frontiers | Research Topics

Implementation of genomic medicine in africa: One continent, one vision

Edited by
Maritha J. Kotze, Tivani Phosa Mashamba-Thompson and Dawn Stephens

Published in
Frontiers in Genetics

ISSN 1664-8714
ISBN 978-2-83251-151-0
DOI 10.3389/978-2-83251-151-0

About Frontiers

Frontiers is more than just an open access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers journal series

The Frontiers journal series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the *Frontiers journal series* operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews. Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the *Frontiers journals series*: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area.

Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers editorial office: frontiersin.org/about/contact

Implementation of genomic medicine in africa: One continent, one vision

Topic editors

Maritha J. Kotze — National Health Laboratory Service, Tygerberg Hospital, South Africa

Tivani Phosa Mashamba-Thompson — University of Pretoria, South Africa

Dawn Stephens — Technology Innovation Agency (TIA), South Africas

Citation

Kotze, M. J., Mashamba-Thompson, T. P., Stephens, D., eds. (2023). *Implementation of genomic medicine in africa: One continent, one vision*. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-83251-151-0

Table of contents

frontiers | Frontiers in Genetics

TYPE Editorial
PUBLISHED 10 January 2023
DOI 10.3389/fgene.2022.1133118

OPEN ACCESS

EDITED AND REVIEWED BY
Jared C. Roach,
Institute for Systems Biology (ISB),
United States

*CORRESPONDENCE
Maritha J. Kotze,
maritha@sun.ac.za

SPECIALTY SECTION
This article was submitted to Human and Medical Genomics,
a section of the journal
Frontiers in Genetics

RECEIVED 28 December 2022
ACCEPTED 30 December 2022
PUBLISHED 10 January 2023

CITATION
Kotze MJ, Mashamba-Thompson TP and Stephens D (2023), Editorial: Implementation of genomic medicine in Africa: One continent, one vision.
Front. Genet. 13:1133118.
doi: 10.3389/fgene.2022.1133118

Editorial: Implementation of genomic medicine in Africa: One continent, one vision

Maritha J. Kotze[1]*, Tivani P. Mashamba-Thompson[2] and Dawn Stephens[3]

[1]Division of Chemical Pathology, Department of Pathology, Faculty of Medicine and Health Sciences, Stellenbosch University and National Health Laboratory Service, Tygerberg Hospital, Cape Town, South Africa, [2]Faculty of Health Sciences, University of Pretoria, Pretoria, South Africa, [3]Technology Innovation Agency, Pretoria, South Africa

KEYWORDS

pathology, personalised medicine, precision medicine, policy, translational research, genomics platform, data, african genomes

Editorial on the Research Topic
Implementation of genomic medicine in Africa: One continent, one vision

The translation of genetic research from bench to bedside involves multiple choices related to clinician-patient shared decision-making, with ethical implications at every step of the way. Given the evolving evidence base and limited genomic data from African genomes, we were delighted to receive three reviews, one opinion article, three perspectives, six original papers, and a brief research report with a checklist for assessment of the readiness to implement public health genomics. The latter contribution by Jongeneel et al. included survey results previously generated in parallel to the development of a framework for implementation of genomic medicine in Africa, which was published in February 2021 on commission of the African Academy of Sciences (https://www.aasciences.africa/publications/policy-paper-framework-implementation-genomic-medicine-public-health-africa). The Policy Brief summarized this framework for personalized genomic medicine as the foundation of the current translational Research Topic, showcasing collated evidence of applied knowledge in Africa to enable translation of research into clinical practice, as the study endpoint. The wide range of methodologies used, and implementation approaches presented, were evaluated for evidence of transition from population to individualised risk stratification required for the application of personalised genomic medicine.

In the first review submitted, Govender et al. highlighted the need for development of preventive intervention strategies that would delay the onset and slow progression of chronic kidney disease through treating the underlying aetiologies in at-risk individuals. Screening and early detection of disease pathways utilizing biomarkers at critical control points in pharmacogenetic algorithms were recommended for future implementation of targeted therapy and preventative interventions. Zondo et al. furthermore urged the need to better understand the mechanisms of drug availability and metabolism of antiretrovirals, as the effectiveness of therapy depends on adequate drug delivery and availability to sites of HIV infection. Genetic polymorphisms and genital inflammation may influence drug transport and lead to contradictory results in clinical trials, posing a significant implementation barrier that needs to be addressed by pharmacogenetic studies in genetically diverse populations to improve the efficacy of drug dosing and delivery. In

Egypt the Reference Genome Project has initialized the era of personalised medicine as reported by El-Attar et al. These authors highlighted the challenges and first steps taken to fulfil the promise of the "right drug administered to the right patient at the right time". The link between serum 25 (OH) Vitamin D level and breast cancer prognosis was used as an example of where the pathology-supported genetic testing framework implemented as a case study in South Africa may be useful to overcome limitations, given the effect on musculoskeletal health.

In a summary of the current state of esophageal squamous cell carcinoma (ESCC) genomic research in Africa, Simba et al. expressed the opinion that implementation of genomic medicine for ESCC remains elusive due to a lack of multi-omics studies in African populations. The finding that several mutational signatures in ESCC have been linked to environmental exposures necessitates the incorporation of known environmental and lifestyle risk factors when screening for genetic factors. No standardized methods for data analysis and reporting exist to allow the implementation of a targeted multi-disciplinary intervention plan. Small sample sizes and omission of controlling for population substructure in admixed populations are major limitations together with the complexities associated with data sharing, transfer and storage. Similar perspectives were shared by Ghoorah et al. from the Southwest Indian Ocean region that is in urgent need of evidence from local genetic data to assess the benefits of implementing genomics in healthcare. While well-established genetic testing of the major cancer susceptibility genes, *BRCA1* and *BRCA2*, is not yet available in the public sector, transcriptional gene profiling (70-gene MammaPrint test) has been introduced in the private sector following medical scheme reimbursement in Mauritius to prevent chemotherapy overtreatment in patients with early-stage breast cancer. Hurrell et al. provided the rationale for cellular hepatic models as a useful tool to validate African relevant gene-drug interactions. A convincing use case was presented for researchers to improve the transferability of global research findings to an African relevant context. In a further perspective from the same research group, Nkera-Gutabar et al. make a strong argument in favour of the human microbiome as a valuable complementary approach to traditional genomic medicine. Less than 2% of microbiome diversity is explained by host genetics, while environmental factors associated with diet and lifestyle account for ~20% of gut microbiome variance. Rapid transition to an increasingly westernized environment over the last 2 decades has reshaped the human microbiome in parallel with an increased prevalence of non-communicable diseases (NCDs) such as obesity, neurological disorders and cancer.

Translation of the research findings of two cross-sectional microRNA-related epigenetic studies performed in South African patients with type 2 diabetes mellitus (Weale et al.) and hypertension (Matshazi et al.), is hampered by a lack of clinical data on complications pertaining to these NCDs. This led to recommendations to clarify the effect of non-coding RNAs by building on the current knowledge base and expertise developed within the research team. As a result of this project, it may be possible in future to develop a panel of biomarker targets for cardiometabolic risk stratification offering potentially novel prognostic and therapeutic avenues at the individual level. A retrospective study reported by Tshabalala et al. provided a unique and large HLA dataset of South Africans, which may be a useful resource for future research despite significant limitations imposed by data missingness, imbalance of sample sizes and methodological difficulties. In the cross-sectional study from Kenya, Gatua et al. described the cytogenetic and molecular abnormalities of acute myeloid leukemia in ten patients treated at the hemato-oncology unit of Kenyatta National Hospital over a 3-month period. Comparison between targeted next-generation sequencing (NGS) and whole exome sequencing (WES) after variant filtering for African ancestry, showed high concordance with many variants of uncertain clinical significance detected by both methods. Although exon capturing did not cover all regions of interest, WES allowed for detection of clinically relevant variants implicated in drug resistance and comorbidities such as hypercholesterolemia as an added benefit when integrating research and service delivery. Van der Merwe et al. and Van der Merwe et al. reported on targeted *BRCA1/2* gene screening available in the public sector of South Africa since 1998, and stakeholders' views on broad-spectrum WES first used as a discovery tool in the private sector in South Africa, respectively. Both research groups expressed their support for future implementation of first-tier point-of-care screening and online genetic counselling platforms. This will increase access to appropriate genetic testing and with addition of WES in uninformative cases, offer the potential to better differentiate between inherited, lifestyle-triggered and therapy-associated disease manifestation. Low uptake of *BRCA*/other cost-effective genomic tests despite proven clinical utility is concerning.

Implementation of genomic medicine in Africa will remain an elusive goal unless the researchers who best understand the benefits and limitations of their results take action to make personalised genomic medicine a reality. The Open Genome project is just one of many initiatives acting out the insights gained through developing and applying the readiness checklist published by Jongeneel et al. This research translation tool listing eight requirements for implementation of genomic medicine programs in Africa, proved useful to evaluate the practicalities of recommendations made in 1) the *review papers* to bridge existing clinical implementation gaps identified, 2) to assess examples of knowledge application based on previous experiences reflected in the *opinion piece* and *perspectives*, and 3) to screen *original articles* for evidence of inventive steps taken or recommended to spearhead the implementation of genomic healthcare solutions in Africa and beyond. Deep knowledge gained during these studies should be translated responsibly without delay into practical applications that will benefit study participants and the population at large.

We echo the idea set forward by Nkera-Gutabara et al. to position the microbiome as the "second genome", giving researchers a second chance to prioritise the development of technologies and datasets that drive the African health agenda together (pamoja!) in a way that embodies the "One Continent: One vision" ethos of this Research Topic.

Author contributions

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

Acknowledgements

Gabrielle Thompson, Lindiwe Whati, Jan-Roux Kotze and Dr Nerina van der Merwe are acknowledged for their individual contributions during the idea generation and preparation phase of this Research Topic.

Conflict of interest

MK is a non-executive director and shareholder of Gknowmix (Pty) Ltd.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

frontiers in Genetics

ORIGINAL RESEARCH
published: 08 September 2021
doi: 10.3389/fgene.2021.702410

Expression Profiles of Circulating microRNAs in South African Type 2 Diabetic Individuals on Treatment

Cecil J. Weale[1], Don M. Matshazi[1], Saarah F. G. Davids[1], Shanel Raghubeer[1], Rajiv T. Erasmus[2], Andre P. Kengne[3,4], Glenda M. Davison[1] and Tandi E. Matsha[1]**

[1]SAMRC/CPUT/Cardiometabolic Health Research Unit, Department of Biomedical Sciences, Faculty of Health and Wellness Science, Cape Peninsula University of Technology, Cape Town, South Africa, [2]National Health Laboratory Service (NHLS), Division of Chemical Pathology, Faculty of Health Sciences, University of Stellenbosch, Cape Town, South Africa, [3]Non-Communicable Diseases Research Unit, South African Medical Research Council, Cape Town, South Africa, [4]Department of Medicine, University of Cape Town, Cape Town, South Africa

OPEN ACCESS

Edited by:
Dawn Stephens, Technology Innovation Agency (TIA), South Africa

Reviewed by:
Zhitong Bing, Institute of Modern Physics, Chinese Academy of Sciences (CAS), China
Francesco Prattichizzo, MultiMedica Scientific Institute for Research, Hospitalization and Healthcare (IRCCS), Italy

***Correspondence:**
Cecil J. Weale
cecilweale@gmail.com
Tandi E. Matsha
matshat@cput.ac.za

Specialty section:
This article was submitted to Human and Medical Genomics, a section of the journal Frontiers in Genetics

Received: *29 April 2021*
Accepted: *13 August 2021*
Published: *08 September 2021*

Citation:
Weale CJ, Matshazi DM, Davids SFG, Raghubeer S, Erasmus RT, Kengne AP, Davison GM and Matsha TE (2021) Expression Profiles of Circulating microRNAs in South African Type 2 Diabetic Individuals on Treatment. Front. Genet. 12:702410. doi: 10.3389/fgene.2021.702410

Aim: The influence of disease duration and anti-diabetic treatment on epigenetic processes has been described, with limited focus on interactions with microRNAs (miRNAs). miRNAs have been found to play key roles in the regulation of pathways associated with type 2 diabetes mellitus (T2DM), and expression patterns in response to treatment may further promote their use as therapeutic targets in T2DM and its associated complications. We therefore aimed to investigate the expressions of circulating miRNAs (miR-30a-5p, miR-1299, miR-182-5p, miR-30e-3p and miR-126-3p) in newly diagnosed and known diabetics on treatment, in South Africa.

Methods: A total of 1254 participants with an average age of 53.8 years were included in the study and classified according to glycaemic status (974 normotolerant, 92 screen-detected diabetes and 188 known diabetes). Whole blood levels of miR-30a-5p, miR-1299, miR-182-5p, miR-30e-3p and miR-126-3p were quantitated using RT-qPCR. Expression analysis was performed and compared across groups.

Results: All miRNAs were significantly overexpressed in subjects with known diabetes when compared to normotolerant individuals, as well as known diabetics vs. screen-detected ($p < 0.001$). Upon performing regression analysis, of all miRNAs, only miR-182-5p remained associated with the duration of the disease after adjustment for type of treatment (OR: 0.127, CI: 0.018–0.236, $p = 0.023$).

Conclusion: Our findings revealed important associations and altered expression patterns of miR-30a-5p, miR-1299, miR-182-5p, miR-30e-3p and miR-126-3p in known diabetics on anti-diabetic treatment compared to newly diagnosed individuals. Additionally, miR-182-5p expression decreased with increasing duration of T2DM. Further studies are, however, recommended to shed light on the involvement of the miRNA in insulin signalling and glucose homeostasis, to endorse its use as a therapeutic target in DM and its associated complications.

Keywords: South Africa, miR-30a-5p, miR-1299, miR-182-5p, miR-126-3p, treatment, type 2 diabetes

INTRODUCTION

MicroRNAs (miRNAs) are a family of short, noncoding RNA molecules, averaging 22 nucleotides in length, responsible for regulating gene expression by repressing the translation of messenger RNA (mRNA) molecules, as well as by destabilization of the mRNA molecules (Zhou et al., 2015; O'Brien et al., 2018). Since their discovery, miRNAs have been found to play key roles in the regulation of pathways associated with various diseases, including cancers (Liu et al., 2017; Zhang et al., 2019), cardiovascular diseases (Maciejak et al., 2018; Zhu et al., 2019), and diabetes mellitus (DM; Zampetaki et al., 2010; Massaro et al., 2019; Wang et al., 2019). These small noncoding transcripts have been shown to modulate insulin biosynthesis, pancreatic β-cell development and survival, as well as glucose and lipid metabolism (Mao et al., 2013). Investigations have illustrated altered expression levels of miRNAs, such as miR-30a-5p and miR-126-3p across different glycaemic states, suggesting their potential use as novel biomarkers for early detection of diabetes (Qin et al., 2017; Jiménez-Lucena et al., 2018a). With increased exploration of human miRNAs in the setting of disease, strategies for diagnosis have been the primary focus. However, the focus has progressively extended toward assessing miRNA expression levels and treatment of disease (Walayat et al., 2018).

The interactions of medications, such as metformin with epigenetic processes, have been detailed, including influences in histone modifications, as well as DNA methylation. DNA methylation-induced changes have been the most widely addressed, with reduced methylation due to metformin treatment reported at the insulin gene promoter in a beta-cell line cultured using high glucose concentrations (Ishikawa et al., 2015). Similarly, reduced methylation of transporter genes (SLC22A1, SLC22A3 and SLC47A1) was reported in the livers of diabetics on metformin therapy, compared to diabetics not receiving anti-diabetic medication (García-Calzón et al., 2017). Although not as widely investigated, key evidence surrounding the associations between DM therapies and miRNA expression also exists, with reports of metformin-induced alterations in miRNA expression in diabetic humans and mice, due to increased levels of DICER enzymes, which are essential in miRNA processing (Noren Hooten et al., 2016). Additionally, plasma miR-222 has been linked with insulin action, and positive associations have been identified with type 2 diabetes mellitus (T2DM; Coleman et al., 2013). In two separate studies, both insulin infusion and metformin treatment led to reduced circulating levels of the miRNA in patients with T2DM (Coleman et al., 2013; Ortega et al., 2014). In the same way, miRNA expressions aid in the mediation of these processes in disease development and progression, and changes in their expression in response to anti-diabetic treatment may pave the way for new therapeutic strategies. In view of this, we aimed to investigate the expression of a panel of miRNAs, previously demonstrating altered expressions (Matsha et al., 2018), in a South African population with T2DM, receiving either metformin treatment, insulin and/or both. Furthermore, we intended to assess the expression levels in association with the duration of T2DM since the date of first diagnosis. Our findings in addition to other investigations focused on miRNA dysregulation in T2DM development and progression may lay the foundation for new targets for disease management and therapy.

MATERIALS AND METHODS

Ethical Approval

The study was approved by the Research Ethics Committees of the Cape Peninsula University of Technology and Stellenbosch University (NHREC: REC—230, 408–014 and N14/01/003, respectively). Ethical clearance for this sub-study was sought from and granted by the Cape Peninsula University of Technology Ethics Committee (CPUT/HW-REC 2019/H3). The study was conducted in accordance with the Declaration of Helsinki, and all participants voluntarily signed written informed consent after all the procedures were fully explained in their language of choice.

Study Design

This study was cross-sectional in design, and data were obtained from the ongoing Cape Town Vascular and Metabolic Health (VMH) study, as previously described (Kengne et al., 2017). Data collection took place between April 2014 and November 2016, involving only South Africans from Cape Town. In a total of 1989 participants that were enrolled in the VMH study, 1,254 met the inclusion criteria for this study, including 92 screen-detected diabetes, 188 known diabetics on treatment and 974 normotolerant individuals. Similarly, a recently conducted cross-sectional investigation, sourced their sample from a much larger, previously reported cohort, with a total sample size of 218 test subjects (Prattichizzo et al., 2021a). With ours being larger, this then validates the size of the chosen sample in this study. Participants with unknown glucose tolerance status underwent a 75 g oral glucose tolerance test in accordance with WHO guidelines (Alberti and Zimmet, 1998), and study procedures, such as anthropometric and blood pressure measurements, were also assessed.

Biochemical parameters were immediately analysed at an ISO 15189 accredited pathology practice (PathCare, Reference Laboratory, Cape Town, South Africa). Blood glucose levels (mmol/L) were determined using an enzymatic hexokinase method (Beckman AU, Beckman Coulter, South Africa), and HbA1c levels were determined by high-performance liquid chromatography (BioRad Variant Turbo, BioRad, South Africa). Serum insulin was assessed using a paramagnetic particle chemiluminescence assay (Beckman DXI, Beckman Coulter, South Africa). Serum high-density lipoprotein cholesterol (cholesterol HDL-S; mmol/L) was measured by enzymatic immune-inhibition – End Point (Beckman AU, Beckman Coulter, South Africa), and serum low-density lipoprotein cholesterol (cholesterol LDL-S; mmol/L) by enzymatic selective protection – End Point (Beckman AU, Beckman Coulter, South Africa) and serum triglycerides (triglycerides-S; mmol/L) were estimated by glycerol phosphate oxidase-peroxidase, End Point (Beckman AU, Beckman Coulter, South Africa). Ultrasensitive C-reactive

protein (usCRP) was measured by Latex Particle immunoturbidimetry. Serum γ-glutamyltransferase (gamma GT-S) was measured using the International Federation of Clinical Chemistry and Laboratory Medicine standardized reagents on a Beckman AU (Beckman Coulter, South Africa). Serum cotinine was determined using the Competitive Chemiluminescent (Immulite 2000, Siemens, South Africa).

Total RNA Isolation

Total RNA, including miRNAs, was isolated from 3 ml of whole blood, which was collected in Tempus RNA tubes that were stored at −80°C before RNA isolation (Applied Biosystems). The MagMAX™ for Stabilized Blood Tubes RNA Isolation Kit was used to perform the RNA extraction, as per manufacturer's specifications (Life Technologies, South Africa). RNA purity and integrity were then evaluated using a nanodrop (Nanodrop Technologies, Wilmington, United States), and only samples with a concentration > 15 ng/ml, and an OD (optical density) ratio $^{A260}/_{A280} > 1.8$, were deemed adequate for further processing.

cDNA Conversion and Reverse Transcriptase qPCR

Subsequent RNA samples were then converted to cDNA, using the TaqMan Advanced miRNA cDNA Synthesis Kit, following manufacturer guidelines (Applied Biosystems, Thermo Fisher Scientific, South Africa). This protocol converts RNA to cDNA in four separate reactions, namely, poly(A) tailing, adapter ligation, reverse transcription, and finally a miR-Amp reaction. The poly(A) tailing reaction involved the addition of a 3`-adenosine tail to the miRNA, catalysed by the enzyme poly(A) polymerase. 2 µl of each RNA sample was aliquoted into individual wells of a MicroAmp™ Optical 96-Well Reaction Plate. A reaction mix was then prepared with the poly(A) tailing reagents, as per the manufacturer's protocol, after which, 3 µl transferred into each well of the plate. The reaction plate was sealed with adhesion film, mixed and centrifuged. Following that, the plate was incubated in a QuantStudio™ 7 Flex Real-Time PCR System (Applied Biosystems, Life Technologies Corporation, Johannesburg, South Africa). The following settings were used to configure the QuantStudio™ 7 Flex – polyadenylation at 37°C for 45 min, then a stop reaction at 65°C for 10 min and finally an infinite hold step at 4°C. Following incubation, the plate was removed and the adapter ligation step commenced. The miRNA with poly(A) tails underwent adaptor ligation at their 5` end. In accordance with manufacturer guidelines, a new reaction mix was prepared with the adapter ligation reagents, following that 10 µl of the mix was transferred into each well of the reaction plate containing the poly(A) tailing reaction product. After sealing the plate, it was briefly mixed, centrifuged and then incubated in the QuantStudio™ 7 Flex using the following settings – ligation at 16°C for 60 min, followed by an infinite hold step at 4°C. The miRNA was then reverse-transcribed into cDNA. This reaction entailed the binding of a universal RT primer to the 3`-poly(A) tails of the miRNA. Sufficient reaction mix was prepared and 15 µl added to each well of the 96-well reaction plate containing the adapter ligation reaction products. After sealing the plate, it was briefly mixed, centrifuged and incubated – reverse transcription at 42°C for 15 min, then a stop reaction at 85°C for 5 min, followed by an infinite hold at 4°C. The miR-Amp reaction step then followed, in which universal forward and reverse primers amplified the number of cDNA templates present in each sample. The reaction mix was prepared as per manufacturer specifications and 45 µl transferred into a new 96-well reaction plate. A total volume of 5 µl of the reverse transcription reaction product from the previous reaction step was added to each well of the new plate containing the miR-Amp reaction mix and incubation commenced at the following cycling conditions – enzyme activation at 95°C for 5 min, for 1 cycle, then denaturation at 95°C for 3 s, for 14 cycles, thereafter annealing/extension at 60°C for 30 s, for 14 cycles, a stop reaction at 99°C for 10 min, for 1 cycle, and a hold step at 4°C, for 1 cycle. After successful completion of cDNA synthesis, samples were stored at −20°C until required for reverse transcriptase qPCR (RT-qPCR) analysis.

Prior to performing RT-qPCR, resultant cDNA samples were diluted 1:10 for optimum quantitative analysis. Thereafter, miRNA expression levels were evaluated, as per manufacturer instructions, using pre-designed TaqMan Advanced miRNA Assay primers for the investigated miRNAs: miR-30a-5p (assay ID: 479448_mir; catalogue number: A25576), miR-1299 (assay ID: 478696_mir; catalogue number: A25576), miR-182-5p (assay ID: 477935_mir; catalogue number: A25576), miR-30e-3p (assay ID: 478388_mir; catalogue number: A25576) and miR-126-3p (assay ID: 477887_mir; catalogue number: A25576; Applied Biosystems, Thermo Fisher Scientific, Johannesburg, South Africa). Data were obtained as Ct values and normalized to an endogenous control, miR-16-5p (assay ID: 477860_mir; catalogue number: A25576; Applied Biosystems, Johannesburg, Thermo Fisher Scientific, South Africa). Minimal differences were observed in the expression of the endogenous control between varied glycaemic groups, validating its stability, particularly between newly diagnosed vs. treated diabetes (Weale et al., 2020). The $2^{-\Delta Ct}$ method was used to evaluate the miRNA expression level in each sample, whilst the $2^{-\Delta\Delta Ct}$ value was used as the measure of the miRNA expression in each sample analysed compared with the control sample (Livak and Schmittgen, 2001).

Statistical Analysis

Analysis of data was ++performed using SPSS v.25 (IBM Corp, 2011). The data were tested for normality using Normal Q-Q Plots. The results for categorical variables were presented as count (and percentages), whilst continuous variables were presented as mean (and standard deviation) for normally distributed variables, and median with 25th–75th percentiles was presented for skewed distributions. For comparison between glucose tolerance groups, analysis of the (ANOVA) and the Kruskal-Wallis test were used for continuous variables, whilst the chi-square test was used to assess categorical variables. Spearman's partial correlations, adjusted for age, sex and body mass index (BMI), were performed to assess the relationship

between miRNA expression and other variables. Furthermore, multivariate regression analysis was conducted in order to assess differences in miRNA expression with duration of disease. Various models were used, which were Model 1: Crude; Model 2: adjusted for age and sex; Model 3: adjusted for age, sex and type of medication; and Model 4: adjusted for adjusted for age, sex and type of medication, serum total cholesterol (cholesterol-S), usCRP, log gamma GT-S and log S-creatinine. A *value of* $p < 0.05$ was used to characterize statistically significant results.

RESULTS

Basic Characteristics of Participants

As illustrated in **Table 1**, participants were 53.8 years old on average, with the majority being female (72.7%), and most being normotolerant ($n = 974$). As expected, glycaemic parameters (fasting blood glucose, 2-h blood glucose and HbA1c) were significantly higher in the screen-detected and known diabetes groups vs. the normotolerant group ($p < 0.001$). Body mass index (BMI), waist circumference and hip circumference were significantly higher in both screen-detected DM and known DM, compared to the normotolerant group ($p < 0.001$). Additionally, lipid variables, such as triglycerides-S and cholesterol LDL-S, increased significantly across the glycaemic groups ($p < 0.001$), whilst cholesterol HDL-S exhibited a significant reduction from the normotolerant group through to the known DM group ($p = 0.039$). Both systolic blood pressure and diastolic blood pressure measurements were observed to increase significantly across the glycaemic groups (all of which, $p < 0.001$). Inflammatory markers usCRP and gamma GT-S were significantly higher in the diabetic groups in contrast to the normotolerant group, with reduced levels in known DM vs. screen-detected. Of the known diabetics on treatment, 165 (87.8%) were on oral medication, whilst 23 (12.2%) were on either both oral and/or insulin treatment.

Relative miRNA Expression

All miRNAs were significantly overexpressed in subjects with known diabetes when compared to normotolerant individuals ($p < 0.001$), and additionally, a significant elevation was observed in known diabetics vs. screen-detected ($p < 0.001$; **Figure 1**). miR-30a-5p, miR-1299, miR-182-5p, miR-30e-3p and miR-126-3p were all significantly upregulated in screen-detected DM compared to the normotolerant group ($p \leq 0.013$), with the exception of miR-30e-3p ($p = 0.145$), and miR-30a-5p exhibited the most significant increase in expression ($p = 0.001$) between the two groups.

Comparisons Between miRNA Expression With Treatment and Duration of Disease

The median duration of condition in individuals with known diabetes and on anti-diabetic treatment was 8 years. Although not significant, all other miRNAs with the exception of miR-182-5p were increased in individuals who had had diabetes for over 8 years with miR-30e-3p and miR-126-3p nearing significance, $p \leq 0.095$. miR-182-5p was significantly reduced in individuals who had had diabetes for over 8 years, $p = 0.033$ (**Table 2**). No significant differences were observed between the expression of miRNAs and the type of treatment, that is, oral or combination of oral and/or insulin, $p > 0.05$ (**Table 2**).

Partial Spearman's Correlation Coefficients Adjusted for Age, Sex and Body Mass Index

Spearman's non-parametric correlations, adjusted for age, sex and BMI, were conducted in order to identify significant associations between the expression of the selected miRNAs and other parameters (**Table 3**). All miRNAs observed significant positive correlations with each other (all, $p \leq 0.002$), with the exception of miR-1299 and miR-126-3p ($p = 0.075$). Correlations with glucose indices revealed significant negative correlations between miR-30a-5p and miR-30e-3p with fasting blood glucose ($r = -0.308$, $p = 0.045$ and $r = -0.315$, $p = 0.040$, respectively). Likewise, miR-30e-3p demonstrated a significant negative correlation with HbA1c, $r = -0.318$, $p = 0.037$. The other target miRNAs, however, did not observe noteworthy correlations with any of the abovementioned glycaemic indices. With respect to obesity markers, only miR-30a-5p and miR-30e-3p correlated significantly and negatively, with triglycerides-S ($r = -0.319$, $p = 0.037$ and $r = -0.371$, $p = 0.014$, respectively), whilst the remaining miRNAs demonstrated no such associations. All miRNAs exhibited significant negative correlations with cholesterol HDL-S (all of which, $p \leq 0.046$), whilst none demonstrated any significant links with cholesterol LDL-S. Although, with respect to cholesterol-S, all miRNAs, with the exception of miR-1299, showed notable negative correlations, all of which were $p \leq 0.036$. Correlations with known inflammatory markers revealed significant negative associations between all miRNAs and both usCRP and gamma GT-S (all, $p \leq 0.046$); miR-1299, however, did not observe any significant correlation with gamma-GT-S. Additionally, all miRNAs demonstrated significant correlations with markers of kidney function, namely, S-creatinine, MDRD eGFR (Modification of Diet in Renal Disease estimated glomerular filtration rate) and CKD-EPI eGFR (Chronic Kidney Disease Epidemiology Collaboration estimated glomerular filtration rate), all of which were $p \leq 0.047$.

Multivariate Regression Analysis of miRNAs for the Duration of Diabetes

The log miRNA $2^{-\Delta Ct}$ was used to perform multivariate regression analysis, in order to demonstrate the relationships between the expression of the investigated miRNAs and the duration of T2DM (**Table 4**). The expressions of miRNA-1,299, miR-182-5p and miR-126-3p were associated with duration of diabetes when the model was adjusted for age and sex [odds ratio (OR) ≥ 0.076, 95% confidence interval (CI): 0.001–0.151, $p \leq 0.046$]; however, after adjustment for type of treatment,

TABLE 1 | Characteristics of participants according to diabetic status.

	Normal	Screened DM	Known DM	*Value of p*
	***n* = 974**	***n* = 92**	***n* = 188**	
miR-30a-5p(2-ΔCt)	0.0252 ± 0.0549	0.0494 ± 0.0916	0.1577 ± 0.2512	<0.001
miR-30a-5p (2-ΔCt)*	0.0086 (0.0023;0.0262)	0.0169 (0.0041;0.0576)	0.0664 (0.011;0.1858)	<0.001
miR-1299 (2-ΔCt)	0.0033 ± 0.0108	0.0042 ± 0.009	0.0264 ± 0.062	<0.001
miR-1299 (2-ΔCt)*	0.0007 (0.0001;0.0023)	0.0009 (0.0003;0.0036)	0.0046 (0.0012;0.0202)	<0.001
miR-182-5p (2-ΔCt)	1.457 ± 2.0279	2.4347 ± 4.742	8.979 ± 11.1365	<0.001
miR-182-5p (2-ΔCt)*	0.7994 (0.3065;1.7989)	1.3223 (0.3966;2.4185)	4.6079 (1.3219;13.0158)	<0.001
miR-30e-3p (2-ΔCt)	0.0047 ± 0.0071	0.0056 ± 0.0067	0.0291 ± 0.0478	<0.001
miR-30e-3p (2-ΔCt)*	0.0019 (0.0006;0.0062)	0.0023 (0.0008;0.0079)	0.0119 (0.0025;0.041)	<0.001
miR-126-3p (2-ΔCt)	1.0375 ± 0.9964	1.6141 ± 1.324	7.2226 ± 7.1824	<0.001
miR-126-3p (2-ΔCt)*	0.7251 (0.2917;1.5107)	1.4188 (0.4404;2.4671)	4.7558 (1.4091;10.7017)	<0.001
Age (years)	45.22 ± 15.3	58.15 ± 10.62	57.88 ± 11.97	<0.001
Sex, n (%)				
Female	688 (70.6)	73 (79.3)	151 (80.3)	0.008
Male	286 (29.4)	19 (20.7)	37 (19.7)	
Body mass index (kg/m^2)	27.4 ± 7.8	31.5 ± 8.0	30.7 ± 6.4	<0.001
Waist circumference (cm)	87.9 ± 16.5	100.2 ± 15.5	99.3 ± 16.5	<0.001
Hip circumference (cm)	100.9 ± 16.5	108.0 ± 15.4	107.0 ± 14.0	<0.001
Systolic blood pressure (mmHg)	131 ± 25	146 ± 26	148 ± 26	<0.001
Diastolic blood pressure (mmHg)	84 ± 15	90 ± 14	88 ± 15	<0.001
Glucose (Fasting) (mmol/L)*	4.7 (4.4;5.1)	7.3 (5.7;8.9)	9.1 (6.6;13.3)	<0.001
Glucose 2 h (mmol/L)*	5.4 (4.5;6.3)	12.8 (11.5;16.6)		<0.001**
HbA1c (%)	5.6 ± 0.5	7.3 ± 1.9	8.9 ± 2.4	<0.001
HbA1c (mmol/mol)	37.5 ± 5.0	56.5 ± 21.2	73.8 ± 26.6	<0.001
Insulin Fasting (mIU/L)*	5.8 (3.7;9.0)	9.4 (5.45;16.6)	9.25 (5.48;15.43)	<0.001
Insulin 2 h (mIU/L)*	30.5 (15.9;53.6)	51.2 (29.2;80.4)		<0.001**
Triglycerides-S (mmol/L)*	1.1 (0.8;1.5)	1.4 (1.1;2.4)	1.6 (1.2;2.2)	<0.001
Cholesterol HDL-S (mmol/L)	1.4 ± 0.4	1.3 ± 0.5	1.3 ± 0.3	0.039
Cholesterol LDL-S (mmol/L)	3.1 ± 1,0	3.5 ± 1.1	3.2 ± 1.1	0.001
Cholesterol-S (mmol/L)	5.0 ± 1.2	5.7 ± 1.3	5.3 ± 1.2	<0.001
usCRP (mg/L)*	3.4 (1.3;7.8)	6.5 (3.2;13.1)	4.9 (2.3;10.3)	<0.001
Gamma GT-S * (IU/L)*	27.0 (19.0;42.0)	42.5 (26.3;76.0)	33.0 (20.0;61.0)	<0.001
S-Creatinine (umol/L)*	59.0 (52.0;68.0)	61.0 (52.0;77.0)	60.5 (52.0;75.0)	0.127
MDRD eGFR (ml/min)	107.3 ± 30.0	93.2 ± 30.7	93.9 ± 32.5	<0.001
CKD-EPI eGFR (ml/min)	106.3 ± 20.6	90.3 ± 23.1	90.8 ± 24.1	<0.001
Education level, n (%)				
<7 years	271 (28)	39 (43.3)	75 (40.3)	<0.001
≥7 years	698 (72)	51 (56.7)	111 (59.7)	
Tobacco use, n (%)				
Non-smoker	393 (42.1)	58 (66.7)	128 (70.3)	<0.001
Current smoker	540 (57.9)	29 (33.3)	54 (29.7)	
Alcohol use, n (%)				
Non-drinker	645 (66.5)	76 (83.5)	161 (86.6)	<0.001
Current drinker	325 (33.5)	15 (16.5)	25 (13.4)	

*Cholesterol HDL-S, serum high-density lipoprotein cholesterol; cholesterol LDL-S, serum low-density lipoprotein cholesterol; cholesterol-S, serum total cholesterol; usCRP, ultrasensitive C-reactive protein; gamma GT-S, serum gamma-glutamyl transferase; S-creatinine, serum creatinine; MDRD eGFR, Modification of Diet in Renal Disease estimated glomerular filtration rate; and CKD-EPI eGFR, Chronic Kidney Disease Epidemiology Collaboration estimated glomerular filtration rate. *data represented as median (25th–75th percentile). **value of p between normal group and screen-detected diabetes only.*

only the expression of miR-182-5p remained independently associated with the duration of the disease (OR: 0.127, CI: 0.018–0.236, $p = 0.023$). Upon further adjustment for usCRP, log gamma GT-S and log S-creatinine, miR-182-5p expression retained significance in its association with duration of T2DM (OR: 0.145, CI: 0.034–0.255, $p = 0.011$); moreover, miR-30e-3p as well as miR-126-3p also exhibited notable associations [(OR: 0.103, CI: 0.007–0.199, $p = 0.036$) and (OR: 0.145, CI: 0.020–0.270, $p = 0.024$), respectively].

DISCUSSION

In this study, the expressions of circulating miRNAs (miR-30a-5p, miR-1299, miR-182-5p, miR-30e-3p and miR-126-3p) were investigated in newly diagnosed and known diabetic individuals on treatment. Our data show that these miRNAs are differentially expressed in individuals with diabetes, as well as between newly diagnosed and those on treatment. All target miRNAs correlated significantly with each other, with the exception of miR-1299

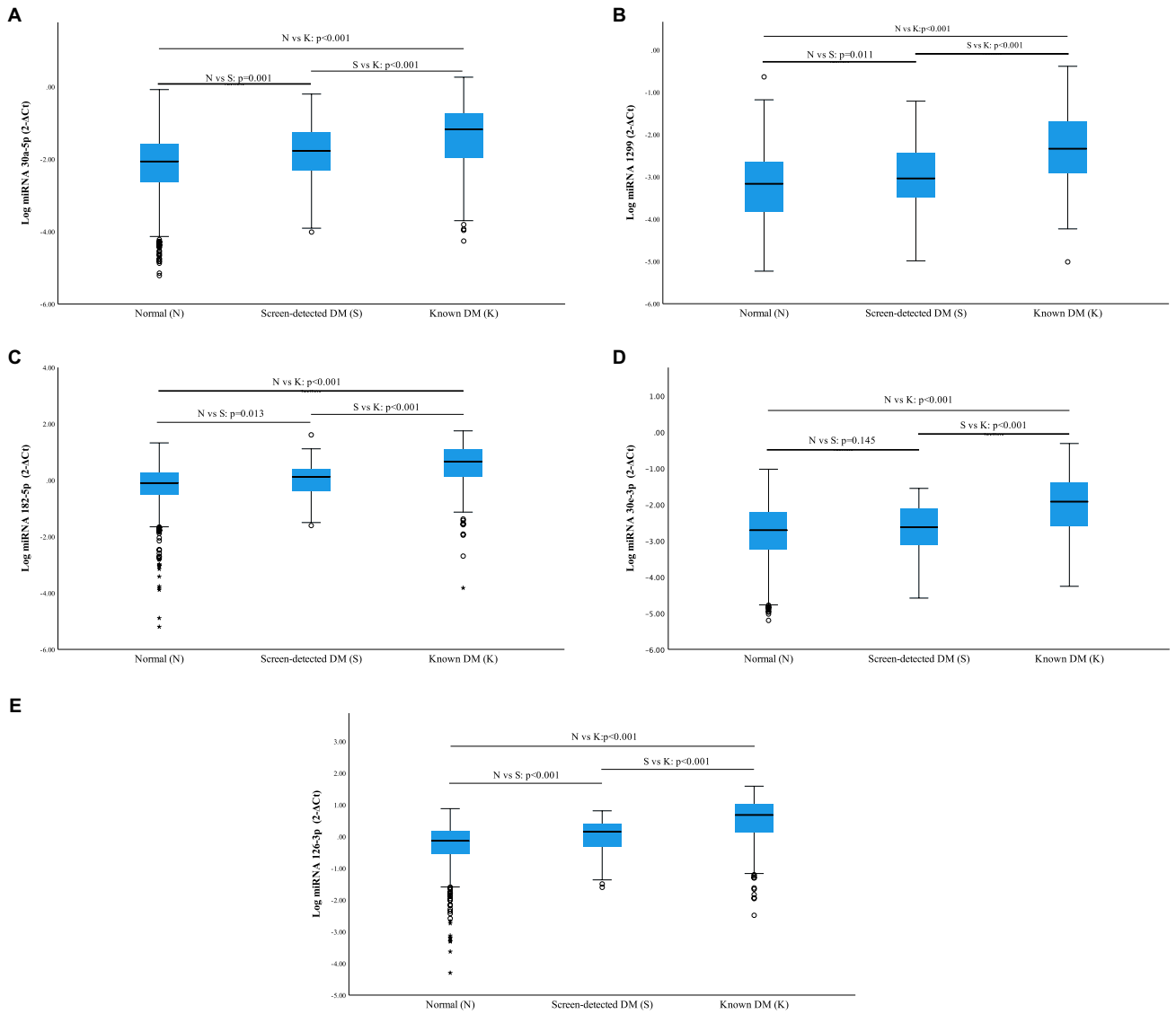


FIGURE 1 | Relative expression of miRNAs according to glycaemic status using log miRNA $2^{-\Delta Ct}$. Normalization was relative to the expression of miR-16-5p. Data are shown as median (25th–75th percentile). **(A)** miR-30a-5p; **(B)** miR-1299; **(C)** miR-182-5p; **(D)** miR-30e-3p; and **(E)** miR-126-3p.

TABLE 2 | Comparisons between miRNA expression with treatment type and duration of disease.

	Types of medication			Duration of disease		
	Both/Insulin	**Oral**	***value of p***	**<8 years**	**≥8 years**	***value of p***
miR-30a-5p ($2^{-w\Delta Ct}$)	0.0641 (0.0292; 0.2787)	0.0664 (0.0098; 0.184)	0.280	0.0581 (0.0069; 0.1787)	0.0607 (0.0182; 0.1906)	0.567
miR-1299 ($2^{-\Delta Ct}$)	0.0048 (0.0015; 0.0233)	0.0045 (0.0011; 0.0196)	0.665	0.0041 (0.0006; 0.0207)	0.0075 (0.0015; 0.0243)	0.155
miR-182-5p ($2^{-\Delta Ct}$)	7.4965 (2.183; 14.1241)	4.5773 (1.1516; 12.9642)	0.427	3.2669 (1.0255; 10.3437)	6.4257 (2.2623; 13.4441)	0.033
miR-30e-3p ($2^{-\Delta Ct}$)	0.0098 (0.003; 0.0375)	0.0128 (0.0025; 0.0431)	0.765	0.0073 (0.0019; 0.0295)	0.0128 (0.0034; 0.0437)	0,065
miR-126-3p ($2^{-\Delta Ct}$)	4.1712 (1.765; 9.1078)	4.8205 (1.3649; 10.9543)	0.873	3.8398 (1.3521; 8.4545)	5.749 (2.0933; 11.674)	0.095

and miR-126-3p. Correlations were illustrated between miR-30a-5p and miR-30e-3p with fasting blood glucose and HbA1c, as well as with triglycerides-S. Furthermore, all miRNAs, except miR-1299, correlated significantly with HDL-cholesterol and serum total cholesterol, as well as with inflammatory markers (usCRP and γ-glutamyltransferase), with similar

TABLE 3 | Partial Spearman's correlation coefficients adjusted for age, sex and body mass index (BMI).

	miR-30a-5p ($2^{-\Delta Ct}$)		miR-1299 ($2^{-\Delta Ct}$)		miR-182-5p ($2^{-\Delta Ct}$)		miR-30e-3p ($2^{-\Delta Ct}$)		miR-126-3p ($2^{-\Delta Ct}$)	
Adjusted for age, BMI and sex	**r**	***value of p***	**r**	***value of p***	**r**	***value of p***	**r**	***value of p***	**r**	***value of p***
miR-30a-5p ($2^{-\Delta Ct}$)	1.000	–	0.497	0.001	0.739	<0.001	0.668	<0.001	0.587	<0.001
miR-1299 ($2^{-\Delta Ct}$)	0.497	0.001	1.000	–	0.582	<0.001	0.452	0.002	0.275	0.075
miR-182-5p ($2^{-\Delta Ct}$)	0.739	<0.001	0.582	<0.001	1.000	–	0.488	0.001	0.620	<0.001
miR-30e-3p ($2^{-\Delta Ct}$)	0.668	<0.001	0.452	0.002	0.488	0.001	1.000	–	0.562	<0.001
miR-126-3p ($2^{-\Delta Ct}$)	0.587	<0.001	0.275	0.075	0.620	<0.001	0.562	<0.001	1.000	–
Waist circumference	−0.078	0.618	−0.002	0.988	−0.134	0.392	−0.067	0.670	−0.200	0.198
Hip circumference (cm)	−0.076	0.628	0.000	0.998	−0.130	0.406	−0.065	0.681	−0.197	0.205
Systolic blood pressure (mmHg)	−0.173	0.266	−0.164	0.293	−0.227	0.143	−0.151	0.335	−0.164	0.292
Diastolic blood pressure (mmHg)	−0.155	0.320	−0.149	0.340	−0.206	0.186	−0.135	0.389	−0.150	0.338
Glucose (Fasting) (mmol/L)	−0.308	0.045	−0.207	0.183	−0.227	0.144	−0.315	0.040	−0.247	0.110
HbA1c (%)	−0.260	0.093	−0.247	0.110	−0.247	0.111	−0.318	0.037	−0.232	0.134
Insulin Fasting (mIU/L)	−0.171	0.273	−0.181	0.245	−0.200	0.198	−0.221	0.154	−0.211	0.175
Triglycerides-S (mmol/L)	−0.319	0.037	−0.281	0.068	−0.297	0.053	−0.371	0.014	−0.280	0.069
Cholesterol HDL-S (mmol/L)	−0.381	0.012	−0.306	0.046	−0.326	0.033	−0.388	0.010	−0.312	0.041
Cholesterol LDL-S (mmol/L)	−0.297	0.053	−0.280	0.069	−0.281	0.068	−0.300	0.051	−0.222	0.153
Cholesterol-S (mmol/L)	−0.385	0.011	−0.293	0.056	−0.320	0.036	−0.434	0.004	−0.326	0.033
usCRP (mg/L)	−0.388	0.010	−0.306	0.046	−0.323	0.035	−0.435	0.004	−0.325	0.033
Gamma GT-S (IU/L)	−0.385	0.011	−0.294	0.056	−0.320	0.037	−0.434	0.004	−0.325	0.033
S-Creatinine (umol/L)	−0.390	0.007	−0.303	0.039	−0.328	0.025	−0.427	0.003	−0.328	0.024
MDRD eGFR (ml/min)	−0.383	0.011	−0.306	0.046	−0.318	0.038	−0.425	0.004	−0.319	0.037
CKD-EPI eGFR (ml/min)	−0.382	0.011	−0.305	0.047	−0.316	0.039	−0.424	0.005	−0.318	0.038

Cholesterol HDL-S, serum high-density lipoprotein cholesterol; cholesterol LDL-S, serum low-density lipoprotein cholesterol; cholesterol-S, serum total cholesterol; usCRP, ultrasensitive C-reactive protein; gamma GT-S – serum gamma-glutamyl transferase; S-creatinine – serum creatinine; MDRD eGFR, Modification of Diet in Renal Disease estimated glomerular filtration rate; and CKD-EPI eGFR, Chronic Kidney Disease Epidemiology Collaboration estimated glomerular filtration rate.

TABLE 4 | Univariate regression analysis of miRNAs for the duration of diabetes.

	B	Std error	95% Confidence Interval (CI)		*value of p*
			Lower	Upper	
Log miR-30a-5p ($2^{-\Delta Ct}$)					
Model 1	−0.021	0.037	−0.094	0.053	0.581
Model 2	0.023	0.036	−0.047	0.094	0.511
Model 3	0.029	0.040	−0.049	0.108	0.464
Model 4	0.049	0.041	−0.032	0.130	0.234
Log miR-1299 ($2^{-\Delta Ct}$)					
Model 1	0.058	0.040	−0.022	0.138	0.151
Model 2	0.076	0.038	0.001	0.151	0.046
Model 3	0.064	0.027	−0.010	0.138	0.088
Model 4	0.068	0.038	−0.007	0.142	0.075
Log miR-182-5p ($2^{-\Delta Ct}$)					
Model 1	0.078	0.040	0.000	0.156	0.051
Model 2	0.101	0.037	0.029	0.174	0.007
Model 3	0.127	0.055	0.018	0.236	0.023
Model 4	0.145	0.056	0.034	0.255	0.011
Log miR-30e-3p ($2^{-\Delta Ct}$)					
Model 1	0.040	0.045	−0.050	0.130	0.384
Model 2	0.070	0.042	−0.013	0.153	0.097
Model 3	0.086	0.047	−0.008	0.179	0.072
Model 4	0.103	0.048	0.007	0.199	0.036
Log miR-126-3p ($2^{-\Delta Ct}$)					
Model 1	0.083	0.043	−0.012	0.177	0.086
Model 2	0.102	0.044	0.014	0.189	0.023
Model 3	0.119	0.062	−0.003	0.241	0.056
Model 4	0.145	0.063	0.020	0.270	0.024

Model 1 Crude; Model 2: adjusted for Age and Sex; Model 3: adjusted for Age, Sex, Type of Medication; and Model 4: adjusted for Age, Sex, Type of Medication, Cholesterol-S, usCRP, log Gamma GT-S and log S-creatinine.

associations exhibited with kidney function indicators, MDRD eGFR and CKD-EPI eGFR. We observed that the duration of the disease is the most determining factor in the expression of these miRNAs. For instance, miR-182-5p was significantly decreased in individuals who had diabetes for over 8 years, but no such differences were observed when anti-diabetic treatment was taken into account. Moreover, miR-1299, miR-182-5p and miR-126-3p were significantly associated with T2DM in regression analysis adjusted for age and sex, but when medication was included in the model, that association was lost for miR-1299 and miR-126-3p and only retained by miR-182-5p.

Advances in whole-genome sequencing as well as epigenome profiling technologies have contributed toward accelerated growth in research aimed at unveiling the roles of epigenetics in human disease (Hawkins et al., 2010; Laird, 2010; Afjeh and Ghaderian, 2013; Koperski et al., 2017) particularly the roles of miRNAs in DM (Zampetaki et al., 2010; Karolina et al., 2011; Jiménez-Lucena et al., 2018a). miRNAs target-specific messenger RNA (mRNA), hindering translation of the subsequent proteins the mRNA encode for (Bartel, 2009), as such, understanding these potential targets and the resultant knock-on effects of altered miRNA expression, is essential in unravelling the intricate subclinical molecular mechanisms that drive disease development. For instance, a previous rat-based study demonstrated miR-30a-5p to be a mediator of beta-cell dysfunction-induced glucotoxicity, by suppressing the expression of Beta2/NeuroD, a gene responsible for regulating pancreatic beta-cell maintenance and insulin gene transcription (Noguchi et al., 2005; Kim et al., 2013). The investigation further established that an overexpression of the miRNA led to pancreatic beta-cell dysfunction (Kim et al., 2013). Additionally, in a follow-up study spanning 5 years, Jiménez-Lucena and co-workers assessed plasma miRNAs, *via* RT-qPCR, in non-diabetic patients, in order to ascertain whether the combined use of these miRNAs with clinical and biochemical measures was more beneficial in predicting type 2 diabetes (Jiménez-Lucena et al., 2018b). Nine dysregulated miRNAs were identified, which when added to HbA1c, yielded acceptable predictive values in early detection of type 2 diabetes, one of which was miR-30a-5p. Elevated expression of the miRNA was found to be associated with increased risk of T2DM development, with higher levels of miR-30a-5p observed in participants diagnosed with T2DM in the first year of follow-up, in comparison with those diagnosed in later years. These findings were consistent with ours, in that we also observed marked miR-30a-5p expression in newly diagnosed diabetics vs. the normotolerant; however, the cross-sectional nature of our investigation thwarted us in accurately assessing the effects of prolonged disease on its expression.

Another promising candidate, miR-126-3p, has been heralded as a potential biomarker for early detection of dysglycaemia (Zampetaki et al., 2010; Zhang et al., 2013; Rezk et al., 2016) and its role as a regulator of angiogenesis is known (Wang et al., 2008). A recent report has shown insulin receptor substrate 1 (IRS1) as a target for miR-126-3p, demonstrating overexpression of the miRNA mediating marked reduction in IRS1 in smooth muscle cells. IRS1 plays an essential role in insulin signalling in hyperglycaemic states; thus, a reduction in its expression negatively impacts glucose metabolism (Tryggestad, 2019). miR-126-3p expression has also been investigated in complications associated with long-term diabetes, such as diabetic kidney disease (DKD; KDIGO, 2013; Bijkerk et al., 2015; Assmann et al., 2018). Bijkerk et al. (2015) determined significantly increased expression of circulating miR-126-3p in the plasma of subjects with diabetic nephropathy (with a mean duration of diabetes of 29 years), in comparison with those without (Bijkerk et al., 2015). In accordance with these findings, a separate systematic review, which aimed at identifying studies comparing miRNA expression in persons with DKD and diabetics without the complication, showed that of the studies included, miR-126-3p was found to be overexpressed in either whole blood, plasma, urine or kidney tissue in patients with DKD (Assmann et al., 2018). These reports may offer a plausible explanation for our findings of altered miR-126-3p expression in diabetic states (both screen-detected and long-term diabetes), as well as the significant associations demonstrated with known measures of renal dysfunction. Although, in our sample, the median duration of diabetes was less, the findings of Bijkerk et al. (2015) corroborate the moderate associations we observed between dysregulated miR-126-3p and duration

of diabetes. Moreover, Prattichizzo et al aimed to identify specific miRNA signatures from plasma extracellular vesicles (EVs) of diabetic individuals with and without complications (Prattichizzo et al., 2021b). These EVs are minute membranous particles that have been shown to transport miRNAs, with postulations that they may elucidate inter-tissue communications in disease states, through the expression patterns of the miRNA signatures they shuttle (Guay and Regazzi, 2017; Xiao et al., 2019). Prattichizzo and colleagues identified miRNAs influenced by diabetes, one of which was miR-126-3p; however, the miRNA did not witness significant differential expression in diabetics with complications (mean duration of disease being 20 years). In our study, miR-126-3p expression observed similar susceptibility to dysglycaemic states; furthermore, although we do not report data pertaining to diabetic complications, we did however demonstrate moderate relationships between long-term T2DM (median duration of 8 years) and its expression. In a similarly study, which included a treatment aspect, significantly decreased miR-126-3p levels were found in untreated diabetics as well as diabetics on metformin treatment, in comparison with levels in the control group (Ghai et al., 2019). We, however, did not witness any notable associations when treatment was accounted for. In light of these findings, it is relevant to mention that whole blood, which was the starting sample in our study, contains collective miRNAs from circulating plasma-specific EVs, blood-cell-derived miRNAs, as well as those bound to carriers such as HDL and RNA-binding proteins (Prattichizzo et al., 2021b). Pre-analytical centrifugation of samples is necessary for retrieving plasma from whole blood, and this may influence the expression profiles of miRNAs, due to the removal of blood-cell-specific miRNAs (Felekkis and Papaneophytou, 2020). All things considered, although whole blood provides better yields of miRNA, caution should still be approached when cross-comparing studies, due to the expression variations between tissues (Prattichizzo et al., 2021b).

Reports have also described the involvement of candidate miRNA, miR-182, in regulating glucose homeostasis, chiefly by targeting FOXO1 (Karolina et al., 2011; Zhou et al., 2014; Zhang et al., 2016). Mammalian cells express four FOXO variants, namely, FOXO1, FOXO3, FOXO4 and FOXO6, of which FOXO1 is the most abundantly expressed in liver, adipose tissue as well as in pancreatic beta-cells (Kitamura et al., 2002). FOXO1 is vital in regulating pancreatic beta-cell replication and differentiation, as well as maintenance in states of metabolic stress (Kitamura, 2013). Furthermore, FOXO1 is involved in stimulating hepatic gluconeogenesis in states of hypoglycaemia *via* the phosphoinositide-3-kinase/protein kinase B (PI3K/Akt) signalling pathway, whilst in hyperglycaemia, insulin signalling *via* insulin, like growth factor-1 (IGF-1) and its receptor (IGF-1R), stimulates PI3K/Akt-dependent phosphorylation of FOXO1, leading to subsequent suppression of gluconeogenesis (Tsuchiya and Ogawa, 2017). In this regard, in early stages of dysglycaemia, increased levels of miR-182 in individuals with prediabetes compared to those with newly diagnosed diabetes have been linked with attempted inhibition of hepatic gluconeogenesis, whilst pathological downregulation in diabetes promoted gluconeogenesis (Karolina et al., 2011). With respect to the direct impact of long-term hyperglycaemia and treatment, there is a paucity of information relating to miR-182-5p expression. Albeit, in a cancer-related study, Li et al. (2012) aimed at evaluating miRNAs connected with the anti-tumour effects of metformin in human gastric cancer cells. Upregulation of miR-182 was revealed in the cultured cells and cancer tissues treated with metformin when compared to untreated cells (Li et al., 2012). In our study, we did not observe any influences of anti-diabetic treatment (oral metformin and/or insulin) on the expression of the selected miRNAs, but rather marked differences with duration of the disease, especially with respect to miR-182-5p. miR-182-5p has, however, demonstrated links with diabetes-related complications, such as DKD, which is a common complication of diabetes (Ming et al., 2019). Ming and co-workers illustrated overexpression of miR-182-5p in the podocytes of individuals with DKD, as opposed to non-diabetic controls, alluding that this increased expression was linked to a reduction in CD2-associated protein, which is crucial in podocyte apoptosis and subsequent development of chronic kidney disease in diabetic persons (Ming et al., 2019). Ming et al. (2019) did not include information pertaining to the duration of diabetes in their study subjects, however, their findings corroborating the associations we found between miR-182-5p expression and key indicators of kidney function. Based on literature, our study mirrors previous findings in that miR-182-5p expression is indeed altered in hyperglycaemic states, though the underlying molecular mechanisms remain unclear. Furthermore, in light of the scarcity of reports pertaining to long-term T2DM and miR-182-5p expression, we have shown that the time span of disease in known diabetics has more of an influence on the expression of the miRNA, as opposed to treatment. This justifies the need for further research ventures, in order to ascertain these associations, particularly in diabetics with complications.

This is the first study of its kind to be conducted in an African setting, and findings may contribute toward curbing the increasing burden of T2DM in Africa. The study was, however, limited by the disproportionate representation of normotolerant vs. screen-detected diabetics and known diabetics on treatment. Additionally, the cross-sectional nature of the study limits accurate evaluation of anti-diabetic induced epigenetic changes; hence, longitudinal studies are advised. The absence of information pertaining to diabetes-associated complications in known diabetics hampered our ability to comprehensively assess the influences of long-term diabetes on miRNA expression, as such, it would be more beneficial to include such data in future. A further limitation was that the expression of the target miRNAs was assessed, without predicting and measuring potential target mRNA levels. Hence, future recommendations would be to explore and measure potential target mRNA levels to elucidate the knock-on effects of altered miRNA expression on downstream proteins. This would aid in more accurate comparisons of the underlying molecular interplays between controlled and uncontrolled diabetics.

CONCLUSION

Our study has revealed important associations and altered expression patterns of miR-30a-5p, miR-1299, miR-182-5p, miR-30e-3p and miR-126-3p in individuals with diabetes on anti-diabetic treatment compared to newly diagnosed cases. Furthermore, we show that miR-182-5p in particular decreases with increasing duration of T2DM. Longitudinal and functional investigations are recommended to elucidate the involvement of the miRNA in insulin signalling and glucose homeostasis, to endorse its use as a therapeutic target in DM and its associated complications.

DATA AVAILABILITY STATEMENT

The datasets presented in this article are not readily available because of the terms of consent to which participants agreed, but are available from the principal investigator of the main study on reasonable request. Requests to access the datasets should be directed to TM, matshat@cput.ac.za.

ETHICS STATEMENT

This study involving human participants was reviewed and approved by the Research Ethics Committees of the Cape Peninsula University of Technology and Stellenbosch University (NHREC: REC—230,408–014 and N14/01/003, respectively). Ethical clearance for this sub-study was sought from and granted by the Cape Peninsula University of Technology Ethics Committee (CPUT/HW-REC 2019/H3). The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

All authors contributed significantly to this project. CW: wrote the first draft, experimental procedures, data analysis and interpretation. DM: experimental procedures, data analysis and interpretation. SD: recruitment and screening of cohort, statistical analysis and interpretation of data. SR: interpretation of data, editing and revising it for intellectual content. RE: conception, interpretation of the data, revising it for intellectual content and final approval of the version to be published. AK: conception, interpretation of the data, revising it for intellectual content and final approval of the version to be published. GD: editing and revising it for intellectual content and final approval of the version to be published. TM: conception and design of the study, analysis and interpretation of the data, revising it for intellectual content and final approval of the version to be published.

FUNDING

This research project was supported by grants from the South African Medical Research Council (SAMRC), with funds from National Treasury under its Economic Competitiveness and Support Package (MRC-RFA-UFSP-01-2013/VMH Study), South African National Research Foundation (SANRF) grant no. 115450). Any opinions, findings, conclusions or recommendations expressed in this article are those of the authors, and the SAMRC and/or SANRF do not accept any liability in this regard.

ACKNOWLEDGMENTS

We thank the Bellville South community and their community Health Forum for supporting the study.

REFERENCES

Afjeh, S. S. A., and Ghaderian, S. M. H. (2013). The role of microRNAs in cardiovascular disease. *Int. J. Mol. Cell. Med.* 2, 50–57.

Alberti, K. G. M. M., and Zimmet, P. Z. (1998). Definition, diagnosis and classification of diabetes mellitus and its complications. Part 1: diagnosis and classification of diabetes mellitus. Provisional report of a WHO consultation. *Diabet. Med.* 15, 539–553. doi: 10.1002/(SICI)1096-9136(199807)15:7<539::AID-DIA668>3.0.CO;2-S

Assmann, T. S., Recamonde-Mendoza, M., de Souza, B. M., Bauer, A. C., and Crispim, D. (2018). MicroRNAs and diabetic kidney disease: systematic review and bioinformatic analysis. *Mol. Cell. Endocrinol.* 477, 90–102. doi: 10.1016/j.mce.2018.06.005

Bartel, D. P. (2009). MicroRNAs: target recognition and regulatory functions. *Cell* 136, 215–233. doi: 10.1016/j.cell.2009.01.002

Bijkerk, R., Duijs, J. M. G. J., Khairoun, M., Ter Horst, C. J. H., Van Der Pol, P., Mallat, M. J., et al. (2015). Circulating MicroRNAs associate with diabetic nephropathy and systemic microvascular damage and normalize after simultaneous pancreas-kidney transplantation. *Am. J. Transplant.* 15, 1081–1090. doi: 10.1111/ajt.13072

Coleman, C. B., Lightell, D. J., Moss, S. C., Bates, M., Parrino, P. E., and Woods, T. C. (2013). Elevation of miR-221 and-222 in the internal mammary arteries of diabetic subjects and normalization with metformin. *Mol. Cell. Endocrinol.* 374, 125–129. doi: 10.1016/j.mce.2013.04.019

Felekkis, K., and Papaneophytou, C. (2020). Challenges in using circulating micro-RNAs as biomarkers for cardiovascular diseases. *Int. J. Mol. Sci.* 21:561. doi: 10.3390/ijms21020561

García-Calzón, S., Perfilyev, A., Männistö, V., de Mello, V. D., Nilsson, E., Pihlajamäki, J., et al. (2017). Diabetes medication associates with DNA methylation of metformin transporter genes in the human liver. *Clin. Epigenetics* 9:102. doi: 10.1186/s13148-017-0400-0

Ghai, V., Kim, T. K., Etheridge, A., Nielsen, T., Hansen, T., Pedersen, O., et al. (2019). Extracellular vesicle encapsulated microRNAs in patients with type 2 diabetes are affected by metformin treatment. *J. Clin. Med.* 8:617. doi: 10.3390/jcm8050617

Guay, C., and Regazzi, R. (2017). Exosomes as new players in metabolic organ cross-talk. *Diabetes Obes. Metab.* 19, 137–146. doi: 10.1111/dom.13027

Hawkins, D., Hon, G., Ren, B., Hawkins, R. D., Hon, G. C., and Ren, B. (2010). Next-generation genomics: an integrative approach. *Nat. Rev. Genet.* 11, 476–486. doi: 10.1038/nrg2795

IBM Corp. (2011). IBM SPSS Statistics for Windows, Version 20.0. Armonk, NY: IBM Corp.

Ishikawa, K., Tsunekawa, S., Ikeniwa, M., Izumoto, T., Iida, A., Ogata, H., et al. (2015). Long-term pancreatic beta cell exposure to high levels of glucose but not palmitate induces DNA methylation within the insulin gene promoter and represses transcriptional activity. *PLoS One* 10:e0115350. doi: 10.1371/journal.pone.0115350

Jiménez-Lucena, R., Camargo, A., Alcalá-Diaz, J. F., Romero-Baldonado, C., Luque, R. M., Van Ommen, B., et al. (2018a). A plasma circulating miRNAs profile predicts type 2 diabetes mellitus and prediabetes: from the CORDIOPREV study. *Exp. Mol. Med.* 50, 1–12. doi: 10.1038/s12276-018-0194-y

Jiménez-Lucena, R., Rangel-Zúñiga, O. A., Alcalá-Díaz, J. F., López-Moreno, J., Roncero-Ramos, I., Molina-Abril, H., et al. (2018b). Circulating miRNAs as predictive biomarkers of type 2 diabetes mellitus development in coronary heart disease patients from the CORDIOPREV study. *Mol. Ther. Nucleic. Acids.* 12, 146–157. doi: 10.1016/j.omtn.2018.05.002

Karolina, D. S., Armugam, A., Tavintharan, S., Wong, M. T. K., Lim, S. C., Sum, C. F., et al. (2011). MicroRNA 144 impairs insulin signaling by inhibiting the expression of insulin receptor substrate 1 in type 2 diabetes mellitus. *PLoS One* 6:e22839. doi: 10.1371/annotation/698b7123-174f-4a09-95c9-fd6f5017d622

Kidney Disease Improving Global Outcomes (KDIGO) (2013). KDIGO 2012 clinical practice guideline for the evaluation and management of chronic kidney disease. *Kidney. Int.* 3, 1–150. doi: 10.1038/ki.2013.243

Kengne, A. P., Erasmus, R. T., Levitt, N. S., and Matsha, T. E. (2017). Alternative indices of glucose homeostasis as biochemical diagnostic tests for abnormal glucose tolerance in an African setting. *Prim. Care. Diabetes.* 11, 119–131. doi: 10.1016/j.pcd.2017.01.004

Kim, J. W., You, Y. H., Jung, S., Suh-Kim, H., Lee, I. K., Cho, J. H., et al. (2013). MiRNA-30a-5p-mediated silencing of Beta2/NeuroD expression is an important initial event of glucotoxicity-induced beta cell dysfunction in rodent models. *Diabetologia* 56, 847–855. doi: 10.1007/s00125-012-2812-x

Kitamura, T. (2013). The role of FOXO1 in β-cell failure and type 2 diabetes mellitus. *Nat. Rev. Endocrinol.* 9, 615–623. doi: 10.1038/nrendo.2013.157

Kitamura, T., Nakae, J., Kitamura, Y., Kido, Y., Biggs, W. H., Wright, C. V. E., et al. (2002). The forkhead transcription factor Foxo1 links insulin signaling to Pdx1 regulation of pancreatic beta cell growth. *J. Clin. Invest.* 110, 1839–1847. doi: 10.1172/JCI200216857

Koperski, Ł., Kotlarek, M., Świerniak, M., Kolanowska, M., Kubiak, A., Górnicka, B., et al. (2017). Next-generation sequencing reveals microRNA markers of adrenocortical tumors malignancy. *Oncotarget* 8, 49191–49200. doi: 10.18632/oncotarget.16788

Laird, P. W. (2010). Principles and challenges of genome-wide DNA methylation analysis. *Nat. Rev. Genet.* 11, 191–203. doi: 10.1038/nrg2732

Li, W., Yuan, Y., Huang, L., Qiao, M., and Zhang, Y. (2012). Metformin alters the expression profiles of microRNAs in human pancreatic cancer cells. *Diabetes Res. Clin.* 96, 187–195. doi: 10.1016/j.diabres.2011.12.028

Liu, M. M., Li, Z., Han, X. D., Shi, J. H., Tu, D. Y., Song, W., et al. (2017). MiR-30e inhibits tumor growth and chemoresistance via targeting IRS1 in breast cancer. *Sci. Rep.* 7:15929. doi: 10.1038/s41598-017-16175-x

Livak, K. J., and Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2-ΔΔCT method. *Methods* 25, 402–408. doi: 10.1006/meth.2001.1262

Maciejak, A., Kostarska-Srokosz, E., Gierlak, W., Dluzniewski, M., Kuch, M., Marchel, M., et al. (2018). Circulating miR-30a-5p as a prognostic biomarker of left ventricular dysfunction after acute myocardial infarction. *Sci. Rep.* 8:9883. doi: 10.1038/s41598-018-28118-1

Mao, Y., Mohan, R., Zhang, S., and Tang, X. (2013). MicroRNAs as pharmacological targets in diabetes. *Pharmacol. Res.* 75, 37–47. doi: 10.1016/j.phrs.2013.06.005

Massaro, J. D., Polli, C. D., Costa E Silva, M., Alves, C. C., Passos, G. A., Sakamoto-Hojo, E. T., et al. (2019). Post-transcriptional markers associated with clinical complications in type 1 and type 2 diabetes mellitus. *Mol. Cell. Endocrinol.* 490, 1–14. doi: 10.1016/j.mce.2019.03.008

Matsha, T. E., Kengne, A. P., Hector, S., Mbu, D. L., Yako, Y., and Erasmus, R. T. (2018). MicroRNA profiling and their pathways in South African individuals with prediabetes and newly diagnosed type 2 diabetes mellitus. *Oncotarget* 9, 30485–30498. doi: 10.18632/oncotarget.25271

Ming, L., Ning, J., Ge, Y., Zhang, Y., and Ruan, Z. (2019). Excessive apoptosis of podocytes caused by dysregulation of microRNA-182-5p and CD2AP confers to an increased risk of diabetic nephropathy. *J. Cell. Biochem.* 120, 16516–16523. doi: 10.1002/jcb.28911

Noguchi, H., Bonner-Weir, S., Wei, F. Y., Matsushita, M., and Matsumoto, S. (2005). BETA2/NeuroD protein can be transduced into cells due to an arginine-and lysine-rich sequence. *Diabetes* 54, 2859–2866. doi: 10.2337/diabetes.54.10.2859

Noren Hooten, N., Martin-Montalvo, A., Dluzen, D. F., Zhang, Y., Bernier, M., and at al., (2016). Metformin-mediated increase in DICER1 regulates microRNA expression and cellular senescence. *Aging Cell* 15, 572–581. doi: 10.1111/acel.12469

O'Brien, J., Hayder, H., Zayed, Y., and Peng, C. (2018). Overview of microRNA biogenesis, mechanisms of actions, and circulation. *Front. Endocrinol.* 9:402. doi: 10.3389/fendo.2018.00402

Ortega, F. J., Mercader, J. M., Moreno-Navarrete, J. M., Rovira, O., Guerra, E., Esteve, E., et al. (2014). Profiling of circulating microRNAs reveals common microRNAs linked to type 2 diabetes that change with insulin sensitization. *Diabetes. Care.* 37, 1375–1383. doi: 10.2337/dc13-1847

Prattichizzo, F., De Nigris, V., Sabbatinelli, J., Giuliani, A., Castaño, C., Párrizas, M., et al. (2021a). CD31+ extracellular vesicles from patients with type 2 diabetes shuttle a miRNA signature associated with cardiovascular complications. *Diabetes* 70, 240–254. doi: 10.2337/db20-0199

Prattichizzo, F., Matacchione, G., Giuliani, A., Sabbatinelli, J., Olivieri, F., de Candia, P., et al. (2021b). Extracellular vesicle-shuttled miRNAs: a critical appraisal of their potential as nano-diagnostics and nano-therapeutics in type 2 diabetes mellitus and its cardiovascular complications. *Theranostics* 11:1031. doi: 10.7150/thno.51605

Qin, L.-L., An, M.-X., Liu, Y.-L., Xu, H.-C., and Lu, Z.-Q. (2017). MicroRNA-126: a promising novel biomarker in peripheral blood for diabetic retinopathy. *Int. J. Ophthalmol.* 10:530. doi: 10.18240/ijo.2017.04.05

Rezk, N. A., Sabbah, N. A., and Saad, M. S. S. (2016). Role of microRNA 126 in screening, diagnosis, and prognosis of diabetic patients in Egypt. *IUBMB Life* 68, 452–458. doi: 10.1002/iub.1502

Tryggestad, J. B. (2019). 188-LB: MicroRNA-126-3p impacts insulin Signaling in smooth muscle cells. *Diabetes* 68:188-LB. doi: 10.2337/db19-188-LB

Tsuchiya, K., and Ogawa, Y. (2017). Forkhead box class O family member proteins: The biology and pathophysiological roles in diabetes. *J. Diabetes Investig.* 8, 726–734. doi: 10.1111/jdi.12651

Walayat, A., Yang, M., and Xiao, D. (2018). "Therapeutic Implication of miRNA in Human Disease," in. *Antisense Therapy.* eds. S. Sharad and S. Kapur (London, UK: IntechOpen).

Wang, S., Aurora, A. B., Johnson, B. A., Qi, X., McAnally, J., Hill, J. A., et al. (2008). The endothelial-specific microRNA miR-126 governs vascular integrity and angiogenesis. *Dev. Cell.* 15, 261–271. doi: 10.1016/j.devcel.2008.07.002

Wang, J., Wang, G., Liang, Y., and Zhou, X. (2019). Expression profiling and clinical significance of plasma microRNAs in diabetic nephropathy. *J. Diabetes Res.* 2013:5204394. doi: 10.1155/2019/5204394

Weale, C. J., Matshazi, D. M., Davids, S. F. G., Raghubeer, S., Erasmus, R. T., Kengne, A. P., et al. (2020). Circulating miR-30a-5p and miR-182-5p in prediabetes and screen-detected diabetes mellitus. *Diabetes Metab. Syndr.* 13, 5037–5047. doi: 10.2147/DMSO.S286081

Xiao, Y., Zheng, L., Zou, X., Wang, J., Zhong, J., and Zhong, T. (2019). Extracellular vesicles in type 2 diabetes mellitus: key roles in pathogenesis, complications, and therapy. *J. Extracell. Vesicles.* 8:1625677. doi: 10.1080/20013078.2019.1625677

Zampetaki, A., Kiechl, S., Drozdov, I., Willeit, P., Mayr, U., Prokopi, M., et al. (2010). Plasma microRNA profiling reveals loss of endothelial miR-126 and other MicroRNAs in type 2 diabetes. *Circ. Res.* 107, 810–817. doi: 10.1161/CIRCRESAHA.110.226357

Zhang, D., Li, Y., Yao, X., Wang, H., Zhao, L., Jiang, H., et al. (2016). miR-182 regulates metabolic homeostasis by modulating glucose utilization in muscle. *Cell Rep.* 16, 757–768. doi: 10.1016/j.celrep.2016.06.040

Zhang, F. B., Du, Y., Tian, Y., Ji, Z. G., and Yang, P. Q. (2019). MiR-1299 functions as a tumor suppressor to inhibit the proliferation and metastasis of prostate cancer by targeting NEK2. *Eur. Rev. Med. Pharmacol. Sci.* 23, 530–538. doi: 10.26355/eurrev_201901_16865

Zhang, T., Lv, C., Li, L., Chen, S., Liu, S., Wang, C., et al. (2013). Plasma miR-126 is a potential biomarker for early prediction of type 2 diabetes mellitus in susceptible individuals. *Biomed. Res. Int.* 2013:761617. doi: 10.1155/2013/761617

Zhou, J., Meng, Y., Tian, S., Chen, J., Liu, M., Zhuo, M., et al. (2014). Comparative microRNA expression profiles of cynomolgus monkeys, rat, and human reveal that miR-182 is involved in T2D pathogenic processes. *J. Diabetes Res.* 2014:760397. doi: 10.1155/2014/760397

Zhou, J. Y., Xu, B., and Li, L. (2015). A new role for an old drug: metformin targets microRNAs in treating diabetes and cancer. *Drug Dev. Res.* 76, 263–269. doi: 10.1002/ddr.21265

Zhu, L., Chen, T., Ye, W., Wang, J. Y., Zhou, J. P., Li, Z. Y., et al. (2019). Circulating miR-182-5p and miR-5187-5p as biomarkers for the diagnosis of unprotected left main coronary artery disease. *J. Thorac. Dis.* 11, 1799–1808. doi: 10.21037/jtd.2019.05.24

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Genetics

ORIGINAL RESEARCH
published: 14 September 2021
doi: 10.3389/fgene.2021.710438

Circulating Levels of MicroRNAs Associated With Hypertension: A Cross-Sectional Study in Male and Female South African Participants

Don M. Matshazi[1], Cecil J. Weale[1], Rajiv T. Erasmus[2], Andre P. Kengne[3,4], Saarah F. G. Davids[1], Shanel Raghubeer[1], Glenda M. Davison[1] and Tandi E. Matsha[1]**

[1] SAMRC/CPUT/Cardiometabolic Health Research Unit, Department of Biomedical Sciences, Faculty of Health and Wellness Science, Cape Peninsula University of Technology, Cape Town, South Africa, [2] Division of Chemical Pathology, Faculty of Health Sciences, National Health Laboratory Service (NHLS) and Stellenbosch University, Cape Town, South Africa, [3] Non-Communicable Diseases Research Unit, South African Medical Research Council, Cape Town, South Africa, [4] Department of Medicine, University of Cape Town, Cape Town, South Africa

OPEN ACCESS

Edited by:
Dawn Stephens,
Technology Innovation Agency (TIA), South Africa

Reviewed by:
Said El Shamieh,
Beirut Arab University, Lebanon
Zhipeng Liu,
Purdue University, United States

***Correspondence:**
Don M. Matshazi
matshazid@gmail.com
Tandi E. Matsha
matshat@cput.ac.za;
tandimatsha@gmail.com

Specialty section:
This article was submitted to Human and Medical Genomics, a section of the journal Frontiers in Genetics

Received: 16 May 2021
Accepted: 13 August 2021
Published: 14 September 2021

Citation:
Matshazi DM, Weale CJ, Erasmus RT, Kengne AP, Davids SFG, Raghubeer S, Davison GM and Matsha TE (2021) Circulating Levels of MicroRNAs Associated With Hypertension: A Cross-Sectional Study in Male and Female South African Participants. Front. Genet. 12:710438. doi: 10.3389/fgene.2021.710438

MicroRNAs are non-coding, post-transcriptional regulators of gene expression and their dysregulation has been associated with development of various diseases, including hypertension. Consequently, understanding their role in the pathogenesis and progression of disease is essential. Prior research focusing on microRNAs in disease has provided a basis for understanding disease prognosis and offered possible channels for therapeutic interventions. Herein, we aimed to investigate possible differences in the expression profiles of five microRNAs in the blood of participants grouped on the basis of their hypertension status. This was done to elucidate the possible roles played by these microRNAs in the development of hypertension. Using quantitative reverse transcription polymerase chain reaction, we evaluated the expression levels of miR-126-3p, 30a-5p, 182-5p, 30e-3p, and 1299 in the whole blood of 1456 participants, normotensive (n = 573), screen-detected hypertensive (n = 304) and known hypertensive (n = 579). The expression of miR-126-3p and 182-5p was significantly higher in known hypertensives relative to both screen-detected hypertensives and normotensives, and also in screen-detected hypertensives vs normotensives. A significant association between the expression of miR-126-3p, 182-5p, and 30a-5p and known hypertension was also evident. This study demonstrated dysregulated miR-126-3p, 182-5p, and 30a-5p expression in hypertension, highlighting the possible efficacy of these microRNAs as targets for the diagnosis of hypertension as well as the development of microRNA-based therapies.

Keywords: hypertension, microRNA, RT-qPCR, Africa, blood pressure, non-coding RNA

INTRODUCTION

Hypertension (HPT) is a complex and multifactorial disease responsible for considerable loss of life globally (Campbell et al., 2016a; Mills et al., 2020). It is an important, modifiable risk factor for cardiovascular disease [1, 3], whose prevalence varies globally, but is on an upward trajectory in sub-Saharan Africa (Campbell et al., 2016a; Mills et al., 2016). The prevalence rates of HPT in some sub-Saharan African countries currently rank among the highest on the globe, in stark

contrast to the picture from a few decades ago when the region had the lowest blood pressure levels (Danaei et al., 2011). Inroads into understanding the pathophysiology of HPT and advancing treatment options have been made over the years. The pathogenesis has been linked to various biological processes, including endothelial dysfunction, impaired angiogenesis, dysregulation of the renin-angiotensin-aldosterone axis and platelet activation (Taddei et al., 2001; Gkaliagkousi et al., 2010; Touyz et al., 2018). Sizeable financial investments have been made to study the genetic and environmental determinants of HPT and some of these studies have linked the Liddle phenotype (a hereditary disorder characterized by overactivity of the renal tubular epithelial sodium channel as a result of mutations in the *SCNN1B* and *SCNN1G* genes), and salt and water retention to the high HPT prevalence levels currently observed in sub-Saharan Africa (Spence and Rayner, 2018). However, in up to 95% of HPT cases, the etiology remains unknown and primary HPT continues to be a leading cause of morbidity and premature mortality globally (Carretero and Oparil, 2000; He and Macgregor, 2007). Described as a "silent killer," its symptoms manifest in the later stages of the condition when hypertension-mediated target organ damage has possibly taken place (Moore, 2005).

Processes involved in blood pressure homeostasis are tightly regulated by various systems in the body. Amongst others, cellular processes like differentiation, growth and metabolism, are known to be under the control of microRNAs (miRNAs) (Vidigal and Ventura, 2015). These miRNAs are 18-25 nucleotide long, non-coding, post-transcriptional regulators of gene expression and their dysregulation has been linked to the development of cancer, essential HPT, viral disease and endothelial dysfunction (Chen et al., 2008; Li and Kowdley, 2012; Nemecz et al., 2016; Silambarasan et al., 2016). It is plausible that dysregulation of miRNAs may lead to disturbances in the body's blood pressure regulatory mechanisms and play an important role in the development of HPT. Herein, we investigated circulating levels of miR-1299, miR-30a-5p, miR-30e-3p, miR-126-3p, and miR-182-5p in participants with normal blood pressure, as well as screen-detected and known hypertensives on anti-hypertensive treatment. These miRNAs were chosen as possible targets in our study as next generation sequencing data had previously shown their association with diabetes and HPT (Matsha et al., 2018; Matshazi et al., 2021).

MATERIALS AND METHODS

Design of the Study and Description of Procedures

This study included males and females from the Vascular and Metabolic Health (VMH) study, a sub-study of the Cape Town Bellville South Study (Matsha et al., 2012). The strategies followed with regards to data collection and the conduction of various procedures in this study were previously described by Matsha et al. (2018). In summary, various anthropometric measurements were taken from each participant and for each variable, the average of three separate readings taken was reported. Body Mass Index was calculated as weight per square meter (kg/m^2) where kg is a participant's weight in kilograms and m^2 is the square of their height in meters. World Health Organization guidelines (Chalmers et al., 1999) informed the process of blood pressure measurement. In brief, blood pressure was measured from the right arm of a participant who was in a sitting position and had rested for at least 10-min, using a semi-automatic digital blood pressure monitor (Omron M6 comfort-preformed cuff blood pressure monitor, China). The blood pressure was measured thrice, at three-min intervals, and the lowest systolic blood pressure and corresponding diastolic blood pressure values reported. Participants were then put into their respective blood pressure categories based on; the use of anti-hypertensive medication as known HPT, blood pressure measurement of 140/90 mm Hg and above as screen-detected HPT and normal blood pressure measurement (less than 140/90 mm Hg) as normotensive.

A number of molecular methods were utilized to measure several biochemical parameters in blood samples collected from study participants. These measurements were conducted in an ISO 15189 accredited pathology practice (PathCare Reference Laboratory, Cape Town, South Africa). The molecular methods used included High Performance Liquid Chromatography (BioRad Variant Turbo, BioRad, Hercules, CA, United States), for measurement of glycated hemoglobin, whilst a paramagnetic particle chemiluminescence assay (Beckman DXI, Beckman Coulter, South Africa) was used to measure serum insulin. The Competitive Chemiluminescent (Immulite 2000, Siemens, Munich, Germany) was used to determine levels of cotinine in serum whilst plasma glucose concentration was measured by the enzymatic hexokinase method (Beckman AU, Beckman Coulter, Brea, CA, United States). An enzymatic immunoinhibition—end point assay (Beckman AU, Beckman Coulter, Brea, CA, United States) was used to determine both total cholesterol and high-density lipoprotein cholesterol, whilst triglycerides were measured using a glycerol phosphate oxidase-peroxidase, end point assay (Beckman AU, Beckman Coulter, Brea, CA, United States). The enzymatic selective protection—end point (Beckman AU, Beckman Coulter, Brea, CA, United States) assay was used to determine the levels of low-density lipoprotein cholesterol. In order to determine ultrasensitive C-reactive protein (CRP) levels, a Latex Particle Immunoturbidimetry (Beckman AU, Beckman Coulter, Brea, CA, United States) assay was used. The analysis of miRNA expression through total RNA extraction and quantitative reverse transcription polymerase chain reaction (RT-qPCR) assays was conducted in blood samples that had previously been collected in Tempus Blood RNA tubes (ThermoFisher Scientific, Waltham, MA, United States) and stored at −80°C.

Total RNA Isolation

The MagMax Total RNA isolation kit (ThermoFisher Scientific, Waltham, MA, United States) was used to extract total RNA (including miRNAs) as per the recommendations of the manufacturer. Total RNA extraction was conducted on thawed, 3 mL whole blood samples. As this was a semi-automated procedure, the Kingfisher Flex system (ThermoFisher Scientific,

TABLE 1 | Study participant characteristics based on blood pressure status.

	Normotensive (*n* = 573) mean ± SD	Screen-detected HPT (*n* = 304) mean ± SD	Known HPT (*n* = 579) mean ± SD	All *p*-value	Normotensive vs Screen-detected HPT *p*-value	Normotensive vs Known HPT *p*-value	Screen-detected HPT vs Known HPT *p*-value
Gender				< 0.001	0.676	< 0.001	<0.001
Female, *n* (%)	386 (67.4)	209 (68.8)	478 (82.6)				
Male, *n* (%)	187 (32.6)	95 (31.2)	101 (17.4)				
Age (years)	39.3 ± 13.7	48.9 ± 13.4	58.9 ± 11.0	< 0.001	<0.001	< 0.001	<0.001
Body mass index (kg/m^2)	25.9 ± 7.0	28.4 ± 8.5	31.4 ± 7.7	< 0.001	<0.001	< 0.001	<0.001
Waist circumference (cm)	84.5 ± 15.3	91.2 ± 16.3	98.5 ± 16.4	< 0.001	<0.001	< 0.001	<0.001
Hip circumference (cm)	98.2 ± 15.0	102.1 ± 16.4	108.5 ± 16.2	< 0.001	<0.001	< 0.001	<0.001
Waist to Hip ratio	0.9 ± 0.1	0.9 ± 0.1	0.9 ± 0.1	< 0.001	<0.001	< 0.001	0.006
Systolic blood pressure (mmHg)	115.5 ± 12.7	151.7 ± 19.5	148.2 ± 26.4	< 0.001	<0.001	< 0.001	0.041
Diastolic blood pressure (mmHg)	74.7 ± 9.1	95.8 ± 11.6	90.7 ± 15.4	< 0.001	<0.001	< 0.001	<0.001
Fasting Blood glucose(mmol/L)*	4.6 (4.3; 5.0)	4.9 (4.6; 5.5)	5.3 (4.9; 6.8)	< 0.001	<0.001	< 0.001	<0.001
2-h glucose (mmol/L)*	5.3 (4.3; 6.6)	6.1 (5.0; 7.7)	6.9 (5.5; 8.6)	< 0.001	<0.001	< 0.001	<0.001
HbA1c (%)	5.7 ± 1.0	5.9 ± 1.3	6.8 ± 1.9	0.032	0.006	< 0.001	<0.001
HbA1c (mmol/mol)	38.6 ± 11.0	40.9 ± 13.8	50.7 ± 21.1	0.032	0.006	< 0.001	<0.001
Fasting Insulin (mIU/L)*	5.6 (3.4; 8.8)	6.1 (3.9; 9.6)	8.0 (5.0; 13.1)	< 0.001	0.065	< 0.001	<0.001
2-h Insulin (mIU/L)*	28.8 (15.0; 53.4)	34.7 (19.1; 64.0)	48.7 (26.5; 88.2)	< 0.001	0.014	< 0.001	<0.001
Triglycerides (mmol/L)*	1.0 (0.7; 1.4)	1.2 (0.9; 1.7)	1.4 (1.0; 1.9)	< 0.001	<0.001	< 0.001	<0.001
Total Cholesterol (mmol/L)	4.8 ± 1.1	5.2 ± 1.1	5.4 ± 1.2	< 0.001	<0.001	< 0.001	<0.001
HDL-cholesterol (mmol/L)	1.3 ± 0.4	1.4 ± 0.5	1.3 ± 0.3	< 0.001	<0.001	0.359	< 0.001
LDL-cholesterol (mmol/L)	3.0 ± 1.0	3.2 ± 1.0	3.3 ± 1.0	< 0.001	0.025	< 0.001	0.021
C-Reactive Protein (mg/L)*	2.7 (1.1; 6.9)	3.7 (1.5; 9.3)	5.2 (2.5; 10.1)	< 0.001	0.012	< 0.001	<0.001
Gamma GT (IU/L)*	25 (18; 39)	30 (21; 50)	32 (22; 53)	< 0.001	<0.001	< 0.001	<0.001
Serum Creatinine (μmol/L)	59.5 ± 12.0	62.7 ± 28.9	71.8 ± 61.0	< 0.001	0.021	< 0.001	0.014

Values presented as mean ± SD unless marked with an asterisk, in which case the median and (25th–75th percentiles) are reported.*
The Kruskal–Wallis test and analysis of variance (ANOVA) were used to compare the median and mean baseline characteristics, respectively, across blood pressure groups. All p-value shows a comparison of the three blood pressure groups together (i.e., normotensives vs screen-detected HPT vs known HPT).
HPT, hypertension; SD, standard deviation; HbA1c, glycated hemoglobin.

Waltham, MA, United States) conducted the RNA washing and elution steps. Quality checks were conducted on the extracted nucleic acid by determining the RNA concentration and purity on the NanoDrop One spectrophotometer (ThermoFisher Scientific, Waltham, MA, United States). An extracted total RNA sample with a 260/280 value between 1.8 and 2.0, and whose concentration was greater than 20 ng/μl passed the quality check and was therefore used for downstream applications like RT-qPCR.

Quantitative Reverse Transcription PCR

In order for miRNA expression to be determined using RT-qPCR, the extracted RNA had to be converted to cDNA first and this conversion was done using the TaqMan Advanced miRNA cDNA synthesis kit (Applied Biosystems, ThermoFisher Scientific, Waltham, MA, United States) as per the manufacturer's recommendations. In summary, by sequentially conducting poly-A tailing, adaptor ligation, reverse transcription and miR-Amp steps, we converted 2 μl of total RNA into cDNA, which was the starting material in the succeeding PCR step used to determine miRNA expression levels. This PCR analysis was conducted on the QuantStudio 7 Flex real-time PCR instrument (Life Technologies, Carlsbad, CA, United States) using TaqMan miRNA Assay primers (Applied Biosystems, ThermoFisher Scientific, Waltham, MA, United States). The determination of relative miRNA expression in a sample was done using the $2^{-\Delta Ct}$, whilst fold change differences in miRNA expression between the study groups were computed with the use of the $2^{-\Delta\Delta Ct}$ method. In order to normalize miRNA quantification, miR-16-5p (ThermoFisher Scientific, Waltham, MA, United States) was used as the endogenous control (Livak and Schmittgen, 2001).

Statistical Analysis

The Statistical Product and Service Solutions (SPSS) v.26 software (IBM Corp, United States) was used to conduct data analyses. For normally distributed variables, results are reported as count (and percentages), mean (and standard deviation) whilst for asymmetrically distributed variables, the results are reported as median (25th-75th percentiles). In order to compare the mean and median baseline characteristics across blood pressure groups, the analysis of variance (ANOVA) and Kruskal–Wallis tests were used, respectively. Age, gender and BMI-adjusted Spearman's

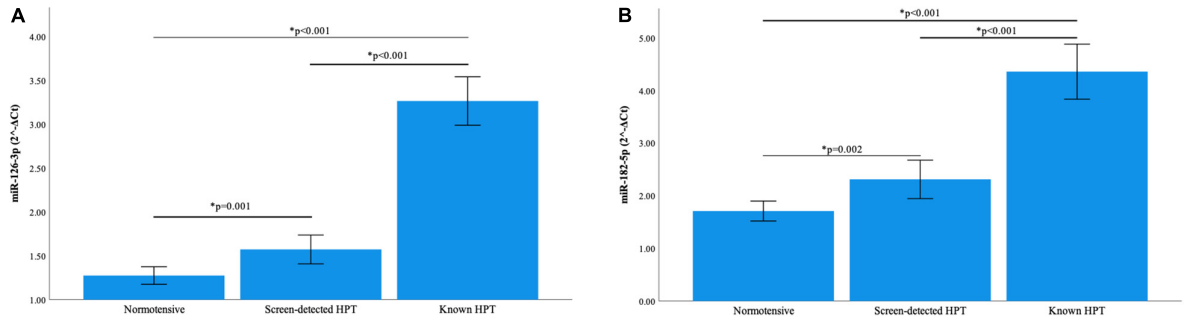


FIGURE 1 | Comparison of the relative expression of miR-126-3p and 182-5p in the three blood pressure groups. **(A)** The expression of miR-126-3p was significantly higher in known hypertensives when compared to screen-detected hypertensives ($p < 0.001$) and the normotensives ($p < 0.001$). There was also a significant difference in expression when screen-detected hypertensives were compared to the normotensives ($p = 0.001$). **(B)** There was a significantly higher expression of miR-182-5p in known hypertensives when compared to screen-detected hypertensives ($p < 0.001$) and the normotensives ($p < 0.001$). When screen-detected hypertensives were compared to the normotensives, that significant difference in expression remained ($p = 0.002$). The relative miRNA expressions were calculated using the $2^{-\Delta Ct}$ method and a one-way ANOVA used to compare differences in miRNA expression between groups. The symbol * denotes a statistically significant *p*-value.

partial correlations were used to assess the relationship between miRNAs and other cardiovascular risk profile variables whilst multivariable logistic regression models were used to assess the association of miRNAs with screen-detected and known HPT. A *p*-value less than 0.05 signified statistically significant findings.

RESULTS

Study Participant Characteristics

A total of 1,456 participants were included in this study. A summary of the participants' characteristics is shown in **Table 1**. There were 386 (67.4%) females and 187 (32.6%) males in the normotensive group with an average age of 39.3 ± 13.7 years, 209 (68.8%) females and 95 (31.2%) males in the screen-detected HPT group with an average age of 48.9 ± 13.4 years whilst the known HPT group comprised of 478 (82.6%) females and 101 (17.4%) males whose average age was 58.9 ± 11.0 years. As expected for HPT, there was a significant difference in the average age of participants across the three groups ($p < 0.001$). Serum gamma-Glutamyltransferase (Gamma GT-S) levels differed significantly between the three groups and the other expected differences (body mass index (BMI), age, systolic blood pressure, waist circumference, high density lipoprotein cholesterol (HDL-c) and low-density lipoprotein (LDL) cholesterol by HPT status in the cardiovascular risk profile were apparent between the participant groups as shown in **Table 1**.

Relative MicroRNA Expression

Normalization of relative miRNA expression was done with reference to the expression of miR-16-5p. Overall, the expression of miR-126-3p and 182-5p differed significantly across all blood pressure groups. **Figure 1A** shows that in participants with known HPT, the relative expression ($2^{-\Delta Ct}$) of miR-126-3p (3.266) was significantly higher than that of the normotensives (1.275) and screen-detected hypertensives (1.573), $p < 0.001$). For miR-182-5p, the relative expression in known HPT was significantly higher (4.360) compared to both normotensives (1.711) and screen detected hypertensives (2.312), both $p < 0.001$ as shown in **Figure 1B**. However, for miR-30a-5p, 1299 and 30e-3p, whilst there were significant differences in their relative expression ($2^{-\Delta Ct}$) in known HPT (0.068; 0.011 and 0.014) vs normotensives (0.033; 0.004 and 0.006), respectively, (all $p < 0.001$), that significance was not seen when screen-detected hypertensives were compared to the normotensives, all $p \geq 0.114$ as shown in **Figures 2A–C**.

Fold Change Computation

When compared to the normotensives, there was at least a 2.2-fold increase in expression of miR-126-3p, 182-5p, 30a-5p, 1299, and 30e-3p in the known HPT group. MiR-126-3p was the most expressed of the five miRNAs, particularly in known hypertensives versus normotensives (fold change = 2.64) and known hypertensives versus screen-detected hypertensives (fold change = 1.99). However, the fold difference in expression of all five miRNAs in the screen-detected hypertensive participants compared to the normotensives was ≤ 1.65-fold, with the lowest fold difference in expression between these two blood pressure groups observed for miR-30e-3p, whose expression was 1.17-fold higher in screen-detected hypertensives compared to the normotensives.

Correlation of MicroRNA Expression With Anthropometric Measurements

There was a significant positive correlation between the expression of the five miRNAs across the blood pressure groups ($r \geq 0.71$, $p < 0.001$), with the highest correlation in expression seen between miR-126-3p and miR-182-5p ($r = 0.983$, $p < 0.001$) as shown in **Table 2**. Detailed correlations between the expression of each miRNA and biochemical parameters are shown in the **Supplementary Tables** (see **Supplementary Tables 1–5**). The expression of all miRNAs correlated negatively with waist circumference, with the highest correlation coefficients

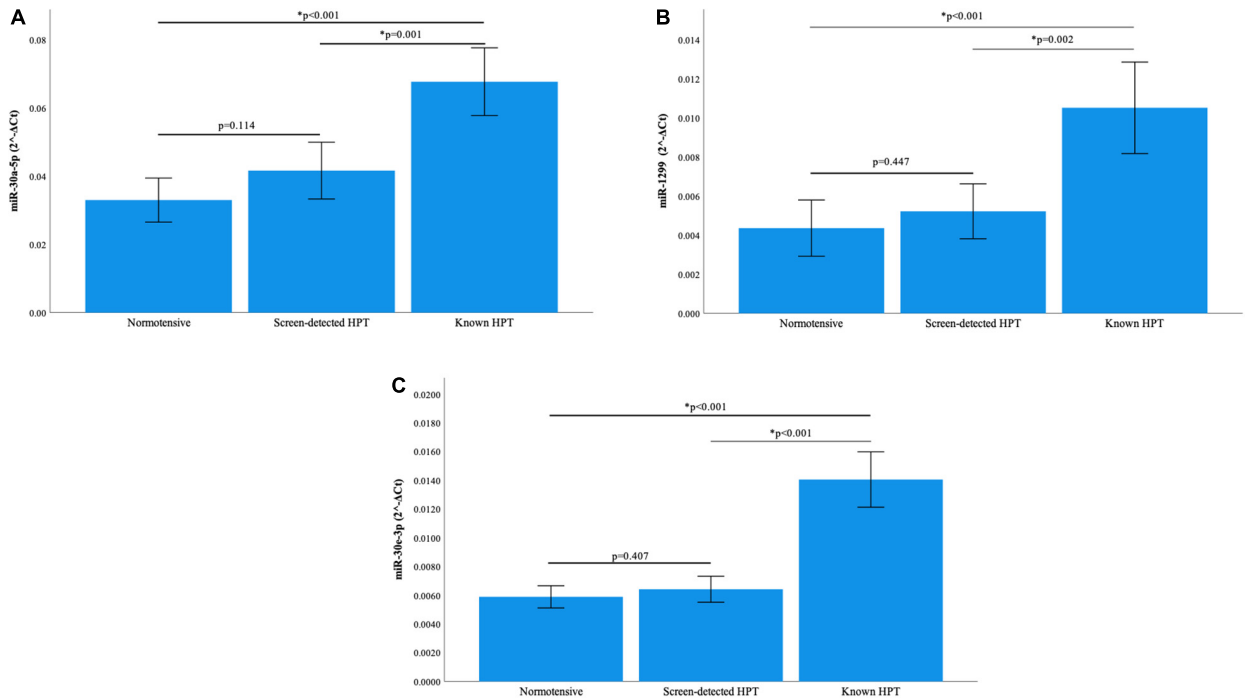


FIGURE 2 | Comparison of the relative expression of miR-30a-5p, miR-1299 and miR-30e-3p. **(A)** The expression of the miR-30a-5p was significantly higher in known hypertension compared to screen-detected hypertension ($p = 0.001$) and the normotensives ($p < 0.001$). However, there was no significant difference in expression between screen-detected hypertensives and normotensives ($p = 0.114$). **(B)** There was a significantly higher expression of miR-1299 in known hypertension when compared to screen-detected hypertension ($p = 0.002$) and the normotensives ($p < 0.001$). There was no significant difference in expression when screen-detected hypertensives were compared to the normotensives ($p = 0.447$). **(C)** The expression of miR-30e-3p was significantly higher in known hypertension when compared to screen-detected hypertension ($p < 0.001$) and the normotensives ($p < 0.001$). There was no significant difference in expression when screen-detected hypertensives were compared to the normotensives ($p = 0.407$). The relative miRNA expressions were calculated using the $2^{-\Delta Ct}$ method and a one-way ANOVA used to compare differences in miRNA expression between groups. The symbol * denotes a statistically significant *p*-value.

seen in the screen-detected HPT group for both miR-126-3p ($r = -0.748$, $p < 0.001$) and miR-30e-3p ($r = -0.729$, $p < 0.001$). There was a significant, though weak correlation between the expression of both miR-182-5p and miR-30a-5p with systolic blood pressure in the known HPT group. However, there was no significant association between the expression of miR-30e-3p, miR-126-3p and 1299 with systolic blood pressure regardless of blood pressure status. Whilst there was no correlation between the expression of any of the five miRNAs and total cholesterol (TC), there was a significant positive correlation with HDL-c, with the highest correlation coefficients seen in the known HPT group with respect to miR-182-5p ($r = 0.629$, $p = 0.001$), miR-30a-5p ($r = 0.615$, $p = 0.002$), miR-30e-3p ($r = 0.608$, $p = 0.002$) miR-126-3p ($r = 0.595$, $p = 0.003$) and finally miR-1299 ($r = 0.508$, $p = 0.013$). Gamma GT-S also showed a significantly positive correlation with the expression of all but miR-1299 in the screen-detected HPT group.

Multivariable Regression Analysis

The results of multivariable regression analysis are shown in **Table 3**. With regards to miR-126-3p, the crude odds (age and gender adjusted only) ratio (OR) was 1.16 (95% confidence interval (CI): 1.05-1.27, '$p = 0.003$) for screen-detected HPT, whilst for known HPT, the OR was 1.58 (95% CI: 1.46-1.71, $p < 0.001$). For miR-30a-5p, the OR was 1.24 (95% CI: 1.00-1.55, $p = 0.053$) for screen-detected HPT whilst for known HPT, the OR was 1.63 (95% CI: 1.36-1.95, $p < 0.001$). The crude odds ratio for miR-182-5p was 1.11 (95% CI: 1.05-1.17, $p < 0.001$) for screen-detected HPT whilst for known HPT, the OR was 1.24 (95% CI: 1.18-1.30, $p < 0.001$). The associations between the expression of these miR-126-3p, 182-5p and 30a-5p and HPT (both screen-detected and known) remained significant even when the model was adjusted for BMI, TC, triglycerides (TG) and glycated hemoglobin (HbA1c). When further adjusted for duration of known HPT diagnosis, the associations also remained significant for these three miRNAs. As for miR-30e-3p, the only significant association was with known HPT which had an OR of 1.71 (95% CI: 1.46–2.01, $p < 0.001$) whilst for screen-detected HPT, the OR was 1.07 (95% CI: 0.91–1.26, $p = 0.393$). There was no significant association between the expression of miR-1299 and screen-detected or known HPT.

DISCUSSION

This study demonstrated a significantly higher expression of miR-126-3p and miR-182-5p in hypertensives (both screen-detected and known) when compared to the normotensives. However, there was no significant difference in the expression

TABLE 2 | Partial correlations between miRNA relative expression and anthropometric and biochemical parameters.

	miR-30a-5p $2^{-\Delta Ct}$		miR-126-3p $2^{-\Delta Ct}$		miR-182-5p $2^{-\Delta Ct}$		miR-30e-3p $2^{-\Delta Ct}$		miR-1299 $2^{-\Delta Ct}$	
	r	p-value	r	p-value	r	p-value	r	p-value	r	p-value
miR-30a-5p $2^{-\Delta Ct}$	1.000		0.901	< 0.001	0.937	< 0.001	0.901	< 0.001	0.710	< 0.001
miR-1299 $2^{-\Delta Ct}$	0.710	< 0.001	0.731	< 0.001	0.738	< 0.001	0.721	< 0.001	1.000	
miR-182-5p $2^{-\Delta Ct}$	0.937	< 0.001	0.983	< 0.001	1.000		0.968	< 0.001	0.738	< 0.001
miR-30e-3p $2^{-\Delta Ct}$	0.901	< 0.001	0.973	< 0.001	0.968	< 0.001	1.000		0.721	< 0.001
miR-126-3p $2^{-\Delta Ct}$	0.901	< 0.001	1.000		0.983	< 0.001	0.973	< 0.001	0.731	< 0.001
Waist circumference (cm)	−0.485	0.019	−0.535	0.009	−0.518	0.011	−0.538	0.008	−0.521	0.011
Hip circumference (cm)	−0.154	0.482	−0.053	0.809	−0.100	0.651	−0.084	0.703	0.048	0.828
Waist to Hip ratio	0.001	0.997	−0.113	0.608	−0.060	0.787	−0.083	0.705	−0.131	0.553
Systolic Blood Pressure (mmHg)	0.313	0.146	0.293	0.174	0.304	0.158	0.275	0.204	0.268	0.216
Diastolic Blood Pressure (mmHg)	0.224	0.304	0.217	0.319	0.228	0.295	0.202	0.354	0.233	0.284
Fasting Blood Glucose (mmol/L)	0.133	0.546	0.077	0.728	0.109	0.621	0.119	0.590	0.040	0.856
2-h glucose (mmol/L)	0.046	0.834	0.019	0.932	0.044	0.841	0.058	0.792	0.102	0.644
HbA1c (%)	−0.201	0.357	−0.170	0.439	−0.190	0.385	−0.153	0.485	−0.130	0.555
Fasting insulin (mIU/L)	0.272	0.209	0.241	0.267	0.267	0.218	0.250	0.250	0.216	0.322
2-h insulin (mIU/L)	0.049	0.826	0.024	0.913	0.051	0.815	0.045	0.838	0.078	0.723
Triglycerides-S (mmol/L)	0.037	0.868	0.018	0.935	0.008	0.971	−0.024	0.915	−0.079	0.719
Total cholesterol (mmol/L)	0.108	0.624	0.019	0.930	0.039	0.858	0.015	0.945	0.041	0.854
HDL-cholesterol (mmol/L)	0.445	0.033	0.401	0.058	0.431	0.040	0.443	0.034	0.421	0.045
LDL-cholesterol (mmol/L)	0.065	0.769	−0.017	0.939	−0.003	0.991	−0.025	0.908	0.021	0.925
C-Reactive Protein (mg/L)	0.145	0.509	0.139	0.528	0.144	0.511	0.173	0.429	0.152	0.487
Gamma GT (IU/L)	0.094	0.669	0.047	0.830	0.078	0.722	0.063	0.775	0.043	0.845
Serum creatinine (μmol/L)	−0.349	0.443	−0.390	0.387	−0.377	0.404	−0.385	0.394	−0.337	0.460

Age, gender, and BMI adjusted Spearman's partial correlations were used to assess the association between miRNA relative expression and anthropometric and biochemical parameters across the three blood pressure groups.

levels of miR-30a-5p, 30e-3p, and 1299 in normotensives, relative to screen-detected hypertensives. Although multivariable logistic regressions showed no association between HPT and the expression of both miR-30e-3p and miR-1299, we observed an association between the expression of miR-126-3p, 182-5p and 30a-5p with screen-detected and known HPT, particularly in the latter. Even after adjustment of the crude model for age, sex, BMI, HbA1c, TG, and TC, the associations remained significant.

Angiogenesis and maintenance of vascular integrity are vital blood pressure regulatory processes in which miR-126 has an essential role (Wang et al., 2008). During hypertensive states, there is loss of endothelial cell function and blood perfusion to the capillaries becomes limited, leading to capillary disappearance. This is described as microvascular rarefaction and is a distinctive characteristic of HPT (Humar et al., 2009). Its effects contribute to HPT-related complications such as organ damage and stroke. It has been reported that miR-126, through the stimulation of proangiogenic activity of vascular endothelial growth factor (VEGF) and fibroblast growth factor (FGF), counteracts microvascular rarefaction by encouraging the formation of blood vessels. This is achieved through the quelling of Spred-1 expression, a known inhibitor of angiogenic signaling (Wang et al., 2008). Our study demonstrated a higher expression of miR-126-3p in hypertensive participants when compared to the normotensives, and this finding has also been reported in another study (Liu et al., 2018). It is plausible that the high expression of miR-126 in hypertensives represents a response to lower the blood pressure through promotion of blood vessel formation, whilst repressing the effects of anti-angiogenic Spred-1. Whilst our study demonstrated a higher expression of miR-126-3p in the whole blood of hypertensives compared to the normotensives, another study demonstrated lower expression of miR-126 in hypertensives when compared to healthy controls, albeit in peripheral blood mononuclear cells (PBMCs) (Kontaraki et al., 2014). There was no significant difference in miR-126 expression between hypertensive and normotensive participants in the study by Chen and co-workers (Chen et al., 2018). These discrepancies could in part be explained by the different sample types and participant recruitment criteria used. It has also been demonstrated that anti-hypertensive medications like nebivolol and atenolol affect miRNA expression (Ye et al., 2013). Kontaraki and colleagues excluded anyone on anti-hypertensive medication from their study, while Chen and colleagues' hypertensive group included any participants who had been on anti-hypertensive therapy for longer than 3 months (Chen et al., 2018).

The efficacy of miR-182 against glioblastoma multiforme, a therapy-resistant cancer of the brain has been previously reported (Kouri et al., 2015). It has been established as an oncogenic miRNA and its interactions in several types of cancers have been reviewed (Wei et al., 2015). The expression of miR-182-5p was significantly higher in both hypertensive groups when compared to the normotensives in our study. However, there is a paucity of studies indicating a role for miR-182-5p in essential HPT. Nonetheless, in a study evaluating the miRNA

TABLE 3 | Multivariable regression analysis of miRNAs for the presence of screen-detected and known hypertension.

	Screen-detected HPT			Known HPT		
	OR	**95% CI**	***p-value***	**OR**	**95% CI**	***p-value***
miR 30a-5p**						
Model 1	1.24	(1.00; 1.55)	0.053	1.63	(1.36;1.95)	< 0.001
Model 2	1.35	(1.07; 1.72)	0.013	1.98	(1.58;2.48)	< 0.001
Model 3	1.35	(1.06; 1.71)	0.015	1.97	(1.57;2.48)	< 0.001
Model 4	1.37	(1.07; 1.74)	0.011	1.99	(1.57;2.51)	< 0.001
Model 5	1.35	(1.06; 1.72)	0.014	1.97	(1.56;2.48)	< 0.001
Model 6	–	–	–	1.81	(1.4;2.34)	< 0.001
miR 30e-3p***						
Model 1	1.07	(0.91; 1.26)	0.393	1.58	(1.4;1.78)	< 0.001
Model 2	1.14	(0.96; 1.36)	0.130	1.71	(1.46;1.99)	< 0.001
Model 3	1.15	(0.96; 1.37)	0.121	1.71	(1.46;2.01)	< 0.001
Model 4	1.16	(0.97; 1.39)	0.094	1.74	(1.48;2.05)	< 0.001
Model 5	1.17	(0.98; 1.39)	0.089	1.74	(1.48;2.05)	< 0.001
Model 6	–	–	–	1.81	(1.42;2.30)	< 0.001
miR 126-3p*						
Model 1	1.16	(1.05; 1.27)	0.003	1.58	(1.46;1.71)	< 0.001
Model 2	1.23	(1.11; 1.37)	< 0.001	1.72	(1.55;1.9)	< 0.001
Model 3	1.21	(1.09; 1.35)	< 0.001	1.69	(1.52;1.87)	< 0.001
Model 4	1.23	(1.11; 1.38)	< 0.001	1.71	(1.54;1.91)	< 0.001
Model 5	1.23	(1.1; 1.37)	< 0.001	1.71	(1.53;1.91)	< 0.001
Model 6	–	–	–	1.93	(1.65;2.26)	< 0.001
miR 1299**						
Model 1	1.07	(0.94; 1.21)	0.329	1.22	(1.11;1.35)	< 0.001
Model 2	1.59	(0.42; 5.96)	0.495	1.18	(1.05;1.33)	0.005
Model 3	1.05	(0.92; 1.2)	0.466	1.19	(1.05;1.34)	0.006
Model 4	1.08	(0.94; 1.24)	0.308	1.21	(1.06;1.37)	0.005
Model 5	1.09	(0.94; 1.25)	0.254	1.22	(1.07;1.39)	0.004
Model 6	–	–	–	1.23	(1.05;1.44)	0.009
miR 182-5p*						
Model 1	1.11	(1.05; 1.17)	< 0.001	1.24	(1.18;1.3)	< 0.001
Model 2	1.14	(1.07; 1.21)	< 0.001	1.31	(1.24;1.39)	< 0.001
Model 3	1.14	(1.07; 1.21)	< 0.001	1.31	(1.23;1.39)	< 0.001
Model 4	1.14	(1.08; 1.22)	< 0.001	1.31	(1.23;1.4)	< 0.001
Model 5	1.14	(1.07; 1.22)	< 0.001	1.31	(1.23;1.39)	< 0.001
Model 6	–	–	–	1.31	(1.21;1.41)	< 0.001

*Model 1: Crude; Model 2: included age and sex; Model 3: included age, sex and BMI; Model 4: included age, sex, BMI, HbA1c; Model 5: included age, sex, BMI, HbA1c, triglycerides, total cholesterol; Model 6: included age, sex, BMI, HbA1c, triglycerides, total cholesterol; duration of disease *calculated for 0.1-unit increase; ** calculated for 0.01-unit increase; ***calculated for 0.001-unit increase.*

expression profiles in the placenta of pregnant participants, the expression of miR-182 was found to be elevated in participants with pre-eclampsia, a mid-term complication of pregnancy characterized by proteinuria and HPT, when compared to normal, pregnant controls (Pineles et al., 2007). Hypertension is a known risk factor for coronary artery disease (CAD) (Weber et al., 2016) and in participants with unprotected left main CAD, the plasma expression of miR-182-5p was significantly higher compared to the non-CAD control group and further analyses indicated its high diagnostic power for uncontrolled left main CAD (Zhu et al., 2019). Whether miR-182-5p plays a role in hemodynamic regulation or indeed the pathogenesis of HPT remains to be elucidated.

In our study, the expression of miR-30a-5p was significantly higher in known hypertensives compared to normotensives. The upregulation of miR-30a in hypertensive participants compared to normal controls was also reported in another study and identified as a possible biomarker target for differentiating white coat HPT from essential HPT and normotensives (Huang et al., 2016). Delta-like ligand 4 (dll4) is mainly expressed in the vascular endothelium and exerts its effects through Notch signaling by attachment to its receptor Notch1. The ligand is a key inhibitor of angiogenesis and the expression of miR-30a works against the expression of dll4 and in doing so, promotes angiogenesis (Jiang et al., 2013). The promotion of angiogenesis through the expression of miR-30a-5p can lead to lowering of the blood pressure and could have been the case in our cohort of known hypertensives.

A limitation of our study was that participants making up the three blood pressure groups were not age and gender-matched. Although these variables were adjusted for in the multivariable regression analysis, their effect may not have been completely eliminated. However, the overall large number of participants involved in the project provided sufficient power to the study as each blood pressure group was adequately represented. Whilst the cross-sectional nature of the study precludes inference about causal relationships between miRNA expression and development of HPT, it provided a basis for setting up longitudinal cohorts in which functional studies can be conducted to further clarify the roles that these non-coding ribonucleic acids (RNAs) play in the pathogenesis of HPT.

In conclusion, we demonstrated for the first time in an African population, the differential expression of miRNAs in the whole blood of participants on the basis of their blood pressure status. These miRNAs could form a panel of biomarker targets for HPT diagnostic purposes. Furthermore, the study validated prior findings on miR-126 and the miR-30 family and highlighted their possible involvement in the pathogenesis of HPT, warranting further investigation into these non-coding RNAs as they could offer potential prognostic and therapeutic avenues for cardiovascular diseases.

DATA AVAILABILITY STATEMENT

Some of the datasets used and/or analyzed during the current study are available from the National Centre for Biotechnology Information (NCBI) Sequence Read Archive on the following link: https://www.ncbi.nlm.nih.gov/sra/PRJNA680302. Other datasets presented are not readily available because of the terms of consent to which participants agreed, but are available from the principal investigator of the main study on reasonable request, which should be directed to TM, matshat@cput.ac.za.

ETHICS STATEMENT

This investigation was based on the Cape Town Vascular and Metabolic Health (VMH) study, which was approved

by the Research Ethics Committees of the Cape Peninsula University of Technology (CPUT) and Stellenbosch University (respectively, NHREC: REC — 230 408 — 014 and N14/01/003). Ethical approval was also obtained for this cross-sectional sub-study from the CPUT Health and Wellness Sciences Research Ethics Committee (CPUT/HW-REC 2019/H7). The study was conducted as per the provisions of the Declaration of Helsinki. All procedures were explained to the participants in their language of choice. Once the participants fully understood their participation, they signed informed consent forms to allow the collection of blood and anthropometric data.

AUTHOR CONTRIBUTIONS

TM, RE, and AK: conceptualization and funding acquisition. DM, SR, and CW: methodology. DM and SD: formal analysis. DM and CW: investigation. TM: resources. DM, SR, and SD: data curation. DM: writing—original draft preparation. TM, RE, AK, GD, and SR: writing—review and editing. DM, CW, and SR: validation. DM, SD, and TM: visualization. TM and GD: supervision. TM and SD: project administration. All authors have read and agreed to the published version of the manuscript.

FUNDING

This research was funded by the South African Medical Research Council (SAMRC), with funds from National Treasury under its Economic Competitiveness and Support Package (MRC-RFA-UFSP-01-2013/VMH Study) and the South African National Research Foundation (SANRF) (Grant no. 115450).

ACKNOWLEDGMENTS

We would like to thank the Bellville South Ward 009 community for partaking in the study as well as the Bellville South Community Health Forum for supporting the engagement with the Bellville South Community.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.710438/full#supplementary-material

REFERENCES

Bromfield, S., and Muntner, P. (2013). High Blood Pressure: the Leading Global Burden of Disease Risk Factor and the Need for Worldwide Prevention Programs. *Curr. Hypertens. Rep.* 15, 134–136. doi: 10.1007/s11906-013-0340-9

Campbell, N. R. C., Bovet, P., Schutte, A. E., Lemogoum, D., and Nkwescheu, A. S. (2016). High Blood Pressure in Sub-Saharan Africa: why Prevention, Detection, and Control are Urgent and Important. *J. Clin. Hypertens.* 17, 663–667. doi: 10.1111/jch.12599

Carretero, O. A., and Oparil, S. (2000). Essential Hypertension. Part I: definition and Etiology. *Circulation* 101, 329–335. doi: 10.1161/01.cir.101.3.329

Chalmers, J., MacMahon, S., Mancia, G., Whitworth, J., Beilin, L., Hansson, L., et al. (1999). 1999 World Health Organization-International Society of Hypertension Guidelines for the Management of Hypertension. *Clin. Exp.* 21, 1009–1060. doi: 10.3109/10641969909061028

Chen, S., Chen, R., Zhang, T., Lin, S., Chen, Z., Zhao, B., et al. (2018). Relationship of Cardiovascular Disease Risk Factors and Noncoding RNAs with Hypertension: a Case-Control Study. *BMC Cardiovasc. Disord.* 18:58. doi: 10.1186/s12872-018-0795-3

Chen, X., Ba, Y., Ma, L., Cai, X., Yin, Y., Wang, K., et al. (2008). Characterization of microRNAs in Serum: a Novel Class of Biomarkers for Diagnosis of Cancer and Other Diseases. *Cell Res.* 18, 997–1006.

Danaei, G., Finucane, M. M., Lin, J. K., Singh, G. M., Paciorek, C. J., Cowan, M. J., et al. (2011). National, Regional, and Global Trends in Systolic Blood Pressure Since 1980: systematic Analysis of Health Examination Surveys and Epidemiological Studies with 786 Country-years and 5.4 Million Participants. *Lancet* 377, 568–577. doi: 10.1016/s0140-6736(10)62036-3

Gkaliagkousi, E., Passacquale, G., Douma, S., Zamboulis, C., and Ferro, A. (2010). Platelet Activation in Essential Hypertension: implications for Antiplatelet Treatment. *Am. J. Hypertens.* 23, 229–236. doi: 10.1038/ajh.2009.247

He, F. J., and Macgregor, G. A. (2007). Blood Pressure is the Most Important Cause of Death and Disability in the World. *Eur. Heart J. Suppl.* 9, B23–B28.

Huang, Y. Q., Huang, C., Chen, J. Y., Li, J., and Feng, Y. Q. (2016). The Association of Circulating miR-30a, miR-29 and miR-133 with White-Coat Hypertension. *Biomark. Med.* 10, 1231–1239. doi: 10.2217/bmm-2016-0215

Humar, R., Zimmerli, L., and Battegay, E. (2009). Angiogenesis and Hypertension: an Update. *J. Hum. Hypertens.* 23, 773–782. doi: 10.1038/jhh.2009.63

Jiang, Q., Lagos-Quintana, M., Liu, D., Shi, Y., Helker, C., Herzog, W., et al. (2013). MiR-30a Regulates Endothelial Tip Cell Formation and Arteriolar Branching. *Hypertension* 62, 592–598. doi: 10.1161/hypertensionaha.113.01767

Kontaraki, J. E., Marketou, M. E., Zacharis, E. A., Parthenakis, F. I., and Vardas, P. E. (2014). MicroRNA-9 and microRNA-126 Expression Levels in Patients with Essential Hypertension: potential Markers of Target-Organ Damage. *J. Am. Soc. Hypertens.* 8, 368–375. doi: 10.1016/j.jash.2014.03.324

Kouri, F. M., Hurley, L. A., Daniel, W. L., Day, E. S., Hua, Y., Hao, L., et al. (2015). MiR-182 Integrates Apoptosis, Growth, and Differentiation Programs in Glioblastoma. *Genes Dev.* 29, 732–745. doi: 10.1101/gad.257394.114

Li, Y., and Kowdley, K. V. (2012). MicroRNAs in Common Human Diseases. *Genomics Proteomics Bioinformatics* 10, 246–253. doi: 10.1016/j.gpb.2012.07.005

Liu, J., Liu, J., Shi, L., Zhang, F., Yu, L., Yang, X., et al. (2018). Preliminary Study of microRNA-126 as a Novel Therapeutic Target for Primary Hypertension. *Int. J. Mol. Med.* 41, 1835–1844.

Livak, K. J., and Schmittgen, T. D. (2001). Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the $2^{-\Delta\Delta CT}$ Method. *Methods* 25, 402–408. doi: 10.1006/meth.2001.1262

Matsha, T. E., Hassan, M. S., Kidd, M., and Erasmus, R. T. (2012). The 30-Year Cardiovascular Risk Profile of South Africans with Diagnosed Diabetes, Undiagnosed Diabetes, Pre-diabetes or Normoglycaemia: the Bellville, South Africa Pilot Study. *Cardiovasc. J. Afr.* 23, 5–11. doi: 10.5830/cvja-2010-087

Matsha, T. E., Kengne, A. P., Hector, S., Mbu, D. L., Yako, Y. Y., and Erasmus, R. T. (2018). MicroRNA Profiling and Their Pathways in South African Individuals with Prediabetes and Newly Diagnosed Type 2 Diabetes Mellitus. *Oncotarget* 9, 30485–30498. doi: 10.18632/oncotarget.25271

Matshazi, D. M., Weale, C. J., Erasmus, R. T., Kengne, A. P., Davids, S. F. G., Raghubeer, S., et al. (2021). MicroRNA Profiles in Normotensive and Hypertensive South African Individuals. *Front. Cardiovasc. Med.* 8:645541. doi: 10.3389/fcvm.2021.645541

Mills, K. T., Bundy, J. D., Kelly, T. N., Reed, J. E., Kearney, P. M., Reynolds, K., et al. (2016). Global Disparities of Hypertension Prevalence and Control: a Systematic Analysis of Population-Based Studies from 90 Countries. *Circulation* 134, 441–450. doi: 10.1161/circulationaha.115.018912

Mills, K. T., Stefanescu, A., and He, J. (2020). The Global Epidemiology of Hypertension. *Nat. Rev. Nephrol.* 16, 223–237.

Moore, J. (2005). Hypertension: catching the Silent Killer. *Nurse Pract.* 30, 16–35. doi: 10.1097/00006205-200510000-00005

Nemecz, M., Alexandru, N., Tanko, G., and Georgescu, A. (2016). Role of MicroRNA in Endothelial Dysfunction and Hypertension. *Curr. Hypertens. Rep.* 18:87.

Pineles, B. L., Romero, R., Montenegro, D., Tarca, A. L., Han, Y. M., Kim, Y. M., et al. (2007). Distinct Subsets of microRNAs are Expressed Differentially in the Human Placentas of Patients with Preeclampsia. *Am. J. Obstet. Gynecol.* 196, 261.e1–6. doi: 10.1016/j.ajog.2007.01.008

Silambarasan, M., Tan, J. R., Karolina, D. S., Armugam, A., Kaur, C., and Jeyaseelan, K. (2016). MicroRNAs in Hyperglycemia Induced Endothelial Cell Dysfunction. *Int. J. Mol. Sci.* 17:518. doi: 10.3390/ijms17040518

Spence, J. D., and Rayner, B. L. (2018). Hypertension in Blacks. Individualized Therapy Based on Renin/Aldosterone Phenotyping. *Hypertension* 72, 263–269. doi: 10.1161/hypertensionaha.118.11064

Taddei, S., Virdis, A., Ghiadoni, L., Sudano, I., and Salvetti, A. (2001). Endothelial Dysfunction in Hypertension. *J. Cardiovasc. Pharmacol.* 38, S11–S14.

Touyz, R. M., Alves-Lopes, R., Rios, F. J., Camargo, L. L., Anagnostopoulou, A., Arner, A., et al. (2018). Vascular Smooth Muscle Contraction in Hypertension. *Cardiovasc. Res.* 114, 529–539.

Vidigal, J. A., and Ventura, A. (2015). The Biological Functions of miRNAs: lessons from in vivo Studies. *Trends Cell Biol.* 25, 137–147. doi: 10.1016/j.tcb.2014.11.004

Wang, S., Aurora, A. B., Johnson, B. A., Qi, X., McAnally, J., Hill, J. A., et al. (2008). The Endothelial-Specific MicroRNA miR-126 Governs Vascular Integrity and Angiogenesis. *Dev. Cell* 15, 261–271. doi: 10.1016/j.devcel.2008.07.002

Weber, T., Lang, I., Zweiker, R., Horn, S., Wenzel, R. R., Watschinger, B., et al. (2016). Hypertension and Coronary Artery Disease: epidemiology, Physiology, Effects of Treatment, and Recommendations: a Joint Scientific Statement from the Austrian Society of Cardiology and the Austrian Society of Hypertension. *Wien. Klin. Wochenschr.* 128, 467–479. doi: 10.1007/s00508-016-0998-5

Wei, Q., Lei, R., and Hu, G. (2015). Roles of miR-182 in Sensory Organ Development and Cancer. *Thorac. Cancer* 6, 2–9. doi: 10.1111/1759-7714.12164

Ye, H., Ling, S., Castillo, A. C., Thomas, B., Long, B., Qian, J., et al. (2013). Nebivolol Induces Distinct Changes in Profibrosis MicroRNA Expression Compared with Atenolol, in Salt-sensitive Hypertensive Rats. *Hypertension* 61, 1008–1013. doi: 10.1161/hypertensionaha.111.00892

Zhu, L., Chen, T., Ye, W., Wang, J.-Y., Zhou, J.-P., Li, Z.-Y., et al. (2019). Circulating miR-182-5p and miR-5187-5p as Biomarkers for the Diagnosis of Unprotected Left Main Coronary Artery Disease. *J. Thorac. Dis.* 11, 1799–1808. doi: 10.21037/jtd.2019.05.24

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Genetics

SYSTEMATIC REVIEW
published: 08 November 2021
doi: 10.3389/fgene.2021.682929

The Use of 'Omics for Diagnosing and Predicting Progression of Chronic Kidney Disease: A Scoping Review

Melanie A. Govender[1,2]*, *Jean-Tristan Brandenburg*[2], *June Fabian*[3] and *Michèle Ramsay*[1,2]*

[1]Division of Human Genetics, National Health Laboratory Service and School of Pathology, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa, [2]Sydney Brenner Institute for Molecular Bioscience, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa, [3]Wits Donald Gordon Medical Centre, School of Clinical Medicine, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa

Globally, chronic kidney disease (CKD) contributes substantial morbidity and mortality. Recently, various 'omics platforms have provided insight into the molecular basis of kidney dysfunction. This scoping review is a synthesis of the current literature on the use of different 'omics platforms to identify biomarkers that could be used to detect early-stage CKD, predict disease progression, and identify pathways leading to CKD. This review includes 123 articles published from January 2007 to May 2021, following a structured selection process. The most common type of 'omic platform was proteomics, appearing in 55 of the studies and two of these included a metabolomics component. Most studies (n = 91) reported on CKD associated with diabetes mellitus. Thirteen studies that provided information on the biomarkers associated with CKD and explored potential pathways involved in CKD are discussed. The biomarkers that are associated with risk or early detection of CKD are SNPs in the *MYH9/APOL1* and *UMOD* genes, the proteomic CKD273 biomarker panel and metabolite pantothenic acid. Pantothenic acid and the CKD273 biomarker panel were also involved in predicting CKD progression. Retinoic acid pathway genes, *UMOD*, and pantothenic acid provided insight into potential pathways leading to CKD. The biomarkers were mainly used to detect CKD and predict progression in high-income, European ancestry populations, highlighting the need for representative 'omics research in other populations with disparate socio-economic strata, including Africans, since disease etiologies may differ across ethnic groups. To assess the transferability of findings, it is essential to do research in diverse populations.

Keywords: 'omics, biomarkers, early detection, diagnosis, Sub-Saharan Africa, diabetes, chronic kidney disease, hypertension

OPEN ACCESS

Edited by:
Maritha J. Kotze,
Stellenbosch University, South Africa

Reviewed by:
Ronny Lesmana,
Universitas Padjadjaran, Indonesia
Jianxiang Shi,
Zhengzhou University, China

***Correspondence:**
Melanie A. Govender
melanieanngovender@gmail.com
Michèle Ramsay
Michele.Ramsay@wits.ac.za

Specialty section:
This article was submitted to Human and Medical Genomics, a section of the journal Frontiers in Genetics

Received: *19 March 2021*
Accepted: *18 October 2021*
Published: *08 November 2021*

Citation:
Govender MA, Brandenburg J-T, Fabian J and Ramsay M (2021) The Use of 'Omics for Diagnosing and Predicting Progression of Chronic Kidney Disease: A Scoping Review. Front. Genet. 12:682929. doi: 10.3389/fgene.2021.682929

INTRODUCTION

The global prevalence of chronic kidney disease (CKD) is estimated to be 13.4%, and thought to be higher in low- and middle-income settings, such as sub-Saharan Africa (SSA) (15.8%) (Hill et al., 2016; Kaze et al., 2018). End stage kidney disease (ESKD) is the most severe stage of CKD, at which point the damage is irreversible and support in the form of chronic dialysis or a kidney transplant is required to sustain life. Major risk factors for CKD vary by setting, with hypertension (HT) being the most common in east Asia, tropical Latin America, and Western and Southern SSA, whereas diabetes predominates as a risk factor in the remaining regions of the world (Bikbov et al., 2020). Endemic and other infectious

diseases, for example schistosomiasis, malaria, and human immunodeficiency virus (HIV) still contribute substantial additional risk for CKD in high prevalence countries (Couser et al., 2011; Ekrikpo et al., 2018). The diagnosis of CKD is currently based on estimated glomerular filtration rate (eGFR) and/or albuminuria (Levey et al., 2002). However, low GFR defined as eGFR<60 ml/min/1.73 m^2, is a relatively late marker of kidney disease, as organ damage may precede functional changes such as filtration impairment. Even though albuminuria is suggested to be an early marker of kidney disease, kidney disease may still progress despite the absence of albuminuria (Miller et al., 2009). Thus, there is a need for alternative markers of kidney disease that identify early disease and predict risk for progression. Such markers would enable interventions to prevent or slow progression of CKD, which is particularly relevant in resource-limited settings where access to kidney replacement therapy for ESKD remains severely restricted.

'Omics technologies refer to collective analyses of cell populations, tissues, organs or the whole organism at the molecular level (Horgan and Kenny, 2011). In the last decade, there have been significant advances in use of 'omics platforms to investigate susceptibility to, and detection of CKD, and related risk factors (Good et al., 2010; Abbiss et al., 2019; Cañadas-Garre et al., 2019). 'Omics technologies permit high-throughput, comprehensive exploration of the genome (DNA sequence), epigenome (epigenetic modifications), proteome (proteins), transcriptome (transcribed RNA) and metabolome (metabolites), through a wide range of platforms including next-generation sequencing, protein and mRNA arrays and mass spectrometry (Papadopoulos et al., 2016). The past decade has seen significant advances in genome-wide association studies (GWAS) for CKD and kidney function traits, with study sample sizes increasing from >100,000 to 1,000,000 facilitating the identification of >250 genetic loci robustly associated with CKD in different ethnic groups (Tin and Köttgen, 2020). Proteomic approaches to identify urinary biomarkers for early detection and prediction of progression of CKD are also on the rise, attractive (in part) for their non-invasive use in clinical settings (Mischak et al., 2010). For example, capillary electrophoresis coupled with mass spectrometry (CE-MS) has been used to develop a proteome-based urine biomarker panel of 273 peptides with profiles that differed significantly between individuals with CKD and healthy controls (Good et al., 2010). These peptides were combined to develop a single score known as the CKD273 risk score (Good et al., 2010). Previous studies defined a CKD273 risk score cut-off of 0.343 to predict diabetic nephropathy (DN) (Alkhalaf et al., 2010). This threshold was lowered to 0.154 to accommodate the early detection of CKD (Lindhardt et al., 2016).

The overall aim of this scoping review was to evaluate existing literature for potential biomarkers–aside from those currently used, namely eGFR and albuminuria–that might facilitate early detection of CKD, and predict its progression. The specific objectives were to 1) describe potential biomarkers identified through different 'omics platforms (genomics, proteomics, metabolomic, transcriptomics and epigenomics) that are associated with CKD, and 2) to explore potential pathways involved in CKD that could inform novel treatment strategies for CKD. This is pertinent for developing affordable diagnostic, prevention, and therapeutic strategies for countries with high prevalence of CKD and limited resources, such as are found in SSA.

METHODS AND MATERIALS

Protocol Registration

This review was conducted using the Preferred Reporting Items for Systematic reviews and Meta-Analyses extension for Scoping Reviews (PRISMA-ScR) approach (de Oliveira, 2018). The background and methods were documented in a protocol in the Open Science Framework (OSF) database (Doi:10.17605/OSF.IO/QM64P, https://osf.io/sw8ne/).

Search Strategy

Relevant studies published from January 2007 to 18 May 2021 (Search date: 18 May 2021) were identified through a comprehensive search of electronic databases including Embase, Web of Science, PubMed, SCOPUS and African Index Medicus. The electronic database search was conducted using medical subject heading (MeSH) terms, keywords pertinent to the topic, and Boolean operators such as AND/OR to obtain the maximum number of studies (**Table 1**).

Screening and Eligibility Criteria

Database search records were combined and all duplicate records removed. One investigator (MG) screened all records by title and abstract for inclusion, and articles with irrelevant topics were excluded. Investigators (J-TB and JF) independently screened records by title and abstract for verification. Any inconsistencies or disagreements were resolved by review with investigator MR. This was followed by screening full-text articles according to the eligibility criteria. The main inclusion criterion was that a study had to be an original study containing 'omics data used to detect CKD and related risk factors. The studies had to report on the characteristics of the study population, type of 'omics technique used, CKD or kidney function traits, relevant risk factors (defined in **Table 1**), and potential biomarker or indicator of early detection of CKD and prediction of progression (methylation profile, genes/genetic variants, transcripts, proteins and/or metabolites). Since the first large GWAS was conducted in 2007 (Craddock and Jones, 2007), we chose to include only studies published from January 2007 to May 2021. The age limit was restricted to participants 18 years and older and only studies published in English were included. Studies were excluded if the study was a review, meta-analysis, conference abstract or presentation, book chapter or response to treatment. Studies involving animals and paediatric participants were also excluded. The reasons for excluding each article were recorded.

Data Extraction

A data extraction sheet was created on Microsoft Excel (version 16.42, available at: https://office.microsoft.com/excel) to ensure all relevant data were extracted. Data extraction was performed by MG, and any discrepancies were resolved by J-TB, JF and MR. Data extracted for each article included the digital object identifier (DOI), name of first author, year of publication and title, participant/population characteristics (sample size,

TABLE 1 | Search strategy used in database searches.

Search no	Search query
#1	"chronic kidney disease" OR "CKD" OR "GFR" OR "glomerular filtration rate" OR "nephropathy" OR "serum creatinine" OR "urinary albumin" OR "ACR" OR "albumin creatinine ratio" OR "albuminuria" OR "urinary albumin creatinine ratio" OR "UACR" OR "proteinuria" OR "protein creatinine ratio" OR "urinary protein creatinine ratio" OR "UPCR"
#2	"diabetes" OR "hypertension" OR "infection" OR "infectious disease" OR "HIV" OR "human immunodeficiency virus" OR "malaria" OR "schistosomiasis" OR "cytomegalovirus" OR "CMV" OR "JC virus" OR "John Cunningham virus" OR "polyoma virus" OR "*tuberculosis*" OR "TB" OR "hepatitis B virus" OR "hepatitis C virus"
#3	"genomic" OR "metabolomic" OR "proteomic" OR "transcriptomic" OR "epigenomic"
#4	1# AND #2 AND #3

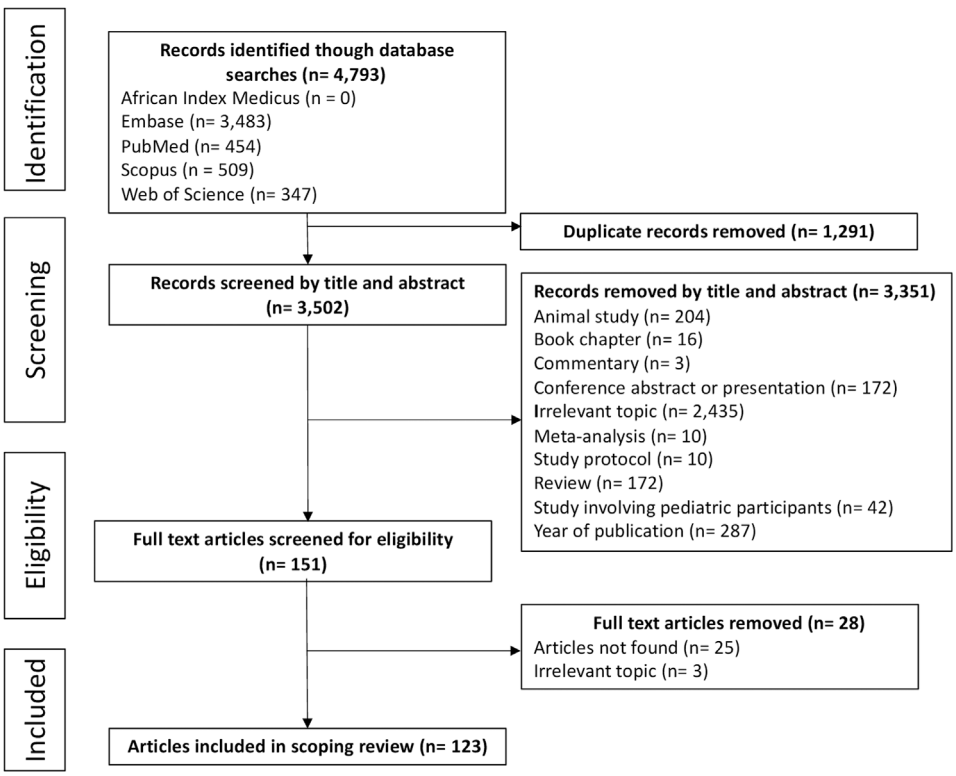


FIGURE 1 | Article selection process according to PRISMA-ScR guidelines.

ethnicity, stage of CKD, presence of risk factors and/or infections), type of 'omics technology, 'omics technique, biological sample and potential biomarkers (genes, variants, miRNAs, mRNAs, proteins, and peptides) associated with early detection of CKD and prediction of progression. As PRISMA-ScR guidelines recommend, we conducted a qualitative synthesis to summarise the main components of the CKD 'omics field.

RESULTS

The database searches identified 4,793 articles, of which 1,291 were duplicates. Through title and abstract screening, we excluded 3,351 articles mainly due to the following reasons: irrelevant topic, year of publication outside the period for this review, and animal studies. A further 28 articles were excluded after assessing the full-text and applying the eligibility criteria. Thus, one hundred and twenty-three full-text articles were included in this review (**Figure 1**, **Supplementary Appendix A**).

Participant and Study Characteristics

The participant and study characteristics of all studies (n = 123) included in this scoping review are shown in **Table 2**. There was an increase in publications from 2015. Majority of studies (28.5%) had sample sizes greater than 500. Most of the participants (42.3%) were between 51 and 60 years of age. Most of the studies (79.7%) included male and female participants. The

TABLE 2 | Participant and study characteristics reported in 123 studies.

Characteristic	Number of studies *n* (%)
Year of publication	
2007–2010	22 (17.9)
2011–2015	28 (22.8)
2016–May 2021	73 (59.3)
Number of participants per study	
1–10	2 (1.6)
11–50	22 (17.9)
51–100	20 (16.3)
101–200	26 (21.1)
201–500	18 (14.6)
>500	35 (28.5)
Mean age of participants	
18–30 years	0 (0)
31–40 years	11 (8.9)
41–50 years	11 (8.9)
51–60 years	52 (42.3)
61–70 years	32 (26.1)
>70 years	2 (1.6)
Unknown	15 (12.2)
Sex	
Males only	5 (4.0)
Females only	0 (0)
Both	98 (79.7)
Not stated	20 (16.3)
Participant ethnicity	
Asian only	15 (12.2)
African ancestry only	8 (6.5)
European ancestry only	19 (15.4)
Indian only	4 (3.3)
Studies including multiple ethnicities (European ancestry, African ancestry, Hispanic, Indian and Asian)	13 (10.6)
Unknown	64 (52.0)

most common ethnic group represented was European ancestry (15.4%), followed by Asian (12.2%), African ancestry (6.5%), Indian (3.3%), and studies reporting on multiple ethnicities (10.6%). It is important to note that ethnicity was not reported in over half of the studies (52.0%). Of eight studies that included participants of self-reported African ethnicity, only one included African individuals residing in Africa (Chen et al., 2007).

Omics' Platforms and Techniques

We reported on five different 'omics technologies including genomic, epigenomic, proteomic, metabolomic and transcriptomic (**Table 3**). The proteomic approach was the most common 'omics platform (n = 53), followed by metabolomics (n = 30), genomics (n = 17), epigenomics (n = 10), transcriptomics (n = 5) and eight studies used a combination of 'omics platforms. A wide range of techniques were used to analyse the 'omics data (**Table 3**). Genomic analysis was mainly performed by arrays including Illumina's High-throughput Bead array, Human Core Exome Bead array, HumanHap550-Duo Bead Chip and Omni Express Exome array. Proteomic analysis was mainly done by liquid chromatography coupled with mass spectrometry (LC-MS) and CE-MS. Gas chromatography-mass spectrometry (GC-MS) and LCMS were important analytical techniques for metabolomic analysis. Methylation arrays (including Illumina's Infinium 27, 450 and 850K methylation arrays) were key for epigenomic analysis and transcriptomic analysis was mainly performed by RNA sequencing.

CKD Risk Categories

The most common risk factors associated with CKD were diabetes mellitus (DM) (either type 1 diabetes, type 2 diabetes or both) (n = 91), followed by hypertension (n = 4), cardiovascular disease (n = 2), and infection (HIV) (n = 2). Sixteen studies reported on more than one risk factor and eight studies did not report on any risk factors (**Figure 2**).

Biomarkers Associated With Risk, Early Detection, and Progression of CKD

Whilst it was not feasible to discuss each biomarker identified for each study across all 'omics platforms, a list of selected studies that identified biomarkers/indicators involved in risk, early detection and prediction of progression, and those that explored potential disease pathways, are shown in **Table 4**.

Genetic Variants Associated With Risk for CKD

The genes implicated in risk for CKD were non–muscle myosin heavy chain 9 (*MYH9*), apolipoprotein L1 (*APOL1*) and

TABLE 3 | The different 'omics technologies and techniques reported in this scoping review for detection of CKD and associated risk in 123 studies.

Type of omic technology	Number of studies *n* (%)
Genomic	**17 (13.8)**
Affymetrix 370K and 500k arrays	
Illumina High-throughput Bead array	
Illumina HumanCoreExome Bead array	
Illumina HumanHap550-Duo BeadChip	
Illumina's IV panel	
llumina OmniExpressExome array	
Panel of 372 autosomal short tandem repeat markers	
Variety of GWAS arrays	
Epigenomic	**10 (8.1)**
Illumina Infinium 27K methylation array	
Illumina Infinium HumanMethylation450 Beadchip	
Illumina Infinium MethylationEPIC BeadChip (HM850K)	
Reverse phase high pressure liquid chromatography (RP-HPLC) analysis and bisulfite sequencing	
SuperTAG methylation-specific digital karyotyping	
Proteomic	**53 (43.1)**
1 and 2-dimensional gel electrophoresis (1 and 2DE) and mass spectrometry (MS)	
Capillary electrophoresis coupled with mass spectrometry (CE-MS)	
Fluorescence two-dimensional (2-D) differential in-gel electrophoresis (DIGE)	
Isobaric tag for relative and absolute quantitation iTRAQ MS	
Liquid chromatography coupled with mass spectrometry (LC-MS)	
Matrix-assisted laser desorption/ionization mass spectrometry (MALDI-MS)	
Olink cardiovascular panel	
Protein chip arrays	
Proximity Extension Assay (PEA)	
Surface-enhanced laser desorption/ionization mass spectrometry (SELDI-MS)	
Metabolomic	**30 (24.4)**
Commercial kits	
CE-MS	
Gas chromatography coupled with mass spectroscopy (GC-MS)	
LC-MS	
Nuclear magnetic resonance (NMR)	
Transcriptomic	**5 (4.1)**
GeneChip human genome series U133A and Plus 2.0 Array	
RNA sequencing	
Genomic and metabolomic	**2 (1.6)**
Infinium Multi-Ethnic Global BeadChip array	
Exome BeadChip array	
LC-MS	
Genomic and epigenomic	**1 (0.8)**
Illumina Human MethylationEPIC BeadChip	
Proteomic and metabolomic	**2 (1.6)**
ELISA Assays	
LC-MS	
SOMAscan	
Transcriptomic and epigenomic	**1 (0.8)**
Illumina 450K methylation array	
RNA sequencing	
Transcriptomic and metabolomic	**2 (1.6)**
Affymetrix transcriptomics arrays	
NMR	

uromodulin (*UMOD*). Kao et al. (2008) performed a genome-wide admixture scan in 2,178 individuals with European and African ancestry. They identified three SNPs (rs16996674, rs16996677, and rs5756152) in the *MYH9* gene that were associated with a ~3 times greater risk of nondiabetic ESKD in African Americans compared to European Americans. Similarly, Shlush et al. (2010) identified two SNPs (rs7286127 and rs5756133) associated with nondiabetic ESKD on chromosome 22 in the region that contains the *MYH9* gene, as well as the neighbouring *APOL1* gene. In addition, Igo et al. (2011) identified one SNP (rs5769116) 10 Mb distal to *MYH9/APOL1* that was associated with DN. Morris et al. (2019) performed an association study of eGFR, a measure of kidney function, in 312,468 individuals of diverse ancestry. They identified a highly significant SNP (rs77924615; $p = 1.5 \times 10^{-54}$) in the *UMOD* gene associated with eGFR.

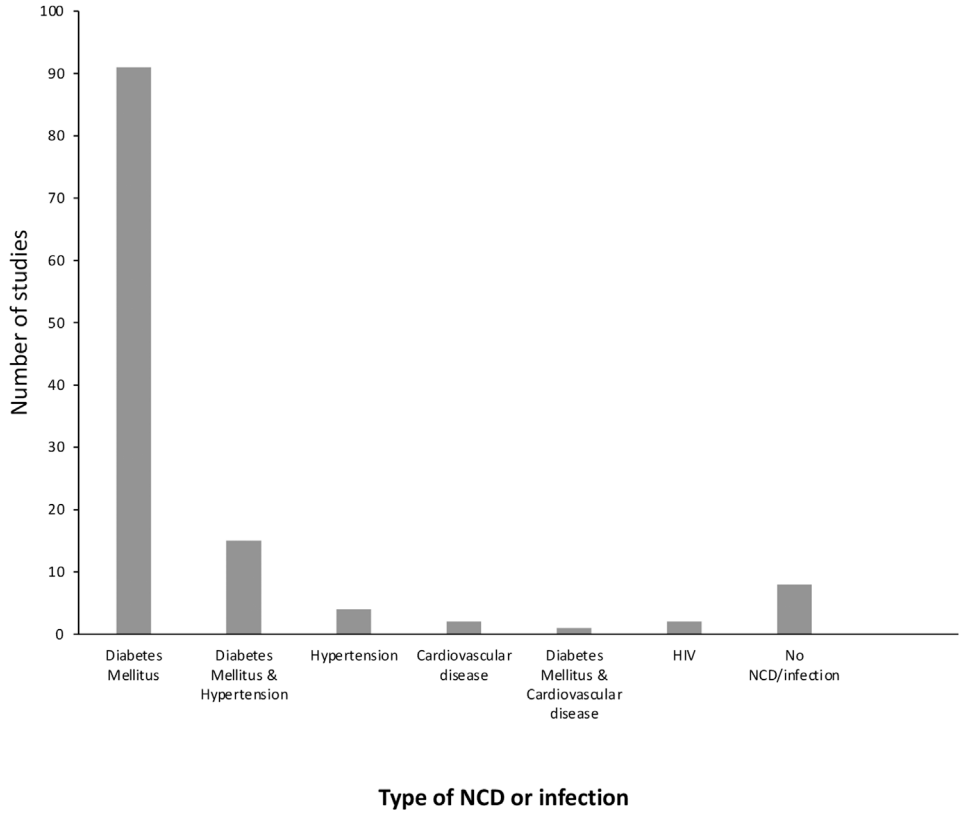


FIGURE 2 | The number of papers describing each of the comorbidities associated with CKD.

Biomarkers Associated With Early Detection of CKD

The proteomic CKD273 biomarker panel and metabolite pantothenic acid (PA) could potentially play a role in the early detection of CKD. Three proteomic studies by Siwy et al. (2014), Øvrehus et al. (2015), and Lindhardt et al. (2017) assessed the ability of a proteomic CKD273 biomarker panel to predict DN with diagnostic accuracy, measured as area under receiver operating characteristic curve, ranging from 0.84 to 1.0. All three proteomic studies used the CE-MS approach with sample sizes ranging from 35 to 737. In addition, Gonzalez-Calero et al. (2016) identified the metabolite PA as a potential predictor of albuminuria development in 118 hypertensive patients.

Biomarkers Associated With Progression of CKD

The proteomic CKD273 biomarker panel, metabolite PA, and methylation probe cg17944885 were suggested to be involved in progression of CKD. Two proteomic studies, Roscioni et al. (2013) and Tofte et al. (2020) found that a high CKD273 risk score was able to predict progression of microalbuminuria with an odds ratio (OR) of 1.35 and p value of $<1.0 \times 10^{-4}$, respectively. In addition to the metabolite PA being a potential predictor of albuminuria development in hypertensive patients, it was also associated with DN progression ($p = 5.0 \times 10^{-3}$) (Ma et al., 2020). Lecamwasam et al. (2020) compared methylation profiles between early and late stages of DN. They found that the methylation probe cg17944885 was associated with progressive renal dysfunction ($p = 0.01$ after adjustment for multiple testing).

Biomarkers Exploring Potential Disease Pathways

Genes of the retinoic acid pathway, *UMOD* and the metabolite PA provided insight into potential pathways involved in CKD. *UMOD* was specifically expressed in epithelial cells of the ascending loop of Henle in the kidney. This points to the role of uromodulin in kidney physiology (Morris et al., 2019). Genes of the retinoic acid pathway are essential for retinoic acid (RA) synthesis and impairment of RA synthesis leads to progression of kidney disease (Li et al., 2014). Pathway analysis of PA revealed that altered PA synthesis and PA-related metabolism could lead to DN (Ma et al., 2020).

DISCUSSION

In this scoping review the first objective was to identify studies on potential biomarkers associated with CKD. The biomarkers that were most commonly associated with risk or early detection of CKD were genes and genetic variants on chromosomes 16 and 22 (Kao et al., 2008; Shlush et al., 2010; Igo et al., 2011; Liao et al., 2014; Taira et al., 2018; Morris et al., 2019), the CKD273 biomarker panel (consisting of proteins such as collagen

TABLE 4 | Studies that identified biomarkers/indicators for early detection of CKD, prediction of CKD progression, and disease pathways.

Author and year	Type of 'omic platform	'Omic technique	Sample size	Biomarker/ indicator of disease	Effect estimate OR/ *p* value	Clinical phenotype/ outcome	What does the biomarker inform on?
Kao et al. (2008)	Genomic	1,354 SNP panel	2,178	rs16996674 (*MYH9*) rs16996677 (*MYH9*) rs5756152 (*MYH9*)	3.10 3.03 2.59	Non-diabetic ESKD	Susceptibility/risk
Shlush et al. (2010)	Genomic	2,016 SNP panel	833	rs7286127 (chromosome 22) rs5756133 (chromosome 22)	$\leq 10^{-3}$ $\leq 10^{-3}$	Non-diabetic ESKD	Susceptibility/risk
Igo Jr et al. (2011)	Genomic	Illumina IV panel of 5,500 SNPs	3,972	rs5769116 (10 Mb distal to *MYH9/ APOL1*)	1.5×10^{-3}	DN	Susceptibility/risk
Morris et al. (2019)	Genomic	A variety of GWAS arrays	312,468	rs77924615 (*UMOD*)	1.5×10^{-54}	CKD	Susceptibility/risk/ disease pathway
Fan et al. (2020)	Transcriptomic	RNA-sequencing	37	*RDH8* *RDH12* *RBP4*	0.02 0.03 0.03	DN	Disease pathway
Øvrehus et al. (2015)	Proteomic	CE-MS	35	CKD273	AUC of 0.977[a]	DN and Hypertension associated kidney disease	Early detection
Siwy et al. (2014)	Proteomic	CE-MS	165	CKD273	AUC of 0.95–1.0[a] <0.05	DN	Early detection
Lindhardt et al. (2017)	Proteomic	CE-MS	737	CKD273	AUC of 0.84[a] < 1.0×10^{-4}	DN	Early detection
Roscioni et al. (2013)	Proteomic	CE-MS	88	CKD273	1.35	DN	Progression
Tofte et al. (2020)	Proteomic	CE-MS	1,775	CKD273	$<1.0 \times 10^{-4}$	DN	Progression
Gonzalez-Calero et al. (2016)	Metabolomic	Nuclear magnetic resonance (NMR)	118	Pantothenate	$<1.0 \times 10^{-4}$	Hypertension associated kidney disease	Early detection
Ma et al. (2020)	Metabolomic	Liquid chromatography-mass spectrometry	68	Pantothenate	5.0×10^{-3}	DN	Progression/ disease pathway
Lecamwasam et al. (2020)	Epigenomic	Illumina Infinium MethylationEPIC BeadChip	119	cg17944885	0.01	DN	Progression

[a]*Diagnostic accuracy measured as area under receiver operating characteristic curve.*

fragments, uromodulin, α-1-antitrypsin, transthyretin and β-2-microglobulin) (Good et al., 2010; Siwy et al., 2014; Øvrehus et al., 2015; Lindhardt et al., 2017), and the metabolite PA (also known as pantothenate or vitamin B5) (Gonzalez-Calero et al., 2016). Majority of these potential biomarkers are exploratory, having been identified in small discovery studies and have not been validated, with the exception of the CKD273 biomarker panel.

It is well known from studies in the United States that individuals with African ancestry are at greater risk for developing incident CKD and progressing to ESKD, compared to other population groups (Perneger et al., 1995; Tarver-Carr et al., 2002; Saran et al., 2020). Much of this risk was attributed to genetic variants in the *MYH9/APOL1* gene region on chromosome 22 that were strongly associated with non-diabetic ESKD (Kao et al., 2008; Shlush et al., 2010). The *MYH9* and *APOL1* genes are tightly linked on chromosome 22, and fine mapping studies revealed that the risk variants were actually in the *APOL1* gene and not in the *MYH9* gene, as originally thought (Genovese et al., 2010; Tzur et al., 2010). Zhang et al. (2016) developed and validated an *APOL1* genotyping assay as a precision medicine genetic test for non-diabetic CKD risk prediction (Zhang et al., 2016). There are still ongoing clinical trials (http://www.clinicaltrials.gov; identifiers NCT02234063, NCT04191824 and NCT04910867), assessing the clinical utility of *APOL1* genetic testing. Implementation of an *APOL1* genotyping assay in a clinical setting could lead to a cost-effective screening program for African Americans who are at high risk for developing CKD and progressing to ESKD.

Studies have suggested the *UMOD* gene located on chromosome 16 as a candidate gene for CKD (Köttgen et al., 2009; Köttgen et al., 2010; Pattaro et al., 2016). Not only does it have a large effect on eGFR decline and CKD risk, the consistency of the effect is seen across different ethnicities (Devuyst et al., 2017). Morris et al. (2019) identified rs77924615 ($p = 1.5 \times 10^{-54}$), located in *UMOD*, as the main variant driving eGFR association. This variant was associated with decreased eGFR and increased *UMOD* expression (Morris et al., 2019).

Due to the multifactorial nature of CKD, a combination approach of several biomarkers is likely to be more predictive than a single biomarker. Compared to currently available markers such as reduced eGFR and albuminuria, the CKD273 biomarker panel enables more accurate diagnosis and earlier detection of CKD. For example, in one study, CKD273 enabled detection of CKD with 95% sensitivity, 100% specificity and an excellent

diagnostic accuracy of 0.977 (95% confidence interval (CI) 0.930–1.000) (Øvrehus et al., 2015). In another study, CKD273 had the ability to detect DN in individuals with type 2 diabetes, with diagnostic accuracy ranging from 0.95 to 1.0 across nine study sites (Siwy et al., 2014). This same biomarker panel was able to predict the development of microalbuminuria in normoalbuminuric type 2 diabetic patients, with a diagnostic accuracy of 0.79 (95% CI 0.75–0.84; $p < 1.0 \times 10^{-4}$) (Lindhardt et al., 2017). The CKD273 biomarker panel includes clinically relevant proteins such as uromodulin and β-2-microglobulin, which are suggested to be markers of early renal damage (Torffvit and Agardh, 1993; Torffvit et al., 1999; Schlatzer et al., 2012; Zeng et al., 2014; Alaje Abiodun et al., 2016). These findings demonstrate the utility of CKD273 in early non-invasive renal risk assessment and may help to prevent unnecessary biopsies. While CKD273 has been validated for DN, its utility for hypertension-associated CKD is yet to be established. The CKD273 biomarker can assist clinicians in decision-making for more frequent check-ups in high-risk patients and guide therapeutic options to prevent long-term chronic conditions. Although the CKD273 biomarker panel has been supported by the FDA (https://www.fda.gov/files/drugs/published/Biomarker-Letter-of-Support_Mischak.pdf), its utility has yet to be established in African populations and the African diaspora. Furthermore, the current cost of screening using the CKD273 biomarker panel is €3,053 per patient making it unaffordable in resource-poor settings such as SSA (Critselis et al., 2018).

CKD273, PA, and methylation probe cg17944885 are suggested to be potential biomarkers associated with CKD progression. In addition to PA being used to predict albuminuria in hypertensive normoalbuminuric patients, it is suggested to be a promising biomarker for predicting diabetes-induced kidney impairment (Gonzalez-Calero et al., 2016; Ma et al., 2020). Methylation marker cg17944885 showed an increasing gradient of methylation with progressive renal dysfunction (Lecamwasam et al., 2020). A high CKD273 risk score was independently associated with faster progression of albuminuria in two separate studies (Roscioni et al., 2013; Tofte et al., 2020). The ability to identify individuals at high-risk for progression has important clinical implications. For those patients at a high-risk for progression, it is suggested that blood pressure be controlled to less than 130/80 mmHg. There are many therapeutic options to control blood pressure including angiotensin-converting enzyme (ACE) inhibitors, angiotensin-receptor blockers (ARBs) or a combination of ACE inhibitors or ARBs with a diuretic (Nguyen et al., 2010). Meanwhile, management of glucose and lipid levels are also suggested to delay the development of chronic complications in DN patients.

The second objective of the scoping review was to explore potential pathways involved in CKD. Using expression quantitative trait loci (eQTL) analysis, *UMOD* expression was mapped to epithelial cells lining the thick ascending limb of the loop of Henle (Morris et al., 2019). Studies on transgenic mice suggested that overexpression of uromodulin causes abnormal activation of the sodium-potassium-chloride transporter (NKCC2) involved in sodium chloride (NaCl) reabsorption by the thick ascending limb segment, leading to salt-sensitive hypertension and renal lesions (Trudu et al., 2013). Mice treated with a single dose of furosemide, a diuretic that targets NKCC2, exhibited a significant reduction in blood pressure (Trudu et al., 2013). This finding suggests that NKCC2 is a potential target for hypertension-induced CKD intervention.

Fan et al. (2020) compared gene expression profiles in three groups; controls, early DN and advanced DN (Fan et al., 2020). They identified genes of the retinoic acid pathway including retinol dehydrogenase 8 (*RDH8*), retinol dehydrogenase 12 (*RDH12*), and retinol binding protein 4 (*RBP4*) with increased expression in early DN but decreased expression in advanced DN. These findings suggest that RA might have kidney protective effects in the early stage of DN that slow the progression to advanced DN. A previous study suggested that downregulation of enzymes retinol dehydrogenase 9 (RDH9) and retinol dehydrogenase 1 (RDH1) essential for RA synthesis causes impaired local synthesis of RA in the kidney that contributes to the progression of kidney disease (Li et al., 2014). These results point to restoration of retinoic acid synthesis as a potential therapy to slow the progression of kidney disease. There is an ongoing clinical trial assessing the efficacy of RA isotretinoin for treatment of various podocyte diseases including collapsing glomerulopathy (http://www.clinicaltrials.gov; identifier NCT00098020).

The metabolite PA is suggested to be a predictive biomarker for development of CKD and progression in hypertensive and diabetic patients (Gonzalez-Calero et al., 2016; Ma et al., 2020). PA (also known as vitamin B5) is an essential vitamin and dietary factor, and has long been recognised as an essential cofactor of biochemical reactions in numerous physiological processes (Patassini et al., 2019). Pathway analysis results showed PA synthesis and PA-related metabolism were disturbed in individuals with DN (Ma et al., 2020). Since mammalian cells are unable to synthesize vitamin B5, they require it to be generated by the intestinal microbiome or to be part of the diet, resulting in a potential therapeutic target for DN treatment (Daugherty et al., 2002). A previous study in a type 2 diabetes rat model has shown that a vitamin B5 supplement restored body weight and blood pressure whilst also decreasing blood glucose levels (Demirci et al., 2014). In addition, this study also reported that a vitamin B5 supplement may reduce diabetes-related oxidative stress and prevent endothelial cell injury (Demirci et al., 2014). However, further studies are required to determine the efficacy of a vitamin B5 supplement in reducing blood pressure and glucose levels in humans.

The translation of 'omics research into disease-related biomarkers for use in clinical practice is a complicated, time-consuming, and expensive process. Some challenges include clinical heterogeneity of individuals with a particular diagnosis, different standards of evidence to evaluate the validity of the biomarker, and a lengthy developmental pathway from biomarker discovery to clinical utility (Omenn et al., 2012). The diagnostic and therapeutic potential of the biomarker has to be carefully assessed before being considered for introduction into the clinic, and additional uses should be considered. For example, the CKD273 biomarker panel is available as a commercial *in vitro* diagnostic test and is a good example of a

"bench-to-bedside" approach (Pontillo and Mischak, 2017). In addition to its diagnostic and prognostic value, CKD273 also has the potential to assess the impact (outcome) of therapeutic interventions (Haubitz et al., 2009). For example, in one study, CKD273 risk scores were significantly reduced (from 0.721 at visit 2 to 0.277 at visit 9), below the previously defined threshold (0.343) after 2 years of treatment with irbesartan (an antihypertensive and antiproteinuric drug) (Andersen et al., 2010).

Low-cost genomics-informed approaches to disease risk prediction pose great promise for low- and middle-income countries, where the morbidity and mortality of non-communicable diseases such as CKD are disproportionately high. However, there are many factors that hinder the implementation of a genomic medicine healthcare model for CKD in resource limited settings. These include population-specific research to validate predictive accuracy and clinical utility, a low critical mass of skilled scientists, insufficiently equipped diagnostic research laboratories and a lagging public health policy framework (Tekola-Ayele and Rotimi, 2015). For example, the paucity of genomics research in individuals from Africa has limited knowledge of clinically relevant CKD genetic variants in these populations. The clinical implementation of genomic medicine requires more than just developing a diagnostic or screening test. To have a functional genomic medicine healthcare model for CKD, costly technology and skill-strengthening among scientists in the fields of medical genetics, genetic counselling, genetic epidemiology, bioinformatics and computational biology are required. In addition, legal and ethical frameworks also need to be strengthened in local contexts (Tekola-Ayele and Rotimi, 2015; Pereira et al., 2021).

The strengths of the review include the novelty of the topic and the assessment of the literature on candidate biomarkers for early detection of CKD and prediction of progression. Several potential biomarkers are highlighted and may be useful in developing screening programs for CKD, however, they need validation in different clinical settings and ethnic backgrounds. In addition, this review has shown that 'omics technologies can be used to explore potential disease mechanisms and therapeutic targets for treatment. There were some limitations in this scoping review. Despite internationally accepted guidelines, studies used various metrics to define CKD and potential sex differences in the manifestation of CKD were not examined. The sample sizes were usually small, and results may not be transferable to populations that were not included in the studies, for example African populations. Furthermore, it was difficult to compare the outcomes across studies because each study design was unique.

A few knowledge gaps were identified. While DM is a major risk factor for CKD in some regions of the world, reflected in the number of DN studies reported in the 'omics literature, other regions such as Asia and Africa have other dominant risk factors such as hypertension and other contributing factors that remain to be identified. The major underlying causes of kidney disease in Africa are poorly understood and often attributed to hypertension which is likely to be secondary hypertension as a consequence of numerous insults. In this regard, there are relatively few 'omics studies that focus on hypertension in the context of CKD in African populations (Alicic et al., 2017). There was a single study with individuals from West Africa, highlighting the lack of 'omics research in continental Africans (Chen et al., 2007), despite many studies documenting the well-established genetic link to CKD in African Americans (*APOL1*-associated nephropathy) (George et al., 2018).

In sub-Saharan Africa, there are limited resources for treating individuals with CKD who progress to ESKD, rendering this condition fatal for those who are not treated. At least half of the patients who require ESKD treatment do not have access to treatment, and only 1.5% are predicted to receive renal replacement therapy (Ashuntantang et al., 2017). Approximately 75% of patients who begin dialysis will subsequently demise due to poor quality of dialysis or termination of dialysis based on its cost (Ashuntantang et al., 2017). This makes screening, early detection and the implementation of preventative interventions for CKD particularly relevant. Clinically relevant biomarkers will be useful to develop screening programs and potential therapeutic approaches to treat CKD in sub-Saharan Africa. Ideally the goal is to identify individuals at risk for developing CKD in the future and develop preventative intervention strategies that would delay onset and slow progression, by treating the underlying etiologies.

The reported 'omics studies identified potential biomarkers associated with CKD. These biomarkers provide insight into the pathways involved in CKD and could possibly be used to guide diagnostic and treatment options. The biomarkers that have emerged through synthesis of existing literature provide evidence for effective approaches for early detection of CKD and prediction of progression in European populations. Their transferability to Africans and other ethnicities needs to be verified through research in relevant populations and assessment of diagnostic implementation in low-resourced healthcare settings. There is extensive regional and ethnic diversity in Africa, and the major causes of CKD appear to differ substantively from those in high-income settings, requiring population representative studies that address the local context and causes of CKD. In addition, many factors need to be addressed to facilitate effective genomic medicine healthcare models for CKD.

DATA AVAILABILITY STATEMENT

The search terms and methodology are fully described in the paper and the list of papers included in the review are shown in the **Supplementary Material**.

AUTHOR CONTRIBUTIONS

MG screened all records by title and abstract for inclusion. J-TB and JF independently screened records by title and abstract for verification. Any inconsistencies or disagreements were resolved by mutual agreement including investigator MR. Writing of the

paper was done by MG. MR, JF, and J-TB contributed to editing the paper.

FUNDING

MG has a bursary from the South African National Research Foundation (NRF); J-TB's postdoctoral fellowship was funded by the NRF and MR is a South African Research Chair in Genomics and Bioinformatics of African populations hosted by the University of the Witwatersrand, funded by the Department of Science and Technology, and administered by the NRF.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.682929/full#supplementary-material

REFERENCES

Abbiss, H., Maker, G., and Trengove, R. (2019). Metabolomics Approaches for the Diagnosis and Understanding of Kidney Diseases. *Metabolites* 9, 34. doi:10.3390/metabo9020034

Alaje Abiodun, K., Benedict, N., and Idogun Sylvester, E. (2016). β2 Microglobulin as A Marker of Early Renal Damage in Patients with Sickle Cell Nephropathy. *IOSR J. Dental Med. Sci.* 15, 118–121. doi:10.9790/0853-150805118121

Alicic, R. Z., Rooney, M. T., and Tuttle, K. R. (2017). Diabetic Kidney Disease. *Cjasn* 12, 2032–2045. doi:10.2215/cjn.11491116

Alkhalaf, A., Zürbig, P., Bakker, S. J. L., Bilo, H. J. G., Cerna, M., Fischer, C., et al. (2010). Multicentric Validation of Proteomic Biomarkers in Urine Specific for Diabetic Nephropathy. *PloS one* 5, e13421. doi:10.1371/journal.pone.0013421

Andersen, S., Mischak, H., Zürbig, P., Parving, H. H., and Rossing, P. (2010). Urinary Proteome Analysis Enables Assessment of Renoprotective Treatment in Type 2 Diabetic Patients with Microalbuminuria. *BMC Nephrol.* 11, 29–9. doi:10.1186/1471-2369-11-29

Ashuntantang, G., Osafo, C., Olowu, W. A., Arogundade, F., Niang, A., Porter, J., et al. (2017). Outcomes in Adults and Children with End-Stage Kidney Disease Requiring Dialysis in Sub-saharan Africa: a Systematic Review. *Lancet Glob. Health* 5, e408–e417. doi:10.1016/s2214-109x(17)30057-8

Bikbov, B., Purcell, C. A., Levey, A. S., Smith, M., Abdoli, A., Abebe, M., et al. (2020). Global, Regional, and National burden of Chronic Kidney Disease, 1990-2017: a Systematic Analysis for the Global Burden of Disease Study 2017. *Lancet* 395, 709–733. doi:10.1016/S0140-6736(20)30045-3

Cañadas-Garre, M., Anderson, K., Cappa, R., Skelly, R., Smyth, L. J., Mcknight, A. J., et al. (2019). Genetic Susceptibility to Chronic Kidney Disease - Some More Pieces for the Heritability Puzzle. *Front. Genet.* 10, 453. doi:10.3389/fgene.2019.00453

Chen, G., Adeyemo, A. A., Zhou, J., Chen, Y., Doumatey, A., Lashley, K., et al. (2007). A Genome-wide Search for Linkage to Renal Function Phenotypes in West Africans with Type 2 Diabetes. *Am. J. kidney Dis.* 49, 394–400. doi:10.1053/j.ajkd.2006.12.011

Couser, W. G., Remuzzi, G., Mendis, S., and Tonelli, M. (2011). The Contribution of Chronic Kidney Disease to the Global burden of Major Noncommunicable Diseases. *Kidney Int.* 80, 1258–1270. doi:10.1038/ki.2011.368

Craddock, N. J., and Jones, I. R. (2007). Genome-wide Association Study of 14,000 Cases of Seven Common Diseases and 3,000 Shared Controls. *Nature* 447, 661–678. doi:10.1038/nature05911

Critselis, E., Vlahou, A., Stel, V. S., and Morton, R. L. (2018). Cost-effectiveness of Screening Type 2 Diabetes Patients for Chronic Kidney Disease Progression with the CKD273 Urinary Peptide Classifier as Compared to Urinary Albumin Excretion. *Nephrol. Dial. Transplant.* 33, 441–449. doi:10.1093/ndt/gfx068

Daugherty, M., Polanuyer, B., Farrell, M., Scholle, M., Lykidis, A., de Crécy-Lagard, V., et al. (2002). Complete Reconstitution of the Human Coenzyme A Biosynthetic Pathway via Comparative Genomics. *J. Biol. Chem.* 277, 21431–21439. doi:10.1074/jbc.m201708200

De Oliveira, B. G. R. B. (2018). Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews (PRISMA-ScR) Checklist.

Demirci, B., Demir, O., Dost, T., and Birincioglu, M. (2014). Protective Effect of Vitamin B5 (Dexpanthenol) on Cardiovascular Damage Induced by Streptozocin in Rats. *Bll* 115, 190–196. doi:10.4149/bll_2014_040

Devuyst, O., Olinger, E., and Rampoldi, L. (2017). Uromodulin: from Physiology to Rare and Complex Kidney Disorders. *Nat. Rev. Nephrol.* 13, 525–544. doi:10.1038/nrneph.2017.101

Ekrikpo, U. E., Kengne, A. P., Bello, A. K., Effa, E. E., Noubiap, J. J., Salako, B. L., et al. (2018). Chronic Kidney Disease in the Global Adult HIV-Infected Population: a Systematic Review and Meta-Analysis. *PLoS One* 13, e0195443. doi:10.1371/journal.pone.0195443

Fan, Y., Yi, Z., D'Agati, V. D., Sun, Z., Zhong, F., Zhang, W., et al. (2020). Erratum. Comparison of Kidney Transcriptomic Profiles of Early and Advanced Diabetic Nephropathy Reveals Potential New Mechanisms for Disease Progression. Diabetes 2019;68:2301-2314. *Diabetes* 69, db20er04b. doi:10.2337/db20-er04b

Genovese, G., Friedman, D. J., Ross, M. D., Lecordier, L., Uzureau, P., Freedman, B. I., et al. (2010). Association of Trypanolytic ApoL1 Variants with Kidney Disease in African Americans. *Science* 329, 841–845. doi:10.1126/science.1193032

George, C., Yako, Y. Y., Okpechi, I. G., Matsha, T. E., Kaze Folefack, F. J., and Kengne, A. P. (2018). An African Perspective on the Genetic Risk of Chronic Kidney Disease: a Systematic Review. *BMC Med. Genet.* 19, 187. doi:10.1186/s12881-018-0702-x

Gonzalez-Calero, L., Martin-Lorenzo, M., Martínez, P. J., Baldan-Martin, M., Ruiz-Hurtado, G., Segura, J., et al. (2016). Hypertensive Patients Exhibit an Altered Metabolism. A Specific Metabolite Signature in Urine Is Able to Predict Albuminuria Progression. *Translational Res.* 178, 25–37. doi:10.1016/j.trsl.2016.07.003

Good, D. M., Zürbig, P., Argilés, À., Bauer, H. W., Behrens, G., Coon, J. J., et al. (2010). Naturally Occurring Human Urinary Peptides for Use in Diagnosis of Chronic Kidney Disease. *Mol. Cell Proteomics* 9, 2424–2437. doi:10.1074/mcp.m110.001917

Haubitz, M., Good, D. M., Woywodt, A., Haller, H., Rupprecht, H., Theodorescu, D., et al. (2009). Identification and Validation of Urinary Biomarkers for Differential Diagnosis and Evaluation of Therapeutic Intervention in Anti-neutrophil Cytoplasmic Antibody-Associated Vasculitis. *Mol. Cell Proteomics* 8, 2296–2307. doi:10.1074/mcp.m800529-mcp200

Hill, N. R., Fatoba, S. T., Oke, J. L., Hirst, J. A., O'Callaghan, C. A., Lasserson, D. S., et al. (2016). Global Prevalence of Chronic Kidney Disease - A Systematic Review and Meta-Analysis. *PloS one* 11, e0158765. doi:10.1371/journal.pone.0158765

Horgan, R. P., and Kenny, L. C. (2011). 'Omic' Technologies: Genomics, Transcriptomics, Proteomics and Metabolomics. *The Obstetrician & Gynaecologist* 13, 189–195. doi:10.1576/toag.13.3.189.27672

Igo, Jr., R. P., Iyengar, S. K., Nicholas, S. B., Goddard, K. A. B., Langefeld, C. D., Hanson, R. L., et al. (2011). Genomewide Linkage Scan for Diabetic Renal Failure and Albuminuria: the FIND Study. *Am. J. Nephrol.* 33, 381–389. doi:10.1159/000326763

Kao, W. H. L., Klag, M. J., Klag, M. J., Meoni, L. A., Reich, D., Berthier-Schaad, Y., et al. (2008). MYH9 Is Associated with Nondiabetic End-Stage Renal Disease in African Americans. *Nat. Genet.* 40, 1185–1192. doi:10.1038/ng.232

Kaze, A. D., Ilori, T., Jaar, B. G., and Echouffo-Tcheugui, J. B. (2018). Burden of Chronic Kidney Disease on the African Continent: a Systematic Review and Meta-Analysis. *BMC Nephrol.* 19, 1–11. doi:10.1186/s12882-018-0930-5

Köttgen, A., Pattaro, C., Böger, C. A., Fuchsberger, C., Olden, M., Glazer, N. L., et al. (2010). New Loci Associated with Kidney Function and Chronic Kidney Disease. *Nat. Genet.* 42, 376–384. doi:10.1038/ng.568

Köttgen, A., Glazer, N. L., Dehghan, A., Hwang, S.-J., Katz, R., Li, M., et al. (2009). Multiple Loci Associated with Indices of Renal Function and Chronic Kidney Disease. *Nat. Genet.* 41, 712–717. doi:10.1038/ng.377

Lecamwasam, A., Novakovic, B., Meyer, B., Ekinci, E. I., Dwyer, K. M., and Saffery, R. (2020). DNA Methylation Profiling Identifies Epigenetic Differences between Early versus Late Stages of Diabetic Chronic Kidney Disease. *Nephrol. Dial. Transplant.*, gfaa226. doi:10.1093/ndt/gfaa226

Levey, A. S., Coresh, J., Bolton, K., Culleton, B., Harvey, K. S., Ikizler, T. A., et al. (2002). K/DOQI Clinical Practice Guidelines for Chronic Kidney Disease: Evaluation, Classification, and Stratification. *Am. J. Kidney Dis.* 39, S1–S266.

Li, X., Dai, Y., Chuang, P. Y., and He, J. C. (2014). Induction of Retinol Dehydrogenase 9 Expression in Podocytes Attenuates Kidney Injury. *Jasn* 25, 1933–1941. doi:10.1681/asn.2013111150

Liao, L. N., Chen, C. C., Wu, F. Y., Lin, C. C., Hsiao, J. H., Chang, C. T., et al. (2014). Identified Single-Nucleotide Polymorphisms and Haplotypes at 16q22.1 Increase Diabetic Nephropathy Risk in Han Chinese Population. *BMC Genet.* 15, 1–11. doi:10.1186/s12863-014-0113-8

Lindhardt, M., Persson, F., Zürbig, P., Stalmach, A., Mischak, H., De Zeeuw, D., et al. (2017). Urinary Proteomics Predict Onset of Microalbuminuria in Normoalbuminuric Type 2 Diabetic Patients, a Sub-study of the DIRECT-Protect 2 Study. *Nephrol. Dial. Transpl.* 32, 1866–1873. doi:10.1093/ndt/gfw292

Lindhardt, M., Persson, F., Currie, G., Pontillo, C., Beige, J., Delles, C., et al. (2016). Proteomic Prediction and Renin Angiotensin Aldosterone System Inhibition Prevention of Early Diabetic nephRopathy in TYpe 2 Diabetic Patients with Normoalbuminuria (PRIORITY): Essential Study Design and Rationale of a Randomised Clinical Multicentre Trial. *BMJ open* 6, e010310. doi:10.1136/bmjopen-2015-010310

Ma, T., Liu, T., Xie, P., Jiang, S., Yi, W., Dai, P., et al. (2020). UPLC-MS-based Urine Nontargeted Metabolic Profiling Identifies Dysregulation of Pantothenate and CoA Biosynthesis Pathway in Diabetic Kidney Disease. *Life Sci.* 258, 118160. doi:10.1016/j.lfs.2020.118160

Miller, W. G., Bruns, D. E., Hortin, G. L., Sandberg, S., Aakre, K. M., Mcqueen, M. J., et al. (2009). Current Issues in Measurement and Reporting of Urinary Albumin Excretion. *Clin. Chem.* 55, 24–38. doi:10.1373/clinchem.2008.106567

Mischak, H., Delles, C., Klein, J., and Schanstra, J. P. (2010). Urinary Proteomics Based on Capillary Electrophoresis-Coupled Mass Spectrometry in Kidney Disease: Discovery and Validation of Biomarkers, and Clinical Application. *Adv. chronic kidney Dis.* 17, 493–506. doi:10.1053/j.ackd.2010.09.004

Morris, A. P., Le, T. H., Wu, H., Akbarov, A., Van Der Most, P. J., Hemani, G., et al. (2019). Trans-ethnic Kidney Function Association Study Reveals Putative Causal Genes and Effects on Kidney-specific Disease Aetiologies. *Nat. Commun.* 10, 29–14. doi:10.1038/s41467-018-07867-7

Nguyen, Q., Dominguez, J., Nguyen, L., and Gullapalli, N. (2010). Hypertension Management: an Update. *Am. Health Drug Benefits* 3, 47–56.

Omenn, G. S., Nass, S. J., and Micheel, C. M. (2012). *Evolution of Translational Omics: Lessons Learned and the Path Forward*. Washington (DC): National Academies Press.

Øvrehus, M. A., Zürbig, P., Vikse, B. E., and Hallan, S. I. (2015). Urinary Proteomics in Chronic Kidney Disease: Diagnosis and Risk of Progression beyond Albuminuria. *Clin. Proteomics* 12, 21–29. doi:10.1186/s12014-015-9092-7

Papadopoulos, T., Krochmal, M., Cisek, K., Fernandes, M., Husi, H., Stevens, R., et al. (2016). Omics Databases on Kidney Disease: where They Can Be Found and How to Benefit from Them. *Clin. Kidney J.* 9, 343–352. doi:10.1093/ckj/sfv155

Patassini, S., Begley, P., Xu, J., Church, S., Kureishy, N., Reid, S., et al. (2019). Cerebral Vitamin B5 (D-Pantothenic Acid) Deficiency as a Potential Cause of Metabolic Perturbation and Neurodegeneration in Huntington's Disease. *Metabolites* 9, 113. doi:10.3390/metabo9060113

Pattaro, C., Teumer, A., Gorski, M., Chu, A. Y., Li, M., Mijatovic, V., et al. (2016). Genetic Associations at 53 Loci Highlight Cell Types and Biological Pathways Relevant for Kidney Function. *Nat. Commun.* 7, 1–19. doi:10.1038/ncomms10023

Pereira, L., Mutesa, L., Tindana, P., and Ramsay, M. (2021). African Genetic Diversity and Adaptation Inform a Precision Medicine Agenda. *Nat. Rev. Genet.* 22, 284–306. doi:10.1038/s41576-020-00306-8

Perneger, T. V., Whelton, P. K., and Klag, M. J. (1995). Race and End-Stage Renal Disease. Socioeconomic Status and Access to Health Care as Mediating Factors. *Arch. Intern. Med.* 155, 1201–1208. doi:10.1001/archinte.155.11.1201

Pontillo, C., and Mischak, H. (2017). Urinary Peptide-Based Classifier CKD273: towards Clinical Application in Chronic Kidney Disease. *Clin. kidney J.* 10, 192–201. doi:10.1093/ckj/sfx002

Roscioni, S. S., De Zeeuw, D., Hellemons, M. E., Mischak, H., Zürbig, P., Bakker, S. J. L., et al. (2013). A Urinary Peptide Biomarker Set Predicts Worsening of Albuminuria in Type 2 Diabetes Mellitus. *Diabetologia* 56, 259–267. doi:10.1007/s00125-012-2755-2

Saran, R., Robinson, B., Abbott, K. C., Bragg-Gresham, J., Chen, X., Gipson, D., et al. (2020). US Renal Data System 2019 Annual Data Report: Epidemiology of Kidney Disease in the United States. *Am. J. Kidney Dis.* 75, A6–A7. doi:10.1053/J.AJKD.2019.09.003

Schlatzer, D., Maahs, D. M., Chance, M. R., Dazard, J.-E., Li, X., Hazlett, F., et al. (2012). Novel Urinary Protein Biomarkers Predicting the Development of Microalbuminuria and Renal Function Decline in Type 1 Diabetes. *Diabetes care* 35, 549–555. doi:10.2337/dc11-1491

Shlush, L. I., Bercovici, S., Wasser, W. G., Yudkovsky, G., Templeton, A., Geiger, D., et al. (2010). Admixture Mapping of End Stage Kidney Disease Genetic Susceptibility Using Estimated Mutual Information Ancestry Informative Markers. *BMC Med. Genomics* 3, 47–12. doi:10.1186/1755-8794-3-47

Siwy, J., Schanstra, J. P., Argiles, A., Bakker, S. J. L., Beige, J., Boucek, P., et al. (2014). Multicentre Prospective Validation of a Urinary Peptidome-Based Classifier for the Diagnosis of Type 2 Diabetic Nephropathy. *Nephrol. Dial. Transplant.* 29, 1563–1570. doi:10.1093/ndt/gfu039

Taira, M., Imamura, M., Takahashi, A., Kamatani, Y., Yamauchi, T., Araki, S.-I., et al. (2018). A Variant within the FTO Confers Susceptibility to Diabetic Nephropathy in Japanese Patients with Type 2 Diabetes. *PloS one* 13, e0208654. doi:10.1371/journal.pone.0208654

Tarver-Carr, M. E., Powe, N. R., Eberhardt, M. S., Laveist, T. A., Kington, R. S., Coresh, J., et al. (2002). Excess Risk of Chronic Kidney Disease Among African-American versus white Subjects in the United States: a Population-Based Study of Potential Explanatory Factors. *Jasn* 13, 2363–2370. doi:10.1097/01.asn.0000026493.18542.6a

Tekola-Ayele, F., and Rotimi, C. N. (2015). Translational Genomics in Low- and Middle-Income Countries: Opportunities and Challenges. *Public health genomics* 18, 242–247. doi:10.1159/000433518

Tin, A., and Köttgen, A. (2020). Genome-Wide Association Studies of CKD and Related Traits. *Clin. J. Am. Soc. Nephrol.* 15, 1643–1656. doi:10.2215/cjn.00020120

Tofte, N., Lindhardt, M., Adamova, K., Bakker, S. J., Beige, J., Beulens, J. W., et al. (2020). Early Detection of Diabetic Kidney Disease by Urinary Proteomics and Subsequent Intervention with Spironolactone to Delay Progression (PRIORITY): a Prospective Observational Study and Embedded Randomised Placebo-Controlled Trial. *Lancet Diabetes Endocrinol.* 8, p301–p312.

Torffvit, O., Agardh, C.-D., and Thulin, T. (1999). A Study of Tamm-Horsfall Protein Excretion in Hypertensive Patients and Type 1 Diabetic Patients. *Scand. J. Urol. Nephrol.* 33, 187–191. doi:10.1080/003655999750015970

Torffvit, O., and Agardh, C.-D. (1993). Tubular Secretion of Tamm-Horsfall Protein Is Decreased in Type 1 (Insulin-dependent) Diabetic Patients with Diabetic Nephropathy. *Nephron* 65, 227–231. doi:10.1159/000187479

Trudu, M., Janas, S., Janas, S., Lanzani, C., Debaix, H., Schaeffer, C., et al. (2013). Common Noncoding UMOD Gene Variants Induce Salt-Sensitive Hypertension and Kidney Damage by Increasing Uromodulin Expression. *Nat. Med.* 19, 1655–1660. doi:10.1038/nm.3384

Tzur, S., Rosset, S., Shemer, R., Yudkovsky, G., Selig, S., Tarekegn, A., et al. (2010). Missense Mutations in the APOL1 Gene Are Highly Associated with End Stage Kidney Disease Risk Previously Attributed to the MYH9 Gene. *Hum. Genet.* 128, 345–350. doi:10.1007/s00439-010-0861-0

Zeng, X., Hossain, D., Bostwick, D. G., Herrera, G. A., and Zhang, P. L. (2014). Urinary β2-microglobulin Is a Good Indicator of Proximal Tubule Injury: a Correlative Study with Renal Biopsies. *J. biomarkers* 2014, 492838. doi:10.1155/2014/492838

Zhang, J., Fedick, A., Wasserman, S., Zhao, G., Edelmann, L., Bottinger, E. P., et al. (2016). Analytical Validation of a Personalized Medicine APOL1 Genotyping Assay

for Nondiabetic Chronic Kidney Disease Risk Assessment. *J. Mol. Diagn.* 18, 260–266. doi:10.1016/j.jmoldx.2015.11.003

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Genetics

PERSPECTIVE
published: 23 November 2021
doi: 10.3389/fgene.2021.758563

Population Structure of the South West Indian Ocean Islands: Implications for Precision Medicine

*Anisah W. Ghoorah[1], Toto Chaplain[2], Rakotoarivony Rindra[3], Smita Goorah[1], Ganessen Chinien[1] and Yasmina Jaufeerally-Fakim[1]**

[1]University of Mauritius, Reduit, Mauritius, [2]University of Toamasina, Barikadimy, Toamasina, Madagascar, [3]University of Antananarivo, Antananarivo, Madagascar

Precision medicine has brought new hopes for patients around the world with the applications of novel technologies for understanding genetics of complex diseases and their translation into clinical services. Such applications however require a foundation of skills, knowledge and infrastructure to translate genetics for health care. The crucial element is no doubt the availability of genomics data for the target populations, which is seriously lacking for most parts of Africa. We discuss here why it is vital to prioritize genomics data for the South West Indian Ocean region where a mosaic of ethnicities co-exist. The islands of the SWIO, which comprise Madagascar, La Reunion, Mauritius, Seychelles and Comoros, have been the scene for major explorations and trade since the 17th century being on the route to Asia. This part of the world has lived through active passage of slaves from East Africa to Arabia and further. Today's demography of the islands is a diverse mix of ancestries including European, African and Asian. The extent of admixtures has yet to be resolved. Except for a few studies in Madagascar, there is very little published data on human genetics for these countries. Isolation and small population sizes have likely resulted in reduced genetic variation and possible founder effects. There is a significant prevalence of diabetes, particularly in individuals of Indian descent, while breast and prostate cancers are on the rise. The island of La Reunion is a French overseas territory with a high standard of health care and close ties to Mauritius. Its demography is comparable to that of Mauritius but with a predominantly mixed population and a smaller proportion of people of Indian descent. On the other hand, Madagascar's African descendants inhabit mostly the lower coastal zones of the West and South regions, while the upper highlands are occupied by peoples of mixed African-Indonesian ancestries. Historical records confirm the Austronesian contribution to the Madagascar genomes. With the rapid progress in genomic medicine, there is a growing demand for sequencing services in the clinical settings to explore the incidence of variants in candidate disease genes and other markers. Genome sequence data has become a priority in order to understand the population sub-structures and to identify specific pathogenic variants among the different groups of inhabitants on the islands. Genomic data is increasingly being used to advise families at risk and propose diagnostic screening measures to enhance the success of therapies. This paper discusses the complexity of the islands' populations and argues for the needs for genotyping and understanding the genetic

OPEN ACCESS

Edited by:
Tivani Phosa Mashamba-Thompson, University of Pretoria, South Africa

Reviewed by:
Martin Maiers, National Marrow Donor Program, United States
Juan Carlos Fernandez-Lopez, Instituto Nacional de Medicina Genómica (INMEGEN), Mexico

***Correspondence:**
Yasmina Jaufeerally-Fakim
yasmina@uom.ac.mu

Specialty section:
This article was submitted to Human and Medical Genomics, a section of the journal Frontiers in Genetics

Received: *14 August 2021*
Accepted: *28 October 2021*
Published: *23 November 2021*

Citation:
Ghoorah AW, Chaplain T, Rindra R, Goorah S, Chinien G and Jaufeerally-Fakim Y (2021) Population Structure of the South West Indian Ocean Islands: Implications for Precision Medicine. Front. Genet. 12:758563. doi: 10.3389/fgene.2021.758563

factors associated with disease risks. The benefits to patients and improvement in health services through a concerted regional effort are depicted. Some private patients are having recourse to external facilities for molecular profiling with no return of data for research. Evidence of disease variants through sequencing represents a valuable source of medical data that can guide policy decisions at the national level. There are presently no such records for future implementation of strategies for genomic medicine.

Keywords: precision medicine, genetics, South West Indian Ocean, demography, population structure, disease variants

INTRODUCTION

The SWIO covers four island states off the shores of Eastern and Southern Africa with the largest one being Madagascar; the others include Mauritius, Seychelles, Comoros and a French island territory (La Reunion). Geographical proximity has contributed to all the islands having similar historical paths and were at times considered as a single entity during the colonial period. This region has been an important trading route between Europe and Asia since the 18th century. There were many events of migration for various reasons but mainly as importation of labor from mainland Africa and India. This activity has shaped the present day demography of the region resulting in a unique diversity and cultural traditions.

All the islands share some common history of sugarcane plantations, the use of both French and English language and the Creole dialects of French origin. The demographics have been shaped through colonization by the Dutch, British and French since the 17th century, by the slave trade in the 18th and 19th centuries and by the arrival of laborers and traders from India. According to Gervese Clarence-Smith (1989) more than 200,000 slaves were brought to sugarcane plantations from the coasts of East Africa and Madagascar. The population structure of this part of Africa is unique in its blend of genetics, lifestyle and environment which calls for an in-depth investigation drive innovation in the health sector. The demography of Mauritius reflects the origin of its settlers who were mostly Indian, African, Chinese, European (**Table.1**). This has led to an important admixed section of the population. Rodrigues, an island part of the Republic of Mauritius, has a higher proportion of African ancestry. Seychelles consists of 115 islands scattered over 1.3 m sq km of ocean. With a population of just over 96,000. with ethnicities being mostly African, and some European, Chinese and Indian. The Comoros consist of three islands

TABLE 1 | Population diversity.

Country	Males[a]	Females[a]	Population (thousands)[a]	Ethnic diversity (DELF)[b]	Ethnic composition %[c]
Comoros	448	440	888	0.041 (rank 192 on 210)	97 Comorian 1.6 Makua
Madagascar	14,184	14,244	28,427	0.255 (rank 94 on 210)	34.3 Malagasy 24 Merina Malagasy 13.4 Betsimisaraka Malagasy 11.3 Betsileo Malagasy 7.0 Tsimihety Malagasy 5.9 Sakalava Malagasy 1.1 Makua Rest French/Other
Mauritius[d]	628	646	1,273	0.560 (rank 6 on 210)	67 South Asian 27.4 Mixed 3 Chinese 2.6 Other (eg European)
Reunion	436	465	902	NA	25.6 European 23 South Asian 3.4 Chinese 42.6 Mixed
Seychelles	51	48	99	0.070 (rank 184 on 210)	91.4 Seychelles 4.4 South Asian

[a]*United Nations, Department of Economic and Social Affairs, Population Division (2019). World Population Prospects 2019, custom data acquired via website. https://population.un.org/wpp/DataQuery/*

[b]*Kolo P (2012) "Measuring a New Aspect of Ethnicity - The Appropriate Diversity Index," Ibero America Institute for Econ. Research (IAI) Discussion Papers 221, Ibero-America Institute for Economic Research. Definition: The DELF measures the expected dissimilarity between two individuals randomly drawn from each country.*

[c]*britannica.com/place.*

[d]*Including islands of Agalega, Rodrigues and Saint Brandon.*

with a total size of 1,660 sq km lying between the East African coast and Madagascar and has mostly Afro-Arab and Malagasy groups. Madagascar has a population of over 27 m with ethnicities from mostly a mix of Austronesian and African and to a lesser extent European and Indian. The presence of Arabs is noted from the 9th to the 15th century on the North-West and oriental coast of Madagascar (Vérin, 1975). First, the Indonesian migrants reached residents of the Indian Ocean through East of Africa, where they would have been in contact with other migrants including the Bantus (Deschamps, 1965; Boiteau, 1982). Studies have shown that at least three great African migrant waves have contributed to the Malagasy population (Rakotoarivony et al., 2019).

Like in other African countries, non-communicable diseases are increasingly being reported in this region and are the cause of high rates of mortality. Promising avenues through precision medicine can bring a radical change in the health sector with more precise diagnostics tools, better preventive care and improved prognosis. Targeted therapies adapted to particular groups of the populations are more likely to succeed than traditional treatments. The potential benefits of precision public health far outweigh the initial investments for its implementation. One key area is pharmacogenomics, where knowledge of the patient's genetic data allows for accurate drug dosage to reduce side effects and enhance efficacy. African genomes have diverse sets of variants in their pharmacogenes that can impact on drug metabolism and efficacy (Tshabalala et al., 2019). Drug metabolizing enzymes that include the CYP family, determine the amount of drug circulating in the bloodstream. Variations in their genes influence their enzymatic activities and therefore the kinetics of drug catabolism.

Precision medicine has brought new hopes for the management of many diseases such as diabetes, cardiovascular diseases and cancer through new methods for identifying those at risk. Patients in many low-income countries often seek medical assistance when their disease state is well advanced. Recent development has culminated in the discovery of large numbers of genetic variants that contribute to disease risks (Ala-Korpela and Holmes, 2020). These are used to calculate and stratify the risks of developing a particular disease. However, it is primordial to have access to the genetic data for different populations in order to use such methods. Hence the need for large-scale sequencing efforts that would have a high return for health care improvement.

Early screening can significantly improve recovery and prognosis of cancer patients. Molecular profiling of cancer and germline testing are becoming the norms for therapeutic interventions (Birner et al., 2016). Conventional subtyping methods are based on immunochemistry and pathological evidence. They assist in treatment choices and disease management; however when combined with identification of genetic variants through sequencing, optimal treatment options can be provided. Currently many countries offer such services, which have resulted in a dramatic improvement in treatment outcomes Alongside the next generation sequencing (NGS) methods, public resources providing information on molecular profiles and cancer subtypes have amplified the accessible data that clinicians can use to make decisions on diagnosis and treatment. The American College of Medical Genetics and Genomics (ACMG) provides regular updates on guidelines for the use of NGS and interpretation of variants for clinical use (Rehder et al., 2021).

HEALTH AND DISEASE

SWIO island states have reported increasing incidence of non-communicable diseases (NCDs). In Madagascar the health care system is organized into different centers at the district, regional, community and university levels namely: 1) CHRD, Centre Hospitalier de Référence de District; 2) CHRR, Centre Hospitalier de Référence Régionale; 3) CSB, Centre de Santé de Base; 4) CHU, Centre Hospitalier Universitaire. Each center records the incidences of pathologies by age groups. In 2016, the Ministry of Health reported the following cases regarding NCDs. There were 7,553 recorded cases of hypertension for individuals aged 5–24 years and around 172,293 cases for people aged 25 years and over at the CSB level (Ministry of Health, Madagascar). In CHRR/CHU settings, 2077 cases causing 55 deaths in adults were reported. In addition, diabetes is also present, being more prevalent in adults aged 25 and above at the community health centers level with 13,690 cases. At the CHRD level, diabetes is quite common from an early age with 1,621 cases for 15–24 years old and 1756 cases in adults aged 25 and over. Cardiovascular diseases are more frequent as age progresses and at the CSB level, 3,324 cases from the age of 25, amounting to a total of 3,466 cases during the year. At the level of the CHRD, the number of outpatient cases reached 3,804 of which 3,053 cases were in adults aged 25 and over.

Cancers are diagnosed at the CHRD, CHRR, and CHU. In Madagascar, several types of cancer are encountered, the most frequent being breast, cervical, bronchial and skin cancer. At the CHRR and CHU 4,118 known cases of breast cancer were reported with 1,694 deaths. Cervical cancer cases recorded were 109 cases in CHRD settings. For bronchial cancers (primary and secondary), 707 cases are recorded from children in outpatient clinic. Skin cancer is more discreet with only 55 outpatient cases and six inpatient cases at the level of the CHRR and CHU. Given the general conditions in the health services, these figures are likely to be lower than the real figures. Many rural communities are far from health centers and cannot reach the medical services.

Mauritius has a rapidly ageing population with 15% of its residents above the age of 70. This is reflected in its epidemiological distribution of diseases. Noncommunicable diseases (NCDs) account for about 84% of its total disease burden (**Table.2**). Death rate due to NCDs for those aged 30–70, was 411 deaths per 100,000 in 2017. Cardiovascular diseases are the main cause of death (33.2%) followed by diabetes (predominately of type 2) and cancer with 23.5 and 12.8% of total deaths, respectively, in 2016. According to the 2015 National NCD Survey, the prevalence of type 2 diabetes in the Mauritian population aged 20–74 years was 20.5%: 19.6% (Male) 21.3% (Female). The ratio of known diabetes to newly diagnosed diabetes was 2:1. Metabolic control was moderately poor in 33% (HBA1c > 9) of the diabetic population. This population group is

TABLE 2 | Prevalence of diseases.

Country	Population (thousands)	Probability of dying from any of cardiovascular disease (CVD), cancer, diabetes, chronic respiratory disease (CRD) between age 30 and 70 (%)[a]	Cancer cases per 100,000 population[b]	Diabetes Prevalence (% of population ages 20–79)[c]	CVD[d]
Comoros	888	20.6	NA	12.3	8.62%
Madagascar	28,427	26.0	NA	4.5	20%
Mauritius[e]	1,273	23.2	153.2	22.0	33%
Reunion	902	NA	213.9	NA	NA
Seychelles	99	21.1	NA	12.3	34%

[a]*https://apps.who.int/iris/bitstream/handle/10665/342703/9789240027053-eng.pdf*
[b]*https://canceratlas.cancer.org/the-burden/sub-saharan-africa/*
[c]*United Nations, Department of Economic and Social Affairs, Population Division (2019). World Population Prospects 2019, custom data acquired via website. https://population.un.org/wpp/DataQuery/*
[d]*https://www.who.int/nmh/countries/syc_en.pdf?ua=1; https://www.who.int/nmh/countries/mus_en.pdf; https://www.who.int/nmh/countries/mdg_en.pdf?ua=1.*
[e]*Including Agalega, Rodrigues and Saint Brandon.*

at high risk of developing complications of diabetes, with the prevalence of pre-diabetes at 19.4%. The prevalence of hypertension was 28.4%: 27.0% for women and 30.3% for men. Only 52.6% of individuals were currently on medication and 70.6% of those continued to have elevated blood pressure (i.e., above 140/90 mmHg). There is a high prevalence level of modifiable risk factors, including overweight/obesity (54.3%); alcohol consumption (52.8%) and tobacco consumption (19.3%) and low prevalence of physical activity (23.7%).

According to the National Cancer Registry report2018[1], there were 2,380 new cancers in 2018: 959 males and 1,421 females. There has been a slight decrease of 3.3% from 2017. Among males, prostate cancer (n = 191, 19.9%) ranks first ahead of colorectal (n = 124, 12.9%) and lung (n = 65, 6.8%) cancers. The number of female breast cancer cases (n = 570, 40.1%) is far ahead followed by colorectal (n = 104, 7.3%), ovarian (n = 99, 7.0%) and uterus (n = 87, 6.1%) cancer. Prostate cancer (n = 99, 15.1%) remains the leading cause of cancer deaths in males, followed by lung (n = 94, 14.4%) and colorectal cancer (n = 80, 12.2%) cancers. Breast cancer (n = 173, 25.6%) is still the main cause of cancer deaths in females followed by cervix (n = 63, 9.3%) and colorectal cancer (n = 58, 8.6%).

PRECISION MEDICINE SCOPE AND PROMISES

Precision medicine (PM) requires an understanding of the interactions of genetic variability, lifestyle and environment, and the individual's health (Ramsay, 2018). It enables timely prevention and assists health professionals to make more informed therapy decisions, thus minimizing side effects and maximizing treatment efficacy. To implement precision medicine, Desmond-Hellmann (2016) emphasized the need for more accurate detection, identification, and tracking of unique traits in subpopulations since individuals or populations are unique (Fatumo, 2020; Pereira et al., 2021). However, genetic factors associated with NCDs have been unraveled by studies targeting mostly European populations.

The last 10 years have seen significant progress in elucidating African genomes and exploring their genetic diversity and disease susceptibility (Chikowore et al., 2021). Efforts from various consortia such as H3Africa (Human Heredity Health in Africa), MADCaP (Men of African Descent Carcinoma of the Prostate) and others have fueled progress in understanding the genetic determinants of complex traits via GWAS studies in African populations. Gurdasani et al. (2019) obtained and analyzed whole-genome data from 14,126 individuals across Africa and observed statistical differences in heritability for traits between African and European populations. Their study revealed novel genetic variants associated with several traits/diseases.

Similar studies carried out by H3Africa provide a robust research and health platform for data analysts and health care professionals to generate and analyse reference datasets from control and disease cohorts (Mulder, 2017). However, participant recruitment in Africa remains a big challenge for researchers in precision medicine (Adebamowo et al., 2018) mainly due to fears and mistrust. Although progress has been made, pharmacogenomics in African populations is still an underrepresented area of research in precision medicine (Mulder, 2017). For these reasons, the African Academy of Sciences has issued a policy paper to promote the implementation of genomic medicine for public health in Africa (African Academy of Sciences, 2021).

[1]https://health.govmu.org/Documents/Statistics/Documents/National%20Cancer%20Registry%20%20-%202018.pdf

TABLE 3 | Research institutions for genomics.

Country	Local institutions that partner with international institutions to conduct genomics research	Genome sequencing of local individuals
Comoros	Laboratoire de Biologie de l'Hôpital El-Maarouf, Moroni	Yes; https://dx.doi.org/10.1038%2Fejhg.2010.128-Y and M chromosomes
Madagascar	Université d'Antananarivo, Pasteur Institute of Madagascar	Yes; https://doi.org/10.1073/pnas.1704906114.
Mauritius	Ministry of Health and Wellness University of Mauritius	NA
Reunion	CIRAD, Université de La Reunion	NA
Seychelles	NA	NA

Bentley et al. (2020) summarized the achievements of recent genomics projects in Africa. For examples, new loci, new variants within known loci, and new pharmacogenomic loci all relevant for individuals with African ancestry were discovered. However, although progress has been made, there are still many challenges the most important being the diversity of African populations and the reduced linkage disequilibrium across the genome (Choudhury et al., 2020, 2018; Fan et al. 2019). For this reason, studies of African populations require larger sample sizes in order to properly interrogate the variation present.

Though NCDs represent 84% of disease burden in Mauritius, the only actions that have been taken so far are mainly prevention through lifestyle changes. Very few studies have been done with regards to genetic screening locally. The most prevalent cancers in Mauritius are breast, prostate, colorectal and ovarian[2]. Cancer result partly from genetic components and genomics is imperative for screening, investigation, risk stratification and targeted therapeutics. The only study published in Mauritius on breast cancer genetics showed a mutation in the BRCA2 6503delTT in two sisters of the same family of Indian origin (Khittoo et al., 2001). The fact that 25% of the breast cancer identified are in women under the age of 50 points towards a strong genetic predisposition. Genomics provides crucial information towards risk-stratification of patients which has an important bearing on therapeutics. Currently, molecular profiling using limited arrays such as MammaPrint are being used to determine whether early-stage breast cancer patients need to go on chemotherapy. In Mauritius, breast cancer is typed only for estrogen receptor, progesterone receptor and HER2. Screening for BRCA1 and BRCA2 is not available in the public sector despite concomitant high incidence of ovarian cancer. One private health care center has recently advertised for cancer screening through genetic testing.

Diabetes has both genetic and environmental predisposing risk factors. Most cases of diabetes are polygenic. Genome wide association studies have identified a number of susceptibility loci for type 2 diabetes. These genetic variants have been used to develop genetic risk scores as a tool for prediction of type 2 diabetes. Monogenic forms of the disease have also been identified. These include HNF1A mutation for maturity onset diabetes of the young (MODY) and KCNJ11 mutations for neonatal diabetes. Both these conditions are highly responsive to sulfonylurea therapy. Though the incidence of MODY is only about 2%, it is critical to identify those patients in order to achieve a good glycaemic control. No genome wide study or any screening for MODY has been carried out in Mauritius so far even though 45% of the population being either diabetic or prediabetic.

Few studies on genomes from Madagascar have been carried out to investigate the population structure. Genetic exploration of the 80 or so groups from across the island have been genotyped for >2.6 m markers on an array chip (Pierron et al., 2017). Other samples for 21 Mikea (hunter-gatherers), 24 Vezo (seminomadic fishermen), and 24 Temoro (farmers) have also been studied through genotyping (Pierron et al., 2014). The African ancestry was shown to be about 67% and the main gene flow from Africa was mainland Bantu speakers. The Austronesian gene flow is about 30%. The source of the Malagasy DNA was found to originate from southern Sulawesi, the Lesser Sunda islands and eastern Borneo (Kusuma et al., 2015). The sequences of mitochondrial DNA from 257 villages across the island together with paternal Y chromosome genotypes and genome-wide SNP typing information have confirmed the Bantu and Indonesian contributions to the Malagasy genomes; while smaller components originate from European and Arab ancestries (Pierron et al., 2017).

CONCLUSION AND RECOMMENDATIONS

There is a dearth of genetic data from this part of Africa, where medical research has been mostly neglected (Rotimi, 2017; Tucci and Akey, 2019). Much of the work in Madagascar has focused on determining ancestry while none seem to have investigated disease related variants. Genomics technologies are in use but mainly for pathogens genomes at the Pasteur Institute of Madagascar and in a few laboratories in Mauritius (**Table 3**). This gap in scientific knowledge should be urgently addressed if the SWIO region is to benefit from the rapidly developing tools for genomic medicine. Given the complexities and costs of genomics research, one possible approach is to launch a regional initiative that would address the needs of all participating countries through a concerted effort of existing institutions (**Table 3**). Common training and infrastructure (eg

[2]https://gco.iarc.fr/today/data/factsheets/populations/480-mauritius-fact-sheets.pdf

biorepository and computing resources) can be shared. There is hope that the countries can leverage on the available resources and skills in some parts of Africa such as South Africa and West Africa where there has been considerable investment in genomics and bioinformatics through local and international efforts. The current pandemic has brought to light the unpreparedness of the scientific communities and the government bodies to address a serious public health issue. NGS platforms have made sequencing more affordable and accessible to even remote parts of Africa. Governments have to push this forward to bring the benefits of genomic medicine to the communities. It is often argued that PM is for rich countries while others will not afford it. There is however a clear economic gain in implementing PM within an electronic health care system that will allow better tracking of individuals for disease monitoring. The return on investment will be the immediate benefits to the community of innovative diagnostics and targeted treatment. Judicious use of drugs, based on pharmacogenomic evidence is far more effective and less costly than applying dosages recommended for European cohorts. There is a soaring demand for molecular profiling among families of cancer patients in Mauritius and elsewhere; most are turning to other countries to access the services. This does not guarantee return of the sequence data to the country. The scientific communities in SWIO have high hopes that molecular profiling through genomics will become the norm for health improvement. For this to happen, governments will need evidence from genetic data to assess the benefits of implementing genomics in health. Ethical issues and community engagement form an integral part of the process.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

YJ-F and AG compiled the manuscript; GC and SG provided contents on diseases for Mauritius; RR and CT provided input for Madagascar.

ACKNOWLEDGMENTS

Authors wish to acknowledge the University of Mauritius, Universities of Antananarivo and Toamasina for support.

REFERENCES

Adebamowo, S. N., Francis, V., Tambo, E., Diallo, S. H., Landouré, G., Nembaware, V., et al. (2018). Implementation of Genomics Research in Africa: Challenges and Recommendations. *Glob. Health Action.* 11 (1), 1419033. doi:10.1080/16549716.2017.1419033

African Academy of Sciences (2021). *A Framework for the Implementation of Genomic Medicine for Public Health in Africa*. ASP Policy Paper 1. https://www.aasciences.africa/publications/policy-paper-framework-implementation-genomic-medicine-public-health-africa.

Ala-Korpela, M., and Holmes, M. V. (2020). Polygenic Risk Scores and the Prediction of Common Diseases. *Int. J. Epidemiol.* 49 (1), 1–3. doi:10.1093/ije/dyz254

Bentley, A. R., Callier, S. L., and Rotimi, C. N. (2020). Evaluating the Promise of Inclusion of African Ancestry Populations in Genomics. *Npj Genom. Med.* 5, 5. doi:10.1038/s41525-019-0111-x

Birner, P., Prager, G., and Streubel, B. (2016). Molecular Pathology of Cancer: How to Communicate with Disease. *ESMO Open* 1 (5), e000085. doi:10.1136/esmoopen-2016-000085

Boiteau, P. (1982). *Contribution à l'histoire de la nation malgache*. Antananarivo: Sociales, 445.

Chikowore, T., Kamiza, A. B., Oduaran, O. H., Machipisa, T., and Fatumo, S. (2021). Non-Communicable Diseases Pandemic and Precision Medicine: Is Africa Ready? *EBioMedicine* 65, 103260. doi:10.1016/j.ebiom.2021.103260

Choudhury, A., Aron, S., Botigué, L. R., Sengupta, D., Botha, G., Bensellak, T., et al. (2020). High-Depth African Genomes Inform Human Migration and Health. *Nature* 586 (7831), 741–748. doi:10.1038/s41586-020-2859-7

Choudhury, A., Aron, S., Sengupta, D., Hazelhurst, S., and Ramsay, M. (2018). African Genetic Diversity Provides Novel Insights into Evolutionary History and Local Adaptations. *Hum. Mol. Genet.* 27 (R2), R209–R218. doi:10.1093/hmg/ddy161

Deschamps, H. (1965). *Histoire de Madagascar*. Paris: Berger-Levrault, 343. p + Illustrations.

Desmond-Hellmann, S. (2016). Progress Lies in Precision. *Science* 353, 6301731. doi:10.1126/science.aai7598

Fan, S., Kelly, D. E., Beltrame, M. H., Hansen, M. E. B., Mallick, S., Ranciaro, A., et al. (2019). African Evolutionary History Inferred from Whole Genome Sequence Data of 44 Indigenous African Populations. *Genome Biol.* 20 (1), 82. doi:10.1186/s13059-019-1679-2

Fatumo, S. (2020). The Opportunity in African Genome Resource for Precision Medicine. *EBioMedicine* 54, 102721. doi:10.1016/j.ebiom.2020.102721

Gervese Clarence-Smith, W. (1989). *The Economics of the Indian Ocean Slave Trade in the Nineteenth Century*. London: Frank Cass.

Gurdasani, D., Carstensen, T., Fatumo, S., Chen, G., Franklin, C. S., Prado-Martinez, J., et al. (2019). Uganda Genome Resource Enables Insights into Population History and Genomic Discovery in Africa. *Cell* 179 (4), 984–1002. e36. doi:10.1016/j.cell.2019.10.004

Khittoo, G., Manning, A., Mustun, H., Appadoo, J., Venkatasamy, S., Fagoonee, I., et al. (2001). Mutation Analysis of a Mauritian Hereditary Breast Cancer Family Reveals the BRCA2 6503delTT Mutation Previously Found to Recur in Different Ethnic Populations. *Hum. Hered.* 52 (1), 55–58. doi:10.1159/000053354

Kusuma, P., Cox, M. P., Pierron, D., Razafindrazaka, H., Brucato, N., Tonasso, L., et al. (2015). Mitochondrial DNA and the Y Chromosome Suggest the Settlement of Madagascar by Indonesian Sea Nomad Populations. *BMC Genomics* 16, 191. doi:10.1186/s12864-015-1394-7

Mulder, N. (2017). Development to Enable Precision Medicine in Africa. *Personalized Med.* 14 (6), 467–470. doi:10.2217/pme-2017-0055

Pereira, L., Mutesa, L., Tindana, P., and Ramsay, M. (2021). African Genetic Diversity and Adaptation Inform a Precision Medicine Agenda. *Nat. Rev. Genet.* 22 (5), 284–306. doi:10.1038/s41576-020-00306-8

Pierron, D., Heiske, M., Razafindrazaka, H., Rakoto, I., Rabetokotany, N., Ravololomanga, B., et al. (2017). Genomic Landscape of Human Diversity across Madagascar. *Proc. Natl. Acad. Sci. USA* 114 (32), E6498–E6506. doi:10.1073/pnas.1704906114

Pierron, D., Razafindrazaka, H., Pagani, L., Ricaut, F.-X., Antao, T., Capredon, M., et al. (2014). Genome-Wide Evidence of Austronesian-Bantu Admixture and Cultural Reversion in a Hunter-Gatherer Group of Madagascar. *Proc. Natl. Acad. Sci.* 111, 936–941. doi:10.1073/pnas.1321860111

Rakotoarivony, R., Spiral, G. J., Ravelonjanahary, S., Ranaivoarisoa, J. F., Hodgson, J., Racimo, F., et al. (2019). *Genomic Investigation of the Ancestry Proportion of Six Malagasy Ethnic Groups*. Antananarivo: Recherche pour le Développement. Université d'Antananarivo Madagascar, CIDST d'Antananarivo Madagascar. ISSN: 1022-8691.

Ramsay, M. (2018). Precision Medicine for Africa: Challenges and Opportunities. *Quest* 14, 3, 2018 . Available at: www.assaf.org.xa.

Rehder, C., Bean, L. J. H., Bick, D., Chao, E., Chung, W., Das, S., et al. (2021). Next-Generation Sequencing for Constitutional Variants in the Clinical Laboratory, 2021 Revision: A Technical Standard of the American College of Medical Genetics and Genomics (ACMG). *Genet. Med.* 23 (8), 1399–1415. doi:10.1038/s41436-021-01139-4

Rotimi, C. N., Bentley, A. R., Doumatey, A. P., Chen, G., Shriner, D., and Adeyemo, A. (2017). The Genomic Landscape of African Populations in Health and Disease. *Hum. Mol. Genet.* 26 (R2), R225–R236. doi:10.1093/hmg/ddx253

Tshabalala, S., Choudhury, A., Beeton-Kempen, N., Martinson, N., Ramsay, M., and Mancama, D. (2019). Targeted Ultra-Deep Sequencing of a South African Bantu-Speaking Cohort to Comprehensively Map and Characterize Common and Novel Variants in 65 Pharmacologically-Related Genes. *Pharmacogenet Genomics* 29 (7), 167–178. doi:10.1097/fpc.0000000000000380

Tucci, S., and Akey, J. M. (2019). The Long Walk to African Genomics. *Genome Biol.* 20, 130. doi:10.1186/s13059-019-1740-1

Vérin, P. (1975). *Les échelles anciennes du commerce sur les côtes Nord de Madagascar Tome II*. Lille: Service de reproduction des Thèses, 519.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Genetics

ORIGINAL RESEARCH
published: 04 March 2022
doi: 10.3389/fgene.2022.711944

High Resolution HLA ~A, ~B, ~C, ~DRB1, ~DQA1, and ~DQB1 Diversity in South African Populations

Mqondisi Tshabalala[1,2], *Juanita Mellet*[1,2], *Kuben Vather*[3], *Derrick Nelson*[3], *Fathima Mohamed*[3], *Alan Christoffels*[4] *and Michael S. Pepper*[1,2]*

[1]*Department of Immunology, Institute for Cellular and Molecular Medicine, Faculty of Health Sciences, University of Pretoria, Pretoria, South Africa,* [2]*South African Medical Research Council (SAMRC) Extramural Unit for Stem Cell Research and Therapy, Faculty of Health Sciences, University of Pretoria, Pretoria, South Africa,* [3]*South African National Blood Service (SANBS), Roodepoort, South Africa,* [4]*SAMRC Bioinformatics Unit, South African National Bioinformatics Institute, University of the Western Cape, Bellville, South Africa*

Background: Lack of HLA data in southern African populations hampers disease association studies and our understanding of genetic diversity in these populations. We aimed to determine HLA diversity in South African populations using high resolution HLA ~A, ~B, ~C, ~DRB1, ~DQA1 and ~DQB1 data, from 3005 previously typed individuals.

Methods: We determined allele and haplotype frequencies, deviations from Hardy-Weinberg equilibrium (HWE), linkage disequilibrium (LD) and neutrality test. South African HLA class I data was additionally compared to other global populations using non-metrical multidimensional scaling (NMDS), genetic distances and principal component analysis (PCA).

Results: All loci strongly ($p < 0.0001$) deviated from HWE, coupled with excessive heterozygosity in most loci. Two of the three most frequent alleles, HLA ~DQA1*05:02 (0.2584) and HLA ~C*17:01 (0.1488) were previously reported in South African populations at lower frequencies. NMDS showed genetic distinctness of South African populations. Phylogenetic analysis and PCA clustered our current dataset with previous South African studies. Additionally, South Africans seem to be related to other sub-Saharan populations using HLA class I allele frequencies.

Discussion and Conclusion: Despite the retrospective nature of the study, data missingness, the imbalance of sample sizes for each locus and haplotype pairs, and induced methodological difficulties, this study provides a unique and large HLA dataset of South Africans, which might be a useful resource to support anthropological studies, disease association studies, population based vaccine development and donor recruitment programs. We additionally provide simulated high resolution HLA class I data to augment the mixed resolution typing results generated from this study.

Keywords: high resolution typing, HLA diversity, South Africa, haplotype frequencies, allele frequencies, human leukocyte antigen (HLA)

OPEN ACCESS

Edited by:
Maritha J. Kotze,
Stellenbosch University, South Africa

Reviewed by:
Milena Camargo,
Colombian Institute of Immunology Foundation, Colombia
José Nunes,
Université de Genève, Switzerland

***Correspondence:**
Michael S. Pepper
michael.pepper@up.ac.za

Specialty section:
This article was submitted to Human and Medical Genomics, a section of the journal Frontiers in Genetics

Received: *19 May 2021*
Accepted: *17 January 2022*
Published: *04 March 2022*

Citation:
Tshabalala M, Mellet J, Vather K, Nelson D, Mohamed F, Christoffels A and Pepper MS (2022) High Resolution HLA ~A, ~B, ~C, ~DRB1, ~DQA1, and ~DQB1 Diversity in South African Populations. Front. Genet. 13:711944. doi: 10.3389/fgene.2022.711944

Abbreviations: AFND, allele frequencies net database; EM, expectation maximisation; FST, population differentiation; HLA, human leukocyte antigen; HWE, hardy-weinberg equilibrium; NHLS, national health laboratory services; NJ, neighbour-joining; NMDS, non-metrical multidimensional scaling; PCA, principal component analysis; SABMR, south african bone marrow registry; SANBS, south african national blood transfusion services.

INTRODUCTION

The human leukocyte antigen (HLA) gene region is considered to be one of the most polymorphic regions in the human genome (Mungall et al., 2003; Wong et al., 2013). Currently, there are 30 862 HLA alleles listed (https://www.ebi.ac.uk/ipd/imgt/hla/stats.html) in the IMGT/HLA database (3.45.0 release of July 2021). HLA genes encode proteins involved in antigen presentation (Robinson et al., 2013), and play a key determining role in transplantation clinical outcomes (Beatty et al., 2000; Carrington and O'Brien, 2003; Ndung'u et al., 2005; Brander et al., 2006; Ovsyannikova and Poland, 2011; Chen et al., 2012; Ramsay, 2012; Garamszegi, 2014). Despite the growing documented evidence of genetic diversity in Africans (Chen et al., 1995; Zietkiewicz et al., 1997; Jorde et al., 2000; Prugnolle et al., 2005; Disotell, 2012), there remains an information gap on HLA diversity in these populations (reviewed in Tshabalala et al. (2015)). This lack of HLA data hampers disease association studies (reviewed in Dyer et al. (2013), population-specific vaccine development (Gourraud et al., 2015) and programs aimed at donor recruitment into registries (Edinur et al., 2016). Additionally, there is a high disease burden in these populations (WHO, 2013). Understanding HLA diversity will compliment efforts to eliminate these health challenges.

In addition to its key role in the human immune system, HLA has been used to understand human genetic diversity, population genetics and anthropology. HLA has been widely used to understand genetic relatedness of different populations as well as demographic events in those populations (Sanchez-Mazas and Meyer, 2014). The HLA genetic makeup of populations provides insight into their histories including selective pressures by pathogens (Prugnolle et al., 2005), migration, admixture and changes in population size (Parham and Ohta, 1996; Kijak et al., 2009; Buhler and Sanchez-Mazas, 2011; Sanchez-Mazas et al., 2011). The availability of population HLA data is thus critical to understanding peopling history and general evolution of the human immune system (Burrell and Disotell, 2009; Meyer et al., 2018).

The South African population comprises 59.6 million people (Statistics South Africa, 2022). The presence of a high disease burden in the population is one factor which may drive high genetic diversity, including HLA diversity. Additionally, a large proportion of the population harbors residual DNA sequences from *Homo naledi* (Berger et al., 2015), one of the oldest Hominid ancestors, allowing accumulation of polymorphisms over thousands of years. Additionally, new and low frequency HLA alleles have been reported in South African populations (Paximadis et al., 2012; Hayhurst et al., 2015) supporting the idea of high genetic diversity in these populations (May et al., 2013; Choudhury et al., 2017). We previously described allele and haplotype frequencies from the South African Bone Marrow Registry (SABMR) (Tshabalala et al., 2018) in an effort to understand HLA diversity in South Africans. The current study is aimed at improving our understanding of HLA diversity in South Africans using retrospectively typed individuals in the National Health Laboratory Services (NHLS) and the South African National Blood Services (SANBS). We additionally sought to compare HLA data from South Africans with other global populations using population genetics approaches.

MATERIALS AND METHODS

Study Population, Human Leukocyte Antigen Data Access and Ethics

Approval for this study was granted by the Research Ethics Committee of the University of Pretoria, Faculty of Health Sciences (approval no. 220/2015), the SANBS Human Research Ethics Committee (SANBS HREC) and NHLS Academic Affairs and Research. We analyzed a combined total of 3005 high resolution (four digit typing HLA ~A, HLA ~B, HLA ~C, HLA ~DRB1, HLA ~DQA1 and HLA ~DQB1) results from the SANBS and the NHLS. The retrospective high resolution typing dataset (defined in the context of this study as four digit typing resolution) has been assembled from higher resolution DNA based methods in the SANBS and the NHLS. All available HLA data from the SANBS (up to 20 November 2016) plus the NHLS data (05 June 2003 to 12 April 2016) was accessed. The NHLS offers national diagnostic pathology services (http://www.nhls.ac.za/) whilst the SANBS aims to supply safe blood and blood products to the local population (https://sanbs.org.za/). Only HLA data was accessed; no additional data was accessed due to ethical considerations. Participants' personal identifiers were not accessed to maintain confidentiality following the Helsinki ethical guidelines (Association, 2013). All the accessed HLA data was checked for allele validity, and all pre-2010 nomenclature designations were converted using current nomenclature conversion tables and conversion tools provided by IMGT/HLA (https://www.ebi.ac.uk) based on the IMGT/HLA database (3.45.0 release of July 2021) (https://www.ebi.ac.uk/ipd/imgt/hla/stats.html). HLA data missingness in our dataset is defined as the lack of typing methods to call two alleles at a given locus, resulting in one allele for that individual at that particular locus. Unfortunately, a distinction between homozygous typing and data missingness could not be established due to the retrospective nature of the study.

Statistical Analysis

High (four digit) resolution data was analyzed to estimate linkage disequilibrium (LD), Hardy-Weinberg equilibrium (HWE) proportions, homozygosity test of neutrality, and allele and haplotype frequencies. Allele and haplotype frequencies were estimated by resolving phase and allelic ambiguities using the expectation-maximization (EM) algorithm (Excoffier and Slatkin, 1995; Eberhard et al., 2013) both implemented in Python for population genomics (PyPop) version 0.7.0 (Lancaster et al., 2007) and gene [RATE] tools (https://hla-net.eu/tools/basic-statistics/) (Nunes, 2016). Excoffier and Slatkin (1995) allows estimation of random haplotypes based on sample allele frequencies. For pairwise LD, we used Hedrick's D' (Hedrick, 1987) and Cramer's V Statistic (W_n) (Cramér, 1946), all implemented in PyPop version 0.7.0 (Lancaster et al.

TABLE 1 | HWE parameters for high resolution typing. Exact Test using Markov chain for all loci with 100 000 dememorization steps.

Locus	#Genotypes	Obs Het	Exp Het	p-HWE
HLA ~A	111	0.07207	0.96714	<0.0001**
HLA ~B	345	0.27536	0.95592	<0.0001**
HLA ~C	128	0.03906	0.86489	<0.0001**
HLA ~DRB1	1927	0.10223	0.94003	0.0015*
HLA ~DQA1	104	0.12500	0.71363	<0.0001**
HLA ~DQB1	325	0.55077	0.93905	<0.0001**

*#Genotypes, number of genotypes; Obs Het, observed heterozygosity; Exp Het, expected heterozygosity; p-HWE, p-value for HWE deviation; ** highly significant (p < 0.0001); * significant (p < 0.01) difference between observed and expected heterozygosity.*

, 2007). HLA genotypes were converted to Arlequin version 3.5.2 (Lancaster et al., 2007) input files using CREATE version 1.37 software (Coombs et al., 2008) to assess deviations from HWE [modified hidden Markov chain (Guo and Thompson, 1992) with 100 000 dememorization steps]. Slatkin's implementation of Ewens-Watterson homozygosity test of neutrality (Slatkin, 1994; 1996) was done in PyPop version 0.7.0 (Lancaster et al., 2007). In addition to allele frequencies, cumulative allele frequencies from the South African population were plotted for high resolution typing data sets.

Population Comparison

To better understand HLA diversity in our dataset, we compared our findings to other global populations. Our data was compared with multiple population datasets from gene [RATE] tools (Nunes, 2016) defined world regions by non-metrical multidimensional scaling (NMDS) analysis. Due to the HLA mixed resolution typing nature and data missingness in our dataset, we performed HLA ~A, ~B and ~C (HLA class I) completion of our data set to get high resolution (four digit typing) data using the PhyloD tool as previously described (Listgarten et al., 2008). The PhyloD HLA completion tool uses statistical *in silico* methods to probabilistically predict four digit HLA class I alleles (Listgarten et al., 2008). We further compared our class I HLA allele frequency data with PhyloD generated allele frequency data (Listgarten et al., 2008), and 28 other publicly available HLA ~A, ~B and ~C allele frequencies (four digit resolution) as well as sub-Saharan African data from the Allele Frequencies Net Database (AFND) (González-Galarza et al., 2015) including previous South African studies (Loubser et al., 2017a; b;Grifoni et al., 2018; Tshabalala et al., 2018). Specifically, our HLA data (RSA) was compared with the following AFND defined populations (population codes we used for phylogenetic analysis): Burkina Faso Fulani (BFF) (Modiano et al., 2001), Burkina Faso Mossi (BFM) (Modiano et al., 2001), Burkina Faso Rimaibe (BFR) (Modiano et al., 2001), Cameroon Baka Pygmy (CBP) (Torimiro et al., 2006), Cameroon Bakola Pygmy (CBkP) (Bruges Armas et al., 2003), Cameroon Bamileke (CaB) (Torimiro et al., 2006), Cameroon Beti (CBt) (Torimiro et al., 2006), Cameroon Sawa (CSw) (Torimiro et al., 2006), Central African Republic Mbenzele Pygmy (CARMP) (Bruges Armas et al., 2003), Ghana Ga-Adangbe (GGA) (Norman et al., 2013), Kenya (KEN) (Luo et al., 2002), Kenya Luo (KENL) (Cao et al., 2004), Kenya Nandi (KENN) (Cao et al., 2004), Kenya, Nyanza Province, Luo tribe (KENNy) (Arlehamn et al., 2017), PhyloD generated data (PSA) (Listgarten et al., 2008), Rwanda (RWA) (Tang et al., 2000), Senegal Niokholo Mandenka (SenMAND) (Sanchez-Mazas et al., 2000), South Africa Black (SoAB) (Paximadis et al., 2012), South Africa Caucasians (SoAC) (Paximadis et al., 2012), South Africa Natal Tamil (SANT) (Hammond and Anley, 2006), South Africa Natal Zulu (SANZ) (Hammond et al., 2006), South Africa Worcester (WOR) (Grifoni et al., 2018), South African Bone Marrow Registry (SAB) (Tshabalala et al., 2018), South African Indian population (SAI) (Loubser et al., 2017a), South African Mixed ancestry (RMX) (Loubser et al., 2017b), Uganda Kampala (UgaKam) (Cao et al., 2004), Uganda Kampala pop 2 (UgaKam2) (Kijak et al., 2009), Zambia Lusaka (ZaL) (Cao et al., 2004) and Zimbabwe Harare Shona (ZiHS) (Louie et al., 2006). HLA class I allele frequencies from the above 30 populations were used to compute pairwise population differentiation (F_{ST}) and Nei's genetic distances (Nei, 1972) in POPTREE software (Takezaki et al., 2010; 2014). An unrooted tree was constructed based on the Neighbour-Joining (NJ) method (Saitou and Nei, 1987) implemented in POPTREE software (Takezaki et al., 2010; 2014) using Nei's genetic distances. The pairwise F_{ST} matrix was used for PCA in ClustVis (a web tool for visualizing clustering of multivariate data using PCA and heat map) (Metsalu and Vilo, 2015). Additionally, the South African HLA ~A, ~B and ~C

TABLE 2 | Slatkin's implementation of Ewens-Watterson homozygosity test of neutrality (Slatkin, 1994;1996). Observed homozygosity (homozygosity F statistic ~ a sum of squared allele frequencies) compared to expected homozygosity (simulated under neutrality/equilibrium expectations for the same sample taking into account unique alleles).

Locus	Observed F	Expected F	Variance in F	Fnd	Fp
HLA ~A	0.0362	0.0657	0.0003	−1.7622	<0.0001**
HLA ~B	0.0461	0.0367	0.0001	1.2062	0.8965
HLA ~C	0.1385	0.1496	0.0026	−0.2165	0.5070
HLA ~DRB1	0.0602	0.0446	0.0001	1.3792	0.9163
HLA ~DQA1	0.2898	0.4738	0.0262	−1.1368	0.0960
HLA ~DQB1	0.0626	0.1091	0.0013	−1.3042	0.0133*

*Observed F (observed homozygosity F statistic); Expected F (expected homozygosity F statistic); Fnd (Normalised deviate of F statistic); Fp (p-value F statistic); ** highly statistically significant (p < 0.0001); * significant (p < 0.05).*

TABLE 3 | Top 20 HLA alleles by locus and typing resolution (Full list in **Supplementary Table S1**).

Four-digit	
Locus	**Frequency**
DQA1*05:02	0.258
DQA1*04:02	0.194
C*17:01	0.149
DQA1*02:01	0.145
A*43:01	0.107
C*16:01	0.087
DQB1*03:19	0.087
DRB1*15:03	0.081
DRB1*15:01	0.071
DQB1*03:01	0.068
B*15:10	0.067
DQB1*05:03	0.063
C*16:02	0.063
C*03:04	0.062
DQB1*06:03	0.061
B*15:01	0.060
DQB1*05:02	0.058
C*14:02	0.051
DQB1*06:02	0.049
DQB1*02:01	0.047

cumulative allele frequencies (four digit resolution) generated in this study were compared to Kenyan, Ugandan and Zambian cumulative frequencies from the AFND (González-Galarza et al., 2015). All HLA alleles were sorted in descending order according to their frequencies, and cumulative frequencies were plotted according to the total number of alleles at a particular locus.

RESULTS

HWE Proportions and Neutrality Test

All loci showed a strong significant deviation from the expected HWE proportions ($p < 0.0001$) as detailed in **Table 1**. The Ewens-Watterson neutrality test showed negative and significant F*nd* values for HLA ~A ($p < 0.0001$) and ~DQB1 ($p = 0.0133$) (**Table 2**). This indicates homozygosity which is suggestive of balancing selection at these loci (**Table 2**). Homozygosity ($p > 0.05$) was detected in HLA ~B, ~C, ~DRB1 and ~DQA1 (**Table 2**).

Allele Frequencies

The full list of alleles is detailed in **Supplementary Table S1** which includes all typing frequencies from the South African population. The top 20 most frequent alleles across the different loci are summarized in **Table 3**. HLA ~DQA1*05:02 (0.258), ~DQA1*04:02 (0.194) and ~ C*17:01 (0.149) were the three most common alleles in our dataset (**Table 3**). From the 3005 individuals in our data set, complete HLA data for each locus were as follows: HLA ~A (111), HLA ~B (345), HLA ~C (128), HLA ~DRB1 (1927), HLA ~DQA1 (104) and HLA ~DQB1 (325). There was profound data missingness which we attempted to address in our quest to highlight HLA diversity from South African populations. **Figure 1** summarizes the cumulative allele frequencies from the South African populations described in this study. We additionally include PhyloD generated (Listgarten et al., 2008) HLA ~A, ~B and ~C estimated genotypes (with probabilities) and allele frequencies in **Supplementary Table S2** for population comparison and as a future resource for other researchers.

Haplotype Frequencies and Linkage Disequilibrium

The most common estimated two, three and four loci haplotypes were A*02:05~C*14:02 (0.500), A*30:02~B*45:01~DRB1*15:03 (1.00) and A*30:02~B*45:01~DRB1*15:03~DQB1*05:01 (0.500), respectively, as summarized in **Table 4** and **Supplementary Table S3**. PyPop version 0.7.0 (Lancaster et al., 2007) could not estimate any five and six loci haplotypes at high resolution (**Supplementary Table S3**) due to lack of data after filtering. Pairwise LD measured by Hedrick's

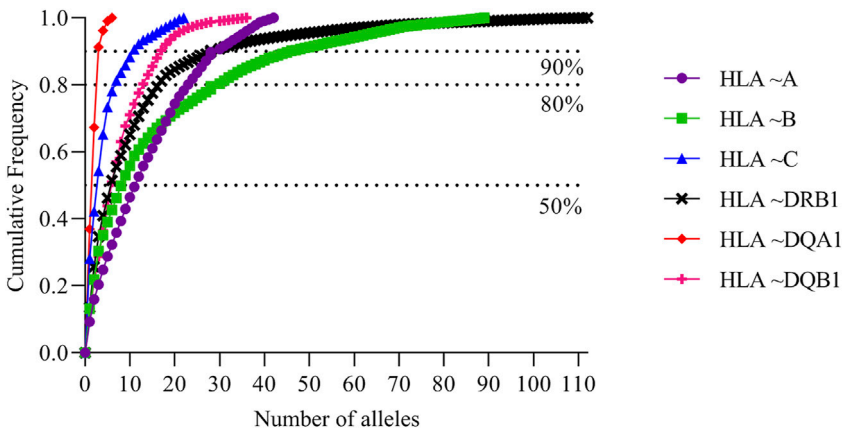


FIGURE 1 | South African cumulative allele frequencies. Cumulative allele frequencies indicating population coverage of South African HLA ~A, ~B, ~C, ~DRB1 and ~DQB1 alleles. HLA alleles were sorted according to their allele frequencies in descending order; cumulative frequencies were plotted according to the number of alleles.

TABLE 4 | The twenty most frequent two, three, four, five and six loci haplotype frequencies (Full list in **Supplementary Table S3**). No data was available after filtering to compute five and six loci haplotype frequencies in Pypop (Lancaster et al., 2007). Only 18 four~loci haplotypes were identified.

Two loci	HF	Three loci	HF	Four loci	HF
A*02:05~C*14:02	0.500	A*30:02~B*45:01~DRB1*15:03	1.000	A*30:02~B*45:01~DRB1*15:03~DQB1*05:01	0.500
A*29:02~C*17:01	0.500	DRB1*11:02~DQA1*05:02~DQB1*03:19	1.000	A*30:02~B*45:01~DRB1*15:03~DQB1*06:02	0.500
C*17:01~DQA1*04:02	0.579	C*17:01~DRB1*11:02~DQB1*03:19	0.667	B*42:01~C*17:01~DRB1*15:03~DQA1*04:02	0.571
B*42:01~DQA1*04:02	0.556	B*42:01~C*17:01~DQA1*04:02	0.657	B*42:02~C*17:01~DRB1*11:02~DQB1*03:19	0.333
A*23:01~DQA1*02:01	0.500	A*23:01~DQA1*02:01~DQB1*02:01	0.500	B*15:10~C*17:01~DRB1*11:02~DQB1*03:19	0.167
A*80:01~DQA1*02:01	0.500	A*80:01~DQA1*02:01~DQB1*02:01	0.500	B*52:02~C*03:04~DRB1*11:02~DQB1*03:19	0.167
B*42:01~C*17:01	0.406	B*42:01~DRB1*15:03~DQA1*04:02	0.444	B*41:02~C*17:01~DRB1*11:02~DQB1*03:19	0.167
C*17:01~DQB1*03:19	0.313	C*17:01~DRB1*15:03~DQA1*04:02	0.444	B*41:02~C*17:01~DRB1*15:03~DQB1*03:19	0.167
C*17:01~DQB1*04:01	0.313	A*30:02~B*45:01~DQB1*05:01	0.400	B*42:01~C*17:01~DRB1*11:02~DQA1*04:02	0.143
A*30:02~DRB1*15:03	0.267	A*30:02~B*45:01~DQB1*06:02	0.400	B*42:01~C*17:01~DRB1*03:02~DQA1*04:02	0.071
DQA1*02:01~DQB1*02:01	0.222	C*17:01~DQA1*04:02~DQB1*04:01	0.313	B*57:03~C*17:01~DRB1*03:02~DQA1*04:02	0.071
C*03:04~DQA1*05:02	0.218	C*17:01~DQA1*05:02~DQB1*04:01	0.312	B*15:10~C*17:01~DRB1*15:03~DQA1*05:02	0.071
A*30:02~B*45:01	0.211	A*30:02~DRB1*15:03~DQB1*05:01	0.250	B*42:02~C*03:04~DRB1*15:03~DQA1*05:02	0.071
A*30:02~DQB1*06:02	0.208	A*30:02~DRB1*15:03~DQB1*06:02	0.250	B*42:01~C*17:01~DQA1*04:02~DQB1*04:01	0.345
DRB1*15:03~DQA1*02:01	0.207	A*68:01~DRB1*03:01~DQB1*02:01	0.250	B*15:10~C*03:04~DQA1*05:02~DQB1*04:01	0.220
B*42:02~C*17:01	0.188	A*68:01~DRB1*11:01~DQB1*03:01	0.250	B*42:01~C*17:01~DQA1*05:02~DQB1*04:01	0.155
C*03:04~DQB1*04:01	0.188	B*42:01~C*17:01~DQB1*04:01	0.250	B*15:10~C*03:04~DQA1*04:02~DQB1*04:01	0.155
C*17:01~DRB1*15:03	0.178	B*42:01~DRB1*11:02~DQA1*04:02	0.222	B*15:10~C*17:01~DQA1*05:02~DQB1*04:01	0.125
A*30:02~DQB1*05:01	0.167	B*42:01~C*17:01~DRB1*15:03	0.211		
A*68:01~DQB1*02:01	0.167	B*42:02~DRB1*11:02~DQB1*03:19	0.200		

haplotype frequency (HF).

TABLE 5 | Pairwise linkage disequilibrium.

Locus pair	D'	Wn	*p*-value
A:B	0.0310	0.9501	<0.0001**
A:C	1.0000	1.000	<0.0001**
A:DRB1	1.0000	1.000	<0.0001**
A:DQA1	0.0000	0.9721	<0.0001**
A:DQB1	0.9583	0.7958	<0.0001**
A:DPB1	+	+	+
B:C	0.9842	0.8967	<0.0001**
B:DRB1	0.8110	0.7693	<0.0001**
B:DQA1	0.7458	0.6177	0.0050*
B:DQB1	0.9328	0.8895	<0.0001**
B:DPB1	+	+	+
C:DRB1	0.7771	0.6520	<0.0001**
C:DQA1	0.5335	0.5335	0.0070*
C:DQB1	0.4583	0.7253	0.1061^{NS}
C:DPB1	+	+	+
DRB1:DQA1	0.5978	0.6758	0.0130*
DRB1:DQB1	0.8669	0.7042	<0.0001**
DRB1:DPB1	+	+	+
DQA1:DQB1	0.6288	0.6693	<0.0001**
DQA1:DPB1	+	+	+
DQB1:DPB1	+	+	+

*D', Hedrick's statistic (Hedrick, 1987); Wn, Cramer's V statistic (Cramér, 1946) for global LD; **highly statistically significant (p < 0.0001); *significant (p < 0.05); ^{NS} not significant (p > 0.05); +, no high resolution HLA ~DPB1 data; #, no data after filtering in Pypop.*

D′ (Hedrick, 1987) and Cramer's V Statistic (W_n) (Cramér, 1946) were strongly significant ($p < 0.0001$) and significant ($p < 0.05$) except for C:DQB1 loci pairs (**Table 5**).

Population Comparison

NMDS analysis implemented in gene [RATE] tools (Nunes, 2016) suggests high genetic diversity in the HLA ~DRB1 locus amongst the global populations referred to (**Figure 2**). Global populations show less diversity in HLA ~A loci, with only two clusters (our data set and other populations) shown by NMDS (**Figure 2**). Additionally, our dataset distinctly clustered away from other global populations (**Supplementary Figure S1**). Usually, closely related populations cluster together while non-related populations form distinct clusters. Tight clusters separated from the rest suggest population sub-structure in the dataset. NMDS analysis suggests high genetic diversity in HLA ~B, ~DQA1, ~DRB1, ~DQB1 (**Supplementary Figure S1**). The NJ generated tree (**Figure 3**) shows a close relationship of the current data (RSA) with other previously described South African studies: SoAC (Paximadis et al., 2012), SoAB (Paximadis et al., 2012) and SANT (Hammond and Anley, 2006), but not with SANZ (Hammond et al., 2006), SAB (Tshabalala et al., 2018), SAI (Loubser et al., 2017a), RMX (Loubser et al., 2017b) and WOR (Grifoni et al., 2018). Interestingly, although our PhyloD generated probability simulated data (PSA) did not cluster with the data generated from RSA, it was closely related to a previous South African study, SAB (Tshabalala et al., 2018) (**Figure 3**). Pairwise F_{ST} based PCA showed 69.6 and 11.1% total population variability explained by PC1 and PC2, respectively (**Figure 4**). PCA suggests Central African Republic Mbenzele Pygmy (CARMP) are completely different from other sub-Saharan populations (**Figure 4**). Additional outliers include Cameroon Baka Pygmy (CBP) and Cameroon Sawa (CSw). Our data (RSA) seem to cluster together with Cameroon Bakola Pygmy (CBkP) and South Africa Natal Tamil (SANT). Our PhyloD generated PSA clustered with the other remaining populations, with Ghana Ga-Adangbe (GGA), Senegal Niokholo Mandenka (SenMAND) and Zambia Lusaka (ZaL) forming a small separate cluster (**Figure 4**). Cumulative frequency comparison between sub-Saharan

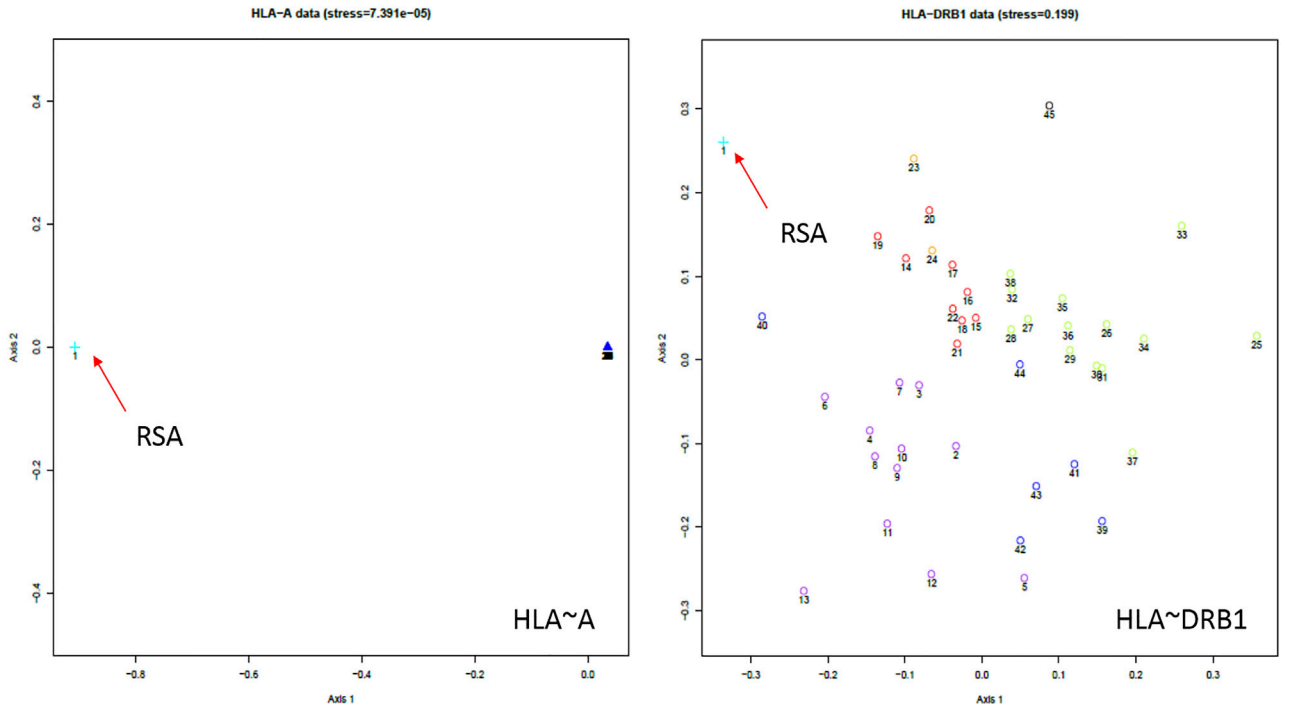


FIGURE 2 | South African HLA ~A and ~DRB1 NMDS analysis using gene[RATE] tools (Nunes, 2016). The distances between each population correlate to the HLA profile dissimilarity in those populations. For example, in HLA ~A, South Africans are distinctly different from the other global populations (clumped together in the far right of the HLA ~A graph). The orientation of axes in NMDS plots is arbitrary and can be rotated in any direction. South African data = orange arrows. NMDS for all loci and description of populations compared are detailed in **Supplementary Figure S1**. NE-EUR (Northeast Europe), CW-EUR (Central and West Europe), SE-EUR (Southeast Europe), WASI (Western Asia), NAFR (Northern Africa), OTH (other European populations of recent origin), USER (South African). Full list in **Supplementary Figure S1**.

populations in **Figure 5** suggests high HLA ~A and ~B diversity amongst South Africans compared to others. Cumulative HLA ~C allelic diversity in South Africans and Kenyan Nandi (KENN) (Cao et al., 2004) was more comparable to Kenyan Luo (KENL) (Cao et al., 2004), Ugandan Kampala pop 2 (UgaKam2) (Kijak et al., 2009) and Zambia Lusaka (ZaL) (Cao et al., 2004) (**Figure 5**).

DISCUSSION

This study applied several population genetic approaches to improve our understanding of HLA diversity in the South African population using retrospectively typed high resolution HLA data.

Ewens-Watterson neutrality test (Watterson, 1978) detected homozygosity ($p < 0.0001$) in HLA ~A and ~DRB1, which is suggestive of balancing selection at these loci (**Table 2**). Balancing selection is well documented to maintain HLA diversity within populations (Barreiro et al., 2008). Although the Ewens-Watterson neutrality test (Watterson, 1978) was designed for non-recombining data, the test has been evaluated to be insensitive to recombination (Zeng et al., 2007). As a result, this test may confidently be used to detect selection in HLA genes, which are known to have a high recombination rate. Deviations from neutrality due to recombination are expected to decrease haplotype homozygosity (Wright et al., 2006; Sanchez-Mazas et al., 2012a) but not influence balancing selection driven allele diversity. The exact mechanism of how balancing selection promotes HLA diversity is poorly understood (Barreiro et al., 2008). Generally, excessive homozygosity is not the result of population sub-structure, but is more common in datasets from admixed (genetically diverse) populations (Sinnock, 1975). This phenomenon has been termed the Wahlund effect (Sinnock, 1975).

The three most frequent alleles detected in South Africans have previously been reported in different AFND populations at varying frequencies (González-Galarza et al., 2015). Interestingly, HLA ~DQA1*05:02 with a frequency of 0.258 in the current study was previously reported at lower frequencies of 0.013 and 0.004 in South African Worcester~WOR (Grifoni et al., 2018) and Harare Zimbabwean Shona~ZiHS (González-Galarza et al., 2015) populations, respectively. This allele has likewise been reported in African Americans at a low frequency of 0.017, and is present at an even lower frequency (0.005) in people in the Wielkopolska Region in Poland (González-Galarza et al., 2015). Additionally, our most common class I allele, HLA ~C*17:01 (**Supplementary Table S1**) with a frequency of 0.149, has previously been reported at lower frequencies in

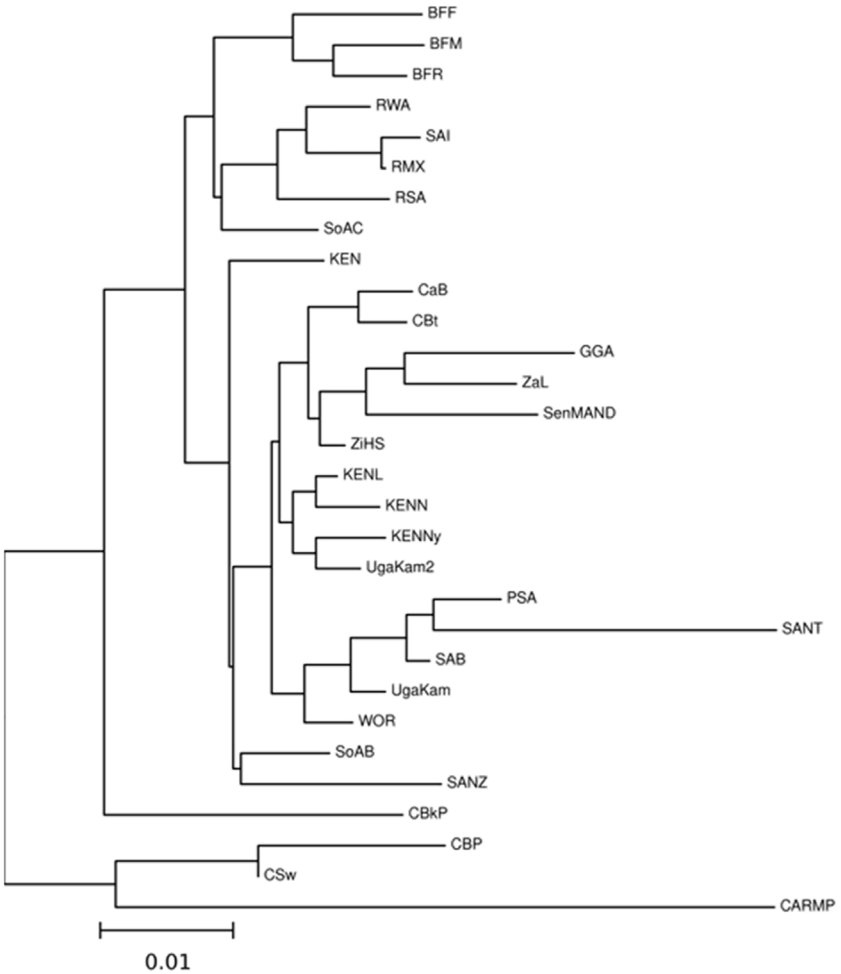


FIGURE 3 | Neighbour-Joining tree based on Nei's genetic distance for HLA ~A, ~B and ~C calculated from sub-Saharan populations. High resolution (four digit typing) HLA ~A, ~B and ~C allele frequencies were used to determine phylogenetic relatedness. Populations include Burkina Faso Fulani (BFF) (Modiano et al., 2001) Burkina Faso Mossi (BFM) (Modiano et al., 2001), Burkina Faso Rimaibe (BFR) (Modiano et al., 2001), Cameroon Baka Pygmy (CBP) (Torimiro et al., 2006), Cameroon Bakola Pygmy (CBkP) (Bruges Armas et al., 2003), Cameroon Bamileke (CaB) (Torimiro et al., 2006), Cameroon Beti (CBt) (Torimiro et al., 2006), Cameroon Sawa (CSw) (Torimiro et al., 2006), Central African Republic Mbenzele Pygmy (CARMP) (Bruges Armas et al., 2003), Ghana Ga-Adangbe (GGA) (Norman et al., 2013), Kenya (KEN) (Luo et al., 2002), Kenya Luo (KENL) (Cao et al., 2004), Kenya Nandi (KENN) (Cao et al., 2004), Kenya, Nyanza Province, Luo tribe (KENNy) (Arlehamn et al., 2017), PhyloD generated data (PSA) (Listgarten et al., 2008), RSA (current study), Rwanda (RWA) (Tang et al., 2000), Senegal Niokholo Mandenka (SenMAND) (Sanchez-Mazas et al., 2000), South Africa Black (SoAB) (Paximadis et al., 2012), South Africa Caucasians (SoAC) (Paximadis et al., 2012), South Africa Natal Tamil (SANT) (Hammond and Anley, 2006), South Africa Natal Zulu (SANZ) (Hammond et al., 2006), South Africa Worcester (WOR) (Grifoni et al., 2018), South African Bone Marrow Registry (SAB) (Tshabalala et al., 2018), South African Indian population (SAI) (Loubser et al., 2017a), South African Mixed ancestry (RMX) (Loubser et al., 2017b), Uganda Kampala (UgaKam) (Cao et al., 2004), Uganda Kampala pop 2 (UgaKam2) (Kijak et al., 2009), Zambia Lusaka (ZaL) (Cao et al., 2004) and Zimbabwe Harare Shona (ZiHS) (Louie et al., 2006). Current NHLS and SANBS data (RSA) showed phylogenetic relatedness to some previous South African studies i.e. SoAC (Paximadis et al., 2012), SoAB (Paximadis et al., 2012) and SANT (Hammond and Anley, 2006), but not with SANZ (Hammond et al., 2006) SAB (Tshabalala et al., 2018), SAI (Loubser et al., 2017a), RMX (Loubser et al., 2017b), and WOR (Grifoni et al., 2018) using the Nei's genetic distances (Nei, 1972).

other South African populations. These include South African Worcester~WOR (Grifoni et al., 2018), black South Africans~SoAB (Paximadis et al., 2012), Caucasian South Africans~SoAC (Paximadis et al., 2012) and in the South African Bone Marrow Registry~SAB (Tshabalala et al., 2018) with frequencies of 0.053, 0.111, 0.005 and 0.028, respectively. This allele (HLA ~C*17:01) is present at lower frequencies (<0.01) in Caucasian, Asian and Hispanic populations residing in the USA, while observed at higher frequencies (>0.06) in Africans, African Americans and Caribbeans (González-Galarza et al., 2015). HLA ~DQA1*04:02 (frequency of 0.194 and second most common in this study), has not previously been reported in any other South African study, but lower frequencies of 0.006 and 0.001 have been reported in Czech Republic (Europe) and San Diego (USA) populations, respectively (González-Galarza et al., 2015; Zajacova et al., 2016; Moore et al., 2018).

The top three haplotypes detected in the South African population have not been reported in any population in the AFND (González-Galarza et al., 2015). There was a strong global

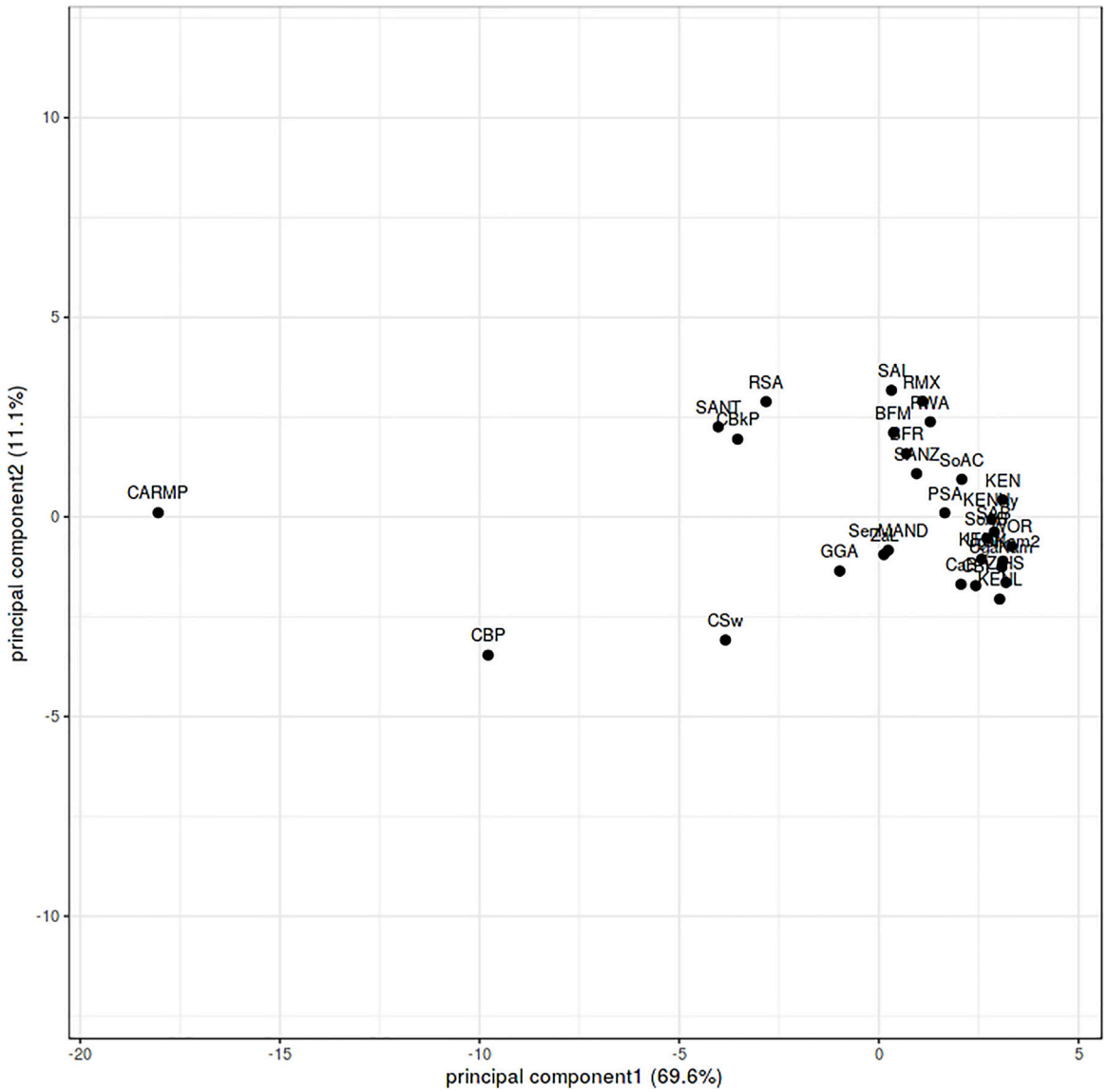


FIGURE 4 | F_{ST} based principal component analysis of HLA ~A, ~B and ~C calculated from sub-Saharan populations. Burkina Faso Fulani (BFF) (Modiano et al., 2001) Burkina Faso Mossi (BFM) (Modiano et al., 2001), Burkina Faso Rimaibe (BFR) (Modiano et al., 2001), Cameroon Baka Pygmy (CBP) (Torimiro et al., 2006), Cameroon Bakola Pygmy (CBkP) (Bruges Armas et al., 2003), Cameroon Bamileke (CaB) (Torimiro et al., 2006), Cameroon Beti (CBt) (Torimiro et al., 2006), Cameroon Sawa (CSw) (Torimiro et al., 2006), Central African Republic Mbenzele Pygmy (CARMP) (Bruges Armas et al., 2003), Ghana Ga-Adangbe (GGA) (Norman et al., 2013), Kenya (KEN) (Luo et al., 2002), Kenya Luo (KENL) (Cao et al., 2004), Kenya Nandi (KENN) (Cao et al., 2004), Kenya, Nyanza Province, Luo tribe (KENNy) (Arlehamn et al., 2017), PhyloD generated data (PSA) (Listgarten et al., 2008), RSA (current study), Rwanda (RWA) (Tang et al., 2000), Senegal Niokholo Mandenka (SenMAND) (Sanchez-Mazas et al., 2000), South Africa Black (SoAB) (Paximadis et al., 2012), South Africa Caucasians (SoAC) (Paximadis et al., 2012), South Africa Natal Tamil (SANT) (Hammond and Anley, 2006), South Africa Natal Zulu (SANZ) (Hammond et al., 2006), South Africa Worcester (WOR) (Grifoni et al., 2018), South African Bone Marrow Registry (SAB) (Tshabalala et al., 2018), South African Indian population (SAI) (Loubser et al., 2017a), South African Mixed ancestry (RMX) (Loubser et al., 2017b), Uganda Kampala (UgaKam) (Cao et al., 2004), Uganda Kampala pop 2 (UgaKam2) (Kijak et al., 2009), Zambia Lusaka (ZaL) (Cao et al., 2004) and Zimbabwe Harare Shona (ZiHS) (Louie et al., 2006).

LD between all locus pairs in our study except for C:DQB1~p = 0.1061 (**Table 5**). Haplotype diversity coupled with highly significant LD might provide insight into purifying selection (Alter et al., 2017) in the HLA genomic region. Due to data missingness, allele frequencies were computed for individual loci with all double blank alleles removed. However, for haplotype frequencies, we could not filter missing data since no data was available for computations after attempted filtering. As a result, the reported haplotype frequencies in this study might be higher than their respective allele frequencies. This limitation and the retrospective nature of the study (it is not possible to access some data that might correct this limitation) does not reduce its potential usefulness particularly given the important need for HLA data from these populations.

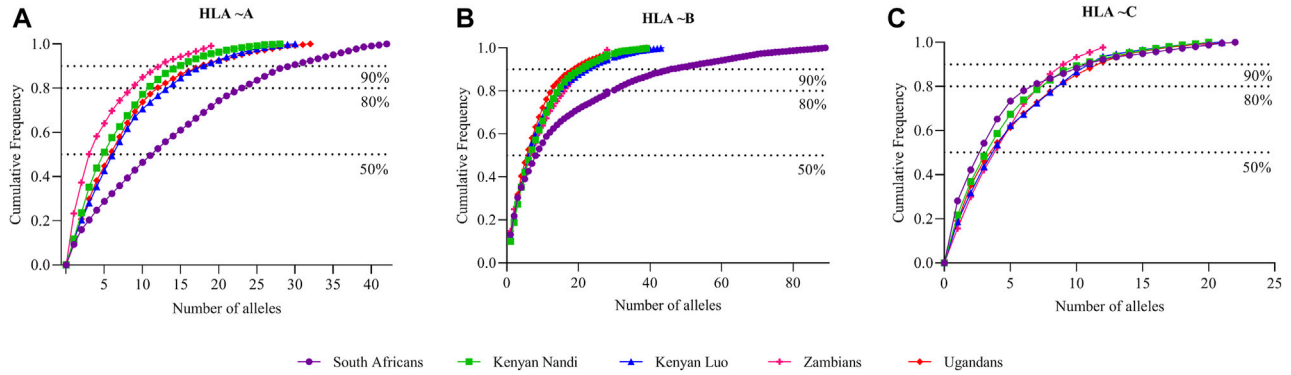


FIGURE 5 | Comparison of cumulative allele frequencies from South Africa, Kenyan, Ugandan and Zambian populations. Cumulative allele frequency indicating population coverage of South African, Kenyan, Ugandan and Zambian **(A)** HLA ~A, **(B)** ~B and **(C)** ~C alleles at high resolution. HLA alleles were sorted according to their allele frequencies in descending order; cumulative frequencies were plotted according to the number of alleles. HLA allele frequency data for African populations was obtained from the Allele Frequencies Net Database (AFND) (González-Galarza et al., 2015), Kenyan Luo (KENL) and Nandi (KENN) (Cao et al., 2004), Ugandan Kampala pop 2 (UgaKam2) (Kijak et al., 2009) and Zambia Lusaka (ZaL) (Cao et al., 2004).

Population comparisons based on allele frequencies using NMDS showed distinct differences between South Africans and other, mostly European populations. This further supports high genetic diversity in Africans in general (Chen et al., 1995; Zietkiewicz et al., 1997; Jorde et al., 2000; Prugnolle et al., 2005; Disotell, 2012), with higher diversity in some HLA loci (HLA ~B, ~DQA1, ~DRB1, ~DQB1) than others. High genetic diversity was further confirmed through cumulative frequencies (**Figure 1**) with an increased number of alleles required to cover the same combined cumulative frequency. Cumulative frequencies for HLA ~A, ~B and ~C alleles were compared with other sub-Saharan African populations including diverse Kenyans (KENN and KENL) and Ugandans (UgaKam2) (**Figure 5**). South Africans displayed high diversity at HLA ~A and HLA ~B when compared to KENN, KENL, UgaKam2 and ZaL populations while these comparator populations showed similarities in frequency between themselves. Less diverse distribution of HLA ~C alleles is observed in South Africans and other sub-Saharan African populations. Data from the current study (RSA) was related to other South African data sets using the Nei's' genetic distance (Nei, 1972) and NJ method (Saitou and Nei, 1987) unrooted tree (**Figure 3**). We expected all the studied South African populations to cluster together, or show more phylogenetic closeness; however, this was not the case. Other South African studies including South Africa Natal Zulu ~SANZ (Hammond et al., 2006), South African Bone Marrow Registry ~SAB (Tshabalala et al., 2018), South African Indian ~SAI (Loubser et al., 2017a), South African Mixed ancestry ~RMX (Loubser et al., 2017b) and South Africa Worcester ~WOR (Grifoni et al., 2018) were more related to other sub-Saharan populations than our current study (RSA). This is once again suggestive of high HLA diversity in South African populations, and their genetic relatedness to other African populations. Generally, if HLA data do not show the expected relatedness amongst populations (geographically, ethnolinguistically, anthropologically and linguistically related), this suggests diversification of the studied loci amongst those populations (Mack et al., 2012). Genetic distance computation assumes that genetic drift drives population differentiation, but there is strong evidence of balancing selection driving differentiation in HLA loci (Hedrick and Thomson, 1983; Hughes and Nei, 1988; Lawlor et al., 1988; Meyer and Thomson, 2001). Caution should thus be exercised when interpreting HLA genetic distance analysis between populations.

Although the expected genetic relatedness was not observed between the current study and other South African studies as mentioned above, PCA confirmed the genetic relatedness of South Africans (current RSA study) to other sub-Saharan populations (**Figure 4**). There is however limited high resolution data for nations neighboring South Africa for comparison, as previously reviewed (Tshabalala et al., 2015). Only data from Zambia Lusaka (ZaL) (Cao et al., 2004) and Zimbabwe Harare Shona (ZiHS) (Louie et al., 2006) was included; as a result, interpretation of this result needs to be done with caution.

The dataset had some missing alleles for some participants (data missingness). However, due to the retrospective nature of the study, we could not distinguish between missing data and blank alleles. We attempted to address this by using our dataset to simulate high resolution (four digit) class I data (Listgarten et al., 2008). Bioinformatics tools have been key in simulating high resolution typing to further understand HLA diversity (Listgarten et al., 2008; Gragert et al., 2013). There is confidence in our simulated data as it clustered with some South African HLA data (Tshabalala et al., 2018) (**Figure 3**) and other sub-Saharan populations (**Figure 4**). This provides hope in using simulated high resolution data from populations like South Africa, which currently have limited HLA data. The PhyloD tool used to address data missingness does not have an "African" representative dataset as a reference, which would provide simulations that are more accurate. Instead, an "African American" dataset was used as a reference in our simulations. We acknowledge that this reference dataset might not be ideal for all African populations since it is based on African Americans

which have a particular geographic origin (west Africa). However, this is the closest available population to the South African population. Ethnic information on our study participants would have further facilitated simulating missing HLA data, but this was not available.

Highly significant deviations from HWE were observed which might be explained by the high data missingness or the presence of family members in the dataset. Other potential causes of the significant deviation from HWE include data heterogeneity, admixture, population sub-structure, a highly endogamous population and a strong selection pressure (Wills, 1991). HWE approximation may give insights into HLA genotyping quality and sampling errors. Due to the retrospective nature of the study, we acknowledge the potential of genotyping errors or failure to detect some alleles (blank allele) which might have contributed to the homozygosity observed, and which could have contributed to the deviation from HWE (Mack et al., 2012).

Additionally, the highly significant HWE deviations (as seen in this study) have been reported to influence allele and haplotype estimations (Single et al., 2002). Global LD considers all possible allele combinations from two loci studied (Klitz et al., 1995); in our case, Hedrick's D′ (Hedrick, 1987) weights alleles in each haplotype and Cramer's V Statistic (W_n) (Cramér, 1946) is a multi-allelic correlation measure between pairs of loci. Haplotype frequency is influenced by LD, sample size, completeness of HLA data and allele frequency (Lewontin, 1964), especially if gamete phase is unknown (reviewed in Mack et al. (2012)). Other reported confounders to haplotype estimation include typing ambiguity (Castelli et al., 2010) and sample size (Gourraud et al., 2015).

We also note the limitation of not having access to demographic information and disease status of the study participants, as these factors contribute to HLA diversity. Although an individual's inherited HLA genotype does not change due to disease state, continuous exposure to pathogens in a population result in increased HLA diversity over an evolutionary time period (Prugnolle et al., 2005). Generally, HLA allele frequencies provide insight into population history and not necessarily information on selection (Blagitko et al., 1997). HLA data has been widely used to understand genetic relatedness of different populations as well as demographic events in those populations (Sanchez-Mazas and Meyer, 2014). The large sample size of the current study might shed light on some demographic events in South Africa and how these relate to other sub-Saharan populations. Population allele frequencies may be used in disease association studies and provide insight into genetic relatedness (Mack et al., 2009; Romphruk et al., 2010; Sanchez-Mazas et al., 2012b). They may additionally be used to track population evolutionary processes including migration, selection and admixture (Fernandez Vina et al., 2012).

CONCLUSION

We provide insight into HLA diversity in South Africans. This constitutes part of our ongoing efforts to fully understand HLA diversity in Africans, and to build a resource for future studies. Generally, HLA genetic makeup of populations provides insight into their population history including selective pressures by pathogens (Prugnolle et al., 2005), migration, admixture, and changes in population size (Parham and Ohta, 1996; Kijak et al., 2009; Buhler and Sanchez-Mazas, 2011; Sanchez-Mazas et al., 2011). Comparison of HLA data at a population level suggests genetic differences and uniqueness of South Africans relative to other global populations. We acknowledge the limitation of the retrospective nature of the data and data missingness, the imbalance of sample sizes for each locus and haplotype pairs and methodological difficulties. Despite these limitations, this study provides a unique and large HLA dataset of South Africans, which could be a useful future resource to support anthropological studies, disease association studies, population based vaccine development and donor recruitment programs.

STUDY LIMITATIONS

The study had limitations accessing the demographic data of individuals, which could have been beneficial in understanding the HLA diversity in South African populations characterized by ethnic, linguistic and racial diversity. Additionally, due to the retrospective nature of the study, we could not distinguish between homozygous typing and/or missingness of one allele. For allele frequency estimation, we filtered out the data to exclude individuals who did not have data for some particular loci, as compared to haplotype estimation where the input included the whole dataset with missing data. The imbalance of sample sizes for each locus and haplotype pairs induced methodological difficulties resulting in no tallying of allele frequencies vs. haplotype frequencies. However, we believe these limitations are far outweighed by the critical importance of understanding HLA data in these populations.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Research Ethics Committee of the University of Pretoria, Faculty of Health Sciences (approval no. 220/2015); SANBS Human Research Ethics Committee (SANBS HREC) and the NHLS Academic Affairs and Research. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

MT designed the study, compiled data, analyzed the data and wrote the manuscript. JM analyzed the data and wrote the

manuscript. KV, DN, and FM recruited the study participants, provided HLA data and contributed in manuscript writing. AC provided and supervised data analysis plan and provided critical review of the manuscript. MP conceived the study, obtained funding for and supervised the study, and provided critical review during the project and of the manuscript.

FUNDING

This research and the publication thereof is the result of funding provided by the South African Medical Research Council (SAMRC) in terms of the MRC's Flagships Awards Project (SAMRC-RFA-UFSP-01-2013/STEM CELLS), the SAMRC Extramural Unit for stem cell Research and Therapy, the Institute for Cellular and Molecular Medicine of the University of Pretoria, and the National Research Foundation of South Africa. The above-mentioned funding bodies had no role in the design of the study, collection, analysis, interpretation of data and in writing the manuscript.

ACKNOWLEDGMENTS

We acknowledge the NHLS (through the NHLS Corporate Data Warehouse) and the SANBS for providing access to HLA genotype data.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.711944/full#supplementary-material

REFERENCES

Alter, I., Gragert, L., Fingerson, S., Maiers, M., and Louzoun, Y. (2017). HLA Class I Haplotype Diversity Is Consistent with Selection for Frequent Existing Haplotypes. *PLoS Comput. Biol.* 13, e1005693. doi:10.1371/journal.pcbi.1005693

Arlehamn, C. S. L., Copin, R., Leary, S., Mack, S. J., Phillips, E., Mallal, S., et al. (2017). Sequence-based HLA-A, B, C, DP, DQ, and DR Typing of 100 Luo Infants from the Boro Area of Nyanza Province, Kenya. *Hum. Immunol.* 78, 325–326. doi:10.1016/j.humimm.2017.03.007

Association, W. M. (2013). World Medical Association Declaration of Helsinki: Ethical Principles for Medical Research Involving Human Subjects. *JAMA* 310, 2191–2194. doi:10.1001/jama.2013.281053

Barreiro, L. B., Laval, G., Quach, H., Patin, E., and Quintana-Murci, L. (2008). Natural Selection Has Driven Population Differentiation in Modern Humans. *Nat. Genet.* 40, 340–345. doi:10.1038/ng.78

Beatty, P. G., Boucher, K. M., Mori, M., and Milford, E. L. (2000). Probability of Finding HLA-Mismatched Related or Unrelated Marrow or Cord Blood Donors. *Hum. Immunol.* 61, 834–840. doi:10.1016/s0198-8859(00)00138-5

Berger, L. R., Hawks, J., De Ruiter, D. J., Churchill, S. E., Schmid, P., Delezene, L. K., et al. (2015). Homo Naledi, a New Species of the Genus Homo from the Dinaledi Chamber, South Africa. *eLife* 4, e09560. doi:10.7554/eLife.09560

Blagitko, N., O'hUigin, C., Figueroa, F., Horai, S., Sonoda, S., Tajima, K., et al. (1997). Polymorphism of the HLA-DRB1 Locus in Colombian, Ecuadorian, and Chilean Amerinds. *Hum. Immunol.* 54, 74–81. doi:10.1016/s0198-8859(97)00005-0

Brander, C., Frahm, N., and Walker, B. D. (2006). The Challenges of Host and Viral Diversity in HIV Vaccine Design. *Curr. Opin. Immunol.* 18, 430–437. doi:10.1016/j.coi.2006.05.012

Bruges Armas, J., Destro-Bisol, G., López-Vazquez, A., Couto, A. R., Spedini, G., Gonzalez, S., et al. (2003). HLA Class I Variation in the West African Pygmies and Their Genetic Relationship with Other African Populations. *Tissue Antigens* 62, 233–242. doi:10.1034/j.1399-0039.2003.00100.x

Buhler, S., and Sanchez-Mazas, A. (2011). HLA DNA Sequence Variation Among Human Populations: Molecular Signatures of Demographic and Selective Events. *PloS ONE* 6, e14643. doi:10.1371/journal.pone.0014643

Burrell, A. S., and Disotell, T. R. (2009). Panmixia Postponed: Ancestry-Related Assortative Mating in Contemporary Human Populations. *Genome Biol.* 10, 245. doi:10.1186/gb-2009-10-11-245

Cao, K., Moormann, A. M., Lyke, K. E., Masaberg, C., Sumba, O. P., Doumbo, O. K., et al. (2004). Differentiation between African Populations Is Evidenced by the Diversity of Alleles and Haplotypes of HLA Class I Loci. *Tissue Antigens* 63, 293–325. doi:10.1111/j.0001-2815.2004.00192.x

Carrington, M., and O'brien, S. J. (2003). The Influence of HLA Genotype on AIDS. *Annu. Rev. Med.* 54, 535–551. doi:10.1146/annurev.med.54.101601.152346

Castelli, E. C., Mendes-Junior, C. T., Veiga-Castelli, L. C., Pereira, N. F., Petzl-Erler, M. L., and Donadi, E. A. (2010). Evaluation of Computational Methods for the Reconstruction of HLA Haplotypes. *Tissue Antigens* 76, 459–466. doi:10.1111/j.1399-0039.2010.01539.x

Chen, C. H., Matthews, T. J., Mcdanal, C. B., Bolognesi, D. P., and Greenberg, M. L. (1995). A Molecular Clasp in the Human Immunodeficiency Virus (HIV) Type 1 TM Protein Determines the Anti-HIV Activity of Gp41 Derivatives: Implication for Viral Fusion. *J. Virol.* 69, 3771–3777. doi:10.1128/jvi.69.6.3771-3777.1995

Chen, H., Ndhlovu, Z. M., Liu, D., Porter, L. C., Fang, J. W., Darko, S., et al. (2012). TCR Clonotypes Modulate the Protective Effect of HLA Class I Molecules in HIV-1 Infection. *Nat. Immunol.* 13, 691–700. doi:10.1038/ni.2342

Choudhury, A., Ramsay, M., Hazelhurst, S., Aron, S., Bardien, S., Botha, G., et al. (2017). Whole-genome Sequencing for an Enhanced Understanding of Genetic Variation Among South Africans. *Nat. Commun.* 8, 2062–00663. doi:10.1038/s41467-017-00663-9

Coombs, J. A., Letcher, B. H., and Nislow, K. H. (2008). CREATE: a Software to Create Input Files from Diploid Genotypic Data for 52 Genetic Software Programs. *Mol. Ecol. Resour.* 8, 578–580. doi:10.1111/j.1471-8286.2007.02036.x

Cramér, H. (1946). *Mathematical Methods of Statistics*. Princeton: Princeton University Press.

Disotell, T. R. (2012). Archaic Human Genomics. *Am. J. Phys. Anthropol.* 149, 24–39. doi:10.1002/ajpa.22159

Dyer, P., Mcgilvray, R., Robertson, V., and Turner, D. (2013). Status Report from 'double Agent HLA': Health and Disease. *Mol. Immunol.* 55, 2–7. doi:10.1016/j.molimm.2012.08.016

Eberhard, H.-P., Madbouly, A. S., Gourraud, P. A., Balère, M. L., Feldmann, U., Gragert, L., et al. (2013). Comparative Validation of Computer Programs for Haplotype Frequency Estimation from Donor Registry Data. *Tissue Antigens* 82, 93–105. doi:10.1111/tan.12160

Edinur, H. A., Manaf, S. M., and Mat, N. F. C. (2016). Genetic Barriers in Transplantation Medicine. *Wjt* 6, 532–541. doi:10.5500/wjt.v6.i3.532

Excoffier, L., and Slatkin, M. (1995). Maximum-likelihood Estimation of Molecular Haplotype Frequencies in a Diploid Population. *Mol. Biol. Evol.* 12, 921–927. doi:10.1093/oxfordjournals.molbev.a040269

Garamszegi, L. Z. (2014). Global Distribution of Malaria-Resistant MHC-HLA Alleles: the Number and Frequencies of Alleles and Malaria Risk. *Malar. J.* 13, 349–2875. doi:10.1186/1475-2875-13-349

González-Galarza, F. F., Takeshita, L. Y. C., Santos, E. J. M., Kempson, F., Maia, M. H. T., Silva, A. L. S. d., et al. (2015). Allele Frequency Net 2015 Update: New Features for HLA Epitopes, KIR and Disease and HLA Adverse Drug Reaction Associations. *Nucleic Acids Res.* 43, D784–D788. doi:10.1093/nar/gku1166

Gourraud, P.-A., Pappas, D. J., Baouz, A., Balère, M.-L., Garnier, F., and Marry, E. (2015). High-resolution HLA-A, HLA-B, and HLA-DRB1 Haplotype Frequencies from the French Bone Marrow Donor Registry. *Hum. Immunol.* 76, 381–384. doi:10.1016/j.humimm.2015.01.028

Gragert, L., Madbouly, A., Freeman, J., and Maiers, M. (2013). Six-locus High Resolution HLA Haplotype Frequencies Derived from Mixed-Resolution DNA Typing for the Entire US Donor Registry. *Hum. Immunol.* 74, 1313–1320. doi:10.1016/j.humimm.2013.06.025

Grifoni, A., Sidney, J., Carpenter, C., Phillips, E., Mallal, S., Scriba, T. J., et al. (2018). Sequence-based HLA-A, B, C, DP, DQ, and DR Typing of 159 Individuals from the Worcester Region of the Western Cape Province of South Africa. *Hum. Immunol.* 79, 143–144. doi:10.1016/j.humimm.2018.01.004

Guo, S. W., and Thompson, E. A. (1992). Performing the Exact Test of Hardy-Weinberg Proportion for Multiple Alleles. *Biometrics* 48, 361–372. doi:10.2307/2532296

Hammond, M. G., and Anley, D. (2006). "Tamil from Natal Province, South Africa," in *Proceedings of the 13th International Histocompatibility Workshop*. Editor J. A. Hansen (Seattle, Washington: International Histocompatibility Working Group Press), 609.

Hammond, M. G., Middleton, D., and Anley, D. (2006). "Zulu from Natal Province, South Africa," in *Proceedings of the 13th International Histocompatibility Workshop and Conference*. Editor J. A. Hansen (Seattle, Washington: IHWG Press), 590–591.

Hayhurst, J. D., Du Toit, E. D., Borrill, V., Schlaphoff, T. E. A., Brosnan, N., and Marsh, S. G. E. (2015). Two Novel HLA alleles,HLA-A*30:02:01:03andHLA-C*08:113, Identified in a South African Bone Marrow Donor. *Tissue Antigens* 85, 291–293. doi:10.1111/tan.12542

Hedrick, P. W. (1987). Gametic Disequilibrium Measures: Proceed with Caution. *Genetics* 117, 331–341. doi:10.1093/genetics/117.2.331

Hedrick, P. W., and Thomson, G. (1983). Evidence for Balancing Selection at Hla. *Genetics* 104, 449–456. doi:10.1093/genetics/104.3.449

Hughes, A. L., and Nei, M. (1988). Pattern of Nucleotide Substitution at Major Histocompatibility Complex Class I Loci Reveals Overdominant Selection. *Nature* 335, 167–170. doi:10.1038/335167a0

Jorde, L. B., Watkins, W. S., Bamshad, M. J., Dixon, M. E., Ricker, C. E., Seielstad, M. T., et al. (2000). The Distribution of Human Genetic Diversity: a Comparison of Mitochondrial, Autosomal, and Y-Chromosome Data. *Am. J. Hum. Genet.* 66, 979–988. doi:10.1086/302825

Kijak, G. H., Walsh, A. M., Koehler, R. N., Moqueet, N., Eller, L. A., Eller, M., et al. (2009). HLA Class I Allele and Haplotype Diversity in Ugandans Supports the Presence of a Major East African Genetic Cluster. *Tissue Antigens* 73, 262–269. doi:10.1111/j.1399-0039.2008.01192.x

Klitz, W., Stephens, J. C., Grote, M., and Carrington, M. (1995). Discordant Patterns of Linkage Disequilibrium of the Peptide-Transporter Loci within the HLA Class II Region. *Am. J. Hum. Genet.* 57, 1436–1444.

Lancaster, A. K., Single, R. M., Solberg, O. D., Nelson, M. P., and Thomson, G. (2007). PyPop Update - a Software Pipeline for Large-Scale Multilocus Population Genomics. *Tissue Antigens* 69, 192–197. doi:10.1111/j.1399-0039.2006.00769.x

Lawlor, D. A., Ward, F. E., Ennis, P. D., Jackson, A. P., and Parham, P. (1988). HLA-A and B Polymorphisms Predate the Divergence of Humans and Chimpanzees. *Nature* 335, 268–271. doi:10.1038/335268a0

Lewontin, R. C. (1964). The Interaction of Selection and Linkage. I. General Considerations; Heterotic Models. *Genetics* 49, 49–67. doi:10.1093/genetics/49.1.49

Listgarten, J., Brumme, Z., Kadie, C., Xiaojiang, G., Walker, B., Carrington, M., et al. (2008). Statistical Resolution of Ambiguous HLA Typing Data. *PLoS Comput. Biol.* 4, e1000016. doi:10.1371/journal.pcbi.1000016

Loubser, S., Paximadis, M., and Tiemessen, C. T. (2017a). Human Leukocyte Antigen Class I (A, B and C) Allele and Haplotype Variation in a South African Indian Population. *Hum. Immunol.* 78, 468–470. doi:10.1016/j.humimm.2017.04.010

Loubser, S., Paximadis, M., and Tiemessen, C. T. (2017b). Human Leukocyte Antigen Class I (A, B and C) Allele and Haplotype Variation in a South African Mixed Ancestry Population. *Hum. Immunol.* 78, 399–400. doi:10.1016/j.humimm.2017.04.006

Louie, L., Mather, K., Meyer, D., Hollenbach, J., Jackman, R., Schultz, K., et al. (2006). "Shona from Harare, Zimbabwe," in *Immunobiology of the Human MHC: Proceedings of the 13th International Histocompatibility Workshop and Conference*. Editor J. A. Hansen (Seattle, Washington: IHWG Press), 587–589.

Luo, M., Embree, J., Ramdahin, S., Ndinya-Achola, J., Njenga, S., Bwayo, J. B., et al. (2002). HLA-A and HLA-B in Kenya, Africa: Allele Frequencies and Identification of HLA-B*1567 and HLA-B*4426. *Tissue Antigens* 59, 370–380. doi:10.1034/j.1399-0039.2002.590503.x

Mack, S. J., Gourraud, P.-A., Single, R. M., Thomson, G., and Hollenbach, J. A. (2012). Analytical Methods for Immunogenetic Population Data. *Methods Mol. Biol. (Clifton, N.J.)* 882, 215–244. doi:10.1007/978-1-61779-842-9_13

Mack, S. J., Tu, B., Lazaro, A., Yang, R., Lancaster, A. K., Cao, K., et al. (2009). HLA-A, -B, -C, and -DRB1 Allele and Haplotype Frequencies Distinguish Eastern European Americans from the General European American Population. *Tissue Antigens* 73, 17–32. doi:10.1111/j.1399-0039.2008.01151.x

May, A., Hazelhurst, S., Li, Y., Norris, S. A., Govind, N., Tikly, M., et al. (2013). Genetic Diversity in Black South Africans from Soweto. *BMC Genomics* 14, 644–2164. doi:10.1186/1471-2164-14-644

Metsalu, T., and Vilo, J. (2015). ClustVis: a Web Tool for Visualizing Clustering of Multivariate Data Using Principal Component Analysis and Heatmap. *Nucleic Acids Res.* 43, W566–W570. doi:10.1093/nar/gkv468

Meyer, D., C. Aguiar, V. R., Bitarello, B. D., C. Brandt, D. Y., and Nunes, K. (2018). A Genomic Perspective on HLA Evolution. *Immunogenetics* 70, 5–27. doi:10.1007/s00251-017-1017-3

Meyer, D., and Thomson, G. (2001). How Selection Shapes Variation of the Human Major Histocompatibility Complex: a Review. *Ann. Hum. Genet* 65, 1–26. doi:10.1046/j.1469-1809.2001.6510001.x

Modiano, D., Luoni, G., Petrarca, V., Sodiomon Sirima, B., De Luca, M., Simporé, J., et al. (2001). HLA Class I in Three West African Ethnic Groups: Genetic Distances from Sub-saharan and Caucasoid Populations. *Tissue Antigens* 57, 128–137. doi:10.1034/j.1399-0039.2001.057002128.x

Moore, E., Grifoni, A., Weiskopf, D., Schulten, V., Arlehamn, C. S. L., Angelo, M., et al. (2018). Sequence-based HLA-A, B, C, DP, DQ, and DR Typing of 496 Adults from San Diego, California, USA. *Hum. Immunol.* 79, 821–822. doi:10.1016/j.humimm.2018.09.008

Mungall, A. J., Palmer, S. A., Sims, S. K., Edwards, C. A., Ashurst, J. L., Wilming, L., et al. (2003). The DNA Sequence and Analysis of Human Chromosome 6. *Nature* 425, 805–811. doi:10.1038/nature02055

Ndung'u, T., Gaseitsiwe, S., Sepako, E., Doualla-Bell, F., Peter, T., Kim, S., et al. (2005). Major Histocompatibility Complex Class II (HLA-DRB and -DQB) Allele Frequencies in Botswana: Association with Human Immunodeficiency Virus Type 1 Infection. *Clin. Vaccin. Immunol* 12, 1020–1028. doi:10.1128/cdli.12.9.1020-1028.2005

Nei, M. (1972). Genetic Distance between Populations. *The Am. Naturalist* 106, 283–292. doi:10.1086/282771

Norman, P. J., Hollenbach, J. A., Nemat-Gorgani, N., Guethlein, L. A., Hilton, H. G., Pando, M. J., et al. (2013). Co-evolution of Human Leukocyte Antigen (HLA) Class I Ligands with Killer-Cell Immunoglobulin-like Receptors (KIR) in a Genetically Diverse Population of Sub-saharan Africans. *PLoS Genet.* 9, e1003938. doi:10.1371/journal.pgen.1003938

Nunes, J. M. (2016). Using UNIFORMAT and GENE[RATE] to Analyze Data with Ambiguities in Population Genetics. *Evol. Bioinform Online* 11, 19–26. doi:10.4137/EBO.S32415

Ovsyannikova, I. G., and Poland, G. A. (2011). Vaccinomics: Current Findings, Challenges and Novel Approaches for Vaccine Development. *Aaps J.* 13, 438–444. doi:10.1208/s12248-011-9281-x

Parham, P., and Ohta, T. (1996). Population Biology of Antigen Presentation by MHC Class I Molecules. *Science* 272, 67–74. doi:10.1126/science.272.5258.67

Paximadis, M., Mathebula, T. Y., Gentle, N. L., Vardas, E., Colvin, M., Gray, C. M., et al. (2012). Human Leukocyte Antigen Class I (A, B, C) and II (DRB1) Diversity in the Black and Caucasian South African Population. *Hum. Immunol.* 73, 80–92. doi:10.1016/j.humimm.2011.10.013

Prugnolle, F., Manica, A., Charpentier, M., Guégan, J. F., Guernier, V., and Balloux, F. (2005). Pathogen-Driven Selection and Worldwide HLA Class I Diversity. *Curr. Biol.* 15, 1022–1027. doi:10.1016/j.cub.2005.04.050

Ramsay, M. (2012). Africa: Continent of Genome Contrasts with Implications for Biomedical Research and Health. *FEBS Lett.* 586, 2813–2819. doi:10.1016/j.febslet.2012.07.061

Robinson, J., Halliwell, J. A., Mcwilliam, H., Lopez, R., Parham, P., and Marsh, S. G. (2013). The IMGT/HLA Database. *Nucleic Acids Res.* 41, D1222–D1227. doi:10.1093/nar/gks949

Romphruk, A. V., Romphruk, A., Kongmaroeng, C., Klumkrathok, K., Paupairoj, C., and Leelayuwat, C. (2010). HLA Class I and II Alleles and Haplotypes in Ethnic Northeast Thais. *Tissue Antigens* 75, 701–711. doi:10.1111/j.1399-0039.2010.01448.x

Saitou, N., and Nei, M. (1987). The Neighbor-Joining Method: a New Method for Reconstructing Phylogenetic Trees. *Mol. Biol. Evol.* 4, 406–425. doi:10.1093/oxfordjournals.molbev.a040454

Sanchez-Mazas, A., and Meyer, D. (2014). The Relevance of HLA Sequencing in Population Genetics Studies. *J. Immunol. Res.* 2014, 971818. doi:10.1155/2014/971818

Sanchez-Mazas, A., Vidan-Jeras, B., Nunes, J. M., Fischer, G., Little, A. M., Bekmane, U., et al. (2012b). Strategies to Work with HLA Data in Human Populations for Histocompatibility, Clinical Transplantation, Epidemiology and Population Genetics: HLA-NET Methodological Recommendations. *Int. J. Immunogenet.* 39, 459–476. doi:10.1111/j.1744-313X.2012.01113.x

Sanchez-Mazas, A., Fernandez-Viña, M., Middleton, D., Hollenbach, J. A., Buhler, S., Di, D., et al. (2011). Immunogenetics as a Tool in Anthropological Studies. *Immunology* 133, 143–164. doi:10.1111/j.1365-2567.2011.03438.x

Sanchez-Mazas, A., Lemaître, J.-F., and Currat, M. (2012a). Distinct Evolutionary Strategies of Human Leucocyte Antigen Loci in Pathogen-Rich Environments. *Phil. Trans. R. Soc. B* 367, 830–839. doi:10.1098/rstb.2011.0312

Sanchez-Mazas, A., Steiner, Q.-G., Grundschober, C., and Tiercy, J.-M. (2000). The Molecular Determination of HLA-Cw Alleles in the Mandenka (West Africa) Reveals a Close Genetic Relationship between Africans and Europeans. *Tissue Antigens* 56, 303–312. doi:10.1034/j.1399-0039.2000.560402.x

Single, R. M., Meyer, D., Hollenbach, J. A., Nelson, M. P., Noble, J. A., Erlich, H. A., et al. (2002). Haplotype Frequency Estimation in Patient Populations: the Effect of Departures from Hardy-Weinberg Proportions and Collapsing over a Locus in the HLA Region. *Genet. Epidemiol.* 22, 186–195. doi:10.1002/gepi.0163

Sinnock, P. (1975). The Wahlund Effect for the Two-Locus Model. *Am. Naturalist* 109, 565–570. doi:10.1086/283027

Slatkin, M. (1996). A Correction to the Exact Test Based on the Ewens Sampling Distribution. *Genet. Res.* 68, 259–260. doi:10.1017/s0016672300034236

Slatkin, M. (1994). An Exact Test for Neutrality Based on the Ewens Sampling Distribution. *Genet. Res.* 64, 71–74. doi:10.1017/s0016672300032560

Statistics South Africa (2022). *Statistics South Africa*. Online, Available at: www.statssa.gov.za (Accessed.

Takezaki, N., Nei, M., and Tamura, K. (2010). POPTREE2: Software for Constructing Population Trees from Allele Frequency Data and Computing Other Population Statistics with Windows Interface. *Mol. Biol. Evol.* 27, 747–752. doi:10.1093/molbev/msp312

Takezaki, N., Nei, M., and Tamura, K. (2014). POPTREEW: Web Version of POPTREE for Constructing Population Trees from Allele Frequency Data and Computing Some Other Quantities. *Mol. Biol. Evol.* 31, 1622–1624. doi:10.1093/molbev/msu093

Tang, J., Naik, E., Costello, C., Karita, E., Rivers, C., Allen, S., et al. (2000). Characteristics of HLA Class I and Class II Polymorphisms in Rwandan Women. *Exp. Clin. Immunogenet.* 17, 185–198. doi:10.1159/000019138

Torimiro, J. N., Carr, J. K., Wolfe, N. D., Karacki, P., Martin, M. P., Gao, X., et al. (2006). HLA Class I Diversity Among Rural Rainforest Inhabitants in Cameroon: Identification of A*2612-B*4407 Haplotype. *Tissue Antigens* 67, 30–37. doi:10.1111/j.1399-0039.2005.00527.x

Tshabalala, M., Ingram, C., Schlaphoff, T., Borrill, V., Christoffels, A., and Pepper, M. S. (2018). Human Leukocyte Antigen-A, B, C, DRB1, and DQB1 Allele and Haplotype Frequencies in a Subset of 237 Donors in the South African Bone Marrow Registry. *J. Immunol. Res.* 2018, 2031571. doi:10.1155/2018/2031571

Tshabalala, M., Mellet, J., and Pepper, M. S. (2015). Human Leukocyte Antigen Diversity: A Southern African Perspective. *J. Immunol. Res.* 2015, 746151. doi:10.1155/2015/746151

Vina, M. A. F., Hollenbach, J. A., Lyke, K. E., Sztein, M. B., Maiers, M., Klitz, W., et al. (2012). Tracking Human Migrations by the Analysis of the Distribution of HLA Alleles, Lineages and Haplotypes in Closed and Open Populations. *Phil. Trans. R. Soc. B* 367, 820–829. doi:10.1098/rstb.2011.0320

Watterson, G. A. (1978). Homozygosity Test of Neutrality. *Genetics* 88, 405–417. doi:10.1093/genetics/88.2.405

WHO (2013). in *Global Health Report* (Geneva: World Health Organisation).

Wills, C. (1991). Maintenance of Multiallelic Polymorphism at the MHC Region. *Immunol. Rev.* 124, 165–220. doi:10.1111/j.1600-065x.1991.tb00621.x

Wong, L.-P., Ong, R. T.-H., Poh, W.-T., Liu, X., Chen, P., Li, R., et al. (2013). Deep Whole-Genome Sequencing of 100 Southeast Asian Malays. *Am. J. Hum. Genet.* 92, 52–66. doi:10.1016/j.ajhg.2012.12.005

Wright, S. I., Foxe, J. P., Derose-Wilson, L., Kawabe, A., Looseley, M., Gaut, B. S., et al. (2006). Testing for Effects of Recombination Rate on Nucleotide Diversity in Natural Populations of Arabidopsis Lyrata. *Genetics* 174, 1421–1430. doi:10.1534/genetics.106.062588

Zajacova, M., Kotrbova-Kozak, A., and Cerna, M. (2016). HLA-DRB1, -DQA1 and -DQB1 Genotyping of 180 Czech Individuals from the Czech Republic Pop 3. *Hum. Immunol.* 77, 365–366. doi:10.1016/j.humimm.2016.02.003

Zeng, K., Mano, S., Shi, S., and Wu, C.-I. (2007). Comparisons of Site- and Haplotype-Frequency Methods for Detecting Positive Selection. *Mol. Biol. Evol.* 24, 1562–1574. doi:10.1093/molbev/msm078

Zietkiewicz, E., Yotova, V., Jarnik, M., Korab-Laskowska, M., Kidd, K. K., Modiano, D., et al. (1997). Nuclear DNA Diversity in Worldwide Distributed Human Populations. *Gene* 205, 161–171. doi:10.1016/s0378-1119(97)00408-3

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Genetics

OPINION
published: 25 March 2022
doi: 10.3389/fgene.2022.864575

Esophageal Cancer Genomics in Africa: Recommendations for Future Research

Hannah Simba[1,2], Gerard Tromp[3,4,5,6,7], Vikash Sewram[1], Christopher G Mathew[8,9], Wenlong C. Chen[8,10] and Helena Kuivaniemi[3,4,5]*

[1]African Cancer Institute, Department of Global Health, Faculty of Medicine and Health Sciences, Stellenbosch University, Cape Town, South Africa, [2]Environment and Lifestyle Epidemiology Branch, International Agency for Research on Cancer (IARC/WHO), Lyon, France, [3]Division of Molecular Biology and Human Genetics, Department of Biomedical Sciences, Faculty of Medicine and Health Sciences, Stellenbosch University, Cape Town, South Africa, [4]DSI–NRF Centre of Excellence for Biomedical Tuberculosis Research, Stellenbosch University, Cape Town, South Africa, [5]South African Medical Research Council Centre for Tuberculosis Research, Stellenbosch University, Cape Town, South Africa, [6]Bioinformatics Unit, South African Tuberculosis Bioinformatics Initiative, Stellenbosch University, Cape Town, South Africa, [7]Centre for Bioinformatics and Computational Biology, Stellenbosch University, Stellenbosch, South Africa, [8]Sydney Brenner Institute for Molecular Bioscience, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa, [9]Division of Human Genetics, National Health Laboratory Service and School of Pathology, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa, [10]National Cancer Registry, National Health Laboratory Service, Johannesburg, South Africa

Keywords: esophageal cancer, esophageal squamous cell carcinoma, genomics, genomic medicine, African esophageal cancer corridor

OPEN ACCESS

Edited by:
Dawn Stephens, Technology Innovation Agency (TIA), South Africa

Reviewed by:
Fazlur Rahman Talukdar, International Agency For Research On Cancer (IARC), France

****Correspondence:***
Hannah Simba
simbah@iarc.fr

Specialty section:
This article was submitted to Human and Medical Genomics, a section of the journal Frontiers in Genetics

Received: *28 January 2022*
Accepted: *14 March 2022*
Published: *25 March 2022*

Citation:
Simba H, Tromp G, Sewram V, Mathew CG, Chen WC and Kuivaniemi H (2022) Esophageal Cancer Genomics in Africa: Recommendations for Future Research. Front. Genet. 13:864575. doi: 10.3389/fgene.2022.864575

INTRODUCTION

Esophageal cancer (EC) is an aggressive malignancy and a major health burden documented as the sixth most common cause of cancer mortality worldwide (Bray et al., 2018). Over 80% of EC cases and deaths are reported in developing countries, where the esophageal squamous cell carcinoma (ESCC) subtype is more common, compared to the adenocarcinoma (EAC) subtype (Bray et al., 2018). EC has a peculiar geographical distribution, with high incidence rates reported in Asia and the African ESCC corridor (Abnet et al., 2017; McCormack et al., 2017; Bray et al., 2018). Malawi has the highest ESCC incidence rate globally for both men and women, followed by Kenya and Zimbabwe in Africa (Bray et al., 2018). In South Africa, ESCC is the 10th most common cancer for men and the 11th most common cancer for women (National Cancer Registry, 2019), and has the 10th highest incidence of ESCC in Africa (Bray et al., 2018). Incidence rates are, however, disproportionately higher in the Eastern Cape Province, where it is the most common cancer for men, and the second most common cancer for women (Somdyala et al., 2015). Due to a lack of ESCC early detection markers, late diagnosis and poor prognosis are the norm. Additionally, distinct ESCC molecular subtypes have not been identified, which could provide opportunities for targeted and novel therapies.

The African ESCC corridor (**Figure 1**), which spans from the eastern to the southern part of Africa is characterized by high incidence rates, young age at presentation, delayed presentation, as well as poor outcomes and survival (Van Loon et al., 2018; Asombang et al., 2019). Risk factors associated with ESCC in high-risk areas include tobacco smoking, alcohol consumption, polycyclic aromatic hydrocarbon exposure, poor diet, hot beverages, poor oral hygiene, microbiome, and genetic factors (Asombang et al., 2019; Chetwood et al., 2019; Simba et al., 2019; Simba, 2021).

Limited Number of Genomic Studies on ESCC in Africa

There is an apparent lack of genomic studies on ESCC on African populations, therefore the genetic etiology is poorly understood and implementation of genomic medicine for ESCC remains elusive. The striking geographical distribution of ESCC suggests that ESCC etiology is multifactorial, with

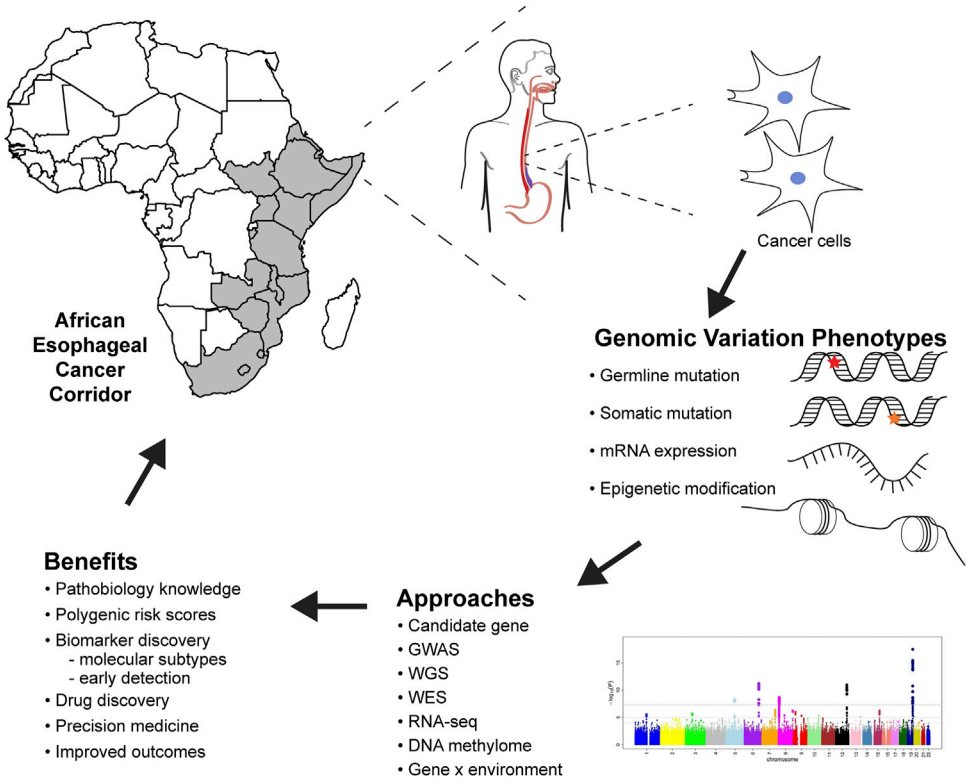


FIGURE 1 | Future of esophageal squamous cell carcinoma genomic research in Africa, methods, analysis, and benefits to genomic medicine. GWAS, genome-wide association studies; WGS, whole-genome sequencing; WES, whole-exome sequencing; RNA-seq, RNA sequencing.

shared and locally relevant environmental and genetic risk factors. There are several mutational signatures in ESCC that have been linked to environmental exposures, including tobacco smoking and alcohol consumption (Talukdar et al., 2013; Matejcic and Parker, 2015). It is, therefore, important to incorporate environmental and lifestyle risk factors when investigating genetic factors. Identification and quantification of these gene-environment (GxE) interactions provides a platform for targeted interventions and may explain the high incidence and mortality rates in Africa. In this opinion piece, we summarize the current state of ESCC genomic research in Africa, which include candidate gene studies, whole-exome studies and genome-wide association studies (GWAS). The first genome-wide methylation profiling of ESCC, with 327 tissues from nine African, Asian and South American populations, identified novel DNA methylation events in ESCC tumors (Talukdar et al., 2021). Importantly, there is a lack of multi-omics studies.

Candidate Gene Studies in African ESCC

We recently performed a systematic review of candidate gene studies carried out for ESCC in African populations (Simba et al., 2019). Studies were done on South African, Malawian and Kenyan populations only (Simba et al., 2019). Only one exome sequencing study from Malawi on ESCC tumors was found (Liu et al., 2016). In the germline, 25 SNPs were reported to be associated with ESCC in 20 different genes, and 22 genes had somatic mutations (Simba et al., 2019). These studies had several limitations: 1) small sample sizes leading to poor statistical power; 2) no cohort studies; and 3) no standardized methods of data collection, reporting, or analysis. In addition, the studies did not correct for population structure in cases and controls, which is particularly relevant in African populations known for high levels of genomic diversity (Choudhury et al., 2020).

Genome-Wide Association Studies in African ESCC are Needed

All GWAS for ESCC published to date were conducted in Asian or European populations (Cui et al., 2009; Wang et al., 2010; McKay et al., 2011; Wu et al., 2012; Wu et al., 2013; Wu et al., 2014; Abnet et al., 2017). These studies identified several associated loci, including *PLCE1*, *RUNX1* and *CHEK2*. However, the transferability and replication of these loci in African populations were limited, even in a well-powered study (Chen et al., 2019). Thus there is a need for ESCC GWAS studies in populations from high-risk regions of sub-Saharan Africa. GWAS are currently in progress in the ERICA-SA study (https://www.samrc.ac.za/intramural-research-units/evolving-risk-factors-cancers-african-populations-erica-sa), Johannesburg Cancer study (Chen et al., 2020), and the African Esophageal Cancer Consortium (AfrECC) (Van Loon et al.,

2018). These studies are likely to provide a much clearer understanding of the genetic etiology of African ESCC. GWAS data can also be used to develop polygenic risk scores for stratifying disease risk (Fritsche et al., 2021).

Somatic Mutations in Tumor Tissue From African ESCC Patients

Genomic analysis of tumor tissues by DNA and RNA sequencing in large-scale global projects such as the Cancer Genome Atlas (TCGA) (The Cancer Genome Atlas Research Network, 2017) has led to a huge increase in knowledge of the genes and somatic mutations which drive tumor development. This in turn has fueled major progress in the development of cancer therapies that target the molecular pathways important for tumor development. No samples from African ESCC cases were included in the TCGA project. Information on the genomic profiles of African cancers is, however, very limited. In ESCC, whole-exome and RNA sequencing of tumors from 59 Malawian patients (Liu et al., 2016) observed similar genetic aberrations to those reported in Asian and North American patients and included mutations of well-established driver genes such as *TP53, CDKN2A, NFE2L2, CHEK2, NOTCH1, FAT1*, and *FBXW7*. Analyses also detected signatures associated with aging and cytidine deaminase (APOBEC) activity but, surprisingly, not of tobacco smoke.

A recent study (Moody et al., 2021) investigated mutational signatures in 552 ESCC patients from high-incidence regions (Iran, Kenya, Tanzania, China and Malawi) including three African countries, and low-incidence regions (Brazil, Japan and the UK) using whole-genome sequencing. Similar mutational profiles across all countries were found. Specific mutational signatures and ESCC risk factors were detected for tobacco, alcohol, opium and germline variants, and also highlighted APOBEC activation as an important step in tumor development. No evidence of an unknown exogenous mutagen associated with a mutational signature, which could explain ESCC variation in incidence, was found.

DISCUSSION

Benefits of Genomic Data in ESCC Research

Genomic studies provide critical information about the pathobiology of diseases, which improves our understanding of the risk and heritability of ESCC, risk prediction for populations and individuals, and contributes to cancer prevention (**Figure 1**). ESCC genomic research must follow rigorous guidelines to ensure reproducibility and reliability of results. Prioritization of ESCC genomic medicine in African populations will help elucidate the genetic etiology of ESCC, giving insights on variants, biological mechanisms, and the GxE interactions associated with ESCC development. This information will facilitate developing algorithms for predicting ESCC prognosis and survival of the patients.

Genomics can be used as a tool to address health disparities in cancer (Balogun and Olopade, 2021). There are differences in ESCC incidence and mortality between European and African populations. Whilst inadequate health care systems, lack of access, and poor quality of care have an obvious role, genomics can also be used to address these gaps in incidence and mortality. Genomic research can identify variants, pathways and biomarkers associated with increased risk and mortality, to be used in precision medicine (**Figure 1**). Ultimately, ESCC genomic research in African populations should not only contribute to understanding the etiology but also generate evidence that can be translated to prevention and therapeutics.

A recent study (Choudhury et al., 2020) explored human migration and the breadth of genomic diversity in 426 African individuals from 50 ethnolinguistic groups. Over three million previously undescribed single nucleotide variants were identified. The authors recommended "*broader characterization of the genomic diversity of African individuals to understand human ancestry and improve health*" (Choudhury et al., 2020). It remains to be determined if ancestral events, migration, the admixture of populations, and adaptation to exposures in the African ESCC corridor play a role in the demographic and geographic aspects of ESCC.

The Pan-Cancer Analysis of Whole Genomes Consortium performed the most comprehensive meta-analysis of cancer genomes to date, using 2,658 tumors and 38 tissues (Cieslik and Chinnaiyan, 2020). Of these, 98 were EAC tumors, but none were ESCC samples. The study assessed key aspects of cancer genomics, i.e., cancer drivers (Campbell et al., 2020), non-coding changes (Rheinbay et al., 2020), mutational signatures (Li et al., 2020), structural variants (Alexandrov et al., 2020), cancer evolution (Gerstung et al., 2020), and RNA alterations (Calabrese et al., 2020). However, none of the investigators were from Africa. Genomic analysis of substantial numbers of African ESCC patients together with detailed epidemiological data is needed to further explore the origins of African ESCC.

Barriers to ESCC Genomics Research in Africa

The lack of investment in genomic medicine in Africa has led to most genomic medicine knowledge being founded on genomes of European ancestry, despite African populations displaying higher levels of genetic diversity. A major barrier to implementing genomic research in Africa is inadequate infrastructure including poorly equipped facilities, erratic power supply, inadequate biotechnology and information technology infrastructure as well as the high cost of genomics tools and the costs associated with implementing genomic medicine (Adebamowo et al., 2018; Munung et al., 2018). These aspects further impact biospecimen collection, transportation and storage, which is pivotal in planning and conducting genomic studies. In Africa, the ethical aspects surrounding genomics research are also challenging and obtaining informed consent is compounded by language barriers and low literacy levels.

Recommendations for Future ESCC Studies and Furthering Genomics Research in Africa

African genomes harbor the most genetic diversity and variation, and yet are the least genetically characterized. This means that genetic variants of medical relevance remain unknown (Choudhury et al., 2020). For ESCC this impedes progress on applying genomics in understanding etiology, tailored screening and therapeutic interventions, promoting health equity and ultimately reducing the burden of ESCC in the African ESCC corridor. To fully understand the etiology of ESCC and provide tools for genomic medicine, ESCC research in Africa should: 1) follow a multidisciplinary approach to study interactions between genomic, environmental and lifestyle factors; 2) foster collaborations and data sharing to accelerate progress; 3) use standardized methods for analysis and reporting; 4) use larger study samples to adequately detect GxE interactions; and 5) control for population stratification in admixed populations. Additionally, openness and precision in reporting of methods are needed to improve reproducibility. In a cancer biology study aimed at replicating 193 experiments from 53 high-impact papers, none of the published studies had sufficient information for repeating the experiments (Errington et al., 2021). For genetic association studies, the STrengthening the REporting of Genetic Association studies (STREGA) statement (Little et al., 2009) should be used as a checklist to assess quality of reporting and methods.

Furthering genomic medicine in Africa requires leveraging existing infrastructure and learning from the extensive experience of current genomic medicine implementations in other countries (AESA, 2020). This requires significant infrastructure, including access to clinical facilities and high-throughput genotyping and sequencing facilities. Creating a unified ESCC genomics research hub in Africa will require standardized sample collection and well-managed biorepositories with the capacity to store and manage biospecimens. Strong information technology infrastructure is needed that is capable of managing, storing and analyzing big data (Wonkam and Mayosi, 2014). In addition, capacity building is needed to create a critical mass of bioinformaticians, and provide genomic medicine training programs for healthcare professionals.

CONCLUSION

Genomics is an invaluable approach in providing unbiased information about the pathogenesis of ESCC. The information could be used to predict risk, screen asymptomatic individuals, diagnose more accurately and develop targeted treatments. We are still far from being able to implement genomic medicine for ESCC in Africa since genomic information on African ESCC patients is very limited.

AUTHOR CONTRIBUTIONS

HS, HK, and GT conceptualized the idea for the "Opinion" manuscript. HS, HK, GT, VS, CGM, and WCC wrote sections of the manuscript. All authors reviewed and approved the final manuscript.

FUNDING

HS acknowledges the Beit Trust fund for providing a Doctoral Scholarship (https://beittrust.org.uk/beit-trust-scholarships), the Collaboration for Evidence-based Healthcare and Public Health in Africa (CEBHA+; https://www.cebha-plus.org/), as part of the Research Networks for Health Innovation in Sub-Saharan Africa, Funding Initiative of the German Federal Ministry of Education and Research, Margaret McNamara Education Grant (https://www.mmeg.org/), and the L'Oreal UNESCO for Women in Science Award (https://en.unesco.org/news/unesco-and-foundation-loreal-recognize-20-young-women-scientists-sub-saharan-africa). HK was supported by the Faculty of Medicine and Health Sciences, Stellenbosch University, South Africa. VS was supported by the African Cancer Institute and National Research Foundation (UID:132783). GT was supported by the South African Tuberculosis Bioinformatics Initiative (SATBBI), a Strategic Health Innovation Partnership grant from the South African Medical Research Council (https://www.samrc.ac.za/) and South African Department of Science and Innovation (https://www.dst.gov.za/). WCC and CGM were supported by the South African Medical Research Council (with funds received from the South African National Department of Health) and the UK Medical Research Council (with funds from the UK Government's Newton Fund) (MRC-RFA-SHIP 01-2015).

REFERENCES

Abnet, C. C., Arnold, M., and Wei, W.-Q. (2018). Epidemiology of Esophageal Squamous Cell Carcinoma. *Gastroenterology* 154 (2), 360–373. doi:10.1053/j.gastro.2017.08.023

Adebamowo, S. N., Francis, V., Tambo, E., Diallo, S. H., Landouré, G., Nembaware, V., et al. (2018). Implementation of Genomics Research in Africa: Challenges and Recommendations. *Glob. Health Action.* 11 (1), 1419033. doi:10.1080/16549716.2017.1419033

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

Alexandrov, L. B., Kim, J., Kim, J., Haradhvala, N. J., Huang, M. N., Tian Ng, A. W., et al. (2020). The Repertoire of Mutational Signatures in Human Cancer. *Nature* 578 (7793), 94–101. doi:10.1038/s41586-020-1943-3

Asombang, A. W., Chishinga, N., Nkhoma, A., Chipaila, J., Nsokolo, B., Manda-Mapalo, M., et al. (2019). Systematic Review of Esophageal Cancer in Africa: Epidemiology, Risk Factors, Management and Outcomes. *Wjg* 25 (31), 4512–4533. doi:10.3748/wjg.v25.i31.4512

Balogun, O. D., and Olopade, O. I. (2021). Addressing Health Disparities in Cancer with Genomics. *Nat. Rev. Genet.* 22 (10), 621–622. doi:10.1038/s41576-021-00390-4

Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A., and Jemal, A. (2018). Global Cancer Statistics 2018: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA: a Cancer J. clinicians* 68 (6), 394–424. doi:10.3322/caac.21492

Calabrese, C., Davidson, N. R., Davidson, N. R., Demircioğlu, D., Fonseca, N. A., He, Y., et al. (2020). Genomic Basis for RNA Alterations in Cancer. *Nature* 578 (7793), 129–136. doi:10.1038/s41586-020-1970-0

Campbell, P. J., Getz, G., Korbel, J. O., Stuart, J. M., Jennings, J. L., Stein, L. D., et al. (2020). Pan-cancer Analysis of Whole Genomes. *Nature* 578 (7793), 82–93. doi:10.1038/s41586-020-1969-6

Chen, W. C., Bye, H., Matejcic, M., Amar, A., Govender, D., Khew, Y. W., et al. (2019). Association of Genetic Variants in CHEK2 with Oesophageal Squamous Cell Carcinoma in the South African Black Population. *Carcinogenesis* 40 (4), 513–520. doi:10.1093/carcin/bgz026

Chen, W. C., Singh, E., Muchengeti, M., Bradshaw, D., Mathew, C. G., Babb de Villiers, C., et al. (2020). Johannesburg Cancer Study (JCS): Contribution to Knowledge and Opportunities Arising from 20 Years of Data Collection in an African Setting. *Cancer Epidemiol.* 65, 101701. doi:10.1016/j.canep.2020.101701

Chetwood, J. D., Garg, P., Finch, P., and Gordon, M. (2019). Systematic Review: the Etiology of Esophageal Squamous Cell Carcinoma in Low-Income Settings. *Expert Rev. Gastroenterol. Hepatol.* 13 (1), 71–88. doi:10.1080/17474124.2019.1543024

Choudhury, A., Aron, S., Botigué, L. R., Sengupta, D., Botha, G., Bensellak, T., et al. (2020). High-depth African Genomes Inform Human Migration and Health. *Nature* 586 (7831), 741–748. doi:10.1038/s41586-020-2859-7

Cieslik, M., and Chinnaiyan, A. M. (2020). Global Genomics Project Unravels Cancer's Complexity at Unprecedented Scale. *Nature* 578 (7793), 39–40. doi:10.1038/d41586-020-00213-2

Cui, R., Kamatani, Y., Takahashi, A., Usami, M., Hosono, N., Kawaguchi, T., et al. (2009). Functional Variants in ADH1B and ALDH2 Coupled with Alcohol and Smoking Synergistically Enhance Esophageal Cancer Risk. *Gastroenterology* 137 (5), 1768–1775. doi:10.1053/j.gastro.2009.07.070

Errington, T. M., Denis, A., Perfito, N., Iorns, E., and Nosek, B. A. (2021). Challenges for Assessing Replicability in Preclinical Cancer Biology. *Elife* 10. doi:10.7554/eLife.67995

Fritsche, L. G., Ma, Y., Zhang, D., Salvatore, M., Lee, S., Zhou, X., et al. (2021). On Cross-Ancestry Cancer Polygenic Risk Scores. *Plos Genet.* 17 (9), e1009670. doi:10.1371/journal.pgen.1009670

Gerstung, M., Jolly, C., Jolly, C., Leshchiner, I., Dentro, S. C., Gonzalez, S., et al. (2020). The Evolutionary History of 2,658 Cancers. *Nature* 578 (7793), 122–128. doi:10.1038/s41586-019-1907-7

Li, Y., Roberts, N. D., Roberts, N. D., Wala, J. A., Shapira, O., Schumacher, S. E., et al. (2020). Patterns of Somatic Structural Variation in Human Cancer Genomes. *Nature* 578 (7793), 112–121. doi:10.1038/s41586-019-1913-9

Little, J., Higgins, J. P. T., Ioannidis, J. P. A., Moher, D., Gagnon, F., von Elm, E., et al. (2009). Strengthening the Reporting of Genetic Association Studies (STREGA)-an Extension of the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement. *J. Clin. Epidemiol.* 62 (6), 597–608. e594. doi:10.1016/j.jclinepi.2008.12.004

Liu, W., Snell, J. M., Jeck, W. R., Hoadley, K. A., Wilkerson, M. D., Parker, J. S., et al. (2016). Subtyping Sub-saharan Esophageal Squamous Cell Carcinoma by Comprehensive Molecular Analysis. *JCI Insight* 1 (16), e88755. doi:10.1172/jci.insight.88755

Matejcic, M., and Iqbal Parker, M. (2015). Gene-environment Interactions in Esophageal Cancer. *Crit. Rev. Clin. Lab. Sci.* 52 (5), 211–231. doi:10.3109/10408363.2015.1020358

McCormack, V. A., Menya, D., Munishi, M. O., Dzamalala, C., Gasmelseed, N., Leon Roux, M., et al. (2017). Informing Etiologic Research Priorities for Squamous Cell Esophageal Cancer in Africa: A Review of Setting-specific Exposures to Known and Putative Risk Factors. *Int. J. Cancer* 140 (2), 259–271. doi:10.1002/ijc.30292

McKay, J. D., Truong, T., Gaborieau, V., Chabrier, A., Chuang, S. C., Byrnes, G., et al. (2011). A Genome-wide Association Study of Upper Aerodigestive Tract Cancers Conducted within the INHANCE Consortium. *Plos Genet.* 7 (3), e1001333. doi:10.1371/journal.pgen.1001333

Moody, S., Senkin, S., Islam, S. M. A., Wang, J., Nasrollahzadeh, D., Cortez Cardoso Penha, R., et al. (2021). Mutational Signatures in Esophageal Squamous Cell Carcinoma from Eight Countries with Varying Incidence. *Nat. Genet.* 53 (11), 1553–1563. doi:10.1038/s41588-021-00928-6

Munung, N. S., Mayosi, B. M., and de Vries, J. (2018). Genomics Research in Africa and its Impact on Global Health: Insights from African Researchers. *Glob. Health Epidemiol.* 3, e12. doi:10.1017/gheg.2018.3

National Cancer Registry (2019). *Cancer in South Africa*. Johannesburg: National Health Laboratory Services. (NCR).

Rheinbay, E., Nielsen, M. M., Nielsen, M. M., Abascal, F., Wala, J. A., Shapira, O., et al. (2020). Analyses of Non-coding Somatic Drivers in 2,658 Cancer Whole Genomes. *Nature* 578 (7793), 102–111. doi:10.1038/s41586-020-1965-x

Simba, H., Kuivaniemi, H., Lutje, V., Tromp, G., and Sewram, V. (2019). Systematic Review of Genetic Factors in the Etiology of Esophageal Squamous Cell Carcinoma in African Populations. *Front. Genet.* 10, 642. doi:10.3389/fgene.2019.00642

Simba, H. (2021). *The Role of Genetic and Environmental Factors in the Aetiology of Esophageal CancerPhD (Public Health) Stellenbosch University*. https://scholar.sun.ac.za/handle/10019.1/123653.

Somdyala, N. I. M., Parkin, D. M., Sithole, N., and Bradshaw, D. (2015). Trends in Cancer Incidence in Rural Eastern Cape Province; South Africa, 1998-2012. *Int. J. Cancer* 136 (5), E470–E474. doi:10.1002/ijc.29224

Talukdar, F. R., Ghosh, S. K., Laskar, R. S., and Mondal, R. (2013). Epigenetic, Genetic and Environmental Interactions in Esophageal Squamous Cell Carcinoma from Northeast India. *PLoS One* 8 (4), e60996. doi:10.1371/journal.pone.0060996

Talukdar, F. R., Soares Lima, S. C., Khoueiry, R., Laskar, R. S., Cuenin, C., Sorroche, B. P., et al. (2021). Genome-Wide DNA Methylation Profiling of Esophageal Squamous Cell Carcinoma from Global High-Incidence Regions Identifies Crucial Genes and Potential Cancer Markers. *Cancer Res.* 81 (10), 2612–2624. doi:10.1158/0008-5472.CAN-20-3445

The Cancer Genome Atlas Research Network (2017). Integrated Genomic Characterization of Oesophageal Carcinoma. *Nature* 541 (7636), 169–175. doi:10.1038/nature20805

Van Loon, K., Mwachiro, M. M., Abnet, C. C., Akoko, L., Assefa, M., Burgert, S. L., et al. (2018). The African Esophageal Cancer Consortium: A Call to Action. *Jgo*, 1–9. doi:10.1200/JGO.17.00163

Wang, L. D., Zhou, F. Y., Li, X. M., Sun, L. D., Song, X., Jin, Y., et al. (2010). Genome-wide Association Study of Esophageal Squamous Cell Carcinoma in Chinese Subjects Identifies Susceptibility Loci at PLCE1 and C20orf54. *Nat. Genet.* 42 (9), 759–763. doi:10.1038/ng.648

Wonkam, A., and Mayosi, B. M. (2014). Genomic Medicine in Africa: Promise, Problems and Prospects. *Genome Med.* 6 (2), 11. doi:10.1186/gm528

Wu, C., Wang, Z., Song, X., Feng, X. S., Abnet, C. C., He, J., et al. (2014). Joint Analysis of Three Genome-wide Association Studies of Esophageal Squamous Cell Carcinoma in Chinese Populations. *Nat. Genet.* 46 (9), 1001–1006. doi:10.1038/ng.3064

Wu, C., Kraft, P., Zhai, K., Chang, J., Wang, Z., Li, Y., et al. (2012). Genome-wide Association Analyses of Esophageal Squamous Cell Carcinoma in Chinese Identify Multiple Susceptibility Loci and Gene-Environment Interactions. *Nat. Genet.* 44 (10), 1090–1097. doi:10.1038/ng.2411

Wu, C., Li, D., Jia, W., Hu, Z., Zhou, Y., Yu, D., et al. (2013). Genome-wide Association Study Identifies Common Variants in SLC39A6 Associated with Length of Survival in Esophageal Squamous-Cell Carcinoma. *Nat. Genet.* 45 (6), 632–638. doi:10.1038/ng.2638

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewer FRT declared a shared affiliation with the author HS to the handling editor at the time of review.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Genetics

PERSPECTIVE
published: 05 April 2022
doi: 10.3389/fgene.2022.869610

Microbiomics: The Next Pillar of Precision Medicine and Its Role in African Healthcare

*C. K. Nkera-Gutabara[1,2,3], R. Kerr[3], J. Scholefield[2], S. Hazelhurst[1,4] and J. Naidoo[2]**

[1]Sydney Brenner Institute for Molecular Bioscience (SBIMB), Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa, [2]Bioengineering and Integrated Genomics Research Group, Council for Scientific and Industrial Research (CSIR), Pretoria, South Africa, [3]Division of Human Genetics, National Health Laboratory Service and School of Pathology, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa, [4]School of Electrical and Information Engineering, University of the Witwatersrand, Johannesburg, South Africa

Limited access to technologies that support early monitoring of disease risk and a poor understanding of the geographically unique biological and environmental factors underlying disease, represent significant barriers to improved health outcomes and precision medicine efforts in low to middle income countries. These challenges are further compounded by the rich genetic diversity harboured within Southern Africa thus necessitating alternative strategies for the prediction of disease risk and clinical outcomes in regions where accessibility to personalized healthcare remains limited. The human microbiome refers to the community of microorganisms (bacteria, archaea, fungi and viruses) that co-inhabit the human body. Perturbation of the natural balance of the gut microbiome has been associated with a number of human pathologies, and the microbiome has recently emerged as a critical determinant of drug pharmacokinetics and immunomodulation. The human microbiome should therefore not be omitted from any comprehensive effort towards stratified healthcare and would provide an invaluable and orthogonal approach to existing precision medicine strategies. Recent studies have highlighted the overarching effect of geography on gut microbial diversity as it relates to human health. Health insights from international microbiome datasets are however not yet verified in context of the vast geographical diversity that exists throughout the African continent. In this commentary we discuss microbiome research in Africa and its role in future precision medicine initiatives across the African continent.

Keywords: microbiome, precision medicine, gut health and disease, African health care, microbiome mapping

OPEN ACCESS

Edited by:
Dawn Stephens, Technology Innovation Agency (TIA), South Africa

Reviewed by:
Francesco Strati, European Institute of Oncology (IEO), Italy

****Correspondence:***
J. Naidoo
jnaidoo@csir.co.za

Specialty section:
This article was submitted to Human and Medical Genomics, a section of the journal Frontiers in Genetics

Received: *04 February 2022*
Accepted: *04 March 2022*
Published: *05 April 2022*

Citation:
Nkera-Gutabara CK, Kerr R, Scholefield J, Hazelhurst S and Naidoo J (2022) Microbiomics: The Next Pillar of Precision Medicine and Its Role in African Healthcare. Front. Genet. 13:869610. doi: 10.3389/fgene.2022.869610

THE HUMAN MICROBIOME AS A FOCAL POINT FOR PRECISION MEDICINE IN AFRICA

Genomic medicine catalysed a change in ideology around the significance of genetic diversity in the provision of healthcare, giving rise to precision medicine, which guides pharmaceutical intervention to improve clinical outcomes (McCarthy et al., 2013; Schork, 2015; Kuntz and Gilbert, 2017). While human genomics continues to remain central to most current precision strategies, there is also an increasing appreciation for the value of integrated multi-omics approaches in better understanding clinical diversity and developing a more comprehensive outlook towards precision medicine (Olivier et al., 2019).

The human microbiome refers to the collection of symbiotic, pathogenic and commensal microorganisms (e.g., bacteria, archaea, viruses and fungi) that cohabit discrete anatomical sites across the human body (Lederberg, 2000). The total number of human cells in the average body is surpassed by the number of microbes it houses, while the unique genetic information encoded by these microbial communities (~3 million genes) also dwarf that of the human genome (~23,000 genes) (Qin et al., 2010; Sender et al., 2016). Increasing accessibility to next generation sequencing (NGS) technologies have catalysed advances in microbiome research and illuminated its role in human health, prompting its moniker as our "second genome" (Grice and Segre, 2012; Cani, 2018; Tierney et al., 2021).

While the general composition of the gut microbiome remains quasi-stable and unique for an individual throughout adult life, it is also heavily shaped by factors like diet, lifestyle practices and environmental exposures (Tidjani Alou et al., 2016; Vangay et al., 2018). Notably, environmental factors associated with diet and lifestyle account for ~20% of gut microbiome variance where genetically unrelated individuals residing in close proximity are more likely to exhibit similar microbiome profiles than genetically related individuals who did not (Rothschild et al., 2018). In fact less than 2% of microbiome diversity can be explained by host genetics thus further strengthening the argument that the human microbiome may serve as a valuable complementary approach to traditional genomic medicine (Rothschild et al., 2018).

With continued industrialisation of human habitats and the widespread adoption of westernised lifestyles and diets, the human microbiome is being reshaped to reflect these changes (Rothschild et al., 2018; Vallès and Francino, 2018; Vangay et al., 2018). Over the last two decades African populations have experienced rapid transitions to an increasingly westernized nutrition and lifestyle, associated with an increased prevalence of non-communicable diseases (NCDs) like obesity, cancer and various neurological disorders (Abrahams et al., 2011). South Africa was recently ranked as the unhealthiest country in the world by the Indigo Wellness Index which is based on metrics including healthy life expectancy, obesity, depression and government expenditure on healthcare (Indigo Wellness Index, 2019). The added burden of infectious disease in addition to NCDs, and limited resources therefore necessitate innovative strategies to democratise access to precision healthcare on the continent.

Due to the dynamic nature of the microbiome, the timescale at which it reacts to environmental and health status changes is significantly shorter than the onset of the associated effects on health (Vallès and Francino, 2018; Vangay et al., 2018). Importantly microbiome profiling of an individual's gut is also significantly more cost effective (~10x) than sequencing an individual's genome. Unlike our base genetics, our "second genome" can also be readily and specifically modulated by a number of non-invasive, acute interventions like the use of probiotics, prebiotics, dietary inputs, lifestyle changes etc. (Dethlefsen and Relman, 2011; Maurice et al., 2013). This malleable property of the microbiome together with its strong valuable proposition as a biomarker, position it as an attractive module for precision medicine strategies (Tidjani Alou et al., 2016; Thursby and Juge, 2017). This is of particular relevance in the African continent where the use of the microbiome as both 1) a cost effective biomarker with the ability to predict early risk for multiple diseases and 2) a readily amenable target for non-invasive and cost effective therapeutic intervention (e.g probiotics) could prove highly impactful. In this commentary we discuss the human gut microbiome, its relevance to healthcare challenges faced by the African continent and the requirement for a coordinated and unified approach to large-scale microbiome research efforts across Africa.

LINKING THE HUMAN MICROBIOME TO AFRICAN HEALTH ISSUES

Dysbiosis, the perturbation of the natural balance of microbes within the human body, is associated with several human pathologies ranging from immune disorders such as asthma (Frati et al., 2018; Stokholm et al., 2018) and inflammatory bowel diseases (IBDs) (Xavier and Podolsky, 2007; Kostic et al., 2014) to metabolic disorders like obesity, malnutrition, diabetes and even gluten intolerance (Larsen et al., 2010; Musso et al., 2010; Nylund et al., 2014; Tidjani Alou et al., 2016).

Of particular relevance to the African continent, significant alterations to the gut microbiome have been noted in: infectious diseases like Human Immunodeficiency Virus (HIV) (reviewed by (Koay et al., 2018) and tuberculosis (TB) (reviewed by (Liu Y. et al., 2021); as well as other emerging health threats like Parkinson's disease (Scheperjans et al., 2015; Tremlett et al., 2017; Vogt et al., 2017); mental health (Bhattacharjee and Lukiw, 2013; Foster and McVey Neufeld, 2013; Kelly et al., 2016); cardiovascular disease (Feng et al., 2016; Jie et al., 2017) and cancer (reviewed by (Lee et al., 2021). While further studies are still required to validate causality versus consequence with regards to dysbiosis and certain diseases, there is clear potential for the microbiome as either a biomarker, therapeutic target or both for some of the major health challenges faced by the African continent, as reviewed below.

Obesity

The prevalence of obesity in African populations doubled between 1990 and 2016, to over 10 million individuals (World Health Organization, 2016; Biadgilign et al., 2017). Obesity has been directly linked to various NCDs including type II diabetes which has increased by over 120% in Africa since 1980, with a predicted economic consequence of $60 billion United State Dollars (USD) by 2030 in sub-Saharan Africa alone (Jaffar and Gill, 2017). Clinical and experimental studies in human cohorts have established correlations between gut microbiota and obesity, through baseline microbial composition having a direct impact on improvement in insulin sensitivity and halting progression of type I diabetes (Kootte et al., 2017; de Groot et al., 2021). Longitudinal studies in the United Kingdom (United Kingdom) and the United States have also shown that there is an increased risk of neurocognitive deficits and autism spectrum disorders (ASDs) associated with maternal obesity

(Basatemur et al., 2013; Pugh et al., 2015). Murine models revealed that maternal high-fat diets (mHFD) induced long-term cognitive deficits that span across several generations (Sarker and Peleg-Raibstein, 2018). This was further confirmed by recent evidence that maternal obesity disrupted the gut microbiome in offspring and that this was strongly associated with cognitive and social dysfunctions (Liu X. et al., 2021). Given the significant prevalence of obesity (>20%) of women above 20 years of age in South Africa (Lobstein and Brinsden, 2017), obesity could pose a serious long term challenge to health on the African continent.

Microbiome-linked therapies in response to maternal obesity-related deficits have provided promising interventions using high-fibre diet supplementation to produce microbiota-derived short-chain fatty acids that alleviate behavioural deficits via the gut-brain axis (Liu X. et al., 2021). Additionally mouse models of diet-induced obesity have demonstrated the bacterium, *A. muciniphila*'s, inverse correlation with obesity, where its presence correlated with the restoration of damaged gut structure, improved hepatic function and glucose homeostasis in high fat-diet-fed mice (Everard et al., 2013; de Vos, 2017; Yang et al., 2020). The value of early interventions against childhood immune disorders have also been successfully demonstrated with *Bifidobacterium infantis*, whereby stable and persistent colonization and remodelling of the infant intestinal microbiome was accomplished, resulting in reduced enteric inflammation (Grześkowiak et al., 2012; Huda et al., 2014; Frese et al., 2017; Henrick et al., 2019). The innovative utilisation of microbiome modulation strategies, like faecal microbiota transplantation and probiotic intervention along with microbiome profiling have great potential for the early diagnosis and treatment of metabolic disorders like obesity, prior to the onset of disease (Khanna et al., 2017; Depommier et al., 2019; Lam et al., 2019; Sorbara and Pamer, 2022).

Neurodegenerative Disorders and the Gut-Brain Axis

African populations have experienced an increase in life expectancy, and with an aging population comes the increased burden of age-related disorders (Dekker et al., 2020). Aging is accompanied by chronic inflammation, increased intestinal permeability, disrupted nutrient absorption and impaired digestion, all exhibiting bidirectional interactions with the gut microbiome, termed the gut-brain axis (An et al., 2018).

Studies of neurodegenerative diseases (ND) such as amyotrophic lateral sclerosis (ALS) and Parkinson's disease have revealed distinct disease associated gut microbial signatures in cohort studies along with evidence of microbial protection against disease progression caused by antibiotic-induced dysbiosis (Petrov et al., 2017; Peng et al., 2018; Blacher et al., 2019; Fang et al., 2020). Animal models have shown that microbiota transplantation and probiotic treatment have the potential to ameliorate ND-related pathology and behaviour (Sun et al., 2018, 2019; Kim et al., 2020). At least 1 in 6 people suffer from anxiety and depression in South Africa, affecting a staggering 41% of pregnant women—more than three times higher than in developing countries (SADAG, 2019). Murine model research has demonstrated that decreased gut microbiota richness and diversity is associated with depression and may play a causal role in its development, further highlighting the potential of the microbiome as a therapeutic target (Kelly et al., 2016).

Human Microbiome and SARS-CoV-2

One of the greatest global health challenges in recent times was driven by the coronavirus disease 2019 (COVID-19) pandemic. The severity of COVID-19 infection as well as the longevity of symptoms, termed "post-acute COVID-19 syndrome" (PACS), has been suggested to be heavily influenced by an individual's gut microbiome profile (Liu et al., 2021a). Bacteria such as *Firmicutes* play a role in the enhanced expression of angiotensin-converting enzyme 2 (ACE2) in the gut, which acts as a receptor for viral entry and cell infection (Zuo et al., 2020). The findings show that at admission to hospital, distinct gut microbial signatures are present within patients and are associated with PACS at 6 months post admission (Liu et al., 2021a).

The above observation was further supported by evidence that low levels of the gut bacteria *Collinsella* coincides with COVID-19 related mortality. This bacteria produces ursodeoxycholate, which is reported to have diverse beneficial effects such as the inhibition of SARS-CoV-2 binding to the ACE2 receptor, suppression of inflammatory responses, and amelioration of acute respiratory distress syndrome in COVID-19 patients (Hirayama et al., 2021). This supports the potential for precision microbiome-based profiling, as an early warning system for the detection of PACS, as well as a target for the prevention of COVID-19 infection (Liu et al., 2021a; Hirayama et al., 2021; Yeoh et al., 2021).

The antiviral properties of the microbiome have been investigated as an alternative therapeutic approach against COVID-19 (Campbell et al., 2021; Chevallier et al., 2021; Kuzmina et al., 2021). Bacterial metabolites are able to inhibit SARS-CoV-2 infection (Piscotta et al., 2021). These metabolites were shown to exhibit structural and functional similarities to synthetic drugs examined for use against the virus, most especially N6-(D2-isopentenyl) adenosine (IPA), which exhibited greater potency and inhibitory effects than interventions such as remdesivir, fluvoxamine and favipiravir (Piscotta et al., 2021). These findings further strengthen the valuable proposition for the inclusion of the microbiome in future precision medicine efforts so that we may unlock its full potential in response to threats like COVID-19.

Pharmacomicrobiomics and Adverse Drug Reactions

Adverse drug reactions (ADRs) refer to unintended consequences that arise from therapeutic agents administered at normal dosages and can include both inefficacy and harmful side effects. ADRs represent a significant source of patient noncompliance and clinical failure, hence a reduction in ADRs is an important goal of precision medicine (Edwards and Aronson, 2000). "Pharmacomicrobiomics" has emerged as the

study of drug-microbiome interactions, or how the variations in microbial profiles can affect a drugs' pharmacokinetics, pharmacodynamics and ultimately efficacy (Rizkallah et al., 2010; ElRakaiby et al., 2014). Evidence exists for bacterial drug metabolism as a general mechanism through which the microbiome in the gut, reproductive tracts and even disease tissue alters drug response (Zimmermann et al., 2019). In turn human-targeted (non-antibiotic) drugs have antibiotic-like side effects on gut bacteria in human cohorts, where 27% of the non-antibiotics studied inhibited the growth of at least one bacterial species (Maier et al., 2018). Over 115 drug-microbiome interactions have been documented in the PharmacoMicrobiomics database (Aziz et al., 2011). Drug-microbiome interactions modulate the bioavailability of drugs, which is important to consider for the appropriate dosing in precision medicine (Kuntz and Gilbert, 2017). Maier et al. (2018) highlight how continued pharmacomicrobiomics research aids in the achievement of a drug-microbe network that could guide drug development in line with precision medicine therapies. Much like pharmacogenomics revolutionised healthcare and led to the birth of precision medicine, pharmacomicrobiomics is quickly positioning itself as another pillar of personalized treatment (Saad et al., 2012; Maurice et al., 2013; ElRakaiby et al., 2014).

MICROBIOME RESEARCH EFFORTS GLOBALLY AND ACROSS AFRICA

Geography has been purported as an important determinant of microbiome composition where geography itself may represent the confluence of multiple variables like regional dietary preferences, unique environmental exposures and ethnically diverse lifestyle practices. Li and colleagues (2014) noted geospatial clustering of gut microbial signatures within 1,200 samples of European and Asian origin while He et al. (2018) found location to be the most significant driver of gut microbial diversity across more than 7,000 participants, with common ancestry and cultural practices, within a single Chinese province. These findings would suggest that the extrapolation of complex microbiome research trends, like the characterisation of healthy and diseased states, may be confounded by geographical diversity especially in low to middle income countries (LMICs) which exhibit increased diversity and unique phyla in comparison to samples from North America and Europe (Porras and Brito, 2019). A number of large-scale efforts have thus sought to characterise microbiome diversity at various regions across the globe in order to better understand microbiome-health dynamics within specific geographical constraints.

Large scale consortium studies in North America, Europe and Asia have driven advancements in microbiome research globally (Turnbaugh et al., 2007; Ehrlich, 2011; McDonald et al., 2018; Proctor et al., 2019). The Human Microbiome Project (HMP) was a pioneering initiative that launched in the United States in 2007 with around 300 participants (Turnbaugh et al., 2007). This project made significant strides in providing communities and researchers with a wealth of data, analytical and bio-specimen resources, whilst also positioning other researchers to continue work on microbiome-linked conditions: such as pregnancy and preterm birth; IBDs; and stressors affecting pre-diabetic individuals (Fettweis et al., 2019; NIH, 2019; Proctor et al., 2019).

As a result of the ground-breaking discoveries from the HMP, the value of continued explorations to uncover the full potential of the microbiome led to other large scale projects such as: the Asian Microbiome Project launched in 2009, the American Gut Project (AGP) in 2012, the British Gut Project in 2014, and the Dutch microbiome project (2016). These projects are collaborative efforts between the Earth Microbiome Project and Human Food Project, in order to discover global microbial taxonomic and functional diversity "in the wild" and across human populations (McDonald et al., 2018). The AGP initiative used a "self-selected citizen-scientist" cohort of over 10,000 participants, collecting microbial sequence data from over 15,000 stool samples. This unprecedented scale of characterizing the human microbiome is continuously revealing novel relationships between health, lifestyle and dietary factors (Nakayama et al., 2015; McDonald et al., 2018). The momentum from these large-scale global efforts could see clinical microbiome profiling become routine practice at point of care in the developed world within the next 5 years (Harrison-Dunn, 2021). However, the lack of representation of LMICs and Africa as a whole in these early large-scale efforts would suggest a longer horizon for the implementation of such practices in these regions in the absence of proactive measures to drive microbiome research in LMICs, like those recently adopted by both China and India (Aziz et al., 2018; Porras and Brito, 2019).

Allali and colleagues (2021) compiled an illuminating systematic review of published, data-driven research articles that included human microbiome samples of African origin and the use of NGS technologies. Only 168 research articles met these criteria with ~67% of these being published from 2017 onwards. More than 60% of countries on the African continent were represented in terms of sampling where SA, Uganda and Kenya collectively accounted for more than 40% of all studies. Intriguingly ~80% of all studies using African microbiome samples have not been led by African research institutes or African researchers. Less than 5% of studies confirmed the recruitment of urban populations while 31% of all studies focused on recruitment in rural settings. Notably less than 4% of studies included more than 500 participants with only ~1% of studies including more than 1,000 participants (Allali et al., 2021). These findings highlight a clear need for coordinated intra-continental, African researcher-driven microbiome initiatives that focus on addressing the future health needs of the continent. Interestingly, they also reveal gaps in current African microbiome research efforts like the under-representation of well-developed African urban environments in sampling, and the current scarcity of large-scale microbiome profiling initiatives represented in publications.

A number of African-directed initiatives have recently been launched to investigate the role of the microbiome in human health and disease on the African continent. These studies are however traditionally limited in terms of either their scale or scope when compared to large scale microbiome profiling efforts

like the AGP (Allali et al., 2021). The South African Microbiome Initiative in Neuroscience, based at Stellenbosch University in SA, is a pioneering African microbiome profiling initiative (saNeuroGut, 2017). The project aims to specifically investigate the gut-brain axis and thus the role of the gut microbiome in South African neurocognitive health, particularly links to post traumatic stress disorder.

The Human Heredity and Health in Africa (H3Africa) initiative was initially established to enhance the research capacity in institutions across Africa, for establishing projects which support precision medicine for Africans (H3Africa Consortium et al., 2014). Recognising the value of the interplay between microbes and human health has led to several ongoing microbiome projects facilitated by H3Africa (Mulder et al., 2016; Mulder et al.,2018). The details of ongoing and prospective microbiome projects are outlined on the H3Africa web-interface (H3Africa, 2010) and generally involve investigating the role of the microbiome within the context of specific diseases in women and children. This includes initiatives such as the African Collaborative Center for Microbiome and Genomics Research (ACCME) based in Nigeria, studying the association between human papillomavirus (HPV) infection and the vaginal microbiome in HIV-negative and HIV-positive African women (Adebamowo et al., 2017). Also, the Respiratory Microbiota of African Children (ReMAC) based at the University of Cape Town in SA.

The South African Council for Scientific and Industrial research (CSIR) in collaboration with the Sydney Brenner Institute for Molecular Bioscience at the University of the Witwatersrand have recently (2021) launched the CSIR Microbiome Mapping Initiative (CMMI) which aims to combine machine learning, 3rd generation sequencing technologies, environmental modelling and bioinformatics to better understand the relationships between gut health and over 100 lifestyle, health and environmental factors in South Africa by recruiting ~500 participants based in the metros of Johannesburg and Pretoria (CMMI, 2021). It is our hope that the CMMI project will provide a framework for the coordination of larger national microbiome profiling efforts that could ultimately adopt a citizen science funding model akin to the AGP that would support its sustainability and also promote public awareness around microbiome research and its significance to the future of human health.

FUTURE PERSPECTIVES AND LESSONS FROM OUR PAST

Publication of the first complete human genome sequence (Venter et al., 2001) launched genomics into the 21st century, and with it the promise of genomic medicine. Precision medicine became a reality by extracting unique sequence-specific variants which deviated from the "baseline" of the initial sequences published—a baseline extracted from just five samples of European origin. However with accessibility to more diverse datasets, it became clear that the human genome harboured significant inter-individual and inter-ethnic diversity which would in turn have profound effects on disease risk and clinical outcomes. This is best exemplified by the initial 1,000 Genomes Project revealing that African genomes contain 25% more genetic variants than any other ethnic group (Gurdasani et al., 2015). Despite this revelation less than 2% of all currently analysed genomic datasets are of African origin (Sirugo et al., 2019). Thus, while precision medicine has been rapidly making improvements to the health space, much of the most advanced interventions are still based on well-established genomic profiles of Caucasian ancestry which do not adequately reflect global diversity, especially African diversity (Sirugo et al., 2019). The consequence of which is that under-researched genetic variants with higher prevalence in developing countries have significantly contributed to undesirable outcomes and perpetuated historical inequalities in healthcare across the African continent. With the microbiome positioned as the "second genome", it also represents a second chance for African researchers to take stewardship of the continent's healthcare narrative and prioritise the development of technologies and datasets that drive the African health agenda.

This will require a concerted investment from African researchers, funders and policy makers alike to avoid perpetuating the historical marginalisation of our continent and its unique healthcare needs within the evolving global healthcare arena. An important first step towards this goal would be the establishment of an African microbiome mapping initiative that embodies the "One Continent: One vision" ethos of this special issue. In contrast to the position the continent found itself in at the start of the genomic medicine revolution decades ago—improved access to NGS technologies and infrastructure along with established expertise in bioinformatics and microbiomics across numerous African research institutes—the continent is currently well poised to successfully meet this challenge. We invite all who share this vision to build the future of African precision medicine with us, pamoja!

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

JN and CKN-G conceptualised the paper. CKN-G, JN, JS, RK and SH contributed to the compilation of the manuscript.

FUNDING

Funding provided by the Postgraduate Merit Award from the University of the Witwatersrand, the Council for Scientific and Industrial Research (CSIR) and the National Research Foundation (NRF) Thuthuka Funding Instrument (TTK190208414825; Grant No:121949).

REFERENCES

Abrahams, Z., Mchiza, Z., and Steyn, N. P. (2011). Diet and Mortality Rates in Sub-saharan Africa: Stages in the Nutrition Transition. *BMC Public Health* 11, 801. doi:10.1186/1471-2458-11-801

Adebamowo, S. N., Dareng, E. O., Famooto, A. O., Offiong, R., Olaniyan, O., Obende, K., et al. (2017). Cohort Profile: African Collaborative Center for Microbiome and Genomics Research's (ACCME's) Human Papillomavirus (HPV) and Cervical Cancer Study. *Int. J. Epidemiol.* 46, 1745–1745j. doi:10.1093/ije/dyx050

Allali, I., Abotsi, R. E., Tow, L. A., Thabane, L., Zar, H. J., Mulder, N. M., et al. (2021). Human Microbiota Research in Africa: a Systematic Review Reveals Gaps and Priorities for Future Research. *Microbiome* 9, 241. doi:10.1186/s40168-021-01195-7

An, R., Wilms, E., Masclee, A. A. M., Smidt, H., Zoetendal, E. G., and Jonkers, D. (2018). Age-dependent Changes in GI Physiology and Microbiota: Time to Reconsider? *Gut* 67, 2213–2222. doi:10.1136/gutjnl-2017-315542

Aziz, R. K., Hegazy, S. M., Yasser, R., Rizkallah, M. R., and ElRakaiby, M. T. (2018). Drug Pharmacomicrobiomics and Toxicomicrobiomics: from Scattered Reports to Systematic Studies of Drug-Microbiome Interactions. *Expert Opin. Drug Metab. Toxicol.* 14, 1043–1055. doi:10.1080/17425255.2018.1530216

Aziz, R. K., Saad, R., and Rizkallah, M. R. (2011). PharmacoMicrobiomics or How Bugs Modulate Drugs: an Educational Initiative to Explore the Effects of Human Microbiome on Drugs. *BMC Bioinformatics* 12, A10. doi:10.1186/1471-2105-12-S7-A10

Basatemur, E., Gardiner, J., Williams, C., Melhuish, E., Barnes, J., and Sutcliffe, A. (2013). Maternal Prepregnancy BMI and Child Cognition: a Longitudinal Cohort Study. *Pediatrics* 131, 56–63. doi:10.1542/peds.2012-0788

Bhattacharjee, S., and Lukiw, W. J. (2013). Alzheimer's Disease and the Microbiome. *Front. Cel. Neurosci.* 7, 153. doi:10.3389/fncel.2013.00153

Biadgilign, S., Mgutshini, T., Haile, D., Gebremichael, B., Moges, Y., and Tilahun, K. (2017). Epidemiology of Obesity and Overweight in Sub-saharan Africa: a Protocol for a Systematic Review and Meta-Analysis. *BMJ Open* 7, e017666. doi:10.1136/bmjopen-2017-017666

Blacher, E., Bashiardes, S., Shapiro, H., Rothschild, D., Mor, U., Dori-Bachash, M., et al. (2019). Potential Roles of Gut Microbiome and Metabolites in Modulating ALS in Mice. *Nature* 572, 474–480. doi:10.1038/s41586-019-1443-5

Campbell, F., Archer, B., Laurenson-Schafer, H., Jinnai, Y., Konings, F., Batra, N., et al. (2021). Increased Transmissibility and Global Spread of SARS-CoV-2 Variants of Concern as at June 2021. *Eurosurveillance* 26. doi:10.2807/1560-7917.ES.2021.26.24.2100509

Cani, P. D. (2018). Human Gut Microbiome: Hopes, Threats and Promises. *Gut* 67, 1716–1725. doi:10.1136/gutjnl-2018-316723

Chevallier, C., Hacquin, A.-S., and Mercier, H. (2021). COVID-19 Vaccine Hesitancy: Shortening the Last Mile. *Trends Cogn. Sci.* 25, 331–333. doi:10.1016/j.tics.2021.02.002

CMMI (2021). Available at: https://cmmi.csir.co.za/(Accessed February 4, 2022).

de Groot, P., Nikolic, T., Pellegrini, S., Sordi, V., Imangaliyev, S., Rampanelli, E., et al. (2021). Faecal Microbiota Transplantation Halts Progression of Human New-Onset Type 1 Diabetes in a Randomised Controlled Trial. *Gut* 70, 92–105. doi:10.1136/gutjnl-2020-322630

de Vos, W. M. (2017). Microbe Profile: Akkermansia Muciniphila: a Conserved Intestinal Symbiont that Acts as the Gatekeeper of Our Mucosa. *Microbiol. Read. Engl.* 163, 646–648. doi:10.1099/mic.0.000444

Dekker, M. C. J., Coulibaly, T., Bardien, S., Ross, O. A., Carr, J., and Komolafe, M. (2020). Parkinson's Disease Research on the African Continent: Obstacles and Opportunities. *Front. Neurol.* 11, 512. doi:10.3389/fneur.2020.00512

Depommier, C., Everard, A., Druart, C., Plovier, H., Van Hul, M., Vieira-Silva, S., et al. (2019). Supplementation with Akkermansia Muciniphila in Overweight and Obese Human Volunteers: a Proof-Of-Concept Exploratory Study. *Nat. Med.* 25, 1096–1103. doi:10.1038/s41591-019-0495-2

Dethlefsen, L., and Relman, D. A. (2011). Incomplete Recovery and Individualized Responses of the Human Distal Gut Microbiota to Repeated Antibiotic Perturbation. *Proc. Natl. Acad. Sci. U.S.A.* 108, 4554–4561. doi:10.1073/pnas.1000087107

Edwards, I. R., and Aronson, J. K. (2000). Adverse Drug Reactions: Definitions, Diagnosis, and Management. *The Lancet* 356, 1255–1259. doi:10.1016/S0140-6736(00)02799-9

Ehrlich, S. D. (2011). "MetaHIT: The European Union Project on Metagenomics of the Human Intestinal Tract," in *Metagenomics of the Human Body*. Editor K. E. Nelson (New York, NY: Springer), 307–316. doi:10.1007/978-1-4419-7089-3_15

ElRakaiby, M., Dutilh, B. E., Rizkallah, M. R., Boleij, A., Cole, J. N., and Aziz, R. K. (2014). Pharmacomicrobiomics: The Impact of Human Microbiome Variations on Systems Pharmacology and Personalized Therapeutics. *OMICS: A J. Integr. Biol.* 18, 402–414. doi:10.1089/omi.2014.0018

Everard, A., Belzer, C., Geurts, L., Ouwerkerk, J. P., Druart, C., Bindels, L. B., et al. (2013). Cross-talk between Akkermansia Muciniphila and Intestinal Epithelium Controls Diet-Induced Obesity. *Proc. Natl. Acad. Sci. U.S.A.* 110, 9066–9071. doi:10.1073/pnas.1219451110

Fang, P., Kazmi, S. A., Jameson, K. G., and Hsiao, E. Y. (2020). The Microbiome as a Modifier of Neurodegenerative Disease Risk. *Cell Host & Microbe* 28, 201–222. doi:10.1016/j.chom.2020.06.008

Feng, Q., Liu, Z., Zhong, S., Li, R., Xia, H., Jie, Z., et al. (2016). Integrated Metabolomics and Metagenomics Analysis of Plasma and Urine Identified Microbial Metabolites Associated with Coronary Heart Disease. *Sci. Rep.* 6, 22525. doi:10.1038/srep22525

Foster, J. A., and McVey Neufeld, K.-A. (2013). Gut-brain axis: How the Microbiome Influences Anxiety and Depression. *Trends Neurosciences* 36, 305–312. doi:10.1016/j.tins.2013.01.005

Frati, F., Salvatori, C., Incorvaia, C., Bellucci, A., Di Cara, G., Marcucci, F., et al. (2018). The Role of the Microbiome in Asthma: The Gut-Lung Axis. *Ijms* 20, 123. doi:10.3390/ijms20010123

Frese, S. A., Hutton, A. A., Contreras, L. N., Shaw, C. A., Palumbo, M. C., Casaburi, G., et al. (2017). Persistence of Supplemented Bifidobacterium Longum Subsp. Infantis EVC001 in Breastfed Infants. *mSphere* 2, e00501–17. doi:10.1128/mSphere.00501-17

Grice, E. A., and Segre, J. A. (2012). The Human Microbiome: Our Second Genome. *Annu. Rev. Genom. Hum. Genet.* 13, 151–170. doi:10.1146/annurev-genom-090711-163814

Grześkowiak, Ł., Collado, M. C., Mangani, C., Maleta, K., Laitinen, K., Ashorn, P., et al. (2012). Distinct Gut Microbiota in southeastern African and Northern European Infants. *J. Pediatr. Gastroenterol. Nutr.* 54, 812–816. doi:10.1097/MPG.0b013e318249039c

Gurdasani, D., Carstensen, T., Tekola-Ayele, F., Pagani, L., Tachmazidou, I., Hatzikotoulas, K., et al. (2015). The African Genome Variation Project Shapes Medical Genetics in Africa. *Nature* 517, 327–332. doi:10.1038/nature13997

H3Africa Consortium, Rotimi, C., Matovu, E., Bucheton, B., Chisi, J., Enyaru, J., Hertz-Fowler, C., et al. (2014). Enabling the Genomic Revolution in Africa. *Science* 344, 1346–1348. doi:10.1126/science.1251546

H3Africa (2010). H3Africa. Available at: https://h3africa.org/index.php/about/(Accessed February 4, 2022).

Harrison-Dunn, A. (2021). Five Years This Will Be Routine": Map My Gut Start-Up Offers Personalised Microbiome Advice. nutraingredients.com. Available at: https://www.nutraingredients.com/Article/2016/11/22/In-five-years-this-will-be-routine-Map-My-Gut-start-up-offers-personalised-microbiome-advice (Accessed January 28, 2022).

Henrick, B. M., Chew, S., Casaburi, G., Brown, H. K., Frese, S. A., Zhou, Y., et al. (2019). Colonization by B. Infantis EVC001 Modulates Enteric Inflammation in Exclusively Breastfed Infants. *Pediatr. Res.* 86, 749–757. doi:10.1038/s41390-019-0533-2

Hirayama, M., Nishiwaki, H., Hamaguchi, T., Ito, M., Ueyama, J., Maeda, T., et al. (2021). Intestinal Collinsella May Mitigate Infection and Exacerbation of COVID-19 by Producing Ursodeoxycholate. *PLOS ONE* 16, e0260451. doi:10.1371/journal.pone.0260451

Huda, M. N., Lewis, Z., Kalanetra, K. M., Rashid, M., Ahmad, S. M., Raqib, R., et al. (2014). Stool Microbiota and Vaccine Responses of Infants. *Pediatrics* 134, e362–e372. doi:10.1542/peds.2013-3937

Indigo Wellness Index (2019). BusinessInsider. Available at: https://www.businessinsider.co.za/most-unhealthy-countries-in-the-world-ranked-2019-3 (Accessed February 4, 2022).

Jaffar, S., and Gill, G. (2017). The Crisis of Diabetes in Sub-saharan Africa. *Lancet Diabetes Endocrinol.* 5, 574–575. doi:10.1016/S2213-8587(17)30219-X

Jie, Z., Xia, H., Zhong, S.-L., Feng, Q., Li, S., Liang, S., et al. (2017). The Gut Microbiome in Atherosclerotic Cardiovascular Disease. *Nat. Commun.* 8, 845. doi:10.1038/s41467-017-00900-1

Kelly, J. R., Borre, Y., O' Brien, C., Patterson, E., El Aidy, S., Deane, J., et al. (2016). Transferring the Blues: Depression-Associated Gut Microbiota Induces Neurobehavioural Changes in the Rat. *J. Psychiatr. Res.* 82, 109–118. doi:10.1016/j.jpsychires.2016.07.019

Khanna, S., Vazquez-Baeza, Y., González, A., Weiss, S., Schmidt, B., Muñiz-Pedrogo, D. A., et al. (2017). Changes in Microbial Ecology after Fecal Microbiota Transplantation for Recurrent *C. difficile* Infection Affected by Underlying Inflammatory Bowel Disease. *Microbiome* 5, 55. doi:10.1186/s40168-017-0269-3

Kim, M.-S., Kim, Y., Choi, H., Kim, W., Park, S., Lee, D., et al. (2020). Transfer of a Healthy Microbiota Reduces Amyloid and Tau Pathology in an Alzheimer's Disease Animal Model. *Gut* 69, 283–294. doi:10.1136/gutjnl-2018-317431

Koay, W. L. A., Siems, L. V., and Persaud, D. (2018). The Microbiome and HIV Persistence. *Curr. Opin. HIV AIDS* 13, 61–68. doi:10.1097/COH.0000000000000434

Kootte, R. S., Levin, E., Salojärvi, J., Smits, L. P., Hartstra, A. V., Udayappan, S. D., et al. (2017). Improvement of Insulin Sensitivity after Lean Donor Feces in Metabolic Syndrome Is Driven by Baseline Intestinal Microbiota Composition. *Cel Metab.* 26, 611–619.e6. doi:10.1016/j.cmet.2017.09.008

Kostic, A. D., Xavier, R. J., and Gevers, D. (2014). The Microbiome in Inflammatory Bowel Disease: Current Status and the Future Ahead. *Gastroenterology* 146, 1489–1499. doi:10.1053/j.gastro.2014.02.009

Kuntz, T. M., and Gilbert, J. A. (2017). Introducing the Microbiome into Precision Medicine. *Trends Pharmacol. Sci.* 38, 81–91. doi:10.1016/j.tips.2016.10.001

Kuzmina, A., Khalaila, Y., Voloshin, O., Keren-Naus, A., Boehm-Cohen, L., Raviv, Y., et al. (2021). SARS-CoV-2 Spike Variants Exhibit Differential Infectivity and Neutralization Resistance to Convalescent or post-vaccination Sera. *Cell Host & Microbe* 29, 522–528.e2. doi:10.1016/j.chom.2021.03.008

Lam, K. N., Alexander, M., and Turnbaugh, P. J. (2019). Precision Medicine Goes Microscopic: Engineering the Microbiome to Improve Drug Outcomes. *Cell Host & Microbe* 26, 22–34. doi:10.1016/j.chom.2019.06.011

Larsen, N., Vogensen, F. K., van den Berg, F. W. J., Nielsen, D. S., Andreasen, A. S., Pedersen, B. K., et al. (2010). Gut Microbiota in Human Adults with Type 2 Diabetes Differs from Non-diabetic Adults. *PLOS ONE* 5, e9085. doi:10.1371/journal.pone.0009085

Lederberg, J. (2000). Infectious History. *Science* 288, 287–293. doi:10.1126/science.288.5464.287

Lee, K. A., Luong, M. K., Shaw, H., Nathan, P., Bataille, V., and Spector, T. D. (2021). The Gut Microbiome: what the Oncologist Ought to Know. *Br. J. Cancer* 125, 1197–1209. doi:10.1038/s41416-021-01467-x

Liu, Q., Mak, J. W. Y., Su, Q., Yeoh, Y. K., Lui, G. C.-Y., Ng, S. S. S., et al. (2022a). Gut Microbiota Dynamics in a Prospective Cohort of Patients with post-acute COVID-19 Syndrome. *Gut* 71, 544–552. doi:10.1136/gutjnl-2021-325989

Liu, X., Li, X., Xia, B., Jin, X., Zou, Q., Zeng, Z., et al. (2021b). High-fiber Diet Mitigates Maternal Obesity-Induced Cognitive and Social Dysfunction in the Offspring via Gut-Brain axis. *Cel Metab.* 33, 923–938. e6. doi:10.1016/j.cmet.2021.02.002

Liu, Y., Wang, J., and Wu, C. (2021c). Microbiota and Tuberculosis: A Potential Role of Probiotics, and Postbiotics. *Front. Nutr.* 8. Available at: https://www.frontiersin.org/article/10.3389/fnut.2021.626254 (Accessed February 4, 2022). doi:10.3389/fnut.2021.626254

Lobstein, C. T., and Brinsden, H. (2017). We Are Grateful for the Compilers of Online Databases and Information on Obesity Especially the World Health Organization's Global Health Observatory, the NCD Risk Factor Collaboration Based at Imperial College London and OECD. Please see the Data Sources section of this report in Appendix 1 for more details. 242.

Maier, L., Pruteanu, M., Kuhn, M., Zeller, G., Telzerow, A., Anderson, E. E., et al. (2018). Extensive Impact of Non-antibiotic Drugs on Human Gut Bacteria. *Nature* 555, 623–628. doi:10.1038/nature25979

Maurice, C. F., Haiser, H. J., and Turnbaugh, P. J. (2013). Xenobiotics Shape the Physiology and Gene Expression of the Active Human Gut Microbiome. *Cell* 152, 39–50. doi:10.1016/j.cell.2012.10.052

McCarthy, J. J., McLeod, H. L., and Ginsburg, G. S. (2013). Genomic Medicine: a Decade of Successes, Challenges, and Opportunities. *Sci. Transl. Med.* 5, 189sr4. doi:10.1126/scitranslmed.3005785

McDonald, D., Hyde, E., Debelius, J. W., Morton, J. T., Gonzalez, A., Ackermann, G., et al. (2018). American Gut: an Open Platform for Citizen Science Microbiome Research. *mSystems* 3, e00031–18. doi:10.1128/mSystems.00031-18

Mulder, N., Abimiku, A. l., Adebamowo, S. N., de Vries, J., Matimba, A., Olowoyo, P., et al. (2018). H3Africa: Current Perspectives. *Pgpm* Vol. 11, 59–66. doi:10.2147/PGPM.S141546

Mulder, N. J., Adebiyi, E., Alami, R., Benkahla, A., Brandful, J., Doumbia, S., et al. (2016). H3ABioNet, a Sustainable Pan-African Bioinformatics Network for Human Heredity and Health in Africa. *Genome Res.* 26, 271–277. doi:10.1101/gr.196295.115

Musso, G., Gambino, R., and Cassader, M. (2010). Obesity, Diabetes, and Gut Microbiota. *Diabetes Care* 33, 2277–2284. doi:10.2337/dc10-0556

Nakayama, J., Watanabe, K., Jiang, J., Matsuda, K., Chao, S.-H., Haryono, P., et al. (2015). Diversity in Gut Bacterial Community of School-Age Children in Asia. *Sci. Rep.* 5. doi:10.1038/srep08397

Nylund, L., Nermes, M., Isolauri, E., Salminen, S., De Vos, W. M., Satokari, R., et al. (2015). Severity of Atopic Disease Inversely Correlates with Intestinal Microbiota Diversity and Butyrate-Producing Bacteria. *Allergy* 70, 241–244. doi:10.1111/all.12549

Olivier, M., Asmis, R., Hawkins, G. A., Howard, T. D., and Cox, L. A. (2019). The Need for Multi-Omics Biomarker Signatures in Precision Medicine. *Ijms* 20, 4781. doi:10.3390/ijms20194781

Peng, C., Gathagan, R. J., Covell, D. J., Medellin, C., Stieber, A., Robinson, J. L., et al. (2018). Cellular Milieu Imparts Distinct Pathological α-synuclein Strains in α-synucleinopathies. *Nature* 557, 558–563. doi:10.1038/s41586-018-0104-4

Petrov, V. A., Saltykova, I. V., Zhukova, I. A., Alifirova, V. M., Zhukova, N. G., Dorofeeva, Y. B., et al. (2017). Analysis of Gut Microbiota in Patients with Parkinson's Disease. *Bull. Exp. Biol. Med.* 162, 734–737. doi:10.1007/s10517-017-3700-7

Piscotta, F. J., Hoffmann, H.-H., Choi, Y. J., Small, G. I., Ashbrook, A. W., Koirala, B., et al. (2021). Metabolites with SARS-CoV-2 Inhibitory Activity Identified from Human Microbiome Commensals. *mSphere* 6. doi:10.1128/mSphere.00711-21

Porras, A. M., and Brito, I. L. (2019). The Internationalization of Human Microbiome Research. *Curr. Opin. Microbiol.* 50, 50–55. doi:10.1016/j.mib.2019.09.012

Proctor, L. M., Creasy, H. H., Fettweis, J. M., Lloyd-Price, J., Mahurkar, A., Zhou, W., et al. (2019). The Integrative Human Microbiome Project. *Nature* 569, 641–648. doi:10.1038/s41586-019-1238-8

Pugh, S. J., Richardson, G. A., Hutcheon, J. A., Himes, K. P., Brooks, M. M., Day, N. L., et al. (2015). Maternal Obesity and Excessive Gestational Weight Gain Are Associated with Components of Child Cognition. *J. Nutr.* 145, 2562–2569. doi:10.3945/jn.115.215525

Qin, J., Li, R., Li, R., Raes, J., Arumugam, M., Burgdorf, K. S., et al. (2010). A Human Gut Microbial Gene Catalogue Established by Metagenomic Sequencing. *Nature* 464, 59–65. doi:10.1038/nature08821

Rizkallah, M., Saad, R., and Aziz, R. (2010). The Human Microbiome Project, Personalized Medicine and the Birth of Pharmacomicrobiomics. *Cppm* 8, 182–193. doi:10.2174/187569210792246326

Rothschild, D., Weissbrod, O., Barkan, E., Kurilshikov, A., Korem, T., Zeevi, D., et al. (2018). Environment Dominates over Host Genetics in Shaping Human Gut Microbiota. *Nature* 555, 210–215. doi:10.1038/nature25973

Saad, R., Rizkallah, M. R., and Aziz, R. K. (2012). Gut Pharmacomicrobiomics: the Tip of an Iceberg of Complex Interactions between Drugs and Gut-Associated Microbes. *Gut Pathog.* 4, 16. doi:10.1186/1757-4749-4-16

SADAG (2019). Available at: https://www.sadag.org/(Accessed February 4, 2022).

saNeuroGut (2017). Available at: https://www.saneurogut.org/(Accessed February 4, 2022).

Sarker, G., and Peleg-Raibstein, D. (2018). Maternal Overnutrition Induces Long-Term Cognitive Deficits across Several Generations. *Nutrients* 11, 7. doi:10.3390/nu11010007

Scheperjans, F., Aho, V., Pereira, P. A. B., Koskinen, K., Paulin, L., Pekkonen, E., et al. (2015). Gut Microbiota Are Related to Parkinson's Disease and Clinical Phenotype. *Mov Disord.* 30, 350–358. doi:10.1002/mds.26069

Schork, N. J. (2015). Personalized Medicine: Time for One-Person Trials. *Nature* 520, 609–611. doi:10.1038/520609a

Sender, R., Fuchs, S., and Milo, R. (2016). Revised Estimates for the Number of Human and Bacteria Cells in the Body. *PLOS Biol.* 14, e1002533. doi:10.1371/journal.pbio.1002533

Sirugo, G., Williams, S. M., and Tishkoff, S. A. (2019). The Missing Diversity in Human Genetic Studies. *Cell* 177, 26–31. doi:10.1016/j.cell.2019.02.048

Sorbara, M. T., and Pamer, E. G. (2022). Microbiome-based Therapeutics. *Nat. Rev. Microbiol.* doi:10.1038/s41579-021-00667-9

Stokholm, J., Blaser, M. J., Thorsen, J., Rasmussen, M. A., Waage, J., Vinding, R. K., et al. (2018). Maturation of the Gut Microbiome and Risk of Asthma in Childhood. *Nat. Commun.* 9, 141. doi:10.1038/s41467-017-02573-2

Sun, J., Xu, J., Ling, Y., Wang, F., Gong, T., Yang, C., et al. (2019). Fecal Microbiota Transplantation Alleviated Alzheimer's Disease-like Pathogenesis in APP/PS1 Transgenic Mice. *Transl. Psychiatry* 9, 189. doi:10.1038/s41398-019-0525-3

Sun, M.-F., Zhu, Y.-L., Zhou, Z.-L., Jia, X.-B., Xu, Y.-D., Yang, Q., et al. (2018). Neuroprotective Effects of Fecal Microbiota Transplantation on MPTP-Induced Parkinson's Disease Mice: Gut Microbiota, Glial Reaction and TLR4/TNF-α Signaling Pathway. *Brain Behav. Immun.* 70, 48–60. doi:10.1016/j.bbi.2018.02.005

Thursby, E., and Juge, N. (2017). Introduction to the Human Gut Microbiota. *Biochem. J.* 474, 1823–1836. doi:10.1042/BCJ20160510

Tidjani Alou, M., Lagier, J.-C., and Raoult, D. (2016). Diet Influence on the Gut Microbiota and Dysbiosis Related to Nutritional Disorders. *Hum. Microbiome J.* 1, 3–11. doi:10.1016/j.humic.2016.09.001

Tierney, B. T., Tan, Y., Kostic, A. D., and Patel, C. J. (2021). Gene-level Metagenomic Architectures across Diseases Yield High-Resolution Microbiome Diagnostic Indicators. *Nat. Commun.* 12, 2907. doi:10.1038/s41467-021-23029-8

Tremlett, H., Bauer, K. C., Appel-Cresswell, S., Finlay, B. B., and Waubant, E. (2017). The Gut Microbiome in Human Neurological Disease: A Review. *Ann. Neurol.* 81, 369–382. doi:10.1002/ana.24901

Turnbaugh, P. J., Ley, R. E., Hamady, M., Fraser-Liggett, C. M., Knight, R., and Gordon, J. I. (2007). The Human Microbiome Project. *Nature* 449, 804–810. doi:10.1038/nature06244

Vallès, Y., and Francino, M. P. (2018). Air Pollution, Early Life Microbiome, and Development. *Curr. Envir Health Rpt* 5, 512–521. doi:10.1007/s40572-018-0215-y

Vangay, P., Johnson, A. J., Ward, T. L., Al-Ghalith, G. A., Shields-Cutler, R. R., Hillmann, B. M., et al. (2018). US Immigration Westernizes the Human Gut Microbiome. *Cell* 175, 962–972.e10. doi:10.1016/j.cell.2018.10.029

Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., et al. (2001). The Sequence of the Human Genome. *Science*. doi:10.1126/science.1058040

Vogt, N. M., Kerby, R. L., Dill-McFarland, K. A., Harding, S. J., Merluzzi, A. P., Johnson, S. C., et al. (2017). Gut Microbiome Alterations in Alzheimer's Disease. *Sci. Rep.* 7, 13537. doi:10.1038/s41598-017-13601-y

World Health Organization (2016). Available at: https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight (Accessed November 26, 2021).

Xavier, R. J., and Podolsky, D. K. (2007). Unravelling the Pathogenesis of Inflammatory Bowel Disease. *Nature* 448, 427–434. doi:10.1038/nature06005

Yang, M., Bose, S., Lim, S., Seo, J., Shin, J., Lee, D., et al. (2020). Beneficial Effects of Newly Isolated Akkermansia Muciniphila Strains from the Human Gut on Obesity and Metabolic Dysregulation. *Microorganisms* 8, 1413–1426. doi:10.3390/microorganisms8091413

Yeoh, Y. K., Zuo, T., Lui, G. C.-Y., Zhang, F., Liu, Q., Li, A. Y., et al. (2021). Gut Microbiota Composition Reflects Disease Severity and Dysfunctional Immune Responses in Patients with COVID-19. *Gut* 70, 698–706. doi:10.1136/gutjnl-2020-323020

Zimmermann, M., Zimmermann-Kogadeeva, M., Wegmann, R., and Goodman, A. L. (2019). Mapping Human Microbiome Drug Metabolism by Gut Bacteria and Their Genes. *Nature* 570, 462–467. doi:10.1038/s41586-019-1291-3

Zuo, T., Zhang, F., Lui, G. C. Y., Yeoh, Y. K., Li, A. Y. L., Zhan, H., et al. (2020). Alterations in Gut Microbiota of Patients with COVID-19 during Time of Hospitalization. *Gastroenterology* 159, 944–955.e8. doi:10.1053/j.gastro.2020.05.048

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Genetics

ORIGINAL RESEARCH
published: 08 April 2022
doi: 10.3389/fgene.2022.834265

Prevalence of Clinically Relevant Germline *BRCA* Variants in a Large Unselected South African Breast and Ovarian Cancer Cohort: A Public Sector Experience

Nerina C. Van der Merwe[1,2]*, *Herkulaas MvE Combrink*[3,4], *Kholiwe S. Ntaita*[1,2] and *Jaco Oosthuizen*[1,2]

[1]Division of Human Genetics, National Health Laboratory Service, Bloemfontein, South Africa, [2]Division of Human Genetics, Faculty of Health Sciences, University of the Free State, Bloemfontein, South Africa, [3]Economic and Management Sciences, University of the Free State, Bloemfontein, South Africa, [4]Interdisciplinary Centre for Digital Futures, University of the Free State, Bloemfontein, South Africa

OPEN ACCESS

Edited by:
Dawn Stephens,
Technology Innovation Agency (TIA), South Africa

Reviewed by:
Rong Bu,
King Faisal Specialist Hospital & Research Centre, Saudi Arabia
Peter Hulick,
NorthShore University HealthSystem, United States

***Correspondence:**
Nerina C. Van der Merwe
vanderMerweNC@ufs.ac.za

Specialty section:
This article was submitted to Human and Medical Genomics, a section of the journal Frontiers in Genetics

Received: 13 December 2021
Accepted: 23 February 2022
Published: 08 April 2022

Citation:
Van der Merwe NC, Combrink HMvE, Ntaita KS and Oosthuizen J (2022) Prevalence of Clinically Relevant Germline BRCA Variants in a Large Unselected South African Breast and Ovarian Cancer Cohort: A Public Sector Experience. Front. Genet. 13:834265. doi: 10.3389/fgene.2022.834265

Breast cancer is a multifaceted disease that currently represents a leading cause of death in women worldwide. Over the past two decades (1998–2020), the National Health Laboratory Service's Human Genetics Laboratory in central South Africa screened more than 2,974 breast and/or ovarian cancer patients for abnormalities characteristic of the widely known familial breast cancer genes, Breast Cancer gene 1 (*BRCA1*) and Breast Cancer gene 2 (*BRCA2*). Patients were stratified according to the presence of family history, age at onset, stage of the disease, ethnicity and mutation status relative to *BRCA1/2*. Collectively, 481 actionable (likely-to pathogenic) variants were detected in this cohort among the different ethnic/racial groups. A combination of old (pre-2014) and new (post-2014) laboratory techniques was used to identify these variants. Additionally, targeted genotyping was performed as translational research revealed the first three recurrent South African pathogenic variants, namely *BRCA1* c.1374del (legacy name 1493delC), *BRCA1* c.2641G>T (legacy name E881X) and *BRCA2* c.7934del (legacy name 8162delG). This initial flagship study resulted in a cost-effective diagnostic test that enabled screening of a particular ethnic group for these variants. Since then, various non-Afrikaner frequent variants were identified that were proven to represent recurrent variants. These include *BRCA2* c.5771_5774del (legacy name 5999del4) and *BRCA2* c.582G>A, both Black African founder mutations. By performing innovative translational research, medical science in South Africa can adopt first-world technologies into its healthcare context as a developing country. Over the past two decades, the progress made in the public sector enabled a pivotal shift away from population-directed genetic testing to the screening of potentially all breast and ovarian cancer patients, irrespective of ethnicity, family history or immunohistochemical status. The modifications over the years complied with international standards and guidelines aimed at universal healthcare for all. This article shares all the cohort stratifications and the likely-to pathogenic variants detected.

Keywords: familial breast and ovarian cancer, predisposing genes *BRCA1/2*, central South Africa, National Health Laboratory Service, public sector

1 INTRODUCTION

The two familial breast cancer (BC) genes, Breast Cancer gene 1 (*BRCA1*) and Breast Cancer gene 2 (*BRCA2*), are highly penetrant and contribute to various cellular events ranging from the response to DNA damage to control of the cell cycle and apoptosis (Yoshida and Miki, 2004). Germline pathogenic variants in these genes create genetic instability impacting their capacity to repair the damage. Likely- to pathogenic variants in these genes are associated with hypersensitivity in BC patients. Their presence results in potentially severe radiotherapy complications during treatment due to spontaneous and enhanced radio-sensitivity (Chistiakov et al., 2008; Kan and Zhang, 2015). Disruptive *BRCA1/2* variants are associated with a predisposition to breast and ovarian cancer, and although at a lower frequency, prostate, pancreas and other cancer types are also linked to pathogenic *BRCA1/2* variants. Although these two high-impact cancer-predisposing genes were discovered more than two decades ago, they have dominated the field of BC genetics ever since.

BC is the most common cancer and the leading cause of cancer-related death in females worldwide (Bray et al., 2018). Most cases are sporadic. However, 5%–10% can be attributed to a hereditary component (Larsen et al., 2014). The disease was mostly considered an illness of the affluent; however, the incidence in developing countries, such as South Africa (SA), is rapidly increasing (Joffe et al., 2018). The age-standardized annual BC incidence rate (ASR) per 100,000 ranges between 52.92 and 79.3 for Asian, Caucasian and mixed-race SA women, compared to 29.1 for Black women. The average ASR, however, is currently 1 in 32 for a SA female to develop the disease (Francies et al., 2015).

The complex history of sub-Saharan Africa has highlighted the diverse populations of SA regarding the field of medical and population genetics (Oosthuizen et al., 2021). Although SA harbors over 60 million people, each of its main population groups has a unique origin. This diversity resulted from various migration events from all over the globe, such as European colonialism from predominantly north-western Europe, which gave rise to the Afrikaner with its Anglo-European descent (Attlee, 1947). Simultaneously, the indigenous expansion of Black Africans to the southern tip of the African continent resulted in approximately 80% of the entire SA population being neither culturally, linguistically, nor genetically homogenous (Van der Merwe et al., 2012).

Additional genetic lineages were introduced by laborers arriving from south Asia. Their arrival resulted in admixture between various groups already residing in SA, including the indigenous African people such as the Khoikhoi, the San, and the African Xhosa tribe (Attlee, 1947). These groups were eventually absorbed into the mixed ancestry group (Oosthuizen et al., 2021). Finally, the last major grouping (Asian population) originated from admixture of individuals from mainland India, neighboring countries such as Bangladesh, and the Mixed Ancestry population of SA (Vishnu and Morrell, 1991). Therefore, the modern-day Asian (specifically the SA Indian) population comprises mostly of people who migrated from mainland India to SA over 300 years, with admixture involving countries from Eurasia and Africa (Mesthrie, 2006; Isaacs et al., 2013). As our genomes reflect a record of historical events, so too does the genetic diversity in the field of hereditary breast and ovarian cancer (HBOC) reflect the complexity of the SA population (Van der Merwe et al., 2020; Combrink et al., 2021; Oosthuizen et al., 2021).

Patients with likely- to pathogenic germline variants in these high-risk genes have an increased predisposition to develop BC and/or ovarian cancer (OVC) throughout their lifetime. According to global statistics, the cumulative risk for *BRCA1* and *BRCA2* mutation carriers to develop BC before 80 years of age is 40%–87% and 27%–84%, respectively. The associated risk for OVC varies from 16%–68% and 11%–30%, respectively (Kuchenbaecker et al., 2017). The etiology related to hereditary BC and OVC in SA derived great benefit from population-based genetic research (Reeves et al., 2004; Agenbag, 2005; Van der Merwe and van Rensburg, 2009; Sluiter and Van Rensburg, 2011; Van der Merwe et al., 2012; Peter, 2014; Chen, 2015; Francies et al., 2015; Combrink, 2016; Moeti, 2016; Oosthuizen, 2016; Van der Merwe et al., 2020; Combrink et al., 2021; Mampunye et al., 2021; Oosthuizen et al., 2021), resulting in the identification of five founder variants representative of three of the four major ethnic groups in the country (*BRCA1* c.1374del [rs397508862], *BRCA1* c.2641G>T [rs39750888] (Reeves et al., 2004); *BRCA2* c.7934del [rs80359688] (Van der Merwe and van Rensburg, 2009); *BRCA2* c.5771_5774del [80359535] (Van der Merwe et al., 2012); *BRCA2* c.582G>A [rs80358810] (Oosthuizen et al., 2021)).

The SA studies performed to date reflect substantial variation in the yield of actionable (likely-pathogenic and pathogenic) germline variants identified in the country. As only 4.5% of the SA total budget is allocated to healthcare expenditure, it burdens an already stressed public sector to seek cost-effective alternatives for routine diagnostic testing of familial breast and ovarian cancer patients. This single-institution public sector study aimed to determine the range and positive mutation percentage of *BRCA1* and *BRCA2* actionable variants in an unselected large cohort of BC and OVC patients. These patients were screened using various technologies ranging from targeted genotyping to comprehensive screening, employing both older and new technologies. Our results prompted us to contemplate the most appropriate workflow for SA state-owned pathology laboratories to provide cost-effective genetic assessment in a financially constraint health sector.

2 MATERIALS AND METHODS

2.1 Study Population

A total of 2,975 BC and/or OVC patients were consulted at the National Health Laboratory Service (NHLS) Human Genetics

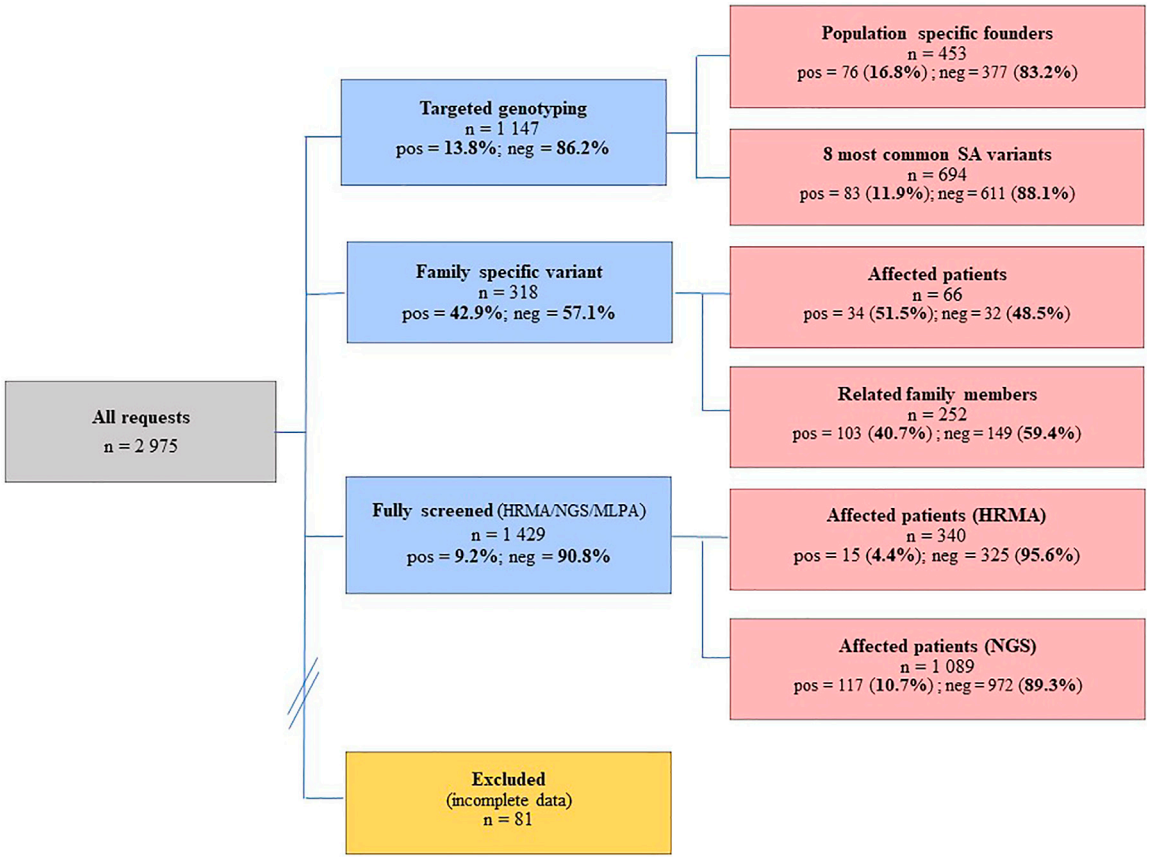


FIGURE 1 | Individuals screened at the NHLS Human Genetics laboratory in Bloemfontein between 1998 and 2020, using a combination of investigative methods ranging from targeted genotyping to comprehensive screening. The numbers screened together with the success rates are indicated. Pos, positives; neg, negatives; HRMA, high resolution melting analysis; NGS, next-generation sequencing; MLPA, multiplex ligation-dependent probe amplification.

TABLE 1 | Self-identified ethnicity of consecutive BC and/or OVC patients included in the study (*n* = 2,896).

Major ethnic classification	Sub-ethnic classification	Count
Asian	South African Indian, Malay, Pakistani, Chinese, Japanese, Bengali, Mongolian	297
Black African	All Black African nationalities	753
Caucasian	South African Afrikaner, British, and European descent	975
Mixed ancestry	A South African subgroup of people that comprise a mixture of any of the abovementioned ethnic classifications	669
Unknown	Not specified	200

Laboratory in Bloemfontein between 01/01/1998 and 31/12/2020, of whom 81 patients had incomplete data and were excluded from the analysis. The remaining 2,894 BC and/or OVC patients (2,733 females and 161 males) were examined by targeted genotyping or comprehensive screening of *BRCA1/2* (**Figure 1**). These BC and/or OVC patients attended their closest genetic clinic at a regional or provincial hospital, where they were referred for diagnostic genetic testing through their local genetic counselor or attending physician. Indications for testing included BC diagnosed at age 45 years or less or a significant family history (at least one first-degree family member with premenopausal BC or OVC, or multiple second-degree family members with premenopausal BC and/or OVC, or males with BC at any age).

Ethnicity was determined by patients' self-identification at the time of consultation. The major ethnic group classifications have sub-classifications, some contextualized specifically to the SA demographic profile (**Table 1**). The majority (*n* = 975) was Caucasian, followed by the Black African (*n* = 753) and mixed ancestry groups (*n* = 669), with the Asian group being the minority (*n* = 297). The ethnicities for the remainder of patients (*n* = 200) were either not indicated or unknown. The distribution of the patients across the various ethnic groups reflected the demographic profile of the breast clinic only and did not reflect

BC incidence within each of these groups. Additionally, patient reports generated over the years were retrospectively analyzed to collect information on epidemiological characteristics, which specifically focused on 1) presence of a family history of BC and/or OVC; 2) age at onset; 3) stage of the disease; 4) ethnicity; and 5) mutation status.

Majority of individuals included in this cohort were diagnosed with BC and/or OVC. All the individuals had received pre- and post-test counseling at their respective referral centers. Information regarding personal and family history, as well as written informed consent for DNA testing, were obtained. The cohort included patients with a positive family history (two or more affected family members) for BC and/or OVC, with most representing low-risk patients who had no prior knowledge of a family history of either condition but were diagnosed at an early age of onset (<40 years). The Health Sciences Research Ethics Committee (HSREC) of the University of the Free State in Bloemfontein, SA, approved all study protocols submitted since 1995 (ETOVS 31/95; ECUFS 31B/95; ECUFS 31C/95; ETOVS 49/06; ETOVS 65/08; ECUFS 107/2014; ECUFS 108/2014; UFS-HSD2019/1835/2910 and UFS-HSD2020/0194/3006). The NHLS permitted the use of the data (reference PR2110611).

2.2 Laboratory Methods

DNA was extracted from peripheral blood using a standard extraction method. The concentration and purity were determined using spectrophotometry. Targeted genotyping for the founder and recurrent actionable variants was performed using the LightCycler® 480 Genotyping Master Mix (Roche Diagnostics, Germany) on the LightCycler® 480 II real-time instrument (Roche Diagnostics, Germany). These assays employ hybridization and simple probe technology described by Oosthuizen (Oosthuizen, 2016). Targeted genotyping was performed using a standard real-time PCR regime, followed by melt curve analysis. Conventional mutation screening for single nucleotide variants (SNVs) and smaller indels was initially performed for a subset of patients described previously (Combrink, 2016; Oosthuizen, 2016). This approach involved a combination of older technology-based techniques including High-Resolution Melt Analysis (HRMA), the protein truncation test (PTT) and Sanger sequencing (Van der Merwe et al., 2020). A total of 340 patients were screened using a combination of these older techniques (**Figure 1**).

Individuals screened using the conventional approach were subjected to copy number variants (CNVs) analysis using Multiplex Ligation-dependent Probe Amplification (MLPA). The assays used included the SALSA® MLPA® P002-C1 and SALSA® MLPA® P002-D1 for *BRCA1*, with SALSA® MLPA® P045-B3 used for *BRCA2* (MRC-Holland, Amsterdam, Netherlands). The products were run together with a size standard on an ABI 3130XL Genetic analyzer (Applied Biosystems, Carlsbad, CA, United States). MLPA-positive results, especially in the case of single exon deletions, were corroborated using an alternative confirmation assay for each of the genes, namely SALSA® MLPA® P087-C1 for *BRCA1* and SALSA® MLPA® P077-A3 for *BRCA2*. All the data were analyzed using GeneMarker® software version 2.6.4 (SoftGenetics, LLC, State College, PA, United States). The CNVs were named according to the Human Genome Variation Society (http://www.HGVS.org/varnomen) guidelines and classified using the adapted recommendations of the American Society of Medical Genetics and Genomics (ACMG) for the interpretation and reporting of single-gene CNVs (Brandt et al., 2020).

Once introduced, next-generation sequencing (NGS) was performed to screen the remainder of samples (n = 1,089), using the Oncomine™ BRCA Research Assay (Life Technologies, Carlsbad, CA, United States). The primer pools targeted both genes' entire coding region and splice-site junctions. Multiplexed primer pools were used to construct the amplicon library using PCR-based targeted amplification. Sequencing was performed on the Ion Proton and S5 Platforms (Life Technologies, Carlsbad, CA, United States), and the Ion Reporter™ Software was used to filter out artifacts and annotate the variants. Raw signal data were analyzed as described by Van der Merwe et al. (2020). The average read-depths obtained were 581×, ranging between 148 and 1,965×. Using the depth per read, quartile statistics were applied to calculate average depth distribution around the mean across the NGS samples. Samples that were located within the 2nd and 3rd quartiles were selected to construct a CNV baseline with the Ion Reporter CNV VCIB 4.0.0.1 algorithm. CNV detection was performed using an algorithm that normalized depth coverage across amplicons to predict the copy number or ploidy states. The computed baseline included a minimum of 100 control samples (each with an average of 24 million bases called and a read count of 215,000), using regions with known ploidy states (https://assets.thermofisher.com/TFS-Assets/LSG/brochures/CNV-Detection-by-Ion.pdf). MLPA was performed to confirm all CNVs detected using NGS. Novel or complex sequence variants were confirmed by means of Sanger DNA sequencing (ABI Prism BigDye® Terminator v3.1 cycle sequencing kit, Foster City, CA, United States), using an Applied Biosystems 3130 automated sequencer (Life Technologies, Foster City, CA, United States).

The clinical significance of variants was determined based on the American College of Medical Genetics guidelines (ACMG, classification confirmed on 02/12/2021) (Richards et al., 2015) and evaluated from freely accessible public databases such as ClinVar and the genomic search engine VarSome (Kopanos et al., 2019). The variant nomenclature was used according to Human Genome Variation Society (HGVS) recommendations (http://www.hgvs.org/rec.html). *BRCA1/2* variants were numbered and annotated using the National Center for Biotechnology Information (NCBI) chromosomes and transcript reference sequences (NC_000017.11, NM_007294.4 and NC_000013.11, NM_000059.3), respectively. To prevent potential biases between the different laboratory techniques utilized for mutation screening throughout the years, all potential nucleotide changes were confirmed using Sanger sequencing. The analyses were confirmed using the same mutation detection databases and reference sequences.

TABLE 2 | Comparison between the number of mutation-positive versus mutation-negative patients (reflected in percentages) observed per ten-year intervals.

Age group	Mutation negative (*n* = 2,413)	Mutation positive (*n* = 481)	Total (%)
0–19 (*n* = 3)	0.10	0.00	0.10
20–29 (*n* = 62)	1.66	0.48	2.14
30–39 (*n* = 480)	13.96	2.63	16.59
40–49 (*n* = 800)	23.15	4.49	27.64
50–59 (*n* = 616)	17.45	3.84	21.29
60–69 (*n* = 467)	13.41	2.73	16.14
70–79 (*n* = 278)	8.05	1.55	9.61
80–89 (*n* = 108)	3.25	0.48	3.73
90+ (*n* = 21)	0.59	0.14	0.73
Unknown (*n* = 59)	1.76	0.28	2.04
Total (%)	83.38	16.62	100.00

3 RESULTS

3.1 Composition of Cohort

The influence of various SA founder variants is reflected in the relatively high mutation-positive rates observed for targeted genotyping (**Figure 1**), with 16.8% observed using population-directed targeted genotyping (*n* = 453). Although the initial screening was based on the individual's ethnicity, the positive mutation rate decreased when cases were screened using a broader approach that included all of the commonly occurring SA variants (11.9%, *n* = 694). Of the 1,429 patients comprehensively screened, 137 (9.2%) carried a likely- to pathogenic *BRCA1/2* variant. Although the majority represented SNV changes consisting predominantly of substitutions and deletions, eight CNVs were identified. Regarding the SA mutation spectrum, the Afrikaner founder variant *BRCA2* c.7934del (rs80359688) was the most common, followed by the Black African founder variant *BRCA2* c.5771_5774del (rs80359535).

An inherited susceptibility was confirmed in 51.5% of affected family members tested (*n* = 66), that carried the actionable variant segregating in the family (**Figure 1**). A relatively low number of unaffected family members (*n* = 252) were genotyped for various family-specific variants. The positive mutation rate for this group was high (40.7%). By knowing their mutation status, these patients were included in various cancer screening programs to facilitate earlier detection and a potentially better prognosis in case of a cancer diagnosis.

3.2 Epidemiology

The epidemiological data were analyzed according to the five variables highlighted, namely a family history of BC and/or OVC, age at onset, stage of the disease, ethnicity and mutation status. Regarding the age at onset, the majority of patients tested fell in the 40–49 age group (collectively 27.7%), followed by the 50–59 age group (21.2%) (**Table 2**). These two age intervals also delivered the highest percentage mutation-positive results compared to the other age groups, namely 4.5% and 3.8%, respectively. Only three (0.5%) of the patients identified with a *BRCA1/2* actionable variant in the 20–29 age group were observed. These positive patients represented approximately one-third of patients tested in this age group. These patients either represented high-risk BC/OVC families or were diagnosed with aggressive disease early and were therefore genetically screened (**Table 2**).

Of the 481 mutation carriers identified, the majority were Caucasian. However, the data for this grouping were skewed due to the translational research performed before 1998, revealing the presence of three common founder variants in the Caucasian sub-category Afrikaner group (**Table 1**). This research initially involved mostly Afrikaner patients with a positive family history of BC and/or OVC being screened. This group served as the ideal research group due to proven extended high linkage disequilibrium with various founder effects reported. This SA group is considered a fruitful "hunting ground" for pathogenic variants associated with disease (Hall et al., 2002; Van der Merwe et al., 2012). The presence of these *BRCA1/2* founder variants increased the mutation positivity rate to a remarkable 8.12%, delivering the highest positivity rate for Caucasians, namely 24.1%. The mutation positivity rate is an indicator used as a proxy for the relative percentage of patients that tested positive out of the total sample population (**Table 3**).

Information regarding the presence or absence of a family history of BC and/or OVC was available for most patients (**Supplementary Table S1**), with 5.8% designated as unknown (*n* = 141). Patients adopted as children contributed to the unknown category, as they had no prior family information. More than 80% of patients (*n* = 392) carrying an actionable *BRCA1/2* variant reported family members affected with BC and other cancer types. Although the information regarding family structure varied from being limited (lacking maternal or paternal lineages and ages at diagnoses) to extensive (three-generation pedigrees with both paternal and maternal lineages indicated), the percentage highlights the consistent importance of determining the family history as an effective selection criterion for *BRCA1/2* genetic testing. Despite reporting a positive family history of BC and/OVC, no actionable variants were detected for 59.9% (*n* = 1,445) of patients screened (**Supplementary Table S1**).

As the African continent has previously been associated with more aggressive breast disease and higher mortality rates due to late-stage presentation, the stage at diagnosis was compared between the ethnicities for 455 patients for whom the relevant information was available (**Supplementary Table S2**). The stage of disease at diagnosis for two groups (the Caucasians and individuals of mixed ancestry) was similar, with most patients diagnosed with Stage 2 disease. These two groups had the highest percentage of Stage 1 BC (10.6%), indicating increased community awareness and successful BC screening programs in the public health sector. These groups also exhibited a low number of patients diagnosed with Stage 4 disease (8.5% and 6.1%, respectively; **Supplementary Table S2**). This pattern was similar to that for the Black African group, except Stage 3 disease being the most prevalent. This finding still hints towards a later stage at presentation. However, it has improved significantly as a mere 7.6% of patients had Stage 4 disease at diagnosis. The Asian population of SA was the most alarming of the four ethnic groups due to the high percentage of patients diagnosed with Stage 4

TABLE 3 | Illustration of the mutation detection and positivity rate per major population group (reflected in percentages) observed for the major groups.

Ethnic group	Mutation positive (*n* = 481)	Mutation negative (*n* = 2,413) (%)	Mutation positivity rate (%)	Total (%)
Asian (*n* = 297)	1.62	8.64	15.8	10.26
Black African (*n* = 753)	3.77	22.25	14.5	26.02
Caucasian (*n* = 975)	8.12	25.57	24.1	33.69
Mixed ancestry (*n* = 669)	2.70	20.42	11.7	23.12
Unknown (*n* = 200)	0.41	6.50	6.0	6.91
Total (%)	16.62	83.38	16.6	100.00

TABLE 4 | Summary of variants detected during comprehensive screening of 1,429 patients.

Variant category	Patients (*n*)	Variants (*n*)	Homozygous variants (*n*)	Heterozygous variants (*n*)
BRCA1 actionable variants	55	34	0	34
BRCA2 actionable variants	76	39	0	39
Total # of *BRCA1* variants	na	146	22	124
Total # of *BRCA2* variants	na	198	26	172
Total # of VUSes in *BRCA1*	30	17	0	17
Total # of VUSes in *BRCA2*	96	30	0	30
BRCA1 variants with MAF <0.01	na	84	1	83
BRCA2 variants with MAF <0.01	na	112	0	112
Splice-site variants in *BRCA1*	na	11	0	11
Splice-site variants in *BRCA2*	na	11	0	11
Novel variants in *BRCA1*	na	13	1	12
Novel variants in *BRCA2*	na	17	0	17
Copy number variants in *BRCA1*	13	6	1	6
Copy number variants in *BRCA2*	2	2	0	2

VUS, variant of unknown clinical significance; na, not applicable.

disease (16.2%). However, this percentage could have been skewed due to the small sample size of this particular group (*n* = 37).

Patients affected with BC and/or OVC were divided according to unilateral and bilateral disease related to the presence or absence of a *BRCA1/2* variant (**Supplementary Table S3**). The majority of patients presented with unilateral BC, with only 220 cases affected with bilateral disease and a further 91 affected with OVC. A small number of patients (*n* = 62) were also diagnosed with a secondary cancer type not specified here (**Supplementary Table S3**). The mutation-positive rates in this cohort varied between the two BC groups, with each group exhibiting similar success rates. For the OVC cases, the detection rate was considerably higher (21.9%), as 20 patients in total carried *BRCA1/2* actionable variants (9 in *BRCA1* and 11 in *BRCA2*).

3.3 Mutation Spectrum

Apart from targeted genotyping, mutation screening of 1,429 patients revealed a wide range of variants across the SA population groups. The data were generated from 340 patients screened using older mutation screening techniques, with a further 1,089 assessed by NGS. The data of the two sets were incorporated and are presented in **Table 4**. A total of 132 (9.2%) patients representing 73 likely- to pathogenic variants were identified (*n* = 117 for NGS and *n* = 15 using older technologies), with 57.6% (76/132) representing *BRCA2*. Twenty-two of these actionable variants were classified as splice-site variants, mainly located in the intronic splice site boundaries. Various CNVs were detected, ranging from single exon to complete gene deletions. These CNVs have been previously described by Van der Merwe et al. (2020). Across the genes, NGS detected 344 variants, with only 14% present in a homozygous state. Unique variation in the SA population resulted in 196 variants identified with a minor allele frequency (MAF) below 0.01, with 47 consequently being classified as variants of unknown clinical significance (VUS), predominantly in *BRCA2*. Of these 47 variants, 30 were completely novel and not detected in international databases used for variant interpretation (**Table 4**).

The mutation positivity rates for the main SA ethnicities varied, with 9.2% reported for the Black African group (44/479), 6.6% for the Asians (12/180), 18.1% for the Caucasians (25/138), 13.2% for the mixed ancestry group (22/167), with 11.8% allocated to the group of unknown ethnicity (14/118). From the 59 actionable variants detected across the populations using NGS, 23 were detected in the Black African group, 10 in the Asian group, 12 in the Caucasian group, 13 in the mixed ancestry group, with 12 falling into the group of unknown ethnicity. Only seven of the 59 variants were detected in two separate populations, with a single pathogenic variant (*BRCA2* c.582G>A [rs80358810]) observed across all four main ethnic groups. However, the remainder of the actionable variants (51/59) were restricted to a single population group. Recurrence of

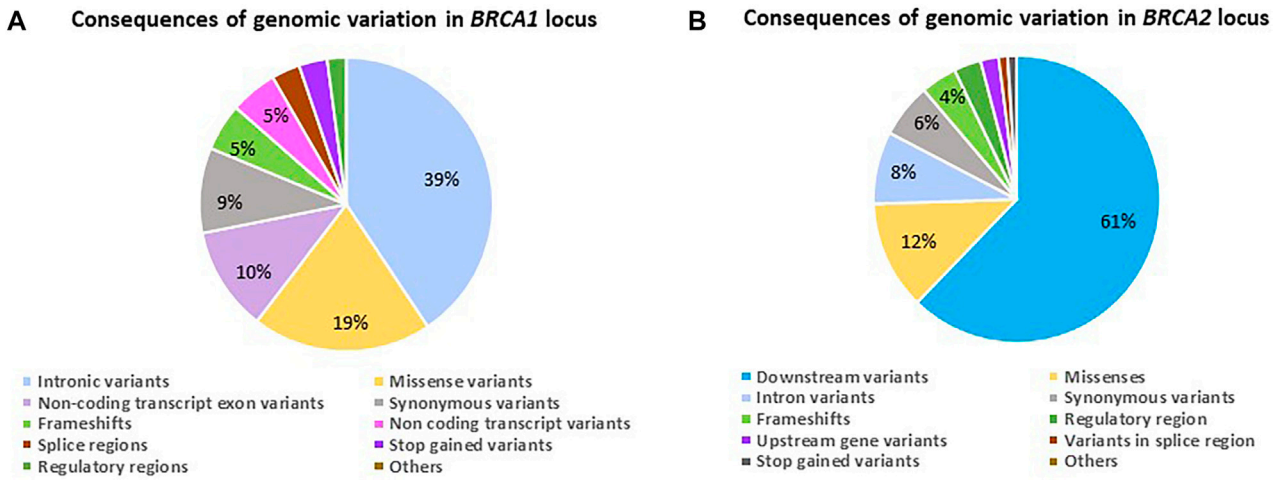


FIGURE 2 | Diagrams indicating the genomic variant frequencies for the *BRCA1* and *BRCA2* loci amongst 1,089 patients screened using NGS. **(A)** The consequence ratios relative to variant frequencies observed for the *BRCA1* locus. **(B)** The consequence ratios relative to variant frequencies observed for the *BRCA2* locus.

these likely- to pathogenic variants was low in the NGS cohort with 66.1% (39/59) observed for a single patient. Another 18.6% of the variants (11/59) were detected twice, followed by 6.8% (4/59) identified in three patients each. A small percentage (8.5%) was common and represented the three most common founder variants, namely *BRCA2* c.5771_5774del (detected in eight patients), *BRCA2* c.582G>A (10 patients) and *BRCA2* c.7934del (detected 17 times in the NGS cohort alone).

3.4 Genomic Consequence

When investigating the consequences of the variation observed, it varied for the two genes (**Figures 2A,B**). The genomic variant frequencies and associated consequences were only based on the NGS data. Although the bulk of variation for *BRCA1* was represented by non-coding transcript variants primarily present in the intronic regions (**Figure 2A**), the consequence of these variants has the potential to affect both overlapping genes, namely *BRCA1* (NM_000294.4 [43,044 295–43,125 364] and *Homo sapiens* Rho family GTPase 2 (*RND2*—NM_005440.5 [43,025 231–43,032 041]), involving a total of eight overlapping transcripts and five regulatory features. For *BRCA2*, the variation was present in the form of downstream changes observed in the 3′ untranslated region (**Figure 2B**). In comparison, the genomic variation observed in this section of chromosome 13 has the potential to affect three overlapping genes, namely *BRCA2* (NM_000059.4 [32,315 508–32,400 268]), *Homo sapiens* zygote arrest 1 like (*ZAR1L*—NM_001136571.2 [32,303 699–32,315 363]) and NEDD4 Binding Protein 2 Like 2 (*N4BP2L2*—NM_001,387,001.1 [32,432 485–32,538 795]), encompassing 28 overlapping transcripts and four regulatory features.

On DNA level, only a small percentage of the observed *BRCA1* changes represented variants potentially having an impact, namely frameshift (9%), splice- (3%), and stop-gained variants (5%) (**Figure 3A**). Their contribution increased on protein level (**Figure 3B**), directly impacting the protein and consequently efficient DNA repair, as 23% resulted in a prematurely truncated peptide (15% frameshift and 8% stop-gained variants) together with 5% missense variants. The majority of the actionable variants were detected in *BRCA1* exon 10. Approximately half of the variants observed at protein level represented missense variants, of which 14/55 changes were classified as VUSes. The missense variants were distributed throughout the gene, with the majority located outside of functional or disordered domains (**Table 4**).

Although there was an enrichment of actionable *BRCA2* variants in our populations (**Table 4**), the profile regarding the composition and impact of the variants on DNA and protein level corresponded between the two genes (**Figures 3A–D**), with the exception that the majority missenses of unknown significance were detected in a functional domain or disordered region. *BRCA2* did reveal a smaller margin of intronic variants (**Figure 3C**), as the bulk of non-coding variation was in the 3′ untranslated region. The high percentage of missense variants in the heterozygous form illustrated the unique composition of the SA populations, contributing to the rate of VUSes identified due to a lack of African data in the international context (**Table 4**).

4 DISCUSSION

4.1 Targeted Genotyping

Translational research performed at the University of Free State in collaboration with Professor L Jansen van Rensburg at the University of Pretoria resulted in the identification of the first two recurrent SA pathogenic variants in *BRCA1*, namely c.1374del (legacy name 1493delC) and c.2641G>T (legacy name E881X) observed for the Afrikaner population. Based on this research, *BRCA1* c.2641G>T was the first SA pathogenic variant to receive founder status, proven by haplotype analysis that indicated a

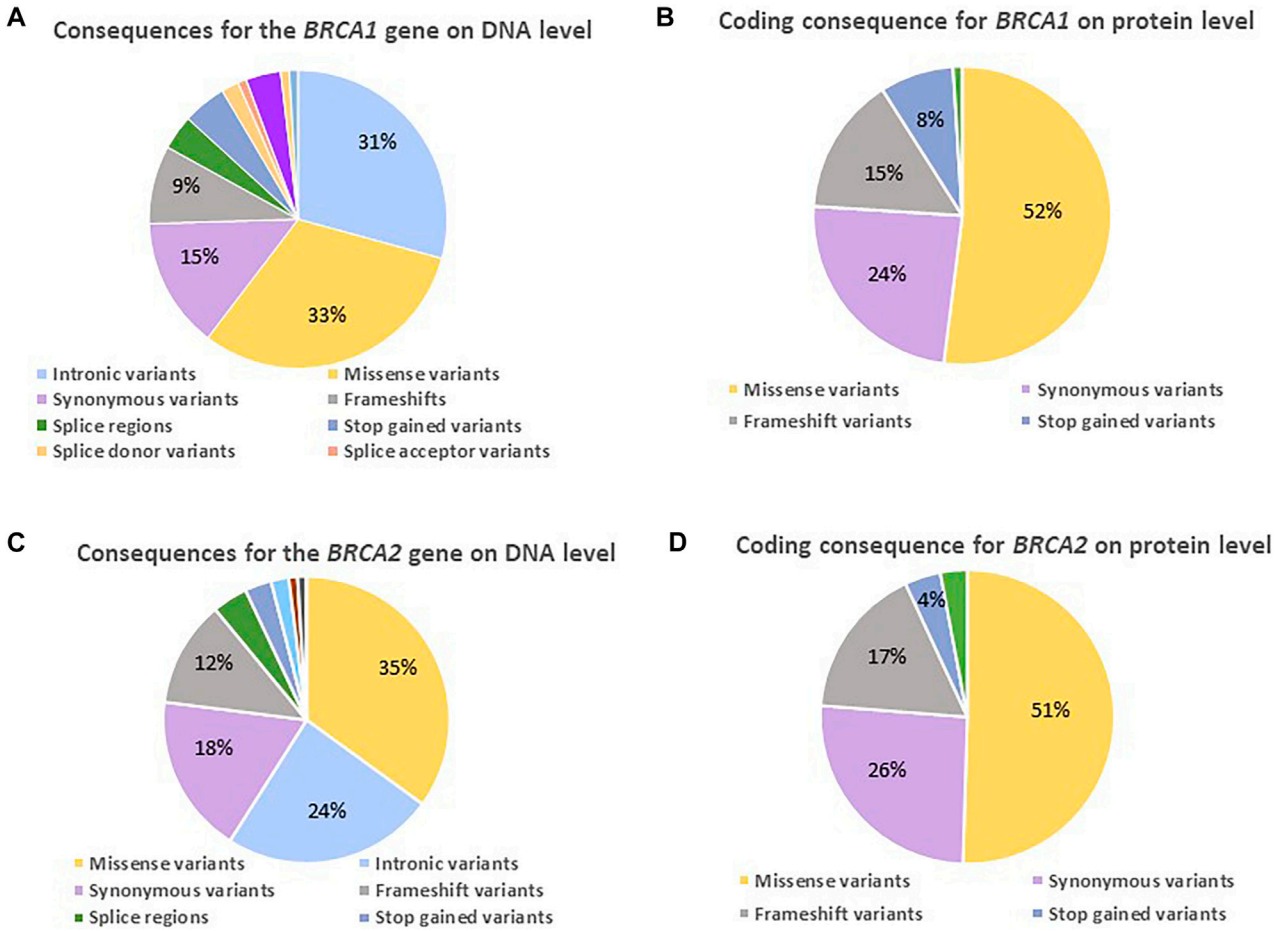


FIGURE 3 | Diagrams indicating the variants detected for the two familial BC genes on DNA and protein level amongst 1,089 BC and/or OVC patients screened using NGS. **(A)** The consequence ratios (in percentages) predicted for *BRCA1* on a DNA level. **(B)** The consequence ratios predicted for *BRCA1* on a protein level. **(C)** The consequence ratios predicted for *BRCA2* on a DNA level. **(D)** The consequence ratios predicted for *BRCA2* on a protein level.

single mutational event (Reeves et al., 2004). This variant was soon followed with founder status for *BRCA1* c.1374del and *BRCA2* c.7934del (legacy 8162delG). These three founder variants (with *BRCA2* c.7934del being the most common) were initially restricted to familial BC families with an Afrikaner heritage. Their founder status was corroborated with genealogical evidence dating back over 13 generations. For each variant, a single founding couple was identified based on genealogical findings traced back to France (*BRCA1* c.2641G>T), the Netherlands (*BRCA2* c.7934del) and Belgium (*BRCA1* c.1374del) (Reeves et al., 2004; Van der Merwe and van Rensburg, 2009).

The research resulted in the first diagnostic test (entailing screening for the three Afrikaner founder variants) offered to patients in the public sector in 1998. Diagnostic testing was initially restricted to Afrikaner individuals as translational research was performed in parallel to identify pathogenic variants in the other SA population groups (Agenbag, 2005; Sluiter and Van Rensburg, 2011; Van der Merwe et al., 2012; Peter, 2014; Chen, 2015; Combrink, 2016; Moeti, 2016; Oosthuizen, 2016). With time, various non-Afrikaner recurrent variants had been identified that were proven to represent recurrent variants, which include *BRCA2* c.5771_5774del (legacy 5999del4) and *BRCA2* c.582G>A, both Black African founder variants (Van der Merwe et al., 2012; Oosthuizen et al., 2021).

As the demand for diagnostic testing increased over the years, patients were collectively tested for the eight most commonly occurring SA variants, irrespective of ethnicity. This mutation set included the three Ashkenazi Jewish/European founder variants based on the African Lemba tribe's affinity with the Ashkenazi Jews (Spurdle and Jenkins, 1996) and the SA Indian community to mainland India and Europe (Combrink, 2016). This targeted genotyping assay served as a first-tier test for all patients that proved justified, because with time, the admixture of the SA populations became evident based on the genotyping results. The founder variants were no longer restricted to a single ethnicity but were shared among groups. This situation was expected, as recently illustrated by the findings of Hollfelder et al. (2020). With the use of five million genome-wide markers, 77 Afrikaner individuals were genotyped. Although the bulk of the genetic

information represented Europeans, the study indicated a contribution of 1.7% from South Asia or India, 1.3% from the Khoikhoi and the San and 0.8% representing West and East Africa (Hollfelder et al., 2020).

The results obtained from this series indicate that not all variants currently included in the first-tier genotyping assay are worthy of inclusion due to their low prevalence. This applies to the three Ashkenazi Jewish founder/European variants and *BRCA1* c.1374del, the rarest Afrikaner founder variant. The maximal financial benefit will be obtained once the first-tier assay has been re-designed to include the most commonly occurring variants in the SA population. It will aid the financially depleted healthcare system by first screening patients for the most common *BRCA1/2* variants in the population before advancing to more costly NGS (Oosthuizen et al., 2021).

4.2 Epidemiology

The epidemiology data highlighted three important aspects to be considered for the SA population. The first entailed the relatively high percentage of young individuals (20–29 years) identified with an actionable variant (0.5%, **Table 2**). Fifty-six percent of the group ($n = 62$) was unaffected and was referred for carrier testing only. For 22.8% of these individuals, targeted genotyping revealed a positive test result. By knowing their mutation status so early in life, these women were empowered by informed screening decisions and potentially alternative management options. Maximum benefit was therefore gained to reduce their risk of developing cancer types associated with *BRCA1/2* pathogenic variants. Asymptomatic young mutation carriers can reduce their risk of BC mortality by being regularly monitored and taking up interventions such as preventive surgery and/or chemoprevention (McCarthy and Armstrong, 2014).

The remainder (44%) of the 20–29 group was affected with early-onset BC ($n = 27$), with ages at onset ranging from 20 to 29 (average 26.7) years. The patients represented the Black African ($n = 15$), mixed ancestry ($n = 5$) and Caucasian ($n = 6$) ethnicities. Twenty-two percent ($n = 6$) of these patients received a positive test result (average age at onset 22.5 years), with the majority self-identified as Black African females. Half of these mutation-positive patients presented with a family history of BC and other cancer types, such as prostate cancer, melanoma and pancreatic cancer. Although these females were diagnosed at such a young age, knowing their *BRCA1/2* status promoted informed decision-making regarding treatment options and potential surgery. With a familial *BRCA* variant identified so early, cascade testing of related family members can proceed to assist with risk-reducing interventions themselves.

According to the latest clinical guidelines released for BC control and management by the SA Department of Health in 2018 (National Department of Health of the Republic of South Africa, 2018), all these patients qualified for genetic services, yet the numbers genetically screened are meager (**Table 2**). With a positive detection rate between 6% and 24%, attention should be drawn to genetic screening, as the benefits would outweigh the financial costs by far (**Table 3**).

The epidemiological evidence highlighted the ever-important value of a family history as a potential predictor of a positive test result (**Supplementary Table S1**), as 90% of patients with an actionable *BRCA1/2* variant reported family members affected with BC and other cancer types. Identifying an actionable *BRCA1/2* variant in an index acts as a key to the entire family. Once a mutation carrier has been identified, the benefits of genetic testing can be magnified through sharing positive test results with at-risk related family members. Doing so will ensure that they benefit from this information and secure their access to predictive testing according to the latest guidelines. This approach is of tremendous importance from a public health perspective, pushing for improved survival and quality of life through earlier detection and optimal management.

Conley et al. (2020) recently investigated the family disclosure patterns of *BRCA* genetic test results among young Black women with invasive BC in the United States of America. The study examined whether a disclosure to relatives was made, and if so, to whom it was disclosed. The authors reported that the most significant benefit of genetic testing, namely to inform family members of a hereditary predisposition, is not being realized in Black American families. Their findings revealed a reluctance of mutation-positive patients to disclose a positive test result even to their daughters (Conley et al., 2020).

The reluctance to share genetic test results with family members is also evident in SA. Here it is not restricted to the Black African population but clearly occurs among all ethnic groups in SA apart from the Caucasians. Despite 481 individuals receiving a positive test result, a mere 318 predictive or carrier tests had been performed over 24 years. The majority was performed before 2003 as a post-research initiative involving Afrikaner families. From the epidemiological results, it appeared that Caucasians tended to value and discussed the outcome of genetic testing to a greater extent than the other main ethnic groups (**Table 1**). This trend was also noted by Armstrong et al. (2003), who found that women pursuing *BRCA1/2* genetic testing in the United States were significantly more likely to be Caucasian.

This issue has been investigated internationally and although the public attitudes towards genetic testing for the risk of diseases, including cancer, are generally positive, various studies highlighted areas of concern. These include factors such as language barriers, fear of discrimination against those with a genetic predisposition for illness (Haga et al., 2013), being labeled as an individual or family with "good" or "bad" genes (Henneman et al., 2013) and fatalistic views of cancer (Allford et al., 2014). In SA, Schoeman et al. (2013) previously reported a low level of awareness regarding genetic testing for BC and other cancer types. Despite genetic testing being available since 1998, recognizing the value of a predictive test remains low. SA, therefore, needs to investigate innovative approaches to increase awareness among patients and communities, starting with related family members of mutation carriers. As the workforce of genetic providers is extremely low both in SA and globally, the international genetics profession has attempted to adapt to the situation by offering genetic counseling via alternative new methods, including service

delivery via telephone, telegenetics and group genetic counseling. Although a face-to-face consultation is always the ideal model to strive for, innovative service delivery models such as group counseling can improve access and contribute to community awareness.

In the paper by Mampunye et al. (2021), one such innovation was described for SA, which investigated the clinical value of rapid point-of-care (POC) genetic testing performed in combination with genetic counseling. As the ParaDNA workflow involves an integrated system from sample collection to report generation, prospective validation using a non-invasive cheek swab or saliva as input DNA is warranted (Mampunye et al., 2021). This approach would be ideal for use as a first-tier test performed by trained healthcare providers in parallel with genetic counseling in rural primary health clinics. This will provide the opportunity to improve the care process by delivering on-demand psychosocial support directly to the patient and indirectly to the community where it is needed. This approach could have many benefits: 1) reducing the lack of early healthcare-seeking due to limited financial resources and transport difficulties; 2) increasing community knowledge and cancer awareness resulting in less stigmatization; 3) creating an opportunity to explain the benefits of knowing your *BRCA1/2* genetic status for evidence-based cancer treatment options; and 4) highlighting the benefits of testing for at-risk family members in the case of a positive test result, thereby increasing preventative actions and early detection (Mampunye et al., 2021).

Unfortunately, the burden of BC-related death is ever increasing due to persisting misconceptions surrounding the disease and various other socio-economic factors such as poverty, cultural and religious beliefs (Van der Merwe et al., 2020; Mampunye et al., 2021; Oosthuizen et al., 2021). Despite the efforts of the SA Department of Health's BC development plan establishing multiple Regional Breast Units (RBUs, 28 distributed throughout the various provinces) at primary and secondary state hospitals, and 14 additional Specialized Breast Units (SBUs) located in mostly tertiary hospitals (National Department of Health of the Republic of South Africa, 2018), the uptake of breast screening remains low in women 30 years and older (Phaswana-Mafuya and Peltzer, 2018). The authors of this large study (n = 10,831) estimated a low prevalence of BC screening similar to that obtained in a 2008 survey involving older SA adults, 50 years and older (Peltzer and Phaswana-Mafuya, 2014). Their estimation corresponded with the prevalence observed in international low-income countries such as Thailand (Mukem et al., 2016) and Turkey (Sözmen et al., 2016), but was lower than that of Brazil (Theme Filha et al., 2016). For SA to advance in this battle and attack BC energetically and effectively, we have to invest in community-health educational out-reaches performed in parallel with highly specialized science to achieve the ultimate goals set for the country.

The epidemiological data finally emphasized the value of screening all OVC patients for actionable variants in *BRCA1/2*, as screening resulted in a mutation-positive rate of 28.2% (**Supplementary Table S3**). Nine (45.0%) of the 20 OVC mutation carriers exhibited one of the SA founder variants, with the remainder of actionable variants being family-specific. These founder variants could easily have been identified cost-effectively using first-tier targeted genotyping, before proceeding to comprehensive genetic analysis. A high number of these OVC *BRCA1/2* mutation carriers, however, was referred for identification of a rare family-specific variant. Although patients affected with OVC before the age of 60 are included in the national clinical guidelines for the control and management of BC (National Department of Health of the Republic of South Africa, 2018), only 91 OVC patients were received between 1998–2021. This number accounts for 3.9 patients screened per year in the public sector. Currently, the age-standardized incidence rate for OVC in southern Africa is 3.9 per 100,000 (The Council for Medical Schemes, 2019). Based on these statistics, the majority of OVC patients are currently not referred for any *BRCA1/2* screening.

Screening these patients could have a two-fold advantage. International ongoing clinical trials involving poly (ADP-ribose) polymerase (PARP) inhibitors have recently resulted in the approval of various inhibitors by the Food and Drug Administration within the US Department of Health and Human Services for clinical use in specifically epithelial OVC patients (Loizzi et al., 2020). This new therapeutic approach for the management of OVC has been suggested in particular for patients with assessed defects in the homologous recombination DNA repair process such as *BRCA1/2* (Sunada et al., 2018; Elias et al., 2018). By knowing their mutation status, patients will be appropriately selected for this new revolutionary treatment option. Unfortunately, the use of PARP inhibitors is not yet registered by the South African Health Products Regulatory Authority and their use in treatment requires Section 21 approval for the private sector (The Council for Medical Schemes, 2019). It is currently not available for the clinical treatment of patients in the public sector.

4.3 Mutation Spectrum

We aimed to describe clinically relevant germline *BRCA1/2* variants and their distribution across ethnicities in the most extensive unselected African series to date for the public sector. The data represent the current status after 23 years of *BRCA* testing within our state laboratory. Although testing commenced with a tiered approach in the form of research-based translational targeted genotyping, the total number of test requests and the number of variants identified soared within the past decade. This could be attributed to a heightened public awareness after the Angelina Jolie revelations in 2013 (Troiano et al., 2017) and the incorporation of NGS into our diagnostic workflow, which enabled a more rapid throughput and resulted in more effective variant discovery.

A total of 481 individuals were identified as high-risk *BRCA1/2* mutation carriers, with 69.6% (n = 335) representing patients affected by BC and/or OVC, accounting for 9.2% of patients comprehensively screened (132/1,429). The data revealed an extreme distribution for each gene, with only a few frequent pathogenic variants identified. The majority of variants were extremely rare and primarily family-specific (**Table 5**).

TABLE 5 | Actionable *BRCA1/2* variants (likely- to pathogenic) identified for the entire SA cohort (*n* = 2,896).

Variant	Protein	Cancer type in index or family	Exon	#Of families	rs number
BRCA1					
NC_000017.11:g.(?_43045584)_(43125327_?)del		BC	1 – 23	1	no rs
NC_000017.11:g.(?_43123946)_(43125327_?)del		BC & OVC	1 – 2	5	no rs
NC_000017.11:g.(?_43104032)_(43106675_?)del		BC & OVC	4 – 6	1	no rs
NC_000017.11:g.(?_43082330)_(43082599_?)dup		BC & OVC	12	2	no rs
NC_000017.11:g.(?_43063781)_(43064034_?)del		BC & OVC	17	3	no rs
NC_000017.11:g.(?_43048992)_(43049260_?)del		BC & OVC	21	1	no rs
NM_007294.4(BRCA1): c.45dup	NP_009225.1: p.Asn16Ter	BC	2	4	rs730881457
NM_007294.4(BRCA1): c.66dup	NP_009225.1: p.Glu23ArgfsTer18	BC	2	5	rs80357783
NM_007294.4(BRCA1): c.68_69del	NP_009225.1: p.Glu23ValfsTer17	BC & Gastric ca	2	8	rs80357914
NM_007294.4(BRCA1): c.71G>C	NP_009225.1: p.Cys24Ser	BC	3	2	no rs
NM_007294.4(BRCA1): c.110C>A	NP_009225.1: p.Thr37Lys	BC & OVC	3	1	rs80356880
NM_007294.4(BRCA1): c.135-1G>T	Splicing defect	BC	4	1	rs80358158
NM_007294.4(BRCA1): c.181T>G	NP_009225.1: p.Cys61Gly	BC & OVC	5	2	no rs
NM_007294.4(BRCA1): c.191G>A	NP_009225.1: p.Cys64Tyr	BC	5	2	rs55851803
NM_007294.4(BRCA1): c.212G>A	NP_009225.1: p.Arg71Lys (Splicing defect)	BC	5	1	rs80356913
NM_007294.4(BRCA1): c.212+1G>A	Splicing defect	BC	5	1	rs80356913
NM_007294.4(BRCA1): c.415C>T	NP_009225.1: p.Gln139Ter	BC	6	2	rs80357372
NM_007294.4(BRCA1): c.431dup	NP_009225.1: p.Asn144LysfsTer15	BC	6	3	rs397509162
NM_007294.4(BRCA1): c.1016dup	NP_009225.1: p.Val340LysfsTer6	BC	10	1	rs80357569
NM_007294.4(BRCA1): c.1360_1361delAG	NP_009225.1: p.Ser454Ter	BC & OVC	10	5	rs80357969
NM_007294.4(BRCA1): c.1374del	NP_009225.1: p.Asp458GlufsTer17	BC	10	9	rs397508862
NM_007294.4(BRCA1): c.1504_1508delTTAAA	NP_009225.1: p.Leu502AlafsTer2	BC & colon ca	10	3	rs80357888
NM_007294.4(BRCA1): c.2008G>T	NP_009225.1: p.Glu670Ter	BC & OVC	10	2	no rs
NM_007294.4(BRCA1): c.2070_2073delAAGA	NP_009225.1: p.Lys690AsnfsTer10	BC	10	1	no rs
NM_007294.4(BRCA1): c.2568T>G	NP_009225.1: p.Tyr856Ter	BC	10	1	rs80356832
NM_007294.4(BRCA1): c.2597G>A	NP_009225.1: p.Arg866His	BC	10	1	rs80356911
NM_007294.4(BRCA1): c.2599G>T	NP_009225.1: p.Gln867Ter	BC	10	1	rs886038001
NM_007294.4(BRCA1): c.2641G>T	NP_009225.1: p.Glu881Ter	BC & OVC	10	28	rs397508988
NM_007294.4(BRCA1): c.3108del	NP_009225.1: p.Phe1036LeufsTer12	BC	10	1	rs80357841
NM_007294.4(BRCA1): c.3228_3229delAG	NP_009225.1: p.Gly1077AlafsTer8	BC	10	3	rs80357635
NM_007294.4(BRCA1): c.3288_3289delAA	NP_009225.1: p.Leu1098SerfsTer4	BC	10	1	rs80357686
NM_007294.4(BRCA1): c.3331_3334delCAAG	NP_009225.1: p.Gln1111AsnfsTer5	BC	10	1	rs80357701
NM_007294.4(BRCA1): c.3400G>T	NP_009225.1: p.Glu1134Ter	BC	10	1	no rs
NM_007294.4(BRCA1): c.3496_3497insT	NP_009225.1: p.Ala1166ValfsTer2	BC	10	1	no rs
NM_007294.4(BRCA1): c.3549_3550delAGinsT	NP_009225.1: p.Lys1183AsnfsTer27	Male BC	10	1	rs273899709
NM_007294.4(BRCA1): c.3593T>A	NP_009225.1: p.Leu1198Ter	Male BC	10	1	rs397509095
NM_007294.4(BRCA1): c.3732_3733delTA	NP_009225.1: p.His1244GlnfsTer10	BC	10	1	no rs
NM_007294.4(BRCA1): c.3756_3759delGTCT	NP_009225.1: p.Ser1253ArgfsTer10	BC & Melanoma	10	2	rs80357868
NM_007294.4(BRCA1): c.3947_3950delTCTT	NP_009225.1: p.Phe1316Ter	BC	10	2	rs886040177
NM_007294.4(BRCA1): c.4308_4309delTT	NP_009225.1: p.Ser1437CysfsTer3	BC	12	2	no rs
NM_007294.4(BRCA1): c.4327C>T	NP_009225.1: p.Arg1443Ter	BC	12	1	rs41293455
NM_007294.4(BRCA1): c.4524G>A	NP_009225.1: p.Trp1508Ter	BC	15	1	rs80356885
NM_007294.4(BRCA1): c.4838_4839insC	NP_009225.1: p.Pro1614SerfsTer8	BC	15	2	rs397509200
NM_007294.4(BRCA1): c.4868C>T	NP_009225.1: p.Ala1623Val	BC	15	1	no rs
NM_007294.4(BRCA1): c.4987-5T>A	Splicing defect	BC	16	1	rs397509214
NM_007294.4(BRCA1): c.5095C>T	NP_009225.1: p.Arg1699Trp	BC	17	2	rs55770810
NM_007294.4(BRCA1): c.5096G>A	NP_009225.1: p.Arg1699Gln	BC	17	3	rs41293459
NM_007294.4(BRCA1): c.5177_5180delGAAA	NP_009225.1: p.Arg1726LysfsTer3	BC	19	1	rs80357867

(Continued on following page)

TABLE 5 | (*Continued*) Actionable *BRCA1/2* variants (likely- to pathogenic) identified for the entire SA cohort (*n* = 2,896).

Variant	Protein	Cancer type in index or family	Exon	#Of families	rs number
NM_007294.4(BRCA1): c.5229_5230delAA	NP_009225.1: p.Arg1744LysfsTer85	BC	19	4	rs80357852
NM_007294.4(BRCA1): c.5240_5243delGAAA	NP_009225.1: p.Arg1747LysfsTer3	BC	19	1	no rs
NM_007294.4(BRCA1): c.5266dup	NP_009225.1: p.Gln1756ProfsTer74	BC & OVC	19	6	rs80357906
NM_007294.4(BRCA1): c.5332+1G>C	Splicing defect	BC	21	1	rs80358041
NM_007294.4(BRCA1): c.5365_5366delGCinsA	NP_009225.1: p.Ala1789IlefsTer4	BC	21	1	no rs
NM_007294.4(BRCA1): c.5467+2T>G	Splicing defect	BC	21	3	rs80358009
NM_007294.4(BRCA1): c.5468-1G>A	Splicing defect	BC	23	1	rs80358048
Total				**143**	
BRCA2					
NC_000013.11:g.(?_32313776)_(32398795_?)del		BC	1 – 27	1	no rs
NC_000013.11:g.(?_32370334)_(32371115_?)del		BC & OVC	19 – 20	1	no rs
NM_000059.4(BRCA2): c.67+3A>G	Splicing defect	BC, Fanconi anemia	2	1	rs1593880835
NM_000059.4(BRCA2): c.93G>A	NP_000050.3: p.Trp31Ter	BC	3	1	rs80359214
NM_000059.4(BRCA2): c.516G>A	NP_000050.3: p.Lys172=	BC	6	2	rs80359790
NM_000059.4(BRCA2): c.582G>A	NP_000050.3: p.Trp194Ter	BC	7	13	rs80358810
NM_000059.4(BRCA2): c.771_775delTCAAA	NP_000050.3: p.Asn257LysfsTer17	BC	9	1	rs80359671
NM_000059.4(BRCA2): c.1261C>T	NP_000050.3: p.Gln421Ter	BC	10	1	rs80358419
NM_000059.4(BRCA2): c.1813dup	NP_000050.3: p.Ile605AsnfsTer11	Male BC	10	1	rs80359308
NM_000059.4(BRCA2): c.2636_2637delCT	NP_000050.3: p.Ser879Ter	BC	11	1	rs276174826
NM_000059.4(BRCA2): c.2806_2809delAAAC	NP_000050.3: p.Ala938ProfsTer21	BC	11	1	rs80359351
NM_000059.4(BRCA2): c.2826_2829delAATT	NP_000050.3: p.Ile943LysfsTer16	BC	11	1	rs397507643
NM_000059.4(BRCA2): c.2828_2831delTTAA	NP_000050.3: p.Ile943LysfsTer16	BC	11	1	rs397507643
NM_000059.4(BRCA2): c.3553dup	NP_000050.3: p.Thr1185AsnfsTer3	BC	11	1	no rs
NM_000059.4(BRCA2): c.3723del	NP_000050.3: p.Phe1241LeufsTer18	BC	11	1	rs886040491
NM_000059.4(BRCA2): c.3847del	NP_000050.3: p.Val1283LysfsTer2	BC	11	1	rs80359405
NM_000059.4(BRCA2): c.3881T>A	NP_000050.3: p.Leu1294Ter	BC	11	1	rs80358632
NM_000059.4(BRCA2): c.4003G>T	NP_000050.3: p.Glu1335Ter	BC	11	3	rs747070579
NM_000059.4(BRCA2): c.4456del	NP_000050.3: p.Val1486LeufsTer6	BC	11	1	no rs
NM_000059.4(BRCA2): c.4482_4483insAAAG	NP_000050.3: p.Ser1494LysfsTer20	BC & OVC	11	1	no rs
NM_000059.4(BRCA2): c.4568del	NP_000050.3: p.Gly1523ValfsTer20	BC	11	1	no rs
NM_000059.4(BRCA2): c.4936G>T	NP_000050.3: p.Glu1646Ter	BC	11	1	rs886038111
NM_000059.4(BRCA2): c.5082dup	NP_000050.3: p.Glu1695ArgfsTer5	BC	11	1	no rs
NM_000059.4(BRCA2): c.5213_5216delCTTA	NP_000050.3: p.Thr1738IlefsTer2	BC & OVC	11	2	rs80359493
NM_000059.4(BRCA2): c.5279C>G	NP_000050.3: p.Ser1760Ter	BC	11	1	rs80358751
NM_000059.4(BRCA2): c.5344C>T	NP_000050.3: p.Gln1782Ter	BC & OVC	11	1	rs80358757
NM_000059.4(BRCA2): c.5564C>G	NP_000050.3: p.Ser1855Ter	OVC	11	1	no rs
NM_000059.4(BRCA2): c.5771_5774delTTCA	NP_000050.3: p.Ile1924ArgfsTer38	Male BC, BC, OVC, prostate ca, endometrial ca, Fanconi aneamia	11	61	rs80359535
NM_000059.4(BRCA2): c.5946del	NP_000050.3: p.Ser1982ArgfsTer22	BC	11	5	rs80359550
NM_000059.4(BRCA2): c.6082_6086delGAAGA	NP_000050.3: p.Glu2028LysfsTer19	BC	11	1	rs80359558
NM_000059.4(BRCA2): c.6228del	NP_000050.3: p.Lys2077ArgfsTer4	BC	11	1	no rs
NM_000059.4(BRCA2): c.6393_6396delATTA	NP_000050.3: p.Lys2131AsnfsTer5	BC	11	1	rs397507849
NM_000059.4(BRCA2): c.6393del	NP_000050.3: p.Lys2131AsnfsTer6	BC	11	1	rs886038145
NM_000059.4(BRCA2): c.6447_6448dupTA	NP_000050.3: p.Lys2150IlefsTer19	BC & OVC	11	5	rs397507858
NM_000059.4(BRCA2): c.6623del	NP_000050.3: p.Asn2208IlefsTer2	BC	11	1	rs886038150
NM_000059.4(BRCA2): c.6937+2delT	Splicing defect	BC	12	1	no rs
NM_000059.4(BRCA2): c.7934del	NP_000050.3: p.Arg2645AsnfsTer3	Male BC, BC, OVC, prostate cancer, melanoma	17	176	rs80359688
NM_000059.4(BRCA2): c.7955T>G	NP_000050.3: p.Val2952Gly	BC	17	2	rs1555286868

(Continued on following page)

TABLE 5 | (*Continued*) Actionable *BRCA1/2* variants (likely- to pathogenic) identified for the entire SA cohort (n = 2,896).

Variant	Protein	Cancer type in index or family	Exon	#Of families	rs number
NM_000059.4(BRCA2): c.8067T>A	NP_000050.3: p.Cys2689Ter	BC	18	4	rs80359046
NM_000059.4(BRCA2): c.8165C>G	NP_000050.3: p.Thr2722Arg	BC	18	1	rs80359062
NM_000059.4(BRCA2): c.8167G>C	NP_000050.3: p.Asp2723His	BC	18	3	rs41293511
NM_000059.4(BRCA2): c.8168A>T	NP_000050.3: p.Asp2723Val	BC	18	1	rs41293513
NM_000059.4(BRCA2): c.8331+2T>C	Splicing defect	BC	18	1	rs309122602
NM_000059.4(BRCA2): c.8504C>G	NP_000050.3: p.Ser2835Ter	BC	18	1	rs80359102
NM_000059.4(BRCA2): c.8686del	NP_000050.3: p.Arg2896ValfsTer13	BC	21	2	no rs
NM_000059.4(BRCA2): c.8696_8712del17	NP_000050.3: p.Gln2899LeufsTer2	BC & OVC	21	1	no rs
NM_000059.4(BRCA2): c.8754+1G>A	Splicing defect	BC	21	4	rs397508006
NM_000059.4(BRCA2): c.8954-2A>C	Splicing defect	BC	23	1	rs1135401928
NM_000059.4(BRCA2): c.8961_8964delGAGT	NP_000050.3: p.Ser2988PhefsTer12	BC	23	2	rs80359734
NM_000059.4(BRCA2): c.9105T>G	NP_000050.3: p.Tyr3035Ter	BC	23	3	rs886040819
NM_000059.4(BRCA2): c.9117G>A	NP_000050.3: p.Pro3039= (Splicing defect)	BC	23	1	rs28897756
NM_000059.4(BRCA2): c.9138del	NP_000050.3: p.Gln3047ArgfsTer15	BC	24	3	no rs
NM_000059.4(BRCA2): c.9154C>T	NP_000050.3: p.Arg3052Trp	BC	24	1	rs45580035
NM_000059.4(BRCA2): c.9351del	NP_000050.3: p.His3117GlnfsTer3	BC	25	8	no rs
NM_000059.4(BRCA2): c.9382C>T	NP_000050.3: p.Arg3128Ter	BC	25	1	no rs
NM_000059.4(BRCA2): c.9833_9842del	NP_000050.3: p.Pro3278HisfsTer32	BC	25	1	no rs
NM_000059.4(BRCA2): c.9435_9436delGT	NP_000050.3: p.Ser3147CysfsTer2	BC & OVC	25	5	rs80359763
Total				**342**	

Although the proven founder or recurrent variants for the Ashkenazi Jewish, Afrikaner and Black African/mixed ancestry variants were the most common variants observed in terms of the highest frequencies, they only represented <10% of the variants identified (4/55 for *BRCA1* and 5/57 for *BRCA2*, **Table 5**).

Both the number of variants and their mutation spectrum differed for the various population groups. From the NGS data, it seemed as if the Black African group (n = 479) exhibited the largest diversity in both actionable and novel variants (44/479), as approximately double the number of variants were observed compared to the other groups (Asians 12/180; Caucasians 25/138; mixed ancestry 22/167 and individuals of unknown ethnicity 14/118). These numbers, however, do not accurately reflect the contribution of pathogenic variants to this group, as considerably more Black African patients were tested (at a ratio of approximately 3:1). Despite their higher diversity of pathogenic variants, the Black African group had the second-lowest positive detection rate (9.2%), apart from the Asian population with 6.6%. This can partly be attributed to patients being referred for genetic testing based on an early age at diagnosis alone, as most Black African patients were unaware of the accumulation of cancer occurrences in their families. In contrast, the Caucasian population exhibited the highest detection rate, namely 18.1%, despite a much lower number of patients tested using NGS. These patients seemed to be more appropriately selected as the majority of patients reported an intermediate to strong family history of BC and/or OVC. As the majority of mutation-positive patients carried one of the Afrikaner founder variants, the contribution of the Afrikaner founder variants to this group was evident.

The Caucasian and Black African detection rates declined from 18.1% to 9.2% to an overall 10% and 6.6%, respectively, once the founder variants detected during NGS were excluded. This finding indicates an ultimate positive mutation detection rate below 10% for NGS, which is currently not cost-effective. If these patients were screened using the first-tier targeted genotyping assay, costs could have been reduced by excluding these patients prior to NGS analysis. The difference in the positive detection rate between these two ethnic groups with well-characterized variants iterates the importance of family history and genetic cancer registries. By updating these registries, testing centers can keep track of related family members carrying actionable familial *BRCA1/2* variants, with the sole purpose to identify at-risk symptomatic-free family members.

This study attempted to report the mutation detection rates over the past two decades from a single institutional series' perspective, with some biases due to: 1) the various techniques used; 2) disproportionate numbers of multiple ethnic groups studied; and 3) the minimum clinical criteria for BRCA testing changing. Therefore, the mutation detection rates presented per ethnic group does not accurately represent the positive predictive value of each technique and the national mutational burden of the respective groups. It merely reflects the frequencies of actionable variants detected at the time and within

the performance specifications regarding the sensitivity and specificity of each mutation screening technique. The genotyping approach only identified selected pathogenic variants and did not enable the reporting of VUSes. Compared to NGS, screening for pathogenic variants using HRMA could have missed pathogenic variants and VUSes due to possible inadequate sensitivity during melt curve analysis in suboptimal PCR conditions (despite rigorous optimization and running each reaction in duplicate) (Combrink, 2016; Oosthuizen, 2016). In addition to a potential reduction in sensitivity of HRMA, the largest exons namely exon 10 of *BRCA1* and exons 10 and 11 of *BRCA2* were screened for protein-truncating variants only using PTT. The technique would therefore have missed various missense, synonymous or splice-site variants which could have represented actionable variants.

Sanger sequencing was utilized only for the confirmation of variants detected using the various mutation screening techniques. It was therefore not employed in this series for sequencing entire coding regions and splice site boundaries of samples. Moreover, biases in the detection frequencies of ethnic groups could have been introduced due to the disproportionate number of individuals in each group being screened with NGS, the most modern and sensitive technique in the test repertoire. Lastly, the criteria for *BRCA1/2* screening broadened with time and became more inclusive throughout the decades with more individuals currently meeting criteria than did a decade or two ago.

This series represents the most extensive report involving the *BRCA1/2* mutation spectrum on the African continent, surpassing the Nigerian study involving 1,136 patients (**Supplementary Table S4**). The positive mutation rate, however, was similar (9.2% versus a collective 11.1% in Nigeria), although the contribution of the genes was reversed, with *BRCA2* being more prevalent in SA (**Supplementary Table S4**). This observation can be attributed to the prevalence of three SA founder variants in *BRCA2*, representing three of the four SA ethnic groups (Black African, Caucasian and mixed ancestry groups). Although some African countries reported extremely high *BRCA1/2* mutation-positive rates above 15% (such as Egypt, Morocco, SA, Sudan and Tunisia), the majority of African studies involved small sample sizes based on very strict selection criteria.

Twenty-seven of the actionable variants were novel, with no unique identifier listed. The total number of novel variants increased when the complete variant list was considered (including benign variants to VUSes), with most present in a heterozygous form (**Table 4**). These novel variants were mostly observed for the Black African group, which was expected due to the high degree of variation evident in the African genome when compared to that of the Asian, African-American and European genomes. Recent genomic studies (Choudhury et al., 2020) revealed the presence of more than 3 million previously undescribed variants and predicted that only a fraction of the genetic diversity among individuals on the African continent has thus far been uncovered. This study exposed complex patterns of ancestral admixture, as both intra- and inter-population variations were observed. Although the authors did not observe a multitude of pathogenic variants in medically relevant genes, a significant number of variants denoted as likely-pathogenic in other genes were present in the ClinVar database (Choudhury et al., 2020). Such a high degree of genomic variation complicates the general approach of Mendelian classifications for variant interpretation, as for many variants, the MAF is either not known or very low, immediately classifying them as a rare variant, possibly also absent from population databases such as GnomAD and others.

This complexity highlights the necessity of functional assays performed in parallel with haplotype analysis. Haplotype inference for SA based on NGS data was performed by Oosthuizen et al. (2021). The authors reported several variants at low frequency to be in linkage disequilibrium in specific SA population groups, which emphasized the importance of long-range PCR confirmation for phasing. Suspected benign missense variants co-segregating with pathogenic variants or SNV-based VUSes, albeit at a low frequency, could act as potential modifiers regarding disease penetrance. Even subtle influences such as these can possibly contribute to the value of risk scores unique to population groups.

Collectively, these factors contribute to variants not being fully classified as actionable due to a lack of evidence (using the ACMG classification criteria). This resulted in potentially actionable variants currently classified as VUSes because of a paucity of evidence in the international literature. Finally, despite the increase in throughput and the extended scope for *BRCA1/2* variant discovery, less than 10% of patients with a personal history of cancer diagnosed at an early age, or who had a positive family history, received a positive test result, leaving the remainder of patients still in the dark regarding alternative management and therapeutic options involving poly (ADP-ribose) (PARP) inhibitors. As various other genes associated with *BRCA1/2* have been indicated to contribute to homologous recombination and DNA repair (such as ATM Serine/Threonine Kinase [*ATM*], BRCA1 Interacting Helicase I [*BRIP1*], Checkpoint Kinase 2 [*CHEK2*], RAD51 Paralog C [*RAD51C*] and Partner And Localizer of BRCA2 [*PALB2*]), the search for actionable variants responsible for hereditary BC and/or OVC needs to be expanded. With NGS already implemented, we propose moving towards multigene panel testing in the future. It could result in the identification of additional role players contributing to the disease burden in SA. By performing multipanel testing, we will be able to identify deleterious variants in multiple cancer susceptibility genes, which will allow us to identify eligible patients and related family members for clinical interventions, surveillance screening, targeted therapy and potential prevention strategies, according to the National Comprehensive Cancer Network (NCCN) guidelines.

5 CONCLUSION

The vision of better health systems for African countries is encompassed in the health-related sustainable development goals set by the World Health Organization (WHO) for Africa in 2018 (World Health Organization Africa, 2021). The SA Department of Health has recognized this initiative and pledged to reform this critical sector by releasing updated

clinical guidelines for BC management and control during the same year. With the advances made in genetic testing for familial BC and OVC in state laboratories in SA, it has the potential to contribute immensely to the identification of high-risk *BRCA1/2* and non-*BRCA* germline actionable variants in patients. Given the magnitude of the disease, knowing a patient's mutation status can aid in the individualization of their treatment, which is of great benefit for the attending physician as it contributes to the patient's overall survival and will be of importance for related family members.

Due to the potential and far-reaching impact genetic testing has on augmenting the risk in a family with a positive family history of BC and/or OVC or in a patient diagnosed at an early age, it is imperative that the search is broadened to include other non-*BRCA* genes. Together, all these genes play an integral role in multiple signaling pathways inside the cell, with crosstalk between the associated proteins. If one of these signaling molecules becomes nonfunctional, the balance could be disturbed and may contribute to the progression of carcinogenesis.

The large number of novel and the abundance of heterozygous variants detected in this series reflect a high degree of genomic diversity. This highlights the existence of an immense gap in available naturally occurring population-specific knowledge due to a lack of African genomes in public genetic archives. Many diagnostic laboratories rely heavily on MAF and *in silico* predictions for variant interpretation and classification. In SA, as in many other African countries, this gap results in an unfavorable amount of VUSes classified. The lack of reference genomes increases the struggle to keep up with the rapid evolution in genetic variant screening for the confirmation of diagnosis. Although major strides have been made in the past decade in an attempt to catch up with first-world countries, uptake of genetic diagnostic services will not reach its full potential unless it becomes more affordable and a substantial number of African genomes is available to assist with variant interpretation and classification.

The SA scientific community is therefore compelled to continue with translational research in order to adopt first-world technologies into its healthcare context as a developing country. The vast progress made over the past two decades enabled a vital shift away from population-directed genetic testing to potentially comprehensive screening for all BC and OVC cancer patients. Consequently, the medical and scientific community in SA will continuously strive to comply with international standards and guidelines aimed at universal healthcare for all patients regardless of ethnicity, financial status or continent of birth. For centuries, the people of Africa have been marginalized and disadvantaged in many aspects, including optimal health care. With the WHO focusing on Africa, the health and well-being of its people are improving, resulting in the people of Africa currently sharing a vision for the future that is filled with optimism and hope.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary Material**.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Health Sciences Research Ethics Committee Faculty of Health Sciences University of the Free State. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

NM, JO, and HC made substantial contributions to the conception, design, and completion of this publication. NM obtained the ethics approvals and selected the data for inclusion. JO, HC, KN, and NM performed the genetic studies together with previous medical scientists involved in *BRCA* testing in this NHLS laboratory over the years. NM, JO and HC were responsible for the optimization of various techniques utilized. KN reviewed all the request forms for the collation of data and assisted with the population of an in-house database designed by HC. NM framed the study in relation to past achievements. All authors contributed to the article and approved the submitted version.

FUNDING

The translational research performed between 1995–2020 was supported by the South African Medical Research Council (SAMRC—self-initiated research grant 2012), with funds also received from the National Health Laboratory Service Research Trust (NHLSRT—2008, 2011, 2013, 2016, 2017) and the National Research Foundation (NRF—98291, 92192).

ACKNOWLEDGMENTS

The authors acknowledge all previous postgraduate students and medical scientists that participated in the translational research phase of the project or who were directly involved with the testing, namely N. Peter (MMedSc and NHLS Intern Medical Scientist), P. Moeti (MMedSc), C. Schultz, B. Dajee and A. Diedericks. Dr. Daleen Struwig, medical writer/editor, Faculty of Health Sciences, University of the Free State, for the final language editing of the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.834265/full#supplementary-material

REFERENCES

AbdelHamid, S. G., Zekri, A.-R. N., AbdelAziz, H. M., and El-Mesallamy, H. O. (2021). *BRCA1* and *BRCA2* Truncating Mutations and Variants of Unknown Significance in Egyptian Female Breast Cancer Patients. *Clinica Chim. Acta* 512, 66–73. doi:10.1016/j.cca.2020.11.023

Adedokun, B., Zheng, Y., Ndom, P., Gakwaya, A., Makumbi, T., Zhou, A. Y., et al. (2020). Prevalence of Inherited Mutations in Breast Cancer Predisposition Genes Among Women in Uganda and Cameroon. *Cancer Epidemiol. Biomarkers Prev.* 29, 359–367. doi:10.1158/1055-9965.EPI-19-0506

Agenbag, G. (2005). *Molecular Genetic Analysis of Familial Breast Cancer in South Africa*. Stellenbosch: MSc dissertationUniversity of Stellenbosch. Available at: https://scholar.sun.ac.za/handle/10019.1/1521 (Accessed November 16, 2021).

Allford, A., Qureshi, N., Barwell, J., Lewis, C., and Kai, J. (2014). What Hinders Minority Ethnic Access to Cancer Genetics Services and what May Help? *Eur. J. Hum. Genet.* 22, 866–874. doi:10.1038/ejhg.2013.257

Armstrong, K., Weber, B., Stopfer, J., Calzone, K., Putt, M., Coyne, J., et al. (2003). Early Use of Clinical *BRCA1/2* Testing: Associations with Race and Breast Cancer Risk. *Am. J. Med. Genet.* 117A, 154–160. doi:10.1002/ajmg.a.10928

Ashour, M., and Ezzat Shafik, H. (2019). Frequency of Germline Mutations in *BRCA1* and *BRCA2* in Ovarian Cancer Patients and Their Effect on Treatment Outcome. *Cmar* Vol. 11, 6275–6284. doi:10.2147/CMAR.S206817

Attlee, M. (1947). The Coloured People of South Africa. *Afr. Aff.* 46, 148–151. doi:10.1093/oxfordjournals.afraf.a093570

Awadelkarim, K. D., Aceto, G., Veschi, S., Elhaj, A., Morgano, A., Mohamedani, A. A., et al. (2007). *BRCA1* and *BRCA2* Status in a Central Sudanese Series of Breast Cancer Patients: Interactions with Genetic, Ethnic and Reproductive Factors. *Breast Cancer Res. Treat.* 102, 189–199. doi:10.1007/s10549-006-9303-z

Bakkach, J., Mansouri, M., Derkaoui, T., Loudiyi, A., El Fahime, E., Barakat, A., et al. (2020). Contribution of BRCA1 and BRCA2 Germline Mutations to Early Onset Breast Cancer: a Series from north of Morocco. *BMC Cancer* 20, 859. doi:10.1186/s12885-020-07352-9

Ben Ayed-Guerfali, D., Ben Kridis-Rejab, W., Ammous-Boukhris, N., Ayadi, W., Charfi, S., Khanfir, A., et al. (2021). Novel and Recurrent *BRCA1/BRCA2* Germline Mutations in Patients with Breast/ovarian Cancer: a Series from the South of Tunisia. *J. Transl Med.* 19, 108. doi:10.1186/s12967-021-02772-y

Bensam, M., Hafez, E., Awad, D., El-Saadani, M., and Balbaa, M. (2014). Detection of New Point Mutations of *BRCA1* and *BRCA2* in Breast Cancer Patients. *Biochem. Genet.* 52, 15–28. doi:10.1007/s10528-013-9623–810.1007/s10528-013-9623-8

Biancolella, M., Ouédraogo, N. L. M., Zongo, N., Zohoncon, T. M., Testa, B., Rizzacasa, B., et al. (2021). Breast Cancer in West Africa: Molecular Analysis of *BRCA* Genes in Early-Onset Breast Cancer Patients in Burkina Faso. *Hum. Genomics* 15, 65. doi:10.1186/s40246-021-00365-w

Boulenouar, A. C. S., Coulet, F., Bendiab, F. M. T., Boudinar, F. Z., and Senhadji, R. (2018). BRCA1 and BRCA2 Germline Mutation Screening in Western Algeria Using High Resolution Melting Analysis (HRM). *Gulf J. Oncolog* 1, 31–37.

Brandt, T., Sack, L. M., Arjona, D., Tan, D., Mei, H., Cui, H., et al. (2020). Adapting ACMG/AMP Sequence Variant Classification Guidelines for Single-Gene Copy Number Variants. *Genet. Med.* 22, 336–344. doi:10.1038/s41436-019-0655-2

Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A., and Jemal, A. (2018). Global Cancer Statistics 2018: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 countriesErratum in. *Ca: A. Cancer J. CliniciansCA Cancer J Clin* 6870, 394313–394424. doi:10.3322/caac.21492

Chen, W. (2015). *The Molecular Aetiology of Inherited Breast Cancer in the South African Black populationMScMed Dissertation*. Johannesburg: University of the Witwatersrand. Available at: http://hdl.handle.net/10539/19755 (Accessed September 11, 2021).

Cherbal, F., Bakour, R., Adane, S., Boualga, K., Benais-Pont, G., and Maillet, P. (2010). BRCA1andBRCA2Germline Mutations Screening in Algerian Breast/Ovarian Cancer Families. *Dis. Markers* 28, 377–384. doi:10.1155/2010/585278

Chistiakov, D. A., Voronova, N. V., and Chistiakov, P. A. (2008). Genetic Variations in DNA Repair Genes, Radiosensitivity to Cancer and Susceptibility to Acute Tissue Reactions in Radiotherapy-Treated Cancer Patients. *Acta Oncologica* 47, 809–824. doi:10.1080/02841860801885969

Choudhury, A., Aron, S., Botigué, L. R., Sengupta, D., Botha, G., Bensellak, T., et al. (2020). High-depth African Genomes Inform Human Migration and Health. *Nature* 586, 741–748. doi:10.1038/s41586-020-2859-7

Combrink, H. (2016). *Molecular Screening of the South African Indian Population for* BRCA1 *and* BRCA2 *Using High-Resolution Melting Analysis*. M.Med.Sc. Dissertation. Bloemfontein: University of the Free State. http://hdl.handle.net/11660/4563 (Accessed September 14, 2021).

Combrink, H. M., Oosthuizen, J., Visser, B., Chabilal, N., Buccimazza, I., Foulkes, W. D., et al. (2021). Mutations in *BRCA*-Related Breast and Ovarian Cancer in the South African Indian Population: a Descriptive Study. *Cancer Genet.* 258-259, 1–6. doi:10.1016/j.cancergen.2021.06.002

Conley, C. C., Ketcher, D., Reblin, M., Kasting, M. L., Cragun, D., Kim, J., et al. (2020). The Big Reveal: Family Disclosure Patterns of BRCA Genetic Test Results Among Young Black Women with Invasive Breast Cancer. *Jrnl of Gene Coun* 29, 410–422. doi:10.1002/jgc4.1196

El Ansari, F. Z., Jouali, F., Marchoudi, N., Bennani, M. M., Ghailani, N. N., Barakat, A., et al. (2020). Screening of *BRCA1/2* Genes Mutations and Copy Number Variations in Patients with High Risk for Hereditary Breast and Ovarian Cancer Syndrome (HBOC). *BMC Cancer* 20, 747. doi:10.1186/s12885-020-07250-0

El Khachibi, M., Diakite, B., Hamzi, K., Badou, A., Senhaji, M. A., Bakhchane, A., et al. (2015). Screening of Exon 11 of *BRCA1* Gene Using the High Resolution Melting Approach for Diagnosis in Moroccan Breast Cancer Patients. *BMC Cancer* 15, 81. doi:10.1186/s12885-015-1040-4

Elias, K. M., Guo, J., and Bast, R. C., Jr (2018). Early Detection of Ovarian Cancer. *Hematology/Oncology Clin. North America* 32, 903–914. doi:10.1016/j.hoc.2018.07.003

Elnour, A. M., Elderdery, A. Y., Mills, J., Mohammed, B. A., ElbietAbdelaal, D., and Mohamed, A. O. (2012). BRCA 1 and 2 Mutations in Sudanese Secondary School Girls with Known Breast Cancer in Their Families. *Ijhs* 6 (1), 63–71. doi:10.12816/0005974

Fourati, A., Louchez, M.-M., Fournier, J., Gamoudi, A., Rahal, K., El May, M.-V., et al. (2014). Screening for Common Mutations in *BRCA1* and *BRCA2* Genes: Interest in Genetic Testing of Tunisian Families with Breast And/or Ovarian Cancer. *Bull. du Cancer* 101, E36–E40. doi:10.1684/bdc.2014.2049

Francies, F. Z., WainsteinDe Leeneer, T., De Leeneer, K., Cairns, A., Murdoch, M., Nietz, S., et al. (2015). BRCA1, BRCA2 and PALB2 Mutations and CHEK2 c.1100delC in Different South African Ethnic Groups Diagnosed with Premenopausal And/or Triple Negative Breast Cancer. *BMC Cancer* 15, 912–921. doi:10.1186/s12885-015-1913-6

Gomaa Mogahed, S. H., Hamed, Y. S., Ibrahim Moursy, Y. E., and Mahomoud Saied, M. H. (2020). Analysis of Heterozygous *BRCA1* 5382ins Founder Mutation in a Cohort of Egyptian Breast Cancer Female Patients Using Pyrosequencing Technique. *Asian Pac. J. Cancer Prev.* 21, 431–438. doi:10.31557/APJCP.2020.21.2.431

Hadiji-Abbes, N., Trifa, F., Choura, M., Khabir, A., Sellami-Boudawara, T., Frikha, M., et al. (2015). A Novel BRCA2 in Frame Deletion in a Tunisian Woman with Early Onset Sporadic Breast Cancer. *Pathologie Biologie* 63, 185–189. doi:10.1016/j.patbio.2015.07.009

Haga, S. B., Barry, W. T., Mills, R., Ginsburg, G. S., Svetkey, L., Sullivan, J., et al. (2013). Public Knowledge of and Attitudes toward Genetics and Genetic Testing. *Genet. Test. Mol. Biomarkers* 17, 327–335. doi:10.1089/gtmb.2012.0350

Hall, D., Wijsman, E. M., Roos, J. L., Gogos, J. A., and Karayiorgou, M. (2002). Extended Intermarker Linkage Disequilibrium in the Afrikaners. *Genome Res.* 12, 956–961. doi:10.1101/gr.136202

Hamdi, Y., Mighri, N., Boujemaa, M., Mejri, N., Nasr, S. N., Rekaya, M. B., et al. (2021). Identification of Eleven Novel BRCA Mutations in Tunisia: Impact on the Clinical Management of *BRCA* Related Cancers. *Res. Sq.* doi:10.3389/fonc.2021.674965)

Henneman, L., Vermeulen, E., van El, C. G., Claassen, L., Timmermans, D. R. M., and Cornel, M. C. (2013). Public Attitudes towards Genetic Testing Revisited: Comparing Opinions between 2002 and 2010. *Eur. J. Hum. Genet.* 21, 793–799. doi:10.1038/ejhg.2012.271

Henouda, S., Bensalem, A., Reggad, R., Serrar, N., Rouabah, L., and 1 Pujo, P. (2016). Contribution of *BRCA1* and *BRCA2* Germline Mutations to Early Algerian Breast Cancer. *Dis. Markers*. doi:10.1155/2016/7869095)

Hollfelder, N., Erasmus, J. C., Hammaren, R., Vicente, M., Jakobsson, M., Greeff, J. M., et al. (2020). Patterns of African and Asian Admixture in the Afrikaner Population of South Africa. *BMC Biol.* 18, 1–16. doi:10.1186/s12915-020-0746-1

Isaacs, S., Geduld-Ullah, T., and Benjeddou, M. (2013). Reconstruction of Major Maternal and Paternal Lineages of the Cape Muslim Population. *Genet. Mol. Biol.* 36, 167–176. doi:10.1590/S1415-47572013005000019

Joffe, M., Ayeni, O., Norris, S. A., McCormack, V. A., Ruff, P., Das, I., et al. (2018). Barriers to Early Presentation of Breast Cancer Among Women in Soweto, South Africa. *PLoS ONE* 13, e0192071. doi:10.1371/journal.pone.0192071

Jouali, F., Laarabi, F.-Z., Marchoudi, N., Ratbi, I., Elalaoui, S. C., Rhaissi, H., et al. (2016). First Application of Next-Generation Sequencing in Moroccan Breast/ovarian Cancer Families and Report of a Novel Frameshift Mutation of the *BRCA1* Gene. *Oncol. Lett.* 12, 1192–1196. doi:10.3892/ol.2016.4739

Kan, C., and Zhang, J. (2015). *BRCA1* Mutation: a Predictive Marker for Radiation Therapy? *Int. J. Radiat. Oncology*Biology*Physics* 93, 281–293. doi:10.1016/j.ijrobp.2015.05.037

Kopanos, C., Tsiolkas, V., Kouris, A., Chapple, C. E., Albarca Aguilera, M., Meyer, R., et al. (2019). VarSome: the Human Genomic Variant Search Engine. *Bioinformatics* 35, 1978–1980. doi:10.1093/bioinformatics/bty897

Kuchenbaecker, K. B., Hopper, J. L., Barnes, D. R., Phillips, K.-A., Mooij, T. M., Roos-Blom, M.-J., et al. (2017). Risks of Breast, Ovarian, and Contralateral Breast Cancer for *BRCA1* and *BRCA2* Mutation Carriers. *JAMA* 317, 2402–2416. doi:10.1001/jama.2017.7112

Laarabi, F.-Z., Ratbi, I., Elalaoui, S. C., Mezzouar, L., Doubaj, Y., Bouguenouch, L., et al. (2017). High Frequency of the Recurrent c.1310_1313delAAGA *BRCA2* Mutation in the North-East of Morocco and Implication for Hereditary Breast-Ovarian Cancer Prevention and Control. *BMC Res. Notes* 10, 188. doi:10.1186/s13104-017-2511-2

Laraqui, A., Uhrhammer, N., Lahlou-Amine, I., Rhaffouli, H. E., Baghdadi, J. E., Dehayni, M., et al. (2013). Mutation Screening of the *BRCA1* Gene in Early Onset and Familial Breast/Ovarian Cancer in Moroccan Population. *Int. J. Med. Sci.* 10, 60–67. doi:10.7150/ijms.5014

Larsen, M. J., Thomassen, M., Gerdes, A-M., and Kruse, T. A. (2014). *Breast Cancer: Basic Clin Res.* doi:10.4137/bcbcr.s18715Hereditary Breast Cancer: Clinical, Pathological and Molecular Characteristics)

Loizzi, V., Ranieri, G., Laforgia, M., Gadaleta, C., Gargano, G., Kardhashi, A., et al. (2020). PARP Inhibitors and Epithelial Ovarian Cancer: Molecular Mechanisms, Clinical Development and Future Prospective (Review). *Oncol. Lett.* 20, 1. doi:10.3892/ol.2020.11951

Mahfoudh, W., Bouaouina, N., Ahmed, S. B., Gabbouj, S., Shan, J., Mathew, R., et al. (2012). Hereditary Breast Cancer in Middle Eastern and North African (MENA) Populations: Identification of Novel, Recurrent and Founder *BRCA1* Mutations in the Tunisian Population. *Mol. Biol. Rep.* 39, 1037–1046. doi:10.1007/s11033-011-0829-8

Mampunye, L., van der Merwe, N. C., Grant, K. A., Peeters, A. V., Torrorey-Sawe, R., French, D. J., et al. (2021). Pioneering BRCA1/2 Point-Of-Care Testing for Integration of Germline and Tumor Genetics in Breast Cancer Risk Management: A Vision for the Future of Translational Pharmacogenomics. *Front. Oncol.* 11, 619817. doi:10.3389/fonc.2021.619817

Mansouri, M., Derkaoui, T., Bakkach, J., Loudiyi, A., Ghailani Nourouti, N., Barakat, A., et al. (2020). Screening of BRCA1 and BRCA2 Germline Mutations in Unselected Triple-negative Breast Cancer Patients: A Series from north of Morocco. *Precision Med. Sci.* 9, 43–48. doi:10.1002/prm2.12009

McCarthy, A. M., and Armstrong, K. (2014). The Role of Testing forBRCA1andBRCA2Mutations in Cancer Prevention. *JAMA Intern. Med.* 174 (7), 1023–1024. doi:10.1001/jamainternmed.2014.1322

Mehemmai, C., Cherbal, F., Hamdi, Y., Guedioura, A., Benbrahim, W., Bakour, R., et al. (2020). *BRCA1* and *BRCA2* Germline Mutation Analysis in Hereditary Breast/ovarian Cancer Families from the Aures Region (Eastern Algeria): First Report. *Pathol. Oncol. Res.* 26, 715–726. doi:10.1007/s12253-019-00586-4

Mesthrie, R. (2006). World Englishes and the Multilingual History of English. *World Englishes* 25, 381–390. doi:10.1111/j.1467-971x.2006.00477.x

Mestiri, S., Monastiri, K., Bouaouina, N., Presnea, N., Bignon, Y. J., Khairi, H., et al. (2000). *Mutational Analysis of Breast/ovarian Cancer Hereditary Predisposition Gene* BRCA1 *in Tunisian Women*, 77. Arch Institut Pasteur Tunis, 11–15.

Mighri, N., Hamdi, Y., Boujemaa, M., Othman, H., Nasr, S. B., El Benna, H., et al. (2020). Identification of Novel *BRCA1* and *RAD50* Mutations Associated with Breast Cancer Predisposition in Tunisian Patients. *Front. Genet.* doi:10.3389/fgene.2020.552971

Moeti, P. J. (2016). *Molecular Screening for the Presence of Large Deletions or Duplications in* BRCA *Using Multiplex Ligation-dependent Probe Amplification in South Africa. M.Med.Sc. Dissertation.*. Bloemfontein: University of the Free State. Available at: https://scholar.ufs.ac.za/handle/11660/5392 (Accessed September 14, 2021).

Msolly, A., and Asma, K. (2015). *BRCA1* and *BRCA2* Mutations Are They Related to Breast Cancer in a Sample of Tunisian Population? *Ctoij* 1. doi:10.19080/CTOIJ.2015.01.555551

Mukem, S., Meng, Q., Sriplung, H., and Tangcharoensathien, V. (2016). Low Coverage and Disparities of Breast and Cervical Cancer Screening in Thai Women: Analysis of National Representative Household Surveys. *Asian Pac. J. Cancer Prev.* 16, 8541–8551. doi:10.7314/apjcp.2015.16.18.8541

National Department of Health of the Republic of South Africa (2018). *Clinical Guidelines for Breast Cancer Control and Management*. Pretoria, South Africa: Department of Health, 1–123. Available at: https://cansa.org.za/files/2019/08/DOH-Breast-Cancer-Guidelines-Final.pdf.

Ndiaye, R., Diop, J. P. D., Bourdon-Huguenin, V., Dem, A., Diouf, D., Dieng, M. M., et al. (2020). Evidence for an Ancient *BRCA1* Pathogenic Variant in Inherited Breast Cancer Patients from Senegal. *Npj Genom. Med.* 5, 8. doi:10.1038/s41525-020-0114-7

Oosthuizen, J. (2016). *Molecular Screening of Coloured South African Breast Cancer Patients for the Presence of* BRCA *Mutations Using High Resolution Melting analysis. MMedSc Dissertation*. Bloemfontein: University of the Free State. Available at: https://scholar.ufs.ac.za/handle/11660/6426 (Accessed September 13, 2021).

Oosthuizen, J., Kotze, M. J., van der Merwe, N., Myburgh, E. J., Bester, P., and van der Merwe, N. C. (2021). Globally Rare *BRCA2* Variants with Founder Haplotypes in the South African Population: Implications for point-of-care Testing Based on a Single-Institution *BRCA1/2* Next-Generation Sequencing Study. *Front. Oncol.* 10. doi:10.3389/fonc.2020.619469

Opolade, O. I., Gao, Q., Adebamowo, C. A., Fackenthal, J., Das, S., Sveen, L., et al. (2000). Protein Truncating *BRCA1* and *BRCA2* Mutations in African Women with Pre-menopausal Breast Cancer. *Hum. Genet.* 107, 192–194. doi:10.1007/s004390000342

Peltzer, K., and Phaswana-Mafuya, N. (2014). Breast and Cervical Cancer Screening and Associated Factors Among Older Adult Women in South Africa. *Asian Pac. J. Cancer Prev.* 15, 2473–2476. doi:10.7314/apjcp.2014.15.6.2473

Peter, N. (2014). *Screening of Young And/or Familial African Breast Cancer Patients for the Presence of* BRCA *Mutations. M.Med.Sc. Dissertation*. Bloemfontein: University of the Free State. Available at: https://scholar.ufs.ac.za/handle/11660/1613 (Accessed September 14, 2021)

Phaswana-Mafuya, N., and Peltzer, K. (2018). Breast and Cervical Cancer Screening Prevalence and Associated Factors Among Women in the South African General Population. *Asian Pac. J. Cancer Prev.* 19, 1465–1470. doi:10.22034/apjcp.2018.19.6.1465

Reeves, M. D., Yawitch, T. M., Van der Merwe, N. C., Van den Berg, H. J., Dreyer, G., and Van Rensburg, E. J. (2004). *BRCA1* Mutations in South African Breast And/or Ovarian Cancer Families: Evidence of a Novel Founder Mutation in Afrikaner Families. *Int. J. Cancer* 110, 677–682. doi:10.1002/ijc.20186

Riahi, A., Ghourabi, M. e., Fourati, A., and Chaabouni-Bouhamed, H. (2016). Family History Predictors of BRCA1/BRCA2 Mutation Status Among Tunisian Breast/ovarian Cancer Families. *Breast Cancer* 24, 238–244. doi:10.1007/s12282-016-0693-4

Riahi, A., Kharrat, M., Ghourabi, M. E., Khomsi, F., Gamoudi, A., Lariani, I., et al. (2014). Mutation Spectrum and Prevalence ofBRCA1 andBRCA2 Genes in Patients with Familial and Early-Onset Breast/ovarian Cancer from Tunisia. *Clin. Genet.* 87, 155–160. doi:10.1111/cge.12337

Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., et al. (2015). Standards and Guidelines for the Interpretation of Sequence Variants: a Joint Consensus Recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet. Med.* 17, 405–424. doi:10.1038/gim.2015.30

Salmi, F., Maachi, F., Tazzite, A., Aboutaib, R., Fekkak, J., Azeddoug, H., et al. (2021). Next-generation Sequencing of *BRCA1* and *BRCA2* Genes in Moroccan Prostate Cancer Patients with Positive Family History. *PLoS ONE* 16, e0254101. doi:10.1371/journal.pone.0254101

Schlebusch, C. M., Dreyer, G., Sluiter, M. D., Yawitch, T. M., van den Berg, H. J., and van Rensburg, E. J. (2010). Cancer Prevalence in 129 Breast-Ovarian Cancer Families Tested for *BRCA1* and *BRCA2* Mutations. *S Afr. Med. J.* 100, 113–117. doi:10.7196/samj.3235

Schoeman, M., Apffelstaedt, J. P., Baatjes, K., and Urban, M. (2013). Implementation of a Breast Cancer Genetic Service in South Africa - Lessons Learned. *S Afr. Med. J.* 103, 529–533. doi:10.7196/SAMJ.6814

Seymour, H. J., Wainstein, T., Macaulay, S. S., Haw, T., and Krause, A. (2016). Breast Cancer in High-Risk Afrikaner Families: Is *BRCA* Founder Mutation Testing Sufficient? *S Afr. Med. J.* 106, 264–267. doi:10.7196/SAMJ.2016.v106i3.10285

Sluiter, M. D., and Van Rensburg, E. J. (2011). Large Genomic Rearrangements of the *BRCA1* and *BRCA2* Genes: Review of the Literature and Report of a Novel *BRCA1* Mutation. *Breast Cancer Res. Treat.* 125, 325–349. doi:10.1007/s10549-010-0817-z

Sözmen, K., Unal, B., Sakarya, S., Dinc, G., Yardim, N., Keskinkilic, B., et al. (2016). Determinants of Breast and Cervical Cancer Screening Uptake Among Women in Turkey. *Asia Pac. J. Public Health* 28, 528–538. doi:10.1177/1010539516654541

Spurdle, A. B., and Jenkins, T. (1996). The Origins of the Lemba "Black Jews" of Southern Africa: Evidence from p12F2 and Other Y-Chromosome Markers. *Am. J. Hum. Genet.* 59, 1126–1133.

Sunada, S., Nakanishi, A., and Miki, Y. (2018). Crosstalk of DNA Double-strand Break Repair Pathways in poly(ADP-ribose) Polymerase Inhibitor Treatment of Breast Cancer Susceptibility Gene 1/2-mutated Cancer. *Cancer Sci.* 109, 893–899. doi:10.1111/cas.13530

Tazzite, A., Jouhadi, H., Nadifi, S., Aretini, P., Falaschi, E., Collavoli, A., et al. (2012). *BRCA1* and *BRCA2* Germline Mutations in Moroccan Breast/ovarian Cancer Families: Novel Mutations and Unclassified Variants. *Gynecol. Oncol.* 125, 687–692. doi:10.1016/j.ygyno.2012.03.007

The Council for Medical Schemes (2019). *Final PMB Definition Guidelines for Ovarian Cancer*. Pretoria, South Africa: Council for Medical Schemes, 1–27. Available at: https://www.medicalschemes.com/files/PMB/Project/FinalPMBdefinitionguidelineforovariancancer.pdf.

Theme Filha, M. M., Leal, M. d. C., Oliveira, E. F. V. d., Esteves-Pereira, A. P., and Gama, S. G. N. d. (2016). Regional and Social Inequalities in the Performance of Pap Test and Screening Mammography and Their Correlation with Lifestyle: Brazilian National Health Survey, 2013. *Int. J. Equity Health* 15, 136. doi:10.1186/s12939-016-0430-9

Troiano, G., Nante, N., and Cozzolino, M. (2017). The Angelina Jolie Effect - Impact on Breast and Ovarian Cancer Prevention A Systematic Review of Effects after the Public Announcement in May 2013. *Health Edu. J.* 76, 707–715. doi:10.1177/0017896917712300

Troudi, W., Loueslati, B., Baccar, A., Ben Ayed, F., and Ben Ammar El Gaaied, A. (2009). Penetrance of BRCA1 Gene Mutation and DNA Mitochondrial in Tunisian Breast Cancer Occurrence. *Tunis Med.* 87, 494–498.

Troudi, W., Uhrhammer, N., Sibille, C., Dahan, C., Mahfoudh, W., Bouchlaka Souissi, C., et al. (2007). Contribution of the BRCA1 and BRCA2 Mutations to Breast Cancer in Tunisia. *J. Hum. Genet.* 52, 915–920. doi:10.1007/s10038-007-0195-5

Van der Merwe, N. C., and van Rensburg, E. J. (2009). Hereditary Breast/ovarian Cancer and BRCA Mutations: A South African Perspective. *Curr. Oncol.* 16, 91.

Van der Merwe, N. C., Oosthuizen, J., Theron, M., Chong, G., and Foulkes, W. D. (2020). The Contribution of Large Genomic Rearrangements in *BRCA1* and *BRCA2* to South African Familial Breast Cancer. *BMC Cancer* 20, 391. doi:10.1186/s12885-020-06917-y

Van der Merwe, N., Hamel, N., Schneider, S.-R., Apffelstaedt, J., Wijnen, J., and Foulkes, W. (2012). A Founder *BRCA2* Mutation in Non-afrikaner Breast Cancer Patients of the Western Cape of South Africa. *Clin. Genet.* 81, 179–184. doi:10.1111/j.1399-0004.2010.01617.x

Vishnu, P., and Morrell, R. (1991). Indian Merchants and Dukawallahs in the Natal Economy, C1875-1914. *J. South. Afr. Stud.* 17, 71–102. https://www.jstor.org/stable/2637287.

World Health Organization Africa *The State of Health in the WHO African Region: An Analysis of the Status of Health*. Available at: https://reliefweb.int/report/world/state-heatlh-who-african-region-analysis-status-health-services-and-health (Accessed November 14, 2021).

Yawitch, T. M. (2005). *Inherited Predisposition to Breast/ovarian Cancer: The Frequency and Nature of* BRCA1 *Gene Mutations in South African familiesMScMed Dissertation*. Pretoria: University of Pretoria. Available at: https://repository.up.ac.za/handle/2263/30470 (Accessed September 14, 2021).

Yoshida, K., and Miki, Y. (2004). Role of *BRCA1* and *BRCA2* as Regulators of DNA Repair, Transcription, and Cell Cycle in Response to DNA Damage. *Cancer Sci.* 95, 866–871. doi:10.1111/j.1349-7006.2004.tb0219510.1111/j.1349-7006.2004.tb02195.x

Zhang, B., Fackenthal, J. D., Niu, Q., Huo, D., Sveen, W. E., AdebamowoDeMarco, C. A., et al. (2009). Evidence for an Ancient *BRCA1* Mutation in Breast Cancer Patients of Yoruban Ancestry. *Fam. Cancer* 8, 15–22. doi:10.1007/s10689-008-9205-9

Zhang, J., Fackenthal, J. D., Zheng, Y., Huo, D., Hou, N., Niu, Q., et al. (2012). Recurrent *BRCA1* and *BRCA2* Mutations in Breast Cancer Patients of African Ancestry. *Breast Cancer Res. Treat.* 134, 889–894. doi:10.1007/s10549-012-2136-z

Zheng, Y., Walsh, T., Gulsuner, S., Casadei, S., Lee, M. K., Ogundiran, T. O., et al. (2018). Inherited Breast Cancer in Nigerian Women. *Jco* 36, 2820–2825. doi:10.1200/JCO.2018.78.3977

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Genetics

PERSPECTIVE
published: 14 April 2022
doi: 10.3389/fgene.2022.864725

Hepatic Models in Precision Medicine: An African Perspective on Pharmacovigilance

Tracey Hurrell[1†], *Jerolen Naidoo*[1†] *and Janine Scholefield*[1,2*]

[1]*Bioengineering and Integrated Genomics Group, Next Generation Health Cluster, Council for Scientific and Industrial Research, Pretoria, South Africa,* [2]*Department of Human Biology, Faculty of Health Sciences, University of Cape Town, Cape Town, South Africa*

Pharmaceuticals are indispensable to healthcare as the burgeoning global population is challenged by diseases. The African continent harbors unparalleled genetic diversity, yet remains largely underrepresented in pharmaceutical research and development, which has serious implications for pharmaceuticals approved for use within the African population. Adverse drug reactions (ADRs) are often underpinned by unique variations in genes encoding the enzymes responsible for their uptake, metabolism, and clearance. As an example, individuals of African descent (14–34%) harbor an exclusive genetic variant in the gene encoding a liver metabolizing enzyme (CYP2D6) which reduces the efficacy of the breast cancer chemotherapeutic Tamoxifen. However, CYP2D6 genotyping is not required prior to dispensing Tamoxifen in sub-Saharan Africa. Pharmacogenomics is fundamental to precision medicine and the absence of its implementation suggests that Africa has, to date, been largely excluded from the global narrative around stratified healthcare. Models which could address this need, include primary human hepatocytes, immortalized hepatic cell lines, and induced pluripotent stem cell (iPSC) derived hepatocyte-like cells. Of these, iPSCs, are promising as a functional *in vitro* model for the empirical evaluation of drug metabolism. The scale with which pharmaceutically relevant African genetic variants can be stratified, the expediency with which these platforms can be established, and their subsequent sustainability suggest that they will have an important role to play in the democratization of stratified healthcare in Africa. Here we discuss the requirement for African hepatic models, and their implications for the future of pharmacovigilance on the African continent.

Keywords: pharmacovigilance, iPSCs, hepatocyte, African precision medicine, ADRs

OPEN ACCESS

Edited by:
Tivani Phosa Mashamba-Thompson, University of Pretoria, South Africa

Reviewed by:
Jim Kaput, Independent researcher

***Correspondence:**
Janine Scholefield
JScholefield@csir.co.za

†These authors have contributed equally to this work

Specialty section:
This article was submitted to Human and Medical Genomics, a section of the journal Frontiers in Genetics

Received: *28 January 2022*
Accepted: *29 March 2022*
Published: *14 April 2022*

Citation:
Hurrell T, Naidoo J and Scholefield J (2022) Hepatic Models in Precision Medicine: An African Perspective on Pharmacovigilance. Front. Genet. 13:864725. doi: 10.3389/fgene.2022.864725

GENETIC DIVERSITY IN AFRICA AND ITS IMPACT ON HEALTHCARE

The global population is becoming increasingly reliant on pharmaceuticals as healthcare systems are burdened by lifestyle diseases, drug resistance, mental health issues, and innumerable orphan diseases. While substantial progress has been made to ensure patients in Africa have access to treatments, the majority of pharmaceuticals are trialed on narrow genetic populations using both preclinical models and population cohorts that are not inclusive of global, especially African, diversity. This is well exemplified by the fact that the African continent currently contributes only ~3.3% of the estimated 393 000 active clinical trials globally (ClinicalTrials.gov, 2021). Yet, despite its

disproportionately high disease burden, Africa tolerates suboptimal treatment regimens which affect a significant portion of the continent (Mpye et al., 2017).

It has long been understood that genetic variants in drug metabolizing genes are a major factor resulting in altered drug responses (Weinshilboum, 2003; Schärfe et al., 2017). Consequently, the extraordinary genetic diversity in the African population, compounded by the underrepresentation of this diversity in preclinical development, is likely to be driving a high prevalence of severe adverse drug reactions (ADRs) within the populus (Rajman et al., 2017). This may in turn decrease adherence to treatment regimens, often in vulnerable population groups. Understanding the magnitude of this issue is further complicated by the likely under-reporting of African ADRs on Vigibase - the global Individual Case Safety Report database (Ampadu et al., 2016). Despite increasing inclusion in studies/resources such as the Genome Aggregation Database (gnomAD), the 1000 Genomes Project, the African Genome Variation Project (AGVP), and the H3Africa project (Gurdasani et al., 2015), the clinical impact of Africa's genetic diversity on ADRs is only now coming to the forefront (da Rocha, 2021). While still striving to address the inequality in the global representation of African genomes (Wonkam, 2021), how then do we best serve the people of Africa, and successfully implement precision medicine on the continent?

In the context of ADRs, precision medicine aims to maximize the probability of clinical success, by matching patients with predicted best treatment options, either for individuals (personalized medicine) or well-defined subsets of the population (stratified medicine). The mitigation of ADRs through regulatory interventions represents a valuable and tenable strategy for the implementation of stratified medicine within Africa, especially for drugs that are highly utilized across the continent. These regulatory changes must first be driven by rigorous scientific efforts to bridge the disconnect that exists between the identification of African relevant genetic variants, and the functional validation of their potential clinical impact. Here we outline why we believe cellular hepatic models would be a useful tool to validate African relevant gene-drug interactions.

THE ROLE OF THE LIVER IN PHARMACOVIGILANCE

Hepatocytes, accounting for 60–70% of the total liver cell population, play a central role in the biotransformation, intermediary and energetic metabolism of endogenous and exogenous xenobiotics, including pharmaceuticals (LeCluyse et al., 2012). Hepatic metabolism of xenobiotics to more polar and hydrophilic counterparts facilitate their excretion and prevents toxic accumulation. This specialized detoxification machinery relies on the selectivity, abundance, expression, and interplay of sequentially coordinated (Phase 0, I, II and III) metabolizing enzymes (Zhang and Surapaneni, 2012). The CYP450 family, is a membrane-associated, hydrophobic enzyme system, functionally intended to be cyto-protective. However, CYP450 enzymes are also responsible for forming pharmaceutically active metabolites of prodrugs, and can generate toxic reactive intermediary products (Liebler and Guengerich, 2005).

CYP450 enzymes are of relevance to pharmaceuticals as their induction/inhibition potential, genetic polymorphisms, epigenetic regulation and non-genetic host factors such as gender, age, disease(s), and polypharmacy can contribute to functional disparities. CYP450 enzymes are responsible for approximately 80% of hepatic Phase I metabolism, with 12 isoforms being responsible for ~75% of Phase I oxidation reactions (Evans and Relling, 1999). In a literature survey for drug metabolism pathways with known CYP450 involvement (248 drugs), the fraction of clinically used drugs metabolized by CYP450 isoforms approximates the following: 3A4/5: 30.2%, 2D6: 20%, 2C9: 12.8%, 1A2: 8.9%, 2B6: 7.2%, and 2C19: 6.8%. However enzyme functionality does not correlate linearly with hepatic enzyme expression, for example, CYP2D6 accounts for 1.3–4.3% of the hepatic pool and metabolizes 20% of clinically used drugs, whereas CYP2B6 accounts for 1.7–5.3% while metabolizing ~7% (Zanger and Schwab, 2013).

The functional consequence of genetic polymorphisms leads to the classification of pharmacokinetic phenotypes as poor, intermediate, extensive/normal, and ultra-rapid metabolizers. Thus, contextualizing major genetic determinants of drug metabolism in different populations is essential to the provision of safe and efficacious pharmaceuticals (Belle and Singh, 2008). Investment in, and implementation of, national genomic-medicine initiatives is driving transformation in healthcare (Stark et al., 2019). However, the adoption of pharmacogenetics testing in sub-Saharan Africa (SSA) faces numerous clinical, scientific, technical, socio-economic, and governance barriers (Tata et al., 2020).

CLINICAL IMPLICATIONS AND CHALLENGES OF GENOTYPING PHARMACOGENES

Given the genetic diversity in Africa, executing precision medicine strategies could be considered beyond the capabilities of the healthcare infrastructure, in the predominantly developing nations which constitute the region. Yet, there are well defined examples of gene-drug pairs where ADRs could be minimized or resolved thereby significantly alleviating socio-economic impact.

One example of such a gene-drug pair, for a non-communicable disease, is that of the breast cancer drug Tamoxifen and its key liver metabolizing gene. Tamoxifen is a selective estrogen receptor modulator, which relies on CYP2D6 to catalyze the formation of primary and secondary metabolites which have higher anti-estrogenic activity. The CYP2D6 gene locus is highly polymorphic, with the allelic variant CYP2D6*17 prevalent in 21.7% of the African population [14–34% in African sub-cohorts (Nemaura et al., 2012; Masimirembwa and Hasler, 2013)]; yet it is nearly absent in Europeans, Asian, and admix American populations. African individuals harboring CYP2D6*17 have decreased enzyme expression and activity, resulting in reduced efficacy and increased ADRs, consequently promoting the risk of cancer recurrence and

diminishing the likelihood of positive clinical outcomes (Higgins and Stearns, 2010).

Another concerning gene-drug pair is that of CYP2B6 and the antiretroviral Efavirenz. CYP2B6 variants, with a higher prevalence in the African population, include functionally deficient haplotypes CYP2B6*6 and CYP2B6*18 (Langmia, 2021). In comparison to European populations, even a small increase in the prevalence of these variants in the context of SSA has a significant impact. These poor metabolizers when administered Efavirenz, which is used as first-line antiretroviral therapy for adults in South Africa, have elevated plasma concentrations and decreased clearance which can precipitate severe neurotoxicities including catatonia, suicidal ideation, and psychosis (Masimirembwa and Hasler, 2013). With the prevalence of HIV in European and SSA populations at 5 and 25% respectively, it is easy to see the positive impact that a pharmacovigilance-based precision medicine strategy would have on the disease burden in SSA.

Here we outlined two of the many well-established gene-drug pairs within the African context. Yet, data from the 1,000 Genomes Project indicates that SSA contains 25% more genetic diversity than the rest of the world (1000 Genomes Project Consortium 2015) rendering the absolute requirement for studies which validate the relationship(s) between host genetic and pharmaceutical interactions. Furthermore, since genes such as CYP2D6 and CYP2B6 are responsible for the metabolism of ~27% of clinically approved drugs (Zanger and Schwab, 2013), it begs the question, how can we model the contribution of these and other genetic variants to ADRs and in doing so, provide best treatment outcomes.

The African Pharmacogenomics Consortium (APC) was launched in 2018, with the mandate to educate, build capacity, capability, governance, and technologies to promote the use of pharmacogenomics for the clinical benefit of African patients (Dandara, 2019). However, pharmacogenetic screening using commercial genotyping applications have historically been ineffective due to variants which are exclusive to the African population not being adequately represented (Dodgen et al., 2013). Driving impact into the clinical space, by the APC and other such initiatives, will require sustained contributions from stakeholders and investment in local population-relevant biotechnology. Given that pharmacogenomics is fundamental to precision medicine, and that screening for genetic variants is not yet considered a point-of-care prior to dispensing drugs (Tata et al., 2020), Africa needs to invest further in technologies which directly contribute to stratified healthcare for its people. With clinical trials incurring the largest financial expenditure of the drug development pipeline, technologies which provide an opportunity to screen drugs within a genetically and physiologically relevant background must be established.

HEPATOCYTE MODELS IN PRECISION MEDICINE

Safety pharmacology informs risk-benefit relationships by assessing adverse effect liability and safety margins. Hepatotoxicity and aberrant xenobiotic metabolism are major contributors to post-marketing drug withdrawal, failure of investigational new drugs, and drug inefficacy. Population risk-benefit ratios, derived from clinical trial data, inform regulatory decisions despite knowing that patient-level responses will differ in terms of efficacy and adverse outcome risk. Importantly, as the contribution of an individual's genetics to the risk-benefit relationship is better understood it is possible that reducing attrition rates during development, using the current testing paradigms, will not be synonymous with global improvement in patient outcomes (Ahuja and Sharma, 2014; Dambach et al., 2016; Atienzar and Nicolas, 2018; Babai et al., 2018).

Predicting treatment inefficacy and ADRs is becoming more challenging as our understanding of the impact which inter-ethnic and inter-individual genetics has on healthcare improves. Establishing effective strategies to validate the impact of Africa's unique population genetics on clinical outcomes is imperative for improved pharmacovigilance, and the successful implementation of precision medicine on the African continent. While numerous preclinical models of xenobiotic metabolism exist, those with the potential to validate the impact of individual, African-relevant genetic variants on xenobiotic metabolism are lacking. Preclinical hepatocyte and liver models include: 1) primary human hepatocytes (PHHs); 2) subcellular fractions including microsomes or S9 fractions; 3) cell lines; 4) *ex vivo* tissue slices; 5) isolated perfused organs; 6) genetically engineered cells expressing metabolizing enzymes; 7) cell-free and in silico approaches; and 8) animal models (Zeilinger et al., 2016; Yamasaki et al., 2020). However, ethical, economic and practical considerations dictate the choice of model, with trade-offs between functional complexity, applicability, cost, scalability, expedience, and accessibility (Costa et al., 2014). The focus here will be on the potential use case for *in vitro* cell-based models of xenobiotic metabolism.

Primary Human Hepatocytes

PHHs are the "gold standard" or "historical standard" for modeling xenobiotic metabolism (LeCluyse et al., 2012). PHHs express all major Phase I and Phase II drug-metabolizing enzymes and transporter proteins at functional levels, with their value being exemplified by their application as models of drug-induced liver injury (Bell et al., 2016), liver metabolism (Vorrink et al., 2017), and liver diseases; including cholestasis (Hendriks et al., 2016), steatosis (Kozyra et al., 2018; Prill et al., 2019) and fibrosis (Pingitore et al., 2019; Hurrell et al., 2020). In this respect, PHHs have a clear advantage over other cellular models in terms of the fidelity with which they can recapitulate liver physiology and function (Ingelman-Sundberg and Lauschke, 2022).

PHHs are however prone to rapidly declining metabolic function and dedifferentiate, within hours, in traditional monolayer cultures (Lauschke et al., 2016; Heslop et al., 2017). The adoption of 3-dimensional (3D) culture methods has largely overcome this limitation, improving the functional longevity of PHHs in culture (>21 days) through enhanced mimicry of endogenous architecture and physiological cues. This has been

achieved using 3D spheroid models (Bell et al., 2016; Vorrink et al., 2017), microfluidic liver-on-a-chip systems (Beckwitt et al., 2018), microscale bioreactors (Freyer et al., 2018), and 3D bioprinted livers (Otieno et al., 2018).

However, donor age, genetics and pathophysiology, as well as the impact of cryopreservation on attachment efficiency/viability and consequently their functional properties, are important factors to consider when using PHHs (Ölander et al., 2019). Furthermore, the inability of PHHs to self-renew and their inaccessibility to the broader research community, present significant challenges to the scalability and use of PHH models, especially for research applications which require population-level interrogation of biological phenotypes (Otieno et al., 2018; Lauschke et al., 2019).

Immortalized Hepatic Cell Lines

Unlike PHHs, immortalized hepatic cell lines harbor a potentially unlimited capacity for self-renewal. Numerous cell lines have been used to model liver and its associated function and pathologies including; human hepatocellular carcinomas (HCC) such as HepG2 and HepB3, Fa2N4 cells, bipotent hepatic progenitors (HepaRG), and HepG2 clonal variants, HepG2/C3A (Qiu et al., 2015). Due to the ease of use and broad accessibility, HepG2 cells are among the most commonly used *in vitro* hepatocyte model despite exhibiting reduced basal gene expression of Phase I and Phase II drug-metabolizing enzymes, as well as functional and phenotypic responses that are inconsistent with those of PHHs (LeCluyse et al., 2012; Zhou and Fan, 2019).

HepaRG cells can be differentiated to express numerous metabolizing enzymes and transporters at levels that are superior to those of other carcinoma lines (Szabo et al., 2013; Mayati et al., 2018). They are functionally stable for weeks following differentiation and like other HCCs have also been used as a surrogate for PHHs to model various liver pathologies, hepatotoxicity and xenobiotic metabolism (Andersson et al., 2012; Mayati et al., 2018). However, the propensity of oncogenic lines to cumulatively acquire genetic perturbations and chromosomal rearrangements (e.g. HepG2s show distinct aneuploidy; Wong, 2000) presents a challenge in their use, particularly in the extrapolation of clinically relevant assertions (Zhou and Fan, 2019). These factors can significantly impact the underlying biology and impartiality of these models in addressing specific questions, for example, assessing the contribution of genetic variance to functional phenotypes.

The generation of immortalized PHH lines, with the capacity for self-renewal, is an interesting prospect for future research efforts as various immortalization strategies and use cases have been described (Ramboer et al., 2014, 2015; Collins et al., 2020). These lines however share some of the disadvantages of their carcinoma-derived counterparts in terms of diminished and (or) limited functionality in comparison to PHHs, along with the negative impact of sustained proliferative cues on genomic stability (Ramboer et al., 2014, 2015). The generation of immortalized lines that overexpress specific CYP450 enzymes and hepatocyte specification factors, alongside conditional immortalization strategies, have also been explored as compensatory mechanisms to enhance the robustness of these models (Ramboer et al., 2015).

Induced Pluripotent Stem Cell Derived Hepatocyte Models

Reprogramming of somatic cells to induced pluripotent stem cells (iPSCs) is achieved by the stochastic overexpression of key stem cell transcription factors (Takahashi et al., 2007). Similar to differentiation protocols for embryonic stem cells, hepatocyte differentiation from iPSCs is mimicked using a multistage cascade via endoderm, anterior definitive endoderm, and hepatocyte commitment through to hepatocyte-like cells (HLCs) (Si-Tayeb et al., 2010; Touboul et al., 2010; Hannan et al., 2013; Mathapati et al., 2016). Failure to express drug metabolizing enzymes, at levels comparable to PHH (Baxter et al., 2015; Sampaziotis et al., 2015) is being improved as the complex mimicry of liver development is better recapitulated by differentiation protocols which direct cellular fate with greater fidelity (Ouchi et al., 2019; Raggi et al., 2022; Takeishi et al., 2020). In addition to these more physiologically relevant models, the improved benchmarking of liver metabolizing enzymes compared to both fetal and adult counterparts (Zabulica et al., 2019), is better defining how HLCs can be appropriately applied to biological questions.

Similar to PHHs, iPSCs retain the original genetic complement of the individual from which they were derived, but with the added advantage of self-renewal. This characteristic further confers genome engineering capabilities to iPSCs (Hockemeyer and Jaenisch, 2016; Sanjurjo-Soriano, 2021). Consequently, they represent a confluence of enabling methodologies in a single model i.e. an infinite source of cellular material representing an individual genome amenable to gene editing and capable of lineage specific differentiation to all three germ layers.

While immortalized cells are also amenable to genome engineering, the improved chromosomal stability and physiological relevance of iPSC derived HLCs lends a significant advantage relevant to the African context. Human iPSC-HLC models have been used to model the disruption of the ER in α1-antitrypsin deficiency (Yusa et al., 2011; Segeritz et al., 2018), and urea cycle defects (Zabulica et al., 2021) by correcting patient-specific cells using ZFN and CRISPR/Cas technologies respectively, to erase the disease signature. The success of these bioengineered models, along with the functional activity of iPSC-HLCs being improved by scalable 3D technologies such as spheroids (Takebe et al., 2017; Heidariyan et al., 2018; Rashidi et al., 2018) and liver-on-a-chip approaches (Kamei et al., 2019), is diversifying their applications in basic and translational research.

How do These Models Interplay and What is Their Value in Pharmacovigilance

In recent years, it has been increasingly recognized that improving preclinical predictions is dependent on the use of relevant cellular material, modeling native architecture, genetic background, and the stringency of defining a model's suitability to specific functional applications. Disparities exist in access to

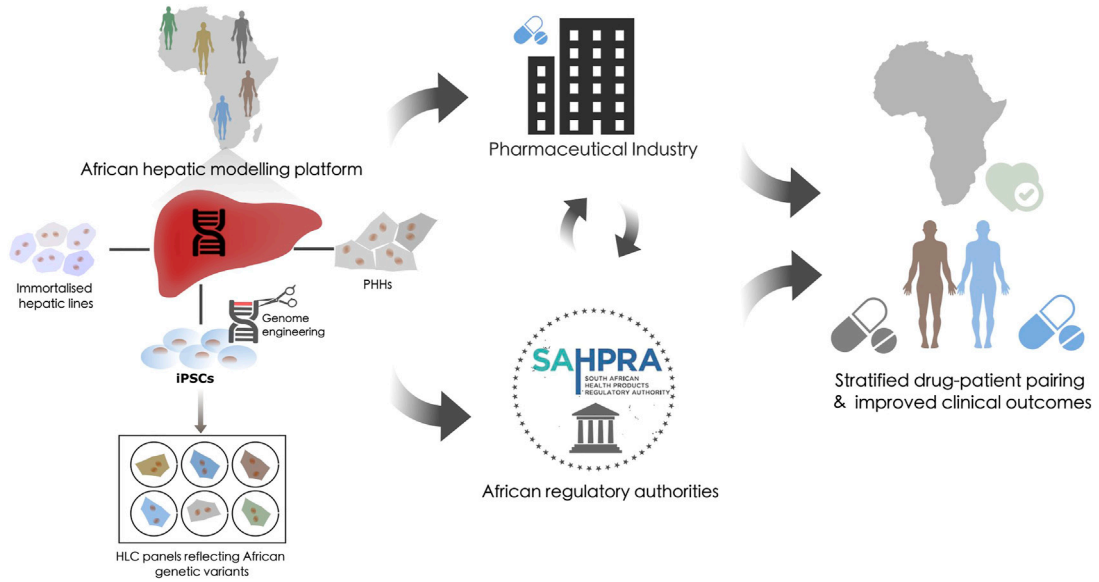


FIGURE 1 | The potential role of a hepatic modeling platform in guiding pharmacovigilance. Proposal for an integrative *in vitro* African pharmacovigilance platform that houses various hepatic models. Such a modeling platform would utilize immortalized hepatic lines, primary human hepatocytes (PHH), and induced pluripotent stem cells (iPSCs) from donors of African origin. iPSC could be genome engineered to generate panels of hepatocyte-like cells (HLCs) to specifically validate African-relevant genetic variants against efficacy/ADRs within an isogenic background. This platform would allow these models to be applied independently, interdependently, or sequentially to recapitulate and validate the xenobiotic metabolism of the liver across the genetic diversity of the African continent. This could then be used to inform decision-making in the global pharmaceutical industry and within the African regulatory landscape to stratify drug-patient pairing and improve clinical outcomes.

in vitro models from diverse genetic backgrounds, as donor demographics of PHHs are overwhelmingly in favor of Caucasian populations ((2 Caucasian, 1 other (Vorrink et al., 2018); 45 Caucasian, 1 other (Baze et al., 2018); 8 Caucasian (Hurrell et al., 2020)). Similarly, of the 2,912 globally registered iPSC lines in the Human Pluripotent Stem Cell Registry (HPSC registry, 2021) only 35 lines from 18 individuals are recorded to have potential African ancestry (annotated as black/African-American/mixed ethnicity), and both HepG2 and HepaRG cell lines were derived from individuals of Caucasian descendant. This highlights the lack of genetic diversity within the laboratory models and remains one of the reasons that bias is still perpetuated in current scientific literature. These circumstances highlight the need for the expansion of an *in vitro* repertoire of "African" liver models (Moore et al., 2021).

Bioengineered iPSC-HLCs derived from individuals of African ancestry, could provide an ideal platform to empirically assess the relationship between individual and cumulative African specific genetic variants and pharmacokinetic phenotypes within an isogenic background. While PHHs have an invaluable role to play in assessing inter-individual variation, and the validation of pharmacokinetic phenotypes, population stratification metrics need to be more stringently and singularly assessed. PHHs may remain the "gold standard" for determining pharmacokinetic phenotypes, however, access to the number of clinically relevant PHHs required to mirror the diversity in Africa is currently not feasible in terms of scalability, accessibility, and cost. Likewise, while the use of immortalized and carcinoma derived hepatic cell lines may present specific challenges for modeling pharmacokinetically relevant genotypes; the derivation of such lines from individuals of African ancestry would facilitate a readily accessible way for researchers to reshape the global underrepresentation of African genetics in science and improve the transferability of global search findings to an African relevant context. The necessity for these resources is perhaps most easily exemplified by the assessment of a few of our own iPSC lines where we focus on generating lines from donors of African origin. Here we identified a line which is both a poor metabolizer for CYP2B6 (CYP2B6*6/*18) as well as a reduced metabolizer for CYP2D6 (CYP2D6*1/*17), leading to almost complete ablation of Efavirenz metabolism and aberrant production of Tamoxifen's potent bioactive metabolites respectively. This is not remarkable within the context of the prevalence in Africa but would be exceptionally rare globally as iPSC lines are predominantly derived from Caucasians.

ADDRESSING PRECISION MEDICINE CHALLENGES FACED BY AFRICA

New scientific breakthroughs, and improved access to these technologies on the African continent, has positioned scientists to establish platforms which directly impact the burden of clinical healthcare. Access to enabling technologies including; genome engineering, iPSC differentiation, and malleable 3D biomimetic microenvironments, provides synthetic biology with tools to develop preclinical models with improved capacity for predictive extrapolation. While individualized treatments might remain an

unrealistic goal in Africa, comprehensive profiling of genetic variants which impact pharmacokinetic/dynamic profiles would aid in the implementation of subpopulation stratification and subsequently improve healthcare strategies on the African continent. With the proposed landscape, independent, interdependent or sequentially applied models could be used to provide empirical evidence for the efficacy or inefficacy of marketed/prescribed drugs in clinical practice.

To address these challenges, we propose an integrative *in vitro* pharmacovigilance platform that houses various hepatic models that recapitulate the xenobiotic metabolism of the liver to specifically link African-relevant genetic variants to efficacy/ADRs. These tools would support pharmacogenetic based decision-making and reduce the use of inadequate or harmful pharmaceutical interventions (**Figure 1**). While the requirement for pharmacogenetic-based treatment stratification has been clinically evidenced in African populations for over a decade (Ngaimisi et al., 2011; Shrif et al., 2011; Masimirembwa and Hasler, 2012; van der Merwe et al., 2012), few national genomic-medicine initiatives/studies, or national regulatory guidelines exist (Radouani et al., 2020) which ingrain pharmacogenetics into the African clinical landscape. The consequence of which is that for some populations we will continue to fail in pairing the right drug, at the right dose, with the right patient to address their healthcare needs. Successful utilization of a platform that can address xenobiotic metabolism from multiple vantage points could lead to intervention in multiple avenues: 1) providing definitive diagnostic assays from proven genetic variant/drug relationships, 2) stratification of drug-patient pairing with improved efficacy outcomes, and 3) providing a model to evaluate redesigned pharmaceutical compounds.

Our research group, which successfully derived one of the first human iPSC lines on the African continent, has applied genome engineering strategies to edit iPSCs, and has been actively engaged in modeling iPSC-HLCs and other "disease-in-a-dish" models. Further to this, isolating PHHs from individuals of African descent, on the continent itself, would represent a significant milestone for the accessibility and utilization of genetically representative PHHs within the African research landscape. However, this will require key stakeholders from multidisciplinary backgrounds to collectively establish such an Afrocentric resource within the global pharmacovigilance arena. Solidifying the foundation of an African hepatic modeling platform will require concerted efforts to establish a new paradigm for preclinical and clinical research collaborations that collectively drive impact in healthcare. The convergence of a number of these traditionally siloed technologies and methodologies, within an established African research infrastructure, would support the feasibility of a hepatic modeling platform and serve as a launchpad for national and ultimately continent-wide initiatives.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

TH and JN conceptualized the manuscript. TH, JN, and JS contributed to the writing of the manuscript.

FUNDING

Funding provided by the Council of Scientific and Industrial Research (CSIR) and the National Research Foundation (NRF) Thuthuka Funding Instrument (TTK200320510301; Grant No: 129,507).

REFERENCES

Ahuja, V., and Sharma, S. (2014). Drug Safety Testing Paradigm, Current Progress and Future Challenges: an Overview. *J. Appl. Toxicol.* 34 (6), 576–594. doi:10.1002/jat.2935

Ampadu, H. H., Hoekman, J., de Bruin, M. L., Pal, S. N., Olsson, S., Sartori, D., et al. (2016) Adverse Drug Reaction Reporting in Africa and a Comparison of Individual Case Safety Report Characteristics between Africa and the Rest of the World: Analyses of Spontaneous Reports in VigiBase®, *Drug Saf.*, 39. 335–345. doi:10.1007/s40264-015-0387-4

Andersson, T. B., Kanebratt, K. P., and Kenna, J. G. (2012). The HepaRG Cell Line: a Uniquein Vitrotool for Understanding Drug Metabolism and Toxicology in Human. *Expert Opin. Drug Metab. Toxicol.* 8 (7), 909–920. doi:10.1517/17425255.2012.685159

Atienzar, F. A., and Nicolas, J.-M. (2018). "Prediction of Human Liver Toxicity Using *In Vitro* Assays: Limitations and Opportunities," in *Drug-Induced Liver Toxicity*. Editors M. Chen and Y. Will (New York, NY: Springer Methods in Pharmacology and Toxicology), 125–150. doi:10.1007/978-1-4939-7677-5_7

Babai, S., Auclert, L., and Le-Louët, H. (2021). Safety Data and Withdrawal of Hepatotoxic Drugs. *Therapies* 76, 715–723. doi:10.1016/j.therap.2018.02.004

Baxter, M., Withey, S., Harrison, S., Segeritz, C.-P., Zhang, F., Atkinson-Dell, R., et al. (2015). Phenotypic and Functional Analyses Show Stem Cell-Derived Hepatocyte-like Cells Better Mimic Fetal rather Than Adult Hepatocytes. *J. Hepatol.* 62 (3), 581–589. doi:10.1016/j.jhep.2014.10.016

Baze, A., Parmentier, C., Hendriks, D. F. G., Hurrell, T., Heyd, B., Bachellier, P., et al. (2018). Three-Dimensional Spheroid Primary Human Hepatocytes in Monoculture and Coculture with Nonparenchymal Cells. *Tissue Eng. C: Methods* 24 (9), 534–545. doi:10.1089/ten.tec.2018.0134

Beckwitt, C. H., Clark, A. M., Wheeler, S., Taylor, D. L., Stolz, D. B., Griffith, L., et al. (2018). Liver 'organ on a Chip'. *Exp. Cel Res.* 363 (1), 15–25. doi:10.1016/j.yexcr.2017.12.023

Bell, C. C., Hendriks, D. F. G., Moro, S. M. L., Ellis, E., Walsh, J., Renblom, A., et al. (2016). Characterization of Primary Human Hepatocyte Spheroids as a Model System for Drug-Induced Liver Injury, Liver Function and Disease. *Sci. Rep.* 6 (1), 25187. doi:10.1038/srep25187

Belle, D. J., and Singh, H. (2008). Genetic Factors in Drug Metabolism. *Am. Fam. Physician* 77 (11), 1553–1560.

B. Tata, E., A. Ambele, M., and Pepper, S. (2020). Barriers to Implementing Clinical Pharmacogenetics Testing in Sub-saharan Africa. A Critical Review. *Pharmaceutics* 12 (9), 809. doi:10.3390/pharmaceutics12090809

ClinicalTrials.gov (2021). clinicaltrials.gov. Available at: https://clinicaltrials.gov/ct2/search/map (Accessed October 29, 2021).

Collins, D. P., Hapke, J. H., Aravalli, R. N., and Steer, C. J. (2020). Development of Immortalized Human Hepatocyte-like Hybrid Cells by Fusion of Multi-Lineage Progenitor Cells with Primary Hepatocytes. *PLOS ONE* 15 (6), e0234002. doi:10.1371/journal.pone.0234002

Costa, A., Sarmento, B., and Seabra, V. (2014). An Evaluation of the Latestin Vitrotools for Drug Metabolism Studies. *Expert Opin. Drug Metab. Toxicol.* 10 (1), 103–119. doi:10.1517/17425255.2014.857402

da Rocha, J. E. B. (2021). The Extent and Impact of Variation in ADME Genes in Sub-saharan African Populations. *Front. Pharmacol.* 12. Available at: https://www.frontiersin.org/article/10.3389/fphar.2021.634016 (Accessed: 21 January 2022). doi:10.3389/fgene.2021.626954

Dambach, D. M., Misner, D., Brock, M., Fullerton, A., Proctor, W., Maher, J., et al. (2016). Safety Lead Optimization and Candidate Identification: Integrating New Technologies into Decision-Making. *Chem. Res. Toxicol.* 29 (4), 452–472. doi:10.1021/acs.chemrestox.5b00396

Dandara, C. (2019). African Pharmacogenomics Consortium: Consolidating Pharmacogenomics Knowledge, Capacity Development and Translation in Africa. *AAS Open Res.* doi:10.12688/aasopenres.12965.1

Dodgen, T. M., Hochfeld, W. E., Fickl, H., Asfaha, S. M., Durandt, C., Rheeder, P., et al. (2013). Introduction of the AmpliChip CYP450 Test to a South African Cohort: a Platform Comparative Prospective Cohort Study. *BMC Med. Genet.* 14 (1), 20. doi:10.1186/1471-2350-14-20

Evans, W. E., and Relling, M. V. (1999). Pharmacogenomics: Translating Functional Genomics into Rational Therapeutics. *Science* 286 (5439), 487–491. doi:10.1126/science.286.5439.487

Freyer, N., Greuel, S., Knöspel, F., Gerstmann, F., Storch, L., Damm, G., et al. (2018). Microscale 3D Liver Bioreactor for *In Vitro* Hepatotoxicity Testing under Perfusion Conditions. *Bioengineering* 5 (1), 24. doi:10.3390/bioengineering5010024

Gurdasani, D., Carstensen, T., Tekola-Ayele, F., Pagani, L., Tachmazidou, I., Hatzikotoulas, K., et al. (2015). The African Genome Variation Project Shapes Medical Genetics in Africa. *Nature* 517 (7534), 327–332. doi:10.1038/nature13997

Hannan, N. R. F., Segeritz, C.-P., Touboul, T., and Vallier, L. (2013). Production of Hepatocyte-like Cells from Human Pluripotent Stem Cells. *Nat. Protoc.* 8 (2), 430–437. doi:10.1038/nprot.2012.153

Heidariyan, Z., Ghanian, M. H., Ashjari, M., Farzaneh, Z., Najarasl, M., Rezaei Larijani, M., et al. (2018). Efficient and Cost-Effective Generation of Hepatocyte-like Cells through Microparticle-Mediated Delivery of Growth Factors in a 3D Culture of Human Pluripotent Stem Cells. *Biomaterials* 159, 174–188. doi:10.1016/j.biomaterials.2018.01.005

Hendriks, D. F. G., Fredriksson Puigvert, L., Messner, S., Mortiz, W., and Ingelman-Sundberg, M. (2016). Hepatic 3D Spheroid Models for the Detection and Study of Compounds with Cholestatic Liability. *Sci. Rep.* 6 (1), 35434. doi:10.1038/srep35434

Heslop, J. A., Rowe, C., Walsh, J., Sison-Young, R., Jenkins, R., Kamalian, L., et al. (2017). Mechanistic Evaluation of Primary Human Hepatocyte Culture Using Global Proteomic Analysis Reveals a Selective Dedifferentiation Profile. *Arch. Toxicol.* 91 (1), 439–452. doi:10.1007/s00204-016-1694-y

Higgins, M. J., and Stearns, V. (2010). CYP2D6 Polymorphisms and Tamoxifen Metabolism: Clinical Relevance. *Curr. Oncol. Rep.* 12 (1), 7–15. doi:10.1007/s11912-009-0076-5

Hockemeyer, D., and Jaenisch, R. (2016). Induced Pluripotent Stem Cells Meet Genome Editing. *Cell Stem Cell* 18 (5), 573–586. doi:10.1016/j.stem.2016.04.013

HPSC registry (2021). HPSC Registry. Available at: https://hpscreg.eu/(Accessed: October 29, 2021).

Hurrell, T., Kastrinou-Lampou, V., Fardellas, A., Hendriks, D. F. G., Nordling, Å., Johansson, I., et al. (2020). Human Liver Spheroids as a Model to Study Aetiology and Treatment of Hepatic Fibrosis. *Cells* 9 (4), 964. doi:10.3390/cells9040964

Ingelman-Sundberg, M., and Lauschke, V. M. (2022). 3D Human Liver Spheroids for Translational Pharmacology and Toxicology. *Basic Clin. Pharma Tox.* 130 (S1), 5–15. doi:10.1111/bcpt.13587

Kamei, K.-i., Yoshioka, M., Terada, S., Tokunaga, Y., and Chen, Y. (2019). Three-dimensional Cultured Liver-On-A-Chip with Mature Hepatocyte-like Cells Derived from Human Pluripotent Stem Cells. *Biomed. Microdevices* 21 (3), 73. doi:10.1007/s10544-019-0423-8

Kozyra, M., Johansson, I., Nordling, Å., Ullah, S., Lauschke, V. M., and Ingelman-Sundberg, M. (2018). Human Hepatic 3D Spheroids as a Model for Steatosis and Insulin Resistance. *Sci. Rep.* 8 (1), 14297. doi:10.1038/s41598-018-32722-6

Langmia, I. M. (2021). 'CYP2B6 Functional Variability in Drug Metabolism and Exposure across Populations—Implication for Drug Safety, Dosing, and Individualized Therapy', Frontiers In Genetics, 12. Available at: https://www.frontiersin.org/article/10.3389/fgene.2021.692234 (Accessed: January 21, 2022).

Lauschke, V. M., Shafagh, R. Z., Hendriks, D. F. G., and Ingelman-Sundberg, M. (2019). 3D Primary Hepatocyte Culture Systems for Analyses of Liver Diseases, Drug Metabolism, and Toxicity: Emerging Culture Paradigms and Applications. *Biotechnol. J.* 14 (7), 1800347. doi:10.1002/biot.201800347

Lauschke, V. M., Vorrink, S. U., Moro, S. M. L., Rezayee, F., Nordling, Å., Hendriks, D. F. G., et al. (2016). Massive Rearrangements of Cellular MicroRNA Signatures Are Key Drivers of Hepatocyte Dedifferentiation. *Hepatology* 64 (5), 1743–1756. doi:10.1002/hep.28780

LeCluyse, E. L., Witek, R. P., Andersen, M. E., and Powers, M. J. (2012). Organotypic Liver Culture Models: Meeting Current Challenges in Toxicity Testing. *Crit. Rev. Toxicol.* 42 (6), 501–548. doi:10.3109/10408444.2012.682115

Liebler, D. C., and Guengerich, F. P. (2005). Elucidating Mechanisms of Drug-Induced Toxicity. *Nat. Rev. Drug Discov.* 4 (5), 410–420. doi:10.1038/nrd1720

Masimirembwa, C., and Hasler, J. (2013). Pharmacogenetics in Africa, an Opportunity for Appropriate Drug Dosage Regimens: on the Road to Personalized Healthcare. *CPT: Pharmacometrics Syst. Pharmacol.* 2 (5), 45. doi:10.1038/psp.2013.17

Mathapati, S., Siller, R., Impellizzeri, A. A., Lycke, M., Vegheim, K., Almaas, R., et al. (2016). Small-Molecule-Directed Hepatocyte-like Cell Differentiation of Human Pluripotent Stem Cells. *Curr. Protoc. Stem Cel Biol* 38 (1), 1G–18G. doi:10.1002/cpsc.13

Mayati, A., Moreau, A., Le Vée, M., Bruyère, A., Jouan, E., Denizot, C., et al. (2018). Functional Polarization of Human Hepatoma HepaRG Cells in Response to Forskolin. *Sci. Rep.* 8 (1), 16115. doi:10.1038/s41598-018-34421-8

Moore, E., Allen, J. B., Mulligan, C. J., and Wayne, E. C. (2021). Ancestry of Cells Must Be Considered in Bioengineering. *Nat. Rev. Mater.* 7, 2–4. doi:10.1038/s41578-021-00397-7

Mpye, K. L., Matimba, A., Dzobo, K., Chirikure, S., Wonkam, A., and Dandara, C. (2017). Disease burden and the Role of Pharmacogenomics in African Populations. *Glob. Health Epidemiol.* 2. doi:10.1017/gheg.2016.21

Nemaura, T., Nhachi, C., and Masimirembwa, C. (2012). Impact of Gender, Weight and CYP2B6 Genotype on Efavirenz Exposure in Patients on HIV/AIDS and TB Treatment: Implications for Individualising Therapy. *Afr. J. Pharm. Pharmacol.* 6 (29), 2188–2193. doi:10.5897/AJPP12.076

Ngaimisi, E., Mugusi, S., Minzi, O., Sasi, P., Riedel, K.-D., Suda, A., et al. (2011). Effect of Rifampicin and CYP2B6 Genotype on Long-Term Efavirenz Autoinduction and Plasma Exposure in HIV Patients with or without Tuberculosis. *Clin. Pharmacol. Ther.* 90 (3), 406–413. doi:10.1038/clpt.2011.129

Ölander, M., Wiśniewski, J. R., Flörkemeier, I., Handin, N., Urdzik, J., and Artursson, P. (2019). A Simple Approach for Restoration of Differentiation and Function in Cryopreserved Human Hepatocytes. *Arch. Toxicol.* 93 (3), 819–829. doi:10.1007/s00204-018-2375-9

Otieno, M. A., Gan, J., and Proctor, W. (2018). "Status and Future of 3D Cell Culture in Toxicity Testing," in *Drug-Induced Liver Toxicity*. Editors M. Chen and Y. Will (New York, NY: Springer New York Methods in Pharmacology and Toxicology), 249–261. doi:10.1007/978-1-4939-7677-5_12

Ouchi, R., Togo, S., Kimura, M., Shinozawa, T., Koido, M., Koike, H., et al. (2019). Modeling Steatohepatitis in Humans with Pluripotent Stem Cell-Derived Organoids. *Cel Metab.* 30 (2), 374–384. e6. doi:10.1016/j.cmet.2019.05.007

Pingitore, P., Sasidharan, K., Ekstrand, M., Prill, S., Lindén, D., and Romeo, S. (2019). Human Multilineage 3D Spheroids as a Model of Liver Steatosis and Fibrosis. *Ijms* 20 (7), 1629. doi:10.3390/ijms20071629

Prill, S., Caddeo, A., Baselli, G., Jamialahmadi, O., Dongiovanni, P., Rametta, R., et al. (2019). The TM6SF2 E167K Genetic Variant Induces Lipid Biosynthesis and Reduces Apolipoprotein B Secretion in Human Hepatic 3D Spheroids. *Sci. Rep.* 9 (1), 11585. doi:10.1038/s41598-019-47737-w

Qiu, G.-H., Xie, X., Xu, F., Shi, X., Wang, Y., and Deng, L. (2015). Distinctive Pharmacological Differences between Liver Cancer Cell Lines HepG2 and Hep3B. *Cytotechnology* 67 (1), 1–12. doi:10.1007/s10616-014-9761-9

Radouani, F., Zass, L., Hamdi, Y., Rocha, J. D., Sallam, R., Abdelhak, S., et al. (2020). A Review of Clinical Pharmacogenetics Studies in African Populations. *Personalized Med.* 17 (2), 155–170. doi:10.2217/pme-2019-0110

Raggi, C., M'Callum, M. A., Pham, Q. T., Gaub, P., Selleri, S., Baratang, N. V., et al. (2022). Leveraging Interacting Signaling Pathways to Robustly Improve the Quality and Yield of Human Pluripotent Stem Cell-derived Hepatoblasts and Hepatocytes. *Stem Cell Rep.* 17 (3), 584–598. doi:10.1016/j.stemcr.2022.01.003

Rajman, I., Knapp, L., Morgan, T., and Masimirembwa, C. (2017). African Genetic Diversity: Implications for Cytochrome P450-Mediated Drug Metabolism and Drug Development. *EBioMedicine* 17, 67–74. doi:10.1016/j.ebiom.2017.02.017

Ramboer, E., De Craene, B., De Kock, J., Vanhaecke, T., Berx, G., Rogiers, V., et al. (2014). Strategies for Immortalization of Primary Hepatocytes. *J. Hepatol.* 61 (4), 925–943. doi:10.1016/j.jhep.2014.05.046

Ramboer, E., Vanhaecke, T., Rogiers, V., and Vinken, M. (2015). Immortalized Human Hepatic Cell Lines for *In Vitro* Testing and Research Purposes. *Methods Mol. Biol. (Clifton, N.J.)* 1250, 53–76. doi:10.1007/978-1-4939-2074-7_4

Rashidi, H., Luu, N.-T., Alwahsh, S. M., Ginai, M., Alhaque, S., Dong, H., et al. (2018). 3D Human Liver Tissue from Pluripotent Stem Cells Displays Stable Phenotype *In Vitro* and Supports Compromised Liver Function *In Vivo*. *Arch. Toxicol.* 92 (10), 3117–3129. doi:10.1007/s00204-018-2280-2

Sampaziotis, F., Segeritz, C. P., and Vallier, L. (2015). Potential of Human Induced Pluripotent Stem Cells in Studies of Liver Disease. *Hepatology* 62 (1), 303–311. doi:10.1002/hep.27651

Sanjurjo-Soriano, C. (2021). "CRISPR/Cas9-Mediated Genome Editing to Generate Clonal iPSC Lines," in *Methods in Molecular Biology*. Editors N. Erkilic, D. Mamaeva, and V. Kalatzis. [Preprint]. doi:10.1007/7651_2021_362

Schärfe, C. P. I., Tremmel, R., Schwab, M., Kohlbacher, O., and Marks, D. S. (2017). Genetic Variation in Human Drug-Related Genes. *Genome Med.* 9 (1), 117. doi:10.1186/s13073-017-0502-5

Segeritz, C.-P., Rashid, S. T., de Brito, M. C., Serra, M. P., Ordonez, A., Morell, C. M., et al. (2018). hiPSC Hepatocyte Model Demonstrates the Role of Unfolded Protein Response and Inflammatory Networks in α1-antitrypsin Deficiency. *J. Hepatol.* 69 (4), 851–860. doi:10.1016/j.jhep.2018.05.028

Shrif, N. E. M. A., Won, H.-H., Lee, S.-T., Park, J.-H., Kim, K.-K., Kim, M.-J., et al. (2011). Evaluation of the Effects of VKORC1 Polymorphisms and Haplotypes, CYP2C9 Genotypes, and Clinical Factors on Warfarin Response in Sudanese Patients. *Eur. J. Clin. Pharmacol.* 67 (11), 1119–1130. doi:10.1007/s00228-011-1060-1

Si-Tayeb, K., Noto, F. K., Nagaoka, M., Li, J., Battle, M. A., Duris, C., et al. (2010). Highly Efficient Generation of Human Hepatocyte-like Cells from Induced Pluripotent Stem Cells. *Hepatology* 51 (1), 297–305. doi:10.1002/hep.23354

Stark, Z., Dolman, L., Manolio, T. A., Ozenberger, B., Hill, S. L., Caulfied, M. J., et al. (2019). Integrating Genomics into Healthcare: A Global Responsibility. *Am. J. Hum. Genet.* 104 (1), 13–20. doi:10.1016/j.ajhg.2018.11.014

Szabo, M., Veres, Z., Baranyai, Z., Jakab, F., and Jemnitz, K. (2013). Comparison of Human Hepatoma HepaRG Cells with Human and Rat Hepatocytes in Uptake Transport Assays in Order to Predict a Risk of Drug Induced Hepatotoxicity. *PLOS ONE* 8 (3), e59432. doi:10.1371/journal.pone.0059432

Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., Tomoda, K., et al. (2007). Induction of Pluripotent Stem Cells from Adult Human Fibroblasts by Defined Factors. *Cell* 131 (5), 861–872. doi:10.1016/j.cell.2007.11.019

Takebe, T., Sekine, K., Kimura, M., Yoshizawa, E., Ayano, S., Koido, M., et al. (2017). Massive and Reproducible Production of Liver Buds Entirely from Human Pluripotent Stem Cells. *Cel Rep.* 21 (10), 2661–2670. doi:10.1016/j.celrep.2017.11.005

Takeishi, K., Collin de l'Hortet, A., Wang, Y., Handa, K., Guzman-Lepe, J., Matsubara, K., et al. (2020). Assembly and Function of a Bioengineered Human Liver for Transplantation Generated Solely from Induced Pluripotent Stem Cells. *Cel Rep.* 31 (9), 107711. doi:10.1016/j.celrep.2020.107711

Touboul, T., Hannan, N. R. F., Corbineau, S., Martinez, A., Martinet, C., Branchereau, S., et al. (2010). Generation of Functional Hepatocytes from Human Embryonic Stem Cells under Chemically Defined Conditions that Recapitulate Liver Development. *Hepatology* 51 (5), 1754–1765. doi:10.1002/hep.23506

van der Merwe, N., Bouwens, C. S. H., Pienaar, R., van der Merwe, L., Yako, Y. Y., Geiger, D. H., et al. (2012). CYP2D6 Genotyping and Use of Antidepressants in Breast Cancer Patients: Test Development for Clinical Application. *Metab. Brain Dis.* 27 (3), 319–326. doi:10.1007/s11011-012-9312-z

Vorrink, S. U., Ullah, S., Schmidt, S., Nandania, J., Velagapudi, V., Beck, O., et al. (2017). Endogenous and Xenobiotic Metabolic Stability of Primary Human Hepatocytes in Long-term 3D Spheroid Cultures Revealed by a Combination of Targeted and Untargeted Metabolomics. *FASEB j.* 31 (6), 2696–2708. doi:10.1096/fj.201601375R

Vorrink, S. U., Zhou, Y., Ingelman-Sundberg, M., and Lauschke, V. M. (2018). Prediction of Drug-Induced Hepatotoxicity Using Long-Term Stable Primary Hepatic 3D Spheroid Cultures in Chemically Defined Conditions. *Toxicol. Sci.* 163 (2), 655–665. doi:10.1093/toxsci/kfy058

Weinshilboum, R. (2003). Inheritance and Drug Response. *N. Engl. J. Med.* 348 (6), 529–537. doi:10.1056/NEJMra020021

Wong, N. (2000). A Comprehensive Karyotypic Study on Human Hepatocellular Carcinoma by Spectral Karyotyping. *Hepatology* 32 (5), 1060–1068. doi:10.1053/jhep.2000.19349

Wonkam, A. (2021). Sequence Three Million Genomes across Africa. *Nature* 590 (7845), 209–211. doi:10.1038/d41586-021-00313-7

Yamasaki, C., Ishida, Y., Yanagi, A., Yoshizane, Y., Kojima, Y., Ogawa, Y., et al. (2020). Culture Density Contributes to Hepatic Functions of Fresh Human Hepatocytes Isolated from Chimeric Mice with Humanized Livers: Novel, Long-Term, Functional Two-Dimensional *In Vitro* Tool for Developing New Drugs. *PLOS ONE* 15 (9), e0237809. doi:10.1371/journal.pone.0237809

Yusa, K., Rashid, S. T., Strick-Marchand, H., Varela, I., Liu, P.-Q., Paschon, D. E., et al. (2011). Targeted Gene Correction of α1-antitrypsin Deficiency in Induced Pluripotent Stem Cells. *Nature* 478 (7369), 391–394. doi:10.1038/nature10424

Zabulica, M., Jakobsson, T., Ravaioli, F., Vosough, M., Gramignoli, R., Ellis, E., et al. (2021). Gene Editing Correction of a Urea Cycle Defect in Organoid Stem Cell Derived Hepatocyte-like Cells. *Ijms* 22 (3), 1217. doi:10.3390/ijms22031217

Zabulica, M., Srinivasan, R. C., Vosough, M., Hammarstedt, C., Wu, T., Gramignoli, R., et al. (2019). Guide to the Assessment of Mature Liver Gene Expression in Stem Cell-Derived Hepatocytes. *Stem Cell Dev.* 28 (14), 907–919. doi:10.1089/scd.2019.0064

Zanger, U. M., and Schwab, M. (2013). Cytochrome P450 Enzymes in Drug Metabolism: Regulation of Gene Expression, Enzyme Activities, and Impact of Genetic Variation. *Pharmacol. Ther.* 138 (1), 103–141. doi:10.1016/j.pharmthera.2012.12.007

Zeilinger, K., Freyer, N., Damm, G., Seehofer, D., and Knöspel, F. (2016). Cell Sources for *In Vitro* Human Liver Cell Culture Models. *Exp. Biol. Med. (Maywood)* 241 (15), 1684–1698. doi:10.1177/1535370216657448

Zhang, D., and Surapaneni, S. (2012). "Appendix: Drug Metabolizing Enzymes and Biotransformation Reactions," in *ADME-enabling Technologies in Drug Design and Development*. Editors N. Penner, C. Woodward, and C. Prakash (Hoboken, NJ, USA: John Wiley & Sons), 545–565. doi:10.1002/9781118180778.app1

Zhou, D., and Fan, J.-G. (2019). Microbial Metabolites in Non-alcoholic Fatty Liver Disease. *Wjg* 25 (17), 2019–2028. doi:10.3748/wjg.v25.i17.2019

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Genetics

BRIEF RESEARCH REPORT
published: 27 April 2022
doi: 10.3389/fgene.2022.769919

A View on Genomic Medicine Activities in Africa: Implications for Policy

*C. Victor Jongeneel[1], Maritha J. Kotze[2], Archana Bhaw-Luximon[3], Faisal M. Fadlelmola[4], Yasmina J. Fakim[5], Yosr Hamdi[6,7], Samar Kamal Kassim[8], Judit Kumuthini[9], Victoria Nembaware[10], Fouzia Radouani[11], Nicki Tiffin[12,13] and Nicola Mulder[12,13]**

[1]Independent Consultant, Rolle, Switzerland, [2]Division of Chemical Pathology, Department of Pathology, Faculty of Medicine and Health Sciences, Stellenbosch University and National Health Laboratory Service, Tygerberg Hospital, Cape Town, South Africa, [3]Center for Biomedical and Biomaterials Research, University of Mauritius, Reduit, Mauritius, [4]Centre for Bioinformatics and Systems Biology, Faculty of Science, University of Khartoum, Khartoum, Sudan, [5]Biotechnology Unit, Faculty of Agriculture, University of Mauritius, Reduit, Mauritius, [6]Laboratory of Biomedical Genomics and Oncogenetics, Institut Pasteur de Tunis, Universite' Tunis El Manar, Tunis, Tunisia, [7]Laboratory of Human and Experimental Pathology, Institut Pasteur de Tunis, Tunis, Tunisia, [8]Medical Biochemistry and Molecular Biology Department, Faculty of Medicine, Ain Shams University and MASRI Research Institute, Ain Shams University, Cairo, Egypt, [9]South African National Bioinformatics Institute (SANBI), Life Sciences Building, University of Western Cape (UWC), Cape Town, South Africa, [10]Division of Human Genetics, Department of Pathology, Faculty of Health Sciences, University of Cape Town, Cape Town, South Africa, [11]Chlamydiae and Mycoplasmas Laboratory, Research Department, Institut Pasteur du Maroc, Casablanca, Morocco, [12]Computational Biology Division, Department of Integrative Biomedical Sciences, IDM, University of Cape Town Faculty of Health Sciences, Cape Town, South Africa, [13]Wellcome CIDRI-Africa, University of Cape Town Faculty of Health Sciences, Cape Town, South Africa*

OPEN ACCESS

Edited by:
Mandeep Kaur, University of the Witwatersrand, South Africa

Reviewed by:
Robert Fred Clark, RTI International, United States
Tanvir Alam, Hamad bin Khalifa University, Qatar

***Correspondence:**
Nicola Mulder
nicola.mulder@uct.ac.za

Specialty section:
This article was submitted to Human and Medical Genomics, a section of the journal Frontiers in Genetics

Received: *02 September 2021*
Accepted: *24 March 2022*
Published: *27 April 2022*

Citation:
Jongeneel CV, Kotze MJ, Bhaw-Luximon A, Fadlelmola FM, Fakim YJ, Hamdi Y, Kassim SK, Kumuthini J, Nembaware V, Radouani F, Tiffin N and Mulder N (2022) A View on Genomic Medicine Activities in Africa: Implications for Policy. Front. Genet. 13:769919. doi: 10.3389/fgene.2022.769919

Genomics policy development involves assessing a wide range of issues extending from specimen collection and data sharing to whether and how to utilize advanced technologies in clinical practice and public health initiatives. A survey was conducted among African scientists and stakeholders with an interest in genomic medicine, seeking to evaluate: 1) Their knowledge and understanding of the field. 2) The institutional environment and infrastructure available to them. 3) The state and awareness of the field in their country. 4) Their perception of potential barriers to implementation of precision medicine. We discuss how the information gathered in the survey could instruct the policies of African institutions seeking to implement precision, and more specifically, genomic medicine approaches in their health care systems in the following areas: 1) Prioritization of infrastructures. 2) Need for translational research. 3) Information dissemination to potential users. 4) Training programs for specialized personnel. 5) Engaging political stakeholders and the public. A checklist with key requirements to assess readiness for implementation of genomic medicine programs is provided to guide the process from scientific discovery to clinical application.

Keywords: Africa, genomic medicine, infrastructure, capacity development, readiness checklist, precision medicine stakeholders, translational research, Pathology-supported genomics

INTRODUCTION

In February 2021, on commission from the African Academy of Sciences, a group of scientists including the authors of this Policy Brief published a White Paper entitled "A Framework for the Implementation of Genomic Medicine for Public Health in Africa" (Policy Paper: A Framework for the Implementation of Genomic Medicine for Public Health in Africa | The AAS). This framework outlines the challenges faced by African stakeholders aiming to implement genomic medicine in their health care systems, and makes specific recommendations to address these challenges in the areas of infrastructure, the selection of participants, the collection of clinical and demographic data, the actionability of linkages between

genotypes and phenotypes, ethical legal and social implication (ELSI) issues and data governance, education and training, translation of research to clinical practice, and stakeholder engagement. As a supplement to the preparation of the framework document, a Task Force was formed to conduct a survey during 2019 and 2020 that would provide a glance of the state of genomic medicine programs in Africa.

Precision medicine for improved public health relies on implementation of prevention and treatment strategies informed by the combination of genetic, environmental, and social factors, which could be targeted to individual genomes and/or the population genetic background (Molster et al., 2018). Based on the responses of representative stakeholders to 25 questions, the questionnaire sought to evaluate the knowledge of the respondents, their institutional environment, the penetration of genomic medicine approaches in their home institution, the awareness of the field in their country, and their opinion on the obstacles faced during implementation initiatives. The full questionnaire is available at http://j.mp/37YDNuw and included in the **Supplementary Material**. Ethics approval for this research was obtained from the Human Research Ethics Committee of the University of Cape Town in South Africa. Although a request to take the survey was sent to many African stakeholders from academic institutions, government and private industry, only 78 responses were received, of which 55 were complete enough to be included in our analysis. While this is not sufficient to draw any truly robust conclusions, we believe that enough information was provided by the respondents to detect useful trends and, along with the aforementioned White Paper, to develop some recommendations to inform policy.

Here, the policy-related implications of responses that were received from twelve African countries are discussed, with the aim to provide a checklist with key requirements to assess readiness for implementation of genomic medicine programs. A more quantitative analysis of the survey results, as much as was possible with the limited responses, is included as **Supplementary Material**, and only summary results are included here as they relate to informing policy recommendations. While the survey probed the broader field of precision medicine, the recommendations focus specifically on genomic medicine.

RESPONSES TO THE SURVEY

Knowledge About Precision Medicine

The respondents included physicians, postgraduate students, laboratory heads, scientists engaged in research activities, and an administrator. Most indicated that they were familiar with the concepts underlying precision medicine, and when asked to provide a definition of the field, the 45 individuals who answered gave reasonable responses. Over 80% indicated that they were already applying advanced genomic technologies in their research or were planning to do so in the foreseeable future. Many project descriptions agreed only marginally with the accepted definition of precision medicine, but incorporated technologies often associated with it, such as next generation sequencing (NGS). Overall, the data collected in the survey indicated that the participating scientists were cognizant of the scope of public health genomics, and that they were eager to engage in research that encompasses its premises. Since it was clear that genomic rather than precision medicine programs have been introduced in several African countries, this study focused on findings related to genomic medicine.

Institutional Environment and Infrastructure

Respondents were from twelve African countries representing most of the Continent's geographical areas, with the exception of Central Africa. **Figure 1** shows a geographical map of Africa, indicating the sample sizes of respondents from individual countries who reported on graduate programs and core sequencing facilities supported by government laboratories, universities and other higher education institutions, and some hospitals and non-governmental organizations (NGOs). Less than half (43%) of the respondents had access to local infrastructures needed to conduct genomic medicine research, the rest either relied on external collaborators (15%) or did not have adequate access (42%). Less than half had access to a biobank (21%), a clinical laboratory (26%), a genomic analysis or sequencing facility (28%), a computational facility for data analysis (28%), or a data storage and archiving facility (26%). Very few had access to all of these essential infrastructures, suggesting that lack of access to proper infrastructure is a major roadblock to the spread of genomic medicine on the African Continent. There are, however, some state of the art research facilities in several African countries that are well placed to support genomic medicine and may be able to provide at least some of these services at a regional level.

State of the Field and Awareness

Approximately one third of the respondents indicated that their home institution hosted a program or project focused on aspects of genomic medicine, while another third stated that there was no such program locally. The rest did not respond to this question or were unsure. The programs that were mentioned addressed a wide variety of diseases. When asked about the data that were collected on patients, only 20% of respondents stated that both demographic and clinical data were being collected. About one third of respondents collected or stored genomic/genetic data to enable sharing with other stakeholders for either research or clinical care purposes. It appears that at least five African countries (South Africa, Tunisia, Zimbabwe, Egypt, and Rwanda) have started or are planning to start national genomic medicine programs. Awareness of the field among clinicians and researchers is still limited as only 14% of respondents thought that scientists in their country were ready to implement precision/genomic medicine concepts. Most felt that either they were cognizant of the concepts but not ready to implement them (36%), or still needed to be better informed or educated (41%).

Barriers to Implementation

Finally, the survey sought to understand the major barriers impeding the implementation of genomic medicine in African countries. The most commonly cited barrier was the lack of proper infrastructures and technologies to support research and clinical translation (81%). Limited information about

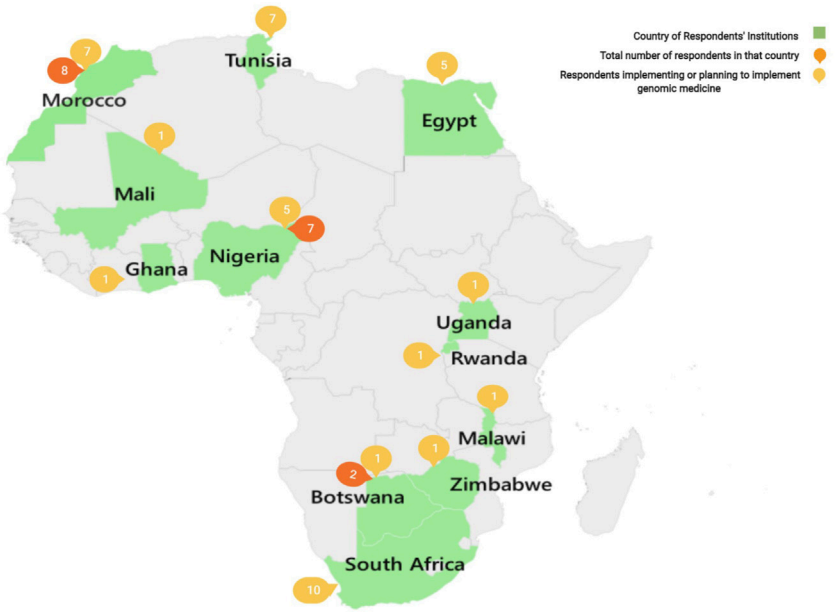


FIGURE 1 | Map showing the countries from which responses were received, including the total number of respondents within the country (dark orange) and out of these, the number who are currently or planning to implement genomic medicine activities (light orange).

determinants of disease susceptibility in local populations (64%), a lack of genetics-focused education of physicians and caregivers (67%), and a perception that this was a low-priority investment (55%) were also mentioned. Less important were the potential low-cost effectiveness (24%) and the fear that it would benefit only the most affluent patients (31%). Most respondents (83%) agreed that additional training would be required in their countries for genomic medicine to become a reality. Specialized training could target university students, house officers, physicians, genetic counselors, nurses and other caregivers, and laboratory technicians.

ACTIONABLE RECOMMENDATIONS

The framework that we previously produced in the form of a policy paper (Policy Paper: A Framework for the Implementation of Genomic Medicine for Public Health in Africa | The AAS) outlined in some detail the required steps for a successful introduction of genomic medicine concepts in the African context. Due to the breadth of precision medicine and areas of expertise of the authors, the paper focused only on genomic medicine. Despite the limited sample size, the results of the survey presented in this study offer an additional window on knowledge gaps identified by respondents from twelve African countries, which should receive priority from a policy perspective as outlined below.

Prioritization of Infrastructures

The results of the survey clearly indicate that lack of access to proper infrastructure is a major roadblock to the implementation of genomic medicine in most African countries and ensuring that such access is put in place is therefore a policy priority. The infrastructure that should be available includes simple point-of-care (POC) platforms to enable sample acquisition and accurate data capture during first-tier genetic testing; more sophisticated facilities for high-throughput data generation and analyses; biobanks for long-term storage and retrieval of patient samples; databases for demographic and clinical data capture, archiving and retrieval that can be accessed from mobile devices; and computational facilities for data analysis. Some such resources already exist in most geographical areas of the Continent, for example through the Human Heredity and Health in Africa (H3Africa) program. Specialized training in data management for return of research results is furthermore provided through the Open Genome Project (https://www.gknowmix.org/opengenome/), which explores the feasibility of combining POC genomics with genetic counselling (Oosthuizen et al., 2021). Pathology-supported genetic testing of multiple non-communicable (NCD) pathways enables genetic counselors to assess different aspects of the same disease in a patient report. Policy efforts should therefore seek to establish collaborative channels for sharing existing facilities wherever possible, and to build new ones only if absolutely necessary and in the context of a regional sharing policy.

Need for More Translational Research

While not directly addressed in the survey, the question of whether existing scientific knowledge is sufficient to guide a path to genomic medicine in Africa is raised by many of the respondents. Most existing data on the relationships between genetic background, environmental effects and lifestyle on the

susceptibility to disease, and on how to best target therapies based on detailed knowledge about a patient's background, have been collected in populations of European or East Asian descent. The dearth of information and patient registries on such relationships among African individuals, especially in light of the high level of genetic diversity on the continent, is a major obstacle. It should therefore be an important policy objective to ensure that high quality epidemiological, clinical and omics research on the same study cohorts in their resident populations is being conducted, be it through local research initiatives or by participating in international efforts such as H3Africa, the SA-UK Newton Collaborative Research and Development Program in Precision Medicine (e.g., https://gtr.ukri.org/projects?ref=103993), or the International Hundred K + Cohorts Consortium (IHCC). There has been recent interest in establishing a longitudinal African population cohort, which would offer excellent opportunities for expanding this knowledge (https://cms.wellcome.org/sites/default/files/2021-03/APCC%20Scoping%20paper_%20FINAL_EN_0.pdf). At the very least, governments should facilitate participation of their citizens and researchers in translational research studies, and access to their health care systems, within a responsible legal framework, to enable genomic medicine.

Information Dissemination to Potential Users

Most of the respondents to the survey thought that healthcare workers in their countries were insufficiently informed of the principles of precision medicine at the genomics level, and thus unable to participate fully in any implementation efforts. In this context, one of the respondents highlighted the benefit of positioning pathology as a bridging science at the interface between the research laboratory and clinical practice, thereby gaining collective knowledge through comparative effectiveness studies that may translate into immediate clinical benefit to patients. This suggests that informing and educating the healthcare workforce about genomic medicine and other areas of precision medicine should be prioritized before starting a national program. Existing frameworks for continuing education (e.g., clinician professional development, good clinical practice courses) should be leveraged, and targeted information campaigns (leaflets, posters, lectures, workshops, etc.) should be developed and distributed. A common theme emerging from the survey is the need to define where genomic tests add value to standard pathology routinely applied in precision medicine.

Training Programs for Specialized Personnel

Successful implementation of a precision medicine program will require a new workforce with updated skills, able to respond to the challenges associated with the implementation. This requires upskilling of existing workers and adjustment of current curricula to produce an interdisciplinary workforce. This is especially true in Africa, where training is often focused on technologies in current use rather than new, leading-edge knowledge; and this is exacerbated where resources for new technologies are not yet in place. Lack of a highly trained specialized workforce as a significant obstacle was clearly described by respondents. The list of required specialties is extensive: researchers in genetics, omics, statisticians, genetic/genomic counselors, bioinformaticians, computer systems administrators, engineers capable of maintaining sophisticated instruments and technicians to operate them. There is also a deficit in soft skills including leadership, management, and governance. The genomics innovation ecosystem should therefore engage with clinicians, patients, researchers, service providers, data curators, data consumers, funding agencies, pharmaceutical and technical/manufacturing companies, policymakers, medical insurance companies, civil society, learned societies, etc., All these stakeholders should acquire specific skills and require specific training activities. While it may be necessary to create some training programs from scratch (including medical scientist professional development courses and relevant degree programs), the need for specialized personnel can probably be largely met by updating curricula in existing educational institutions and encouraging students to pursue training in these new fields. There should be a clearly stated policy encouraging institutions to update their curricula and to train the required workforce in advance or in parallel to putting in place genomic medicine and genetic pathology programs that will safeguard personal information processing. Due consideration should be given to leveraging training opportunities provided by the Wellcome Connecting Science, H3Africa and others offered in-country by African scientists.

Engage Political Stakeholders and the Public

Successful implementation of genomic medicine approaches in healthcare systems critically depends on the engagement of knowledgeable politicians, who in turn will need the support of their voters. Genomic programs should be tightly integrated with existing governmental health and legal systems. Also, in the absence of understanding and support from the medical profession and the general public, uptake will be limited, and such programs are more likely to fail. Themes to be highlighted include: how each patient will receive care better adapted to their personal circumstances; how more emphasis will be given to prevention of advanced disease; how treatments that are not efficacious will be avoided, and that the overall cost of providing such care will be less. It should also be emphasized that everyone will benefit, not only the privileged classes, and that population-level research using genomic medicine approaches can also ensure general improvements to health care provided to specific communities. Leaders of genomic programs should ensure that from the start they engage policy makers and the general public, not only to provide financial and logistical means, but also to ensure broad popular support for the initiatives and to proactively address questions and concerns raised by the general public and healthcare clients.

TABLE 1 | Checklist with key requirements to assess readiness of African countries prior to implementation of genomic medicine programs.

Key elements	Processes required	Readiness assessment
Patient selection: Clinical facilities for patient counselling, screening, treatment and monitoring	Informed consent of participants, obtain relevant previous pathology/other test results from health records, data translation into an adaptable report, genetic counselling	Clinical infrastructure for patient enrollment, collection and analysis of biosamples linked to patient data to enable treatment recommendations and monitoring of clinical outcome
Sample selection: Sample collection, processing and storage facilities, data acquisition tools to prevent operational fragmentation	Sample type selection (e.g., blood, saliva, biopsy) and metadata collection, sample transfer and preparation applying good clinical practice	Biorepositories for sample preparation and storage to enable retrieval and re-analysis of patient samples, acquisition and storage of metadata including clinical data from different sources
Data generation: Genetic testing, genomics data generation and storage	DNA/RNA extraction or direct swab-specimen application, quality control, data generation through genetic testing or omics technologies	Data generation instruments for generation of results on portable devices (point-of-care/other genotyping tests) and/or large scale (microarrays, high-throughput sequencing)
Data analysis: Data storage, curation, analysis and interpretation by assessing clinical relevance of genetic findings	Data processing, analysis, variant classification, identification of actionable gene variant(s), analytical validation using gold standard methodology, alignment of clinical characteristics with familial vs. lifestyle risk and/or treatment response	Data and computing infrastructures for acquiring and storage of genomic data and to enable efficient and secure transfer of data, complex software environment for running research-informed pipelines to enable analysis and interpretation of high throughput genomic data, integrated data systems for analysis, interpretation and report generation, and AI to facilitate clinical decision making
Knowledge databases: Up to date information on genotype-phenotype links and evidence for actionability	Compare variants with reference and disease datasets and prior evidence, extract additional clinically relevant data (e.g., medication use, comorbidities) for clinical interpretation	Reference genomics datasets for relevant populations, sufficient evidence for actionability based on data generated in-country and reported in other patients with similar clinical phenotypes based on well-established scientific literature
Research facilities: Increase knowledge on genomics in African populations, enable translation of results	Data generation and sharing for research translation, gain collective knowledge for precision medicine applications, validation and transition to clinical application	Data generation and sharing capability, research infrastructure, facilities for transitioning to applied research involving feasibility and proof of principle studies on assay validation, clinical utility and health economics to enable up-scaling for clinical translation and implementation
Training: Genomic medicine training programs for healthcare professionals and support personnel	Training to develop a multi-disciplinary service delivery team, case-based learning to achieve learning objectives	Training facilities and curricula for new degrees and professional development courses, online platforms to practice implementation ideas (e.g. pathology-supported genomics)
Regulatory framework: Policy development for governing relevant activities, informed by standard operating procedures (SOPs) and instructions for use (IFUs)	Data and sample governance policy, informed consent documents and tracking, material transfer and data sharing agreements, participant engagement and intellectual property disclosure to authorities	Genomic medicine framework with data and sample governance, and ethical oversight, long-term participant engagement to close the gap between expectation and reality

IMPLEMENTATION STRATEGY

Translation of evidence-based guidelines to health practice is one of the most challenging aspects of applied research and genetic testing service delivery, especially in resource-limited settings (Mitropoulos et al., 2015; Kristensen et al., 2016; Fontes Marx et al., 2021). Implementation of genomic medicine requires a dynamic process informed by the nature of the innovation and the manner in which the type and strength of the evidence is communicated to potential users. The slow pace of adoption of novel findings largely relates to poor understanding of the role of genomics in risk stratification and targeted treatment. Practice-changing solutions require inventive steps which demonstrate clinical utility beyond standard pathology tests, as evidenced in the surgical oncology field using a cost-minimization pathology-supported genetic testing strategy (Mampunye et al., 2021; Myburgh et al., 2021). Incorporation of omics solutions into existing treatment algorithms is therefore needed to prevent misdirection and fragmentation of genetic services relevant to individual patients or their healthy, at-risk family members (Solomon et al., 2011; Moremi et al., 2021). This may involve molecular profiling of not only a few genes, but also whole exomes and genomes in uninformative cases. Highly dense datasets are therefore becoming the norm for disease screening, risk stratification and therapeutic interventions.

Important aspects to consider for successful implementation of genomic medicine and policy development in African countries, are summarized in **Table 1**. This information is partly based on the pathology-supported genetic testing method being introduced as a case study across the illness and wellness domains (Okunola et al., 2019; Moremi et al., 2021). The list of requirements for implementation of genomic medicine in clinical care was informed by our survey, the previously mentioned policy framework compiled by 38 African scientists, as well as insights gained from literature studies and genome-scale sequencing recently performed in African facilities (Torrorey-Sawe et al., 2020; Glanzmann et al., 2021). The checklist provided in **Table 1** may be used to guide the

process from scientific discovery to clinical application, while accommodating infrastructure barriers and knowledge gaps that may be encountered in the research translation process.

Molecular genetic testing has raised a variety of policy issues (Klein 2020), ranging from privacy to reimbursement challenges, including the development of medical insurance coverage policies for genetic tests. For regulatory approval of laboratory-developed genomic solutions, we need to leverage the strengths of both precision medicine and public health genomics (Roberts et al., 2021). Moving from basic research to translation requires effective management of the genomic service delivery process from sample collection to report generation, which should fit seamlessly into existing clinical workflows and treatment algorithms. Given the potential of biochemical methods and other health assessments to uncover the genetic component of complex, multifactorial diseases in a clinical context, we envisage a future where research databases (e.g., https://www.cansa.org.za/establishing-cancer-genomics-registry-to-support-implementation-of-personalised-medicine) are developed in parallel with patient care for differential diagnosis and long-term pharmacogenomics informed clinical outcome studies. Use of the multi-assay pathology-supported genetic testing platform towards seamless integration of research and service delivery has recently been recognized as best practice by the International Consortium of Personalized Medicine (ICPerMED).

CONCLUSION

This policy brief highlights important requirements for successfully implementing genomic medicine programs in Africa. The survey that we conducted for feedback from stakeholders about their perceptions on this healthcare model was limited by sample size and lack of broad representation across sectors. However, the results we did receive from multiple institutions in twelve countries of varying economic levels (low- and middle-income countries), as well as our broader effort to outline the challenges faced when putting in place genomic medicine programs, have allowed us to formulate reasonably actionable recommendations that are broadly applicable. The finding that only 10% of scientist respondents were ready to or have already started a process of implementing genomic medicine in their country was concerning. This led to inclusion of a checklist (**Table 1**) as an educational tool to assist scientist entrepreneurs with identification of infrastructural and other gaps they need to address in their translational research projects. The pathology-supported genetic testing strategy, for example, makes use of a dynamic screening algorithm for eligibility assessment and clinical interpretation of genomic tests by guiding treatment decision-making beyond a single objective. The policy recommendations we provide would not necessitate major investments, although aspects such as data sharing and integration platforms may require stronger cross-border cooperation and joint investments than are currently ongoing in the region. Broad buy-in from healthcare providers, educational institutions, and government services as well as coordinated activities between them will be key to success.

DATA AVAILABILITY STATEMENT

The datasets presented in this article are not readily available because the raw results of the survey cannot be shared to protect the privacy of participants. Summary results are available in the **Supplementary Material**.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the University of Cape Town Human Research Ethics Committee. The participants provided their informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

CVJ wrote the main text of the article. MK assisted with analysis and interpretation of the survey results, compiled **Table 1** with input from all authors and contributed to scientific content editing throughout the writing process. NM coordinated the project and contributed to the table and scientific content of the manuscript. AB-L, FF, CVJ, YF, YH, NT and NM led the development of core components of the framework document, SK, JK, VN and FR contributed to analysis of the survey results. CVJ and MK are acknowledged as co-first authors. All authors approved the final manuscript.

FUNDING

Funding sources are acknowledged by NT and NM from the Wellcome Trust 203135/Z/16/Z and by NT, NM, JK, YF, FF, SK, FR and CVJ from the NIH Common Fund (H3ABioNet, U24HG006941); by MK from the Strategic Health Innovation Partnerships Unit of the South African Medical Research Council (SAMRC), with funds received from the South African Department of Science and Innovation (S003665, S006652), the Cancer Association of South Africa (CANSA) and the South African Technology Innovation Agency (grant number 401/01). We acknowledge Ichrak Benamri and Wiame Belbellaj who generated the figures for the survey results and the paper, respectively.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.769919/full#supplementary-material

REFERENCES

Fontes Marx, M., Ataguba, J. E., Vries, J. D., and Wonkam, A. (2021). Systematic Review of the Economic Evaluation of Returning Incidental Findings in Genomic Research. *Front. Public Health* 9, 697381. doi:10.3389/fpubh.2021.697381

Glanzmann, B., Jooste, T., Ghoor, S., Gordon, R., Mia, R., Mao, J., et al. (2021). Human Whole Genome Sequencing in South Africa. *Sci. Rep.* 11, 606. doi:10.1038/s41598-020-79794-x

Klein, R. D. (2020). Current Policy Challenges in Genomic Medicine. *Clin. Chem.* 66 (1), 61–67. doi:10.1373/clinchem.2019.308775

Kristensen, N., Nymann, C., and Konradsen, H. (2016). Implementing Research Results in Clinical Practice- the Experiences of Healthcare Professionals. *BMC Health Serv. Res.* 16, 48. doi:10.1186/s12913-016-1292-y

Mampunye, L., Grant, K. A., Peeters, A. V., Torrorey-Sawe, R., French, D. J., Moremi, K. E., et al. (2021). MammaPrint Risk Score Distribution in South African Breast Cancer Patients with the Pathogenic *BRCA2* c.7934delG Founder Variant: Towards Application of Genomic Medicine at the Point-of-Care. *Breast* 56 (Suppl. 1), S17–S90. doi:10.1016/s0960-9776(21)00120-x

Mitropoulos, K., Al Jaibeji, H., Forero, D. A., Laissue, P., Wonkam, A., Lopez-Correa, C., et al. (2015). Success Stories in Genomic Medicine from Resource-Limited Countries. *Hum. Genomics* 9 (1), 11. doi:10.1186/s40246-015-0033-3

Molster, C. M., Bowman, F. L., Bilkey, G. A., Cho, A. S., Burns, B. L., Nowak, K. J., et al. (2018). The Evolution of Public Health Genomics: Exploring its Past, Present, and Future. *Front. Public Health* 6, 247. doi:10.3389/fpubh.2018.00247

Moremi, K. E., Scott, C. J., de Jager, L. J., Pienaar, R., Zemlin, A. E., and Kotze, M. J. (2021). Implementation of a Pathology-Supported Genetic Testing Framework for Return of Research Results to Family Members of Deceased Breast Cancer Patients with Somatic *TP53* Variants. *The Breast* 56 (Suppl. 1), S64–S65. doi:10.1016/s0960-9776(21)00205-8

Myburgh, E. J., de Jager, J. J., Murray, E., Grant, K. A., Kotze, M. J., and de Klerk, H. (2021). The Cost Impact of Unselective vs Selective MammaPrint Testing in Early-Stage Breast Cancer in Southern Africa. *The Breast* 59, 87–93. ISSN 0960-9776. doi:10.1016/j.breast.2021.05.010

Okunola, A., Torrorey-Sawe, R., Baatjes, K. J., Zemlin, A. E., Erasmus, R. T., and Kotze, M. J. (2019). Investigation of the Role of Vitamin D Metabolism in South African Breast Cancer Patients Using Whole Exome Sequencing. *The Breast* 44 (Suppl. 1), S36. doi:10.1016/s0960-9776(19)30162-6

Oosthuizen, J., Kotze, M. J., Van Der Merwe, N., Myburgh, E. J., Bester, P., and van der Merwe, N. C. (2021). Globally Rare BRCA2 Variants with Founder Haplotypes in the South African Population: Implications for Point-of-Care Testing Based on a Single-Institution BRCA1/2 Next-Generation Sequencing Study. *Front. Oncol.* 10, 619469. doi:10.3389/fonc.2020.619469

Roberts, M. C., Fohner, A. E., Landry, L., Olstad, D. L., Smit, A. K., Turbitt, E., et al. (2021). Advancing Precision Public Health Using Human Genomics: Examples from the Field and Future Research Opportunities. *Genome Med.* 13 (1), 97. doi:10.1186/s13073-021-00911-0

Solomon, B. D., Pineda-Alvarez, D. E., Hadley, D. W., Teer, J. K., Cherukuri, P. F., Hansen, N. F., et al. (2011). Personalized Genomic Medicine: Lessons from the Exome. *Mol. Genet. Metab.* 104 (1-2), 189–191. doi:10.1016/j.ymgme.2011.06.022

Torrorey-Sawe, R., van der Merwe, N., Mining, S. K., and Kotze, M. J. (2020). 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.* 11 (170), 170–176. doi:10.3389/fgene.2020.00170

Conflict of Interest: MK is a non-executive director and shareholder of Gknowmix (Pty) Ltd., that developed a database resource for research translation under the auspices of the South African Research Council.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Genetics

REVIEW
published: 13 May 2022
doi: 10.3389/fgene.2022.797465

Genomics in Egypt: Current Status and Future Aspects

Eman Ahmed El-Attar[1], Rasha Mohamed Helmy Elkaffas[2], Sarah Ahmed Aglan[1], Iman S. Naga[3], Amira Nabil[4] and Hoda Y. Abdallah[5,6]*

[1]Chemical Pathology Department, Medical Research Institute, Alexandria University, Alexandria, Egypt, [2]Clinical and Chemical Pathology Department, Faculty of Medicine, Cairo University, Cairo, Egypt, [3]Department of Microbiology, Medical Research Institute, Alexandria University, Alexandria, Egypt, [4]Department of Human Genetics, Medical Research Institute, Alexandria University, Alexandria, Egypt, [5]Medical Genetics Unit, Histology and Cell Biology Department, Faculty of Medicine, Suez Canal University, Ismailia, Egypt, [6]Center of Excellence in Molecular and Cellular Medicine, Faculty of Medicine, Suez Canal University, Ismailia, Egypt

Egypt is the third most densely inhabited African country. Due to the economic burden and healthcare costs of overpopulation, genomic and genetic testing is a huge challenge. However, in the era of precision medicine, Egypt is taking a shift in approach from "one-size-fits all" to more personalized healthcare via advancing the practice of medical genetics and genomics across the country. This shift necessitates concrete knowledge of the Egyptian genome and related diseases to direct effective preventive, diagnostic and counseling services of prevalent genetic diseases in Egypt. Understanding disease molecular mechanisms will enhance the capacity for personalized interventions. From this perspective, we highlight research efforts and available services for rare genetic diseases, communicable diseases including the coronavirus 2019 disease (COVID19), and cancer. The current state of genetic services in Egypt including availability and access to genetic services is described. Drivers for applying genomics in Egypt are illustrated with a SWOT analysis of the current genetic/genomic services. Barriers to genetic service development in Egypt, whether economic, geographic, cultural or educational are discussed as well. The sensitive topic of communicating genomic results and its ethical considerations is also tackled. To understand disease pathogenesis, much can be gained through the advancement and integration of genomic technologies via clinical applications and research efforts in Egypt. Three main pillars of multidisciplinary collaboration for advancing genomics in Egypt are envisaged: resources, infrastructure and training. Finally, we highlight the recent national plan to establish a genome center that will aim to prepare a map of the Egyptian human genome to discover and accurately determine the genetic characteristics of various diseases. The Reference Genome Project for Egyptians and Ancient Egyptians will initialize a new genomics era in Egypt. We propose a multidisciplinary governance system in Egypt to support genomic medicine research efforts and integrate into the healthcare system whilst ensuring ethical conduct of data.

Keywords: genomics, Egypt, cancer, communicable disease, COVID 19, governance

OPEN ACCESS

Edited by:
Maritha J. Kotze, Stellenbosch University, South Africa

Reviewed by:
Yesim Aydin Son, Middle East Technical University, Turkey
Rajiv Erasmus, Stellenbosch University, South Africa

****Correspondence:***
Eman Ahmed El-Attar
eman.elattar@alexu.edu.eg

Specialty section:
This article was submitted to Human and Medical Genomics, a section of the journal Frontiers in Genetics

Received: *18 October 2021*
Accepted: *04 April 2022*
Published: *13 May 2022*

Citation:
El-Attar EA, Helmy Elkaffas RM, Aglan SA, Naga IS, Nabil A and Abdallah HY (2022) Genomics in Egypt: Current Status and Future Aspects. Front. Genet. 13:797465. doi: 10.3389/fgene.2022.797465

INTRODUCTION

Egypt is a densely occupied country with a total population exceeding 100,388,000, a life expectancy of 71.8 years and a healthy life expectancy of 63 years. Maternal mortality ratio approaches 37 per 100 000 live births. Approximately 26.2% of the population has health expenditures of more than 10% of household income. Both the basic health sector and medical research receive around 0.10 US dollars per capita as official development assistance. Under-five mortality rate is 20 per 1000 live births. These figures delineate the huge economic burden of healthcare costs due to overpopulation. In spite of all challenges, universal health coverage (service coverage index) reached 66 out of 100 in 2017 (World Health Organization, 2021).

Ever since the first draft of the Human genome project was published in 2001 and fully completed and published in 2003, the genomic field has been enormously growing, with huge discoveries and great applications improving healthcare and managing diseases that were once before deemed incurable. Genetic diseases are of particular concern in Egypt owing to the high consanguinity rate in the population (Shawky et al., 2011). Congenital genetic defects were noticed as anomalies thousands of years ago, since the discovery of sculptures and mummies in temples revealed some inherited disorders as those observed in Tutankhamun; cleft palate, oligodactyly, clubfoot (Hawass et al., 2010) as well as evidence of dwarfism, osteogenesis imperfecta and others that could be seen in mummies (Kozma, 2008).

This article presents the status of genomic knowledge, services, applications, needs, regulation and population acceptance in Egypt. It highlights some important challenges and efforts to overcome them, shedding the light on more aspects that need to be addressed and fulfilled.

Diseases and Health Indices in Egypt

Disease burden is the main driver for healthcare service prioritization and research efforts. In Egypt, several diseases are of particular concern such as hepatitis, specifically hepatitis C viral (HCV) infection, which was estimated to be 4.5—6.7% prevalent in 2016 (World Health Organization, 2016). A huge national screening program was initiated in 2018 for screening of the population for HCV and optimal treatment requires a great effort to eliminate the disease from Egypt (Abdel-Razek et al., 2019). Chronic respiratory, cardiovascular diseases, diabetes mellitus and cancer account for 28% of the mortality causes in 2019. It was reported that cancer's five-year prevalence was 278165, with new cases reaching 134632 in 2020. Liver cancer, the most reported among cancer cases, accounted for 20.7% of all cancer types and it is the most reported in males as well. Breast cancer comes next (16.4%) as the first among cancers that affect females, followed by non-Hodgkin lymphoma (5.4%), bladder cancer (7.9%) and lung cancer (4.9%) (Iarc.fr, 2022).

Current Need for Genomic Medicine and Applications in Egypt

Burden of Genetic Diseases

Genetic diseases have been reported to affect 2–5% of all live births. Two main factors lead to increased genetic disease burden in Egypt including old maternal age and consanguineous marriages.

Although the rate of consanguineous marriage is decreasing worldwide, most Middle Eastern Arabs, still have the custom of preferring consanguineous marriage (Hamamy and Alwan, 1994). Shawky et al., 2011 revealed that the overall frequency of consanguinity within Egyptians is around 35% which is considered quite high; however, this frequency varies by region (59.9% in rural areas & 17% in urban areas) (Shawky et al., 2011). The adverse health effects of consanguineous marriage include a greater risk not only on producing offspring which are homozygous for a deleterious recessive gene variant, but also individuals with increased susceptibility for polygenic disorders, still births, spontaneous abortions, child, or neonatal deaths, as well as congenital anomalies (Capmas, 2022). There is also a noticeable relationship between consanguinity and specific physical defects, behavior and psychiatric disorders (Bittles, 2001; Shawky et al., 2011). Accordingly, consanguinity represents a risk factor and a public health concern in Egypt for many unfavorable outcomes (Shawky et al., 2011). Variable genetic disorders are frequent in Egypt. Neurological disorders are the most diagnosed genetic disorders among Egyptian children followed by neuromuscular diseases, skeletal anomalies and inborn errors of metabolism as reported by Shawky et al., 2012 (Shawky et al., 2012) and by Afifi et al., 2010 (Afifi et al., 2010).

Availability and Accessibility to Genetic Services and Research Centers

The awareness of the importance of medical genetics in Egypt was well appreciated in the early 1960s by establishing the first medical genetics units at the pediatric departments of Cairo and Ain Shams Universities (https://www.asu.edu.eg/ce/84/page) (Temtamy and Hussen, 2017) In 1966, the specialty of Human genetics at the National Research Centre (NRC) (https://www.nrc.sci.eg/human-genetics-genome-research-division/) was established and in 1967, the medical genetics unit at the Medical Research Institute (MRI) in Alexandria commenced (https://mri.alexu.edu.eg/). This was followed by the establishment of medical genetics units in many other universities (Temtamy et al., 2010). Those centers include multiple specialized teams covering variable genetic disorders and offer genetic counseling, clinical diagnosis of patients with genetic disorders, follow up for patients with known or suspected genetic disorders and other supportive services for patients and their families, with a registry for genetic disorders including thousands of cases. To date, more than 21,000 families with rare genetic disorders or seeking genetic counseling, are registered since 1981 in the Medical Research Institute in Alexandria alone. It is noteworthy that there are continuous international collaborations between the aforementioned Institutions; NRC and MRI and research teams worldwide in attempt to reach the correct diagnosis and unravel the etiology and pathogenesis of rare disorders and identify causative mutations implicated in rare genetic syndromes. These research studies are either conducted on a cohort or per individual case and many of the past studies solved many cases and led to very promising and interesting results in the

studied rare genetic syndromes in different genetic subspecialities; neuromuscular disorders, growth disorders, ciliopathies, ocular anomalies, congenital heart diseases, skeletal disorders, dental genetic abnormalities, inherited metabolic disorders, craniofacial disorders, dermatological anomalies and other variable rare genetic disorders with multiple congenital anomalies or disorders with underlying chromosomal aberrations or microdeletions (Toomes et al., 1999; Zaki et al., 2007; Bielas et al., 2009; Handley et al., 2013; Rice et al., 2013; Traverso et al., 2013; Abdalla et al., 2014; Novarino et al., 2014; Abdalla et al., 2015; El-Hattab et al., 2016; Scott et al., 2016; Seifi et al., 2016; Abdalla et al., 2017; Fassad et al., 2018; Maddirevula et al., 2018; Patel et al., 2018; Fassad et al., 2020a; Fassad et al., 2020b; Chatron et al., 2020; Nabil et al., 2020; Shamseldin et al., 2020; Donato et al., 2021; Essawi et al., 2021; Horn et al., 2021; Meyer et al., 2021; Patel et al., 2021; Thomas et al., 2021). Moreover, several clinical genetic studies on categories of syndromes or rare disorders were conducted in Egypt. One example is the study conducted by Shawky et al., 2010 on a cohort of patients with limb anomalies to reveal the prevalence of isolated limb anomalies versus those with well-defined genetic syndromes or chromosomal aberrations (Shawky et al., 2022). Also comprehensive studies of the clinical, cytogenetic or molecular aspects of some rare disorders as Robinow syndrome and Roberts syndrome were acknowledged (Aglan et al., 2015; Ismail et al., 2016).

In 2002, the first national committee for community genetic services was settled and its main mission was setting prevention and management policy guidelines of genetic disorders. In the same year, the committee established a national genetic counseling program. The main objective was promoting free of charge genetic services. The program involved a hierarchical system of referrals from primary health care (PHC) to secondary care, then tertiary care level of services. The Ministry of Health and Population (MoH&P) started delivering primary and secondary care services through health facilities, where early detection is done at PHC level, followed by referral of cases to the secondary care services which include genetic counseling clinics in the catchment area to be consulted by well-trained secondary care physicians who offer diagnosis, counseling, treatment and referral for rehabilitation services. Complex cases are then referred to tertiary care provided through university and institutional genetic departments and units (Raouf, 2008).

Public services are also provided by non-governmental organizations (NGOs) for those who have genetic disorders at a reasonable cost. Moreover, private laboratories provide testing services involving several technologies including and not limited to cytogenetic, molecular genetic tests, metabolic diagnostic tests for screening inborn errors of metabolism. Some of the tests are either performed in those private laboratories or outsourced for the lack of technology or high cost and low demand, rendering the test highly expensive. In addition to the aforementioned centers offering genetic services, other centers offer research services as well. The Molecular biology Research and Diagnostics unit (MRDU) in the Clinical and Chemical Pathology Department, Cairo University (https://clinchemd.kasralainy.edu.eg/clinchem-labs is one of the top-notch research facilities for postgraduate studies. The Genomics Centre (GC) in Zewail city (https://zewailcity.edu.eg/) employs a multidisciplinary approach to better understand how genomes function in health and disease. Aswan Heart Centre (https://myf-egypt.org/aswan-heart-centre/) is one of the largest heart centers in Egypt and the Middle East. The center aims at defining the Egyptian genetic landscape, linking genotypic, phenotypic data based on individual levels for cardiovascular disorders (Shawky et al., 2012).

The Center of Excellence in Molecular & Cellular Medicine (http://fom.suez.edu.eg/) at Suez Canal University, Ismailia, serves genomic research starting from diagnosis to treatment in the area of Eastern Egypt including the Suez Canal area and Sinai. Although there is increased awareness for the need of genetic services, a gap between the need and the availability is present. In 2008, it was reported that only a small percentage (20%) of more than 4000 primary health facilities across the country offer the services of detection and referral of genetic disorders and only 11 genetic counseling clinics were distributed over 7 out of 27 governorates (Cairo, Giza, Alexandria, Portsaid, Elsharquia, ElMenia, Assiut) (Raouf, 2008). Unfortunately, there are no updated data on the current situation. A growingly dense population as in Egypt mandates continuous efforts to extend and expand service delivery as well as updating diagnostic technology to meet healthcare needs. Thorough planning for accessibility, extension of testing services and availability of advanced technologies at a reasonable cost and/or full coverage by insurance is of utmost importance. Thus, the role of national genomic medicine initiatives is to address these barriers and challenges (Stark et al., 2019).

Newborn Screening

Early detection of diseases is crucial for their management. In 1991, the first newborn screening for hypothyroidism in Egypt was conducted by human genetics experts from Cairo university, Medical Research Institute of Alexandria, Ain Shams and Mansoura universities. In 1996, the first newborn screening for metabolic disorders in Alexandria governorate took place through a study conducted by the Medical Research Institute where newborns were screened for three treatable inborn errors of metabolism, namely phenylketonuria "PKU", galactosemia and congenital hypothyroidism, aiming at early detection of these diseases and providing therapeutic regimens to guard against development of mental retardation (Ismail et al., 1996). Collectively, all the conducted studies have created an important awareness among physicians, scientists, decision makers, and the public about the importance of human genetics as a medical science in Egypt. The results of this neonatal screening of 15,000 newborns showed a high frequency of Phenylketonuria (1:7000), and of hypothyroidism (1:3000) (Temtamy, 1998), which convinced the health authorities in Egypt to start a mass neonatal screening program for hypothyroidism in all Egyptian governorates in 2003; screening for PKU was later added as well (Temtamy, 2019).

Rare diseases face the challenges of being appropriately diagnosed, receiving adequate care and affording personalized

medications (Stoller, 2018). However, a landmark in raising awareness on rare diseases among the public was in August 2021, since a presidential initiative to treat muscular atrophy with the government bearing its high cost and urged the early diagnosis of the disease. This implied the establishment of a national registry for spinal muscular atrophy and encouraged many genetic laboratories in the public and private sector to offer genetic testing for different types of the diseases (State Information Service, 2022).

Non-Communicable Diseases (NCDs) Genomics Initiatives

In 2018, NCDs were reported to be responsible for 84% of deaths of the Egyptian population with cardiovascular diseases (CVDs) representing the most common cause of mortality (40%) and cancer was next (13%). NCD-related premature mortality is strikingly high and affecting young people in initiating their economic independence through associated disabilities which frequently ends with death (World Health Organization, 2022). According to a review performed as part of the development of a framework for implementation of public health genomics in Africa, it is estimated that these dramatic figures are mostly due to changes in lifestyle associated with urbanization and sedentary lifestyle together with intake of high-salt and high-lipid diets as well as genetic factors (Policy paper: A framework for the implementation of genomic medicine for public health in Africa, 2022). A collaborative work between Egypt Ministry of health and population and the World Health Organisation (WHO) had resulted in developing a five-years multisectoral action plan for NCD prevention and control (2018–2022), with a primary goal to achieve a 15% reduction in premature mortality from NCDs by 2022 (Egyptian Ministry of Health and Population and World Health Organization, 2022).

Cardiac Genomics

CVDs are heterogeneous pathological conditions which have complex genetic etiology as well as environmental-molecular interactions. In ancient Egypt, atherosclerotic vascular disease had been reported after thorough examination of many mummies in various studies, one of them studied 44 mummies and showed that nearly half had evidence atherosclerosis (Butrous et al., 2020).

The role of genetics in CVDs in populations of African ancestry has been recently studied in a number of projects (Scott et al., 2016; Owolabi et al., 2019) However, the need for a national reference population data to identify the healthy state and to use it in variant interpretation was addressed in 2016 by the Egyptian Collaborative Cardiac Genomics (ECCO-GEN) Project through the Aswan Heart Center and the Magdi Yacoub Global Heart Foundation support. The project aimed to recruit 1,000 Egyptian volunteers free of CVDs (EHVols) from the local population and to identify genetic variation in genes previously shown to be involved in inherited cardiac conditions (ICCs), especially those involved in hypertrophic cardiomyopathy (HCM) and dilated cardiomyopathy (DCM) (Aguib et al., 2020). The first 391 samples of an Egyptian healthy volunteers cohort from the EHVol study were sequenced using a targeted panel and 1262 variants were identified in 27 cardiomyopathy genes, their sequencing data were deposited on the European Genome-Phenom Archive (EGA) under the accession code: EGAD00001006160 and the dataset was published on May 20,2020. The ECCO-GEN project aims at providing individual-level genetic and phenotypic data to support future studies in CVD and population genetics.

In 2020, a pilot study was published that investigated sarcomeric and non-sarcomeric variants in a cohort of idiopathic pediatric hypertrophic cardiomyopathy patients using next generation sequencing and suggested that due to the population's high consanguinity burden, a particular genetic background contributes significantly to pediatric hypertrophic cardiomyopathy patients in Egypt (Darwish et al., 2020).

Cancer

According to the WHO Globocan 2020 cancer report about cancer prevalence in Egypt, the commonest sites for cancers were the liver (20.7%) and breast (16.4%) for both sexes (Cancer today, 2022). Ibrahim et al. developed a mathematical model and estimated a 3-fold increase in cancer incidence in Egypt by 2050 compared to 2013 (Ibrahim et al., 2014). These figures have drawn the attention of the Egyptian authorities to prioritizing cancer as a national health problem that needs more efforts among both research and services sectors. The genetic background of cancer has directed Egyptian researchers concerned with this discipline to focus more on genetics and genomics. The Egyptian population is a heterogeneous population (Tishkoff et al., 2009) and accordingly this necessitates implementing genomic/genetic studies to reflect the diversity among its different subpopulations. In this section, we will shed the light on genetic/genomic research efforts relating to most prevalent two types of cancers in Egypt: breast and liver cancers.

Breast Cancer and the Egyptian Women's Health Initiative

Breast cancer (BC) history in Egypt dates to around 3000 BC in the ancient Egyptian medical documents - the Edwin Smith Papyrus, the first and oldest cancer record as shown in **Figure 1** (Ades et al., 2017). It stated that bulging lesions of the breast were a grave disease, and the ancient Egyptians tried to treat these lesions with cautery, knives, and salts.

The earliest genomic research on BC in Egypt focused on unraveling the role of genetic variations in the tumor suppressor genes *BRCA1/2* and *TP53* (El- A Helal et al., 2000; Hussein and Hassan, 2004; Swellam et al., 2004; Saleh et al., 2004) Other research focused on identification of novel *BRCA1*/2 variants and their role in early-onset and sporadic breast and/or ovarian cancer in Egypt (Abdel-Mohsen et al., 2016). As the studies on the *BRCA1/2* genes have expanded, new population-based large genomic rearrangements have been revealed (Hagag et al., 2013; Eid et al., 2017). Considering *TP53* variant research, several researchers initially focused on a common codon 72 (Hussein and Hassan, 2004; Saleh et al., 2004; El-Ghannam et al., 2011). For

other DNA repair genes, *XPD* was the most studied one in Egypt (Hussien et al., 2012; Ramadan et al., 2014). Although *BRCA1*/2 are the most studied genes in Egyptian cancer research, clearly there is still a knowledge gap in determining the mutational effects that determine disease incidence and clinical outcomes in BC. Epigenetic aspects have been studied as important contributors to the risk and prognosis of BC reported in Egypt, including the role of microRNAs in BC tumor tissues, circulating free miRNA, circulating long non-coding RNA (Hafez et al., 2012; Zidan et al., 2018) in addition to DNA methylation status of known susceptibility genes such as *FHIT* (Zaki et al., 2015).

As a part of the country's endeavors for advancing Egyptian population health, the President's "Egyptian Women's Health Initiative" was launched in 2020 to promote the health of Egyptian women (Presidency.eg, 2022). This initiative was a major step against the alarming increase in BC among Egyptian women aiming at screening 28 million women in all Egyptian governorates; emphasizing on regular self-examination; and providing free-of-charge treatment.

However, due to the high expenses of these testing procedures, BC genetic testing could not be provided in this initiative. Genetic testing services for BC are mostly available via the private sector or non-governmental organizations (NGOs), and is utilized in BC prevention, recurrence risk prediction, metastatic BC and BC survivors. Examples of bodies and private labs providing this service are the Breast Cancer Foundation of Egypt, Genetic Diagnostic Center, Hassan Healthcare LLC and Generations Genetic Laboratory.

Liver Cancer

Liver cancer (LC) is the most prevalent cancer among males in Egypt (Ibrahim et al., 2014). It is reported that chronic inflammation of the liver like non-alcoholic fatty liver disease (NAFLD), non-alcoholic steatohepatitis (NASH), and liver cirrhosis are major predisposing factors for the development of hepatocellular carcinoma (HCC) in Egypt. Viral hepatitis is an important causative agent of these liver diseases finally leading to LC in Egypt (Rotimi et al., 2020).

Most of the genetic studies on LC in Egypt focused on the contribution of *TP53* expression on different genetic mutations, and other tumor suppressor genes like *TP73, RB, KLF6, and CTNNB1*, to liver carcinogenesis (El-Kafrawy et al., 2005; El Far et al., 2006; Wahab et al., 2010).

Regarding epigenetics studies, Egyptian researchers have focused on discovering novel biomarkers for LC like microRNAs. These studied the role of serum miRNA-224, miRNA-215, miRNA-143, miRNA-122, miRNA-199a, and miRNA-16 (El-Abd et al., 2015; Mamdouh et al., 2017). Motawi et al. reported the contributing role of miRNA-122 and miRNA-222 as a discriminating biomarker for distinguishing liver injury from LC (Motawi et al., 2016). They also unraveled the possible role of the genetic variants LncRNA HULC rs7763881 and MALAT rs619586 in the progression of

FIGURE 1 | Edwin Smith papyrus (Image credit — U.S. National Library of Medicine).

hepatitis virus-persistent carriers' progression to LC (Motawi et al., 2019).

Communicable Diseases and SARS-CoV-2 Research Efforts

Coronavirus disease 2019

The Corona virus disease 2019 "COVID-19" started in Wuhan, China by the end of December 2019 (Shigemura et al., 2020) and has quickly spread throughout the world (Ojha et al., 2020). The causative agent of this infectious disease of the respiratory tract is Severe Acute Respiratory Syndrome Coronavirus 2 "SARS-CoV-2" (Lake, 2020). On March 11th, 2020, COVID-19 was declared a pandemic by the WHO.

The importance of disease monitoring systems and genome sequencing technology in improving public health has been highlighted by COVID-19 (Institute of Pathogen Genomics, 2020). In Africa a total of 140 disease outbreaks are anticipated to occur each year (World Health Organization, 2020) which are additional to the threats of endemic infectious diseases, which make up approximately 35 percent of Africa's 10 million annual fatalities (Roser and Ritchie, 2016). These figures highlight the crucial need for Africa's epidemic preparedness and surveillance systems to be expanded, including genomic and digital surveillance capabilities (Eib.org, 2020) and manufacturing capabilities for diagnostics and treatments in the local area (Nkengasong, 2020).

Pathogen genomics can revolutionize public health surveillance through enhancing outbreak detection, following up on transmission routes, searching for genetic changes that affect infectivity, diagnosis, treatments, and vaccinations, as well as evaluating the efficacy of interventions (Armstrong et al., 2019). As a consequence, public health experts will be able to stay on top of new micro-organisms and re-emerging diseases. It will enable health systems to more effectively combat pathogens that cause diseases such as COVID-19, polio, malaria, tuberculosis, HIV/AIDS, cholera, Ebola, and other emerging health issues as well as antibiotic resistance (Institute of Pathogen Genomics, 2020).

Rapid advances in sequencing technology have led to development of dependable next-generation sequencing (NGS) equipment that can resolve pathogens at a high resolution. Despite the considerable need to reduce Africa's substantial burden of infectious diseases, NGS application is limited (Inzaule et al., 2021). The capacity for sequencing in Africa is scant and limited. About 71% of next-generation sequencers are clustered in just five African countries: South Africa (n = 79; 38%), Kenya (n = 28; 14%), Nigeria (n = 13; 6%), Morocco (n = 18; 9%), and Egypt (n = 10; 5%). Sanger-based assays are used in the majority of African countries. Due to the high prices of reliable NGS equipment and techniques, financial constraints, very few resources and inadequate infrastructure and framework for data-sharing, its full potential has yet to be realized in Africa. Democratizing the use of NGS technologies in Africa and other low- and middle-income regions will enable partnerships and overturn the habit of relying on partners from the global north for cutting-edge global health innovation (Institute of Pathogen Genomics, 2020).

Africa CDC Institute of Pathogen Genomics, through the Africa Pathogen Genomics Initiative (Africa PGI) intends to improve disease surveillance and public health collaborations (Institute of Pathogen Genomics, 2020). Various approaches are being implemented in collaboration with the Africa CDC-led African Task Force for Coronavirus Preparedness and Response (AFTCOR) laboratory technical working group.

First, resources are being mobilized to speed SARS-CoV-2 sequencing across Africa, exploiting the continent's available next-generation sequencing capabilities, and 16 countries are receiving resources and technical assistance. Second, a network of sequencing laboratories, genomics, and bioinformatics professionals is in the making to pool resources, information and expertise. Third, policies are being created to ensure that data is representative and repeatable. Fourth, sample referral methods are being developed with regional laboratories to network countries having insufficient sequencing capacity (Tessema et al., 2020).

SARS-CoV-2 has evolved over time, and genomic data has been used to identify and track nosocomial infection spread in South Africa (Giandhari et al., 2021), sources of introduction and lineages (Tessema et al., 2020), including the emergence and disperse of more transmissible variants with the potential to affect disease activity and countermeasures such as vaccines (Toyoshima et al., 2020; World Health Organization, 2016).

Omicron variant of SARS-CoV-2 has been first reported in Egypt December the 18th 2021 in three cases (Al-Youm, 2021). In order to respond more quickly to the Omicron variant and the increase in cases, WHO is assisting countries in improving genomic surveillance to trace the virus and find additional potential variants that could arise. A regional genome sequencing laboratory in South Africa is assisting 14 southern African countries and has considerably increased sequencing capacity. Southern African countries sequenced only 5500 samples in the first half of 2021. They're currently sequencing the same number every month (World Health Organization, 2022).

The first country in the Middle East to be struck by the COVID-19 pandemic was Iran, followed by Turkey, Saudi Arabia, Qatar, and Egypt (Temtamy et al., 2010). Although the number of cases was low in the beginning, it has accelerated at an unprecedented rate, hitting all Middle Eastern countries. This could be a reflection of the readiness and response measures taking place in Middle Eastern countries. The Middle East, which sits at the crossroads of Eurasia and Africa, has played a significant role in shaping modern civilizations and religions. It is a highly sensitive area in terms of geography, economics, politics, culture, and religion (Baloch et al., 2020). There are significant obstacles in limiting the spread of COVID-19 in the Middle East. Compromised healthcare systems, long-term regional conflicts and humanitarian catastrophes, lack of transparency and cooperation, and frequent religious gatherings are just a few examples. These variables are intertwined, and their combined effect determines how this region responds to the epidemic (Baloch et al., 2020).

In Egypt, from January 3rd, 2020 to September 3rd, 2021, there have been 288,732 cases of COVID-19 with 16,743 deaths,

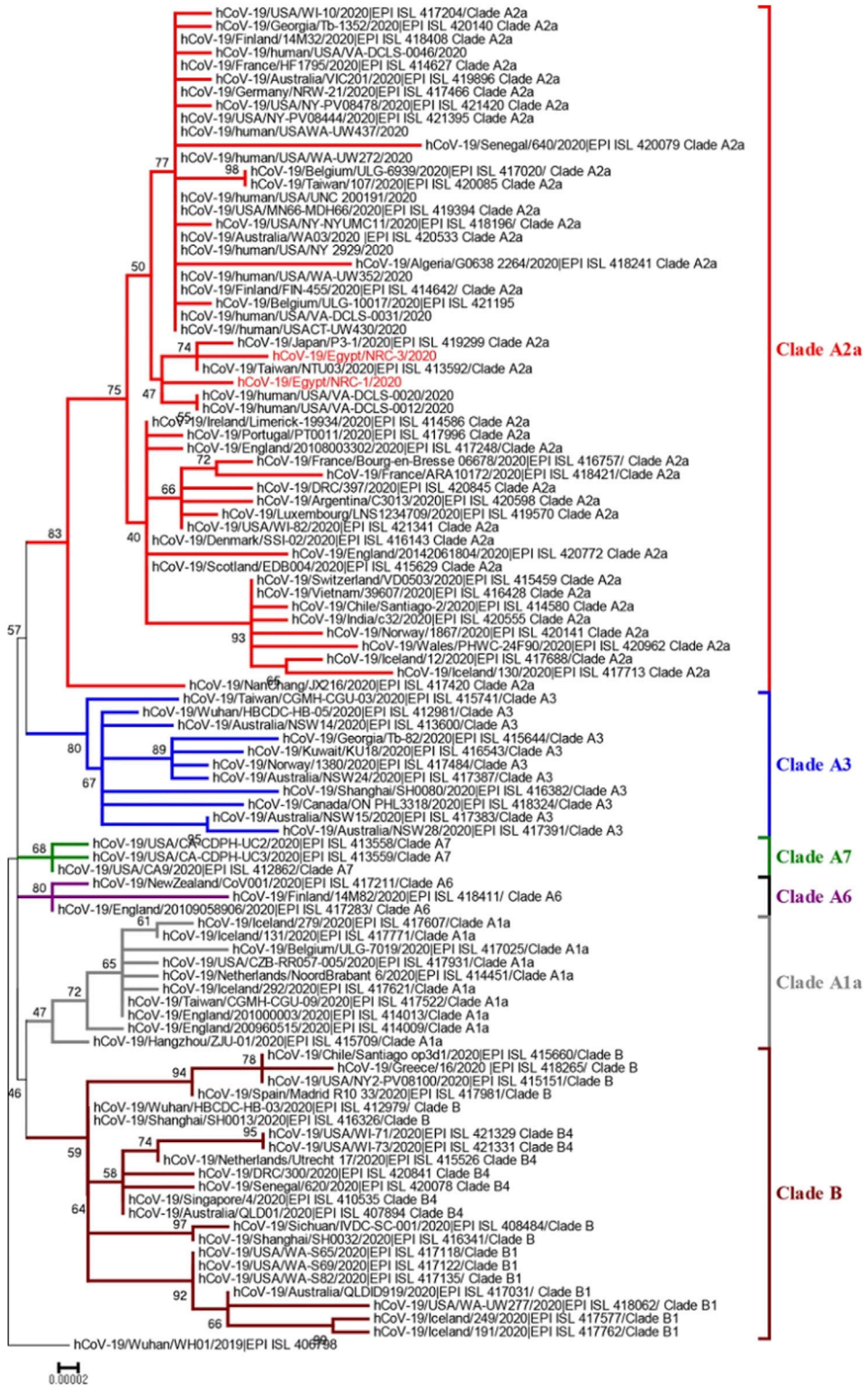


FIGURE 2 | Neighbor-joining phylogenetic tree of a SARS-CoV-2 strain from Egypt and other global strains. The two Egyptian strains are shown in red. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1,000 replicates) is shown at the dendrogram nodes. The phylogenetic analysis was performed using MEGA7 (Kandeil et al., 2020). Source: Coding-Complete Genome Sequences of Two SARS-CoV-2 Isolates from Egypt. Permission Request ID: 600053988.

reported to WHO and as of August 27th, 2021, a total of 8,741,005 vaccine doses have been given (WHO (COVID-19) Dashboard with Vaccination Data, 2022).

Using next generation sequencing; interpretation of viral genome sequences presents insight about the pattern of distribution around the world, the genetic diversity during epidemics and pandemics as well as the kinetics of the development of different subtypes. SARS-CoV-2 database, such as "GISAID (www.gisaid.org)" and the "NCBI SARS-CoV-2 database (https://www.ncbi.nlm.nih.gov/sars-cov-2)" allow the genomic data, as well as epidemiological data for the sequenced isolates, to be made public (Zekri et al., 2021a).

Until September 8th,2021 on GISAID index.php, there are 977 genomes shared on GISAID from Egypt. Few genomic studies were conducted in Egypt due to the high economic burden.

In a study by Kandeil et al. (Kandeil et al., 2020) in May 2020 they announced the genome sequences of two SARS-CoV-2 isolates found in patients from Egypt "The isolate hCoV-19/Egypt/NRC-1/2020 and hCoV-19/Egypt/NRC-3/2020". These sequences are available at the GISAID-EpiCoV newly emerging coronavirus SARS-CoV-2 platform with the listed identifiers "EPI_ISL_430819 and EPI_ISL_430820" available at https://www.gisaid.org/. These strains belong to clade A2a, which cover strains from different countries like Asia, Europe, the United States, Australia, and Africa (**Figure 2**). Another study by Zekri et al. (Zekri et al., 2021a) in November 2020 conducted a study on 61 whole genome sequences from Egyptian patients. They detected 204 unique sequence variations in Egyptian patients' genomes. It was concluded that most Egyptian genomic strains sequenced were close to isolates originating from the United States of America, Austria, Sweden, Saudi Arabia and France.

Further work by Zekri et al. (Zekri et al., 2021b) in April 2021 used next-generation sequencing to search for mutation hotspots in Egyptian COVID-19 patients and identify dominant variations that may be linked to differences in clinical presentations. They concluded that D614G/spike-glycoprotein and P4715L/RNA-dependent-RNA polymerase were the most common mutations among Egyptian isolates. These mutations were linked to transmissibility regardless of symptom changeability. E3909G-nsp7 could answer why children recover so quickly. Nsp6-L3606fs, spike-glycoprotein-V6fs, and nsp13-S5398L variants may be associated with clinical manifestations intensifying.

Furuse Y et al. (Furuse, 2021) in his recent article stated that forty-nine countries have already published >100 SARS-CoV-2 genomic sequences with large number of all published genomic sequences (n = 93,817) originating from United Kingdom (38.9%) and United States (22.7%). However, only two sequences originated from Egypt which emphasizes the need to increase sequencing capacities and funding possibilities.

Egypt's death rate from infectious diseases has gradually decreased, from 20.87 percent in 2000 to 9.63 percent in 2019 (Knoema.com, 2022). Some of the major communicable diseases in Egypt include Typhoid fever, hepatitis virus and tuberculosis.

There are huge efforts dedicated to genomic technology in the field of microbiology in Egypt. The microbiology field has been completely altered using genomics. Genomic analyses are yielding unparalleled insights into microbial evolution and diversity and are explaining the complexity of the genetic variation in both host and pathogens that underlies disease (Knoema.com, 2019).

In Egypt, Shoeib et al. (Shoeib et al., 2020) performed whole-genome analysis on a fecal specimen obtained from a hospitalized 6-month-old child with acute gastroenteritis in 2012. They unexpected rotavirus group A (RVA) strain "RVA/Human-tc/EGY/AS997/2012/G9 [14]". Phylogenetic analysis designated that the strain AS997 had the consensus P[14] genotype cluster with G9, T1 and H1 reassortment. This finding proposes either a mixed gene arrangement developed from a human Wa-like strain with a P[14]-containing animal virus, or that this P[14] was obtained only through the reassortment of human strains. The study highlights the significance of whole-genome analysis in surveillance studies of rotavirus strains and reveals the potential roles of interspecies transmission and numerous reassortments resulting in the development of new rotavirus genotypes.

Hepatitis viruses

Viral hepatitis is a major public health issue and a cause of mortality in Egypt. Around 8–10 millions are living with viral hepatitis in Egypt (Centers for Disease Control and Prevention, 2012; Globalhep.org, 2014). The most common causes of viral hepatitis in Egypt are Hepatitis A virus (HAV) and hepatitis E virus (HEV). By the age of 15, approximately half of the Egyptian population had been infected with HAV. Moreover, more than 60% of the Egyptian population are anti-HEV positive in the first ten years of life. Hepatitis B virus (HBV), hepatitis C virus (HCV), and hepatitis D virus (HDV) are the most significant causes for chronic hepatitis, liver cirrhosis, and hepatocellular carcinoma (HCC) in Egypt (Elbahrawy et al., 2021).

Egypt has a moderate endemic level for HBV with only around 3.3 million people infected. In 1992 after the initiation of mass immunization program for infants in Egypt; the prevalence rate has declined (Sherif et al., 1985). Statistically significant decrease in anti-HCV prevalence has been reported among the general population ranging from 0 to 51% (median value: 13%) (Kouyoumjian et al., 2018) The prevalence of HCV was 8.7–40.3% (Kandeel et al., 2017; Bayomy Helal et al., 2018; El-Ghitany and Farghaly, 2019; Soliman et al., 2019) before the National HCV treatment program was carried out. A large-scale national study conducted during 2018–2019 outlined a much lower rate of HCV prevalence (4.6%) (Waked et al., 2020).

In 2016, 194 member states of the WHO had devoted themselves by 2030 viral hepatitis will no longer be a public health hazard, focusing mainly on HBV and HCV infection. A "National Committee for the Control of Viral Hepatitis (NCCVH)" was established by the MOH&P in 2006. By April 2008, this committee had developed a national control strategy for viral hepatitis, which mandated improvements in preventive actions regarding reducing the incidence of HBV and HCV infections and offered more efficient treatment accessibility for patients with chronic hepatitis. Egypt held the largest mass screening and treatment campaign for HCV in late 2018,

which allowed screening 50 million people for HCV infection as a step toward disease eradication (Waked et al., 2020).

Tuberculosis

Tuberculosis (TB) is another major public health threat worldwide, competing with HIV as the cause of death due to infectious diseases worldwide (Sulis et al., 2014). According to the WHO, around 8.6 million TB cases were estimated in 2012. Most cases are in Asia and Africa (58 and 27%, respectively) (World Health Organization, 2013). TB is still a health problem in Egypt, affecting the young active age group. TB is a significant challenge to public health officials, particularly "drug-resistant (DR) and multidrug-resistant (MDR)" isolates of *Mycobacterium tuberculosis*. Alyamani et al. (Alyamani et al., 2019) reported that around 35% of TB isolates are MDR in Egypt which is relatively high. Phylogenetic and molecular dating analyses showed that lineages coming from Egypt recently diverged (~78 years). In contrast to drug-sensitive isolates, drug-resistant isolates are not clustered or largely disseminating, proposing deficient treatment as the main reason for antimicrobial resistance emergence rather than higher virulence or more capacity to survive.

We are witnessing nowadays a revolution in sequencing technology rendering it rapid, and robust. The advancement in this technology impacted greatly COVID-19 global response. The rapid sequence data sharing has allowed countries to develop assays for detection, helped in deeper understanding of pathogenesis and better prediction of action as well as fast production of vaccines. It has become clear in this pandemic that global pathogen genomic surveillance strategy is needed indeed to enhance preparedness for pathogens with pandemic or epidemic potentials. The WHO has taken an initiative setting a draft for Global genomic surveillance strategy and invited various stakeholders; academia, member states, national health authorities, industry, and civil societies to review the draft in December 2021. The draft represents an important step towards empowering public health actions in response to pandemics and epidemics around the world, supporting and encouraging collaborative work and data and specimens sharing (World Health Organization, 2022). Genomic surveillance for pathogens is still costly for many countries including Egypt. However, given its impact on public and global health, governments should advocate pathogen genomic surveillance establishment through direction of funds towards capacity building and training.

Challenges and Barriers to Genomic Testing

The huge leap in genomic information and applications is continuing to push the research efforts and regulations worldwide to keep pace with new inputs and information processing. Genomic medicine impacts almost every aspect of medical practice starting from preconception carrier risk through prenatal genetic diseases diagnosis and newborn screening to cancer risk prediction, precision medicine and gene therapy. Such an ever-growing field mandates continuous training, close monitoring and regulating as well as re-examining data gathering ethics and results conduction methods to patients and families. Identification of challenges is fundamental to improving genomic services as well as establishment of strategic plans to fill the gap towards advancement (**Figure 3**).

Service Availability and Resources

Availability of genetic and genomic services remains a big challenge in Egypt. Although reasonably available in urban areas, rural areas might not have easy access to those services. Moreover, the financial aspect is a fundamental obstacle especially for rural areas, given its high cost and insurance coverage issues. Private sector remains a key player in providing such services, however, some tests are not yet readily available and are outsourced outside the country, which makes the service even more expensive, nevertheless the low demand on such services makes the situation more challenging.

Geographical Barriers and Access to Genetic Services

Most University and institutional genetic services are free of charge or of low cost. The national health insurance organization which covers elementary school students provides reimbursement for investigations as well as treatment and rehabilitation services. However, the coverage of genetic tests by the public sector is limited and services for low-income citizens are frequently provided through donations from NGOs or charity. Genetic services provided by the private sector are covered mainly by out-of-pocket payments or reimbursed by private health insurance companies.

A huge step towards standardization, improvement and coverage of healthcare services is expected after the establishment of the new health insurance system in Egypt. Founded in 2019, it will be implemented over six phases to include all governorates and every Egyptian. It aims at maintaining health records for all patients with coverage of most of the service costs and availability of services even in private sectors. Healthcare service providers shall be monitored and accredited by the corresponding accrediting body to ensure quality of service (Gov.eg, 2022).

Economic Barriers

The validation of laboratory genetic testing requires the lab to join an external quality assurance (EQA) program, which is usually an international one and difference in currency makes the registration in these programs of high cost. Moreover, the expensive genetic analyzers and kits all add to the cost of genetic testing, rendering it unaffordable for a large sector of the population. Lack of reimbursement is also an issue that faces a large portion of patients and the dependence on out-of-pocket services.

Population Acceptance and Cultural Barriers

An important aspect of conducting genetic testing is public acceptance of decisions relying on those test results, otherwise

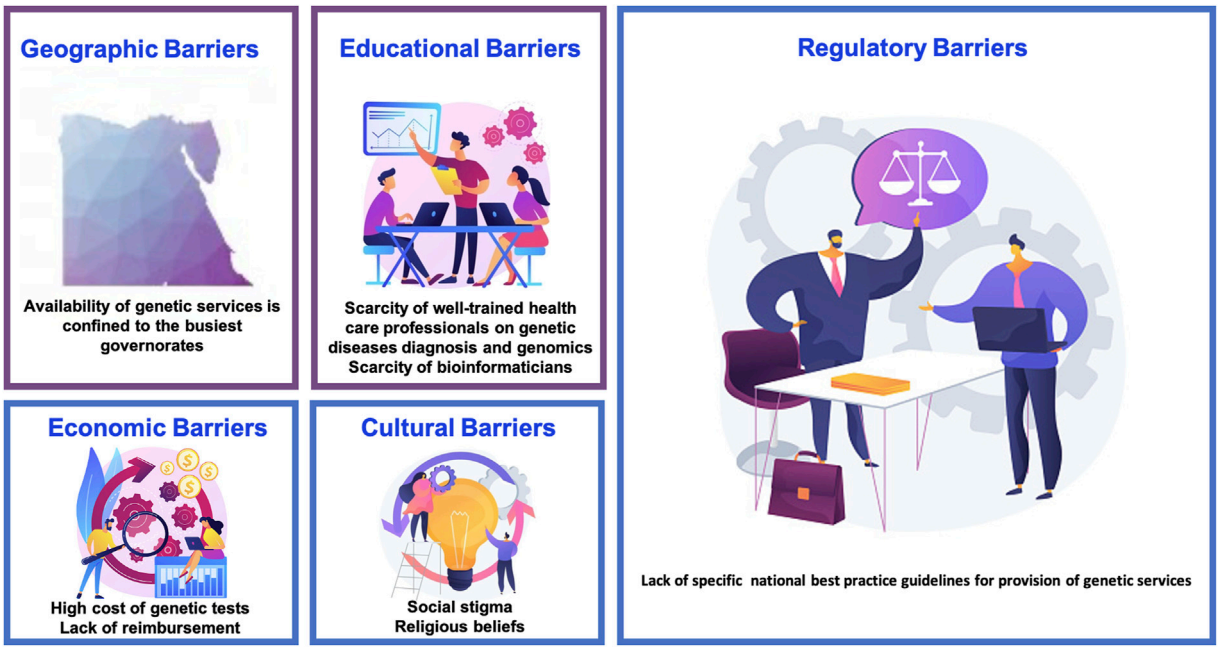


FIGURE 3 | Barriers to genomic/genetic services in Egypt.

the whole benefit from testing would be questionable. A study conducted by Elgawhary et al. (2008) offered 30 Egyptian females with high-risk pregnancies for beta thalassemia DNA sequencing for fetal tissue prenatally. Out of 22 cases with affected fetuses 14 couples refused to terminate pregnancy reflecting cultural and religious beliefs (Elgawhary et al., 2008) Fear of being socially stigmatized for using genetic services or having a genetic condition as well is particularly common in rural areas, highlighting the cultural considerations for Egyptians (Raouf, 2008).

Public awareness of genetic testing importance and applications should be addressed in media as well as incorporating religious entities to refute any erroneous cultural or religious beliefs that might hinder healthcare conduct.

Educational Barriers

Genomic medicine is a relatively new field, knowledge of its importance and application is a cornerstone in providing health care. However, there is still a gap between research studies, application and attitude towards this field in Egypt. Despite the mounting research efforts as aforementioned in the article, physician awareness and knowledge of genomic applications still needs improvement. A cross sectional study by Nagy et al. (2020) was conducted in one of the largest children's cancer management hospitals; Children's cancer hospital 57357 through a survey questioning knowledge and attitude towards pharmacogenetic testing. It highlighted a lack of knowledge of pharmacogenomics and test unavailability due to limited funding. The study attributed defective knowledge due to non-integration of such fields in the curricula of undergraduates and postgraduates as well as lack of training and funding (Nagy et al., 2020).

Communication and Regulation

One of the pillars of genomic medicine and healthcare management is reporting and communicating test results. However, it is particularly complex in genomic testing settings. It could be partially attributed to the fact that some results could be ambiguous, providing uncertainty more than answers such as reporting of variants of uncertain significance. Such result reporting and communicating to patients is indeed challenging and a well -trained counselor should conduct such results within the frame of a guideline on how to deal with such results (Medendorp et al., 2020) A national standard should be present to set guidance for who to conduct results and how.

"The right not to know" remains a great ethical concern and an area of conflict when it comes to reporting and communicating secondary or incidental findings that would affect management of patients. Also, communicating results that would affect family members as well as the patient. Patients could choose not to know such important data, leaving the physician in a difficult situation that needs ethical or legal consultation (Marchant et al., 2020). Such situations need to be identified, addressed, and regulated with thorough training of physicians and the judiciary system by the Egyptian government. Nevertheless, regulating direct to consumer genetic tests availability, reporting and result communication and reliability is of equal importance as well.

Complex testing procedures and reports should be governed by an accrediting body and continuously monitored starting from test selection by setting criteria for diagnostic versus research-based tests, what should be offered and what should not be offered, result reporting and

communication. General authority for healthcare accreditation and regulation (GAHAR) was established in 2018 under law No. Two for that year, it provides quality standards for laboratories and healthcare facilities by setting standards and inspection services. Currently 19 healthcare facilities are accredited, and many are awaiting accreditation. Such a great step towards standardization and improvement of quality in healthcare in Egypt will reshape health service and give hope to tackling such important aspects of genomic testing in the country (The General Authority for Health Accreditation and Control, 2022).

Establishing national policy initiatives on genetic services is essential. Some laboratories follow the internationally published guidelines as those of the American College of Medical Genetics and Genomics (ACMG), however it is essential to point out that the needs and priorities in developing countries are to a great extent different than those in developed countries (Mohamed, 2015). An important progress by policymakers was in 2020 when the Egyptian parliament approved two important laws that can benefit genetics research as well as genetics services: the data protection law (Privacylaws.com, 2022) and the clinical medical research law (Egypt Independent, 2020). However, the lack of specific national best practice guidelines for provision of genetic services constitutes a major problem in providing high quality genetic/genomic testing services as well as high impact academic research.

Current Status and Improvement Efforts

The Egyptian MoH&P was keen on implementing a holistic plan related to the upgrading of genetic services in Egypt with special consideration to the prevention and early management of disabilities via addressing three important dimensions:

- Creating registries through an integrated system for treatment and rehabilitation.
- Various programs for prevention and early detection of diseases as (World Health Organization, 2021) premarital care counseling; (Shawky et al., 2011) motherhood program covering all stages from antenatal care, childbirth care, neonatal care and post-natal care (with a focus on congenital anomalies, and early detection of causes of mental retardation); (Hawass et al., 2010) newborn screening program; (Kozma, 2008) childcare program for monitoring child growth and development, immunization and nutrition.
- Supporting rehabilitation centers and introducing the concept of community-based rehabilitation (CBR).

Egypt has also increased investments to fund research in medical genetics and genomics. The main funder is the governmental sector from the ministry of higher education and scientific research *via* the Science, Technology and Innovation Funding Agency (STIFA) and Academy of Scientific Research and Technology (ASRT).

The African Society of Human Genetics (AfSHG), founded in 2003, aims to address the public health burden of diseases across the continent and to develop a road map for translating genomic knowledge across the continent. Through this society, a Pan-African research consortium was established, namely The Human Heredity and Health in African Consortium (H3Africa). It was funded largely by the NIH in the United States and the Wellcome Trust in the United Kingdom and aimed to focus on human health–related genomics and genetics research in Africa. H3Africa funded the implementation of large-scale genomic technologies through collaborative centers in 27 African countries (http://h3africa.org) (El-Kamah et al., 2020).

The "African Genomic Medicine Training Initiative, established in 2016, provides Africa-wide virtual distance learning genomic medicine courses with the aim of developing competencies in Genomic medicine and exploring its practical application targeting African health professionals (H3abionet.org, 2022). The course is hosted by different African countries including pre-recorded lectures & supported by local facilitators linking attendees to the trainer Since 2017, increasing number of "classrooms" are joining this initiative. In 2022, the third iteration was hosted by 27 classrooms in 11 countries, five of these classrooms were in Egypt (Kasralainy hospitals, Cairo University, Suez Canal University, Faculty of Medicine, Modern University for Technology and Information, Faculty of Medicine, Galala University, Faculty of Nursing, Ain Shams University). Egyptian classrooms were the second most frequent to host the course after Nigerian classrooms (Nembaware and Mulder, 2019). This shows the increased interest among Egyptian health professionals to learn about Genomic medicine and to build a capacity of Egyptian trainers who can develop further genomic courses tailored for the Egyptian context in terms of disease burden and available genomic health care services.

The tenth conference of the AfSHG was held in partnership with the National Society of Human Genetics in Egypt and the H3Africa Consortium in Cairo in November 2017. An important outcome of these young research forums (YRFs) at the AfSHG meetings is the development of formal and informal networks to enhance relationships and develop future partnerships among emerging researchers. In the same context, through the 3rd Conference of the Chemical Pathology department, Medical Research Institute, Alexandria University in 2019 a scientific collaboration channel was created with Stellenbosch University in South Africa supported by the Technology Innovation Agency in South Africa through a travel grant. The common interest was implementation of personalized medicine in Breast cancer. Research performed over more than a decade in South Africa (136) led to the development of a rapid BRCA1/2 founder/recurrent mutation assay for improved clinical management of breast cancer and associated co-morbidities. Their vision seeks to move basic research to translation by up-scaling of clinical uptake and implementation of point-of-care genetic testing using an adaptive pathology-supported genetic testing (PSGT) framework (Kotze et al., 2013; Kotze, 2016; van der Merwe and Kotze, 2018). Much of the work in health-related genomics on the African continent is currently being carried out in the context of research. However, to ensure health benefits, advances in diagnosis and treatment of cancer and other NCDs need to shift towards implementation in healthcare settings. In November 2019,

TABLE 1 | Comparison between the current genomic projects in Egypt and Africa.

	Pilot egyptian human genome project (EgyptRef)	The national reference genome project for egyptians and ancient egyptians	H3Africa international Genome project
Duration	2018 and ongoing	2021–2025 (first phase)	2010 and ongoing
Aims and scope	To establish a reference genome for Egyptian and North African populations to complement the Genome Reference Consortium human genome (GRCh)	To establish a benchmark genome center for creating the map of the Egyptian human genome to discover and accurately determine the genetic characteristics of various diseases aiming to help Egypt in entering the era of precision medicine. The project's scope is classified into three categories: A- The population genome B- The Genome of the ancient Egyptians, and C- Diseased genome.	H3Africa empowers African researchers to be competitive in genomic sciences, establishes and nurtures effective collaborations among African researchers on the African continent, and generates unique data that could be used to improve both African and global health.
Website	https://www.egyptian-genome.org	Under Construction	https://h3africa.org
Coordinator	The genetics and systems biology divisions of LIED, Lübeck University, Germany	Medical Research and Regenerative Medicine Center at the Ministry of Defense, Egypt	The African Society of Human Genetics
Partners	The Center for Experimental Medical Research (MERC), Mansoura University, Egypt	Thirteen Temtamy and Hussen, (2017) Egyptian universities and research centers from the Ministries of Defense, Higher Education, Scientific Research, Health and Communications, and a number of civil society institutions.	The National Institutes of Health (NIH) The Wellcome Trust (WT) The African Academy of Sciences (The AAS)
Funding	The German Science Foundation, excellence program (EXC 306) and the DAAD	Egyptian governmental funding a total of 62,500,000 dollars for first phase over 5 years	NIH Director's Common Fund. Till now, H3Africa activities has been supported with 176 million dollars investment by NIH/WT
Data sharing	Public access	No data Yet	Public access

the Technology Innovation Agencies in South African and Egypt signed a collaboration agreement that is intended to further enable the promotion of targeted market-oriented research cooperation and technology innovation partnerships between the two countries.

The PSGT framework for moving to a translational phase was implemented as a case study in South Africa. This framework is based on providing a personalized patient experience which aims to distinguish between inherited and environmental causal factors such as lifestyle (epigenetic) or therapy-induced (pharmacogenetic) NCDs for risk stratification and clinical management of multi-factorial and polygenic disorders (Policy Paper: A Framework for the Implementation of Genomic Medicine for Public Health in Africa, 2022). Genomic medicine requires expensive laboratory equipment, and it may take many days to weeks for result generation. In a resource limited African setting such as Egypt, this could lead to loss-in-follow-up and thereby ineffective risk management. The South African experience from using rapid, cost-effective point of care (PoC) DNA assays to speed up the process of genetic testing and decrease cost demonstrated the feasibility of this approach in a pilot study (Mampunye et al., 2021) POC *BRCA1/2* testing combined with genetic counselling to inform patients about the limitations and benefits of mutation-specific PoC tests versus comprehensive sequencing methods will increase access to genomic medicine, given the clinical dilemma caused by a relatively high frequency of variants of uncertain clinical significance (VUS) uncovered by gene panel testing and whole exome/genome sequencing (WES/WGS). PSGT is used to identify the target group most likely to benefit from germline DNA testing and/or tumour gene profiling and facilitate clinical interpretation of genetic results to add value in combination and beyond pathology test results. Incorporating WES/WGS as the research component of PSGT provides clinicians with the option to request additional genomic information pertinent to the patient (Policy Paper: A Framework for the Implementation of Genomic Medicine for Public Health in Africa, 2022; Kotze, 2016; van der Merwe and Kotze, 2018).

There have been a few studies in Egypt aiming for a similar approach by associating serum 25 (OH) Vitamin D level and breast cancer prognosis (El Shorbagy et al., 2017; Ismail et al., 2018; Abdel-Razeq, 2019). However, we need to shift from efforts done in a research context to actual implementation in a healthcare setting. A pipeline moving from primary research findings to validation of results in a wide population segment to translation into clinical application is required. Application in routine diagnostic, prognostic and therapeutic care should be followed by population level implementation.

The Reference Genome Project for Egyptians and Ancient Egyptians

In 1990, the first Human Genome Project worldwide was started aiming to decode the sequence of the human genome (Watson, 1990; Cantor, 1990). After 8 years, the Icelandic deCode Project was initiated to link genomic data with other medical and non-medical data (Palsson and Rabinow, 1999). In 2010, the UK10K project was announced for where it established a collaboration among many United Kingdom public and private entities, to identify genetic causes of rare diseases (The International Genome Sample Resource, 2022). In 2015, the United States and China started their large precision medicine initiatives based on genomic medicine (Cyranoski, 2016; Stark et al., 2019; All of

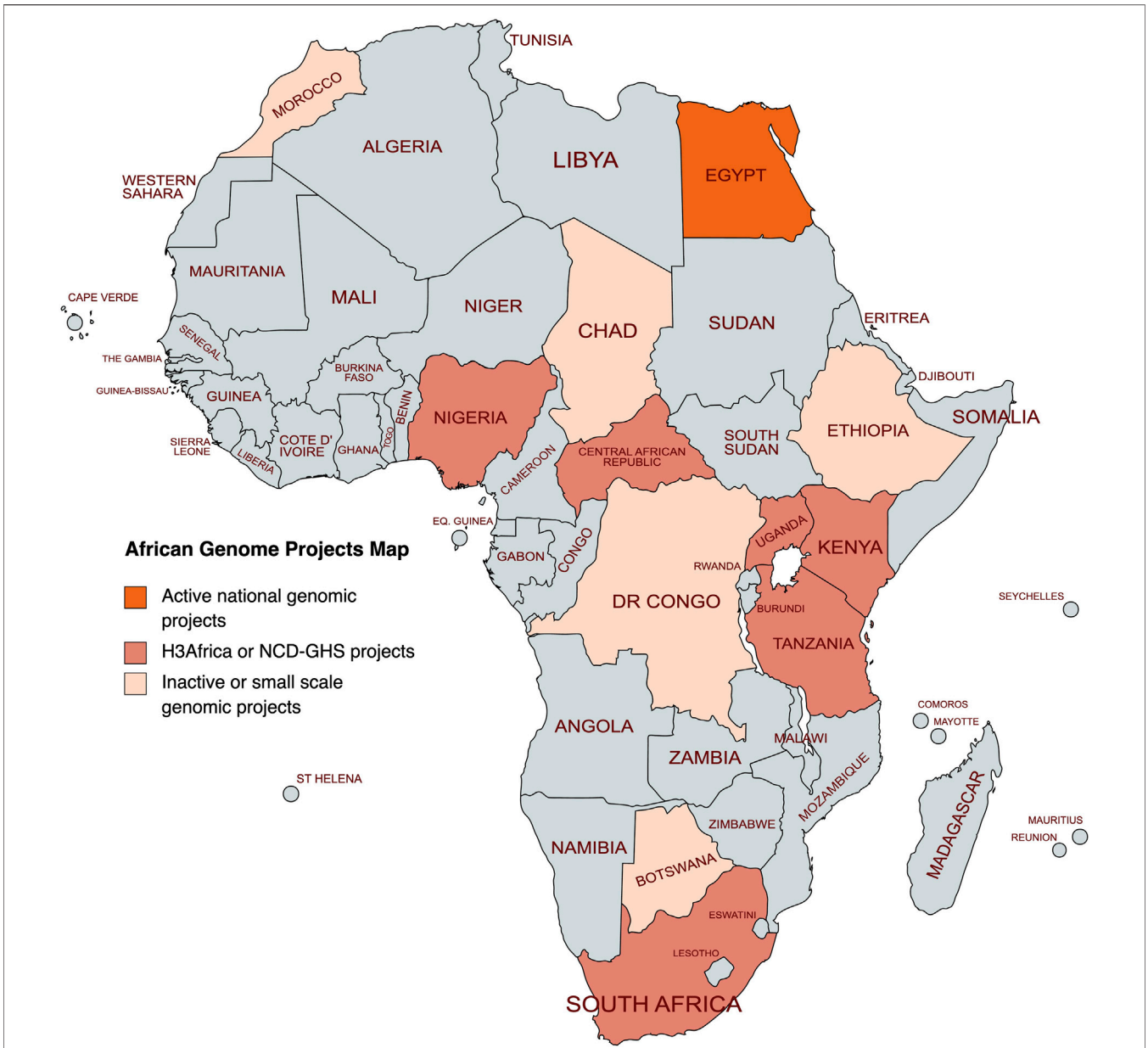


FIGURE 4 | African genome projects map. Created with mapchart.net.

Us Research Program, 2022; Globaltimes.cn., 2022). In 2018, Europe announced an initiative entitled "Towards access to at least 1 million sequenced genomes in the EU by 2022" aiming to share genomic information and best practices among the European union countries (Saunders et al., 2019; 1+ Million Genomes, 2022). All these major genomic initiatives were far away from Africa but owing to the importance of decoding the African human genome, efforts and collaborations have started to achieve this purpose since 2010 via establishing the H3Africa international project. According to our literature search through the Egyptian presidential online resources and the different African genomic initiatives websites, a total of 13 African countries were found to have either genomic projects and/or genomic databases among the 54 countries constituting the African continent (Kovanda et al., 2021), of which only the Reference Genome Project for Egyptians and Ancient Egyptians is currently active (Kovanda et al., 2021; Ain Shams University Official Website, 2022) (**Figure 4**). The remaining 12 projects were either not currently active or were part of the H3Africa international project (Adoga et al., 2014; Ramsay, 2015) or the Nigerian 100K Non-Communicable Diseases Genetic Heritage Study (NCD-GHS) (Genomeweb.com., 2022) and hence cannot be classified as a national project. In Nigeria, beside the NCD-GHS study, a venture called 54Gene, named to reflect the 54 countries in Africa, has started its activity aiming to build the continent's largest

FIGURE 5 | SWOT analysis of current genetic/genomics services in Egypt (MOH&P: Ministry of Health and Population; NRC: National Research Center; SCU: Suez Canal University).

biobank. The first step in this venture was a study, launched in 2020, to sequence and analyze the genomes of 100,000 Nigerians with a promise of bringing precision medicine to Nigeria (Maxmen, 2020; Genomeweb.com., 2022).

In Egypt, the current National Reference Genome Project for Egyptians and Ancient Egyptians was preceded by a smaller scale project "EgyptRef" in 2018 with the aim to establish a reference genome for Egyptian and North African populations to complement the Genome Reference Consortium human genome (GRCh) (Egyptian-genome.org., 2022). The most important details and comparisons related to the previous and the current Egyptian genome projects are clarified in **Table 1** and compared also to the major African genomics endeavors in Africa which is H3Africa International genome Project. Currently, there are high expectations on the National Reference Genome Project for Egyptians and Ancient Egyptians in terms of entering the era of precision medicine to Egypt, improving cost-effective diagnostics, with more targeted prevention and treatment for Egyptians.

As one of the main goals of this huge national project is to upgrade the infrastructure of genetics and genomics in Egypt, the project acquired a number of NGS platforms and formed a network with previously existing platforms in different universities or research centers which will enable the project to identify enormous number of causal mutations in a short period of time and at relatively low cost. The availability of sequence variants from the pilot EgyptRef project started in 2018 will also be made use of when an available reference genome is required. In this way, NGS based genetic screening will be possible for the identification and mapping of causal mutations among the Egyptian population. However, data obtained from NGS is highly complicated and requires more sophisticated data processing (Pereira et al., 2020). NGS based genetic approaches are vastly reliant on compatible bioinformatics tools and pipelines to obtain the final outputs and desired information. Numerous bioinformatics tools and pipelines have been developed to perform the different functions. Utilization of these tools/pipelines depends on the NGS methodology, type of experimental materials, and the genome of organisms (Sahu et al., 2020). Many Egyptian bioinformaticians are sharing their experience in this national project and will help in all analysis phases.

Advances in sequencing technologies provided the world with a great tool to decode the human genome efficiently, at much lower costs than before and at a significantly faster rate. Having said that, still the reference genome lacks population specific variations. A great milestone in advancing health care services was achieved by the establishment of the "Egyptian reference human genome" through sequencing Egyptian individuals' genome, a project conducted by Mansoura University in collaboration with Lübeck university and the German science foundation, DFG, excellence program (EXC 306) and DAAD, www.egyptian-genome.org. The results of the sequencing showed Egyptian specific gene variants, paving the way towards precision medicine implementation and adding greater value for genetic research as well (Wohlers et al., 2020). The next step that is underway is the implementation of "Reference Genome Project for Egyptians and Ancient Egyptians", which should aim at

further characterization of the genetic variants in Egyptians and evaluation of their impact and relevance to diseases.

On the 31st of January 2022, a detailed disclosure of the Egyptian genome project was revealed at an event "The Reference Genome Project for the Egyptians and the Ancient Egyptian" held at the Dubai Expo 2020, United Arab Emirates. The event was organized by the Egyptian Ministry of Higher Education and Scientific Research, represented by the Academy of Scientific Research and Technology ASRT, in partnership with the Sheikh Zayed Center for Genetic Research, affiliated with the Emirates Society for Genetic Diseases. It was announced that according to the executive plan, the implementation of the project will end by the end of 2025, with an initial cost of two billion Egyptian pounds (128 million dollars)." With the first phase starting with a sample of 20,000 Egyptians to identify the most common diseases, and then increasing to 100,000 over 5 years for the first phase, until the project covers all of Egypt. The ASRT began implementing the recommendations of the specific councils in the academy to launch the reference genome program for the Egyptians, in cooperation with the Center for Research and Regenerative Medicine Center (ECRRM) of the Armed Forces being the main research body, and a number of scientific and executive bodies in the Egyptian state are participating in the project as well, represented by the Ministries of Defense, Health and Communication, more than 15 universities, research centers and civil society institutions (Teller Report, 2022; Egypt Center for Research and Regenerative Medicine, 2022; Emirati Foundation, 2022; Archyde, 2022; Newsbeezer, 2022).

SWOT Analysis of Current Genetic/ Genomics Services in Egypt

The presented SWOT analysis (**Figure 5**) is based on the results of the Erasmus project report entitled "Genetic Testing in Emerging Economies (GenTEE)" in the period from 2010 to 2012 (Genetic Testing in Emerging Economies (GenTEE) Project, 2022). The GenTEE partners were from European and non-European partner countries, where Egypt was among the partners in this project. The GenTEE report relied upon implementing a wide range survey in each partner country based on a common method/framework for data ascertainment, thus allowing examination and comparison of the following: Service development in relation to the existing health care systems, available genetic services resources, national genetic testing services within the available health policies, and factors that affect the state of genetic service delivery such as the demographic, socio-economic and legal factors. This survey addressed nine dimensions using a set of indicators pre-defined by the project's consortium which were: demography and health indicators, health expenditure and financing, indicators of congenital/genetic "disorder burden", availability of genetic services, access to genetic services, state of genetic services, research priorities in genetics/genomics, patient organizations and public education in genetics, and future outlook for service development in each country. The survey targeted specific outcomes specifically genetic testing services within the health care settings, which were: PND and PGD, newborn screening, carrier screening, diagnostic testing for congenital and genetic disorders, including testing for common disorders with a major single gene subgroup, pharmacogenetic testing, and genetic susceptibility testing (e.g., for infectious diseases). For data ascertainment purposes, the data collection of the survey was based on published data including grey-literature, accessible unpublished reports/data, and expert opinion. Before a country report was accepted by the GenTEE consortium it had to be submitted to an external expert review for validation.

From this GenTEE report, it was obvious that the urban, upper-middle classes in most developing countries including Egypt are the main beneficiaries of the genetic services. This is mostly attributed to the economic class of this category, where they can afford these services in the private sector. Also, the scarcity of health care professionals specialized in genetics is another important problem facing the genetics services in Egypt.

In addition, we have a huge mismatch in Egypt between the non-availability of enough services in the genetic/genomic field and in developing research capacities for improving the Egyptian patients care. Therefore, to achieve a rapid development of genetic/genomic Egyptian services, the current service infrastructure needs to be upgraded in Egypt. This step will support the successful driving of genetic/genomic facilities and academic research into accredited reliable services and will thereby be reflected on the overall population health outcomes.

Multidisciplinary Collaboration for Advancing Genomics in Egypt: Recommendations and Future Directions

A multi-disciplinary approach in the field of genomics is currently emerging world-wide. This approach for ensuring engagement, informed decision-making and supporting clinicians has been made possible by rapid advances in DNA sequencing, bioinformatics and digital applications. All these evolving technologies along with cornerstone directions and initiatives represented in the "Egyptian genome" and the "comprehensive Egyptian health insurance" projects recently adopted presidentially in Egypt paves the way for successful genomic medicine integration in healthcare system in Egypt. Multidisciplinary collaboration is the mainstay of genomic medicine integration and we recommend it to be implemented under the umbrella of governance body (Adebamowo et al., 2018) in a stepwise approach (**Figure 6**). A task force of experts in the field of genomics is recommended to be assigned to identify and consult stakeholders (Tindana et al., 2019). This pivotal task is a principal key for acceptance, consent and collaboration, ensuring successful implementation. Consultation with stakeholders could take place in different forms such as workshops, interviews and/or surveys to form a multidisciplinary governance body (Tindana et al., 2019). Inclusion of all stakeholders in the governance body is advisable for lean, resistance free implementation. Stakeholders could include the Ministry of Health and Population which is a main primary care provider, Ministry of Higher education and Research, a main training and research entity, Egypt Center for Research and regenerative Medicine (ECRRM), a main contributor in the Reference Genome Project for Egyptians and Ancient Egyptians, as well as investors, religious and judiciary entities, non-profit organizations and civil societies.

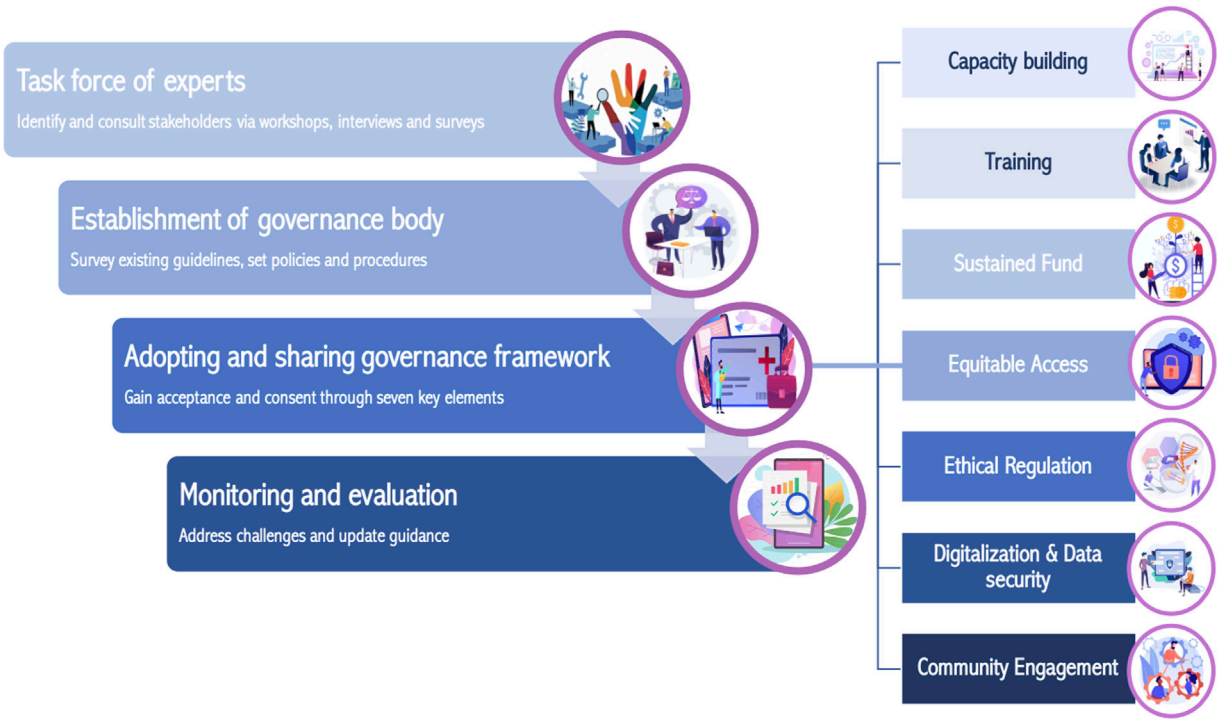


FIGURE 6 | Suggested stepwise approach for genomic medicine governance establishment.

The governing body should set policies and guidelines for every sector involved in genomic medicine establishment and integration in healthcare as well as adopting genomic surveillance strategy for pathogens with pandemic or epidemic potentials. Sharing guidelines with international and national communities (Tindana et al., 2019) for broader acceptance and refinement is advisable and could be established through conferences and meetings. The governing body should address the needs of the Egyptian community, respecting the cultural and religious beliefs, avoiding stigmas and harm, addressing fears and establishing genuine Egyptian genomic medicine services. It should practice as well good governance principles; accountability, transparency, responsiveness, equitability, inclusiveness, consensus oriented, effectiveness, efficiency, follow the rule of law and participatory (Governance for Sustainable Human Development, 1997). We recommend that the governance should adopt a framework (Policy Paper: A Framework for the Implementation of Genomic Medicine for Public Health in Africa, 2022; O'Doherty et al., 2021) that fulfills the above-mentioned principles as well as be inclusive of key elements for genomic medicine establishment. We suggest the following elements to be considered in the framework.

Key Elements to be Considered in the Governance Framework

Capacity building (Yakubu et al., 2018) is the main pillar of genomic medicine establishment and should be planned for, governed and monitored. The governance should ensure the presence of laboratories with latest technologies as well as fully functional **biorepository** that will house all the biological specimens that will be collected. This infrastructure will ensure that biological specimens remain in the country and will encourage large-scale collaboration between Egyptian scientists while fostering collaborations between Egyptian and international scientists as well.

Infrastructure and technology are particularly important in a high paced, rapidly evolving field such as genomics, yet alone never ensure needs are met to population. Complex technology and huge amount of data generated mandates high level of expertise in several fields such as; genetic counseling, bioinformatics and technical operation of various instruments. Investment in individuals through **training** (Policy Paper: A Framework for the Implementation of Genomic Medicine for Public Health in Africa, 2022) is the mainstay for sustainable and high standard quality of service. It could be achieved through scholarships for post graduate students, training programs and workshops for physicians, data scientists, scientists and genetic counselors in collaboration with experts in the field from local and international universities. Training program should be in multiple disciplines, including genomics (high-throughput technologies), genetics, epidemiology, bioinformatics, statistical genetics, and Ethical Legal and Social Issues (ELSI).

Integration of introductory genomic medicine course in undergraduate medical student curriculum as well as establishment of genomic medicine and bioinformatics post

graduate master's degrees is important to prepare good calibers with the needed expertise to deliver the best of service. However, encouraging young physicians and scientists to hold relevant degrees could be challenging, especially if there is not a rewarding salary, opportunities for hiring in the system are scarce as well as few mentorships available as in the case of African countries (Policy Paper: A Framework for the Implementation of Genomic Medicine for Public Health in Africa, 2022), such concerns should be identified and addressed.

The governance should **encourage research** in genomics and **raise funds** allocating them towards this field and promote collaboration among different entities and investors under its auspices, as well as setting financial plans for sustainable funding (Cazabon et al., 2021) of genomic medicine related services and research. Moreover, it should facilitate and encourage sharing expertise with universities and organizations around the globe.

The governance should support and facilitate collaborative efforts to integrate genomic medicine service in the healthcare system including the public and private sectors, monitor performance and identify challenges. Ensure thorough training to all involved personnel as well as public genomic medicine literacy for efficient and effective deployment of service.

Although genomic medicine is present in the medical field for years now, the service remains highly expensive especially for low-income societies, to ensure **equitable access**, insurance system should cover expenses. Reimbursement of service should be for both public and private sector with cost predefined and controlled by the governance. Also, the type of service covered by insurance, whether testing, treatment or counseling according to preset tiers defined by the governance, are equally applied on both public and private sectors. Having insurance coverage to private sectors as well shall expand the availability and accessibility of genomic medicine, decreasing the burden on public sector and better allocating ministries' budget to other demanding sources. The new comprehensive insurance system to be established in Egypt is a fundamental step towards advancing genomic medicine integration in healthcare system.

In the digitalized era and with the ever-growing populations, paperwork became more and more difficult to catch advancements in several fields. For easier, rapid and standardized archiving, simple and fast access to data and reaching out to people in underprivileged areas, **digitalization of service** is indispensable.

Reaching out for patients remotely is widely spread nowadays through health applications that offer simple information, appointment booking, test results and immunity passports as in Covid-19 pandemic nowadays in certain countries. Applications are emerging to help patients receive their genomic test results with greater understanding of the nature of test, result, actions to be taken, impact on health and family, giving them choice of what information to receive or not. Genomics ADvISER (Bombard et al., 2018) and Genetics Navigator are examples for digital applications offering guidance in incidental findings and entire genomic testing respectively (Bombard and Hayeems, 2020). Although fast, readily accessible, updated and can notify patient with updates, alone it is not enough to deliver all information, consult and conduct result. In-person meeting with genetic counselor would be essential in Egypt to ensure good understanding and provide emotional support.

Digitalization of service either through hospital, clinic information systems, medical records, applications and websites, generate enormous amount of **data that should be secure and confidential** (Policy Paper: A Framework for the Implementation of Genomic Medicine for Public Health in Africa, 2022) under law and its abidance must be monitored. Legislation and enforcement of this law is the responsibility of the governance in collaboration with the judiciary and legal entities.

For community acceptance, service must be assured it is within the **religious and ethical frame** (Policy Paper: A Framework for the Implementation of Genomic Medicine for Public Health in Africa, 2022) of the community. Formulation of tiers of contests (broad, tiered, conditioned, restricted) to the usage and sharing of data and specimens for research purposes gives more confidence and acceptance in genomic medicine services and research especially if the archiving is within the country. Governance should collaborate with religious entities as well to encourage community to endorse genomic medicine and assure adherence to values.

Encouraging and supporting the establishment of civil societies promoting **community engagement** (Cazabon et al., 2021) in decision making, addressing their needs and resolving challenges. Such efforts would promote community acceptance bringing consensus and ease implementation.

Monitoring and evaluation is indispensable for a project to be successful and sustainable, the governance should continuously monitor and evaluate the establishment of genomic medicine to identify challenges and update guidance accordingly.

CONCLUSION

Genomic medicine is a new emerging field in clinical care, having large impact on communicable diseases, as seen in the ongoing pandemic, as well as pharmacological treatments, rare and undiagnosed diseases and oncology. There is enormous amount of genomic data that is generated continuously from around the world. However, data from other countries could lack special features that are specific and relatable to Egypt and African countries. The recent launch of Reference Genome project for Egyptians and Ancient Egyptians represents a new era of medicine in Egypt and is a corner stone in genomic medicine establishment. In this review we presented the needs of our community as regards to genomic medicine in several fields, pointing out the challenges and barriers towards advancing this type of clinical care into our healthcare system. We also suggested establishment of a governance body which would be essential in regulating the genomic medicine services in Egypt. The suggested governance body would improve the existing

regulatory process, develop and periodically review guidelines and policies as well as collaborating efforts among stakeholders. Observing the thirst for genomic medicine in Egypt, it is hopeful that a revolutionary era of healthcare service would be seen in the country soon.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

REFERENCES

1+ Million Genomes (2022). Shaping Europe's Digital Future. [Cited 2022 Feb 5]. Available from: https://digital-strategy.ec.europa.eu/en/policies/1-million-genomes.

Abdalla, E. M., Mostowska, A., Jagodziński, P. P., Dwidar, K., and Ismail, S. R. (2014). A Novel WNT10A Mutation Causes Non-Syndromic Hypodontia in an Egyptian Family. *Arch. Oral Biol.* 59 (7), 722–728. doi:10.1016/j.archoralbio.2014.04.004

Abdalla, E. M., Rohrbach, M., Bürer, C., Kraenzlin, M., El-Tayeby, H., Elbelbesy, M. F., et al. (2015). Kyphoscoliotic Type of Ehlers-Danlos Syndrome (EDS VIA) in Six Egyptian Patients Presenting with a Homogeneous Clinical Phenotype. *Eur. J. Pediatr.* 174 (1), 105–112. doi:10.1007/s00431-014-2429-9

Abdalla, E., Ravenscroft, G., Zayed, L., Beecroft, S. J., and Laing, N. G. (2017). Lethal Multiple Pterygium Syndrome: A Severe Phenotype Associated with a Novel Mutation in the Nebulin Gene. *Neuromuscul. Disord.* 27 (6), 537–541. doi:10.1016/j.nmd.2017.01.013

Abdel-Mohsen, M. A., Ahmed, O. A., and El-Kerm, Y. M. (2016). BRCA1 Gene Mutations and Influence of Chemotherapy on Autophagy and Apoptotic Mechanisms in Egyptian Breast Cancer Patients. *Asian Pac. J. Cancer Prev.* 17 (3), 1285–1292. doi:10.7314/apjcp.2016.17.3.1285

Abdel-Razek, W., Hassany, M., El-Sayed, M. H., El-Serafy, M., Doss, W., Esmat, G., et al. (2019). Hepatitis C Virus in Egypt: Interim Report from the World's Largest National Program. *Clin. Liver Dis. (Hoboken)* 14 (6), 203–206. doi:10.1002/cld.868

Abdel-Razeq, H. (2019). Prognostic Significance of Serum Vitamin D Levels in Egyptian Females with Breast Cancer. *Asian Pac. J. Cancer Prev.* 20 (4), 983. doi:10.31557/APJCP.2019.20.4.983

Adebamowo, S. N., Francis, V., Tambo, E., Diallo, S. H., Landouré, G., Nembaware, V., et al. (2018). Implementation of Genomics Research in Africa: Challenges and Recommendations. *Glob. Health Action.* 11 (1), 1419033. [Internet]. doi:10.1080/16549716.2017.1419033

Ades, F., Tryfonidis, K., and Zardavas, D. (2017). The Past and Future of Breast Cancer Treatment-From the Papyrus to Individualised Treatment Approaches. *ecancer* 11, 746. doi:10.3332/ecancer.2017.746

Adoga, M. P., Fatumo, S. A., and Agwale, S. M. (2014). H3Africa: A Tipping point for a Revolution in Bioinformatics, Genomics and Health Research in Africa. *Source Code Biol. Med.* 9 (1), 10. doi:10.1186/1751-0473-9-10

Afifi, H. H., El-Ruby, M. O., El-Bassyouni, H. T., Ismail, S. I., Aglan, M. S., El-Harouni, A. A., et al. (2010). The Most Encountered Groups of Genetic Disorders in Giza Governorate, Egypt. *Bratisl Lek Listy* 111 (2), 62–69.

Africa's Digital Solutions to Tackle Covid-19 (2020). Eib.org. [Cited 2022 Feb 5]. Available from: https://www.eib.org/attachments/country/africa_s_digital_solutions_to_tackle_covid_19_en.pdf.

Aglan, M., Amr, K., Ismail, S., Ashour, A., Otaify, G. A., Mehrez, M. A. I., et al. (2015). Clinical and Molecular Characterization of Seven Egyptian Families with Autosomal Recessive Robinow Syndrome: Identification of Four novel *ROR2* gene Mutations. *Am. J. Med. Genet.* 167 (12), 3054–3061. doi:10.1002/ajmg.a.37287

Aguib, Y., Allouba, M., Afify, A., Halawa, S., El-Khatib, M., Sous, M., et al. (2020). The Egyptian Collaborative Cardiac Genomics (ECCO-GEN) Project: Defining a Healthy Volunteer Cohort. *Npj Genom. Med.* 5 (1), 46. doi:10.1038/s41525-020-00153-w

Ain Shams University Official Website (2022). The President of the University Participates in the Second Meeting of the Board of Directors of the National Reference Genome Project for Egyptians and Ancient Egyptians. Edu.eg. 2022 [Cited 2022 Feb 5]. Available from: https://www.asu.edu.eg/3568/news/the-president-of-the-university-participates-in-the-second-meeting-of-the-board-of-directors-of-the-national-reference-genome-project-for-egyptians-and-ancient-egyptians.

Al-Youm, A-M. (2021). Egypt Reports First Three Omicron Cases. *Egypt Independent*. [Cited 2022 Feb 5]. Available from: https://egyptindependent.com/egypt-reports-first-three-omicron-cases/.

All of Us Research Program (2022). *National Institutes of Health (NIH) — All of Us*. [Cited 2022 Feb 5]. Available from: https://allofus.nih.gov/.

Alyamani, E. J., Marcus, S. A., Ramirez-Busby, S. M., Hansen, C., Rashid, J., El-Kholy, A., et al. (2019). Genomic Analysis of the Emergence of Drug-Resistant Strains of *Mycobacterium tuberculosis* in the Middle East. *Sci. Rep.* 9 (1), 4474. doi:10.1038/s41598-019-41162-9

Archyde (2022). *The Egyptian Genome Project Is an Important Step in Scientific Research*. [Internet]. Archyde. 2022 [Cited 2022 Feb 13]. Available from: https://www.archyde.com/nahyan-bin-mubarak-the-egyptian-genomeproject-is-an-important-step-in-scientific-research/.

Armstrong, G. L., MacCannell, D. R., Taylor, J., Carleton, H. A., Neuhaus, E. B., Bradbury, R. S., et al. (2019). Pathogen Genomics in Public Health. *N. Engl. J. Med.* 381 (26), 2569–2580. doi:10.1056/NEJMsr1813907

Baloch, Z., Ma, Z., Ji, Y., Ghanbari, M., Pan, Q., and Aljabr, W. (2020). Unique Challenges to Control the Spread of COVID-19 in the Middle East. *J. Infect. Public Health* 13 (9), 1247–1250. doi:10.1016/j.jiph.2020.06.034

Bayomy Helal, H. E., Yuonis, A., Shaker, R. H. M., and Elawady, M. A. (2018). Prevalence of HCV Infection in Household Contacts of Chronic Liver Diseases Cases in Egypt. *J. Environ. Public Health* 2018, 1–9. [Internet]. doi:10.1155/2018/2153537

Bielas, S. L., Silhavy, J. L., Brancati, F., Kisseleva, M. V., Al-Gazali, L., Sztriha, L., et al. (2009). Mutations in INPP5E, Encoding Inositol Polyphosphate-5-Phosphatase E, Link Phosphatidyl Inositol Signaling to the Ciliopathies. *Nat. Genet.* 41 (9), 1032–1036. doi:10.1038/ng.423

Bittles, A. (2001). Consanguinity and its Relevance to Clinical Genetics. *Clin. Genet.* 60 (2), 89–98. doi:10.1034/j.1399-0004.2001.600201.x

Bombard, Y., Clausen, M., Mighton, C., Carlsson, L., Casalino, S., Glogowski, E., et al. (2018). The Genomics ADvISER: Development and Usability Testing of a Decision Aid for the Selection of Incidental Sequencing Results. *Eur. J. Hum. Genet.* 26 (7), 984–995. doi:10.1038/s41431-018-0144-0

Bombard, Y., and Hayeems, R. Z. (2020). How Digital Tools Can Advance Quality and Equity in Genomic Medicine. *Nat. Rev. Genet.* 21 (9), 505–506. doi:10.1038/s41576-020-0260-x

Butrous, G., Maron, B., and Yacoub, M. (2020). The Lamp of Medicine of Ancient Egypt Is Still Burning. *Gcsp* 2020 (1), e202016. doi:10.21542/gcsp.2020.16

Cancer Today (2022). Iarc.fr. [Cited 2022 Feb 4]. Available from: https://gco.iarc.fr/today.

Cantor, C. R. (1990). Orchestrating the Human Genome Project. *Science* 248 (4951), 49–51. doi:10.1126/science.2181666

Capmas, E. (2022). The Central Agency for Public Mobilisation and Statistics. *Annu. Marriage Divorce Stat.* Gov.eg. [Cited 2022 Feb 4]. Available from: https://censusinfo.capmas.gov.eg/Metadata-en-v4.2/index.php/catalog/279/accesspolicy.

Cazabon, D., Farrell, M., Gupta, R., Joseph, L., Pathni, A. K., Sahoo, S., et al. (2021). A Simple Six-Step Guide to National-Scale Hypertension Control Program Implementation. *J. Hum. Hypertens.* doi:10.1038/s41371-021-00612-6

Centers for Disease Control and Prevention (2012). Progress Toward Prevention and Control of Hepatitis C Virus infection–Egypt, 2001-2012. *MMWR Morb. Mortal Wkly Rep.* 61 (29), 545–549.

Chatron, N., Becker, F., Morsy, H., Schmidts, M., Hardies, K., Tuysuz, B., et al. (2020). Bi-Allelic GAD1 Variants Cause a Neonatal Onset Syndromic Developmental and Epileptic Encephalopathy. *Brain [Internet]* 143 (5), 1447–1461. doi:10.1093/brain/awaa085

Cross-Sectional Survey of Genetics and Genomics Training in Africa (2022). H3abionet.org. [Cited 2022 Mar 19]. Available from: http://training.h3abionet.org/AGMC_2016/wp-content/uploads/2017/01/TrainingSurveyAfrica-Upload.pdf

Cyranoski, D. (2016). China Embraces Precision Medicine on a Massive Scale. *Nature* 529 (7584), 9–10. doi:10.1038/529009a

Darwish, R. K., Haghighi, A., Seliem, Z. S., El-Saiedi, S. A., Radwan, N. H., El-Gayar, D. F., et al. (2020). Genetic Study of Pediatric Hypertrophic Cardiomyopathy in Egypt. *Cardiol. Young* 30 (12), 1910–1916. doi:10.1017/S1047951120003157

Privacylaws.com (2022). Data Protection Law Personal Data Protection Act by Promulgating. [Cited 2022 Feb 5]. Available from: https://privacylaws.com/media/3263/egypt-data-protection-law-151-of-2020.pdf

Donato, L., Abdalla, E. M., Scimone, C., Alibrandi, S., Rinaldi, C., Nabil, K. M., et al. (2021). Impairments of Photoreceptor Outer Segments Renewal and Phototransduction Due to a Peripherin Rare Haplotype Variant: Insights from Molecular Modeling. *Ijms* 22 (7), 3484. doi:10.3390/ijms22073484

Egypt Center for Research and Regenerative Medicine (2022). Ecrrm.ac.eg. [Cited 2022 Feb 13]. Available from: http://www.ecrrm.ac.eg/Default.aspx.

Egypt Independent (2020). *Sisi Passes Law Regulating Clinical Medical Research in Egypt*. [Internet]. Egypt Independent. 2020 [Cited 2022 Feb 5]. Available from: https://egyptindependent.com/sisi-passes-law-regulating-clinical-medical-research-in-egypt/.

Egypt Plan of Action for the Prevention, Care and Treatment of Viral Hepatitis 2014-2018 (2014). Globalhep.org. [Internet]. [Cited 2022 Feb 5]. Available from: https://www.globalhep.org/evidence-base/egypt-plan-action-prevention-care-treatment-viral-hepatitis-2014-2018.

Egypt (2022). WHO Coronavirus Disease (COVID-19) Dashboard with Vaccination Data. *Who.int*. [Cited 2022 Feb 5]. Available from: https://covid19.who.int/region/emro/country/eg.

Egyptian-genome.org. (2022). Egyptian Genome. [Cited 2022 Feb 5]. Available from: https://www.egyptian-genome.org.

Egyptian Ministry of Health and Population, World Health Organization (2022). *Egypt Multisectoral Action Plan for Noncommunicable Diseases Prevention and Control*. Who.int. [Cited 2022 Feb 4]. Available from: https://www.who.int/ncds/governance/policies/Egypt-NCD-MAP-2018-2022.pdf.

Eid, O. M., El Ghoroury, E. A., Eid, M. M., Mahrous, R. M., Abdelhamid, M. I., Aboafya, Z. I., et al. (2017). Evaluation of BRCA1 Large Genomic Rearrangements in Group of Egyptian Female Breast Cancer Patients Using MLPA. *Gulf J. Oncolog.* 1 (25), 64–69.

El Far, M. A., Atwa, M. A., Yahya, R. S., and El Basuni, M. A. (2006). Evaluation of Serum Levels of P53 in Hepatocellular Carcinoma in Egypt. *Clin. Chem. Lab. Med. [Internet]* 44 (5), 653–656. doi:10.1515/CCLM.2006.091

El Shorbagy, S., Haggag, R., Ebian, H. F., Labi, H. A., and Harb, O. A. (2017). Prognostic Impact of 25-Hydroxyvitamin D Levels in Egyptian Patients with Breast Cancer. *J. Cancer Sci. Ther.* 09 (06). doi:10.4172/1948-5956.1000466

el-A Helal, T., Khalifa, A., and Kamel, A. S. (2000). Immunohistochemical Expression of P53 and C-erbB2 Proteins in Breast Cancer in Egypt. *Anticancer Res.* 20 (3B), 2145–2150.

El-Abd, N. E., Fawzy, N. A., El-Sheikh, S. M., and Soliman, M. E. (2015). Circulating miRNA-122, miRNA-199a, and miRNA-16 as Biomarkers for Early Detection of Hepatocellular Carcinoma in Egyptian Patients with Chronic Hepatitis C Virus Infection. *Mol. Diagn. Ther.* 19 (4), 213–220. doi:10.1007/s40291-015-0148-1

El-Ghannam, D. M., Arafa, M., and Badrawy, T. (2011). Mutations of P53 Gene in Breast Cancer in the Egyptian Province of Dakahliya. *J. Oncol. Pharm. Pract.* 17 (2), 119–124. doi:10.1177/1078155209356130

El-Ghitany, E. M., and Farghaly, A. G. (2019). Geospatial Epidemiology of Hepatitis C Infection in Egypt 2017 by Governorate. *Heliyon* 5 (8), e02249. doi:10.1016/j.heliyon.2019.e02249

El-Hattab, A. W., Shaheen, R., Hertecant, J., Galadari, H. I., Albaqawi, B. S., Nabil, A., et al. (2016). On the Phenotypic Spectrum of Serine Biosynthesis Defects. *J. Inherit. Metab. Dis.* 39 (3), 373–381. doi:10.1007/s10545-016-9921-5

El-Kafrawy, S. A., Abdel-Hamid, M., El-Daly, M., Nada, O., Ismail, A., Ezzat, S., et al. (2005). P53 Mutations in Hepatocellular Carcinoma Patients in Egypt. *Int. J. Hyg. Environ. Health* 208 (4), 263–270. doi:10.1016/j.ijheh.2005.02.002

El-Kamah, G. Y., Mohamed, A. M., Gad, Y. Z., Abdelhak, S., Hennig, B. J., Ramesar, R. S., et al. (2020). Developing a Road Map to Spread Genomic Knowledge in Africa: 10th Conference of the African Society of Human Genetics, Cairo, Egypt. *Am. J. Trop. Med. Hyg. [Internet]* 102 (4), 719–723. doi:10.4269/ajtmh.19-0408

Elbahrawy, A., Ibrahim, M. K., Eliwa, A., Alboraie, M., Madian, A., and Aly, H. H. (2021). Current Situation of Viral Hepatitis in Egypt. *Microbiol. Immunol.* 65 (9), 352–372. doi:10.1111/1348-0421.12916

Elgawhary, S., Elbaradie Sahar, M. Y., Rashad, W. M., Mosaad, M., Abdalla, M. A. H., Ezzat, G., et al. (2008). PRENATAL DIAGNOSIS of β-THALASSEMIA IN EGYPT: Implementing Accurate High-Tech Methods Did Not Reflect Much on the Outcome. *Pediatr. Hematol. Oncol.* 25 (6), 541–548. doi:10.1080/08880010802313509

Emirati Foundation (2022). The Egyptian Genome Is a Promising Project for the Entire Arab Region - Al Shorouk Gate. [Internet]. Shorouknews.com. [Cited 2022 Feb 13]. Available from: https://www.shorouknews.com/mobile/news/view.aspx?cdate=01022022&id=8cf48cf4-c91a-4e12-8173-275ede6ddf85.

Essawi, M. L., Fateen, E. M., Atia, H. A., Eissa, N. R., Aboul-Ezz, E. H., Ibrahim, M. M., et al. (2021). Quaternary Diagnostics Scheme for Mucolipidosis II and Detection of Novel Mutation in GNPTAB Gene. *J. Genet. Eng. Biotechnol.* 19 (1), 111. doi:10.1186/s43141-021-00204-4

Fassad, M. R., Patel, M. P., Shoemark, A., Cullup, T., Hayward, J., Dixon, M., et al. (2020). Clinical Utility of NGS Diagnosis and Disease Stratification in a Multiethnic Primary Ciliary Dyskinesia Cohort. *J. Med. Genet.* 57 (5), 322–330. doi:10.1136/jmedgenet-2019-106501

Fassad, M. R., Shoemark, A., Legendre, M., Hirst, R. A., Koll, F., le Borgne, P., et al. (2018). Mutations in Outer Dynein Arm Heavy Chain DNAH9 Cause Motile Cilia Defects and Situs Inversus. *Am. J. Hum. Genet.* 103 (6), 984–994. doi:10.1016/j.ajhg.2018.10.016

Fassad, M. R., Shoman, W. I., Morsy, H., Patel, M. P., Radwan, N., Jenkins, L., et al. (2020). Clinical and Genetic Spectrum in 33 Egyptian Families with Suspected Primary Ciliary Dyskinesia. *Clin. Genet.* 97 (3), 509–515. doi:10.1111/cge.13661

Furuse, Y. (2021). Genomic Sequencing Effort for SARS-CoV-2 by Country During the Pandemic. *Int. J. Infect. Dis.* 103, 305–307. doi:10.1016/j.ijid.2020.12.034

Genetic Testing in Emerging Economies (GenTEE) Project (2022). Eshg.org. [Cited 2022 Feb 5]. Available from: https://www.eshg.org/fileadmin/www.eshg.org/NHGS2012/4_Nippert__presentation_GenTEE_project.pdf.

Giandhari, J., Pillay, S., Wilkinson, E., Tegally, H., Sinayskiy, I., Schuld, M., et al. (2021). Early Transmission of SARS-CoV-2 in South Africa: An Epidemiological and Phylogenetic Report. *Int. J. Infect. Dis.* 103, 234–241. doi:10.1016/j.ijid.2020.11.128

Gov.eg (2022). The Actual Start of the Comprehensive Health-Insurance Law in Egypt. [Cited 2022 Feb 5]. Available from: https://www.sis.gov.eg/Story/211365/The actual start of the comprehensive health-insurance law in Egypt?lang=ar.

Governance for Sustainable Human Development (1997). *An Integrated Paper on the Highlights of Four Regional Consultation Workshops on Governance for Sustainable Human Development*. Makati, Philippines: United Nations Development Programme. [Cited 2022 Mar 19]. Available from: https://digitallibrary.un.org/record/3831662?ln=en.

Hafez, M. M., Hassan, Z. K., Zekri, A. R. N., Gaber, A. A., Rejaie, S. S. A., Sayed-Ahmed, M. M., et al. (2012). MicroRNAs and Metastasis-Related Gene Expression in Egyptian Breast Cancer Patients. *Asian Pac. J. Cancer Prev.* 13 (2), 591–598. doi:10.7314/apjcp.2012.13.2.591

Hagag, E., Shwaireb, M., Coffa, J., and El Wakil, A. (2013). Screening for BRCA1 Large Genomic Rearrangements in Female Egyptian Hereditary Breast Cancer Patients. *East. Mediterr. Health J.* 19 (03), 255–262. doi:10.26719/2013.19.3.255

Hamamy, H., and Alwan, A. (1994). Hereditary Disorders in the Eastern Mediterranean Region. *Bull. World Health Organ.* 72 (1), 145–154.

Handley, M. T., Morris-Rosendahl, D. J., Brown, S., Macdonald, F., Hardy, C., Bem, D., et al. (2013). Mutation Spectrum in *RAB3GAP1*, *RAB3GAP2*, and *RAB18* and Genotype-Phenotype Correlations in Warburg Micro Syndrome and Martsolf Syndrome. *Hum. Mutat.* 34 (5), 686–696. doi:10.1002/humu.22296

Hawass, Z., Gad, Y. Z., Ismail, S., Khairat, R., Fathalla, D., Hasan, N., et al. (2010). Ancestry and Pathology in King Tutankhamun's Family. *Jama* 303 (7), 638–647. doi:10.1001/jama.2010.121

Horn, D., Fernández-Núñez, E., Gomez-Carmona, R., Rivera-Barahona, A., Nevado, J., Schwartzmann, S., et al. (2021). Biallelic Truncating Variants in MAPKAPK5 Cause a New Developmental Disorder Involving Neurological, Cardiac, and Facial Anomalies Combined with Synpolydactyly. *Genet. Med.* 23 (4), 679–688. doi:10.1038/s41436-020-01052-2

Hussein, M. R., and Hassan, H. I. (2004). Alterations of P53, Bcl-2, and hMSH2 Protein Expression in the normal Breast, Benign Proliferative Breast Disease, *In Situ* and Infiltrating Ductal Breast Carcinomas in the Upper Egypt. *Cancer Biol. Ther.* 3 (10), 983–988. doi:10.4161/cbt.3.10.1136

Hussien, Y. M., Gharib, A. F., Awad, H. A., Karam, R. A., and Elsawy, W. H. (2012). Impact of DNA Repair Genes Polymorphism (XPD and XRCC1) on the Risk of Breast Cancer in Egyptian Female Patients. *Mol. Biol. Rep.* 39 (2), 1895–1901. doi:10.1007/s11033-011-0935-7

Iarc.fr (2022). Incidence, Mortality and Prevalence by Cancer Site. Available from: https://gco.iarc.fr/today/data/factsheets/populations/818-egypt-fact-sheets.pdf.

Ibrahim, A. S., Khaled, H. M., Mikhail, N. N., Baraka, H., and Kamel, H. (2014). Cancer Incidence in Egypt: Results of the National Population-Based Cancer Registry Program. *J. Cancer Epidemiol.*, 2014, 437971. doi:10.1155/2014/437971

Institute of Pathogen Genomics (2020). Africa CDC. [Cited 2022 Feb 5]. Available from: https://africacdc.org/institutes/ipg/.

Inzaule, S. C., Tessema, S. K., Kebede, Y., Ogwell Ouma, A. E., and Nkengasong, J. N. (2021). Genomic-Informed Pathogen Surveillance in Africa: Opportunities and Challenges. *Lancet Infect. Dis.* 21 (9), e281–e289. doi:10.1016/S1473-3099(20)30939-7

Ismail, A., El-Awady, R., Mohamed, G., Hussein, M., and Ramadan, S. S. (2018). Prognostic Significance of Serum Vitamin D Levels in Egyptian Females with Breast Cancer. *Asian Pac. J. Cancer Prev.* 19 (2), 571–576. doi:10.22034/APJCP.2018.19.2.571

Ismail, S., Essawi, M., Sedky, N., Hassan, H., Fayez, A., Helmy, N., et al. (2016). Roberts Syndrome: Clinical and Cytogenetic Studies in 8 Egyptian Patients and Molecular Studies in 4 Patients with Genotype/Phenotype Correlation. *Genet. Couns.* 27 (3), 305–323.

Ismail, S. R., Abdel-Rahim, N., Hashishe, M. M., and Abdallah, E. M. (1996). Newborn Screening for Certain Treatable Inborn Errors of Metabolism in Alexandria. *J. Egypt. Public Health Assoc.* 71 (5–6), 495–520.

Kandeel, A., Genedy, M., El-Refai, S., Funk, A. L., Fontanet, A., and Talaat, M. (2017). The Prevalence of Hepatitis C Virus Infection in Egypt 2015: Implications for Future Policy on Prevention and Treatment. *Liver Int.* 37 (1), 45–53. doi:10.1111/liv.13186

Kandeil, A., Mostafa, A., El-Shesheny, R., Shehata, M., Roshdy, W. H., Ahmed, S. S., et al. (2020). Coding-Complete Genome Sequences of Two SARS-CoV-2 Isolates from Egypt. *Microbiol. Resour. Announc.* 9 (22). doi:10.1128/MRA.00489-20

Knoema.com (2019). A Genomic Approach to Microbiology. *Nat. Rev. Genet.* 20 (6), 311. doi:10.1038/s41576-019-0131-5

Knoema.com (2022). Deaths from Communicable Diseases in Egypt. Available from: https://knoema.com/data/communicable-diseases+egypt+deaths.

Kotze, M. (2016). Application of Advanced Molecular Technology in the Diagnosis and Management of Genetic Disorders in South Africa. *S Afr. Med. J.* 106 (6 Suppl. 1), 114–118. doi:10.7196/SAMJ.2016.v106i6.11012

Kotze, M. J., van Velden, D. P., Botha, K., Badenhorst, C. H., Avenant, H., van Rensburg, S. J., et al. (2013). Pathology-Supported Genetic Testing Directed at Shared Disease Pathways for Optimized Health in Later Life. *Personalized Med.* 10 (5), 497–507. doi:10.2217/pme.13.43

Kouyoumjian, S. P., Chemaitelly, H., and Abu-Raddad, L. J. (2018). Characterizing Hepatitis C Virus Epidemiology in Egypt: Systematic Reviews, Meta-Analyses, and Meta-Regressions. *Sci. Rep.* 8 (1), 1661. doi:10.1038/s41598-017-17936-4

Kovanda, A., Zimani, A. N., and Peterlin, B. (2021). How to Design a National Genomic Project-A Systematic Review of Active Projects. *Hum. Genomics* 15 (1), 20. doi:10.1186/s40246-021-00315-6

Kozma, C. (2008). Skeletal Dysplasia in Ancient Egypt. *Am. J. Med. Genet.* 146A (23), 3104–3112. doi:10.1002/ajmg.a.32501

Lake, M. A. (2020). What We Know So Far: COVID-19 Current Clinical Knowledge and Research. *Clin. Med.* 20 (2), 124–127. doi:10.7861/clinmed.2019-coron

Maddirevula, S., Alsahli, S., Alhabeeb, L., Patel, N., Alzahrani, F., Shamseldin, H. E., et al. (2018). Expanding the Phenome and Variome of Skeletal Dysplasia. *Genet. Med.* 20 (12), 1609–1616. doi:10.1038/gim.2018.50

Mamdouh, S., Khorshed, F., Aboushousha, T., Hamdy, H., Diab, A., Seleem, M., et al. (2017). Evaluation of Mir-224, Mir-215 and Mir-143 as Serum Biomarkers for HCV Associated Hepatocellular Carcinoma. *Asian Pac. J. Cancer Prev.* 18 (11), 3167–3171. doi:10.22034/APJCP.2017.18.11.3167

Mampunye, L., van der Merwe, N. C., Grant, K. A., Peeters, A. V., Torrorey-Sawe, R., French, D. J., et al. (2021). Pioneering BRCA1/2 Point-of-Care Testing for Integration of Germline and Tumor Genetics in Breast Cancer Risk Management: A Vision for the Future of Translational Pharmacogenomics. *Front. Oncol.* 11, 11. doi:10.3389/fonc.2021.619817

Marchant, G., Barnes, M., Evans, J. P., LeRoy, B., Wolf, S. M., and LawSeq Liability Task Force (2020). From Genetics to Genomics: Facing the Liability Implications in Clinical Care. *J. L. Med. Ethics* 48 (1), 11–43. doi:10.1177/1073110520916994

Maxmen, A. (2020). The Next Chapter for African Genomics. *Nature* 578 (7795), 350–354. doi:10.1038/d41586-020-00454-1

Medendorp, N. M., van Maarschalkerweerd, P. E. A., Murugesu, L., Daams, J. G., Smets, E. M. A., and Hillen, M. A. (2020). The Impact of Communicating Uncertain Test Results in Cancer Genetic Counseling: A Systematic Mixed Studies Review. *Patient Edu. Couns.* 103 (9), 1692–1708. doi:10.1016/j.pec.2020.03.015

Meyer, R., Begemann, M., Hübner, C. T., Dey, D., Kuechler, A., Elgizouli, M., et al. (2021). One Test for All: Whole Exome Sequencing Significantly Improves the Diagnostic Yield in Growth Retarded Patients Referred for Molecular Testing for Silver-Russell Syndrome. *Orphanet J. Rare Dis.* 16 (1), 42. doi:10.1186/s13023-021-01683-x

Mohamed, W. M. Y. (2015). Neurogenomics: An Egyptian Perspective. *Appl. Translational Genomics* 5, 15–17. doi:10.1016/j.atg.2015.04.003

Motawi, T. M. K., El-Maraghy, S. A., Sabry, D., and Mehana, N. A. (2019). The Expression of Long Non Coding RNA Genes Is Associated with Expression with Polymorphisms of HULC Rs7763881 and MALAT1 Rs619586 in Hepatocellular Carcinoma and HBV Egyptian Patients. *J. Cell. Biochem.* 120 (9), 14645–14656. doi:10.1002/jcb.28726

Motawi, T. M. K., Sadik, N. A. H., Shaker, O. G., and Ghaleb, M. H. (2016). Elevated Serum microRNA-122/222 Levels Are Potential Diagnostic Biomarkers in Egyptian Patients with Chronic Hepatitis C but Not Hepatic Cancer. *Tumor Biol.* 37 (7), 9865–9874. doi:10.1007/s13277-016-4884-6

Nabil, A., El Shafei, S., El Shakankiri, N. M., Habib, A., Morsy, H., Maddirevula, S., et al. (2020). A Familial PLCB4 Mutation Causing Auriculocondylar Syndrome 2 with Variable Severity. *Eur. J. Med. Genet.* 63 (6), 103917. doi:10.1016/j.ejmg.2020.103917

Nagy, M., Lynch, M., Kamal, S., Mohamed, S., Hadad, A., Abouelnaga, S., et al. (2020). Assessment of Healthcare Professionals' Knowledge, Attitudes, and Perceived Challenges of Clinical Pharmacogenetic Testing in Egypt. *Personalized Med.* 17 (4), 251–260. doi:10.2217/pme-2019-0163

Nembaware, V., and Mulder, N. (2019). The African Genomic Medicine Training Initiative (AGMT): Showcasing a Community and Framework Driven Genomic Medicine Training for Nurses in Africa. *Front. Genet.* 10, 1209. doi:10.3389/fgene.2019.01209

Globaltimes.cn (2022). New Research Center Seeks to Map out China's Genes. [Cited 2022 Feb 5]. Available from: http://www.globaltimes.cn/content/1072485.shtml.

Newsbeezer (2022). Emirates News Agency – 'Reference Genome Project for Egyptians and Ancient Egyptians' to Be Held at Expo 2020 Dubai. [Cited 2022 Feb 13]. Available from: https://newsbeezer.com/uaeeng/emirates-news-agency-reference-genome-project-for-egyptians-and-ancient-egyptians-to-be-held-at-expo-2020-dubai/.

Genomeweb.com. (2022). Nigerian Research Consortium Begins Collecting Genetic. *Clinical Data from 100 K Subjects*. Genomeweb.com. [Cited 2022 Feb 5]Available at: https://www.genomeweb.com/sequencing/nigerian-research-consortium-begins-collecting-genetic-clinical-data-100k-subjects?

Nkengasong, J. (2020). Let Africa into the Market for COVID-19 Diagnostics (2020). *Nature* 580 (7805), 565. doi:10.1038/d41586-020-01265-0

Noncommunicable Diseases Egypt 2018 Country Profile (2018). Who.int. [Cited 2022 Feb 4]. Available from: https://www.who.int/publications/m/item/noncommunicable-diseases-egy-country-profile-2018.

Novarino, G., Fenstermaker, A. G., Zaki, M. S., Hofree, M., Silhavy, J. L., Heiberg, A. D., et al. (2014). Exome Sequencing Links Corticospinal Motor Neuron Disease to Common Neurodegenerative Disorders. *Science* 343 (6170), 506–511. doi:10.1126/science.1247363

O'Doherty, K. C., Shabani, M., Dove, E. S., Bentzen, H. B., Borry, P., Burgess, M. M., et al. (2021). Toward Better Governance of Human Genomic Data. *Nat. Genet.* 53 (1), 2–8. doi:10.1038/s41588-020-00742-6

Ojha, V., Mani, A., Pandey, N. N., Sharma, S., and Kumar, S. (2020). CT in Coronavirus Disease 2019 (COVID-19): A Systematic Review of Chest CT Findings in 4410 Adult Patients. *Eur. Radiol.* 30 (11), 6129–6138. doi:10.1007/s00330-020-06975-7

Omicron Spreads but Severe Cases Remain Low in South Africa (2022). WHO | Regional Office for Africa. [Cited 2022 Feb 5]. Available from: https://www.afro.who.int/news/omicron-spreads-severe-cases-remain-low-south-africa.

Owolabi, M. O., Akpa, O. M., Made, F., Adebamowo, S. N., Ojo, A., Adu, D., et al. (2019). Data Resource Profile: Cardiovascular H3Africa Innovation Resource (CHAIR). *Int. J. Epidemiol. [Internet]* 48 (2), 366–367g. doi:10.1093/ije/dyy261

Palsson, G., and Rabinow, P. (1999). Iceland: The Case of a National Human Genome Project. *Anthropol. Today* 15 (5), 14–18. doi:10.2307/2678370

Patel, N., Khan, A. O., Alsahli, S., Abdel-Salam, G., Nowilaty, S. R., Mansour, A. M., et al. (2018). Genetic Investigation of 93 Families with Microphthalmia or Posterior Microphthalmos. *Clin. Genet.* 93 (6), 1210–1222. doi:10.1111/cge.13239

Patel, N., Nabil, A., Alshammari, M., and Alkuraya, F. S. (2021). Hoarse Voice in Children as the Presenting Feature of ECM1-Related Lipoid Proteinosis. *Am. J. Med. Genet. Pt A* 185 (12), 3924–3925. doi:10.1002/ajmg.a.62406

Pereira, R., Oliveira, J., and Sousa, M. (2020). Bioinformatics and Computational Tools for Next-Generation Sequencing Analysis in Clinical Genetics. *Jcm* 9 (1), 132. doi:10.3390/jcm9010132

Policy Paper: A Framework for the Implementation of Genomic Medicine for Public Health in Africa (2022). Aasciences.africa. [Cited 2022 Feb 4]. Available from: https://www.aasciences.africa/publications/policy-paper-framework-implementation-genomic-medicine-public-health-africa.

Presidency.eg (2022). Presidency/Initiative-President-of-the-Republic-to Support-Women-Health-Egypt. https://www.presidency.eg/en/Presidency/initiative-president-of-the-republic-to support-women-health-Egypt/.

Ramadan, R. A., Desouky, L. M., Elnaggar, M. A., Moaaz, M., and Elsherif, A. M. (2014). Association of DNA Repair Genes XRCC1 (Arg399Gln), (Arg194Trp) and XRCC3 (Thr241Met) Polymorphisms with the Risk of Breast Cancer: A Case-Control Study in Egypt. *Genet. Test. Mol. Biomarkers* 18 (11), 754–760. doi:10.1089/gtmb.2014.0191

Ramsay, M. (2015). Growing Genomic Research on the African Continent: The H3Africa Consortium. *S Afr. Med. J.* 105 (12), 1016–1017. doi:10.7196/SAMJ.2015.v105i12.10281

Raouf, R. (2008). Capacity Building for the Transfer of Genetic Knowledge into Practice and Prevention. *Genet. Serv. Egypt: Curr. Situation Needs Assess.* Capabilitynet.eu. [Cited 2022 Feb 4]. Available from: http://www.capabilitynet.eu/Documents/Needs_Assesement_CAPABILITY_Egypt.pdf.

Rice, G. I., Forte, G. M. A., Szynkiewicz, M., Chase, D. S., Aeby, A., Abdel-Hamid, M. S., et al. (2013). Assessment of Interferon-Related Biomarkers in Aicardi-Goutières Syndrome Associated with Mutations in TREX1, RNASEH2A, RNASEH2B, RNASEH2C, SAMHD1, and ADAR: A Case-Control Study. *Lancet Neurol.* 12 (12), 1159–1169. doi:10.1016/S1474-4422(13)70258-8

Roser, M., and Ritchie, H. (2016). Burden of Disease. *Our World in Data*. [Cited 2022 Feb 5]. Available from: https://ourworldindata.org/burden-of-disease.

Rotimi, S. O., Rotimi, O. A., and Salhia, B. (2020). A Review of Cancer Genetics and Genomics Studies in Africa. *Front. Oncol.* 10, 606400. doi:10.3389/fonc.2020.606400

Sahu, P. K., Sao, R., Mondal, S., Vishwakarma, G., Gupta, S. K., Kumar, V., et al. (2020). Next Generation Sequencing Based Forward Genetic Approaches for Identification and Mapping of Causal Mutations in Crop Plants: A Comprehensive Review. *Plants* 9 (10), 1355. doi:10.3390/plants9101355

Saleh, E. M., Wahab, A. H., Elhouseini, M. E., and Eisa, S. S. (2004). Loss of Heterozygosity at BRCA1, TP53, Nm-23 and Other Loci on Chromosome 17q in Human Breast Carcinoma. *J. Egypt. Natl. Canc Inst.* 16 (1), 62–68.

Saunders, G., Baudis, M., Becker, R., Beltran, S., Béroud, C., Birney, E., et al. (2019). Leveraging European Infrastructures to Access 1 Million Human Genomes by 2022. *Nat. Rev. Genet.* 20 (11), 693–701. doi:10.1038/s41576-019-0156-9

Scott, E. M., Greater Middle East Variome ConsortiumHalees, A., Itan, Y., Spencer, E. G., He, Y., et al. (2016). Characterization of Greater Middle Eastern Genetic Variation for Enhanced Disease Gene Discovery. *Nat. Genet.* 48 (9), 1071–1076. doi:10.1038/ng.3592

Seifi, M., Footz, T., Taylor, S. A. M., Elhady, G. M., Abdalla, E. M., and Walter, M. A. (2016). Novel PITX2 Gene Mutations in Patients with Axenfeld-Rieger Syndrome. *Acta Ophthalmol.* 94 (7), e571–e579. doi:10.1111/aos.13030

Shamseldin, H. E., Shaheen, R., Ewida, N., Bubshait, D. K., Alkuraya, H., Almardawi, E., et al. (2020). The Morbid Genome of Ciliopathies: An Update. *Genet. Med.* 22 (6), 1051–1060. doi:10.1038/s41436-020-0761-1

Shawky, R. M., Karara, H. M. F., and Salem, M. S. Z. (2022). Genetic Study of Congenital Limb Anomalies Among Egyptian Children. *Ajol.info*. [Cited 2022 Feb 4]. Available from: https://www.ajol.info/index.php/ejhg/article/view/56147/44593.

Shawky, R. M., El-Awady, M. Y., Elsayed, S. M., and Hamadan, G. E. (2011). Consanguineous Matings Among Egyptian Population. *Egypt. J. Med. Hum. Genet.* 12 (2), 157–163. doi:10.1016/j.ejmhg.2011.07.001

Shawky, R. M., Elsayed, N. S., Ibrahim, D. S., and Seifeldin, N. S. (2012). Profile of Genetic Disorders Prevalent in Northeast Region of Cairo, Egypt. *Egypt. J. Med. Hum. Genet.* 13 (1), 45–62. doi:10.1016/j.ejmhg.2011.10.002

Sherif, M. M., Abou-Aita, B. A. S., Abou-Elew, M. H., and el-Kafrawi, A. O. M. M. (1985). Hepatitis B Virus Infection in Upper and Lower Egypt. *J. Med. Virol.* 15 (2), 129–135. doi:10.1002/jmv.1890150205

Shigemura, J., Ursano, R. J., Morganstein, J. C., Kurosawa, M., and Benedek, D. M. (2020). Public Responses to the Novel 2019 Coronavirus (2019-nCoV) in Japan: Mental Health Consequences and Target Populations. *Psychiatry Clin. Neurosci.* 74 (4), 281–282. doi:10.1111/pcn.12988

Shoeib, A., Velasquez Portocarrero, D. E., Wang, Y., and Jiang, B. (2020). First Isolation and Whole-Genome Characterization of a G9P[14] Rotavirus Strain from a Diarrheic Child in Egypt. *J. Gen. Virol.* 101 (9), 896–901. doi:10.1099/jgv.0.001455

Soliman, G., Elzalabany, M. S., Hassanein, T., and Miller, F. D. (2019). Mass Screening for Hepatitis B and C in Southern Upper Egypt. *BMC Public Health* 19 (1), 1326. doi:10.1186/s12889-019-7640-1

Stark, Z., Dolman, L., Manolio, T. A., Ozenberger, B., Hill, S. L., Caulfied, M. J., et al. (2019). Integrating Genomics into Healthcare: A Global Responsibility. *Am. J. Hum. Genet.* 104 (1), 13–20. doi:10.1016/j.ajhg.2018.11.014

State Information Service (2022). Egypt Launches Initiative for Early Detection of 19 Genetic Diseases in Newborns, Begins Free Treatment of Spinal Muscular Atrophy. Gov.eg. [Cited 2022 Feb 4]. Available from: https://www.sis.gov.eg/Story/158530/Egypt-launches-initiative-for-early-detection-of-19-genetic-diseases-in-newborns%2C-begins-free-treatment-of-spinal-muscular-atrophy?lang=en-us.

Stoller, J. K. (2018). The Challenge of Rare Diseases. *Chest* 153 (6), 1309–1314. doi:10.1016/j.chest.2017.12.018

Sulis, G., Roggi, A., Matteelli, A., and Raviglione, M. C. (2014). Tuberculosis: Epidemiology and Control. *Mediterr. J. Hematol. Infect. Dis.* 6 (1), e2014070. doi:10.4084/MJHID.2014.070

Swellam, M., Ismail, M., Eissa, S., Hamdy, M., and Mokhtar, N. (2004). Emerging Role of P53, Bcl-2 and Telomerase Activity in Egyptian Breast Cancer Patients. *IUBMB Life (International Union Biochem. Mol. Biol. Life)* 56 (8), 483–490. doi:10.1080/15216540400010834

Temtamy, S. (1998). Prevention of Genetic Diseases and Malformations in Newborns. *Health Popul. Scientific J. Ministry Health Popul.* 2, 22–27.

Temtamy, S. A., Aglan, M. S., and Meguid, N. A. (2010). "Genetic Disorders in Egypt," in *Genetic Disorders Among Arab Populations* (Berlin, Heidelberg: Springer Berlin Heidelberg), 219–272. doi:10.1007/978-3-642-05080-0_8

Temtamy, S. A., and Hussen, D. F. (2017). Genetics and Genomic Medicine in Egypt: Steady Pace. *Mol. Genet. Genomic Med.* 5 (1), 8–14. doi:10.1002/mgg3.271

Temtamy, S. A. (2019). The Development of Human Genetics at the National Research Centre, Cairo, Egypt: A Story of 50 Years. *Annu. Rev. Genom. Hum. Genet.* 20 (1), 1–19. doi:10.1146/annurev-genom-083118-015201

Tessema, S. K., Inzaule, S. C., Christoffels, A., Kebede, Y., de Oliveira, T., Ouma, A. E. O., et al. (2020). Accelerating Genomics-Based Surveillance for COVID-19 Response in Africa. *The Lancet Microbe* 1 (6), e227–e228. doi:10.1016/S2666-5247(20)30117-8

The General Authority for Health Accreditation and Control (2022). [Internet]. Gov.eg. [Cited 2022 Feb 5]. Available from: https://gahar.gov.eg/.

The International Genome Sample Resource (2022). *Genomes Project Dataset*. London, United Kingdom: Genomics England. [Cited 2022 Feb 5]. Available from: https://www.genomicsengland.co.uk/about-gecip/for-gecip-members/data-and-data-access/.

Thomas, M. M., Ashaat, E. A., Otaify, G. A., Ismail, S., Essawi, M. L., Abdel-Hamid, M. S., et al. (2021). First Report of Two Egyptian Patients with Desbuquois

Dysplasia Due to Homozygous CANT1 Mutations. *Mol. Syndromol.* 12 (5), 279–288. doi:10.1159/000516607

Tindana, P., as members of the H3Africa ConsortiumYakubu, A., Staunton, C., Matimba, A., Littler, K., et al. (2019). Engaging Research Ethics Committees to Develop an Ethics and Governance Framework for Best Practices in Genomic Research and Biobanking in Africa: The H3Africa Model. *BMC Med. Ethics* 20 (1). doi:10.1186/s12910-019-0398-2

Tishkoff, S. A., Reed, F. A., Friedlaender, F. R., Ehret, C., Ranciaro, A., Froment, A., et al. (2009). The Genetic Structure and History of Africans and African Americans. *Science* 324 (5930), 1035–1044. doi:10.1126/science.1172257

Toomes, C., James, J., Wood, A. J., Wu, C. L., McCormick, D., Lench, N., et al. (1999). Loss-of-Function Mutations in the Cathepsin C Gene Result in Periodontal Disease and Palmoplantar Keratosis. *Nat. Genet.* 23 (4), 421–424. doi:10.1038/70525

Toyoshima, Y., Nemoto, K., Matsumoto, S., Nakamura, Y., and Kiyotani, K. (2020). SARS-CoV-2 Genomic Variations Associated with Mortality Rate of COVID-19. *J. Hum. Genet.* 65 (12), 1075–1082. doi:10.1038/s10038-020-0808-9

Traverso, M., Assereto, S., Gazzerro, E., Savasta, S., Abdalla, E. M., Rossi, A., et al. (2013). Novel FAM126A Mutations in Hypomyelination and Congenital Cataract Disease. *Biochem. Biophysical Res. Commun.* 439 (3), 369–372. doi:10.1016/j.bbrc.2013.08.077

van der Merwe, N., and Kotze, M. J. (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* (South Africa: University of Johannesburg), 21–23.

Wahab, A. H., Kassem, A. M., Matter, S., El Deen, A. F., Helmy, A. S., Ismaeil, M. M., et al. (2010). Role of KLF6 Tumor Suppressor Gene Mutations in the Development of Colorectal Carcinoma in an Egyptian Population. *Hepatogastroenterology* 57 (104), 1405–1410.

Waked, I., Esmat, G., Elsharkawy, A., El-Serafy, M., Abdel-Razek, W., Ghalab, R., et al. (2020). Screening and Treatment Program to Eliminate Hepatitis C in Egypt. *N. Engl. J. Med.* 382 (12), 1166–1174. doi:10.1056/NEJMsr1912628

Watson, J. D. (1990). The Human Genome Project: Past, Present, and Future. *Science* 248 (4951), 44–49. doi:10.1126/science.2181665

World Health Organization (2022). Consultation on the Global Genomic Surveillance Strategy for Pathogens with Pandemic and Epidemic Potential. [Internet]. Who.int. [Cited 2022 Mar 19]. Available from: https://www.who.int/news-room/events/detail/2021/12/08/default-calendar/who-consultation-on-the-global-genomic-surveillance-strategy-for-pathogens-with-pandemic-and-epidemic-potential.

World Health Organization (2013). *Global Tuberculosis Report 2013*. Geneva: World Health Organization. Available from: https://apps.who.int/iris/handle/10665/91355.

World Health Organization (2016). Key Facts on Hepatitis C Treatment. [Cited 2022 Feb 4]. Available from: http://www.who.int/medicines/areas/access/hepCtreat_key_facts/en/.

World Health Organization (2021). SARS-CoV-2 Variants. [Cited 2022 Feb 5]. Available from: https://www.who.int/csr/don/31-december-2020-sars-cov2-variants/en/.

Wohlers, I., Künstner, A., Munz, M., Olbrich, M., Fähnrich, A., Calonga-Solís, V., et al. (2020). An Integrated Personal and Population-Based Egyptian Genome Reference. *Nat. Commun.* 11 (1), 4719. doi:10.1038/s41467-020-17964-1

World Health Organization (2020). Regional Office for Africa. Emergency Operations Annual Report: Saving Lives and Reducing Suffering WHO's Work in Emergency Response Operations in the WHO African Region in 2018. Brazzaville: World Health Organization Regional Office for Africa. Available from: https://apps.who.int/iris/handle/10665/330396.

World Health Organization (2021). *World Health Statistics 2021: Monitoring Health for the SDGs, Sustainable Development Goals*. Geneva, Switzerland: World Health Organization. Available from: https://www.who.int/publications/i/item/9789240027053/

Yakubu, A., Tindana, P., Matimba, A., Littler, K., Munung, N. S., Madden, E., et al. (2018). Model Framework for Governance of Genomic Research and Biobanking in Africa - A Content Description. *AAS Open Res.* 1, 13. doi:10.12688/aasopenres.12844.210.12688/aasopenres.12844.1

Zaki, M., Shehab, M., El-Aleem, A. A., Abdel-Salam, G., Koeller, H. B., Ilkin, Y., et al. (2007). Identification of a Novel Recessive *RELN* Mutation Using a Homozygous Balanced Reciprocal Translocation. *Am. J. Med. Genet.* 143A (9), 939–944. doi:10.1002/ajmg.a.31667

Zaki, S. M., Abdel-Azeez, H. A., El Nagar, M. R., Metwally, K. A.-A., and Ahmed, M. M. S. S. M. M. S. (2015). Analysis of FHIT Gene Methylation in Egyptian Breast Cancer Women: Association with Clinicopathological Features. *Asian Pac. J. Cancer Prev.* 16 (3), 1235–1239. doi:10.7314/apjcp.2015.16.3.1235

Zekri, A.-R. N., Mohanad, M., Hafez, M. M., Soliman, H. K., Hassan, Z. K., Abouelhoda, M., et al. (2021). Genome Sequencing of SARS-CoV-2 in a Cohort of Egyptian Patients Revealed Mutation Hotspots that Are Related to Clinical Outcomes. *Biochim. Biophys. Acta (Bba) - Mol. Basis Dis.* 1867 (8), 166154. doi:10.1016/j.bbadis.2021.166154

Zekri, A.-R. N., Easa Amer, K., Hafez, M. M., Hassan, Z. K., Ahmed, O. S., Soliman, H. K., et al. (2021). Genomic Characterization of SARS-CoV-2 in Egypt. *J. Adv. Res.* 30, 123–132. doi:10.1016/j.jare.2020.11.012

Zidan, H. E., Karam, R. A., El-Seifi, O. S., and Abd Elrahman, T. M. (2018). Circulating Long Non-Coding RNA MALAT1 Expression as Molecular Biomarker in Egyptian Patients with Breast Cancer. *Cancer Genet.* 220, 32–37. doi:10.1016/j.cancergen.2017.11.005

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers | Frontiers in Genetics

ORIGINAL RESEARCH
published: 08 June 2022
doi: 10.3389/fgene.2022.864822

Whole Exome Sequencing in South Africa: Stakeholder Views on Return of Individual Research Results and Incidental Findings

Nicole Van Der Merwe[1,2]*, *Raj Ramesar*[1] and *Jantina De Vries*[3,4]

[1]UCT/MRC Genomic and Precision Medicine Research Unit, Division of Human Genetics, Institute for Infectious Diseases and Molecular Medicine, Department of Pathology, Faculty of Medicine and Health Sciences, University of Cape Town, Cape Town, South Africa, [2]Department of Pathology, Faculty of Medicine and Health Sciences, Stellenbosch University, Tygerberg, South Africa, [3]Department of Medicine, Faculty of Health Sciences, University of Cape Town, Cape Town, South Africa, [4]Neuroscience Institute, Faculty of Health Sciences, University of Cape Town, Observatory, South Africa

OPEN ACCESS

Edited by:
Dawn Stephens,
Technology Innovation Agency (TIA), South Africa

Reviewed by:
Shiri Shkedi-Rafid,
Hadassah Medical Center, Israel
Varun Sharma,
NMC Healthcare, India

***Correspondence:**
Nicole Van Der Merwe
genetic.counselling@famgen.co.za

Specialty section:
This article was submitted to Human and Medical Genomics, a section of the journal Frontiers in Genetics

Received: 28 January 2022
Accepted: 30 March 2022
Published: 08 June 2022

Citation:
Van Der Merwe N, Ramesar R and De Vries J (2022) Whole Exome Sequencing in South Africa: Stakeholder Views on Return of Individual Research Results and Incidental Findings. Front. Genet. 13:864822. doi: 10.3389/fgene.2022.864822

The use of whole exome sequencing (WES) in medical research is increasing in South Africa (SA), raising important questions about whether and which individual genetic research results, particularly incidental findings, should be returned to patients. Whilst some commentaries and opinions related to the topic have been published in SA, there is no qualitative data on the views of professional stakeholders on this topic. Seventeen participants including clinicians, genomics researchers, and genetic counsellors (GCs) were recruited from the Western Cape in SA. Semi-structured interviews were conducted, and the transcripts analysed using the framework approach for data analysis. Current roadblocks for the clinical adoption of WES in SA include a lack of standardised guidelines; complexities relating to variant interpretation due to lack of functional studies and underrepresentation of people of African ancestry in the reference genome, population and variant databases; lack of resources and skilled personnel for variant confirmation and follow-up. Suggestions to overcome these barriers include obtaining funding and buy-in from the private and public sectors and medical insurance companies; the generation of a locally relevant reference genome; training of health professionals in the field of genomics and bioinformatics; and multidisciplinary collaboration. Participants emphasised the importance of upscaling the accessibility to and training of GCs, as well as upskilling of clinicians and genetic nurses for return of genetic data in collaboration with GCs and medical geneticists. Future research could focus on exploring the development of stakeholder partnerships for increased access to trained specialists as well as community engagement and education, alongside the development of guidelines for result disclosure.

Keywords: whole exome sequencing, incidental findings, secondary findings, return of results, genetic counselling, South Africa

INTRODUCTION

The past two decades have seen a considerable increase in the use of genomic methods in health research and clinical care. Studies have assessed the use of whole exome sequencing (WES) for a broad spectrum of disorders towards increasing the diagnostic yield in patient populations restricted to specific phenotypes, to up to 80% (Neveling et al., 2013; Yadava and Ashkinadze, 2017). In South Africa (SA), WES is currently used only in the research setting at approximately 10 facilities/institutions, few of whom have published exome data (Roberts et al., 2016; van der Merwe, 2016; Pierrache et al., 2017; Baynam et al., 2020; Sawe et al., 2020). An advantage of WES over individual gene or next generation sequencing (NGS) panel testing is that the stored data enables re-analysis of new as well as pharmacogenomic genes that are linked to the patient's phenotype. WES is furthermore reported to be beneficial in patients with atypical presentations as it may reduce the time to diagnosis as well as the psychological and financial burdens associated with prolonged investigation (Monroe et al., 2016; Harris et al., 2017; Stark et al., 2017; Walsh et al., 2017).

Although WES has great potential to improve diagnostic accuracy, health outcomes and resource utilization, several limitations contribute to a widening gap between the generation of complex genetic data and its use in daily clinical practice. Return of individual WES results, notably unanticipated or incidental findings (IFs) generated in the research setting in particular, remains a complex and highly debated issue (Wolf et al., 2008; Berg et al., 2011; Bredenoord et al., 2011; Ortiz-Osorno et al., 2015). One of the key challenges relates to what to do with unanticipated individual research results. Internationally, various recommendations on how to deal with findings that are outside of the test indication [secondary findings (SFs)], and guidance for their return, have been established (ACMG policy statement, 2021). However, in Africa, no standardised guidelines exist. While the H3Africa consortium (https://h3africa.org/) has developed a policy for the return of findings, little empirical data exists about the views of African professional stakeholders on the topic. The Individual Findings in Genomics Research (IFGeneRA) project conducted under the umbrella of H3Africa has established a working group tasked to generate and publish an African-specific list of reportable genes/variants (Wonkam & De Vries, 2020). These authors proposed that the African genetics community in collaboration with international experts, generate three priority lists. Until such guidelines are developed and publically available, we refer to IFs.

To date, this is the first South African study conducted to map the experiences of health professionals regarding the practice of WES and return of IFs. Mwaka et al. (2021) conducted a qualitative study on Ugandan researchers' perspectives on return of individual genetic findings to research participants. They asserted that community engagement, reconsenting and adequate preparation of participants to safely receive individual results, may be achieved by building capacity and increasing access to clinical genetics and genetic counselling (GC). Given the possibility of future reinterpretation of data, non-disclosure of actionable research results may be considered unethical (Bombard et al., 2019), however, the no-return criterion may apply when results from genomic studies are intended as generalisable findings instead of those demonstrating clinical utility (AESA, 2020 policy paper, 2020). Several studies conducted in high income countries that have examined the expectations of research participants and attitudes toward the return of individual research results with potential clinical significance, have shown that a majority of individuals want genetic data returned to them (Christenhusz et al., 2013; Wolf, 2013). In addition, some studies revealed that many research participants want all these data regardless of its "actionability" or clinical significance (Bollinger et al., 2013; Harris et al., 2013; Mackley et al., 2017). Alternatively, the right of participants (not) to know has also been extensively discussed (Herring and Foster, 2012). The distinction between research and clinical sequencing has served as a justification for not returning IFs unless they meet the criteria of clinical utility, actionability, and clinical urgency (Clift et al., 2015).

Limited work has been done to investigate obligations to return IFs in lower-and middle income countries, including those in Africa (Kerasidou, 2015). The obligation of healthcare professionals to patients - to improve patients' health, avoid psychological and social harms and provide prolonged disease-free survival - should arguably extend to the return of any findings which would facilitate the latter, even in the African setting where the healthcare system is limited and heavily under-resourced. In the local setting, pertinent issues that affect WES are further muddled by the challenges inherent to the public healthcare system i.e. the lack of infrastructure and adequately trained personnel, high burden of communicable disease and high incidence of poverty experienced by patients (Mayosi et al., 2012; Barron and Padarath, 2017; Manyisa and Van Aswegen, 2017; Bradshaw et al., 2019). Another important challenge hindering the return of IFs is insufficient validation of genomic findings for genetically diverse African populations. Most of the currently well-established bioinformatics tools and variant calling pipelines are benchmarked using non-African genomic data, and available reference genomes are biased toward non-African populations, resulting in mislabelling of rare variants and potential misclassification by in silico prediction tools (Bao et al., 2014; Martin et al., 2018; Bope et al., 2019; Wonkam and De Vries, 2020).

Against this backdrop, this study aims to explore the experiences and perspectives of pertinent stakeholders on the return of IFs in the South African research context. Specifically, we conducted a study with clinicians, genomics researchers, and genetic counsellors (GCs) who are involved with WES in the research and/or clinical setting.

MATERIALS AND METHODS

Essential aspects of the study context and methods, research team, findings, analysis and interpretations are reported according to the 32-point consolidated criteria for reporting qualitative research (COREQ) checklist (Tong et al., 2007).

Study Design

This study was designed as an exploratory qualitative study which enables exploration of the essential qualities of complex phenomena. Phenomenology is a qualitative research method that involves a detailed examination of participants' perceptions of their shared and lived experiences (Smith and Firth, 2011).

Research Team

This research was conducted to fulfil the requirements for an MSc (Med) Genetic Counsellor degree at the University of Cape Town, SA. NvdM and JdV are female, RR is male. Whilst being an intern Genetic Counsellor at the time, NvdM had also obtained a PhD degree in Pathology from Stellenbosch University. RR is a Genetic Counsellor and Professor in Human Genetics; JdV is an Associate Professor in Bioethics with a background in Sociology. NvdM obtained training in qualitative research methods as part of her genetic counselling degree; she also received guidance from JdV throughout the process. As a scientist with experience in WES variant interpretation and reporting, NvdM had a tendency to focus on the technical rather than the ethical aspects of WES during the interviews. JdV and NvdM worked together to minimise such bias and ensure that interviews covered broader aspects of WES beyond the technical ones only.

Participants

A purposive sampling method was used to select participants most likely to have the expertise or experience to provide valuable insights on the research topic. Some participants were known to members of the research team and were recruited from the University of Cape Town, Stellenbosch University, Red Cross War Memorial Children's Hospital as well as two private medical facilities from September 2016 to March 2017. Pertinent stakeholders such as clinicians, genomics researchers and GCs involved in any stage of the WES process, both in the clinical and research settings, were included in the study. Participants were selected based on their (presumed) insight relating to the return of individual WES results. Stakeholders were invited to participate in this study by means of an email that included a description of the research to be conducted, the ethics approval document and consent form. While twenty-eight participants were invited to participate, 17 were recruited as part of the study. Interview dates and times were subsequently established with those who responded, and face-to-face, semi-structured interviews conducted at their offices/consultation rooms. Interviews were conducted by NvdM with no-one else present. All interviews took place between October 2016 and April 2017, were conducted in English and lasted between 45 minutes and an hour. Saturation was reached by the 14th interview and an additional three interviews were conducted to ensure this.

Instrumentation and Procedures

The topic guide was designed based on the literature, including theoretical work that identifies and describes issues pertaining to WES/return of findings, as well as empirical studies conducted elsewhere in the world. Questions were carefully framed to remove bias and any suggestive wording, and to ensure neutrality. They related to stakeholder experiences and understanding of various aspects pertaining to WES; examples of its use; the role of WES in clinical practice; consent relating to return of findings; return of WES results and data storage and reuse. Finally, the topic guide was reviewed by the two project supervisors and piloted twice to elucidate potential weaknesses, flaws or limitations in its design. This enabled adaptations to be made to the topic guide prior to implementation of the study. During interviews, closed-ended questions were used to capture demographic data, followed by a series of open-ended questions using the topic guide. Interviews were audio-recorded and no repeat interviews were conducted.

Data Analysis

The framework approach was selected to analyse the data as it allows researchers to trace, map and categorise the perspectives of individual participants across themes or other key attributes. This approach was initially developed in the 1980's for social policy research and is a useful tool for qualitative data analysis in the healthcare setting (Smith and Firth, 2011). A systematic, four-step process was used to facilitate rigorous data analysis and transparent data management, ultimately resulting in the categorisation of data into themes and interrelated sub-themes (Spencer et al., 2003).

The interview recordings were transcribed verbatim by the researcher (NvdM). An alphanumeric code (e.g., P1-P17) was assigned to each participant to ensure participant confidentiality and any identifiable information removed during transcription. This was followed by randomisation of this code to disrupt the sequential order of interviews. Transcripts were not returned to participants for verification. They were initially open-coded to generate an open-coding scheme, which was subsequently discussed with supervisors to develop a hierarchical coding scheme. This was followed by importing transcripts into a data management software program called NVivo11 (QSR International). The entire dataset was coded and analysed using the Framework Table function in NVivo. Initial steps of this analysis occurred simultaneously with data collection. Field-notes were used to facilitate interpretation of the data and were subsequently discussed with the research team.

RESULTS

A total of 17 participants were interviewed in this study (**Table 1**). They had various levels of experience with WES, being either directly or indirectly involved in a WES process in the research, or clinical, or both settings (**Table 2**). "Direct" involvement refers to direct interaction with WES data or the patients whom WES was performed in, as opposed to "indirect" involvement where the use of WES is limited to the environment the participant operates in. The themes and sub-themes identified by transcript analysis, as described in the methods section, are summarised in **Table 3**. For the purpose of this study, Only extracts of the data pertaining

TABLE 1 | Themes and sub-themes generated during the data analysis phase of the study.

Themes	Sub-themes
Whole exome sequencing (WES) practice	WES selection criteria WES approach Barriers and opportunities
Incidental findings in WES	Knowledge and experience Type of results that should be returned Result validation Community engagement
Ethical considerations and implications	Risks and benefits Consent models Type of information that should be provided Data ownership, accessibility and storage Local guidelines: views and recommendations
Return of findings	Who should return findings The role of the genetic counsellor The need for training in the field of genomics

to the themes "WES practice" and "incidental findings", are presented in this paper.

Participants' Understanding of Various Aspects Pertaining to Whole Exome Sequencing

Participants displayed various levels of understanding of WES technology, with genomics researchers articulating a greater depth of knowledge and understanding of the technical scope and limitations of WES compared to clinicians. When discussing how and when WES is used in clinical practice and health research, most participants who were directly involved in WES asserted that the selection of patients is based on the complexity or heterogeneity of the disease and the lack of findings using standard genetic testing approaches. Other interviewees, particularly those in the research setting, based the selection of patients for WES on the occurrence of treatment failure, novelty or "publishability" of the condition. With regard to the sequencing approach, many interviewees in both settings described that their first port of call would be to use a NGS or a virtual WES panel, and only to turn to full exome sequencing analysis when these options have been exhausted. A virtual gene panel refers to restricted analysis of disease-associated genes sequenced using WES. The reasons for using targeted gene panels were ascribed to the fact that it narrows the spectrum of findings (both WES and NGS panels) and increases the sequencing coverage (NGS panels).

> "Targeted WES seems to be, from a clinical point of view, perhaps a more viable option, because then if you have at least some narrowing down of what you are looking for, the ability to perhaps target certain genes seems to be useful. So, we have found that successful in a group of patients. . ." - P7
>
> ". . .there's two very important reasons to pursue this test (WES) and the one is, if there's treatment failure and secondly, if there's a family history that cannot be accounted for by the initial BRCA screening, whether it's the founder mutations in a high risk group or a full BRCA1 or 2 screen." – P9

Some participants ascribed their preference of a virtual panel to the lower cost of WES, compared to constructing/using a multi-gene NGS panel. Other interviewees described their point of departure to be full WES analysis, to avoid preclusion of novel findings.

TABLE 2 | Demographics of the study participants.

Participant code	Years of practice in field: 1) 1–5; 2) 6–10; 3) 11–15; 4) 16–20; 5) >21	Sector: private practice (PP)/public institution (PI)	Formal training in genetics (yes/no)	Direct/indirect use of WES	Setting (research, clinical or both)
P1	4	PI	Yes	Direct	Research
P2	2	PI	Yes	Direct	Both
P3	2	PI	Yes	Direct	Research
P4	5	PP	No	Indirect	Research
P5	3	PI	Yes	Direct	Research
P6	2	PI	Yes	Direct	Research
P7	5	PI & PP	No	Indirect	Research
P8	4	PP	No	Indirect	Research
P9	5	PI & PP	Yes	Direct	Research
P10	2	PI	Yes	Direct	Research
P11	2	PI & PP	Yes	Direct	Both
P12	5	PI	No	Direct	Research
P13	5	PI	No	Direct	Both
P14	3	PI	Yes	Direct	Research
P15	3	PI	Yes	Direct	Research
P16	3	PI	Yes	Direct	Research
P17	2	PI	Yes	Indirect	Research

TABLE 3 | Number and proportion of study participants, listed by profession.

Stakeholder		Participants	
		N = 17	% of total
Clinician	Other	5	29
	Medical Geneticist	3	18
	Genetic Counselor	3	18
Researcher	Medical Scientist	6	35

While various challenges have been described in the literature with regard to the technical and ethical aspects of WES, interviewees described that specific barriers such as resource scarcity (including finances, skilled professionals and population-specific genetic knowledge) and lack of infrastructure are pronounced in the South African setting and impede the routine use of WES.

> "Resources, definitely. Human resources, mainly. As I understand it, the biggest limiting factor is, in South Africa, well, one thing is the data storage, apparently, and the other thing is the bioinformatics, the manpower . . . obviously being able to that right. From our experience here, it does also sound like there has been issues with that, in that two labs both doing WES, are coming up with different things." - P6

> "...when we do have the policies and the data management capacity in place, the bioinformatics support in place - which hopefully will happen but I am not holding my breath because that would mean getting a significant increase in staffing and funding - then we can start thinking about opening up the scope that we offer [returning IFs/SFs]." – P2

An array of opportunities were identified by interviewees to overcome such barriers, including the generation of a knowledge database that serves as a source of genetic information pertaining to local, indigenous populations. In addition, multidisciplinary collaboration and training of health professionals in the field of genomics were suggested as a means of understanding the incorporation of genomic data into clinical practice.

> "If you are working out of European population (databases), you are going to find a lot of novel variants. So, if we have to implement it in the clinical practice, we need to enrich first of all the exome databases, and the reference example from all the populations. Otherwise you are going to have a misrepresentation if you use the exome database that is enriched just for the population of European descent, which is the case now." – P14

Participant Views on What Findings Should Be Returned

More than 80% of stakeholders had not themselves encountered IFs, since they had wilfully avoided the possibility of obtaining them by using more targeted approaches.

> "...if you're not using virtual panels, or you're not just looking at the genes of interest, in that way you would complicate your life." – P6

> "...I would never ever look at everything, simply because I do not want to know. That could point to incidental findings. [...] So, you cannot just go and open up this can of worms if you are not ethically prepared for it. And you need to have policies in place to say, if we find something, what are we going to do with it. ...If we have the capacity and we have the infrastructure and the policies in place then we can start thinking about opening up the scope that we offer. [...] I don't even want to touch it at the moment and I avoid it at all costs." – P2

A range of views were expressed with regards to the kind of WES results that should be returned, echoing the five principles related to actionability as well as the availability of resources to return such results. Considerations were given to the likely clinical and psychosocial impact of the research result, and the autonomy of the patient. Generally, participants expressed that there is an obligation to return any results with life-saving potential or those that will impact on the clinical management of the patient in terms of prevention, surveillance and medical intervention - whether obtained in the research or the clinical setting. Approximately 33% of interviewees (mostly researchers) described an obligation to return results pertaining only to the disease that is being studied, while mainly "other" clinicians advocated return of any actionable results, including those outside of the disease in question. Overall, medical scientists seemed to be most unfavourably inclined to returning IFs, compared to clinicians who responded more favourably in this regard.

> "I believe that feeding back of important results to families is the only thing. Whether we do it the same way they do it in America; the answer is: definitively not. [...] Important for me would be a life-threatening expectation given in an individual, and for which medical care can provide some therapeutic solutions, or at least some follow-up and some preventive solution. I believe that the [56] actionable genes are a good start. We need to define, according to the context, which type of results will have to be fed back." – P14

Participants made a distinction between the clinical and research setting in relation to returning IFs. Some believe that in the clinical setting, there is an obligation to disclose any IF that points to a disease or the predisposition thereof, particularly due to the legal implications associated with non-disclosure – whether actionable or not. Lastly, some participants held that there should exist two categories of results for disclosure: a standardised list for all participants as well as a patient-specific list that is to be informed by the personal, medical and family history of the patient. Two

participants contended that the ACMG's list of reportable SFs is a good point of departure, but that eliciting the view of the general population in terms of what they think should be returned, is of utmost importance.

> "...I believe that the [now 73] actionable genes are a good start. We need to define, according to the context, which type of result will have to be fed back. [...] But before we do that, it will be always nice to know what the people think about that: what does the stakeholder, what does the general population think about that? So, I do think that there is a fair amount of qualitative type of research to be performed in the population to know, first of all, their knowledge of genetics, and second, their knowledge of all this new emerging technology, and their willingness eventually to have those results back. [...] We need to know what our people think first." – P14

Next Generation Sequencing Result Validation

Study participants interpreted the concept of result validation as confirming that the variant detected by WES is actually present (exclusion of false positives), as well as confirming the clinical significance or pathogenicity of the variant (disease-causing potential). While 90% of participants thought that it is crucial to confirm WES results with Sanger sequencing, some were of the opinion that it is merely necessary to confirm results that were obtained in an unaccredited setting and/or those that will be acted upon.

> "I would never send out a result if it hasn't been confirmed with Sanger sequencing at all..." – P2,
>
> "I think that it's just probably not practical to Sanger sequence every single variant. But I do think if one is doing something - in a not carefully quality-controlled environment - that is going to have a healthcare intervention, you should probably double check that result before you act on it." – P16

While some participants contended that results should be confirmed strictly in an accredited environment, others believe that the lack of accredited laboratories in (South) Africa needs to be taken into account.

> "Validation on sanger sequencing should be done in a diagnostic kind of environment as far as possible." – P5
>
> "The gold standard of diagnostic validation is something that has been set in the Western context. But if we have the context of the African continent, we have so few labs that can actually do molecular diagnostics - and I will be fairly confident that even an unaccredited lab that is working in a reasonable environment, clinical environment, can be trustworthy from the beginning. We have to have a two-stage approach. One stage is that we work with the resources that we have, and in the second stage, we can move forward to our own type of system of validation. But the fact that we cannot validate should not stop us from feeding back that result to the patient." – P14

DISCUSSION

Recent years have seen a local increase in the use of WES for the generation of genetic reference data from historically marginalized populations to help distinguish real from spurious findings and to improve the diagnosis of rare diseases with a low diagnostic rate including primary immunodeficiency, retinal degenerative (Roberts et al., 2016; Pierrache et al., 2017) and mitochondrial disorders, neurodevelopmental conditions, hearing loss (Walsh et al., 2017; Wonkam et al., 2021), and cancers (van der Merwe et al., 2017; Sawe et al., 2020). In the breast cancer setting, van der Merwe et al. (2017) and Sawe et al. (2020) described the use of WES with appropriate consent for simultaneous assessment of inherited-, lifestyle- and therapy-induced risk, toward improved patient management across the continuum of care. While WES and whole genome sequencing (WGS) (Glanzmann et al., 2021) is still only used in research, NGS panels have recently been adopted in the clinical setting, particularly for the diagnosis of cancers. Less than 40 molecular biology medical laboratories across SA including some of the centres from which participants were recruited, are accredited with the South African National Accreditation System (https://www.sanas.co.za/), and very few laboratories' NGS gene panels are currently accredited for diagnostic use. Interviewees, and particularly the researchers, seemed to be well versed in the benefits, risks and limitations of WES and were optimistic about its future implementation in clinical practice.

An interesting observation was that in the absence of local policies on how to manage IFs, stakeholders have largely avoided obtaining them. The majority of participants expressed a preference for targeted approaches such as the use of disease-specific NGS panels, for the purpose of reducing the likelihood of IFs and variants of uncertain significance (VUSes). This is in line with previous recommendations of the American Society of Human Genetics (ASHG) to use targeted sequencing or selective sequence analysis, to minimise the likelihood of discovering IFs particularly during adolescent testing (Botkin et al., 2015). Conversely, ACMG policy statements are largely in favour of disclosing SFs in clinical exome and genome sequencing regardless of patient age. The ACMG furthermore has not explicitly considered reporting of SFs with NGS or WES virtual panel testing until 2019/2020, when they described various challenges related to obtaining consent and the additional workload incurred when using these test methods (Miller et al., 2021).

The development of African guidelines for disclosure of genomic findings - and in particular IFs - are imperative given the novelty and high level of variation of African DNA, and the impact of its absence on local implementation of WES. Due to lower costs,

customisable testing options and confidence in laboratory processes, a number of stakeholders are routinely making use of overseas-based sequencing (NGS) facilities. This phenomenon has been somewhat criticised due to the embargo and otherwise unlimited access of African data to non-African scientists, but not opposed in lieu of the lack of standards and limited clinical implementation of this technology in SA. The lack of guidelines for IFs and the complexities relating to variant interpretation poses some of the main challenges that are impeding the widespread adoption of WES.

On a technical level, low coverage of certain genomic regions, the discordance between variant-calling pipelines and bioinformatics tools, and biased reference genomes used in WES further obstructs its routine use in the clinical setting. Genetic heterogeneity is a major limitation when dealing with African genomic data as it may increase the number of false positive and false negative findings in relation to the reference genome used. When working with an ill-representative reference genome, the increased amount of rare (especially missense) variation detected with lack of corresponding functional information makes it difficult to establish pathogenicity. Equally, if the depth of germline sequencing is too low, false positive variation due to systematic errors cannot be distinguished from the presence of somatic variation. Several studies evaluated variant filtering tools and criteria for categorising novel variants, reviewed available databases and in-silico prediction tools, and proposed recommendations for analysing variation in African genomes (Kessler et al., 2016; Bope et al., 2019). Variant annotation is furthermore a crucial step in the analysis of NGS data and results can have a strong influence on variant classification. As may be the case when working with African genomic data, incorrect or incomplete annotations may cause researchers to miss, overlook or dilute note-worthy DNA variants in a pool of false positives (McCarthy, et al., 2014). Since less than 2% of human genomes analysed comprises that of African individuals (Sirugo et al., 2019), variant annotation is expected to improve as the availability of African genomic data increases, with subsequent increase in classification accuracy. One of the study participants reported that ~70% of causative variants detected using WES in African patients were novel and not found in publically available databases. They emphasised the need for exome databases and the reference genome to be enriched for African data in order for WES to be successfully implemented in clinical practice. Approximately 70% of the current human reference genome, GRCh38, is derived from a single individual and therefore it fails to capture the genetic diversity of most populations (Ballouz et al., 2019). Methods proposed to overcome this issue include nucleotide additions and extensions to GRCh38, graph-based references that simultaneously represent multiple, diverse populations and the generation of population-specific consensus sequences *via de novo* assembly of raw read data (Huang et al., 2013; Duan et al., 2019; Li et al., 2020). Population-specific reference genomes/panels developed to date include European, East Asian and African (Yoruban) major allele reference sequences Dewey et al. (2011), Vietnamese (Nguyen et al., 2015), Danish (Maretty et al., 2017), Chinese (Zhang et al., 2021), Japanese (Takayama et al., 2021) and Arab genome sequences (Fakhro et al., 2016; Elbait et al., 2021), and an African American reference panel (O'Connell et al., 2021). In 2021, Ebert et al. (2021) published a new, more comprehensive reference dataset reflecting 64 assembled human genomes representing 25 different human populations from across the globe. In a Nature commentary, Wonkam (2021) shared the vision of H3Africa to sequence three million African genomes (3MAG) in order to build a more representative reference genome. van der Merwe et al. (2017) previously compared the GRCh37 (hg19) reference genome to a European major allele reference sequence during WES in breast cancer patients and controls. The authors demonstrated that using an ethnically concordant reference genome increases the specificity and sensitivity of WES results.

WES is furthermore susceptible to multiple source errors including sequencing errors, incorrect mapping ("mismapping"), and random sampling (Kiezun et al., 2012). In light of these caveats, WES data is routinely validated with Sanger sequencing which is considered to be the gold standard for result confirmation (McCourt et al., 2013). According to the study participants, three aspects pertaining to the validation of WES data are important to consider: whether and what results should in fact be confirmed using Sanger sequencing; whether validation should occur in an accredited laboratory; and whether or not this depends on the setting. Nearly 90% of participants expressed that it is crucial to confirm all reported WES results with Sanger sequencing (or PCR), especially if the result is to be used for family screening or prenatal testing. However, not all of these stakeholders contended that variant confirmation should be done in an accredited laboratory. The rest were of the opinion that the technology has improved to the extent that we may be able to trust it, and that only findings obtained in an unaccredited environment that lead to a healthcare intervention, should be confirmed. It appears that even in the literature there is no absolute consensus regarding the need for NGS variant confirmation using Sanger sequencing. While some studies highlight the importance of Sanger confirmation, others do not (Strom et al., 2014; Beck et al., 2016; Mu et al., 2016).

Interviewees emphasised the need for context-specific practices, i.e., practices that take into consideration limited local capacity and the lack of accredited laboratories in the (South) African setting. While some participants held that validation should be done exclusively in an accredited facility, others contended that they are fairly confident that results generated in an unaccredited laboratory operating in a reasonable clinical environment, may be trustworthy, and that the lack of accreditation should not hamper or obstruct the return of results to patients.

Additional challenges mentioned by stakeholders that are echoed in the literature (Bertier et al., 2016) include the lack of data management, storage and bioinformatics capacity, funding and our understanding of gene-gene/gene-environment interactions and protective effects. An array of opportunities were identified to overcome the barriers to widespread adoption of WES in clinical practice. These include: obtaining buy-in from the private sector, government and medical insurance companies for funding

tests, and the generation of a locally relevant reference genome or knowledge database. The ability to perform exome sequencing and functional studies in parallel would furthermore aid interpretation of rare/population-specific missense variation if accompanied by family and case-control studies. In addition to this, multidisciplinary collaboration and training of health professionals in the field of genomics were suggested as a means of facilitating better understanding, interpretation and incorporation of genomic data into clinical care. This is corroborated by the literature, which suggests that variant interpretation requires the collaborative intervention of different highly-trained specialists (Newman and Black, 2014) including bioinfomaticians, biologists and clinicians (Sisodiya, 2015). Krause (2019) highlighted the requirements for an improved genetic testing service in SA. These include correct indications for testing (appropriate selection criteria), the generation of a local database resource, implementation of state-of-the-art testing based on research, and increased academic training including active training of GCs and medical geneticists.

What Findings Should be Returned?

IFs obtained with WES is a contentious subject (Bredenoord et al., 2011; Wolf, 2013). While a limited number of interviewees preempted IFs alongside extension of their WES analyses, most participants did not foresee obtaining IFs, largely because they confine analysis and return of findings to more targeted approaches. In addition to avoiding IFs due to the lack of standards for dealing with them, concerns regarding available infrastructure and capacity for managing them also arose. Specifically, stakeholders emphasized that the lack of resources for returning primary findings is exaggerated by the prospect of delivering an even greater volume of (unanticipated) results with potential clinical importance. This larger volume and wider scope of results calls for an increase in several resources, which would place additional strain on an already-stretched healthcare system. While an average of 30% of patients who undergo diagnostic WES obtain a molecular diagnosis (Valencia et al., 2015; Trujillano et al., 2017), ~3% of individuals are also left with an IF of unclear significance or secondary genomic finding (Yang et al., 2014; eMERGE Clinical Annotation Working Group, 2020). VUSes account for 40% of variants discovered to date (Federici and Soddu, 2020). WES of 6503 individuals of European and African ancestry have identified actionable variants in 2% and 1.1% in these groups, respectively (Amendola et al., 2015). The uniqueness and missing heritability of African populations raised the question of whether we should use existing recommendations developed by the ACMG or 100,000 Genomes Project (www.genomicsengland.co.uk) and customise these to the local context, or develop African-specific actionable targets prior to the implementation of local policies (Wonkam and De Vries, 2020). As local institutions are expanding WES efforts, determining the relevance and application of international recommendations across diverse populations have become paramount. Currently, the IFGeneRA project involving both local and international experts in the field, is developing African-specific guidelines for returning genomic findings obtained using African DNA (Wonkam and De Vries, 2020) (https://h3africa.org/). This is a significant endeavour since ethnicity is likely to be an important predictor of penetrance of pathogenic variants (Trinh et al., 2014). Stakeholders contended that penetrance studies require the analyses of diverse populations, since allelic background is likely to influence the phenotypic consequences of IFs.

In relation to the type of WES results that ought to be reported, a range of views were expressed by stakeholders. Although varied in terms of the setting in which WES results are obtained, views were largely grounded in the clinical and psychosocial impact of the research result, and the autonomy of the patient. Participants expressed that in both the clinical and research settings, there is an obligation to deliver results with life-saving potential or those that will impact the clinical management of the patient in relation to prevention, surveillance or medical intervention - if related to the disease in question (or test requested). The literature largely supports the disclosure of variants taking into consideration actionability, pathogenicity, phenotypic severity, analytic validity, and participant consent (Fabsitz et al., 2010; Cassa et al., 2012). While stakeholders agreed upon the criteria of actionability, 33% of participants (mainly researchers) contended that only pertinent research results should be delivered, while clinicians (other than GCs and medical geneticists) expressed that any actionable results, whether primary or incidental, should be returned. The literature supports the latter (Clift et al., 2015). Interestingly, among stakeholders, the clinicians appeared to be more focused on the potential clinical impact of IFs and their duty to return meaningful results whereas scientists were more focused on the lack of guidelines and limited resources available for seeking, confirming and reporting IFs. This disparity potentially relates to the differential allocation and/or availability of resources within the two settings as well as to the ACMG recommendations which states that clinicians have an obligation to deliver IFs that pertain to their patients' health and management (American College of Medical Genetics and Genomics Board of directors, 2021). Hallowell et al. (2015) argued that issues concerning the return of IFs are influenced by the context in which WES is performed, namely the clinical or research setting, and that decisions about the disclosure of IFs generated in the clinical context are much less ethically contentious than those detected in the research setting.

Five principles for determining whether research results should be returned to participants were outlined in the literature and reflected in this study, including obtaining appropriately informed consent, analytical validity of results, the possibility of result-based intervention and adequate mechanisms and resources in place to return results (McGuire et al., 2008; Fabsitz et al., 2010; Munung et al., 2016). GCs may be in the ideal position to proactively define their roles in the WES process. These may include an essential role in the consenting process where they are able to obtain informed consent in a staged manner through ongoing patient interaction, thereby potentially mitigating some of the concerns related to the high volumes of information to be relayed and the potential negative emotions they may provoke. Ormond (2013) discussed several areas in which practice will likely change as we move from genetic to

"genomic" counselling, while Patch and Middleton (2018) outlined its current implementation in GC training curricula. The progressive adoption of WES into the clinic furthermore allows for the awareness, education and training of various stakeholders (Brazas and Ouellette, 2016). In (South) Africa where the number of GCs and medical geneticists are few, we contend that laboratories, researchers, public and private institutions housing clinicians, genetic nurses and other healthcare workers ought to work in partnership with GCs to disseminate genetic information (Baynam et al., 2020). The establishment of an online genetic counselling platform (FamGen Counselling™) to facilitate nation and ultimately continent-wide access through multidisciplinary collaboration and integrated GC networks, was an important consideration towards reducing the widening gap between increased amounts of genomic data being produced and insufficient access to/underutilisation of trained genetics specialists who are able to make sense of it. Efforts are furthermore geared towards upscaling and certifying the conventional and unconventional GC workforce, respectively, given the pace at which genomics is incorporated into clinical practice. To furthermore promote multidisciplinary stakeholder training and interaction, various online courses and webinars attended by local and international genomics researchers, clinicians and GCs are periodically conducted to share and identify gaps in current knowledge and address relevant clinical questions encountered in oncogenomic practice. Although knowing which reportable findings stakeholders deem important, is essential for developing local guidelines, a researcher and clinician highlighted the need to determine the views of the general population in this regard. To achieve this, a fair amount of qualitative, community or patient-focused research is required.

STRENGTHS AND LIMITATIONS OF THE STUDY

This study is the first of its kind to employ qualitative methodology to explore stakeholder views and experiences of WES IFs in the South African setting. It highlights the barriers related to the local context but also identifies valuable opportunities to overcome these barriers – which may be relevant to other lower and middle-income countries. As there are no other qualitative research regarding the role of GCs in returning WES findings, this research is valuable for the growing field of GC in SA.

Being the researcher's first introduction to qualitative research methodology, it is possible that the quality of the interviews has improved over the course of the study. Furthermore, purposive sampling may have lent itself to ascertainment bias. Lastly, the views of other health professionals who are not involved in WES, patients and the community as a whole, were not included in this study.

In conclusion, this study provides valuable insights into the views and experiences of local stakeholders involved in the WES process in SA, which have not previously been explored. Appraisal of stakeholder experiences and practices facilitated the understanding the current operations, challenges and barriers within respective settings, but also identified opportunities and potential solutions which may ultimately inform the adoption of WES into clinical practice. Current roadblocks that impede this process include a lack of local guidelines for returning IFs; complexities related to variant interpretation due to lack of functional studies and underrepresentation of people of African ancestry in the reference genome, population and variant databases; the lack of resources both in terms of skilled personnel and infrastructure for variant confirmation and follow-up. The demand for adequately skilled professionals will likely be met in two ways: the upskilling of health professionals in the field of genomics and the upscaling of GC use and accessibility to a larger volume of the (South) African population.

DATA AVAILABILITY STATEMENT

The data supporting the findings and conclusions of this article will be made available by the authors without undue reservation.

ETHICS STATEMENT

This study was reviewed and approved by University of Cape Town's Faculty of Health Sciences Health Research Ethics Committee (FHS HREC/REF: 226/2010). Study participants provided written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

All authors have made substantial contributions to the conception and design of the work. NvdM, RR, and JdV together conceptualised the study and designed the interview instrument. NvdM is responsible for the acquisition, analysis, and interpretation of data and has drafted this article. NvdM and JdV together developed the coding scheme; NvdM coded all the data. NvdM, RR, and JdV together discussed themes and insights emerging from data analysis. All authors have revised this article critically for important intellectual content. JdV and RR have provided final approval of the version to be published and all authors have agreed to be accountable for all aspects of the work and ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

FUNDING

JdV is supported by IFGeneRA, an H3Africa ELSI Collaborative Centre (Grant No. 1U54HG009790-01, administered by the National Human Genome Research Institute).

REFERENCES

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

Amendola, L. M., Dorschner, M. O., Robertson, P. D., Salama, J. S., Hart, R., Shirts, B. H., et al. (2015). Actionable Exomic Incidental Findings in 6503 Participants: Challenges of Variant Classification. *Genome Res.* 25 (3), 305–315. doi:10.1101/gr.183483.114

American College of Medical Genetics and Genomics (ACMG) Board of directors (2021). Recommendations for Reporting of Secondary Findings in Clinical Exome and Genome Sequencing, 2021 Update: a Policy Statement of the American College of Medical Genetics and Genomics (ACMG). *Genet. Med.* 23, 1391–1398.

Ballouz, S., Dobin, A., and Gillis, G. A. (2019). Is it Time to Change the Reference Genome?. *Genome Biology* 20, 1–9.

Bao, R., Huang, L., Andrade, J., Tan, W., Kibbe, W. A., Jiang, H., et al. (2014). Review of Current Methods, Applications, and Data Management for the Bioinformatics Analysis of Whole Exome Sequencing. *Cancer Inform.* 13, 67–82. doi:10.4137/CIN.S13779

Barron, P., and Padarath, A. (2017). *Twenty Years of the South African Health Review, South African Health Review 2017*. Durban: Health Systems Trust.

Baynam, G. S., Groft, S., van der Westhuizen, F. H., Gassman, S. D., du Plessis, K., Coles, E. P., et al. (2020). A Call for Global Action for Rare Diseases in Africa. *Nat. Genet.* 52, 21–26. doi:10.1038/s41588-019-0552-2

Beck, T. F., Mullikin, J. C., and Biesecker, L. G. (2016). NISC Comparative Sequencing ProgramSystematic Evaluation of Sanger Validation of Next-Generation Sequencing Variants. *Clin. Chem.* 62 (4), 647–654. doi:10.1373/clinchem.2015.249623

Berg, J. S., Khoury, M. J., and Evans, J. P. (2011). Deploying Whole Genome Sequencing in Clinical Practice and Public Health: Meeting the challenge One Bin at a Time. *Genet. Med.* 13 (6), 499–504. doi:10.1097/gim.0b013e318220aaba

Bertier, G., Hétu, M., and Joly, Y. (2016). Unsolved Challenges of Clinical Whole-Exome Sequencing: a Systematic Literature Review of End-Users' Views. *BMC Med. Genomics* 9 (1), 52. doi:10.1186/s12920-016-0213-6

Bollinger, J. M., Green, R. C., and Kaufman, D. (2013). Attitudes about Regulation Among Direct-To-Consumer Genetic Testing Customers. *Genet. Test. Mol. Biomarkers* 17 (5), 424–428. doi:10.1089/gtmb.2012.0453

Bombard, Y., Brothers, K. B., Fitzgerald-Butt, S., Garrison, N. A., Garrison, L., James, C. A., et al. (2019). The Responsibility to Recontact Research Participants after Reinterpretation of Genetic and Genomic Research Results. *Am. J. Hum. Genet.* 104 (4), 578–595. doi:10.1016/j.ajhg.2019.02.025

Bope, C. D., Chimusa, E. R., Nembaware, V., Mazandu, G. K., de Vries, J., and Wonkam, A. (2019). Dissecting In Silico Mutation Prediction of Variants in African Genomes: Challenges and Perspectives. *Front. Genet.* 10, 601. doi:10.3389/fgene.2019.00601

Botkin, J. R., Belmont, J. W., Berg, J. S., Berkman, B. E., Bombard, Y., Holm, I. A., et al. (2015). Points to Consider: Ethical, Legal, and Psychosocial Implications of Genetic Testing in Children and Adolescents. *Am. J. Hum. Genet.* 97 (1), 6–21. doi:10.1016/j.ajhg.2015.05.022

Bradshaw, D., Nannan, N. N., Pillay-van Wyk, V., Laubscher, R., Groenewald, P., and Dorrington, R. E. (2019). Burden of Disease in South Africa: Protracted Transitions Driven by Social Pathologies. *S Afr. Med. J.* 109 (11b), 69–76. doi:10.7196/SAMJ.2019.v109i11b.14273

Brazas, M. D., and Ouellette, B. F. F. (2016). Continuing Education Workshops in Bioinformatics Positively Impact Research and Careers. *Plos Comput. Biol.* 12 (6), e1004916. doi:10.1371/journal.pcbi.1004916

Bredenoord, A. L., Kroes, H. Y., Cuppen, E., Parker, M., and van Delden, J. J. M. (2011). Disclosure of Individual Genetic Data to Research Participants: the Debate Reconsidered. *Trends Genet.* 27 (2), 41–47. doi:10.1016/j.tig.2010.11.004

Cassa, C. A., Savage, S. K., Taylor, P. L., Green, R. C., McGuire, A. L., and Mandl, K. D. (2012). Disclosing Pathogenic Genetic Variants to Research Participants: Quantifying an Emerging Ethical Responsibility. *Genome Res.* 22 (3), 421–428. doi:10.1101/gr.127845.111

Christenhusz, G. M., Devriendt, K., Dierickx, K., Littler, K., Masiye, F., Ouwe-Missi-Oukem-Boyer, O., et al. (2013). To Tell or Not to Tell? A Systematic Review of Ethical Reflections on Incidental Findings Arising in Genetics Contexts. *Eur. J. Hum. Genet.* 21 (3), 248–255. doi:10.1038/ejhg.2012.130

Clift, K. E., Halverson, C. M. E., Fiksdal, A. S., Kumbamu, A., Sharp, R. R., and McCormick, J. B. (2015). Patients' Views on Incidental Findings from Clinical Exome Sequencing. *Appl. Translational Genomics* 4, 38–43. doi:10.1016/j.atg.2015.02.005

Daw Elbait, G., Henschel, A., Tay, G. K., and Al Safar, H. S. (2021). A Population-specific Major Allele Reference Genome from the United Arab Emirates Population. *Front. Genet.* 12, 660428. doi:10.3389/fgene.2021.660428

Dewey, F. E., Chen, R., Cordero, S. P., Ormond, K. E., Caleshu, C., Karczewski, K. J., et al. (2011). Phased Whole-Genome Genetic Risk in a Family Quartet Using a Major Allele Reference Sequence. *Plos Genet.* 7 (9), e1002280. doi:10.1371/journal.pgen.1002280

Duan, Z., Qiao, Y., Lu, J., Lu, H., Zhang, W., Yan, F., et al. (2019). Hupan: A Pan-Genome Analysis Pipeline for Human Genomes. *Genome Biology* 20, 149.

Ebert, P., Audano, P. A., Zhu, Q., Rodriguez-Martin, B., Porubsky, D., Bonder, M. J., et al. (2021). Haplotype-Resolved Diverse Human Genomes and Integrated Analysis of Structural Variation. *Science* 372 (6537), eabf7117. doi:10.1126/science.abf7117

eMERGE Clinical Annotation Working Group (2020). Frequency of Genomic Secondary Findings Among 21,915 eMERGE Network Participants. *Genet. Med.* 22, 1470–1477. doi:10.1038/s41436-020-0810-9

Fabsitz, R. R., McGuire, A., Sharp, R. R., Puggal, M., Beskow, L. M., Biesecker, L. G., et al.National Heart, Lung (2010). Ethical and Practical Guidelines for Reporting Genetic Research Results to Study Participants: Updated Guidelines from a National Heart, Lung, and Blood Institute Working Group. *Circ. Cardiovasc. Genet.* 3 (6), 574–580. doi:10.1161/CIRCGENETICS.110.958827

Fakhro, K., Staudt, M., Ramstetter, M., Robay, A., Badii, A., Khalil, C., et al. (2016). The Qatar Genome: A Population-Specific Tool for Precision Medicine in the Middle East. *Hum. Genome. Var.* 3, 16016. doi:10.1038/hgv.2016.16

Federici, G., and Soddu, S. (2020). Variants of Uncertain Significance in the Era of High-Throughput Genome Sequencing: a Lesson from Breast and Ovary Cancers. *J. Exp. Clin. Cancer Res.* 39, 46. doi:10.1186/s13046-020-01554-6

Glanzmann, B., Jooste, T., Ghoor, S., Gordon, R., Mia, R., Mao, J., et al. (2021). Human Whole Genome Sequencing in South Africa. *Sci. Rep.* 11, 606. doi:10.1038/s41598-020-79794-x

Hallowell, N., Hall, A., Alberg, C., and Zimmern, R. (2015). Revealing the Results of Whole-Genome Sequencing and Whole-Exome Sequencing in Research and Clinical Investigations: Some Ethical Issues: Table 1. *J. Med. Ethics* 41 (4), 317–321. doi:10.1136/medethics-2013-101996

Harris, A., Kelly, S. E., and Wyatt, S. (2013). Counseling Customers: Emerging Roles for Genetic Counselors in the Direct-To-Consumer Genetic Testing Market. *J. Genet. Counsel* 22 (2), 277–288. doi:10.1007/s10897-012-9548-0

Harris, E., Topf, A., Barresi, R., Hudson, J., Powell, H., Tellez, J., et al. (2017). Exome Sequences versus Sequential Gene Testing in the UK Highly Specialised Service for Limb Girdle Muscular Dystrophy. *Orphanet. J. Rare Dis.* 12, 151. doi:10.1186/s13023-017-0699-9

Herring, J., and Foster, C. (2012). "Please Don't Tell Me". *Camb Q. Healthc. Ethics* 21 (1), 20–29. doi:10.1017/s0963180111000466

Huang, L., Popic, V., and Batzoglou, S. (2013). Short Read Alignment With Populations of Genomes. *Bioinformatics* 29, i361–i370.

Kerasidou, A. (2015). Sharing the Knowledge: Sharing Aggregate Genomic Findings with Research Participants in Developing Countries. *Developing World Bioeth.* 15 (3), 267–274. doi:10.1111/dewb.12071

Kessler, M. D., Yerges-Armstrong, L., Yerges-Armstrong, L., Taub, M. A., Shetty, A. C., Maloney, K., et al. (2016). Challenges and Disparities in the Application of Personalized Genomic Medicine to Populations with African Ancestry. *Nat. Commun.* 7, 12521. doi:10.1038/ncomms12521

Kiezun, A., Garimella, K., Do, R., Stitziel, N. O., Neale, B. M., McLaren, P. J., et al. (2012). Exome Sequencing and the Genetic Basis of Complex Traits. *Nat. Genet.* 44 (6), 623–630. doi:10.1038/ng.2303

Krause, A. (2019). New Genetic Testing Technologies: Advantages and Limitations. *S Afr. Med. J.* 109 (4), 207–209. doi:10.7196/samj.2019.v109i4.13990

Li, H., Feng, X., and Chu, C. (2020). The Design and Construction of Reference Pangenome Graphs With Minigraph. *Genome Biology* 21, 1–9.

Mackley, M. P., Fletcher, B., Parker, M., Watkins, H., and Ormondroyd, E. (2017). Stakeholder Views on Secondary Findings in Whole-Genome and Whole-Exome Sequencing: a Systematic Review of Quantitative and Qualitative Studies. *Genet. Med.* 19 (3), 283–293. doi:10.1038/gim.2016.109

Manyisa, Z. M., and Van Aswegen, E. J. (2017). Factors Affecting Working Conditions in Public Hospitals: A Literature Review. *Int. J. Africa Nurs. Sci.* 6, 28–38. doi:10.1016/j.ijans.2017.02.002

Maretty, L., Jensen, J., Petersen, B., Sibbesen, J. A., Liu, S., Villesen, P., et al. (2017). Sequencing and *de novo* Assembly of 150 Genomes From Denmark as a Population Reference. *Nature* 548, 87–91. doi:10.1038/nature23264

Martin, A. R., Teferra, S., Möller, M., Hoal, E. G., and Daly, M. J. (2018). The Critical Needs and Challenges for Genetic Architecture Studies in Africa. *Curr. Opin. Genet. Develop.* 53, 113–120. doi:10.1016/j.gde.2018.08.005

Mayosi, B. M., Lawn, J. E., van Niekerk, A., Bradshaw, D., Abdool Karim, S. S., and Coovadia, H. M. (2012). Health in South Africa: Changes and Challenges since 2009. *The Lancet* 380 (9858), 2029–2043. doi:10.1016/s0140-6736(12)61814-5

McCarthy, D. J., Humburg, P., Kanapin, A., Rivas, M. A., Gaulton, K., Cazier, J.-B., et al. (2014). Choice of Transcripts and Software Has a Large Effect on Variant Annotation. *Genome Med.* 6, 26. doi:10.1186/gm543

McCourt, C. M., McArt, D. G., Mills, K., Catherwood, M. A., Maxwell, P., Waugh, D. J., et al. (2013). Validation of Next Generation Sequencing Technologies in Comparison to Current Diagnostic Gold Standards for BRAF, EGFR and KRAS Mutational Analysis. *PLoS ONE* 8 (7), e69604. doi:10.1371/journal.pone.0069604

McGuire, A. L., Caulfield, T., and Cho, M. K. (2008). Research Ethics and the challenge of Whole-Genome Sequencing. *Nat. Rev. Genet.* 9 (2), 152–156. doi:10.1038/nrg2302

Miller, D. T., Lee, K., Gordon, A. S., Amendola, L. M., Adelman, K., Bale, S. J., et al.ACMG Secondary Findings Working Group (2021). Recommendations for Reporting of Secondary Findings in Clinical Exome and Genome Sequencing, 2021 Update: a Policy Statement of the American College of Medical Genetics and Genomics (ACMG). *Genet. Med.* 23 (8), 1391–1398. doi:10.1038/s41436-021-01171-4

Monroe, G. R., Frederix, G. W., Savelberg, S. M. C., de Vries, T. I., Duran, K. J., van der Smagt, J. J., et al. (2016). Effectiveness of Whole-Exome Sequencing and Costs of the Traditional Diagnostic Trajectory in Children with Intellectual Disability. *Genet. Med.* 18, 949–956. doi:10.1038/gim.2015.200

Mu, W., Lu, H.-M., Chen, J., Li, S., and Elliott, A. M. (2016). Sanger Confirmation Is Required to Achieve Optimal Sensitivity and Specificity in Next-Generation Sequencing Panel Testing. *J. Mol. Diagn.* 18 (6), 923–932. doi:10.1016/j.jmoldx.2016.07.006

Munung, N. S., Marshall, P., Campbell, M., Littler, K., Masiye, F., Ouwe-Missi-Oukem-Boyer, O., et al. (2016). Obtaining Informed Consent for Genomics Research in Africa: Analysis of H3Africa Consent Documents. *J. Med. Ethics* 42 (2), 132–137. doi:10.1136/medethics-2015-102796

Mwaka, E. S., Sebatta, D. E., Ochieng, J., Munabi, I. G., Bagenda, G., Ainembabazi, D., et al. (2021). Researchers' Perspectives on Return of Individual Genetics Results to Research Participants: a Qualitative Study. *Glob. Bioeth.* 32 (1), 15–33. doi:10.1080/11287462.2021.1896453

Neveling, K., Feenstra, I., Gilissen, C., Hoefsloot, L. H., Kamsteeg, E.-J., Mensenkamp, A. R., et al. (2013). A post-hoc Comparison of the Utility of sanger Sequencing and Exome Sequencing for the Diagnosis of Heterogeneous Diseases. *Hum. Mutat.* 34 (12), 1721–1726. doi:10.1002/humu.22450

Newman, W., and Black, G. (2014). Delivery of a Clinical Genomics Service. *Genes* 5, 1001–1017. doi:10.3390/genes5041001

O'Connell, J., Yun, T., Moreno, M., Li, H., Litterman, N., and Kolesnikov, A. (2021). A Population-Specific Reference Panel for Improved Genotype Imputation in African Americans. *Commun. Biol.* 4, 1269. doi:10.1038/s42003-021-02777-9

Ormond, K. E. (2013). From Genetic Counseling to "genomic Counseling". *Mol. Genet. Genomic Med.* 1 (4), 189–193. doi:10.1002/mgg3.45

Ortiz-Osorno, A., Ehler, L. A., and Brooks, J. (2015). Considering Actionability at the Participant's Research Setting Level for Anticipatable Incidental Findings from Clinical Research. *J. L. Med. Ethics* 43 (3), 619–632. doi:10.1111/jlme.12304

Patch, C., and Middleton, A. (2018). Genetic Counselling in the Era of Genomic Medicine. *Br. Med. Bull.* 126 (1), 27–36. doi:10.1093/bmb/ldy008

Pierrache, L. H. M., Kimchi, A., Ratnapriya, R., Roberts, L., Astuti, G. D. N., Obolensky, A., et al. (2017). Whole-Exome Sequencing Identifies Biallelic IDH3A Variants as a Cause of Retinitis Pigmentosa Accompanied by Pseudocoloboma. *Ophthalmology* 124 (7), 992–1003. doi:10.1016/j.ophtha.2017.03.010

Roberts, L., Ratnapriya, R., du Plessis, M., Chaitankar, V., Ramesar, R. S., and Swaroop, A. (2016). Molecular Diagnosis of Inherited Retinal Diseases in Indigenous African Populations by Whole-Exome Sequencing. *Invest. Ophthalmol. Vis. Sci.* 57 (14), 6374–6381. doi:10.1167/iovs.16-19785

Sirugo, G., Williams, S. M., and Tishkoff, S. A. (2019). The Missing Diversity in Human Genetic Studies. *Cell* 177 (1), 26–31. doi:10.1016/j.cell.2019.02.048

Sisodiya, S. M. (2015). Genetic Screening and Diagnosis in Epilepsy? *Curr. Opin. Neurol.* 28, 136–142. doi:10.1097/wco.0000000000000180

Smith, J., and Firth, J. (2011). Qualitative Data Analysis: the Framework Approach. *Nurse Res.* 18 (2), 52–62. doi:10.7748/nr2011.01.18.2.52.c8284

Spencer, L., Ritchie, J., Lewis, J., and Dillon, L. (2003). *Quality in Qualitative Evaluation: A Framework for Assessing Research Evidence*. London: Government Chief Social Researcher's OfficeCabinet Office. Retrieved from http://dera.ioe.ac.uk/21069/.

Stark, Z., Schofield, D., Alam, K., Wilson, W., Mupfeki, N., Macciocca, I., et al. (2017). Prospective Comparison of the Cost-Effectiveness of Clinical Whole-Exome Sequencing with that of Usual Care Overwhelmingly Supports Early Use and Reimbursement. *Genet. Med.* 19, 867–874. doi:10.1038/gim.2016.221

Strom, S. P., Lee, H., Das, K., Vilain, E., Nelson, S. F., Grody, W. W., et al. (2014). Assessing the Necessity of Confirmatory Testing for Exome-Sequencing Results in a Clinical Molecular Diagnostic Laboratory. *Genet. Med.* 16 (7), 510–515. doi:10.1038/gim.2013.183

Takayama, J., Tadaka, S., Yano, K., Katsuoka, F., Gocho, C., Funayama, T., et al. (2021). Construction and Integration of Three *de novo* Japanese Human Genome Assemblies Toward a Population-Specific Reference. *Nat. Commun.* 12, 226. doi:10.1038/s41467-020-20146-8

Thanh, N. D., Trang, P. T. M., Hai, D. T., Tuan, N. H. A., Quang, L. S., Minh, B. Q., et al. (2015). *Building Population-specific Reference Genomes: A Case Study of Vietnamese Reference Genome*. Seventh International Conference on Knowledge and Systems Engineering KSE, 97–102. doi:10.1109/KSE.2015.49

Tong, A., Sainsbury, P., and Craig, J. (2007). Consolidated Criteria for Reporting Qualitative Research (COREQ): a 32-item Checklist for Interviews and Focus Groups. *Int. J. Qual. Health Care* 19, 349–357. doi:10.1093/intqhc/mzm042

Torrorey-Sawe, R., van der Merwe, N., Mining, S. K., and Kotze, M. J. (2020). 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.* 11, 170. doi:10.3389/fgene.2020.00170

Trinh, J., Guella, I., and Farrer, M. J. (2014). Disease Penetrance of Late-Onset Parkinsonism. *JAMA Neurol.* 71 (12), 1535–1539. doi:10.1001/jamaneurol.2014.1909

Trujillano, D., Bertoli-Avella, A. M., Kumar Kandaswamy, K., Weiss, M. E., Köster, J., and Marais, A. (2017). Clinical Exome Sequencing: Results From 2819 Samples Reflecting 1000 Families. *EJHG* 25 (2), 176–182. doi:10.1038/ejhg.2016.146

Valencia, C. A., Husami, A., Holle, J., Johnson, J. A., Qian, Y., Mathur, A., et al. (2015). Clinical Impact and Cost-Effectiveness of Whole Exome Sequencing as a Diagnostic Tool: A Pediatric Center's Experience. *Front. Pediatr.* 3, 67. doi:10.3389/fped.2015.00067

van der Merwe, N. (2016). *Development of a Pathology Supported Pharmacogenetics Test for Improved Clinical Management of South African Patients with Breast Cancer and Associated Comorbidities*. Tygerberg, South Africa: Stellenbosch University. http://scholar.sun.ac.za/handle/10019.1/98481.

van der Merwe, N., Peeters, A., Pienaar, F., Bezuidenhout, J., van Rensburg, S., and Kotze, M. (2017). Exome Sequencing in a Family with Luminal-type Breast Cancer Underpinned by Variation in the Methylation Pathway. *Ijms* 18 (2), 467. doi:10.3390/ijms18020467

Walsh, M., Bell, K. M., Chong, B., Creed, E., Brett, G. R., Pope, K., et al. (2017). Diagnostic and Cost Utility of Whole Exome Sequencing in Peripheral Neuropathy. *Ann. Clin. Transl. Neurol.* 4, 318–325. doi:10.1002/acn3.409

Wolf, S. M., Lawrenz, F. P., Nelson, C. A., Kahn, J. P., Cho, M. K., Clayton, E. W., et al. (2008). Managing Incidental Findings in Human Subjects Research: Analysis and Recommendations. *J. L. Med. Ethics* 36 (2), 219211–219248. doi:10.1111/j.1748-720x.2008.00266.x

Wolf, S. M. (2013). Return of Individual Research Results and Incidental Findings: Facing the Challenges of Translational Science. *Annu. Rev. Genom. Hum. Genet.* 14, 557–577. doi:10.1146/annurev-genom-091212-153506

Wonkam, A., and De Vries, J. (2020). Returning Incidental Findings in African Genomics Research. *Nat. Genet.* 52, 17–20. doi:10.1038/s41588-019-0542-4

Wonkam, A., Lebeko, K., Mowla, S., Noubiap, J. J., Chong, M., and Pare, G. (2021). Whole Exome Sequencing Reveals a Biallelic Frameshift Mutation in GRXCR2 in Hearing Impairment in Cameroon. *Mol. Genet. Genomic Med.* 9 (3), e1609. doi:10.1002/mgg3.1609

Wonkam, A. (2021). Sequence Three Million Genomes across Africa. *Nature* 590, 209–211. doi:10.1038/d41586-021-00313-7

Yadava, S. M., and Ashkinadze, E. (2017). 125: Whole Exome Sequencing (WES) in Prenatal Diagnosis for Carefully Selected Cases. *Am. J. Obstet. Gynecol.* 216, S87–S88. doi:10.1016/j.ajog.2016.11.029

Yang, Y., Muzny, D. M., Xia, F., Niu, Z., Person, R., Ding, Y., et al. (2014). Molecular Findings Among Patients Referred for Clinical Whole-Exome Sequencing. *JAMA* 312 (18), 1870–1879. doi:10.1001/jama.2014.14601

Zhang, P., Luo, H., Li, Y., Wang, Y., Wang, J., Zheng, Y., et al. (2021). NyuWa Genome Resource: A Deep Whole-Genome Sequencing-Based Variation Profile and Reference Panel for the Chinese Population. *Cell Reports* 37 (7), 110017. doi:10.1016/j.celrep.2021.110017

Conflict of Interest: Authors NvdM is employed by FamGen Counselling, Bloemfontein, South Africa.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Genetics

ORIGINAL RESEARCH
published: 28 June 2022
doi: 10.3389/fgene.2022.843705

Molecular Profiling of Kenyan Acute Myeloid Leukemia Patients

*Mercy Gatua[1,2], Mohsen Navari[3,4,5], Matilda Ong'ondi[2], Noel Onyango[6], Serah Kaggia[7], Emily Rogena[7], Giuseppe Visani[8], Nicholas A. Abinya[6] and Pier Paolo Piccaluga[1,7,9,10]**

[1]Biobank of Research, IRCCS S. Orsola-Malpighi Academic Hospital, Department of Experimental, Diagnostic and Specialty Medicine, Institute of Hematology and Medical Oncology "L. and A. Seràgnoli", University of Bologna School of Medicine, Bologna, Italy, [2]Kenyatta National Hospital, Nairobi, Kenya, [3]Department of Medical Biotechnology, School of Paramedical Sciences, Torbat Heydariyeh University of Medical Sciences, Torbat Heydariyeh, Iran, [4]Bioinformatics Research Group, Mashhad University of Medical Sciences, Mashhad, Iran, [5]Research Center of Advanced Technologies in Medicine, Torbat Heydariyeh University of Medical Sciences, Torbat Heydariyeh, Iran, [6]Nairobi Hospital, University of Nairobi, Nairobi, Kenya, [7]Department of Pathology, School of Medicine, Jomo Kenyatta University of Agriculture and Technology, Juja, Kenya, [8]Hematology and Stem Cell Transplantation, AORMN, Pesaro, Italy, [9]Istituto Euro-Mediterraneo di Scienza e Tecnologia (IEMEST), Palermo, Italy, [10]Nanchang University, Nanchang, China

Acute myeloid leukemia (AML) is an infrequent disease, and it is associated with high morbidity and mortality. It harbors a unique configuration of cytogenetic abnormalities and molecular mutations that can be detected using microscopic and molecular methods respectively. These genetic tests are core elements of diagnosis and prognostication in high-income countries. They are routinely incorporated in clinical decision making, allowing for the individualization of therapy. However, these tests are largely inaccessible to most patients in Kenya and therefore no data has been reported on this group of patients. The main purpose of this study is to describe the cytogenetic and molecular abnormalities of acute myeloid leukemia patients seen at the hemato-oncology unit of Kenyatta National Hospital. A cross-sectional descriptive study was carried out over a 3-month period on ten patients with a diagnosis of AML. Social demographics and clinical data were collected through a study proforma. A peripheral blood sample was collected for conventional metaphase G-banding technique and next generation sequencing. Particularly, targeted DNA sequencing (Illumina myeloid panel) and whole exome sequencing (WES) were performed. Cytogenetic analysis failed in 10/10 cases. Targeted sequencing was successfully obtained in 8 cases, whereas WES in 7. Cytogenetic studies yielded no results. There were 20 mutations detected across 10 commonly mutated genes. All patients had at least one clinically relevant mutation. Based on ELN criteria, NGS identified three patients with high-risk mutations, affecting *TP53* (n = 2) and *RUNX1* (n = 1). One patient was classified as favorable (*PML-RARA*) while 4 were standard risk. However, *WT1* mutations associated with unfavorable prognosis were recorded in additional 2 cases. WES showed concordant results with targeted sequencing while unveiling more mutations that warrant further attention. In conclusion, we provide the first molecular profiling study of AML patients in Kenya including application of advanced next generation sequencing technologies, highlighting current limitations of AML diagnostics and treatment while confirming the relevance of NGS in AML characterization.

Keywords: acute myeloid leukemia, next generation sequencing, illumina, myeloid panel, ELN, cytogenetics, targeted therapy

OPEN ACCESS

Edited by:
Maritha J Kotze,
Stellenbosch University, South Africa

Reviewed by:
Hamid Dolatshad,
University of Oxford, United Kingdom
Rispah Torrorey,
Moi University, Kenya

***Correspondence:**
Pier Paolo Piccaluga
pierpaolo.piccaluga@unibo.it

Specialty section:
This article was submitted to Human and Medical Genomics, a section of the journal Frontiers in Genetics

Received: *26 December 2021*
Accepted: *13 May 2022*
Published: *28 June 2022*

Citation:
Gatua M, Navari M, Ong'ondi M, Onyango N, Kaggia S, Rogena E, Visani G, Abinya NA and Piccaluga PP (2022) Molecular Profiling of Kenyan Acute Myeloid Leukemia Patients. Front. Genet. 13:843705. doi: 10.3389/fgene.2022.843705

INTRODUCTION

Acute Myeloid Leukemia (AML) is an infrequent disease (De Kouchkovsky and Abdul-Hay, 2016). Although treatment outcomes continue to improve over time, AML is still a significant cause of mortality (Fiegl, 2016). Despite an improvement in the treatment associated mortality, chemo-resistance and post-transplant disease relapse, account for one of the most challenging aspects of AML management (Döhner et al., 2015). It is a heterogeneous clonal disorder that arises from a malignant myeloid stem cell that has acquired genetic and epigenetic mutations that have accumulated in a stepwise fashion. These acquired genetic alterations cause proliferative and survival advantage with reduced apoptosis leading to a buildup of abnormal, poorly differentiated neoplastic cells in the blood and bone marrow with resultant suppression of the normal hematopoietic process (Meyer and Levine, 2014; Döhner et al., 2015).

In Sub-Saharan Africa, acute leukemia causes high mortality. In a 2012 population-based study, the age-standardized rates in East Africa were 3.8 and 3.4 per 100,000 in men and women respectively (Miranda-Filho et al., 2018). Latest Globocan data estimates that in Kenya, the incidence rates of leukemia are 4.8 and 4.5/100,000 in men and women respectively and is listed among the top ten causes of cancer mortality (World Health Organization, 2020). Despite an increasing disease burden in Sub-Saharan Africa, there is limited infrastructure and finances that deters the use of recommended genetic testing (Gopal et al., 2012; Wanjiku et al., 2018). Cytogenetic and molecular techniques, which are core elements of diagnosis in the developed world, are nonexistent in most Sub-Saharan countries (Gopal et al., 2012). Significant financial challenges do exist in emerging economies, where majority of healthcare costs are personal expenditures with many falling below the poverty line. This occurs even in those emerging countries with highly skilled specialists and state of the art facilities that mirror those in developed countries (Philip et al., 2015).

AML cells harbor a unique configuration of cytogenetic and molecular mutations that involve critical genes that regulate the normal hematopoietic process (Löwenberg and Rowe, 2016). This accounts for the phenotypic heterogeneity of the disease (Lagunas-Rangel et al., 2017). Understanding the pathobiology of AML has provided a framework for risk stratification, development of novel treatment approaches with individualization of therapy as well as detection of post treatment minimal residual disease (Döhner and Gaidzik, 2011; Stein, 2015; Dombret and Gardin, 2016; Stein and Tallman, 2016). The urgent need for therapeutic advancement has come at the backdrop of a dismal 5-year overall survival of 50% and 20% for those below and above 60 years old, respectively, with traditional cytotoxic therapies (Stein and Tallman, 2016). A landmark novel anti-leukemic agent should successfully eradicate the malignant founding clone and its sub-clones, eradicating a potential niche for recurrence (Löwenberg and Rowe, 2016). WHO classification of AML incorporates clinical features, morphological assessment of bone marrow specimens, cytochemical studies, immune-phenotyping, cytogenetic and molecular testing to distinguish distinct biological subgroups with clinical importance (Arber et al., 2016). The major categories included in the 2016 WHO classification include; AML with recurrent genetic abnormalities, AML with myelodysplasia-related changes, Therapy-related myeloid neoplasms, AML Not otherwise specified, Myeloid Sarcoma and Myeloid proliferations related to Down syndrome (Arber et al., 2016).

Cytogenetic abnormalities are analyzed using conventional metaphase G-banding techniques and fluorescence *in situ* hybridization (FISH) whereas molecular mutations are detected using next generation sequencing molecular methods (He et al., 2015; Muhammad Ilyas et al., 2015). These technologies are able to comprehensively identify genetic lesions that are critical in the process of leukamogenesis. These genetic abnormalities are the single most powerful prognostic factors and risk stratifies the patient into Favorable, Intermediate and Adverse risk groups (Döhner et al., 2017). Prognostic classification is critical in the management of AML patients, particularly in respect to establishing those with poor prognostic features who are likely to relapse or have chemo-resistant disease. It's also important for category-specific treatment (Kumar, 2011). However, the prognostic impact of these genetic groups may change with targeted therapy (Gill et al., 2016).

Data from an ongoing prospective study by Prof N. A. Othieno-Abinya (2013 onwards) shows that about 30 patients are diagnosed with AML annually. The median age at diagnosis is 30 years with a male: female ratio of 1.2:1. AML diagnosis at the hemato-oncology unit of KNH is mainly through morphological assessment of bone marrow specimens with few or none of the patients undergoing karyotyping or molecular assessment.

In this study, we therefore aimed to identify the cytogenetic and molecular abnormalities found in patients' diagnosed with AML at the adult hemato-oncology unit of KNH and prognosticate them according to the EuropeanLeukemiaNet risk stratification model.

METHODS

Case Selection

The study included patients diagnosed and treated at the adult hemato-oncology unit of Kenyatta National Hospital, the biggest public referral hospital in Nairobi, Kenya. The molecular analyses were carried on at the Department of Experimental, Diagnostic, and Specialty Medicine, Bologna University Italy.

Institutional consent was obtained from the Department of Clinical Medicine and Therapeutics, University of Nairobi (UON) and Ethics and Research Committee of KNH. Request for shipment of the samples for cytogenetic and molecular tests, was obtained from the Ministry of Health subject to the fulfillment of the requirements of KNH-ERC. Material Transfer Agreement between the two institutions was obtained. All patients were informed of the study and a consent/assent obtained in either English/Kiswahili, the national languages of Kenya. For patients aged below 18 years, a legal representative (parent/guardian) gave the consent.

TABLE 1 | Panel of genes tested.

Hot Spot genes	Full genes	Fusion driver genes
ABL1	*ASXL*	*ABL*
BRAF	*BCOR*	*BCL2*
CBL	*CALR*	*BRAF*
CSF3R	*CEPBA*	*ALK*
DNMT3A	*ETV6*	*CCND1*
FLT3	*EZH2*	*CREBBP*
GATA2	*IKZF1*	*EGFR*
HRAS	*NF1*	*ETV6*
GATA2	*PHF6*	*FGFR2*
IDH1	*PRPF8*	*FGFR1*
IDH2	*RB1*	*FUS*
JAK2	*RUNX1*	*HMGA2*
KIT	*SH2B3*	*JAK2*
K-RAS	*STAG2*	*KMT2A*
MPL	*TET2*	*MECOM*
MYD88	*TP53*	*MET*
NPM1	*ZRSR2*	*MLLTI0*
N-RAS		*MYBL1*
PTPN11	*MYH11*	
SETBP1		*NTRK3*
SRSF2	*NUP214*	
U2AF1		*PDGFRA*
WT1	*PDGFRB*	
		RARA
		RBM15
		RUNX1
		TCF3
		TFE3

Cytogenetic Analysis

Cell culture: The cells were cultured using a protocol for harvesting chromosomes from whole blood. In brief, 0.25mls of fresh whole blood was collected in 10 mls of RPMI media containing L-glutamine (20% fetal bovine serum, 1% Penicillin/streptomycin, 1% fungizone, and 1% PHA) and incubated for 48 h at 37°C with 5% CO_2. Cell counting prior to harvesting of cells revealed that the final seeding densities for each of the samples were less than 1 × 106/ul below the optimum of 1-3 × 106/ul for adequate metaphases. Harvesting was done according to standard protocol and all 10 samples had no evaluable metaphases (Hastings et al., 2013; Howe et al., 2014).

Next Generation Sequencing

Total DNA was extracted from 8 samples (MG2, MG4i, MG5, MG5i, MG6, MG6i, MG7, MG8) with QIAamp DNA mini kit Qiagen according to the manufacturer's procedure (Qiagen, Italy). Qubit was used for DNA quality control assessment (ThermoFisher, Italy).

Thereafter, based on Illumina's TruSeq DNA Sample Preparation, DNA libraries were pre-enriched, according to manufacturer's instructions (Illumina, United States). Quant-it PicoGreen dsDNA Assay Kit was eventually used for libraries quantification, according to the manufacturer's protocol (Invitrogen, Life Technologies, United States).

Using Illumina iSeq2500 (Illumina, San Diego, United States), we sequenced the paired-end libraries (2 × 150 base pair), following the manufacturer's instructions. On average, about 82.9 million 151 bp PF reads were generated, and the theoretical coverage was c143.2, calculated based on hg19 RefSeq non redundant exome length; the median target coverage at 50× was 87.2% (range, 60.7–87.5%). Details on sequencing statistics are described in **Supplementary Table S1**.

FastQC V0.10.0 tool (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) was used for quality control, The reads were mapped to Homo sapiens (UCSC hg19) as reference genome, using Burrows-Wheeler Aligner version 2.12.0, while the targeted regions defined by 11062019_ALLEXONV7-NEW-TXT (Li and Durbin, 2009).

To remove potential PCR duplicates, we used SAMtools command rmdup to detect and collapse multiple mapped reads pairs with identical external coordinates (Li and Durbin, 2009). Mapping quality score recalibration and local realignment around insertions and deletions (InDels) was performed using Genome Analysis Toolkit (GATK—v1.6-23-gf0210b3) (McKenna et al., 2010). Single-nucleotide variants (SNVs) and small insertions and deletions (InDels) were called separately using GATK Unified-Genotyper.

All the mutations detected were filtered using thresholds based on quality, coverage, and strand of the mapped reads and according to variants already present in public databases (Hapmap, dbSNP and 1000genome project [The 1000 Genomes Project 2010].

Targeted DNA sequencing utilized the AmpliSeq for illumina myeloid panel. It's a targeted panel that investigates 62 genes associated with myeloid cancers (**Table 1**). Library preparation and sequencing were performed as previously described, according to the manufacturer instructions (AmpliSeq for Illumina Myeloid Panel, 2021). Lacking high quality RNA, the gene expression part of the assay was omitted.

Raw NGS data are available on request (Prof. Piccaluga).

Clinically relevant lesions were defined according to ESMO, NCCN, FDA, and EMA guidelines, as relevant for prognostic and diagnostic significance, or therapy (Li et al., 2017; Heuser et al., 2020; Pollyea et al., 2021).

RESULTS

Patients Characteristics

Fifteen AML patients were treated at Kenyatta National Hospital during the study period. Ten consecutives gave the consent and were enrolled. DNA was successfully collected and analyzed in 8/10 cases. The main clinical characteristics of those 8 patients are summarized in **Table 2**. Briefly, their mean age was 35 years (13-60); 5/8 were males; mean WBC count was 14.3 × 10^9/L (0.77–20.92); average Hb level was 6.65 g/dl (3.1–10.8); average PLT count was 26.65 × 10^9/L (6-53). Only one secondary case was reported, previously affected and treated for aplastic anemia.

NGS Revealed Clinically Relevant Genetic Lesions in All Patients

Cytogenetic analysis technically failed in all instances, probably due to the latency between sample collection and analysis (which

TABLE 2 | Patient characteristics.

Case	Age (years)	Gender	% of Marrow Blasts	WBC ($\times 10^9$/L)	Hb (g/dl)	Platelets (×109/L)	Prior Cytotoxic therapy Or Radiotherapy	Antecedent Hematological Disorder
MG2	19	M	30	20.92	6.8	17	NONE	NONE
MG5i	13	F	62	0.77	7.3	46	NONE	NONE
MG4i	50	F	82	44.8	8.2	20	NONE	NONE
MG5	36	M	40	2.86	5.8	11	NONE	APLASTIC ANEMIA
MG6	34	F	44	2.63	4.3	6	NONE	NONE
MG6i	60	M	75	2.60	3.1	44	NONE	NONE
MG7	53	M	70	36	10.8	53	NONE	NONE
MG8	13	M	60	3.49	6.9	16	NONE	NONE

PDF report ID	MG2	MG4i	MG5	MG5i	MG6	MG6i	MG7	MG8
PML-RARA				1				
TP53					1			1
WT1	2						1	1
DNMT3A		1						
IDH2		1*				1		
FLT3-TKD		1*						
NRAS	1						1	1
RUNX1			1					
CEBPA				1			1	
STAG2					2*		1*	
Total	**3**	**3**	**1**	**2**	**3**	**1**	**4**	**3**

FIGURE 1 | Mutational analysis by NGS—Clinically relevant mutations/Translocations. * not detected at WES.

requires alive cells). DNA quality was, by contrast assessed and confirmed by Qubit (median green RFU 27190.09; range 9697.21–13,262.96).

As far as NGS analysis was concerned, Among the 62 analyzed genes, 20 clinically relevant (as reported by previous publications and common databases) mutations across 10 genes were detected. Of these, 8 were missense mutations, 9 were frame-shift insertions, 1 was a nonsense mutation, and 1 was a fusion gene. The remaining one, affecting FLT3, was a synonymous change. However, it was included among clinically relevant as accepted for accessing clinical trials with FLT3 inhibitors. All patients had at least one clinically relevant mutation with 1/8 cases showing 2 mutations, 4/8 showing 3 mutations each, 2/8 showing 1 mutation each and 1/8 showing 4 (**Figure 1**).

The most mutated gene was *WT1* (4 mutations in 3 cases), followed by *NRAS* and *STAG2* (3 lesions in 3 cases and 3 lesions in 2 cases, respectively), *TP53*, *IDH2*, and *CEBPA* (2 lesions in 2 cases), and *DNMT3A*, *FLT3*, *RUNX1* (1 lesion). One patient presented with *PML/RARA* rearrangement.

Pairwise mutations to assess for co-mutations could not be undertaken due to the small sample size.

Despite cytogenetics unavailability, NGS results allowed to identify one patient with favorable genetics (*PML/RARA*+), 3 likely to be high risk (*RUNX1* and *TP53*), and 4 standard-risk. Among these, detection of mutations in *IDH2* (N = 2 patients) and *FLT3* (N = 1) allowed to candidate those patients to specific targeted treatments.

NGS Revealed Clinically Relevant Genetic Lesions of Unknown Significance

In addition to clinically relevant lesions, a series of additional genetic lesions, the significance of which is still undefined, have been identified (**Supplementary Table S1**).

Briefly, 7/8 patients showed at least one single nucleotide variant (SNV) affecting *SH2B3*, 6/8 *ASXL1* and *TET2* genes, 3/8 *TP53*, 2/8 *ETV6*, *FLT3*, *NF1*, *PTPN11*, while 1/8 *CEBPA*, *GATA2*, *IKZF1*, *KIT*, *PRPF8, RB1,* and *SF3B1.* Despite not being reported as clinically relevant, those disrupting mutations occurred in genes well-known to be associated with myeloid malignancies; therefore, it is conceivable that they might have some pathogenic activity.

Whole Exome Sequencing Unveiled SNVs Never Reported in AML

Following targeted sequencing of myeloid malignancies associated genes (see above), we sought to perform whole exome sequencing (WES) to further describe the genetic

TABLE 3 | Clinical outcome.

\Patients ID	MG2	MG5i	MG4i	MG5	MG6	MG6i	MG7	MG8
Induction treatment	3 + 7	Dead before treatment	Dead before treatment	3–+7	Dead before treatment	Azax9	3–+7	3–+7
Outcome	DDI (d10)	—	—	DDI (d30)	—	Stable disease	RES	DDI (d15)
OS days	40	30	30	60	30	Alive with disease	90	45

DDI, death during induction.
RES, Resistance.
3 + 7, daunorubicin plus cytarabine conventional chemotherapy regimen.

landscape of Kenyan AML patients. The procedure was successful in 7/8 cases.

First, WES confirmed the presence of the clinically relevant lesions observed at targeted sequencing, with the only exceptions of *STAG2* mutations that were not detected in both patients supposed to carry them.

In addition, several SNVs and SNPs were identified. Most of them were not likely associated with cancer and were filtered out. By contrast, a couple of patients showed additional lesions that warrants attention. Patient MG_2 showed an overall quite complex scenario, characterized by several SNVs affecting other 12 genes whose association with cancer (or anti-cancer drug response) is well known. Particularly, *AURKA*, which is associated to colon cancer and was recently found to be over expressed also in hematological malignancies, as well as the drug resistant associated genes *ABCB1, XRCC1* and *CBR3* turned out to be affected. Furthermore, we observed mutations affecting *FGFR4*, *KLC1* and *XRCC3*, associated to colon cancer and melanoma (**Supplementary Table S2**). Interestingly, *NAT2* might be involved in epigenetic deregulation. Finally, *TP53* turned out to be largely altered, though the identified SNVs are not formally associated with leukemias.

In patient MG_5, known for the previous history of aplastic anemia (AA), we found a *TERC* mutation, known to be associated with AA pathogenesis. By contrast, *TERT*, the other gene associated with AA, presented a synonymous mutation only.

Clinical Correlates

The limited sample size didn't allow a proper clinic-pathological correlation. Furthermore, it should be noted that 3/8 patients died before any treatment, including the young acute promyelocytic leukemia patient. Among the remaining 5 patients, 3 died from infection during induction (day 10, day 15 and day 30, respectively), 1 was resistant to induction and died during supportive treatment shortly after, and 1 is alive with disease after treatment with azacytidine (9 courses) (**Table 3**).

Overall, this scenario is representative of the major current limitation in Kenyan hospitals, meaning the delay in diagnosis and treatment initiation when AML patients are often septic and in very poor clinical conditions.

DISCUSSION

This study looked at the cytogenetic and molecular abnormalities among AML patients presenting to the hemato-oncology unit of Kenyatta National Hospital. There was no reportable data on the karyotype status of the patients as no cells were cultured from peripheral blood. Increased transit time to the laboratory was the most likely cause for this failure, confirming the lack of feasibility for metaphase cytogenetics in this setting. Of note, despite cytogenetics being obviously relevant for AML prognostication, NGS analysis allowed us to identify lesions that could assign patients to the high risk ELN group independently from karyotyping. We sought to combine WES and targeted DNA sequencing to ensure the highest sensitivity and specificity for detecting mutations associated to myeloid malignancies but still retaining the capability of identify a broader spectrum of mutations by WES.

The study demonstrated, in fact, that patients with AML in KNH do have deleterious mutations that are well-known to be associated with AML pathogenesis. In this regard, although the limited sample size didn't allow a proper statistical evaluation and cannot be extended to all African cases, the mutational spectrum seemed not significantly different from what has been reported in Western Countries series. Nonetheless, consistent with the overall poor clinical outcome, genetic lesions associated with unfavorable outcome seemed quite common. Three out of eight patients were classified as high risk according to the ELN score for the presence of *RUNX1* (N = 1 patient) and *TP53* (N = 2) mutations. Four were recorded as standard risk, and 1 as favorable (*PML/RARA*+) (Döhner et al., 2017). If also considering *WT1* as adverse risk factor (Virappane et al., 2008; Renneville et al., 2009; Hou et al., 2010), 2 additional patients could be regarded as high risk. Finally, we observed a significantly high occurrence of SNVs affecting *TP53* (overall 5/8 patients) and *ASXL1* (5/8 patients). Even if the specific SNVs were not yet associated with clinical relevance, they cannot be excluded as having potential deleterious role. Similarly, genetic lesions potentially affecting epigenetic regulation (one of the main mechanisms of myeloid malignancies transformation) were common, even if the specific SNVs are not currently associated with a clinical phenotype. Seven out of eight patients presented with *TET2* SNVs, while *DNMT3A* was affected in one out of eight patients.

Overall, the study population seemed to reflect a slightly different scenario from what is commonly observed in AML series. Patients were younger (median age around 30 years vs 68) and the overall treatment response quite poor. Despite younger mean age (43 excluding children), however, only two patients aged below 18 were studied and therefore the series have to be regarded as referring to adults. Pediatric cases definitely need further investigation. This is in line, however, with what was observed in a large series of patients treated at Kenyatta National

Hospital and Nairobi Hospital (N. O. Abinya, manuscript in preparation).

This is the first study exploring in depth, the molecular features of Kenyan AML patients, while very few are available on African AML in general. Kappala et al. in South Africa using a microarray-based assessment of molecular variables on AML patients noted that 5.7% had inv (6)/t (16; 16) (p13; q22), 11.3% had t (8; 21) (q22; q22), 3.8% had t (15; 17) (q24; q21), 1.9% had double mutant *CEBPA*, 9.4% had *NPM1*-ABD mutations and 18.9% had a high expression of *EVI1*. *NPM1* mutations were reported at a lower frequency whereas *EVI1* over expression occurred at a higher frequency compared to world data, which would indicate age and racial differences (Kappala et al., 2017). Lower prevalence of *NPM1* and *FLT3*-ITD mutations compared to world data were documented by Marshall et al. looking at a South African cohort with *de novo* AML (Marshall et al., 2014). Awad et al. investigated the prevalence of *FLT3*-ITD mutations of 346 patients with AML in Egypt. 9.2% had t (8; 21), 2.3% had inv (16) and <1% had both t (9; 11) and inv (3) (102). *FLT3*-ITD mutations were present in 18.5% of the total population mirroring the lower frequency of *FLT3* mutation in African studies (Adnan-Awad et al., 2017). Shamaa et al. in Egypt noted a frequency of 34.6 and 28.8% in *FLT3*-ITD and *NPM1* mutations respectively in a cohort of AML patients with a normal karyotype, like western studies. *DNMT3A* and *IDH1*-R132 are frequently mutated in AML patients in Egypt whereas *TET2* overexpression is not a frequent finding (Hamed et al., 2015; Salem et al., 2017; El Gammal et al., 2019). The cytogenetic and molecular patterns of acute myeloid leukemia patients in Africa are difficult to elucidate and compare due to lack of large-scale studies and differing study designs. In India, a country with similar demographics, a large scale analysis of the cytogenetic profile of patients with *de novo* acute myeloid leukemia showed that 15% had t (8; 21), 9% had t (15; 17), 8% had 8+, 6% had −7/del 7q, 5% had *KMT2A* rearrangements, 4.4% had inv (16)/t (16/16), 3% had −5/del 5q, 2% has −17/abn17p and 1.5% had inv (3) in order of decreasing frequency (Amare et al., 2016). Other studies from India show that *FLT3*-ITD, *CEBPA* and *NPM1* mutations occur in 22.3, 8.3 and 8% of patients respectively, a frequency that's lower than that reported in western data (Ahmad et al., 2012; Khera et al., 2017).

In our study, we could associate targeted DNA sequencing and WES. The concordance between the two was remarkably high, as expected based on the relative specificity and the use of similar chemistry. Only one gene, *STAG2*, turned out to be mutated at targeted sequencing but not at WES in 2 cases. Not being possible to apply a third independent method, we could only speculate that, considering the high specificity of Illumina myeloid panel and the possible lack of sensitivity of WES, it's indeed more likely that exon capturing didn't cover the regions of *STAG2* affected by mutations.

On the other hand, WES allowed to detect many SNVs associated with associated comorbidities (e.g., hypercholesterolemia, data not presented). In one case, WES added significant information concerning the molecular pathogenesis of leukemic cells; SNVs potentially associated with cellular transformation as well as drug resistance were, in fact, detected (MG_2). The affected genes included *AURKA*, *ABCB1* and *CBR3* (the two latter associated with drug resistance), *XRCC3, FGFR4, KLC1*, and *NAT2*. Aurora Kinase A (AURKA) has been documented to have some oncogenic activity in the microenvironment milieu of leukemic cells, necessitating search for small molecule inhibitors with anti-leukemic activity (Wang et al., 2020; Du et al., 2021). While MG_2 died during induction he had mutations detected in the *ABCB1* and *CBR3* gene. Evidence have shown that higher expression of *ABCB1* gene (a member of the ATP binding cassette transporters family) results in an increased efflux of chemotherapeutic agents, resulting in drug resistance (Thorn et al., 2011; Shaffer et al., 2012), whereas genetic polymorphisms of *CBR3* (anthracycline metabolizing enzyme) results in differing pharmacokinetics that influence treatment efficacy (Bains et al., 2010). In addition to the aforementioned, mutations in FGFR4 which is a tyrosine kinase receptor, *XRCC3*, *KLC1* and *NAT2* were detected in the same patient. Dysregulated activation of this TKR has been reported to be a significant oncogenic pathway in various solid tumors (Helsten et al., 2016; Liu et al., 2020). Genotypic variants of the DNA repair gene, X-ray cross complementary group 3 (*XRCC3*) have been associated with a significantly higher cancer risk including AML (Bănescu et al., 2013), whereas polymorphisms of *NAT2* an acetylator with epigenetic influences have been described as a modifier of tumorigenesis in various solid tumors (; Zhu et al., 2021). In addition, WES allowed identification of a pathogenic *TERC* mutation in a patient previously affected by AA (Young, 2018; Brzeźniakiewicz-Janus et al., 2020). In this contest, the evolution to AML was accompanied by the acquisition of *RUNX1* mutations.

The present study carries some limitations. First, in this exploratory study, the patients' population was small and limited to one center, and proper statistical evaluations couldn't be performed. Second, the paucity of available material didn't allow to match NGS studies with other techniques exploring both gene mutations (as validation) or protein expression. Indeed, the two applied system validated each other somehow, but further studies are warranted. Finally, despite filtering the analyses for African ethnicity, we cannot exclude that some population-specific genetic lesions could not be captured; in fact, Africans are still underrepresented in genetic databases.

In conclusion, we provided the first study on NGS molecular profiling on Kenyan AML patients, highlighting current limitations of AML diagnostics and treatment in this setting and confirming the relevance of this approach in AML characterization.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Department of Clinical Medicine and Therapeutics,

University of Nairobi (UON) and Ethics and Research Committee of Kenyatta National Hospital. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

MG, NA, MO, and PP conceptualized and designed the project. MG, MO, and NA did the clicical care of patients. MG, MN, MO, NO, SK, ER, GV, and PP analyzed the data. All the authors read and approved the manuscript.

FUNDING

This work was supported by AIL-Pesaro Onlus (2019-2021, GV), BolognAIL (2020, PP), RFO DIMES (2015–2020, PP), FIRB Futura 2011 RBFR12D1CB (PP).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.843705/full#supplementary-material

REFERENCES

Adnan-Awad, S., Gaber, O., Eltokhy, S. A., Mourad, D., Amer, M., Kandeel, E. Z., et al. (2017). FLT3-ITD Mutations in Egyptian Patients of Acute Myeloid Leukemia: Correlation with Cytogenetic, FAB Subgroups and Prognosis. *Clin. Lab.* 63 (5), 1027–1034. doi:10.7754/Clin.Lab.2017.170121

Ahmad, F., Rajput, S., Mandava, S., and Das, B. R. (2012). Molecular Evaluation ofCEBPAGene Mutation in Normal Karyotype Acute Myeloid Leukemia: A Comparison of Two Methods and Report of NovelCEBPAMutations from Indian Acute Myeloid Leukemia Patients. *Genet. Test. Mol. Biomarkers* 16 (7), 707–715. doi:10.1089/gtmb.2011.0317

Amare, P. S. K., Jain, H., Kabre, S., Deshpande, Y., Pawar, P., Banavali, S., et al. (2016). Cytogenetic Profile in 7209 Indian Patients with <i>De Novo</i> Acute Leukemia: A Single Centre Study from India. *Jct* 07, 530–544. doi:10.4236/jct.2016.77056

AmpliSeq for Illumina Myeloid Panel *Documentation [Internet]*. [cited 2021 Dec 18]. Available from: https://support.illumina.com/sequencing/sequencing_kits/ampliseq-for-illumina-myeloid-panel/documentation.html.

Arber, D. A., Orazi, A., Hasserjian, R., Thiele, J., Borowitz, M. J., Le Beau, M. M., et al. (2016). The 2016 Revision to the World Health Organization Classification of Myeloid Neoplasms and Acute Leukemia. *Blood* 127 (20), 2391–2405. doi:10.1182/blood-2016-03-643544

Bains, O. S., Karkling, M. J., Lubieniecka, J. M., Grigliatti, T. A., Reid, R. E., and Riggs, K. W. (2010). Naturally Occurring Variants of Human CBR3 Alter Anthracycline *In Vitro* Metabolism. *J. Pharmacol. Exp. Ther.* 332 (3), 755–763. LP – 763. Available from: http://jpet.aspetjournals.org/content/332/3/755.abstract. doi:10.1124/jpet.109.160614

Bănescu, C., Tilinca, M., Benedek, E. L., Demian, S., Macarie, I., Duicu, C., et al. (2013). XRCC3 Thr241Met Polymorphism and Risk of Acute Myeloid Leukemia in a Romanian Population. *Gene* 526 (2), 478–483. Gene [Internet]Available from: https://www.sciencedirect.com/science/article/pii/S037811191300704X. doi:10.1016/j.gene.2013.05.054

Brzeźniakiewicz-Janus, K., Rupa-Matysek, J., and Gil, L. (2020). Acquired Aplastic Anemia as a Clonal Disorder of Hematopoietic Stem Cells. *Stem Cell. Rev. Rep.* 16 (3), 472–481. [Internet]Available from: https://pubmed.ncbi.nlm.nih.gov/32270433. doi:10.1007/s12015-020-09971-y

De Kouchkovsky, I., and Abdul-Hay, M. (2016). 'Acute Myeloid Leukemia: a Comprehensive Review and 2016 Update', *Blood Cancer J.*, 6, e441. doi:10.1038/bcj.2016.50

Döhner, H., Estey, E., Grimwade, D., Amadori, S., Appelbaum, F. R., Büchner, T., et al. (2017). Diagnosis and Management of AML in Adults: 2017 ELN Recommendations from an International Expert Panel. *Blood* 129 (4), 424–447. doi:10.1182/blood-2016-08-733196

Döhner, H., and Gaidzik, V. I. (2011). Impact of Genetic Features on Treatment Decisions in AML. *ASH Educ. Progr B* 2011 (1), 36–42.

Döhner, H., Weisdorf, D. J., and Bloomfield, C. D. (2015). Acute Myeloid Leukemia. *N. Engl. J. Med.* 373 (12), 1136–1152.

Dombret, H., and Gardin, C. (2016). An Update of Current Treatments for Adult Acute Myeloid Leukemia. *Blood* 127 (1), 53–61. doi:10.1182/blood-2015-08-604520

Du, R., Huang, C., Liu, K., Li, X., and Dong, Z. (2021). Targeting AURKA in Cancer: Molecular Mechanisms and Opportunities for Cancer Therapy. *Mol. Cancer* 20 (1), 15. doi:10.1186/s12943-020-01305-3

El Gammal, M. M., Ebid, G. T., Madney, Y. M., Abo-Elazm, O. M., Kelany, A. K., Torra, O. S., et al. (2019). Clinical Impact of Combined Mutations in DNMT3A, FLT3-ITD and NPM1 Among Egyptian Acute Myeloid Leukemia Patients. *Clin. Lymphoma Myeloma Leuk.*19(6):e281-e290. doi:10.1016/j.clml.2019.02.001

Fiegl, M. (2016). "Epidemiology, Pathogenesis, and Etiology of Acute Leukemia," in *Handbook of Acute Leukemia*. 25, 3–13.doi:10.1007/978-3-319-26772-2

Gill, H., Leung, A. Y., and Kwong, Y.-L. (2016). Molecularly Targeted Therapy in Acute Myeloid Leukemia. *Future Oncol.* 12 (6), 827–838. doi:10.2217/fon.15.314

Gopal, S., Wood, W. A., Lee, S. J., Shea, T. C., Naresh, K. N., Kazembe, P. N., et al. (2012). Meeting the Challenge of Hematologic Malignancies in Sub-saharan Africa. *Blood* 119 (22), 5078–5087. doi:10.1182/blood-2012-02-387092

Hamed, N., El Halawani, N., Kassem, H., Ayad, M., and Dammag, E. (2015). TET2 Expression in a Cohort of Egyptian Acute Myeloid Leukemia Patients. *Egypt J. Haematol.* 40 (4), 159.

Hastings, R., Howell, R., Betts, D., Porter, S., Haferlach, C., Dastugue, N., et al. (2013). *Guidelines and Quality Assurance for Acquired CytogeneticsA*. Permanent Working Group for Cytogenetics and Society Index.

He, R., Wiktor, A. E., Hanson, C. A., Ketterling, R. P., Kurtin, P. J., Van Dyke, D. L., et al. (2015). Conventional Karyotyping and Fluorescence *In Situ* Hybridization. *Am. J. Clin. Pathol.* 143 (6), 873–878. doi:10.1309/ajcpp6lvmqg4lnck

Hein, D. W., Doll, M. A., Fretland, A. J., Leff, M. A., Webb, S. J., Xiao, G. H., et al. (2000). Molecular Genetics and Epidemiology of the NAT1 and NAT2 Acetylation Polymorphisms. *Cancer Epidemiol. Biomarkers & Prev. [Internet]* 9 (1), 29–42. Available from: http://cebp.aacrjournals.org/content/9/1/29.abstract.

Helsten, T., Elkin, S., Arthur, E., Tomson, B. N., Carter, J., and Kurzrock, R. (2016). The FGFR Landscape in Cancer: Analysis of 4,853 Tumors by Next-Generation Sequencing. *Clin. Cancer Res.* 22 (1), 259–267. Clin Cancer Res [Internet]LP – 267. Available from: http://clincancerres.aacrjournals.org/content/22/1/259.abstract. doi:10.1158/1078-0432.ccr-14-3212

Heuser, M., Ofran, Y., Boissel, N., Brunet Mauri, S., Craddock, C., Janssen, J., et al. (2020). Acute Myeloid Leukaemia in Adult Patients: ESMO Clinical Practice Guidelines for Diagnosis, Treatment and Follow-Up. *Ann. Oncol.*, 31. 697–712. doi:10.1016/j.annonc.2020.02.018

Hou, H.-A., Huang, T.-C., Lin, L.-I., Liu, C.-Y., Chen, C.-Y., Chou, W.-C., et al. (2010). WT1 Mutation in 470 Adult Patients with Acute Myeloid Leukemia: Stability during Disease Evolution and Implication of its Incorporation into a Survival Scoring System. *Blood* 115 (25), 5222–5231. doi:10.1182/blood-2009-12-259390

Howe, B., Umrigar, A., and Tsien, F. (2014). Chromosome Preparation from Cultured Cells. *J. Vis. Exp.* 83, e50203. doi:10.3791/50203

Kappala, S. S., Alessandrini, M., Matlhako, T., Beltchev, E., Pool, R., and Pepper, M. S. (20172017). Application of the AMLprofiler Diagnostic Microarray in the South African Setting. *Stem Cells Int.* 2017, 1–8. doi:10.1155/2017/2560191

Khera, R., Ahmed, F., Mundada, M., Devi, S., Murthy, S., Lavanya, N., et al. (2017). Multiplex Approach in Classification, Diagnosis, and Prognostication in Acute

Myeloid Leukemia: An Experience from Tertiary Cancer Center in South India. *Indian J. Med. Paediatr. Oncol.* 38 (3), 266. doi:10.4103/ijmpo.ijmpo_89_16

Kumar, C. C. (2011). Genetic Abnormalities and Challenges in the Treatment of Acute Myeloid Leukemia. *Genes. & Cancer* 2 (2), 95–107. doi:10.1177/1947601911408076

Lagunas-Rangel, F. A., Chávez-Valencia, V., Gómez-Guijosa, M. Á., and Cortes-Penagos, C. (2017). Acute Myeloid Leukemia-Genetic Alterations and Their Clinical Prognosis. *Int. J. Hematol. Oncol. Stem Cell. Res.* 11 (4), 328–339.

Li, H., and Durbin, R. (2009). Fast and Accurate Short Read Alignment with Burrows-Wheeler Transform. *Bioinformatics* 25 (14), 1754–1760. doi:10.1093/bioinformatics/btp324

Li, M. M., Datto, M., Duncavage, E. J., Kulkarni, S., Lindeman, N. I., Roy, S., et al. (2017). Standards and Guidelines for the Interpretation and Reporting of Sequence Variants in Cancer. *J. Mol. Diagnostics* 19 (1), 4–23. doi:10.1016/j.jmoldx.2016.10.002

Liu, Y., Cao, M., Cai, Y., Li, X., Zhao, C., and Cui, R. (2020). Dissecting the Role of the FGF19-FGFR4 Signaling Pathway in Cancer Development and Progression. *Front. Cell. Dev. Biol. [Internet]*. ;8:95. Available from: https://pubmed.ncbi.nlm.nih.gov/32154250. doi:10.3389/fcell.2020.00095

Löwenberg, B., and Rowe, J. M. (2016). Introduction to the Review Series on Advances in Acute Myeloid Leukemia (AML). *Blood* 127 (1), 1.doi:10.1182/blood-2015-10-662684

Marshall, R. C., Tlagadi, A., Bronze, M., Kana, V., Naidoo, S., Wiggill, T. M., et al. (2014). Lower Frequency of NPM1 and FLT3-ITD Mutations in a South African Adultde novoAML Cohort. *Int. Jnl. Lab. Hem.* 36 (6), 656–664. doi:10.1111/ijlh.12204

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., et al. (2010). The Genome Analysis Toolkit: a MapReduce Framework for Analyzing Next-Generation DNA Sequencing Data. *Genome Res.* 20 (9), 1297–1303. Available from: https://pubmed.ncbi.nlm.nih.gov/20644199. doi:10.1101/gr.107524.110

Meyer, S. C., and Levine, R. L. (2014). Translational Implications of Somatic Genomics in Acute Myeloid Leukaemia. *Lancet Oncol.* 15 (9), e382–e394. doi:10.1016/s1470-2045(14)70008-7

Miranda-Filho, A., Piñeros, M., Ferlay, J., Soerjomataram, I., Monnereau, A., and Bray, F. (2018). Epidemiological Patterns of Leukaemia in 184 Countries: a Population-Based Study. *Lancet Haematol.* 5 (1), e14–e24. doi:10.1016/s2352-3026(17)30232-6

Muhammad Ilyas, A., Ahmad, S., Faheem, M., Imran Naseer, M., Kumosani, T. A., Hussain Al-Qahtani, M., et al. (2015). Next Generation Sequencing of Acute Myeloid Leukemia: Influencing Prognosis, *BMC Genomics*, 16. doi:10.1186/1471-2164-16-s1-s5

Philip, C., George, B., Ganapule, A., Korula, A., Jain, P., Alex, A. A., et al. (2015). Acute Myeloid Leukaemia: Challenges and Real World Data from India. *Br. J. Haematol.* 170 (1), 110–117. doi:10.1111/bjh.13406

Pollyea, D. A., Bixby, D., Perl, A., Bhatt, V. R., Altman, J. K., Appelbaum, F. R., et al. (2021). NCCN Guidelines Insights: Acute Myeloid Leukemia, Version 2.2021. *J. Natl. Compr. Canc Netw.* 19 (1), 16–27. doi:10.6004/jnccn.2021.0002

Renneville, A., Boissel, N., Zurawski, V., Llopis, L., Biggio, V., Nibourel, O., et al. (2009). Wilms Tumor 1 Gene Mutations Are Associated with a Higher Risk of Recurrence in Young Adults with Acute Myeloid Leukemia. *Cancer* 115 (16), 3719–3727. doi:10.1002/cncr.24442

Salem, D., El-Aziz, S. A., El-Menshawy, N., Abouzeid, T., and Ebrahim, M. (2017). Prevalence and Prognostic Value of IDH1 R132 Mutation in Newly Diagnosed AML Egyptian Patients with Normal Karyotype. *Indian J. Hematol. Blood Transfus.* 33 (1), 49–55. doi:10.1007/s12288-016-0649-z

Shaffer, B. C., Gillet, J. P., Patel, C., Baer, M. R., Bates, S. E., and Gottesman, M. M. (2012). Drug Resistance: Still a Daunting Challenge to the Successful Treatment of AML. *Drug Resist Updat* 15 (1–2), 62–69. Available at: https://pubmed.ncbi.nlm.nih.gov/22409994. doi:10.1016/j.drup.2012.02.001

Stein, E. M. (2015). Molecularly Targeted Therapies for Acute Myeloid Leukemia. *Hematol. Am. Soc. Hematol. Educ. Progr* 2015 (1), 579–583. doi:10.1182/asheducation-2015.1.579

Stein, E. M., and Tallman, M. S. (2016). Emerging Therapeutic Drugs for AML. *Blood* 127 (1), 71–78. doi:10.1182/blood-2015-07-604538

Thorn, C. F., Oshiro, C., Marsh, S., Hernandez-Boussard, T., McLeod, H., Klein, T. E., et al. (2011). Doxorubicin Pathways. *Pharmacogenet Genomics [Internet* 21 (7), 440–446. Available from: https://pubmed.ncbi.nlm.nih.gov/21048526. doi:10.1097/fpc.0b013e32833ffb56

Virappane, P., Gale, R., Hills, R., Kakkas, I., Summers, K., Stevens, J., et al. (2008). Mutation of the Wilms' Tumor 1 Gene Is a Poor Prognostic Factor Associated with Chemotherapy Resistance in Normal Karyotype Acute Myeloid Leukemia: The United Kingdom Medical Research Council Adult Leukaemia Working Party. *Jco* 26 (33), 5429–5435. doi:10.1200/jco.2008.16.0333

Wang, J.-D., Zhang, W., Zhang, J.-W., Zhang, L., Wang, L.-X., Zhou, H.-S., et al. (2020). A Novel Aurora Kinase Inhibitor Attenuates Leukemic Cell Proliferation Induced by Mesenchymal Stem Cells. *Mol. Ther. - Oncolytics* 18, 491–503. –503. Available from. doi:10.1016/j.omto.2020.08.001

Wanjiku, C. M., Saliba, A., and Njiru, E. W. (2018). Acute Myeloid Leukemia: Does One Size Fit All? A Retrospective Analysis of Outcomes of Therapy at Moi Teaching and Referral Hospital in Eldoret, Kenya. *Jco* 36 (15_Suppl. l)–e18529. doi:10.1200/JCO.2018.36.15_suppl.e18529

World Health Organization, *Cancer Today* (2020), Available at: https://gco.iarc.fr/today/data/factsheets/populations/404-kenya-fact-sheets.pdf.

Young, N. S. (2018). Aplastic Anemia. *N. Engl. J. Med.* 379 (17), 1643–1656. [Internet]Available from: https://pubmed.ncbi.nlm.nih.gov/30354958. doi:10.1056/nejmra1413485

Zhu, C., Wang, Z., Cai, J., Pan, C., Lin, S., Zhang, Y., et al. (2021). VDR Signaling via the Enzyme NAT2 Inhibits Colorectal Cancer Progression. *Front. Pharmacol.* 12, 727704. Front Pharmacol [Internet]Available from: https://pubmed.ncbi.nlm.nih.gov/34867333. doi:10.3389/fphar.2021.727704

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Genetics

TYPE Review
PUBLISHED 29 September 2022
DOI 10.3389/fgene.2022.940661

OPEN ACCESS

EDITED BY
Dawn Stephens,
Technology Innovation Agency (TIA), South Africa

REVIEWED BY
Susana Alicia Pesoa,
LACE Laboratorios, Argentina
Milcah Dhoro,
University of Zimbabwe, Zimbabwe

*CORRESPONDENCE
Derseree Archary,
Desh.Archary@caprisa.org

SPECIALTY SECTION
This article was submitted to Human and Medical Genomics, a section of the journal Frontiers in Genetics

RECEIVED 10 May 2022
ACCEPTED 15 September 2022
PUBLISHED 29 September 2022

CITATION
Zondo NM, Sobia P, Sivro A, Ngcapu S, Ramsuran V and Archary D (2022), Pharmacogenomics of drug transporters for antiretroviral long-acting pre-exposure prophylaxis for HIV.
Front. Genet. 13:940661.
doi: 10.3389/fgene.2022.940661

Pharmacogenomics of drug transporters for antiretroviral long-acting pre-exposure prophylaxis for HIV

Nomusa M. Zondo[1,2], Parveen Sobia[1,2], Aida Sivro[1,2], Sinaye Ngcapu[1,2], Veron Ramsuran[1,2] and Derseree Archary[1,2]*

[1]Centre for the AIDS Programme of Research in South Africa (CAPRISA), Mucosal Immunology Department, Durban, South Africa, [2]Department of Medical Microbiology, School of Laboratory Medicine and Medical Sciences, University of KwaZulu-Natal, Durban, South Africa

The use of antiretrovirals (ARVs) as oral, topical, or long-acting pre-exposure prophylaxis (PrEP) has emerged as a promising strategy for HIV prevention. Clinical trials testing Truvada® [tenofovir disoproxil fumarate (TDF)/tenofovir (TFV) and emtricitabine (FTC)] as oral or topical PrEP in African women showed mixed results in preventing HIV infections. Since oral and topical PrEP effectiveness is dependent on adequate drug delivery and availability to sites of HIV infection such as the blood and female genital tract (FGT); host biological factors such as drug transporters have been implicated as key regulators of PrEP. Drug transporter expression levels and function have been identified as critical determinants of PrEP efficacy by regulating PrEP pharmacokinetics across various cells and tissues of the blood, renal tissues, FGT mucosal tissues and other immune cells targeted by HIV. In addition, biological factors such as genetic polymorphisms and genital inflammation also influence drug transporter expression levels and functionality. In this review, drug transporters and biological factors modulating drug transporter disposition are used to explain discrepancies observed in PrEP clinical trials. This review also provides insight at a pharmacological level of how these factors further increase the susceptibility of the FGT to HIV infections, subsequently contributing to ineffective PrEP interventions in African women.

KEYWORDS

African women, PrEP (pre-exposure prophylaxis), female genital tract (FGT), drug transporters, single nucleotid polymorphism (SNP), inflammation

1 Introduction

HIV remains a formidable public health challenge, with currently over 37.7 million people living with HIV globally (UNAIDS, 2021). At the end of 2020 globally, an estimated 1.5 million new HIV infections were reported, of which 60% occurred in sub-Saharan African (SSA) countries (UNAIDS, 2021). The use and coverage of (antiretrovirals) ARVs has, however, had a positive impact in many SSA countries

including South Africa. Despite the advent and use of ARVs, new HIV infections (~4,000 daily) continue and remain a major concern (UNAIDS, 2021).

The increasing challenge of providing ARVs to a rapidly growing HIV population prompts the need for new interventions to decrease HIV incidence rates (Nicol et al., 2018). Previously tested HIV prevention methods have included the use of ARVs as oral, topical gels or long-acting pre-exposure prophylaxis (PrEP) formulations in uninfected individuals (Abdool Karim et al., 2010; Janes et al., 2018; Nicol et al., 2018). Clinical trials using oral and topical PrEP regimens in high risk heterosexual HIV-serodiscordant couples (Baeten et al., 2012; Thigpen et al., 2012) and men who have sex with men (MSM) (Grant et al., 2010; Molina et al., 2015; Mccormack et al., 2016) reported high levels of protection against HIV acquisition ranging from 44 to 86% (Grant et al., 2010; Thigpen et al., 2012; Mccormack and Dunn, 2015; Molina et al., 2015). However, clinical trials using the same PrEP regimens that focused primarily on at-risk African women, produced inconsistent levels of protection against HIV, ranging from −49 to 39%, with a majority leading to trial termination (Abdool Karim et al., 2010; Van Damme et al., 2012; Marrazzo et al., 2015; Delany-Moretlwe et al., 2018).

The main contributory factor for these low efficacies was identified as low to no adherence to PrEP. However, underlying biological factors beyond adherence have been proposed to play an integral role in low PrEP efficacies (Hu et al., 2015; Nicol et al., 2018). These include drug transporters, which are transmembrane proteins that are expressed ubiquitously in various cells and tissues of the body. Various ARVs used as PrEP have been identified as substrates of different drug transporters (Hu et al., 2015; Taneva et al., 2016; Reznicek et al., 2017). Therefore, drug transporter expression levels and functionality are considered essential for optimal PrEP delivery and for maintaining optimal drug concentrations in cells and tissues targeted by HIV (Hu et al., 2015; Nicol et al., 2018). In addition, there are also host biological factors such as inflammation (Saib and Delavenne, 2021) and genetic polymorphisms (Shenfield, 2004; Arruda et al., 2016) affecting drug transporter disposition, which subsequently affects drug efficacy (Shenfield, 2004; Arruda et al., 2016; Saib and Delavenne, 2021). These findings underscore drug transporters as critical determinants of drug pharmacokinetics. However, there is limited data on drug transporter expression profiles and host factors affecting drug transporter expression and function in anatomical compartments such as the FGT, the predominant site for HIV infection in women during heterosexual intercourse (Hu et al., 2015; Nicol et al., 2018). This warrants the need for further studies that will evaluate these factors, especially in high-risk groups such as African women. The current review, therefore, evaluates various biological factors affecting PrEP pharmacokinetics [absorption, distribution, metabolism and excretion (ADME)] to better understand inconsistencies in PrEP effectiveness observed in clinical trials with at-risk African women.

2 Biological, behavioural, and socio-economic factors that increase women's susceptibility to HIV

Despite noticeable reductions in HIV infections and increases in ARVs accessibility, there are several biological, behavioural and social factors that contribute to higher HIV prevalence rates in women (Ramjee and Daniels, 2013; Abdool Karim et al., 2020). Socio-economic factors that drive high HIV incidence rates in women include sexual abuse, lack of education, lack of food security and the lack of proper social services such as education on HIV and insufficient provision of health services; especially in highly affected regions (Abdool Karim et al., 2012; Ramjee and Daniels, 2013; Nicol et al., 2018; Durevall et al., 2019).

Behavioural factors also play an integral role in high rates of HIV acquisition in young women. These include early age of sexual debut (Mabaso et al., 2018), multiple concurrent sex partners, intergenerational sexual partnering with older men and transactional sexual encounters (Maartens et al., 2014; De Oliveira et al., 2017; Mabaso et al., 2018). Other factors include low marriage rates (Alcaide et al., 2014), intravaginal practices, and low to no condom use due to the inability to negotiate safe sexual practices with their male partners (Ramjee and Daniels, 2013; De Oliveira et al., 2017; Mabaso et al., 2018). Additionally, the use of injectable drugs and alcohol have also been associated with increased HIV transmission through shared needles and high-risk sexual behaviour, respectively (Maartens et al., 2014). Together, these factors suggest that the economic and social disempowerment of young women especially in a developing country such as South Africa contributes largely to high HIV prevalence rates within this population.

Apart from behavioural and socio-economic factors that fuel HIV infections, biological factors also drive higher rates of HIV infections in women. The greater mucosal surface area of the female genital tract (FGT) makes this surface highly susceptible through increased opportunities for target $CD4^+$ T cells to become infected with HIV and other sexually transmitted infections (STIs) during sexual intercourse (Ramjee and Daniels, 2013). Other biological factors that increase women's susceptibility to HIV include bacterial vaginosis (BV) (Heffron et al., 2017; Klatt et al., 2017), vaginal micro-abrasions (Stanley, 2009), cervical ectopy (Critchlow et al., 1995) and genital inflammation (Masson et al., 2015; Mckinnon et al., 2018). Additionally, the use of long-acting injectable progestin hormonal contraceptives (particularly DMPA) has also been associated with increased women's susceptibility to HIV, however, this remains a topic of ongoing debate, with some studies showing an increased HIV risk (Heffron et al., 2012; Hapgood, 2020) while, others showed no differences (Myer et al., 2007; Shen et al., 2017).

TABLE 1 PrEP clinical trials demonstrating various efficacies in high-risk populations from different regions.

Clinical trials	Study population (regions)	PrEP drugs	PrEP efficacy -reduction in HIV incidence (%)	References
CAPRISA 004	African women (South Africa)	1% TFV gel	39	Abdool Karim et al. (2010)
Partners PrEP	Heterosexual couples (Kenya and Uganda)	Oral TDF-FTC	75	Baeten et al. (2012)
		Oral TDF alone	67	
TDF2	Heterosexual couples (Botswana)	Oral TDF-FTC	62	Thigpen et al. (2012)
iPrEx	MSM (South America, the United States, South Africa, and Thailand)	Oral TDF-FTC	44	Grant et al. (2010)
PROUD	MSM (England)	Oral TDF-FTC	86	McCormack and Dunn, (2015)
IPERGAY	MSM (France and Canada)	Oral TDF-FTC	86	Molina et al. (2015)

3 PrEP in HIV prevention

Since the main route of HIV infection in women is through sexual intercourse, many prevention strategies are aimed at protecting the FGT (Celum, 2011; Nicol et al., 2018). As a result, many different modalities have been tested, which include a vaginal ring containing dapivirine (Baeten et al., 2016), various microbicide gel formulations such as Carraguard (Skoler-Karpoff et al., 2008) and PRO2000 vaginal gels (McCormack et al., 2010) and spermicide gel formulations such as nonoxynol-9 (N-9) (Wilkinson et al., 2002). Besides topical gels and rings, implants containing tenofovir alafenamide (TAF) (Gunawardana et al., 2015) and long-acting injectable formulations containing for example cabotegravir (CAB) (Landovitz et al., 2018) are currently being tested for HIV prevention in PrEP clinical trials. Truvada® which is the co-formulated tenofovir disoproxil fumarate (TDF)/TFV and emtricitabine (FTC)) is a licensed oral PrEP drug that is also used for HIV prevention (Alvarez et al., 2011).

The concept of using TFV as PrEP in preventing HIV infections was initially investigated in macaques with simian immunodeficiency virus (SIV) (Person et al., 2012). During the 1990s, studies on TFV (a nucleotide reverse transcriptase inhibitor of HIV) previously known as (R)-9-(2-phosphonylmethoxypropyl)adenine (PMPA) demonstrated evidence of complete protection against SIV infections (Tsai et al., 1995; Person et al., 2012). The success observed in animal models was, however not fully translated in human HIV prevention clinical trials (Person et al., 2012). One of the major factors that attributed to these inconsistencies was low to no adherence, which limited PrEP exposure to tissues and cells targeted by HIV in areas such as FGT (Rohan and Sassi, 2009; Hu et al., 2015).

4 PrEP clinical trials: Efficacy in African women

Significant breakthroughs in using PrEP to prevent HIV infections have been observed in PrEP trials focused-on high-risk HIV-serodiscordant heterosexual couples and MSM (Baeten et al., 2012; Thigpen et al., 2012) (Table 1). The Partners PrEP trial, performed in HIV-serodiscordant heterosexual couples in Kenya and Uganda showed significant HIV reductions of 75 and 67% respectively, with oral TDF-FTC and with TDF alone (Baeten et al., 2012). The TDF2 trial which evaluated the effectiveness of TDF-FTC drugs in sexually active HIV negative heterosexual adults from Botswana; showed that TDF-FTC prevented new HIV infections, by demonstrating a 62% reduction in HIV incidence (Thigpen et al., 2012). In the Pre-exposure Prophylaxis Initiative trial (iPrEX) in MSM from South America, the United States, South Africa and Thailand, a daily single oral dose of Truvada® demonstrated a 44% reduction in HIV incidence (Grant et al., 2010). Similarly, in other MSM European studies testing Truvada® the Pragmatic open-label randomised trial of pre-exposure prophylaxis (PROUD) (Mccormack and Dunn, 2015) and the On-Demand Antiretroviral Pre-exposure Prophylaxis for HIV Infection (IPERGAY) in men who have sex with men (MSM) (Molina et al., 2015), both showed an 86% reduction in HIV incidence (Mccormack and Dunn, 2015; Molina et al., 2015). Results from these studies further supported the effectiveness of PrEP among MSM who are at risk of acquiring HIV (Mccormack and Dunn, 2015; Molina et al., 2015). Currently, in African women, the CAPRISA 004 trial remains the only trial that showed an overall 39% efficacy with a topical gel containing ARV (Table 1) (Abdool Karim et al., 2010). Furthermore, when the data was stratified according to

TABLE 2 PrEP clinical trials demonstrating low PrEP efficacies in high-risk populations from different regions.

Clinical trials	Study population (regions)	PrEP drugs	PrEP efficacy -reduction in HIV incidence %	References
FEM-PrEP	African women (South Africa, Kenya and Tanzania)	Oral TDF-FTC	4.7%	Van Damme et al. (2012)
VOICE	African women (South Africa, Uganda and Zimbabwe)	Oral TDF-FTC	−4.4%	Marrazzo et al. (2015)
		Oral TDF alone	−49%	
		1% TFV gel	14.5%	
FACTS-001	African women (South Africa)	1% TFV gel	6.52%	Delany-Moretlwe et al. (2018)

degree of adherence, women with high adherence had a corresponding reduction in HIV incidence by 54%, while women with low adherence had only a 28% reduction in HIV incidence (Abdool Karim et al., 2010).

Other PrEP clinical trial studies that focused primarily on at-risk African women demonstrated inconsistent levels of protection showing HIV incidence of −49% to 14.5% (Table 2). These include the FEM-PrEP trial (Van Damme et al., 2012) which evaluated, daily Truvada® in women from high-risk areas in South Africa, Kenya and Tanzania. This trial was, however, terminated following low efficacy largely attributed to lack of adherence (Corneli et al., 2016) and low drug concentrations in the FGT (Abdool Karim et al., 2011). The Vaginal and Oral Interventions to Control the Epidemic (VOICE) Microbicide Trial Network (MTN 003) trial was also conducted in women from high HIV prevalence areas in South Africa, Uganda and Zimbabwe (Marrazzo et al., 2015). Women were randomised to either oral TDF, oral TDF-FTC, TFV gel, or respective oral or vaginal placebos. Similar to the FEM-PrEP study, results from this trial showed no efficacy (Marrazzo et al., 2015). The moderate success of the topical CAPRISA 004 1% TFV gel trial led to the Follow on African Consortium for Tenofovir Studies (FACTS-001) (Delany-Moretlwe et al., 2018). The study was conducted at nine community-based clinical trial sites where it assessed the safety and efficacy of the precoitally applied 1% TFV gel in high-risk South African women. Here too, the FACTS-001 trial showed no significant reduction in HIV incidence between the active arm and the control (Delany-Moretlwe et al., 2018).

A potential explanation for the disparities in efficacy observed in these PrEP clinical trials could be due to differential drug penetration levels in the rectal compared to the vaginal mucosal tissues (Patterson et al., 2011; Janes et al., 2018). Data from Cottrell et al. (2015), observed that similar levels of adherence of two doses per week of Truvada® reduced HIV incidence by 90% in the MSM population of the iPrEX study, whereas in heterosexual women populations of the FEM-PrEP and VOICE studies low to no protection was observed (Cottrell et al., 2015). Additionally, the complex composition of the vagina's microbiome and inflammation may affect PrEP disposition in women (Patterson et al., 2011; Klatt et al., 2017; Janes et al., 2018; Mckinnon et al., 2018). These findings urge the need to better understand the mechanisms of drug availability and metabolism within the area of vulnerability, the FGT (Rohan and Sassi, 2009; Hu et al., 2015).

5 Compartmental heterogeneity in PrEP drug disposition

The FGT is a highly active and diverse immune environment, with a wide range of heterogeneous immune cells such as macrophages, dendritic cells, Langerhans cells, natural-killer cells, B and T-cells, making it highly susceptible to HIV infections (Hu et al., 2015; Nicol et al., 2018). The increased vulnerability of the FGT to HIV infections is also largely ascribed to the presence of a single columnar epithelium cell layer in the endocervix as opposed to the multi-layered squamous epithelium of the ectocervix (Nicol et al., 2018). These cell layers are vulnerable to micro-abrasions caused by friction during heterosexual intercourse allowing for easy access of HIV (Nicol et al., 2018). A previous study by Shen et al. (2014) also observed that even with intact epithelium, vaginal myeloid dendritic cells expressing HIV receptors can facilitate the capture and dissemination of HIV into the deeper mucosal tissue layers (Shen et al., 2014). The vulnerability of the FGT prompts the need for HIV prevention interventions that provide sufficient protection within all compartments exposed to the virus (Nicol et al., 2018).

Interventions such as PrEP should therefore provide optimal ARV drug concentrations that are sufficient in preventing HIV viral entry, transcription, and replication in HIV targeted cells (Cottrell et al., 2015). Drug exposure of PrEP in the FGT has previously shown variability in drug delivery and availability (Cottrell et al., 2015). To understand this variability; a 14-days open-labelled study by Patterson et al. (2011), demonstrated varying concentration levels of TFV, FTC and their respective active metabolites TFV-DP and FTC triphosphate (FTC-TP) in rectal, vaginal and cervical tissues, following a single dose of Truvada®. In rectal tissues, TFV and TFV-DP were detected throughout the 14-day period and concentrations were 100-fold higher when compared to cervical and vaginal tissues; while FTC

concentrations were 10 to 15- fold higher in vaginal and cervical tissue when compared to rectal tissue. The active metabolite FTC-TP was, however, only detected for 2 days in all tissues (Patterson et al., 2011). A prior study also compared ARVs drug exposure in cervicovaginal fluid and blood. In the FGT, NRTIs lamivudine (3TC), zidovudine (ZDV), FTC and TFV exhibited high drug concentrations relative to the blood. However, non-nucleoside reverse transcriptase inhibitor (NNRTI) efavirenz (EFV) and protease inhibitors (PIs) lopinavir (LPV) and atazanavir (ATV) exhibited low concentrations in the FGT when compared to the blood (Dumond et al., 2007). These results indicate the heterogeneity of drug disposition and their respective metabolites within sub-compartments of the same anatomical surface or different compartments (Patterson et al., 2011). These studies suggested that certain ARVs may or may not be good PrEP candidates according to their ability to penetrate certain areas (Dumond et al., 2007). In addition, these data further underscore the importance of understanding factors affecting drug pharmacokinetics in tissues that are highly susceptible to HIV infections.

These variations show differential drug penetration levels, giving an insight into the varying levels of protection against HIV observed in some PrEP trials (Cottrell et al., 2015). These discrepancies may be due to the interplay between these PrEP drugs and various membrane-bound proteins that mediate drug transport and availability (Cottrell et al., 2015). For example ARVs such as TFV, FTC and ZDV have been previously shown to be substrates of drug transporters P-glycoprotein (P-pg), multi-drug resistance protein-1 (MRP-1) and organic anion transporters-1 (OAT-1), respectively (Kis et al., 2010; Hu et al., 2015). These data indicated that intracellular and extracellular ARV drug levels can be predominantly regulated by certain drug transporters (Kis et al., 2010; Hu et al., 2015). Therefore, understanding the distribution and biological characteristics of drug transporters may help further define their roles in affecting PrEP efficacy.

6 Drug transporters involved in PrEP pharmacokinetics

Drug transporters are types of transmembrane proteins that are ubiquitously expressed in the human body in areas such as the gastrointestinal tract, in epithelial cells in the FGT, lungs, blood-brain barrier, endothelial cells, and liver cells (Zhou et al., 2013; Arruda et al., 2016). Drug transporters comprise of two superfamilies: the ATP-binding cassette (ABC) and Solute Carrier (SLC) proteins (Hu et al., 2015). The ABC proteins are a large family of efflux pumps that bind ATP and utilize its hydrolysis energy to transport molecules across and out of the cell membrane. This family comprises seven subfamilies of which three are the most relevant in the efflux of PrEP drugs (Petzinger and Geyer, 2006; Hu et al., 2015). These include (Figure 1):

i) the *ABCB* subfamily that comprises P-glycoprotein (P-gp),
ii) the *ABCC* subfamily that comprises multidrug and toxin extrusion proteins (MATE) and MRPs, and,
iii) the *ABCG* subfamily that comprises breast cancer resistance protein (BCRP) (Petzinger and Geyer, 2006; Hu et al., 2015; Nicol et al., 2018). The SLC proteins influx or uptake molecules across and into the cell membrane *via* ATP energy dependant carriers or through an electrochemical gradient (Lin et al., 2015). Subfamilies that are the most relevant in the uptake or influx of PrEP drugs include (Figure 1):

i) the *SLC22* subfamily that comprises OAT and organic cation transporters (OCTs),
ii) the *SLC28* subfamily that comprises concentrative nucleoside transporters (CNTs),
iii) the *SLC29* subfamily that comprises equillibrative transporters (ENTs) and
iv) the *SLCO* subfamily that comprises organic anion-transporting polypeptides (OATPs) (Petzinger and Geyer, 2006; Hu et al., 2015; Nicol et al., 2018).

Most of these drug transporters are localised on polarized cells, and regulate substrate distribution on the apical or basolateral surfaces of cells, contributing to the pharmacokinetics of several ARVs (Sissung et al., 2012). In previous studies, TFV and FTC have been shown as substrates of these ABC and SLC drug transporters (Hu et al., 2015; Nicol et al., 2018). This data indicated that the delivery and absorption of these drugs in cells is facilitated by drug transporters, establishing their emerging role as critical determinants in drug pharmacokinetics. This interaction has been noted especially in HIV target cells such as the immune cells of the FGT that include macrophages, vaginal epithelial cells, T cells, and dendritic cells expressing CD4 receptors (Hu et al., 2015; Taneva et al., 2016), and also in peripheral blood mononuclear cells (PBMCs) and epithelial cells of the intestinal and renal system (Kerb, 2006).

6.1 Role of drug transporters in the FGT

The extracellular accumulation of TFV and FTC in cells overexpressing certain drug transporters has been demonstrated in various studies focused on the FGT. Findings from these studies suggested that the delivery of effective PrEP drug concentrations to cells and tissues in the FGT is highly associated with the mRNA expression level and functionality of drug transporters (Grammen et al., 2014; Nicol et al., 2014; Hijazi et al., 2015; Taneva et al., 2016). Zhou et al. (2013) showed varying drug transporter expression levels in the FGT (vaginal and ectocervix tissues) and liver. mRNA expression was defined as ≤2% (undetectable), 2%–10% (low expression), 10%–50% (moderate expression) and 50%–100% (high expression)

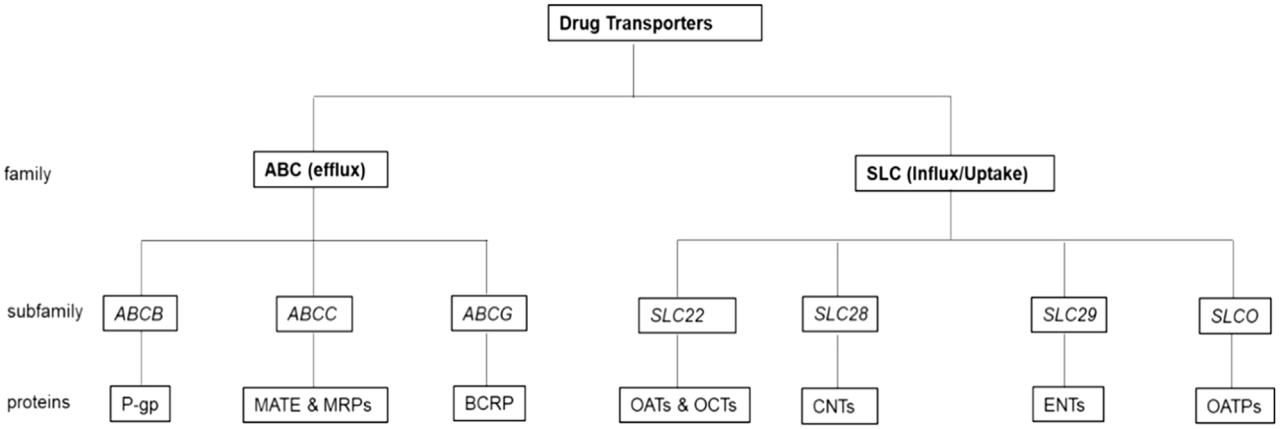


FIGURE 1
Drug transporters most relevant in the efflux and influx of PrEP drugs. In the ABC (efflux) family there are three-drug transporter subfamilies along with their respective proteins that are the most relevant in the efflux of PrEP drugs, while in the SLC (influx/uptake) family, there are four drug transporter subfamilies along with their respective proteins that are the most relevant in the uptake or influx of PrEP.

(Zhou et al., 2013). In the FGT, high mRNA expression of drug transporters (MRP-1, MRP-4, P-gp, BCRP, ENT-1 and OCT-2) was observed when compared to the liver, while MRP-2 and influx drug transporters OAT-1 and OAT-3 showed moderate and low mRNA expression as compared to the liver (Zhou et al., 2013). Similarly, a study by Taneva et al. (2016) also showed significantly low expression of uptake drug transporters OAT-1 and OAT-3 in vaginal epithelial cells and T-cells, which accounted for the poor permeability of TFV across the cell membranes and into the cells. Additionally, this study showed that the *in-vitro* transfection of T cells with the drug transporter OAT-1 increased TFV uptake, resulting in high intracellular drug accumulation (Taneva et al., 2016). These studies indicated that there is variability in drug transporter expression levels within different tissues. Therefore, analysing the expression levels of drug transporters in the FGT could aid in better understanding their role in the pharmacokinetics of drugs (Zhou et al., 2013; Taneva et al., 2016).

Nicol et al. (2014) showed high mRNA expression levels of efflux drug transporters P-gp and MRP-2 in vaginal tissues compared to colorectal tissue, while MRP-4 was only highly expressed in colorectal tissues. In contrast, uptake drug transporters OAT-1, OAT-3 and OATP1B1 exhibited extremely low to no expression in colorectal and vaginal tissues, respectively (Nicol et al., 2014). Additionally, immunohistochemistry that informed on the localisation of these drug transporters revealed high protein expression of P-gp and MRP-2 in vaginal epithelial cells compared to colorectal epithelial cells, while low to no protein expression of OAT-1 was observed in colorectal epithelial and vaginal cells, respectively (Nicol et al., 2014). Differences in protein localisation and expression suggested an increased expression of efflux drug transporters in vaginal tissues compared to colorectal tissues (Nicol et al., 2014). These data show that more drug is pumped out of cells in the vagina, while an increased expression of uptake drug transporters in colorectal tissues promoted an uptake of drugs (Nicol et al., 2014). These findings highlighted that inter-tissue variability in drug transporter expression may contribute to the greater intracellular accumulation of ARVs such as TFV and maraviroc in colorectal tissues compared to vaginal tissues (Nicol et al., 2014). High expression levels of efflux drug transporters P-gp, MRP-2 and BCRP in vaginal and endocervical tissues was also reported by Grammen et al. (2014). The study further established in intestinal cell lines-Caco-2 and vaginal epithelial cell lines-SiHa using specific drug transporter inhibitors, that ARV drugs darunavir, maraviroc and saquinavir are substrates of efflux drug transporters P-gp and MRP-2, which are likely to contribute to lower intracellular levels of these respective drugs (Grammen et al., 2014).

To further understand the role of drug transporters in the mucosal compartment, the relationship between the accumulation of topically applied PrEP drugs dapivirine, darunavir and TFV, and the expression of drug transporters was characterised in cervicovaginal cell lines (Hijazi et al., 2015). These included HeLa cell lines, VK2/E6E7, Ect1/E6E7 and End1/E6E7 derived from human cervical epithelial adenocarcinoma, primary vaginal, ectocervical and endocervical epithelial cells, respectively (Hijazi et al., 2015). Tenofovir significantly downregulated the mRNA expression of MPR5 in VK2/E6E7, while dapivirine significantly upregulated most MRP drug transporters in all cell lines. Darunavir stimulation also significantly upregulated the uptake drug transporter CNT3 in

all cells, while MRP3 was only significantly unregulated in VK2/E6E7 cell line (Hijazi et al., 2015). This characterisation by Hijazi et al. (2015) provided insight not only on the type of drug transporters present in the FGT but also how drug transporter disposition may be altered by the presence of certain drugs; which could assist in the assessment of ARV pharmacokinetics in the FGT. Furthermore, these findings could assist in the determination of suitable PrEP drug formulations that could provide sufficient drug concentrations to susceptible tissues and cells of the FGT (Hijazi et al., 2015).

6.2 Role of drug transporters in peripheral blood mononuclear cells

A study by Turriziani et al. (2008) determined the mRNA expression levels of drug transporters in PBMCs (isolated from buffy coats) from HIV infected individuals failing ARV therapy and HIV negative individuals. The mRNA expression levels of P-gp, MRP (-1,-4 and -5) was significantly higher in HIV infected individuals compared to HIV negative individuals (Turriziani et al., 2008). A higher inter-individual mRNA expression variability was also observed in HIV infected individuals, indicating a correlation between the presence of ARVs and drug transporter expression levels (Turriziani et al., 2008). Similarly, Bousquet et al. (2009) investigated if the singular or combined (dual or triple) use of TFV, FTC and EFV on PBMCs isolated from healthy donors disrupts mRNA drug transporter expression levels (Bousquet et al., 2009). Following a 20-h *in-vitro* incubation, a singular use of FTC induced MRP5, while TFV reduced MRP (-1,-5,-6) and P-gp mRNA expression in PBMCs (Bousquet et al., 2009). FTC was also shown to exhibit an inhibitory effect on the mRNA expression of efflux drug transporter MRP-1 in a dose-responsive manner. These findings suggest a correlation between the presence of FTC with MRP-1 expression (Bousquet et al., 2008). The use of ZDV was also previously shown to be associated with the upregulation of efflux drug transporters MRP-1 and MRP5 expressed on PBMCs (Jorajuria et al., 2004). Findings from these studies showed that an interaction between ARVs and drug transporters may alter drug transporter disposition by affecting mRNA expression levels; subsequently affecting intracellular drug accumulation (Jorajuria et al., 2004; Bousquet et al., 2009).

Contrary to these studies, Falasca et al. (2011) and Giraud et al. (2010) showed no correlations between the mRNA expression levels of efflux drug transporters and the presence of ARVs ritonavir (RTV), ATV, and LPV in PBMCs isolated from HIV infected patients (Giraud et al., 2010; Falasca et al., 2011). The study found no variation in the mRNA expression levels of P-gp and MRP drug transporters before and after ARV intake (Falasca et al., 2011). However, in a more recent study, a significant association was observed between ARVs and drug transporters P-gp, BCRP, MRP-1, ENT-2 and OCT-1 expressed on monocytes and monocyte-derived macrophages isolated from HIV negative individuals and HIV infected individuals receiving ARV therapy containing either abacavir, ATV, EFV, rilpivirine, TFV, 3TC, FTC, elvitegravir, dolutegravir, and cobicistat (Hoque et al., 2021). These findings showed that these associations could lead to sub-optimal intracellular drug concentrations, subsequently allowing HIV infections in HIV negative individuals or further HIV replication in HIV infected individuals (Hoque et al., 2021).

6.3 Role of drug transporters in the renal system

The entry of TFV into epithelial cells of the kidney tubule is mediated by the uptake drug transporters OAT-1 and OAT-3 expressed on its basolateral membrane (Cihlar et al., 2007), while the efflux of TFV into urine is mediated by efflux drug transporter MRP-4 expressed at the apical side of renal proximal tubules (Ray et al., 2006). These data together provide evidence that TFV is a substrate of OAT-1, OAT-3 and MRP-4 drug transporters expressed in renal tubules (Ray et al., 2006; Cihlar et al., 2007). TFV is also a substrate of the efflux drug transporter MRP-8 expressed in renal proximal tubules since higher cytotoxic concentrations of the drug were observed in cells overexpressing MRP-8 (Tun-Yhong et al., 2017). The uptake of ARVs cidofovir, adefovir and TFV was evaluated in human embryonic kidney (HEK293) cells transfected with uptake drug transporters OCT-2, OAT-1 and OAT-3 (Uwai et al., 2007). Results showed higher uptake of all ARVs through OAT-1 compared to OAT-3, while OCT-2 exhibited no uptake, indicating that OAT-1 plays a significant role in renal transport of these ARVs (Uwai et al., 2007). Similarly, renal secretion of FTC was mediated by MATE-1 which functionally acts as an efflux drug transporter, expressed on the apical side of renal proximal tubules (Reznicek et al., 2017).

These studies collectively provide insight that ARV drug levels are not only determined by drug adherence but also by other factors such as the presence of specific drug transporters and their expression levels. However, definitive conclusions on the full effects of drug transporters on ARV pharmacokinetics in at-risk groups such as young women especially in Africa have not been drawn. The paucity of data on African women warrants the need for new studies to fully understand:

i) the effect of ARVs on drug transporters expression,
ii) how varying drug transporter expression levels influence ARV penetration in vulnerable areas such as the FGT, and
iii) how different biological factors such as inflammation and polymorphisms may also affect drug transporter expression and function.

7 Biological factors modulating drug transporter expression and function

7.1 Genetic polymorphisms

Pharmacogenetic research has been used as a tool to determine individuals' susceptibility to certain diseases and for the customisation of drug therapies according to patient's genetic blueprint (Sissung et al., 2012; Castellanos-Rubio and Ghosh, 2019). As such, sequencing and genotyping technology have been widely used to identify and determine the effect of variants such as genetic polymorphisms in various genes. There are four types of genetic polymorphisms that have been shown to regulate genes (Ismail and Essawi, 2012). These include:

I. small insertions and deletions (InDels) which is a deletion or insertion in the DNA sequence (Boschiero et al., 2015),
II. interspaced or tandem repeat polymorphisms which are tandemly repeated nucleotides of approximately ≥2 base pairs (bp) in DNA sequences (Ismail and Essawi, 2012),
III. structure or copy-number variations (CNVs), polymorphisms which are various copies of differently sized segments of nucleotides in DNA sequences (Stankiewicz and Lupski, 2010) and
IV. single nucleotide polymorphisms (SNPs), which are point mutations of nucleotide bases within DNA sequences (Sissung et al., 2012; Yee et al., 2018).

Types of SNP variations include missense mutations or nonsynonymous substitutions which is a single nucleotide change within a codon, subsequently resulting in the coding of a different amino acid (Hunt et al., 2009). The presence of such mutations on protein binding sites may affect substrate binding, while those not found on protein binding sites may affect protein expression levels. For example, the missense mutation rs2273697 located on the efflux drug transporter gene *ABCC2* encoding MRP-2 results in a change from valine to isoleucine, on exon 417 (V417I), affecting its expression levels (Yee et al., 2018). Another type of SNP mutation is a silent mutation or synonymous substitutions, which are single nucleotide point mutations on a codon that do not result in an amino acid change (Hunt et al., 2009). However, these may still affect RNA transcription and stability that may affect mRNA expression levels and protein binding (Yee et al., 2018). For example, the SNP rs1045642 located on the efflux drug transporter *ABCB1* gene (3435C/T Ile1145Ile) encoding P-gp is a type of silent mutation that has been highly studied in drug pharmacokinetics (Sissung et al., 2012; Yee et al., 2018). This SNP has been also previously associated with low P-gp expression levels in the duodenum which correlated with an increase of digoxin plasma concentrations (Hoffmeyer et al., 2000). The presence of certain genetic variations in drug transporter genes has sparked a huge interest in further understanding their functional effect; especially since SNPs in certain drug transporter genes have been shown to modulate their function by affecting protein folding, expression levels, and their ability to bind substrates and regulate drug pharmacokinetics (Sissung et al., 2012; Yee et al., 2018).

SNPs involved in the pharmacokinetics of ARVs have led to adverse effects and varied ARV therapy outcomes amongst HIV infected patients (Shenfield, 2004; Arruda et al., 2016). Arruda et al. (2016) showed the association between SNPs in drug transporter genes and intolerance to ARVs in a cohort of HIV infected Brazilian participants (Arruda et al., 2016). Results showed an association between variations in *ABCC2* genes (rs3740066 and rs4148396) encoding MRP-2 and intolerance in patients taking regimens containing either LPV, RTV, indinavir or ATV PIs; while variations in *SLCO2B1* genes (rs2712816, rs12422149, rs1676885 and rs949069) encoding OATP2B1 caused intolerance in patients taking regimens containing stavudine or ZDV nucleotide reverse transcriptase inhibitors (NRTIs) (Arruda et al., 2016). The presence of the C allele on the *ABCC1* gene 198217C/T (rs212091) encoding MRP-1 and the TT genotype on the *ABCB1* gene 3435C/T (rs1045642) encoding P-gp; was also shown to be possibly associated with reduced gene expression in an HIV infected Brazilian participants receiving highly active antiretroviral therapy (HAART) regimens; subsequently affecting the efflux of ARV regimens containing PIs, leading to an increased risk of virological failure (Table 3) (Coelho et al., 2013).

Fellay et al. (2002) showed that the TT genotype on the *ABCB1* gene 3435C/T in an HIV infected Caucasian population was associated with low P-gp expression in PBMCs, affecting ARV concentrations (Fellay et al., 2002), However, a subsequent study showed that virological failure was associated with the CC genotype of the *ABCB1* gene 3435C/T instead of the TT genotype in HIV infected patients from the province of British Columbia in Canada (Brumme et al., 2003). To elucidate variations of the *ABCB1* 3435C/T SNP observed in these studies; prior results by Ameyaw et al. (2001) that assessed the frequency of this SNP in ten ethnic groups can be used (Ameyaw et al., 2001). Results showed noticeable differences in the SNP frequencies between African, Asian and European populations. The C allele was highly present in the African populations compared to Asian and European populations which exhibited high frequencies for the CT and TT genotypes (Ameyaw et al., 2001). Schaeffeler et al. (2001) also supported these findings by reporting a high frequency of the CC genotype in the *ABCB1* gene 3435C/T of West African and African American populations compared to the T allele (Schaeffeler et al., 2001). These findings could imply possible variations in drug transporter genes which could lead to varied ARV therapy outcomes in the African vs. Caucasian populations (Ameyaw et al., 2001; Schaeffeler et al., 2001).

Pharmacogenetic studies conducted with African populations have also shown high genetic diversity, which subsequently leads to varied drug transporter function and

TABLE 3 Effects of SNPs in drug transporter genes involved in the pharmacokinetics of ARVs in different ethnic groups.

SNPs	Ethnic group	ARVs	SNPs effect	Genotype causing effect	Genotype frequency Number of patients n (%)	References
ABCC2 224C/T (rs717620)	Thailand	TFV, Lamivudine, Efavirenz	Increased TFV plasma concentration	CC	CC 67 (57); CT 45 (39); TT 5 (4) n = 117	Manosuthi et al. (2014)
	Japanese	TFV, Emtricitabine, Darunavir, Ritonavir	TFV induced-KTD	CC	CC 18 (94.7); CT 1 (5.3); TT 0 (0) n = 19	Nishijima et al. (2012)
	Caucasian	TFV	TFV induced-KTD	CC	CC 9 (60.0); TT 1 (6.7); CT 5 (33.3) n = 15	Danjuma et al. (2018)
ABCC4 4131T/G (rs3742106)	Thailand	TFV	Increased TFV plasma concentration	TG/GG	TT 34 (22.7); TG 80 (53.3); GG 36 (24.0) n = 150	Rungtivasuwan et al. (2015)
ABCC1 198217C/T (rs212091)	Brazilian	Zidovudine, Lamivudine, Efavirenz/Nevirapine; Lopinavir/Ritonavir	Increased risk of virological failure	CC	CC 62 (84.9); TC 10 (13.7); CC 1 (1.4) n = 73	Coelho et al. (2013)
ABCB1 3435C/T (rs1045642)				TT	CC 37 (50.7); CT 25 (34.2); TT 11 (15.1) n = 73	
ABCC4 4976C/T (rs1059751)	Thailand	TFV	TFV induced-KTD	CC	CC 20 (37.0); TT 9 (16.7); CT 25 (46.3) n = 54	Likanonsakul et al. (2016)
ABCC4 3436A/G (rs1751034)	Caucasian	TFV	TFV induced-KTD	GG	AA 27 (64.3.); AG 9 (21.4); GG 6 (14.3) n = 42	Salvaggio et al. (2017)
SLCO1B1 463C/A (rs11045819)	African	Rifampin	Low plasma concentrations	CC	CC 30 (81); CA 7 (19) n = 37	Weiner et al. (2010)
	Ghanaian	Rifampin	High plasma concentrations	CC	CC 95 (84.1); CA 17 (15); AA 1 (0.09) n = 113	Dompreh et al. (2018)

expression levels; impacting drug pharmacokinetics differently as reviewed by Rajman et al. (2020). The presence of the *SLCO1B1* SNP 463C/A rs11045819 encoding the OATP1B1 protein was shown to impact rifampin pharmacokinetics differently in African populations (Weiner et al., 2010; Dompreh et al., 2018). A study by Weiner et al. (2010) showed that a high frequency of the CC genotype for the *SLCO1B1* 463C/A (rs11045819) gene was associated with low rifampin concentrations in African individuals during multidrug intensive therapy against TB (Weiner et al., 2010). However, in an African Ghanaian population also exhibiting a high frequency CC genotyping for the same gene taking standard first-line TB therapy; no effect on rifampin was observed (Dompreh et al., 2018) Table 3. Similarly studies by Chigutsa et al. (2011) and Gengiah et al. (2014) on the *SLCO1B1* (rs4149032) SNP both reported an associated between high SNP frequency and low rifampin plasma concentrations in TB and HIV-TB co-infected South African individuals taking rifampin (Chigutsa et al., 2011; Gengiah et al., 2014). This association was however, not observed in a TB infected Ghanaian population taking standard first-line TB therapy containing rifampin which also exhibited a high frequency for this SNP (Dompreh et al., 2018). The effect of the *ABCB1* SNP 4036G/G (rs3842) encoding P-gp on efavirenz was evaluated in different African populations. In an HIV infected South African population the AG and GG genotypes were significantly associated with decreased efavirenz plasma concentrations (Swart et al., 2012), however the GG genotypes in a healthy Ugandan population was associated with higher efavirenz plasma concentrations (Mukonzo et al., 2009). Similarly in Ethiopian and Tanzanian HIV infected populations the presence of the G allele was associated with higher efavirenz plasma concentrations, with higher frequency of the G allele observed in Tanzanians (Ngaimisi et al., 2013). These data indicated that the effects of SNPs may differ among African populations; therefore, in order to make definitive conclusions that a SNP affects the African population in a certain way, the functional or expressional effect SNPs should be tested among a wide range of different African populations as reviewed by Rajman et al. (2020). Despite the small sample size and sparsity of these data in various studies with African populations as reviewed in Rajman et al. (2020), these data do add to the understanding of how SNPs can impact drug pharmacokinetics in the African population. Together these studies could be used to adjust the standard recommended dose of ARV and TB drug for the African population that accounts for the presence of SNPs (Dandara et al., 2011; Rajman et al., 2020).

The effects of SNPs on drug transporter genes have also been associated with increased plasma concentrations of TFV. Studies on an HIV-infected cohort in Thailand showed higher TFV plasma concentrations in patients with the CC genotype on the *ABCC2* 224C/T gene (rs717620) encoding MRP-2 compared to patients with the *ABCC2* TT or CT genotypes (Table 3) (Manosuthi et al., 2014). Similarly, another Thailand

study by Rungtivasuwan et al. (2015) reported higher TFV plasma concentrations in HIV infected patients with the *ABCC4* 4131 (rs3742106) TG or GG genotypes (encoding MRP-4) compared to patients with the *ABCC4* TT genotype (Table 3) (Rungtivasuwan et al., 2015). These studies proposed that polymorphisms in these drug transporter genes may alter their gene expression or function in renal tubules leading to more effluxed drug and reduced glomerular filtration which is involved in TFV renal clearance; resulting in higher plasma concentrations (Manosuthi et al., 2014; Rungtivasuwan et al., 2015). A more recent study also showed in an HIV infected Caucasian population a significant association of the CC genotype in the *ABCC2* 224C/T gene with high TFV plasma concentrations, resulting in an increased risk of TFV induced-kidney tubular dysfunction (KTD) (Table 3) (Danjuma et al., 2018). Nishijima et al. (2012) previously confirmed that the CC genotype in the *ABCC2* 224C/T gene leads to high TFV plasma concentrations resulting in the induction of KTD or renal toxicity in Japanese patients (Table 3) (Nishijima et al., 2012). While the presence of the TT genotype in the *ABCC4* 4131T/G gene was not associated with TFV induced-KTD, the study attributed these findings to inter-individual variability in genetic backgrounds, which may cause patients to respond differently to the same drug (Kerb, 2006; Nishijima et al., 2012). Other SNPs on the *ABCC4* gene that have been associated with increased plasma TFV concentrations were evaluated in two studies; in an infected population from Thailand with the C allele on the ABCC4 4976C/T gene (Likanonsakul et al., 2016), and in a Caucasian population with GG genotype on the *ABCC4* 3436A/G gene (Salvaggio et al., 2017).

Reports obtained from these studies highlight the importance of understanding how the presence of SNPs may affect the efficacy of ARVs by affecting drug transporters' expression and function. Furthermore, these findings could be used to identify populations who are at a higher risk of developing adverse effects due to the presence of certain SNPs. However, most of these studies on SNPs in drug transporter genes affecting ARVs have been performed in non-African populations. Since SNP frequency differs significantly among different ethnic populations, more comprehensive investigations of SNPs in drug transporter genes are required, especially in the populations of African ethnicity. Data from populations of African descent will help us better understand how genetic diversity within these populations and SNPs influence drug transporter genes and subsequently lead to effective or ineffective therapy.

Pharmacogenetic research on polymorphisms present in drug transporter and drug-metabolizing genes is also vital in precision medicine, which enables the tailoring of effective therapies based on patients' genetic backgrounds (Hockings et al., 2020). The advantage of a precision medicine approach is the ability to predict putative ineffective therapies and possibly reduce adverse reactions (Hockings et al., 2020). Since there are reports of increased adverse reactions in patients in populations of African ethnicity taking ARVs, precision medicine is highly important in HIV prevention and treatment (Hockings et al., 2020). Patients taking ARV regimens containing EFV in SSA were predisposed to EFV-induced neuropsychiatric adverse reactions, due to specific genetic variants that reduced the functionality of cytochrome P450 2B6 (*CYP2B6*) the enzyme involved in EFV metabolism (Masimirembwa et al., 2016). One of the genetic variants of *CYP2B6* 516G>T (rs3745274) reported a frequency of 34%–50% in African populations compared to 15 and 20% in white populations (Masimirembwa et al., 2016). However, when the EFV dosages in ARV regimens were further titrated and reduced, there was improved EFV metabolism leading to significantly reduced neuropsychiatric adverse reactions (Gatanaga et al., 2007; Masimirembwa et al., 2016). These disparities in frequencies between the populations could lead to varied enzyme metabolism when similar drugs are used which may lead to ineffective drug metabolism and availability. This data highlights the importance of using pharmacogenetic research in guiding the development of precision medicine, especially in highly affected populations ensuring effective drug dosing, delivery, and metabolism.

7.2 Genital inflammation

Genital tract inflammation has been identified as an elevated profile of five of any of the nine pro-inflammatory cytokines (MIP-1α, MIP-1β, IP-10, IL-8, MCP-1, IL-1α, IL-1β, IL-6, and TNF-α) above the 75th percentile for each cytokine in a previous CAPRISA study (Abdool Karim et al., 2010; Masson et al., 2015) Genital inflammation creates an environment conducive for HIV infection and replication (McKinnon et al., 2018) increasing the risk for HIV by more than three-fold (Masson et al., 2015). The role of genital inflammation in undermining PrEP efficacy was demonstrated in a study by McKinnon et al. (2018). A 57% protective efficacy was found in women with no genital inflammation compared to 3% in women with genital inflammation (McKinnon et al., 2018). Although the mechanisms to explain why some people have comparatively high levels of genital inflammation while others do not are not fully understood, a likely driver of genital inflammation is bacterial vaginosis (BV), a microbial dysbiosis common in reproductively active women (Klatt et al., 2017). BV also plays a role in significantly modifying PrEP efficacy (Klatt et al., 2017). Klatt et al. (2017) showed that TFV gel reduced HIV incidence by 61% in women with a *Lactobacillus* dominant vaginal microbiome compared to only 18% in women with a *non-Lactobacillus* dominant vaginal microbiome (Klatt et al., 2017). Furthermore, sexually transmitted infections (STIs) (Masson et al., 2014) and exogenous hormonal contraceptives (HCs) (Deese et al., 2015) are also significantly associated with genital inflammation, through the secretion of pro-inflammatory

cytokines (Masson et al., 2014; Deese et al., 2015). The mechanisms by which all these factors individually or collectively interplay with drug transporter disposition, drug levels and in turn PrEP efficacy remains less well defined.

7.2.1 Role of inflammation-induced cytokines, in modulating drug transporter expression and function

The impact of inflammation on drug transporter expression and function has been examined in tissues of the intestines, kidneys, and blood-brain barrier (Petrovic et al., 2007; Saib and Delavenne, 2021). Despite the lack of data regarding the direct mechanisms involved; inflammation-mediated changes in drug transporter expression and function have been implicated in significantly impacting drug pharmacokinetics (Petrovic et al., 2007; Cressman et al., 2012). In an *in vitro* study, human brain cell lines (hCMEC/D3) treated with IL-6 and IL-1β, resulted in the downregulation of BCRP and P-gp expression levels (Poller et al., 2010). Additionally, the induction of IL-6 and IFN-γ on primary human hepatocytes was also shown to downregulate the mRNA expression levels of the efflux drug transporters BCRP, MRP-2, and MRP-3 and influx/uptake drug transporters OATP (-2B1,-1B1,-1B3) (Le Vee et al., 2009; Le Vee et al., 2011). Previous studies corroborated similar findings of inflammation IL-6 induced downregulation of P-gp expression on rat hepatocytes and human hepatoma cell lines (Sukhai et al., 2000). Similarly human cell line Caco-2 pre-treated with TNF-α significantly decreased intestinal P-gp expression, while IFN-γ had no effect (Belliard et al., 2004). In rats with endotoxemia, high levels of IL-6 and IL-1β reduced the mRNA expression levels of P-gp and MRP-2 in intestinal tissues (Arana et al., 2017). This lipopolysaccharide-induced endotoxemia in rats model showed that there was IL-1β induced downregulation of MRP-2 in enterocytes (Arana et al., 2020). These various cellular and small animal models demonstrate how infection and inflammation-induced cytokines can modulate drug transporter disposition. The caveat to the methods used in these models is that mRNA expression levels may not directly reflect functional proteins expressed. Future investigations are therefore required and should include both mRNA expression to its corresponding protein. There are also other biological factors related to inflammation that could also affect drug transporter disposition.

7.3 Role of toll-like receptors and pH in modulating drug transporter expression and function

7.3.1 Toll-like receptors-induced inflammation

TLRs are pattern recognition receptors, these receptors recognise pathogen-associated molecular patterns located on various microbes for example Pam3CSK-4 and lipopolysaccharide (LPS) which are TLR-2 and TLR-4 agonists, respectively (Cario, 2016; Suzuki et al., 2017). TLRs are activated by the binding to their respective agonists. This interaction causes the stimulation of appropriate signalling pathways in innate and adaptive immune cells which then regulate drug transporter expression levels (Cario, 2016; Suzuki et al., 2017). These TLR-mediated changes in drug transporter expression have been evaluated in the progression of atherosclerosis. To determine which downstream transcriptional signalling pathways were involved in this interaction; *in-vitro* testing using mouse macrophage cell lines (Raw 264.7) stimulated with TLR-2 and TLR-4 agonists Pam3CSK-4 and Lipid-A, respectively; were performed (Suzuki et al., 2017). Expression of myeloid differentiation primary-response protein 88 (MyD88), Toll/interleukin-1-domain-containing adapter-inducing interferon β (TRIF), liver X receptors (LXR), interferon regulatory factor 3 (IRF3), and the phosphorylation of nuclear factor kappa B (NF-kb) were determined with TLR-2 and TLR-4 activation. These results showed a differential pattern of significantly increased MyD88, LXR and NF-kb expression and low TRIF and IRF3 expression (Suzuki et al., 2017). This coincided with the significant upregulation of *ABCA1* expression levels, while ABCG1 expression levels were downregulated. TLR-2 stimulated cells pre-treated with NF-kb and p38 inhibitors MG-132 and SB203580, respectively suppressed the expression of *ABCA1*. These data provided evidence of the sensitivity of drug transporter expression to signal transduction–the MyD88, LXR, NF-kb and p38 pathways. These data provide support to the hypothesis that inflammation modulates the expression of drug transporters which can then lead to disease pathogenesis (Suzuki et al., 2017).

7.3.2 Sensitivity of drug transporter function to pH

The level of acidity or alkalinity (pH) in extracellular fluids is an additional factor that has been shown to modulate drug transporter function (Breedveld et al., 2007). The function of the efflux drug transporter BCRP was determined in Madin-Darby canine kidney (MDCK II) cells grown in pH adjusted media and exposed to methotrexate (MTX). At acidic pH levels, the efflux transporter BCRP pumped out MTX more efficiently when compared to physiological and basic pH levels. This data highlighted the possible clinical implications that the function of BCRP is pH-sensitive in the extracellular environment, thereby affecting intracellular concentrations and the effectiveness of MTX (Breedveld et al., 2007). These data suggest that pH is an additional factor that can also modulate drug transporter function which can then affect the effectiveness of drugs (Breedveld et al., 2007).

Collectively these studies demonstrate how inflammation-induced cytokines and TLRs are involved in regulating the expression of drug transporters in various tissues;

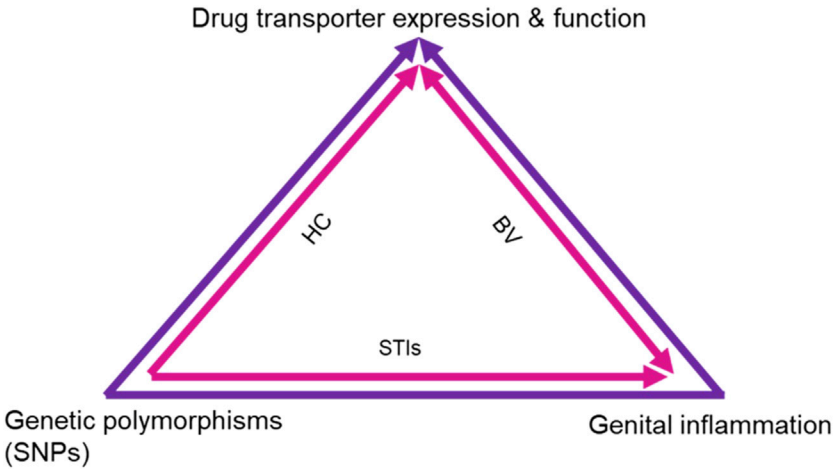


FIGURE 2
Proposed mechanism of effects on drug transporter expression and function. The schematic shows the intersection of different biological factors and SNPs in drug transporter genes that affect drug transporter expression in the FGT, renal system and blood, subsequently affecting PrEP efficacy. Genital inflammation and SNPs are known to directly affect drug transporter expression and functionality, while the combined use of HCs and ARVs also affects drug transporter expression and function. Additionally, the presence of STIs and BV are shown to contribute to genital inflammation which in turn affects drug transporter expression and function. HC, Hormonal contraceptives; BV, Bacterial vaginosis; STIs, Sexually transmitted infections.

subsequently altering intracellular and plasma drug concentrations, thereby affecting drug pharmacokinetics and efficacy. Inflammation mediated changes in drug transporter expression are however mostly based on animal models and cell lines (Cressman et al., 2012; Saib and Delavenne, 2021), thereby warranting the need for comparative *in-vivo* human studies. Future studies should also elucidate how BV, HC and STIs-induced genital inflammation contribute to drug transporter expression and function, subsequently, predisposing women to HIV infections, even during PrEP intake. Therefore, additional studies are needed to understand the interplay between inflammation and drug transporter expression, especially in sites highly susceptible to HIV such as the FGT and blood. Findings from such studies would provide a better understanding of how the presence of systemic and genital inflammation may alter drug transporters subsequently affecting ARV pharmacokinetics. Further elucidation of these factors either individually or collectively will aid in understanding disparities in PrEP efficacies observed in PrEP trials. This is especially important in highly susceptible groups such as African women from HIV endemic settings where PrEP is advocated as the standard of care for HIV prevention.

Together these studies show evidence that there may be an inextricable link between the expression and function of drug transporters with genetic polymorphisms, TLRs, pH and genital inflammation, which is further influenced by the presence of BV, STIs and HCs (Figure 2). Subsequently these factors may significantly affect drug concentrations and potentially drug efficacies.

8 Conclusion

The current review provides evidence that the FGT, renal system and blood are subject to a variety of host biological factors that may undermine PrEP efficacy by affecting drug transporter expression levels and function. These aforementioned studies show how drug transporters are increasingly recognised as key determinants in drug pharmacokinetics and response. However, their contributions to the inconsistent efficacies seen in PrEP clinical studies in African women from regions with high HIV infection rates such as South Africa, have not been elucidated. Characterising the expression level of drug transporters in the blood and FGT from a vulnerable population will better define the biological factors underlying compartment variation in drug exposure during oral PrEP in at-risk African women. In turn, we may be able to better understand why African women remain susceptible to HIV despite PrEP interventions. Additionally, findings from such studies will shed an important light on how the genetics and the biology of the mucosal environment may play a pivotal role in modifying drug transporter expression, subsequently modulating HIV risk. Understanding these data may also aid in the development of more effective, safe and optimal delivery systems that facilitate consistent effective dosage and usage of appropriate PrEP drugs.

Author contributions

Study conceptualisation was by DA, NZ, and PS. Manuscript writing (original draft) was conducted by NZ. Reviewing and edit suggestions for final manuscript was conducted by DA, PS, AS, SN, and VR. Final manuscript edited by NZ and accepted by all authors.

Funding

NZ, PS and DA were funded by the Department of Science and Innovation (DSI)-National Research Foundation (NRF) Centre of Excellence in HIV Prevention (Grant No. 96354). PS was also funded by the South African Medical Research Council (SAMRC) Special Initiative (Grant no. 96151) and S&F Scarce Skills Postdoctoral Fellowships (Grant no. 132714). AS is supported by EDCTP Career Development Fellowship (Grant No. TMA2016CDF-1582). DA was also funded through the SAMRC Self-Initiated Grant and the NRF of South Africa Thuthuka (Grant No. TTK160517165310), the NRF Research Career Advancement Fellowship (Grant No. RCA13101656388), the Polio Research Foundation of South Africa (PRF 17/02) and an EDCTP fellowship (Grant No. TMA2017SF-1960). SN was funded by the Columbia University-Southern African Fogarty AITRP Programme (Grant No. D43TW00231), National Research Fund Thuthuka Research Grant (Grant No. TTK160510164586), and Poliomyelitis Research Foundation Research Grant (Grant No. 16/17). VR was funded as a FLAIR Research Fellow (the Future Leader in African Independent Research (FLAIR) Fellowship Programme was a partnership between the African Academy of Sciences (AAS) and the Royal Society that was funded by the United Kingdom Government as part of the Global Challenge Research Fund (GCRF) (Grant No. FLAIR-FLR\R1\190204); supported by the South African Medical Research Council (SAMRC) with funds from the Department of Science and Technology (DST); and VR was also supported in part through the Sub-Saharan African Network for TB/HIV Research Excellence (SANTHE), a DELTAS Africa Initiative (Grant No. DEL-15-006) by the AAS; VR was also supported by the Grants, Innovation and Product Development unit of the South African MedicalResearch Council with funds received from Novartis and GSK R&D (Grant No. GSKNVS2/202101/005).

Acknowledgments

Special thanks to the Centre for the AIDS Programme of Research in South Africa (CAPRISA) and the University of KwaZulu-Natal Medical Microbiology Department for research support.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Abdool Karim, Q., Abdool Karim, S. S. A., Frohlich, J. A., Grobler, A. C., Baxter, C., Mansoor, L. E., et al. (2010). Effectiveness and safety of tenofovir gel, an antiretroviral microbicide, for the prevention of HIV infection in women. *science* 329, 1168–1174. doi:10.1126/science.1193748

Abdool Karim, Q., Havlir, D., and Phanuphak, N. (2020). Putting women in the centre of the global HIV response is key to achieving epidemic control. *J. Int. AIDS Soc.* 23, 25473. doi:10.1002/jia2.25473

Abdool Karim, Q., Kharsany, A. B., Frohlich, J. A., Werner, L., Mlotshwa, M., Madlala, B. T., et al. (2012). HIV incidence in young girls in KwaZulu-Natal, South Africa-public health imperative for their inclusion in HIV biomedical intervention trials. *AIDS Behav.* 16, 1870–1876. doi:10.1007/s10461-012-0209-y

Alcaide, M. L., Chisembele, M., Mumbi, M., Malupande, E., and Jones, D. (2014). Examining targets for HIV prevention: Intravaginal practices in urban lusaka, Zambia. *AIDS Patient Care STDs* 28, 121–127. doi:10.1089/apc.2013.0309

Alvarez, E., Morello, J., Soriano, V., Labarga, P., and Rodriguez-Nóvoa, S. (2011). Critical appraisal and update on tenofovir in management of human immunodeficiency virus infection. *Virus Adapt. Treat.* 3, 55. doi:10.2147/VAAT.S12708

Ameyaw, M. M., Regateiro, F., Li, T., Liu, X., Tariq, M., Mobarek, A., et al. (2001). MDR1 pharmacogenetics: Frequency of the C3435T mutation in exon 26 is significantly influenced by ethnicity. *Pharmacogenetics* 11, 217–221. doi:10.1097/00008571-200104000-00005

Arana, M. R., Dominguez, C. J., Zecchinati, F., Tocchetti, G. N., Mottino, A. D., and Villanueva, S. S. M. (2020). Role of interleukin 1 beta in the regulation of rat intestinal multidrug resistance-associated protein 2 under conditions of experimental endotoxemia. *Toxicology* 441, 152527. doi:10.1016/j.tox.2020.152527

Arana, M. R., Tocchetti, G. N., Zecchinati, F., Londero, A. S., Dominguez, C., Perdomo, V., et al. (2017). Glucagon-like peptide 2 prevents down-regulation of intestinal multidrug resistance-associated protein 2 and P-glycoprotein in endotoxemic rats. *Toxicology* 390, 22–31. doi:10.1016/j.tox.2017.08.007

Arruda, M. B., Campagnari, F., de Almeida, T. B., Couto-Fernandez, J. C., Tanuri, A., and Cardoso, C. C. (2016). Single nucleotide polymorphisms in cellular drug transporters are associated with intolerance to antiretroviral therapy in Brazilian HIV-1 positive individuals. *PloS one* 11, e0163170. doi:10.1371/journal.pone.0163170

Baeten, J. M., Donnell, D., Ndase, P., Mugo, N. R., Campbell, J. D., Wangisi, J., et al. (2012). Antiretroviral prophylaxis for HIV prevention in heterosexual men and women. *N. Engl. J. Med.* 367, 399–410. doi:10.1056/nejmoa1108524

Baeten, J. M., Palanee-Phillips, T., Brown, E. R., Schwartz, K., Soto-Torres, L. E., Govender, V., et al. (2016). Use of a vaginal ring containing dapivirine for HIV-1 prevention in women. *N. Engl. J. Med.* 375, 2121–2132. doi:10.1056/nejmoa1506110

Belliard, A. M., Lacour, B., Farinotti, R., and Leroy, C. (2004). Effect of tumor necrosis factor-α and interferon-γ on intestinal P-glycoprotein expression, activity, and localization in Caco-2 cells. *J. Pharm. Sci.* 93, 1524–1536. doi:10.1002/jps.20072

Boschiero, C., Gheyas, A. A., Ralph, H. K., Eory, L., Paton, B., Kuo, R., et al. (2015). Detection and characterization of small insertion and deletion genetic variants in modern layer chicken genomes. *BMC Genomics* 16, 562. doi:10.1186/s12864-015-1711-1

Bousquet, L., Pruvost, A., Didier, N., Farinotti, R., and Mabondzo, A. (2008). Emtricitabine: Inhibitor and substrate of multidrug resistance associated protein. *Eur. J. Pharm. Sci.* 35, 247–256. doi:10.1016/j.ejps.2008.06.017

Bousquet, L., Pruvost, A., Guyot, A.-C., Farinotti, R., and Mabondzo, A. (2009). Combination of tenofovir and emtricitabine plus efavirenz: *In vitro* modulation of ABC transporter and intracellular drug accumulation. *Antimicrob. Agents Chemother.* 53, 896–902. doi:10.1128/aac.00733-08

Breedveld, P., Pluim, D., Cipriani, G., Dahlhaus, F., van Eijndhoven, M. A., de Wolf, C. J., et al. (2007). The effect of low pH on breast cancer resistance protein (ABCG2)-mediated transport of methotrexate, 7-hydroxymethotrexate, methotrexate diglutamate, folic acid, mitoxantrone, topotecan, and resveratrol in *in vitro* drug transport models. *Mol. Pharmacol.* 71, 240–249. doi:10.1124/mol.106.028167

Brumme, Z. L., Dong, W. W., Chan, K. J., Hogg, R. S., Montaner, J. S., O'Shaughnessy, M. V., et al. (2003). Influence of polymorphisms within the CX3CR1 and MDR-1 genes on initial antiretroviral therapy response. *Aids* 17, 201–208. doi:10.1097/00002030-200301240-00010

Cario, E. (2016). Toll-like receptors in the pathogenesis of chemotherapy-induced gastrointestinal toxicity. *Curr. Opin. Support Palliat. Care* 10, 157–164. doi:10.1097/spc.0000000000000202

Castellanos-Rubio, A., and Ghosh, S. (2019). Disease-associated SNPs in inflammation-related lncRNAs. *Front. Immunol.* 10, 420. doi:10.3389/fimmu.2019.00420

Celum, C. L. (2011). HIV preexposure prophylaxis: New data and potential use. *Top. Antivir. Med.* 19, 181–185.

Chigutsa, E., Visser, M. E., Swart, E. C., Denti, P., Pushpakom, S., Egan, D., et al. (2011). The SLCO1B1 rs4149032 polymorphism is highly prevalent in South Africans and is associated with reduced rifampin concentrations: Dosing implications. *Antimicrob. Agents Chemother.* 55, 4122–4127. doi:10.1128/aac.01833-10

Cihlar, T., Ray, A. S., Laflamme, G., Vela, J. E., Tong, L., Fuller, M. D., et al. (2007). Molecular assessment of the potential for renal drug interactions between tenofovir and HIV protease inhibitors. *Antivir. Ther.* 12, 267–272. doi:10.1177/135965350701200211

Coelho, A. V., Silva, S. P., de Alencar, L. C., Stocco, G., Crovella, S., Brandão, L. A., et al. (2013). ABCB1andABCC1variants associated with virological failure of first-line protease inhibitors antiretroviral regimens in Northeast Brazil patients. *J. Clin. Pharmacol.* 53, 1286–1293. doi:10.1002/jcph.165

Corneli, A., Perry, B., McKenna, K., Agot, K., Ahmed, K., Taylor, J., et al. (2016). Participants' explanations for nonadherence in the FEM-PrEP clinical trial. *JAIDS J. Acquir. Immune Defic. Syndromes* 71, 452–461. doi:10.1097/qai.0000000000000880

Cottrell, M. L., Srinivas, N., and Kashuba, A. D. (2015). Pharmacokinetics of antiretrovirals in mucosal tissue. *Expert Opin. Drug Metabolism Toxicol.* 11, 893–905. doi:10.1517/17425255.2015.1027682

Cressman, A. M., Petrovic, V., and Piquette-Miller, M. (2012). Inflammation-mediated changes in drug transporter expression/activity: Implications for therapeutic drug response. *Expert Rev. Clin. Pharmacol.* 5, 69–89. doi:10.1586/ecp.11.66

Critchlow, C. W., Wölner-Hanssen, P., Eschenbach, D. A., Kiviat, N. B., Koutsky, L. A., Stevens, C. E., et al. (1995). Determinants of cervical ectopia and of cervicitis: Age, oral contraception, specific cervical infection, smoking, and douching. *Am. J. Obstetrics Gynecol.* 173, 534–543. doi:10.1016/0002-9378(95)90279-1

Dandara, C., Lombard, Z., Du Plooy, I., McLellan, T., Norris, S. A., and Ramsay, M. (2011). Genetic variants in CYP (-1A2, -2C9, -2C19, -3A4 and -3A5), VKORC1 and ABCB1 genes in a black South African population: A window into diversity. *Pharmacogenomics* 12, 1663–1670. doi:10.2217/pgs.11.106

Danjuma, M. I., Egan, D., Abubeker, I. Y., Post, F., and Khoo, S. (2018). Polymorphisms of tenofovir disoproxil fumarate transporters and risk of kidney tubular dysfunction in HIV-positive patients: Genetics of tenofovir transporters. *Int. J. STD AIDS* 29, 1384–1389. doi:10.1177/0956462418786562

de Oliveira, T., Kharsany, A. B., Gräf, T., Cawood, C., Khanyile, D., Grobler, A., et al. (2017). Transmission networks and risk of HIV infection in KwaZulu-natal, South Africa: A community-wide phylogenetic study. *Lancet HIV* 4, e41–e50. doi:10.1016/s2352-3018(16)30186-2

Deese, J., Masson, L., Miller, W., Cohen, M., Morrison, C., Wang, M., et al. (2015). Injectable progestin-only contraception is associated with increased levels of pro-inflammatory cytokines in the female genital tract. *Am. J. Reprod. Immunol.* 74, 357–367. doi:10.1111/aji.12415

Delany-Moretlwe, S., Lombard, C., Baron, D., Bekker, L.-G., Nkala, B., Ahmed, K., et al. (2018). Tenofovir 1% vaginal gel for prevention of HIV-1 infection in women in South Africa (FACTS-001): A phase 3, randomised, double-blind, placebo-controlled trial. *Lancet Infect. Dis.* 18, 1241–1250. doi:10.1016/S1473-3099(18)30428-6

Dompreh, A., Tang, X., Zhou, J., Yang, H., Topletz, A., Adu Ahwireng, E., et al. (2018). Effect of genetic variation of NAT2 on isoniazid and SLCO1B1 and CES2 on rifampin pharmacokinetics in Ghanaian children with tuberculosis. *Antimicrob. Agents Chemother.* 62, 17. doi:10.1128/AAC.02099-17

Dumond, J. B., Yeh, R. F., Patterson, K. B., Corbett, A. H., Jung, B. H., Rezk, N. L., et al. (2007). Antiretroviral drug exposure in the female genital tract: Implications for oral pre- and post-exposure prophylaxis. *Aids* 21, 1899–1907. doi:10.1097/qad.0b013e328270385a

Durevall, D., Lindskog, A., and George, G. (2019). Education and HIV incidence among young women in KwaZulu-Natal: An association but no evidence of a causal protective effect. *PloS one* 14, e0213056. doi:10.1371/journal.pone.0213056

Falasca, F., Maida, P., Montagna, C., Antonelli, L., d'Ettorre, G., Monteleone, K., et al. (2011). Expression of the mRNA levels for MDR1, MRP1, MRP4, and MRP5 in HIV antiretroviral naive patients: Follow-up at 48 weeks after the beginning of therapy. *J. Acquir Immune Defic. Syndr.* 56, e54–e56. doi:10.1097/qai.0b013e3181fe4c89

Fellay, J., Marzolini, C., Meaden, E. R., Back, D. J., Buclin, T., Chave, J.-P., et al. (2002). Response to antiretroviral treatment in HIV-1-infected individuals with allelic variants of the multidrug resistance transporter 1: A pharmacogenetics study. *Lancet* 359, 30–36. doi:10.1016/s0140-6736(02)07276-8

Gatanaga, H., Hayashida, T., Tsuchiya, K., Yoshino, M., Kuwahara, T., Tsukada, H., et al. (2007). Successful efavirenz dose reduction in HIV type 1-infected individuals with cytochrome P450 2B6 *6 and *26. *Clin. Infect. Dis.* 45, 1230–1237. doi:10.1086/522175

Gengiah, T. N., Botha, J. H., Soowamber, D., Naidoo, K., and Abdool Karim, S. S. (2014). Low rifampicin concentrations in tuberculosis patients with HIV infection. *J. Infect. Dev. Ctries.* 8, 987–993. doi:10.3855/jidc.4696

Giraud, C., Manceau, S., Declèves, X., Goffinet, F., Morini, J. P., Chappuy, H., et al. (2010). Influence of development, HIV infection, and antiretroviral therapies on the gene expression profiles of ABC transporters in human lymphocytes. *J. Clin. Pharmacol.* 50, 226–230. doi:10.1177/0091270009343696

Grammen, C., Baes, M., Haenen, S., Verguts, J., Augustyns, K., Zydowsky, T., et al. (2014). Vaginal expression of efflux transporters and the potential impact on the disposition of microbicides *in vitro* and in rabbits. *Mol. Pharm.* 11, 4405–4414. doi:10.1021/mp5005004

Grant, R. M., Lama, J. R., Anderson, P. L., McMahan, V., Liu, A. Y., Vargas, L., et al. (2010). Preexposure chemoprophylaxis for HIV prevention in men who have sex with men. *N. Engl. J. Med.* 363, 2587–2599. doi:10.1056/nejmoa1011205

Gunawardana, M., Remedios-Chan, M., Miller, C. S., Fanter, R., Yang, F., Marzinke, M. A., et al. (2015). Pharmacokinetics of long-acting tenofovir alafenamide (GS-7340) subdermal implant for HIV prophylaxis. *Antimicrob. Agents Chemother.* 59, 3913–3919. doi:10.1128/aac.00656-15

Hapgood, J. P. (2020). Is the injectable contraceptive depo-medroxyprogesterone acetate (DMPA-IM) associated with an increased risk for HIV acquisition? The jury is still out. *AIDS Res. Hum. Retroviruses* 36, 357–366. doi:10.1089/aid.2019.0228

Heffron, R., Donnell, D., Rees, H., Celum, C., Mugo, N., Were, E., et al. (2012). Use of hormonal contraceptives and risk of HIV-1 transmission: A prospective cohort study. *Lancet Infect. Dis.* 12, 19–26. doi:10.1016/s1473-3099(11)70247-x

Heffron, R., McClelland, R. S., Balkus, J. E., Celum, C., Cohen, C. R., Mugo, N., et al. (2017). Efficacy of oral pre-exposure prophylaxis (PrEP) for HIV among women with abnormal vaginal microbiota: A post-hoc analysis of the randomised, placebo-controlled partners PrEP study. *Lancet HIV* 4, e449–e456. doi:10.1016/s2352-3018(17)30110-8

Hijazi, K., Cuppone, A. M., Smith, K., Stincarelli, M. A., Ekeruche-Makinde, J., De Falco, G., et al. (2015). Expression of genes for drug transporters in the human female genital tract and modulatory effect of antiretroviral drugs. *PloS one* 10, e0131405. doi:10.1371/journal.pone.0131405

Hockings, J. K., Pasternak, A. L., Erwin, A. L., Mason, N. T., Eng, C., and Hicks, J. K. (2020). Pharmacogenomics: An evolving clinical tool for precision medicine. *Ccjm* 87, 91–99. doi:10.3949/ccjm.87a.19073

Hoffmeyer, S., Burk, O., von Richter, O., Arnold, H. P., Brockmöller, J., Johne, A., et al. (2000). Functional polymorphisms of the human multidrug-resistance gene:

Multiple sequence variations and correlation of one allele with P-glycoprotein expression and activity *in vivo*. *Proc. Natl. Acad. Sci. U.S.A.* 97, 3473–3478. doi:10.1073/pnas.97.7.3473

Hoque, T. M. D., Cattin, A., Whyte-Allman, S. K., Winchester, L., Fletcher, C. V., Routy, J. P., et al. (2021). Antiretroviral drug transporters and metabolic enzymes in circulating monocytes and monocyte-derived macrophages of ART-treated people living with HIV and HIV-uninfected individuals. *J. Acquir Immune Defic. Syndr.* 87, 1093–1101. doi:10.1097/qai.0000000000002682

Hu, M., Patel, S. K., Zhou, T., and Rohan, L. C. (2015). Drug transporters in tissues and cells relevant to sexual transmission of HIV: Implications for drug delivery. *J. Control. Release* 219, 681–696. doi:10.1016/j.jconrel.2015.08.018

Hunt, R., Sauna, Z. E., Ambudkar, S. V., Gottesman, M. M., and Kimchi-Sarfaty, C. (2009). Silent (synonymous) SNPs: Should we care about them? *Methods Mol. Biol.* 578, 23–39. doi:10.1007/978-1-60327-411-1_2

Ismail, S., and Essawi, M. (2012). Genetic polymorphism studies in humans. *Middle East J. Med. Genet.* 1, 57–63. doi:10.1097/01.mxe.0000415225.85003.47

Janes, H., Corey, L., Ramjee, G., Carpp, L. N., Lombard, C., Cohen, M. S., et al. (2018). Weighing the evidence of efficacy of oral PrEP for HIV prevention in women in southern Africa. *AIDS Res. Hum. Retroviruses* 34, 645–656. doi:10.1089/aid.2018.0031

Jorajuria, S., Dereuddre-Bosquet, N., Becher, F., Martin, S., Porcheray, F., Garrigues, A., et al. (2004). ATP binding cassette multidrug transporters limit the anti-HIV activity of zidovudine and indinavir in infected human macrophages. *Antivir. Ther.* 9, 519–528. doi:10.1177/135965350400900403

Karim, S. S., Kashuba, A. D., Werner, L., and Karim, Q. A. (2011). Drug concentrations after topical and oral antiretroviral pre-exposure prophylaxis: Implications for HIV prevention in women. *Lancet* 378, 279–281. doi:10.1016/s0140-6736(11)60878-7

Kerb, R. (2006). Implications of genetic polymorphisms in drug transporters for pharmacotherapy. *Cancer Lett.* 234, 4–33. doi:10.1016/j.canlet.2005.06.051

Kis, O., Robillard, K., Chan, G. N., and Bendayan, R. (2010). The complexities of antiretroviral drug-drug interactions: Role of ABC and SLC transporters. *Trends Pharmacol. Sci.* 31, 22–35. doi:10.1016/j.tips.2009.10.001

Klatt, N. R., Cheu, R., Birse, K., Zevin, A. S., Perner, M., Noël-Romas, L., et al. (2017). Vaginal bacteria modify HIV tenofovir microbicide efficacy in African women. *science* 356, 938–945. doi:10.1126/science.aai9383

Landovitz, R. J., Li, S., Grinsztejn, B., Dawood, H., Liu, A. Y., Magnus, M., et al. (2018). Safety, tolerability, and pharmacokinetics of long-acting injectable cabotegravir in low-risk HIV-uninfected individuals: HPTN 077, a phase 2a randomized controlled trial. *PLoS Med.* 15, e1002690. doi:10.1371/journal.pmed.1002690

Le Vee, M., Jouan, E., Moreau, A., and Fardel, O. (2011). Regulation of drug transporter mRNA expression by interferon-γ in primary human hepatocytes. *Fundam. Clin. Pharmacol.* 25, 99–103. doi:10.1111/j.1472-8206.2010.00822.x

Likanonsakul, S., Suntisuklappon, B., Nitiyanontakij, R., Prasithsirikul, W., Nakayama, E. E., Shioda, T., et al. (2016). A single-nucleotide polymorphism in ABCC4 is associated with tenofovir-related beta2-microglobulinuria in Thai patients with HIV-1 infection. *PloS one* 11, e0147724. doi:10.1371/journal.pone.0147724

Lin, L., Yee, S. W., Kim, R. B., and Giacomini, K. M. (2015). SLC transporters as therapeutic targets: Emerging opportunities. *Nat. Rev. Drug Discov.* 14, 543–560. doi:10.1038/nrd4626

Maartens, G., Celum, C., and Lewin, S. R. (2014). HIV infection: Epidemiology, pathogenesis, treatment, and prevention. *Lancet* 384, 258–271. doi:10.1016/s0140-6736(14)60164-1

Mabaso, M., Sokhela, Z., Mohlabane, N., Chibi, B., Zuma, K., and Simbayi, L. (2018). Determinants of HIV infection among adolescent girls and young women aged 15-24 years in South Africa: A 2012 population-based national household survey. *BMC Public Health* 18, 183. doi:10.1186/s12889-018-5051-3

Manosuthi, W., Sukasem, C., Thongyen, S., Nilkamhang, S., and Sungkanuparph, S. (2014). ABCC2*1C and plasma tenofovir concentration are correlated to decreased glomerular filtration rate in patients receiving a tenofovir-containing antiretroviral regimen. *J. Antimicrob. Chemother.* 69, 2195–2201. doi:10.1093/jac/dku129

Marrazzo, J. M., Ramjee, G., Richardson, B. A., Gomez, K., Mgodi, N., Nair, G., et al. (2015). Tenofovir-based preexposure prophylaxis for HIV infection among African women. *N. Engl. J. Med.* 372, 509–518. doi:10.1056/nejmoa1402269

Masimirembwa, C., Dandara, C., and Leutscher, P. D. (2016). Rolling out efavirenz for HIV precision medicine in Africa: Are we ready for pharmacovigilance and tackling neuropsychiatric adverse effects? *OMICS A J. Integr. Biol.* 20, 575–580. doi:10.1089/omi.2016.0120

Masson, L., Mlisana, K., Little, F., Werner, L., Mkhize, N. N., Ronacher, K., et al. (2014). Defining genital tract cytokine signatures of sexually transmitted infections and bacterial vaginosis in women at high risk of HIV infection: A cross-sectional study. *Sex. Transm. Infect.* 90, 580–587. doi:10.1136/sextrans-2014-051601

Masson, L., Passmore, J.-a. S., Liebenberg, L. J., Werner, L., Baxter, C., Arnold, K. B., et al. (2015). Genital inflammation and the risk of HIV acquisition in women. *Clin. Infect. Dis.* 61, 260–269. doi:10.1093/cid/civ298

McCormack, S., and Dunn, D. (2015). Pragmatic open-label randomised trial of preexposure prophylaxis: The PROUD study. *Lancet* 387, 23–26. doi:10.1016/S0140-6736(15)00056-2

McCormack, S., Dunn, D. T., Desai, M., Dolling, D. I., Gafos, M., Gilson, R., et al. (2016). Pre-exposure prophylaxis to prevent the acquisition of HIV-1 infection (PROUD): Effectiveness results from the pilot phase of a pragmatic open-label randomised trial. *Lancet* 387, 53–60. doi:10.1016/s0140-6736(15)00056-2

McCormack, S., Ramjee, G., Kamali, A., Rees, H., Crook, A. M., Gafos, M., et al. (2010). PRO2000 vaginal gel for prevention of HIV-1 infection (microbicides development Programme 301): A phase 3, randomised, double-blind, parallel-group trial. *Lancet* 376, 1329–1337. doi:10.1016/s0140-6736(10)61086-0

McKinnon, L. R., Liebenberg, L. J., Yende-Zuma, N., Archary, D., Ngcapu, S., Sivro, A., et al. (2018). Genital inflammation undermines the effectiveness of tenofovir gel in preventing HIV acquisition in women. *Nat. Med.* 24, 491–496. doi:10.1038/nm.4506

Molina, J.-M., Capitant, C., Spire, B., Pialoux, G., Chidiac, C., Charreau, I., et al. (2015). On demand PrEP with oral TDF-FTC in MSM: Results of the ANRS ipergay trial. *N. Engl. J. Med.* 373, 2237–2246. doi:10.1056/NEJMoa1506273

Mukonzo, J. K., Röshammar, D., Waako, P., Andersson, M., Fukasawa, T., Milani, L., et al. (2009). A novel polymorphism in ABCB1 gene, CYP2B6*6 and sex predict single-dose efavirenz population pharmacokinetics in Ugandans. *Br. J. Clin. Pharmacol.* 68, 690–699. doi:10.1111/j.1365-2125.2009.03516.x

Myer, L., Denny, L., Wright, T. C., and Kuhn, L. (2007). Prospective study of hormonal contraception and women's risk of HIV infection in South Africa. *Int. J. Epidemiol.* 36, 166–174. doi:10.1093/ije/dyl251

Ngaimisi, E., Habtewold, A., Minzi, O., Makonnen, E., Mugusi, S., Amogne, W., et al. (2013). Importance of ethnicity, CYP2B6 and ABCB1 genotype for efavirenz pharmacokinetics and treatment outcomes: A parallel-group prospective cohort study in two sub-saharan Africa populations. *PloS one* 8, e67946. doi:10.1371/journal.pone.0067946

Nicol, M. R., Corbino, J. A., and Cottrell, M. L. (2018). Pharmacology of antiretrovirals in the female genital tract for HIV prevention. *J. Clin. Pharmacol.* 58, 1381–1395. doi:10.1002/jcph.1270

Nicol, M. R., Fedoriw, Y., Mathews, M., Prince, H. M., Patterson, K. B., Geller, E., et al. (2014). Expression of six drug transporters in vaginal, cervical, and colorectal tissues: Implications for drug disposition in HIV prevention. *J. Clin. Pharmacol.* 54, 574–583. doi:10.1002/jcph.248

Nishijima, T., Komatsu, H., Higasa, K., Takano, M., Tsuchiya, K., Hayashida, T., et al. (2012). Single nucleotide polymorphisms in ABCC2 associate with tenofovir-induced kidney tubular dysfunction in Japanese patients with HIV-1 infection: A pharmacogenetic study. *Clin. Infect. Dis.* 55, 1558–1567. doi:10.1093/cid/cis772

Patterson, K. B., Prince, H. A., Kraft, E., Jenkins, A. J., Shaheen, N. J., Rooney, J. F., et al. (2011). Penetration of tenofovir and emtricitabine in mucosal tissues: Implications for prevention of HIV-1 transmission. *Sci. Transl. Med.* 3, 112re4. doi:10.1126/scitranslmed.3003174

Person, K., Hicks, C. B., Hick, B., and Charles, S. (2012). Pre-exposure prophylaxis--one more tool for HIV prevention. *Curr. HIV Res.* 10, 117–122. doi:10.2174/157016212799937254

Petrovic, V., Teng, S., and Piquette-Miller, M. (2007). Regulation of drug transporters: During infection and inflammation. *Mol. Interv.* 7, 99–111. doi:10.1124/mi.7.2.10

Petzinger, E., and Geyer, J. (2006). Drug transporters in pharmacokinetics. *Naunyn Schmied. Arch. Pharmacol.* 372, 465–475. doi:10.1007/s00210-006-0042-9

Poller, B., Drewe, J., Krähenbühl, S., Huwyler, J., and Gutmann, H. (2010). Regulation of BCRP (ABCG2) and P-glycoprotein (ABCB1) by cytokines in a model of the human blood-brain barrier. *Cell Mol. Neurobiol.* 30, 63–70. doi:10.1007/s10571-009-9431-1

Rajman, I., Knapp, L., and Hanna, I. (2020). Genetic diversity in drug transporters: Impact in african populations. *Clin. Transl. Sci.* 13, 848–860. doi:10.1111/cts.12769

Ramjee, G., and Daniels, B. (2013). Women and HIV in sub-saharan Africa. *AIDS Res. Ther.* 10, 30. doi:10.1186/1742-6405-10-30

Ray, A. S., Cihlar, T., Robinson, K. L., Tong, L., Vela, J. E., Fuller, M. D., et al. (2006). Mechanism of active renal tubular efflux of tenofovir. *Antimicrob. Agents Chemother.* 50, 3297–3304. doi:10.1128/aac.00251-06

Reznicek, J., Ceckova, M., Cerveny, L., Müller, F., and Staud, F. (2017). Emtricitabine is a substrate of MATE1 but not of OCT1, OCT2, P-gp, BCRP or MRP2 transporters. *Xenobiotica* 47, 77–85. doi:10.3109/00498254.2016.1158886

Rohan, L. C., and Sassi, A. B. (2009). Vaginal drug delivery systems for HIV prevention. *Aaps J.* 11, 78–87. doi:10.1208/s12248-009-9082-7

Rungtivasuwan, K., Avihingsanon, A., Thammajaruk, N., Mitruk, S., Burger, D. M., Ruxrungtham, K., et al. (2015). Influence of ABCC2 and ABCC4 polymorphisms on tenofovir plasma concentrations in Thai HIV-infected patients. *Antimicrob. Agents Chemother.* 59, 3240–3245. doi:10.1128/aac.04930-14

Saib, S., and Delavenne, X. (2021). Inflammation induces changes in the functional expression of P-gp, BCRP, and MRP2: An overview of different models and consequences for drug disposition. *Pharmaceutics* 13, 1544. doi:10.3390/pharmaceutics13101544

Salvaggio, S. E., Giacomelli, A., Falvella, F. S., Oreni, M., Meraviglia, P., Atzori, C., et al. (2017). Clinical and genetic factors associated with kidney tubular dysfunction in a real-life single centre cohort of HIV-positive patients. *BMC Infect. Dis.* 17, 396. doi:10.1186/s12879-017-2497-3

Schaeffeler, E., Eichelbaum, M., Brinkmann, U., Penger, A., Asante-Poku, S., Zanger, U. M., et al. (2001). Frequency of C3435T polymorphism of MDR1 gene in African people. *Lancet* 358, 383–384. doi:10.1016/s0140-6736(01)05579-9

Shen, R., Kappes, J. C., Smythies, L. E., Richter, H. E., Novak, L., and Smith, P. D. (2014). Vaginal myeloid dendritic cells transmit founder HIV-1. *J. Virol.* 88, 7683–7688. doi:10.1128/jvi.00766-14

Shen, Z., Rodriguez-Garcia, M., Patel, M. V., Bodwell, J., Kashuba, A. D., and Wira, C. R. (2017). Hormonal contraceptives differentially suppress TFV and TAF inhibition of HIV infection and TFV-DP in blood and genital tract CD4+ T cells. *Sci. Rep.* 7, 17697–17712. doi:10.1038/s41598-017-18078-3

Shenfield, G. M. (2004). Genetic polymorphisms, drug metabolism and drug concentrations. *Clin. Biochem. Rev.* 25, 203–206.

Sissung, T. M., Troutman, S. M., Campbell, T. J., Pressler, H. M., Sung, H., Bates, S. E., et al. (2012). Transporter pharmacogenetics: Transporter polymorphisms affect normal physiology, diseases, and pharmacotherapy. *Discov. Med.* 13, 19–34.

Skoler-Karpoff, S., Ramjee, G., Ahmed, K., Altini, L., Plagianos, M. G., Friedland, B., et al. (2008). Efficacy of carraguard for prevention of HIV infection in women in South Africa: A randomised, double-blind, placebo-controlled trial. *Lancet* 372, 1977–1987. doi:10.1016/s0140-6736(08)61842-5

Stankiewicz, P., and Lupski, J. R. (2010). Structural variation in the human genome and its role in disease. *Annu. Rev. Med.* 61, 437–455. doi:10.1146/annurev-med-100708-204735

Stanley, M. (2009). Early age of sexual debut: A risky experience. *J. Fam. Plann Reprod. Health Care* 35, 118–120. doi:10.1783/147118909787931979

Sukhai, M., Yong, A., Kalitsky, J., and Piquette-Miller, M. (2000). Inflammation and interleukin-6 mediate reductions in the hepatic expression and transcription of the mdr1a and mdr1b Genes. *Mol. Cell Biol. Res. Commun.* 4, 248–256. doi:10.1006/mcbr.2001.0288

Suzuki, K., Kawakami, Y., and Yamauchi, K. (2017). Impact of TLR 2, TLR 4-activation on the expression of ABCA1 and ABCG1 in raw cells. *Ann. Clin. Lab. Sci.* 47, 436–446.

Swart, M., Ren, Y., Smith, P., and Dandara, C. (2012). ABCB1 4036A>G and 1236C>T polymorphisms affect plasma efavirenz levels in South African HIV/AIDS patients. *Front. Gene.* 3, 236. doi:10.3389/fgene.2012.00236

Taneva, E., Crooker, K., Park, S. H., Su, J. T., Ott, A., Cheshenko, N., et al. (2016). Differential mechanisms of tenofovir and tenofovir disoproxil fumarate cellular transport and implications for topical preexposure prophylaxis. *Antimicrob. Agents Chemother.* 60, 1667–1675. doi:10.1128/aac.02793-15

Thigpen, M. C., Kebaabetswe, P. M., Paxton, L. A., Smith, D. K., Rose, C. E., Segolodi, T. M., et al. (2012). Antiretroviral preexposure prophylaxis for heterosexual HIV transmission in Botswana. *N. Engl. J. Med.* 367, 423–434. doi:10.1056/nejmoa1110711

Tsai, C. C., Follis, K. E., Sabo, A., Beck, T. W., Grant, R. F., Bischofberger, N., et al. (1995). Prevention of SIV infection in macaques by (R)-9-(2-Phosphonylmethoxypropyl)adenine. *science* 270, 1197–1199. doi:10.1126/science.270.5239.1197

Tun-Yhong, W., Chinpaisal, C., Pamonsinlapatham, P., and Kaewkitichai, S. (2017). Correction for Tun-Yhong et al., "Tenofovir Disoproxil Fumarate Is a New Substrate of ATP-Binding Cassette Subfamily C Member 11". *Antimicrob. Agents Chemother.* 61, 17. doi:10.1128/AAC.01753-17

Turriziani, O., Gianotti, N., Falasca, F., Boni, A., Vestri, A. R., Zoccoli, A., et al. (2008). Expression levels of MDR1, MRP1, MRP4, and MRP5 in peripheral blood mononuclear cells from HIV infected patients failing antiretroviral therapy. *J. Med. Virol.* 80, 766–771. doi:10.1002/jmv.21152

UNAIDS (2021). UNAIDS data 2021. Retrieved from: https://www.unaids.org/sites/default/files/media_asset/JC3032_AIDS_Data_book_2021_En.pdf (Accessed Juanuary 31, 2021).

Uwai, Y., Ida, H., Tsuji, Y., Katsura, T., and Inui, K.-I. (2007). Renal transport of adefovir, cidofovir, and tenofovir by SLC22A family members (hOAT1, hOAT3, and hOCT2). *Pharm. Res.* 24, 811–815. doi:10.1007/s11095-006-9196-x

Van Damme, L., Corneli, A., Ahmed, K., Agot, K., Lombaard, J., Kapiga, S., et al. (2012). Preexposure prophylaxis for HIV infection among African women. *N. Engl. J. Med.* 367, 411–422. doi:10.1056/nejmoa1202614

Vee, M., Lecureur, V., Stieger, B., and Fardel, O. (2009). Regulation of drug transporter expression in human hepatocytes exposed to the proinflammatory cytokines tumor necrosis factor-α or interleukin-6. *Drug Metab. Dispos.* 37, 685–693. doi:10.1124/dmd.108.023630

Weiner, M., Peloquin, C., Burman, W., Luo, C. C., Engle, M., Prihoda, T. J., et al. (2010). Effects of tuberculosis, race, and human gene SLCO1B1 polymorphisms on rifampin concentrations. *Antimicrob. Agents Chemother.* 54, 4192–4200. doi:10.1128/aac.00353-10

Wilkinson, D., Tholandi, M., Ramjee, G., and Rutherford, G. W. (2002). Nonoxynol-9 spermicide for prevention of vaginally acquired HIV and other sexually transmitted infections: Systematic review and meta-analysis of randomised controlled trials including more than 5000 women. *Lancet Infect. Dis.* 2, 613–617. doi:10.1016/s1473-3099(02)00396-1

Yee, S. W., Brackman, D. J., Ennis, E. A., Sugiyama, Y., Kamdem, L. K., Blanchard, R., et al. (2018). Influence of transporter polymorphisms on drug disposition and response: A perspective from the international transporter Consortium. *Clin. Pharmacol. Ther.* 104, 803–817. doi:10.1002/cpt.1098

Zhou, T., Hu, M., Cost, M., Poloyac, S., and Rohan, L. (2013). Short communication: Expression of transporters and metabolizing enzymes in the female lower genital tract: Implications for microbicide research. *AIDS Res. Hum. retroviruses* 29, 1496–1503. doi:10.1089/aid.2013.0032

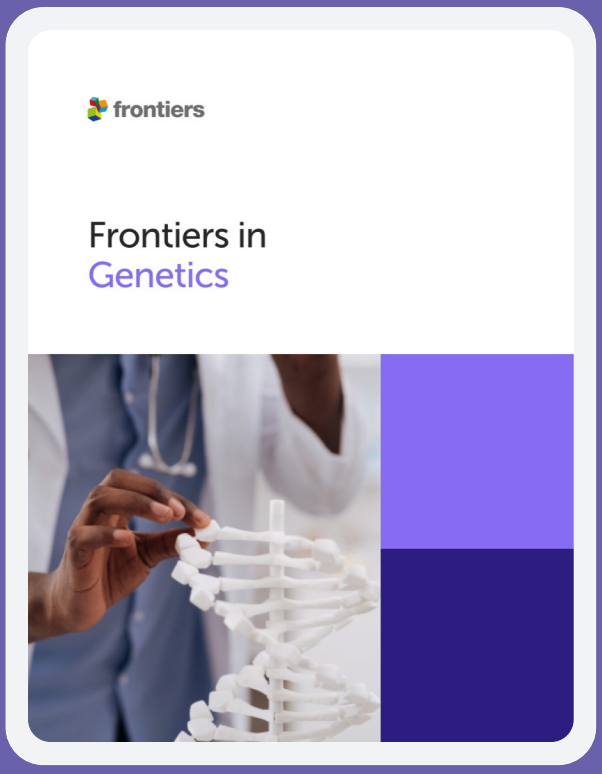
Frontiers in Genetics

Highlights genetic and genomic inquiry relating to all domains of life

The most cited genetics and heredity journal, which advances our understanding of genes from humans to plants and other model organisms. It highlights developments in the function and variability of the genome, and the use of genomic tools.

Discover the latest Research Topics

See more →



Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact