world level [43]. Egypt, ranked fourth in this recent report [43], is among the leading Middle Eastern countries in scientific output. Out of the 16 Middle Eastern countries, Egypt is ranked third by the H-index, with Israel and Turkey taking the first and second positions, respectively [44]. Importantly, while the scientific contribution of Egypt relative to the rest of the Middle East has slightly decreased, it steadily increased by more than three-fold in the category of "Biochemistry, Genetics and Molecular Biology" over the past 15 years [44]. Moreover, international collaboration is a major aspect of the Egyptian

scientific output: about 40% of all publications are contributed by authors from two or more countries; this number reaches 50% in the category "Biochemistry, Genetics, and Molecular Biology".

In spite of this stupendous growth in Egyptian biochemical and genetics research, the contributions to genome-wide studies are still restricted to scattered individual attempts, mostly relying on international collaborations, where Egyptian researchers single-handedly travel and carry out most of the experimental work abroad [45], or in collaboration with a large regional institution such as the Red Sea Metagenomics Project (launched as a collaboration between the American University in Cairo and King Abdullah University in Saudi Arabia [46]).

5.4. Turkey

Turkey has a high rate of consanguineous marriages that is one of the major causes of recessive diseases in the country. For example, in the southern population of Turkey, the prevalence of consanguinity in rural areas is quite high, estimated to be around 40%, and is associated, for example, with beta-thalassaemia major [47]. Because personalized medicine has long been framed in the international arena rather narrowly around common complex diseases, we wish to underscore that in countries such as Turkey with a high level of consanguineous marriages, single gene and recessive disorders have great public health significance; novel diagnostics can make an important contribution for early detection and genetic counseling of vast numbers of families in Turkey. Hence, it will be important to highlight the similarities between both common diseases and single gene diseases and that both fall under the rubric of personalized medicine as indicated recently for the CPPM readership by Burton *et al.* [1].

An inaugural conference on personalized medicine was held in Istanbul in 2009 with large participation from scientists, patient interest groups and publics [48]. Among the Middle Eastern countries, Turkey stands in second rank in scientific impact according to the H-index, as noted above. Still, the much-needed global public health genomics lens [4] is missing in Turkey as in many other LMICs. Incorporation of the ACCE public health evidentiary framework to emerging R&D efforts for personalized healthcare in Turkey is essential and would be a concrete way forward (see also discussion in Section 8) [4]. Regional capacity building efforts in genomics and personalized medicine, its integration with genetic counseling, as well as education of scientists, healthcare professionals and publics are underway at the Personalized Genomics Healthcare Center, Pharmvation-BİGEM at Anadolu University in Eskişehir, Turkey. Incorporation of human genomics variation in routine therapeutic drug monitoring and pharmacovigilance is an immediately actionable priority area for Turkey.

5.5. India

India has the world's second largest population, consisting of over 4500 communities with vast human genetic/genomic variation. The tradition of consanguineous marriages is common in certain communities, as are

hereditary diseases. The most prominent single gene disorders include haemoglobinopathies such as beta-thalassemia (with 4% prevalence), sickle cell anemia, cystic fibrosis and Wilson's disease. To identify the carriers in the population and at risk groups, screening programs have been initiated [49]. The Council of Scientific and Industrial Research, along with seven research institutes, started the Indian Genome Variation (IGV) project in 2003 to understand the inherent genetic variability of subpopulations, identify SNPs in thousands of genes, construct haplotype maps and determine functional consequence of genetic markers [50]. Furthermore, the IGV database (IGVdb) project is documenting SNPs and repeat polymorphisms using high-throughput platforms. The New Millennium Indian Technology Leadership and the National Biotechnology Development Strategy have also been initiated. The Department of Biotechnology has constituted a task force on the Human Genetics and Genome Analysis, which, among several initiatives, have established several units for genetic diagnosis and counseling services.

It is interesting to note that although India could not participate in the Human Genome Project extensively, the IGV and other initiatives outlined above illustrate the growing investments in genomics and postgenomics R&D in the country. Insofar as the HVP, RaDiCAL and DELSA are concerned, notable priority data-enabled science fields in India include proteomics, pharmacoproteomics and nanoproteomic technologies [6]. Indeed, in 2005, Dr. Abdul Kalam, the former President of India, aptly observed that "India has the potential to tap research opportunities in proteomics and biochips to help understand the biological processes and treat diseases. This is possible even though the country has missed the opportunity to partner in the Human Genome Project" [6]. The Indian Institute of Technology Bombay and various other leading centers and universities in India have developed extensive expertise in the field of pharmacoproteomics, proteomics and nanotechnology. We refer the readers to the editorial analysis in the September issue of CPPM, where we published a detailed "innovation map" of proteomics and pharmacoproteomics in India with a list of "hot spots" [6].

5.6. Sri Lanka

The island-state of Sri Lanka is situated closely to the Indian subcontinent. While there are a very limited number of experts in molecular genetics in the country, public health genomics services are necessary for population health in Sri Lanka, particularly to better understand the epidemiology and transmission of life-threatening common tropical infections such as dengue hemorrhagic fever in the Indian subcontinent as well as globally. However, aid and development agencies have unfortunately not kept Sri Lanka in their priority list thus far, despite a recent rise in congenital anomalies and NCDs. In response to the changing epidemiology of disease burden in the country, national health policy is being targeted towards the common causes of neonatal and infant deaths, tropical diseases and NCDs.

Molecular genetic/cytogenetic testing for human diseases is available from the private sector and state institutions for a large number of conditions in Sri Lanka. There are three

state-of-the-art molecular technology units attached to university centers geared towards developing new diagnostics and performing research in molecular genetics, and a cancer screening center offering advanced care from initial risk assessment to appropriate screening for cancer. At present, data on the prevalence of genetic diseases and associated genetic markers are limited. Population-based prevalence studies have been done only for genetic conditions where there is a known high prevalence in a particular area, for example, thalassaemia in the Kurunegala district, or in the case of Down's syndrome. Equity of access to genetic tests and accreditation of genetic laboratories remain very problematic in Sri Lanka. The Sri Lanka Accreditation Board has recently started the process of voluntary accreditation of state and private laboratories offering diagnostic investigations in the country.

With the end of the civil war in 2009, Sri Lanka is an important context for global health practitioners for post-war capacity building in health R&D that is informed by data-enabled sciences such as genomics and proteomics. The Sri Lankan Personal Genome Project is a noteworthy and timely development in this regard [51].

6. BEYOND TECHNOLOGY

6.1. Open Source and Collaborative Knowledge Innovation Platforms

Despite the great successes achieved through genomics and related omics sciences, the field suffers from a "transfer problem", whereby its important pool of fundamental scientific knowledge has not materialized in concrete applications. Various reasons have been provided to explain this disappointing outlook, including biotechnology patent thickets, university business models focused on a narrow commercialization framework, inefficient regulatory approval processes, and overly constraining material transfer agreements [52, 53].

Promising avenues to solve this transfer problem could be based on the open source model implemented through collaborative knowledge innovation platforms (e.g., crowdsourcing, expert sourcing, cloud computing, social networking technologies). Their potential benefits, though too soon to empirically evaluate, are considerable and include: promoting the interoperability and harmonization of research platforms and tools, reducing financial risks, democratizing access to innovation, preventing harmful duplication of research, and facilitating the technology transfer process. Nonetheless, shifting from a classic commercialization and individual-centered model of innovation to one that stresses collective innovation, transparency and collaboration requires an acknowledgement of more than just the ensuing (and obvious) technical and economic changes. It also requires an exploration of the ability of LMICs to have the necessary IT infrastructure to access openly shared data and actively participate in collaborations, the nature of the "informed" consent provided by participants for open research, the role of ownership and commercialization of genomic data, and the ability to uphold the duty of confidentiality and respect of participants' privacy. As postgenomic era science advances

in its march towards collective innovation to create novel diagnostics, democratize omics research and move from silos to systems, we need to consider that open source, along with collaborative knowledge innovation platforms, will play an essential role.

Since the HVP, RaDiCAL and DELSA are international large-scale initiatives, much can be learned from the experience of population biobanks, in addition to the open source model described above. For example, the Public Population Project in Genomics (P³G) is a not-for-profit international consortium that aims to provide the international population genomics community with access to tools and resources for biobanks to encourage collaboration between researchers and biobankers, promote harmonization of information and optimize the design, set-up and research activities of population-based biobanks [54]. Due to their global nature, the collective innovation cases discussed in this editorial, as well as the policy infrastructure (*e.g.*, universal consent, open/controlled access, *etc.*) of future biobanks in LMICs, can build on such publicly available tools.

7. THE DARK SIDE OF THE MOON: HOW DO PEOPLE ACTUALLY COLLABORATE?

7.1. A Lesser-Known Side of Collective Innovation: Complex Collaboration Across Knowledge Boundaries

It is interesting to note that we know very little about how individuals and stakeholder groups are organizing themselves in the practice of data-enabled sciences, ways in which collaborations (*e.g.*, cooperation, competition, pre-competitive collaboration) are developed, sustained or devolve; what human values guide and inform the knowledge ecologies emergent from collective innovation; and ultimately, how knowledge is created, transformed and transferred between innovators and users. That is, the large-scale collective innovation projects described in this editorial require not only the collection of genomic/proteomic and human variome data on a world-wide scale but also an understanding of its social science and humanities dimensions. As we move from a science where data collection was an essential locus of the scientific endeavor to one where data collection is automated or at least made easier and is available in digital form, there is a need to understand the new collaborative possibilities that emerge.

It is our contention that traditional modes of knowledge sharing, such as international meetings, pre-planned teams, and the traditional review process are possibly becoming constraints on the collaboration needed to bring the Fourth Paradigm of scientific inquiry to fruition. Current research in the organizational sciences has developed important insights on how complex collaboration can be managed. It begins with a recognition that complex collaboration is needed when knowledge crosses organizational or national boundaries, when those involved have differences in methods, research practices, ontologies, human values and epistemologies (*i.e.*, ways of knowing). Further, because the knowledge is typically produced in a local context within a specific community of practice, knowledge is hard to transfer across boundaries and combine effectively [55-57].

Therefore, we suggest that clear attention to complex collaboration is needed in order to make this new form of research, *i.e.*, data-enabled sciences for collective innovation, successful. Much can be learned from adapting basic organizing principles, norms, and rewards from collective action [58].

7.2. Collaborating with Publics in LMICs

In order to move ahead for global pharmacogenomics and personalized medicine, we need to work with the public(s) in developed countries and LMICs, recalling that experts, too, are part of the publics. Each community within a country or region may have different priorities with respect to the major infectious diseases (malaria, tuberculosis and HIV/ AIDS), neglected tropical diseases (helminth infections), NCDs (cardiovascular diseases, cancer, diabetes and chronic lung diseases) [8], and maternal, reproductive and mental health. These priorities will need to be determined by the aid recipient countries to identify the solutions that are required in the short-term versus the long-term while in parallel guiding the R&D funding directions and development of new diagnostic tests for personalized healthcare for each community. Setting these priorities collaboratively will also encourage transparency and accessibility of genomics research findings for community members and directly contribute to the measurement of health outcomes in postgenomics personalized medicine.

8. EVIDENTIARY FRAMES FOR DATA EVALUATION

In this issue of CPPM, Ptolemy describes the extant difficulties in translation of putative biomarkers into a routine clinical setting [59]. To this end, those within the public health genomics community have deployed the ACCE framework [4]. This framework respectively evaluates the: (a) analytical validity, (b) clinical validity, (c) clinical utility and (d) ethical, legal and social implications of a genetic test and/or putative biomarker. Through the implementation of this evidentiary frame (which has a strong public health lens), researchers would not only be able to challenge the validity of their data, but also objectively determine the true clinical utility of a derived result, including its benefit to global society.

Including the experimental and public health tenets, such as those embodied in the ACCE framework, in the global consortia from their outset, we can ensure that the massive amounts of data obtained from global collective innovation efforts are not only valid, but also truly applicable and appropriate for their stated public health and societal goals.

CONCLUDING REMARKS

While an editorial analysis cannot provide the possibility to cover all global regions, we trust that the original synthesis presented herein crucially informs future debates on global personalized medicine, including the often-neglected views and advances from developing countries. Recent applications of collective open-source genomics analysis, data release and crowdsourcing approaches [60, 61] support the notion that the Fourth Paradigm of data-enabled science has much to offer personalized medicine and novel diagnostics. We therefore express our collective optimism

that this new era will benefit global society in both developed and developing countries.

ABBREVIATIONS

DELSA	=	Data Enabled Life Sciences Alliance International
HIV/AIDS	=	Human Immunodeficiency Virus/Acquired Immunodeficiency Syndrome
HVP	=	Human Variome Project International
IGV	=	Indian Genome Variation
LMICs	=	Low and middle income countries
LSDB	=	Locus-specific database
NCDs	=	Noncommunicable diseases
NSF	=	United States National Science Foundation
OECD	=	Organisation for Economic Co-operation and Development
RaDiCAL	=	Rare Disease Consortium for Autosomal Loci
UNESCO	=	United Nations Educational, Scientific and Cultural Organization
WHO	=	World Health Organization

DUALITY/CONFLICT OF INTERESTS

None declared/applicable.

ACKNOWLEDGEMENTS

The analysis and concepts presented herein were supported by the following research grants to the authors: a career investigator salary for science-in-society research in personalized medicine from the Fonds de recherche du Québec – Santé, and an operating research grant from the Canadian Institutes of Health Research (CIHR) (#84620 to V. Ozdemir); the CIHR operating grant entitled “*Inborn errors of cobalamin and folate metabolism*” (D.S. Rosenblatt); the Canada Research Chair in Technology, Management & Healthcare (S.A. Faraj); Seattle Children’s Research Institute and the US National Science Foundation (#DBI-0969929 and #DBI 0544757 to E. Kolker) and the US National Institutes of Health (NIGMS grant 5R01 GM076680-02); Australian National Health and Medical Research Council (NHMRC) (#350418 to R.G.H. Cotton). The views expressed in this article are the personal opinions of the authors and do not necessarily represent the positions of their affiliated institutions or the funding agencies. This editorial paper was peer-reviewed by the CPPM editorial board.

REFERENCES

- Burton H, Nacul L, Sanderson S. Personalised medicine: lessons from birth defects and single gene disorders. *Curr Pharmacogenomics Person Med* 2011; 9(2): 80-3.
- Borda-Rodriguez A, Huzair F. Revisiting the *Paris Declaration*: pharmacogenomics and personalized medicine as accelerators for development aid and effectiveness. *Curr Pharmacogenomics Person Med* 2011; 9(4): 240-2.
- OECD. Paris Declaration and Accra Agenda For Action 2005.
- Khoury MJ, Muin J. Khoury discusses the future of public health genomics and why it matters for personalized medicine and global health. *Curr Pharmacogenomics Person Med* 2009; 7(3): 158-63. Available from: <http://www.benthamscience.com/cppm/openaccessarticles/cppm7-3/003AF.pdf> [Accessed October 27, 2010].
- Warnich L, Drögemöller BI, Pepper MS, *et al.* Pharmacogenomic research in South Africa: Lessons learned and future opportunities in the rainbow nation. *Curr Pharmacogenomics Person Med* 2011; 9(3): 191-207. Available from: <http://www.benthamdirect.org/pages/content.php?CPPM/2011/00000009/00000003/007AF.SGM> [Accessed October 27, 2010].
- Reddy PJ, Jain R, Paik YK, *et al.* Personalized medicine in the age of pharmacoproteomics: A close up on India and need for social science engagement for responsible innovation in post-proteomic biology. *Curr Pharmacogenomics Person Med* 2011; 9(3): 159-67.
- Ozdemir V, Muljono DH, Pang T, *et al.* Asia-Pacific Health 2020 and genomics without borders: co-production of knowledge by science and society partnership for global personalized medicine. *Curr Pharmacogenomics Person Med* 2011; 9(1): 1-5.
- Editorial. Two days in New York: reflections on the UN NCD Summit. *Lancet Oncol* 2011; 12(11): 981.
- Kickbusch I. Governance for health in the 21st century. Geneva: WHO Regional Office for Europe 2011. Available from: http://www.euro.who.int/_data/assets/pdf_file/0010/148951/RC61_InfDoc6.pdf. [Accessed October 27, 2011].
- Wynne B. Lab work goes social, and vice versa: Strategising public engagement processes: commentary on: “What Happens in the Lab Does Not Stay in the Lab: Applying Midstream Modulation to Enhance Critical Reflection in the Laboratory”. *Sci Eng Ethics* 2011 Oct 21 [Epub ahead of print].
- Cotton RG, Vihinen M, den Dunnen JT, *et al.* Genetic tests need the Human Variome Project. *Genet Test Mol Biomarkers* 2011; 15(1-2): 3.
- Patrinis GP, Al Aama J, Al Aqeel A, *et al.* Recommendations for genetic variation data capture in developing countries to ensure a comprehensive worldwide data collection. *Hum Mutat* 2011; 32(1): 2-9.
- Rosenblatt DS. A RaDiCAL approach to gene discovery. *J Med Genet* 2011; 48(9): 577-8.
- Ozdemir V. Pharmacogenomics: Reflections on the old and new social, ethical and policy issues in postgenomics medicine. In: *Pharmacogenomics in Psychiatry*. Schwab M, Kaschka W, Spina E. (Eds.), Basel: S. Karger AB 2010; pp. 12-29.
- Fisher E, Mahajan RL, Mitcham C. Midstream modulation of technology: Governance from within. *Bullet Science Technol Soc* 2006; 26: 485-96.
- Barben D, Fisher E, Selin C, *et al.* Anticipatory governance of nanotechnology: Foresight, engagement and integration. In: *The Handbook of Science and Technology Studies*. Hackett EJ, Amsterdamska O, Lynch M, Wajzman J, Eds. Cambridge, MA: MIT Press 2008; pp. 979-1000.
- Kickbusch I. Healthy Societies: addressing 21st century health challenges. Adelaide: Department of the Premier and Cabinet, State of South Australia 2008.
- Knoppers BM, Harris JR, Tassé AM, *et al.* Towards a data sharing Code of Conduct for international genomic research. *Genome Med* 2011; 3(7): 46.
- Dove ES, Joly Y, Knoppers BM. Trade-secret model: legal limitations. *Science*. 2011; 333(6049): 1575.
- Ozdemir V, Smith C, Bongiovanni K, *et al.* Policy and data-intensive scientific discovery in the beginning of the 21st century. *OMICS* 2011; 15(4): 221-5.
- Hotez PJ. New antipoverty drugs, vaccines, and diagnostics: a research agenda for the US President’s Global Health Initiative (GHI). *PLoS Negl Trop Dis* 2011; 5: e1133.
- Hey T, Tansley S, Tolle K. The Fourth Paradigm: Data-Intensive Scientific Discovery. Redmond, WA: Microsoft 2009. Available from: <http://research.microsoft.com/en-us/collaboration/fourthparadigm/> [Accessed October 27, 2011].
- Hather GJ, Haynes W, Higdon R, *et al.* The United States of America and scientific research. *PLoS One* 2010; 5(8): e12203.
- Kolker E. A vision for 21st century US Policy to support sustainable advancement of scientific discovery and technological innovation. *OMICS* 2010; 14(4): 333-5.

- [25] The Human Variome Project. Available from: www.humanvariomeproject.org [Accessed October 27, 2011].
- [26] Watkins D, Schwartzenuber JA, Ganesh J, *et al.* Novel inborn error of folate metabolism: identification by exome capture and sequencing of mutations in the MTHFD1 gene in a single proband. *J Med Genet* 2011; 48(9): 590-2.
- [27] Schofield PN, Eppig J, Huala E, *et al.* Research funding. Sustaining the data and bioresource commons. *Science* 2010; 330(6004): 592-3.
- [28] US National Science Foundation (NSF). Task Force Report Washington DC: NSF 2011. Available from: <http://www.nsf.gov/od/oci/taskforces/> [Accessed October 27, 2011].
- [29] Kolker E. The 2nd Data-Intensive Science Workshop Report (DISW2). Transition of bioinformatics to clouds: Opportunities and challenges. Washington DC: National Science Foundation 2011.
- [30] Abbott A. Rare-disease project has global ambitions. *Nature* 2011; 472: 17.
- [31] Editorial. What is the Human Variome Project? *Nat Genet* 2007; 39: 423.
- [32] Cotton RGH, Participants of the 2006 Human Variome Project Meeting. Recommendations of the 2006 Human Variome Project meeting. *Nat Genet* 2007; 39: 433-6.
- [33] Kohonen-Corish MR, Al-Aama JY, Auerbach AD, *et al.* How to catch all those mutations--the report of the third Human Variome Project Meeting, UNESCO Paris, May 2010. *Hum Mutat* 2010; 31(12): 1374-81.
- [34] Cotton RG, Al Aqeel AI, Al-Mulla F, *et al.* Capturing all disease-causing mutations for clinical and research use: toward an effortless system for the Human Variome Project. *Genet Med* 2009; 11(12): 843-9.
- [35] Kohonen-Corish MR, Macrae F, Genuardi M, *et al.* Deciphering the colon cancer genes - report of the InSiGHT-Human Variome Project Workshop, UNESCO, Paris 2010. *Hum Mutat* 2011; 32(4): 491-4.
- [36] Kolker E. Special issue on data-intensive science, *OMICS* 2011; 15(4): 197-8.
- [37] Abdelhak S, Adebamowo C, Adeyemo A, *et al.* Harnessing genomic technologies toward improving health in Africa: opportunities and challenges. Available from: http://h3africa.org/h3africa_whitepaper.pdf [Accessed Oct 27, 2011].
- [38] Human Variome Project Working Groups. Available from: <http://www.humanvariomeproject.org/index.php/activities/working-groups> [Accessed October 27, 2011].
- [39] Pepper MS. Launch of the Southern African Human Genome Programme. *S Afr Med J* 2011; 101(5): 287-8.
- [40] Editorial. Helping Hand. *Nature* 2011; 474(7353): 542.
- [41] Bissar-Tadmouri N, Tadmouri GO. Bibliometric analyses of biomedical research outputs in Lebanon and the United Arab Emirates (1988-2007). *Saudi Med J* 2009; 30(1): 130-9.
- [42] The Catalogue for Transmission Genetics in Arabs (CTGA) Database 2011. Available from: http://www.cags.org.ac/ctga_search.html [Accessed October 27, 2011].
- [43] Science-Metrix. 30 Years in Science. Montreal: Science-Metrix 2010. Available from: <http://www.science-metrix.com/30years-Paper.pdf> [Accessed October 27, 2011].
- [44] SCImago. SJR — SCImago Journal & Country Rank 2010. Available from: <http://www.scimagojr.com> [Accessed October 27, 2011].
- [45] Aziz RK, Kansal R, Abdeltawab NF, *et al.* Susceptibility to severe streptococcal sepsis: use of a large set of isogenic mouse lines to study genetic and environmental factors. *Genes Immun* 2007; 8(5): 404-15.
- [46] King Abdullah University of Science and Technology. Red Sea Expedition 2011. Available from: <http://krse.kaust.edu.sa/index.html> [Accessed October 27, 2011].
- [47] Keser I, Sanlioglu AD, Manguoglu E, *et al.* Molecular analysis of beta-thalassemia and sickle cell anemia in Antalya. *Acta Haematol* 2004; 111(4): 205-10.
- [48] Hizel C, Gök S, Sardas S, *et al.* Personalized and predictive medicine in Turkey: A symposium report of the Istanbul Working Group on Personalized Medicine, Istanbul, Turkey, September 10-12, 2009. *Curr Pharmacogenomics Person Med* 2009; 7(4): 297-301.
- [49] Gorakshakar AC, Colah RB. Cascade screening for beta-thalassemia: A practical approach for identifying and counseling carriers in India. *Indian J Community Med* 2009; 34(4): 354-6.
- [50] Hardy BJ, Séguin B, Singer PA, *et al.* From diversity to delivery: the case of the Indian Genome Variation initiative. *Nat Rev Genet* 2008; 9(Suppl 1): S9-14.
- [51] Dissanayake VHW. The Sri Lankan Personal Genome Project (SLPGP). Human Genome Meeting 2011. 4th Pan Arab Human Genetics Conference, Dubai, U.A.E., March 16, 2011.
- [52] Joly Y. Open biotechnology: licenses needed. *Nat Biotechnol* 2010; 28(5): 417-9.
- [53] Joly Y, Livingstone A, Dove ES. Moving beyond commercialization: Strategies to maximize the economic and social impact of genomics research. *Genome Canada GE3LS Series Policy Brief (No. 5) 2011 (in press)*.
- [54] Public Population Project in Genomics (P3G) 2011. Available from: <http://www.p3g.org> [Accessed on October 27, 2011].
- [55] Faraj S, Xiao Y. Coordination in fast-response organizations. *Management Science*, 2006; 52(8): 1155-69.
- [56] Majchrzak A, More PHB, Faraj S. Transcending knowledge differences in cross-functional teams. *Organization Science* 2011; orsc. 1110.0677 v1111.
- [57] Vaast E, Walsham G. Trans-situated learning: supporting a network of practice with an information infrastructure. *Information Systems Research* 2009; isre. 1080.0228 v1081.
- [58] Ostrom E. Collective Action and the Evolution of Social Norms. *Journal Econ Perspectives* 2000; 14(3): 137-58.
- [59] Ptolemy AS. Biomarker diversity, validation and clinical translation: Tenets for emergent personalized medicine initiatives. *Curr Pharmacogenomics Person Med* 2011; 9(4): 286-298.
- [60] Rohde H, Qin J, Cui Y, *et al.* Open-source genomic analysis of Shiga-toxin-producing *E. coli* O104:H4. *N Engl J Med* 2011; 365(8): 718-24.
- [61] Aziz RK, Saad R, Rizkallah MR. PharmacoMicrobiomics or how bugs modulate drugs: an educational initiative to explore the effects of human microbiome on drugs. *BMC Bioinformatics* 2011; 12(Suppl 7): A10.