Current Epidemiology Reports (2022) 9:376–386
https://doi.org/10.1007/s40471-022-00303-x

GENETIC EPIDEMIOLOGY (C I AMOS, SECTION EDITOR)

# Examining Barriers and Opportunities of Conducting Genome-Wide Association Studies in Developing Countries

**Gerard G. Dumancas[1,2] · Megan Rachal[3] · Pia Regina Fatima C. Zamora[4] · Romulo de Castro[2,4]**

Accepted: 28 July 2022 / Published online: 25 August 2022


## Abstract

**Purpose of Review** Genome-wide association studies (GWASs) allow correlations between genetic variations and diseases that will benefit the field of public health in developing countries. It is considered to be the primary step toward a journey from gene discovery to improving health outcomes of individuals. The aim of this manuscript was to examine barriers and opportunities of conducting GWASs in selected developing nations by considering a holistic set of elements including demand, resources, implementation/practicality, and adaptation.

**Recent Findings** Consideration of various aforementioned elements shows that while some developing countries exhibit great potential in performing GWAS research, several factors are still hindering progress toward this direction including adaptation/integration, proper governance, and availability of human and financial resources.

**Summary** This informative analysis provided information that determined and considered various factors that make it challenging to conduct GWAS in developing nations. United efforts from genetic research institutions (i.e., multicenter collaboration) and government health agencies are recommended to successfully implement GWASs in developing economies.

**Keywords** Genome-wide association study · GWAS · Developing country · Feasibility · Phenotype

## Introduction

A genome-wide association study (GWAS) is an approach that involves rapidly scanning markers across the complete sets of DNA or genomes of many people to find genetic variations associated with a particular trait or disease [1]. After the first GWAS was published in October 2002, such studies have then made a huge impact in the underlying world of genetics [2]. It has provided an avenue for researchers to identify variations in a genomic sequence, then consequently use that data to determine correlations between such genetic variations (i.e., single nucleotide polymorphisms (SNPs)) and human diseases. GWASs have already contributed to the identification of SNPs associated with various conditions such as diabetes, heart abnormalities, Parkinson's disease, Crohn's disease, asthma, and mental illnesses [3–7]. Thus, the prevention and treatment of diseases could be vastly enhanced due to GWASs. In the field of public health, GWAS plays an important role in improving disease intervention strategies in the community by predicting individuals at risk, designing targeted biologic interventions, and establishing deeper insights into the etiology of a disease [8].

Despite the successes of GWASs, most studies involving such innovation are primarily focused on developed countries. These countries have evolved in the development of sophisticated DNA chips that consequently allowed the identification of tiny differences between individuals' genes that predict genetic risks in infancy for which preventable measures (i.e., lifestyle changes) could be adopted [6].

This article is part of the Topical Collection on *Genetic Epidemiology*.

✉ Gerard G. Dumancas
gerard.dumancas@scranton.edu

✉ Romulo de Castro
rdecastro@usa.edu.ph

1 Department of Chemistry, Loyola Science Center, The University of Scranton, Scranton, PA 18510, USA

2 Balik Scientist Program, Department of Science and Technology-Philippine Council for Health Research and Development, 1631 Bicutan, Taguig City, Philippines

3 Department of Biological Sciences, Louisiana State University-Alexandria, LA 71302 Alexandria, USA

4 Center for Informatics, University of San Agustin, Gen. Luna St, 5000 Iloilo City, Philippines

However, far fewer GWASs have been conducted in developing countries, where disparities in burden of disease, financial resources, educational attainment, and health outcomes prevail [6]. In research involving complex diseases, race and ethnicity can provide information about social, cultural, and environmental factors that affect disease risk. The lack of diversity in genetic studies is problematic for a variety of ethic and scientific reasons. Continued reliance on samples that only represent a fraction of genomic, sociocultural, and environmental diversity limits our understanding of disease biology and may ultimately contribute to widening global health disparities. Greater ancestral diversity in study samples has the potential to accelerate discovery of causal risk variants and is critical for a greater understanding of the biological causes of disease, including gene and environment interactions. Further, broadening diversity of studied populations will improve our understanding and effectiveness of genomic medicine by expanding the collection of known human genetic variations and eventually bolstering our understanding of disease etiology [9••, 10]. In this manuscript, we have highlighted the challenges and benefits of working with diverse populations, recommended practices based on current methods, and noted specific areas that are in need of further methodological development. In summarizing progress, remaining challenges, and requisite next steps, we considered three main domains: *(1)* researcher participation, *(2)* data resources, and *(3)* analytic methods.

It is imperative to perform GWASs in developing countries since significant problems posed by communicable diseases (e.g., tuberculosis, HIV) as well as rapid increases in non-communicable diseases including cancer and heart disease emanate from these countries [11]. Investing in genotyping projects including GWASs can provide developing countries with the necessary tools to better understand drug response, disease susceptibility, and disease mechanisms in their own [12, 13]. Further, such investment is also critical as a country's potential entry point into the global knowledge-based economy [12]. This manuscript aimed to determine the feasibility of conducting GWASs in selected developing nations, specifically identifying barriers and opportunities of establishing such studies in these countries. Developing countries within the context of this study refers to those nations which are classified as developing economies by the United Nations [14].

## Current State of GWAS in Developing Countries

A PubMed database search by using the keywords "genome-wide association study" generated 47,995 studies, while that using the keywords "genome-wide associate study," "X," where "X" refers to the name of the country with a developing economy as classified by the United Nations [14] garnered 20,795 results within the last decade (2011–2021) (Figs. 1 and 2). Overall, there is an increasing trend in GWASs among developing economies (Fig. 2).

### General Criteria of Examining Factors Affecting the Feasibility of Establishing GWAS in Developing Nations

The feasibility of conducting GWASs in developing nations is affected by numerous factors including demand, availability of resources, implementation/practicality, and adaptation/integration, acceptability, and ease of implementation and integration.

To determine if there is a demand, aspects related to diseases affecting the population will need to be assessed. If there is a high demand for studies of specific diseases in the country, GWAS will definitely play a critical role in elucidating the genetic component of such diseases. Next, resources needed to perform GWAS include technological and financial resources, samples, human resources, and

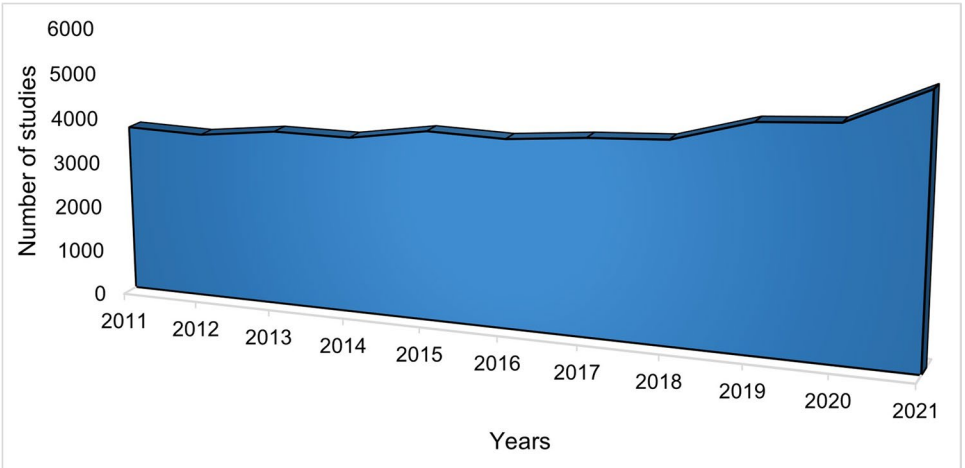


**Fig. 1** PubMed search results using the keywords "genome-wide association study." Non-human studies maybe included in the search results

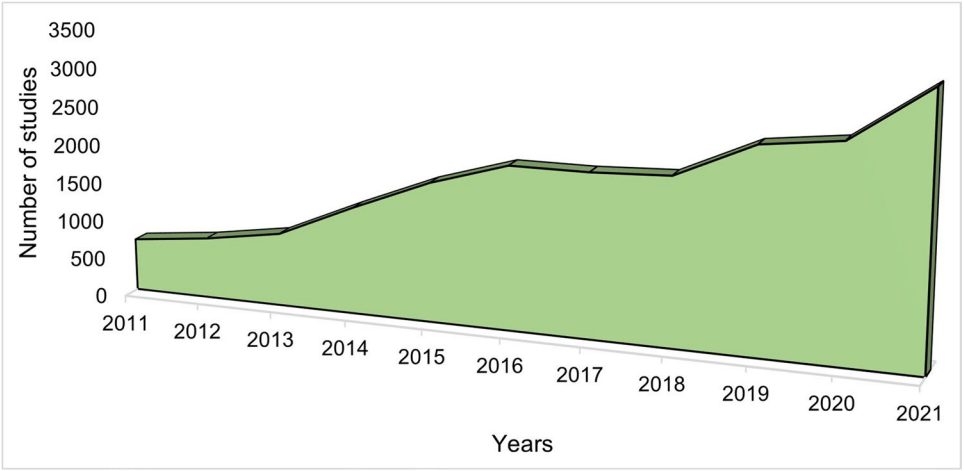


**Fig. 2** PubMed search results using the keywords "genome-wide association study" and "X" where "X" refers to the name of the developing country as classified by the United Nations. Non-human studies maybe included in the search results

infrastructure among others, must also be available. Along with these needs, researchers will need to have practical means of obtaining biological samples as well as other pertinent information from participants. This will include getting volunteers to donate samples and ensuring that there is written, informed consent before using participant data in research. Lastly, in the implementation stage, practicality, and integration with existing healthcare practices are critical factors that should be taken into consideration.

These criteria were assessed in selected developing countries or regions including the Philippines, Malaria Genomic Epidemiology Network (MalariaGen) (i.e., a research consortium of 21 developing countries in Africa, Asia, and Oceania), China, India, and the Human Heredity and Health in Africa (H3Africa) Initiative (i.e., a research consortium of more than 30 African countries). MalariaGen's overall goal was to determine how human genetic variations, Plasmodium parasites, and Anopheles mosquitoes affect the transmission of malaria [15]. H3Africa, on the other hand, is an initiative borne out of partnerships among the Wellcome Trust (United Kingdom), the African Society of Human Genetics, and the US National Institutes of Health (NIH) with the ultimate goal of facilitating research into diseases on the African continents while also developing resources, training and ethical guidelines, as well as infrastructure, to assist in the development of a sustainable African research enterprise [16, 17].

## A. Demand

There is a demand for GWASs in developing countries because of endemic and/or prevalent diseases affecting its population as mentioned in earlier. In many developing countries, citizens are at high risk for dementia, diabetes, cardiovascular disease (CVD), and depression among other diseases [18••, 19•, 20, 21]. Using GWASs, medical professionals would be able to use data to determine if a patient has genetic risk factors associated with the aforementioned diseases. GWAS could revolutionize precision medicine in these developing countries as it has done for other highly developed countries. For example, the Philippines currently faces the issue of infant birth defects that are causing a rise in the mortality rate and there are few geneticists to diagnose and study these birth defects. The rise in birth defects has accelerated genetic testing and studies of genetic abnormalities for diseases afflicting the population. However, there is still more to be done [22]. GWAS could be useful for determining the causes behind these birth defects. The Philippines is also faced with the threat of HIV. The rate of infection growth in the Philippines is much higher than other countries in the Asia–Pacific region [23]. There are currently GWASs being performed in other countries to determine genetic risk factors and SNPs related to susceptibility to HIV. The rise in HIV in the Philippines makes it a practical avenue to conduct GWAS in the country. Another illness plaguing the Philippines is the dengue virus. GWAS could be used to identify genetic risk factors that cause the high fatality rate associated with this virus [24••]. Similarly, the establishment of the MalariaGEN consortium is a result of a need to perform GWASs in order to study the evolutionary processes that affect malaria transmission and disease [15].

The presence of rare diseases endemic to a specific region is another particular motivation to pursue GWAS in developing nations. For example, in the Philippines the X-linked dystonia–parkinsonism, also referred to as DYT3 dystonia or "Lubag" disease characterized by severe and progressive dystonia followed by overt parkinsonism in the later years of life, is a major motivation to pursue GWAS in the country [25]. In India and China, an asymptomatic infection with Visceral Leishmaniasis in the Bihar area warrants the demand for a possible GWAS study to assess the evolutionary responses toward such infection [26, 27]. Lastly, the H3Africa consortium facilitates research into diseases on

the African continent which constitutes 51 projects ranging from heart and renal disease, to tuberculosis, which also warrants further investigation for GWAS research. An analysis of abstracts in the H3Africa consortium as well as other international meetings/conferences suggests that genome-wide data analysis projects for many H3Africa studies are nearing completion with two already published last 2020 [28, 29•, 30•].

## B. Resources

Conducting GWAS requires financial, manpower, and infrastructure resources. Below are some of the resources that must be accessed to ensure successful implementation and sustainability of GWASs. Bioinformatic tools are one of the vital resources needed to successfully carry out GWASs. Many of these bioinformatic tools used for GWASs are publicly available or would cost a small fraction of the overall GWAS budget [31]. These tools are typically executed using an R package or by using C/C + + language or Python language, which may require additional skillset to successfully run these advanced data analytic software packages.

**Sample Sources and Publicly Available Datasets** A biobank or a large database of patient samples can increase the number of participants needed for GWASs, leading to an adequate sample size which will yield results representative of a chosen cohort [32]. After a biobank has been established, it will also need to be maintained, thus, implying involvement of financial resources. Further, there is also a need to follow up with participants in prospective studies, as well as admission of new subjects. The use of Electronic Health Record (EHR) information will facilitate the establishment of a biobank. As records become digitized, the information are more organized, and electronic records are preserved in a safer manner than paper records complementing the existing biobank [33, 34]; however, not all developing countries use EHRs.

In the Philippines, publicly available GWAS datasets are very limited and are particularly focused on a specific phenotype. For example, the Cebu Longitudinal Health and Nutrition Survey is focused on the long-term effects of prenatal and early childhood nutrition and health on later adult outcomes, including development of chronic disease risk factors [35]. Most of the publicly available Indian databases for GWAS, on the other hand, are incomplete. Examples of Indian specific databases include Index-DB (i.e., a database of exonic variants from normal individuals of sub-Indian continent), TMC-SNPdb (i.e., contains variants generated from exome of normal samples derived from cancer patients of Indian origin), and GWAS—Central India (i.e., a genotype–phenotype association database with summary level findings from genetic association studies) [36••, 37••, 38]. The aforementioned databases have several limitations including limited sample size, lack of disease-variant associations, and lack of regular updating of the database. This can be attributed to numerous factors including financial, legal, ethical, and administrative procedures that lead to a lot of important parameters not being recorded [36••]. Perhaps the largest SNP genotyping database of Asian population is the HUGO Pan-Asian SNP consortium. It is comprised of Asians by sampling 1,719 unrelated individuals among 71 populations from China, India, Indonesia, Japan, Malaysia, the Philippines, Singapore, South Korea, Taiwan, and Thailand [39]. Another large catalogue of genomic database comprises samples of the Han Chinese population (i.e., the PGG.Han). It archives the whole-genome sequences or high-density genome-wide single-nucleotide variants of 114,783 Han Chinese population representing geographical sub-populations of China, as well as Singapore [40••]. In Africa, on the other hand, The H3Africa CVD Working Group is pooling resources and harmonizing data to establish the Cardiovascular H3Africa Innovation Resource (CHAIR) that will house more than 50,000 participants to study environmental and genomic contributions to CVD in Africa. Further, CHAIR is ideally placed for meta-analysis of GWAS data [41]. As the H3Africa genotyping array only became available in early 2018, majority of the GWASs were delayed which explains the scarcity of full-scale published research studies under the consortium [42]. Overall, there seems to be an effort in the establishment of publicly available genomic databases for developing countries although such efforts are also jointly in collaboration and/or funded by other developed countries.

EHRs and biobanks provide excellent sources for the necessary genotype data (as discussed earlier) as well as phenotype information necessary to perform GWASs. Specifically, the sources of phenotype information may include clinical examinations conducted by trained, ideally calibrated examiners, clinical records (e.g., electronic patient record data), administrative claims data, as well as self-reported data (e.g., obtained via written or electronic questionnaires or via telephone). The sources for genotype information, on the other hand, may include blood samples, saliva samples, buccal swabs, as well as blood spots [43••].

**Sample Sizes** GWASs require large sample sizes to achieve adequate statistical power. Testing a single SNP-marker disease association requires only 248 cases, while testing 500,000 SNPs or 1 million markers requires 1,206 cases and 1,255 cases, respectively, under the assumption of an odds ratio of 2, 5% disease prevalence, 5% minor allele frequency, complete linkage disequilibrium, 1:1 case/control ratio, and a 5% experiment-wise error rate in an allelic test [44]. However, it should also be noted that using very large

sample sizes for GWAS may not always be necessary. For example, if the polygenic contribution is 50% of the phenotypic variance, 500 individuals are sufficient to detect a quantitative trait locus explaining 5% of the phenotypic variance. Extremely large sample sizes are needed to detect rare variants that are often important in such cases [45••]. In general, the power to detect association between genetic variation and disease is a function of multiple factors including but not limited to the frequency of the risk allele or genotype, the relative risk conferred by the disease-associated allele or genotype, the correlation between the genotyped marker and the risk allele, sample size, disease prevalence, genetic heterogeneity of the sample population as well as study design. Power studies have shown that at least 2,000 to 5,000 samples are required for most diseases when using general populations [46].

New statistical method using *pleiotropy* analysis may also be necessary to exponentially increase the ability to discover genetic insights. This analysis provides insights on how individual genes result in multiple characteristics and has become increasingly valuable in mining genes to inform disease treatments. However, there is a challenge related to privacy stipulations in performing comprehensive pleiotropy analysis because individual patient data cannot be easily and regularly shared between sites. A statistical method called "sum-share," however, can integrate multiple electronic health records [47••].

The nature of GWAS has its inherent limitations, especially in developing economies. That is, since the technology is a non-candidate gene approach, it is hypothesis free. This leads to the need for large resources (e.g., sample size, technology, etc.) to make it successful in developing countries. Furthermore, replication of the associated loci should also be considered as well as deep sequencing followed by functional studies to elucidate the underpinning biological mechanism of the disease [48].

**DNA Extraction, Quantitation, and Quality Assessment** A necessary starting point to obtain genetic information needed for GWAS may include automated DNA extraction platform from the whole blood using either high-salt extraction or automated magnetic-bead extraction methods. Alternatively, an automated DNA extraction from saliva, buccal brushes, or blood spots with automated magnetic-bead extraction methods may also be necessary. DNA quantitation including quality assessment may employ various spectrophotometric and chemical assay procedures [43••]. Overall, the use of these techniques requires a skilled researcher. The same is also true for the resources mentioned below.

**Genotyping Supplies, Equipment, and Software** SNP genotyping, a requirement to obtain genetic information for GWAS, would require the use of high-density genotyping arrays (e.g., Illumina Infinium Omni5Exome-4 BeadChip array, offering ~ 4.3 million variants and exome content) or targeted genotyping arrays (e.g., Illumina Metabochip or Immunochip). An array scanning platform (e.g., Illumina iScan) is also needed. Lastly, a variant calling software, Illumina GenomeStudio, which is freely available is also necessary [43••]. Variant calling is a necessary step to detect sequence variants in clinical samples and also serves as an avenue by which virtually all downstream analysis and interpretation processes rely [49•].

Various software which are publicly available are necessary in the imputation, genetic association, visualization, meta-analysis, subsequent quality control and post-processing of GWAS results, as well as genomic context and functional annotations [43••, 49•]. However, it should be noted that the use of this software requires a highly skilled bioinformatics researcher with a clear understanding of its purpose as well as issues arising from it.

**Genotype Storage, Transfer, and Management** A cloud- or intranet-based storage with secure File Transfer Protocol capabilities is required for genotype storage, transfer, and management. A high-performance computing cluster should allow multi-threading and large memory jobs. A server or workstations with common data management and programming suites (e.g., R, SAS, Stata) are also necessary requirements for genotype data storage and analysis [43••].

*Other considerations that should be taken into account when conducting GWASs* include capital, human, medical, and industrial resources. Sustainability of GWASs is another consideration that should be examined. Resources such as the Human Genome Project and the HapMap Project have greatly contributed to the success of GWASs. Since GWAS rely heavily on data analytic platforms as well as the size of data to be analyzed, it may also be necessary to consider strong information and communication technology resources, also known as digital health. Specifically, GWASs require considerable broadband speed. This is unfortunate since developing countries such as the Philippines, India, and Indonesia are considered to be among the top 3 countries with the lowest broadband speed in Asia in 2017 [50]. In order to mitigate this limitation, efforts are needed toward strong digital health governance which involve a holistic approach as well as the involvement of several stakeholders across agencies. This involves the following steps*: 1)* defining the digital health enterprise and identify stakeholders, *2)* identifying and agreeing on what needs to be governed such as assets, architecture, standards, and applications, *3)* identifying instruments to be utilized for governance, *4)* convening entities and agencies under an identified leadership, *5)* adopting a digital health governance framework, *6)* identifying performance measures and monitoring processes for the adopted framework, and *7)* maintaining an active governance framework according to changing requirements

of the digital health enterprise with advancing technological adoptions.

**Financial** The availability of financial resources is a very important consideration in establishing a GWAS research in developing countries. In the Philippines, the budget for science and research has increased in recent years. The Department of Science and Technology (DOST)—Philippines is the primary agency for allocating budget for various research projects to all investigators in the country. Under the DOST, the Philippine Council for Health Research, and Development is the sectoral council responsible for coordinating and monitoring health research activities in the country. Generally, there is limited support for R&D in the Philippines and many investors are not looking to put forth money towards R&D right now [51••]. Overall, the Philippines only spends 0.14% of its gross domestic product on research and development as of 2013 according to the World Bank [52, 53]. This limited financial priority limits the potential of the Philippines to pursue GWASs. The MalariaGen project, on the other hand, originated from the work as a result of the 2003 funding from the Bill & Melinda Gates Foundation and by the UK Medical Research Council. In 2005, the project was formally organized with joint funding from the aforementioned foundation and the Wellcome Trust, as part of the Grand Challenges in Global Health initiative [15, 54]. Similar to MalariaGen, the H3Africa has been funded by institutes coming from high developed nations, the Wellcome Trust and the US NIH, in partnership with the Alliance for Accelerating Excellence in Science in Africa, and the African Academy of Sciences through its funding platform [55]. Further, while many developing economies such as South Africa, Mexico, Brazil, and India were able to make significant strides in utilizing genomic sequencing technologies due to availability of research funds, the situation remains unchanged for the rest of the developing nations and the world especially, Africa [56].

**Human Capital** GWAS will require more human resources like geneticists, biochemical scientists, and scientists studying infectious diseases in developing countries. Among the scientists studying genetics, few would study human subjects which can be considered an obstacle standing in the way of implementing GWAS in developing countries [57]. Human resources may be gathered through programs such as the Balik Scientist Program (BSP) in the Philippines [58]. A Balik Scientist is a science and technology expert who is a Filipino citizen or a foreigner of Filipino descent, residing abroad and contracted by the Philippine Government to return and work in the Philippines along his/her field of expertise [59]. Despite the institutionalization of the BSP, only few Filipino scientists abroad, however, are willing to establish a long-term career in the Philippines due to lucrative offers abroad. By recruiting scientists from various regions of the world as well as exchanging information with other scientists, the possibility of pursuing GWAS in the Philippines would become more feasible. Similar to the BSP of the Philippines, China has also implemented a Young Thousand Talent Program with the overall intent of recruiting globally bred talent to return to China and further enhance the country's research capacity [60]. In the context of MalariaGEN, the challenge of tapping experts in the area of GWASs was addressed via a training program in which junior researchers from participating research centers received intensive training in the analyses of genetic data. Specifically, these researchers participated in the annual data analysis workshops and also in the annual MalariaGEN meetings where they present site specific analyses [15, 61]. For the H3Africa, in order to circumvent the lack of bioinformatics expertise, the network developed a large number of tools, resources, as well as training to support the exploration and analysis of the data generated by the consortium. Since developing countries have limited capacities in education and human development, insufficient training, therefore, would lead to a formidable obstacle in the use of state-of-the-art GWAS facilities [16, 56].

**Infrastructure** Laboratory spaces will be required for genotyping via a collaboration with various neighboring institutions. However, even with the help of an academic institution, GWAS will likely still require more buildings and laboratories to place all the necessary equipment. In the Philippines, several initiatives are in place to leverage resources from universities to manage and analyze the data. In particular, the (DOST) and the Philippine Council for Health Research and Development (PCHRD) BSP has the primary objective of developing research informatics tools and provide mentorship to researchers engaged in a wide range of fields [59]. To date, the Philippine Genome Center remains to be the primary institution performing GWAS, albeit, in a nascent stage [10]. Where research takes place in an international collaboration as in the case of MalariaGEN, the network should identify strengths and weaknesses for all partner sites and be committed to sharing these to others [62]. In African countries, the development of infrastructure support through the H3ABioNet (i.e., a pan-African bioinformatics network designed to enable H3Africa researchers to analyze their data in Africa) has strongly contributed to the establishment of data analysis facilities, African scientific network, and training programs [16, 63].

## Ethical Challenges in Developing Countries

A biobank would be ideal for GWAS in developing countries; however, there are challenges associated with establishing a biobank including barriers related to continuous

collection of genetic material, ethical standards (i.e., developing appropriate processes for valid consent, privacy, and data release), processing and storage of samples, and infrastructure. Biobanks can be specialized based on the goal of scientific research and design [61, 64, 65]. Many ethical issues are raised when genomics research is conducted in areas of lower income as well as literacy levels (e.g., how to clearly explain complex biomedical concepts) including inclusion and reuse of archived samples, export of samples, capacity building, and ethical review [62]. In India, with the release of Personal Data Protection Bill 2019, informed consent, data minimization, and storing a copy of data within India are some of the essential requirements under the bill for collection and usage of personal data [36••, 66]. In the Philippines, the Republic Act 10,173 or the Data Privacy Act of 2012 of the Philippine National Privacy Commission and the Philippine Medical Association's Code of Ethics require confidentiality of any information related to the patient's identity [67]. In many developing economies, however, there are weak or nonexistent regulatory frameworks related to establishing legal and ethical conventions [56]. Beyond obtaining informed consent from individual participants, it is also essential to conduct a process of engagement or consultation with relevant communities and key stakeholders to obtain their views on various social, ethical, and cultural issues the study raises for them [68]. The actions to move forward in this field constitute strengthening bioethics capacity in developing nations; increasing communications between scientists and ethicists in developing and industrialized countries; and linking health research to community needs in a participatory and transparent manner [69].

Another aspect that should be given consideration is to ensure that researchers in developing countries, who generate samples and data for GWAS, are not placed in a scientific disadvantage especially when collaborating with other large type of collaborative research partners in order to accomplish the GWAS project [61, 65].

## C. Implementation/ Practicality

There is also another factor that should be considered—geography, that could affect the implementation and practicality of GWAS, particularly in archipelagic regions. In the Philippines, many of the provinces are separated by waters, creating a physical barrier that could affect the sharing of physical information and resources. In the MalariaGen study, one major challenge for GWAS implementation is the development of research capacity across all participating research sites. In order to ensure that all research sites are capable of conducting site-specific analyses, considerations related to central data repository, local infrastructure, and network infrastructure need to be considered [62].

## D. Adaptation/Integration

The implementation of a GWAS will be more efficient when there is an existing electronic patient database to work on. Currently, medical records in developing countries are transitioning from paper to electronic format, but it is a relatively recent change, and few hospitals have integrated EHR into their systems [70, 71]. Adapting hospital systems to use EHRs is a primary step toward making GWAS feasible in developing countries. EHR has aided many countries in the pursuit of precision medicine. If all hospitals in these developing countries integrated EHRs into their system, it would allow for easier access to information that could be beneficial for GWAS of their population. Overall, the wide adoption of EHRs requires significant effort than what is assumed in developing countries because of challenges related to current poor level of technological advancement, lack of required computer skills, limited resources, the research capacity to use the current technologies, as well as shift in research investment priorities in order to reduce the inequity in international research that currently exists among others [72, 73].

## Moving Forward

Since GWAS requires significant resources to be carried out, a multi-center international collaboration is usually recommended [48]. That is, current challenges can be met by united efforts from government health agencies and genetic research institutions by performing large scale sequencing projects as well as detailed documentation on patients' clinical features and family history. Data privacy and safety by educating the people involved in aspects related to research objectives, subject de-identification, as well as data security should be mandatory [36••]. Lastly, it should be noted that identifying genetic associations via GWAS is just the first step of a long journey. Follow-up analyses, which include focusing on translation of the discovered genetic loci into new biological insights as well as aiming to implement the new knowledge into clinical care, are other major significant challenges that must be tackled past the GWAS stage [74••]. Such challenges are particularly hindered by the premise that the ~ 90% of GWAS-identified genetic variants are located in the noncoding parts of the genome [74••].

To summarize, various challenges and opportunities arise for establishing GWAS in a developing country (Table 1).

**Table 1** Summary of challenges and opportunities for establishing genome-wide association study (GWAS) in a developing country

| Challenges | Opportunities |
|---|---|
| • The adoption of electronic health records (EHRs) including both technology and organizational (stakeholders, policy, resources) barriers<br>• Lack of a uniform EHR or biorepository<br>• Lack of GWAS experts<br>• Some participating hospitals may hesitate to collaborate due to confidentiality issues | • Presence of endemic diseases<br>• Cardiovascular disease (CVD), diabetes, dementia, and HIV are leading causes of death in many developing countries<br>• Research centers and hospitals have started developing patient registries, electronic medical records, and EHRs<br>• Multicenter affiliation of medical specialists in most hospitals<br>• Collaboration between government health agencies and genetic research institutions<br>• Support via consortium from developed nations |

## Conclusion

Several challenges exist among various developing countries that are hindering their capacity to successfully implement GWASs. To establish GWAS in these countries, demand, availability of resources, implementation/practicality, and adaptation/integration will need to be considered. If these considerations are met, then GWAS could potentially be feasible in many developing nations. However, application of the aforementioned factors would take years of work and would involve significant financial, manpower, technological, and infrastructure resources. Further, the possibility of GWAS being used in these countries would depend on the demand. Until human genetics becomes a priority in these nations (i.e., hindered by challenges related to the length of time that it takes for results of GWASs to be relevant for improving health as well as such studies not considered a priority compared to other more immediate translational research), there will be a paucity of GWAS implementation in these economies.

**Acknowledgements** Dr. Gerard Dumancas would like to acknowledge the University of San Agustin—Center for Informatics for hosting his short-term stint to carry out this research. The authors Dr. Gerard Dumancas and Dr. Romulo de Castro wish to thank the Balik Scientist Program of the Department of Science and Technology—Philippine Council for Health, Research, and Development for the opportunity to serve the Filipino community through science, technology, and innovation. The Balik (Filipino word for "Return" or "Repatriate") Scientist Program (BSP) seeks highly trained Filipino scientists, technologists, experts, and professionals residing abroad to return to the Philippines and transfer their expertise to the local community for the acceleration of scientific, agro-industrial, and economic development of the country.

**Funding** The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the Department of Science and Technology—Balik Scientist Program.

## Compliance with Ethical Standards

**Conflict of Interest** The authors declare no conflict of interest regarding this work.

**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. National Human Genome Research Institute. Genome-Wide Association Studies Fact Sheet [Internet]. Genome.gov. 2022 [cited 2022 Jul 7]. Available from: https://www.genome.gov/about-genomics/fact-sheets/Genome-Wide-Association-Studies-Fact-Sheet
2. Ikegawa S. A short history of the genome-wide association study: where we were and where we are going. Genomics & informatics. Korea Genome Organization; 2012;10:220.
3. Sanghera DK, Blackett PR. Type 2 Diabetes Genetics: Beyond GWAS. J Diabetes Metab. 2012;3:6948.
4. Deng X, Sabino EC, Cunha-Neto E, Ribeiro AL, Ianni B, Mady C, et al. Genome Wide Association Study (GWAS) of Chagas Cardiomyopathy in Trypanosoma cruzi Seropositive Subjects. PLoS ONE. 2013;8: e79629.
5. Tan E-K, Kwok H-K, Tan LC, Zhao W-T, Prakash KM, Au W-L, et al. Analysis of GWAS-linked loci in Parkinson disease reaffirms PARK16 as a susceptibility locus. Neurology AAN Enterprises. 2010;75:508–12.
6. Bloom BR, Trach DD. Genetics and developing countries: the human genome has unifying messages as well as the potential to divide. BMJ. British Medical Journal Publishing Group; 2001. p. 1006–7.
7. National Library of Medicine. What are genome-wide association studies?: MedlinePlus Genetics [Internet]. 2022 [cited 2022 Jul 7]. Available from: https://medlineplus.gov/genetics/understanding/genomicresearch/gwastudies/
8. Fallin MD, Duggal P, Beaty TH. Genetic Epidemiology and Public Health: The Evolution From Theory to Technology. Am J Epidemiol. 2016;183:387–93.
9.•• Peterson RE, Kuchenbaecker K, Walters RK, Chen C-Y, Popejoy AB, Periyasamy S, et al. Genome-wide Association Studies in Ancestrally Diverse Populations: Opportunities, Methods, Pitfalls, and Recommendations. Cell. 2019;179:589–603. **This study shows that despite the perception that analyzing**

**genetic data from diverse populations comes with a challenge, it is still imperative to pursue such avenue scientifically and ethically.**

10. Cutiongco-De la Paz EM. Genomic Association Studies in Filipinos on Treatment, Diagnosis and Risk Assessment of Type 2 Diabetes Mellitus and Its Related Medical Condition [Internet]. Philippine Genome Center. 2022 [cited 2022 Jul 7]. Available from: https://pgc.up.edu.ph/research-development/program-on-health/genomic-association-studies-in-filipinos-on-treatment-diagnosis-and-risk-assessment-of-type-2-diabetes-mellitus-and-its-related-medical-condition/
11. Weatherall DJ. The global problem of genetic disease. Annals of human biology. Taylor & Francis; 2005;32:117–22.
12. Séguin B, Hardy B-J, Singer PA, Daar AS. Genomic medicine and developing countries: creating a room of their own. Nat Rev Genet Nature Publishing Group. 2008;9:487–93.
13. Collins FS, Green ED, Guttmacher AE, Guyer MS. A vision for the future of genomics research. Nature Nature Publishing Group. 2003;422:835–47.
14. United Nations. World Economic Situation Prospects [Internet]. 2021. Available from: https://www.un.org/development/desa/dpad/wp-content/uploads/sites/45/WESP2021_ANNEX.pdf
15. Malaria Genomic Epidemiology Network. A global network for investigating the genomic epidemiology of malaria. Nature. 2008;456:732–7.
16. Choudhury A, Sengupta D, Aron S, Ramsay M. The H3Africa Consortium: Publication Outputs of a Pan-African Genomics Collaboration (2013 to 2020) [Internet]. Brill; 2021 [cited 2022 May 2]. p. 257–304. Available from: https://brill.com/view/book/9789004500228/BP000021.xml
17. Fogarty International Center. Human Heredity and Health in Africa (H3Africa) [Internet]. Fogarty International Center. 2021 [cited 2022 May 2]. Available from: https://www.fic.nih.gov:443/Funding/Pages/collaborations-h3africa.aspx
18.•• Mattap SM, Mohan D, McGrattan AM, Allotey P, Stephan BC, Reidpath DD, et al. The economic burden of dementia in low- and middle-income countries (LMICs): a systematic review. BMJ Global Health. BMJ Specialist J. 2022;7:e007409. **This review demonstrated that the estimated costs of dementia in LMICs are lower than in high-income countries with indirect costs contributing mostly to such costs in LMICs. It is suggested that management of comorbidities and early detection of dementia in LMICs can greatly reduce costs.**
19.• Misra A, Gopalan H, Jayawardena R, Hills AP, Soares M, Reza-Albarrán AA, et al. Diabetes in developing countries. J. Diabetes. 2019;11:522–39. **This study discussed the current situation of type 2 diabetes in developing countries. It also discussed various challenges and opportunities for the management of such disease in such countries.**
20. Celermajer DS, Chow CK, Marijon E, Anstey NM, Woo KS. Cardiovascular Disease in the Developing World: Prevalences, Patterns, and the Potential of Early Disease Detection. J Am Coll Cardiol. 2012;60:1207–16.
21. Addressing Depression in Developing Countries [Internet]. 2017 [cited 2022 Apr 8]. Available from: https://www.borgenmagazine.com/depression-in-developing-countries/
22. Padilla CD, Cutiongco-de la Paz EM. Genetics and genomic medicine in the Philippines. Mol Genet Genomic Med. 2016;4:494–503.
23. Chen Y, Hora B, DeMarco T, Berba R, Register H, Hood S, et al. Increased predominance of HIV-1 CRF01_AE and its recombinants in the Philippines. J Gen Virol. 2019;100:511–22.
24.•• Galarion MaJ, Schwem B, Pangilinan C, dela Tonga A, Petronio-Santos JA, delos Reyes E, et al. Genotypic persistence of dengue virus in the Philippines. Infection, Genetics and Evolution. 2019;69:134–41. **The study aimed to characterize all available Philippine genetic molecular sequences related to the dengue virus in the Philippines. Different DENV genotypes may warrant continuous molecular surveillance for signs of genotypic shifts that can lead to an increased severity risks as well as cause local outbreak related to the dengue virus in the country.**
25. Kawarai T, Morigaki R, Kaji R, Goto S. Clinicopathological Phenotype and Genetics of X-Linked Dystonia-Parkinsonism (XDP; DYT3; Lubag). Brain Sci. 2017;7:72.
26. Topno RK, Das VNR, Ranjan A, Pandey K, Singh D, Kumar N, et al. Asymptomatic Infection with Visceral Leishmaniasis in a Disease-Endemic Area in Bihar. India Am J Trop Med Hyg. 2010;83:502–6.
27. Lun Z-R, Wu M-S, Chen Y-F, Wang J-Y, Zhou X-N, Liao L-F, et al. Visceral Leishmaniasis in China: an Endemic Disease under Control. Clinical Microbiology Reviews. Am Soc Microbiol. 2015;28:987–1004.
28. H3Africa. H3Africa – Human Heredity & Health in Africa [Internet]. 2022 [cited 2022 May 2]. Available from: https://h3africa.org/
29.• Gulsuner S, Stein DJ, Susser ES, Sibeko G, Pretorius A, Walsh T, et al. Genetics of schizophrenia in the South African Xhosa. Science. American Association for the Advancement of Sci. 2020;367:569–73. **Results of this study show that the greater genetic variation among African Americans than that of Europeans provides the former more power to detect relationships of genes to phenotypes.**
30.• Boua PR, Brandenburg J-T, Choudhury A, Hazelhurst S, Sengupta D, Agongo G, et al. Novel and Known Gene-Smoking Interactions With cIMT Identified as Potential Drivers for Atherosclerosis Risk in West-African Populations of the AWI-Gen Study. Frontiers in Genetics [Internet]. 2020 [cited 2022 May 2];10. Available from: https://doi.org/10.3389/fgene.2019.01354. **This is the first study to show a gene-smoking interaction for cIMT, as a risk factor for atherosclerosis in sub-Saharan African populations**
31. 129 Free Genome-wide Association Study (GWAS) Tools - Software and Resources [Internet]. 2022 [cited 2022 Jul 7]. Available from: https://bioinformaticshome.com/tools/gwas/gwas.html
32. Bonte C, Makri E, Ardeshirdavani A, Simm J, Moreau Y, Vercauteren F. Towards practical privacy-preserving genome-wide association study. BMC Bioinformatics. 2018;19:537.
33. The Health Privacy Code of the Philippines - Part 1 [Internet]. 2019. Available from: https://www.privacyph.net/2019/07/09/the-health-privacy-code-of-the-philippines-part-1/
34. Genes in Life. Registries and biobanks | Genes in Life [Internet]. 2022 [cited 2022 Jul 7]. Available from: http://www.genesinlife.org/research/registries-and-biobanks
35. Adair L, Mohlke K, Popkin B, Bollen K, Guilkey D, Gultiano S, et al. The Cebu Longitudinal Health and Nutrition Survey: dbGaP accession phs000523.v1.p1 [Internet]. [cited 2022 Apr 8]. Available from: https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000523.v1.p1&phv=173308&phd=&pha=&pht=2922&phvf=&phdf=&phaf=&phtf=&dssp=1&consent=&temp=1
36.•• Pemmasani SK, Raman R, Mohapatra R, Vidyasagar M, Acharya A. A Review on the Challenges in Indian Genomics Research for Variant Identification and Interpretation. Frontiers in Genetics [Internet]. 2020 [cited 2022 Apr 3];11. Available from: https://www.frontiersin.org/article/10.3389/fgene.2020.00753. **The study discussed various challenges faced by genomic testing facilities and genetic researchers in India across various areas. The manuscript also discussed potential solutions to enhance genomic research in the country.**
37.•• Ahmed P H, V V, More RP, Viswanath B, Jain S, Rao MS, et al. INDEX-db: The Indian Exome Reference Database (Phase I).

J Comput Biol. 2019;26:225–34. **The study demonstrated a genetic reference map from the Indian population.**
38. Upadhyay P, Gardi N, Desai S, Sahoo B, Singh A, Togar T, et al. TMC-SNPdb: an Indian germline variant database derived from whole exome sequences. Database (Oxford). 2016;2016:baw104.
39. Ngamphiw C, Assawamakin A, Xu S, Shaw PJ, Yang JO, Ghang H, et al. PanSNPdb: the Pan-Asian SNP genotyping database. PLoS ONE. 2011;6: e21451.
40.•• Gao Y, Zhang C, Yuan L, Ling Y, Wang X, Liu C, et al. PGG. Han: the Han Chinese genome database and analysis platform. Nucleic Acids Research. 2020;48:D971–6. **The study showed a population genomic database that will serve as the central repository for the genomic data emanating from the Han Chinese Genome Initiative (Phase I).**
41. Mulder N, Abimiku A, Adebamowo SN, de Vries J, Matimba A, Olowoyo P, et al. H3Africa: current perspectives. Pharmgenomics Pers Med. 2018;11:59–66.
42. Schneider MJ. Chapter Two Pyridine and piperidine alkaloids: An update. In: Pelletier SW, editor. Alkaloids: Chemical and Biological Perspectives [Internet]. Pergamon; 1996 [cited 2022 Feb 12]. p. 155–299. Available from: https://www.sciencedirect.com/science/article/pii/S0735821096800264
43.•• Agler CS, Shungin D, Ferreira Zandoná AG, Schmadeke P, Basta PV, Luo J, et al. Protocols, Methods, and Tools for Genome-Wide Association Studies (GWAS) of Dental Traits. Methods Mol Biol. 2019;1922:493–509. **This interesting study showed the protocols, methods, and tools for GWAS of dental traits which can serve as a template for future genome investigations of other traits.**
44. Hong EP, Park JW. Sample size and statistical power calculation in genetic association studies. Genomics Inform. 2012;10:117–22.
45.•• Wang M, Xu S. Statistical power in genome-wide association studies and quantitative trait locus mapping. Heredity (Edinb). 2019;123:287–306. **This study demonstrated that large samples are not necessary to detect a biologically meaningful quantitative trait loci.**
46. Keith T. Human Genome-Wide Association Studies [Internet]. GEN - Genetic Engineering and Biotechnology News. 2007 [cited 2022 Jul 7]. Available from: https://www.genengnews.com/magazine/63/human-genome-wide-association-studies/
47.•• Pennsylvania PS of M at the U of. New statistical method exponentially increases ability to discover genetic insights [Internet]. 2021 [cited 2022 Jul 7]. Available from: https://medicalxpress.com/news/2021-01-statistical-method-exponentially-ability-genetic.html. **This report showed that a statistical method called Sum-Share can exponentially increase ability to discover genetic insights by pulling summary information from many different sites to generate significant insights.**
48. Du Y, Xie J, Chang W, Han Y, Cao G. Genome-wide association studies: inherent limitations and future challenges. Front Med. 2012;6:444–50.
49.• Schurz H, Müller SJ, van Helden PD, Tromp G, Hoal EG, Kinnear CJ, et al. Evaluating the Accuracy of Imputation Methods in a Five-Way Admixed Population. Frontiers in Genetics [Internet]. 2019 [cited 2022 Jul 7];10. Available from: https://www.frontiersin.org/articles/10.3389/fgene.2019.00034. **The study compared the imputation results obtained using three freely-accessible methods for accuracy and quality.**
50. Marcelo A, Medeiros D, Ramesh K, Roth S, Wyatt P. Transforming health systems through good digital health governance. Asian Development Bank; 2018. https://doi.org/10.22617/WPS189244-2
51.•• Jalea G. DOST budget for research cut by ₱76 million [Internet]. CNN Philippines. 2020 [cited 2022 Jul 7]. Available from: https://www.cnnphilippines.com/news/2020/9/9/DOST-budget-P76-million-research.html. **This study was a recent report showing that the research and development budget of the Philippine government was cut by substantial amount, thus, affecting the advancement of science in the Philippine nation.**
52. Panela S. 5 things to make PH a better place for scientists [Internet]. Rappler. 2017 [cited 2022 Mar 21]. Available from: https://www.rappler.com/science/169871-philippines-better-place-scientists/
53. UNESCO Institute for Statistics. Research and development expenditure (% of GDP) | Data [Internet]. 2021 [cited 2022 Mar 21]. Available from: https://data.worldbank.org/indicator/GB.XPD.RSDV.GD.ZS
54. Varmus H, Klausner R, Zerhouni E, Acharya T, Daar AS, Singer PA. PUBLIC HEALTH: Grand Challenges in Global Health. Science. 2003;302:398–9.
55. Human Hereditary and Health in Africa. Funding – H3Africa [Internet]. 2022 [cited 2022 Jul 7]. Available from: https://h3africa.org/index.php/resource/funding/
56. Helmy M, Awad M, Mosa KA. Limited resources of genome sequencing in developing countries: Challenges and solutions. Applied & Translational Genomics. 2016;9:15–9.
57. Official Gazette. List of National Scientists of the Philippines | GOVPH [Internet]. Official Gazette of the Republic of the Philippines. 2022 [cited 2022 Jul 7]. Available from: https://www.officialgazette.gov.ph/lists/national-scientists-of-the-philippines/
58. Kang JJ. The Thousand Talents Plan is part of China's long quest to become the global scientific leader [Internet]. The Conversation. 2020 [cited 2022 Jul 7]. Available from: http://theconversation.com/the-thousand-talents-plan-is-part-of-chinas-long-quest-to-become-the-global-scientific-leader-145100
59. Department of Science and Technology of the Philippines. Balik Scientist Program [Internet]. 2020 [cited 2022 Jul 7]. Available from: https://bspms.dost.gov.ph/
60. Yang L, Marini G. Research Productivity of Chinese Young Thousand Talents. International Higher Education. 2019;17–8.
61. Parker M, Bull SJ, de Vries J, Agbenyega T, Doumbo OK, Kwiatkowski DP. Ethical Data Release in Genome-Wide Association Studies in Developing Countries. PLoS Med. 2009;6: e1000143.
62. de Vries J, Bull SJ, Doumbo O, Ibrahim M, Mercereau-Puijalon O, Kwiatkowski D, et al. Ethical issues in human genomics research in developing countries. BMC Med Ethics. 2011;12:5.
63. Mulder NJ, Adebiyi E, Adebiyi M, Adeyemi S, Ahmed A, Ahmed R, et al. Development of Bioinformatics Infrastructure for Genomics Research. Global Heart Ubiquity Press. 2017;12:91–8.
64. Lohmann E, Gasser T, Grundmann K. Needs and Requirements of Modern Biobanks on the Example of Dystonia Syndromes. Front Neurol. 2017;8:9.
65. Chokshi DA, Thera MA, Parker M, Diakite M, Makani J, Kwiatkowski DP, et al. Valid consent for genomic epidemiology in developing countries. PLoS Med. 2007;4: e95.
66. Prasad RS. The Personal Data Protection Bill, 2019 [Internet]. PRS Legislative Research. 2019 [cited 2022 Apr 3]. Available from: https://prsindia.org/billtrack/the-personal-data-protection-bill-2019
67. National Privacy Commission. Republic Act 10173 – Data Privacy Act of 2012. 2016. https://www.privacy.gov.ph/data-privacy-act/
68. H3Africa Consortium. Guideline for Informed Consent [Internet]. H3Africa; 2018. Available from: https://h3africa.org/wp-content/uploads/2019/07/H3Africa-IC-Guideline-3rd-Edition-.pdf
69. Bhutta ZA. Ethics in international health research: a perspective from the developing world. Bulletin of the World Health Organization. SciELO Public Health; 2002;80:114–20.

70. Montemayor MaT. Going digital to enhance UHC implementation: health care expert [Internet]. Republic of the Philippines News Agency. 2019 [cited 2022 Jul 7]. Available from: https://www.pna.gov.ph/articles/1065473
71. Kalogriopoulos NA, Baran J, Nimunkar AJ, Webster JG. Electronic medical record systems for developing countries: Review. 2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society. 2009. p. 1730–3.
72. Odekunle FF, Odekunle RO, Shankar S. Why sub-Saharan Africa lags in electronic health record adoption and possible strategies to increase its adoption in this region. Int J Health Sci (Qassim). 2017;11:59–64.
73. Rudan I, Marušić A, Campbell H. Developing biobanks in developing countries. J Glob Health. 2011;1:2–4.
74.•• Loos RJF. 15 years of genome-wide association studies and no signs of slowing down. Nat Commun. Nature Publishing Group. 2020;11:5900. **This communication reported that over the past 15 years, the field of GWAS has tremendously matured with many cutting-edge technologies developed, with the ultimate goal of translating GWAS into new opportunities for clinical care.**