



Dissecting *in silico* Mutation Prediction of Variants in African Genomes: Challenges and Perspectives

Christian Domilongo Bope^{1,2*}, Emile R. Chimusa¹, Victoria Nembaware¹, Gaston K. Mazandu¹, Jantina de Vries³ and Ambroise Wonkam^{1,3,4}

¹ Department of Pathology, Faculty of Health Sciences, University of Cape Town, Cape Town, South Africa, ² Departments of Mathematics and Computer Sciences, Faculty of Sciences, University of Kinshasa, Kinshasa, Democratic Republic of Congo, ³ Department of Medicine, Faculty of Health Sciences, University of Cape Town, Cape Town, South Africa, ⁴ Institute of Infectious Diseases and Molecular Medicine, Faculty of Health Sciences, University of Cape Town, Cape Town, South Africa

OPEN ACCESS

Edited by:

Tuo Zhang,
Cornell University, United States

Reviewed by:

Erik Garrison,
Wellcome Sanger Institute,
United Kingdom
Marcelo Adrian Marti,
University of Buenos Aires, Argentina

*Correspondence:

Christian Domilongo Bope
christian.bope@uct.ac.za

Specialty section:

This article was submitted to
Bioinformatics and Computational
Biology,
a section of the journal
Frontiers in Genetics

Received: 28 February 2019

Accepted: 05 June 2019

Published: 25 June 2019

Citation:

Bope CD, Chimusa ER,
Nembaware V, Mazandu GK,
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

Genomic medicine is set to drastically improve clinical care globally due to high throughput technologies which enable speedy *in silico* detection and analysis of clinically relevant mutations. However, the variability in the *in silico* prediction methods and categorization of functionally relevant genetic variants can pose specific challenges in some populations. *In silico* mutation prediction tools could lead to high rates of false positive/negative results, particularly in African genomes that harbor the highest genetic diversity and that are disproportionately underrepresented in public databases and reference panels. These issues are particularly relevant with the recent increase in initiatives, such as the Human Heredity and Health (H3Africa), that are generating huge amounts of genomic sequence data in the absence of policies to guide genomic researchers to return results of variants in so-called actionable genes to research participants. This report (i) provides an inventory of publicly available Whole Exome/Genome data from Africa which could help improve reference panels and explore the frequency of pathogenic variants in actionable genes and related challenges, (ii) reviews available *in silico* prediction mutation tools and the criteria for categorization of pathogenicity of novel variants, and (iii) proposes recommendations for analyzing pathogenic variants in African genomes for their use in research and clinical practice. In conclusion, this work proposes criteria to define mutation pathogenicity and actionability in human genetic research and clinical practice in Africa and recommends setting up an African expert panel to oversee the proposed criteria.

Keywords: African genome, incidental findings, actionable variants, whole exome sequencing, whole genome sequencing, precision medicine, pathogenicity

INTRODUCTION

High throughput technologies in “omics” research are expected to improve clinical care globally through genomic medicine. However, the categorization and criteria to infer variants’ pathogenicity differs around the world and can pose specific challenges in some populations (Dorschner et al., 2013; Green et al., 2013; MacArthur et al., 2014; Amendola et al., 2015;

Hunter et al., 2016; Ichikawa et al., 2017; Kwak et al., 2017; Lacaze et al., 2017; Tang et al., 2018). Particularly, in African genomes that harbor the highest genetic diversity, it is possible that most *in silico* prediction tools could lead to the highest rate of false positive/negative results (Martin et al., 2018). The H3Africa Consortium has significantly contributed to reducing the dearth of genomic research on the African continent by supporting African genomics researchers and developing policies (Dandara et al., 2014; H3Africa, 2017). However, in the current genomics landscape, it is particularly challenging to interpret some variants found in African genomes, i.e., to determine whether that variant is common or rare, benign or pathogenic. Firstly, approaches to determine the rareness of a variant are based on exploring publicly available genome reference databases in which African data are under-represented (Lek et al., 2016; Popejoy and Fullerton, 2016). In addition, most of the current well-established bioinformatics tools, variant calling pipelines, are benchmarked using non-African populations and most of the variants deposited in the public database are from non-African populations (Pabinger et al., 2012; Bao et al., 2014). Secondly, the high genetic diversity of African populations means that genomic studies are likely to detect many novel variants that are yet to be described in current public databases (Lebeko et al., 2017). Thirdly, there is a lack of evidence-based policies and guidelines to inform the characterization of actionable genes in African genomic research. A guideline on feeding back findings was recently developed by H3Africa; while this is a commendable achievement, it lacks the support of published empirical evidence¹. This latter point is particularly important given the recent call from the American College of Medical Genetics (ACMG) to investigate pathogenic variants in so-called actionable genes that could potentially have direct clinical benefit, and to return the results to research participants (ACMG, 2013). This will open up a series of ethically relevant questions (Kiezun et al., 2012; MacArthur et al., 2014; Parker and Kwiatkowski, 2016), such as the definition of actionability and relevance to personalized medicine in a context of often scarce human and material resources, and ill-equipped healthcare systems (Masimirembwa et al., 2014).

To address these multiple challenges, and particularly that of variant interpretation in African genomes, it is appropriate to develop new pipelines using African genetics data or to benchmark existing bioinformatics pipeline tools using African populations to account for African genetic diversity. This paper aims to (i) provide an inventory of existing Whole Exome/Genome data from Africans that could help develop an African reference genome build, improve reference panels, and explore the frequency of pathogenic variants in actionable genes and related challenges; (ii) review available *in silico* prediction mutation tools and criteria for categorization of pathogenicity of novel variants; and (iii) propose recommendations for analyzing pathogenic

variants in African genomes for their use in research and clinical practice.

CURRENT CHALLENGES OF WES/WGS DATA INTERPRETATION IN AFRICANS

Mastering of genome sequencing pipelines and downstream analysis are important for inferring meaningful information, such as detection of variants in medically relevant genes, from high throughput data such as Next Generation Sequencing (NGS), Whole Exome Sequencing (WES), or Whole Genome Sequencing (WGS). However, data processing, deep sequencing, and meticulous downstream analysis of WES/WGS still constitute a challenge in most of the current pipelines and tools. In addition, there are still some challenges, such as the interpretation of rare missense variants, reliability, and accuracy of pipelines for sequence alignments, variant calling, and data analysis, for the WES and WGS data of African populations (Wang et al., 2013; Rabbani et al., 2014; Bertier et al., 2016; Popejoy and Fullerton, 2016). To address some of these challenges, a plethora of bioinformatics algorithms and pipelines have been developed (Pabinger et al., 2012; Hentzschke et al., 2016; Xu, 2018). Current practice is to use existing variant calling pipelines, but this raises a number of questions, including how are universally reliable and accurate current WES/WGS bioinformatics tools and pipelines benchmarked using non-African data? What is the true proportion of African population data in the current reference genome builds that are publicly available, taking into account the variable level of admixture of African Americans who tend to be considered proxies of Africans in these databases [the Genome Reference Consortium Human Genome (GRCh3) and University of California, Santa Cruz (UCSC)] (Kuhn et al., 2009; Fujita et al., 2011; Leipzig, 2017)? Addressing these challenges will require that genomic research communities from the African continent develop an African benchmark bioinformatics pipeline to analyze genomic data that includes genetic diversity found in the African populations, and engage in a major effort in constructing an African-specific reference panel.

African populations in current reference panels are not representative of more differentiated population groups within Africa. Variant calling from NGS data is based on alignment to a single reference genome, which is problematic for diverse regions or populations, such as African populations. There is great opportunity in improving read alignment and variant calling for African genomes. A genome reference graph for alignment and variant calling may capture natural variation among populations, particularly populations of high diversity with low level of linkage disequilibrium.

Repetitive DNA sequences are abundant in a broad range of species, from bacteria to mammals, and they cover nearly half of the human genome. The other main issue is that repeats have always presented technical challenges for sequence alignment and assembly programs. NGS projects, with their short read lengths and high data volumes, have made these challenges more difficult. From a computational perspective, repeats create ambiguities in

¹<https://h3africa.org/wp-content/uploads/2018/05/H3Africa%20Feedback%20of%20Individual%20Genetic%20Results%20Policy.pdf>

alignment and assembly, which, in turn, can produce biases and errors when interpreting results. Simply ignoring repeats is not an option, as this creates problems of its own and may mean that important biological phenomena are missed. Variation in repeats can alter the expression of genes, and changes in the number of repeats have been linked to certain human diseases. Unfortunately, the molecular characterization of these repeats has been hampered by technical limitations related to cloning, sequencing techniques, and alignment algorithms (Dilthey et al., 2014; Marcus et al., 2014; Church et al., 2015; Paten et al., 2017).

Fortunately, the number of genomic researchers in Africa is on the rise, which has led to an increase in African genomic data and publications (The H3Africa Consortium, 2014; Uthman et al., 2015; Mulder et al., 2016; Ndiaye Diallo et al., 2017). The increase in African genomic research has the potential to narrow the research gap between Africa and the rest of the world and can also improve implementation of genomic medicine. Therefore, we propose to use the available data to (i) develop Bioinformatics tools using African data, particularly for populations from sub-Saharan Africa who have the highest genetic diversity and low levels of admixture with European or Asian populations; (ii) benchmark existing tools using available African population data; and (iii) there is an urgent need for a centralized repository of publicly available African genomic data with annotated variants based on their pathogenicity, in order to increase our understanding of continental genomic diversity (Jongeneel et al., 2017; Mulder et al., 2017; Ahmed et al., 2018). To help initiate such endeavors we have provided here an inventory of African Whole Exome and Whole Genome data that are currently available to our knowledge (Table 1).

IN SILICO PREDICTION OF MUTATIONS AND CHALLENGES

The accuracy of variant calling pipelines (Li et al., 2009; DePristo et al., 2011; Wei et al., 2011; Garrison and Marth, 2012; Koboldt et al., 2012; Wilm et al., 2012; Lai et al., 2016) is a major step prior to the downstream *in silico* prediction of mutations. Nevertheless, a challenge remains in downstream NGS variant calling analysis, i.e., to distinguish pathogenic mutations and rare non-pathogenic variants from most of the annotating variant calling pipelines. The accuracy of *in silico* prediction of rare and actionable disease-causing genetic variants for the detection of pathogenic rare mutations and polymorphisms is the greatest challenge. Variant calling pipelines generate large numbers variations erroneously, which may contain rare, common genetic variants, false positives, and false negatives (Dong et al., 2015). Further downstream analysis such as variant annotations, variant filtrations, and prioritization methods are conducted to annotate variant genomic features, gene symbols, exonic functions, and amino acid modifications (Bao et al., 2014). Different *in silico* prediction algorithms are implemented to annotate disease-causing mutations based on the following information from the variants: (i) sequence homology (Reva et al., 2011), (ii) protein structure (Ng and Henikoff, 2006; Teng et al., 2009), (iii) evolutionary conservation (Cooper et al., 2010), (iv) the

frequency of pathogenicity (Kobayashi et al., 2017), and (v) change in ancestry. Most of the *in silico* prediction methods interact with public databases to incorporate updated variant information in order to enhance annotation prediction efficiency. The incorporated information is mainly the minor allele frequency (MAF), experimental clinical assay information and deleterious prediction of variants (Pabinger et al., 2012). The majority of *in silico* prediction tools provide a reduced number of annotations from large background errors of detected variants. To annotate, filter, and prioritize accurately variant calling, researchers developed pipelines combined with different annotation tools and databases. Germline and somatic mutation databases, such as ANNOVAR (Wang et al., 2010; Yang and Wang, 2015), Human Gene Mutation Database², dbSNP³ (Sherry et al., 2001), and GENEKEEPER⁴ and others are important for evaluating variants. Liu et al. (2011) developed a robust database called dbNSFP, which combines the prediction scores of six prediction algorithms namely SIFT (Kumar et al., 2009), PolyPhen-2 (Adzhubei et al., 2010), LRT (Chun and Fay, 2009), MutationTaster (Schwarz et al., 2010), Mutation Assessor (Reva et al., 2011), FATHMM (Chun and Fay, 2009; Shihab et al., 2013), and conservative score tools namely GERP++ (Davydov et al., 2010), SiPhy (Garber et al., 2009), and PhyloP (Doerks et al., 2002) and then compiles the scores of these tools into one (Liu et al., 2011). ClinVar is a commonly used database for germline variants, namely pathogenic and benign and provides related clinical and experimental information⁵ (Landrum et al., 2016).

After annotation, it is recommended to filter annotated variants from many tools using two approaches (i) free hypothesis, to cast the vote of the annotated variant filters for “Deleterious or damaging disease-causing (D)” or “disease-causing automatic (A)” among annotation prediction tools based on a defined cut-off (~50%); and (ii) non-free hypothesis, which provides a list of known genes of the studies with another level of prediction cut-off (~25%). The cut-off for both hypotheses is study related.

In silico prediction of mutations in the context of African populations introduces additional specific challenges that are partly related to the use of non-African populations to benchmark *in silico* prediction pipelines and the low proportion of African population data in most of the interrogated databases. Another challenge when working with African population data is the annotation of common variants specific to African populations, which can be considered as pathogenic variants when using public databases. This emphasizes the need for a guideline, which defines approaches to infer pathogenicity variants in African populations.

Predicting Pathogenic Variants and Challenges

In the literature and in most annotation databases, the classification of pathogenicity differs (Sherry et al., 2001;

²<http://www.hgmd.cf.ac.uk/ac/index.php>

³<https://www.ncbi.nlm.nih.gov/projects/SNP/>

⁴<https://kewinc.com/analytics/>

⁵<https://www.ncbi.nlm.nih.gov/clinvar/>

TABLE 1 | Published whole exome and genomes data from sub-Saharan Africa.

Country	Region	Individuals	References
Exomes			
Botswana	Southern Africa	164	Retshabile et al., 2018
Uganda	Southern Africa	150	Retshabile et al., 2018
Ghanaian	Western Africa	1032	Kodaman et al., 2017
Tunisia	North Africa	7 19	Hamdi et al., 2018 Ben Rekaya et al., 2018
Morocco	North Africa	3	Bousfiha et al., 2017
Cameroonian	Central Africa	179	Sickle Cell Disease Project (Unpublished)
Congo	Central Africa	23	Sickle Cell Disease Project (Unpublished)
Black Xhosa (SA)	Southern Africa	25	Hearing Impairment Project (Unpublished)
African Caribbean (ACB)	Caribbean	98	1000 Genomes project
Esan in Nigeria (ESN)	Western Africa	111	1000 Genomes project
Mende in Sierra Leone (MSL)	Western Africa	98	1000 Genomes project
Yoruba in Ibadan, Nigeria (YRI)	Western Africa	100	1000 Genomes project
Luhya in Webuye, Kenya (LWK)	East Africa	106	1000 Genomes project
African Ancestry in Southwest USA (ASW)	North America	68	1000 Genomes project
Gambia in Western Division, The Gambia (GWD)	Western Africa	120	1000 Genomes project
African American	North America	761	Auer et al., 2012
Genomes			
Baganda, Banyarwanda, Barundi (Uganda); Luhya, Kikuyu, Kalenjin (Kenya); Sotho, Zulu (South Africa); Yoruba, Igbo (Nigeria); Ga-Adangbe (Ghana); Jola, Fula, Wolof, Mandika (Gambia); Amhara, Oromo, Somali (Ethiopia)	Sub Sahara (Eastern Africa, Southern Africa, Western Africa, Southern Africa)	320	Gurdasani et al., 2015
Kombo (Gambia)	Western Africa	2560	Jallow et al., 2009
South Africa	Southern Africa	13	Kramvis et al., 2002
Ovamboland (Namibia) Angola Madagascar	Southern Africa Southern Africa Madagascar	23 2 1	Kramvis et al., 2005
Algeria, Morocco, Libya, Tunisia, Tuareg Congo, Gabon, Cameroun, Nigeria	Northern Africa Sub Sahara	25 59	Fadhlaoui-Zid et al., 2013
Uganda Zimbabwean	Sub Sahara	112 174	Venner et al., 2016
Mandinka II, Serehule, Bambara, Malike, Fulall, Fulal, Mandikal, Wollof, Serere, Manjogo, Jola Mossi, Kasem, Yoruba, Namkam, Semi-Bantu, Akans, Bantu Kaurna, Chonyi, Wabondel, Kambe, Luhya, Maasai, Wasambaa, Giriama, Mzigua Ari, Anuak, Sudanese, Gumuz Oromo, Somali, Wolayta, Afar, Tigray, Amhara Nama, Karretjie, Khomani, Malawi, Herero, Khwe, Ixu, HU/Hoansi, Amaxhosa Sebantú	West African Niger-Congo Central West African Niger-Congo East Africa Niger-Congo East Africa Nilo-Saharan East Africa Afroasiatic Khoesan	2504	Busby et al., 2016
African American	African American	35370	Ng et al., 2014
Ethiopian (Weth)	Sub Sahara	120	Tekola-Ayele et al., 2015
South Africa	Southern Africa	24 (8 Colored, 16 Black)	Choudhury et al., 2017

Wang et al., 2010; Yang and Wang, 2015; Landrum et al., 2016; McLaren et al., 2016). Nevertheless, a common strategy to define pathogenicity involves combining results from many annotation pipelines (Lebeko et al., 2017). Further downstream analyses are gene network analysis and gene enrichment. The purpose of these analyses is to investigate the level of interactions between genes and the annotated variants associated with human phenotypes and then mine affected biological processes, networks, pathways, and molecular functions (Bindea et al., 2009; Warde-Farley et al., 2010; Lebeko et al., 2017).

In the comprehensive standards and guidelines, ACMG and the Association for Molecular Pathology (AMP) define the nomenclature for variants (Table 2). Recommendations for laboratories and clinicians to return incidental findings (IFs) has led to interest toward defining criteria and mechanisms for evaluating pathogenicity and the frequencies of IFs in different populations. For example, Dorschner et al. (2013) analyzed actionable pathogenic variants in 500 European and 500 participants of African descent using exome data. The classifications for pathogenicity (Table 2) included allele

TABLE 2 | Variants pathogenicity categorization.

References	Categorization	Interpretation
Dorschner et al., 2013	Pathogenic	Allele frequency of the identified variant is below cutoff AND segregation can be found in at least two unrelated families
	Likely pathogenic variant of uncertain significance (VUS)	Allele frequency of the identified variant is below cutoff AND identified in at least three unrelated individuals
	VUS	Allele frequency of the discovered variant is below cutoff AND present in less than three unrelated affected individuals
	Likely benign VUS	Allele frequency of the identified variant is below cutoff AND/OR seen in combination with a known pathogenic mutation
Richards et al., 2015 ACGM and AMP	Pathogenic very strong	Null variant (non-sense, frameshift, initiation codon, single or multiexon deletion) in a gene is a known mechanism of disease
	Pathogenic strong	Similar amino acid modification which was previously considered as pathogenic variant independent of nucleotide change
	Pathogenic moderate	Localize in a mutational critical region and very important functional domain without benign variation
	Pathogenic supporting	Cosegregation with disease located in many affected family members in a gene well known to be the cause the disease
	Benign standalone	Allele frequency is greater than 5% in Exome Sequencing Project, 1000 Genomes Project, or Exome Aggregation Consortium
	Benign strong	Allele frequency is greater than expected for disorder
	Benign supporting	Missense variant in a gene for which primarily truncating variants are known to cause disease

frequency of the variants, segregation evidence, and the number of the patients affected with the variants and their status as a *de novo* mutation. The results showed major discrepancies in the frequencies of pathogenic variants among Europeans versus Africans, with an estimated frequency of ~3.4% for those of European descent and ~1.2% for those of African descent. In a similar study, Amendola et al. (2015) investigated IFs in 6503 with 4300 Europeans and 2203 individuals of African descent. In addition, functionally disruptive variant categories were added which represent the expected pathogenic variants as truncating and misplace-causing variants. To validate the results, a comparative analysis was conducted with other clinical and research genetic laboratories and *in silico* pathogenicity scores. The results also showed that those of African descent had a scientifically lower proportion (nearly 50%) of a pathogenic variant in actionable genes compared to European participants. This lower proportion found in both studies could be due to the underrepresentation of populations of African descent in the literature and publicly available databases.

Taking into account the high level of admixture of European ancestry among African Americans and the highest level of diversity among Africans, and poor representativity in public databases as well little clinical genetic research from Africa that is publicly available, it is likely that a similar study could even lead to a much lower proportion of IFs in sub-Saharan African populations. This indicates that there is an urgent need to improve criteria to categorize the pathogenicity when studying African populations, stressing for example investigating an appropriate number of ethnically matched control populations.

Variants Actionability and Challenges

The Clinical Genome Resource (ClinGen) defines actionability as clinically prescribed interventions specific to the genetic disorder

under consideration that is effective for prevention or delay of clinical disease, lowered clinical burden, or improved clinical outcomes in a previously undiagnosed adult and suggested a metric to score clinical actionability (Hunter et al., 2016). Interventions include patient management (e.g., risk-reducing surgery), surveillance, or specific circumstances the patients should avoid (e.g., certain types of anesthesia). The actionability includes interventions to improve outcomes for at-risk family members. Genetic testing recommendations for at-risk family members alone, however, were not considered sufficient to meet the criteria for actionability. In addition, actionability did not include reproductive decision-making.

Alternatively, the 100,000 Genomes Project protocol defines actionable genes as variants with a significant potential to prevent disease morbidity and mortality, if identified before symptoms become apparent. The variants with potentially severe impacts are clinically actionable causes of rare disease, where a healthcare intervention or screening programs might prevent an untoward outcome. The variants are known to result in illness or disability that is clinically significant, severely or moderately life threatening and clinically actionable. It should be emphasized that the exact criteria for considering whether a variant is considered actionable or not, and serious or not, is context-dependent and in some instances only emerges during the process of seeking ethical approval for the study (Genomics England, 2017).

The accepted process consists of defining actionability of the variant and a pathogenicity classification criterion. Both processes are evaluated, inspected and validated by a group of experts (Richards et al., 2015; Hunter et al., 2016). In the African context with highly genetically diverse populations, there is a need to update the proposed scoring metric to take into account the scarcity of health care professionals with medical genetics and

genetic counseling skills, poorly equipped health facilities with a major disparity between urban and rural setting, and generally inadequate health systems.

RETURN OF INCIDENTAL FINDINGS AND CHALLENGES IN AFRICA

Next Generation Sequencing analysis could contribute to the improvement of patient care. This development has blurred the line between genomics and healthcare; the global recommendations on the identification and the return of IFs have raised some ethical concerns for genomic researchers, clinicians, and the public health authority. Prior to returning IFs, there is a need to have clear guidelines and recommendations on a list of potentially actionable genes and define how, what and when IFs should be returned (Ness, 2008; ACMG, 2013, 2015; Souzeau et al., 2016; Nowak et al., 2018). Wolf et al. (2008) published a paper, proposing a framework supporting disclosure of IFs to guide researchers particularly on informed consent, the handling process and the responsibility of institutional review boards. The process on informed consent regarding incidental findings returns is a separate ethical debate that will require appropriate consideration by various stakeholders through, for example, an African and international experts panel meeting with the aim to address (a) the definition of actionability in the context of Africa, (b) the priority list of conditions and related gene variants that are actionable in Africa, (c) the criteria for molecular validation of the variants found in genomic research for clinical use, (d) the clinical environment necessary for returning such results and by which category of health professionals, as most African settings do not have medical genetic services, and (e) the process of wording and integrating informed consent for incidental findings in genomic research in Africa. In the United States, the ACMG has provided a guideline and recommendations to evaluate the cost-effectiveness of returning pathogenic variants for 56 specific genes considered medically actionable (ACMG, 2013, 2015). In Europe, the EuroGenTest and the European Society of Human Genetics recently presented guidelines for diagnostic NGS, including a rating system for diagnostic tests (Matthijs et al., 2016). In the United Kingdom, the Association for Clinical Genetic Science (ACGS) has also released a guideline for the evaluation of pathogenicity and reporting of sequence variants in clinical molecular genetics (Wallis et al., 2013). To the best of our knowledge, there are no evidence-based recommendations for African researchers and clinicians on how to report IFs (de Vries and Pepper, 2012; Sookrajh et al., 2015). This is not a surprise due to the fact that African populations and the diaspora are underrepresented in most of the genetics studies, which questions the universal applicability of the genetic findings in large genome studies, disease association and evolutionary genetic studies (Need and Goldstein, 2009; Rosenberg et al., 2010; Dorschner et al., 2013; Tiffin, 2014; Manrai et al., 2016).

Prior to proposing guidelines on the return of IFs for the African populations, researchers and clinicians should first conduct multiple genetics studies to characterize the nature of

genes for both monogenic and complex diseases on multiple African populations. The results of such studies should first identify the frequency of pathogenic variants in actionable gene lists as defined, e.g., by the ACMG, annotate, and filter genes. An expert panel should validate the list of pathogenic and actionable variants, then conduct a comparative analysis with results from non-African populations (ACMG, 2013; Green et al., 2013; Kalia et al., 2017). The next step could be to define novel actionable genes and variants that are relevant to Africa, e.g., sickle cell disease or *APOL1* variants. Only after completing the aforementioned steps, African researchers and clinicians will be able to provide a comprehensive and clear guideline on which putative pathogenic genes may be returned. It should be noted that the framework on the return of IFs should cover different aspects such as ethical guidelines and genetic counseling. Due to the high diversity in the African population, the classification of pathogenic and actionable variants for the return of secondary findings is more challenging due to the following additional factors: (i) contextualizing the African definition of pathogenicity and actionable genes, (ii) the choice of control cohort for the validation among African populations (iii) the power of the sample size for the case and control cohort, and (iv) a list of actionable genes of the most prominent diseases in the African populations. These questions need to be considered and addressed prior to the development of African actionable gene standards and guideline for IFs. The guidelines and the list of African populations' actionable genes to be returned as IFs is a major milestone toward personalized medicine.

CONCLUSION AND PERSPECTIVES

The power of high-throughput genomic technologies, particularly DNA sequencing, has potential to bridge the gap between genomic research and clinical care. However, this blurry line has opened several technical and ethical questions and concerns, especially in the context of African genomic research. With the highest genetic diversity found in individuals and communities across the African continent, the use of personalized medicine will be beneficial both to the continent and worldwide. The state of WES and WGS on the continent is in the early stages in terms of available genetic data, publications on genetic conditions, appropriately designed pipelines and bioinformatics tools. The process of handling IFs should be clearly discussed and defined by the African research community, clinicians, specifically on the categorization of the pathogenicity, and actionability of genes and variants in order to take advantage of the genomic technology.

We have provided a list of available WES and WGS data that can help in initiating, the development of bioinformatics pipelines suitable for African population genomic data, quantify the frequency of pathogenic and so-called actionable genes, and to develop appropriate policies for their investigation in genomic research. This requires African researchers and experts to be encouraged to share and make data available in public databases. This once again is an urgent call to set an African expert panel to categorize and refine criteria for pathogenicity

and African actionability in human genetic research in Africa. We recommend that experts should prioritize the following steps: (1) define better criteria for classification of pathogenicity, and actionability, including relevant genes lists, that can be explored and return as IFs to research participant in Africa; (2) benchmark existing variant calling and *in silico* prediction pipelines for African genomic data or develop new pipelines using African data; (3) use hypothesis and non-hypothesis approaches *in silico* mutation prediction to avoid false positive mutation; (4) develop an African reference panel; and (5) Sanger sequencing to be done on the new variants for validation.

DATA AVAILABILITY

All datasets analyzed for this study are cited in the manuscript and the Supplementary Files.

REFERENCES

- ACMG (2013). ACMG practice guidelines: incidental findings in clinical genomics: a clarification. *Genet. Med.* 15, 664–666.
- ACMG (2015). ACMG policy statement: updated recommendations regarding analysis and reporting of secondary findings in clinical genome-scale sequencing. *Genet. Med.* 17, 68–69. doi: 10.1038/gim.2013.82
- Adzhubei, I. A., Schmidt, S., Peshkin, L., Ramensky, V. E., Gerasimova, A., Bork, P., et al. (2010). A method and server for predicting damaging missense mutations. *Nat. Methods* 7, 248–249. doi: 10.1038/nmeth0410-248
- Ahmed, A. E., Mpangase, P. T., Panji, S., Baichoo, S., Botha, G., Fadlilmola, F. M., et al. (2018). Organizing and running bioinformatics hackathons within Africa: the H3ABioNet cloud computing experience. *AAS Open Res.* 1, 1–14. doi: 10.12688/aasopenres.12847.1
- 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, 305–315. doi: 10.1101/gr.183483.114
- Auer, P. L., Johnsen, J. M., Johnson, A. D., Logsdon, B. A., Lange, L. A., Nalls, M. A., et al. (2012). Imputation of exome sequence variants into population-based samples and blood-cell-trait-associated loci in African Americans: NHLBI GO exome sequencing project. *Am. J. Hum. Genet.* 91, 794–808. doi: 10.1016/j.ajhg.2012.08.031
- 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.RECEIVED
- Ben Rekaya, M., Naouali, C., Messaoud, O., Jones, M., Bouyacob, Y., Nagara, M., et al. (2018). Whole Exome Sequencing allows the identification of two novel groups of Xeroderma pigmentosum in Tunisia. XP-D and XP-E: impact on molecular diagnosis. *J. Dermatol. Sci.* 89, 172–180. doi: 10.1016/j.jdermsci.2017.10.015
- Bertier, G., Hetu, M., and Joly, Y. (2016). Unsolved challenges of clinical whole-exome sequencing: a systematic literature review of end-users' views. *BMC Med. Genomics.* 9:52. doi: 10.1186/s12920-016-0213-6
- Bindea, G., Mlecnik, B., Hackl, H., Charoentong, P., Tosolini, M., Kirilovsky, A., et al. (2009). ClueGO: a cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics* 25, 1091–1093. doi: 10.1093/bioinformatics/btp101
- Bousfha, A., Bakhchane, A., Charoute, H., Detsouli, M., Rouba, H., Charif, M., et al. (2017). Novel compound heterozygous mutations in the GPR98 (USH2C) gene identified by whole exome sequencing in a Moroccan deaf family. *Mol. Biol. Rep.* 44, 429–434. doi: 10.1007/s11033-017-4129-9
- Busby, G. B., Band, G., Si Le, Q., Jallow, M., Bougama, E., Mangano, V. D., et al. (2016). Admixture into and within sub-Saharan Africa. *Elife* 5:e15266. doi: 10.7554/eLife.15266

AUTHOR CONTRIBUTIONS

All authors contributed in conceiving, preparing, and revising the manuscript, approved the manuscript, and agreed to be accountable for all aspects of the presented work.

FUNDING

This work was supported by NIH Common Fund H3Africa Initiative, award number U54HG009790 to AW and JdV, Wellcome Trust/AAS Ref: H3A/18/001 to AW, and the NIH, United States, Grant Numbers U01HG009716, U01HG007459, and NIH/NHLBI U24HL135600 to AW. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

- 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. doi: 10.1038/s41467-017-00663-9
- Chun, S., and Fay, J. C. (2009). Identification of deleterious mutations within three human genomes. *Genome Res.* 19, 1553–1561. doi: 10.1101/gr.092619.109
- Church, D. M., Schneider, V. A., Steinberg, K. M., Schatz, M. C., Quinlan, A. R., Chin, C. S., et al. (2015). Extending reference assembly models. *Genome Biol.* 16:13. doi: 10.1186/s13059-015-0587-3
- Cooper, G. M., Goode, D. L., Ng, S. B., Sidow, A., Bamshad, M. J., Shendure, J., et al. (2010). Single-nucleotide evolutionary constraint scores highlight disease-causing mutations. *Nat. Methods* 7, 250–251. doi: 10.1038/nmeth0410-250
- Dandara, C., Huzair, F., Borda-Rodriguez, A., Chirikure, S., Okpechi, I., Warnich, L., et al. (2014). H3Africa and the African life sciences ecosystem: building sustainable innovation. *OMICS* 18, 733–739. doi: 10.1089/omi.2014.0145
- Davydov, E. V., Goode, D. L., Sirota, M., Cooper, G. M., Sidow, A., and Batzoglou, S. (2010). Identifying a high fraction of the human genome to be under selective constraint using GERP++. *PLoS Comput. Biol.* 6:e1001025. doi: 10.1371/journal.pcbi.1001025
- de Vries, J., and Pepper, M. (2012). Genomic sovereignty and the African promise: mining the African genome for the benefit of Africa. *J. Med. Ethics* 38, 474–478. doi: 10.1136/medethics-2011-100448
- DePristo, M. A., Banks, E., Poplin, R., Garimella, K. V., Maguire, J. R., Hartl, C., et al. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* 43, 491–498. doi: 10.1038/ng.806492VOLUME
- Dilthey, A., Cox, C. J., Iqbal, Z., Cox, C., Nelson, M. R., and McVean, G. (2014). Improved genome inference in the MHC using a population reference graph. *BioRxiv*
- Doerks, T., Copley, R. R., Schultz, J., Ponting, C. P., and Bork, P. (2002). Systematic identification of novel protein domain families associated with nuclear functions. *Genome Res.* 12, 47–56. doi: 10.1101/gr.203201
- Dong, C., Wei, P., Jian, X., Gibbs, R., Boerwinkle, E., Wang, K., et al. (2015). Comparison and integration of deleteriousness prediction methods for nonsynonymous SNVs in whole exome sequencing studies. *Hum. Mol. Genet.* 24, 2125–2137. doi: 10.1093/hmg/ddu733
- Dorschner, M. O., Amendola, L. M., Turner, E. H., Robertson, P. D., Shirts, B. H., Gallego, C. J., et al. (2013). Actionable, pathogenic incidental findings in 1,000 participants' exomes. *Am. J. Hum. Genet.* 93, 631–640. doi: 10.1016/j.ajhg.2013.08.006
- Fadhlaoui-Zid, K., Haber, M., Martinez-Cruz, B., Zalloua, P., Benammar Elgaaid, A., and Comas, D. (2013). Genome-wide and paternal diversity reveal a recent origin of human populations in North Africa. *PLoS One* 8:e80293. doi: 10.1371/journal.pone.0080293

- Fujita, P. A., Rhead, B., Zweig, A. S., Hinrichs, A. S., Karolchik, D., Cline, M. S., et al. (2011). The UCSC Genome Browser database: update 2011. *Nucleic Acids Res.* 39, D876–D882. doi: 10.1093/nar/gkq963
- Garber, M., Guttman, M., Clamp, M., Zody, M. C., Friedman, N., and Xie, X. (2009). Identifying novel constrained elements by exploiting biased substitution patterns. *Bioinformatics* 25, i54–i62. doi: 10.1093/bioinformatics/btp190
- Garrison, F., and Marth, G. (2012). Haplotype-based variant detection from short-read sequencing. *arXiv*
- Genomics England (2017). *Genomics England K1000 Project Protocol*. Charterhouse Square: Genomics England.
- Green, R. C., Berg, J. S., Grody, W. W., Kalia, S. S., Korf, B. R., Martin, C. L., et al. (2013). ACMG recommendations for reporting of incidental findings in clinical exome and genome sequencing. *Genet. Med.* 15, 565–574. doi: 10.1038/gim.2013.73
- 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 (2017). *Ethics and Governance Framework for Best Practice in Genomic Research and Biobanking in Africa*. Available at: https://www.sun.ac.za/english/faculty/healthsciences/rds/Documents/Final%20Framework%20for%20Africa%20genomics%20and%20biobanking_SC_February%202017%20II%20.pdf
- Hamdi, Y., Boujemaa, M., Ben Rekaya, M., Ben Hamda, C., Mighri, N., El Benna, H., et al. (2018). Family specific genetic predisposition to breast cancer: results from Tunisian whole exome sequenced breast cancer cases. *J. Transl. Med.* 16:158. doi: 10.1186/s12967-018-1504-9
- Hentsche, J. D., Robinson, W. A., and Tan, A. C. (2016). A survey of computational tools to analyze and interpret whole exome sequencing data. *Int. J. Genomics* 2016:7983236. doi: 10.1155/2016/7983236
- Hunter, J. E., Irving, S. A., Biesecker, L. G., Buchanan, A., Jensen, B., Lee, K., et al. (2016). A standardized, evidence-based protocol to assess clinical actionability of genetic disorders associated with genomic variation. *Genet. Med.* 18, 1258–1268. doi: 10.1038/gim.2016.40
- Ichikawa, H., Nagahashi, M., Shimada, Y., Hanyu, T., Ishikawa, T., Kameyama, H., et al. (2017). Actionable gene-based classification toward precision medicine in gastric cancer. *Genet. Med.* 9:93. doi: 10.1186/s13073-017-0484-3
- Jallow, M., Teo, Y. Y., Small, K. S., Rockett, K. A., Deloukas, P., Clark, T. G., et al. (2009). Genome-wide and fine-resolution association analysis of malaria in West Africa. *Nat. Genet.* 41, 657–665. doi: 10.1038/ng.388
- Jongeneel, C. V., Achinike-Oduaran, O., Adebisi, E., Adebisi, M., Adeyemi, S., Akanle, B., et al. (2017). Assessing computational genomics skills: our experience in the H3ABioNet African bioinformatics network. *PLoS Comput. Biol.* 13:e1005419. doi: 10.1371/journal.pcbi.1005419
- Kalia, S. S., Adelman, K., Bale, S. J., Chung, W. K., Eng, C., Evans, J. P., et al. (2017). Recommendations for reporting of secondary findings in clinical exome and genome sequencing, 2016 update (ACMG SF v2.0): a policy statement of the American college of medical genetics and genomics. *Genet. Med.* 19, 249–255. doi: 10.1038/gim.2016.190
- Kiezun, A., Garimella, K., Do, R., Stitzel, N. O., Neale, M. B., McLaren, P. J., et al. (2012). Exome sequencing and the genetic basis of complex traits. *Nat. Genet.* 44, 623–630. doi: 10.1038/ng.2303
- Kobayashi, Y., Yang, S., Nykamp, K., Garcia, J., Lincoln, S. E., and Topper, S. E. (2017). Pathogenic variant burden in the ExAC database: an empirical approach to evaluating population data for clinical variant interpretation. *Genome Med.* 9:13. doi: 10.1186/s13073-017-0403-7
- Koboldt, D. C., Zhang, Q., Larson, D. E., Shen, D., McLellan, M. D., Lin, L., et al. (2012). VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res.* 22, 568–576. doi: 10.1101/gr.129684.111
- Kodaman, N., Sobota, R. S., Asselbergs, F. W., Oetjens, M. T., Moore, J. H., Brown, N. J., et al. (2017). Genetic effects on the correlation structure of CVD risk factors: exome-wide data from a Ghanaian population. *Glob. Heart* 12, 133–140. doi: 10.1016/j.gheart.2017.01.013
- Kramvis, A., Restorp, K., Norder, H., Botha, J. F., Magnius, L. O., and Kew, C. M. (2005). Full genome analysis of hepatitis B virus genotype E strains from South-Western Africa and Madagascar reveals low genetic variability. *J. Med. Virol.* 77, 47–52. doi: 10.1002/jmv.20412
- Kramvis, A., Weitzmann, L., Owiredo, K. B. A., and Kew, C. M. (2002). Analysis of the complete genome of subgroup AH hepatitis B virus isolates from South Africa. *Gen. Virol.* 83, 835–839. doi: 10.1099/0022-1317-83-4-835
- Kuhn, R. M., Karolchik, D., Zweig, A. S., Wang, T., Smith, K. E., Rosenbloom, K. R., et al. (2009). The UCSC genome browser database: update 2009. *Nucleic Acids Res.* 37, D755–D761. doi: 10.1093/nar/gkn875
- Kumar, P., Henikoff, S., and Ng, P. C. (2009). Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat. Protoc.* 4, 1073–1081. doi: 10.1038/nprot.2009.86
- Kwak, S. H., Chae, J., Choi, S., Kim, M. J., Choi, M., Chae, J. H., et al. (2017). Findings of a 1303 Korean whole-exome sequencing study. *Exp. Mol. Med.* 49:e356. doi: 10.1038/emmm.2017.142
- Lacaze, P., Ryan, J., Woods, R., Winship, I., and McNeil, J. (2017). Pathogenic variants in the healthy elderly: unique ethical and practical challenges. *J. Med. Ethics* 43, 714–722. doi: 10.1136/medethics-2016-103967
- Lai, Z., Markovets, A., Ahdesmaki, M., Chapman, B., Hofmann, O., McEwen, R., et al. (2016). VarDict: a novel and versatile variant caller for next-generation sequencing in cancer research. *Nucleic Acids Res.* 44:e108. doi: 10.1093/nar/gkw227
- Landrum, M. J., Lee, J. M., Benson, M., Brown, G., Chao, C., Chitipiralla, S., et al. (2016). ClinVar: public archive of interpretations of clinically relevant variants. *Nucleic Acids Res.* 44, D862–D868. doi: 10.1093/nar/gkv1222
- Lebeko, K., Manyisa, N., Chimusa, E. R., Mulder, N., Dandara, C., and Wonkam, A. (2017). A genomic and protein-protein interaction analyses of nonsyndromic hearing impairment in Cameroon using targeted genomic enrichment and massively parallel sequencing. *OMICS* 21, 90–99. doi: 10.1089/omi.2016.0171
- Leipzig, J. (2017). A review of bioinformatic pipeline frameworks. *Brief. Bioinform.* 18, 530–536. doi: 10.1093/bib/bbw020
- Lek, M., Karczewski, K. J., Minikel, E. V., Samocha, K. E., Banks, E., Fennell, T., et al. (2016). Analysis of protein-coding genetic variation in 60,706 humans. *Nature* 536, 285–291. doi: 10.1038/nature19057
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078–2079. doi: 10.1093/bioinformatics/btp352
- Liu, X., Jian, X., and Boerwinkle, E. (2011). dbNSFP: a lightweight database of human nonsynonymous SNPs and their functional predictions. *Hum. Mutat.* 32, 894–899. doi: 10.1002/humu.21517
- MacArthur, D. G., Manolio, T. A., Dimmock, D. P., Rehms, H. L., Shendure, J., Abecasis, G. R., et al. (2014). Guidelines for investigating causality of sequence variants in human disease. *Nature* 508, 469–475. doi: 10.1038/nature13127
- Manrai, K. A., Funke, B. H., Rehms, H. L., Olesen, M. S., Maron, B. A., Szolovits, P., et al. (2016). Genetic misdiagnoses and the potential for health disparities. *N Engl. J. Med.* 375, 655–665. doi: 10.1056/NEJMs1507092
- Marcus, S., Lee, H., and Schatz, M. C. (2014). SplitMEM: a graphical algorithm for pan-genome analysis with suffix skips. *Bioinformatics* 30, 3476–3483. doi: 10.1093/bioinformatics/btu756
- Martin, A. R., Tefferri, S., Moller, M., Hoal, E. G., and Daly, M. J. (2018). The critical needs and challenges for genetic architecture studies in Africa. *Curr. Opin. Genet. Dev.* 53, 113–120. doi: 10.1016/j.gde.2018.08.005
- Masimirembwa, C., Dandara, C., and Hasler, J. (2014). “Population diversity and pharmacogenomics in Africa,” in *Handbook of Pharmacogenomics and Stratified Medicine*, ed. S. Padmanabhan (Amsterdam: Elsevier Science), 971–998. doi: 10.1016/b978-0-12-386882-4.00043-8
- Matthijs, G., Souche, E., Alders, M., Corveleyn, A., Eck, S., Feenstra, I., et al. (2016). Guidelines for diagnostic next-generation sequencing. *Eur. J. Hum. Genet.* 24, 2–5. doi: 10.1038/ejhg.2015.226
- McLaren, W., Gil, L., Hunt, S. E., Riat, H. S., Ritchie, G. R., Thormann, A., et al. (2016). The ensembl variant effect predictor. *Genome Biol.* 17:122. doi: 10.1186/s13059-016-0974-4
- Mulder, N. J., Adebisi, E., Adebisi, M., Adeyemi, S., Ahmed, A., Ahmed, R., et al. (2017). Development of bioinformatics infrastructure for genomics research. *Glob. Heart* 12, 91–98. doi: 10.1016/j.gheart.2017.01.005
- Mulder, N. J., Adebisi, 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
- Ndiaye Diallo, R., Gadji, M., Hennig, B. J., Gueye, M. V., Gaye, A., Diop, J. P. D., et al. (2017). Strengthening human genetics research in Africa: report of the 9th

- meeting of the African society of human genetics in Dakar in May 2016. *Glob. Health Epidemiol. Genom.* 2:e10. doi: 10.1017/gheg.2017.3
- Need, A. C., and Goldstein, D. B. (2009). Next generation disparities in human genomics: concerns and remedies. *Trends Genet.* 25, 489–494. doi: 10.1016/j.tig.2009.09.012
- Ness, B. V. (2008). Genomic research and incidental findings. *J. Law Med. Ethics* 36, 292–312. doi: 10.1111/j.1748-720X.2008.00272.x
- Ng, M. C., Shriner, D., Chen, B. H., Li, J., Chen, W. M., Guo, X., et al. (2014). Meta-analysis of genome-wide association studies in African Americans provides insights into the genetic architecture of type 2 diabetes. *PLoS Genet.* 10:e1004517. doi: 10.1371/journal.pgen.1004517
- Ng, P. C., and Henikoff, S. (2006). Predicting the effects of amino acid substitutions on protein function. *Annu. Rev. Genomics Hum. Genet.* 7, 61–80. doi: 10.1146/annurev.genom.7.080505.115630
- Nowak, K. J., Bauskis, A., Dawkins, H. J., and Baynam, G. (2018). Incidental inequity. *Eur. J. Hum. Genet.* 26, 616–617. doi: 10.1038/s41431-018-0101-y
- Pabinger, S., Dander, A., Fischer, M., Snajder, R., Sperk, M., Efreanova, M., et al. (2012). A survey of tools for variant analysis of next-generation genome sequencing data. *Brief. Bioinform.* 15, 256–278. doi: 10.1093/bib/bbs086
- Parker, M., and Kwiatkowski, D. P. (2016). The ethics of sustainable genomic research in Africa. *Genome Biol.* 17:44. doi: 10.1186/s13059-016-0914-3
- Paten, B., Novak, M. A., Eizenga, M. J., and Garrison, E. (2017). Genome graphs and the evolution of genome inference. *Genome Res* 27, 665–676. doi: 10.1101/gr.214155.116
- Popejoy, A. B., and Fullerton, S. M. (2016). Genomics is failing on diversity. *Nature* 538, 161–164. doi: 10.1038/538161a
- Rabbani, B., Tekin, M., and Mahdieh, N. (2014). The promise of whole-exome sequencing in medical genetics. *J. Hum. Genet.* 59, 5–15. doi: 10.1038/jhg.2013.114
- Retshabile, G., Mlotshwa, B. C., Williams, L., Mwesigwa, S., Mboowa, G., Huang, Z., et al. (2018). Whole-exome sequencing reveals uncaptured variation and distinct ancestry in the Southern African population of Botswana. *Am. J. Hum. Genet.* 102, 731–743. doi: 10.1016/j.ajhg.2018.03.010
- Reva, B., Antipin, Y., and Sander, C. (2011). Predicting the functional impact of protein mutations: application to cancer genomics. *Nucleic Acids Res.* 39:e118. doi: 10.1093/nar/gkr407
- 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
- Rosenberg, N. A., Huang, L., Jewett, E. M., Szpiech, Z. A., Jankovic, I., and Boehnke, M. (2010). Genome-wide association studies in diverse populations. *Nat. Rev. Genet.* 11, 356–366. doi: 10.1038/nrg2760
- Schwarz, J. M., Rodelsperger, C., Schuelke, M., and Seelow, D. (2010). MutationTaster evaluates disease-causing potential of sequence alterations. *Nat. Methods* 7, 575–576. doi: 10.1038/nmeth0810-575
- Sherry, S. T., Ward, M. H., Kholodov, M., Baker, J., Phan, L., Smigielski, E. M., et al. (2001). dbSNP- the NCBI database of genetic variation. *Nucleic Acids Res.* 29, 308–311. doi: 10.1093/nar/29.1.308
- Shihab, H. A., Gough, J., Cooper, D. N., Day, I. N., and Gaunt, T. R. (2013). Predicting the functional consequences of cancer-associated amino acid substitutions. *Bioinformatics* 29, 1504–1510. doi: 10.1093/bioinformatics/btt182
- Sookrajh, Y., Naidoo, S., Ramjee, G., and Team, M. D. P. (2015). Shared responsibility for ensuring appropriate management of incidental findings: a case study from South Africa. *J. Med. Ethics* 41, 281–283. doi: 10.1136/medethics-2013-101561
- Souzeau, E., Burdon, K. P., Mackey, D. A., Hewitt, A. W., Savarirayan, R., Otlowski, M., et al. (2016). Ethical Considerations for the Return of Incidental Findings in Ophthalmic Genomic Research. *Transl. Vis. Sci. Technol.* 5, 1–11. doi: 10.1167/tvst.5.1.3
- Tang, C.-M., Dattani, S., So, M.-T., Cherny, S. S., Tam, P. K. H., Sham, P. C., et al. (2018). Actionable secondary findings from whole-genome sequencing of 954 East Asians. *Hum. Genet.* 137, 31–37. doi: 10.1007/s00439-017-1852-1
- Tekola-Ayele, F., Adeyemo, A., Aseffa, A., Hailu, E., Finan, C., Davey, G., et al. (2015). Clinical and pharmacogenomic implications of genetic variation in a Southern Ethiopian population. *Pharmacogenomics J.* 15, 101–108. doi: 10.1038/tpj.2014.39
- Teng, S., Madej, T., Panchenko, A., and Alexov, E. (2009). Modeling effects of human single nucleotide polymorphisms on protein-protein interactions. *Biophys. J.* 96, 2178–2188. doi: 10.1016/j.bpj.2008.12.3904
- The H3Africa Consortium (2014). Enabling the genomic revolution in Africa. *Science* 344, 1347–1348. doi: 10.1126/science.1251546
- Tiffin, N. (2014). Unique considerations for advancing genomic medicine in African populations. *Per. Med.* 11, 187–196. doi: 10.2217/pme.13.105
- Uthman, O. A., Wiysonge, C. S., Ota, M. O., Nicol, M., Hussey, G. D., Ndumbe, P. M., et al. (2015). Increasing the value of health research in the WHO African Region beyond 2015—reflecting on the past, celebrating the present and building the future: a bibliometric analysis. *BMJ Open* 5:e006340. doi: 10.1136/bmjopen-2014-006340
- Venner, C. M., Nankya, I., Kyeyune, F., Demers, K., Kwok, C., Chen, P. L., et al. (2016). Infecting HIV-1 Subtype Predicts Disease Progression in Women of Sub-Saharan Africa. *EBioMedicine* 13, 305–314. doi: 10.1016/j.ebiom.2016.10.014
- Wallis, Y., Payne, S., McNulty, C., Bodmer, D., Siermans, E., Robertson, K., et al. (2013). *Practice Guidelines for Evaluations of Pathogenicity and the Reporting of Sequencing Variants in Clinical Molecular Genetics*. Birmingham: Association for Clinical Genetic Science.
- Wang, K., Li, M., and Hakonarson, H. (2010). ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* 38:e164. doi: 10.1093/nar/gkq603
- Wang, Z., Liu, X., Yang, B.-Z., and Gelernter, J. (2013). The role and challenges of exome sequencing in studies of human diseases. *Front. Genet.* 4:160. doi: 10.3389/fgene.2013.00160
- Warde-Farley, D., Donaldson, S. L., Comes, O., Zuberi, K., Badrawi, R., Chao, P., et al. (2010). The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. *Nucleic Acids Res.* 38, W214–W220. doi: 10.1093/nar/gkq537
- Wei, Z., Wang, W., Hu, P., Lyon, G. J., and Hakonarson, H. (2011). SNVer: a statistical tool for variant calling in analysis of pooled or individual next-generation sequencing data. *Nucleic Acids Res.* 39:e132. doi: 10.1093/nar/gkr599
- Wilm, A., Aw, P. P., Bertrand, D., Yeo, G. H., Ong, S. H., Wong, C. H., et al. (2012). LoFreq: a sequence-quality aware, ultra-sensitive variant caller for uncovering cell-population heterogeneity from high-throughput sequencing datasets. *Nucleic Acids Res.* 40, 11189–11201. doi: 10.1093/nar/gks918
- 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. Law Med. Ethics* 36, 219–248.
- Xu, C. (2018). A review of somatic single nucleotide variant calling algorithms for next-generation sequencing data. *Comput. Struct. Biotechnol. J.* 16, 15–24. doi: 10.1016/j.csbj.2018.01.003
- Yang, H., and Wang, K. (2015). Genomic variant annotation and prioritization with ANNOVAR and wANNOVAR. *Nat. Protoc.* 10, 1556–1566. doi: 10.1038/nprot.2015.105

Conflict of Interest Statement: 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.

Copyright © 2019 Bope, Chimusa, Nembaware, Mazandu, de Vries and Wonkam. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.