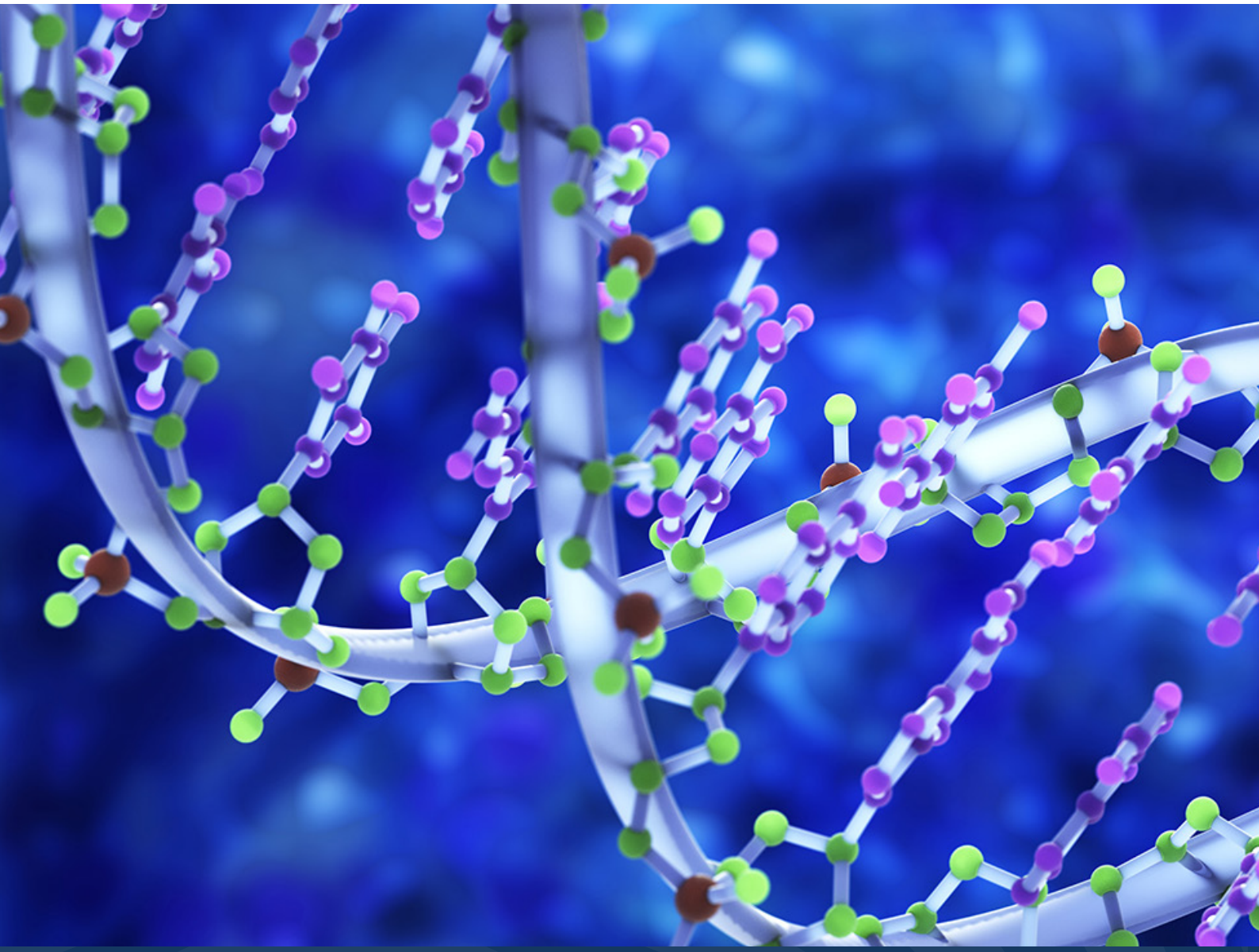
The African Academy of Sciences

AESA AN INITIATIVE OF THE AAS & AUDA (AFRICAN UNION DEVELOPMENT AGENCY)

AUDA-NEPAD AFRICAN UNION DEVELOPMENT AGENCY



A Framework for the Implementation of Genomic Medicine for Public Health in Africa

8 February 2021

Contents

Appendix 6: Acknowledgements

This document is part of the outputs of the science prioritisation exercises being conducted under the Africa's Scientific Priorities Programme (ASP), a five-year initiative being implemented by the AAS-AESA in partnership with AUDA-NEPAD and with financial support from the Bill and Melinda Gates Foundation. We acknowledge the contributions of the Genomics and Precision Committee (Appendix 4) who contributed extensively to this document. The previous versions of this document have been reviewed by external reviewers and presented for validation at AAS and AUDA-NEPAD meetings. Staff who contributed to the prioritisation process and preparation of the document include Isayvani Naicker, Jennifer Maroa, Grace Mwaura, Rosalyn Wamuyu, Olivia Osula, Gladys Akinyi, Juliette Mutheu-Asego, Deborah-Fay Ndlovu, and Davies Mbela.

Acronyms

AACR	American Association for Cancer Research
AADM	Africa America Diabetes Mellitus
ACGME	Accreditation Council for Graduate Medical Education
ACMG	American College of Medical Genetics and Genomics
ADME	Absorption, Distribution, Metabolism and Elimination
AGMO	Alkylglycerol Monooxygenase
AMR	Anti-Microbial Resistance
ART	Anti-retroviral Therapy
BRCA1	Breast Cancer Type 1 Susceptibility Protein
BRCA2	Breast Cancer Type 2 Susceptibility Protein
BSP	Bantu Speaking Populations
CAD	Coronary Arterial Disease
CGA	Candidate Gene Approach
CHAIR	Cardiovascular H3Africa Innovation Resource
CHPC	Centre for High-Performance Computing
CNVs	Copy Number Variants
COVID19	Coronavirus Disease 19
CPIC	Clinical Pharmacogenetics Implementation Consortium
CVD	Cardiovascular Disease
CWL	Common Workflow Language
DALYs	Disability-adjusted life years
DMZ	“demilitarised zone”, section of a computer network that allows communications that bypass firewalls and routers
DNA	Deoxyribonucleic acid
EDCTP	European & Developing Countries Clinical Trials Partnership
EGFR	Epidermal growth factor receptor
ELSI	Ethical, Legal and Social Implications
ENIGMA	Evidence based Network for the Interpretation of Germline Mutant Alleles
ERSPC	European Randomised Study of Screening for Prostate Cancer
eQTL	Expression Quantitative Trait Locus
ExAC	Exome Aggregation Consortium
FAIR	Findable, Accessible, Interoperable and Reusable
GDPR	General Data Protection Regulation (EU)
GenCC	Gene Curation Coalition
GENIE	Genomics Evidence Neoplasia Information Exchange
GM	Genomic Medicine
gnomAD	Genome Aggregation Database
GTR	Genetic Testing Registry
GWAS	Genome Wide Association Studies
H3A	H3Africa
HGMD	Human Gene Mutation Database
HICs	High Income Countries
HIV	Human Immunodeficiency Virus
HPC	High Performance Computing
ICU	Intensive Care Unit
LD	Linkage Disequilibrium
LDL	Low Density Lipoprotein
LMICs	Low- and Middle-Income Countries
MAF	Minor Allele Frequency

MDR	Multiple Drug Resistance
miRNA	Micro RNA
NCD	Non-communicable Disease
NCDs	Non-Communicable Diseases
NEMD	National Ethnic Mutation Databases
NGS	Next Generation Sequencing
NIH	National Institutes of Health
noSQL	Non-relational Database Architecture
NSCLC	Non-Small-Cell Lung Carcinoma
PARP	Poly ADP Ribose Polymerase
PB	PetaBytes
PCR	Polymerase Chain Reaction
PCV	Pneumococcal Conjugate Vaccine
PDE	Phosphodiesterase
PfEMP	*Plasmodium falciparum* Erythrocyte Membrane Protein
PM	Personalised Medicine
POPIA	Protection of Personal Information Act
RAID	Redundant Array of Inexpensive Disks
REDCap	Research Electronic Data Capture
RPPA	Reverse-Phase Protein Arrays
SCD	Sickle Cell Disease
SNP	Single Nucleotide Polymorphism
SSA	Sub Saharan Africa
TB	Tuberculosis
TCGA	The Cancer Genome Atlas
UTRs	Untranslated Regions
VEGF-A	Vascular Endothelial Growth Factor-A
VUS	Variant of Unknown Significance
WACREN	West and Central African Research and Education Network
WGS	Whole Genome Sequencing

Executive Summary

The term "Precision Medicine" describes the use of specific information about an individual to facilitate a more precise approach to their healthcare. The term is very broad with wide reaching approaches and implementations. One of the more specific approaches is genomic medicine, where knowledge about a person or population's genetic makeup can be used to derive the most appropriate diagnosis, treatment and, where possible, disease prevention strategy. Genomic medicine has been applied to a number of areas of healthcare, including perinatal testing and newborn screening for monogenic diseases, cancer screening and treatment, rare disease diagnosis and treatment, pharmacogenomics and sequencing of pathogen genomes. However, most of these activities have been restricted to developed countries. Some of the key barriers to the implementation of genomic medicine world-wide, but more pronounced in developing countries include clinical and data infrastructure, regulatory environments, integration of new technologies into clinical practice, and cost. Knowledge and evidence generation are also essential elements which are constantly evolving as new genomes are sequenced and genome wide association studies are undertaken. In Africa, due to under-representation of African populations in such studies, knowledge and evidence for actionability on genomics data are lagging behind the developed world. This, along with poorer resources and infrastructure limitations, mean that African countries need to overcome greater barriers to implement genomic medicine, but this is by no means an impossible task, challenges can be addressed in a stepwise manner.

The aim of this framework document is to highlight these elements within the African context and provide some recommendations on how African countries can work on putting them in place by building on existing infrastructure. This will enable resource limited countries to start implementing appropriate genomics-based health-related interventions, drawing on experiences from elsewhere, but adapting to the African context where necessary. The document is written by volunteers, who are not necessarily experts in genomic or precision medicine but have a variety of data, science and clinical backgrounds. It is not aiming to be prescriptive or comprehensive, but rather to provide some information and recommendations to enable the implementation of genomic medicine. The document focuses on the clinical context, while indicating barriers to implementation (situational and needs), which includes additional research that is required for knowledge and evidence generation.

Some of the key elements required for implementation of genomic medicine in clinical care that have previously been described include:

- **Clinical facilities for patient counselling, screening, treatment and monitoring**
- **Sample collection, processing and data generation facilities**
- **Data storage, curation, analysis, interpretation and sharing infrastructure**
- **Knowledge bases with up to date information on genotype-phenotype link and actionability**
- **Research facilities to increase knowledge on genomics in African populations**
- **Genomic medicine training programmes for healthcare professionals**
- **Regulatory, data governance and ethics consent processes governing all the above activities**

Translational genomics has blurred the distinction between research and service delivery across the continuum of care, from health maintenance to disease management, so, ideally, research and healthcare should remain interconnected. Additionally, since research on African populations is under-represented compared to well-studied northern populations, it cannot be decoupled from the clinical implementation.

The framework is divided into sections that address the different elements required for implementation of genomic medicine. Each section aims to highlight the major requirements, challenges and any relevant African-specific considerations. The sections conclude with a set of practical recommendations. A summary of these are provided below:

Introduction

The introduction defines precision and genomic medicine and the benefits to healthcare provision, highlighting the need for genomic medicine in African countries. Africa has a disproportionate burden of diseases, and yet has been excluded from many large-scale genomics studies in the past, as well as from clinical trials. The age and diversity of African populations, together with exposure to different and extreme environmental pressures means that genomics knowledge generated from other world populations is not appropriate for application to the provision of healthcare in Africa. African countries need to take ownership of generating new knowledge on the genetic landscape of its people to improve the precision and accuracy with which healthcare decisions can be made.

Infrastructure Requirements

There is significant infrastructure required for genomic medicine implementation, including clinical facilities, biorepositories, data generation (sequencing, genotyping, etc.) facilities, and data storage, management and analysis platforms. This section describes these in more detail together with existing infrastructures that can be leveraged.

Key recommendations

- Develop a national data capture infrastructure (e.g. REDCap) for entering and managing research or clinical data about patients that complements clinical records used in standard healthcare.
- Equip points of care facilities to extract and quality check DNA and run simple genetic tests.
- Establish or source access to high-throughput genotyping and sequencing facilities. Implement accreditation systems for clinical laboratory systems to conduct genotyping and sequencing.
- Establish/use a well-run biorepository that has the capacity to store, distribute and manage the samples being collected.
- Develop security protocols for samples and data and adhere to relevant data sharing and data storage laws and guidelines.
- Ensure computer networks and internet have the capacity to move large datasets reliably between data generation, analysis, and long-term storage locations.
- Leverage existing or develop new computational resources for large-scale data analysis.

Cohorts and diseases

The focus of this section is on the disease priority areas and the need for additional reference data. As mentioned above, there is a dearth of genomics data describing the current genetic landscape of African populations for both the healthy and disease state. With migration, admixture and multiple different ethnic groups residing in several countries, data on broader African population groups rather than just country-specific information is needed. However, disease focus areas should relate to the specific disease burdens within a country and applicability of genomics to addressing those burdens. Specific examples of non-communicable and infectious diseases are discussed, as well as the use of genomics for pathogen surveillance and monitoring outbreaks.

Key recommendations

- Generate and analyse sequence and genotyping data for more diverse African human reference genomes.
- Establish longitudinal cohorts with inclusive participation from different geographical regions for surveillance and to detect and monitor diseases.
- Carry out GWA studies in African populations to identify and confirm variants relevant for African populations for specific diseases. Findings should be functionally validated.
- Assess relevance of disease-causing variants in databases (e.g. ClinVar) for African populations and diseases.
- Determine polygenic risk scores for specific populations for complex traits and develop disease panels accordingly.
- Implement PCR or other assays for single mutations, Arrays (CGH, SNPs/pathogenic mutations, pharmacogenomic variants) where disease-specific panels are available, or NGS (single genes, gene panels, WES, WGS) for more complex diseases and undiagnosed disorders in healthcare settings.
- In addition to traditional typing tools, implement WGS for pathogen surveillance and monitoring, routine pathogen diagnosis and drug resistance testing.

Phenotype and clinical data

Clinical data collection should be standardised and integrated into healthcare systems. The data can come from either research cohorts or electronic (or paper) health records and can be collected through electronic capture forms or mobile devices. In addition to following standards to ensure high quality and consistency, collection and use of clinical data should follow ethics and data governance guidelines.

Key recommendations

- Build governance systems for clinical data, such as informed consent processes for research use and standard operating procedures for research access to data.
- Build relationships to facilitate ethical and consented sharing of routine health data for genomic medicine research and ensure feedback of findings and results to health services.
- Ensure that new clinical data collection is done in collaboration with existing health infrastructure and using data standardisation with existing, commonly used standards
- Leverage existing REDCap databases and/or database templates, using federated data storage where needed.

Linking genes and diseases

Genomic medicine relies on evidence for a link between a genetic feature and a clinical action. This section addresses the correlation between genotype and phenotype data, describing some of the existing approaches for identifying and validating disease-causing (pathogenic) variants. It highlights what information is already available and actionable, and how new disease-associated variants can be identified in African populations. Pharmacogenomics is discussed as a tangible implementation.

Key recommendations

- Implement genetic testing for known alleles of pharmacogenes in Africa, as this requires only DNA extraction and PCR facilities.
- Generate more African genotype data linked to drug response information in pharmacogenomic studies and clinical trials.
- Harness existing genotype-phenotype data to prioritise variants for follow-up and study design for further research.
- Generate new genotype-phenotype data for African populations through large-scale genetic studies on larger African cohorts and under-represented African subpopulations.
- Establish an African expert review panel to work with existing international panels to review novel variants and evidence for actionability.
- Establish guidelines on the use of genomics data in clinical practice in Africa.

ELSI and data governance

The use of personal genomic data has ethical, social and legal implications that need to be addressed to ensure that participants understand the risks and benefits of participating in genomic medicine-based research and clinical care, and that their data are used appropriately. This section discusses the considerations associated with consent and participant privacy protection. The key recommendations are:

- Catalogue the national-level legislation across African countries to build a matrix describing personal data protection through legislation in these countries.
- Draft an African treaty on research integrity and addressing digital disparity in working with vulnerable populations.
- Establish guidelines for the informed consent process to ensure appropriate participant information, informed consent processes and community education is provided.
- Develop processes to evaluate ethics implementation in ongoing projects, as well

as in clinical settings and ensure quality assurance

- Provide training for counsellors on genomics, researchers on community engagement, and Institutional Ethics Review Boards who evaluate informed consent processes.
- Create an experts' platform to catalogue ethics and data governance human resources, skills and expertise which can be accessible to countries where such resources are limited.

Education and Training

The genomic medicine ecosystem involves a number of different stakeholders including patients, physicians, genetic counsellors, pathologists, laboratory technicians, data generators and data analysts, among others. Many of these have not been trained in genetics or genomics, highlighting the need to restructure existing training curricula and promote up-skilling of healthcare professionals and other professions involved in this ecosystem. There is thus a need to develop competencies specific to Africa to guide genomic medicine training for this broad range of stakeholders.

Key recommendations

- Create and design job descriptions for new healthcare and other professionals required for implementation of genomic medicine in Africa.
- Identify and adapt existing competencies to guide curriculum development for these professions, both for new degrees and professional development.
- Conduct a curriculum review of the existing Genetic Counselling and Medical Genetics speciality areas and adapt them for other institutes.
- Increase learning of basic science subjects in medical curricula, focusing on how this knowledge can directly impact clinical practice, include health informatics education where applicable.
- Establish accreditation bodies in genomic medicine and align to existing health professional bodies.
- Assess policies and lobby policy makers to recognise genomic medicine as a speciality field

Moving from Research to Translation

Genomics activities on the continent are currently being carried out predominantly in the context of research. There is a need to shift the balance towards implementation of genomics in healthcare settings, but also to bring research and healthcare closer together and to promote translation of basic research into health benefits. This section describes some of the key considerations for translation into clinical practice using an existing pathology-supported testing framework being implemented in South Africa as a case study.

Key recommendations

- Develop capacity of a new cadre of geneticists at MSc and PhD level, participating in ethics approved translational research projects.
- Develop test validation and feasibility studies to demonstrate the value of low-cost PoC tests in comparison with traditional laboratory-based genetic testing of well-established biomarkers with proven clinical utility.
- Conduct research on the development of novel and relevant genomic medicine test panels for use on PoC DNA devices.
- For new tests, assess analytical and scientific validity and clinical utility, including the sensitivity and specificity of the tests.
- Incorporate novel findings into an existing body of knowledge towards engagement of African scientists in the commercial manufacture of test panels for clinical use.
- Implement integrated data systems for analysis, interpretation and report generation, and AI for clinical decision making.

Stakeholder Engagement

In addition to stakeholders involved in genomic medicine at the cold face, there are other parties involved at different levels. These include members of the general public, researchers in genetics, genomics, bioinformatics and public health, biomedical data analytics companies, private insurers and government departments. This section describes the potential role of different stakeholders in the genomic medicine ecosystem, which may differ across countries. Though this framework is designed for a country-level implementation, we provide suggestions on how country-level implementations can benefit

from cross-border collaboration.

Key recommendations

- Regional and national efforts should define who is responsible for engaging stakeholders and understanding their requirements.
- Undertake stakeholder mapping, active community and stakeholder engagement, and meaningful qualitative research to align the agendas of all stakeholders.
- Define regional/national processes to hold stakeholders accountable and ensure that genomic medicine activities are appropriate, equitable and ethical.
- Encourage governments to work together across departments and engage with other stakeholders to ensure effective genomic medicine implementation.
- Keep communication and processes across different country-level efforts open to enable long-term collaboration between countries.

Conclusion

The elements described above should ideally be well integrated with the healthcare system, including electronic health records to ensure the clinicians have access to all the information required to provide the most appropriate healthcare plan for their patients. The requirements for implementation of genomic medicine seem vast, complex and expensive, but in some countries, some of the pieces are already in place at a basic level through existing health systems or as a result of recently developed research infrastructures.

An important question though, is who is responsible for funding the individual components? A collaboration between government departments would be beneficial as the ecosystem requires elements of health, science, education and policy, but industry partnerships and medical insurance companies can also have an important role to play. Through national coordination of efforts and taking small steps with pilot projects, this large, unimaginable undertaking can be realised, particularly in some countries with reasonable infrastructure already in place.

Section 1. Introduction

The most accurate prevention, diagnosis and treatment of diseases should be available to all of humanity, not just those who can afford it. Precision medicine (PM) is being increasingly used in the developed world to improve the health of individuals. Although some precise methods for diagnosis and treatment are in practice in Africa, these are predominantly "generalised" to reach as wide a community as cheaply as possible. With a high disease burden on the continent, precision public health, which aims to more accurately diagnose and treat diseases at a population or sub-population level is likely to be more readily adopted than an individual-level approach. Precision medicine (defined later) and precision public health are broad concepts and can encompass the application of many different technologies (including all the omics), together with health, social and environmental data for applications in health. Genomic medicine, on the other hand narrows the focus to the use of genomics in healthcare.

This document proposes a framework incorporating many of the elements required for implementation of genomic medicine (GM) within countries in Africa, which use genetic or genomic information about individuals or populations to improve their health care. Therefore, though some aspects of precision medicine (PM) may be eluded to throughout the document, we acknowledge that the content is more specific to genomics applications. While we recognise their importance, trying to cover all aspects of PM, including the use of other omics technologies and incorporation of environmental data is beyond the scope of this document. The document has been written by volunteers from different sectors of science and health who do not claim to be experts in GM; there are undoubtedly omissions, but the hope is to provide a useful discussion point from which an African country or region can develop a genomic medicine/precision public health implementation plan. We also acknowledge that there is an abundance of materials and resources available on the subject, not all of which are covered here. There are also now well established national precision medicine initiatives (Figure 1) for location of such programmes, noting the complete absence of any in Africa), such as Genomics England in the UK, All of Us in the USA and Australian Genomics, among others, who have developed documents, processes and standard operating procedures, some of which are referenced in the document. These initiatives have experiences that should be drawn from to enable new initiatives to leapfrog in their implementations.

This introduction provides some background on the disease burden in Africa, the need for precision medicine, and some of the global and local challenges to implementation. Thereafter, each section covers a different element or infrastructural requirement for genomic medicine and provides some key recommendations. While the focus is on the clinical implementation of GM, there is also an emphasis on research due to the limited data and evidence for genetic associations with diseases in African populations. Ideally this basic research should lead to applied research and then translation through proof of principle for clinical utility, economic and feasibility studies, including validation and potential clinical uptake, and finally clinical implementation.

1.1. The Disease Burden in Africa

Africa is home to over 1.2 billion people in 54 countries, living in very diverse settings and socioeconomic environments. The current landscape of African populations has been shaped by many historical factors including migration, slavery and colonisation, and more recent changes in the region with a rising wave of emigration from the continent, particularly to Europe. However, the past decades have seen an increase in economic growth and an emerging middle class in many African countries driven by factors such as more stable political regimes, fewer armed conflicts, agricultural and industrial development. Despite these improvements, however, the disease burden remains inordinately high.

Active interventions in health services, including improved immunisation coverage, have in many areas reduced the incidence of infectious diseases. Overall, however, death rates due to disease remain very high in Africa. Malaria, tuberculosis and HIV-related illnesses are among the primary causes of death. Not only are infectious diseases still prevalent – for example, recent data indicate an increase in malaria cases possibly related to climate change and increasing temperatures influencing vector populations, and viral outbreaks such as Ebola in West Africa are still ongoing - but non-communicable diseases (NCDs) are increasing concurrently at an alarming rate. Using data from the Global Burden of Diseases, Injuries, and Risk Factors Study 2017, (Gouda et al. 2019) report that the total disability-adjusted life-years (DALYs) resulting from NCDs increased by 67% between 1990 and 2017. Whilst partially due to population growth and ageing, by 2017 the age-

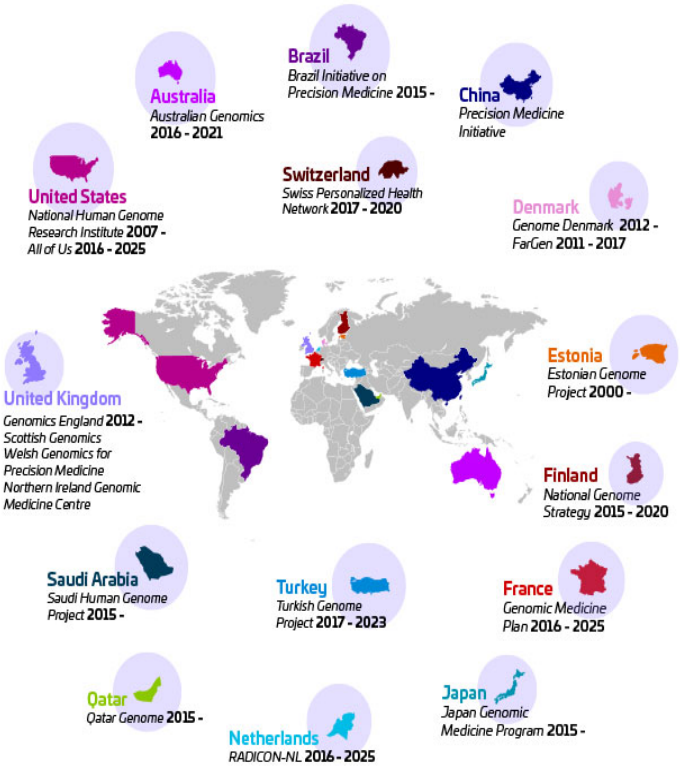


Figure 1.Map showing existing genomic medicine initiatives as of 2019. Source: Stark, et al (2019). Image source: http://www.bio-itworld.com/2019/08/12/national-genomic-data-initiatives-worldwide-update.aspx.

standardised NCD-attributable DALY rate was on a par with that of communicable, maternal, neonatal, and nutritional diseases. Cardiovascular diseases (CVD) have become particularly prevalent in many countries in Africa, and cancer prevalence is equally on the rise. These changes reflect epidemiological transition and associated dual burden of disease in Africa (Mayosi et al. 2009; Gouda et al. 2019).

Health care provision varies dramatically across and within African countries, and acute shortages of resources, disparate health services and poor socioeconomic conditions continue to underlie high death rates. Health systems support and improvements in primary health care remain a top priority, but in an era of evidence-based health care it is also essential to ensure that patients receive efficient and accurate diagnoses and the most effective and affordable treatments for their conditions.

1.2. Genetic diversity within African populations

GM implementation in Africa requires consideration of the diverse ethnic origin of patient groups within and between countries. A high degree of genetic diversity and population stratification has been observed across Africa, with nearly as many ancestries identified within Africa as in the rest of the world combined (Shriner et al. 2014). According to the 'Out of Africa' hypothesis, this reflects the origins of modern man in Africa: from an established, diverse African population only a small subset subsequently populated the rest of the globe, this original human migrational bottle-neck resulted in less genetic diversity in human populations outside of Africa. Also, because humans have spent most of their evolutionary lives on the continent of Africa, the genomes of Africans display unique characteristics including shorter haploblocks, higher linkage disequilibrium and a greater array of genetic variation than rest-of-world populations (Campbell and Tishkoff 2008).

This genetic diversity offers opportunities but is also a real challenge for the replication of global genetic studies in African sub-populations; and population effects must be carefully managed through appropriate study design to reduce the effects of population bias and admixture in the identification of candidate aetiological variants. Genetic variations also arise from the combined effects of population size, migrations, admixture

Box 1. Demographics of African populations

Demographics of African populations

Africa, the cradle of mankind, is where humans originated and has the most genetically diverse populations in the world. In addition to the large number of diverse populations, migration out of and across the continent and its colonisation history have resulted in admixed populations that have adapted to local prevailing climatic and environmental conditions. Genetic studies of the various ethnic groups help to understand the past events that shaped the present-day populations and their possible predisposition to some diseases. Hence the choice of individuals for genetic studies will be influenced by their geographic, ethnic and linguistic origins.

The demographic history of Africa indicates major expansions between 150 000 - 250000; 60 000 - 80000; 25 000 – 40 000 and 12 000 years ago. The movement out of Africa towards Asia and Europe happened 100 000 years ago with a back-migration of ancient Eurasian farmers into Eastern Africa about <5 000 years ago (Gallego Llorente et al. 2015). African populations have significant Eurasian ancestry (up to 23 % in the isiZulu and Sotho) (Gurdasani et al. 2015). Admixture between hunter gatherers and Eurasian happened between 7 500 – 10 500 years ago (Yoruba; Ancient admixture), Yoruba 2 400 – 3 200 years ago (Ethiopian populations; Old admixture) and 150 – 1 500 years ago (East African populations; Recent admixture). Some Neanderthal traits were introgressed into Yoruba through Eurasian admixture from back migration.

Haplotype-based studies (Patin et al. 2017) have demonstrated that Bantu-speaking (BS) people migrated from Western Central Africa into Eastern and Southern Africa around 5000 years ago. The migration meant that they adapted to the different geographical, climatic environments and engendered new genetic structures. One third of the sub-Saharan populations are BS. BS populations moved southwards first, then into Eastern and Southern Africa. In doing so they encountered local populations, which resulted in admixtures and therefore acquisition of gene variants that could be positively selected. In the east they came across an Afro Asiatic group from Ethiopia and in the south the Ju/'hoansi San from Namibia. BS populations were also, more recently, forced to move across the Atlantic from West Central Africa to North America. Hence today's population diversity across Africa is the result of many events that brought people of different geographic regions together as well as the impact of climate and other environmental factors, including diseases.

Box 2. Ethnolinguistic groups in Africa

Ethnolinguistic groups in Africa

Populations on the African continent are grouped into people with the same or similar languages and origin, depending on where they are mostly located (Tishkoff et al. 2009).

There are over 2000 spoken languages in Africa (Eifring and Theil 2005) within the main groups distributed as shown in Figure 2. The principal language families within the continent are: Afroasiatic (Northern Africa), Niger-Congo, also called Niger Kondofanian (Western, Central, Eastern and Southern Africa), Nilo-Saharan (Northern, Central and Eastern Africa) and Khoe-San (South Africa). Others include Afrikaaners, Portuguese, French, Arabic and English (West European ancestry) and Malayo-Polynesians (Malagasy).

There are approximately 84 distinct ethnic groups, although it can be difficult to fully ascertain the limits of each group.

Afroasiatic - Within this phylum are languages of people of Asia and Northern Africa. It groups more than 300 languages and constitutes the second largest language family after Niger-Congo in Africa.

Niger-Congo – This is the largest language family in Africa, and covers part of West Africa, most of Central and Eastern Africa and part of Southern Africa. Bantu languages make up nearly half of this group with 350 million speakers and include Shona, Swahili and Zulu, amongst others.

Nilo-Saharan – This language family belongs to Nilotic people who speak Nilotic languages which are indigenous to the Nile valley and Central Sahara, including South Sudan, Uganda, Kenya and North Tanzania. Groups include: Luo, Sara, Maasai, Kalenjin, Dinka, Nuer, Shiluk, Ateker, and Maa-speaking peoples.

Khoe-San - Languages of the San and KhoiKhoi populations of Southern Africa and are non Bantu-speaking. The San are mainly in Botswana and Namibia; Admixture of various origins including indigenous Africans, European ancestry, Polynesians and Malagasy has resulted in a mosaic landscape.

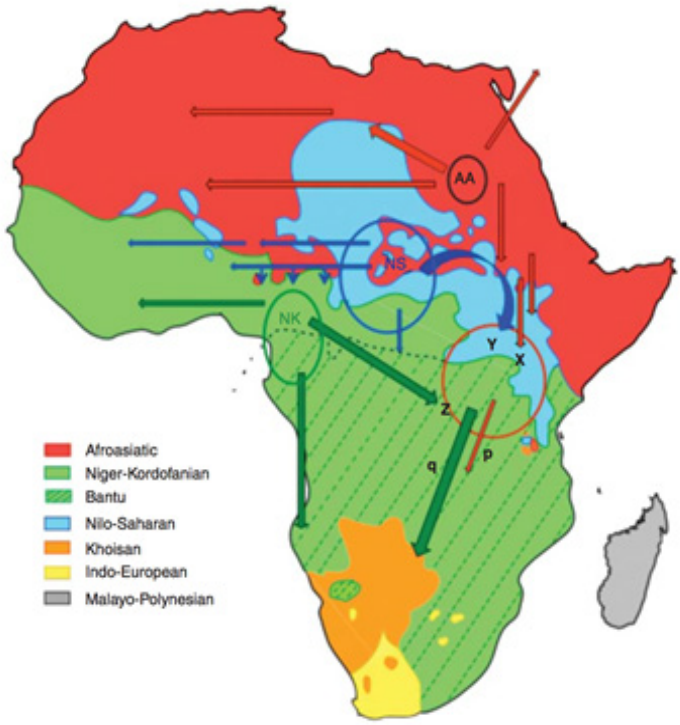


Figure 2. Linguistic groups in Africa. Source: Gomez et al. 2014

events, and random mutations that accumulate over generations, further affected by the positive selection of beneficial variants and loss of harmful ones. These selection patterns were dominated by the effects of environment and lifestyle, including diet (Pennisi 2007), possibly altitude, radiation levels, and exposure to pathogens (Vasseur and Quintana-Murci 2013). Adaptation to environmental conditions or biotic stress could also co-occur with susceptibility to other chronic conditions (Tekola-Ayele et al. 2015; Baker et al. 2017), for example a haemoglobin B variant confers susceptibility to sickle cell anaemia but resistance to malaria (Elguero et al. 2015), an Apolipoprotein variant confers trypanosomiasis protection but increases kidney disease susceptibility (Genovese et al. 2010), and G6PD deficiency syndrome simultaneously protects against malaria (Mbanefo et al. 2017). The short haploblock architecture and genetic variation do offer an opportunity for fine mapping of aetiological variants, and the increasing availability of African-specific genotyping tools and next generation sequencing services on the Continent will harness this advantage (Rotimi et al. 2016, 2017; Chen et al. 2020).

1.3. What is Precision Medicine?

Precision Medicine (PM) can be described as an approach for disease diagnosis and treatment that takes into account the variation in genes, environment, and lifestyle between individuals. PM aims to identify which treatment will be the safest and most effective, for each patient (The Precision Medicine Initiative Working Group (NIH) 2015; National Institutes of Health, USA 2019). PM approaches have the potential to minimise harmful side effects from therapeutics whilst simultaneously maximising their therapeutic effectiveness to ensure a more successful outcome in each patient.

Some key objectives of PM are:

1. To improve sensitivity and specificity of diagnostic and prognostic processes, including preventative medicine, early enough to mitigate a disease's impact on the patient.
2. To implement the most appropriate and effective therapeutic approach to treatment for each patient through a better understanding of their genetic, clinical and environmental risk factors; and
3. To administer the correct dosage of the appropriate drug in order to optimise the chance of successful treatment whilst minimising side effects and adverse events.

PM is enabled by new technologies and scientific strategies that require the integration of medicine, technology and computation with ethical, regulatory, economic, and social factors.

Photo credit: Pinterest

The implementation of PM is data intensive, and requires acquiring complex data about individuals or populations that include extensive individual-level longitudinal information about the transitions between the healthy and disease state as well as their genomic and socioeconomic/demographic information (Galas and Hood 2009). PM objectives can be achieved in part at a population level through stratification of patient populations according to their risk and treatment response profiles, and ensuring that individuals immediately receive the best possible treatment for their specific stratum (Prosperi et al. 2018).

PM has thus brought about a paradigm shift in clinical care: while the individual patient is still the central focus for determining the appropriate disease treatment and management, this aims to be more precise. While conventional clinical assessment, laboratory screening tests and medical imaging form a strong core of diagnostics, additional molecular and genetic data are rapidly adding another layer of parameters to determine disease susceptibility, diagnosis, prognosis and most likely therapeutic response in the individual.

Some PM related terms warrant further clarification (Barker 2017), including: *Personalised Medicine*, which may be interpreted to include tailor-made treatments or preventative medicine developed uniquely for an individual; *Genomic Medicine* – referring to the impact of the entirety of a person's genetic material on their health; *Precision Public Health* – which applies population-level genomic, clinical and health data to deliver the correct intervention to the appropriate population, at a global level; and *Predictive, Preventive, Personalised and Participatory (P4)* medicine - which entails implementing PM in a more predictive, personalised, preventive and participatory, and less reactive way (Galas and Hood 2009).

1.4. Why is precision medicine needed in Africa?

Given the high dual burden of disease in Africa, African health systems are often overwhelmed with addressing infectious diseases and managing treatment of NCDs, applying conventional clinical approaches that examine multiple factors such as epidemiology, medical histories, biomarkers, demographics, exposures, known susceptibilities and other measurements. Not all individuals or populations are the same, however, and we need to increase the accuracy of identifying, and tracking specific traits in subpopulations in order to achieve greater precision in responding to health challenges (Desmond-Hellmann 2016). Increasing the accuracy of diagnosis and treatment whilst simultaneously reducing ineffective treatments, trial-and-error treatment switching and adverse drug reactions is a virtuous circle that can

Photo credit: Shutterstock

improve patients' health outcomes at the same time as reducing resource requirements and costs to the health system (Barker 2016, 2017). PM and precision public health are changing the way clinicians approach disease diagnosis and treatment on other continents, but methods and data applied to Northern Hemisphere populations cannot necessarily be extrapolated to African populations, and local disparities in clinical research investment to date have resulted in a historical evidence base for modern medicine primarily garnered from Caucasian populations (Patra 2018). It is essential for Africa to build an evidence base for clinical practice that addresses the specific needs and profiles of African patients, and PM can provide a fast-track route to tailor clinical care and treatment regimens to best improve health outcomes in Africa. In addition, knowledge of global population health and genetics is essential in the interpretation and validation of data being used in health care across the world, increasing our need for knowledge on currently under-studied populations.

PM research in other countries, such as the 100k Genomes project in the UK (Genomics England) or the "All of Us" project in the USA (All of Us Research Programme Investigators et al. 2019), have illustrated the challenges to overcome in order to build a framework and the necessary infrastructure for PM, keeping in mind that their starting point was better resourced than in developing countries where infrastructure is lacking and national health systems are generally over-burdened and under funded. Some of the global challenges include building a stable and secure data sharing infrastructure with appropriate ethics compliance and regulation; integration of technologies and data with clinical practice; evidence for actionability; affordable pricing of PM healthcare interventions; and community engagement. Many African countries face additional barriers and challenges to overcome. Nevertheless, back in 2009, Galas and Hood suggested that in time P4 medicine will be possible in the developing world and will "become the foundation of global medicine" (Galas and Hood 2009). This could become a reality if the challenges can be overcome through the development of a practical framework that brings together all stakeholders, including scientists, researchers, clinicians, funders, patient groups, patient relatives, OMICs technologists and pharmaceutical industries to collaborate around a common goal of PM implementation. In addition, African PM can build on the global achievements to date, and leapfrog some of the research and development that has already been undertaken elsewhere. Implementation can be achieved in a stepwise manner with pilot projects in better resourced African countries to assess the readiness of healthcare systems. Infrastructure can be developed over time starting with priority areas that may be more easily achievable, such as pharmacogenomics.

Figure 3 provides diagrams of some of the key elements required for the framework, most of which are discussed in more detail in the chapters that follow. As mentioned previously, we have narrowed the scope here to focus on genomic medicine (GM), which is one aspect of PM but still aims to provide more precise, personalised healthcare. Sometimes GM and personalised medicine are used interchangeably in the document, as some of the aspects apply more generally than just to genomic medicine.

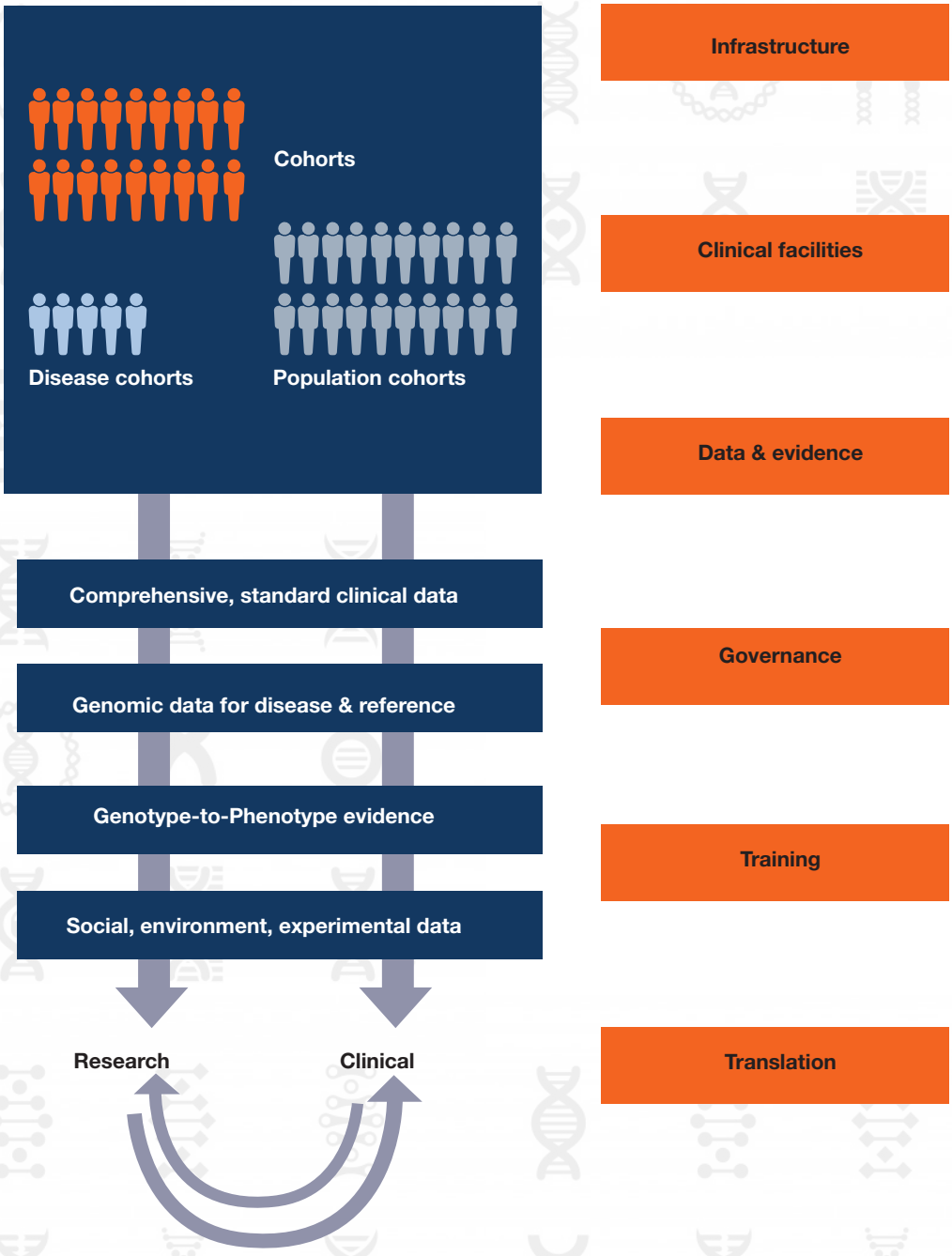


Figure 3. Some of the elements required for a genomic medicine framework.

Section 2. Infrastructure Requirements for Genomic Medicine

2.1. Background

The infrastructure requirements for the development and implementation of a successful national genomic medicine initiative can be extensive, particularly if implementing genome sequencing, but in many cases do not need to be developed from scratch, and some simpler approaches can be adopted. We consider five types of required infrastructures:

- **Clinical infrastructures** are required to enrol patients, collect and analyse biosamples, collect and preserve patient demographics and histories, offer personalised treatments, and monitor clinical outcomes. The required infrastructures at the point of care are relatively simple. An ongoing personalised medicine programme is very similar to a Phase 4 clinical trial, with an emphasis on pharmacovigilance, monitoring of treatment efficacy, and refinement of the criteria used to tailor treatment to the patients' genetic and lifestyle backgrounds. A newly created programme, when the efficacy of the personalised treatments is not yet firmly established, will have to go through a process similar to Phase 2 and 3 clinical trials.
- **Biorepositories** are key resources for any GM or personalised medicine programme, as they allow the retrieval and potential re-analysis of patient samples. Acquisition and storage of rich metadata about the patients and the samples are needed to allow retrieval and re-analysis for targeted studies. Sample preparation, storage conditions and security should be given careful consideration. In many African countries, as well as through international collaborations such as H3Africa, biorepositories for African patient samples have already been created. These should be leveraged as much as possible, to augment the total samples available for analysis, as well as to share infrastructures.
- **Data generation facilities** can deliver results on a small or large scale. **High-throughput technologies** such as sequencing and genotyping are used to identify genetic markers that are predictive of patient susceptibility to disease, disease outcomes, and optimal treatments, and are foundational to genomic medicine. The instruments needed to implement these technologies are expensive and require both technical expertise and a sophisticated environment to operate. If such an infrastructure already exists in-country it should be leveraged. If not, subcontracting to external organisations is often the safest and least onerous initial solution. Some GM interventions can be implemented on a smaller scale with point-of-care tests, PCR, targeted sequencing or targeted genotyping using smaller panels. These facilities and relevant reagents need to be readily available.
- Sophisticated **data and computing infrastructures** are required to acquire, analyse, interrogate and store the data generated by GM programmes. The networks connecting these digital resources have to be configured to allow the efficient, safe and resilient transfer of very large amounts of data. The H3ABioNet network has developed key expertise in this area, which should be leveraged in the design, deployment and testing of computing infrastructures.
- The analysis and interpretation of high-throughput genetic and genomic data requires a complex **software environment**. This includes the use of "containers" that allow deployment on heterogeneous compute resources, workflows implementing complex operations and data flows, and sophisticated databases to store data, metadata, and results. In order to enable all stakeholders to access and use the data, intuitive user interfaces need to be developed. Controlled access supporting strong governance, security, and preservation of the data are paramount.

All these infrastructures are needed to start a successful GM programme, but only the clinical infrastructure needs to be deployed locally. Many basic tests can also be run locally, but for the bigger infrastructure needs (sequencing, computing etc.) at least in the first instance and to the extent possible, these can be shared between multiple countries, and existing efforts and programmes should be leveraged and incorporated.

2.2. Clinical Facilities

A robust and resilient infrastructure adapted to resource-poor environments will be required to collect, process and store clinical samples destined for genomic analysis. A minimal infrastructure for sample collection and initial processing will be able to extract DNA from a variety of sources (blood, buffy coat, saliva, buccal swabs, stools), and to perform simple QC tests prior to sending the DNA to a centralised analytical lab and/or biorepository

Photo credit: AESA 2020

Clinical trials room

for further analysis (Soo et al. 2017). Sample collection and DNA extraction protocols should be standardised for all collection points, ideally using easily handled extraction kits, and technicians should be trained accordingly. After extraction, DNA concentrations should be measured as accurately as possible, using Nanodrop or Qubit technologies. All of these can be implemented in standard clinical laboratory settings. It is advisable to subject aliquots to gel electrophoresis to ensure against degradation, although this can also be done at the central lab or biorepository.

Samples should be stored under ethanol and refrigerated until shipping to the central lab. They should be shipped without undue delay in refrigerated containers if at all possible, although DNA is reasonably stable under ethanol at ambient temperature if refrigerated shipping is not possible. Responsibility for long-term storage and preservation lies with the biorepository.

It is also very important that patient metadata be captured carefully and completely at the point of contact. This is effected most efficiently by providing the primary caretakers with smartphone-based apps that can capture these data and feed them to a centralised database. The increased adoption in Africa of the REDCap environment for managing patient data, coupled with well-documented standards for metadata capture, will make it easy to implement this aspect of clinical data entry.

In building a nationwide GM programmeme, the discovery and implementation of biomarkers of drug and treatment efficacy and safety requires clinical evaluation. Clinical evaluations represent a feasible entry point for capacity building in Africa. Of the four phases of clinical drug development, infrastructure is weakest for Phase 1 studies and most reasonable for Phase 4 studies. In general, Phase 1 studies involve dose finding pharmacokinetics and safety studies in a few healthy volunteers. Phase 2 are proof of concept studies evaluating the efficacy and safety of the treatment intervention in a few patients under controlled disease and clinical settings. Phase 3 studies focus on drug or treatment safety and efficacy and involve thousands of patients in varied disease and clinical settings. Phase 4 clinical studies are product life span evaluations for safety and any observed new indications. The implementation of a PM or GM programme will involve Phase 3 and 4 type studies.

Therefore, in addition to basic clinical facilities, infrastructure for clinical trials is important in Africa if GM approaches are to take Africa's genomic diversity into account during the treatment or drug discovery development and deployment value chain. Whilst Phase 4 clinical studies heavily depend on the existence of strong pharmacovigilance systems, Phase 3 studies depend on the existence of well-equipped and managed hospitals. Through numerous clinical

studies funded by the European & Developing Countries Clinical Trials Partnership (EDCTP), National Institutes of Health (NIH), Wellcome Trust and other development partners, many African countries have developed decent capacity for Phase 3 and 4 studies.

Africa however is seriously lagging with respect to infrastructures for Phase 1 and Phase 2 clinical studies and is thus limited in its ability to conduct clinical pharmacogenetics studies that could be required during these phases. Recent efforts at building capacity for Phase I clinical studies has been noted (Gutierrez et al. 2017), where such facilities have been established in Ghana, South Africa, Kenya, Zimbabwe, and Ethiopia. Driven by Phase 4 observations, some phase 1 clinical pharmacogenetic studies have now been conducted at some of these Phase I clinical trial units (Soko et al. 2016).

Phase I and 2 clinical trial sites are highly specialised facilities that are composed of dedicated health volunteers and/or patient beds, highly specialised monitoring gadgets, laboratory support for the screening of patients/volunteers for biochemical and genetic variables, data capture and analysis systems, research pharmacies, and access to an intensive care unit (ICU). The facilities are run by highly qualified clinical trial teams and all procedures and processes run according to standard operating procedures. Establishing and maintaining such Phase 1 and 2 facilities that meet international standards is very expensive. To be sustainable, such facilities need a healthy pipeline of clinical studies performed each year. To do this, discussions between these units and the international pharmaceutical industry need to be done such that these units can conduct some of the clinical studies when knowledge of pharmacogenomics in African populations presents both a regulatory, scientific, logistical and commercial advantage in the development of their products.

2.3. Biorepositories

A biorepository is defined as a storage facility for well characterised biological material - this can include clinical patient samples, population-based samples, non-human animal samples, environmental samples, plant banks (including biodiversity banks and agricultural banks) and nucleic acids with their associated data. In the context of human research, different biorepository collections exist. A banking model will collect specified material (for example, plasma) and associated clinical datasets from participants. Participants may be recruited in the context of a clinical trial, in a hospital care programme or in a large epidemiological or population-based study. Prospective biorepositories collect samples at clinician or researcher-request. In many cases, there are a combination of models which are used. These models have different advantages and disadvantages. Banking models (including where residual sample material is collected and stored) may not focus on specimen utilisation yet strict quality management principles must be present to ensure that specimen integrity is maintained. Prospective sample collection may not allow for a sufficiently large sample to be present when testing is initiated. Establishment and maintenance of biorepository infrastructure is expensive and the biorepository must adhere to strict guidelines to ensure that appropriate informed consent is granted by donors for sample and data re-use, and quality is maintained (Grizzle et al. 2015)an important aspect of the provision of tissues by biorepositories is the assurance of high quality and consistency of processing specimens. This is best accomplished by a quality management system (QMS. Despite these potential disadvantages, biorepositories form a critical resource to the success of a GM programme.

A key requirement for GM is the presence of high-quality, well-annotated clinical material with associated data. High quality samples are particularly important for sensitive techniques like sequencing where nucleic acid extraction and amplification methodologies can materially influence the final product. Biorepositories implement an accredited process to ensure that sample integrity is optimised. Standardisation and continuing quality improvement also facilitate improved interrogation of the sample stores and minimise disconnects between the sample and associated phenotypic or clinical data. In accredited repositories, mandatory quality control is performed on selected samples. Sample integrity is key to diagnostic accuracy and precision in cases where a diagnostic assay is performed and also allows for scientific rigour and reproducibility. Where these things are in place, the biorepository is invaluable for longitudinal research as well as cross-sectional studies, biomarker discovery in large well-defined clinical cohorts and for rare diseases which can form the nucleus of a successful molecular diagnostic programme.

Key requirements of a biorepository infrastructural resource should be considered in any GM programme. This includes identification of a physical dedicated space with adequate disaster management strategies. Sample processing equipment acquisition should be driven by the sample types the repository manages. Basic molecular biology equipment (like thermocycler,

centrifuge, biosafety cabinet) should be part of a biorepository. All biorepositories should adhere to standard operating protocols and perform internal and external quality control. Ideally, where possible, there should also be a formal accreditation process which oversees the total quality management system. Since the annotation of specimens is key, all biorepositories should operate a robust laboratory information management system which possesses the capacity to interface with other systems.

A number of ethical, legal and societal issues arise in the context of long-term sample storage. This includes the need for appropriate informed consent while adequately protecting participant confidentiality (see section on ELSI and Data Governance). Where industry and international collaborative partnerships are considered for innovation, these should be specified at the outset. Finally, all biorepositories should operate within the regulatory and legal framework of the country in which they are located.

Although some of these requirements may appear to be onerous, the benefits which accrue to a high quality, community and researcher supported repository are significant for researchers who are attempting to leapfrog conventional diagnostic systems, particularly in low-resource settings. Apart from the curation and provision of high-quality samples, biorepositories also have the potential to form a core for other services and a site for training and capacity building: a valuable resource for GM for public health.

2.4. Sequencing and genotyping

We will discuss here mainly the Generation of data that support clinical diagnosis and care, which also largely apply for generating reference data and variant discovery. There are other omics technologies used in PM but here we are focusing on genomics and more specifically high-throughput genomics data generation. The requirements for data generation will be driven by the scope of the chosen personalised medicine programme (Table 1). The programme could rely on the measurement of population-specific pharmacogenomic markers for the determination of appropriate drug dosages and combinations, as is the case in several programmes in South-East Asia. It could aim to measure allelic variants at a number of loci known to be associated with increased risk for specific diseases, and thus warrant additional surveillance for at-risk individuals. It could be directed at discovering causative variants in families with rare genetic diseases or detecting specific genotypes known to cause monogenic diseases. Alternatively, it could measure tumour-specific somatic mutations in known cancer-causing genes to direct optimal courses of treatment. Each of these options, and many others that could be chosen, will require different approaches to data collection and analysis, and different infrastructures (Table

Photo credit: Shutterstock

Genome sequencing

Table 1. Choice of genomic technologies and applications

Type of analysis	Clinical benefit	Technology required	Computational infrastructure required
Pharmacogenomic markers (SNV)	Choosing best drug, optimizing drug dosage, avoiding side effects and adverse events	Genome-wide or targeted genotyping using SNP arrays; design of population-specific SNP arrays	*Storage*: hundreds of GB. *Memory*: tens of GB *Computer class*: high-end workstation or small cluster.
Risk alleles for common diseases and cancer	Improving surveillance and preventive care, recommending lifestyle adjustments	Genome-wide or targeted genotyping using SNP arrays; targeted sequencing of known risk genes	*Storage*: hundreds of GB. *Memory*: tens of GB *Computer class*: high-end workstation or small cluster.
Finding causative variants for and molecular diagnosis of rare diseases	Suggesting possible courses of treatment for diseases of unknown etiology	Whole genome or exome sequencing	*Storage*: in the PB range to keep raw data, hundreds of GB for the results *Memory*: hundreds of GB *Computer class*: HPC cluster / cloud
Finding somatic mutations driving cancer progression	Targeting cancer therapy to specific gene products or pathways	Amplification and targeted sequencing of cancer gene panel from tumor samples	*Storage*: hundreds of GB *Memory*: tens of GB *Computer class*: standard workstation
Discovery of novel population-specific risk alleles	Developing better tools for prevention and monitoring, adapted to local ethnic groups	Whole genome sequencing or SNP genotyping of controls and affected cohorts, GWAS	*Storage*: hundreds of GB (for SNP analysis) *Memory*: hundreds of GB *Computer class*: high-end HPC cluster / cloud
Sequence-based identification of pathogens	Accurate diagnostics, appropriate interventions, epidemiological information	Targeted sequencing of pathogen genes / genomes, for some pathogens, PCR-based diagnostic tests	*Storage*: hundreds of GB *Memory*: tens of GB *Computer class*: standard workstation

1). The programme could, of course, encompass more than one approach.

The required data generation facilities can be divided into two broad categories: those relying on using arrays of probes attached to chips to interrogate variants at a set of predetermined locations in the genome ("SNP chips"), and those using high-throughput sequencing technologies to determine the nucleotide sequence of a subset of the genome, varying from a limited number of critical regions, to all protein-coding exons, to the entire genome. Additional technologies may need to be deployed to explore, e.g., copy number variations or large structural variants, although these are at present of limited clinical value.

From a practical perspective, the minimal infrastructure requirement is a lab equipped for the extraction, quality control, and short-term storage of DNA at each of the locations where patient samples are collected, and a biorepository that can provide long-term preservation of the samples and secure shipping to appropriate service providers. The actual analysis of the DNA by any of the techniques mentioned above requires significant investments, not only in equipment, but also in supporting infrastructures (clean labs, reliable power supplies, high-throughput networking), in setting up reliable delivery of consumables (expedited customs, continued cold storage), and in training a cadre of technicians and engineers to operate and maintain the equipment. There is a non-negligible risk that a sizable investment in sophisticated equipment may not deliver the expected technical capacity. Therefore, we suggest that in some cases, the national genomic medicine programmes could initially rely on carefully chosen external contractors to perform the chosen analytical procedures. Once the programme is well established and successful, a well-planned

investment in instrumentation and supporting services and staff should be considered, keeping in mind that this will require a continued allocation of resources to operate and thrive. It is not a once-off investment and funds should be provisioned for upgrades and ongoing maintenance.

There are several authors of this document who will be available to discuss specifics with the authorities leading national initiatives, to narrow down their requirements, recommend technical solutions and vendors, establish budgets, and help plan the deployment of the necessary instrumentation and services.

2.5. Computing Infrastructure

2.5.1. Computational resources

The storage and analysis of genomic datasets at the scale required for a national genomic medicine initiative requires sizable computational resources, particularly if genome sequencing is implemented. Legitimate concerns about confidentiality and ownership of the data being collected mandate that this computational infrastructure be controlled by the institutions implementing the programme, though in clinical settings, healthcare facilities seldom have access to such infrastructure. The scope of the programme, and decisions on the technologies being deployed will impact the size and complexity of these computational resources: for example, a programme based on extensive sequencing of whole genomes or exomes will require much larger investments than one focusing on variant discovery or diagnostics in a limited number of genes, or on the use of genotyping chips. Planning a suitably sized computational infrastructure should be a priority for all national programmes and should be driven by the chosen technologies underpinning the programme. Currently, computational infrastructures do exist in various African countries such as the ILIFU and CHPC platforms in South Africa, the Uganda Medical Informatics Centre, national HPC centers such as in Kenya and Tanzania, and various university based HPC resources that can be strengthened to cope with national genomics programmes. A key consideration is the development of a cadre of competent bioinformaticians to use the required hardware and software for analysis and production of results.

The use of Cloud based resources should be carefully considered as a potentially attractive option, as they require only modest up-front investments and avoid maintenance and depreciation costs. However, they also come with significant requirements and potential pitfalls: they require a robust, fast, and reliable Internet connectivity; they often have sizable hidden costs (storage, data transfers); and the confidentiality of the data is difficult to guarantee as they need to reside at least temporarily on servers outside the control of the national programme, which in some African countries is not legally permitted. The choice between a fully owned infrastructure and the use of cloud resources needs to take into consideration all of these factors, as well as a detailed financial analysis covering the initiation phase, and both the short-and long-term expected maintenance and expansion costs. Consulting similar national local experiences may provide a helpful template to build upon.

2.5.2. Data Transfers

The successful implementation of a genomic medicine programme requires that a data infrastructure is put into place that can securely move large amounts of encrypted data (TB to PB) across computer networks and architectures. Specifically, data need to be moved between the laboratories where they are being produced, whether in-country or internationally, the databases from which they can be interrogated and distributed to relevant healthcare and research partners, the computer centers or cloud environments where they can be analysed, and the long-term repositories where they will be archived.

It should be stressed that “a good Internet connection” is not enough to fulfill the data transfer requirements. At a minimum, the following elements should be considered:

- If data-generating instruments, data storage facilities and/or data analysis infrastructures are co-localised in the same building or campus, their connectivity can be greatly improved by implementing an internal research network (“DMZ”) that bypasses the routers and firewalls used for general communications (Web browsing, media streaming, email, etc).
- If at all possible, national genomic medicine programmes should work with national research networks (if such an organisation exists) to provide high-throughput paths between its component organisations, as well as direct links to international research networks such as WACREN and UbuntuNet.
- The H3ABioNet network has developed tools and best practices to test and implement data transfer protocols between African countries, as well as with European and North American organisations. These tools could be used to identify potential bottlenecks and to test the infrastructures.
- The software used to transfer data has a significant impact on the speed, error rate,

reliability, security, and fault tolerance of the transfer. H3ABioNet has performed extensive tests and evaluations of several software solutions that include encryption of the data and will be able to make informed recommendations to national initiatives.

2.5.3. Storage/databases

Data generated by GM programmes need to be stored and preserved, ideally according to FAIR principles: **Findable**, in that it should be easy for all stakeholders to identify them unambiguously; **Accessible**, in that they should be in a repository trusted by all parties, and that their structure should be easily understandable; **Interoperable**, in that the metadata should use widely accepted standards and structures, and should be interpretable by commonly available software; **Reusable**, in that the rules for accessing and using the data should be clearly stated, and allow their aggregation with other similar datasets.

From the generation of the data in laboratory settings to their final archiving in international databases, they are likely to be stored in several locations and formats, and to undergo analyses and transformations. During this process, raw data will be transformed into multiple types of information useful to multiple stakeholders: **physicians**, ideally as part of a patient's (electronic) health record; **public health officials**, to extract epidemiological, demographic and statistical information; **bioinformaticians**, to perform detailed analyses; and potentially by many others.

From an infrastructure point of view, a robust storage environment, sized to accommodate not only the raw data but also all of their derivatives, will need to be put in place. It should include redundant data structures (typically a RAID variant), frequent safety backups to allow recovery in case of a system crash, as well as archival backups to maintain access to data no longer in production. Additionally, long-term archival deposition in an appropriate, internationally recognised public database should be planned from the start if consents allow for this. A robust and secure user authentication system, allowing multiple levels of access, should be included to support proper governance of the data. The movement of data, their versioning and association with updated metadata should be tracked with a data management system such as iRODS; failure to do so may result in loss of data, or incertitude about their provenance and identity.

Operationally, the data and metadata collected by the programme should be stored in a structured database that can be easily interrogated using a query language, with an appropriate user-friendly interface. Attention should be paid to the underlying database structure, as classical SQL-based database management systems may not scale gracefully to the size and complexity of the genomic data being acquired. Current thinking favors "noSQL" or non-relational database systems that are not constrained by the classical tabular representation of the data, and often permit their more efficient interrogation.

2.6. Data Analysis and Interpretation

2.6.1. Workflows

The analysis of genomic data typically requires complex, data-intensive workflows that perform multiple interdependent operations on one or more data streams. Computational platforms used for the analyses, whether local or cloud-based, need to be configured in such a way that they can support these complex workflows. In recent years, modular workflows have been commonly implemented using abstracted and portable "containers" such as Docker or Singularity, and workflow languages such as Nextflow or

Photo credit: Shutterstock

Data transfer

Photo credit: GettyImages

Data analysis

the Common Workflow Language (CWL). The H3ABioNet network has developed a number of analysis workflows using these components, which can easily be adapted to support national programmes. We would recommend using the existing expertise in H3ABioNet and collaborating institutions to define and develop the specific workflows required by a national programme, and to install and test them on the programme's chosen computational infrastructure. However, each programme will need to train local computational scientists to run and maintain the workflows as data are coming in and being analysed, and to communicate with the physicians and public health specialists who need to interpret the results (see "2.6.3 Interpretation", below). Workflows used to process genomic data should be well documented and tested. This should include information on the software versions, parameters and reference data used, including their versions.

2.6.2. Reference data for imputation

When working with genomic data, it is important to have access to relevant and up-to-date reference data. For genotyping arrays for research, it is customary to impute additional variants from a reference panel derived from sequence data from relevant populations. For African populations there are several choices, including the Sanger African reference panel, the Michigan reference panels and the H3Africa reference panel, all of which are run as imputation services. Ideally a facility should have an in-house reference panel as it can be used whenever needed, no data need to be transferred/submitted and there is control over the data used to generate the panel. However, large samples sizes have been shown to provide more accurate imputation and most facilities will not have access to a large number of samples with whole genome sequence data, at least in the beginning. For arrays used purely as diagnostic tools imputation would be less relevant as the diagnostic SNPs would be on the array. However, reference data demonstrating the "healthy state" is important for comparison with data from patients.

For sequence data a reference genome is required for alignment of reads and variant calling. There is a standard human reference agreed on by the Genome Reference Consortium (https://www.ncbi.nlm.nih.gov/grc). It is important to document which reference was used. Since this reference genome is biased towards non-African populations and African genomes harbour extensive diversity, it may be worth considering using population-specific reference graphs for alignment and variant calling in the future.

2.6.3. Interpretation - clinical decision support

The results emerging from genomic analyses document the genetic background on which a disease may develop or has developed. In most cases they do not pinpoint the causes of the disease nor quantify the risk of developing the disease. To achieve this, it is necessary to interpret the results of the analyses in the light of other information about the patient: personal and family history, ethnicity, lab results, disease symptoms, and lifestyle, as well as documented effects of similar genomic variants from public databases. It is also important to be able to access the most up to date information on existing knowledge, for example actionable variants of relevance to the population. This data may come from internal resources or public databases. This process has to rely on the combined expertise of physicians, geneticists, epidemiologists, genetic counsellors, and bioinformaticians.

While there are some software environments that provide clinical decision support based on genomic data from patients, their utility at present is limited to recommendations of drug

dosage or choice (pharmacogenomics), or to specific diseases with well-documented genomic correlates. They are also most efficient when linked to a comprehensive electronic health record (EHR) system, which is often not available in resource-limited environments. Therefore, our recommendation would be for a panel of physicians and other domain experts to establish a set of best practices and recommendations, appropriate for the scope of each national programme, for following up on genomic findings. These can be partially automated by flagging features in the data that warrant attention. There are of course different considerations for monogenic versus complex diseases. For the former, a single or small panel of variants can be screened for and clinical decisions made based on the result, while for complex traits, wider screening is required in terms of variants, and one approach is to develop polygenic risk scores based on data from the same or related populations. Ultimately, clinical panels are required that are both disease specific and population appropriate. Some national GM initiatives have developed web interfaces, and reporting templates for clinical data interpretation, for example, Genomics England provide reporting templates and access to their PanelApp (https://panelapp.genomicsengland.co.uk/), which provides gene panels for different diseases. To ensure accurate interpretations, it is also of paramount importance to keep an updated set of databases such as ClinVar that document known effects of genomic variants, and to augment them with data accrued from local patient populations.

2.6.4. Curation

The accuracy and consistency of data collected in a genomic medicine programme are essential to its success, both in delivering quality healthcare and in advancing knowledge about the patient populations. Therefore, it is important that curation activities, designed to test and correct any errors in the data, be deployed throughout the programme. In addition, the knowledge gained by the international scientific community as well as by the national programme should be used to update the metadata collected previously, and to bring all the data used in clinical decision-making to the same standard. Curation requires qualified biomedical data specialists (who will need to be trained) as well as automated sanity checks on the data. It is important not to underestimate the importance of and the resources required for data cleaning and curation. As mentioned above, the curated metadata should be stored and made available and searchable by healthcare professionals and others accessing the GM "facility".

Key Recommendations for infrastructure requirements

- A national data capture infrastructure (REDCap or similar) is essential for entering and managing research or clinical data about patients. This can complement EHRs or other clinical records used in standard healthcare.
- Existing frameworks for conducting clinical trials, especially phases 3 or 4, will be extremely useful and should be leveraged.
- All participating points of care should be equipped to extract DNA and do initial quality checks (QC). Simple genetic tests should be available at these facilities.
- The programme should have access to a well-run biorepository that has the capacity to store, distribute and manage the samples being collected.
- While access to high-throughput genotyping and sequencing facilities is important, developing them locally may not be a priority at first. However, as more resources become available, we should aim to shift this capability to the Continent.
- The security of both samples and data should be a high priority and should be put in place before the programme starts.
- Global and country-specific data sharing and data storage laws and guidelines should be adhered to.
- Computer networks and Internet should have the capacity to move large datasets reliably between data generation, analysis, and long-term storage locations.
- A programme that uses high-throughput technologies will require robust computational resources. To the extent possible, existing resources should be leveraged, but a dedicated secure facility may be required.
- Accreditation systems for clinical laboratory systems to conduct genotyping and sequencing should be acquired.

Section 3. Cohorts and Diseases

3.1. Summary

This section aims to provide guidance on potential disease focus areas and identify some of the challenges in working with populations of high genetic diversity. It highlights the need for reference data from healthy individuals and describes some diseases of high prevalence which are relevant for translation of genomics into clinical care. Some of the most prevalent communicable and non-communicable diseases in Africa are described, providing insight into their underlying genetic factors and the research required to address future applications in medical care. Recent developments in the genetics of CVD and cancers as well as achievements in human genome and pathogen sequencing of African populations are highlighted. For the diseases described there are numerous research publications, but the aim is not to review the literature, only to provide some examples of where genetics or genomics could inform healthcare. Countries should focus on the diseases most relevant to their setting. The apparent gaps in existing data are discussed with recommendations that could enable the implementation of genomics medicine.

3.2. Reference Cohorts

The genetics of some African populations have been studied and revealed a high level of inter-population diversity. Many genetic studies have been carried out within specific genomic regions while more recent investigations have introduced whole genome sequencing and genotyping arrays done for selected populations or individuals with specific phenotypes such as predisposition to certain diseases. Some of the recent efforts at whole genome sequencing for populations in Africa have provided insights into the origins of disease predisposition (Rotimi et al. 2017). Deep analysis of genomic data and variant forms of disease-related genes offer new avenues for GM applications, for example, replacing defective proteins with functional equivalents can potentially manage disease as is routinely done for blood clotting disorders (hemophilia) with replacement by recombinant factors, but this requires distinguishing between pathogenic variants in patients and non-pathogenic variations occurring in healthy individuals.

To date, genome sequencing has been completed for individuals from only a limited number of African populations. Additionally, many of the sequence datasets have restricted access or are private. DNA sequencing and genome-wide association studies (GWAS) have demonstrated that existing data from non-African populations cannot be directly extrapolated to the more genetically diverse African populations, highlighting the need to generate additional data for healthy reference individuals representing the diversity across geographical regions of the Continent. Any national GM programme must create reference cohorts for local populations, whilst recognising multiple population groups as well as admixture with other populations. Reference population data are needed to understand the healthy state and what genetic diversity is “normal” (non-pathogenic), and to use as controls in future GWAS. The data can also be used in reference panels which are required for imputation for genotyping array data. For this, reasonable samples sizes for each population are required.

3.3. Setting up a Structured Cohort Database

In addition to reference datasets, cohorts should be established for surveillance and for specific diseases. Ideally, data collection should include longitudinal studies, which gather data on selected observations over long periods of time (e.g. from infancy to adulthood) to decipher or interpret changes taking place in an individual that will help to provide an appropriate healthcare service. A cohort study looks at a group of individuals over a time period to determine possible association with some outcome. Members of a cohort or group are selected for a particular feature or risk factor to which they are exposed. If they develop the disease, then the link between the feature/marker and the disease can be established. Such studies can show the relationship between risk factors and health status. Alternatively, population cross-sectional cohorts can be followed, where participants are not selected for any specific trait. Here, demographic and health surveys could be a useful staring point.

Most biomedical research in Africa is currently being carried out on a regional level, where individual research groups are focusing on specific populations and diseases. However, it has been demonstrated that in population genetics and disease discovery studies, large samples sizes are needed and comparison across different populations can be very informative. If a national GM initiative wishes to make use of existing studies, then there is a need for a common platform whereby findings generated in different studies are comparable. Data should be harmonised, and sample collections and analytical tests done across regions or studies should be

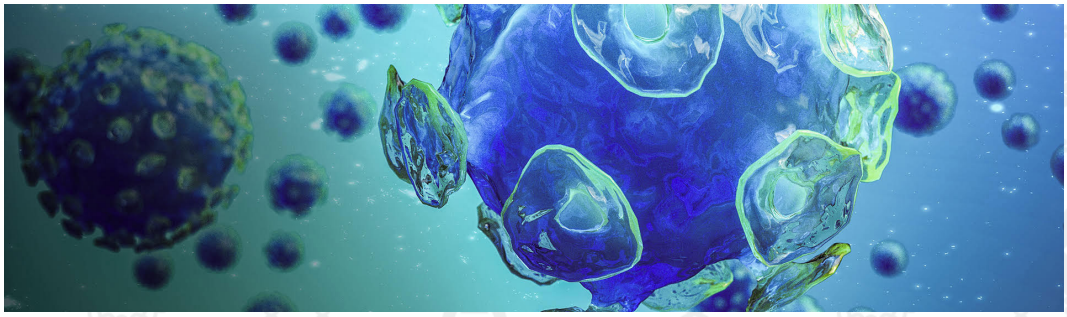


Photo credit Roche.com

Disease illustration

consistent. Ideally, national cohorts should be established with clear guidelines for participation and a policy for building a database for large-scale cohorts. This should gather physical, chemical, serological and genetic information, and collect biological samples for studies and long-term access and sharing. Coordinated implementation of regional projects will ensure that all procedures are done using a common approach that will allow easy data sharing and use. The process should include online registration, surveys for collecting data in a standard form and simultaneous collection of biological samples (blood, serum, tissue fluids, tissue biopsies) at health centres following a medical check.

The surveys should have questionnaires for information on:

- ***Lifestyle:*** Employment details, e.g. occupational hazards, family environment, e.g. number of people living in the same household, recreational activities, physical activities, diet, smoking habits, substance use/abuse (e.g. alcohol) and exposure to traumatic events
- ***Health:*** General health conditions including mental conditions, history of medical treatments, exposure to infectious diseases, inherited diseases, allergies, family history of NCDs
- ***Socio-demographic data:*** Ethno-linguistic group, economic status /professional activity, ancestry information i.e. ethnic group of parents/grand-parents
- At health centres, medical checks should include assessment of ***Anthropometric phenotypes*** such as measurements of different body parts e.g. height, weight, waist size, body fat etc., biochemical profiles to assess functional capacity of organs/systems -hemoglobin, white blood cells count, creatine, microbial detection by application of diagnostics or clinical examination.

All the data should be collected, processed, and stored within a secured, communal system with restricted access (see section 6 – ELSI and Data Governance).

3.4. Diseases in Africa and Relevance to GM

Africa is overwhelmed with disease burdens, both communicable and non-communicable (Amegah 2018) (see Box 3: *Example of disease burdens in Africa*). Many infectious diseases have been controlled through immunisation, which has led to the elimination of several pathogens on the continent. However, poor social and economic conditions are worsening disease situations in low income countries with the re-emergence of pathogens in places where they had been eradicated. These, as well as diet and other lifestyle changes, have also given rise to an increase in non-communicable diseases. Not all diseases can be addressed using GM approaches, and the genetic associations for some are more complex than others. These are discussed below under the two categories of diseases (NCD and CD).

3.4.1. Non-Communicable Diseases (NCDs)

NCDs include a diverse range of diseases and disorders, some of which are monogenic, where the direct causal mutation is known, while others are multifactorial. The monogenic diseases lend themselves to a GM approach as genetic testing can be used for diagnosis or to guide treatment. The same is true in pharmacogenomics and some cancers, where causative variants are known and can be tested for using single tests or panels of SNPs on arrays. However, complex diseases are under the influence of hundreds of loci which often do not individually explain how they can contribute to the pathological manifestations. For complex traits, polygenic risk scores, which sum up the smaller effects of multiple variants associated with a trait, can be calculated (Dudbridge 2013) and used to determine a person's disease risk, however these are increasingly being shown to be population-specific, so it essential that these are determined based on data from the relevant population (see Box 4: *Calculating polygenic risk scores*). A more large-scale genome-wide approach is thus required for the original discovery of associated variants for complex traits, but also for rare diseases. This enables novel discovery

Box 4. Calculating Polygenic Risk Scores (PRS)

Example of disease burdens in Africa

Deaths due to NCDs are increasing at an estimated rate of 27% in Africa (2014), though more recent figures show a more dramatic increase. Lifestyle and genetic factors are likely to be the main causes. (Amegah 2018) reports that CVDs account for 11.3% of deaths in Africa and have been increasing at an alarming rate. This trend seems to be associated with urbanisation and the transition from traditional, healthier eating to high-salt, high-lipid diets, which become more popular with economic development and Westernization of lifestyles. The same high-caloric risk factors are linked to diabetes and some predict a 110% increase in diabetes in Africa between 2013 and 2035 (Hunter-Adams et al., 2017). Social and economic conditions exacerbate the risks of poor urban dwellers for developing NCDs.

The prevalence of diabetes in Africa is as important as elsewhere in the world. Several studies have pointed to the alarming figures of increasing cases of diabetes with an estimated 34 m patients by 2040 (Pastakia et al., 2017). This increase is due to change in lifestyle, as confirmed by the difference in prevalence between rural and urban areas, with the latter being 2-4 times higher; and to some extent due to better ascertainment through screening. It has also been reported that diabetes patients are three times more likely to contract tuberculosis (Jeon and Murray, 2008; Young et al., 2009). This has been attributed to impaired immune functions in diabetics who show lowered numbers of leucocytes and T-helper1 cytokine production, the latter being essential for microbial defense. Macrophage functions are reduced in diabetics, leading to an incapacity to phagocytose and kill the bacteria. The association of diabetes and TB inevitably makes it far more difficult to manage both diseases, each aggravating the effects of the other.

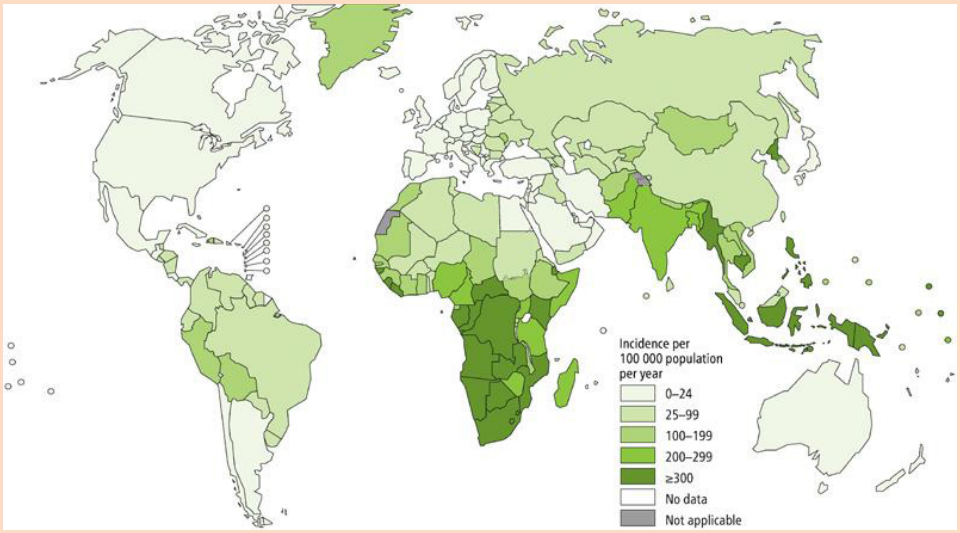


Fig 4. TB incidence rates in 2018. Source: WHO Global Tuberculosis Report (2018)

For infectious diseases, malaria, tuberculosis, HIV and respiratory infections are the biggest threats. Africa has the second highest incidence of tuberculosis (TB) in the world, after South-East Asia. The most affected countries are Nigeria, South Africa, DR Congo and Mozambique, in this order (Figure 4). Similarly, HIV incidence remains high in Africa but frequencies at intra national levels are highly diverse (Dwyer-Lindgren et al., 2019). A study from the Institute of Health Metrics and Evaluation, University of Washington, is the first of its kind to provide fine-scale data on HIV prevalence at district and department levels that could be used for planning, monitoring and control of the infection. Southern Africa (Botswana, Lesotho, Mozambique, Namibia, South Africa, Swaziland, Zambia and Zimbabwe) has the highest prevalence of HIV (over 10%) on the continent, though other countries also show high prevalence, e.g. in Kenya, Uganda, Tanzania and Malawi. The Global Burden of Disease (GBD) and UNAIDS studies confirm the country level variation in HIV prevalence. While there are hotspots of the disease, new infections in specific locations form corridors of rapid transmission to other areas

and population stratification for risk but is more challenging to apply in routine healthcare. Nevertheless, whole genome sequencing or large-scale genotyping are now being applied in healthcare in developed countries with many successes, particularly in rare diseases and cancer, so these should be considered for future application in LMICs.

3.4.1.1. Cardiovascular diseases (CVDs)

CVDs include coronary heart disease, heart attacks, stroke and hypertension and tend to be multifactorial. Patients often have several of these conditions simultaneously, and heart attacks or strokes are usually fatal. CVDs are heterogeneous pathological conditions under the control of complex molecular interactions and gene activities. Multiple loci and common variants (at least 5% prevalence) have been identified through GWAS, but most have relatively small effects. As an example, although roles of individual genes in lipid metabolism have been unravelled, it is the cumulative effect of many variants that is more likely to contribute to the development of CVDs. Hence the need for calculation of polygenic risk scores based on combinations of polymorphic sites in various populations (Wünnemann et al. 2019)their clinical management would benefit from improved prevention and prediction tools. Polygenic risk scores (PRS. In this, the risk posed by a combination of variants in one individual can be assessed through the generation of genetic risk scores, calculated on the aggregation of the weighted effects of single variants (see Box 4: *Calculating Polygenic Risk Scores*). Once provisions are made for the relevant tests to become available to clinicians that assess these risk scores, these will change the diagnostics and prognostics of CVDs.

CVD is now considered an epidemic in sub-Saharan Africa with more than 1 million deaths per year. The risk factors are high blood pressure, unhealthy diet (e.g. high salt intake, high lipids), smoking and lack of physical exercise, and are more prevalent in disadvantaged communities. Obesity, diabetes and age are also associated with higher risks of CVD. Effective therapeutic options for high-risk patients include antiplatelet therapy, ADP-receptor antagonists and phosphodiesterase (PDE) inhibitors. Understanding the central role of platelets in CVD has led to the development of drugs that decrease platelet aggregation, which has greatly improved prognosis (Granger D. Neil et al. 2004; von Hundelshausen Philipp and Weber Christian 2007). Several GWA studies have been carried out on various CVDs, some of which are described in Appendix 1, along with the loci identified. Though less work has been done to date on populations of African ancestry, the H3Africa consortium has a number of projects investigating the role of genetics in CVDs. For example, the AWIGen project aims to find the links between genetic and environmental risk factors and cardiometabolic diseases. It is a continent-wide study with participants from Ghana, Burkina Faso, Kenya, and South Africa. Related H3Africa projects have also united in the CHAIR (Cardiovascular H3Africa Innovation Resource) consortium to harmonise their clinical data, and eventually genomic data too, to increase sample size and facilitate novel discoveries in CVDs (Owolabi et al. 2019).

Calculating Polygenic Risk Scores (PRS)

Different approaches have been used to calculate PRS: (Khera et al. 2018) have developed PRS for several diseases in-cluding CAD. A genome-wide polygenic score (GPS) helped identify individuals at higher risks of CAD based on a large GWAS dataset. The UK Biobank data were used to generate GPS scores using two algorithms: 1. LDPred for generating seven candidate GPS through a Bayesian approach of posterior mean effect size for each variant; and 2. An additional 24 candidate GPS were produced through pruning and thresholding scores obtained using PLINK. Strength of association of risk alleles was used to assign weights to the variants. For individuals identified as having greater than 3 fold risk using GPS, most of them would not have been detect-ed using clinically used parameters such as hypertension or hyperlipidemia.

Another approach by (Inouye et al. 2018) compiled morphological (height, weight, waist/hip ratio), medical (blood pressure, cholesterol level, family history of diseas-es) data to define risk factors. GRS scores were calculated based on previous GWAS markers some of which were specific for cardiometabolic loci. MetaGRS were found to be associated with risk factors while > 1.7 m genetic loci were identified that correlated with CAD heritability. This model was further improved by combining 6 conventional risk factors.

3.4.1.2. Diabetes

Diabetes is a global health challenge with increasing economic and social consequences, reflecting the body's inability to regulate blood glucose levels within a physiologically acceptable range. If untreated, diabetes has serious long-term consequences such as cardiovascular disease,

stroke, diabetic nephropathy, foot ulcers and diabetic retinopathy. Complications include blood acidosis, high osmolarity and diabetic coma. Type 1 Diabetes is an auto-immune disease with the production of anti-beta cells antibodies which prevent the synthesis of insulin. Insulin replacement is therefore the best solution. Type 2 Diabetes (T2D) results from resistance to insulin, where although the pancreas is producing insulin, blood glucose levels do not reduce. This form is associated with CVD. Diabetes also causes an increased burden of atherosclerotic plaques, which are unstable and enhance clot formation with reduced ability to break down the clot and increasing risk of heart attacks. Normally blood clots are treated with anti-platelets such aspirin and clopidogrel. However diabetic patients are resistant to these drugs (Schuette et al. 2015) and different treatment regimens have to be applied. Diabetic clots tend to be larger and more prone to causing stokes. Like for CVD, diabetes is increasingly being reported in younger patients.

Several GWA studies to find genetic markers linked to T2D have discovered candidate SNPs and gene variants associated with the disease. Most of these however are common alleles with minor allele frequency (MAF) >0.5. Recent large-scale initiatives, such as the Genetics of Type 2 Diabetes (GoT2D) and Type 2 Diabetes Genetic Exploration by Next-generation sequencing in multi-Ethnic Samples (T2D-GENES) consortia (Fuchsberger et al. 2016), have further refined the discovery of variants associated with Type 2 diabetes. GoT2D focused on individuals of European descent, while T2D-GENES investigated exome data from 5 ethnicities (Europeans, African-Americans, Hispanics, South Asians and East Asians). 2,657 diverse European genomes and 12,940 exomes of other different ancestral groups were sequenced and an additional 111,548 samples were genotyped. Most of the variants found were in the same regions as previous markers obtained by GWAS. More recent studies on African populations have identified novel variants near the AGMO gene and also confirmed diabetes associated variants near the TCF7L2 locus (Chen et al. 2019). Participants were from the Zulu tribe of South Africa and others from Nigeria, Ghana and Kenya as part of the Africa America Diabetes Mellitus (AADM) study. In another recent study, an African-specific variant in the ZRANB3 locus was found to be associated with T2D in individuals from Nigeria, Ghana and Kenya and validated in functional studies on the zebrafish ortholog (Adeyemo et al. 2019). This highlights the importance of discovery research on relevant populations.

It is unclear whether the markers which have been identified at specific gene loci have high enough predictive potential to be used in GM diagnostic assays for early screening, but they could be used together with other information, such as family history with diabetes. Individuals carrying these alleles would benefit from early detection and treatment to avoid or minimise the damaging effects of high blood glucose such as glaucoma and renal complications. As is current practice, a change in lifestyle and diet are additional preventive measures for improving disease outcome.

3.4.1.3. Mental disorders

There is increasing recognition of the importance of viewing mental disorders within the framework of non-communicable disorders. Firstly, mental disorders are often comorbid with other key NCDs, including cardiovascular disorders, diabetes, and respiratory disorders. Secondly, key risk factors for NCDs, such as tobacco use, unhealthy diet, physical inactivity, and harmful alcohol use, often cluster together, perhaps particularly in people with mental disorders. Thirdly, there is increased understanding of the causal mechanisms that underlie the comorbidity of mental disorders with other NCDs; these range from the neurobiological through to the sociocultural; and fourthly, there has been an increase in evidence-based approaches to the treatment of mental disorders in general medical settings, for example, collaborative care approaches.

There is also growing recognition of the potential importance of GM approaches within psychiatry. This reflects significant growth in our understanding of the genetic architecture of mental disorders over the past decade or two. In particular, it has become increasingly apparent that mental disorders involve multiple genes, and that there is considerable overlap in the genetics of a number of these conditions (The Brainstorm Consortium et al. 2018). Concern has been expressed that unless such progress is based on studies of more diverse populations, including African populations, there will be a further increase in global mental health inequalities. While global mental health efforts focus on closing the treatment gap by advocating for more implementation science, there is also a clear need to complement such work with discovery science, including work in neurogenetics.

3.4.1.4. Cancer

Attempts to comprehend the biology of cancer formation and pathology have been strongly motivated by genetic studies. Cancer cells have a highly elevated number of DNA modifications of somatic origins. These variations are both specific and shared among different types of cancers. Some of the alterations in DNA sequences are extensive and complex. The American College of Medical Genetics and Genomics (ACMG)

recommends clinical exome and genome sequencing for 59 genes associated with different types of diseases (Kalia et al. 2017). Information on presence of genetic variants, if detected early enough can significantly improve prognosis.

Conventional classification of tumours is done on the basis of tissue origin, cell types and histologic features. Molecular characteristics now complement the criteria used for more refined classification. The major efforts of The Cancer Genome Atlas (TCGA) have produced lists of genetic markers linked to clusters of tumour types. Unsupervised clustering using information on aneuploidy, mRNA, miRNA and reverse-phase protein arrays (RPPA) produced 25, 25, 15 and 10 groups respectively (Hoadley et al. 2018). Many of these groups included mixed cancer types (i.e. same profile from different cancer cell types). An integrative clustering approach was used to test the extent to which the different platforms overlapped and resulted in 28 iClusters. Some types of cancer can be identified by the mRNA, miRNA and proteins profiles (e.g. gastrointestinal) but were dispersed into diverse aneuploidy clusters. The outcome of this work points to the potential of molecular features for cancer typing that can assist in precision therapy. Further details on genomic studies on cancer are provided in Appendix 1 and well-known cancer mutations are described in Box 5: Germline cancer mutations and recommended panel testing.

With the increasing number of variants for which we know their likely pathogenicity, it is now possible to have early screening based on family history or known predisposition factors. This would dramatically change the way healthcare is provided with a personalised approach. For this it is important to distinguish monogenic cancer predisposition (see Box 5) from complex polygenic and multifactorial causes of sporadic cancers, for which panels will be more challenging.

3.4.1.5. Cancer Micro RNAs in cancer

miRNAs are small RNA molecules (19-23 nucleotides long) that have a role in post-transcriptional control of gene expression. They function by binding to a corresponding mRNA and start a process of its degradation resulting in inhibition of translation into a protein product. The human genome has over 2000 miRNA genes controlling about one third of the genes (Alles et al. 2019), many of which are disease-related. Transcription of miRNA genes produce a primary RNA (pri-miRNA) which have hairpin structures and are polyadenylated, capped and undergo splicing. They are processed by endonuclease Drosha to form pre-miRNA, which are exported to the cytoplasm for further processing by endonuclease Dicer, which removes the loop part and releases the duplex mature miRNA. One strand of the miRNA is then loaded onto the RNA induced silencing complex (RISC). Several examples of non-canonical pathways have been reported that bypass Drosha or Dicer.

Aberrant expression of miRNA disrupts normal gene expression, bringing about uncontrolled cell division and promotes tumorigenesis. miRNAs associated with cancer are mir-21 and mir-17-92 cluster which are elevated, while let-7 and mir-34 families are downregulated (Esquela-Kerscher and Slack 2006). The family of mir-17-92 are particularly important in lymphomas and leukemias. Reduced expression of miR-374b in osteosarcoma cells, favour angiogenesis because of concomitant increase in the amount of VEGF-A, a vascular endothelial growth factor, that is produced (Liao et al. 2016). Similarly, a decrease in miR-543 also increases angiogenesis by elevated expression of angiopoietin 2. miRNAs in blood are carried complexed with proteins or lipids or in lipid vesicles. In this form, they are stable and not accessible by RNases. They remain intact for some time after blood has been collected for diagnosis purposes.

Approaches for using miRNAs for cancer treatment:

1) To reduce the effects of those miRNAs that are overexpressed in tumour cells, modified synthetic nucleic acids that bind to them are introduced. Examples include locked nucleic acids (LAN) and antagomir. These are not degraded by nucleases. (Liu et al. 2018) reported the effectiveness of newly designed DNA nanotubes for delivering DNA that are complementary to overexpressed miRNAs. These were successfully used to target miR-21 and miR-155 which are well characterised oncogenic miRNAs.

2) miRNAs whose expressions are reduced in cancer can be introduced, but require a carrier system to protect the miRNA, such as nanoparticles or liposomes. Intravenous administration has been shown to prevent metastasis formation.

With the rapid generation of genome sequence data, the number and types of miRNAs characterised has increased exponentially over the past decade. Understanding miRNAs and their involvement in controlling metabolic and pathological conditions is important for targeted therapeutic interventions. Recent progress in this field has demonstrated that they are promising alternatives to conventional drug development. Several databases now host extensive information on miRNAs (e.g. miRBase) and those associated with cancers; e.g. miRCancer database (Xie et al. 2013), dbDEMC provides information on miRNAs that are differentially expressed in cancers (Yang et al. 2017), OncomiR allows searches by

Box 5. Germline cancer mutations and recommended panel testing

Germline cancer mutations and recommended panel testing

In addition to the profiles of somatic mutations that describe different types of cancers, hereditary mutations in several genes have been associated with some tumours. As these are germline mutations, they can be detected well before the development of the symptoms. They are therefore particularly useful for early screening in individuals with family history of the disease. Below are some of those genes and the list of recommended mutations for screening.

BRCA1: a tumour suppressor gene involved in DNA repair and associated with RNA polymerase II and histone deacetylase. It has a role in cell cycle and is required for genome stability.

BRCA2: a tumour suppressor gene involved in DNA repair, and important for genome stability. Mutations in both BRCA1 and BRCA2 are strongly associated with breast and ovarian cancers as well as prostrate and pancreatic cancers. They are recommended for germline screening.

ATM: a serine/threonine kinase involved in control of cell cycle. It acts as a DNA damage sensor and is a regulator of tumour suppressor such as p53 and BRCA1. It is linked to Ataxia Telangectasia and Mantle cell lymphoma.

TP53(p53): a transcription activator and tumour suppressor gene that regulates cell cycle, apoptosis and DNA repair. It also controls an inhibitor of cell-cycle dependent kinase. Mutations in TP53 are found in many cancer types.

CDH1- cell-cell adhesion protein with cadherin repeats: a calcium dependent cell adhesion protein. Mutations in this gene are strongly associated with gastric, breast (germline mutations only implicated in invasive lobular carcinoma), colorectal, thyroid and ovarian (methylation, pleiotropic) cancer.

PLK2- tumour suppressor serine-threonine protein kinase: involved in G1/S phase transition.

MHS6 - member of the DNA-mismatch repair MutS family: identifies a mismatch before repair is effected. It is recruited on chromatin during G1 and S phases and binds to ATP. Is also DNA-binding and is involved in positive regulation of isotype switching in immunoglobulin synthesis, and in host-virus interaction.

PTEN- tumour suppressor- phosphatidylinositol-3,4,5-trisphosphate 3-phosphatase: involved in many biological processes including CNS development, synaptic function, cell proliferation, signalling, and many others. The gene is highly mutated in a large number of cancers.

PALB2: possible role in tumour suppression, and recruits BRCA2 to DNA breaks. It is essential for the formation of BRCA1-PALB2-BRAC2 complex for homologous recombination. Germline mutations are found in patients with pancreatic cancer.

CHEK2: cell cycle regulator and putative tumour suppressor. It is required for DNA repair and apoptosis.

TNFSF10 at 3q26.21 - Tumour Necrosis Factor Superfamily, Member 10: part of an LD block of 100kb on Ch 3. Specific to women of African ancestry for breast cancer. Three specific SNPs for this gene are: 1) rs13074711, 26.5 Kb upstream of TNFSF10 at 3q26.21, strong association with risk of ER (estrogen receptor)-negative breast cancer; 2) rs10069690 best indicator for ER-negative breast cancer at 5p15.33; 3) rs12998806, key indicator for ER-positive breast cancer at 2q35.

Genes recommended for panel screening (Susswein et al. 2016) of germline, hereditary mutations are:

High risk: APC, BMPR1A, BRCA1, BRCA2, CDH1, CDKN2A, EPCAM, MLH1, MSH2, MSH6, MUTYH, PMS2, PTEN, SMAD4, STK11, TP53, VHL

Moderate risk: ATM, CHEK2, PALB2

Increased but less well-defined risk: AXIN2, BARD1, BRIP1, CDK4, FANCC, NBN, RAD51C, RAD51D, XRCC2

cancer type and stage of tumours (Wong et al. 2018) and Human MicroRNA Disease Database (HMDD) hosts data on miRNAs from several diseases including cancer (Huang et al. 2019). These data can potentially be used for diagnosis (biomarkers) or for targeted therapies.

Precision oncology: How NGS applications can assist in surveillance of cancers and improve prognosis

Mutations and other forms of chromosomal aberrations are prolific in most cancers. However, the same types of cancer do not all have the exact same mutations. By using molecular diagnostic approaches to determine which genes have which mutations, the most appropriate treatment can be administered. There are generally several hundreds of mutations in a single type of cancer. The mutation profiles differ across individuals with the equivalent cancer type, and the same treatment cannot be administered, hence the need for screening for mutations to identify the actionable variants that can be targeted by an appropriate drug.

Box 6. EGFR and ALK mutations and non-small cell lung cancer

EGFR and ALK mutations and non-small cell lung cancer

Two important biomarkers for non-small cell lung cancer (NSCLC) are mutations in the genes for epidermal growth factor receptor (EGFR) and anaplastic lymphoma kinase (ALK). EGFR is member of the kinase superfamily and binds to ligands such as EGF. This results in dimerisation and phosphorylation of the receptor, which then activates other proteins downstream of the signaling pathway for cellular responses. EGFR has 31 exons with 11 transcripts. Mutations in exons 21 and 19 affect the kinase domain and result in a constitutive activation of the signaling pathway leading to uncontrolled growth. Three inhibitors of the kinase activity (Tyrosine kinase inhibitors (TKI)- erlotinib, afatinib and gefitinib) are effective at inhibiting this activation and can stop the cancerous growth (Metzger et al. 2011). Patients that do not have these mutations will not respond positively to this treatment. Targeting EGFR is also effective for colorectal cancer through the use of anti-EGFR antibodies administered as cetuximab and panitumumab. The efficiency of EGFR treatment is also dependent on mutation profiles of RAS genes (KRAS and NRAS) and these have to be ascertained (Kafatos et al. 2017)

The detection of specific mutations in genemarkers using conventional PCR and qPCR are current molecular diagnostic tests for cancer screening. These are simply applied on faecal samples for diagnosis of colorectal cancers to target excreted cancer cells. However, NGS methods can accelerate the detection of genetic abnormalities and identify novel ones. As more data become available additional diagnostic markers can be identified. The American Association for Cancer Research (AACR) has launched the GENIE (Genomics Evidence Neoplasia Information Exchange) initiative, a platform where extensive genomic records from 59 types of cancer, are shared. This resource provides genomic data on mutations to help in decision-making for treatment of various types of cancers.

Immune-oncology-targeted therapy with specific antibodies

Antibodies against mutated forms of oncogenes have been shown to be useful for diagnostics as well as therapeutics (Vermaelen et al. 2018). In Non-Small Cell Lung Cancer the following genes are frequently expressed as altered forms due to genomic alterations: ALK, ROS1, EGFR, BRAF, MET, HER2, NTRK, and RET. Immunohistochemistry on tissue biopsies with the specific antibodies allow an assessment of the gene expression for the target protein (Zhiwei et al. 2017). Significant success has been achieved for treatment of several types of cancers through an immunotherapy approach. The aim is to enhance the immune response against tumour thriving cells through re-activation of T cells. Considerable efforts are now directed towards the use of PD-L1/PD1 immune-inhibitors. PD-1 (programmed cell death protein) is a receptor for (its ligand) PD-L1 and is found on many cells and tissues including T-lymphocytes. Expression of the corresponding ligand PD-L1, is increased in tumours under the influence of interferon gamma, a cytokine which is released from infiltrated lymphocytes and natural killer cells. Binding of PD-L1 to its receptor PD-1 on T-cells greatly inhibits the killing function of these cells and allows the cancer cells to thrive and divide. Precision immunotherapy with monoclonal antibodies against PD-L1/PD-1 has resulted in an impressive achievement for the treatment of several types of tumours and more importantly for NSCLC. Pre-treatment screening should establish elevated expression of PD-L1 before the administering the inhibitors (Bell et al. 2018).

3.4.1.5 Sickle Cell Disease

Sickle cell disease (SCD) most commonly occurs in individuals of African descent. SCD is caused predominantly by a single point mutation in the beta-subunit of haemoglobin resulting in a single amino acid substitution (Rees et al. 2010). This is the principal oxygen transporter in red blood cells,

and the mutation causes a sickling of the blood cells and resultant constriction of blood flow. It is one of the most common monogenic diseases in humans, which was fixed in certain populations because of the protective effect of the sickle cell mutation against malaria. SCD therefore has the highest incidence and prevalence in tropical regions, particularly in Sub-Saharan African countries, where it affects approximately 300,000 newborn babies every year (Diallo and Guindo 2014). SCD is a chronic disease of variable phenotypes, associated with high morbidity and mortality. Unfortunately, there are limited drugs or therapy options that are effective in managing the disease, particularly in the developing world, where blood transfusions or gene therapy are unaffordable. The burden on the health-care system of affected countries is therefore high. Since the causative mutations for SCD are well characterised, cost-effective genetics tests have been developed for diagnosis. However, there is a need for more wide-spread newborn screening so that cases can be identified as early as possible to enable successful interventions. Additionally, further research is required to identify other genetic factors predisposing SCD patients to different levels of disease severity or adverse effects.

3.4.2 Communicable/Infectious Diseases in Africa

Africa has the highest burden of infectious diseases worldwide. Microbial and parasitic infectious agents thrive in the tropical environment and are generally maintained in a human or animal reservoir, from which transmission to the host is affected when the right conditions prevail. Pathogen dissemination happens through various routes, including air, soil, food, excrement, insect vectors, animal or human to human. Some control measures specifically target the transmission path, the vectors and reservoirs. Disease risks are determined by the drivers (climate, socio-economic) and the transmission pathways, and pathogen distribution follows that of the insect vector (where relevant). In addition, the genetics of the host underpin the severity of an infection, as in the case of HIV controllers who can host the virus but do not develop the disease. Although there has been progress in reducing incidence of some diseases, most are still prevalent and impacting morbidity and mortality rates. While considerable efforts have been made to minimise pathogen dissemination, economic and social conditions of many countries are adding to the current plight. In addition, antimicrobial resistance (AMR) and resistance of insect vectors to insecticides is rapidly increasing and aggravating the situation. However, there is still a dearth of information on AMR genes in medical, agricultural and environmental settings (see Box 7: Antimicrobial resistance).

Genomic medicine can apply to different components of managing infectious disease, for example: epidemiology and surveillance, diagnostic assays (DNA and other methods) to detect and accurately identify pathogens, and pharmacogenomics to better understand best drugs/medications and dosage. The recent COVID19 outbreak has implemented many of these measures which required collection, analysis and sharing of information. Four of the most important infectious diseases, in terms of burden in Africa are briefly described here along with some example genomics applications.

3.4.2.1 Malaria

Malaria is still the most important communicable disease in the world, but ninety-two percent (92%) of the world's malaria cases are in Africa. The five countries accounting for over 50% of the cases are Nigeria, Uganda, Mozambique, Democratic Republic of Congo and India (World Health Organisation 2018). In 2017-2018, Madagascar had an increase of more than half a million cases. The deployment of insecticide-treated nets has been successful in some parts of the Continent, but the 10 highest burden countries remain highly vulnerable. Research into new vaccine and drug targets are slow in bringing long-lasting solutions for eradication. Novel approaches are therefore urgently needed to find more promising alternatives.

Different genotypes of *Plasmodium falciparum* have been associated with virulence or severe malaria cases. Genetic typing of the parasite strains can help identify such strains and address the cause of resistance to some treatment options. The malaria parasite has evolved a complex strategy of antigenic variation in the variant surface antigen (VSA) with about 60 *var* genes thus making it recalcitrant to immune response. Erythrocytes express these parasite antigens on their surface, and this promote their adherence to host receptor on vascular endothelial cells (Hviid and Jensen 2015). The *var* genes code for PfEMP (*Plasmodium*

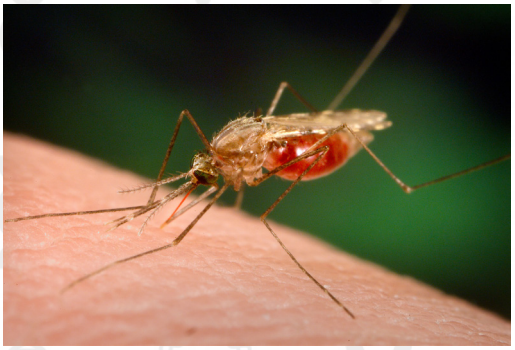


Photo credit: NIH

Mosquito

Box 7. Antimicrobial Resistance

Antimicrobial Resistance

Antibiotics are compounds produced by microorganisms to kill other microorganisms. They have been widely utilised for treating infections in humans and animals. Many have been chemically modified to increase their potency. There have been few new compounds introduced over the last few decades to fight pathogens, hence the limited repertoire of drugs for treating microbial infections. Anti-microbial resistance is a major health issue that is proving extremely challenging to address. The extensive use of antibiotics in healthcare as well as in animal husbandry (Lambrecht et al., 2018), is regarded as being the main cause for the emergence of many pathogens carrying resistance genes, although soil bacteria naturally harbour resistance genes. Some examples are provided below.

Methicillin resistant Staphylococcus aureus (MRSA) carry the mecA gene, whose product (PBP2) is insensitive to beta lactamases. This is an altered form of the antibiotic target. Other mechanisms include efflux pump that removes the antibiotic from the cell. ESBL are extended-spectrum beta lactamases found extensively in Klebsiella pneumoniae, that degrade beta lactam antibiotics. Clinical use of a beta-lactam inhibitor, such as clavulanate, can bypass the resistance. The most frequent ESBLs are of Class A and include CTX-M, TEM and SHV families, all widely found in K pneumoniae (Daehre et al., 2018). Resistance genes are readily acquired through horizontal transfer and often are found on mobile elements such as plasmids. They are therefore easily disseminated within the microbial communities. Mutations also contribute to genetic changes and enhanced resistance. More importantly, there can be multi-drug resistance (MDR) within a single genome, which carries several genes that confer non-susceptibility to the anti-microbials. Complex microbial habitats (e.g. the gut or respiratory microbiomes) are ideal environments that promote gene exchange and rapid spread of resistance. Whole genome sequencing of bacterial genomes produces high-resolution data on resistance genes and can inform which phenotypic tests should be carried out. In clinical cases, such as in intensive care, time is limited and treatments cannot be delayed.

falciparum erythrocyte membrane protein) which is responsible for many of the symptoms of severe malaria such as anemia and vascular congestion. Expression of distinct subsets of PfEMP has been associated with severity of symptoms. The *var* groups A, B and DC (domain cassette) 8, 11,13 are known to be highly virulent (Lavsten et al 2012). Similarly var group C and DC8 genes expression are higher in febrile patients compared to non-febrile (Gupta et al. 2019). Transcript profiling of these variants can indicate the gravity of the disease and help provide the right treatment. This can be done using routine quantitative PCR for gene expression assays on RNA extracted from blood.

Malaria genomics has accelerated discovery into the cellular mechanisms of pathogenicity as well as the understanding of genetic diversity of *Plasmodium falciparum* (pf). Work by a large consortium of malaria researchers (Malaria Genetic Epidemiology Network -MalariaGEN) has led to a deeper insight into the parasite lifecycle and pathogenicity. WGS of both parasite and vector has revealed their intricate interactions and explained appearance of resistance to drugs and insecticides. Plasmodium genomes harbor more than 900 000 variants (SNPs) which are catalogued in the MalariaGEN database (Kirchner et al. 2016) revolutionary progress over the last 3 years in nucleic acid sequencing, reverse genetics, and post-genome analyses has generated step changes in our understanding of malaria parasite (Plasmodium spp.. Close surveillance of the parasite through WGS can identify new strains with altered phenotype such as increased virulence or drug resistance. Pf has a complex strategy for control of gene expression, which includes epigenetic modification of histones. The parasite development is highly dependent on those histone modifiers which could be targeted for drug discovery. An important breakthrough has been the identification of mutations in the gene *kelch13* that are linked to resistance to artemisin. WGS of Pf is recommended in order to closely monitor the appearance of novel resistance mechanism that can result from selection of mutants.

3.4.2.2 Tuberculosis

Africa has the second highest incidence of tuberculosis (TB) in the world, after South-East Asia. TB is a common cause of death in HIV positive patients and the combination of both pathogens in the patient results in a higher death rate, as the immune system is highly perturbed. The urgency to develop a common policy for TB research and management of the disease saw the first WHO global ministerial conference in Russia in 2017, followed by a high-level United Nations meeting in 2018. The main outcome was a commitment to a declaration for ending TB across the world. This global threat to public health can only be

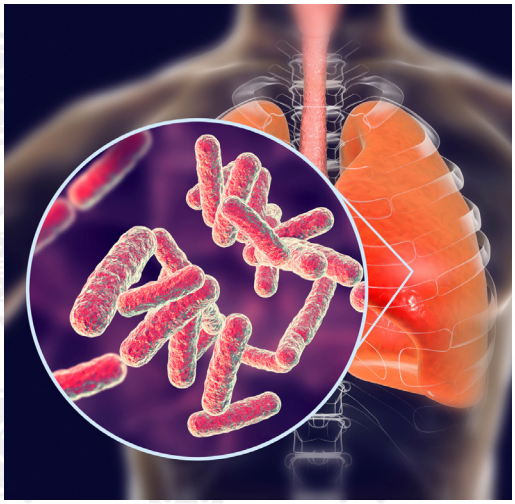


Photo credit: Science magazine

Lungs infected with TB

tackled with the participation of all countries and a coordinated approach. Drug resistance is today the most critical issue with treating TB. Multi drug resistant (MDR) *Mycobacterium tuberculosis* is insensitive to rifampicin and isionazid, the most widely used treatment. Extended-spectrum resistance (XDR) TB strains do not respond to first, or second line antibiotics and cannot be treated. Parts of the world with highest incidence of MDR include Russia, India and China (WHO 2018).

TB diagnosis can be slow due to the need for culture and the slow growth rate of the organism. NGS can increase the speed of diagnosis and antimicrobial resistance testing, particularly if done from sputum, however, sequencing from sputum samples remains a challenge. The turn-around time for NGS-based diagnosis can be as short as 24-48 hours, depending on the sequencing platform used (Dlamini et al. 2019), new benchtop devices can be even quicker. There are several TB drug resistance marker databases and analysis pipelines for NGS data that perform lineage classifications and identification of drug resistance mutations (Gröschel et al. 2018). Though the mutation databases may not be comprehensive, WGS data provides the opportunity to identify novel markers, which standard drug resistance tests do not offer. Early diagnosis of drug resistant TB is essential for improved treatment and reducing the spread.

3.4.2.3 HIV

Similarly, to TB, HIV incidence remains high in Africa and co-infections are common. Significant numbers of people living with HIV (PLHIV) do not have access to anti-retroviral therapy (ART) for various reasons such as inaccessibility to remote, rural areas. HIV is a retrovirus which can integrate into its host's genome. It has the gag (structural proteins), env (envelope) and pol (protease, RT, RNase H, and integrase) genes encoded on the viral RNA. The viral reverse transcriptase produces the corresponding cDNA. There are nine genetically diverse subtypes of the common lineage M of HIV, which have likely arisen because of the use of reverse transcriptase during the replication cycle. RT is known to wrongly incorporate nucleotides during synthesis of the complementary DNA, hence the high level of genetic variation in HIV, which, coupled with recombinant formation, expand the range of intraspecific polymorphisms. HIV detection methods must be capable of identifying all forms of the virus to be considered viable for implementation. Molecular detection is carried out by PCR of gag or pol genes. Mutations in gag and pol genes confer resistance to anti-retroviral therapy (Van Laethem et al. 2015); drug resistance mutations are likely being selected when individuals at risk of infection receive prophylactic anti-retrovirals. A large number of such mutations have been identified in many HIV subtypes, making treatment even more challenging. Understanding HIV diversity is complicated by high recombination rates, leading to differing drug susceptibility, as well as compartmentalisation of the virus in immunological tissues where further recombination can occur. Sequencing of virus samples is necessary to identify drug resistance mutations for improved response to treatment and enhanced management of the disease (Dessilly et al. 2018).

Several vaccines have been designed based on the gp120 envelope protein of HIV that can induce effective antibodies, and new recombinant vaccines can be designed for enhanced immunogenicity and protection. HIV targets the immune system by entry through CD4 TH cells and chemokine receptors such as CCR5. Intense interactions of

Photo credit: dw.org

People hold the red ribbon that symbolises awareness and support for people living with HIV

CD4 TH cells with specific T cell receptors (TCR) with the viral capsid epitope Gag293 explain why some individuals do not have active disease even when infected with the virus. These individuals are known as HIV controllers and have a low virus titre. They have specific TCRs that can interact with five different Human Leucocyte Antigen-DR (HLA -DR) for control of the virus. The T cells are able to kill the virus-infected cells (Galperin et al. 2018). This provides new therapeutic applications for the TCR as it works in a broad genetic HLA background.

In addition to tracking possible host protective genetic factors, GM can be used to track HIV epidemiology, evolution and transmission. There is an abundance of sequence data available in the public domain for HIV and databases collating information, for example, the Los Alamos HIV database (https://www.hiv.lanl.gov/content/sequence/HIV/mainpage.html) and the HIV drug resistance database (https://hivdb.stanford.edu/).

3.4.2.4 Respiratory infections

The burden of respiratory infections is the third cause of mortality in Africa after malaria and HIV. Africa and Asia have the highest mortality of children under 5 years due to such infections. Several pathogens can induce lower respiratory tract infections, including Streptococcus pneumoniae, Klebsiella pneumoniae, Staphylococcus aureus, Haemophilus influenza and other species. With the assistance of the Global Alliance for Vaccines and Immunisation, many countries in Africa have benefitted from the pneumococcal conjugate vaccine (PCV).

Specific pathogen detection is essential for diagnosis and molecular techniques are therefore more sensitive for this purpose. Based on the polysaccharide capsule (cps locus), there are 94 known serotypes of S pneumoniae and 11 176 multi locus sequence types. Patients' samples comprise of multiple serotypes and those present in low titres are not easily detected. Currently the recommended detection approach is through identification the autolysin gene (lyt A) and the permease gene (piaB) by real-time PCR. A more recent assay for the SP2020 gene, a transcriptional regulator, has been reported as being very specific. Multiplex qPCR can also detect several serotypes in one assay (Olwagen et al. 2017; Tavares et al. 2019).

Use of PCV vaccines with seven commonly found serotype antigens has resulted in the increased prevalence of other serotypes. Hence there is selective pressure on the emergence of less frequent or new serotypes. Whole genome sequencing (WGS) has been used to show that there are serotype switches that can occur producing altered phenotypes such as penicillin resistance (Makarewicz et al. 2017). This switching occurs within the cps cluster of genes for the synthesis of the pneumococcal polysaccharide, which is a virulence factor. WGS has shown that there are rearrangements and recombinations within the cluster that could lead to loss of functions. A large number of S pneumoniae genome sequences are available in the Streptococcus pneumoniae Genome Database (SPGDB) (Swetha et al. 2014).

3.4.2.5 Genomic Medicine for Infectious Disease Control

Close monitoring of pathogens requires precise tools for their identification and classification. NGS technologies can provide this level of accuracy. The Food and Agricultural Organisation (FAO) and WHO have recommended whole genome sequencing of pathogens in clinical settings for detection and epidemiology, as it provides the highest resolution phylogenetic method for tracking and understanding pathogen dissemination. Microbial genomes evolve at different rates, with some changing rapidly over few generations, while others remain highly conserved over long periods. This is further complicated by horizontal exchange of DNA, a scheme for bacterial species to acquire foreign genetic material from their environment particularly in complex mixed microbial habitats. Integration of foreign DNA into the pathogen's genome results in mosaic patches of insertions, which could be resolved with whole genome sequencing.

Molecular diagnostics through Next Generation Sequencing (NGS) are taking the lead over traditional methods. Whole genome sequencing (WGS) or sub-genomic sequencing has become the norm in microbiology in some developed countries, as this approach is more rapid and accurate for the detection and control of resistant pathogens. Unlike lengthy culture and antibiotic testing methods, WGS can save lives by providing data to guide the appropriate choice of antibiotics or antivirals. HIV treatment is effective when a combination of several anti-retroviral agents are used for therapy, but it is essential to first identify the drug resistance mutations present in the virus in order to decide on the best combination of drugs to use (Mu et al. 2018). For the hepatitis C virus, knowing the genotype is essential to provide guidance in determining the best drug to use as well as the duration of the treatment (Dahiya et al. 2019).

A thorough understanding of the molecular mechanisms of pathogenicity and the

corresponding interactions with host and vector, is essential for designing new therapies. Immunisation and anti microbial compounds have been highly effective in the control of many pathogens, however, genetic changes in the pathogen genome can lead to failure of some vaccines and drugs. Microbial and parasite genomes mutate rapidly and more so under environmental pressure such as exposure to antibiotics or other compounds. Despite the genetic flexibility of the large repertoire of human immune molecules, virulent pathogens display an equally complex, intra-specific, genetic variation. There is also evidence of host-pathogen reciprocal co-evolution whereby interacting proteins from both organisms evolve at comparable rates. Balancing selection from pathogen richness is thought to drive the extreme variation of MHC loci (17 000 alleles in human MHC) whose products are responsible for antigen presentation to T cells (Vossen et al. 2002). MHC loci evolve to increase fitness of the species against infectious agents, while pathogens evolve to escape the immune system using strategies such as 'hiding' in host cells or mimicking host molecules. HLA alleles determine the host's ability to deal with different strains of pathogens and therefore its survival. We can use genomics to improve our understanding of both host and pathogen variability and HLA typing to map susceptibility to diseases.

Another obvious application of genomics is to track the spread of outbreaks and to design interventions to control them. There is no clearer evidence of the need for this than the coronavirus pandemic. For some pathogens, such as viruses, genome sequencing is efficient, while for others (e.g. bacterial pathogens), traditional typing methods are cheaper, faster and therefore more feasible for routine surveillance (Sabat et al. 2013) provide a review of molecular typing methods for bacterial pathogen surveillance and outbreak detection in clinical practice). Nevertheless, the dropping costs of NGS and emergence of desktop sequencing devices are enabling greater adoption of NGS for infectious diseases. A number of tools have been developed to support the analysis, including IDSeq, which enables pathogen identification from metagenomics data (https://idseq.net/), NextStrain (https://nextstrain.org/), and PathogenWatch (https://www.sanger.ac.uk/science/tools/pathogenwatch), among others. There are also international projects which coordinate research and data in the field, such as COMPARE (https://www.compare-europe.eu/) and the numerous COVID-19 data portals which have recently emerged. The tools and portals were designed to facilitate the sharing of data which is essential for global responses, though there are some ethical and legal issues and ongoing debates surrounding openness of the data and benefit sharing.

The application of genomics in infectious disease control requires laboratory, computational and analytical capabilities that are tightly linked to public health systems. Infectious disease materials should be managed in laboratories with trained staff and appropriate biosafety level compliance. Cross-border collaboration and cooperation is important and adhering to standards in sample collection, processing, data generation, analysis and reporting is essential (Schneidman et al. 2018). Countries which are not yet equipped to manage epidemiological surveillance or outbreak detection should develop relevant policies and engage with their local or regional Centre for Disease Control (CDC).

3.4 Conclusions

A Genomics medicine strategy in Africa requires a two-pronged approach to address: 1) filling the gaps in locally relevant genomics research, and 2) implementation in clinical practice for relevant diseases. A policy is required to harmonise procedures for biomedical research that would feed into genomic medicine applications and to generate data for the many diseases/populations that have not yet been studied in the local context. Information is lacking for many NCDs in terms of their prevalence in different geographic regions and contributing factors. The genetics of these diseases for most African populations are still unknown, which is a major challenge for implementing genomic medicine. It is imperative that the African scientific community is mobilised to develop a new strategy to undertake genomic studies using a coordinated approach for the results to feed into clinical implementation. H3Africa projects have paved the way for a more coordinated approach, but more harmonisation is required. Data must be utilised to drive change in healthcare to encourage the use of GM. Early detection of predisposition to diseases using known genetic markers can lead to timely management and improved prognosis. Next generation sequencing and genotyping arrays can bridge the gap in missing data and could be used in medical applications. Genome sequencing is being widely implemented in clinical settings for microbiological diagnostics and surveillance in some countries. Affordable platforms are now available that can be adapted for use in resource-limited locations in many parts of Africa.

3.5 Recommendations for cohorts and diseases

- The generation of more diverse African human reference genomes is urgently needed. This can be achieved through whole genome sequencing of representative populations from the four main families - AfroAsiatic, Nilo-Saharan, Bantu and Khoesan. The data should be consented for use as controls in other studies and in reference panels.
- There is a need to set up longitudinal cohorts with inclusive participation from different geographical regions and ideally across the Continent. These should be used for surveillance and to detect and monitor diseases.
- GWA studies in African populations need to be strengthened in order to identify and confirm variants relevant for African populations for specific diseases. Findings should be functionally validated.
- For complex traits, research data should be used to determine polygenic risk scores for specific populations and disease panels developed or updated accordingly.
- Determination of disease-causing variants should be improved using existing data from other populations that are currently in databases (e.g. ClinVar) and assessment of whether the same variants are associated with the same diseases in Africans.
- Clinical implementation for NCDs should use several approaches depending on the disease: 1) PCR or other assays for single mutations (e.g. sickle and other common mutations and triplet expansion disorders), 2) Arrays (CGH, SNPs/pathogenic mutations, pharmacogenomic variants) where panels are available 3) NGS (single genes, gene panels, WES, WGS) for more complex diseases and undiagnosed disorders.
- We recommend a comprehensive strategy to bring next generation sequencing platforms into clinical settings. This will require a change in (or a complementation of) diagnostic procedures and training of medical /laboratory personnel.
- In addition to traditional typing tools, WGS should be considered for pathogen surveillance and monitoring, as it provides high-resolution genetic profiles for pathogen tracking and evolution.
- WGS should be considered, where feasible, for more routine pathogen diagnosis and drug resistance testing. Known and novel anti-microbial resistance can be detected using WGS and removes the need for multiple tests.
- Genome sequences of pathogens should be stored along with relevant de-identified clinical data and shared.

Section 4. Phenotype and Clinical Data

4.1 Introduction

Clinical data are essential for the discovery phase of GM. These data can take the form of existing clinical data collected during the provision of standard healthcare, or bespoke clinical data collected during a research project. In this section we discuss informed consent issues for the use of routine clinical and health data for GM research; the legislative and governmental context for using clinical data; integration of GM approaches with clinical health data systems; the need for standardisation of clinical case report forms, data coding, data storage and data transfer; feedback loops that integrate the return of actionable results within the routine provision of healthcare; and future directions - including machine learning analysis - to harness clinical data for GM.

Currently, clinical data collected during routine provision of healthcare are seldom used as a primary source of phenotype data for GM research. Instead, research endeavours tend to collect bespoke observational, clinical, laboratory and treatment data in a system that is created in parallel with clinical practice and provision of healthcare. Such parallel healthcare paradigms can weaken existing health delivery structures, overburden healthcare professionals and dilute their efforts to provide health services, and should be avoided: it is important to ensure that GM approaches aim to complement and strengthen existing systems in a collaborative effort with health care providers, rather than competing for scarce resources in parallel efforts that detract from routine healthcare provision. This can be more easily enforced when the GM initiative is national and driven by government. In existing national GM initiatives, clinical data collected for GM implementation in a healthcare plan are well integrated with patient records. Clinical documentation includes rich data about diagnoses, signs and symptoms, risk factors, treatments, family history, exposures, and clinical decision making. Establishing a gold standard validation whether a patient has a diagnosis or a given trait generally involves a review of the clinical notes.

Clinical health data collected from individuals should wherever possible be retained in the health departments so that they can support healthcare provision: currently, there are too many private repositories of research health data from participants that are never used further to improve the health of those individuals. A paradigm shift away from such silo'd research can enable an ecosystem in which health data collected from participants for routine healthcare can be used - with the appropriate informed consent from

Photo credit: Shutterstock

Data display

healthcare clients - to inform GM research; and in turn the findings of GM application can be returned to healthcare clients through existing healthcare provision mechanisms. Furthermore, once GM research has been completed, the interpretation of findings might be translatable into standard of care at a population level without necessarily requiring additional expensive genetic tests for individuals – for example, using pharmacogenomics research into adverse events to establish that a particular medicine is generally not well-tolerated in a particular population. Health Information approaches can ensure genomic data are digitised in a way that is meaningful to clinicians, and encourage wider data sharing where appropriate to maximise the long term utility of test data (Weng and Kahn 2016).

4.2 Existing Clinical Data

The rise in the availability of longitudinal patient information in electronic health records (EHRs) along with DNA biobanks have greatly contributed to genomic research using EHR data for physical traits (Robinson et al. 2018). EHRs can provide a rich data source of health-related phenotypes, e.g. drug response traits or disease severity. Genetic variants are frequently identified as causal factors in disease and are relatively easy and inexpensive to measure. High-throughput genotyping is converging with EHRs, providing opportunities for scientists to use routine healthcare data to accelerate genomic discovery where appropriate consents are in place. In order to achieve this vision, EHR-linked DNA biobanks are being created worldwide, by health institutions (Gudbjartsson et al. 2015; Leitsalu et al. 2015; Genomics England 2018). However, detailed disease and drug-response phenotype information found in EHRs may consist primarily of doctor's notes for provision of healthcare and can be difficult to mine. Despite these difficulties, studies have successfully replicated known associations, and described disease aetiology and drug response traits, demonstrating the potential of EHRs for broad-based phenome-wide association studies (PheWAS) (Wei and Denny 2015).

4.2.1. Building relationships with health service stakeholders

Harnessing routine clinical data to describe disease phenotypes for research will require building relationships between researchers and health systems stakeholders - often government health departments at provincial and national levels; and working closely with core health services and healthcare providers. Such partnerships can also help to ensure that the local appropriate legislative requirements are respected: e.g. Health Acts, Privacy Protection Acts, Protection of Children Acts etc. Creating formal collaboration and data sharing agreements between research organisations and government can also provide assurances of appropriate and ethical data use that will minimise participant risk and maximise autonomy, benefits and quality of healthcare for healthcare clients. The partnership should work both ways though, and the integration of relevant clinical data and results from GM research into provision of healthcare should be encouraged.

4.2.2. How can these data be utilised?

Re-use of routine clinical data for GM research can be more easily facilitated where electronic platforms are already in use for managing routine health records. Standards are often in place already for clinical data capture and for coding diagnoses (e.g. SnoMed or ICD9/10/11 codes), pharmacy data (e.g. ATC coding) and laboratory data (e.g. LOINC coding). SNOMED-CT is a comprehensive, multilingual clinical healthcare standard that encodes meanings used in health information and supports the effective clinical recording of data (http://www.snomed.org/). Specifications for data interoperability that are also often in use include the HL7 (Health Level Seven International: https://www.hl7.org/) and more recently the FHIR (Fast Healthcare Interoperability Resources: https://www.hl7.org/fhir/overview.html) specifications, and these can be adopted for portability and re-use of clinical data for research purposes.

Because of the granularity and repeat measurements often captured during patient care, using existing clinical data also requires the development of statistical methodologies to use longitudinal and observational routine clinical data to describe GWAS phenotypes. These longitudinal data offer opportunities to research more nuanced phenotypes, such as rates of response to medication or the severity of a disease phenotype. These routine clinical data can help to move GM approaches forward from current research methods that analyse a simple binary outcome of having a condition or not.

4.2.3. What checks and balances will be required?

We need to develop strong ethics oversight and robust informed consent processes where individuals accessing healthcare can receive the appropriate information, and should they so wish, specifically agree to their routine health data being used in GM research. Informed consent must also be taken regarding return of findings – both primary outcomes as well as incidental findings, actionable and non-actionable findings; and provision made for the return of findings through standard healthcare facilities and services

to ensure appropriate health benefits reach participants and patients.

Information from clinical records must be carefully compiled, stored and protected, because GM research links highly sensitive detailed health data to genomic data that cannot be anonymised, creating a situation where sensitive routine health data (possibly additional to primary outcomes being researched) are always potentially re-identifiable. Community-level harms are also possible if clinical/health outcomes are linked to specific identifiable populations - often the case for African populations. Technical solutions must be employed to minimise the risk to participants and communities of their health data being exposed in an identifying or stigmatising way, and specific guidelines and policies for the use of routine clinical data for research must be negotiated to protect participants and preserve their autonomy and privacy. Procedural controls and clear data use regulations must be established to protect such clinical data from predatory data acquisition and from data sharing or movement across borders that are contrary to consents given as well as legislature protecting personal health data (often 'special data' in Protection of Privacy legislation) (Tiffin et al. 2019).

4.3. Collecting New Clinical Data

Where routine health data cannot be accessed or repurposed for research, longitudinal clinical data including observational (symptoms and doctors' observations), laboratory results and medication data will need to be collected. This should be done in partnership with healthcare providers to prevent duplication of data collection and overburdening healthcare providers with data capture duties, or other complications that weaken rather than support healthcare provision: the resources available to ensure accurate data collection and the electronic availability of those data remain severely limited (Shaffer et al. 2018).

In some circumstances, regionally centralised database facilities rather than many individual databases may be a more economical and effective way of collecting, standardising, validating and securely storing new clinical data, and a federated model can provide a secure, backed-up, central database platform within which each data contributor has sole access to and control of their own data, but data are all standardised to a common structure and specification. REDCap is a databasing platform that is widely used for clinical research databases, is free to use and has a large user-base and support network; and is well-suited for federated database models (Obeid et al. 2013). Going forward, collaborative agreements and Memoranda of Understanding can be put in place for negotiated data sharing for future meta-analyses, but the clinical data are already harmonised because they have been captured into a single database structure. A centralised database solution and adherence to interoperability and data capture standards can also allow for clinical data from community health workers or researcher to be collected in the field using mobile health apps, and transferred to central storage on the fly, increasing data security.

The granularity of clinical data must be high to allow modelling of the nuances of clinical presentation of conditions and accurate stratification by presentation where appropriate. Research based on a body of clinical and diagnostic data from African populations rather than Caucasian data that have dominated to date can ensure that African individuals receive correct annotation of disease risk variants for a disease as it presents in Africans (Popejoy and Fullerton 2016; Mulder 2017).

Generating new clinical data resources will require significant input of health informatics infrastructure and development of health informatics skills, but also provides an opportunity to ensure clinical data are collected in a standardised way, using standardised phenotype descriptions and ontologies, data capture and coding. Standards include the Human Phenotype Ontology (HPO) and PhenX (consensus measure for Phenotypes and Exposures: https://www.phenx.org/), among others. HPO provides a standardised vocabulary of phenotypic abnormalities associated with human disease. The HPO was developed using various sources of disease data, including the medical literature, Orphanet, DECIPHER and OMIM (Robinson et al. 2008). Developing uniform case report forms, variable descriptions and code books - such as the work done by the H3Africa Phenotype Harmonisation working group and H3ABioNet Health Informatics work package to date - can ensure that clinical data are comparable across different sources and can be used seamlessly in meta-analyses. FAIR principles and populating study meta-data can also enable this (Wilkinson et al. 2016). There is a need to build computing solutions that can integrate heterogeneous data in a seamless manner, using interoperability standards such as the HL7 and FHIR specifications: failure to ensure interoperability may result in costly trials, missing data and reduced opportunity for future analysis.

Using standardisation and standard operating procedures for collecting clinical data and international coding standards can ensure maximal benefit can be derived from datasets through re-use, where appropriate consents are in place (e.g.

ATC codes for pharmacy/medical products, LOINC codes for medical laboratory data, ICD10 codes for health conditions and clinical diagnoses, SNOMED codes for medical terms). Where these sensitive data need to be shared, blockchain technologies may aid secure data transfer and help to track data origin, re-use and provenance to ensure ethical and legal compliance in data use.

4.3.1. Collecting clinical data in the field using mobile devices

Increased coverage and decreasing cost of mobile phone networks has raised the profile of mobile phones and tablets as tools for clinical health data collection in the field. These data usually conform to the format of data from routine clinical care, using equivalent data fields and codes, and can be uploaded on the fly to centralised, secured databases of clinical data. These data should conform to a data transport specification such as the FHIR specification, and data governance structures should be in place to ensure that the data are adequately protected during transfer, and that mobile devices can be remotely wiped should they be stolen or lost. Technology stacks can facilitate transfer of these data from the field; for example, the OpenHIM (www.openHim.com).

4.3.2. Operationalisation of precision medicine clinical solutions in standard of care

Individual genetic data include Individual Research Results (IRRs) and Incidental Findings (IFs). IRRs are results of the research project related to the health status of individual participants, while IFs also concern an individual participant's health, but are findings that lie outside the original objectives of the research project. Returning clinically relevant findings to participants through routine healthcare provision requires collaboration with core health service to enable integration and harmonisation with existing services and clinical practice/standard of care. Engagement with healthcare practitioners is essential, and the clinical relevance and impact of the findings must be clearly articulated. A quality control pipeline must be established to determine if the findings are scientifically valid and confirmed, whether they have significant implications for the health of the individuals, and whether there is a course of action to ameliorate or treat identified health risks (Thorogood et al. 2014; Sitapati et al. 2017). Healthcare professionals must establish formal, interdisciplinary clinical advisory committees to decide on the return of GM findings during routine healthcare, with a medium-term goal to operationalise GM through integration of GM tests and return of findings into routine healthcare, to ensure patient benefits.

Photo credit: AESA 2019

Dahabo Adi of Grand Challenges Africa collecting data from a patient

4.4. Evidence-Based Medicine and Machine Learning using clinical data

The success of GM depends on the availability of healthcare and biomedical data that can be mined to advance GM. Processing such large and often disparate amounts of data requires huge computing power as well as algorithms that can learn by themselves - that is, machine learning (ML) algorithms (Mesko 2017). With the recent developments in ML and its application to medical big data, ML promises to assist clinicians in mining the ever-increasing loads of medical knowledge and patient data into routine care (Forbes Team Insights Contributor). Such techniques can help scientists reveal patterns and relationships in clinical data that might not otherwise be detected or expected (Krumholz 2014). The main aims of ML being to identify similarities and differences in patient phenotypes and genomes, normalise diagnostic approaches, improve existing treatments, identify new drug targets, optimise prediction rules as well as prevent clinical errors due to human cognitive bias and fatigue (e.g. analysing thousands of images in a day), and finally deliver GM (Scott 2018). ML can drive healthcare improvement by using data, algorithms, and models to predict an event and simulate interventions. Machine learning for GM, using clinical data, has been used to integrate big data for GM in acute myeloid leukemia (Lee et al. 2018), for precision psychiatry (Bzdok and Meyer-Lindenberg 2018) amongst others. Medical big data, along with efficient data mining methods, can lead to a better understanding of diseases that can be translated into targeted treatments (Sitapati et al. 2017). An evidence-based approach is essential, otherwise an ML model might exclude features that clinical evidence supports as risk factors for a certain condition and affect the clinical value of the model and result in inappropriate interventions and patient risk. ways to boost the validity of a predictive model, engage data scientists around healthcare projects and save time and money

Key recommendations for phenotype and clinical data

- Build strong relationships with health service stakeholders such as government health departments
- Build governance systems for clinical data, such as informed consent processes for research use and standard operating procedures for research access to data, to ensure the data are used appropriately.
- Build relationships to facilitate ethical and consented sharing of routine health data for GM research and ensure feedback of findings and results to health services.
- Ensure that new clinical data collection is done in collaboration with existing health infrastructure to avoid diluting scarce health resources into parallel data ecosystems, and to ensure that all data are used towards providing better healthcare to the participants.
- Integrate GM approaches and the return of results with routine healthcare provision, at individual, community and population levels as appropriate.
- Use data standardisation with existing, commonly used standards.
- Leverage existing REDCap databases and/or database templates, using federated data storage where needed.
- Ensure most effective research use and meta-analyses from collected data, where consents are in place, e.g. for data coding and data transfer protocols.

Section 5. Linking Genes and Diseases: Genotype-Phenotype Evidence and Actionability

5.1. Summary

This section addresses the correlation between genotype and phenotype data, which is key to understanding genomic-related data for biomedical research, human health and genomic medicine. 'Phenotype' refers to the physical characteristics of the individual, and 'genotype' refers to the combination of all genetic factors belonging to that individual. The term 'DNA variant' refers to a change at a specific site or region in a DNA sequence compared to a given reference sequence. It may be a newly acquired change (somatic mutation) or an inherited one (germline) that forms part of the normal phenotype variation of that individual, or it may have a deleterious effect on their phenotype.

We describe existing approaches to identifying disease-causing (pathogenic) variants, how the identified variants can be validated, what information is already available and actionable, and what tools are available for identifying new disease-associated variants in African populations. These approaches can be used to implement a reproducible and evidence-based method for determining variant impact and actionability for health.

5.2. Measuring Phenotype and Genotype Associations

The combination of genotype and phenotype data can be used to explore complex pathologies as well as rare and ultra-rare diseases (Kent 2009). The approaches for monogenic versus multifactorial complex diseases will, however, differ. Several steps are involved in this process:

- Robust statistical and computational analysis of potential associations between genetic variants (genotype) and disease-related observations in a patient (phenotype) can identify a set of candidate causative genetic variants – most commonly single base changes (Single Nucleotide Polymorphisms, or SNPs), inserted or deleted base pairs (collectively referred to as 'indels'), but also site-specific changes in DNA methylation (the epigenome) or changes in copy number of genes (copy number variation, CNV).
- Prioritised candidate variants are compared to data about known variant-phenotype associations that are catalogued in a variety of databases. This comparison can also be cross-species, as mutations in conserved gene sequence in animal or yeast models, for example, can give clues to similar associations in human diseases.
- Bioinformatics tools can identify which genetic variants identified in association studies are most likely to result in a dysregulated downstream effect and can be prioritised for further analysis; for example, a mutation in the promoter region of a gene can alter a transcription factor binding site on this promoter and induce or inhibit the expression of a downstream gene, leading to a change in the amount of a protein being synthesised.

Thereafter, causal links between identified genetic risk factors (variants) and phenotype observations must be empirically tested and validated before incorporating these findings into the provision of healthcare.

The clinical significance of common variants:

Technical advances coupled with substantial decrease in genotyping costs, have enabled investigators to:

- move beyond evaluating a few candidate variants in key genes
- conduct more comprehensive as well as exploratory evaluations of common genetic variations in candidate pathways
- perform large GWA studies in different populations from different ethnic groups
- refine risk assessment and genetic testing tools.

These advances allowed a rapid evolution of approaches for studying common genetic susceptibility factors. Available data has shown that most of the unresolved fraction of genetic susceptibility to hereditary diseases is likely to be explained by a polygenic model involving a combination of several common variants that have, individually, weak associations with the disease (Antoniou et al. 2004; Tyrer et al. 2006) but in a polygenic model may explain a considerable part of disease heritability.

GWAS and large-scale replication studies have identified several SNPs distributed throughout the genome that show strong evidence of association with hereditary diseases and also show a direct or indirect functional effect (Collins 2015). Some of

these variants are directly genotyped, while others are detected through imputation, a statistical method based on Linkage Disequilibrium (LD) analysis (Ohta and Kimura 1969). However, because of LD blocks, multiple SNPs may have strong LD correlations and therefore, may have equally strong evidence of being the causal variant. Because the pattern of LD often differs among populations, case-control studies from multiple ethnic groups and diverse populations (notably populations of East Asian and African ancestries) can assist with either determining the genotypic identity of causal variants or validating the functional role of common polymorphisms of known genetic identity. This will also help in implementing precise risk assessment tools and polygenic risk scores in order to screen and prevent several complex pathologies such as cancer and cardiovascular diseases.

5.3. Variant Classification and Clinically Actionable Genetic Variants

Different types of mutations have diverse consequences on the function of a gene and the protein it encodes, and therefore the organism phenotype. Variant classification is where the role of a DNA variant in causing a disease phenotype is determined. To apply this in clinical practice, multiple lines of evidence should be used to determine whether the variant is actionable (Carter and He 2016) but translating these advances into reality for improving healthcare outcomes depends essentially on our ability to discover disease- and/or drug-associated clinically actionable genetic mutations. Integration and manipulation of diverse genomic data and comprehensive electronic health records (EHRs). For some variants, tools can predict whether they are pathogenic or benign with reasonable confidence. The most explicable evidence is when the variant is predicted to have a functional effect on the protein (pathogenic), but other evidence can include whether the variant has been seen in other individuals with the same condition but not in individuals without it. Where there is no strong evidence of its role, the variant is usually reported as a Variant of Unknown Significance (VUS) (Hoffman-Andrews 2017).

SNVs (SNPs) or single nucleotide variants/polymorphisms can be silent – **synonymous** - or can cause a change in the amino acid – **non-synonymous**. Alternatively, the presence or absence of several nucleotides at a specific site results in **insertions** or **deletions** (**indels**) which often lead to loss of function of the protein. **Copy number variants** (CNVs) result from gene deletion or amplification. CNVs can be of various types; **large segment alterations** occur at the chromosomal level with regions that get translocated, inverted, deleted or duplicated with particularly serious consequences. Such events can at times bring about fusion of whole segments from unrelated genes into a single transcribed protein (https://tumorfusions.org).

Variants involved in some disease genes are well documented with many studies confirming their roles in pathological conditions. Information on the types of variants, their location, change caused in the corresponding protein structure/function and implication in disease is stored in various databases. There have been several efforts to harmonise the vocabulary for describing variants and their effects. With increasing demand for genetic testing, it is imperative to have a coordinated policy for referring to the type of genetic change that can cause disease for both diagnostic purposes as well as for treatment. The Global Alliance for Genomics and Health is developing standards for variant representation and for representing individual-level genotype-phenotype associations, known as Phenopackets (see: https://www.ga4gh.org/genomic-data-toolkit/).

The American College of Medical Genetics and Genomics (ACMG) guidelines provide 27 variant characteristics for determining whether a variant is pathogenic or benign. They classify the impact of a variant including: “very strong evidence of pathogenicity (PVS1), strong evidence of pathogenicity (PS1–PS4), moderate evidence of pathogenicity (PM1–PM6), supporting evidence of pathogenicity (PP1–PP5), stand-alone evidence of benign impact (BA1), strong evidence of benign impact (BS1–BS4), and supporting evidence of benign impact (BP1–BP6)” (Maxwell et al. 2016)methods for classification of variants resulting from this testing are not well studied. We evaluated the ability of a variant-classification methodology based on American College of Medical Genetics and Genomics (ACMG. These have been used to develop a systematic methodology for variant classification (Richards et al. 2015) sequencing technology has evolved rapidly with the advent of high-throughput next-generation sequencing. By adopting and leveraging next-generation sequencing, clinical laboratories are now performing an ever-increasing catalogue of genetic testing spanning genotyping, single genes, gene panels, exomes, genomes, transcriptomes, and epigenetic assays for genetic disorders. By virtue of increased complexity, this shift in genetic testing has been accompanied by new challenges in sequence interpretation. In this context the ACMG convened a workgroup in 2013 comprising representatives from the ACMG, the Association for Molecular Pathology (AMP).

Based on these variant classification guidelines, four main groups of genetic variants have been identified and are used in genetic testing:

- **Druggable variants:** genomic aberrations for which pharmaceutical agents are available or known to be in development.
- **Actionable variants:** genomic aberrations that may impact treatment prescribed for predictive or prognostic reasons or those that are known to have prognostic or diagnostic implications, including pharmacogenes.
- **Benign phenotype-associated variants:** genomic aberrations that correlate with one or several traits or diseases without showing any clinical impact.
- **Variants of Unknown significance (VUS):** A variation in a genetic sequence for which the association with disease risk is unclear. Also called unclassified variant, variant of uncertain significance and VUS.

Limited evidence of the clinical significance of VUS is one of the major challenges in the interpretation and the application of genetic data in healthcare. To facilitate the incorporation of these genomic data into clinical care, a more accurate review and interpretation of scientific literature by an expert committee is required to determine whether a variant is clinically relevant or actionable. This is time-consuming, and thus it is not feasible to routinely review the large number of variants identified from NGS data (Tavtigian et al. 2008). Tools such as Align GVGD (http://agvgd.iarc.fr/ (Mathe et al. 2006)), POLYPHEN-2 (http://genetics.bwh.harvard.edu/pph2/ (Adzhubei et al. 2013)); VEP (https://www.ensembl.org/info/docs/tools/vep/index.html (McLaren et al. 2016)), Mutation taster (http://www.mutationtaster.org/ (Schwarz et al. 2014)), and SIFT (http://sift.jcvi.org/ (Vaser et al. 2016)) can be used to predict variant effects. However, there are sometimes inaccuracies in these tools resulting in differences in predictions for the same variant, which makes clinical interpretation unclear. Efforts made by the "Evidence based Network for the Interpretation of Germline Mutant Alleles consortium" (ENIGMA) to reclassify variants of uncertain significance in BRCA1 and BRCA2 genes, using clinical and functional assays (such as genome editing), represent an excellent initiative to overcome this challenge (https://enigmaconsortium.org/, (Spurdle et al. 2012)).

5.3.1 Determining and validating deleterious variants

Selection of variants can be based on biological plausibility which may have been demonstrated in previous publications, based on knowledge of gene functions and their relation to the disease or phenotype, or known to be involved in pharmacokinetic and pharmacodynamics pathways (available in PharmGKB). This is known as the candidate gene approach (CGA), in which dozens to thousands of genetic variations within one or more genes are genotyped or sequenced. However, this approach requires prior knowledge about gene function and its relation to a disease or its interaction with a drug or pharmacokinetic pathway (http://www.encepp.eu/standards_and_guidances/methodologicalGuide10_3.shtml). Such an approach would be appropriate for monogenic disorders.

Where there is no prior information on candidate genes or variants, or for complex traits, a hypothesis-generating approach can be used, which looks at variants across the whole genome to identify possible candidates. This is the approach using in genome wide association studies, often comparing cases and controls. In GWAS or pharmacogenomic studies the difference in frequency of variants between cases and controls or those who respond positively to drugs versus those who don't enables the identification of possible genetic determinants. This approach relies on large sample sizes to achieve statistical significance and may not uncover all associated variants. Also, since many variants are associated through linkage disequilibrium, a variant identified through genome-wide approaches may not be the true causal variant, but rather linked to the true variant. Nevertheless, the approach is able to identify novel variants or genes not previously known to be associated with the disease or response. (http://www.encepp.eu/standards_and_guidances/methodologicalGuide10_3.shtml).

Risk loci identified by GWA studies may result in a change in the protein structure/function if in a coding region or regulate gene expression or splicing if in a regulatory region or intron, respectively. When variants in non-coding regions are found to have a regulatory function, long-range interactions with genes have been described, i.e., the SNP forms a regulatory loop with genes involved in pathways which are relevant to the disease. Chromatin organisation can explain the effect of some variants on genes implicated in metabolic pathways. The same gene may have multiple causative variants for the same disease, known as allelic heterogeneity, which can impact molecular diagnostics. Recommendations for a

specific variant or set of variants might be required for testing in diverse populations. For example, there are > 2000 mutations in the gene for Cystic Fibrosis Trans-membrane conductance Regulator (CFTR), and patients of African ancestry have a lower frequency of the F508 deletion (deletion of a phenylalanine), which has a high frequency in those of European ancestry. Therefore, it is important that GWAS results are replicated in other cohorts and validated, preferably in functional studies, to determine their applicability to other populations and to identify the biological mechanism (http://www.encepp.eu/standards_and_guidances/methodologicalGuide10_3.shtml).

5.3.2. Existing resources for variant actionability and pathogenicity assessment

So far, many thousands of genetic variants have been identified as associated with hundreds of human traits and diseases, although defining this set of aetiological variants remains challenging (Sun and Yu 2019). Several mutation-centric databases such as HGMD and NEMDB are used to assess mutation pathogenicity and to interpret next-generation OMICs data. A comprehensive table documenting these resources can be found in Appendix 2. Additional databases such as OMIM contain information on disorders and genes focusing on the relationship between phenotype and genotype. Others such as ClinVar, DisGeNET, PharmGKB and GTR provide information on variant pathogenicity, genotype-phenotype correlations, variants related to pharmacogenetic responses (pharmacogenomics is discussed in more detail in section 5.4) and gene-drug associations. ClinVar is a public archive of correlations between human variations and phenotypes. To address the needs of the medical genetics community, ClinVar provides accurate data on variant actionability and pathogenicity assessment. Data provided by ClinVar is evaluated, interpreted and archived by recognised expert panels based on practice guidelines. DisGeNET is another platform providing publicly available data on genes and variants associated with human diseases. It integrates data from expert curated databases such as the GWAS Catalog and others. This data is then homogeneously annotated to assist the prioritisation of genotype-phenotype relationships. Disease-centric databases are also publicly available, for example oncology databases such as TCGA and OncoKB that contain multi-omics data related to cancer. The ACMG (https://www.acmg.net/) provides a list of potentially actionable variants though not all are applicable to African populations and some relevant African variants are missing. Finally, the Genomics England PanelApp (https://panelapp.genomicsengland.co.uk/) is another publicly available knowledge base that enables the creation of and provides access to virtual gene panels for specific human disorders. It provides an opportunity to standardise these gene panels, and a platform for experts to reach a consensus on gene-disease association based on sufficient evidence. Australian Genomics has also adopted PanelApp for their genomic medicine programme.

Critically evaluating variant evidence will help address several issues related to variant classification and to the use of the right genetic variants for the right diagnostic or therapeutic decision. Initiatives such as the Clinical Genome Resource (ClinGen) include more than 1125 scientists from more than 230 institutions worldwide (www.clinicalgenome.org). ClinGen facilitates reviewing for organisations that provide practice guidelines in ClinVar. So far, scientists from only three African countries are involved in the ClinGen expert Panel (Tunisia, Egypt and South Africa). Before defining a variant's role in disease, researchers and clinicians together must assess the gene level evidence (Strande et al. 2017) the number of reported gene-disease relationships has rapidly expanded. However, the evidence supporting these claims varies widely, confounding accurate evaluation of genomic variation in a clinical setting. Despite the critical need to differentiate clinically valid relationships from less well-substantiated relationships, standard guidelines for such evaluation do not currently exist. The NIH-funded Clinical Genome Resource (ClinGen. The Gene Curation Coalition (GenCC) has been set up to assess gene-disease validity and genotype-to-phenotype curation and to decide which genes should be included on clinical test panels or interpreted for different indications (http://thegencc.org). Involving African partners in these initiatives will be essential to ensure applicability for both African and non-African patients; with SNP frequencies differing widely, causal variants in African populations may differ from other populations for the same diseases.

5.3.3. African resources and data for disease-causing variants

The primary source of genotype-phenotype information is research outputs. The number of genomics studies in Africa is increasing, particularly through large consortia such as MalariaGen and H3Africa. Where these and previous studies are published, genotype-phenotype data can be collected from the literature and some provide controlled access to their phenotype and/or genotype data in public repositories. Examples of some African-specific variants extracted from the literature are listed in the Box 8: *Disease-causing variants relevant in the African context.*

Small-scale African genotypic data can also be collected from public data resources, including the variant databases listed in Appendix 2. Other resources include the GWAS catalogue, which curates published GWAS papers for genotype-phenotype associations from the scientific literature; dbSNP, which includes data on single nucleotide polymorphisms (SNPs) and their frequencies and genotypes in different populations; NCBI Gene, which includes gene-specific data, such as nomenclature, chromosomal localisation, gene products, phenotypes, and links to related resources; and eQTL data from the GTEx programme, which archives and displays associations between genetic variation and high-throughput molecular-level phenotypes. However, as mentioned previously, African representation in these resources is low. Additionally, limitations in the curation of the source of data means that finding and retrieving African-specific data is challenging and time-consuming.

Work is currently underway by the H3ABioNet Genomic Medicine project to address this by developing an African genomic medicine and pharmacogenomics portal, which collates and curates pharmacogenomics and other genomic medicine metadata specific to African populations (www.h3abionet.org). The portal also provides links to existing GM implementation tools and resources in order to support such initiatives within Africa. Another project that will contribute to knowledge on African-specific actionable variants is the H3Africa funded Individual Findings in Genetics Research in Africa project (IFGeneRA, NIH grant no. NIH/U54-HG009790-01), that aims to progressively build an evidence-based context- and country specific set of policies on which genetic results may need to be fed back, when and how. It combines methods from medical geneticists, bioethics, genetic counselling, health economics, bioinformatics, and social science to explore the actionability of newly identified genetic variants that are specific to African populations.

5.3.4. Identification of new African-specific disease-associated variants

Currently the majority of variants in most databases have been identified in populations of European

Disease-causing variants relevant in the African context

- Some mutations in the CD36 receptor gene have been found to be associated with the severity of malaria. The SNPs T1264G in exon 10 (rs3211938) and G1439C in exon 12 result in truncated proteins that were found to be expressed at high levels in patients with severe malaria in The Gambia, Tanzania, and Kenya (de Mendonça et al. 2012)).
- Variants in the Lactase gene (LCT) have been identified as associated with lactase persistence in eastern Bantu Speaking Populations (BSP). These variants are highly frequent in Bakiga eastern BSP (30%) and rare in western BSP (1%) which suggest acquisition of the lactase persistence allele from local eastern African populations ((Ingram et al. 2009).
- Two coding mutations in the C-terminus of Apolipoprotein 1 (APOL1) are associated with certain types of kidney disease with different effects in different populations, though this protein has also been known to protect against African trypanosomiasis. The alleles relating to kidney disease are G1, with two amino acid substitutions (S342G and I384M) and G2, a two amino acid deletion (del388N389Y). The risk alleles, which are in the same functional domain, have only been seen in individuals with recent African ancestry. (Friedman and Pollak 2016).
- Pathogenic mutations in BRCA1 and BRCA2 genes are considered as druggable mutations for breast, ovarian and pancreatic cancer patients, as there is targeted therapy (PARP inhibitors) that may be given to patients that are carriers of BRCA mutations. Several mutations have been identified in the BRCA1 and BRCA2 genes that seem to be specific to African populations such as the c.211 DupA mutation on exon5-BRCA1 that seems to be specific to the Tunisian population *(Laraqui et al. 2015)*

ancestry while the wider human genomic diversity is not represented. This impacts on the translation into practical applications for diagnostics and treatment, because using uninformative or inappropriate variants in specific populations for screening purposes can lead to misdiagnosis of diseases and inappropriate treatment (Manrai et al. 2016). Additionally there are nearly as many ancestries within Africa as in the rest of the world combined (Shriner et al. 2014), and the resulting shorter haplotypes among Africans represents a real challenge for the replication of genetic studies in African sub-populations. *Box 8. Disease-causing variants relevant in the African contex*

As discussed in section 3, to identify new African specific variants, large-scale genotype data on African samples should be generated using whole genome sequencing, whole exome sequencing or other large-scale analysis. Genotypic data can be extracted from the whole African genomes that have been sequenced in the framework of projects such as 1000 Genomes, H3Africa, TopMed and others. However, most of the data are not publicly available and require access requests to data access committees. Sometimes summary data are openly available, for example through gnomAD, which aggregates frequency data for multiple populations.

Figure 5 provides a flow chart showing the high-level steps that should be followed to identify novel African specific variants from new sequence data and the steps required to assess or validate the biological function or the potential actionability of these candidates. The total set of variants generated by sequencing and genotyping projects is filtered against existing genetic variants reported in the public database dbSNP (that includes data from the 1000 Genomes, GO.ESO, ExAc, gnomAD, TOPMED and HLI). The novel variants are those that have not been reported previously in dbSNP. These newly identified variants should be assessed for their functional effect using several *in silico* as well as *in vitro* functional assays such as genome editing. Actionable variants should then be investigated further by a recognised pharmacogenetic or other genetic testing and evaluation framework based on international recommendations to assess their clinical actionability.

The identification of novel variants by sequencing additional African genomes from different ethnic backgrounds will help to increase variant re-classifications. For example, in a recent study, Manrai et al. showed that the lack of access to non-white population data resulted in the misclassification of benign variants as pathogenic and therefore in the misdiagnosis of patients with hypertrophic cardiomyopathy (Manrai et al. 2016) risk stratification for hypertrophic cardiomyopathy has been enhanced by targeted genetic testing. Using sequencing results, clinicians routinely assess the risk of hypertrophic cardiomyopathy in a patient's relatives and diagnose the condition in patients who have ambiguous clinical presentations. However, the benefits of genetic testing come with the risk that variants may be misclassified.\nMETHODS: Using publicly accessible exome data, we identified variants that have previously been considered causal in hypertrophic cardiomyopathy and that are overrepresented in the general population. We studied these variants in diverse populations and re-evaluated their initial ascertainments in the medical literature. We reviewed patient records at a leading genetic-testing laboratory for occurrences of these variants during the near-decade-long history of the laboratory.\nRESULTS: Multiple patients, all of whom were of African or unspecified ancestry, received positive reports, with variants misclassified as pathogenic on the basis of the understanding at the time of testing. Subsequently, all reported variants were recategorised as benign. The mutations that were most common in the general population were significantly more common among black Americans than among white Americans (P<0.001. It should be noted that some variants (mainly the newly identified variants) may only have been observed in a single individual. Choosing not to submit this variant to ClinVar or similar open access databases until other submitters accumulate additional cases, may result in the variant never being shared or in its erroneous classification. These ultra-rare variants are those most often associated with disease pathogenicity and therefore the most valuable to be shared. However, several ethical considerations have to be outlined when re-classifying and sharing these sensitive clinical genomic data for broad access (Azzariti et al. 2018). Reducing patient re-identification risk and respecting an individual's right to privacy and autonomy should be considered as priorities when dealing with clinical genomic data. This can be managed by only reporting the minimal phenotypic information required for another healthcare professional to understand the case. There are resources that enable this, such as MatchMaker Exchange (https://www.matchmakerexchange.org), mostly for rare diseases, and the Beacon project (https://beacon-network.org/).

Reducing the burden of disease relies on availability of policy-driven disease-specific practice guidelines that are based on updated research results and sufficient evidence. 30 clinical practice guidelines, for example for HIV in adults,

malaria, pre-eclampsia, diarrhoea in children and hypertension in primary care, have been reported in 13 different African countries (Kredo et al. 2012). Most of these guidelines have been implemented by African knowledge societies. At the international level, the USA Food and Drug Administration (FDA) and the European Medicines Agency regularly update guidelines for the use of genome-based therapies in healthcare.

5.4 Pharmacogenomics: Linking Genetic Variation to Treatment Phenotype

Phamacogenomics is the study of how genetic variations in key genes result in differential response to drugs. It is presented here as a special case of genotype-phenotype association that is highly relevant to GM and provides many opportunities for rapid translation to clinical practice. There are currently 19 000 drugs approved by FDA for treating various diseases. These have been tested in clinical trials on selected participants for their efficacy and possible side effects or adverse events. In most cases, however, this pre-selection is based on testing on a narrow range of individuals mostly of European ancestry. There is accumulating evidence that the same drug does not necessarily have comparable effects on people of other genetic backgrounds.

Several studies have confirmed the role of more than 360 genes implicated in drug metabolism, which include membrane transporters, enzymes, transcriptional regulators and different types of receptors (Pharmgkb.org; (Man et al. 2010)). *(See Box 9: ADME (absorption, distribution, metabolism and elimination) genes).* Variations in these genes have been found in coding regions, resulting in loss of function, and in non-coding, regulatory regions, UTRs or introns which can result in altered gene expression. The effect of mutations on drug pharmacokinetics or pharmacodynamics must first be established to include them as pharmacogenes. This has been possible through the use of computational tools for predicting changes in protein structure/function (Klein et al. 2019)distribution, metabolism and excretion (ADME).

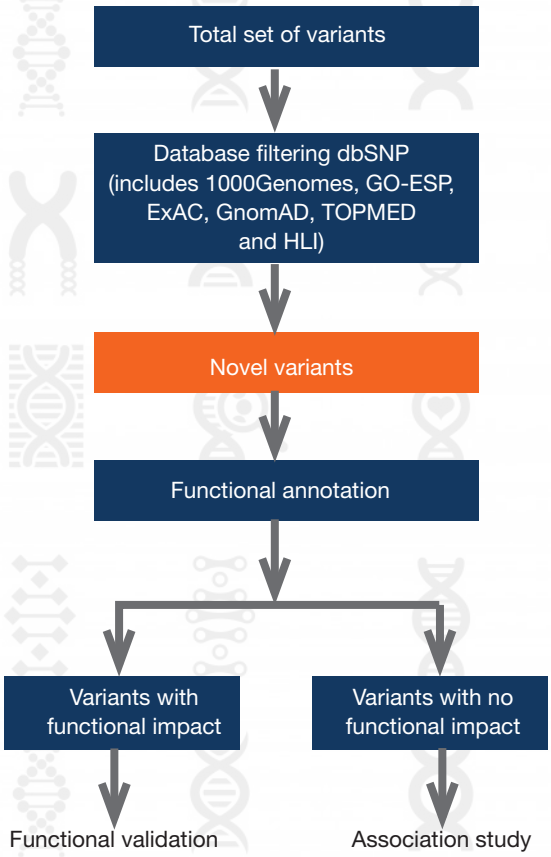


Figure 5. Flow chart showing the high-level steps that are usually followed to identify novel variants from e.g. a GWAS or genome sequencing and predict their function. African populations have many rare variants which should also be identified in variant analysis

Box 9. ADME (absorption, distribution, metabolism and elimination) genes

ADME (absorption, distribution, metabolism and elimination) genes

CYP P450: Cytochrome P450 proteins are a large family of hydrolytic enzymes which metabolise >80% of medical drugs, affect lipid metabolism and mainly localise in the liver. 60 highly polymorphic CYP genes have varying enzymatic activity. Slow drug metabolism, means persisting active drug, requiring smaller doses; faster drug metabolism requires a higher dose (Lynch and Price 2007). CYP P450 activation increases drug metabolism in cancer cells leading to treatment resistance (Rochat 2005). Anti-malarials such as Artemisins also get metabolised by CYP enzymes.

CYP2C19: CYP P450 family 2 subfamily C, polypeptide 19 enzyme metabolises antidepressants, selective serotonin reuptake inhibitors citalopram and sertraline, voriconazole, and antiplatelet agent clopidogrel; the latter is a pro-drug that must be activated by CYP P450. Well-characterised variants include CYP2C19*2, an exon 5 (G-->A) mutation introduces an aberrant splice site and loss of function, leading to drug accumulation; and CYP2C19*17 – a regulatory region (C-->A/(C-->T) mutation causes increased expression and rapid drug degradation. CYP2C19 alleles also affect anti-platelet agent clopidogrel, a pro-drug that must be activated by cytochrome P450. CYP2C19 loss of function alleles occur in 35-45% with African ancestry (Man et al. 2010).

CYP2D6: This protein metabolises around 25% of currently prescribed drugs, including various antidepressants, neuroleptics, beta-blockers, opioids, antiemetics, and antiarrhythmics. Distribution of loss- and gain-of-function alleles of CYP2C19 and CYP2D6 across Europe indicate inter-ethnic differences across the continent with a North-West to South-East gradient (Petrovic et al. 2020).

CYP2C9: Warfarin, an anticoagulant, effectively reduces clot formation risk in cardiovascular disease patients, but can lead to severe bleeding, and requires careful dose titration according to age, sex and genetic factors. CYP2C9 codes for an enzyme that metabolises warfarin with two variants (*2 and *3) exhibiting reduced (40% - 90%) activity. Individuals with these variants are at higher risk of bleeding and require careful monitoring (Niinuma et al. 2014). Warfarin targets vitamin K epoxide reductase, VKORC1, and its variants also affect warfarin activity (Limdi et al. 2008, 2010). A regulatory region mutation in VKORC1 (1639G>A;rs9923231 reduces enzyme production so that less warfarin is needed. Several studies identified alleles of CYP2C9 and of VKORC in African-Americans that impact warfarin dosing for optimal treatment (Limdi et al. 2008; Perera et al. 2011).

UGT1A1: UDP-glucuronosyltransferase mediates glucuronidation of bilirubin leading to detoxification and elimination; and also metabolises irinotecan, an anti-cancer agent. UGT1A1 variants with reduced activity decrease irinotecan inactivation, leading to drug toxicity (Lankisch et al. 2008) and severe neutropenia, which can be lethal.

TPMT: Thiopurine S-methyltransferase metabolises thiopurines commonly used for cancer treatment and immunosuppressants to prevent graft rejections in renal transplants. Thiopurines include thioguanine, mercaptopurine, and azathioprine - all nucleotide analogues interfering with purine biosynthesis. TPMT metabolises Azathioprine to the active compound 6-mercaptopurine (6MP). Key TPMT variants have been described in different populations: TPMT*2 (c.238G>C), TPMT*3A (c.460G>A and c.719A>G), TPMT*3B (c.460G>A) (Wang et al. 2010), TPMT*3C (c.719A>G). TPMT deficiency reduces 6MP production, and shunts production to an alternative metabolite, 6-thioin-5-monophosphate which can cause pantocytopenia (Konstantopoulou et al. 2005). TPMT*3A is most common in Caucasians, and TPMT*3C in African populations.

SLCO1B1: This member of the organic anion transporter family (OATP) is predominantly expressed in the liver and encodes a membrane anion influx transporter for many compounds including xenobiotics. About 190 variants have been reported (http://www.hapmap.org). The SLCO1B1 T521C forms (*5, *15, *16, *17) reduces processing of the cholesterol-lowering drug Atorvastatin (Pasanen et al. 2007) and the rs4149056 (c.521T>C) variant - in 3% of people of African descent -is associated with myopathy, affecting those on long-term statin therapy (Stewart 2013).

The efficacy of any drug depends on its optimal concentration at the site of action and on its interaction with target molecules. This in turn relies on the metabolism of the therapeutic agent and the kinetics of the enzymes, which modulate them. Germline genetic variations can have an effect on the pharmacokinetics or pharmacodynamics and the response to drugs (Alfirevic and Pirmohamed 2017). The need for pharmacogenetic screening in clinical laboratories has become an essential test to assist physicians in their prescriptions. This is based on substantial evidence that relates adverse reactions to drugs or their metabolism due to the presence of genetic variants. Applying such screening more broadly will ensure correct use of medications and avoid unnecessary adverse effects in susceptible individuals.

5.4.1 Pharmacogenomics resources

Data on ADME genes and pharmacogenomics can be found in resources from the Pharmacogenetic Variation Consortium (PharmVar) and PharmGKB, which standardise pharmacogene variation classification and nomenclature for researchers and clinical professionals (Figure 6). PharmGKB (https://www.pharmgkb.org/) is an important resource for curated pharmacogenes, pharmacogenetic variants, drugs and diseases. The Clinical Pharmacogenetics Implementation Consortium (CPIC) has published guidelines (https://cpicpgx.org/publications/) for therapeutic recommendations concerning several drugs such as warfarin, clopidogrel, efavirenz and tamoxifen amongst others. Many of the drugs have FDA-approved risk labels. The Ubiquitous Phamacogenomics Consortium has been set up in Europe to investigate the genetics variants of 13 phamacogenes in participants from seven countries. The outcome of the study will produce genetic information that physicians can use to prescribe the corresponding drugs, and this will lead to more accurate doses and effective treatments. A recent initiative has seen the formation of the African Pharmacogenomics Consortium (Dandara et al. 2019) which shows the commitment of the research community to undertake concrete actions to improve treatments based on genetic information. One of the objectives of this consortium includes "collaborative strategies for the discovery and curation of genetic variants that are of pharmacogenomic importance among African populations" (Mpye et al. 2017; Dandara et al. 2019). Recently, a review on available African pharmacogenomics studies has been published by the H3ABioNet Precision Medicine working group (Radouani F. et al, Personalised Medicine, 2020).

5 .4.1.1 Recommendations for African Pharmacogenomics

Existing information on pharmacogenes can be applied to provide better care to patients in Africa by incorporating genetic testing. In addition, the lack of adequate genomic data on the large number of ethnic groups in Africa must be addressed with a consolidated approach to research in this field. (Mpye et al. 2017) argue for the need to generate pharmacogenomics data for the diverse African populations and the urgent establishment of clinical trials for new drugs in Africa. There is a large amount of data pointing to the importance of genotyping pharmacogenes for more precise drug dosage, based on the patient's allelic profile. Many countries have implemented strategies for doing so for the benefit of the patients and for significant reduction in costs of treatments. Though several clinical centres in Africa have the facilities for genotyping patients before deciding on the use and dose of drugs to be administered, there needs to be broader use of screening prior to treatment or as soon as an adverse drug reaction has been detected. These should also be reported to the pharmacovigilance database VigiBase (https://www.who-umc.org/vigibase/vigibase/). In addition, we recommend the following:

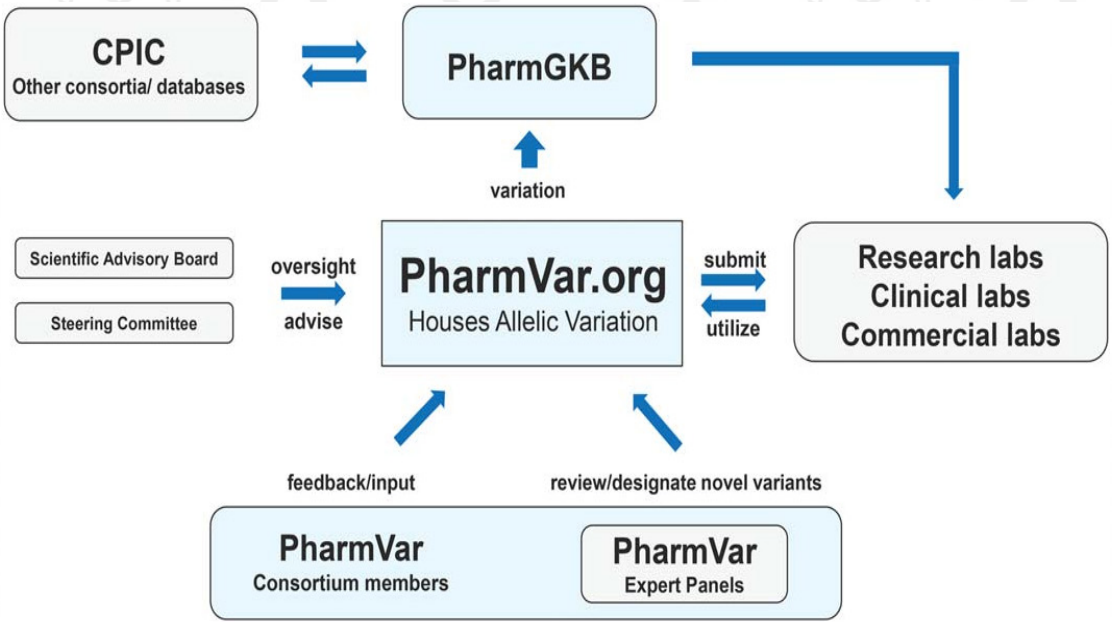


Figure 6. PharmVar Consortium and Database. Source: https://www.pharmvar.org/

- Genetic testing for known alleles of pharmacogenes should be implemented in Africa. This would require only DNA extraction and PCR facilities.
- More specifically, (Baker et al. 2017) make recommendations for the genotyping of alleles in genes that are associated with treatments for malaria, trypanosomiasis, HIV, Lassa fever and tuberculosis.
- Although some data is available on African genotypes, which can be used for drug prescriptions, more information is needed for the diverse populations on the continent.
- It is necessary to generate more genotype data linked to drug response information in pharmacogenomic studies and clinical trials.
- Countries should implement pharmacovigilance programmes to better track adverse drug responses.

5.4.2 Data Integration

Once large-scale multi-OMICs data are generated and complex environmental, clinical and epidemiological data are collected, there is a need for efficient data integration to ensure translation into health benefits. Several strategies could be used in data integration (Figure 7). Infrastructure-as-code strategy is a method for formally defining coded instructions on how a set of computers should be provisioned and managed. By employing infrastructure-as-code strategies to define computers, adopters have the leverage to deploy as many cloned or similar computers as needed to optimise data integration, costs and utilisation (Frey 2018). Other externally validated methods that are used in data integration include adjustment for variations that rely on deep knowledge of the genetic landscape of the studied population. However, only 3% of GWAS have been performed in Africans (Nordling 2017) witnessing the large inequities in genetic data availability between different human populations. Data integration strategies such as advanced computational tools, machine learning and artificial intelligence are more commonly implemented in high income countries (HICs) than in Low and Middle Income Countries (LMICs) (Drake et al. 2018). Therefore, consolidating collective efforts to implement these standardised approaches to data integration can accelerate the application of GM in LMICs for precise risk assessment, accurate prognosis and personalised therapeutic decisions.

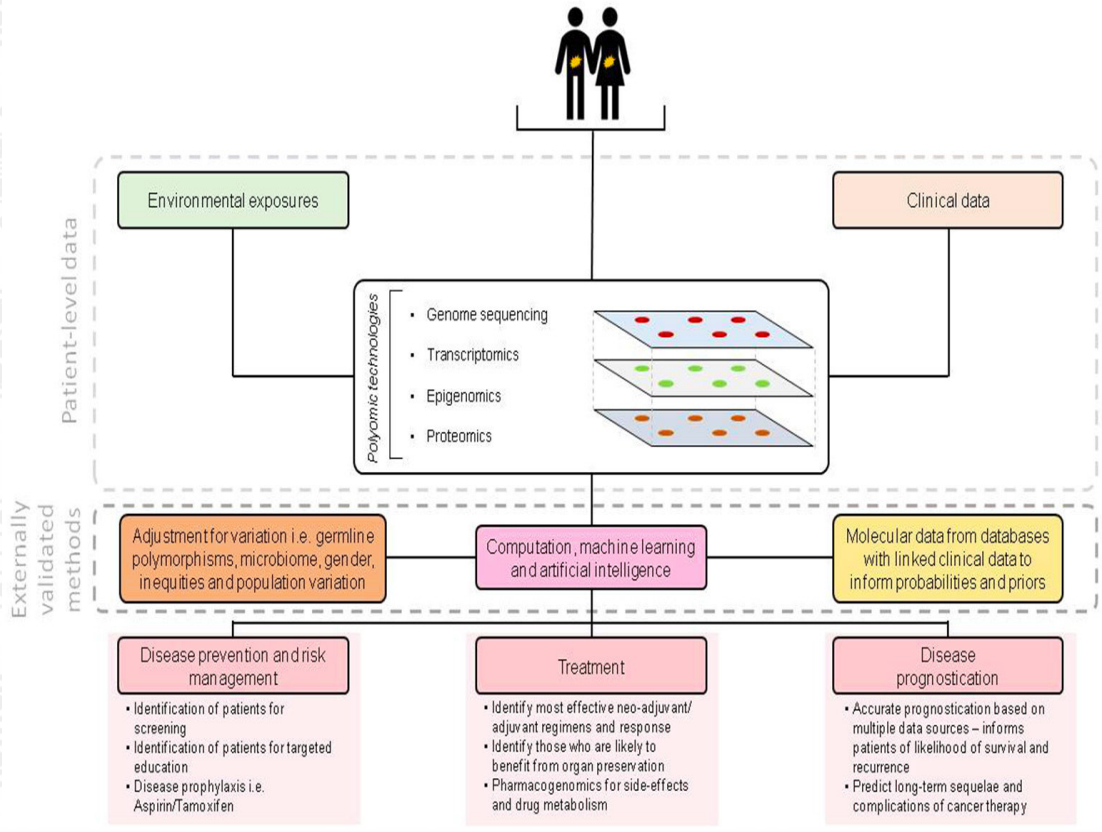


Figure 7. Phenotypic, environmental and multi-omics data integration. Different levels of data management include: correlating phenotype (clinical, environmental and epidemiological) and genotype (genomic, transcriptomic, epigenomic and proteomic) data, implementing personalised screening, and a precise therapeutic decision. Data integration requires archiving and distributing primary association data, and with external validation using available molecular databases and additional methods such as machine learning and artificial intelligence. Source: Drake et al. (2018).

5.4.1.2 Recommendations for data integration

- Harness existing genotype-phenotype data to prioritise variants for follow-up and study design for further research.
- Generate new data for African populations through large-scale genetic studies on larger African cohorts, particularly under-represented African subpopulations. For novel variants, use orthogonal methods to assess their functionality, biological effect and clinical actionability.
- Establish an African expert review panel to work with existing international panels to review novel variants and evidence for actionability.
- Establish African guidelines on the use of genomics data in clinical practice and in GM research in Africa. Joint efforts can be made by the African Society of Human genetics, the Alliance for Excellerating Science in Africa and other African regulatory bodies to set these guidelines.

Section 6. ELSI and data governance

6.1 Summary

GM has ethical, social and legal implications that need to be addressed simultaneously as the required technologies are in the early stages of development - for instance standardising and developing protocols for collection, storage and sharing of clinical data. Participants, who in the healthcare sector are patients, need to understand the risks and benefits of participating in GM-based research and clinical care, which means researchers will require a rigorous process of using participant information and recording their informed consent. It is critical to find ways to protect participants' privacy and the confidentiality of their health information. This is easier to do when the GM initiative is driven by government and integrated into national health systems, For clinical implementation consent should be aligned with current medical practice. Ethical and governance issues also address questions about cost, quality control of data, resource allocation, control and ownership of data, and the role of genetic stratification in treating and monitoring patients.

6.2 Introduction - Ethics and Governance in Precision Medicine

GM investigates individuals or groups of patients based on their specific characteristics in order to develop targeted interventions, raising ethical issues about privacy, informed consent, equitable benefits and social justice. To implement GM in routine healthcare, healthcare providers will need to improve their knowledge of molecular genetics and biochemistry. They will increasingly need to interpret the results of genetic tests, to determine how the information can inform treatment or prevention approaches and convey this information to patients. Thus, sound ethics practices are essential for research and application of GM approaches.

GM as a field encompasses both research undertaken to develop diagnostic, therapeutic, and prognostic tools that are based on patient variants or genomic profiles as well as the provision of care based on the sequencing and analysis of individual patients' DNA to provide individualised preventative, diagnostic, prognostic or therapeutic approaches. While GM approaches require data, in practice these data can be ethically challenging to obtain. There are important ethical issues that need to be considered in the collection, storage, use and sharing of such data. At times, data are collected at financial or other cost to patients, which raises questions about how to balance the benefits and burdens of implementing a more personal approach to their care (Batten 2018).

Genomic medicine is based on the idea that medical conditions, disease susceptibility or response to treatment is caused, regulated or influenced by genetic factors, whereas PM also takes into consideration patients' environment and lifestyle in addition to their genetic profile (Gameiro et al. 2018). Governance of data collected for GM research and practice requires extra protections and considerations as it usually links two separate types of data (genomic and clinical) that are each highly sensitive in their own right: Genomic data is uniquely identifying of an individual and can encode information about all current and future characteristics of that individual. By its very nature, genomic data cannot be de-identified and can always be linked back to an identified individual (Gymrek et al. 2013; Erlich and Narayanan 2014; Erlich et al. 2018). It is likewise not possible to predict future uses of genomic data as scientific understanding of DNA and the human genome continues to evolve. Clinical data contain personal information about an individual's health status, including information that could result in social, economic or community harms (for example, the HIV data breach in Singapore) (Leyl 2019).

Fundamental tenets of ethical research include balancing beneficence and potential harms and ensuring justice and equitable practices. For GM, this means weighing up the risks to the individual and community - e.g. of data breach, stigmatisation, receiving unwanted health news - and potential benefits for individuals and populations of improved healthcare protocols and better health outcomes. Equitable and just research requires that all individuals are equally able to participate in research, and that all populations are represented in the global health research effort: the tenet of justice demands that such research and translated benefits are equal for all sectors of the community. Data governance practices must ensure that the rights to privacy and confidentiality are upheld, and consent processes must ensure the right to participate or decline participation are similarly respected.

6.3 Challenges for Ethics and Informed Consent for Genomic Medicine

6.3.1 Potential for community level harms

African populations have a wide range of genomic diversity, meaning that small population groups are sufficiently genetically distinct to be identifiable and geographically located. This increases the risk of health-related discrimination/stigmatisation for communities and easily-identifiable population groups that may be associated with particular aetiological variants: even if an individual's status is not known for a particular variant, if the allele frequency for a deleterious variant is known to be high for their population group of origin, they may be subject to discrimination based on a perceived elevated risk for all members of their population (Tiffin 2019) . Examples include founder effects for monogenic diseases such variegate porphyria in South Africa (Warnich et al. 1996) and familial hypercholesterolaemia in several countries (Kusters et al. 2011), as well as polymorphic variation in pigment genes associated with vitamin D deficiency linked to many NCDs on a global scale (Datta et al. 2019). Data sensitivity for small population groups or ethnic minorities cannot thus be compared with data protection for other vulnerable groups, in which the patient retains the right to privacy without third party disclosure in a balance with population-level and public health benefits, equitable healthcare and social justice. This balance for community and individual-level benefits and potential harms should be agreed between the patient and the healthcare professional, as well as through wider community engagement, which is crucial in ensuring that not only are patients' rights are respected, but important data also being used for the betterment of the community. For the clinical use of data, governments need to ensure that risks of stigmatisation or discrimination are not elevated through the provision of GM-based clinical care.

6.3.2 Equitable access to healthcare

In many parts of Africa, individuals experience limited access to healthcare, driven by socioeconomic factors, poor infrastructure and failure of governments to provide adequate healthcare. This raises ethical questions around using GM approaches where basic healthcare needs are not met, and therapeutics for particular genetic profiles may not be available and/or affordable for a large proportion of the population. GM research can, however, also facilitate improvements in therapeutic approaches at a population level. This has been demonstrated through recommended protocols for people living with HIV on ART in Africa (Warnich et al. 2011; Masimirembwa et al. 2016). Furthermore, the long-term benefits of improvements in patient outcomes can in some cases warrant the burden of introducing new GM approaches within the context of insufficient healthcare provision.

Socioeconomic vulnerabilities do require community engagement and provision of participant information to be carefully thought through and properly implemented. Intensive field work might be required to assess the level of understanding of participants prior to recruitment and requesting informed consent for study participation, as the tools devised for consenting must fit the needs of all stakeholders. Global policy for GM is lacking, and policies from non-African countries might not be adaptable to the Continent: we need to consider the precise needs of our population when designing community engagement programmes, setting policy and drafting guidelines for ethics and consenting. Such needs include drafting of an African treaty on research integrity, training in working with vulnerable populations, and practical policies to address digital disparity.

6.3.3 Increasing use of digital media for healthcare

Wide-spread access to mobile phones has raised their profile as tools for improving patient-provider communication, access to health services and information and data collection. There is a potential for data misuse, however, through the combination of sensitive health data with digital platforms that are well suited to replication and dissemination of datasets. With a personal computer or mobile device and Internet access, it is possible to copy and disseminate huge datasets almost instantaneously, increasing the risk of inappropriate data sharing, and making it harder to contain or reverse data breaches. There may be serious consequences for digital datasets because once shared, it is almost impossible to track down or delete copies of those data. This is exacerbated by the complexity of the data flow which involves multiple stakeholders and potential points of exposure. Data moves from individuals to data consumers, via data collectors, through mobile devices, interoperability layers and intermediate databases, to the databases where the data are finally stored. The potential for unconsented commodification of the data, whereby individuals cannot control or access how their data are being shared, reused or commercialised, poses a significant risk. Whilst this is a challenge in data collected in the context of medical research, it should also be considered in the context of electronic medical records collected en masse

by governments and insurance companies. These risks are particularly acute in low-resource settings with deep intersectional inequalities, and where governments may be lagging behind in implementing data protection policies and regulatory oversight to ensure the protection of personal information. Governance frameworks need to be established to ensure participant data are adequately protected, as shown in Figure 8 and described in (Tiffin et al. 2019). A concerning example is the proliferation in LMICs of use of the WhatsApp chat application for sharing patient information between clinicians without informed consent or protection of patient confidentiality (Mars and Scott 2016). This highlights the need for improved governance and ethics practices, convenient and secure internal platforms for efficient and appropriate sharing of patient data.

6.4 Personal Protection and the Legislative Environment in Africa

In Africa, there is disparity in the contents of legislation dealing with the protection of data of patients. Examples include the Data Protection Act of Mauritius, which is in line with current relevant international standards; the Protection of Personal Information Act (POPI Act) in South Africa (Information Regulator South Africa, 2013), which will come into full force in 2020 and the Data Protection Bill of 2018 in Kenya which includes medical data. Such Acts are similar in nature to the European Union's General Data Protection Regulation 2016 (GDPR) (Marelli and Testa 2018), which sets a standard for the protection of personal information. In some cases medical information may not be included in the general data regulation, for example the Data Protection Regulation in Nigeria, but similar to many other African countries, confidentiality of health data and the right to privacy around health information is instead enshrined in the National Health Act (Oloyede 2018).

The status of data protection in Africa as defined in a whitepaper by (Deloitte 2017) can be mapped as shown in Figure 9, and indicates that, at the time of writing, there is no unified approach to personal data protection across the African continent, with some countries having comprehensive personal data protection legislation in place whilst others have no legislation or constitutional protection. These data governance standards are be required to ensure that sensitive data are not breached or misused by governments, cybercriminals, insurance companies, medical practitioners, or other parties that may seek to use genomic and health data to their own advantage without consent or ethical clearance to do so.

6.5 Ensuring Research Integrity in Genomic Medicine

Research integrity is defined by the US NIH as "the use of honest and verifiable methods in proposing, performing, and evaluating research, reporting research results with particular attention to adherence to rules, regulations and guidelines, and following commonly accepted professional codes or norms" (https://grants.nih.gov/policy/research_integrity/what-is.htm). In undertaking GM research and implementation, research professionals must continue to uphold the principles of research transparency, data and research quality, research veracity, and data sharing where appropriate - whilst simultaneously respecting participant rights to choice, autonomy and privacy. Undertaking high quality GM research in African countries can

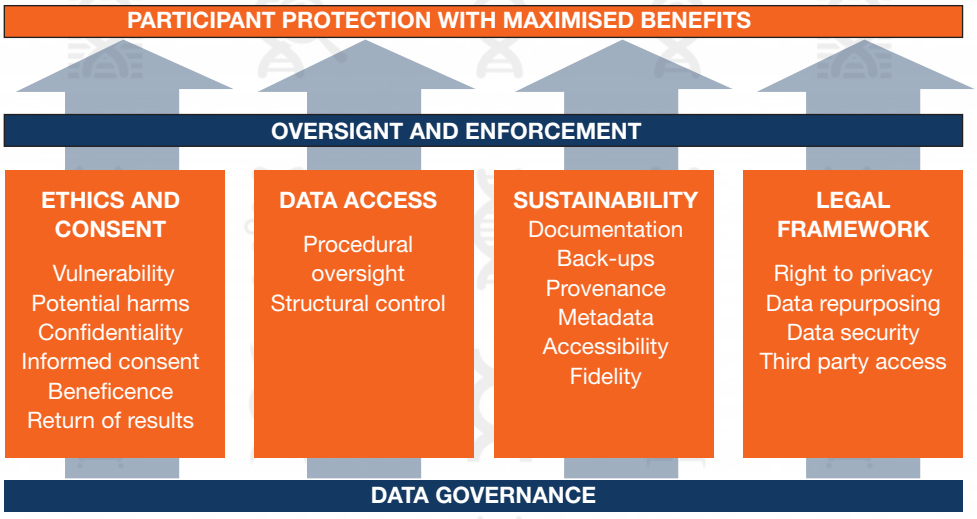
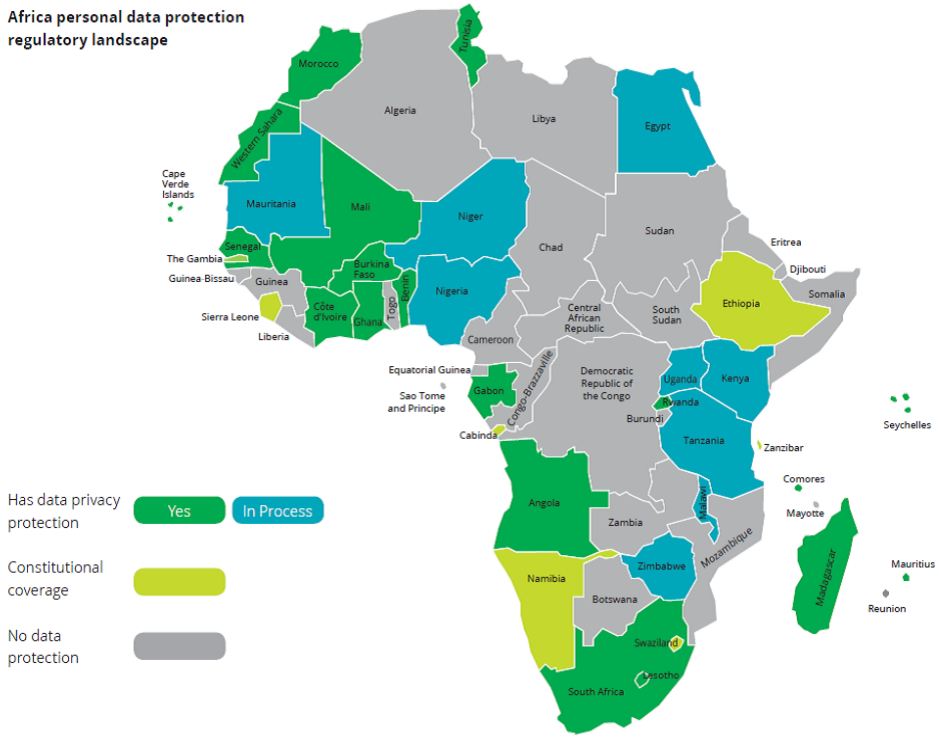


Figure 8. A framework for data governance for health data. Source: Tiffin et al. (2019).

already be challenging due to financial, logistical and skills constraints, but oversight of research integrity at institutional and/or National level is also essential to ensure that only the highest quality of research is conducted and translated to impact health outcomes on the Continent. This is a role that can also be taken by pan-African initiatives such as the African Academy of Sciences.

Map 1

Africa personal data protection regulatory landscape



Map 2

Cross-border data transfer and breach notification requirement in African countries which have adopted personal data protection legislation[1]

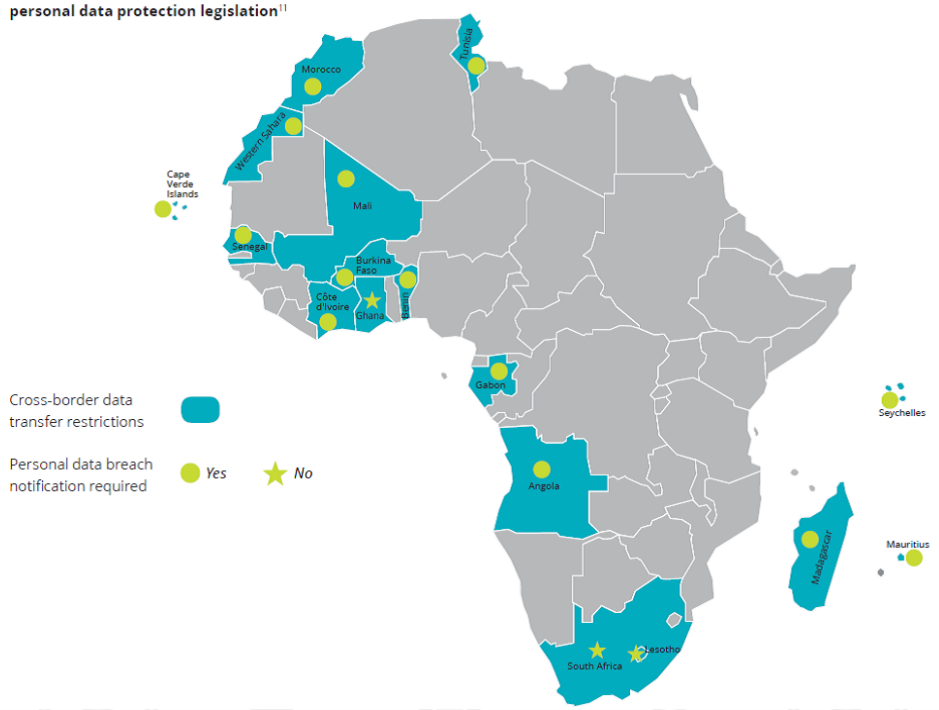


Figure 9. Privacy is Paramount, Personal Data Protection in Africa. Source: Deloitte (2017).

6.6 Informed consent for Genomic Medicine

The process of consent should encompass all the different aspects such as participation in research, respecting participants' choices, providing personal biomedical data over a long period of time and storage of the participant's data (and samples) while protecting the rights of participants and their security. Access to data by other parties to enhance clinical care need to be clearly explained and justified to participants at the time of consent (Wright et al. 2014). For a national GM programme the primary aim is healthcare so patients need to consent for the use of their genetic data for that purpose, but may also be requested to consent for research to further improve GM implementation. Without explicit consent for long term use of samples for other studies in the future, samples and data cannot be repurposed accordingly. The current consenting procedure for clinical trials can provide a baseline for developing informed consent processes, but clarity on the use of genomic material and potential for international sharing must be explicit. Participants might find it challenging to grasp the scope of the research, compromising their informed participation, so participant information should be clearly and completely presented. The informed consent process can be very complex in Africa, given the lack of guidelines and quality assurance due to resource limitations. Preparing for GM implementation on the Continent therefore also requires building capacity for informed consent processes that facilitate best data use for public health benefits whilst respecting participants' preferences. Consideration should also be given to appropriate practices for proxy informed consent and assent for minors and those who are not competent to give consent; and for proxy consents and/or consent for re-contact for those who are too unwell to give informed consent at the time of contact or treatment. These types of consent for medical treatment are often delineated in National Health Acts or similar legislation, for example reviewed in (Worku et al. 2016).

6.6.1 What level of information should be provided to a patient?

Honesty, transparency and respect for participant autonomy is of utmost importance (Tiffin 2018; Nembaware et al. 2019). The benefits, risks and side-effects should be discussed in simple terms and participants given the opportunity to choose the data use scenarios with which they feel comfortable. For example, from prior clinical studies (van der Merwe et al. 2017) some patients only wish to receive genetic results for specific tests for which they consented, and do not wish to obtain results of research findings. They expressed concerns about funding of the various genetic tests that will be included in translational research studies and wanted to know how the results will guide their therapy to make a favourable difference in their clinical outcome. Non-disclosure of actionable research findings may be considered unethical given the possible reinterpretation of data over time (Bombard et al. 2019). However, the no-return criterion may apply when results from genomic tests are generated for the purpose of producing generalisable findings as opposed to demonstrating clinical utility.

The scope of new research areas in genomic medicine is changing rapidly and cannot be forecasted accurately. Additionally, social, medical, economic and political factors can also have unpredictable effects on patient outcomes. It is therefore important to explain and discuss with participants the uncertainty around types of future use of a biomedical sample. Efforts to re-consent participants can also be made based on the new research studies in the future. A qualitative study has highlighted that patients feel happier when they give new consents to further use of their samples and data (Dixon-Woods et al. 2017). Although the genomic data has been stripped completely of personal information, it can always be re-linked to an individual through the genomic sequence; and it remains the right of the participant to know how the sample or data will be re-used in the future, and to withhold consent for unknown future uses should they prefer to do so. The participant should also be informed who is responsible for safeguarding their biospecimens/data, where these will be stored and for how long, when they will be destroyed, and how other parties might get access to the biospecimen or data. Other information provided should include what risks are associated with the storage of biospecimens and data, as well as steps that will be taken to mitigate those risks. A framework for tiered informed consent for genomic health research within Africa has been proposed by (Nembaware et al. 2019).

It is also important to inform participants of unforeseen consequences of their participation. A prevalent example is that non-paternity or adoption could be accidentally disclosed and cause patient and family trauma (Kotze et al. 2006; Wright et al. 2019).

Finally, the informed consent process should also outline what might be given back to the participant in terms of results, and the anticipated benefits to the participant and their community in both the short and long term. The community feedback relates more to research and may be more long term, while short term benefit to the patient applies

in the clinical setting. Other considerations for communication might also include how data might impact health insurance, and how data might be implicated in forensic or criminal investigations (as highlighted by recent use of genotype data for identifying criminals in the USA) (Kolata and Murphy 2018)

6.6.2 How to address language barriers in Africa for proper informed consent

In Africa, low health literacy levels and different language needs can make the application of the current consenting process complex, requiring comprehensive provision of clear and simply presented information about the research requirements and procedures and the collection and dissemination processes for data and biospecimens. This must be thoroughly explained in simple terms by someone who has been trained and has appropriate skills, which may be lacking in the African context. Whilst consenting procedures ultimately result in a signature on the consent form, the process of discussions and clarifications to reach this conclusion can vary significantly. Cultural differences such as communal decision making further influence the outcome of the consenting process; and translation of participant information into the primary language of the participant should also be undertaken, alongside employing local counsellors to undertake the information and consent processes. Local counsellors can ensure that participants do not have to rely only on written documents in their own language, but can have an appropriate discussion about the consent in their own language with someone who understands and respects their social constructs and preferences (Krogstad et al. 2010). If patients are not literate, the presence of a third person who speaks the same primary language, such as a relative, is essential. Adequate time must be assigned for the process, and training and/or informative videos in participants' first languages and filmed within their context can be helpful (Tait et al. 2014). Sufficient time should be allowed between providing participants with study information, and enrolling them, so that they can discuss their participation with family and community members if they wish.

Preliminary data from a breast cancer study performed in Kenya (Sawe et al. 2016) has shown that language barriers may result in decline of research participation, and that the same strategies used for projects involving whole genome/exome sequencing (WGS/WES) in developed countries may not work in rural African settings. This requires developing and testing innovative strategies to improve voluntary participation and explaining clearly the option for return of research results, for example through joint consultation with the researcher and treating clinician. Adhering to the gold standard of using genetic counsellors or medical geneticists are not practical, except if an online system is used. Primary care physicians and primary care nurses need to be empowered with practical tools such as the WES pre-screening algorithm developed in South Africa to enable multidisciplinary input for generation of adaptable patient reports for clinical management and follow-up as appropriate (van der Merwe et al. 2017)due to interpretation challenges caused by variants of uncertain clinical significance (VUS. Proxy decision making, undertaken through an interpreter or family members, should be discouraged as it is hard to confirm how much the participant actually understands (Jodheea-Jutton and Jheelan-Ramchandur 2018). This can be particularly difficult for studies involving children where compensation or financial gains might influence parents or guardians providing proxy consent.

6.6.3 Need to standardise and make routine the coding and capture of consents from individuals

Currently, informed consent practices are often based on broad consent models, whereby participants provide just one consent answer, with a yes/no binary option. This model is often considered problematic as it does not allow for individuals to express their preferences or exercise their autonomy over how their sample/data might be used in research (Tiffin 2018) The consent is seldom digitalised because those who do not give broad consent are excluded from the research study - in this case storing consent choices is at best a pdf scan of the signed consent template. Where tiered consent is used, however, it becomes important to digitally capture the choices made by each individual participant so that it is easy to respect those preferences in onward use of the data (Nembaware et al. 2019). Standardising which informed consent questions are asked, especially for tiered consent, can facilitate this process, help to make routine the coding and capture of consents from individuals, and aid the development of ontologies to capture this information.

6.6.4 Providing counselling with return of results

At present, African countries have limited capacity to counsel individuals participating in genomic projects and need to develop these resources to further the GM agenda (Ormond et al. 2018). It is important, however, to have some kind of counselling provided with results that are returned from GM research or translational implementation, and at least one genetic counsellor should be

involved in any study where genetic results may be returned to individuals. The same is true for providing genetic results in a clinical setting. Difficult issues that have to be resolved include defining whether verified results are actionable or not, and whether this definition should refer to what is actionable in the specific participant's case or what is actionable anywhere in the world (for example, where a treatment does exist but is beyond the financial or national means of the patient; or where a treatment only succeeds for a small number of treated patients).

This highlights the need for tiered consent particularly because of sensitivities around incidental findings and return of findings: PM requires return of results as a primary deliverable - but these could be specific, incidental or secondary findings; and could be actionable or not (Nembaware et al. 2019). National and pan-African guidelines for return of results would be helpful, and Institutional Ethics Review Boards also have an important role to play in guiding return of results and defining what should be considered actionable, on a study by study basis. A first phase qualitative research programme with focus group discussions among academia might help to raise awareness and move the agenda forward for informed consent practices in Africa. The Alliance for Accelerating Excellence in Science in Africa Data and Biospecimens Governance Committee may also be well placed to take this agenda forward.

6.7 Recommendations for ELSI and Data Governance

- Though a GM initiative may be country-specific, cataloguing the national-level legislation across African countries may be useful: Build a matrix for all countries, describing personal data protection through legislation (for example: review privacy acts, health acts, access to information acts, acts protecting vulnerable populations).
- Draft an African treaty on research integrity and addressing digital disparity in working with vulnerable populations. We also need to better understand how, when and under what circumstances GM could lead to, aggravate or alleviate stigma or discrimination for populations groups.
- More work needs to go into understanding how GM can be made available across African countries and communities in a way that is fair and equitable, with particular focus on how existing intellectual property impedes the development of affordable GM innovations.
- Establish guidelines for the informed consent process to ensure appropriate participant information, informed consent processes and community education is provided (e.g. (Nembaware et al. 2019).
- Develop training for community engagement projects, particularly for research projects.
- Develop processes to evaluate ethics implementation in ongoing projects, as well as in clinical settings and ensure quality assurance.
- Provide training for counsellors on genomic projects and Institutional Ethics Review Boards who evaluate informed consent processes, support genetic counselling in clinical practice.
- Create an experts' platform to catalogue ethics and data governance human resources, skills and expertise which can be accessible to countries where such resources are limited.

Box 10. Ensuring appropriate practices respecting patients and their families

Ensuring appropriate practices respecting patients and their families
Given the pace at which genomics is incorporated into clinical practice, a monthly meeting with stakeholders including clinicians, genetic counsellors, medical scientists, dieticians and a genetics trained medical journalist was held during 2018/2019 in Cape Town to discuss how available germline and tumour genomic tests impact treatment of patients and their families. Issues pertaining to interpretation of results, cost implications and new developments in this field are essential aspects of this discussion, which led to consideration of implementing the ECHO model (https://echo.unm.edu/need-to-know/). This guided practice model is ideally positioned to advance medical education and exponentially increases workforce capacity beyond local genomic expertise. It soon became clear that roll-out of these monthly meetings into Africa could provide best-practice specialty care to reduce health disparities.

Section 7. Education and Training

7.1 Summary

Genomic Medicine has the potential to revolutionise healthcare and health outcomes not only for developed countries but also for developing countries including those in Africa (Manolio et al. 2013; Mitropoulos et al. 2015). There is increasing interest from various stakeholders, including patients and physicians, in clinical and consumer-based genomic tools (Chung et al. 2016; Gabriels and Moerenhout 2018; Grünloh et al. 2018). Several examples of successful translation of genomic medicine research into the healthcare systems in developing countries provide concrete evidence of feasibility (Mitropoulos et al. 2015), however, a major hurdle is building the human capacity required to make this a reality at scale. The integration of genomic medicine into medical practice requires extensive restructuring of existing training curricula and continual up-skilling of healthcare professionals given the evolving nature of this field. Genomic medicine encompasses multi and interdisciplinary sub-fields which require healthcare professionals to continually engage with new content and with other professionals outside of their own disciplines.

Ideally, any genomic medicine training should cover all areas pertinent to this field such as knowledge in genetics and genomics, ethical legal and social implications, methods and tools which facilitate clinical application of genetics and genomics, and research (including genetic epidemiology) (Nembaware et al., 2019). Several genomic medicine training courses exist (Nembaware et al. 2016) which could accelerate training in this field if curriculum developers build upon existing knowledge. Though knowledge in genetics and genomics is central in genomic medicine, such knowledge is limited in African healthcare workers such as nurses, doctors and pharmacists (Wonkam and Angwafo 2006; Muzoriana et al. 2017). This lack of knowledge highlights the need for training in Africa, however, there are factors hindering widescale training which need to be systematically documented in order to address them more effectively.

There is recognition that genomic medicine training, like any other skills development training and education, should be guided by well-defined competencies which helps make the curricula relevant to specific contexts and enables more effective evaluations. Therefore, in addition to building trainer capacity, continual up-skilling of existing healthcare professionals to change knowledge, attitudes and behaviours, there is need to develop competencies specific to Africa to guide genomic medicine training for all healthcare workers in this field.

7.2 Genomic Medicine Training Needs Assessment in Africa

To capture the genomic medicine training needs from institutes around the continent, the African Genomic Medicine Training Initiative (AGMT) conducted a survey across 33 different African institutes from 19 countries (H3Africa Bioinformatics Network 2016). The AGMT (AGMT 2016) was established in part by H3ABioNet (www.h3abionet.org) in collaboration with the University

Photo credit: AESA 2019

AAS data and governance committee convening of 2019.

Photo credit: AMREF Kenya

Health care workers undergoing training credit Amref Health Africa

of Cape Town and the H3Africa Consortium Education and Coordinated Training Working Group. Results from this survey highlighted challenges and needs, these included: lack of support from management, finance, equipment and resources, expertise, time and lack of interest of computer science students with lack of expertise and resources being the most common themes. A few responses suggested solutions which included forging partnerships with private or better equipped organisations and increased online training. Results of the survey also made explicit the lack of trained personnel to conduct training and limited training resources, further underscoring the potential benefits of a collaborative training approach in Africa.

Specialised healthcare professionals in genetics such as medical geneticists and genetic counsellors should ideally be at the forefront of the implementation of GM (Kromberg et al. 2013). However, not all African countries recognise medical genetics as a profession, and for countries that do, the number of registered professionals is limited and incapable of supporting large-scale implementation of GM. For example, in 2008 there were only 10 registered medical geneticists and 10 registered genetic counsellors for the 49 million South Africans (Kromberg et al. 2013). Globally, there were an estimated 7000 registered genetic counsellors in 2018 (Abacan et al. 2019). It seems, however, that most of these counsellors are not based in Africa. South Africa, hosts the only formal genetic counselling training programmes to date on the continent, i.e. the University of Witwatersrand (Wits) and the University of Cape Town (UCT) both of whom offer a Masters degree in genetic counselling (Abacan et al. 2019). The Wits programme was established about 30 years ago and the UCT programme was established 5 years later. While these programmes have the potential to produce genetic counsellors amass, the opportunities and salaries for genetic counsellors has led to high drop-out rate. This calls for changes in national policies and the creation of job opportunities in relevant healthcare facilities. The need for experts such as medical geneticists and genetic counsellors is well-recognised, however, the status quo is that public healthcare in Africa is mostly provided by nurses, community healthcare workers and general practitioners. In the interim, it may therefore be useful to upskill the current healthcare professionals so they can provide basic genetic counselling and knowledge.

For the healthcare delivery processes to effectively adapt to the requirements of GM, there is need to recognise new specialty areas which require the formulations of job descriptions. In some developed countries, job descriptions have already been coined for new emerging GM relevant professions such as Clinical Bioinformaticians and Data Stewards. In the African context, there might also be an urgent need to professionalise the role of community engagement personnel as these are key to aligning GM implementation to specific cultural contexts amongst other roles.

7.3 Competency-based Training for Healthcare Workers

Competency-based training aims to increase knowledge and skills required by participants to effectively carry out tasks/activities in the workplace and it also promotes accountability (Frank et al. 2010). While several competency resources relevant to the genomic medicine field exist, the Inter-Society Coordinating Committee for Physician Education in Genomics (ISCC) has one of the most comprehensive sets of competencies which are

freely available online for downloading and adapting based on needs (Korf et al. 2017). The ISCC is convened and coordinated by the National Institute of Genome Research (National Institute of Health, United State of America) and has representatives from a wide range of institutes and countries. The ISCC maintains a resource for competencies in genetics and genomics "Entrustable Professional Activities" EPAs for healthcare professionals (Korf et al., 2014). This resource is freely available online, although tailored for United States of America professionals, it is an ideal starting point for curriculum developers in this field.

7.4 Existing Genomic Medicine Training Frameworks for Healthcare Providers

Several organisations have worked on setting standards for genomic medicine education and establishing the genomics-related skills and knowledge required. In the USA, the Accreditation Council for Graduate Medical Education (ACGME: https://www.acgme.org/) developed six core competencies: 1) patient care, 2) medical knowledge, 3) practice-based learning and improvement, 4) professionalism, 5) interpersonal skills and communication and 6) systems-based practice. These competencies have been adopted by the Association of Professors of Human and Medical Genetics (APHMG) (Hyland et al. 2013) and others to ensure healthcare providers gain skills and knowledge in genetic testing implications and impact (Hyland et al. 2013).

Additional freely available training courses relevant to GM exist, some are summarised in the article (https://globalresearchnurses.tghn.org/articles/preparing-genomic-medicine-nurse-training-africa/) Of note, is the Master's in Genomic Medicine course that was established in 2014 (and adjusted in 2018) and is delivered by the Genomics Education Programme (GEP) in collaboration with the National Health Services (NHS) England and Genomics England.

7.5 Genomic Medicine Course Content

Different healthcare providers get trained with different curricula, most of which do not include basic sciences or genetics. Therefore, genomic medicine training should include these topics as well as more social ones relevant to interactions with patients. Some of the core topics are outlined below, though the health informatics and information technology topics are potentially more appropriate for a data specialist.

Genetics and Genomics: A core understanding of the basic genetics and genomics is required and this includes training in science subjects such as molecular and cell biology, genetics, and pathophysiology, which are important for understanding and practicing GM. They should know about DNA, genes, and mutations which may impact health.

Ethical, Legal and Social Implications: Most ethical and social issues in GM are not unique, however with the advent of novel technologies and increased coverage of sequencing platforms, existing ethical issues have been amplified and new ones have arisen. In addition, the interface of research and healthcare through GM makes the understanding and engagement of ethical, legal, and social implications necessary for healthcare professionals.

Genetic counselling: While genetic counselling is integral to the implementation of GM, the limited resources, scarce job opportunities and training programmes hamper the creation of a critical mass of genetic counsellors. There is therefore need for task-shifting in the African context to allow for non-genetic counsellor professionals to provide counselling. Such a task-shifting model worked for HIV counselling where lay counsellors and health professionals are provided with basic counselling training (Mwisongo et al. 2015).

Clinical Application of Genetics and Genomics: The implementation of GM could be promoted by upskilling all healthcare workers in technology and through use of case-studies which have been successfully implemented in similar settings. Providing training on clinical data interpretation and communication is also essential to ensure those healthcare providers are able to use the new data effectively and appropriately in the provision of evidence-based care.

Research and Genetic Epidemiology: There is need for existing healthcare workers to gain an appreciation of research and the research process.

Health Informatics: Understandably, most health informatics training done in Africa focusses on the health systems that are being utilised in the country. While clinicians should be aware of the potential of health informatics in improving patient care, it would be useful for some staff within the healthcare environment to understand, at least at a basic level, more technical subjects that are not included in the current medical curriculum, such as coding or the use of artificial intelligence in clinical decision support. There may already be some

data scientists and health informatics experts within established health systems, however, they need to remain up to date with new types of data and novel techniques to improve the accuracy of healthcare decisions. There has been some formal institutional health informatics education in Africa as shown in the widget results from the Ghanaian research (Chawani 2014). Countries like South Africa, Nigeria, Ghana, Ethiopia, Cameroon, Rwanda and Malawi are at the forefront, though there may be activities in some other African countries.

Information Technology: The use of Information Technology in healthcare has potential benefits, especially when public health systems are not in close proximity to patients (Tiihonen et al. 2008). Again though, there is a lack of skills to support the development, usage and maintenance of such eHealth systems (World Health Organisation 2014). Therefore, training in this field is required.

Instead of developing curricula and courses from scratch, it is worth adapting these from existing resources. The AGMT developed their curriculum and competencies using modified GM modules from the NHS genomic medicine course (see Box 11). The AGMT made use of case studies and content relevant to the disease burden in Africa. In addition, an extra module of community engagement was also added to cater for the African context. This highlights the feasibility of adapting existing content for the african needs and adding new content areas where necessary. The AGMT course is run as a remote classroom project and by the end of 2019 two iterations of genomic medicine training had been run for nurses.

Box 11. Adaptation of the Precision Medicine Training in the United Kingdom by AGMT

Adaptation of the Precision Medicine Training in the United Kingdom by AGMT

The Master's in Genomic Medicine framework is an educational programme designed to provide healthcare professionals in the UK with a multidisciplinary perspective on genomics and its applications in healthcare. Initiated in 2014, the Master's programme was one of the first major initiatives devised and delivered by the Genomics Education Programme (GEP). The curriculum was first developed by the GEP in collaboration with National Health Services (NHS) England and Genomics England and was fully reviewed and revised in 2018.

The Master's framework is intended to support NHS healthcare professionals in developing their knowledge of genomics and how it can be applied to clinical practice and medical research. AGMT based their competencies and course content on this programme with some adaptations to the African context.

Modules	Courses	AGMT Adaptation
Mandatory	Fundamentals in Human Genetics and Genomics: Omics Techniques and Technologies and their Application to Genomic Medicine; Bioinformatics, Interpretation and Data Quality Assurance in Genomic Analysis	All course material used case studies from Africa and made the content relevant to African settings
Elective	Genomics of Common and Rare Inherited Diseases; Molecular Pathology of Cancer and Application in Cancer Diagnosis, Treatment and Monitoring; Pharmacogenomics and Stratified Healthcare; Application of Genomics in Infectious Disease	All these courses were mandatory. In addition, to making the case studies and course material align to African healthcare problems, participants had to learn and practise basic genetic counselling skills. Given the shortage of genetic counsellors, task shifting is required.
Optional	Ethical, Legal and Social Issues in Applied Genomics; Introduction to the Counselling Skills used in Genomic Medicine; Health Economic Evaluation in Genomics; Professional and Research	Ethical issues pertinent to Africa were included in the training. In addition, an extra course on community engagement was added. This community engagement course is tailored for the african context were villages exist and other sub-tribes and populations.

Recommendations for education and training

- Create and design job descriptions for new healthcare professionals required for implementation of GM in Africa.
- Identify and adapt existing publicly available competencies to help guide and possibly fast-track curriculum development for all healthcare professionals, both for new degrees and professional development.
- Conduct a curriculum review of the existing Genetic Counselling and Medical Genetics speciality areas and adapt them for other institutes.
- Increase learning of basic science subjects in medical curricula, focusing on how this knowledge can directly impact clinical practice, include health informatics education where applicable.
- Map out all key stakeholders and their needs.
- Promote career uptake in the emerging careers for example bringing in experts in the ground-breaking fields that are changing healthcare and, thus, stimulating students to pursue a career in these areas.
- Assess policies and lobby policy makers to recognise genomic medicine as a speciality field.
- Establish accreditation bodies and policies in GM and align to existing health professional bodies.

Section 8. Moving from Research to Translation

8.1 Summary

Much of the work in health-related genomics on the continent is currently being carried out in the context of research. To ensure health benefits these need to shift towards implementation in healthcare settings. To move from basic research to translation requires transitioning to applied research, proof of principle for clinical application, validation, economic and feasibility studies, and up-scaling for clinical uptake and implementation. In some cases, this may be fast-tracked, when there is already sufficient evidence for actionability based on other patients with similar clinical phenotypes or well-established literature, most notably for rare diseases and cancer. This section describes some of the key considerations for translation into clinical practice using an existing pathology-supported testing framework being implemented in South Africa as a case study.

8.2 Introduction

Several frameworks have been proposed for standardisation of GM applications that extend beyond the limitations applicable to genetic tests providing simple yes/no answers. Given the structures required as discussed in the above-mentioned sections and how this should be operational in the genomics era, a decision needs to be taken as to what framework should be used for clinical application of big data. This framework needs to accommodate overemphasis of collection of data from single patients in line with the concept of N-of-1 trials to personalise treatment (Ioannidis and Khoury 2018). It furthermore requires combining individual clinical experience with patient needs and preferences, while obtaining experimental evidence on the validity and utility of patient and population data. This is obtained from systematic identification, analysis and description, whenever possible, in order to move from single- to multi-gene analysis of cancer and other NCDs. Identification of a unified risk factor such as the metabolic syndrome was considered particularly useful to distinguish between inherited and environmental causal factors such as lifestyle or therapy-induced NCDs (Kotze et al. 2015; Kotze et al. 2015; Kotze 2016), as risk stratification and clinical management of multi-factorial and polygenic disorders guided from the genetic background remains a challenge. The translation of personalised genomic information into clinical practice requires large-scale comparative population data that have an impact on public health. This depends on the ability to affect large segments of the population at increased risk of NCDs caused by gene-gene and gene-environment interaction. It is by combining genetic knowledge with real-life experience through application of precision medicine that genomics can make a positive impact on disease prevention and treatment-decision making. This, however, requires significant infrastructure. The South African Medical Research Council (MRC) and Department of Science and Innovation (DSI) are key stakeholders in healthcare initiatives such as the African Genomics Centre recently established in Cape Town. This provides South African scientists and international collaborators with a state-of-the-art facility for research translation into clinical practice. It is paramount that this infrastructure be utilised for development and implementation of genomic solutions that are affordable and accessible to Africans through public-private partnerships.

Current laboratory technologies applied in translational genomics employ a wide variety of methods and tools that require expensive laboratory-based equipment. Return of results to a patient can take several days or weeks, posing a significant challenge in resource-limited African settings where lost to follow-up is a major problem. Large numbers of patient samples are currently leaving African countries from both private and public institutions for genetic or genomic testing abroad, instead of making use of local services if available. This is due to delayed turn-around time, interpretation challenges and reduced prices offered to patients making direct payment to international companies (e.g. Invitae, $250 per gene panel screen). The use of rapid, cost-effective point of care (PoC) DNA assays may curb this trend, if used as first-line pharmacodiagnostics or screening tests that can be extended to next generation sequencing (NGS) in uninformative cases. This can be accomplished by the following steps to establish the evidence base for implementation of personalised GM:

- Primary research: preliminary findings from primary research (including functional studies).
- Validation: replicating basic research findings in a wider population.
- Translation: transforming research into clinical application.
- Incorporation in routine care: diagnostic, prognostic, therapeutic,
- Public health: population level healthcare implementation.
- Clinical trials: long-term follow-up studies.

- Clinical decision support: pre-screen and pharmacogenetics algorithms.
- Improving access to tests and treatment: Point of care tests as 1st entry.

8.3 Genetic counselling combined with point-of-care testing

Genetic counsellors should be suitably qualified to inform patients about the limitations and benefits of mutation-specific PoC tests versus comprehensive sequencing methods, given the relatively high frequency of variants of uncertain clinical significance (VUS) uncovered by gene panel testing and WES/WGS. PoC genetic testing provides a unique opportunity for genetic counselling to take place during the waiting period of 1-2 hours from sample collection to PoC test result for simple tests, following referral by the treating clinician. Figure 10 shows an example of the different aspects or stages of cancer, each with its own set of genes implicated, which may need to be considered during a genetic counselling session. Differences in opinion on the value of genomics beyond standard practice led to the development of a pathology-supported genetic testing (PSGT) platform used to identify the target group most likely to benefit from germline DNA testing and/or tumour gene profiling (van der Merwe et al. 2012, 2017; Kotze et al. 2013; Grant et al. 2013, 2019; Baatjes et al. 2017, 2019).

The use of "virtual" gene panels subsequent to WES/WGS may lead to significant cost saving given the ability to address different aspects of the same disease, as appropriate. Clinical interpretation may however be challenging when high- to moderate- and low-penetrance genes are simultaneously assessed towards disease diagnosis and identification of lifestyle-triggered (epigenetics) or therapy-induced (pharmacogenetics) medical conditions. PSGT incorporating WES/WGS provides clinicians with the option to request additional genomic information pertinent to the patient's clinical management, even in cases where expert groups disagree on the level of evidence

Box 12. Case study: Warfarin pharmacogenetics as proof of concept

Case study: Warfarin pharmacogenetics as proof of concept

The first validated PoC DNA assay developed for warfarin dosing using cheek swabs has recently been implemented in the UK, using the portable ParaDNA device (Jorgensen et al. 2019). Prescription of warfarin is challenging due to its narrow therapeutic index, which led to the development of dosing algorithms incorporating both clinical and genetic data. In the EU-PACT randomised controlled trial, (Pirmohamed et al. 2013) showed that prediction of warfarin dose using age, sex, weight, height, amiodarone use and genotype was superior to using conventional dosing criteria. These studies demonstrated sufficient evidence to justify translational studies of host-environment interaction using PoC DNA tests tailored to the African context. Since warfarin is the most commonly used anticoagulant worldwide for prevention and treatment of thromboembolism, proof of cost-effectiveness demonstrated in the UK and Sweden (Verhoef et al. 2016) was an important consideration for PoC DNA testing in South Africa. A retrospective review to determine the causes and management of warfarin toxicity of patients admitted to Tygerberg Hospital between June 2014 and June 2015, showed that warfarin toxicity carries significant mortality and an average treatment cost of R10 578 for a median of eight days (Jacobs et al. 2017). In a study performed at the Victoria Hospital in Cape Town, a therapeutic international normalised ratio (INR) outcome was achieved in less than 50% of patients (Sonuga et al. 2016).

Correct dosage of warfarin is usually only established after six to eight visits to the clinic. This is partly due to genetic factors shown to predict more than 40% of the daily dose of warfarin as a consequence of increased sensitivity or decreased drug metabolism, as well as increased risk of bleeding (Johnson et al. 2010). The ParaDNA PoC warfarin dosing assay, which provides a genotyping result from a saliva swab or drop of blood within I hour, heralded a new era in personalized medicine (Fitzgerald et al. 2019). This analytically and clinically validated PoC assay using HyBeacon® probes includes three extensively studied clinically useful SNVs: CYP2C9 *2, CYP2C9 *3 and VKORC1 c.-1639G>A. However, this assay has limited use in Africa where the following gene variants may be most relevant for a broad population assay (Ndadza et al. 2019b, a): CYP2C9*2, CYP2C9*3, CYP2C9*5, CYP2C9*6, CYP2C9*8, CYP2C9*11, VKORC1 (-1639G>A; rs9923231), VKORC1 (g.9041G>A; rs7294) and rs12777823 (CYP2C cluster). In addition, the multi-functional APOE 2/3/4 polymorphism that may alter the transport of vitamin K to the liver appears to have an independent effect on warfarin dose requirements and may explain up to 2.5% of warfarin variability in Africans (Kimmel et al. 2008). This gene variant is involved in the lipid and lipoprotein pathway targeted as part of the pharmaco-diagnostic screening process used to distinguish between monogenic familial hypercholesterolaemia (FH) and less severe dyslipidaemia subtypes (Kotze 2016; Marais et al. 2019).

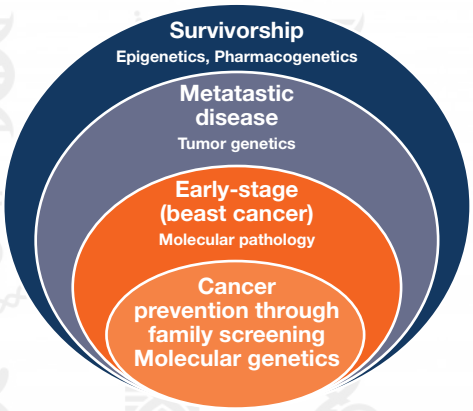


Figure 10. Pathology supported genetic testing platform underpinned by genomic counselling support across the continuum of cancer care.

for clinical utility to justify additional genetic tests (e.g. CYP2D6 genotyping for tamoxifen resistance) (van der Merwe et al. 2012; Baatjes et al. 2017). Lessons learned from pathway analysis highlighted both the benefits and limitations associated with detection of a functional gene variant associated with different clinical conditions and/or treatment response (Delport et al. 2014; Luckhoff et al. 2015; Marais et al. 2019).

8.4 Informed Consent for Integration of Research and Service Delivery

The clinical utility of WES/WGS is limited by interpretation challenges that can be overcome with a three-pronged approach applied from sample collection to report generation (Torrorey-Sawe et al. 2020). By first focusing on the primary reason for genetic testing such as the identification of a causative gene variant with well-established clinical guidelines, insight is gained on what else may be required to optimise clinical management. WES/WGS has the benefit to extend data analysis beyond a single application based on 1) the information provided in the informed consent form signed by the patient at referral/baseline, 2) the usefulness of the data obtained and stored for test selection and clinical interpretation, and 3) infrastructure used to generate a report for recommendation of next steps related to treatment, lifestyle modification or familial risk.

8.5 Pathology-supported Genetic Testing

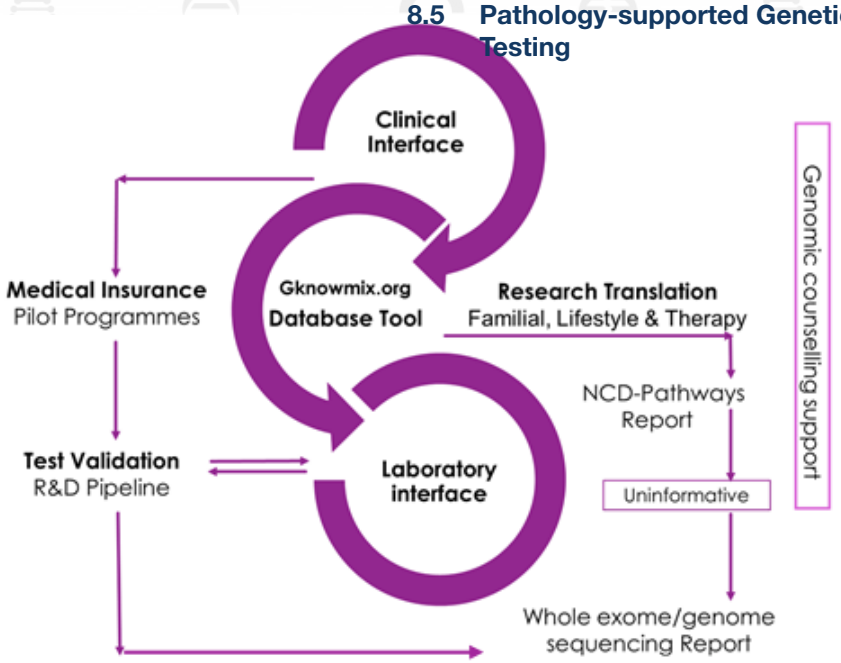


Figure 11. A pathology-supported genetic testing framework for application of a three-pronged approach towards identification of 1) pathogenic mutations associated with familial risk, 2) genetic underpinnings of tumour subtypes, biochemical abnormalities or co-morbidities influenced by modifiable environmental risk factors, and 3) patients at increased risk of medication side effects/Failure due to genetic variation.

PSGT was implemented in South Africa to facilitate the paradigm shift from genetics to genomics, targeting disease pathways shared by cancer and other NCDs (van der Merwe and Kotze 2018). In addition to assessment of combined gene-environment effects causing biochemical abnormalities as the true basis of disease, pathology tests allow monitoring of response to treatment as a vital factor in the differential diagnosis of inherited diseases influenced by lifestyle and other environmental factors. When used as a first-line screening test, PoC DNA testing is ideally suited to provide a definitive diagnostic result or indicate the need for extended testing using gene/genome sequencing in uninformative cases (Figure 12). An automated doctor/patient report generated within approximately 1 hour during a genetic counselling session, should be backed up by a comprehensive patient report authorised by a multi-disciplinary team of domain experts. Research translation from the knowledge database into an informative report should be based on the understanding that clinical **diagnostic** scoring systems and family history often cannot differentiate between adults with extreme medical conditions caused by monogenic versus polygenic risk factors influenced by an environmental component. The same applies to the use of immunohistochemistry for tumour subtyping and biochemistry tests in **screening** algorithms. They frequently fail to identify individuals who may safely avoid invasive or expensive interventions found to be most effective in genetically predisposed individuals. Therapeutic interventions may be ineffective or cause severe side effects due to somatic or germline variants that differ between individuals requiring personalised **treatment** strategies. By converging fragmented information collected at various data points into disease-specific "indications for genetic testing" that are seamlessly translated into adaptable genomic reports, widespread adoption of personalised medicine may become a reality.

8.6 Recommendations for moving from research to translatio

- Capacity development of a new cadre of geneticists at MSc and PhD level, participating in ethics approved translational research projects to conduct research and development of novel and relevant GM test panels for use on PoC DNA devices (e.g. ParaDNA, Genie 3, Idylla, Q-POC and the Flurocylcer).
- Test validation and feasibility studies to demonstrate the value of low-cost PoC tests in comparison with traditional laboratory-based genetic testing of well-established biomarkers with proven clinical utility.
- For new tests there is need to assess analytical and scientific validity and clinical utility, including the sensitivity and specificity of the tests.
- Novel findings should be incorporated into an existing body of knowledge towards engagement of African scientists in the commercial manufacture of test panels for clinical use.
- Implementation of integrated data systems for analysis, interpretation and report generation, and AI for clinical decision making.

Section 9. Stakeholder Engagement

Photo credit: AESA 2019

H3Africa Data Governance Workshop. Nairobi, Kenya, June 2019

9.1 Introduction

The GM ecosystem aims to link patients, care providers, clinical laboratories, the research endeavour, and health systems to ensure that patients receive the most effective healthcare for their individual context, but this can only be achieved through buy-in across the GM ecosystem (Ginsburg and Phillips 2018). Stakeholders in GM fall across all tiers – from micro- through meso- to macro- levels of involvement, with stakeholders from all levels across multiple hierarchies that include members of the general public; healthcare providers; researchers in genetics, genomics, bioinformatics and public health; health service management; biomedical and data analytics companies; private insurers; national departments, regional and global organisations. The impact of GM is also cross-sectoral, with key roles for different sectors that include health and finance, as well as national and regional sectors with an economic interest in product innovation, skills and industrial development. There is concurrently a requirement to engage with educational investment in developing future employees with the right skills for new technological, cutting edge careers in GM in Africa.

Some of the questions that need to be explored and addressed about GM stakeholders are:

- Who are the stakeholders and what are their lenses on GM?
- What are the GM outputs that they are invested in realising?
- What is their level of influence towards achieving those outputs?
- How can stakeholders be engaged to ensure collaborative and complementary roles towards realising equitable and ethical GM research and translation?

This chapter explores these questions, with recognition that each country and region in Africa may have different priorities and challenges in engaging with stakeholders, which should be refined through appropriate stakeholder engagement and analysis. Furthermore, countries and regions may harness existing local strengths to provide a base from which to grow GM. This approach is illustrated in the UK, where technological skills in high-throughput sequencing were linked to existing electronic medical records (Genomics England 2018); in Estonia (Leitsalu et al. 2015) where strong electronic medical records and mature eGovernment were leveraged; and in Switzerland where there was a focus on interoperability and data sharing with coordination through the Swiss Personalised Health Network (Swiss Personalised Health Network).

A generic summary of some key stakeholders is provided in Table 2 and presented as a socio-ecological framework in Figure 2. A potential power/interest grid is proposed in Figure 13.

9.2 Micro Level: Individual Stakeholders

At the micro, or individual level, primary stakeholders are patients, who will directly benefit through anticipated improved health outcomes due to GM approaches. Because GM is based on genetic variations which are in most cases inherited, relatives of patients may also have an investment in the application of GM approaches in routine healthcare. Whilst having a high interest in the application of GM, patients and their families may not have a strong influence on uptake of GM, because of lack of knowledge or agency and/or generally limited access to healthcare. Also, patients in physical ill health may not have the capacity to demand new clinical care paradigms. For these reasons, it is important to ensure appropriate community engagement and accessible public information about what is (and isn't) possible and safe with GM approaches. With the advent of GM translation, it will be equally important to support and engage with patients to ensure that they are sufficiently informed and truly consenting to new GM health interventions. This requires an active partnership with patients, their families, care providers and counsellors to ensure an inclusive and holistic approach to GM within the African context. Having the involvement of local, rather than foreign or non-African care providers and counsellors can also ensure culturally appropriate, inclusive, understandable and accessible engagement between these key stakeholders. In the USA, a 2016 study investigated how prepared for whole genome sequencing diagnostics primary care providers and cardiologists were (Christensen et al. 2016) physicians may be unprepared to use it.\n\nMethods\nPrimary care physicians (PCPs, and similar regional and national quantitative studies will help to identify areas where most stakeholder engagement is most required for GM within Africa.

Researchers who are generating evidence to underpin GM in Africa are also key stakeholders, as they have built careers around this work in the hope of seeing GM improve health outcomes on the Continent. Researchers therefore have high interest in continued promotion and translation of PM in Africa, although individually they have fairly small influence on how research is prioritised in their institutions, by National, regional or international funders, or by governments. Researchers driving GM research in Africa need financial and institutional support to take this research forward, as well as National and regional infrastructure sufficient to support research requirements. At the same time, it should be recognised and respected that researchers inevitably have a personal interest

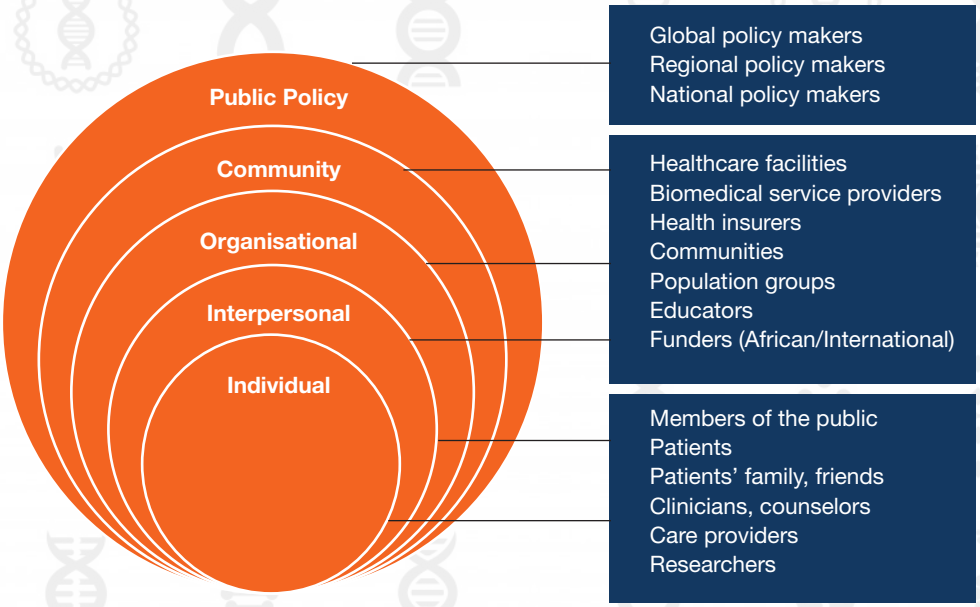


Figure 13. A socio-ecological framework for stakeholders in precision medicine.

in GM approaches to further their careers – both financially (Johnston 2008) and academically (Levinsky 2002; Saver 2012), and Institutional Review Boards and other oversight bodies should take responsibility for ensuring that improved African patient outcomes remain the key priority for all GM research undertaken on the Continent.

9.3 Meso level: Communities, Health Systems Oversight

At the community and institutional/organisational level, key stakeholders in the implementation of GM approaches in Africa include health facilities and the infrastructure that ensures an effective supply chain for appropriate healthcare. This includes government health facilities, and in Africa often a parallel infrastructure of private healthcare facilities and health insurance companies. It would be optimal for public and private healthcare ecosystems, as well as NGO stakeholders, to collaborate in the integration of GM approaches into the standard of care, to ensure that national and population-based resources are most effectively used to provide equitable access to best GM practices, rather than having to duplicate resource investment in parallel systems. Any national thinktank on how to develop and translate GM interventions should include pro-active and pro-collaboration members from both healthcare streams. Any established or nascent biotechnology sector companies or parastatals are also poised to enable, as well as benefit from, GM implementation. Decision-makers at a national or regional level can engage with the biotech sector to promote accessibility to GM services as well as creating employment and driving a new biotech endeavour within Africa: whilst the biotech sector is still nascent in many African countries, a competitive edge on lower facility/property and employment overheads could underpin African growth in this sector. The recent establishment of the African Genomics Centre by the Beijing Genomics Institute at the South African Medical Research Council is a first step towards developing commercially viable high-throughput competencies on the Continent. Health informatics and data analytics require computational hardware but less biotechnological infrastructural investment, and can certainly be undertaken successfully within Africa as demonstrated through the achievements of the H3Africa Bioinformatics Network (Mulder et al. 2016)made possible by new high-throughput genotyping and sequencing technologies and improved data analysis capabilities. Some of the greatest genetic diversity among humans, animals, plants, and microbiota occurs in Africa, yet genomic research outputs from the continent are limited. The H3Africa, providing another entry point for internationally-competitive GM service development on the Continent.

9.4 Macro Level: Governments and Regions

At the macro level, African governments and regional policy makers can be key influencers and promoters of GM development on the Continent. National Departments of Health can work towards developing and harnessing electronic medical records and health data in an ethical and responsible way, using international standards in interoperability and data governance. Within an appropriate ethics and governance framework, data integration of existing health records with genetic data can underpin both research and translation of GM approaches for African patients. Promoters of GM approaches can also include Departments of Science and Technology who provide funding, support and training for research and innovation, as well as ensuring infrastructure and technical support for implementation. Departments of Education can ensure that STEM secondary and tertiary education can meet the needs of GM research and implementation; as well as providing work opportunities for suitably-skilled STEM graduates.

At the regional and pan African level, research consortia have already been established to drive collaborative health and genomics research on the Continent. Regional organisations within Africa, as well as the African Union (https://au.int/), the World Health Organization and the Alliance for Accelerating Excellence in Science in Africa (https://www.aasciences.africa/), are also well positioned to promote GM in Africa and advocate for research funding and translational support for GM in Africa.

There are several advantages of collaborating across borders for GM initiatives. Aside from the need for genetic data from global populations to improve variant interpretation, large sample sizes enable more accurate analysis and interpretation of genomic studies. Additionally, diagnosis and treatment of rare diseases, by the nature of the fact that few cases are available in each country, would benefit from access to cases across the globe. In Africa there is also the reality that population sub-groups are often spread across countries due to history and more recently migration, so sharing of knowledge about population-specific variants as they relate to disease is essential. Europe has recognised this need and established the “1+ Million Genomes” initiative (Saunders et al. 2019), to which 21 member state governments have signed up. The aim of the project is to facilitate

Box 13. Case Study: Pan-African research consortia promote PG

Case Study: Pan-African research consortia promote PG
Pan-African research consortia have leveraged extensive foreign research funding to generate a common vision for health genomics research on the Continent. The Human Health and Heredity in Africa Consortium promotes health genomics research for multiple conditions in projects across many African countries, supported by a collaboration between the National Institutes of Health and the Wellcome Trust (through AESA) and African researchers to ensure a unified effort. The African Academy of Sciences (AAS) also coordinates and supports pan-African PM research on the Continent. Other examples of coordinated PM programmes include the Sickle Africa Data Coordinating Centre (SADaCC)*, and the Global Emerging Pathogens Treatment Consortium (GET Consortium) which was established in 2014 as a direct response to the Ebola outbreak in West Africa* (Get Consortium)*.*

responsible, ethical genomic data sharing that respect government regulations across national borders. By working together, they aim to ensure that the appropriate technical infrastructure that facilitates secure, federated access to data is established across the EU, and that access is line with ethical and legal frameworks. Since the focus is on national precision medicine initiatives, they hope to encourage "uptake by healthcare systems and integration into personalised healthcare".

There are several other international initiatives bringing together representatives from cohorts across the globe, both those within national precision medicine initiatives linked to the healthcare system, and those linked to research projects. The Global Genomic Medicine Collaborative (G2MC: https://g2mc.org/) is a community initiative to bring together groups who are advancing genomic medicine implementation in healthcare, and a spin-off from this is the International HundredK+ Cohorts Consortium (IHCC: https://ihcc.g2mc.org/), which will "create a global platform for translational research". The International Common Disease Alliance (ICDA: https://www.icda.bio/) is more research focused, aiming to increase our knowledge of common diseases for improved diagnosis and treatment. There are also organisations that develop standards and tools for precision medicine, including HL7, FHIR, and the Global Alliance for Genomics and Health (GA4GH: https://www.ga4gh.org/), which is developing tools for secure federated data analysis. These efforts, as well as others, such as the World Economic Forum 'leapfrogging with precision medicine' project (https://www.weforum.org/projects/leapfrogging-with-precision-medicine) are helping to ensure that lessons are learnt from existing initiatives and experiences, as well as ensuring that data can be shared for the benefit of the patients.

9.5 Discussion

Here, we have outlined some of the key stakeholders who may play a role in driving the GM agenda forward in Africa; but regional and national priorities and circumstances need to be explored to understand nuances in the different locations across Africa. Common to all, though, is a primary goal of ethical, effective, equitable and affordable healthcare for all African individuals, maximising best health outcomes whilst minimising cost and risks to individuals and populations. Stakeholders can work collaboratively to find a balance that prioritises health outcomes of individuals, health outcomes of populations and communities, whilst ensuring financial capacity and accessibility of funds for GM research and implementation and logistics and supply chain capacity, to ensure equitable and sustainable provision of GM. Through stakeholder collaboration, this balance can aim to grow the biotech industry whilst keeping GM affordable and equitable for governments and individuals; and must actively reconcile different agendas of researchers, governments and funders to ensure that the outcomes they seek are complementary and respect the ultimate goal of equitable access to improved health for African populations.

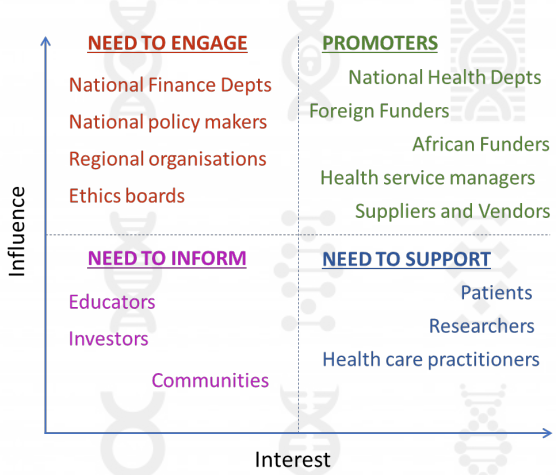


Figure 14. Identifying stakeholders, their interest and influence in GM.

9.6 Recommendations for Stakeholder Engagement

- Regional and national efforts can define who is responsible for engaging stakeholders and understanding their requirements.
- Stakeholder mapping, active community and stakeholder engagement, and meaningful qualitative research can assist in reconciling and aligning the agendas of all PM stakeholders.
- Regional/national processes should be defined to hold stakeholders accountable for how they engage with the GM agenda, and to ensure that GM activities are appropriate, equitable and ethical.
- Such an endeavour will have a greater chance of success if key stakeholders, particularly in government, work together across departments.

Table 2. Summary of potential stakeholder interests

	Stakeholder	Sector	Focus/Lens	Priority outcomes
MICRO	Patient	General public	Diagnosis, cure	Better health (cure)
	Relatives of patient	General public	Diagnosis, prognosis	Better health (prevention, cure)
	Consulting doctors	Health care	Diagnosis, cure	Better health outcomes
	Genetic counsellors	Health care	Patient decision making	Patient empowerment
	Researcher	Biomedical	Innovation, career	Innovation, research output
MESO	Pathology services	Biomedical	Accurate DNA analysis	Accurate, reliable genetic testing
	Healthcare facilities	Health care	Logistics and safety for healthcare provision	Treatment provision without adverse events
	Communities, population groups	General public	Diagnosis, prognosis, cure	Better health
	Educators	Education (tertiary)	Skills development	Trained personnel
	Private health insurance sector	Commercial	Improved health outcomes with cost benefit	Cheaper, better diagnosis and cure, with
	Biotech companies	Commercial	Product innovation	Profitability
	African funders	Not for profit, research, philanthropic	Health of Africans	Improved health outcomes for Africans
	International funders	Not for profit, research, philanthropic	Health of Africans, impact of health on non-African populations	Improved health outcomes for Africans, availability of new data/ knowledge to benefit non-Africans
MACRO	Government health departments	Health care	Affordable, effective healthcare	Affordable, practical diagnosis and treatment resulting in better health outcomes
	Government finance departments	Finance	Better healthcare with cost benefit	Affordability with effectiveness
	National organisations	Regulatory/policy	Regulation for appropriate healthcare	Equitable, safe and ethical processes for improving health outcomes
	Regional organisations	Regulatory/policy	Regulation for appropriate healthcare	Equitable, safe and ethical processes for improving health outcomes
	Global organisations	Regulatory/policy	Regulation for appropriate healthcare	Equitable, safe and ethical processes for improving health outcomes

Section 10. **Conclusion**

This framework document has outlined the key components of a genomic medicine ecosystem, though undoubtebly there are ommissions and the landscape will change as technologies and healthcare evolve. What should be stressed though is the need to have genomic medicine integrated with the healthcare system, so that the balance shifts from the knowledge-generating research space to clinical implementation and thus a direct benefit to patients. This will require buy-in from government health departments and healthcare workers otherwise no traction will be gathered for its roll-out.

The requirements for implementation of genomic medicine seem vast, complex and expensive, but in some countries, some of the pieces are already in place at a basic level through existing health systems or as a result of recently developed research infrastructures. Some basic genomic medicine implementation is likely already occurring as genetic tests exist for some specific diseases that are relevant to African populations. It is important to leverage existing infrastructure to extend these rather than start from nothing, and also to learn from the extensive experience and successes of existing genomic medicine implementations in other countries. The ecosystem can also be broken down into more manageable components that can be developed in parallel with small pilot implementations where some basic infrastructure exists.

The initial and day-to-day recurring costs of data collection and generation for genomic medicine may seem expense, but related efforts in other countries have shown that costs are saved in the long term through faster diagnosis, more precise treatment and the reduction of adverse drug reactions. Government departments can work together to share the costs by taking the responsibility for individual components. A collaboration between government departments would be beneficial as the ecosystem requires elements of health, science, education and policy. The department of health can build the clinical infrastructure, the department of education can train the stakeholders and the department of science and innovation can promote new knowledge generation that is translational and drive the development of innovative health products. There is also a role to play for industry and medical insurance companies who can bring products to market and fund clinical tests for their clients, respectively. Through national coordination of efforts and taking small steps with pilot projects, this large, unimaginable undertaking can be realised, particularly in some countries with reasonable infrastructure already in place.

Section 11. Appendices

Appendix 1. Examples of Genetic and Genomics Studies on Noncommunicable Diseases

Cardiovascular Diseases (CVDs)

Several GWAS studies (Kathiresan et al. 2009; CARDIoGRAMplusC4D Consortium et al. 2013) on CVDs have localised SNPs shown to contribute to different cardiovascular diseases. The International Consortium for Blood Pressure (ICBP) (Evangelou et al. 2018) and the UK Biobank CardioMetabolic Consortium (BP working group) (Warren et al. 2017) have located more than 500 variants that can explain the physiology of blood pressure; among which were targets of antihypertensive drugs indicating that these variants could be druggable. Many of these SNPs were found to be linked to transcription binding sites in vascular tissue, smooth muscle and kidney tissue, all of which are important in controlling blood pressure. GWAS loci for hypertension have been found in the uromodulin (UMOD) gene, which regulates the Na-K cotransporter NKCC2. Mutations are associated with kidney disease and hypertension. A review by (Shukla et al. 2019) provides a comprehensive list of many genes and SNP markers found to be related to susceptibility to CVDs. Of particular interest are those that influence blood lipids levels. Mutations in the ABCA1 gene, which codes for an ATP-binding cassette transporter, impair cholesterol removal and therefore enhance atheroma formation. Other genes involved in triglyceride synthesis have also been described. The MLXIPL gene codes for a transcription factor, with the typical DNA binding leucine zipper motif, that activates promoters of genes responsible for triglycerides synthesis. It is specific for promoters with carbohydrate response elements. Variant forms of the genes result in deregulation of triglycerides synthesis and elevated blood levels (Kooner et al. 2008), increasing risks of CVDs.

GWAS for coronary arterial disease (CAD), have localised variants which impact lipid metabolism, vascular function and innate immunity. Obesity and type 2 diabetes, well known risk factors for CAD, result in chronic inflammation in adipocytes and cause disruption of normal lipid metabolism (Nazare et al. 2012). The CARDIoGRAMplusC4D UK Biobank studies have provided evidence for the involvement of more than 300 variants that could explain 21.2 % of CAD. Most of these, however, were from European populations. Asian and African populations might not share the same SNPs. Attempts to determine the roles of these variants in disease have focused on understanding their effects on functions and regulation of key genes using an omics approach. With data from eQTLs and DEPICT analysis, the genes with SNPs that have a causal effect were identified, and include: APOE, PCSK9, ANGPTL4 and SORT1, all of which are core genes of lipid metabolism. Three genes (TRIM5, CCM2 and FNDC3B) related to immunometabolism were identified as being responsible for enhanced inflammation, increased production of pro-inflammatory cytokines and increased vascular permeability. TRIM5 is therefore crucial in linking the innate immunity and metabolism in CAD (Hughes et al. 2018).

Hypercholesterolemia is an inherited condition that is linked to mutations in LDL receptor (LDLR gene) resulting in a reduced binding of LDL (Ference et al. 2017) and high blood LDL levels. This leads to increased risks of atherosclerosis and CVDs. On the other hand, a mutation in convertase subtilisin/kexin type 9 serine protease gene (PCSK9) has been found to reduce plasma levels of LDL cholesterol, with a resultant decreased risk of coronary heart disease (Cohen et al. 2006). Notably, this variant was predominant in individuals of African ancestry, and has led to anti-PCSK9 therapies to reduce risk of coronary heart disease in other populations (El Khoury et al. 2017).

Cancer

Protein kinases play a role in cancer and can be specifically targeted by drugs to reduce their activity. Identification of upregulated genes in cancers means that their protein products are drug-targetable (e.g. JAK/STAT inhibitors) to inhibit or diminish their activities. JAK or Janus kinases are associated with cytokine receptors and phosphorylate tyrosine residues during signaling pathways. Once phosphorylated, these proteins bind to STAT (signal transducer and activator of transcription) transcription factors which are then able to increase transcription of specific genes or repress that of others. Variant forms of JAK with gain of function are responsible for several myeloproliferative disorders and cancers. Treatment with small molecules acting as JAK inhibitors have been successfully used in many cases (Vainchenker et al. 2018)the search for JAK2 inhibitors continued with the discovery that the other driver mutations (\n CALR\n and\n MPL\n, and a second generation of tyrosine kinase inhibitors has been developed.

A comprehensive analysis of gastrointestinal (GI) adenocarcinomas and their associated genomic alterations has provided insight into which genes and proteins have an important role in pathogenesis (Liu et al. 2018). Genes frequently

mutated in GI tumours are *FBXW7*, *SMAD2*, *SOX9*, *PCBP1, ATM*, *PZP*, *CACNA1C*, and *FBN3.* GI Adenocarcinomas have high frequency of CpG island hypermethylation and a high density of somatic mutations. Amplification of genes is considered gain-of-function while deletions and indels are assumed to lead to loss-of-function, e.g. *CDK6, GATA4, GATA6* are amplified while *APC* and *SOX9* are deleted.

Decreased survival rate for different types of cancer has been linked to mutations in genes *PTEN, TP53* and *BAP1*. Similarly, differential chromosome loss (e.g. 3q, 1, 2, 6, 7, 10, 17) or gain (5q) are characteristic of different forms of renal cancers (Ricketts et al. 2018). Chromosomal translocations lead to repositioning of gene segments which can result in activation of enzyme activities such as with the fibroblast growth factor receptor 1 (FGFR1) gene at chromosome 8p11. Reciprocal translocation with chromosome 9 centrosomal protein gene (*CEP110*) forms a constitutively active tyrosine kinase associated with several myeloproliferative disorders.

Appendix 2. Public Databases Used for Interpreting Next-Generation OMICs data

Resource	Name and URL	Description	African Data
Population databases	Exome aggregation consortium (Exac) (http://exac.broadinstitute.org/)	The Exome Aggregation Consortium (ExAC) is a coalition of investigators seeking to aggregate and harmonise exome sequencing data from a wide variety of large-scale sequencing projects, and to make summary data available for the wider scientific community.	**YES:** African/ African-American data are available for 5,203 Samples (Male samples: 1,888 Female samples : 3,315)
	The Genome Aggregation Database (gnomAD) (http://gnomad.broadinstitute.org/)	The Genome Aggregation Database (gnomAD) is a resource developed by an international coalition of investigators, with the goal of aggregating and harmonising both exome and genome sequencing data from a wide variety of large-scale sequencing projects, and making summary data available for the wider scientific community.	**YES:** African/ African-American data are available. In the latest data release (gnomAD v3), 21 042 whole genome sequence data are available for individuals with African ancestry. In the gnomAD v2 release, there were, 8128 Exomes and 4359 Genomes.
	1000 Genomes (http://www.internationalgenome.org/)	The 1000 Genomes Project represents the largest public catalogue of human variation and genotype data. The final data set contains data for 2,504 individuals from 26 populations.The goal of the 1000 Genomes Project was to find most genetic variants with frequencies of at least 1% in the studied populations.	**YES:** African/ African-American data are available for the following African sub-populations:Yoruba in Ibadan, Nigeria (YRI)Luhya in Webuye, Kenya (LWK) Gambian in Western Divisions in the Gambia (GWD)Mende in Sierra Leone (MSL)Esan in Nigeria (ESN)Americans of African Ancestry in SW USA (ASW) African Caribbeans in Barbados (ACB).
	Exome Variant server (http://evs.gs.washington.edu/EVS/)	The goal of the NHLBI GO Exome Sequencing Project (ESP) is to discover novel genes and mechanisms contributing to heart, lung and blood disorders by pioneering the application of next-generation sequencing of the protein coding regions of the human genome across diverse, richly-phenotyped populations and to share these datasets and findings with the scientific community to extend and enrich the diagnosis, management and treatment of heart, lung and blood disorders.	**YES:** African/ African-American data are available. The last version contains data for 2203 African-American samples.
	The Human Gene Mutation Database (HGMD) (http://www.hgmd.cf.ac.uk/ac/index.php)	Represents an attempt to collate all known (published) gene lesions responsible for human inherited disease	**NO:** African data available
	dbSNP (https://www.ncbi.nlm.nih.gov/projects/SNP/)	A public-domain archive for a broad collection of simple genetic polymorphisms. This collection of polymorphisms includes single-base nucleotide substitutions (SNPs), small-scale multi-base deletions or insertions and retroposable element insertions and microsatellite repeat variations (STRs)	**YES:** African/ African-American data are available for the following African sub-populations:Samples collected from North Africa (including the Sahara desert), East Africa (south to the Equator), Levant, and the Persian Gulf. Sub-Saharan nations bordering the Atlantic north of the Congo River and central/southern Atlantic island nations.
	NCBIb GTR genetic testing registry (https://www.genetests.org)	A genetic testing information resource targeted to health care providers and supported by the National Institutes of Health, strives to incorporate genetic testing into patient care.	**YES:** African/ African-American data are available for the following African sub-populations:Laboratories from South Africa, Egypt and Morocco are participating in the Genetic Testing registry
	Leiden open variant database (links to many locus-specific databases) (http://www.lovd.nl/3.0/home)	A flexible, freely available tool for Gene-centered collection and display of DNA variants. LOVD 3.0 extends this idea to also provide patient-centered data storage and storage of NGS data, even of variants outside of genes. LOVD is open source, released under the GPL license, and is actively being improved.	**NO:** African data available
	OMIM: The Online Mendelian Inheritance in Man (OMIM) (https://www.ncbi.nlm.nih.gov/omim)	The Online Mendelian Inheritance in Man (OMIM) is a comprehensive compendium of human genes and genetic phenotypes that is freely available and updated daily. OMIM contains information on all known mendelian disorders and over 15,000 genes and focuses on the relationship between phenotype and genotype.	**YES:** African/ African-American data are available .

Resource	Name and URL	Description	African Data
Oncology databases	The cancer genome atlas (TCGA) (http://cancergenome.nih.gov/)	A landmark cancer genomics programme, molecularly characterised over 20,000 primary cancer and matched normal samples spanning 33 cancer types. This joint effort between the National Cancer Institute and the National Human Genome Research Institute began in 2006, bringing together researchers from diverse disciplines and multiple institutions.	**YES:** African/ African-American data are available. Of the 5729 samples, 12% (n = 660) were black and/or African American.
	OncoKB (http://oncokb.org/#/)	A precision oncology knowledge base and contains information about the effects and treatment implications of specific cancer gene alterations. It is developed and maintained by the Knowledge Systems group in the Marie Josée and Henry R. Kravis Center for Molecular Oncology at Memorial Sloan Kettering Cancer Center (MSK), in partnership with Quest Diagnostics and Watson for Genomics, IBM.	**NO:** African data available
	Clinical knowledgebase (CKB) (https://www.jax.org/clinical-genomics/ckb)	The CKB is a dynamic digital resource for interpreting complex cancer genomic profiles. Join thousands of clinicians and researchers across the globe, saving time and finding valuable information that connects cancer variants to therapies and clinical trials.	**NO:** African data available
	My cancer genome (https://www.mycancergenome.org/)	Contains information on the clinical impact of molecular biomarkers in cancer-related genes, proteins, and other biomarker types on the use of anticancer therapies in cancer. This information is derived from FDA labels, NCCN and other professional society guidelines, clinical trials, peer-reviewed publications, and more.	**YES:** African/ African-American data are available.
	Catalogue of somatic mutations in cancer (COSMIC) (http://cancer.sanger.ac.uk/cosmic)	A Catalogue Of Somatic Mutations In Cancer, is the world's largest and most comprehensive resource for exploring the impact of somatic mutations in human cancer.	**YES:** African/ African-American data are available.
Data Interpretation	ClinVar (https://www.ncbi.nlm.nih.gov/clinvar/)	A public archive of relationships between sequence variation and human phenotypes. ClinVar database reports human variation, interpretations of the relationship of that variation to human health and the evidence supporting each interpretation	**YES:** African/ African-American data are available.
	DisGeNet (http://www.disgenet.org)	DisGeNET is a discovery platform containing collections of genes and variants associated to human diseases.	**YES:** Genomics African data are available. Around 2500 variants have African specific data. This information is not directly found on the DisGeNet interface (extracted by text mining).
	MedGen (https://www.ncbi.nlm.nih.gov/medgen/)	A free, comprehensive resource for one-stop access to essential information on phenotypic health topics related to medical genetics as collected from established high-quality sources.	**YES:** African/ African-American data are available.
	GTR (Genetic Testing Registry) (https://www.ncbi.nlm.nih.gov/gtr/)	a National Institutes of Health (NIH)-funded resource for disease attributes, actionable genetic variants, patients resources and professional guidelines. It gives information on Mendelian diseases, pharmacogenetic responses, complex diseases and clinical features. a useful platform where to find the clinical features of genetic diseases.	**YES:** African/ African-American data are available.
	ClinGen (https://clinicalgenome.org/)	NIH-funded resource dedicated to building a central resource that defines the clinical relevance of genes and variants for use in precision medicine and research.	**YES:** African/ African-American data are available.

Re-source	Name and URL	Description	African Data
	PharmaGkB (https://www.pharmgkb.org/)	an NIH-funded resource that provides information about how human genetic variation affects response to medications. PharmGKB collects, curates and disseminates knowledge about clinically actionable gene-drug associations and genotype-phenotype relationships.	**YES:** African/ African-American data are available. Data on African population is extracted from Around 165 studies on different African sub-populations. The extracted data are on around 95 Pharmagenes that have been assessed and studies on Africans.
	African Precision Medicine Portal (APMP)	An H3ABioNet web based portal to collate and curate pharmacogenomics and other genomic medicine metadata specific to African populations, as well as links to existing precision medicine implementation tools and resources in order to support precision medicine initiatives within Africa.	**YES:** African/ African-American data are available. 100% data collected in this portal is on African populations
	The Clinical Exome	Clinical Exome Sequencing is a test for identifying disease-causing DNA variants within the 1% of the genome which codes for proteins (exons) or flanks the regions which code for proteins (splice junctions). Indeed, On average, 20,000 DNA variants are detected per Exome. However, the vast majority of these DNA variants are benign polymorphisms. the clinical Exome consists of a list of selected exons known to carry pathogenic and clinically meaningful mutations	**YES:** African/ African-American data are available.
	Evidence-Based Network for the Interpretation of Germline Mutant Alleles (ENIGMA) (https://enigmaconsortium.org/)	An international initiative that determines how variants in breast-ovarian cancer genes are annotated and classified.	**NO:** African data available
	Critical Assessment of Genome Interpretation (CAGI) (https://genomeinterpretation.org/)	CAGI aims to advance phenotypic interpretation of genomic variation. The CAGI experiments depend on the interrogation of data from people whose information has been collected as part of clinical care, following participation in a research project or biorepository, or from healthy volunteers.	**NO:** African data available

Appendix 3. References

Abacan M, Alsubaie L, Barlow-Stewart K, et al (2019) The Global State of the Genetic Counseling Profession. Eur J Hum Genet EJHG 27:183–197. https://doi.org/10.1038/s41431-018-0252-x

ACMG American College of Medical Genetics and Genomics. https://www.acmg.net/. Accessed 31 Dec 2019

Adeyemo AA, Zaghloul NA, Chen G, et al (2019) ZRANB3 is an African-specific type 2 diabetes locus associated with beta-cell mass and insulin response. Nat Commun 10:3195. https://doi.org/10.1038/s41467-019-10967-7

Adzhubei I, Jordan DM, Sunyaev SR (2013) Predicting Functional Effect of Human Missense Mutations Using PolyPhen-2. Curr Protoc Hum Genet Editor Board Jonathan Haines Al 0 7:Unit7.20. https://doi.org/10.1002/0471142905.hg0720s76

AGMT (2016) African Genomic Medicine Training Initiative (AGMT) | African Genomic Medicine Training Initiative. https://training.h3abionet.org/AGMC_2016/. Accessed 27 Feb 2020

Alfirevic A, Pirmohamed M (2017) Genomics of Adverse Drug Reactions. Trends Pharmacol Sci 38:100–109. https://doi.org/10.1016/j.tips.2016.11.003

All of Us Research Program Investigators, Denny JC, Rutter JL, et al (2019) The "All of Us" Research Program. N Engl J Med 381:668–676. https://doi.org/10.1056/NEJMsr1809937

Alles J, Fehlmann T, Fischer U, et al (2019) An estimate of the total number of true human miRNAs. Nucleic Acids Res 47:3353–3364. https://doi.org/10.1093/nar/gkz097

Amegah AK (2018) Tackling the Growing Burden of Cardiovascular Diseases in Sub-Saharan Africa. Circulation 138:2449–2451. https://doi.org/10.1161/CIRCULATIONAHA.118.037367

Antoniou AC, Pharoah PPD, Smith P, Easton DF (2004) The BOADICEA model of genetic susceptibility to breast and ovarian cancer. Br J Cancer 91:1580–1590. https://doi.org/10.1038/sj.bjc.6602175

Azzariti DR, Riggs ER, Niehaus A, et al (2018) Points to consider for sharing variant-level information from clinical genetic testing with ClinVar. Cold Spring Harb Mol Case Stud 4:. https://doi.org/10.1101/mcs.a002345

Baatjes KJ, Conradie M, Apffelstaedt JP, Kotze MJ (2017) Pharmacogenetics of Aromatase Inhibitors in Endocrine Responsive Breast Cancer: Lessons Learnt from Tamoxifen and CYP2D6 Genotyping. Anticancer Agents Med Chem 17:1805–1813. https://doi.org/10.2174/1871521409666170412124226

Baatjes KJ, Kotze MJ, McCaul M, Conradie M (2019) Baseline bone health status in multi-ethnic South African postmenopausal breast cancer patients at initiation of aromatase inhibitor therapy: A descriptive study. PloS One 14:e0214153. https://doi.org/10.1371/journal.pone.0214153

Baker JL, Shriner D, Bentley AR, Rotimi CN (2017) Pharmacogenomic implications of the evolutionary history of infectious diseases in Africa. Pharmacogenomics J 17:112–120. https://doi.org/10.1038/tpj.2016.78

Barker R (2016) Bioscience - lost in translation? how precision medicine closes the innovation gap, First edition. Oxford University Press, Oxford

Barker RW (2017) Is precision medicine the future of healthcare? Pers Med 14:459–461. https://doi.org/10.2217/pme-2017-0060

Batten JN (2018) How Stratification Unites Ethical Issues in Precision Health. AMA J Ethics 20:798–803. https://doi.org/10.1001/amajethics.2018.798.

Bell RB, Feng Z, Bifulco CB, et al (2018) Immunotherapy. In: Oral, Head and Neck Oncology and Reconstructive Surgery. Elsevier, pp 314–340

Bombard Y, Brothers KB, Fitzgerald-Butt S, et al (2019) The Responsibility to Recontact Research Participants after Reinterpretation of Genetic and Genomic Research Results. Am J Hum Genet 104:578–595. https://doi.org/10.1016/j.ajhg.2019.02.025

Bzdok D, Meyer-Lindenberg A (2018) Machine Learning for Precision Psychiatry: Opportunities and Challenges. Biol Psychiatry Cogn Neurosci Neuroimaging 3:223–230. https://doi.org/10.1016/j.bpsc.2017.11.007

Campbell MC, Tishkoff SA (2008) African Genetic Diversity: Implications for Human Demographic History, Modern Human Origins, and Complex Disease Mapping. Annu Rev Genomics Hum Genet 9:403–433. https://doi.org/10.1146/annurev.genom.9.081307.164258

CARDIoGRAMplusC4D Consortium, Deloukas P, Kanoni S, et al (2013) Large-scale association analysis identifies new risk loci for coronary artery disease. Nat Genet 45:25–33. https://doi.org/10.1038/ng.2480

Carter TC, He MM (2016) Challenges of Identifying Clinically Actionable Genetic Variants for Precision Medicine. J Healthc Eng 2016:. https://doi.org/10.1155/2016/3617572

Chawani MS (2014) A Cross-case Analysis of the Effects of EMR Deployment on Antenatal Care Services in Rural Health Centres in Malawi. J Health Inform Afr 2:. https://doi.org/10.12856/JHIA-2014-v2-i1-89

Chen G, Shriner D, Doumatey AP, et al (2020) Refining genome-wide associated loci for serum uric acid in individuals with African ancestry. Hum Mol Genet 29:506–514. https://doi.org/10.1093/hmg/ddz272

Chen J, Sun M, Adeyemo A, et al (2019) Genome-wide association study of type 2 diabetes in Africa. Diabetologia 62:1204–1211. https://doi.org/10.1007/s00125-019-4880-7

Christensen KD, Vassy JL, Jamal L, et al (2016) Are Physicians Prepared for Whole Genome Sequencing? A Qualitative Analysis. Clin Genet 89:228–234. https://doi.org/10.1111/cge.12626

Chung C-F, Dew K, Cole A, et al (2016) Boundary Negotiating Artifacts in Personal Informatics: Patient-Provider Collaboration with Patient-Generated Data. CSCW Proc Conf Comput-Support Coop Work Conf Comput-Support Coop Work 2016:770–786. https://doi.org/10.1145/2818048.2819926

Cohen JC, Boerwinkle E, Mosley TH, Hobbs HH (2006) Sequence variations in PCSK9, low LDL, and protection against coronary heart disease. N Engl J Med 354:1264–1272. https://doi.org/10.1056/NEJMoa054013

Collins A (2015) The genomic and functional characteristics of disease genes. Brief Bioinform 16:16–23. https://doi.org/10.1093/bib/bbt091

Dahiya M, Hussaini T, Yoshida EM (2019) The revolutionary changes in hepatitis C treatment: A concise review. BC Med J 61:72–77

Dandara C, Masimirembwa C, Haffani YZ, et al (2019) African Pharmacogenomics Consortium: Consolidating pharmacogenomics knowledge, capacity development and translation in Africa. AAS Open Res 2:19. https://doi.org/10.12688/aasopenres.12965.1

Datta P, Philipsen PA, Olsen P, et al (2019) Pigment genes not skin pigmentation affect UVB-induced vitamin D. Photochem Photobiol Sci Off J Eur Photochem Assoc Eur Soc Photobiol 18:448–458. https://doi.org/10.1039/c8pp00320c

de Mendonça VRR, Goncalves MS, Barral-Netto M (2012) The host genetic diversity in malaria infection. J Trop Med 2012:940616. https://doi.org/10.1155/2012/940616

*Dean L (2012) Abacavir Therapy and HLA-B*57:01 Genotype. In: Pratt V, McLeod H, Rubinstein W, et al. (eds) Medical Genetics Summaries. National Center for Biotechnology Information (US), Bethesda (MD)*

Deloitte (2017) Privacy is Paramount: Personal Data Protection in Africa

Delport D, Schoeman R, van der Merwe N, et al (2014) Significance of dietary folate intake, homocysteine levels and MTHFR 677 C>T genotyping in South African patients diagnosed with depression: test development for clinical application. Metab Brain Dis 29:377–384. https://doi.org/10.1007/s11011-014-9506-7

Desmond-Hellmann S (2016) Progress lies in precision. Science 353:731–731. https://doi.org/10.1126/science.aai7598

Dessilly G, Goeminne L, Vandenbroucke A, et al (2018) First evaluation of the Next-Generation Sequencing platform for the detection of HIV-1 drug resistance mutations in Belgium. PLOS ONE 13:e0209561. https://doi.org/10.1371/journal.pone.0209561

Diallo DA, Guindo A (2014) Sickle cell disease in sub-Saharan Africa: stakes and strategies for control of the disease. Curr Opin Hematol 21:210–214. https://doi.org/10.1097/MOH.0000000000000038

Dixon-Woods M, Kocman D, Brewster L, et al (2017) A qualitative study of participants' views on re-consent in a longitudinal biobank. BMC Med Ethics 18:22. https://doi.org/10.1186/s12910-017-0182-0

Dlamini MT, Lessells R, Iketleng T, de Oliveira T (2019) Whole genome sequencing for drug-resistant tuberculosis management in South Africa: What gaps would this address and what are the challenges to implementation? J Clin Tuberc Mycobact Dis 16:100115. https://doi.org/10.1016/j.jctube.2019.100115

Drake TM, Knight SR, Harrison EM, Søreide K (2018) Global Inequities in Precision Medicine and Molecular Cancer Research. Front Oncol 8:346. https://doi.org/10.3389/fonc.2018.00346

Dudbridge F (2013) Power and predictive accuracy of polygenic risk scores. PLoS Genet 9:e1003348. https://doi.org/10.1371/journal.pgen.1003348

Eifring H, Theil R (2005) Linguistics for students of Asian and African languages

El Khoury P, Elbitar S, Ghaleb Y, et al (2017) PCSK9 Mutations in Familial Hypercholesterolemia: from a Groundbreaking Discovery to Anti-PCSK9 Therapies. Curr Atheroscler Rep 19:49. https://doi.org/10.1007/s11883-017-0684-8

Elguero E, Délicat-Loembet LM, Rougeron V, et al (2015) Malaria continues to select for sickle cell trait in Central Africa. Proc Natl Acad Sci U S A 112:7051–7054. https://doi.org/10.1073/pnas.1505665112

Erlich Y, Narayanan A (2014) Routes for breaching and protecting genetic privacy. Nat Rev Genet 15:409–421. https://doi.org/10.1038/nrg3723

Erlich Y, Shor T, Pe'er I, Carmi S (2018) Identity inference of genomic data using long-range familial searches. Science 362:690–694. https://doi.org/10.1126/science.aau4832

Esquela-Kerscher A, Slack FJ (2006) Oncomirs - microRNAs with a role in cancer. Nat Rev Cancer 6:259–269. https://doi.org/10.1038/nrc1840

Evangelou E, Warren HR, Mosen-Ansorena D, et al (2018) Genetic analysis of over 1 million people identifies 535 new loci associated with blood pressure traits. Nat Genet 50:1412–1425. https://doi.org/10.1038/s41588-018-0205-x

Ference BA, Ginsberg HN, Graham I, et al (2017) Low-density lipoproteins cause atherosclerotic cardiovascular disease. 1. Evidence from genetic, epidemiologic, and clinical studies. A consensus statement from the European Atherosclerosis Society Consensus Panel. Eur Heart J 38:2459–2472. https://doi.org/10.1093/eurheartj/ehx144

Fitzgerald G, Prince C, Downing J, et al (2019) Processes and barriers to implementation of point-of-care genotype-guided dosing of warfarin into UK outpatient anticoagulation clinics. Pharmacogenomics 20:599–608. https://doi.org/10.2217/pgs-2019-0021

Forbes Team Insights Contributor How Machine Learning Is Crafting Precision Medicine. In: Forbes. https://www.forbes.com/sites/insights-intelai/2019/02/11/how-machine-learning-is-crafting-precision-medicine/. Accessed 31 Dec 2019

Frank JR, Mungroo R, Ahmad Y, et al (2010) Toward a definition of competency-based education in medicine: a systematic review of published definitions. Med Teach 32:631–637. https://doi.org/10.3109/0142159X.2010.500898

Frey LJ (2018) Data integration strategies for predictive analytics in precision medicine. Pers Med 15:543–551. https://doi.org/10.2217/pme-2018-0035

Friedman DJ, Pollak MR (2016) Apolipoprotein L1 and Kidney Disease in African Americans. Trends Endocrinol Metab TEM 27:204–215. https://doi.org/10.1016/j.tem.2016.02.002

Fuchsberger C, Flannick J, Teslovich TM, et al (2016) The genetic architecture of type 2 diabetes. Nature 536:41–47. https://doi.org/10.1038/nature18642

Gabriels K, Moerenhout T (2018) Exploring Entertainment Medicine and Professionalisation of Self-Care: Interview Study Among Doctors on the Potential Effects of Digital Self-Tracking. J Med Internet Res 20:e10. https://doi.org/10.2196/jmir.8040

Galas DJ, Hood L (2009) Systems Biology and Emerging Technologies Will Catalyse the Transition from Reactive Medicine to Predictive, Personalised, Preventive and Participatory (P4) Medicine. Interdiscip Bio Cent 1:1–4. https://doi.org/10.4051/ibc.2009.2.0006

Gallego Llorente M, Jones ER, Eriksson A, et al (2015) Ancient Ethiopian genome reveals extensive Eurasian admixture throughout the African continent. Science 350:820–822. https://doi.org/10.1126/science.aad2879

Galperin M, Farenc C, Mukhopadhyay M, et al (2018) CD4+ T cell–mediated HLA class II cross-restriction in HIV controllers. Sci Immunol 3:eaat0687. https://doi.org/10.1126/sciimmunol.aat0687

Gameiro GR, Sinkunas V, Liguori GR, Auler-Júnior JOC (2018) Precision Medicine: Changing the way we think about healthcare. Clin Sao Paulo Braz 73:e723. https://doi.org/10.6061/clinics/2017/e723

Genomics England (2018) The journey to 100,000 genomes | Genomics England. https://www.genomicsengland.co.uk/the-journey-to-100000-genomes/. Accessed 31 Dec 2019

Genovese G, Friedman DJ, Ross MD, et al (2010) Association of trypanolytic ApoL1 variants with kidney disease in African Americans. Science 329:841–845. https://doi.org/10.1126/science.1193032

Get Consortium GET Africa – ONE AFRICA, ONE HEALTH, ONE DESTINY. In: GET Afr. Brief. https://www.getafrica.org/. Accessed 31 Dec 2019

Ginsburg GS, Phillips KA (2018) Precision Medicine: From Science to Value. Health Aff Proj Hope 37:694–701. https://doi.org/10.1377/hlthaff.2017.1624

Gomez F, Hirbo J, Tishkoff SA (2014) Genetic variation and adaptation in Africa: implications for human evolution and disease. Cold Spring Harb Perspect Biol 6:a008524. https://doi.org/10.1101/cshperspect.a008524

Gouda HN, Charlson F, Sorsdahl K, et al (2019) Burden of non-communicable diseases in sub-Saharan Africa, 1990–2017: results from the Global Burden of Disease Study 2017. Lancet Glob Health 7:e1375–e1387. https://doi.org/10.1016/S2214-109X(19)30374-2

Granger D. Neil, Vowinkel Thorsten, Petnehazy Thomas (2004) Modulation of the Inflammatory Response in Cardiovascular Disease. Hypertension 43:924–931. https://doi.org/10.1161/01.HYP.0000123070.31763.55

Grant KA, Apffelstaedt JP, Wright CA, et al (2013) MammaPrint Pre-screen Algorithm (MPA) reduces chemotherapy in patients with early-stage breast cancer. South Afr Med J Suid-Afr Tydskr Vir Geneeskd 103:522–526. https://doi.org/10.7196/samj.7223

Grant KA, Myburgh EJ, Murray E, et al (2019) Reclassification of early stage breast cancer into treatment groups by combining the use of immunohistochemistry and microarray analysis. South Afr J Sci 115:. https://doi.org/10.17159/sajs.2019/5461

Grizzle WE, Gunter EW, Sexton KC, Bell WC (2015) Quality management of biorepositories. Biopreservation Biobanking 13:183–194. https://doi.org/10.1089/bio.2014.0105

Gröschel MI, Walker TM, van der Werf TS, et al (2018) Pathogen-based precision medicine for drug-resistant tuberculosis. PLoS Pathog 14:e1007297. https://doi.org/10.1371/journal.ppat.1007297

Grünloh C, Myreteg G, Cajander Å, Rexhepi H (2018) "Why Do They Need to Check Me?" Patient Participation Through eHealth and the Doctor-Patient Relationship: Qualitative Study. J Med Internet Res 20:e11. https://doi.org/10.2196/jmir.8444

Gudbjartsson DF, Sulem P, Helgason H, et al (2015) Sequence variants from whole genome sequencing a large group of Icelanders. Sci Data 2:1–11. https://doi.org/10.1038/sdata.2015.11

Gupta H, Galatas B, Matambisso G, et al (2019) Differential expression of var subgroups and PfSir2a genes in afebrile Plasmodium falciparum malaria: a matched case-control study. Malar J 18:326. https://doi.org/10.1186/s12936-019-2963-z

Gurdasani D, Carstensen T, Tekola-Ayele F, et al (2015) The African Genome Variation Project shapes medical genetics in Africa. Nature 517:327–332. https://doi.org/10.1038/nature13997

Gutierrez M, Pillai G, Felix S, et al (2017) Building Capability for Clinical Pharmacology Research in Sub-Saharan Africa. Clin Pharmacol Ther 102:786–795. https://doi.org/10.1002/cpt.695

Gymrek M, McGuire AL, Golan D, et al (2013) Identifying personal genomes by surname inference. Science 339:321–324. https://doi.org/10.1126/science.1229566

H3Africa Bioinformatics Network (2016) Training Survey Africa 2016. https://training.h3abionet.org/AGMC_2016/wp-content/uploads/2017/01/TrainingSurveyAfrica-Upload.pdf. Accessed 27 Feb 2020

Hoadley KA, Yau C, Hinoue T, et al (2018) Cell-of-Origin Patterns Dominate the Molecular Classification of 10,000 Tumors from 33 Types of Cancer. Cell 173:291-304.e6. https://doi.org/10.1016/j.cell.2018.03.022

Hoffman-Andrews L (2017) The known unknown: the challenges of genetic variants of uncertain significance in clinical practice. J Law Biosci 4:648–657. https://doi.org/10.1093/jlb/lsx038

Huang Z, Shi J, Gao Y, et al (2019) HMDD v3.0: a database for experimentally supported human microRNA-disease associations. Nucleic Acids Res 47:D1013–D1017. https://doi.org/10.1093/nar/gky1010

Hughes MF, Lenighan YM, Godson C, Roche HM (2018) Exploring Coronary Artery Disease GWAs Targets With Functional Links to Immunometabolism. Front Cardiovasc Med 5:. https://doi.org/10.3389/fcvm.2018.00148

Hviid L, Jensen ATR (2015) PfEMP1 - A Parasite Protein Family of Key Importance in Plasmodium falciparum Malaria Immunity and Pathogenesis. Adv Parasitol 88:51–84. https://doi.org/10.1016/bs.apar.2015.02.004

Hyland KM, Dasgupta S, Garber K, et al (2013) MEDICAL SCHOOL CORE CURRICULUM IN GENETICS 2013. 10

Ingram CJE, Raga TO, Tarekegn A, et al (2009) Multiple rare variants as a cause of a common phenotype: several different lactase persistence associated alleles in a single ethnic group. J Mol Evol 69:579–588. https://doi.org/10.1007/s00239-009-9301-y

Inouye M, Abraham G, Nelson CP, et al (2018) Genomic Risk Prediction of Coronary Artery Disease in 480,000 Adults: Implications for Primary Prevention. J Am Coll Cardiol 72:1883–1893. https://doi.org/10.1016/j.jacc.2018.07.079

Ioannidis JPA, Khoury MJ (2018) Evidence-based medicine and big genomic data. Hum Mol Genet 27:R2–R7. https://doi.org/10.1093/hmg/ddy065

Jacobs A, Bassa F, Decloedt EH (2017) A preliminary review of warfarin toxicity in a tertiary hospital in Cape Town, South Africa. Cardiovasc J Afr 28:346–349. https://doi.org/10.5830/CVJA-2017-029

Jodheea-Jutton A, Jheelan-Ramchandur A (2018) The Educational Need Among Healthcare Professionals in Mauritius. Int Ann Med 2:. https://doi.org/10.24087/IAM.2018.2.3.432

Johnson RC, Nelson GW, Troyer JL, et al (2010) Accounting for multiple comparisons in a genome-wide association study (GWAS). BMC Genomics 11:724. https://doi.org/10.1186/1471-2164-11-724

Johnston J (2008) Conflict of Interest in Biomedical Research. In: From Birth to Death and Bench to Clinic: The Hastings Center Bioethics Briefing Book for Journalists, Policymakers, and Campaigns. Mary Crowley, Garrison, NY: The Hastings Center, pp 31–34

Jorgensen AL, Prince C, Fitzgerald G, et al (2019) Implementation of genotype-guided dosing of warfarin with point-of-care genetic testing in three UK clinics: a matched cohort study. BMC Med 17:76. https://doi.org/10.1186/s12916-019-1308-7

Kafatos G, Niepel D, Lowe K, et al (2017) RAS mutation prevalence among patients with metastatic colorectal cancer: a meta-analysis of real-world data. Biomark Med 11:751–760. https://doi.org/10.2217/bmm-2016-0358

Kalia SS, ACMG Secondary Findings Maintenance Working Group, Adelman K, et al (2017) Recommendations for reporting of secondary findings in clinical exome and genome sequencing, 2016 update (ACMG SF v2.0): a policy statement of the American College of Medical Genetics and Genomics. Genet Med 19:249–255. https://doi.org/10.1038/gim.2016.190

Kathiresan S, Willer CJ, Peloso GM, et al (2009) Common variants at 30 loci contribute to polygenic dyslipidemia. Nat Genet 41:56–65. https://doi.org/10.1038/ng.291

Kent JW (2009) Analysis of multiple phenotypes. Genet Epidemiol 33 Suppl 1:S33-39. https://doi.org/10.1002/gepi.20470

Khera AV, Chaffin M, Aragam KG, et al (2018) Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. Nat Genet 50:1219–1224. https://doi.org/10.1038/s41588-018-0183-z

Kimmel SE, Christie J, Kealey C, et al (2008) Apolipoprotein E genotype and warfarin dosing among Caucasians and African Americans. Pharmacogenomics J 8:53–60. https://doi.org/10.1038/sj.tpj.6500445

Kirchner S, Power BJ, Waters AP (2016) Recent advances in malaria genomics and epigenomics. Genome Med 8:92. https://doi.org/10.1186/s13073-016-0343-7

Klein K, Tremmel R, Winter S, et al (2019) A New Panel-Based Next-Generation Sequencing Method for ADME Genes Reveals Novel Associations of Common and Rare Variants With Expression in a Human Liver Cohort. Front Genet 10:7. https://doi.org/10.3389/fgene.2019.00007

Kolata G, Murphy H (2018) The Golden State Killer Is Tracked Through a Thicket of DNA, and Experts Shudder. N. Y. Times

Konstantopoulou M, Belgi A, Griffiths KD, et al (2005) Azathioprine-induced pancytopenia in a patient with pompholyx and deficiency of erythrocyte thiopurine methyltransferase. BMJ 330:350–351

Kooner JS, Chambers JC, Aguilar-Salinas CA, et al (2008) Genome-wide scan identifies variation in MLXIPL associated with plasma triglycerides. Nat Genet 40:149–151. https://doi.org/10.1038/ng.2007.61

Korf BR, Blitzer MG, Demmer LA, et al (2017) Report on the Banbury Summit Meeting on medical genetics training in the genomic era, 23-26 February 2014. Genet Med Off J Am Coll Med Genet 19:. https://doi.org/10.1038/gim.2017.38

Kotze M (2016) Application of advanced molecular technology in the diagnosis and management of genetic disorders in South Africa. South Afr Med J Suid-Afr Tydskr Vir Geneeskd 106:S114-118. https://doi.org/10.7196/SAMJ.2016.v106i6.11012

Kotze MJ, Lückhoff HK, Peeters AV, et al (2015) Genomic medicine and risk prediction across the disease spectrum. Crit Rev Clin Lab Sci 52:120–137. https://doi.org/10.3109/10408363.2014.997930

Kotze MJ, Scholtz CL, Opperman P (2006) Health implications and counselling for paternity testing. South Afr Fam Pract 48:34–34. https://doi.org/10.1080/20786204.2006.10873319

Kotze MJ, van Velden DP, Botha K, et al (2013) Pathology-supported genetic testing directed at shared disease pathways for optimised health in later life. Pers Med 10:497–507. https://doi.org/10.2217/pme.13.43

Kredo T, Gerritsen A, van Heerden J, et al (2012) Clinical practice guidelines within the Southern African Development Community: a descriptive study of the quality of guideline development and concordance with best evidence for five priority diseases. Health Res Policy Syst 10:1. https://doi.org/10.1186/1478-4505-10-1

Krogstad DJ, Diop S, Diallo A, et al (2010) Informed Consent in International Research: The Rationale for Different Approaches. Am J Trop Med Hyg 83:743–747. https://doi.org/10.4269/ajtmh.2010.10-0014

Kromberg JGR, Sizer EB, Christianson AL (2013) Genetic services and testing in South Africa. J Community Genet 4:413–423. https://doi.org/10.1007/s12687-012-0101-5

Krumholz HM (2014) Big Data And New Knowledge In Medicine: The Thinking, Training, And Tools Needed For A Learning Health System. Health Aff (Millwood) 33:1163–1170. https://doi.org/10.1377/hlthaff.2014.0053

Kusters DM, Huijgen R, Defesche JC, et al (2011) Founder mutations in the Netherlands: geographical distribution of the most prevalent mutations in the low-density lipoprotein receptor and apolipoprotein B genes. Neth Heart J Mon J Neth Soc Cardiol Neth Heart Found 19:175–182. https://doi.org/10.1007/s12471-011-0076-6

Lankisch TO, Schulz C, Zwingers T, et al (2008) Gilbert's Syndrome and irinotecan toxicity: combination with UDP-glucuronosyltransferase 1A7 variants increases risk. Cancer Epidemiol Biomark Prev Publ Am Assoc Cancer Res Cosponsored Am Soc Prev Oncol 17:695–701. https://doi.org/10.1158/1055-9965.EPI-07-2517

Laraqui A, Uhrhammer N, Rhaffouli HEL, et al (2015) BRCA genetic screening in Middle Eastern and North African: mutational spectrum and founder BRCA1 mutation (c.798_799delTT) in North African. Dis Markers 2015:194293. https://doi.org/10.1155/2015/194293

Lee S-I, Celik S, Logsdon BA, et al (2018) A machine learning approach to integrate big data for precision medicine in acute myeloid leukemia. Nat Commun 9:42. https://doi.org/10.1038/s41467-017-02465-5

Leitsalu L, Alavere H, Tammesoo M-L, et al (2015) Linking a Population Biobank with National Health Registries—The Estonian Experience. J Pers Med 5:96–106. https://doi.org/10.3390/jpm5020096

Levinsky NG (2002) Nonfinancial Conflicts of Interest in Research. N Engl J Med 347:759–761. https://doi.org/10.1056/NEJMsb020853

Leyl S (2019) An HIV data leak has left a community in fear

Liao Y-Y, Tsai H-C, Chou P-Y, et al (2016) CCL3 promotes angiogenesis by dysregulation of miR-374b/ VEGF-A axis in human osteosarcoma cells. Oncotarget 7:4310–4325. https://doi.org/10.18632/oncotarget.6708

Limdi NA, McGwin G, Goldstein JA, et al (2008) Influence of CYP2C9 and VKORC1 1173C/T genotype on the risk of hemorrhagic complications in African-American and European-American patients on warfarin. Clin Pharmacol Ther 83:312–321. https://doi.org/10.1038/sj.clpt.6100290

Limdi NA, Wadelius M, Cavallari L, et al (2010) Warfarin pharmacogenetics: a single VKORC1 polymorphism is predictive of dose across 3 racial groups. Blood 115:3827–3834. https://doi.org/10.1182/blood-2009-12-255992

Liu Y, Sethi NS, Hinoue T, et al (2018) Comparative Molecular Analysis of Gastrointestinal Adenocarcinomas. Cancer Cell 33:721-735. e8. https://doi.org/10.1016/j.ccell.2018.03.010

Luckhoff HK, Brand T, van Velden DP, et al (2015) Clinical relevance of apolipoprotein E genotyping based on a family history of Alzheimer's disease. Curr Alzheimer Res 12:210–217. https://doi.org/10.2174/1567205012666150302154354

Lynch T, Price A (2007) The effect of cytochrome P450 metabolism on drug response, interactions, and adverse effects. Am Fam Physician 76:391–396

Makarewicz O, Lucas M, Brandt C, et al (2017) Whole Genome Sequencing of 39 Invasive Streptococcus pneumoniae Sequence Type 199 Isolates Revealed Switches from Serotype 19A to 15B. PloS One 12:e0169370. https://doi.org/10.1371/journal.pone.0169370

Man M, Farmen M, Dumaual C, et al (2010) Genetic variation in metabolising enzyme and transporter genes: comprehensive assessment in 3 major East Asian subpopulations with comparison to Caucasians and Africans. J Clin Pharmacol 50:929–940. https://doi.org/10.1177/0091270009355161

Manolio TA, Chisholm RL, Ozenberger B, et al (2013) Implementing genomic medicine in the clinic: the future is here. Genet Med 15:258–267. https://doi.org/10.1038/gim.2012.157

Manrai AK, Funke BH, Rehm HL, et al (2016) Genetic Misdiagnoses and the Potential for Health Disparities. N Engl J Med 375:655–665. https://doi.org/10.1056/NEJMsa1507092

Marais AD, Kotze MJ, Raal FJ, et al (2019) Familial hypercholesterolaemia workshop for leveraging point-of-care testing and personalised medicine in association with the Lipid and Atherosclerosis Society of Southern Africa. Cardiovasc J Afr 30:297–304. https://doi.org/10.5830/CVJA-2019-055

Marelli L, Testa G (2018) Scrutinizing the EU General Data Protection Regulation. Science 360:496–498. https://doi.org/10.1126/science.aar5419

Mars M, Scott RE (2016) WhatsApp in Clinical Practice: A Literature Review. Stud Health Technol Inform 231:82–90

Masimirembwa C, Dandara C, Leutscher PDC (2016) Rolling out Efavirenz for HIV Precision Medicine in Africa: Are We Ready for Pharmacovigilance and Tackling Neuropsychiatric Adverse Effects? Omics J Integr Biol 20:575–580. https://doi.org/10.1089/omi.2016.0120

Mathe E, Olivier M, Kato S, et al (2006) Computational approaches for predicting the biological effect of p53 missense mutations: a comparison of three sequence analysis based methods. Nucleic Acids Res 34:1317–1325. https://doi.org/10.1093/nar/gkj518

Maxwell KN, Hart SN, Vijai J, et al (2016) Evaluation of ACMG-Guideline-Based Variant Classification of Cancer Susceptibility and Non-Cancer-Associated Genes in Families Affected by Breast Cancer. Am J Hum Genet 98:801–817. https://doi.org/10.1016/j.ajhg.2016.02.024

Mayosi BM, Flisher AJ, Lalloo UG, et al (2009) The burden of non-communicable diseases in South Africa. Lancet Lond Engl 374:934–947. https://doi.org/10.1016/S0140-6736(09)61087-4

Mbanefo EC, Ahmed AM, Titouna A, et al (2017) Association of glucose-6-phosphate dehydrogenase deficiency and malaria: a systematic review and meta-analysis. Sci Rep 7:45963. https://doi.org/10.1038/srep45963

McLaren W, Gil L, Hunt SE, et al (2016) The Ensembl Variant Effect Predictor. Genome Biol 17:122. https://doi.org/10.1186/s13059-016-0974-4

Mesko B (2017) The role of artificial intelligence in precision medicine. Expert Rev Precis Med Drug Dev 2:239–241. https://doi.org/10.1080/23808993.2017.1380516

Metzger B, Chambeau L, Begon DY, et al (2011) The human epidermal growth factor receptor (EGFR) gene in European patients with advanced colorectal cancer harbors infrequent mutations in its tyrosine kinase domain. BMC Med Genet 12:144. https://doi.org/10.1186/1471-2350-12-144

Mitropoulos K, Al Jaibeji H, Forero DA, et al (2015) Success stories in genomic medicine from resource-limited countries. Hum Genomics 9:11. https://doi.org/10.1186/s40246-015-0033-3

Mpye KL, Matimba A, Dzobo K, et al (2017) Disease burden and the role of pharmacogenomics in African populations. Glob Health Epidemiol Genomics 2:. https://doi.org/10.1017/gheg.2016.21

Mu Y, Kodidela S, Wang Y, et al (2018) The dawn of precision medicine in HIV: state of the art of pharmacotherapy. Expert Opin Pharmacother 19:1581–1595. https://doi.org/10.1080/14656566.2018.1515916

Mulder N (2017) Development to enable precision medicine in Africa. Pers Med 14:467–470. https://doi.org/10.2217/pme-2017-0055

Mulder NJ, Adebiyi E, Alami R, et al (2016) H3ABioNet, a sustainable pan-African bioinformatics network for human heredity and health in Africa. Genome Res 26:271–277. https://doi.org/10.1101/gr.196295.115

Muzoriana N, Gavi S, Nembaware V, et al (2017) Knowledge, Attitude, and Perceptions of Pharmacists and Pharmacy Students towards Pharmacogenomics in Zimbabwe. Pharm Basel Switz 5:. https://doi.org/10.3390/pharmacy5030036

Mwisongo A, Mehlomakhulu V, Mohlabane N, et al (2015) Evaluation of the HIV lay counselling and testing profession in South Africa. BMC Health Serv Res 15:278. https://doi.org/10.1186/s12913-015-0940-y

National Institutes of Health, USA (2019) Precision Medicine

Nazare J-A, Smith JD, Borel A-L, et al (2012) Ethnic influences on the relations between abdominal subcutaneous and visceral adiposity, liver fat, and cardiometabolic risk profile: the International Study of Prediction of Intra-Abdominal Adiposity and Its Relationship

With Cardiometabolic Risk/Intra-Abdominal Adiposity. Am J Clin Nutr 96:714–726. https://doi.org/10.3945/ajcn.112.035758

Ndadza A, Cindi Z, Makambwa E, et al (2019a) Warfarin Dose and CYP2C Gene Cluster: An African Ancestral-Specific Variant Is a Strong Predictor of Dose in Black South African Patients. Omics J Integr Biol 23:36–44. https://doi.org/10.1089/omi.2018.0174

Ndadza A, Thomford NE, Mukanganyama S, et al (2019b) The Genetics of Warfarin Dose-Response Variability in Africans: An Expert Perspective on Past, Present, and Future. Omics J Integr Biol 23:152–166. https://doi.org/10.1089/omi.2019.0018

Nembaware V, Johnston K, Diallo AA, et al (2019) A framework for tiered informed consent for health genomic research in Africa. Nat Genet. https://doi.org/10.1038/s41588-019-0520-x

Nembaware V, Mulder N, Ramesar RS (2016) Article: Preparing for Genomic Medicine Nurse Training in Africa - Global Research Nurses. In: Glob. Res. Nurses Netw. https://globalresearchnurses.tghn.org/articles/preparing-genomic-medicine-nurse-training-africa/. Accessed 27 Feb 2020

Niinuma Y, Saito T, Takahashi M, et al (2014) Functional characterisation of 32 CYP2C9 allelic variants. Pharmacogenomics J 14:107–114. https://doi.org/10.1038/tpj.2013.22

Nordling L (2017) How the genomics revolution could finally help Africa. Nat News 544:20. https://doi.org/10.1038/544020a

Obeid JS, McGraw CA, Minor BL, et al (2013) Procurement of shared data instruments for Research Electronic Data Capture (REDCap). J Biomed Inform 46:259–265. https://doi.org/10.1016/j.jbi.2012.10.006

Ohta T, Kimura M (1969) Linkage disequilibrium at steady state determined by random genetic drift and recurrent mutation. Genetics 63:229–238

Oloyede R (2018) Privacy And Security In Nigeria's Health-Care Sector - Data Protection - Nigeria. http://www.mondaq.com/Nigeria/x/757682 Data+Protection+PrivacyPrivacy+and+Security+In+Nigerias+Healthcare+Sector. Accessed 13 Aug 2019

Olwagen CP, Adrian PV, Madhi SA (2017) Comparison of traditional culture and molecular qPCR for detection of simultaneous carriage of multiple pneumococcal serotypes in African children. Sci Rep 7:4628. https://doi.org/10.1038/s41598-017-04915-y

Ormond KE, Laurino MY, Barlow-Stewart K, et al (2018) Genetic counseling globally: Where are we now? Am J Med Genet C Semin Med Genet 178:98–107. https://doi.org/10.1002/ajmg.c.31607

Owolabi MO, Akpa OM, Made F, et al (2019) Data Resource Profile: Cardiovascular H3Africa Innovation Resource (CHAIR). Int J Epidemiol 48:366–367g. https://doi.org/10.1093/ije/dyy261

Pasanen MK, Fredrikson H, Neuvonen PJ, Niemi M (2007) Different effects of SLCO1B1 polymorphism on the pharmacokinetics of atorvastatin and rosuvastatin. Clin Pharmacol Ther 82:726–733. https://doi.org/10.1038/sj.clpt.6100220

Patin E, Lopez M, Grollemund R, et al (2017) Dispersals and genetic adaptation of Bantu-speaking populations in Africa and North America. Science 356:543–546. https://doi.org/10.1126/science.aal1988

Patra SK (2018) Africa hosts very few clinical trials. Why this is bad for innovation. In: The Conversation. http://theconversation.com/africa-hosts-very-few-clinical-trials-why-this-is-bad-for-innovation-97390. Accessed 3 Jan 2020

Pennisi E (2007) Human Genetic Variation. Science 318:1842–1843. https://doi.org/10.1126/science.318.5858.1842

Perera MA, Gamazon E, Cavallari LH, et al (2011) The missing association: sequencing-based discovery of novel SNPs in VKORC1 and CYP2C9 that affect warfarin dose in African Americans. Clin Pharmacol Ther 89:408–415. https://doi.org/10.1038/clpt.2010.322

Petrović J, Pešić V, Lauschke VM (2020) Frequencies of clinically important CYP2C19 and CYP2D6 alleles are graded across Europe. Eur J Hum Genet EJHG 28:88–94. https://doi.org/10.1038/s41431-019-0480-8

Pirmohamed M, Burnside G, Eriksson N, et al (2013) A randomised trial of genotype-guided dosing of warfarin. N Engl J Med 369:2294–2303. https://doi.org/10.1056/NEJMoa1311386

Popejoy AB, Fullerton SM (2016) Genomics is failing on diversity. Nature 538:161–164. https://doi.org/10.1038/538161a

Prosperi M, Min JS, Bian J, Modave F (2018) Big data hurdles in precision medicine and precision public health. BMC Med Inform Decis Mak 18:139. https://doi.org/10.1186/s12911-018-0719-2

Rees DC, Williams TN, Gladwin MT (2010) Sickle-cell disease. Lancet Lond Engl 376:2018–2031. https://doi.org/10.1016/S0140-6736(10)61029-X

Richards S, Aziz N, Bale S, 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 Off J Am Coll Med Genet 17:405–424. https://doi.org/10.1038/gim.2015.30

Ricketts CJ, De Cubas AA, Fan H, et al (2018) The Cancer Genome Atlas Comprehensive Molecular Characterisation of Renal Cell Carcinoma. Cell Rep 23:313-326.e5. https://doi.org/10.1016/j.celrep.2018.03.075

Robinson JR, Wei W-Q, Roden DM, Denny JC (2018) Defining Phenotypes from Clinical Data to Drive Genomic Research. Annu Rev Biomed Data Sci 1:69–92. https://doi.org/10.1146/annurev-biodatasci-080917-013335

Robinson PN, Köhler S, Bauer S, et al (2008) The Human Phenotype Ontology: A Tool for Annotating and Analysing Human Hereditary Disease. Am J Hum Genet 83:610–615. https://doi.org/10.1016/j.ajhg.2008.09.017

Rochat B (2005) Role of cytochrome P450 activity in the fate of anticancer agents and in drug resistance: focus on tamoxifen, paclitaxel and imatinib metabolism. Clin Pharmacokinet 44:349–366. https://doi.org/10.2165/00003088-200544040-00002

Rotimi CN, Bentley AR, Doumatey AP, et al (2017) The genomic landscape of African populations in health and disease. Hum Mol Genet 26:R225–R236. https://doi.org/10.1093/hmg/ddx253

Rotimi CN, Tekola-Ayele F, Baker JL, Shriner D (2016) The African diaspora: history, adaptation and health. Curr Opin Genet Dev 41:77–84. https://doi.org/10.1016/j.gde.2016.08.005

Sabat AJ, Budimir A, Nashev D, et al (2013) Overview of molecular typing methods for outbreak detection and epidemiological surveillance. Euro Surveill Bull Eur Sur Mal Transm Eur Commun Dis Bull 18:20380. https://doi.org/10.2807/ese.18.04.20380-en

SADaCC SickleInAfrica Consortium | https://www.sickleinafrica.org. In: Sick. Afr. Data Coord. Cent. https://www.sickleinafrica.org/. Accessed 31 Dec 2019

Saunders G, Baudis M, Becker R, et al (2019) Leveraging European infrastructures to access 1 million human genomes by 2022. Nat Rev Genet 20:693–701. https://doi.org/10.1038/s41576-019-0156-9

Saver RS (2012) Is it Really All about the Money? Reconsidering Non-Financial Interests in Medical Research. J Law Med Ethics 40:467–481. https://doi.org/10.1111/j.1748-720X.2012.00679.x

Sawe RT, Kerper M, Badve S, et al (2016) Aggressive breast cancer in western Kenya has early onset, high proliferation, and immune cell infiltration. BMC Cancer 16:204. https://doi.org/10.1186/s12885-016-2204-6

Schneidman M, Matu M, Nkengasong J, et al (2018) Building Cross-Country Networks for Laboratory Capacity and Improvement. Clin Lab Med 38:119–130. https://doi.org/10.1016/j.cll.2017.10.009

Schuette C, Steffens D, Witkowski M, et al (2015) The effect of clopidogrel on platelet activity in patients with and without type-2 diabetes mellitus: a comparative study. Cardiovasc Diabetol 14:15. https://doi.org/10.1186/s12933-015-0182-7

Schwarz JM, Cooper DN, Schuelke M, Seelow D (2014) MutationTaster2: mutation prediction for the deep-sequencing age. Nat Methods 11:361–362. https://doi.org/10.1038/nmeth.2890

Scott IA (2018) Machine Learning and Evidence-Based Medicine. Ann Intern Med 169:44–46. https://doi.org/10.7326/M18-0115

Shaffer JG, Doumbia SO, Ndiaye D, et al (2018) Development of a data collection and management system in West Africa: challenges and sustainability. Infect Dis Poverty 7:125. https://doi.org/10.1186/s40249-018-0494-4

Shriner D, Tekola-Ayele F, Adeyemo A, Rotimi CN (2014) Genome-wide genotype and sequence-based reconstruction of the 140,000 year history of modern human ancestry. Sci Rep 4:6055. https://doi.org/10.1038/srep06055

Shukla H, Mason JL, Sabyah A (2019) Identifying genetic markers associated with susceptibility to cardiovascular diseases. Future Sci OA 5:FSO350. https://doi.org/10.4155/fsoa-2018-0031

Sitapati A, Kim H, Berkovich B, et al (2017) Integrated precision medicine: the role of electronic health records in delivering personalised treatment. Wiley Interdiscip Rev Syst Biol Med 9:. https://doi.org/10.1002/wsbm.1378

Soko ND, Masimirembwa C, Dandara C (2016) Pharmacogenomics of Rosuvastatin: A Glocal (Global+Local) African Perspective and Expert Review on a Statin Drug. OMICS J Integr Biol 20:498–509. https://doi.org/10.1089/omi.2016.0114

Sonuga BO, Hellenberg DA, Cupido CS, Jaeger C (2016) Profile and anticoagulation outcomes of patients on warfarin therapy in an urban hospital in Cape Town, South Africa. Afr J Prim Health Care Fam Med 8:e1-8. https://doi.org/10.4102/phcfm.v8i1.1032

Soo CC, Mukomana F, Hazelhurst S, Ramsay M (2017) Establishing an academic biobank in a resource-challenged environment. S Afr Med J 107:486. https://doi.org/10.7196/SAMJ.2017.v107i6.12099

Spurdle AB, Healey S, Devereau A, et al (2012) ENIGMA--evidence-based network for the interpretation of germline mutant alleles: an international initiative to evaluate risk and clinical significance associated with sequence variation in BRCA1 and BRCA2 genes. Hum Mutat 33:2–7. https://doi.org/10.1002/humu.21628

Stark, et al (2019). Integrating Genomics into Healthcare: A Global Responsibility. AJHG, Volume 104, ISSUE 1, P13-20, 2019 DOI:https://doi.org/10.1016/j.ajhg.2018.11.014,

Stewart A (2013) SLCO1B1 Polymorphisms and Statin-Induced Myopathy. PLoS Curr 5:. https://doi.org/10.1371/currents.eogt.d21e7f0c58463571bb0d9d3a19b82203

Strande NT, Riggs ER, Buchanan AH, et al (2017) Evaluating the Clinical Validity of Gene-Disease Associations: An Evidence-Based Framework Developed by the Clinical Genome Resource. Am J Hum Genet 100:895–906. https://doi.org/10.1016/j.ajhg.2017.04.015

Sun H, Yu G (2019) New insights into the pathogenicity of non-synonymous variants through multi-level analysis. Sci Rep 9:1–11. https://doi.org/10.1038/s41598-018-38189-9

Susswein LR, Marshall ML, Nusbaum R, et al (2016) Pathogenic and likely pathogenic variant prevalence among the first 10,000 patients referred for next-generation cancer panel testing. Genet Med 18:823–832. https://doi.org/10.1038/gim.2015.166

Swetha RG, Sekar DKK, Devi ED, et al (2014) Streptococcus pneumoniae Genome Database (SPGDB): A database for strain specific comparative analysis of Streptococcus pneumoniae genes and proteins. Genomics 104:582–586. https://doi.org/10.1016/j.ygeno.2014.09.012

Swiss Personalised Health Network Swiss Personalised Health Network. https://www.sphn.ch/en/about.html. Accessed 31 Dec 2019

Tait AR, Voepel-Lewis T, Chetcuti SJ, et al (2014) Enhancing patient understanding of medical procedures: evaluation of an interactive multimedia program with in-line exercises. Int J Med Inf 83:376–384. https://doi.org/10.1016/j.ijmedinf.2014.01.011

Tavares DA, Handem S, Carvalho RJ, et al (2019) Identification of Streptococcus pneumoniae by a real-time PCR assay targeting SP2020. Sci Rep 9:3285. https://doi.org/10.1038/s41598-019-39791-1

Tavtigian SV, Greenblatt MS, Lesueur F, et al (2008) In silico analysis of missense substitutions using sequence-alignment based methods. Hum Mutat 29:1327–1336. https://doi.org/10.1002/humu.20892

Tekola-Ayele F, Adeyemo A, Chen G, et al (2015) Novel genomic signals of recent selection in an Ethiopian population. Eur J Hum Genet EJHG 23:1085–1092. https://doi.org/10.1038/ejhg.2014.233

The African Academy of Sciences About Us | The AAS. In: AAS. https://www.aasciences.africa/about. Accessed 31 Dec 2019

The Brainstorm Consortium, Anttila V, Bulik-Sullivan B, et al (2018) Analysis of shared heritability in common disorders of the brain. Science 360:eaap8757. https://doi.org/10.1126/science.aap8757

The Precision Medicine Initiative Working Group (NIH) (2015) The Precision Medicine Initiative Cohort Program – Building a Research Foundation for 21st Century Medicine

Thorogood A, Joly Y, Knoppers BM, et al (2014) An implementation framework for the feedback of individual research results and incidental findings in research. BMC Med Ethics 15:88. https://doi.org/10.1186/1472-6939-15-88

Tiffin N (2019) Potential risks and solutions for sharing genome summary data from African populations. BMC Med Genomics 12:152. https://doi.org/10.1186/s12920-019-0604-6

Tiffin N (2018) Tiered informed consent: respecting autonomy, agency and individuality in Africa. BMJ Glob Health 3:e001249. https://doi.org/10.1136/bmjgh-2018-001249

Tiffin N, George A, LeFevre AE (2019) How to use relevant data for maximal benefit with minimal risk: digital health data governance to protect vulnerable populations in low-income and middle-income countries. BMJ Glob Health 4:e001395. https://doi.org/10.1136/bmjgh-2019-001395

Tiihonen T, Vesisenaho M, Sutinen E (2008) Concept on Context: IS Meeting Context in Developing Countries

Tishkoff SA, Reed FA, Friedlaender FR, et al (2009) The Genetic Structure and History of Africans and African Americans. Science 324:1035–1044. https://doi.org/10.1126/science.1172257

Torrorey-Sawe R, van der Merwe N, Mining SK, Kotze MJ. Pioneering informed consent for return of research results to breast cancer patients facing barriers to implementation of genomic medicine: The Kenyan BRCA1/2 testing experience using whole exome sequencing. Front Genet 2020, March 6. https://doi.org/10.3389/fgene.2020.00170.

Tyrer J, Pharoah PDP, Easton DF (2006) The admixture maximum likelihood test: a novel experiment-wise test of association between disease and multiple SNPs. Genet Epidemiol 30:636–643. https://doi.org/10.1002/gepi.20175

Vainchenker W, Leroy E, Gilles L, et al (2018) JAK inhibitors for the treatment of myeloproliferative neoplasms and other disorders. F1000Research 7:82. https://doi.org/10.12688/f1000research.13167.1

van der Merwe N, Bouwens CSH, Pienaar R, et al (2012) CYP2D6 genotyping and use of antidepressants in breast cancer patients: test development for clinical application. Metab Brain Dis 27:319–326. https://doi.org/10.1007/s11011-012-9312-z

van der Merwe N, Kotze MJ (2018) Implementation of a pathology-supported genetic testing strategy to detect common risk factors shared between breast carcinoma and associated comorbidities. In: Women in Science Without Borders 2018, resilience in diversity: School of Tourism and Hospitality, University of Johannesburg, South Africa, 21-23 March 2018 : Proceedings

van der Merwe N, Peeters AV, Pienaar FM, et al (2017) Exome Sequencing in a Family with Luminal-Type Breast Cancer Underpinned by Variation in the Methylation Pathway. Int J Mol Sci 18:. https://doi.org/10.3390/ijms18020467

Van Laethem K, Theys K, Vandamme A-M (2015) HIV-1 genotypic drug resistance testing: digging deep, reaching wide? Curr Opin Virol 14:16–23. https://doi.org/10.1016/j.coviro.2015.06.001

Vaser R, Adusumalli S, Leng SN, et al (2016) SIFT missense predictions for genomes. Nat Protoc 11:1–9. https://doi.org/10.1038/nprot.2015.123

Vasseur E, Quintana-Murci L (2013) The impact of natural selection on health and disease: uses of the population genetics approach in humans. Evol Appl 6:596–607. https://doi.org/10.1111/eva.12045

Verhoef TI, Redekop WK, Langenskiold S, et al (2016) Cost-effectiveness of pharmacogenetic-guided dosing of warfarin in the United Kingdom and Sweden. Pharmacogenomics J 16:478–484. https://doi.org/10.1038/tpj.2016.41

Vermaelen K, Waeytens A, Kholmanskikh O, et al (2018) Perspectives on the integration of Immuno-Oncology Biomarkers and drugs in a Health Care setting. Semin Cancer Biol 52:166–177. https://doi.org/10.1016/j.semcancer.2017.11.011

von Hundelshausen Philipp, Weber Christian (2007) Platelets as Immune Cells. Circ Res 100:27–40. https://doi.org/10.1161/01.RES.0000252802.25497.b7

Vossen MTM, Westerhout EM, Söderberg-Nauclér C, Wiertz EJHJ (2002) Viral immune evasion: a masterpiece of evolution. Immunogenetics 54:527–542. https://doi.org/10.1007/s00251-002-0493-1

Wang L, Pelleymounter L, Weinshilboum R, et al (2010) Very important pharmacogene summary: thiopurine S-methyltransferase.

Pharmacogenet Genomics 20:401–405. https://doi.org/10.1097/FPC.0b013e3283352860

Warnich L, Drögemöller BI, Pepper MS, et al (2011) Pharmacogenomic Research in South Africa: Lessons Learned and Future Opportunities in the Rainbow Nation. Curr Pharmacogenomics Pers Med 9:191–207. https://doi.org/10.2174/187569211796957575

Warnich L, Kotze MJ, Groenewald IM, et al (1996) Identification of three mutations and associated haplotypes in the protoporphyrinogen oxidase gene in South African families with variegate porphyria. Hum Mol Genet 5:981–984. https://doi.org/10.1093/hmg/5.7.981

Warren HR, Evangelou E, Cabrera CP, et al (2017) Genome-wide association analysis identifies novel blood pressure loci and offers biological insights into cardiovascular risk. Nat Genet 49:403–415. https://doi.org/10.1038/ng.3768

Wei W-Q, Denny JC (2015) Extracting research-quality phenotypes from electronic health records to support precision medicine. Genome Med 7:41. https://doi.org/10.1186/s13073-015-0166-y

Weng C, Kahn MG (2016) Clinical Research Informatics for Big Data and Precision Medicine. Yearb Med Inform 211–218. https://doi.org/10.15265/IY-2016-019

Wilkinson MD, Dumontier M, Aalbersberg IJJ, et al (2016) The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3:160018. https://doi.org/10.1038/sdata.2016.18

Wong NW, Chen Y, Chen S, Wang X (2018) OncomiR: an online resource for exploring pan-cancer microRNA dysregulation. Bioinforma Oxf Engl 34:713–715. https://doi.org/10.1093/bioinformatics/btx627

Wonkam A, Angwafo FF (2006) Prenatal diagnosis may represent a point of entry of genetic science in sub-Saharan Africa: a survey on the attitudes of medical students and physicians from Cameroon. Prenat Diagn 26:760–761. https://doi.org/10.1002/pd.1502

Worku EB, Davis AM, Morrow B (2016) A critical review of health research ethical guidelines regarding caregiver consent for HIV research involving minors in South Africa: Ethical and legal issues. South Afr J Bioeth Law 9:78. https://doi.org/10.7196/SAJBL.2016.v9i2.458

World Health Organisation (2014) Joint Inter-Ministerial Policy Dialogue on eHealth Standardisation and Second WHO Forum on eHealth Standardisation and Interoperability

World Health Organisation (2018) World malaria report 2018

Wright CF, Parker M, Lucassen AM (2019) When genomic medicine reveals misattributed genetic relationships—the debate about disclosure revisited. Genet Med 21:97–101. https://doi.org/10.1038/s41436-018-0023-7

Wright GEB, Adeyemo AA, Tiffin N (2014) Informed consent and ethical re-use of African genomic data. Hum Genomics 8:18. https://doi.org/10.1186/s40246-014-0018-7

Wünnemann F, Sin Lo K, Langford-Avelar A, et al (2019) Validation of Genome-Wide Polygenic Risk Scores for Coronary Artery Disease in French Canadians. Circ Genomic Precis Med 12:e002481. https://doi.org/10.1161/CIRCGEN.119.002481

Xie B, Ding Q, Han H, Wu D (2013) miRCancer: a microRNA-cancer association database constructed by text mining on literature. Bioinforma Oxf Engl 29:638–644. https://doi.org/10.1093/bioinformatics/btt014

Yang S, Wu L, Wang A, et al (2017) dbDEMC 2.0: updated database of differentially expressed miRNAs in human cancers. Nucleic Acids Res 45:D812–D818. https://doi.org/10.1093/nar/gkw1079

Shiwei W, Yuan J, Yihui Y, et al (2017) Ventana immunohistochemistry assay for anaplastic lymphoma kinase gene rearrangement detection in patients with non-small cell lung cancer: A meta-analysis. Thorac Cancer 8:471–476. https://doi.org/10.1111/1759-7714.12468

Appendix 5. Authors and contributors

*Contributors are listed in alphabetical order. **
***Bold names and numbers:** Indicates they were section lead*

Name	Surname	Sections*	Country	Expertise
Samah	Ahmed	5	Sudan	Pharmacogenomics, Precision medicine, Structural bioinformatics, NGS data analysis with focus on variant calling and RNA-Seq analysis
Stuart	Ali	2	South Africa	Regulatory requirements; funding; RUO vs. IVD use
Shakuntala	Baichoo	4,6	Mauritius	Computer science, Bioinformatics, Machine learning
Alia	Benkahla	2,5	Tunisia	Expertise on setting a computing infrastructure/ Standard NGS data analysis pipelines implementation
Archana	Bhaw-Luximon	4,6	Mauritius	Advanced polymers; biomaterials; nanomedicine and tissue engineering
Marie France	Chan Sun	6	Mauritius	Medicine, Public Health
Mahamadou	Diakite	5	Mali	Malaria, immunogenetics
Faisal	**Fadlelmola**	2,4,5,**7**	Sudan	Computational genomics
Yasmina	**Fakim**	**3**,5	Mauritius	Genomics, Bioinformatics
Kais	Ghedira	5	Tunisia	Computational biology and bioinformatics
Smita	Goorah	3	Mauritius	Medicine, education
Yosr	**Hamdi**	4,**5**	Tunisia	Cancer genomics and precision oncology
Abha	Jodheea-Jutton	6	Mauritius	Medicine, Public health education
Moses	Joloba	2	Uganda	Molecular biology, Biorepository science
Victor	**Jongeneel**	**2**	Curacao	High-performance computing, networks, computational genomics
Rogers	Kamulegeya	2	Uganda	Biorepository Science, Genomics
Samar	Kassim	5	Egypt	Molecular Biology of Cancer, Nutrigenomics
Maritha	**Kotse**	6,**8**	South Africa	Genomics, Precision/ Personalised Medicine, Biochemical and Molecular genetics, Cardiovascular and Oncogenomics, Variant interpretation, Point-of-care Technology, Return of research results
Judit	Kumuthini	3,4,5	South Africa	Bioinformatics
Jian	Li	5		
Collen	Masimirembwa	2,5	South Africa/Simbabwe	Pharmacogenomics Genotyping Panels/Arrays, Drug concentration determination by LC-MSMS
Nicky	**Mulder**	**1**,2,5, 1-9	South Africa	Bioinformatics, genomics
Zahra	Mungloo-Dilmohamud	4	Mauritius	Computer science, Bioinformatics
Henry	Musinguzi	2	Uganda	Training, Biorepository science, Public Engagement
Victoria	Nembaware	2,7	South Africa	Genomics, Training, Public Engagement
Tulio	de Oliveira	3	South Africa	Pathogen Genomics, Bioinformatics
Sumir	Panji	2,5	South Africa	Big data analysis, computing infrastructure, storage and reproducible science

Name	Surname	Sections*	Country	Expertise
Michael	Pepper	6	South Africa	Medicine, Genetics
Fouzia	Radouani	3,4,5	Morocco	Microbiology, Bioinformatics
Michele	Ramsay	3,5	South Africa	Genetics
Reem	Sallam	5	Egypt	Molecular Biology, Obesity and Nutritional Therapy
Sunita	Santchum	3		
Rania	Sibira	4,5	Sudan	Precision medicine, Cancer genomics, Neuroscience, Bioinformatics
Dan J	Stein	3,5	South Africa	Medicine, Psychiatry
Nicki	**Tiffin**	**4,6,9,** 1-9	South Africa	Health genomics, human genetics, public health, epidemiology, ethics.
Ozlem	Tastan Bishop	3	South Africa	Structural Bioinformatics
Rispah	Torrorey	6	Kenya	Immunology, Bioethics, Next Generation Sequencing, Bioinformatics, Precision medicine and Oncogenomics
Nicole	Van der Merwe	6	South Africa	Genetic counselling, Precision medicine, Oncogenomics, NGS variant interpretation

Suggested citation:

AESA (2020) *A Framework for the Implementation of Genomic Medicine for Public Health in Africa October 2020.* Nairobi: Alliance for Accelerating Excellence in Science in Africa.

About the Africa Scientific Priority Setting Programme

This outcome document is part of the outputs of the science prioritisation exercises being conducted by the African Academy of Sciences under the Empowering Africa Empowering African Ownership Research and Innovation programme (2019-2023). The programme is a collaborative initiative of AESA and AUDA/NEPAD aimed at engaging key stakeholders in the process of identifying and validating scientific priorities for the African continent. Guided by the SDGs, Africa's Agenda 2063 and STISA 2024, scientific priorities for the continent will be identified and disseminated with the aim to inform investment decisions of major actors including African governments, funders, science leaders and other stakeholders to ensure that resources are directed at the critical gaps identified for the continent. This programme is funded by the Bill and Melinda Gates Foundation.

Alliance for Accelerating Excellence in Science in Africa

No. 8 Miotoni Lane, Karen

P.O. Box 24916 – 00502
Nairobi, Kenya

Tel: +254 20 896 0674/5

aasciences

www.aasciences.africa

The African Academy of Sciences

AESA AN INITIATIVE OF THE AAS & AUDA (AFRICAN UNION DEVELOPMENT AGENCY)

AUDA-NEPAD AFRICAN UNION DEVELOPMENT AGENCY