



A Review on the Challenges in Indian Genomics Research for Variant Identification and Interpretation

*Sandhya Kiran Pemmasani**, *Rasika Raman*, *Rajkishore Mohapatra*,
Mathukumalli Vidyasagar and *Anuradha Acharya*

Research and Development Division, Mapmygenome India Limited, Hyderabad, India

OPEN ACCESS

Edited by:

Gustavo Glusman,
Institute for Systems Biology (ISB),
United States

Reviewed by:

Ignacio Macpherson,
International University of Catalonia,
Spain
Stylios Chatzichronis,
National and Kapodistrian University
of Athens, Greece

*Correspondence:

Sandhya Kiran Pemmasani
drsandhyakiran@mapmygenome.in

Specialty section:

This article was submitted to
ELSI in Science and Genetics,
a section of the journal
Frontiers in Genetics

Received: 23 December 2019

Accepted: 24 June 2020

Published: 22 July 2020

Citation:

Pemmasani SK, Raman R,
Mohapatra R, Vidyasagar M and
Acharya A (2020) A Review on the
Challenges in Indian Genomics
Research for Variant Identification and
Interpretation. *Front. Genet.* 11:753.
doi: 10.3389/fgene.2020.00753

Today, genomic data holds great potential to improve healthcare strategies across various dimensions – be it disease prevention, enhanced diagnosis, or optimized treatment. The biggest hurdle faced by the medical and research community in India is the lack of genotype-phenotype correlations for Indians at a population-wide and an individual level. This leads to inefficient translation of genomic information during clinical decision making. Population-wide sequencing projects for Indian genomes help overcome hurdles and enable us to unearth and validate the genetic markers for different health conditions. Machine learning algorithms are essential to analyze huge amounts of genotype data in synergy with gene expression, demographic, clinical, and pathological data. Predictive models developed through these algorithms help in classifying the individuals into different risk groups, so that preventive measures and personalized therapies can be designed. They also help in identifying the impact of each genetic marker with the associated condition, from a clinical perspective. In India, genome sequencing technologies have now become more accessible to the general population. However, information on variants associated with several major diseases is not available in publicly-accessible databases. Creating a centralized database of variants facilitates early detection and mitigation of health risks in individuals. In this article, we discuss the challenges faced by genetic researchers and genomic testing facilities in India, in terms of dearth of public databases, people with knowledge on machine learning algorithms, computational resources and awareness in the medical community in interpreting genetic variants. Potential solutions to enhance genomic research in India, are also discussed.

Keywords: clinical genomics, variant classification, Indian genomics research, Indian genomic databases, machine learning

INTRODUCTION

Dynamic migration history, ethnic and genetic diversity and a high degree of consanguinity contribute to the complex and heterogeneous nature of the Indian population. There are many known genetic diseases affecting different population subgroups and insufficient scientific resources to diagnose and treat them (Aggarwal and Phadke, 2015; GUARDIAN Consortium et al., 2019). Large-scale genetic studies in Indian patients are required to study disease-causing mutations

and to develop personalized treatment methods. Another important aspect is accurate analysis and interpretation of genetic data. While tried and tested statistical methods work fairly well for biomarker discovery, advanced solutions like machine learning algorithms bring a promise of genomics driven clinical solutions. In this article, we discuss the current scope of Indian genomics in healthcare, challenges in scientific resources and data analysis, and solutions to enhance genomic medicine in India.

CURRENT STATE OF GENETIC TESTING

Genetic testing in India has evolved in leaps and bounds in the past decade. Currently, there exist DNA-based tests that address multiple concerns in healthcare, from disease prevention to molecular diagnosis (Kar and Sivamani, 2016). In the case of preventive healthcare, genetic tests estimate the lifetime risk of disease, predisposition to biological traits and health parameters (Mohan et al., 2011). They also analyze a person's response to drugs in terms of efficacy and risk for adverse reactions. These tests are primarily used as screening tools for establishing an effective strategy to reduce disease risk, delay, or avoid symptoms and manage existing conditions. Diagnostic genetic tests, on the other hand, help in identification of the molecular cause of the disease. These tests are used to confirm known or suspected diagnosis, carrier status determination, identification of at-risk genetic relatives, optimize treatments, and clinical decisions (Gupta et al., 2017; Aravind et al., 2019; Uttarilli et al., 2019). There are different types of diagnostic genetic tests currently available in India such as single-gene and multigene testing, exome, and genome sequencing, carrier and newborn screening (Puri et al., 2017; Singh et al., 2018). Other types of tests include those which assess reproductive risk, such as prenatal testing and preimplantation genetic diagnosis (Dada et al., 2008).

Challenges and Limitations

Understanding the need of the patient is the key for determining the right genetic test. The biggest hurdles faced by clinicians are genetic data interpretation, finding genetic links for complex conditions, and lack of actionable genetic information. In certain cases of complex conditions, such as cancer, an array of genetic tests might be ordered to determine the genetic cause (Prabhash et al., 2019). However, no findings may come to light, thus posing a challenge for the patient and the clinician. The accuracy and precision of genetic tests lies in the translation of genetic findings into clinical outcomes. In the absence of information on genotype-phenotype correlations, genetic test results might be inconclusive.

GENETIC DATA INTERPRETATION FOR INDIAN PATIENTS

Genetic diagnosis via clinical sequencing (e.g., genome-, exome-, single-, or multi-gene) is the front-line test recommended for many inherited diseases (Verma et al., 2018; Ganapathy et al., 2019). Establishing the genetic cause of disease is vital for

patient care and treatment, and hence clinical findings must be reported with high precision and accuracy (Singh et al., 2016). All clinical reporting protocols are required to adhere to standards set by American College of Medical Genetics (ACMG), for proper classification of variants and subsequent disclosure to patient/clinician (Richards et al., 2015). As per ACMG guidelines, in order to differentiate benign and pathogenic variants, a detailed study of the variant's clinical significance is required. This includes multiple criteria such as variant frequency, location in or near the gene, mechanism of said gene, effect of variant on protein domain or function, hotspot, or nearby mutations if any, etc. Apart from these, evidence of the variant having caused the disease in patients with similar clinical phenotype is essential to establish pathogenicity.

Challenges and Limitations

Currently, there is a dearth of publicly available resources that provide an extensive list of clinically significant variants in Indian patients, for several genetic diseases. In the absence of published literature for a particular variant, which clearly is not benign, the variant gets classified as a variant of uncertain significance (VOUS). Interpreting VOUS is often challenging as they are not actionable, yet hold potential for establishing pathogenicity. For accurate classification and high-precision reporting of genetic variants, it is vital that geneticists and scientists have access to information on the complete spectrum of variants and mutations in Indian patients (Rajasimha et al., 2014; Genomics and other Omics tools for Enabling Medical Decision, 2019). Only the most relevant mutations are listed in databases like OMIM, which use selection criteria such as frequency, phenotype, significance, disease mechanism, and inheritance, etc.

Case Study: Retinitis Pigmentosa

Retinitis Pigmentosa (RP) represents a very large group of eye disorders, with different clinical features, and symptoms. RP can be inherited in an autosomal dominant, autosomal recessive, or X-linked manner. Genetic diagnosis of RP helps in establishing genetic cause of disease, screening in at-risk family members and clinical management. But there are limited studies which report population specific mutations in Indian RP patients (Table 1).

Let us examine a case study of a 42 year old male reported with personal medical history of RP, who had been diagnosed at 16 years of age. His sister and two paternal cousins were also affected with RP. Married to a non-consanguineous partner, there were no clinical conditions in his children – 16 year old son and 12 year old daughter. Exome sequencing was done at Mapmygenome (Mapmygenome, 2020) to identify disease causing gene mutations associated with RP. Exome analysis revealed a heterozygous missense variant in exon 4 of the *NR2E3* gene. The observed variant is not reported as a variant in the normal samples of 1000 Genomes database and has a minor allele frequency of 0.018% in the gnomAD database. The variant is conserved across the species and *in silico* prediction by Mutation taster was found to be damaging. Another missense variant in this gene, Pro152Ser, has previously been reported with “retinitis pigmentosa 37,” and “retinitis pigmentosa (recessive)” in Clinvar (Clinvar, 2019) as VOUS. In Indian Genetic Disease Database

TABLE 1 | Information on summary of mutation studies in Indian RP patients.

Number of patients	Significant findings	Gene	Method	Variant information available	References
14 Families (autosomal recessive RP) and 100 cases (sporadic RP)	12 Novel mutations	EYS	Whole exome sequencing of proband followed by targeted analysis in family members	Yes	Di et al., 2016
1 Family and 100 cases (sporadic)	4 Novel mutations	CRB1		Yes	Yang et al., 2016
1 Family (autosomal recessive RP)	1 Novel mutation	MERTK		Yes	Bhatia et al., 2018b
1 Family (non-syndromic autosomal dominant RP)	1 Novel mutation	PRPF31	Targeted sequencing	Yes	Bhatia et al., 2018a
1 Family (autosomal recessive RP)	1 Novel mutation	MERTK	Microarray	Yes	Srilekha et al., 2015
2 Families (autosomal recessive RP) and 100 cases (sporadic RP)	1 Novel mutation	FAM161A	Whole exome sequencing	Yes	Zhou et al., 2015
2 Families (autosomal recessive RP)	1 Novel mutation	NR2E3	Microarray	Yes	Kannabiran et al., 2012
101 Cases (48 isolated cases and 53 autosomal dominant RP cases)	2 Novel mutations	PRPF31	Capillary sequencing	Yes	Gandra et al., 2008

(IGDD), which is the first patient based genetic disease database of India (Pradhan et al., 2011) only seven genes have been mapped for RP. No variant was reported from the *NRE23* gene in IGDD. There are no other databases which have mutation information from Indian RP patients. Since there is no functional or published study of *NR2E3* mutations in Indian RP patients, Mapmygenome had classified the variant in exon 4 of the gene as a VOUS. There is insufficient evidence to establish this variant's pathogenicity.

AVAILABLE RESOURCES WITH GENOTYPE-PHENOTYPE ASSOCIATIONS IN INDIAN POPULATION

Databases which host information on gene variants and associated diseases help genome analysts to make clinically significant and medically actionable inferences. However, most of the publicly available Indian databases are incomplete. This can be attributed to legal, ethical, financial, or administrative procedures due to which a lot of key parameters do not get recorded. Some of the Indian-specific databases, along with their scope and utility, have been discussed below.

Index-dB

A database of exonic variants from normal individuals of Indian sub-continent (Ahmed et al., 2019). It is a user-friendly database with a querying feature and a browser to search for the variants. But the current version is based only on 109 individuals and is still under development.

TMC-SNPdB

Contains variants generated from exome data of normal samples derived from tongue, gall bladder, and cervical cancer patients of Indian origin (Upadhyay et al., 2016). The major limitation of

the database is not only the sample size of 62, but also the way the variants were processed. The COSMIC database was used to filter out somatic variants, because of which some novel Indian variants might have got filtered out.

SAGE

A repository of genetic variants derived through an integration of six datasets comprising 1213 South Asian genomes and exomes (Judith et al., 2018). It contains more than 154 million variants, out of which 69 million are novel variants. Though this a comprehensive database of South Asians, it should be enriched with region or ethnicity specific datasets within South Asia.

Indian Genetic Disease Database (IGDD)

A curated database of variants associated with diseases prevalent in Indian population (Pradhan et al., 2011). Diseases were categorized into different therapeutic areas. The current version of the database covers 104 diseases with a total of ~3500 patients. Further enrichment is required to cover more diseases in the population.

Indian Genome Variation Database (IGVdB)

This was started as a consortium activity in 2003, with the goal to create a variation database of Indian population (Indian Genome Variation Consortium, 2005; Narang et al., 2010). However, this database does not contain disease-variant associations, which are helpful in interpreting the data obtained from genetic tests.

GWAS Central – India

A genotype-phenotype association database with summary level findings from genetic association studies (Indian GWAS, 2010). Lack of regular updates and absence of extensive data points for genetic diagnosis, make this database a less effective

tool for clinicians, or bioinformaticians, thereby limiting its clinical utility.

Indian SNP Data

Contains genotype data of 871,771 SNPs, obtained from 15 Dravidian trios, and 13 Indo-European trios (Indian SNP, 2020). Browser and query features are not available for this database. Files can be downloaded for academic and research purposes only. Although it was initially developed as a reference panel for Indians, it has limited data and the work is still in progress.

Genotype/Phenotype DB

This database contains genotype and phenotype data of Indian population along with their demographic details (CCMB, 2020). Browser and query features are not available for this database. Commercial organizations are strictly prohibited from using the data.

Indigen Project

This is an initiative from Council of Scientific and Industrial Research (CSIR) for whole genome sequencing of 1000 Indian genomes, across diverse ethnic groups, with the goal to enable clinical applications in rare genetic diseases. This is an initiative which is yet to see fruition and is yet to be publicly available for the scientific community (IndiGen, 2020).

The above databases have not been presented in a way that allows the user to understand the pathogenicity of variants. Genomics companies like Mapmygenome (Mapmygenome, 2020), do not have access to most of such databases. A centralized database curated from Indian patients, for different diseases, would help in precise reporting and clinical decision making.

Publicly available data and results generated from genome wide association studies (GWAS) can also be utilized in interpreting the variants and in identifying new variants. There are case-control association studies done on Indian population, for majorly occurring diseases – Type 2 Diabetes, cardiovascular diseases and cancers (Chauhan et al., 2010; Nagrani et al., 2017; Bellary et al., 2019). Polygenic Risk Scores (PRS) developed from GWAS act as prognostic indicators in preventive healthcare. However, reliability of the results depends on the algorithm used and the data available.

MACHINE LEARNING (ML) ALGORITHMS IN INDIAN GENOMICS

With the availability of diverse data types – gene expression, SNP genotypes, demographics, health history, laboratory findings, and images etc. – machine learning algorithms have become the obvious choice for accurate prediction of disease risk and personalized treatment. They can learn patterns underlying complex data and build models that can be used for prediction purposes. Numerous machine learning methods, such as support vector machines, random forests, and Bayesian networks, are being used successfully in genomics research and applications (Libbrecht and Noble, 2015; Xu and Jackson, 2019). Now, deep learning algorithms, a subcategory of machine learning, have

emerged as the most successful algorithms for combining clinical data with genomics (Ching et al., 2018; Zou et al., 2019). They use artificial neural networks to progressively extract novel features from input data and learn from the features (Eraslan et al., 2019).

Deep learning and machine learning algorithms, which come under the umbrella term Artificial Intelligence (AI), are being used in clinical practice through numerous commercial applications involving clinical and genomics data. A well known personal genomics company, 23andme (2020) uses machine learning algorithms in disease risk prediction. IBM's Watson for Oncology (IBM Watson for Oncology, 2020) helps clinicians in identifying most appropriate treatment options based on information collated from medical records, medical journals, genomic journals, and relevant guidelines. Many startups are increasingly using the combination of machine learning algorithms and genomics in creating tools and processes that enhance the healthcare systems. For example, Freenome (2019), Benevolent AI (2020), Cambridge Cancer Genomics (2020), and DeepGenomics (2020) use AI in predicting disease risk, response to therapy and in developing personalized treatment regimens. In India, very few organizations use machine learning algorithms in clinical genomics, with the reasons being lack of awareness and lack of expertise in research and application of AI. Some of the Indian pharmaceutical and genomic organizations that are using AI include Innoplexus (2019), Lantern Pharma (2019), Manipal Group of Hospitals (2019), TCS Innovation Labs (2019), BioXcel Therapeutics Inc. (2020), Mapmygenome (2020), OncoStem (2020), and PierianDx (2020).

Challenges and Limitations

Main technical challenges in the application of ML algorithms are data curation and data pre-processing (Ngiam and Khor, 2019). Different hospitals and laboratories adopt different terminologies to record a disease or a health condition and use different reference ranges. In India, Electronic Health Records Standards were released by the Ministry of Health and Family Welfare in 2016. But sharing of data between the hospitals through a common platform is still a work in progress.

Data sets used in training the machine learning algorithms should clearly represent the target data for which risk predictions are made. For example, genetic algorithms trained on data from North Indians might make less accurate predictions when applied on South Indians. Comprehensive and robust clinical data sets that represent the ethnic differences among the people of India are still unavailable. To facilitate sharing of biological data across various research organizations in India, especially high-throughput data generated by sequencing and microarrays, and to create National Biological Data Centre, Ministry of Science and Technology has released zero draft on Biological data storage, access and sharing policy of India in July 2019 (Department of Science and Technology, 2020). But it is still in its nascent stage. A standard procedure for normalizing the raw data must be developed to maintain uniformity across the research groups.

Lack of understanding among clinicians and patients about the machine learning algorithms and their predictions make them considered as black box algorithms (Vayena et al., 2018). Data scientists should explain the general logic behind the

algorithm-based decisions. Doctors and patients should understand the risk associated with such decisions. Clear communication between data scientists, doctors and patients is required to maintain ethical standards in clinical applications.

Solutions to Overcome the Challenges

NITI Aayog, a policy think tank of the government of India, made several recommendations to address the challenges and to harness the power of AI in India (National Strategy for AI, 2018). They include – establishing Centres of Research Excellence (COREs), increasing R&D resources, supporting Ph.D. researchers, establishing common supercomputing facilities, and creating an ecosystem for development and application of AI. Encouraging institute-industry partnerships, creating investment funds for AI startups and reskilling the existing workforce have also been discussed in detail. Other research agencies like Itihaasa (2018) made similar recommendations.

Institutional review boards, ethical review committees and scientific societies should come up with best practices for application of ML in clinical genomics. Government should start a regulatory body in lines similar to the United States Food and Drug Administration (FDA) to enforce best practices. Data sets used in training the algorithms, variables considered in building the models and accuracy of the predictions should be scrutinized. Updating the models by retraining the algorithms and checking the efficiency of the models should be done in coordination with the clinicians. The Government of India should take initiatives to train clinicians in understanding machine learning algorithms. Certification programs run through premier institutes would encourage the people to take up such courses.

ETHICAL AND LEGAL CONCERNS IN DATA SHARING

Genomic data is sensitive in nature and public sharing of such data brings a fair share of ethical and legal concerns with it. Given the increasing number of direct-to-consumer tests that are available, there is a need to streamline certain processes. The collection, storage and usage of genetic data must enable meaningful outcomes for personalized medicine. Data security and privacy remains one of the major concerns reported by users. The “Personal Genomes: Accessing, Sharing and Interpretation” conference held in the United Kingdom, in April 2019 (Genetics Society, 2019) addressed several conundrums which hinder sharing of genetic and medical data, for the creation and maintenance of genomic databases. There is also a growing segment of users who are open to sharing their de-identified data (Kim et al., 2015; Rubin and Glusman, 2019). They share their data for getting updates on their health reports, for providing social good or for financial compensation (Hendricks-Sturup and Lu, 2020). In India, with the release of Personal Data Protection Bill 2019 (The Personal Data Protection Bill, 2019) certain principles were laid down on collection and usage of personal data. Informed consent, data minimization and storing a copy of data within India are some of the essential requirements under the bill.

The benefits of sharing genomic data in the scientific community are far too many to ignore. Collaborative efforts between sequencing facilities, data scientists, clinics, and healthcare providers must be directed toward building a healthy ecosystem for data sharing. De-identification of the genetic information as well as medical records is essential. Wright et al. (2019) proposes a system wherein genetic variant details and their associated conditions can be shared in online databases, without requiring explicit consent from patients. However, detailed clinical information and case study at a deeper level will require consent from the doctors and their patients. For the Indian scenario, a robust system for data sharing is required. This system must be regulated by measures which protect the patients’ interests as well. Policy makers and leaders must come together to develop a framework that allows more variant databases to become publicly accessible, without breach of privacy.

ROLE OF CLINICIANS IN INDIAN GENOMICS

Clinicians play a very important role in facilitating genomics-driven healthcare. From the time a patient visits the clinic to the time of treatment, there are several stages that require the clinician to relay information related to testing procedures and their possible outcomes. The clinician holds a key responsibility of comprehending the implications of genetic findings and making the necessary correlations for treatment and management. Hence, it is imperative that the clinician is well versed with different genetic mechanisms, inheritance, gene-gene, and gene-environment interaction mechanisms, variants and their pathogenicity. In the clinic, staff must be trained to perform timely reviews of clinical and family history and identify cases which warrant genetic testing. For the current generation of clinicians, training on genetic diseases, testing methodologies, clinical variant interpretation and application in medicine, must be included as part of their continuing education. Policy makers such as Medical Council of India and Board of Education play an important role in training clinicians on utilizing genomics in their practice (Scheuner et al., 2008; Aggarwal and Phadke, 2015).

CONCLUSION

Given the broad spectrum of genetic diseases and their burden on the Indian population, it is essential for genomic researchers to tap Indian genetic data for disease prevention, timely diagnosis, and treatment. Studies show that there are novel mutations in Indian patients, for different phenotypes. Hence, genome analysts need to refer to Indian-specific databases for meaningful translation of genomics data into clinical reporting. Current challenges can be met by united efforts from government health agencies and genetic research institutes by executing large scale sequencing projects, accompanied by detailed documentation

on patients' clinical features and family history. Obtaining informed consent from the patients must be mandatory, to protect their interests including concerns about data privacy and safety. The patients must be educated about protocols such as de-identification, data security and research objectives.

Novel variants must be made available in a centralized database for analysts to refer to, and draw inferences from. Such a database would vastly improve the diagnostic accuracy of genetic diseases. Indian genomics will also greatly benefit by the development of machine learning algorithms for analyzing health trends in the Indian population. Additionally, clinicians

from all walks of medicine must be equipped with technical knowledge on medical genetics and its clinical application, for enhanced patient care.

AUTHOR CONTRIBUTIONS

SP and RR have contributed conception and design of the study. SP, RR, and RM wrote sections of the manuscript. MV and AA have supervised and reviewed the manuscript. All authors contributed to the article and approved the submitted version.

REFERENCES

- 23andme (2020). *23andme*. Available online at: <https://www.23andme.com/> (accessed November, 2019).
- Aggarwal, S., and Phadke, S. R. (2015). Medical genetics and genomic medicine in India: current status and opportunities ahead. *Mol. Genet. Genomic Med.* 3, 160–171. doi: 10.1002/mgg3.150
- Ahmed, P. H., Viswanath, V., More, R. P., Viswanath, B., Jain, S., Rao, M. S., et al. (2019). INDEX-db: the indian exome reference database (Phase I). *J. Comput.* 26, 225–234. doi: 10.1089/cmb.2018.0199
- Aravind, S., Ashley, B., Mannan, A., Ganapathy, A., Ramesh, K., Ramachandran, A., et al. (2019). Targeted sequencing of the DMD locus: a comprehensive diagnostic tool for all mutations. *Indian J. Med. Res.* 150, 282–289. doi: 10.4103/ijmr.IJMR_290_18
- Bellary, K., Dwarkanath, K. M., Nagalla, B., and Mohini, T. A. (2019). Genetic variants of chromosome 9p21.3 region associated with coronary artery disease and premature coronary artery disease in an Asian Indian population. *Indian Heart J.* 71, 263–271. doi: 10.1016/j.ihj.2019.04.005
- Benevolent AI (2020). *Benevolent AI*. Available online at: <https://benevolent.ai/> (accessed November, 2019).
- Bhatia, S., Goyal, S., Singh, I., Singh, D., and Vanita, V. (2018a). A novel mutation in the PRPF31 in a North Indian adRP family with incomplete penetrance. *Doc. Ophthalmol.* 137, 103–119. doi: 10.1007/s10633-018-9654-x
- Bhatia, S., Kaur, N., Singh, I., and Vanita, V. (2018b). A novel mutation in MERTK for rod-cone dystrophy in a North Indian family. *Can. J. Ophthalmol.* 54, 40–50. doi: 10.1016/j.jcjo.2018.02.008
- BioXcel Therapeutics Inc. (2020). *BioXcel Therapeutics Inc.* Available online at: <https://www.bioxceltherapeutics.com/> (accessed November, 2019).
- Cambridge Cancer Genomics (2020). *Cambridge Cancer Genomics*. Available online at: <https://www.ccg.ai/> (accessed November, 2019).
- CCMB (2020). *Genotype/Phenotype dB*. Available online at: https://www.ccmb.res.in/bic/database_pagelink.php?page=genotype (accessed November, 2019).
- Chauhan, G., Spurgeon, C. J., Tabassum, R., Bhaskar, S., Kulkarni, S. R., Mahajan, A., et al. (2010). Impact of common variants of PPARG, KCNJ11, TCF7L2, SLC30A8, HHEX, CDKN2A, IGF2BP2, and CDKAL1 on the risk of type 2 diabetes in 5,164 Indians. *Diabetes Metab. Res. Rev.* 59, 2068–2074. doi: 10.2337/db09-1386
- Ching, T., Himmelstein, D. S., Beaulieu-Jones, B. K., Kalinin, A. A., Do, B. T., Way, G. P., et al. (2018). Opportunities and obstacles for deep learning in biology and medicine. *J. R. Soc. Interface* 15:20170387. doi: 10.1098/rsif.2017.0387
- Clinvar (2019). *Clinvar*. Available online at: <https://www.ncbi.nlm.nih.gov/clinvar> (accessed November, 2019).
- Dada, R., Kumar, R., Shamsi, M. B., Tanwar, M., Pathak, D., Venkatesh, S., et al. (2008). Genetic screening in couples experiencing recurrent assisted procreation failure. *Indian J. Biochem. Biophys.* 45, 116–120.
- DeepGenomics (2020). *DeepGenomics*. Available online at: <https://www.deepgenomics.com/> (accessed November, 2019).
- Department of Science and Technology (2020). *Department of Science, and Technology*. Available online at: <https://dst.gov.in/> (accessed November, 2019).
- Di, Y., Huang, L., Sundaresan, P., Li, S., Kim, R., Ballav Saikia, B., et al. (2016). Whole-exome sequencing analysis identifies mutations in the *eyes* gene in retinitis pigmentosa in the Indian population. *Sci. Rep.* 6:19432. doi: 10.1038/srep19432
- Eraslan, G., Avsec, Ž, Gagneur, J., and Theis, F. J. (2019). Deep learning: new computational modelling techniques for genomics. *Nat. Rev. Genet.* 20, 389–403. doi: 10.1038/s41576-019-0122-6
- Freenome (2019). *Freenome*. Available online at: <https://www.freenome.com/> (accessed November, 2019).
- Ganapathy, A., Mishra, A., Soni, M. R., Kumar, P., Sadagopan, M., Kanthi, A. V., et al. (2019). Multi-gene testing in neurological disorders showed an improved diagnostic yield: data from over 1000 Indian patients. *J. Neurol.* 266, 1919–1926. doi: 10.1007/s00415-019-09358-1
- Gandra, M., Anandula, V., Authiappan, V., Sundaramurthy, S., Raman, R., Bhattacharya, S., et al. (2008). Retinitis pigmentosa: mutation analysis of RHO, PRPF31, RP1, and IMPDH1 genes in patients from India. *Mol. Vis.* 14, 1105–1113.
- Genetics Society (2019). *Genetics Society*. Available online at: <https://genetics.org.uk/events/personal-genomes-accessing-sharing-and-interpretation/> (accessed November, 2019).
- Genomics and other Omics tools for Enabling Medical Decision (2019). *Genomics and other Omics tools for Enabling Medical Decision*. Available online at: <http://gomed.igib.in/> (accessed November, 2019).
- GUARDIAN Consortium, Sivasubbu, S., and Scaria, V. (2019). Genomics of rare genetic diseases-experiences from India. *Hum. Genomics* 14:52. doi: 10.1186/s40246-019-0215
- Gupta, S., Chaurasia, A., Pathak, E., Mishra, R., Chaudhry, V. N., Chaudhry, P., et al. (2017). Whole exome sequencing unveils a frameshift mutation in CNGB3 for cone dystrophy: a case report of an Indian family. *Medicine* 96:e7490. doi: 10.1097/MD.00000000000007490
- Hendricks-Sturupp, R., and Lu, C. (2020). What motivates the sharing of consumer-generated genomic information? *SAGE Open Med.* 8:205031212091540. doi: 10.1177/2050312120915400
- IBM Watson for Oncology (2020). *IBM Watson for Oncology*. Available online at: <https://www.ibm.com/watson-health/oncology-and-genomics> (accessed November, 2019).
- Indian Genome Variation Consortium (2005). The Indian genome variation database (IGVdb): a project overview. *Hum. Genet.* 118, 1–11. doi: 10.1007/s00439-005-0009-9
- Indian GWAS (2010). *Centra*. Available online at: <https://vigeyegpms.in/gpmsv2/gwascentralindia/>.
- Indian SNP (2020). *Database*. Available online at: https://www.ccmb.res.in/bic/database_pagelink.php?page=snpdata (accessed November, 2019).
- IndiGen (2020). *Project*. Available online at: <https://indigen.igib.in/> (accessed November, 2019).
- Innoplexus (2019). *Innoplexus*. Available online at: <https://www.innoplexus.com/> (accessed November, 2019).
- Itihaasa (2018). *Itihaasa*. Available online at: http://www.itihaasa.com/pdf/itihaasa_AI_Research_Report.pdf (accessed November, 2019).
- Judith, M. H., Shamsudheen, K. V., Ankit, V., Anop, S. R., Rijith, J., Rowmika, R., et al. (2018). SAGE: a comprehensive resource of genetic variants integrating South Asian whole genomes and exomes. *Database* 2018:bay080. doi: 10.1093/database/bay080
- Kannabiran, C., Singh, H., Sahini, N., Jalali, S., and Mohan, G. (2012). Mutations in TULP1, NR2E3, and MFRP genes in Indian families with autosomal recessive retinitis pigmentosa. *Mol. Vis.* 18, 1165–1174.

- Kar, B., and Sivamani, S. (2016). Directory of genetic test services and counselling centres in India. *Int J Hum Genet.* 16, 148–157. doi: 10.1080/09723757.2016.11886292
- Kim, K. K., Joseph, J. G., and Ohno-Machado, L. (2015). Comparison of consumers' views on electronic data sharing for healthcare and research. *J. Am. Med. Inform. Assoc.* 22, 821–830. doi: 10.1093/jamia/ocv014
- Lantern Pharma (2019). *Lantern Pharma*. Available online at: <https://www.lanternpharma.com/> (accessed November, 2019).
- Libbrecht, M. W., and Noble, W. S. (2015). Machine learning applications in genetics and genomics. *Nat. Rev. Genet.* 16, 321–332. doi: 10.1038/nrg3920
- Manipal Group of Hospitals (2019). *Collaboration with IBM's Watson for Oncology*. Available online at: <https://www.manipalhospitals.com/> (accessed November, 2019).
- Mapmygenome (2020). *Mapmygenome*. Available online at: <https://mapmygenome.in/> (accessed November, 2019).
- Mohan, V., Goldhaber-Fiebert, J. D., Radha, V., and Gokulakrishnan, K. (2011). Screening with OGTT alone or in combination with the Indian diabetes risk score or genotyping of TCF7L2 to detect undiagnosed type 2 diabetes in Asian Indians. *Indian J. Med. Res.* 133, 294–299.
- Nagrani, R., Mhatre, S., Rajaraman, P., Chatterjee, N., Akbari, M. R., Boffetta, P., et al. (2017). Association of genome-wide association study (GWAS) identified SNPs and risk of breast cancer in an Indian population. *Sci. Rep.* 7:40963. doi: 10.1038/srep40963
- Narang, A., Roy, R. D., Chaurasia, A., Mukhopadhyay, A., and Mukerji, M. (2010). Indian genome variation consortium. Das D. IGVBrowser—a genomic variation resource from diverse Indian populations. *Database* 2010:baq022. doi: 10.1093/database/baq022
- National Strategy for AI (2018). *National Strategy for AI*. Available online at: <https://niti.gov.in/sites/default/files/2019-01/NationalStrategy-for-AI-Discussion-Paper.pdf> (accessed November, 2019).
- Ngiam, K. Y., and Khor, I. W. (2019). Big data and machine learning algorithms for health-care delivery. *Lancet Oncol.* 20, e262–e273. doi: 10.1016/s1470-2045(19)30149-4
- OncoStem (2020). *OncoStem*. Available online at: <https://www.oncostem.com/> (accessed November, 2019).
- PierianDx (2020). *PierianDx*. Available online at: <https://www.pierianDX.com/> (accessed November, 2019).
- Prabhash, K., Advani, S. H., Batra, U., Biswas, B., Chougule, A., Ghosh, M., et al. (2019). Biomarkers in non-small cell lung cancers: Indian consensus guidelines for molecular testing. *Adv. Ther.* 36, 766–785. doi: 10.1007/s12325-019-00903-y
- Pradhan, S., Sengupta, M., Dutta, A., Bhattacharyya, K., Bag, S., Dutta, C., et al. (2011). Indian genetic disease database. *Nucleic Acids Res.* 39(Suppl. 1), D933–D938. doi: 10.1093/nar/gkq1025
- Puri, R. D., Tuteja, M., and Verma, I. C. (2017). genetic approach to diagnosis of intellectual disability. *Indian J. Pediatr.* 83, 1141–1149. doi: 10.1007/s12098-016-2205-0
- Rajasimha, H. K., Shirol, P. B., Ramamoorthy, P., Hegde, M., Barde, S., Chandru, V., et al. (2014). Organization for rare diseases India (ORDI) - addressing the challenges and opportunities for the Indian rare diseases' community. *Genet. Res.* 96:e009. doi: 10.1017/S0016672314000111
- Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., et al. (2015). Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet. Med.* 17, 405–424. doi: 10.1038/gim.2015.30
- Rubin, I. R., and Glusman, G. (2019). Opportunities and challenges in interpreting and sharing personal genomes. *Genes* 10:643. doi: 10.3390/genes10090643
- Scheuner, M. T., Sieverding, P., and Shekelle, P. G. (2008). Delivery of genomic medicine for common chronic adult diseases: a systematic review. *JAMA* 299, 1320–1334. doi: 10.1001/jama.299.11.1320
- Singh, B., Mandal, K., Lallar, M., Narayanan, D. L., Mishra, S., Gampbhir, P. S., et al. (2018). Next generation sequencing in diagnosis of MLPA Negative cases presenting as duchenne/ becker muscular dystrophies. *Indian J. Pediatr.* 85, 309–310. doi: 10.1007/s12098-017-2455-5
- Singh, J., Mishra, A., Pandian, A. J., Mallapatna, A. C., Khetan, V., Sriprिया, S., et al. (2016). Next-generation sequencing-based method shows increased mutation detection sensitivity in an Indian retinoblastoma cohort. *Mol. Vis.* 22, 1036–1047.
- Srilekha, S., Arokiasamy, T., Srikrupa, N. N., Umashankar, V., Meenakshi, S., Sen, P., et al. (2015). Homozygosity mapping in leber congenital amaurosis and autosomal recessive retinitis pigmentosa in south Indian families. *PLoS One* 10:e0131679. doi: 10.1371/journal.pone.0131679
- TCS Innovation Labs (2019). *TCS Innovation Labs*. Available online at: <https://www.tcs.com/reimagining-drug-safety-powered-by-genomics-information-integration-and-emerging-technologies> (accessed November, 2019).
- The Personal Data Protection Bill (2019). *The Personal Data Protection Bill*. Available online at: <https://www.prsindia.org/billtrack/personal-data-protection-bill-2019> (accessed November, 2019).
- Upadhyay, P., Gardi, N., Desai, S., Sahoo, B., Singh, A., Togar, T., et al. (2016). TMC-SNPdb: an Indian germline variant database derived from whole exome sequences. *Database* 2016:baw104. doi: 10.1093/database/baw104
- Uttarilli, A., Shah, H., Bhavani, G. S., Upadhyay, P., Shukla, A., and Girisha, K. M. (2019). Phenotyping and genotyping of skeletal dysplasias: evolution of a center and a decade of experience in India. *Bone* 120, 204–211. doi: 10.1016/j.bone.2018.10.026
- Vayena, E., Blasimme, A., and Cohen, I. G. (2018). Machine learning in medicine: addressing ethical challenges. *PLoS Med.* 15:e1002689. doi: 10.1371/journal.pmed.1002689
- Verma, I. C., Paliwal, P., and Singh, K. (2018). Genetic Testing in Pediatric Ophthalmology. *Indian J. Pediatr.* 85, 228–236. doi: 10.1007/s12098-017-2453-7
- Wright, C. F., Ware, J. S., Lucassen, A. M., Hall, A., Middleton, A., Rahman, N., et al. (2019). Genomic variant sharing: a position statement. *Wellcome Open Res.* 4:22. doi: 10.12688/wellcomeopenres.15090.2
- Xu, C., and Jackson, S. A. (2019). Machine learning and complex biological data. *Genome Biol.* 20:76. doi: 10.1186/s13059-019-1689-0
- Yang, Y., Yang, Y., Huang, L., Zhai, Y., Li, J., Jiang, Z., et al. (2016). Whole exome sequencing identified novel CRB1 mutations in Chinese and Indian populations with autosomal recessive retinitis pigmentosa. *Sci. Rep.* 6:33681. doi: 10.1038/srep33681
- Zhou, Y., Saikia, B., Jiang, Z., Zhu, X., Liu, Y., Huang, L., et al. (2015). Whole-exome sequencing reveals a novel frameshift mutation in the FAM161A gene causing autosomal recessive retinitis pigmentosa in the Indian population. *J. Hum. Genet.* 60:625. doi: 10.1038/jhg.2015.92
- Zou, J., Huss, M., Abid, A., Pejman, M., Ali, T., and Amalio, T. (2019). A primer on deep learning in genomics. *Nat. Genet.* 51, 12–18. doi: 10.1038/s41588-018-0295-5

Conflict of Interest: SP, RR, RM, and AA are employed by the company Mapmygenome India Limited. MV is a member of Scientific Advisory Board at Mapmygenome India Limited.

Copyright © 2020 Pemmasani, Raman, Mohapatra, Vidyasagar and Acharya. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.