Pharmacogenomics for Precision Medicine in the Era of Collaborative Co-creation and Crowdsourcing

Yasha Hasija · Jameel Ahmad Khan ·
Vinod Scaria

Published online: 22 May 2014
© Springer Science + Business Media New York 2014

Abstract The whole gamut of new technologies in the past decade has revolutionized DNA sequencing, making it cheaper, efficient, and scalable. The consequent big-data in genomics have posed new challenges and opportunities. The transformation of internet as a fabric that intertwines multiple technological and social layers and the rise of platforms that can organize and integrate massively parallel human activities have transformed the workplaces in many industries and offers a new opportunity in the area of genomics. In this short review, we discuss the state-of-the-art of crowdsourcing in genomics research with special focus on pharmacogenomics. We discuss the field, starting with an overview of technology and the major challenges. We also discuss a number of ongoing crowdsourcing approaches in the area of pharmacogenomics and personal genomics. We conclude with deliberating on the issues in genomics and how crowdsourcing could offer a plausible alternative to conventional approaches in genomics.

Keywords Personal genomes · Precision medicine · Pharmacogenomics · Crowdsourcing · Genomics · Clinical genomics

Introduction

The sequencing and creation of the blueprint of the Human Genome [1, 2] offered promise toward creating groundbreaking advances in the understanding of the Human Genome and genetic diseases. The decade following the announcement of completion of the Human Genome sequencing has seen drastic improvements in the scale and consequent reduction in the costs of sequencing. This has been particularly contributed by advancements in technology and in scale of nucleic acid sequencing, brought about by a spectrum of novel methodologies to miniaturize and significantly improve the scale of sequencing: dubbed as next-generation sequencing technologies [3]. Moreover, methodologies like microarray have complemented the scale of querying genomic regions. There have also been tremendous developments in understanding genetic variations among populations [4]. Apart from providing valuable mechanistic insight into genetic diseases, it has also contributed to understanding genetic selection, human migration, and disease predisposition [5, 6]. These advancements have also illuminated the organization and architecture of functional elements in the genome, thus leading to a new understanding on the potential role of genetic variations in modulating phenotypes. A major contribution to this acumen has been provided by the Encyclopedia of DNA elements (ENCODE) consortium [7•] and other allied projects including the NIH Epigenomics Roadmap project [8].

Recent years have seen a number of Human genomes sequenced from multiple countries representing multiple ethnic groups. This includes personal genomes from many countries and ethnic populations spread across the globe. Besides, selected sets of population have been represented in the 1000 Genomes Project [9•]. Significant number of

Y. Hasija
Department of Biotechnology, Delhi Technological University,
Bhawana Road, Shahbad Daultpur, New Delhi, India

J. A. Khan
Lifecode Technologies, B-1/E-24, Mohan Co-operative
Industrial Estate, Mathura Road, New Delhi 110 044, India

V. Scaria (✉)
GN Ramachandran Knowledge Center for Genome Informatics,
CSIR Institute of Genomics and Integrative Biology,
Mathura Road, New Delhi 110025, India
e-mail: vinods@igib.in

the world population which resides in the Asian and African continents and their genomic diversity have been poorly represented in the 1000 Genomes initiative [10]. Regional efforts have been initiated to fill in the gap of population genetic diversity. The Pan Asian Population Genomics Initiative (PAPGI) encompasses participants from over 17 countries aiming to create a high resolution map of population level genetic diversity within the Asian continent (www.papgi.org). The coming years would see a drastic improvement in resolution of genomic variations in these populations.

The advances in technology including miniaturization, parallelization, and improved methods of reading bases also promise to drastically reduce the cost of genome sequencing in near future, potentially enabling it as a clinical diagnostic tool. It is widely believed that the price of whole genome sequencing would drop approximately to USD 1000. This would have far-reaching implications in both the understanding of human genomic variations as well as the application of whole genome sequences in clinical applications. It has been speculated that the availability of high-quality human genomes at affordable costs could have far-reaching implications in clinical settings, and is widely believed to herald a new era of personalized medicine [11]. It has been argued that affordable clinical genome sequencing would enormously add to understanding genetic disorders, especially rare and inherited genetic diseases, and contribute largely to the genetic diagnosis of diseases. Moreover, Clinical genome sequencing can make a difference in better management of diseases based on evidence, otherwise called precision medicine.

The affordability and scale of genome sequencing and its application in clinical settings do not come without challenges. The major challenges include, first the management of humongous amount of data generated, which would require systematic efforts to store, organize, and annotate. Second, the analytics challenge, arising out of paucity of tools, resources and appropriate workforce to systematically parse and interpret the genomic information needs to be streamlined. A number of reviews have highlighted these issues in genomics. These two challenges are not unique to genomics and are shared by other data-intensive areas [3]. Many related industries, which rely on data-intensive tasks, have extensively used innovative approaches including crowdsourcing.

The developments in genomics have been in parallel with technological advances in many other areas including connectivity and the rise of collaboration as an enterprise at scale. The recent years have seen the transformation of internet from a communication platform to a fabric that intertwines multiple technological and social layers. This has impacted the way we gather, analyze, interpret data,

and implement decisions. Rise of social media and platforms that can organize and integrate massively parallel human activities has transformed the workplaces in many industries and offers a new opportunity in the area of genomics. It has been widely discussed how massive parallelization of human tasks could be a viable alternative to traditional approaches for collecting, analyzing, interpreting data, and resources [12]. These approaches, involving tasks which are distributed among a large number of individuals, have been popularly called crowdsourcing [13]. A variant of this methodology which includes analytical tasks spawned among a large group of individuals has been popularly called crowd-computing [14].

In this short review, we discuss the state-of-the art of crowdsourcing in genomics research with special focus on pharmacogenomics and precision medicine. We discuss the field in three distinct headings, starting with an overview of technology and the major challenges. We take cues from other areas of technologies where crowdsourcing and crowd-computing have been successfully applied. We also discuss a number of ongoing crowdsourcing approaches in the area of pharmacogenomics and personal genomics. We conclude with deliberating on the issues in genomics and how crowdsourcing could offer a plausible alternative to conventional approaches in genomics.

New Technologies

A number of recent technologies have provided the much needed starting point toward generating the required knowledge base for implementing translational application of genomics. Broadly, these technologies could be grouped into Genotyping technologies and the consequent genome-wide association studies; next-generation sequencing; and third, the computational analysis capability, curation of data sets, and other tools of functional genomics.

The availability of microarray technology that has enabled querying of large number of genetic variants at an ample scale has revolutionized the way we look at genetic variants and their association with diseases. This has led to hypothesis-free, genome-wide estimates of genetic associations with traits or disease phenotypes and discovery of novel/unknown genes related to a particular disease. Recently, the NHGRI GWAS Catalog has compiled over 5,000 variants from over 1,300 publications which have shown evidence of genotype–phenotype correlations using Genome Wide Association Studies [15]. Presently, these datasets encompass over 17 general traits, including a large number of distinct Human traits and diseases. A number of them are of pharmacological relevance with high odds-ratios. Easy availability of genetic variations from genome-scale association studies offers a new opportunity toward

applying them in clinical settings by detecting not only the susceptibility of an individual toward a particular disease/disorder at an early stage in life, but also prioritizing specific health and medical regimens best suited to the individual.

One of the major technological advancements, which followed the Human genome sequencing, is the availability of fast, efficient, and cheap methodologies for re-sequencing human genomes. These gamut of technologies have been based on massively parallel sequencing of short reads. These technologies have been popularly dubbed as Next Generation Sequencing technology. This has also been complemented by availability of faster algorithms and dwindling compute and storage costs. The initial years have seen sequencing of a number of individual genomes from different countries, populations, and ethnic groups. These include genomes from China [16], Japan [17], Korea [18], India [19], Sri Lanka [20], and Malaysia [21]. The 1000 Genome initiative includes 13 population groups in Phase I and additional 7 population groups in Phase II. Other regional initiatives aimed at covering regions/populations that have not been covered by the 1000 Genome initiative have taken shape. For example, the ambitious Pan-Asian Population Genomics Initiative encompasses the entire population of Asia [6]. These initiatives are expected to provide essential raw material for sequence level, and genome wide studies at global level. Together, these advances would pave the way for a new era of evidence-based medicine, known as precision medicine, based on pharmacological landscape of the individual to provide precise dosage and treatment schedule.

Recent years have also seen a paradigm shift in the drug discovery process. Traditional drug discovery processes were heavily dependent on chance and involved tedious screening of large compound libraries. Such methods usually take more than 10 years for the chance discovery of drugs. Also, they utilize a lot of resources, sometimes without giving any productive results. Computational modeling of biological activities has increasingly been used to prioritize molecules for screening [22–24]. Recent research, including from our lab, has extensively used data from previous high-throughput screens of large molecular libraries to create highly accurate in silico structure activity relationship models for prioritizing molecules with requisite activity. Such computational methods of molecular modeling and drug design have the potential to accelerate, automate, and make the drug discovery more successful and less expensive.

New Challenges

The availability of new technologies, both for the discovery of new genetic variants and associations with human traits

and the possibility of re-sequencing human genomes in clinical settings, have also created newer challenges in a number of fields ranging from data management and analysis, methodologies for prediction of phenotypes/traits and interpretation based on genomic data and of course, making technology affordable and easy to use. These challenges should also be seen in context with other challenges generally faced by the pharmaceutical companies on the ever-rising cost of drug discovery and rapidly dwindling rate of discovery of new drugs [25]. Technological challenges apart, the major challenge, which has not been sufficiently discussed, is making advancements in genomic technologies affordable, thus significantly improving the reach and utility of these technologies in improving the quality of life of large underprivileged populations in under-developed and developing worlds.

The availability of affordable whole genome or exome sequences at clinical settings is widely believed to impact healthcare and disease management, but also has been widely discussed to pose new challenges in the area of data management, mining, interpretation, and integration with health care systems. This is primarily contributed by the fact that functional correlates to a majority of human variations are unknown [26]. A large proportion of the markers revealed from Genome-wide association studies, especially for complex traits, do not have enough predictive power to be applied in clinical settings [27]. These studies nevertheless have significantly contributed to the understanding of biological pathways [28]. Furthermore, the numbers of genetic variants that show additional phenotypic correlates have been growing at an exponential phase. This necessitates constant re-mining and re-interpretation of the data necessary in real-time.

Another major challenge is the availability of a systematic-curated resource for genetic variants and their associations. This has majorly been a challenge in past since the datasets and reports are scattered in thousands of publications, and in non-standardized formats and different genome builds and versions. It would be an impossible task for researchers to compile and manually curate this vast information from literature sources, though there have been ample efforts to curate it using computational tools and methods. The creation of locus-specific variation databases (LSDBs), pioneered by the Human Genome Variation Society [29, 30] along with standard guidelines to report and disseminate various variations, has been a major effort toward systematic curation of information in a standardized format. Recent availability of tools and aggregators, including efforts from the NCBI and the Gen2Phen consortium [31], is worth mentioning. Fortunately, in the area of pharmacogenomics, a number of resources like DrugBank [32] and PharmGKB [33] curate complementary evidence for genetic markers and traits. Nevertheless, it has

not escaped our attention that neither of the databases has been able to cope with the deluge of information on pharmacogenetic markers and reports being published in the scientific literature. Apart from the availability of highly curated datasets, appropriate tools to systematically mine and compare variations from personal genomes is a necessity. Integration of the knowledge from various parts of the world, in different formats and versions, has posed a huge challenge in the pharmacogenomics research. It is a herculean task to manually curate such large amount of data and make it available in standardized formats for efficient mining and analysis.

Innovation Through Crowdsourcing and Open Source

The recent years have seen the development of a new organizational framework which is popularly called as crowdsourcing. The crowdsourcing taps into collective competence of the network of large number of individuals present outside of physical organizational frameworks. It is a paradigm shift from the traditional organizational structure and is particularly accelerated by the ubiquitous internet and connectivity. These organizational frameworks or rather networks have been increasingly tapped for tasks that were earlier considered impossible. For example, using crowdsourcing model, Amazon Mechanical Turk (<http://www.mturk.com>) could organize large human networks and spawn activities across the networks. In fact, crowdsourcing has presently become norms in many industries including software development and testing and marketing.

Crowdsourcing may be defined as a product or service delivery pipeline outsourced to volunteers, often amateurs. The defining characteristic of crowdsourcing is that the task is carried out by people unknown to the initiator. In crowdsourcing, initiator (whether individual, companies, or non-profit organizations) do not need to hire many employees to accomplish the task. Instead, major portion of the task is carried out by volunteers in their spare time. Usually, the volunteers may not be even paid for the task they do. Instead, the primary motivators for the participation in crowdsourcing projects are intrinsic factors like peer recognition, social contact, intellectual stimulation, skill development, and autonomy. Therefore, crowdsourcing is turning out to be a very low cost method to accomplish highly complex tasks [13].

Innovative solutions are being achieved through crowdsourcing in areas as diverse as genomics, engineering, predictive analytics, software development, video games, mobile apps, and marketing. The innumerable numbers of apps in Apple, Android, and Nokia stores are probably the best examples of crowdsourcing. The crowd-sourced solutions are not limited to low end product

development. In fact, it has been shown that, in terms of novelty and customer benefit, the crowd-sourced solutions may surpass those of professionals. It is being recognized that only limitation in crowdsourcing is posing right challenges and taking it to the right crowd of volunteers. Successful launch of many businesses based on crowd-sourced design at OpenIDEO is a glaring example of the power of crowdsourcing. A lot can be learned from success stories of other fields and applied into biomedical research.

Crowdsourcing in Biomedical Research

Last few decades have seen steady increase in healthcare costs without significant improvement in clinical outcomes. The condition is even more complex in case of the so-called “neglected diseases” which primarily affect people in the third-world countries. While it is very difficult for Pharma companies to develop drugs for these diseases, which will return the investments, the academic R&D is too fragmented to develop drugs on their own [25, 34]. It shows that the cost burden of traditional model of drug development is not sustainable, anymore. The situation appears to have reached to an inflection point where paradigm shifts in traditional approaches to biomedical research for drug discovery are inevitable.

The researchers in biomedical sciences and the volunteers in crowdsourcing share a lot in common, in terms of motivation to accomplish the tasks. In both fields, the primary motivators are intangible intrinsic factors like peer recognition, intellectual stimulation, skill development, etc. Therefore, biomedical research is one of the ideal areas to explore crowdsourcing model for accomplishing complex tasks. It is not surprising that it is already being used in many areas of biomedical research, especially those in which tasks can be accomplished online. It is becoming an increasingly popular approach to solve highly complex biomedical problems and emerging as a powerful, low cost alternative to the traditional approaches of biomedical research.

Crowdsourcing has been innovatively utilized to accomplish many low cost biomedical solutions. For example, David Baker’s group at University of Washington, have developed a protein-folding video game Foldit. The game utilizes human problem-solving skills to solve complex crystal structures of proteins. The Foldit volunteers, many of them with no training in biology, have produced accurate model of M-PMV retroviral protease [35]. It is claimed that gamers accomplished this task in just 3 weeks which eluded researchers for 15 years. The structure can provide new insights for development of anti-HIV drugs. In another recent accomplishment of Foldit, the users were challenged to computationally remodel the

Diels-Alderase enzyme backbone, for better enzymatic activity. The gamers came up with a 13-residue insertion that could increase enzyme activity by more than 18-fold. Through an approach similar to crowdsourcing the image analysis program, ImageJ [36] (known as NIH Image in its previous incarnation) has emerged as a powerful alternative to many commercial image analysis software. The power of ImageJ relies in its user contributed plugins. The ImageJ and many of its curated plugins are freely available as an Open Source software. In terms of utility, they match or even surpass most commercially available image analysis software which cost thousands of dollars.

As stated earlier, the crowdsourcing model has been especially successful for projects amenable to online collaborations. Therefore, it is not surprising that paradigm of crowdsourcing has shown immense potential in the area of genomics or rather the omics sciences, one of the largest big-data enterprises in life sciences. The Genomics organizations like 23andMe (www.23andme.com), PatientsLikeMe (<http://www.patientslikeme.com>), Quantified Self (www.quantifiedself.com), DIYgenomics (www.diygenomics.org) have pioneered crowd-sourced research. They have started publishing data on disease research, drug response, and user experience for consumer Genomics products. Many researcher-organized genome annotation projects have been successfully completed at a fraction of cost and in significantly less time by large number of voluntary curators. These success stories are just the beginning. As mentioned earlier, the crowdsourcing has unimaginable potential for innovations, limited only to posing right challenges and taking it to the right crowd of volunteers. Some successful examples in Genomics field in more detail are elaborated below.

Crowdsourcing Genomics and Genomic Analysis

The present scale and throughput of genome sequencing far exceeds the analysis capability, and has been increasingly discussed as the next major challenge in this area. Harnessing the power of genomics would require careful and large-scale integration of seemingly disparate datasets derived from multiple-omics technologies to provide insights and further integrate insights or models to provide knowledge [37]. Open initiatives like the ones for Ash Dieback [38] aim at harnessing the expertise and speed of crowd-sourced analysis on Openly available genome and transcriptome datasets. One of the earliest initiatives include the annual Genetic Analysis Workshops (<http://www.gaworkshop.org>) where genetic analysis methodologies of a given genetic problem and dataset have been crowd-sourced and compared to identify the best possible methodology.

Similar efforts, with the help of large-number of students and custom web tools, have been explored for

metagenomic analysis. In addition, crowd-sourced mycobacteriophage identification has been recently explored. Recently, microbiome collection and analysis have been crowd-sourced through probably the first citizen science driven Human microbiome analysis (www.ubiome.com/). In the area of Human genomics, recent report shows an elegant example of crowd-sourced and self-reported traits and genomic datasets could be extensively used for discovering novel associations.

SNPedia and OpenPGx

Annotation and interpretation of personal genomes would also require standardized and well-curated evidence from the peer-reviewed literature. Single-nucleotide polymorphisms (SNPs) have been long known to confer phenotypic changes and hence medically important consequences. Systematic curation and collection of these SNPs is an important part of analyzing personal genomes, and a number of projects largely involving collaborative editing and contribution similar to the Wikipedia model have been spearheaded. SNPedia is one such wiki-based resource which collects information on SNPs from peer-reviewed journals and provides it in a semantic web form for easy access and interpretation [39]. While this manuscript was being written, SNPedia houses 38,492 SNPs in its database presently. SNPedia is already used by many resources for clinical interpretation of data from genome sequencing and other test. While SNPedia only provides genotypic information related to any SNP, OpenPGx [40], in addition to giving genotypic information, also curates and stores pharmacogenomically significant information regarding that SNP. Similar analytical challenges for omics data have been attempted recently including the DREAM challenges (www.the-dream-project.org). A similar effort for genomic data analysis has been put forward by the Boston Children's Hospital called the CLARITY (Children's Leadership Award for the Reliable Interpretation and appropriate Transmission of Your genomic information) challenge (<http://www.childrenshospital.org/research-and-innovation/research-initiatives/clarity-challenge>) for interpretation of genomic data, with the vision to be able to identify best analytic methodologies for genome interpretation. Similar approaches for technological innovations, not limited to genomics, have also been attempted as part of the X-prize (<http://www.xprize.org>) Initiative.

Personal Genome Project

Notwithstanding the privacy concerns, human genome sequence and analysis have also been amenable to crowdsourcing. The personal genome project (PGP) [41] aims to create a publicly available and crowd-sourced

repertoire of human genome information associated with self-reported traits and medical records. The present repertoire encompasses a total of 1,000 volunteers; a small subset of whose genomic information is also available to the public.

OpenSNP

Another such crowd-sourced effort is OpenSNP (www.opensnp.org) where a user freely shares their genotype information over the web for everyone to access. One can simply upload their genotype file from any Direct-to-consumer (DTC) genetic testing company and have their genotypes analyzed. Annotation of the genotypes is done using a host of resources like SNPedia and the peer-reviewed literature.

The Pan-Asian Population Genomics Initiative

The Pan-Asian population genomics initiative (www.papgi.org) is a unique experiment of a loosely knit community of researchers, strongly motivated by the spirit of collaboration and sharing of resources toward realizing the application of genomics to improve the quality of life and health care through understanding genome diversity of Asia. Incidentally, Asia is home to a fifth of world population and home to a rich pool of ethnically, socially, linguistically, and genetically diverse populations with a vivid history of migration and admixture. The open consortium is built upon the strong network of over 100 collaborators from over 16 countries and aims to create one of the most comprehensive curated catalogs of genetic variations in Asian populations. The closely-knit consortium is built on the principles of mutual trust, cooperation, sharing of resources, transparency, and accountability, without the binding of formal memoranda or legal obligations, and thus reflects the true spirit of scientific collaboration in genomics.

The Impact of Crowdsourcing in Clinical Genomics and Precision Medicine

Crowdsourcing of genome data is increasingly being adopted by individuals, though it has not become the mainstream yet. One example worth mentioning is the crowd-sourced “*Corpasome*,” where the genomic and metagenomic data of individuals of a family have been made available for crowd-sourced analysis and interpretation [42•]. The idea stemmed from the fact that no single organization or individual is in a position to provide the most comprehensive analysis and interpretation of personal genome data. The diverse skillsets of the genomics community could provide ways of richly annotating and

interpreting genomic variations. Another example has been the involvement of crowd-sourced cohorts of individuals and genomic datasets for personalized medicine. The impact of such approaches is two-fold; on one end they make use of the rich knowledge base of the community, which was not possible before, and on the other end, they could access the resource at a fraction of cost of the service if it was existent [13]. This would possibly impact the field in distinct ways. The researchers would immensely benefit from the availability of well-curated and characterized genome-scale information for integration, development, and standardization of new methods and baseline data sets for scientific discovery at minimal costs. The individuals/participants and their doctors and care-givers gain from saving significantly on the cost of analysis, while being able to tap an enormous knowledge base and skill set, which was previously not accessible. The funders also stand to gain from saving cost on duplication of efforts, primarily caused by de-identified data sets generated by multiple researchers, while gaining immense benefit from the possibility to integrate datasets for discovery. The planning bodies also stand to gain from the ready availability of epidemiological data which could help make accurate and precise decisions. The drawbacks of crowd-sourcing approaches would also be worth mentioning. The major ones include intentional malicious activity, inadvertent errors, and data inconsistencies which could potentially affect the quality of the final resource, methodology or product. Many internal correction systems, including automated bots for consistency checks and manual regular scanning of articles, have been implemented in many large crowd-sourced and collaborative editing platforms like Wikipedia. It should also be noted that the success of the crowdsourcing initiative would also depend largely on the incentive system and motivation of the stakeholders involved. It should be emphasized that the field has a fair share of examples for failures due to the lack of a proper incentive system and draining motivation.

Future Perspective

The future of pharmacogenomics for personalized disease management would see creation of pharmacogenomic maps with different granularities being created. On one end, population-scale genome sequencing would provide the much necessary baseline for population-scale differences in pharmacogenetic markers and would provide for disease management strategies for clinicians and policy-makers alike. Personal genome maps for population-scale applications could be created by genotyping or sequencing large number of individuals in the population.

Another major area where pharmacogenomic maps could be potentially used would be to rationalize therapies, especially cost-effective regimens and keep low-cost drugs in the market. This approach would attain more significance in coming years, with the increasingly dwindling drug pipelines and increasing cost of research and development in the drug industry. In a more futuristic scenario, computational models could be used to plan combination regimens *in silico* or even used to model effects of new and investigational drugs even before human clinical trials. The computational power made available by volunteers through crowd-computing approaches may be utilized to reduce the costs of *in silico* modeling. This would also have the potential to drastically reduce the cost of failures and subsequently reduce the cost of drug discovery as a whole.

The major impact of pharmacogenomic maps would be in how they could be potentially integrated into Electronic Medical Records, Decision Support Systems and Research Databases [43]. The future could see new standards enabled with the highest encryption, offering tiered release of relevant information to all stakeholders- the patient, physicians, researchers, and planners. Such integration also offers an immense opportunity to provide real-time evidence-based assessment and planning of the disease prevention and management. It has also not escaped our attention that data integration and tiered access to data and crowd-sourced or participatory research could provide for a new framework of collaborative co-creation of knowledge potentially at a fraction of the cost incurred today, while providing enormous health benefits to the community as a whole.

These futuristic developments are not without concerns of privacy. It is imperative that new frameworks for consultative and collaborative ethical and regulatory frameworks emerge, and would increasingly be one of the major challenges in the coming years, as the public knowledge and perception on access to genomic information improve. Nevertheless this could also be compounded by the improving accuracy of predictions of phenotypes based on computational heuristics and well-curated datasets. Keeping the technological challenges aside, the public perception of genomics and its utility to improve quality of life and health care is of significant importance. We also foresee the regional or rather the east–west differences in perceptions of privacy to impact the acceptability of crowdsourcing in genomics.

Acknowledgments The authors acknowledge discussions with Dr. Sridhar Sivasubbu and Dr. S Ramachandran and members of the OpenPGx consortium and PAPGI consortium, which have significantly enriched the content of the Manuscript. This work was funded by the Council of Scientific and Industrial Research, India, through Grant CARDIOMED (BSC0122).

Disclosure Y Hasija, JA Khan, and V Scaria all declare no conflicts of interest.

Human and Animal Rights and Informed Consent This article does not contain any studies with human or animal subjects performed by any of the authors.

References

Papers of particular interest, published recently, have been highlighted as:

- Of importance
- Of major importance

1. Venter JC, Adams MD, Myers EW, et al. The sequence of the human genome. *Science*. 2001;291:1304–51.
2. Consortium” “the International Human Genome Sequencing. Initial sequencing and analysis of the human genome. *Nature*. 2001;409:860–921.
3. Metzker ML. Sequencing technologies—the next generation. *Nat Rev Genet*. 2010;11:31–46.
4. International T, Consortium H. A haplotype map of the human genome. *Nature*. 2005;437:1299–320.
5. Article R. Genetic landscape of the people of India: a canvas for disease gene exploration. *Genome*. 2008;87:3–20.
6. Abdulla MA, Ahmed I, Assawamakin A, et al. Mapping human genetic diversity in Asia. *Science*. 2009;326:1541–5.
7. •• ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature*. 2012;489:57–74 *The ENCODE consortium paper discusses the outline and major results from the large international collaborative initiative.*
8. Bernstein BE, Stamatoyannopoulos JA, Costello JF, et al. The NIH Roadmap Epigenomics Mapping Consortium. *Nat Biotechnol*. 2010;28:1045–8.
9. •• The 1000 Genomes Project Consortium. An integrated map of genetic variation from 1,092 human genomes. *Nature*. 2012;491:56–65 *This landmark paper provides a glimpse of the population level variability of humans.*
10. • Lu D, Xu S. Principal component analysis reveals the 1000 Genomes Project does not sufficiently cover the human genetic diversity in Asia. *Front Genet*. 2013. doi:[10.3389/fgene.2013.00127](https://doi.org/10.3389/fgene.2013.00127). *This paper discusses the issues of comprehensiveness of the 1000 Genomes Project in uncovering the Human genetic diversity.*
11. Highnam G, Mittelman D. Personal genomes and precision medicine. *Genome Biol*. 2012;13:324.
12. Nicholson N (2012) Crowdsourcing. *Manag Today* 18.
13. Malone TW, Laubacher R, Dellarocas C (2009) Harnessing crowds: mapping the genome of collective intelligence. *Elements* 1–20.
14. Murray DG, Yoneki E, Crowcroft J, Hand S (2010) The case for crowd computing. In: *MobiHeld '10 Proceedings of the second ACM SIGCOMM workshop on Networking, systems, and applications on mobile handhelds*.
15. Hindorf LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, Collins FS, Manolio TA. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci USA*. 2009;106:9362–7.
16. Wang J, Wang W, Li R, et al. The diploid genome sequence of an Asian individual. *Nature*. 2008;456:60–5.

17. Fujimoto A, Nakagawa H, Hosono N, et al. Whole-genome sequencing and comprehensive variant analysis of a Japanese individual using massively parallel sequencing. *Nat Genet.* 2010;42:931–6.
18. Ahn S-M, Kim T-H, Lee S, et al. The first Korean genome sequence and analysis: full genome sequencing for a socio-ethnic group. *Genome Res.* 2009;19:1622–9.
19. Patowary A, Purkanti R, Singh M, Chauhan RK, Bhartiya D, Dwivedi OP, Chauhan G, Bharadwaj D, Sivasubbu S, Scaria V. Systematic analysis and functional annotation of variations in the genome of an Indian individual. *Hum Mutat.* 2012;33:1133–40.
20. Dissanayake VHW, Samarakoon PS, Scaria V, Patowary A, Sivasubbu S, Gokhale RS. The Sri Lankan Personal Genome Project: an overview. *Sri Lanka J Bio-Med Informatics.* 2011;2:4–8.
21. Salleh MZ, Teh LK, Lee LS, et al. Systematic pharmacogenomics analysis of a malay whole genome: proof of concept for personalized medicine. *PLoS One.* 2013;8:e71554.
22. Jamal S, Periwal V, Consortium OSDD, Scaria V. Predictive modeling of anti-malarial molecules inhibiting apicoplast formation. *BMC Bioinformatics.* 2013;14:55.
23. Jamal S, Periwal V, Scaria V. Computational analysis and predictive modeling of small molecule modulators of microRNA. *J Cheminform.* 2012;4:1–9.
24. Periwal V, Rajappan JK, others. Predictive models for anti-tubercular molecules using machine learning on high-throughput biological screening datasets. *BMC Res Notes.* 2011;4:504.
25. Munos B. Lessons from 60 years of pharmaceutical innovation. *Nat Rev Drug Discov.* 2009;8:959–68.
26. Scriver CR. After the genome—the phenome? *J Inher Metab Dis.* 2004;27:305–17.
27. Goldstein DB. Common genetic variation and human traits. *N Engl J Med.* 2009;360:1696–8.
28. Hirschhorn JN. Genomewide Association Studies—illuminating biologic pathways. *N Engl J Med.* 2009;360:1699–701.
29. Cotton RGH, Horaitis O (2000) Human Genome Variation Society. eLS.
30. Horaitis O, Cotton RGH. The challenge of documenting mutation across the genome: the human genome variation society approach. *Hum Mutat.* 2004;23:447–52.
31. Thorisson GA, Lancaster O, Free RC, Hastings RK, Sarmah P, Dash D, Brahmachari SK, Brookes AJ. HGVbaseG2P: a central genetic association database. *Nucleic Acids Res.* 2009;37:D797–802.
32. Wishart DS, Knox C, Guo AC, Cheng D, Shrivastava S, Tzur D, Gautam B, Hassanali M. DrugBank: a knowledgebase for drugs, drug actions and drug targets. *Nucleic Acids Res.* 2008;36:D901–6.
33. Sangkuhl K, Berlin DS, Altman RB, Klein TE. PharmGKB: understanding the effects of individual genetic variants. *Drug Metab Rev.* 2008;40:539–51.
34. Munos B. Can open-source drug R&D repower pharmaceutical innovation? *Clin Pharmacol Ther.* 2010;87:534–6.
35. Eiben CB, Siegel JB, Bale JB, Cooper S, Khatib F, Shen BW, Players F, Stoddard BL, Popovic Z, Baker D. Increased Diels-Alderase activity through backbone remodeling guided by Foldit players. *Nat Biotech.* 2012;30:190–2.
36. Schneider CA, Rasband WS, Eliceiri KW. NIH Image to ImageJ: 25 years of image analysis. *Nat Methods.* 2012;9:671–5.
37. Editor. Community engagement. *Nat Rev Microbiol.* 2013;11: 219.
38. Maclean D, Yoshida K, Edwards A, et al. Crowdsourcing genomic analyses of ash and ash dieback—power to the people. *Gigascience.* 2013;2:2.
39. • Cariaso M, Lennon G. SNPedia: a wiki supporting personal genome annotation, interpretation and analysis. *Nucleic Acids Res.* 2012;40:D1308–D1312. *An open source resource for variant annotation.*
40. Pasha A, Scaria V. Pharmacogenomics in the era of personal genomics: a quick guide to online resources and tools. In: Barh D, Dhawan D, Ganguly NK, editors. *Omics for personalized medicine.* Springer; 2013. p. 187–211.
41. Church GM. The personal genome project. *Mol Syst Biol.* 2005;1(2005):0030.
42. • Corpas M. Crowdsourcing the corpasome. *Source Code Biol Med.* 2013;8:13. *An example for crowdsourcing genome analysis.*
43. Kawamoto K, Lobach DF, Willard HF, Ginsburg GS. A national clinical decision support infrastructure to enable the widespread and consistent practice of genomic and personalized medicine. *BMC Med Inform Decis Mak.* 2009;9:17.