frontiers | Research Topics

The path towards precision health: Prospects and challenges

Edited by
Muntaser Eltayeb Ibrahim, Ayman Hussein,
Maria G. Stathopoulou and Sophie Visvikis-Siest

Published in
Frontiers in Genetics
Frontiers in Medicine

ISSN 1664-8714
ISBN 978-2-83251-003-2
DOI 10.3389/978-2-83251-003-2

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

The path towards precision health: Prospects and challenges

Topic editors

Muntaser Eltayeb Ibrahim — University of Khartoum, Sudan
Ayman Hussein — University of Khartoum, Sudan
Maria G. Stathopoulou — Institut National de la Santé et de la Recherche Médicale (INSERM), France
Sophie Visvikis-Siest — University of Lorraine, France

Citation

Ibrahim, M. E., Hussein, A., Stathopoulou, M. G., Visvikis-Siest, S., eds. (2022). *The path towards precision health: Prospects and challenges*.
Lausanne: Frontiers Media SA. doi: 10.3389/978-2-83251-003-2

Table of contents

frontiers | Frontiers in Medicine

TYPE Editorial
PUBLISHED 11 October 2022
DOI 10.3389/fmed.2022.989746

OPEN ACCESS

EDITED BY
Somchai Chutipongtanate,
University of Cincinnati, United States

REVIEWED BY
Debmalya Barh,
Institute of Integrative Omics and Applied Biotechnology (IIOAB), India
Chanitra Thuwajit,
Mahidol University, Thailand

*CORRESPONDENCE
Muntaser E. Ibrahim
mibrahim@iend.org

SPECIALTY SECTION
This article was submitted to Precision Medicine, a section of the journal Frontiers in Medicine

RECEIVED 08 July 2022
ACCEPTED 06 September 2022
PUBLISHED 11 October 2022

CITATION
Ibrahim ME, Hussein A, Stathopoulou MG and Visvikis-Siest S (2022) Editorial: The path towards precision health: Prospects and challenges. *Front. Med.* 9:989746. doi: 10.3389/fmed.2022.989746

Editorial: The path towards precision health: Prospects and challenges

Muntaser E. Ibrahim[1]*, Ayman Hussein[1], Maria G. Stathopoulou[2] and Sophie Visvikis-Siest[2]

[1]Department of Molecular Biology Institute of Endemic Diseases, University of Khartoum, Khartoum, Sudan, [2]Institut National de la Santé et de la Recherche Médicale (INSERM), Paris, France

KEYWORDS

precision medicine, precision health, phenotype, genomics, individualized, population

Editorial on the Research Topic
The path towards precision health: Prospects and challenges

Emerging scientific concepts set sail carried by the wings and tides of history, causing paradigm shifts, changing cultures, dissociating from old beliefs and promoting appropriate technological advancements that render them operable. The emerging concept of "Precision Medicine/health," which has been hailed as the future of medicine for well over a decade is one such concept.

In this Research Topic, "*The path toward precision health prospects and challenges*," nine articles reflect on key elements of the current practice of precision medicine, outlining triumphs as well as potential hurdles and gaps. Some of the most spectacular examples of these triumphs have been in pharmacogenomics. Visvikis-Siest et al., highlight in their article (Milestones in personalized medicine: from ancient time to nowadays-the provocation of COVID-19), examples of these, while emphasizing the dire want of personalized solutions and treatments tragically and ironically underscored by the SARS-CoV-2 pandemic. In fact, the pandemic blatantly exposed the feebleness of health systems that still rely on outdated concepts of "average" patients.

Although genomics was key in freeing science from the clutches of a "one gene one disease" set of mind, introducing us to realms of genomic interactions in the complex milieu of systems biology, science still largely grapples with these within one-dimensional perspectives of the biological world. Considerations of human-microbial host-parasite relationships thus remain beyond the purview of current genomic and transcriptomic approaches and consequently so too do potent modifiers of phenotypic outcomes resulting from complex scenarios in both infectious and non-infectious disease.

Moreover, although precision medicine owes much of its current utility and relative success to genomic advancements and the explosion of genomic databases, this progress does not come without cost. It is accompanied by profound ethical and societal questions, including those of "ownership" of personal or collective information and the labeling of groups and individuals with measured or reported phenotypes. Furthermore, the

increased dependence on specific patient's genetic material lowers the value of knowledge based on "averaged individuals." This use of patient and population specific genomic data to populate the public genetic databases that inform clinical and public health decisions raises, in addition to the above-mentioned questions of "ownership" of genetic data, concerns about the affordability of the technology for certain countries and individuals and thus the issue of equity in benefiting from these approaches.

The highlighted differences between groups and populations underscore the place of inclusiveness in research, publications and databases. Bibliometric and scientometric studies reveal the current scope, focus and state of precision medicine, its disparities and its trends and changes over time. Zhu and Zhang 's work on Emerging Trends and Research Foci in Cataract Genes gives us an example of such work.

Any platform for precision medicine must also include tools for translating relevant research into appropriate, applicable and evidence-based standards for clinical practice.

In "*Legal challenges in precision medicine: What duties arising from genetic and genomic testing does a physician owe to patients?*" McGrath et al., highlight the difficulties that specialist and non-specialist physicians face in understanding, communicating and utilizing genomic data, the challenges in establishing standards of practice in this rapidly changing domain and the ethical and, particularly, legal risks arising when utilizing (or failing to appropriately utilize) genomic technologies.

The spectrum of individualized phenotypes extends from the risk of exposures in healthy individuals through the vast spectrum of clinical phenotypes and up to death. Preventive precision medicine, with its applications and caveats, ideally aims to stratify individuals into finite and well-defined risk groups, incorporating biomarkers predictive of risk to guide pre-emptive interventions. In this collection, the death phenotype, as the ultimate phenotype to prevent, is the subject of two articles published by Chen et al., "*development of a simple risk model to predict mortality in patients with osteosarcoma of the extremity*" and by Primorac et al., "*sudden cardiac death-a new insight into potentially fatal genetic markers.*" Both, by employing genomic or clinical data, attempts with varying degrees of success to predict the risk of death. They reveal the challenges of identifying the underpinnings of extreme phenotypes, even when they are believed to be explainable by a single or few genomic changes. Multi-omic approaches, including epigenomics, transcriptomics, proteomics, metabolomics, and microbiomics, could come handy in refining such disease phenotypes, including death. Such a complex undertaking would require sophisticated tools such as artificial intelligence.

In conditions, classically, attributed to a single input, such as Mendelian diseases, heterogeneity in the causative input, together with genetic and environmental modifiers, can create complex and varied outcomes. It is therefore imperative to revisit paradigms and dictums, such as "one gene-one phenotype."

Even for conditions where a single gene, or even a single mutation, is primarily responsible for disease initiation, significant and impactful variation in clinical phenotype, disease progression and response to therapy can arise from genetic, environmental or lifestyle modifiers.

In this collection, Yahia and Stevanin explore the "*History of gene hunting in hereditary spinocerebellar degeneration*," highlighting the inherent genetic heterogeneity and reviewing the tools used in its elucidation. Ngo-Bitoungui et al., investigate the contribution of previously identified kidney Dysfunction-Related Gene Variants to kidney dysfunction in a cohort of Sickle Cell Disease Patients from Cameroon.

As for the common polygenic noninfectious diseases which occupy the widest space in the disease susceptibility spectrum, such as the phenotypically and genetically interrelated metabolic and cardiovascular diseases, genomics can become instrumental in redefining both their phenotypes and underlying molecular basis. Diabetes is one example that can benefit from such reclassification. The article by Fedotkina et al., (*Novel reclassification of adult diabetes is useful to distinguish stages of B-cell function linked to the risk of vascular complications: The DOLCE study from Northern Ukraine*) gives a glimpse of the utility that such approaches, which tend to greater precision in classification, can achieve.

There is no other place, where the complex interaction between genotype and phenotype as applied to precision medicine could be extensively elucidated and understood than the African continent. Africa is mankind's ancestral home and the store of its greatest wealth of genetic diversity. It has equally rich social and cultural legacies relevant to the understanding and practice of precision medicine. Hussein et al., (*individualized medicine in Africa: Bringing the practice into the realm of population heterogeneity*), present their views on the strengths and deficiencies of Precision Medicine in the African context, emphasizing challenges and opportunities of adopting such approaches in populations of high effective size and suggesting how population genomics could aid the transition between generic and personalized approaches.

Although precision medicine falls under the wider umbrella of precision health (which includes in addition to clinical medicine aspects of public health, community and preventive medicine, and health promotion,) the emphasis in much of the published literature (this topic included), is on clinical precision medicine. This reflects a rather global bias in understanding and managing issues and challenges of health and disease which we hope will become less pervasive as the precision health community grows and expands.

Author contributions

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

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

frontiers in Genetics

REVIEW
published: 30 November 2020
doi: 10.3389/fgene.2020.569175

Milestones in Personalized Medicine: From the Ancient Time to Nowadays—the Provocation of COVID-19

Sophie Visvikis-Siest[1,2]*, *Danai Theodoridou*[1,2], *Maria-Spyridoula Kontoe*[1,2], *Satish Kumar*[1,2] *and Michael Marschler*[1,2]

[1] Université de Lorraine, IGE-PCV, Nancy, France, [2] The Santorini Conferences (SCs) Association, Nancy, France

The first evidence of individual targeting medicine appeared in ancient times thousands of years ago. Various therapeutic approaches have been established since then. However, even nowadays, conventional therapies do not take into consideration individuals' idiosyncrasy and genetic make-up, failing thus to be effective in some cases. Over time, the necessity of a more precise and effective treatment resulted in the development of a scientific field currently known as "personalized medicine." The numerous technological breakthroughs in this field have acknowledged personalized medicine as the next generation of diagnosis and treatment. Although personalized medicine has attracted a lot of attention the last years, there are still several obstacles hindering its application in clinical practice. These limitations have come to light recently, due to the COVID-19 pandemic. This review describes the "journey" of personalized medicine over time, emphasizing on important milestones achieved through time. Starting from the treatment of malaria, as a first more personalized therapeutic approach, it highlights the need of new diagnostic tools and therapeutic regimens based on individuals' genetic background. Furthermore, it aims at raising global awareness regarding the current limitations and the necessity of a personalized strategy to overpass healthcare problems and hence, the current crisis.

Keywords: public health, personalized medicine, pharmacogenomics, inflammation, COVID−19

OPEN ACCESS

Edited by:
Julio Vera González,
University Hospital Erlangen, Germany

Reviewed by:
Emiliano Giardina,
University of Rome Tor Vergata, Italy
Guido Santos-Rosales,
University Hospital Erlangen, Germany

***Correspondence:**
Sophie Visvikis-Siest
sophie.visvikis-siest@inserm.fr

Specialty section:
This article was submitted to Genomic Medicine, a section of the journal Frontiers in Genetics

Received: *03 June 2020*
Accepted: *26 October 2020*
Published: *30 November 2020*

Citation:
Visvikis-Siest S, Theodoridou D, Kontoe M-S, Kumar S and Marschler M (2020) Milestones in Personalized Medicine: From the Ancient Time to Nowadays—the Provocation of COVID-19. Front. Genet. 11:569175. doi: 10.3389/fgene.2020.569175

In ancient times, ∼1550 BC, the first evidence about medicine adapted to individual's health appeared in the *Odyssey* written by Homer (Jouanna, 2012): "*Telemachus, the son of Odysseus, visits Menelaus and Helen in search of news about his father, who has still not returned home after the Trojan War. Reminiscence of the absent Odysseus leads to tears and at this moment Helen puts a drug (φάρμακον) into the crater of wine, which eases grief or anger and makes one forget one's woes. This drug came from Egypt: Homer says*" (Jouanna, 2012). In fact, according to Homer, Egyptians were considered as "the wisest of men" even from Greek physicians because the Egyptians were descendants of Paeon, the doctor of the gods (Jouanna, 2012).

The adaptation of that ancient ***"Egyptian medicine"*** to an individual's health status was further elucidated in the Classical period from Herodotus when the practice of medicine was divided into categories and every doctor was a specialist for one disease, one body part (Jouanna, 2012). This is the first evidence of personalized medicine, as doctors realized that separating the diseases according to the human body parts can help them to achieve a deeper understanding of illness

and consequently attain a better therapeutic outcome. Greeks were fascinated by this approach of medicine, and that is why they constantly mention in their treatises the Egyptian medicine with admiration (Jouanna, 2012). Only after the appearance of Hippocratic medicine in the fifth century, the Egyptian medicine started to be omitted from the Greek books (Jouanna, 2012).

Although **"*Hippocratic medicine*"** shares similarities with the Egyptian, the former does not undermine the latter. In contrast, Hippocrates used the knowledge of Egyptian medicine and advanced it by removing the magico-religious part and making it more rational (Jouanna, 2012). In fact, at that ancient time, physicians had to think about the patient's needs and beliefs in order to have a successful treatment (Fierz, 2004). Hippocrates believed that "diseases might be treated from their origin" and "the treatment carried out should be opposed to the cause of the disease." Therefore, they focused more on the personalized approach of the disease and eliminated all the superstition that was surrounding this time of history (Sykiotis et al., 2005). In light of that breakthrough, Hippocrates was ahead of his time, as he managed to give a direction in the understanding of the genomic medicine by suggesting that every human is distinct, and this affects both the disease prediction and the treatment (Sykiotis et al., 2005).

During the twenty-five centuries, from the so-called Father of Western Medicine Hippocrates to the modern physician, personalized medicine has evolved attracting a lot of attention (Sykiotis et al., 2005).

However, despite the ancient vision and recommendations, medical therapy employed a very broad approach that was based on clinical and genetic/genomic data from heterogeneous populations instead of focusing on each patient, even in recent decades (National Institutes of Health, 2007). Physicians used standardized approaches based on data and knowledge of the earlier patients/diseases to decide on a therapeutic regimen. Clinical trials were only aiming for standardized treatment rooting out 20% of the population that was not going to respond to a treatment or even worse-experience adverse effects possibly due to differences in their genetic make-up (Fierz, 2004). This meant that medicine had no room for idiosyncrasy (from ancient Greek *ἰδιοσυγκρασία*/idiosynkrasia, "a peculiar temperament, habit of the body, e.g., blend of humors." Idiosyncrasy defined the way that physicians conceived diseases in the nineteen century)[1]. This approach began to change in the 1870s, when discoveries made by researchers in Europe allowed the advent of a **"*scientific medicine*,"** a precursor to the evidence-based medicine.

In the early 1950s, scientists started to realize progressively the need for "***evidence-based medicine***." The prediction of drug response to ensure the safety of the patient as well as a better outcome gave birth to the field of today's **"*personalized medicine*."** Discoveries in the field of molecular biology contributed to a better understanding of drug response (Vogenberg et al., 2010). In this regard, human genome mapping was a breakthrough providing a better understanding of people's genetic make-up. Although individuals are 99.1% identical, the remaining 0.9% of interindividual genetic variability is responsible for the observed variability within the humans (National Institutes of Health, 2007).

Overall, the challenge that researchers and physicians face is still immense nowadays. The purpose of personalized medicine is to combine modern medicine with molecular advances in order to target patients separately and improve the efficacy and effectiveness of the therapeutic approach (Mini and Nobili, 2009). The realization that the conventional approach of using candidate genes alone is not sufficient to explain the differences in disease risks between different ethnic groups and also within individuals led to the whole genome approaches. The evolution of different genotyping technologies over the years has allowed focusing on specific regions of the genome enabling deeper coverage and understanding of the variants. Therefore, it enabled medical practitioners to identify and treat patients based on their unique characteristics.

Today, the four humors of Hippocrates, blood, phlegm, yellow bile, and black bile, which determined the treatment of each individual (Hippocrates, 1543), have been replaced with the four building blocks (A, T, G, C) enabling improved medical predictions.

Cutting-edge biochemical advances including single-nucleotide polymorphisms (SNPs), genotyping, and biochips have made personalized medicine a reality justifying the use of the terminology in the last few decades.

Indeed, the unique identity of every person's genome provides valuable information regarding disease onset and progression along with the response to different therapeutic regimens (Agyeman and Ofori-Asenso, 2015). Variations such as SNPs, insertions and deletions, structural variants, and copy number variations in the human genome play a distinctive role in the manifestation and progression of diseases such as cancer, diabetes, and neurodegenerative and cardiovascular diseases (Agyeman and Ofori-Asenso, 2015). Hence, biomarkers are being investigated as a way of predicting certain diseases and also identifying patient subgroups that respond only to specific drugs. However, environmental factors can also act as triggers and/or cofactors. Therefore, predicting response to drugs as well as treatment based only on genetic information without taking into account the environmental determinants can lead to poor or false results (Agyeman and Ofori-Asenso, 2015).

Combining the human genome, environmental factors, disease assessments, and medication in order to achieve a better therapeutic outcome is the exact vision that personalized medicine is aiming to achieve. For the aforementioned, it is obvious that the journey of personalized medicine, as described in our previous article (Visvikis-Siest et al., 2018), has not reached its final destination. A mob of problems is still around the patient's needs, which is a challenge for personalized medicine nowadays.

The present review focuses on the major discoveries from the past to the present, points out milestones that helped on personalized medicine's journey, underlines the current problems that physicians still face, and gives insights into the future, assisting thus healthcare systems globally. More importantly, it gives direction on how to handle epidemic crises such us

[1]Available online at: https://dictionary.cambridge.org/dictionary/english/idiosyncrasy.

coronavirus disease 2019 (COVID-19) that the world is currently facing, regarding diagnostic tools and therapeutic strategies.

UNDERPINNINGS OF TARGETED THERAPY

This chapter points out some important milestones through the history of pharmacotherapy that we must keep in mind and use as a beacon for achieving targeted medicine for individuals' needs, which, unfortunately, are not yet applied.

The Key Treatment of Malaria as the Beginning—Building Knowledge on Pandemics

The first evidence of malaria is found in 2700 BC into ancient Chinese medical records. Even today, malaria is an extremely serious and fatal disease (Talapko et al., 2019). It is estimated that malaria affected 228 million people resulting in 405,000 deaths globally in 2018[2]. There is still a lot of research regarding diagnosis, prevention, and treatment of this disease all over the world including countries where malaria is even more common like Africa and some Asian countries. In these countries, the prevalence of malaria is higher probably due to the tropical climate, which increases the mortality rate from 0.3 to 2.2% globally to 11–30% in tropical environments (Talapko et al., 2019).

Several herbs have been used to treat malaria such as Qinghai in the second century BC in China and the Cinchona tree in the Sixteen century in Peru (Talapko et al., 2019). In 1926, one of the most effective drugs distributed to treat acute malaria was pamaquine an 8-aminoquinoline. At that time, pamaquine was a groundbreaking discovery because of its effectiveness as an antimalarial. However, adverse effects observed after its administration raised concerns about its safety (Howes et al., 2013). More specifically, between 1930 and 1940 at least 250 cases of acute hemolytic anemia were reported after providing the drug to patients (Beutler, 1959). As a result, in 1943, scientists started investigating alternative therapeutic regimens to encounter the adverse effects of pamaquine. Various compounds were tested, among them primaquine, which also belongs to 8-aminoquinolines (Howes et al., 2013). Primaquine, an 8-(4-amino-1-methyl-buty1amino)-6-methoxyquinoline, first appeared in the Korean War as an antimalarial, where soldiers were administered the drug to eliminate the long latency of P. vivax infection (Ashley et al., 2014). Although primaquine was considered the most appropriate candidate, hemolytic anemia was still observed (Howes et al., 2013).

The answer to this problem came later, in 1956 when Carson et al. discovered that the side effects of hemolytic anemia were caused by a deficiency in the G6PD enzyme (Howes et al., 2013). The G6PD deficiency was well-established from 510 BC (Relling and Evans, 2015) when Pythagoras, even though he was not a physician (Luzzatto and Arese, 2018), had observed this side effect after a number of his students consumed fava (Relling and Evans, 2015). The advances in molecular diagnostics revealed that there are a lot of mutations in the gene but people remained asymptomatic and only in a few cases, such as after the administration of primaquine trigger severe side effects (Luzzatto and Arese, 2018).

According to WHO, primaquine is now used to cure the liver infection caused by malaria (P. vivax and P. ovale) and prevent relapse. To eliminate the hemolytic anemia and achieve a better therapeutic outcome, WHO has published guidelines for primaquine administration to reduce the risk of the adverse effect in people with G6PD deficiency. For example, for the prevention of malaria in normal adults the therapeutic approach is to administrate 0.25–0.50 mg/kg body weight daily for a duration of 14 days. However, for people with G6PD deficiency, the dosage is differentiated to 0.75 mg/kg body weight once a week for 8 weeks with close monitoring of the patient's therapy (Policy brief, 2016).

The discovery of the association between antimalarial drugs and G6PD deficiency opened a new perspective regarding the adverse effects of these drugs as well as a more personalized approach to the disease. This was one of the first examples that led to a big step toward the application of a more personalized therapy that was established as a term many years later in 1991 and is currently still quite limited.

But why are the reported cases of malaria still large (228 million globally in 2018) while diseases such as Ebola and Cholera, which were also of a similar magnitude, are being managed properly resulting in a decrease of the infected population[3,4]? One answer to that question could be that 90% of the population infected by malaria is originated in Africa[2]. For example, Ebola appeared in 1976 in central Africa but the outbreak in 2014–2016 was the one that alerted the scientific community as the virus managed to spread quickly from West Africa to urban areas and across borders transforming it into an epidemic[1]. Another aspect could be the inadequacy of resources in developing countries. Lack of food and medical supplies hinders the treatment, proving thus that the environment is also a co-factor in disease progression and cure. So physicians should consider every continent and every patient individually according to his origin and taking into account their environment to achieve improved therapy.

The Pharmacogenetic Evolution—an Important Milestone

Biochemical health sciences started to evolve around the 1940s and 1950s at the same time as the development of instrumentation and new research methods. The scientific field responsible for (1) the research of different patients' responses to the drugs and (2) the minimization of the adverse effects caused also by variability on the metabolizing enzymes is known as

[2]Malaria. Available online at : https://www.who.int/news-room/fact-sheets/detail/malaria (accessed January 14, 2020).

[3]*Years of Ebola Virus Disease Outbreaks*. Available online at: https://www.cdc.gov/vhf/ebola/history/chronology.html?CDC_AA_refVal=https%3A%2F%2Fwww.cdc.gov%2Fvhf%2Febola%2Foutbreaks%2Fhistory%2Fchronology.html (accessed October 25, 2019).

[4]Cholera. Available online at: https://www.who.int/news-room/fact-sheets/detail/cholera (accessed January 17, 2019).

pharmacogenetics (Rogers et al., 2002). After the first appearance of mass spectrometers, science evolved quickly leading to the first observation of cytochrome P450 (CYP450) around 1958. Later, in 1965, the novel drug metabolizing enzyme CYP450 was introduced (Estabrook, 2003). The constant speculation is the gap between physicians and the knowledge regarding pharmacogenetics in order to achieve personalized medicine (Rogers et al., 2002).

One of the first examples that shifted treatment is CYP2D6, an enzyme that belongs to the CYP450 family and is responsible for the metabolism of 20% of the drugs involving anti-arrhythmic, antidepressants, antipsychotics, b-blockers, and analgesics (Maréchal et al., 2008). CYP2D6 was discovered in 1969 due to different plasma concentrations of nortriptyline observed in patients, indicating differences in its metabolism (Yang et al., 2017). Some years later, in 1977 it was observed that debrisoquine, an adrenergic-blocking drug, which was used to treat hypertension (Mahgoub et al., 1977) had also variations in response to treatment (Silas et al., 1977). Today, debrisoquine is mostly used as a marker to determine the activity of the CYP2D6 enzyme in patients. More specifically, debrisoquine and its 4-OH-metabolite are measured in urine with gas chromatography and high-performance liquid chromatography (HPLC) methods (Llerena et al., 2009) in order to find the individual's CYP2D6 genotype. Therefore, by predicting the phenotype of each patient it can give insight into their response to specific drugs metabolized by the CYP2D6 enzyme (Siest et al., 2007).

Although nowadays important evidence about the multiple variants in CYP2D6 exists, there are still not a lot of applications in medicine where the administration of several drugs in patients is depending on the occurrence of specific SNPs on their DNA. This is a future challenge in this field that will bring us a step forward to personalized therapy.

Another important advancement regarding the application of personalized medicine in cancer therapy is the discovery of the HER-2 gene. Breast cancer is a disease that can be divided into different subtypes depending on the tumor as well as on the patients' genetic background predisposition (Burstein, 2005). More specifically, HER-2 gene (Human Epidermal Growth Factor Receptor 2) encodes a tyrosine kinase receptor that takes part in signaling pathways both in normal and in malignant breast cells and is strongly associated with a lower response to cancer treatment and survival rate (Slamon et al., 2011).

The *HER-2* gene became clinically relevant in 1987 when Salmon et al. reported a lower survival rate in women with breast cancer carrying the mutated gene (Slamon et al., 1987). It was discovered that in the majority of patients carrying the mutation, overexpression of this gene was associated with chemotherapy resistance, poor patient prognosis (Harari and Yarden, 2000), high risk of cancer progression, and a low survival rate (Gajria and Chandarlapaty, 2011). Patients that are carrying the *HER-2* gene amplified have a distinctive molecular signature that can distinguish these type of cancers from other breast cancers (Burstein, 2005). A lot of studies were conducted in order to reverse these adverse effects of the *HER-2* gene overexpression and eventually achieved with the use of monoclonal antibodies targeting the tyrosine kinase receptor (Slamon et al., 2011). A great example is the antibody trastuzumab that inhibits tumor growth when used as monotherapy, also when used with cisplatin, carboplatin, docetaxel, and ionizing radiation which have synergistic effects and when used with doxorubicin, cyclophosphamide, methotrexate, and paclitaxel which has additive effects (Slamon et al., 2001).

Since that time, several clinical trials have proved the efficacy of trastuzumab resulting also in establishing routine HER-2 testing in breast cancer patients and changing dramatically the therapeutic approach to those carrying the mutation (Burstein, 2005). This gene is a great milestone of applied personalized medicine, clearly showing that the right choice of a drug, based on the genetic background of a patient, can have positive effects on their life.

Ultimately, an essential public health necessity is to identify, through innovative diagnostic methods, individuals at high risk for a given disease, enabling thus early prognosis and the application of preventive therapies. After the Genome Wide Association Studies (GWAS), thousands of genetic variants were linked to several complex diseases overcoming the statistical limitations (Visvikis-Siest et al., 2018). GWAS studies allowed disease prediction combining multiple genetic factors with the use of polygenic risk scores. Although polygenic risk score is a promising new tool, its application has not reached clinical accuracy yet. However, several discoveries have indicated its potential utility in disease prediction, such as cancer, Alzheimer's disease, Parkinson disease (PD), and cardiovascular disease (CVD) (Chasioti et al., 2019).

From the aforementioned, the impact of personalized medicine on the healthcare community is obvious. There is an important value in understanding the cause of the problem and making health better by solving it.

LIMITATIONS OF TODAY'S MEDICINE—THE CASE OF COVID-19

The limitations of personalized medicine have come to the foreground nowadays due to the pandemic of coronavirus disease 2019 (COVID-19) that emerged in December 2019 in China and managed to spread rapidly in multiple countries at the beginning of February 2020 (Gao et al., 2020). Despite all the worldwide-recognized advances and discoveries that have been achieved, modern medicine still cannot provide a treatment with current therapeutic approaches. It is widely recognized that the genetic background of each patient in the case of COVID-19 pandemic is one of the major contributors of drug effectiveness and toxicity (Cascella et al., 2020). Thereby, the challenge of COVID-19 virus made physicians and healthcare staff to realize the problems that the global healthcare system faces and to acknowledge the crucial role of applied personalized medicine.

Treatment Difficulties

Scientists believe that this virus causes pneumonia by interacting with the ACE2 receptors. But are we sure that the COVID-19 virus attacks the respiratory system and not the circulation of oxygen?

Due to the limitations of existing experimental methods, the pathogenesis of the virus is not clear yet. According to literature, decreased levels of hemoglobin and neutrophil and increased levels of serum ferritin, erythrocyte sedimentation rate, C-reactive protein, albumin, and lactate dehydrogenase were observed in patients with COVID-19 pneumonia. Hemoglobin is a protein contained in red blood cells, and it is responsible for the transportation of oxygen to the tissues. It consists of four units of haem. Therefore, the aforementioned observations indicate that haem also increases (Liu and Li, 2020) as a result of hemoglobin oxidation (Wu et al., 2019) and causes, in turn, the accumulation of many detrimental iron ions (Liu and Li, 2020). The haem release can cause inflammation in two ways: (1) by "intercalating in the membrane and altering cellular structures" and (2) by "activating immune responses and inflammatory reactions which act as the pro-oxidant in endothelial cells, neutrophils, and macrophage" (Wu et al., 2019). Consequently, haem accumulation in the respiratory system results in increased permeability in the membranes of endothelial cells, facilitating, thus, the COVID-19 virus to enter into the endothelial cells of the lung and cause secondary inflammation resulting in pneumonia. The hypoxia that low levels of hemoglobin causes in the lungs can be a co-factor that ultimately leads also to pneumonia. The maintenance of satisfactory levels of hemoglobin is hereby essential for the oxygenation of tissue, and decreased levels of the latter in infected patients result in a limited capacity of red blood cells to transfer oxygen to the tissues.

So far, there are no specific drugs that can be effective in controlling the disease. Due to the global spread of the virus and the non-existent vaccine the global health community has been focused on finding the best antiviral agent to control the disease. Many clinical trials are ongoing in order to establish a course of treatment and prevent the numerous deaths happening daily. Several drugs are being tested for their activity against the COVID-19 virus, and almost 30 agents have been revealed (Dong et al., 2020).

According to the National Health Commission (NHC) of the People's Republic of China 7th edition interferon α (IFN-α), lopinavir/ritonavir, chloroquine phosphate, ribavirin, and arbidol are recommended for empirical therapy against COVID-19 virus (latest edition March 4, 2020) [5](Luo et al., 2020; Zhu et al., 2020). On the 7th of March, the NHC reported that tocilizumab (TCZ) also has been included in the treatment guidelines as an antiviral agent[6]. Other drugs such as azithromycin, an antibiotic, or corticosteroids act in support of the above antiviral agents helping in the overall treatment as concomitant agents (Gautret et al., 2020; Luo et al., 2020). Details about these drugs are shown in **Figure 1**.

The aforementioned guidelines include 6 drugs as principal antiviral agents for the COVID-19 therapy. IFN-a is a broad-spectrum antiviral agent that is described as an inhibitor of the *in vitro* reproduction of the COVID-19 virus (Dong et al., 2020). Lopinavir/ritonavir is an aspartate protease inhibitor used as medication in the human immunodeficiency virus (HIV). Due to its *in vitro* antiviral activity (Dong et al., 2020), lopinavir/ritonavir has been suggested as second-line treatment for COVID-19 according to the only one existing today therapeutic algorithm from Hellenic government[7]. However, its effect on eliminating COVID-19 virus is still controversial (Jienchi and Kome, 2020; Stower, 2020). Chloroquine phosphate (or hydroxychloroquine), a widely used antimalaria drug, might also have positive effects on treating the COVID-19 virus (Dong et al., 2020). Studies suggest that its potential antiviral activity can be attributed to an increase in the endosomal pH required for virus/cell fusion and the disruption of the glycosylation of cellular receptors of the COVID-19 virus (Gao et al., 2020). Due to its antiviral and anti-inflammatory effects, chloroquine phosphate is utilized in the first line of COVID-19 treatment along with azithromycin according to the therapeutic algorithm of the Hellenic government[4] (Gao et al., 2020; Touret and de Lamballerie, 2020) (**Figure 2**). However, its use should be considered with caution due to its potential cardiotoxicity (e.g., QT prolongation and drug-drug interactions)[4,8] as well as its adverse effects on people with G6PD deficiency. The administration of chloroquine phosphate in patients with G6PD deficiency can cause hemolytic anemia proving once again that genes influence reponses to drugs and highlighting the vital importance of personalized medicine. Chloroquine phosphate has also been suggested for prophylaxis in areas with high COVID-19 incidences, but results are inconclusive and further investigation is needed (Principi and Esposito, 2020). Although it is proposed as a treatment against the COVID-19 virus, its benefits have not been proven yet (Cortegiani et al., 2020).

Another antiviral agent suggested on the guidelines is ribavirin. Ribavirin inhibits the replication of multiple viruses, and because of its use in emergency clinical management, it has also been suggested for the management of COVID-19 virus. However, its clinical effectiveness has not yet been established (Khalili et al., 2020). Ribavirin is used in combination with IFN-a or lopinavir/ritonavir for the treatment of COVID-19 virus (Dong et al., 2020). Arbidol is another antiviral drug that is approved for influenza treatment. Arbidol has promising antiviral activity regarding the COVID-19 pandemic (Xu et al., 2020), and it is potentially more effective as a monotherapy compared to lopinavir/ritonavir compound (Zhu et al., 2020). In any case, the drugs described above are not suggested to be administered for more than 10 days (Dong et al., 2020).

A new addition to the guidelines was TCZ as studies mention a positive effect on controlling the COVID-19 virus. So far, findings support that TCZ is effective in preventing or treating

[5]*Chinese Clinical Guidance for COVID-19 Pneumonia Diagnosis and Treatment, 7th Edn 2020.* Available online at: http://kjfy.meetingchina.org/msite/news/show/cn/3337.html(Luo et al., 2020; Zhu et al., 2020).

[6]*Latest developments in epidemic control.* Available online at: http://en.nhc.gov.cn/2020-03/07/c_77436.htm (accessed March 7, 2020).

[7]Coronavirus disease (COVID-19). Available online at: https://eody.gov.gr/neos-koronaios-covid-19/ (accessed April 5, 2020).

[8]*COVID-19 Drug Interactions*. Available online at: https://www.covid19-druginteractions.org/.

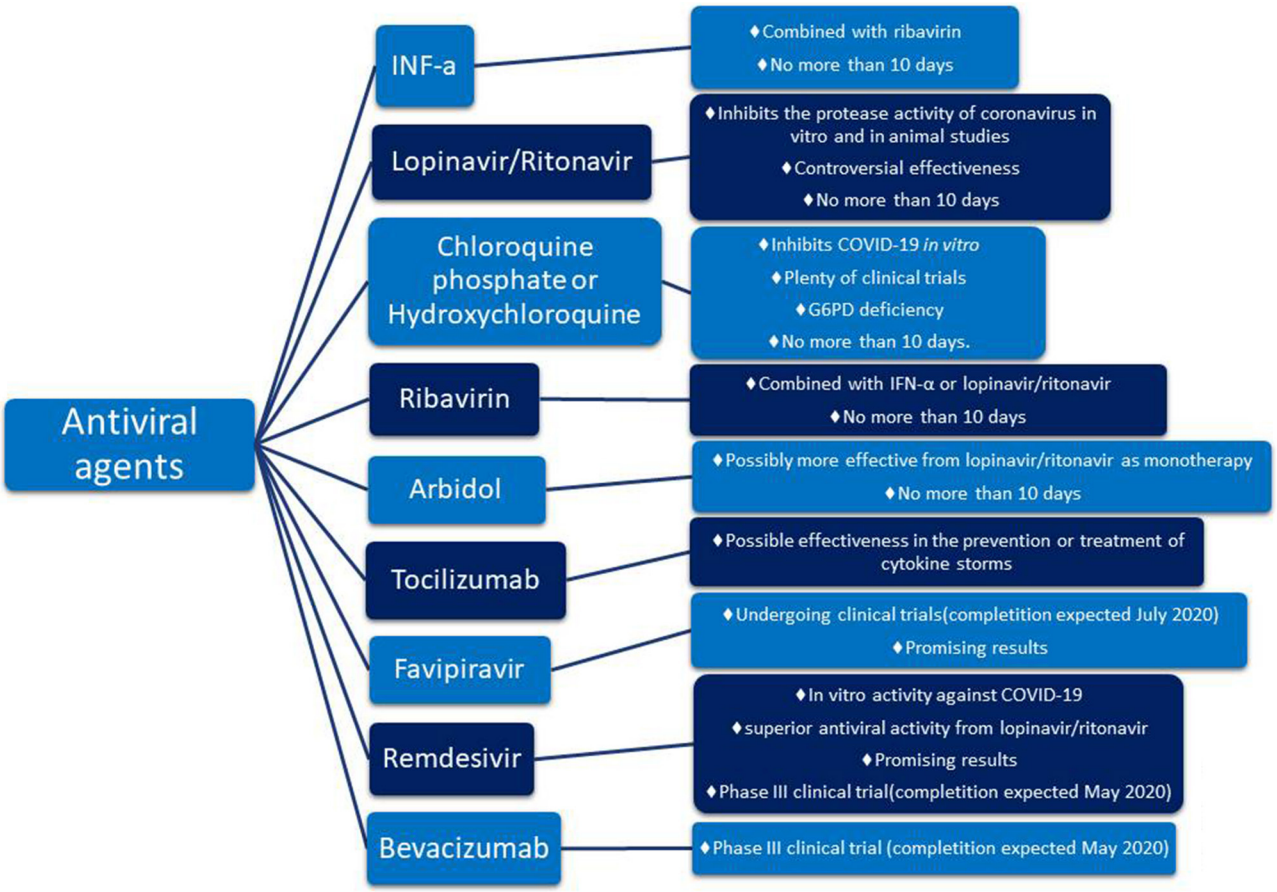


FIGURE 1 | Promising pharmacotherapy for encountering COVID-19 virus.

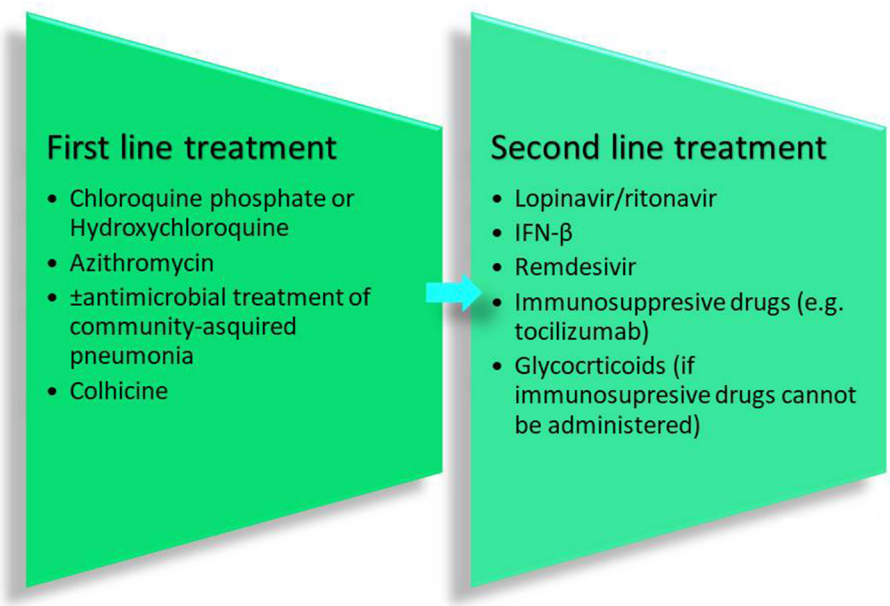


FIGURE 2 | Therapeutic recommendations from the Hellenic government.

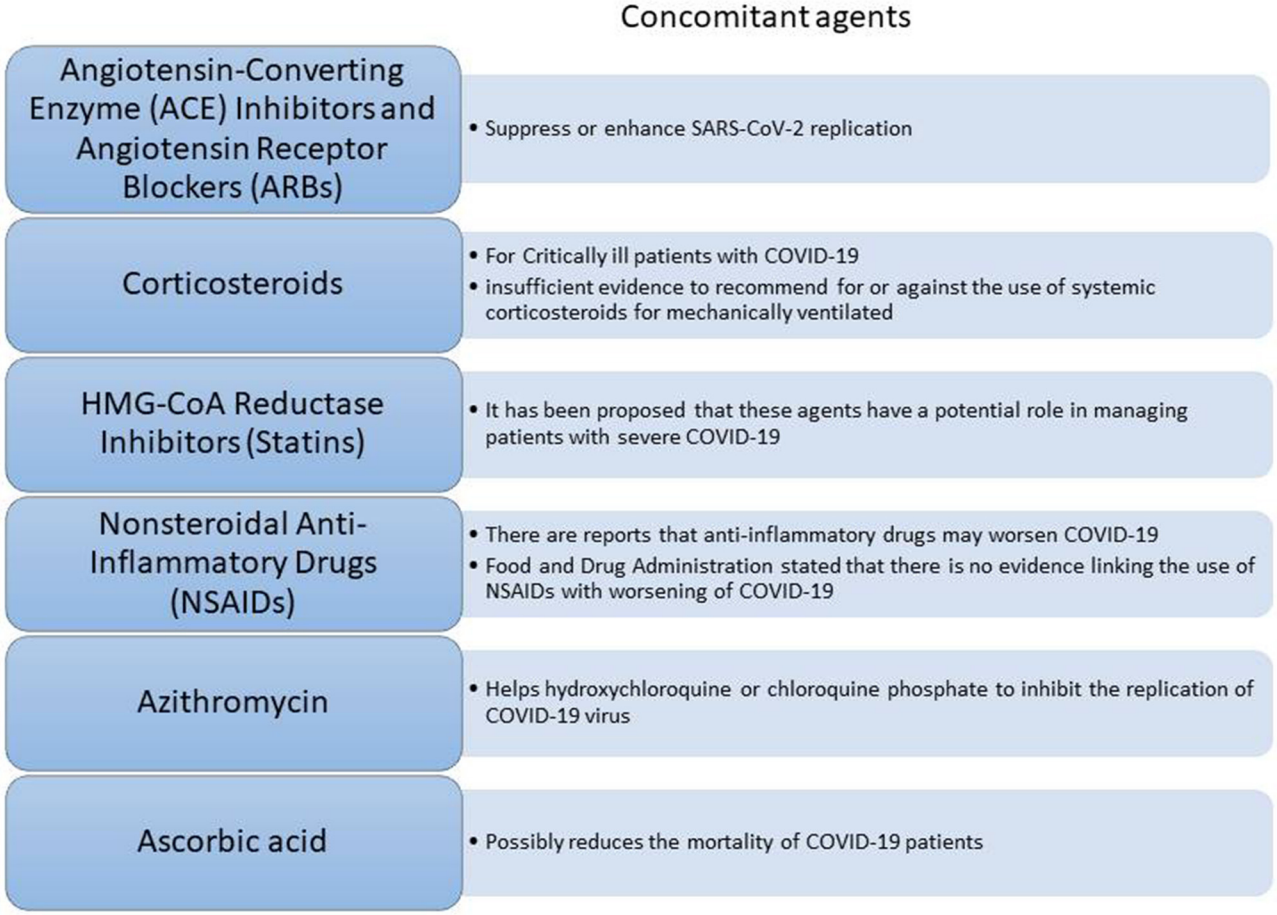


FIGURE 3 | Concomitant agents that can assist on the COVID-19 therapy.

the cytokine storm, observed in patients affected by the COVID-19 virus (Luo et al., 2020). TCZ has been chosen instead of corticosteroids due to its fewer side effects on patients. It is an anti-IL-6 receptor (interleukin-6 receptor) antibody that in combination with glucocorticoid could potentially improve the condition of critically ill patients (Luo et al., 2020). TCZ alone or in combination with anakinra, siltuximab alone or in combination with anakinra, and anakinra alone are currently being tested for their ability to improve the lung function by inhibiting the cytokine storm (expected completion December 2020)[9].

Besides the drugs already included in the guidelines, there are several other drugs worth mentioning. Favipiravir and remdesivir are two promising agents that are currently being tested for their antiviral effect against COVID-19 (Dong et al., 2020). Favipiravir, an RNA-polymerase inhibitor, is in an ongoing clinical trial that is expected to end in July 2020[10] and remdesivir, a nucleoside analog, in an ongoing clinical trial expected to finish in May 2020[11] Another promising drug is bevacizumab that will be further discussed in the next chapter.

Given that a lot of COVID-19 patients are already on treatment for other chronic diseases, concerns have been raised about the potential synergistic effect of commonly used therapeutic agents (**Figure 3**) along with the COVID-19 therapy. Such agents known as concomitant agents are the angiotensin-converting enzyme (ACE) inhibitors and angiotensin receptor blockers (ARBs), HMG-CoA reductase inhibitors, non-steroidal anti-inflammatory drugs (NSAIDs), and corticosteroids. ACE inhibitors and ARBs are used globally from numerous people for various diseases such as hypertension, heart failure, coronary artery disease, or kidney disease[12] Due to considerations of biological plausibility and a large percentage of COVID-19

[9] *Treatment of COVID-19 Patients With Anti-interleukin Drugs (COV-AID 2020.* Available online at: https://clinicaltrials.gov/ct2/show/NCT04330638.

[10] *Clinical Study To Evaluate The Performance And Safety Of Favipiravir in COVID-19.* Available online at: https://clinicaltrials.gov/ct2/show/NCT04336904 (accessed April 7, 2020).

[11] Study to Evaluate the Safety and Antiviral Activity of Remdesivir (GS-5734^{TM}) in Participants With Severe Coronavirus Disease (COVID-19). Available online at: https://clinicaltrials.gov/ct2/show/NCT04292899.

[12] *COVID-19 and the use of angiotensin-converting enzyme inhibitors and receptor blockers.* Available online at: https://www.who.int/news-room/commentaries/detail/covid-19-and-the-use-of-angiotensin-converting-enzyme-inhibitors-and-receptor-blockers.

patients with cardiovascular disease having a poor disease progression, speculation exists about the worse outcome of patients on long-term therapy with these agents. However, there is still a lack of clinical evidence regarding their effects on the COVID-19 infection and further investigation is being conducted[13] HMG-CoA reductase inhibitors (or statins) might lower cardiovascular morbidity related to COVID-19 reducing thus the progression of the disease[10,14] Furthermore, the use of non-steroidal anti-inflammatory drugs (NSAIDs) seems to be controversial. Several reports suggested that NSAIDs might worsen the outcome of the virus by inhibiting antibody production. However, FDA counteracted this belief highlighting that there is no adequate evidence[10] (**Figure 3**). Another concomitant agent is ascorbic acid. Given its involvement in the immune response to viral agents, it has been suggested that ascorbic acid might have additional benefits to COVID-19 patients. Thus, intravenous ascorbic acid administration is being tested in an ongoing clinical trial investigating its potential anti-inflammatory and antioxidant activity[15].

As the novel COVID-19 pandemic continues, the research community works extensively to suggest the best treatment approach. In this regard, WHO launched a promising international clinical trial called "Solidarity" that aims at finding drug combinations for the COVID-19 management[16]. Other clinical trials are also being conducted in order to finally determine the right treatment for COVID-19 patients (Available online). Although a lot of effort has been done from the healthcare community, treatment guidelines are still changing every day worsening thus the circumstances and conditions that physicians have to face. Until now, there is not an effective treatment approach suitable for everyone. Furthermore, disease severity and progression seem to vary even among people with similar phenotypes. Several studies have been conducted aiming to explain different clinical complications and outcomes in patients suffering from COVID-19. Their focus is to identify ACE2 variants that are associated with different disease response. Two ACE2 alleles (i.e., rs73635825 and rs143936283) showed a reduced probability of viral attachment and a possibility of resistance to infection due to the low binding affinity of the virus protein. However, published evidence of such findings is very limited, as it is still difficult to distinguish via testing who have been infected and who have not (Lippi et al., 2020; Teichmann and Regev, 2020). Moreover, recent studies demonstrate the involvement of chromosome 3p21.31 (rs11385942) and chromosome 9q34.2 (rs657152) implying the ABO blood group in disease's behavior in patients with respiratory failure infected with COVID-19 (Ellinghaus et al., 2020). This global pandemic proves once again the necessity for personalized medicine and genetic sequencing for detecting different disease behaviors and potential adverse effects attributed to drugs used for virus treatment. In this regard, the case of chloroquine for malaria in patients with G6PD deficiency, previously described, should be taken as an example.

Obstacles in Diagnostic Test Development

Although new technological advances have helped scientists to move a step forward in the development of novel diagnostic tools, there are still major problems regarding the reliability, sensitivity, and specificity of diagnostic testing. As it is above mentioned, the world faces a pandemic disease, the coronavirus disease 2019 (COVID-19) (Gao et al., 2020). Scientists struggle to balance between urgency and the sensitivity of diagnostic testing to achieve applied effectiveness in any medical diagnostic tool.

So far, several regulatory authorized diagnostic tests have been used to detect the existence of the COVID-19 virus. One of the most utilized is the RT-PCR test (real-time reverse transcription-polymerase chain reaction)[17,18] (Ai et al., 2020). Designed for the qualitative detection of nucleic acids from the COVID-19 virus[14], this test uses samples taken from nasopharyngeal and oropharyngeal swab or sputum of patients that their symptoms suggest the existence of the COVID-19 virus (e.g., fever, tiredness and dry cough and/or symptoms of acute respiratory illness)[14,15].

However, despite the supply challenges due to the increasing demand currently, there are also concerns regarding the performance of the different technologies used (Petherick, 2020). The RT-PCR test lacks the necessary accuracy and sensitivity due to the substantial percentage of "false-negative" results[14,15] (Ai et al., 2020). Hence, depending on the kit's label, the number of samples used and the RT-PCR machine there is a great percentage of error when using the RT-PCR test[14,15]. According to WHO, negative results do not necessarily exclude the possibility of COVID-19 infection and RT-PCR assay should not be the only criterion for COVID-19 diagnosis[14] Literature suggests adopting chest CT (computed tomography) as an additional diagnostic tool in parallel with the RT-PCR test. Chest CT is a non-invasive diagnostic test with great efficiency that can minimize the false-negative cases from RT-PCR assay (Ai et al., 2020).

Nevertheless, the RT-PCR test is suggested to be used in severe cases to determine the need for hospitalization. Due to lack of resources, this test cannot be applied widely and people with less severe symptoms are recommended to stay at home without testing.

Rapid antibody testing has also been authorized by WHO according to the Emergency Use Listing (EUL) for the identification of IgM/IgG antibodies in patients. It is well-known that the presence of IgM antibodies signals the first line of defense

[13] *Considerations for Certain Concomitant Medications in Patients with COVID-19 2020.* Available online at: https://www.covid19treatmentguidelines.nih.gov/concomitant-medications/.

[14] Preventing Cardiac Complication of COVID-19 Disease With Early Acute Coronary Syndrome Therapy: A Randomized Controlled Trial. (C-19-ACS) 2020. Available online at: https://clinicaltrials.gov/ct2/show/NCT04333407.

[15] *Use of Ascorbic Acid in Patients With COVID 19.* Available online at: https://clinicaltrials.gov/ct2/show/NCT04323514.

[16] "*Solidarity*" *Clinical trial for COVID-19 Treatments.* Available online at: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/global-research-on-novel-coronavirus-2019-ncov/solidarity-clinical-trial-for-covid-19-treatments (accessed March 21, 2020).

[17] Coronavirus Disease (COVID-19) Technical Guidance: Laboratory Testing for 2019-nCoV in Humans. Available online at: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/technical-guidance/laboratory-guidance (accessed April 8, 2020).

[18] Emergency Use Authorizations. Available online at: https://www.fda.gov/medical-devices/emergency-situations-medical-devices/emergency-use-authorizations (accessed April 24, 2020).

when a patient is infected and the presence of IgG antibodies signals immunity for an individual (Li et al., 2020). This type of assay is different from the RT-PCR test, as it focuses on the proteins (antibodies) produced by the immune system as a response to a viral infection. Despite that it is a simple, rapid, and highly sensitive test, it lacks specificity (Li et al., 2020), as IgM and IgG antibodies can be detected in various infections and not only in the case of COVID-19 virus. Even though antibody testing is probably the test that can provide the most accurate results, there are not such tests provided at the moment.

Consequently, with the current diagnostic tools several people could be misdiagnosed or mistreated and hence, it is evident that there is a need for more specificity in diagnosis resulting in a better therapeutic outcome.

THE "NEXT GENERATION OF TREATMENT"

Nowadays, research community and healthcare providers try to improve not only the treatment strategies but also the *in vitro* diagnostics including the identification of novel biomarkers or the study of clinical phenotypes for a better disease prediction, response to drugs, etc. (Wurcel et al., 2016). In this regard, significant progress has been accomplished in the field of biomarkers. The term biomarker exists since the 1950s and has been widely used during the 1980s (Albert, 2011). It is defined as a measurable characteristic that can indicate physiological and/or pathophysiological processes or pharmacologic responses to treatment (Landeck et al., 2016). In the past decade, the field of biomarker research and especially in cardiovascular diseases and cancer has been developed rapidly (Albert, 2011). A great example for the revolutionary application of biomarkers in personalized medicine is the vascular endothelial growth factor A (VEGF-A), which is described as an endothelial cell-specific mitogen. Produced by many cell types including tumor cells, macrophages, platelets, keratinocytes, and renal mesangial cells (Khan et al., 2002), the VEGF family and their receptors have a key role in angiogenesis, in tumor growth (Feliz, 2013), and in physiological functions such as hematopoiesis, wound healing, etc. (Khan et al., 2002).

In 1993, Kim et al. (1993) were the first who identified monoclonal antibodies that can target and neutralize VEGF-A, inhibiting thus tumor growth in preclinical studies. This triggered the production of numerous anti-VEGF drugs such as the recombinant humanized VEGF-A-specific monoclonal antibody bevacizumab, which was approved in 2004 by the US Food and Drug Administration (FDA) as the first-line treatment for metastatic colorectal cancer (Ferrara and Adamis, 2016). A new era of treatment, the VEGF-A anti-therapy, has come to light, resulting in better therapeutic outcomes in patients and approaching the ultimate goal, personalized medicine.

As described above, hypoxia and inflammation are two interrelated conditions. Hypoxia causes inflammation in several lung diseases such as acute lung injury (ALI) and infection and vice versa (Ramakrishnan et al., 2014). Hypoxic stress can induce VEGF-A activity in the lungs (Ramakrishnan et al., 2014). Nowadays, regarding the COVID-19 pandemic, it has been observed that COVID-19 patients have higher VEGF levels compared to healthy population and hypothesized that VEGF-A anti-therapy might be applicable in this case as well. Thereby, a clinical trial is being conducted in order to investigate the effect of bevacizumab on disease control and treatment, especially on the acute lung injury (ALI) and acute respiratory distress syndrome (ARDS) associated with COVID-19 virus (estimated study completion: May 2020)[19]. The results showed that patients receiving bevacizumab improved their oxygen support by 92% without leading to any deaths whereas patients receiving standard care the improvement rate of oxygen was only 62% with deterioration rate 19% and 3 deaths. Also, rapid improvement of the PaO2/FiO2 ratio in patients with severe COVID-19, fever reduction, anti-inflammatory action, and increase in peripheral blood lymphocytes were observed (Pang et al., 2020).

Biomarkers and especially VEGF-A family are important tools for personalized medicine that can assist the diagnosis as well as the selection of patients for specific treatments. Unfortunately, the majority of them are not yet clinically used. Given the disputable efficacy of various chronic disease prevention strategies, new, stringent biomarkers, including the genetic predisposition ones, should be urgently identified and established in the clinical practice.

PERSONALIZED MEDICINE STARTS WITH THE PATIENT

Personalized medicine is an ancient vision of rising challenges for healthcare systems and the research community throughout the years. It aims at confronting every obstacle on the prevention, diagnosis, and treatment of diseases by targeting each patient individually. Advances in both science and technology have already contributed to significant improvements regarding disease management and clinical outcomes prediction (Conti et al., 2009; Nimmesgern et al., 2017).

In **Figure 4**, the journey of personalized medicine through time is depicted. Forty-seven centuries passed between two very serious pandemics and the reasonable query is what it really happened over these years. From 2700 BC until the Hippocratic period, medicine developed rapidly to reach a level that even today we face difficulties to attain (Jouanna, 2012). After the Hippocratic period, a significant gap appeared. Eighteen centuries passed without the patient being considered as an individual and with the "one size fits all" approach being the center of attention. This gap, which inevitably slowed down the evolution of personalized medicine, resulted in patients' exposure to a healthcare system that did not consider them as different entities (Fierz, 2004). After the 1950s, personalized medicine gained traction again with impactful discoveries starting with shaping the future of medicine. Since 2005, although a lot of

[19]Bevacizumab in Severe or Critical Patients With COVID-19 Pneumonia (BEST-CP). Available online at: https://clinicaltrials.gov/ct2/show/NCT04275414 (accessed February 19, 2020).

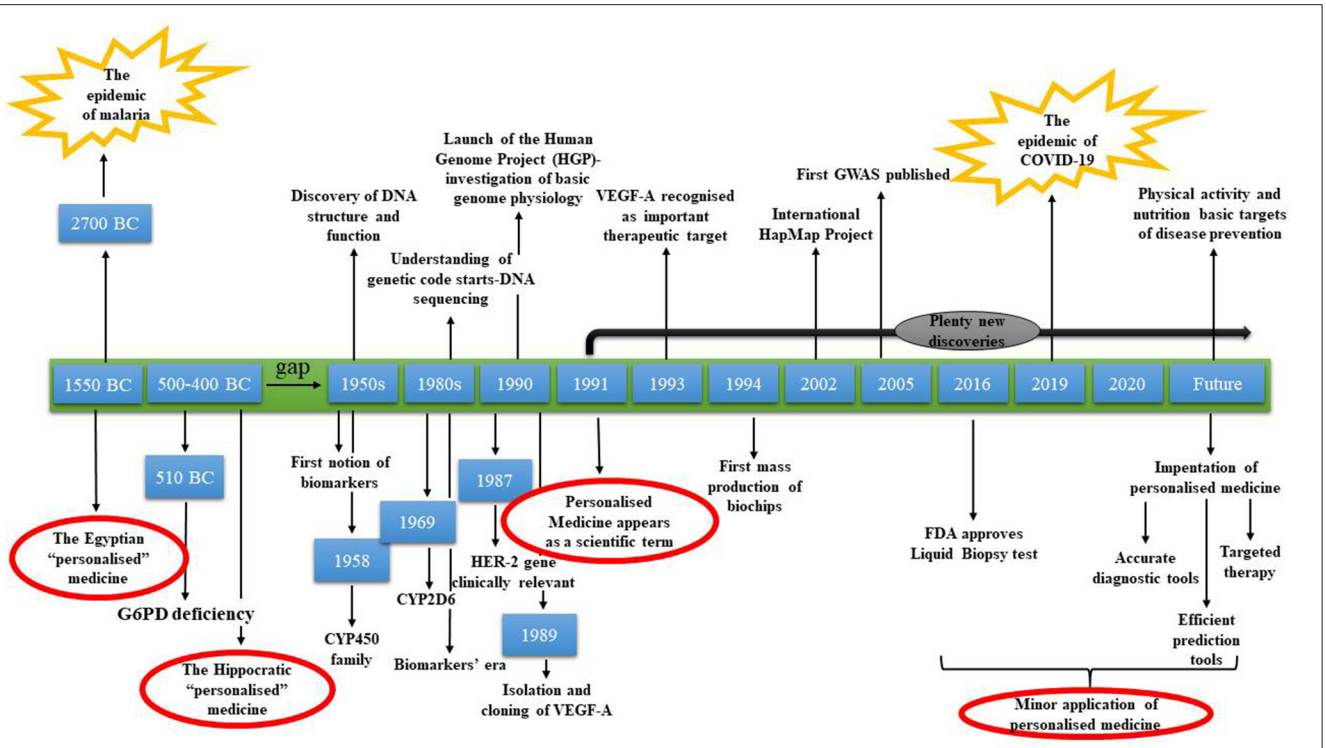


FIGURE 4 | Personalized medicine: from the ancient times to a promising future.

discoveries and technological advances have been accomplished, minor applications have been observed.

Nowadays, 15 years later, personalized medicine still struggles to be applicable. There is no need to leave another 15 years to pass until we understand the necessity of this field in medicine. Overcoming the application obstacles immediately will be an important step toward a new more personalized beginning on medicine. However, this cannot be achieved without providing both to the healthcare community and to the public the necessary information and awareness regarding this field.

At the beginning of the Twenty first century, the "-omics" era arose making possible the detailed examination of organisms and their molecular phenotypes. The "-omics" research together with the advanced sequencing technologies increased the possibilities of early prognosis and targeted therapy for several medical conditions. Although sequencing technologies have identified numerous diseases, millions of variants have not yet been interpreted (Visvikis-Siest et al., 2018).

Furthermore, the human genome is basically the foundation of personalized medicine. It is thought that personalized medicine will improve the practice of medicine based on the better understanding of the characteristics of each individual. Thus, personalized medicine is expected to enhance adherence and minimize harm associated with adverse events (Sorich and McKinnon, 2012). Despite its benefits and potential value, there are still obstacles making the implication of personalized medicine in clinical routine challenging. The number of genetic markers being discovered is increasing; however, their clinical validation is processing slowly. Further analyses for clinical validation may require computational methods in systems biology that involves the use of software tools and human resources expertised in this field. Unfortunately, there is still insufficient support and education for the clinical care professionals. Education on the use and limitations of personalized medicine is also necessary for patients. Moreover, the collection and analysis of bio-specimen which is essential for personalized medicine raise ethical, legal, and social issues that should be addressed prior to its employment in clinical routine (Overby and Tarczy-Hornoch, 2013). In addition, increased emphasis is given by the scientific community nowadays on the cost-effectiveness of some applications of personalized medicine (Sorich and McKinnon, 2012).

The outbreak of the COVID-19 infection is the right occasion and challenge for both research and healthcare professionals to change toward a more personalized approach taking into account individuals' needs. Efficient therapeutic regimens should be discovered and tested in a very short period to minimize the consequences of this infection. New diagnostic tools with increased sensitivity and specificity should be applied and might be the next generation of diagnosis. The combination of new and already established biomarkers could be the key to improving diagnostics tools and treatments, as it happened in the case of VEGF and bevacizumab.

Consequently, the COVID-19 pandemic acknowledged the limitations that health care is facing, underlining the important

issues that medicine is forced to encounter. Thus, the COVID-19 urged the healthcare system to change adapting a new reality of tailored therapy. A lot of advances are expected to be accomplished and a lot of current limitations to be overpassed in the next years. Personalized medicine, having the patient at the center of attention, is going to shape the future in medicine.

AUTHOR CONTRIBUTIONS

SV-S: bibliography search, conception, and writing. DT: bibliography search, writing, and preparation of figures. M-SK: bibliography search and writing. SK: bibliography search and writing. MM: bibliography search and writing. All authors: contributed to the article and approved the submitted version.

REFERENCES

Agyeman, A. A., and Ofori-Asenso, R. (2015). Perspective: does personalized medicine hold the future for medicine? *J. Pharm. Bioallied. Sci.* 7, 239–244. doi: 10.4103/0975-7406.160040

Ai, T., Yang, Z., Hou, H., Zhan, C., Chen, C., Lv, W., et al. (2020). Correlation of chest CT and RT-pcr testing for coronavirus disease 2019 (COVID-19) in China: a report of 1014 cases. *Radiology* 2020:200642. doi: 10.1148/radiol.2020200642

Albert, M. A. (2011). Biomarkers and heart disease. *J. Clin. Sleep Med.* 7, S9–S11. doi: 10.5664/JCSM.1342

Ashley, E. A., Recht, J., and White, N. J. (2014). Primaquine: the risks and the benefits. *Malar J.* 13:418. doi: 10.1186/1475-2875-13-418

Beutler, E. (1959). The hemolytic effect of primaquine and related compounds: a review. *Blood* 14, 103–139. doi: 10.1182/blood.V14.2.103.103

Burstein, H. J. (2005). The distinctive nature of HER2-positive breast cancers. *N. Engl. J. Med.* 353, 1652–1654. doi: 10.1056/NEJMp058197

Cascella, M., Rajnik, M., Cuomo, A., Dulebohn, S. C., and Di Napoli, R. (2020). "Features, evaluation and treatment coronavirus (COVID-19)," in *StatPearls*. (Treasure Island, FL: StatPearls Publishing) Available online at: https://www.ncbi.nlm.nih.gov/books/NBK554776/

Chasioti, D., Yan, J., Nho, K., and Saykin, A. J. (2019). Progress in polygenic composite scores in alzheimer's and other complex diseases. *Trends Genet.* 35, 371–382. doi: 10.1016/j.tig.2019.02.005

Conti, R., Veenstra, D. L., Armstrong, K., Lesko, L. J., and Grosse, S. D. (2009). Personalized therapy and pharmacogenomics: future perspective. *Pharmacogenomics* 10, 927–930. doi: 10.2217/pgs.09.45

Cortegiani, A., Ingoglia, G., Ippolito, M., Giarratano, A., Einav, S. (2020). A systematic review on the efficacy and safety of chloroquine for the treatment of COVID-19. *J. Crit. Care* 57, 279–283. doi: 10.1016/j.jcrc.2020.03.005

Dong, L., Hu, S. and Gao, J. (2020). Discovering drugs to treat coronavirus disease (2019). (COVID-19). *Drug Discov. Ther.* 14, 58–60. doi: 10.5582/ddt.2020.01012

Ellinghaus, D., Degenhardt, F., Bujanda, L., Buti, M., Albillos, A., Invernizzi, P., et al. (2020). Genomewide association study of severe covid-19 with respiratory failure. *N. Engl. J. Med.* 383, 1522–1534. doi: 10.1056/NEJMoa2020283

Estabrook, R.W. (2003). A passion for P450s (rememberances of the early history of research on cytochrome P450). *Drug Metab. Dispos.* 31, 1461–1473. doi: 10.1124/dmd.31.12.1461

Feliz, L. R. and Tsimberidou, A. M. (2013). Anti-vascular endothelial growth factor therapy in the era of personalized medicine. *Cancer Chemother. Pharmacol.* 72, 1–12. doi: 10.1007/s00280-013-2124-y

Ferrara, N., and Adamis, A. P. (2016). Ten years of anti-vascular endothelial growth factor therapy. *Nat. Rev. Drug Discov.* 15, 385–403. doi: 10.1038/nrd.2015.17

Fierz, W. (2004). Challenge of personalized health care: to what extent is medicine already individualized and what are the future trends? *Med. Sci. Monit.* 10, 111–123.

Gajria, D., and Chandarlapaty, S. (2011). HER2-amplified breast cancer: mechanisms of trastuzumab resistance and novel targeted therapies. *Expert Rev. Anticancer Ther.* 11, 263–275. doi: 10.1586/era.10.226

Gao, J., Tian, Z., and Yang, X. (2020). Breakthrough: chloroquine phosphate has shown apparent efficacy in treatment of COVID-19 associated pneumonia in clinical studies. *Biosci. Trends* 14, 72–73. doi: 10.5582/bst.2020.01047

Gautret, P., Lagier, J. C., Parola, P., Hoang, V. T., Meddeb, L., Mailhe, M., et al. (2020). Hydroxychloroquine and azithromycin as a treatment of COVID-19: results of an open-label non-randomized clinical trial. *Int. J. Antimicrob. Agents* 56:105949. doi: 10.1016/j.ijantimicag.2020.105949

Harari, D., and Yarden, Y. (2000). Molecular mechanisms underlying ErbB2/HER2 action in breast cancer. *Oncogene* 19, 6102–6114. doi: 10.1038/sj.onc.1203973

Hippocrates (ca. 460 B.C.Bca. 370 B.C.), Iatrike, Basel (1543). Book: 15 (h) x 21 (open width) (WZ 240 H667 1543).

Howes, R. E., Battle, K. E., Satyagraha, A. W., Baird, J. K., and Hay, S. I. (2013). G6PD deficiency: global distribution, genetic variants and primaquine therapy. *Adv. Parasitol.* 81, 133–201. doi: 10.1016/B978-0-12-407826-0.00004-7

Jienchi, D., and Kome, G. (2020). *Lopinavir/ritonavir: A Rapid Review of Effectiveness in COVID-19*. Available online at: https://www.cebm.net/covid-19/lopinavir-ritonavir-a-rapid-review-of-the-evidence-for-effectiveness-in-treating-covid/

Jouanna, J. (2012). *Greek Medicine from Hippocrates to Galen*. (Jacques: Bril). doi: 10.1163/9789004232549

Khalili, J. S., Zhu, H., Mak, N. S. A., Yan, Y., and Zhu, Y. (2020). Novel coronavirus treatment with ribavirin: Groundwork for an evaluation concerning COVID-19. *J. Med. Virol.* 92, 740–746. doi: 10.1002/jmv.25798

Khan, M. S. S., Asif, M., Basheer, M. K. A., Kang, C. W., Al-Suede, F. S., Ein, O. C., et al. (2002). VEGF and the quest for tumour angiogenesis factors. *Nat Rev Cancer*, 2, 795–803. doi: 10.1038/nrc909

Kim, K. J., et al. (1993). Inhibition of vascular endothelial growth factor-induced angiogenesis suppresses tumor growth *in vivo*. *Nature* 362, 841–844. doi: 10.1038/362841a0

Landeck, L., Kneip, C., Reischl, J., and Asadullah, K. (2016). Biomarkers and personalized medicine: current status and further perspectives with special focus on dermatology. *Exp Dermatol.* 25, 333–339. doi: 10.1111/exd.12948

Li, Z., Yi, Y., Luo, X., Xiong, N., Liu, X., Li, S., et al. (2020). Development and clinical application of a rapid IgM-IgG combined antibody test for SARS-CoV-2 infection diagnosis. *J. Med. Virol.* 92, 1518–1524. doi: 10.1002/jmv.25727

Lippi, G., Lavie, C. J., Henry, B. M., and Sanchis-Gomar, F. (2020). Do genetic polymorphisms in angiotensin converting enzyme 2 (ACE2) gene play a role in coronavirus disease 2019. (COVID-19)? *Clin. Chem. Lab. Med.* 58, 1415–1422. doi: 10.1515/cclm-2020-0727

Liu, W., and Li, H. (2020). COVID-19: attacks the 1-beta chain of hemoglobin and captures the porphyrin to inhibit human heme metabolism. *ChemRxiv*. doi: 10.26434/chemrxiv.11938173.v9

Llerena, A., Dorado, P., and Peñas-Lledó, E. M. (2009). Pharmacogenetics of debrisoquine and its use as a marker for CYP2D6 hydroxylation capacity. *Pharmacogenomics* 10, 17–28. doi: 10.2217/14622416.10.1.17

Luo, P., Liu, Y., Qiu, L., Liu, X., Liu, D., Li, J. (2020). Tocilizumab treatment in COVID-19: a single center experience. *J. Med. Virol.* 92, 814–818. doi: 10.1002/jmv.25801

Luzzatto, L., and Arese, P. (2018). Favism and glucose-6-phosphate dehydrogenase deficiency. *N. Engl. J. Med.* 378, 60–71. doi: 10.1056/NEJMra1708111

Mahgoub, A., Idle, J. R., Dring, L. G., Lancaster, R., and Smith, R. L. (1977). Polymorphic hydroxylation of debrisoquine in man. *Lancet* 2, 584–586. doi: 10.1016/S0140-6736(77)91430-1

Maréchal, J. D., Kemp, C. A., Roberts, G. C., Paine, M. J., Wolf, C. R., Sutcliffe, M. J. (2008). Insights into drug metabolism by cytochromes P450 from modelling studies of CYP2D6-drug interactions. *Br. J. Pharmacol.* 153(Suppl 1), S82–S89. doi: 10.1038/sj.bjp.0707570

Mini, E., and Nobili, S. (2009). Pharmacogenetics: implementing personalized medicine. *Clin. Cases Miner. Bone Metab*. 6, 17–24.

National Institutes of Health (2007). *Biological Sciences Curriculum Study. NIH Curriculum Supplement Series [Internet]. Bethesda (MD): National Institutes of Health (US); Understanding Human Genetic Variation.* Available online at: https://www.ncbi.nlm.nih.gov/books/NBK20363/

Nimmesgern, E., Benediktsson, I., and Norstedt, I. (2017).Personalized medicine in europe. *Clin. Transl. Sci.* 10 61–63. doi: 10.1111/cts.12446

Overby, C. L., and Tarczy-Hornoch, P. (2013). Personalized medicine: challenges and opportunities for translational bioinformatics. *Personal. Med.* 10, 453–462. doi: 10.2217/pme.13.30

Pang, J., Massidda, B., Natale, D., Putzu, C., Sandomenico, C., Filippelli, G., et al. (2020). Efficacy and tolerability of bevacizumab in patients with severe Covid−19. *medRxiv [Preprint]*. doi: 10.1101/2020.07.26.20159756

Petherick, A. (2020). Developing antibody tests for SARS-CoV-2. *Lancet* 395, 1101–1102. doi: 10.1016/S0140-6736(20)30788-1

Policy brief (2016). *Testing for G6PD Deficiency for Safe Use of Primaquine in Radical Cure of P. Vivax and P. ovale*. Available online at: https://www.who.int/malaria/publications/atoz/g6pd-testing-pq-radical-cure-vivax/en/

Principi, N., and Esposito, S. (2020). Chloroquine or hydroxychloroquine for prophylaxis of COVID-19. *Lancet Infect. Dis.* 20:1118. doi: 10.1016/S1473-3099(20)30296-6

Ramakrishnan, S., Anand, V., and Roy, S. (2014). Vascular endothelial growth factor signaling in hypoxia and inflammation. *J. Neuroimmun. Pharmacol.* 9, 142–160. doi: 10.1007/s11481-014-9531-7

Relling, M. V., and Evans, W. E. (2015). Pharmacogenomics in the clinic. *Nature* 526, 343–350. doi: 10.1038/nature15817

Rogers, J. F., Nafziger, A. N., and Bertino, J. S. (2002). Pharmacogenetics affects dosing, efficacy, and toxicity of cytochrome P450-metabolized drugs. *Am. J. Med*. 113, 746–750. doi: 10.1016/S0002-9343(02)01363-3

Siest, G., Jeannesson, E., and Visvikis-Siest, S. (2007). Enzymes and pharmacogenetics of cardiovascular drugs. *Clin. Chim. Acta* 381, 26–31. doi: 10.1016/j.cca.2007.02.014

Silas, J. H., Lennard, M. S., Tucker, G. T., Smith, A. J., Malcolm, S. L., and Marten, T. R. (1977). Why hypertensive patients vary in their response to oral debrisoquine. *Br. Med. J.* 1, 422–425. doi: 10.1136/bmj.1.6058.422

Slamon, D., Eiermann, W., Robert, N., Pienkowski, T., Martin, M., Press, M., et al. (2011). Adjuvant trastuzumab in HER2-positive breast cancer. *N. Engl. J. Med*. 365, 1273–1283. doi: 10.1056/NEJMoa0910383

Slamon, D. J., Clark, G. M., Wong, S. G., Levin, W. J., Ullrich, A., and McGuire, W. L. (1987). Human breast cancer: correlation of relapse and survival with amplification of the HER-2/neu oncogene. *Science* 235, 177–182. doi: 10.1126/science.3798106

Slamon, D. J., Leyland-Jones, B., Shak, S., Fuchs, H., Paton, V., Bajamonde, A., et al. (2001). Use of chemotherapy plus a monoclonal antibody against her2 for metastatic breast cancer that overexpresses HER2. *N. Engl. J. Med.* 344, 783–792. doi: 10.1056/NEJM200103153441101

Sorich, M. J., and McKinnon, R. A. (2012). Personalized medicine: potential, barriers and contemporary issues. *Curr. Drug Metab.* 13, 1000–1006. doi: 10.2174/138920012802138615

Stower, H. (2020). Lopinavir–ritonavir in severe COVID-19. *Nat. Med.* 26:465. doi: 10.1038/s41591-020-0849-9

Sykiotis, G. P., Kalliolias, G. D., and Papavassiliou, A. G. (2005). Pharmacogenetic principles in the hippocratic writings. *J Clin. Pharmacol.* 45, 1218–1220. doi: 10.1177/0091270005281091

Talapko, J., Škrlec, I., Alebić, T., Jukić, M., and Včev, A. (2019). Malaria: the past and the present. *Microorganisms* 7:179. doi: 10.3390/microorganisms7060179

Teichmann, S., and Regev, A. (2020). The network effect: studying COVID-19 pathology with the Human Cell Atlas. *Nat. Rev. Mol. Cell Biol.* 21, 415–416. doi: 10.1038/s41580-020-0267-3

Touret, F., and de Lamballerie, X. (2020). Of chloroquine and COVID-19. *Antiviral Res.* 177:104762. doi: 10.1016/j.antiviral.2020.104762

Visvikis-Siest, S., Gorenjak, V., and Stathopoulou, M. G. (2018). Personalised medicine: the odyssey from hope to practice. *J. Pers. Med.* 8:31. doi: 10.3390/jpm8040031

Vogenberg, F. R., Isaacson Barash, C., and Pursel, M. (2010). Personalized medicine: part 1: evolution and development into theranostics. *P T* 35, 560–576.

Wu, B., Wu, Y., and Tang, W. (2019). Heme catabolic pathway in inflammation and immune disorders. *Front. Pharmacol.* 10:825. doi: 10.3389/fphar.2019.00825

Wurcel, V., Perche, O., Lesteven, D., Williams, D. A., Schäfer, B., Hopley, C., et al. (2016). The value of companion diagnostics: overcoming access barriers to transform personalised health care into an affordable reality in Europe. *Public Health Genom.* 19, 137–143. doi: 10.1159/000446531

Xu, K., Chen, Y., Yuan, J., Yi, P., Ding, C., Wu, W., et al. (2020). Clinical efficacy of arbidol in patients with 2019 novel coronavirus-infected pneumonia: a retrospective cohort study. *Lancet*. doi: 10.2139/ssrn.3542148. [Epub ahead of print]

Yang, Y., Botton, M. R., Scott, E. R., and Scott, S. A. (2017). Sequencing the CYP2D6 gene: from variant allele discovery to clinical pharmacogenetic testing. *Pharmacogenomics* 18, 673–685. doi: 10.2217/pgs-2017-0033

Zhu, Z., Lu, Z., Xu, T., Chen, C., Yang, G., Zha, T., et al. (2020). Arbidol monotherapy is superior to lopinavir/ritonavir in treating COVID-19. *J. Infect.* 81, e21–e23.doi: 10.1016/j.jinf.2020.03.060

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 in Genetics

ORIGINAL RESEARCH
published: 15 March 2021
doi: 10.3389/fgene.2021.595702

Investigations of Kidney Dysfunction-Related Gene Variants in Sickle Cell Disease Patients in Cameroon (Sub-Saharan Africa)

*Valentina J. Ngo-Bitoungui[1,2,3], Suzanne Belinga[4], Khuthala Mnika[2], Tshepiso Masekoameng[2], Victoria Nembaware[1], René G. Essomba[5,6], Francoise Ngo-Sack[7], Gordon Awandare[1], Gaston K. Mazandu[2,8] and Ambroise Wonkam[2]**

[1] West African Centre for Cell Biology of Infectious Pathogens, University of Ghana, Legon-Accra, Ghana, [2] Division of Human Genetics, Department of Medicine, Faculty of Health Sciences, University of Cape Town, Cape Town, South Africa, [3] Department of Microbiology Haematology and Immunology, University of Dschang, Yaoundé, Cameroon, [4] Centre Pasteur of Cameroon, Yaoundé, Cameroon, [5] National Public Health Laboratory, Yaoundé, Cameroon, [6] Department of Microbiology, Parasitology, Haematology, Immunology and Infectious Diseases, Faculty of Medicine and Biomedical Sciences, University of Yaounde 1, Yaounde, Cameroon, [7] Faculty of Medicine and Pharmaceutical Sciences, University of Douala, Douala, Cameroon, [8] African Institute for Mathematical Sciences, Muizenberg, Cape Town, South Africa

OPEN ACCESS

Edited by:
Maria G. Stathopoulou, Institut National de la Santé et de la Recherche Médicale (INSERM), France

Reviewed by:
Leon Mutesa, University of Rwanda, Rwanda
Henu Verma, Consiglio Nazionale Delle Ricerche (CNR), Italy

***Correspondence:**
Ambroise Wonkam
ambroise.wonkam@uct.ac.za

Specialty section:
This article was submitted to Genomic Medicine, a section of the journal Frontiers in Genetics

Received: *17 August 2020*
Accepted: *19 February 2021*
Published: *15 March 2021*

Citation:
Ngo-Bitoungui VJ, Belinga S, Mnika K, Masekoameng T, Nembaware V, Essomba RG, Ngo-Sack F, Awandare G, Mazandu GK and Wonkam A (2021) Investigations of Kidney Dysfunction-Related Gene Variants in Sickle Cell Disease Patients in Cameroon (Sub-Saharan Africa). Front. Genet. 12:595702. doi: 10.3389/fgene.2021.595702

Background: Renal dysfunctions are associated with increased morbidity and mortality in sickle cell disease (SCD). Early detection and subsequent management of SCD patients at risk for renal failure and dysfunctions are essential, however, predictors that can identify patients at risk of developing renal dysfunction are not fully understood.

Methods: In this study, we have investigated the association of 31 known kidney dysfunctions-related variants detected in African Americans from multi-ethnic genome wide studies (GWAS) meta-analysis, to kidney-dysfunctions in a group of 413 Cameroonian patients with SCD. Systems level bioinformatics analyses were performed, employing protein-protein interaction networks to further interrogate the putative associations.

Results: Up to 61% of these patients had micro-albuminuria, 2.4% proteinuria, 71% glomerular hyperfiltration, and 5.9% had renal failure. Six variants are significantly associated with the two quantifiable phenotypes of kidney dysfunction (eGFR and crude-albuminuria): *A1CF-rs10994860* (P = 0.02020), *SYPL2-rs12136063* (P = 0.04208), and *APOL1 (G1)-rs73885319* (P = 0.04610) are associated with eGFR; and *WNT7A-rs6795744* (P = 0.03730), *TMEM60-rs6465825* (P = 0.02340), and *APOL1 (G2)-rs71785313* (P = 0.03803) observed to be protective against micro-albuminuria. We identified a protein-protein interaction sub-network containing three of these gene variants: *APOL1, SYPL2,* and *WNT7A*, connected to the Nuclear factor NF-kappa-B p105 subunit (NFKB1), revealed to be essential and might indirectly influence extreme phenotypes. Interestingly, clinical variables, including body mass index (BMI), systolic

blood pressure, vaso-occlusive crisis (VOC), and haemoglobin (Hb), explain better the kidney phenotypic variations in this SCD population.

Conclusion: This study highlights a strong contribution of haematological indices (Hb level), anthropometric variables (BMI, blood pressure), and clinical events (i.e., vaso-occlusive crisis) to kidney dysfunctions in SCD, rather than known genetic factors. Only 6/31 characterised gene-variants are associated with kidney dysfunction phenotypes in SCD samples from Cameroon. The data reveal and emphasise the urgent need to extend GWAS studies in populations of African ancestries living in Africa, and particularly for kidney dysfunctions in SCD.

Keywords: sickle cell disease, kidney dysfunctions, gene variants, cameroon, Africa

INTRODUCTION

Sickle Cell Disease (SCD) is a monogenic disease with high prevalence and high mortality rates in Africa. Globally, SCD is estimated to affect more than 300,000 births per year, with nearly two-thirds occurring in sub-Sahara Africa (Piel et al., 2013). Cameroon is a sub-Saharan African country with an estimated SCD carrier frequency rate between 8 and 34% (Weatherall and Clegg, 2001). Although Cameroon declared SCD as a public health priority, access to care and treatment is still limited due to lack of a national medical insurance, leaving SCD patients to self-fund or depend on familial financial support. Therefore, medical care costs are often not met (Wonkam et al., 2014) and patients frequently suffer from severe SCD complications such as kidney dysfunction (Geard et al., 2017).

Renal failure caused by recurrent episodes of ischemia-reperfusion injury and haemolytic anaemia, occurs in 5–18% of SCD patients and is associated with an increased risk of early mortality (Platt et al., 1994; Gladwin, 2017). The prevention of renal failure relies on early detection and management of kidney dysfunction. In SCD patients, renal failure can be caused by gradual infiltration of glomerulus, which leads to glomerular sclerosis or promotes progression of micro-albuminuria to macro-albuminuria/proteinuria and finally to nephrotic-range proteinuria (Nath and Hebbel, 2015). Micro-albuminuria is prevalent in 26–68% of adult patients (Ataga et al., 2014; Gosmanova et al., 2014) and is the most sensitive early clinical marker for glomerular damage and other types of kidney dysfunction. Recent studies have demonstrated that the co-inheritance of alpha-thalassemia with SCD and/or specific variants in the HbF promoting loci can delay the clinical progression of kidney disease in African American SCD patients (Saraf et al., 2017). In addition, genetic variations in two coding regions of *Apolipoprotein L1 (APOL1)* and *Heme oxygenase 1 (HMOX1)* genes have been associated to chronic kidney disease (Genovese et al., 2010; Tzur et al., 2010), and to SCD nephropathy (Saraf et al., 2015; Schaefer et al., 2016).

Abbreviations: SCD, sickle cell disease; eGFR, estimated glomerular filtration rate; VOC, vaso-occlusive crisis; GWAS, genome-wide association studies; BMI, Body mass index; BP, blood pressure; HbF, foetal haemoglobin; HBB, beta-globin gene; HBA, alpha-globin gene; PCR, polymerase chain reaction; RFLP, Restriction fragment length polymorphism; WBC, white blood cell; Hb, haemoglobin; MCV, mean corpuscular volume; CKD-EPI, Chronic Kidney Disease-Epidemiology.

In a previous study, we showed that variants in *APOL1* and *HMOX1* variants are associated with kidney dysfunctions using a targeted SNP based approach. Further investigations reveal that these variants are associated with albumin creatinine ratio, micro-/macro-albuminuria and eGFR in a group of SCD patients in Cameroon (Geard et al., 2017). Given the high rate of renal dysfunction in SCD patients in Africa with its high genetic diversity, there is need to explore possible novel genetic variants associated with kidney dysfunction. Ideally, whole genome sequencing and other large-scale gene discovery approaches should be use, expanding the targeted genetic discovery to other variants known to be associated with renal dysfunctions. Given the limited number of gene and variant discovery research in SCD patients, a plausible strategy is using renal dysfunctions associated SNPs from non-SCD affected populations.

Several kidney dysfunction genome-wide association studies (GWAS) have been conducted in many non SCD-affected populations (Kottgen et al., 2009, 2010; Chambers et al., 2010; Pattaro et al., 2012). Furthermore, a recent GWAS meta-analysis integrated 15 GWAS studies of 133,413 individuals from multiple ethnicities and uncovered 53 SNPs associated with renal dysfunction, including 26 SNPs found in individuals of African descent (Pattaro et al., 2016). In this study, we investigated the associations of these 26 SNPs in addition to four previously characterised kidney dysfunction-related variants, including *APOL1* (G1 or G2) for *rs60910145, rs73885319* and *rs71785313*, and *HMOX1* for *rs3074372* and *rs743811*, relevant to populations of African ancestry (Pattaro et al., 2016), e.g., SCD patients from Cameroon.

Several studies, including a previous study from our group (Geard et al., 2017), have shown that, in addition to genetic variants, clinical, and biological factors also contribute to glomerular damage (Audard et al., 2017). This highlights the need to employ a multi-factorial approach in investigating factors associated to renal abnormalities in SCD patients. Therefore, in addition to investigating the contribution of the selected 31 SNPs to renal dysfunctions in SCD patients from Cameroon, we also explored the contribution of clinical factors: socio-demographic, anthropometric, clinical and haematological variables, and employed multifactorial regression models for associating the variable to kidney dysfunction parameters. Finally, systems level bioinformatics analyses were performed,

employing protein-protein interaction networks to further interrogate the putative associations.

MATERIALS AND METHODS

Ethical Approval

The study was performed following the Declaration of Helsinki. This study was approved by the Faculty of Health Sciences Human Research Ethics Committee of the University of Cape Town, South Africa (HREC REF: 661/2015), and the National Ethics Committee of the Ministry of Public Health, Yaoundé, Republic of Cameroon (No. 193/CNE/SG/10). Patients older than 18 years self-consented into the study and informed consent was given by the parents or guardians for participants younger than 18 years old with a requirement for children older than 7 years to also sign assent forms.

Study Participants

Cameroonians living with SCD were prospectively recruited at the Yaoundé Central Hospital and Laquintinie Hospital in Douala, between January 2010 and December 2012. Only patients older than 2 years of age who had not received a blood transfusion in the past 6 months were included. None of the patients were receiving *hydroxycarbamide* treatment. Community-based recruitments were also conducted through two SCD patients' associations who were engaged for collaboration. Additional patients were subsequently recruited during the SCD patient associations' monthly meetings. No incentive was provided for participation in the study. Socio-demographic and clinical events were collected by means of a structured questionnaire administered to parents/guardians or adult patients. Body mass index (BMI) and blood pressure (BP) were also measured. Patients' clinical records were extracted from their medical records covering the past 3 years. These clinical records include blood transfusion history, the occurrence of vaso-occlusive crisis (VOC) and hospitalisation rates per year.

Hematological Phenotypes

Complete blood counts and foetal haemoglobin (HbF) quantifications were conducted during hospital visits. Two methods of HbF detection were employed in this study for patients: initially using the alkali denaturation test in 55% of the cohort and subsequently High-Performance Liquid Chromatography when it became available at the haematological laboratory of the Centre Pasteur of Cameroon (CPC).

Renal Function Measurements

Urinary albumin quantifications were performed using either the Siemens Clinitek Status test or the Hemocue Albumin 20 system on the first morning urine samples during planned hospital visits, when patients were not experiencing VOC, as previously reported by Geard et al. (2017). The presence of albumin in the urine is defined as normal when the concentration is <30 mg/dl, micro-albuminuria (30–300 mg/dl) or macro-albuminuria >300 mg/dl. The glomerular filtration rate (GFR) is estimated (eGFR) using the Chronic Kidney Disease-Epidemiology Collaboration (CKD-EPI-creatinine) formula. Kidney failure is defined as an eGFR <90 ml/min/1.73 m^2, renal hyperfiltration as an eGFR > 130 ml/min/1.73 m^2 for women and >140 ml/min/1.73 m^2 for men, and normal filtration as an eGFR between 90 ml/min/1.73 m^2 and 130/140 ml/min/1.73 m^2 (Haymann et al., 2010).

Molecular Methods

DNA Extraction

DNA was extracted from peripheral blood in EDTA following the manufacturer's instructions (Puregene Blood Kit) at CPC, and Genotype analyses were performed at the Division of Human Genetics, University of Cape Town, South Africa.

Sickle Cell Mutation, Beta-Globin Gene (HBB) Cluster Haplotypes and 3.7 kb Alpha-Globin Gene (HBA1/HBA2) Deletion

Molecular analysis to confirm the presence of the sickle mutation w carried out on 200 ng of DNA by polymerase chain reaction (PCR) to amplify a 770 bp segment of *HBB*, followed by a digestion with DdeI restriction enzyme on the PCR product (Saiki et al., 1985).

Five restriction fragment length polymorphism (RFLP) sites in the *HBB* cluster were amplified using published primers and methods to analyse the HBB haplotype background (Bitoungui et al., 2015).

The 3.7 kb HBA_1/HBA_2 deletion was successfully screened using the expand-long template PCR as previously published (Rumaney et al., 2014).

Kidney Dysfunction -Related Targeted Variants

Twenty-six African American specific kidney dysfunction-related gene variants were genotyped in this SCD cohort after being mapped to 53 single nucleotide polymorphisms (SNPs) identified in a GWAS meta-analysis of kidney diseases (Pattaro et al., 2016). Five additional gene variants in *APOL1* and *HMOX1* from the literature were also considered (Genovese et al., 2010; Tzur et al., 2010; Wonkam et al., 2014; Geard et al., 2017; Saraf et al., 2017). Thus, all targeted SNPs in *A1CF-rs10994860, WNT7A-rs6795744, PTPRO-rs7956634, UMOD-rs4293393, LRP2 -rs4667594 ANXA9-rs267734, GCKR-rs1260326, TFDP2-rs347685, DAB2-rs11959928, SLC22A2-rs2279463, TMEM60-rs6465825, SLC6A13-rs10774021, BCAS1-rs17216707, SKIL-rs9682041, UNCX-rs10277115, KBTBD2-rs3750082, CNQ1-rs163160, AP5B1-rs4014195, NFKB1-rs228611, CACNA1S-rs3850625, SYPL2-rs12136063, ETV5-rs10513801, DPEP1-s164748, SIPA1L3-rs11666497, NFATC1-rs8091180 and IGFBP5-rs2712184,* and in *APOL1 (G1)-rs60910145, APOL1 (G1)-rs73885319, APOL1 (G2)-rs71785313, HMOX1-rs3074372, HMOX1-rs743811,* were genotyped with the iPLEX Gold Sequenom Mass Genotyping Array. Thereafter, the validation of the genotyping results was done by Sanger sequencing using BigDye terminator mix in 10% subset of sample (**Supplementary Figure S1**).

Statistical and Bioinformatics Analysis

We performed association analyses between kidney dysfunction outcomes (characterised by eGFR and crude-albuminuria scores), socio-demographic and clinical variables, and 31 known kidney dysfunction-related variants among this group of SCD patients. First, to ensure genotypic quality of the data, we ran PLINK 1.9 (Purcell et al., 2007), performing a Hardy-Weinberg Equilibrium (HWE) test with significant level, minor allele frequency (MAF), and missing genotype data thresholds of 0.001, 0.05, and 0.1, respectively. A total of 13 out of 31 SNPs in the dataset did not pass quality control (QC) filters, of which 8 were removed due to missing genotype data and 5 others do not meeting the set minor allele thresholds. With R software, we first performed descriptive statistics to provide a general summary of different parameters to be considered in the analysis. Thereafter, two regression analyses were performed: (1) multi-variable regressions for each kidney dysfunction phenotype with different socio-demographic, anthropometric, haematogical, clinical variables and each genetic variant after adjusting or transforming phenotype values to approximate a symmetric (normal) distribution based on their Fisher-Pearson skewness coefficient scores; (2) logistic regressions for each kidney dysfunction phenotype and all variables and genetic variants under consideration after mapping phenotypes to 0 (controls) or 1 (cases: micro- and macro-albuminuria for crude-albuminuria, then renal failure and hyperfiltration for eGFR). Finally, we performed functional and protein–protein interaction network enrichment analyses, using Gene Ontology (GO) process (The Gene Ontology Consortium, 2019), the protein GO Annotation (GOA) mapping (Mi et al., 2019) and the Kyoto Encyclopeadia of Genes and Genomes (KEGG) pathway (Kanehisa et al., 2019) datasets, to identify potential enriched biological processes and pathways in which identified candidate genes are involved. A significance level of 0.05 was considered after adjusting *p*-values (P) for Bonferroni multiple corrections and gene functional annotations were retrieved from the Ensembl database (Yates et al., 2020).

RESULTS

Description of Study Participants

A total of 413 SCD steady state Cameroonian patients were included in the study. The participants' characteristics are described in **Table 1**. There are roughly equal numbers of males and females (M/F = 210/203), with a median age of 15.5 years. The median number of VOC per year was 2 (range: 0–80). 41% (n = 168) of patients had (3 VOC per year and 28.1% (n (=114) had 2 hospitalizations hospitalisations per year. All participants were homozygous HbSS, Benin being the most prevalent (-globin-like gene cluster haplotype (55%; n = 195). It is worth noting that, in this study, we used the modified annotation protocol suggested by Crawford et al. (2002) and Hanchard et al. (2007) for determining β-globin gene cluster haplotypes Benin, Cameroon, Bantu, Senegal and Arab-Indian. Haplotypes that are not conform to these are considered to be "Atypical." We refer to "Other haplotypes," all underrepresented haplotypes in the cohort, covering Bantu, Senegal, and Arab-Indian. 32 and 11% co-inherited a single or double 3.7 kb HBA_1/HBA_2 deletion, respectively.

Clinical and Socio-Demographic Factors Associated With Renal Dysfunctions

For the genotype dataset and two renal phenotypes under consideration, a prior QC and value adjustment processes were performed, respectively. Only 18 out the 31 genetic variants (highlighted in **Table 2**) were considered for further analyses: eight variants were removed due to missing genotype data (genotype call rate 0.9: for cut-off of 0.1) and 5 had minor allele frequency of less than 5%. For each phenotype to approximate a symmetric or normal distribution, we computed the Fisher-Pearson skewness coefficients to select the type of transformation required. For eGFR, no value adjustment was required as its skewness coefficient was 0.27581, comprised in the range of -0.5 and 0.5 fitting a symmetric distribution. For crude-albuminuria, the skewness coefficient is $6.45456 \geq 1$, in which case, the log10 transformation was applied to ensure that crude-albuminuria dataset approximate a symmetric distribution (**Figure 1**).

After the transformation, we performed linear regression models and factors associated with crude albuminuria and eGFR in Cameroonian SCD patients are presented in **Table 3**. Age and gender as well as creatinine were not considered in the eGFR-based regression models to prevent biases in outputs as these two factors are confounders, mixing up with eGFR as an outcome (because age, gender and creatinine contribute to the eGFR calculation).

eGFR

the level of serum creatinine used to estimate GFR had a median of 6.8 mg/l and the eGFR median was 155.4 ml/min/1.73 m^2 (**Table 1**). Up to 71% of the patients had glomerular hyperfiltration and 5.9% renal failure (see overall age-based population distributions in **Figure 2A**). **Figure 2A** suggests that the prevalence of glomerular hyperfiltration is high amongst patients under 21 years old and kidney failure is relatively high up to 30 years, decreasing after 30 years. This agrees with another Canadian study which highlighted that children with SCD between 4 and 11 years have a significantly higher mean eGFR (Mammen et al., 2017). The eGFR was significantly increased in male patients (P = 4.65297e–10). Isolated hyperfiltration was present in 25.8% (n = 105) of patients, while 41.5% (n = 169) and 2.2% (n = 9) were experiencing glomerular hyperfiltration with micro- and macro-albuminuria, respectively. Haemoglobin, HbF, MCV, platelet, lymphocytes, and granulocytes presented highly significant p-values and positive correlation with eGFR, explaining some proportion of eGFR variations (**Table 3**). This suggested that there was a reduced protection of kidney from haemoglobin mediated toxicity. Though high HbF level as well as lymphocytes and granulocytes have beneficial clinical effect on SCD patients and results indicated that an

TABLE 1 | Description of the Cameroonian SCD cohort.

Variables	Median (95% CI) or frequency (%)	Min–max	*P*-value	Observation (*n*)
Age (years)	15.5 (15–16.5)	2–58	0.1123	413
Gender	M/F: 210/203			413
Clinical Events				
VOC (*n*/year)	2 (2–3)	0–80	0.4949	412
Hospitalisation (*n*/year)	1 (1–1)	0–40	0.2676	405
Transfusion	Y/N: 76.8/23.2			410
Anthropometric				
Body mass index (kg/m^2)	17.6 (17.2–17.9)	10.6–32	0.1155	413
Systolic blood pressure (mmHg)	103 (101.5–104)	72–156	0.8637	409
Diastolic blood pressure (mmHg)	58 (57–59)	37–93	0.0162	409
Haematological				
Hb (g/dl)	7.55 (7.4–7.7)	3.5–13.1	0.4905	406
MCV (fl)	91 (90–92)	61–125	0.9905	406
Platelets (10^9/l)	368.5 (355–382)	29–1,078	0.8483	406
Leucocytes (10^9/l)	12.65 (12.2–13.2)	4–49.8	0.0471	406
Lymphocytes (10^9/l)	5.2 (4.95–5.4)	1.4–22.1	0.0219	406
Monocytes (10^9/l)	1.5 (1.45—-1.6)	0.1–8.1	0.0009	406
Granulocytes (10^9/l)	4.5 (4.25–4.75)	0.2–24.3	0.0976	406
HbA2 (%)	3.15 (2.9–3.3)	0–18.2	0.0008	412
HbF (%)	9.95 (9.3–10.65)	0–37.2	0.2026	412
Alpha-thalassemia Genotypes (%)			<0.0001	339
αα/αα	57			194/339*
αα/α3.7	32			109/339*
α3.7/α3.7	11			36/339*
HBB Haplotype (%)			<0.0001	352
Ben/Ben	55			195/352*
Ben/Cam	26			92/352*
Ben/Atypical	7			25/352*
Cam/Cam	7			23/352*
Cam/Atypical	2			7/352*
Atypical	1			3/352*
Other haplotypes	2			7/352*
Renal functions				
Serum creatinine (mg/l)	6.8 (6.5–7)	2–13.8	0.3958	404
Crude-albuminuria (mg/dl)	51.5 (47–55.5)	3–1,180	<0.0001	407
Normal (%)	37			149/407
Micro-albuminuria (mg/dl) (%)	61			248/407
Proteinuria (mg/dl) (%)	2			10/407
eGFR (ml/min/1.73 m^2)	155.4 (151.6–159.2)	58.9–290.7	0.5715	404
Normal (%)	23			93/404
Glomerular hyperfiltration (%)	71			287/404
Kidney failure (%)	6			24/404

**Number of individuals, not alleles. VOC, vaso-occlusive crises; Hb, haemoglobin; MCV, mean corpuscular volume; HbA_2, Haemoglobin A_2; eGFR, estimated glomerular filtration rate. In all cases, Wilcoxon tests were performed to check whether the parameter falls within the confidence interval (CI) for p-value greater than 0.05, except for alpha-thalassemia genotypes and the HBB Haplotype, for which adjusted ($\chi 2$ p-values were computed to check whether these proportions follow a uniform distribution for p-value greater than 0.05. In the second column, median values are followed by CI, whereas proportions are not.*

increased level of these variables beyond the steady state likely increase the risk of clinical complications. Finally, eGFR is also correlated to systolic BP (P = 0.04674) and VOC frequencies (P = 0.02802), explaining 40.55 and 21.73% of the eGFR variation, respectively (**Table 3**), with stable systolic BP being highly protective against glomerular hyperfiltration.

Crude-Albuminuria

The median of crude-albuminuria was 51.5 mg/dl ranging from 3 to 1,180 mg/dl. The prevalence of micro and macro-albuminuria was 61 and 2%, respectively. All age ranges have nearly equal proportions of micro-albuminuria, while proteinuria is shown to increase with age (**Figure 2B**). **Figure 2B** shows similar profile as for eGFR in **Figure 2A** and indicates that

TABLE 2 | Allele frequencies of kidney dysfunction-related gene variants.

Gene	dbSNP ID	SNP position	Allele change	MAF	Proven disease associations (Ensembl)
UNCX	rs10277115	7:1245559	A > T	0.16	Renal function related trait
APOL1 (G1)	rs73885319	22:36265860	T > G	0.13	Renal function related trait
APOL1 (G2)	rs71785313	22:36266000	TTATAA > Deletion	0.082	Renal function related trait
A1CF	rs10994860	10:50885664	C > T	0.24	Glomerular filtration rate
DAB2	rs11959928	5:39397030	T > A	0.32	Chronic kidney disease
SYPL2	rs12136063	1:109471548	G > A	0.29	Glomerular filtration rate
GCKR	rs1260326	2:27508073	C > T	0.057	-
KCNQ1	rs163160	11:2768725	A > G	0.057	Glomerular filtration rate
SLC22A2	rs2279463	6:160247357	A > G	0.23	Chronic kidney disease
NFKB1	rs228611	4:102640552	G > A	0.228	Glomerular filtration rate
IGFBP5	rs2712184	–	C > A	0.46	–
TFDP2	rs347685	3:142088295	A > C	0.248	Chronic kidney disease
UMOD	rs4293393	16:20353266	A > G	0.227	Chronic kidney disease
TMEM60	rs6465825	7:77787122	C > T	0.45	Chronic kidney disease
WNT7A	rs6795744	3:13865353	G > A	0.191	Glomerular filtration rate
PTPRO	rs7956634	12:15168260	C > T	0.45	Glomerular filtration rate
NFATC1	rs8091180	18:79404243	G > A	0.068	Glomerular filtration rate
SKIL	rs9682041	3:170374114	T > C	0.274	Glomerular filtration rate

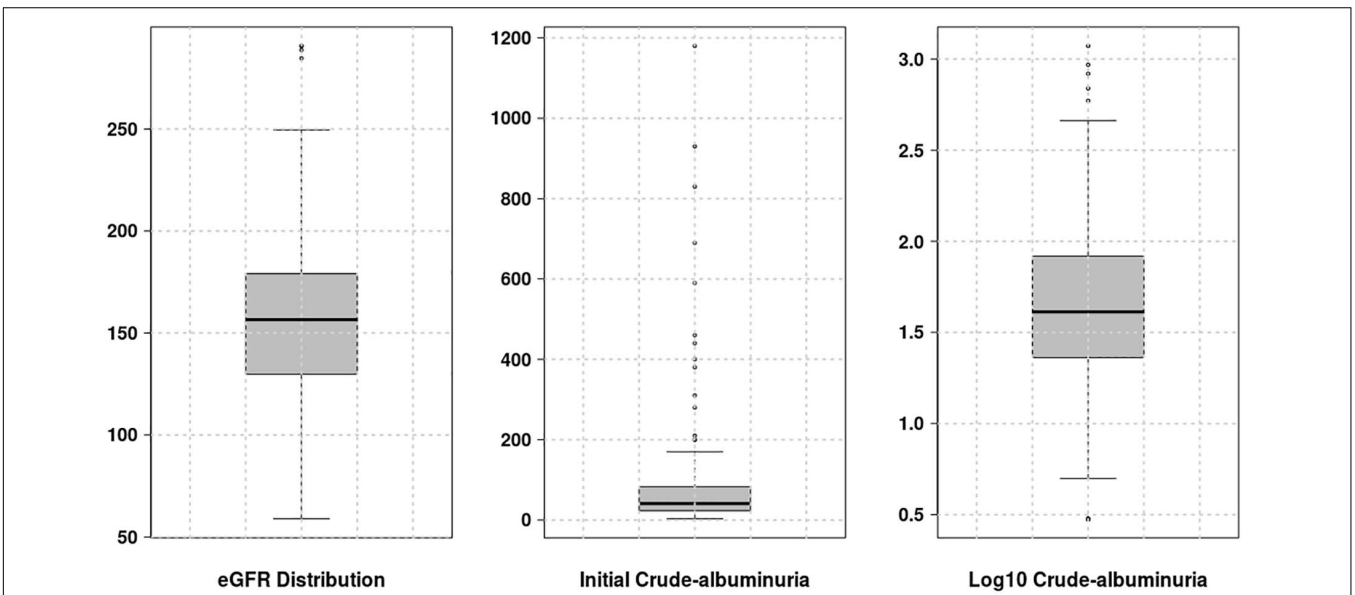


FIGURE 1 | Distribution of different phenotype values. For eGFR, no transformation was required, however, for crude-albuminuria, the initial distribution is highly skewed and log10 transformation was applied to approximate a normal distribution.

the prevalence of microalbuminurea is high amongst patients under 21 years old, with some proteinurea cases that vanish after 30 years of ages. This also agrees with another cohort study in Ghana (Anto et al., 2019), which indicated that the prevalence of renal complications, such as proteinuria, is high in young patients aged between 5 and 12 years. BMI, MCV, haemoglobin and transfusion are significantly associated with crude albuminuria with the highest phenotypic variation explained by BMI (45.34%) and Haemoglobin with reduced risk of micro-albuminuria, though with small effect size (**Table 3**).

Associations of eGFR and Crude-Albuminuria With Kidney Disfunction-Related Gene Variants

Three genetic variants: *Synaptophysin-like protein 2* (*SYPL2-rs12136063*), *APOBEC1 complementation factor* (*A1CF-rs10994860*) and *Apolipoprotein L1* (*APOL1 (G1)-rs73885319*) are significantly associated with eGFR with respective *P*-values of 0.04208, 0.02020, and 0.04610. **Figure 3** shows the eGFR values for three gene variants distributed in homozygous dominant, recessive, and heterozygous genotypes. These distribution

TABLE 3 | Blood pressure, clinical, and haematological variables, and genetic variants associated eGFR and Crude albuminuria.

eGFR (ml/min/1.73 m^2)	Effect size (SE)	Mean variation explained (%)	*P*-values
SBP	−15.01547 (7.51850)	40.55	0.04674
VOC	0.79123 (0.35834)	21.73	0.02802
Hb	3.11861 (1.40541)	16.44	0.02725
MCV (fl)	0.68573 (0.15854)	1.90	2.09130e–05
HbF	1.62340 (0.32960)	1.13	1.40989e–06
Platelet	0.05052 (0.01497)	0.28	8.35290e–04
Granulocytes	2.06830 (0.76779)	0.12	7.47090e–03
Lymphocytes	2.12071 (0.97584)	0.08	0.03057
rs12136063*	0.57907 (0.28487)	11.04	0.04208
rs10994860*	−0.69852 (0.30075)	0.18	0.02020
rs73885319*	0.890305 (0.44638)	0.97	0.04610
Crude albuminuria (mg/dl)			
BMI	0.01872 (0.00596)	45.34	1.87777e–03
MCV (fl)	0.01273 (0.00121)	1.22	1.17630e–07
Transfusion	0.12406 (0.05602)	0.23	0.02759
Hb*	−0.28340 (0.11040)	0.19	0.0103
rs6795744*	−0.59520 (0.28580)	3.70	0.03730
rs6465825*	−0.46910 (0.20690)	0.65	0.02340
rs71785313	−0.12686 (0.06086)	0.10	0.03803

*Effect provides an indication on the influence of a given factor on the phenotype under consideration and mean variation explained the average contribution of a specific factor to the phenotype. SE stands for standard error and * indicates that the association was inferred from logistic model.*

values suggested that the two allele changes (homozygous recessive) of *A1CF-rs10994860* is protective against renal dysfunction (**Table 3**). On the other hand, one allele change or heterozygous genotype showed a protective effect against prevalent hyperfiltration considering the negative correlation with eGFR (conferring about 2 times more protection as compared to a patient with no copy). *SYPL2-rs12136063* confers increased risk of progressing to the renal dysfunction and *APOL1 (G1)-rs73885319* to prevalent hyperfiltration and ultimately to the renal dysfunction (approximately 2 times more likely to progress to renal dysfunction).

Three genetic variants were also identified to be significantly associated with crude-albuminuria: *Protein Wnt-7a (WNT7A-rs6795744)*, *Transmembrane protein 60 (TMEM60-rs6465825)*, *APOL1 (G2)-rs71785313* with *P*-values of 0.03730, 0.02340, and 0.03803, respectively. With knowledge of the crude-albuminuria distribution as shown in **Figure 3**, a change in allele or WT deletion in the case of *APOL1* provides a protective effect against prevalent micro-albuminuria (**Table 3**). These results indicate that patients with *WNT7A-rs6795744* and *TMEM60-rs6465825* changes are approximately 2 times less likely to progress to micro-albuminuria state, and a single and double WT *APOL1* (G2) deletions decrease crude-albuminuria value by 0.12686 and 0.24732 ml/min/1.73/m^2, respectively. Though this phenotype level change is negligible due to the effect size, it can be essential for crude-albuminuria extreme values (e.g., values on the border line of state changes).

Distribution of HbF Levels vs. β-Globin Gene Cluster Haplotypes and α-Globin Gene Deletions

HbF is a major SCD modifier, which is known to modulate the SCD phenotype (Coleman and Inusa, 2007), to ameliorates pathophysiological and clinical manifestations of the sickling process (Adekile, 2020). There is accumulating evidence indicating that this major disease modifier is influenced by β-globin gene cluster haplotypes (Lakkakula et al., 2017; Piel et al., 2017; Adekile, 2020) and (α-globin gene deletions (Piel et al., 2017). Thus, we have looked at distribution of fetal haemoglobin levels vs. representative β-globin gene cluster haplotypes and α-globin gene deletions and results are shown in **Figure 4**.

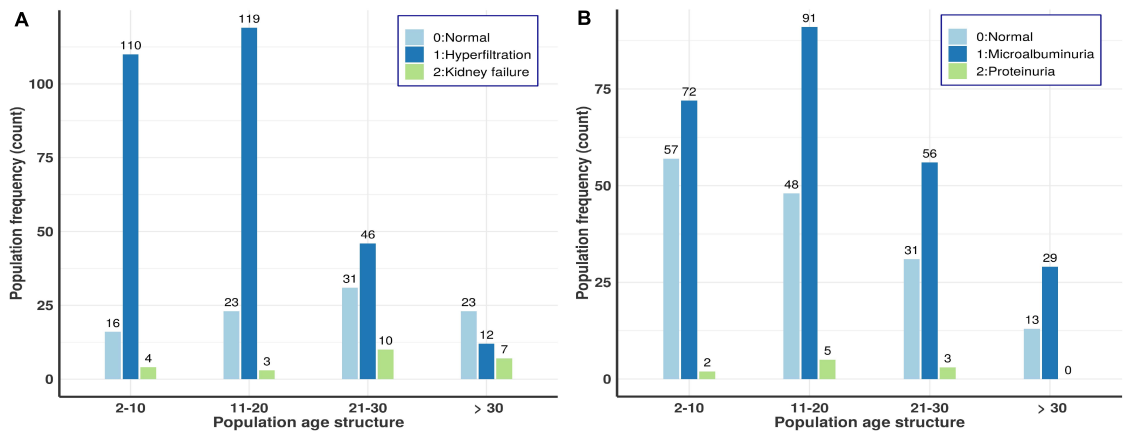


FIGURE 2 | Age-based population distributions for the two kidney dysfunction indicators, i.e., crude-albuminuria and eGFR scores. **(A)** Age-based population distribution using eGFR scores. **(B)** Age-based population distribution using microalbuminuria scores.

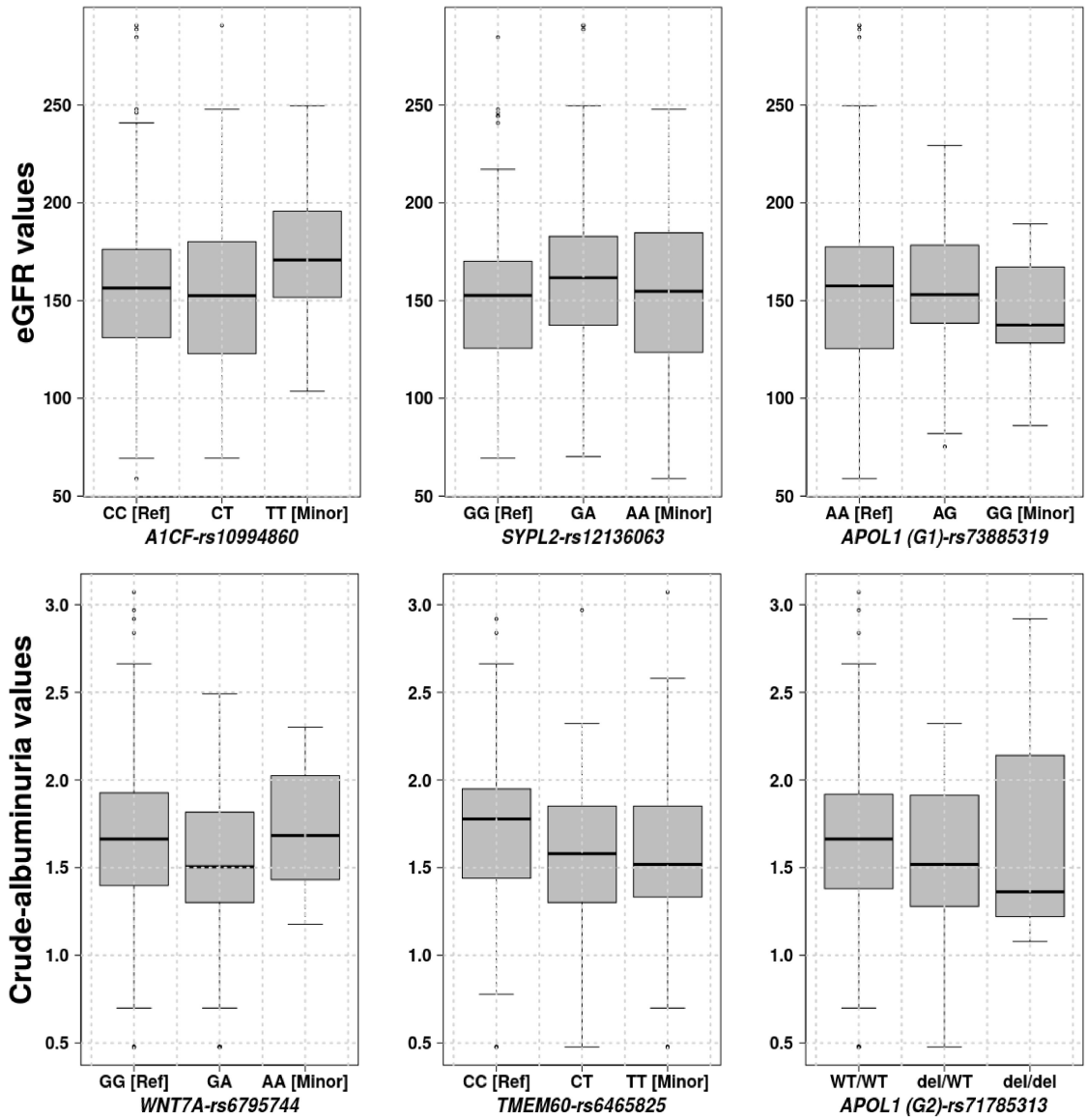


FIGURE 3 | Phenotype values, eGFR and crude-albuminuria values based on the significant gene variants distributed over homozygous dominant, recessive, and heterozygous genotypes.

These results indicate that HbF levels vary in the population based on β-globin gene cluster haplotypes and α-globin gene deletions. Individuals with Benin haplotyte and double 3.7 kb α-globin gene deletions having significantly higher HbF levels with P = 0.0813e–2 and 0.0203, respectively, accepting the alternative hypothesis that the HbF level median is greater than 9.3%, the lower bound of the HbF level confidence interval (**Table 1**), which is slightly higher than the minimal level (8.6%) indicated to improve SCD patient survival (Coleman and Inusa, 2007). Furthermore, we checked whether there is any association between exposure to Benin haplotype or double 3.7 kb α-globin gene deletions and outcomes, namely renal dysfunction phenotypes: eGFR and albuminurea. Results obtained have revealed an association between double 3.7 kb α-globin gene deletions and eGRF showing a protective effect against prevalent hyperfiltration/kidney failure with odds ratio = 0.36320 (95%CI: 0.16695–0.80532, P = 0.01007) under the null hypothesis that odds ratio is equal to 1.

Selecting Optimal Phenotype Proxy for Kidney Dysfunction Prediction

Using the eGFR- and crude-albuminuria-based logistic regression model, we compute areas under the receiver operating characteristic (ROC) and Precision-Recall (PR) curves, as well as accuracy to identify the best phenotype proxy for predicting kidney dysfunction in Cameroonian SCD patients. Different areas are shown in **Figure 5**, with eGFR-based model

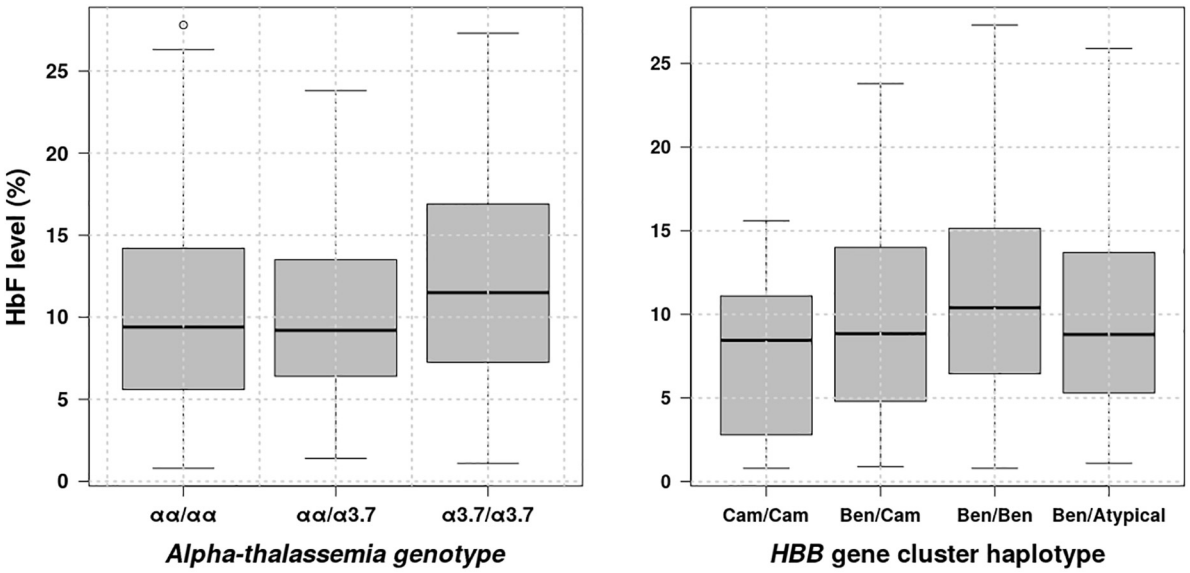


FIGURE 4 | Distribution of HbF levels vs. β-globin gene cluster haplotypes and α-globin gene deletions.

achieving the area under ROC of 0.76 and an accuracy score of 0.78 vs. the area under ROC of 0.73 with an accuracy score of 0.70 for crude-albuminuria-based model. This suggests that it is more effective to use eGFR phenotype as a proxy for predicting kidney dysfunction in Cameroonian SCD population. This is also in agreement with the Akaike's Information Criterion (AIC) scores produced by the two models, 284.33 for eGFR-based model vs. 361.32 for crude-albuminuria-based model, indicating that eGFR-based model fits data better than crude-albuminuria phenotype proxy. This suggests that a simple classification learning algorithm can be designed, taking as inputs, patient age, gender, and level of serum creatinine and predicting kidney dysfunction in patients.

Identifying Essential Genes and Functional Enrichment Analyses

We analysed how interactive genes from knowledge-based Protein-Protein Interaction (PPI) interacted with the 18 genes used in different analyses, focusing specifically on six genes identified to be associated with kidney dysfunction phenotypes. Mapping these 18 genes to a comprehensive human Protein-Protein Interaction (PPI) network (Wu et al., 2009; Mazandu et al., 2018), we identify sub-networks containing these gene variants: *APOL1, SYPL2, WNT7A*, *IGFBP5*, *UNCX*, *NFKB1*, *UMOD,* and *SKIL*. Three gene variants within this sub-network, namely APOL1 (G1)-rs73885319, as well as APOL1 (G2)-rs71785313, WNT7A-rs6795744, and SYPL2-rs12136063, have been identified to influence variation in renal dysfunction phenotypes in SCD patients. These variant genes are connected to *NFKB1*, identified to be essential or a hub based on the network centrality measures within the sub-network via some specific intermediate genes (**Figure 6** and **Table 4**), following a small world property of human PPI network (Mazandu et al., 2020). This *NFKB1* gene might indirectly influence extreme phenotype levels. Moreover, these gene variants are enriched with the *cartilage condensation process* (P = 0.02976) in which *WNT7A*, revealed to be likely implicated in SCD patient renal dysfunction, is involved. This process is possibly involved in the development of VOC complications, resulting in proteinuria and glomerular hyperfiltration, and ultimately in kidney damage.

DISCUSSION

This is the first study to investigate the relevance of kidney dysfunction-related variants identified through a GWAS meta-analysis as well as functional enrichment and protein-protein interaction network analyses in SCD patients. The results highlighted the high prevalence of micro-albuminuria as presented in a previous study in Cameroon by Geard in 2017 (Geard et al., 2017). This prevalence is much higher than the values of 18.5 and 27% observed in paediatric cohorts from several sub-Saharan African countries (Ranque et al., 2014; Aloni et al., 2017), the 13.2% in the multicentric study of children with SCD in the United States (Schaefer et al., 2016) and the 44% in adults from Nigerian and the United States (Bolarinwa et al., 2012; Drawz et al., 2016). These differences likely reflect the lack of appropriate care of SCD or the manifestation of the most severe SCD phenotype in Cameroon. The low proportion of macro-albuminuria found in this study is distinct from 15.1% reported in a cohort of SCD patients in the United Kingdom (UK) (Brewin et al., 2017). This could be due to the difference in age structure between the two cohorts. This study replicated a positive association of crude-albuminuria with increasing age as presented by a multi-center African study (Ranque et al., 2014) and a Nigeria-based study (Brewin et al., 2017). Some haematological variables, such as MCV and Hb level

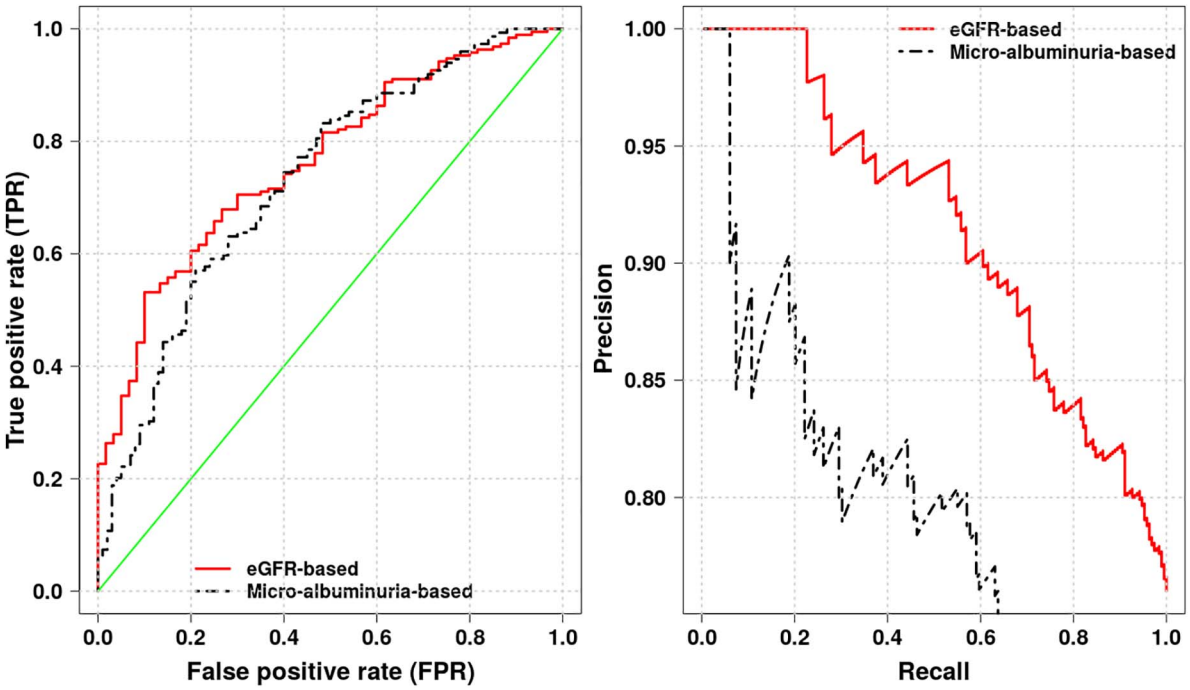


FIGURE 5 | Receiver operating characteristic (ROC) and Precision-Recall curves for eGFR- and crude-albuminuria-based logistic models.

(**Table 3**), influence crude albuminuria among SCD patients from Cameroon, Hb level observed to be protective against micro- albuminuria. This is in accordance with studies in Jamaica and the United States (Aban et al., 2017; Niss et al., 2020) which revealed that lower concentration of Hb is associated with development of micro-albuminuria, leading to relative renal ischemia, ischemia-reperfusion injury, and increased medullary sickling (Aban et al., 2017). BMI provides the highest mean variation explained for crude-albuminuria and may be a major anthropometric factor leading to renal dysfunction. This is likely related to the nutrition of SCD patients who struggle to maintain an adequate quality of life. We have not observed any significant association of crude-albuminuria with WBC counts or BP as previously observed in a Jamaican cohort (Asnani et al., 2016).

The prevalence of glomerular hyperfiltration was similar to the 76% found among SCD children in the United States (Aygun et al., 2011), but higher than that previously reported in France (51%) (2010) (Haymann et al., 2010) and DRC (2017) (Aloni et al., 2017). It was lower than the 98% presented in the United Kingdom (Drawz et al., 2016). These differences may be explained by the variability in the median ages of participants in the other studies; they may also be due to the method used to calculate eGFR, as the Schwartz formula tends to underestimate GFR in children compared to the CKD-EPI used in the present study. Unlike the studies of Asnani in Jamaica (Aban et al., 2017) and Vazquez in the United States (Aygun et al., 2011), no significant associations between eGFR and crude albuminuria were identified in this study. However, 41 and 2% of patients had glomerular hyperfiltration associated with micro-albuminuria and proteinuria, respectively. These results are different from the values highlighted by Haymann et al. (2010) and from the 22% of patients who had both glomerular hyperfiltration and micro-albuminuria in DRC (Aloni et al., 2017). The eGFR was highly associated with haemoglobin, HbF and SBP, as previously observed in patients from the United Kingdom (Wu et al., 2009). The strong association between SBP and eGFR is also consistent with data reported in the study by Niss et al. (2020) in Jamaica confirming the highly protective role of stable SBP against glomerular hyperfiltration (**Table 3**). VOC frequencies are also highly correlated to eGFR, explaining the morbidity due to renal dysfunction among SCD patients in Cameroon. Associations between eGFR with Hb level and WBC (granulocytes and lymphocytes) were previously highlighted by Aban et al. (2017), revealing that a low Hb level and increased WBC are associated with renal failure. The correlations between eGFR with age, BMI and creatinine were observed in patients from the United Kingdom (Drawz et al., 2016), Jamaica (Eke et al., 2012), Nigeria (Ajite et al., 2019), and the United States (Becker et al., 2014). However, age and creatinine were not confirmed by this study as they were confounding factors, mixing up with eGFR. This suggests that previous observations were biased, and results obtained by these studies may be flawed. Our analysis agreed with publications within the Taiwanese (Chang et al., 2018) and Caucasian (Brown et al., 2012) which found no association between BMI and renal dysfunctions.

Results highlighted that only three replicated genetic variants are associated to renal dysfunction: *A1CF-rs10994860* with reduced risk of renal dysfunction for two allele changes and a protective effect against prevalent glomerular hyperfiltration; It confers about 2 times more protection to SCD individuals

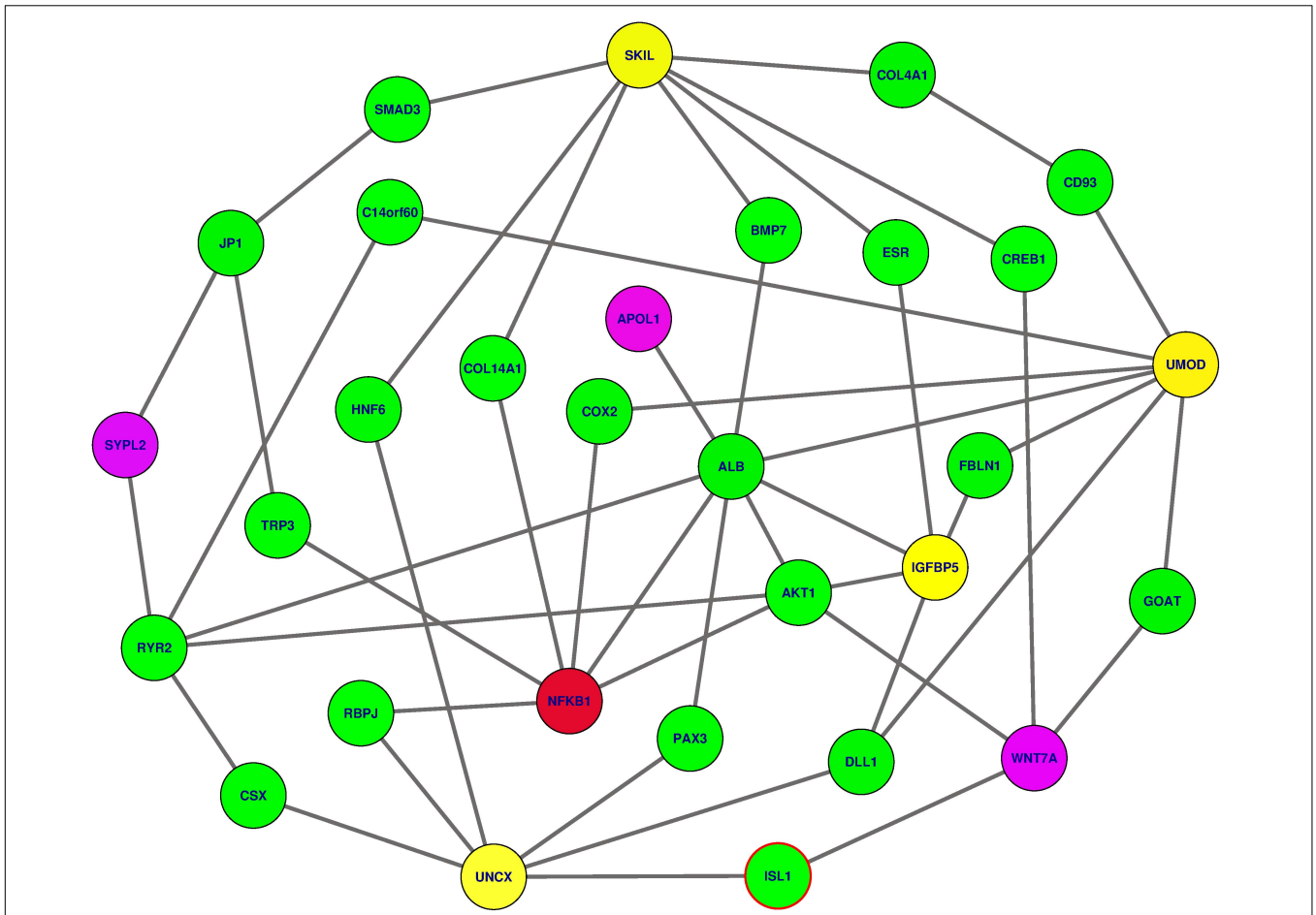


FIGURE 6 | The subnetwork extracted from the human PPI network revealing how predicted gene variants interact together to influence the kidney dysfunction (refer to **Table 4** for gene descriptions). In yellow and magenda are gene variants previously shown to be associated with kidney-dysfunction, three in magenda color have been confirmed. In green color, are intermediate nodes used by kidney-dysfunction gene variants to reach the one in red color (*NFKB1*) indicated to be essential in the PPI network.

with one allele change or heterozygous genotype as compared to a patient with no copy, and explaining only 0.18% of status change to the glomerular hyperfiltration or renal dysfunction. However, *SYPL2-rs12136063* as well as *APOL1 (G1)-rs73885319* conferred increased risk of progressing to the renal dysfunction or to prevalent glomerular hyperfiltration, explaining 11.04 and 0.97% of status changes, respectively. Like eGFR, three gene variants are also identified for crude-albuminuria: *WNT7A-rs6795744, TMEM60-rs6465825,* and *APOL1 (G2)-rs71785313* providing a protective effect against prevalent micro-albuminuria (**Table 3**), explaining less than 4% of status changes. These relative low contributions to status changes suggests that the targeted SNPs may not be relevant to the ancestral African populations, or it is possible SCD specific kidney dysfunctions associated variants are still to be found. Indeed, there is bias in polygenic risk scores (PRSs) regarding usability, and transferability for complex trait, as most PRSs do not account for multiple alleles that are either limited or of high frequency among Africans, due highest genomic variations (Gurdasani et al., 2019). A genome-wide association study (GWAS) on genetic susceptibility to type 2 diabetes (T2D) identified a previously unreported African-specific significant locus, while showing transferability of 32 established T2D loci (Adeyemo et al., 2019). Alternatively, kidney dysfunction in SCD may be mostly driven by the pathophysiology of the disease itself rather than genetic factors. For eGFR, SBP explained this status changes by nearly 40.55%, followed by the number of VOC (21.73%), Hb (16.55%), and the upper out range level of HbF would explain the status changes for nearly 1.13%. The HbF level distribution results indicate that HbF levels vary with β-globin gene cluster haplotypes, as well as with α-globin gene deletions in agreement with the current knowledge. The HbF level in patients with Benin haplotyte and double 3.7 kb α-globin gene deletions is significantly higher as compared to the HbF level cutoff observed to improve the patient survival. Thus, these two genetic events may confer a relatively favorable clinical manifestation (Steinberg, 2009; Bitoungui et al., 2015) and patients with double 3.7 kb α-globin gene deletions are about 3 times less likely to progress to the glomerular hyperfiltration or kidney failure state as compared to a patient with no or one

TABLE 4 | The description of the different genes displayed in **Figure 6** extracted from the UniProt database (https://www.uniprot.org/).

Gene	UniProt description
COL4A1	Collagen alpha-1 (IV) chain (Cleaved into: Arresten)
ISL1	Insulin gene enhancer protein ISL-1 (Islet-1)
CD93	Complement component C1q receptor (C1q/MBL/SPA receptor) (C1qR) (C1qR(p)) (C1qRp) (CDw93) (Complement component 1 q subcomponent receptor 1) (Matrix-remodeling-associated protein 4) (CD antigen CD93)
SMAD3	Mothers against decapentaplegic homolog 3 (MAD homolog 3) (Mad3) (Mothers against DPP homolog 3) (hMAD-3) (JV15-2) (SMAD family member 3) (SMAD 3) (Smad3) (hSMAD3)
AKT1	RAC-alpha serine/threonine-protein kinase (EC 2.7.11.1) (Protein kinase B) (PKB) (Protein kinase B alpha) (PKB alpha) (Proto-oncogene c-Akt) (RAC-PK-alpha)
JP1	Junctophilin-1 (JP-1) (Junctophilin type 1)
FBLN1	Fibulin-1 (FIBL-1)
COL14A1	Collagen alpha-1 (XIV) chain (Undulin)
COX2	Prostaglandin G/H synthase 2 (EC 1.14.99.1) (Cyclooxygenase-2) (COX-2) (PHS II) (Prostaglandin H2 synthase 2) (PGH synthase 2) (PGHS-2) (Prostaglandin-endoperoxide synthase 2)
RBPJ	Recombining binding protein suppressor of hairless (CBF-1) (J kappa-recombination signal-binding protein) (RBP-J kappa) (RBP-J) (RBP-JK) (Renal carcinoma antigen NY-REN-30)
RYR2	Ryanodine receptor 2 (RYR-2) (RyR2) (hRYR-2) (Cardiac muscle ryanodine receptor) (Cardiac muscle ryanodine receptor-calcium release channel) (Type 2 ryanodine receptor)
UNCX	Homeobox protein unc-4 homolog (Homeobox protein Uncx4.1)
C14orf60	Neurexin-3 (Neurexin III-alpha) (Neurexin-3-alpha)
NFKB1	Nuclear factor NF-kappa-B p105 subunit (DNA-binding factor KBF1) (EBP-1) (Nuclear factor of kappa light polypeptide gene enhancer in B-cells 1) (Cleaved into: Nuclear factor NF-kappa-B p50 subunit)
HNF6	Hepatocyte nuclear factor 6 (HNF-6) (One cut domain family member 1) (One cut homeobox 1)
ESR	Estrogen receptor (ER) (ER-alpha) (Estradiol receptor) (Nuclear receptor subfamily 3 group A member 1)
UMOD	Uromodulin (Tamm-Horsfall urinary glycoprotein) (THP) (Cleaved into: Uromodulin, secreted form)
CREB1	Cyclic AMP-responsive element-binding protein 1 (CREB-1) (cAMP-responsive element-binding protein 1)
DLL1	Delta-like protein 1 (Drosophila Delta homolog 1) (Delta1) (H-Delta-1)
SKIL	Ski-like protein (Ski-related oncogene) (Ski-related protein)
PAX3	Paired box protein Pax-3 (HuP2)
IGFBP5	Insulin-like growth factor-binding protein 5 (IBP-5) (IGF-binding protein 5) (IGFBP-5)
BMP7	Bone morphogenetic protein 7 (BMP-7) (Osteogenic protein 1) (OP-1) (Eptotermin alfa)
SYPL2	Synaptophysin-like protein 2
ALB	Fas-binding factor 1 (FBF-1) (Protein albatross)
TRP3	Short transient receptor potential channel 3 (TrpC3) (Transient receptor protein 3) (TRP-3) (hTrp-3) (hTrp3)
APOL1	Apolipoprotein L1 (Apolipoprotein L) (Apo-L) (ApoL) (Apolipoprotein L-I) (ApoL-I)
WNT7A	Protein Wnt-7a
GOAT	Ghrelin O-acyltransferase (EC 2.3.1.-) (Membrane-bound O-acyltransferase domain-containing protein 4) (O-acyltransferase domain-containing protein 4)
CSX	Homeobox protein Nkx-2.5 (Cardiac-specific homeobox) (Homeobox protein CSX) (Homeobox protein NK-2 homolog E)

deletion (odds ratio = 0.36320). For crude-albuminuria, the status change is mainly explained by BMI with approximately 45.34% of variations, followed by MCV with 1.22%, transfusion (0.23%), and Hb (0.19%).

LIMITATIONS

The first limitation of the present study is the cross-sectional design. A longitudinal study would give more precise data on kidney dysfunction in SCD in Cameroon. Another limitation is the use of the CKD-EPI-creatinine equation to estimate GFR. Recent reports indicate that the CKD-EPI original formula overestimates GFR values. The reliability of the CKD-EPI equation was recently adjusted by inclusion of a molecular weight protein, Cystatin C (CysC), which is eliminated exclusively by glomerular filtration (Yee et al., 2017). However, the protein was not quantified for this cohort. The imbalanced distribution of individuals without kidney dysfunction in this group of SCD patients likely affects the performance of the different regression models (Jedrzejowicz et al., 2018), tending to be biased toward the normal ranges (KrishnaVeni and Sobha, 2011) and potentially failing to identify possible signals.

CONCLUSION

This study has replicated *APOL1* gene variants: *(G1)-rs73885319* and *(G2)-rs71785313*, shown to be strongly associated with renal dysfunction in SCD patients, as well as *A1CF-rs10994860*, *SYPL2-rs12136063*, *WNT7A-rs6795744,* and *TMEM60-rs6465825* in Cameroonian SCD patients. Though the protein-protein

interaction network and enrichment analyses have revealed the sub-network, which may influence extreme phenotype levels, enriched with the *cartilage condensation process*, which likely contributes to the development of VOC complications and possibly to the renal dysfunction, these gene variants only explain a small proportion of status changes. The results also suggest that haematological indices, clinical events, anthropometric and socio-demographic variables, are major contributors to the pathophysiology of kidney dysfunction in SCD. This elicits the need for further research to investigate new genetic biomarkers which account for kidney dysfunction risk factors in SCD patients in the African setting.

DATA AVAILABILITY STATEMENT

The data supporting the findings of this study are available from the corresponding author upon reasonable request.

ETHICS STATEMENT

The studies involving human participants were reviewed and the study was performed following the Declaration of Helsinki. This study was approved by the Faculty of Health Sciences Human Research Ethics Committee of the University of Cape Town, South Africa (HREC REF: 661/2015), and the National Ethics Committee of the Ministry of Public Health, Yaoundé, Republic of Cameroon (No. 193/CNE/SG/10). Patients older than 18 years self-consented into the study and informed consent was given by the parents or guardians for participants younger than 18 years old with a requirement for children older than 7 years to also sign assent forms. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

AW, VN-B, KM, TM, and GA conceived and designed the experiments. VN-B, SB, KM, and TM performed the experiments. VN-B, SB, RE, FN-S, TM, KM, and AW needed patient recruitment, samples, and clinical data collection. GM, VN-B, VN, KM, and AW processed and analysed the data. AW, GA, GM, and VN contributed to reagents, materials, and analysis tools. VN-B, KM, GM, and AW wrote the manuscript. GM, VN-B, SB, KM, TM, RE, VN, FN-S, GA, and AW revised and approved the manuscript. All the authors approved the manuscript and agreed to be accountable for all aspects of the presented work.

FUNDING

The data collection, molecular experiments and analysis of the study were supported by following Awards: Welcome Trust, Developing Excellence in Leadership, Training and Science (DELTAS) Africa, Award 107755Z/15/Z to GA, and AW; the NIH/H3Africa, U54HG009790 and U01HG009716 to AW, and NIH/NHLBI U24HL135600 to AW. VN-B was supported by a WACCBIP DELTAS postdoctoral fellowship (DEL-15-007: Awandare). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.595702/full#supplementary-material

REFERENCES

Aban, I., Baddam, S., Hilliard, L. M., Howard, T. H., Feig, D. I., and Lebensburger, J. D. (2017). Severe anemia early in life as a risk factor for sickle-cell kidney disease. *Blood* 129, 385–387. doi: 10.1182/blood-2016-09-738104

Adekile, A. (2020). The genetic and clinical significance of fetal hemoglobin expression in sickle cell disease. *Med. Princ. Pract.* doi: 10.1159/000511342 [Epub ahead of print].

Adeyemo, A. A., Zaghloul, N. A., Chen, G., Doumatey, A. P., Leitch, C. C., Hostelley, T. L., et al. (2019). ZRANB3 is an African-specific type 2 diabetes locus associated with beta-cell mass and insulin response. *Nat. Commun.* 10, 3195.

Ajite, A., Ogundare, E., Oluwayemi, O., Olatunya, O., Oke, O., Tolorunju, K., et al. (2019). The pattern of blood pressure and renal function among children with Sickle cell anaemia presenting in a tertiary health institution in Nigeria. *J. Clin. Nephrol.* 3, 083–092. doi: 10.29328/journal.jcn.1001031

Aloni, M. N., Mabidi, J.-L. L., Ngiyulu, R. M., Ekulu, P. M., Mbutiwi, F. I., Makulo, J. R., et al. (2017). Prevalence and determinants of microalbuminuria in children suffering from sickle cell anemia in steady state. *Clinical Kidney Journal* 10, 479–486. doi: 10.1093/ckj/sfx058

Anto, E. O., Obirikorang, C., Acheampong, E., Adua, E., Donkor, S., Afranie, B. O., et al. (2019). Renal abnormalities among children with sickle cell conditions in highly resource-limited setting in Ghana. *PLoS One.* 14:e0225310. doi: 10.1371/journal.pone.0225310

Asnani, M., Serjeant, G., Royal-Thomas, T., and Reid, M. (2016). Predictors of renal function progression in adults with homozygous sickle cell disease. *British journal of haematology* 173, 461–468. doi: 10.1111/bjh.13967

Ataga, K. I., Derebail, V. K., and Archer, D. R. (2014). The glomerulopathy of sickle cell disease. *American journal of hematology* 89, 907–914.

Audard, V., Bartolucci, P., and Stehle, T. (2017). Sickle cell disease and albuminuria: recent advances in our understanding of sickle cell nephropathy. *Journal of Clinical Nephrology* 10, 475–478. doi: 10.1093/ckj/sfx027

Aygun, B., Mortier, N. A., Smeltzer, M. P., Hankins, J. S., and Ware, R. E. (2011). Glomerular hyperfiltration and albuminuria in children with sickle cell anemia. *Pediatric nephrology* 26, 1285–1290. doi: 10.1007/s00467-011-1857-2

Becker, A. M., Goldberg, J. H., Henson, M., Ahn, C., Tong, L., Baum, M., et al. (2014). Blood pressure abnormalities in children with sickle cell anemia. *Pediatric blood & cancer* 61, 518–522. doi: 10.1002/pbc.24843

Bitoungui, V. J., Pule, G. D., Hanchard, N., Ngogang, J., and Wonkam, A. (2015). Beta-globin gene haplotypes among cameroonians and review of the global distribution: is there a case for a single sickle mutation origin in Africa? *Omics* 19, 171–179. doi: 10.1089/omi.2014.0134

Bolarinwa, R. A., Akinlade, K. S., Kuti, M. A., Olawale, O. O., and Akinola, N. O. (2012). Renal disease in adult Nigerians with sickle cell anemia: a report of prevalence clinical features and risk factors. *Saudi journal of kidney diseases and transplantation* 23, 171–175.

Brewin, J., Tewari, S., Hannemann, A., Al-Balushi, H., Sharpe, C., Gibson, J. S., et al. (2017). Early markers of sickle nephropathy in children with sickle cell anemia are associated with red cell cation transport activity. *HemaSphere* 1:11e2.

Brown, R. N. K. L., Mohsen, A., Green, D., Green, D., Hoefield, R. A., Summers, L. K. M., et al. (2012). Body mass index has no effect on rate of progression of chronic kidney disease in non-diabetic subjects. *Nephrol. Dial. Transplant.* 27, 2776–2780. doi: 10.1093/ndt/gfr757

Chambers, J. C., Zhang, W., Lord, G. M., Van der Harst, P., Lawlor, D. A., Sehmi, J. S., et al. (2010). Genetic loci influencing kidney function and chronic kidney disease. *Nat. Genet.* 42, 373–375.

Chang, T. J., Zheng, C. M., and Wu, M. Y. (2018). Relationship between body mass index and renal function deterioration among the Taiwanese chronic kidney disease population. *Sci. Rep.* 8:6908.

Coleman, E., and Inusa, B. (2007). Sickle cell anemia: targeting the role of fetal hemoglobin in therapy. *Clin. Pediatr. (Phila.)* 46, 386–391. doi: 10.1177/0009922806297751

Crawford, D. C., Caggana, M., Harris, K. B., Lorey, F., Nash, C., Pass, K. A., et al. (2002). Characterization of beta-globin haplotypes using blood spots from a population-based cohort of newborns with homozygous HbS. *Genet. Med.* 4, 328–335. doi: 10.1097/00125817-200209000-00003

Drawz, P., Ayyappan, S., Nouraie, M., Saraf, S., Gordeuk, V., Hostetter, T., et al. (2016). Kidney disease among patients with sickle cell disease haemoglobin SS and SC. *Clin. J. Am. Soc. Nephrol.* 11, 207–215. doi: 10.2215/cjn.03940415

Eke, C. B., Okafor, H. U., and Ibe, B. C. (2012). Prevalence and correlates of microalbuminuria in children with sickle cell anaemia: experience in a tertiary health facility in Enugu Nigeria. *Int. J. Nephrol.* 2021:240173.

Geard, A., Pule, G. D., Chetcha-Chemegni, B., Ngo Bitoungui, V. J., Kengne, A. P., Chimusa, E. R., et al. (2017). Clinical and genetic predictors of renal dysfunctions in sickle cell anaemia in Cameroon. *Br. J. Haematol.* 178, 629–639. doi: 10.1111/bjh.14724

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

Gladwin, M. T. (2017). Cardiovascular complications in patients with sickle cell disease. *Am. Soc. Hematol. Educ. Program* 2017, 423–430. doi: 10.1182/asheducation-2017.1.423

Gosmanova, E. O., Zaidi, S., Wan, J. Y., and Adams-Graves, P. E. (2014). Prevalence and progression of chronic kidney disease in adult patients with sickle cell disease. *J. Investig. Med.* 62, 804–807. doi: 10.1097/01.jim.0000446836.75352.72

Gurdasani, D., Barroso, I., Zeggini, E., and Manjinder, S. S. (2019). Genomics of disease risk in globally diverse populations. *Nat. Rev. Genet.* 20, 520–535. doi: 10.1038/s41576-019-0144-0

Hanchard, N., Elzein, A., Trafford, C., Rockett, K., Pinder, M., Jallow, M., et al. (2007). Classical sickle beta-globin haplotypes exhibit a high degree of long-range haplotype similarity in African and Afro-Caribbean populations. *BMC Genet* 8:52.

Haymann, J. P., Stankovic, K., Levy, P., Avellino, V., Tharaux, P. L., Letavernier, E., et al. (2010). Glomerular hyperfiltration in adult sickle cell anemia: a frequent hemolysis associated feature. *Clin. J. Am. Soc. Nephrol.* 5, 756–761. doi: 10.2215/cjn.08511109

Jedrzejowicz, J., Kostrzewski, R., Neumann, J., and Zakrzewska, M. (2018). Imbalanced data classification using MapReduce and relief. *J. Inform. Telecommun.* 2, 217–230. doi: 10.1080/24751839.2018.1440454

Kanehisa, M., Sato, Y., Furumichi, M., Morishima, K., and Tanabe, M. (2019). New approach for understanding genome variations in KEGG. *Nucleic Acids Res.* 47, D590—-D595.

Kottgen, 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.

Kottgen, A., Pattaro, C., Boger, 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.

KrishnaVeni, C. V., and Sobha, R. T. (2011). On the classification of imbalanced datasets international. *J. Comput. Sci. Technol.* 2, 145–148.

Lakkakula, B. V. K. S., Verma, H. K., Choubey, M., Patra, S., Khodiar, P. K., and Patra, P. K. (2017). Assessment of renal function in Indian patients with sickle cell disease. *Saudi J. Kidney Dis. Transpl.* 28, 524–531. doi: 10.4103/1319-2442.206440

Mammen, C., Bissonnette, M. L., and Matsell, D. G. (2017). Acute kidney injury in children with sickle cell disease—compounding a chronic problem. *Pediatr. Nephrol.* 32, 1287–1291. doi: 10.1007/s00467-017-3650-3

Mazandu, G. K., Chimusa, E. R., Rutherford, K., Zekeng, E. G., Gebremariam, Z. Z., Onifade, M. Y., et al. (2018). Large-scale data-driven integrative framework for extracting essential targets and processes from disease-associated gene data sets. *Brief. Bioinform.* 19, 1141–1152.

Mazandu, G. K., Hooper, C., Opap, K., Makinde, F., Nembaware, V., Thomford, N. E., et al. (2020). IHP-PING—generating integrated human protein–protein interaction networks on-the-fly. *Brief. Bioinform.* bbaa277. doi: 10.1093/bib/bbaa277 [Epub ahead of print].

Mi, H., Muruganujan, A., Ebert, D., Huang, X., and Thomas, P. D. (2019). PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. *Nucleic Acids Res.* 47, D41–D426.

Nath, K. A., and Hebbel, R. P. (2015). Sickle cell disease: renal manifestations and mechanisms. *Nat. Rev. Nephrol.* 11, 161–171. doi: 10.1038/nrneph.2015.8

Niss, O., Lane, A., Asnani, M. R., Yee, M. E., Raj, A., Creary, S., et al. (2020). Progression of albuminuria in patients with sickle cell anemia: a multicenter, longitudinal study. *Blood Adv.* 4, 1501–1511. doi: 10.1182/bloodadvances.2019001378

Pattaro, C., Kottgen, A., Teumer, A., Garnaas, M., Boger, C. A., Fuchsberger, C., et al. (2012). Genome-wide association and functional follow-up reveals new loci for kidney function. *PLoS Genet.* 8:e1002584.

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.* 7:10023.

Piel, F. B., Patil, A. P., Howes, R. E., Nyangiri, O. A., Gething, P. W., Dewi, M., et al. (2013). Global epidemiology of sickle haemoglobin in neonates: a contemporary geostatistical model-based map and population estimates. *Lancet* 381, 142–151. doi: 10.1016/s0140-6736(12)61229-x

Piel, F. B., Steinberg, M. H., and Rees, D. C. (2017). Sickle cell disease. *N Engl J Med.* 376, 1561–1573.

Platt, O. S., Brambilla, D. J., Rosse, W. F., Milner, P. F., Castro, O., Steinberg, M. H., et al. (1994). Mortality in sickle cell disease life expectancy and risk factors for early death. *N. Engl. J. Med.* 330, 1639–1644. doi: 10.1056/nejm199406093302303

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., et al. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81, 559–575. doi: 10.1086/519795

Ranque, B., Menet, A., Diop, I. B., Thiam, M. M., Diallo, D., Diop, S., et al. (2014). Early renal damage in patients with sickle cell disease in sub-Saharan Africa: a multinational prospective cross-sectional study. *Lancet Haematol.* 1, e64–e73.

Rumaney, M. B., Ngo Bitoungui, V. J., Vorster, A. A., Ramesar, R., Kengne, A. P., Ngogang, J., et al. (2014). The co-inheritance of alpha-thalassemia and sickle cell anemia is associated with better hematological indices and lower consultations rate in Cameroonian patients and could improve their survival. *PLoS One* 9:e100516. doi: 10.1371/journal.pone.0100516

Saiki, R. K., Scharf, S., Faloona, F., Mullis, K. B., Horn, G. T., Erlich, H. A., et al. (1985). Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science* 230, 1350—-1354.

Saraf, S. L., Shah, B. N., Zhang, X., Han, J., Tayo, B. O., Abbasi, T., et al. (2017). APOL1 alpha-thalassemia and BCL11A variants as a genetic risk profile for progression of chronic kidney disease in sickle cell anemia. *Haematologica* 102, e1–e6.

Saraf, S. L., Zhang, X., Shah, B., Kanias, T., Gudehithlu, K. P., Kittles, R., et al. (2015). Genetic variants and cell-free haemoglobin processing in sickle cell nephropathy. *Haematologica* 100, 1275–1284. doi: 10.3324/haematol.2015.124875

Schaefer, B. A., Flanagan, J. M., Alvarez, O. A., Nelson, S. C., Aygun, B., Nottage, K. A., et al. (2016). Genetic modifiers of white blood cell count albuminuria and glomerular filtration rate in children with sickle cell anemia. *PLoS One* 11:e0164364. doi: 10.1371/journal.pone.0164364

Steinberg, M. H. (2009). Genetic etiologies for phenotypic diversity in sickle cell anemia. *Sci. World J.* 9, 46–67. doi: 10.1100/tsw.2009.10

The Gene Ontology Consortium (2019). The gene ontology resource: 20 years and still GOing strong. *Nucleic Acids Res.* 47, D330–D338.

Tzur, S., Rosset, S., Shemer, R., Selig, S., Tarekegn, A., Bekele, E., 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

Weatherall, D. J., and Clegg, J. B. (2001). Inherited haemoglobin disorders: an increasing global health problem. *Bul. World Health Organ.* 79, 704–712.

Wonkam, A., Ngo Bitoungui, V. J., Vorster, A. A., Ramesar, R., Cooper, R. S., Tayo, B., et al. (2014). Association of variants at BCL11A and HBS1L-MYB with haemoglobin F and hospitalization rates among sickle cell patients in Cameroon. *PLoS One* 9:e92506. doi: 10.1371/journal.pone.0092506

Wu, J., Vallenius, T., Ovaska, K., Westermarck, J., Mäkelä, T. P., and Hautaniemi, S. (2009). Integrated network analysis platform for protein-protein interactions. *Nat. Methods* 6, 75–77. doi: 10.1038/nmeth.1282

Yates, A. D., Achuthan, P., Akanni, W., Allen, J., Allen, J., Alvarez-Jarreta, J., et al. (2020). Ensembl 2020. *Nucleic Acids Res.* 48, D682–D688.

Yee, M. E. M., Lane, P. A., Archer, D. R., Joiner, C. H., Eckman, J. R., and Guasch, A. (2017). Estimation of glomerular filtration rate using serum cystatin C and creatinine in adults with sickle cell anemia. *Am. J. Hematol.* 92, E598–E599.

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 in Medicine

HYPOTHESIS AND THEORY
published: 22 March 2021
doi: 10.3389/fmed.2021.647412

Sudden Cardiac Death—A New Insight Into Potentially Fatal Genetic Markers

Dragan Primorac[1,2,3,4,5,6,7,8,9], Ljubica Odak[1,10], Vitorio Perić[1], Jasmina Ćatić[1,11], Jozica Šikić[12], Vjekoslav Radeljić[13], Šime Manola[13], Robert Nussbaum[14], Matteo Vatta[14], Swaroop Aradhya[14], Tanja Sofrenović[14], Vid Matišić[1], Vilim Molnar[1], Andrea Skelin[1], Jure Mirat[6] and Johannes Brachmann[4,8]*

[1] St. Catherine Specialty Hospital, Zagreb, Croatia, [2] Eberly College of Science, The Pennsylvania State University, University Park, State College, Philadelphia, PA, United States, [3] The Henry C. Lee College of Criminal Justice and Forensic Sciences, University of New Haven, West Haven, CT, United States, [4] Medical School, University of Split, Split, Croatia, [5] Faculty of Dental Medicine and Health, Josip Juraj Strossmayer University of Osijek, Osijek, Croatia, [6] Faculty of Medicine, Josip Juraj Strossmayer University of Osijek, Osijek, Croatia, [7] Medical School, University of Rijeka, Rijeka, Croatia, [8] Medical School REGIOMED, Coburg, Germany, [9] Medical School, University of Mostar, Mostar, Bosnia and Herzegovina, [10] Children's Hospital Zagreb, Zagreb, Croatia, [11] Department of Cardiology, Clinical Hospital Dubrava, Zagreb, Croatia, [12] Department of Cardiology, Clinical Hospital Sveti Duh, Zagreb, Croatia, [13] Department of Cardiology, Clinical Hospital Center Sestre Milosrdnice, Zagreb, Croatia, [14] Invitae, San Francisco, CA, United States

OPEN ACCESS

Edited by:
Maria G. Stathopoulou, Institut National de la Santé et de la Recherche Médicale (INSERM), France

Reviewed by:
Oscar Campuzano, University of Girona, Spain
Laura Yeates, Royal Prince Alfred Hospital, Australia

***Correspondence:**
Dragan Primorac
draganprimorac2@gmail.com

Specialty section:
This article was submitted to Precision Medicine, a section of the journal Frontiers in Medicine

Received: *29 December 2020*
Accepted: *01 March 2021*
Published: *22 March 2021*

Citation:
Primorac D, Odak L, Perić V, Ćatić J, Šikić J, Radeljić V, Manola Š, Nussbaum R, Vatta M, Aradhya S, Sofrenović T, Matišić V, Molnar V, Skelin A, Mirat J and Brachmann J (2021) Sudden Cardiac Death—A New Insight Into Potentially Fatal Genetic Markers. Front. Med. 8:647412. doi: 10.3389/fmed.2021.647412

Sudden cardiac death (SCD) is an unexpected and dramatic event. It draws special attention especially in young, seemingly healthy athletes. Our scientific paper is based on the death of a young, 23-year-old professional footballer, who died on the football field after a two-year history of cardiac symptoms. In this study we analyzed clinical, ECG and laboratory data, as well as results of genetic testing analysis in family members. To elucidate potential genetic etiology of SCD in this family, our analysis included 294 genes related to various cardiac conditions.

Keywords: sudden cardiac death, genes, coronary artery disease, cardiomyopathy, professional athletes

INTRODUCTION

Definition of Sudden Cardiac Death

Sudden cardiac death (SCD) is an unforeseen, unexpected heart function loss that occurs within 1 h of symptom onset or within 24 h of last being seen alive (1). SCD may occur after a long history of cardiac disease, but it can also be the first manifestation of genetic heart disease in a seemingly healthy person (1, 2). The cause of death is either cessation of heartbeat or irregular heart rhythm, both resulting in decreased tissue perfusion.

SCD in the Population: Epidemiology and Etiology

SCD presents a significant cause of mortality, including ∼20% of all deaths in Western societies (3). It is estimated that 5 million cases of SCD occur worldwide each year. Also, 50% of all deaths associated with cardiovascular disease are caused by SCD (2). Although SCD rates range from 50–100 deaths per 100,000 people in the general population, there are differences in SCD rates comparing different age groups (2). According to a study conducted in Denmark, adults aged 35–49 years have 9.4 times the risk of dying from SCD compared to younger populations aged 1–35 years (4). Coronary artery disease (CAD) is responsible for 70-75% of all SCD cases in the population over the age of 35 (2, 5). In people aged < 35 years, the rate of SCD is approximately 1–2 per 100,000 people in the general population (2, 6). The etiology of SCD varies and is

highly dependent on the demographic characteristics of the patients included in the studies (7). A large number of SCDs at a younger age (<35 years old) have an unexplained etiology and potential underlying genetic basis (1, 8). Furthermore, the etiology of SCD in this group includes hypertrophic cardiomyopathy (HCM), dilated cardiomyopathy (DCM), arrhythmogenic right ventricular cardiomyopathy (ARVC), myocarditis, and different kinds of arrhythmogenic disorders including Brugada syndrome (BrS), congenital long-QT syndrome (LQTS), and catecholaminergic polymorphic ventricular tachycardia (CPVT) (6, 9).

SCD is a particularly traumatic event in seemingly healthy young athletes. Such deaths have a strong impact on the public because being an athlete is considered a healthy habit that prolongs and improves the quality of life.

CAD is the leading cause of SCD among athletes older than 35 years (10). Screening in that age group is much easier because CAD develops gradually throughout one's lifetime and is manifested by specific cardiac symptoms that alert the athlete. The situation is completely different among athletes under the age of 35 where most SCDs are caused by underlying heart disease triggered by intense exertion (8, 11).

Studies conducted in France (12) and the USA (11, 13) showed that young adults who are engaged in sports activities had 3.7–4.5 times the risk of SCD compared to age-matched young adults who were not athletes. Moreover, according to a study conducted in Italy (14), young adults who were not athletes had 0.41 times the rate of SCD compared to age-matched young adults who were athletes.

The cause of SCD is often associated with inherited or congenitally acquired mechanisms that provide the basis for the onset of malignant ventricular arrhythmia (15). HCM and ARVC are the most common causes of SCD in this age group (16, 17).

Other heart diseases that may also lead to SCD include myocarditis, dilated cardiomyopathy, congenital coronary anomalies, mitral valve prolapse, premature atherosclerotic coronary artery disease, WPW syndrome, etc. (17, 18). In a large number of SCD caused deaths, the autopsy findings show a normal cardiac structure, without any anomaly that would indicate pre-existing heart disease. In such cases, the cause of SCD is most commonly related to inherited cardiac ion-channel defects (channelopathies) which are responsible for the electrical activity of the heart (17, 19, 20).

Research Aim

Our investigation sought to identify genetic variants in a family sample in which two individuals under 25 years-of-age died from SCD. The purpose of researching the genetic variants in this family lies in the fact that the brother (IV:4) and cousin (IV:2) of the proband (IV:3) play professional football, but is also important in first-degree relatives who are non-athletes and with potential evidence of significant genetic variants present in the family, prophylaxis could be performed in the form of more frequent thorough cardiac examinations and potential ICD implantation in high-risk individuals.

Case Background

The proband (IV:3) was a professional footballer who suffered an SCD at the age of 23. Over a 2 year period, he lost consciousness 3 times playing in different professional clubs. After each syncope, all clinical examinations and complete cardiac diagnostics were performed. Due to the lack of a pathological finding, he was given a green light each time to return to the football fields. It is also important that his first cousin once removed (III:1) died at the age of 19 by collapsing abruptly on the football field during a match.

MATERIALS AND METHODS

Participants and Case Preparation

This study includes the family of eight of the deceased professional footballers who died from SCD. To join the study, participants signed an informed consent form. Family members included in the study underwent genetic testing at Invitae Corp. clinical diagnostic laboratory. Knowing that the 2020 APHRS/HRS expert consensus statement indicates the required clinical examinations in the closest kin of the deceased, they were invited for a thorough cardiologic assessment. Due to the family's unwillingness to undergo further clinical examinations, only a part of the data was available. However, the proband's brother (IV:4), who is a professional football player, underwent a thorough clinical assessment including exercise test, holter ECG, echocardiography and ECG in order to determine possible underlying heart disease in line with the APHRS/HRS statement, which could potentially increase the risk of SCD (21). The results of his exercise test showed rare monomorphic ventricular extrasystoles during the first 3 min, later blocked with higher heartbeat frequency. Besides that, there was nothing pathological in the test. Holter ECG was recommended to all family members but was performed only in the same brother as mentioned above (IV:4). There were no pathological findings. Echocardiography didn't show any pathological findings. Echocardiography was also done in the father (III:4) and didn't show any pathological findings. An interpretation of the ECG findings of all family members is provided below.

Although the proband passed away 8 years ago, his family is still deeply saddened. Further investigations remind them of a tragic event therefore they are not prone to additional analysis.

Cardiology Genetic Testing

Peripheral blood or saliva was processed and evaluated by Next Generation Sequencing (NGS). Each gene was targeted with oligonucleotide baits (Agilent Technologies, Santa Clara, CA; Roche, Pleasanton, CA; IDT, Coralville, IA) to capture all coding exons, plus 10–20 bases of flanking intronic sequences, and non-coding regions of clinical interest. Baits were balanced to obtain a minimum of 50× and an average of 350× depth-of-sequence read coverage. A bioinformatics pipeline was utilized that incorporated both standard and custom algorithms to identify single-nucleotide variants, small indels, large indels, structural variants with breakpoints in target sequences, and exon-level copy number variants (CNVs). In addition to standard GATK-based alignments and analysis, validated coverage-based

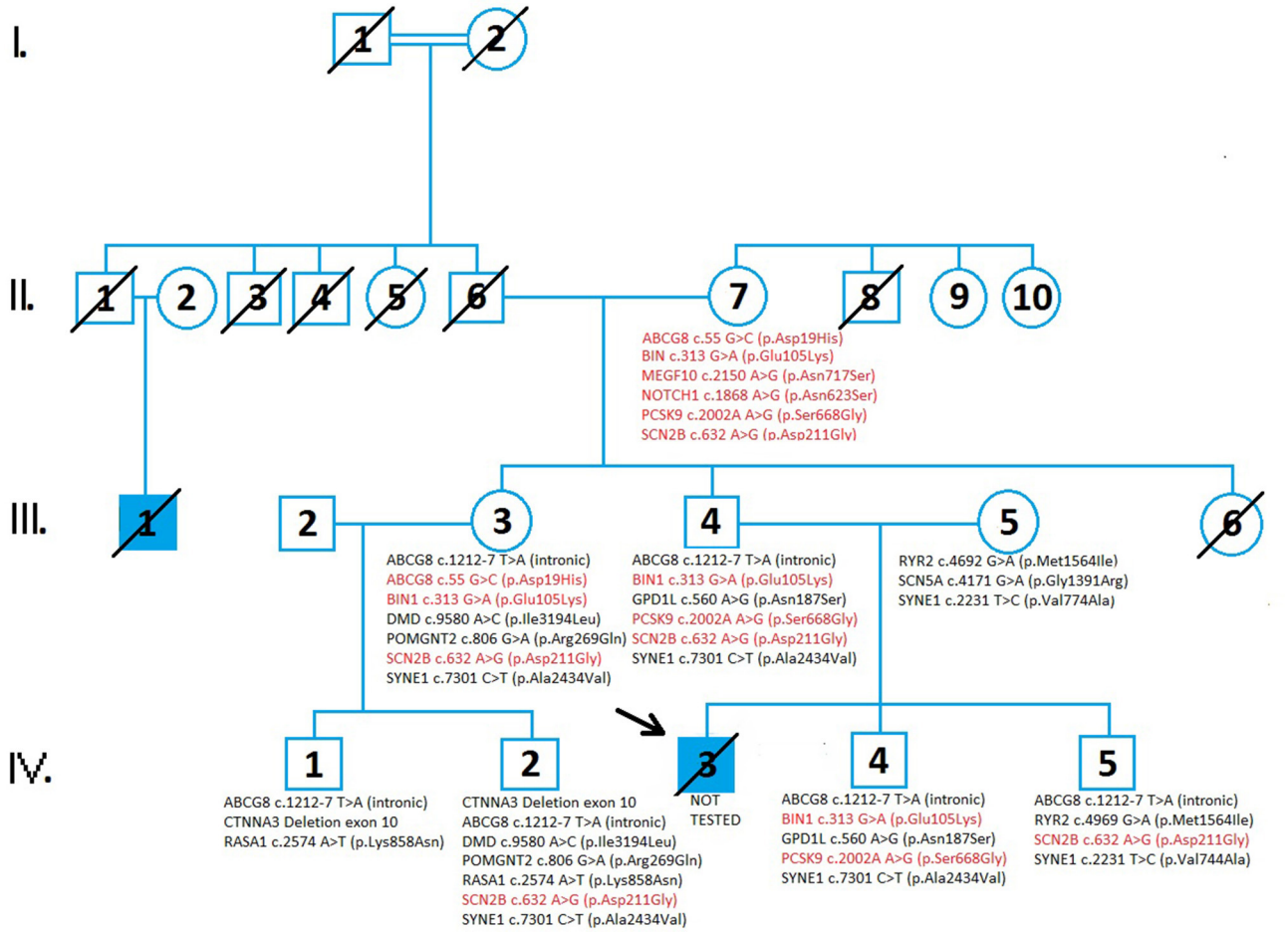


FIGURE 1 | A pedigree analysis demonstrates the cases of sudden cardiac death and the results of genetic testing in four generations of family members. Circles indicate females while the squares indicate males. The diagonal line across the filled square indicated the patients died from sudden cardiac death. Genetic variants inherited from the proband's grandmother (II:7) are represented in red. Genetic variants inherited from the grandfather (II:6) of the proband are shown in black. Genetic testing in family members revealed variants of unknown significance (VUS) and there were no detected pathogenic variants in tested family members. The increased risk allele, ABCG c.55 G>C (p.Asp19His), is presented in II:7 and III:3. A comprehensive genetic test analyzed 294 genes involved in cardiac disease and conditions related to cardiac disease which can lead to sudden cardiac death.

CNV detection algorithms designed to flag possible split-read signals were applied. Once verified, the variant call format was updated and interpreted.

Variants were classified using a point-based scoring system congruent with the system for grading evidence for pathogenicity as recommended by the American College of Medical Genetics and Genomics (22). Orthogonal confirmation of CNVs was performed using gene-centric array-CGH. All testing for the 294-gene panel was performed at Invitae Corp., which is accredited by the Clinical Laboratory Improvement Amendments and College of American Pathologists. The Invitae cardiology genetic test provides a comprehensive analysis of 294 genes involved in the pathogenesis of inherited cardiovascular conditions: arrhythmia, cardiomyopathy, hereditary musculoskeletal diseases, anatomical anomalies of the cardiovascular system, connective tissue disorders with cardiac involvement, familial hypercholesterolemia, pulmonary hypertension, and congenital heart disease.

Review of Family Pedigree

The family pedigree consists of four generations (**Figure 1**). The first generation includes a great-grandmother (I:2) and great-father (I:1) of the proband (in which consanguinity can not be excluded) who had five children. All relatives of the family who could know the information about the degree of consanguinity have passed away and it is not possible to find out what degree of kinship it is. It is an isolated, closed population with only a few inhabitants. All inquiries to living relatives indicate the existence of consaguinity. All five children died in old age from an unknown cause. One of the great-uncles (II:1) had a son, (III:1) who died at the age of 19 by collapsing abruptly on the field during a football match. The grandfather (II:6) and grandmother (II:7) have two living children; aunt of the proband (III:3), father

of the proband (III:4), while one of their daughters died at a young age from leukemia (III:6). Aunt (III:3) and her husband (III:2) have two living children (IV:1 and IV:2). Parents of the proband (III:4 and III:5) have two living sons, brothers of the proband (IV:4 and IV:5) and a proband (IV:3), who died at 23 years old from SCD.

ECGs and Laboratory Blood Tests

A standard 12-channel ECG was performed on all participants. In order to better define possible additional cardiovascular risks (hypercholesterolemia) and based on results of genetic testing, biochemical blood tests were performed in the whole family. Biochemical blood tests included a complete blood count, lipid profile, and creatine kinase level.

RESULTS

Analysis of Genetic Variants

The *ABCG8* gene (ATP binding cassette subfamily G member 8) encodes for sterolin-2, which in combination with sterolin-1 creates sterolin. Sterolin is a transporter protein responsible for eliminating plant sterols. *ABCG8* gene is related to autosomal recessive sitosterolemia, the disorder of cholesterol metabolism characterized by the accumulation of cholesterol in the skin, tendons, and various tissues. Polymorphisms in the *ABCG8* gene are associated with an increased risk of gallstones (23–25), and renal disease in patients with diabetes type 2 (26). Also, *ABCG8* gene polymorphism may contribute to the rapid onset of CAD in patients suffering from familial hypercholesterolemia (27, 28). A unique intronic variant in the *ABCG8* gene, c.1212-7 T>A (intronic) has been found in cousins (IV:2 and IV:1), the aunt (III:3), and brothers (IV:4 and IV:5) of the proband. According to the current prediction algorithm, variant c.1212-7 T>A (intronic) may disrupt the consensus splice site but it is not confirmed by transcriptional studies. An additional variant in the *ABCG8* gene; c.55 G>C (p.Asp19His), classified as an increased risk variant has been identified in the grandmother (II:7) of the proband. This variant has been associated with the development of gallstones.

The *BIN1* gene (Bridging integrator 1) belongs to the N-BAR domain, provides instructions for making membrane-associated protein, essential for biogenesis of plasma membrane invaginations (T-tubules) in muscle tissues, important for muscle contraction and relaxation. Initially, it was identified as a tumor suppressor protein (myc-interacting protein). The *BIN1* gene is associated with autosomal recessive and dominant centronuclear myopathy (CNM), characterized by muscle weakness (myopathy), abnormal localization of nuclei, and growth retardation. Several isoforms of the BIN1 protein are expressed with tissue and disease specificity and therefore some isoform BIN1 was discovered within cardiac transverse tubules (T-tubules) assumed to be important for cardiomyocyte homeostasis (calcium signaling) and is down-regulated in cardiomyopathy (29, 30). Heterozygous missense variant (VUS) c.313G>A (p.Glu105Lys) in the *BIN1* gene has been found in grandmother (II:7), father (III:4), aunt (III:3), and brother (IV:4) of the proband. The algorithm predicting the effect of missense changes on protein structure is contradictory at the moment, suggesting a "tolerated" or "probably damaging" impact.

The *CTNNA3* (catenin alpha 3) gene encodes a protein that belongs to the vinculin/alpha-catenin family, which has a role in cell-cell adhesion, specifically, binding plakophilins in cardiomyocytes. The *CTNNA3* gene is a new candidate gene for arrhythmogenic right ventricular cardiomyopathy (ARVC) (31). ARVC is a congenital heart disease that follows an autosomal dominant pattern, characterized by infiltration of adipose and fibrous tissue into the right ventricle and loss of myocardial cells, resulting in ventricular and supraventricular arrhythmias and RV dysfunction. A gross "in-frame" deletion variant (VUS) encompassing exone 10 of the *CTNNA3* gene results in a protein product lacking 31 amino acids. The functional significance of deleted amino acids is currently unknown and predicting algorithms are not available. *CTNNA3* gene variant was found in cousins (IV:1 and IV:2) of the proband.

The *DMD* (dystrophin) gene is the largest human gene producing cytoskeletal protein dystrophin, located primarily in muscles and in cardiac muscle, that enables the strength, stability, and functionality of myofibers (32). The *DMD* gene is associated with X-linked neuromuscular disorders Duchenne and Becker dystrophies, X linked dilated cardiomyopathy 3B, and familial cardiomyopathy. The underlying pathology of *DMD* related disease is the absence of essential muscle protein dystrophin caused by variants leading to an abnormal gene product. Dystrophin is partially functioning with a shorter protein product (33). A variety of variant types have been reported in the *DMD* gene. The identified heterozygous variant of uncertain significance (VUS) p.Ile3194Leu, is a rare point variant (missense change), with no clear impact on protein structure and/or function. In summary, all evidence is insufficient about the clinical significance of the variant and there is no indication that the variant causes disease. The variant has been identified in aunt (III:3) and her son (IV:2). Further analysis of family history data and detailed clinical examination did not reveal any signs of muscle weakness and muscular pathology in family members.

The *GPD1L* (glycerol-3-phosphate dehydrogenase 1 like) gene encodes a protein that catalyzes the conversion of sn-glycerol 3-phosphate to glycerone phosphate. The encoded protein binds the sodium channel, voltage-gated, type V, alpha subunit (*SCN5A*-coded Nav1.5) (34). It is expressed near the cell membrane and regulates the action potential of cardiomyocytes. Previous literature data suggest that genetic alterations in the *GPD1L* gene are related to cardiac rhythm disorders; Brugada syndrome, sudden infant death syndrome (SIDS), and long QT syndrome (35, 36). On the other hand, Hosseini et al. showed that only *SCN5A* gene alterations have definitive evidence in the pathogenesis of Brugada syndrome (37). The latest clinical study published in 2020 demonstrated that genetic alterations in the *GPD1L* gene led to decreased activation of the sodium channel and consequently early repolarization syndrome, one of the causes of sudden cardiac death (38). In this pedigree, father (III:4) and son (IV:4) share the same VUS in the *GPD1L* gene. According to the prediction algorithm, this variant is likely to be tolerated but additionally, functional and clinical studies for confirmation of its significance are needed.

The *MEGF10* (Multiple EGF-like domains 10) gene encodes multiple epidermal growth factor-like domain protein families, which contains 17 atypical EGF-like domains, each with eight cysteine residues (39). The encoded protein plays a role in cell adhesion, motility, and proliferation, and is a critical mediator of apoptotic cell phagocytosis as well as amyloid-beta peptide uptake in the brain. Expression of this gene may be associated with autosomal recessive early-onset minicore myopathy, areflexia, respiratory distress, and dysphagia (EMARDD) and is also associated with schizophrenia (40, 41). Specific missense change c.2150A>G (p.Asn717Ser) at codon 717 of MEGF10 protein has been identified as heterozygous in only one family member, grandmother (II:7) of the proband. The effect of sequence change on RNA splicing was predicted contradictory by PolyPhen to be "Benign" and by SIFT: to be "Deleterious." The role of this variant in disease is uncertain since the available evidence is insufficient.

The *NOTCH1* (Notch 1) gene encodes for protein Notch1, a member of Notch family receptors. NOTCH is a key signaling pathway in the development of many tissues ensuring crosstalk between different types of cells, their physiological proliferation, differentiation, and cell self-destruction (apoptosis) (42). Despite its involvement in many key developmental systems, variants in the *NOTCH1* gene are mainly associated with autosomal dominant aortic valve disorder and Adams-Oliver Syndrome. Various cardiac disorders (BAV, aortic aneurysm, aortic coarctation), as well as the formation and progression of aortic valve calcification, were reported in correlation with *NOTCH1* variants (43). Heterozygous *NOTCH1* missense variant c.1868A>G (p.Asn623Ser) was also found in one family member, grandmother (II:7). Missense change on protein structure has a SIFT score deleterious and is rated as probably damaging by Poly-Phen 2. The contribution of this variant in predisposing disease remains uncertain and requires further evidence.

The *PCSK9* (proprotein convertase subtilisin/kexin type 9) gene is associated with familial hypercholesterolemia. Genetic variants of *PCSK9* are present in familial hypercholesterolemia and familial hypobetalipoproteinemia (44, 45).

A heterozygous missense variant in the *PCSK9* gene; c.2002A>G (p.Ser668Gly) has been found in the grandmother (II:7) and father (III:4) and brother (IV:4) of the proband. This variant does not disrupt protein function, but its significance is still not confirmed by functional studies.

The *POMGNT2* [Protein O-linked mannose N acetylglucosaminyltransferase 2 (beta 1,4)] gene encodes for endoplasmic reticulum (ER) – a resident protein that catalyzes the second step of the O-mannosyl glycosylation in the mucin-like domain of α-dystroglycan (46). Defect in glycosylation of α-dystroglycan led to a subgroup of muscular dystrophies, known as dystroglycanopathies. Genetic variants of *POMGNT2* are associated with limb-girdle muscular dystrophy (47). Heterozygous, *POMGNT2* missense variant c.806G>A (p.Arg.269Gln) was identified in two members, aunt (III:3) and cousin (IV:2). Missense change does not adversely affect protein structure and function, thus algorithm predictions by PolyPhen-2 "Benign" and SIFT "Tolerated" are aligned.

The *RASA1* (RAS p21 protein activator 1) gene encodes a protein called p120-RasGAP involved in the regulation of the RAS/MAPK signaling pathway from outside the cell to the cell's nucleus. Although its role is not entirely clear, it appears to be essential for the normal development of the vascular system. Variants in the *RASA1* gene are associated with autosomal dominant capillary malformation-arteriovenous malformations (CM-AVM) and Parker Weber syndrome (48, 49). Heterozygous, *RASA1* missense variant (VUS) c.2574A>T (p.Lys858Asn) has been identified in cousins (IV:2 and IV:1) of the proband. The specific missense change on protein structure and function has a SIFT score deleterious and is rated as probably damaging by Poly-Phen 2. The contribution of this variant in predisposing disease remains uncertain and requires further evidence.

The *RYR2* (ryanodine receptor 2) gene codes a protein called ryanodine receptor 2 involved in calcium transport within cells. *RYR2* gene alterations are involved in the pathogenesis of catecholaminergic polymorphic ventricular tachycardia, familial atrial fibrillation, and cardiomyopathy (20, 50). Mother (III:5) and son (IV:5) share the same VUS in the *RYR2* gene, c.4692G>A (p.Met1564Ile). In current literature, this variant has not been described and prediction tools gave inconclusive results regarding their clinical significance. This variant is transmitted from the maternal side so it could not be directly involved in both cases of SCD in this family but some additive effect of this variant is still unknown.

The *SCN5A* (sodium voltage-gated channel alpha subunit 5) gene codes the pore-forming alfa subunit of the primary cardiac channel. Genetic variants of *SCN5A* play role in Brugada syndrome, progressive familial heart block, Romano-Ward syndrome, sick sinus syndrome, familial atrial fibrillation, and familial dilated cardiomyopathy (51–53). In our study, a mother (III:5) without any cardiac condition had a heterozygous (VUS) variant, c.4171G>A (p.Gly1391Arg) in the *SCN5A* gene and it wasn't transmitted to a currently living offspring.

The *SCN2B* (Sodium Voltage-Gated Channel Beta Subunit 2) gene codes the beta II subunit of type II voltage-gated sodium channel. *SCN2B* gene alterations lead to autosomal dominant Brugada syndrome, atrial fibrillation, and sudden infant death syndrome (54). Heterozygous VUS has been found in all generations, including II:7, III:3, III:4, IV:2, and IV:5, suggesting an autosomal dominant mode of inheritance. This *SCN2B* variant, c.632A>G (p.Asp211Gly) in exon 4 has been reported to affect SCN2B protein function. The algorithm predicting the effect of sequence changes on RNA splicing suggests that this variant may create or strengthen a splice site. Unfortunately, this prediction is not yet confirmed by transcriptional studies.

The *SYNE1* (spectrin repeat-containing nuclear envelope protein 1) gene codes Syne-1 protein is mostly expressed in the cerebellum and involved in the coordination of movements. Their genetic alterations can lead to autosomal recessive cerebellar ataxia type 1, Emery-Dreifuss muscular dystrophy, and recently there are some reports of dilated cardiomyopathy in some patients (55, 56). Our study revealed two different VUSs. The heterozygous *SYNE1*: c.7301C>T (p.Ala2434Val) variant has been found in III:3, III:4, IV:2, and IV:4, while the heterozygous c.2231T>C (p.Val744Ala) variant was found in III:5 and IV:5.

Both variants require further transcriptional studies to confirm their effect on protein function. There were no ataxia cases in this family. A summary of the genetic variants is provided in the table below (**Table 1**).

Pedigree Analysis

The pedigree analysis encompassed available data from four generations of the family (**Figure 1**). Available data included data on health conditions from medical history, causes of death, results of laboratory testing, and genetic testing results.

Grandmother (II:7) has hypertension and increased levels of blood cholesterol. Her genetic test revealed six variants of unknown significance that are related to cholesterol metabolism (*ABCG8*, *PCSK9*), muscle contraction (*BIN1*), tissue development, and apoptosis (*NOTCH1*, *MEGF10*) as well as sodium channel in cardiac muscle (*SCN2B*).

In the third generation, genetic testing was performed on the mother (III:5), father (III:4), and aunt (III:3). Father (III:4) and aunt (III:3) had the same variants in genes: *ABCG8*, *BIN1*, *SCN2B*, and *SYNE1* (related to Emery-Dreifuss muscular dystrophy with cardiomyopathy, cardiac conduction defects, and cerebellar ataxia). Person III:3 is a compound heterozygote for ABCG8 gene variants with observed elevated blood cholesterol, potentially induced by the variants genotype. However, her BMI is 48.4 therefore the increase in cholesterol levels can be attributed to her diet and lifestyle habits. Additional variants in aunt were present in *DMD* (Duchenne muscular dystrophy), *POMGNT2* (dystrophy-dystroglycanopathy type A8) that are not involved in cardiac pathology. An additional variant in father (III:4) was present in the *GPD1L* gene, involved in the pathogenesis of early repolarization syndrome, Brugada syndrome type 2 and sudden cardiac death in children. Genetic testing in the mother (III:5) revealed variants in the *RYR2* gene related to cardiomyopathy, autosomal catecholaminergic polymorphic ventricular tachycardia, arrhythmogenic dysplasia of the right ventricle. Additional variants were present in *SYNE1* and *SCN5A* genes related to various cardiac conduction defects (Brugada syndrome, long QT interval type 3, and cardiomyopathy).

In the fourth generation, genetic testing was performed in brothers (IV:4 and IV:5) of the proband. They share the same variants in *ABCG8* and *SYNE1* gene. Additional variants were present in *BIN1*, *GPD1L*, *PCSK9*, and *RYR2*, and *SCN2B*.

According to the test results, cousins (IV:1 and IV:2) have the same variants in *ABCG8* (intronic), *CTNNA3*, and *RASA1* genes, involved in the pathogenesis of arrhythmogenic right ventricular cardiomyopathy (*CTNNA3*) and *RASA1* (capillary malformations, Park Weber syndrome). Also, additional variants of *DMD*, *POMGNT2*, *SCN2*, and *SYNE1* gene were found in IV:2.

Interpretation of Participants' Laboratory Blood Tests and ECGs

The ECGs of all study participants showed an orderly finding of heart action, without any pathological changes suggesting heart disease (**Table 2**). Complete blood counts in all participants showed no major deviations from normal values. Blood cholesterol was increased in grandmother (II:7), mother (III:5), father (III:4), and cousin (IV:1). Interestingly, the highest blood cholesterol level has been found in the mother of an index patient who doesn't have any genetic alterations in these cholesterol metabolism-related genes. Also, it is important to consider a high body mass index in these patients, which contributes to alterations in cholesterol metabolism and blood levels. III:5, II:7, IV:1, and III:3 have an increased cholesterol ratio that matches their phenotype. III:5 and IV:1 have an increased creatine kinase which can also be associated with being overweight (**Table 3**).

DISCUSSION

Interpretation of Genetic Variants

SCD is a catastrophic event not only for the patient and his family but also for the medical staff and the general public. According to the CDC (Center for Disease Control), approximately, more than 2,000 people in the USA experience sudden cardiac death each year. Following recommendations of the European Society of Human genetics (57) our study has been conducted by a multidisciplinary team and took all relevant family history data, clinical investigation and cardiac genetic screening test. In clinical practice, a family history of SCD requires careful medical attention and evaluation. In the era of genomic medicine genetic testing and detection of pathogenic variants enables accurate diagnosis and exact risk assessment for the whole family (57, 58). Unfortunately, in this manuscript as well as in many SCD cases biological samples from deceased patients for genetic analysis were not taken and it was is not possible to define a diagnosis, calculate risk for family members, and provide appropriate medical care.

In this manuscript, we describe four generations of the family with two cases of SCD at an early age.

Pedigree analysis in this family revealed alterations in seven genes related to various cardiac conditions, mostly involved in cardiac rhythm disorders (*GPD1L, RYR2, SCN5A, SCN2B*), structural abnormalities that could involve cardiac muscle (*CTNNA3, DMD*), and regulation of calcium homeostasis in cardiac muscle *(BIN1)* but according to our updated knowledge, none of this variants were pathogenic. In order to get a more comprehensive insight into SCD and potential pathophysiological mechanisms, our analysis included additional genes related to glycosylation (α-dystroglycan, *POMGNT2*) and the development of vascular abnormalities (*RASA1*). In latter cases, the analysis did not reveal pathogenic variants. Considering that SCD at an early age is usually the result of an unrecognized cardiac conduction defect, we paid special attention to the analysis of genes involved in cardiac rhythm disorders. According to modern cardiology concepts (59), the pathogenesis of cardiac arrhythmia includes abnormalities in ion channel activity, cellular signaling, and structural abnormalities mediated by a large number of genes.

As can be seen from the pedigree analysis same variants; *SCN2B* c.632A>G (p.Asp211Gly) and *BIN1* c.313G>A (p.Glu105Lys) were present in II generation (II:7), III generation (III:3 and III:4), and their offspring in IV generation (IV:4 and IV:2). In these patients, there were no cardiac events and the results of an ECG and heart ultrasound were normal.

TABLE 1 | List of genetic biomarkers recorded by INVITAE cardiology genetic test (SCD – sudden cardiac death, SNP – single nucleotide polymorphism, VUS – a variant of uncertain significance).

Gene	Gene function	Association with SCD	ZYGOSITY	SNP/variant	Pathogenic variant or VUS	A family member with genetic variant	gnomADv2.1.1 frequency	V3.1 frequency
ABCG8	Encodes a protein sterolin-2 which transports plant sterols	1. Sitosterolemia 2. Renal disease in patients with diabetes type 2 3. Enhancement of the effect of familial hypercholesterolemia 4. Gallstones	Heterozygous	c.1212-7 T>A (intronic)	VUS	III:3, III:4, IV:1, IV:2, IV:4. IV:5.	0.0001273	0.0001052
ABCG8	Encodes a protein sterolin-2 which transports plant sterols	1. Sitosterolemia 2. Gallstones 3. Renal disease in patients with diabetes type 2 4. Enhancement of the effect of familial hypercholesterolemia	Heterozygous	c.55 G>C (p.Asp19His)	Pathogenic variant	II:7, III:3	0.06640	missing VEP annotations
BIN1	Encodes a BIN1 protein that has a role in endocytosis and apoptosis	1. Skeletal muscle myopathy 2. Regulation of calcium homeostasis 3. Cardiomyopathy	Heterozygous	c.313 G>A (p.Glu105Lys)	VUS	II:7, III:3, III:4, IV:4	absent	absent
CTNNA3	Encodes a protein that belongs to the vinculin/alpha-catenin family	Arrhythmogenic right ventricular dysplasia	Heterozygous	Deletion exon 10	VUS	IV:1, IV:2	N/A	N/A
DMD	Encodes protein dystrophin -has a role in strengthening muscle fibers	1. X-linked dilated cardiomyopathy 2. Familial dilated cardiomyopathy 3. Duchenne muscular dystrophy	Hemizygous	c.9580 A>C (p.Ile3194Leu)	VUS	III:3, IV:2	0.00002480	0.00007145
GPD1L	Encoded a protein that catalyzes the conversion of sn-glycerol 3-phosphate to glycerone phosphate	1. Brugada syndrome 2. Brugada syndrome 2 3. Long QT syndrome 4. Sudden infant death syndrome	Heterozygous	c.560 A>G (p.Asn187Ser)	VUS	III:4, IV:4	0.000003977	absent
MEGF10	Encodes a member of the multiple epidermal growth factor-like domains protein family	1. Respiratory distress 2. Myopathy	Heterozygous	c.2150 A>G (p.Asn717Ser)	VUS	II:7	0.000178	0.0001248
NOTCH1	Encodes a protein called Notch1, a member of the Notch family of receptors	1. Critical congenital heart disease 2. Bicuspid aortic valve 3. Aortic aneurysm 4. Aortic coarctation 5. Formation and progression of aortic valve calcification	Heterozygous	c.1868 A>G (p.Asn623Ser)	VUS	II:7	0.00003260	absent
PCSK9	Encodes a protein that helps regulate cholesterol level in the blood	1. Familial hypercholesterolemia 2. Familial hypobetalipoproteinemia	Heterozygous	c.2002 A>G (p.Ser668Gly)	VUS	II:7, III:4, IV:4	0.00004790	0.00003941
POMGNT2	Encodes a protein with glycosyltransferase activity	Limb-girdle muscular dystrophy	Heterozygous	c.806 G>A (p.Arg269Gln)	VUS	III:3, IV:2	0.00002387	absent

(Continued)

TABLE 1 | Continued

Gene	Gene function	Association with SCD	ZYGOSITY	SNP/variant	Pathogenic variant or VUS	A family member with genetic variant	gnomADv2.1.1 frequency	V3.1 frequency
RASA1	Encodes a protein called p120-RasGAP included in RAS/MAPK signaling pathway	1. Parkes Weber syndrome 2. Capillary malformation-arteriovenous malformation syndrome (CM-AVM)	Heterozygous	c.2574 A>T (p.Lys858Asn)	VUS	IV:1, IV:2	0.000007988	absent
RYR2	Encodes a protein called ryanodine receptor 2 involved in the regulation of calcium channels	1. Catecholaminergic polymorphic ventricular tachycardia (CPVT) 2. Arrhythmogenic right ventricular cardiomyopathy 3. Familial atrial fibrillation	Heterozygous	c.4692 G>A (p.Met1564Ile)	VUS	III:5, IV:5	0.00008902	0.00003287
SCN5A	Produces a protein essential for the regulation of sodium channels	1. Romano-Ward syndrome 2. Brugada syndrome 3. Progressive familial heart block 4. Sick sinus syndrome 5. Familial atrial fibrillation 6. Familial dilated cardiomyopathy	Heterozygous	c.4171 G>A (p.Gly1391Arg)	VUS	III:5	0.00003603	absent
SCN2B	Encodes beta 2 subunits of type II voltage-gated sodium channel	1. Brugada syndrome 2. Atrial fibrillation	Heterozygous	c.632 A>G (p.Asp211Gly)	VUS	II:7, III:3, III:4, IV:2, IV:5	0.000007954	0.00001972
SYNE1	Encodes an SYNE-1 protein present in Purkinje cells responsible for coordinating movement	1. Emery-Dreifuss muscular dystrophy 2. Dilated cardiomyopathy	Heterozygous	c.7301 C>T (p.Ala2434Val)	VUS	III:3, III:4, IV:2, IV:4	absent	missing VEP annotations
SYNE1	Encodes an SYNE-1 protein present in Purkinje cells responsible for coordinating movement	1. Emery-Dreifuss muscular dystrophy 2. Dilated cardiomyopathy	Heterozygous	c.2231 T>C (p.Val744Ala)	VUS	III:5, IV:5	absent	absent

TABLE 2 | Description of ECG records of all participants.

ECG / participants	Rhytm	Electrical axis	ST-T interval description	QT/QTc	Additional findings
IV:4	Sinus rhythm 50/min	Intermediate	No changes	440/402 msec	Incomplete RBBB
IV:5	Sinus rhythm 61/min	Intermediate	No changes	442/427 msec	
IV:2	Sinus rhythm 45/min	Intermediate	No changes	476/411 msec	PVC
IV:1	Sinus rhythm 99/min	Intermediate	No changes	363/419 msec	
III:5	Sinus rhythm 67/min	Intermediate	No changes	404/427 msec	
III:4	Sinus rhythm 48/min	Intermediate	No changes	500/447 msec	
III:3	Sinus rhythm 66/min	Intermediate	No changes	393/406 msec	
II:7	Atrial fibrillation with ventricular response of 91/min	Intermediate	No changes	338/386 msec	

(RBBB- right bundle branch block, PVC- premature ventricular contraction).

TABLE 3 | Findings in laboratory analysis of blood and association with BMI.

Laboratory findings and BMI	Total cholesterol	LDL	VLDL	HDL	Triglycerides	Cholesterol ratio	CK	BMI
Reference range / Participants	< 5,0 mmol/L	< 3,0 mmol/L	0,1-0,5 mmol/L	W >1.2mmol/L M = 1,0-1,9 mmol/L	<1,7 mmol/L	< 3,5	W < 153 U/L M < 177 U/L	<25 kg/m^2
III:5	**7.8**	/	/	**1.2**	**8.7**	**6.5**	**161**	**31,6**
II:7	**6.5**	**4.0**	**0.8**	1.7	**1.7**	**3.8**	51	**32,2**
IV:1	**7.1**	**5.0**	**0.8**	1.3	**1.8**	**5.5**	**284**	**46,8**
III:3	**5.9**	**3.6**	**1.0**	1.3	**2.1**	**4.5**	102	**48,4**
IV:5	3.9	1.9	**0.8**	1.2	**1.8**	3.3	128	**31,1**
IV:4	4.5	2.5	0.4	1.5	1.0	3.0	73	**25,6**
IV:2	4.3	1.9	0.4	**2.1**	0.8	2.0	**240**	23,8
III:4	**5.7**	**3.5**	/	**2.0**	1.4	2.9	107	**42,0**

Elevated values are bolded.

Fields with "/" represent uncalculated values due to technical difficulties. (LDL-low density lipoprotein, VLDL-very low-density lipoprotein, HDL-high density lipoprotein, CK- creatine kinase, BMI-body mass index).

Analysis of other variants in this pedigree showed a variant of unknown significance; *GPD1L* c.560A>G (p.Asn187Ser) present in the father (III:4) and brother (IV:4) of the proband. This variant wasn't present in the II generation (II:7) and other family members suggesting that this variant originates from the paternal side of the II generation [grandfather (II:6) and his brother (II:1)]. Father's (III:4) cousin (III:1) also experienced SCD at an early age and there is a 50% possibility that he had this variant also. Unfortunately, we did not have a possibility to perform genetic testing in person III:1 so we can not confirm this hypothesis. Also, it is very important to emphasize that the proband might have inherited *SCN5A* c.4171 G>A (p.Gly1391Arg) genetic variant from his mother (III:5). She represents the only carrier of this genetic variant in the whole pedigree. The fact that proteins encoded by *SCN5A* and *GPD1L* are included in the same regulatory pathway, in which *SCN5A* encodes sodium channel, voltage-gated, type V, alpha subunit, and *GPD1L* encodes an intracytoplasmic membrane-bound enzyme bonded to that sodium channel, may indicate that there was a chance for the additive effect of the two genetic variants, ie digenic inheritance (34–36). It has been proven that 30% of the causes of Brugada syndrome originate in the polymorphism of the *SCN5A* gene. Genetic variants of other genes, including *GPD1L*, *SCN1B, SCN2B, SCN3B, RANGRF, SLMAP, KCNE3, KCNJ8, HCN4, KCNE5, KCND3, CACNA1C, CACNB2B, CACNA2D1,* and *TRPM4*, may also underlie Brugada syndrome. These genes encode proteins responsible for cardiac sodium, potassium, and calcium channels, as well as proteins involved in the circulation or regulation of these channels (52). When the enzymatic activity of the protein encoded by *GPD1L* is decreased, levels of glycerol 3-phosphate are higher. It activates the *GPD1L*-dependent *SCN5A* phosphorylation pathway. Consequently, sodium current may be decreased. Also, an imbalance of NAD(H) induced by *GPD1L* can result in a reduced-sodium current. Combined with a possible co-present defective protein encoded by *SCN5A*, it may have increased the chance of developing Brugada syndrome (60). Given that the proband had a 2-year history of cardiac symptoms, where he had syncope on three occasions and after each complete cardiac treatment had a proper finding and received the green light to continue playing professional football, it can be assumed that he did not suffer from any structural heart disease which would be seen during the routine ultrasound examination.

Also, according to Snir et al. a regular ECG finding does not rule out the existence of Brugada syndrome, because many patients have an intermittent Brugada syndrome ECG pattern (61).

Various triggers of Brugada syndrome are mentioned in the literature including fever, lithium, antidepressants, sodium channel blockers, and exercise (62–64). Therefore, diagnosis is difficult, and it is easy to overlook the existence of a Brugada pattern. This may explain the fact why the proband did not have any pathological findings on cardiac examinations.

Although the genetic variants *PCSK9* c.2002A A>G (p.Ser668Gly), *ABCG8* c.1212-7 T>A (intronic) and *ABCG8* c.55 G>C (p.Asp19His) present in the grandmother (II:7) and the father (III:4) of the proband are responsible for hereditary hypercholesterolemia, the proband died at the age of 23 without any visible signs of hypercholesterolemia including xanthoma, xanthelasma, jaundice. Therefore, the etiology of CAD in the event of SCD can be ruled out.

The presence of consanguinity in grandfather's side of first-generation increases the risk for the presence of various inherited disorders including cardiac conduction abnormalities. Taken together, the presence of consanguinity in this part of the family, as well as the presence of a unique variant in the *GPD1L* gene not present in the grandmother (II:7) indicates that SCD could be related to genetic alterations in the grandfather (II:1) and his siblings /relatives. Unfortunately, grandfather (II:1) and his relatives died so this assumption cannot be confirmed by genetic testing. Additional functional studies are needed to clarify the role of *GPD1L* c.560A>G (p.Asn187Ser) variant in the pathogenesis of Brugada syndrome.

In the third-generation, the mother (III:5) had VUS variants in two genes involved in cardiac rhythm disorders; *SCN5A* and *RYR2*, and she didn't have any cardiac pathology condition. *SCN5A* genetic alterations are present in 30% of all Brugada syndrome cases while the *RYR2* genetic alterations are reported because of catecholaminergic polymorphic ventricular tachycardia, atrial fibrillation, and arrhythmogenic right ventricular cardiomyopathy.

Current literature data emphasize the high variability of disease phenotype in patients with the same pathogenic variant (*SCN5A*), even within the same family. There are patients with malignant arrhythmias, SCD and on the other hand, there are patients without any clinical signs. Reduced or incomplete penetrance in these families has been observed (65). Also, disease expressivity and severity are regulated by individual-specific factors related to other cardiac ion channels that regulate cardiac action potential. Recent studies suggest that single nucleotide polymorphisms, copy number variations, and their combination can modulate disease expressivity and severity in patients with *SCN5A* variants.

Additional non-genetic modifiers such as gender, age, tobacco, drug and alcohol use, medication, exercise, fever, comorbidities, and lifestyle are also important determinants of disease severity. It is well known that individual *SCNA5* variant carriers develop long QT intervals early in life (at birth) while Brugada syndrome is seen later in life (66). Alcohol, fever, tobacco, exercise, drug, and some medication may modulate the electrical activity of the cardiac channel directly or indirectly and trigger arrhythmia and cardiac events. Obesity and hypertension are related to metabolic disturbances (dyslipidemia) and progressive myocardial remodeling that result in electrical, homeostatic, and structural alterations. All these changes act synergistically and can precipitate cardiac events.

Interestingly, both parents of the proband had variants in two different genes involved in cardiac rhythm diseases. It is assumed that some additive effects of these variants in cardiac disease pathology cannot be excluded. Ultimately, the interpretation of variants depends on their frequency in the population and ethnic background that also has to be taken into consideration.

Lack of visible ECG abnormalities and absence of structural cardiac disorder on a routine ultrasound could obscure cardiac structural or conduction defects in many patients with a previous history of cardiac syncope and family history of sudden cardiac death. Genetic testing in these patients enables the detection of various genetic variants and could modify routine diagnostic workup. Detection of pathogenic variants in genes involved in cardiac conduction (*GPD1L*, *RYR2*, *SCN5A*, *SCN2B*) and structural (*CTNNA3*) defects indicates a more comprehensive approach in the diagnostic workup and requires additional testing; ajmaline provocation tests (Brugada syndrome), electrophysiology cardiac studies as well as magnetic resonance and echocardiographic strain imaging.

Even though our case did not involve mitochondrial DNA analysis, heart diseases that can lead to SCD can be inherited by mitochondrial DNA and this type of inheritance should be considered in further research (67).

As well as in many other cases of SCD, the absence of an autopsy in deceased individuals and lack of available biological samples for genetic testing is the main limitations of our study. Despite the initiative of European Council and Recommendations (68) that enable clear criteria when autopsy is performed, there are still large variations among countries. The lack of available biological samples for genetic testing data does not allow any insight into the genetic basis of SCD in deceased individuals. Recent studies showed a diagnostic yield of 40% in cases where a multidisciplinary team approach conducted postmortem genetic testing, family and clinical investigation (69, 70). In the circumstances of our legislation, our study tried to elucidate the genetic basis of SCD in this family using all available resources. Although we did not detect pathogenic variants in this family that could explain SCD cases, we found many variants whose effect is still unknown and requires additional investigations.

Future Perspective

Based on a comparison between genes of living relatives with the genes of the deceased, significant gene variants could be identified and associated with an increased risk of SCD. Although current clinical guidelines do not recommend genetic testing in patients without symptoms (syncope, arrhythmia) it is not possible to ignore opportunities and benefits (non-invasive method compared to various provocation tests, detection of pathogenic variant) of genetic testing. Since genetic testing is a non-invasive method, it should be a routine part of the diagnostic workup in patients with cardiac syncope/family history of sudden cardiac death. Based on genetic testing results it is possible to change diagnostic workup using electrophysiological studies and more detailed cardiac imaging studies that enable early

detection of cardiac pathology. Such a discovery would provide a revolutionary new indication at the level of genetics, that would provide the possibility of ICD prophylaxis in those patients who haven't yet developed heart disease. Based on risk factors, genotype, and clinical examination results, patients could be stratified into specific disease phenotype groups which would facilitate further patient management. The purpose of this kind of approach is in reducing the incidence and consequently the mortality from SCD. Genetic testing in these patients enables detection of inherited pathogenic variants, but there is also a high probability of uncertain findings, that cannot explain SCD in these families. Appropriate genetic counseling, before and after testing, provides at-risk relatives with relevant information on genetic basis of SCD, possibilities of genetic testing and its impact on further management. During pre-testing genetic counseling, all possible outcomes of genetic testing should be discussed and explained in detail to the patient. In this way, the level of psychological stress is reduced to a minimum.

Genetic testing inherently increases costs of management of relatives who are at risk of SCD, however this cost is lower than medical-cardiovascular regular check-ups every 3–5 years (71). Considering all aspects of SCD management, benefits and limits of genetic testing and limited knowledge on SCD, our study encourages genetic testing in at-risk relatives as a powerful tool that could elucidate cause of SCD and give valuable data needed for appropriate clinical management. This approach has been successfully implemented in diseases where genetic background has significant pathogenetic role. Breast cancer diagnostics and management showed suitable example of this approach. In these patients, use of genetic testing in combination with relevant clinical, radiological and laboratory data has significantly influenced diagnostic and therapeutic procedures (72).

CONCLUSION

In conclusion, our pedigree analysis did not find any pathogenic or likely pathogenic variant according to existing data that could give a valuable explanation of SCD in this family. Both parents had variants in several cardiac-related genes, but their clinical significance is still uncertain and we cannot ignore possible causal relation in the pathogenesis of SCD. For this reason, additional functional *in vitro*/*in vivo* studies are necessary to determine the importance of these variants and related risks. This family study emphasizes the importance of a systematic approach in the diagnostic workup of these patients and their relatives. A detailed approach requires concise family history data, results of clinical testing and laboratory data, and of taking blood samples for genetic testing in every case of SCD. Furthermore, large family cohorts, segregation analysis, additional functional studies of variants, detailed family history, and clinical examination are important for a better definition of disease-related genes and translation variants in clinically relevant data. We should also be aware that the variants of uncertain significance are prone to reclassification with further analysis and should be careful when interpreting their clinical significance.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Ethics Committee of the St. Catherine Specialty Hospital. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

DP: conceptualization. DP, LO, VP, JĆ, and AS: data curation. DP, LO, VP, JĆ, JŠ, VR, ŠM, RN, MV, SA, TS, VMa, VMo, AS, JM, and JB: roles/writing—original draft. DP, VP, VMa, VMo, and JB: writing—review & editing. All authors contributed to the article and approved the submitted version.

ACKNOWLEDGMENTS

We thank St. Catherine Specialty Hospital, Invitae, the Croatian Football Federation, the International Society for Applied Biological Sciences for their support.

REFERENCES

1. Isbister J, Semsarian C. Sudden cardiac death: an update. *Intern Med J.* (2019) 49:826–833. doi: 10.1111/imj.14359
2. Lucena JS. Sudden cardiac death. *Forensic Sci Res.* (2019) 4:199–201. doi: 10.1080/20961790.2019.1622062
3. Wong CX, Brown A, Lau DH, Chugh SS, Albert CM, Kalman JM, et al. Epidemiology of sudden cardiac death: global and regional perspectives. *Hear Lung Circ.* (2019) 28:6–14. doi: 10.1016/j.hlc.2018.08.026
4. Risgaard B, Winkel BG, Jabbari R, Behr ER, Ingemann-Hansen O, Thomsen JL, et al. Burden of sudden cardiac death in persons aged (1) to 49 years nationwide study in denmark. *Circ Arrhythmia Electrophysiol.* (2014) 7:205–11. doi: 10.1161/CIRCEP.113.001421
5. Hayashi M, Shimizu W, Albert CM. The spectrum of epidemiology underlying sudden cardiac death. *Circ Res.* (2015) 116:1887–906. doi: 10.1161/CIRCRESAHA.116.304521
6. Bagnall RD, Weintraub RG, Ingles J, Duflou J, Yeates L, Lam L, et al. A prospective study of sudden cardiac death among children and young adults. *N Engl J Med.* (2016) 374:2441–52 doi: 10.1056/NEJMoa1510687
7. Magi S, Lariccia V, Maiolino M, Amoroso S, Gratteri S. Sudden cardiac death: focus on the genetics of channelopathies and cardiomyopathies. *J Biomed Sci.* (2017) 24:56. doi: 10.1186/s12929-017-0364-6
8. Morentin B, Suárez-Mier MP, Monzó A, Molina P, Lucena JS. Sports-related sudden cardiac death due to myocardial diseases on a population from 1–35 years: a multicentre forensic study in Spain. *Forensic Sci Res.* (2019) 4:257–66. doi: 10.1080/20961790.2019.1633729

9. Semsarian C, Ingles J, Wilde AAM. Sudden cardiac death in the young: the molecular autopsy and a practical approach to surviving relatives. *Eur Heart J.* (2015) 36:1290–6. doi: 10.1093/eurheartj/ehv063
10. Wasfy MM, Hutter AM, Weiner RB. Sudden cardiac death in athletes. *Methodist Debakey Cardiovasc J.* (2016) 12:76–80. doi: 10.14797/mdcj-12-2-76
11. Asif IM, Harmon KG. Incidence and etiology of sudden cardiac death: new updates for athletic departments. *Sports Health.* (2017) 9:268–79. doi: 10.1177/1941738117694153
12. Marijon E, Tafflet M, Celermajer DS, Dumas F, Perier MC, Mustafic H, et al. Sports-related sudden death in the general population. *Circulation.* (2011) 124:672–81. doi: 10.1161/CIRCULATIONAHA.110.008979
13. Toresdahl BG, Rao AL, Harmon KG, Drezner JA. Incidence of sudden cardiac arrest in high school student athletes on school campus. *Hear Rhythm.* (2014) 11:1190–4. doi: 10.1016/j.hrthm.2014.04.017
14. Corrado D, Basso C, Pavei A, Michieli P, Schiavon M, Thiene G. Trends in sudden cardiovascular death in young competitive athletes after implementation of a preparticipation screening program. *J Am Med Assoc.* (2006) 296:1593–601. doi: 10.1001/jama.296.13.1593
15. Bezzina CR, Lahrouchi N, Priori SG. Genetics of sudden cardiac death. *Circ Res.* (2015) 116:1919–36. doi: 10.1161/CIRCRESAHA.116.304030
16. Corrado D, Zorzi A. Sudden cardiac death in young people and athletes. *Ital J Med.* (2018) 12:74–87. doi: 10.4081/itjm.2018.1027
17. Corrado D, Zorzi A. Sudden death in athletes. *Int J Cardiol.* (2017) 237:67–70. doi: 10.1016/j.ijcard.2017.03.034
18. Basso C, Aguilera B, Banner J, Cohle S, d'Amati G, de Gouveia RH, et al. Guidelines for autopsy investigation of sudden cardiac death: (2017) update from the Association for European Cardiovascular Pathology. *Virchows Arch.* (2017) 471:691–705. doi: 10.1007/s00428-017-2221-0
19. Rella V, Parati G, Crotti L. Sudden cardiac death in children affected by cardiomyopathies: an update on risk factors and indications at transvenous or subcutaneous implantable defibrillators. *Front Pediatr.* (2020) 8:139 doi: 10.3389/fped.2020.00139
20. Skinner JR, Winbo A, Abrams D, Vohra J, Wilde AA. Channelopathies that lead to sudden cardiac death: clinical and genetic aspects. *Hear Lung Circ.* (2019) 28:22–3 doi: 10.1016/j.hlc.2018.09.007
21. Stiles MK, Wilde AA, Abrams DJ, Ackerman MJ, Albert CM, Behr ER, et al. 2020 APHRS/HRS Expert Consensus Statement on the Investigation of Decedents with Sudden Unexplained Death and Patients with Sudden Cardiac Arrest, and of Their Families. *J Arrhythmia.* (2021) 18:e1–e50. doi: 10.1002/joa3.12449
22. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, et al. 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.* (2015) 17:405–24. doi: 10.1038/gim.2015.30
23. Bustos BI, Pérez-Palma E, Buch S, Azócar L, Riveras E, Ugarte GD, et al. Variants in ABCG8 and TRAF3 genes confer risk for gallstone disease in admixed Latinos with Mapuche Native American ancestry. *Sci Rep.* (2019) 9:772. doi: 10.1038/s41598-018-35852-z
24. Lammert F, Gurusamy K, Ko CW, Miquel JF, Méndez-Sánchez N, Portincasa P, et al. Gallstones. *Nat Rev Dis Prim.* (2016) 2:16024. doi: 10.1038/nrdp.2016.24
25. Renner O, Lütjohann D, Richter D, Strohmeyer A, Schimmel S, Müller O, et al. Role of the ABCG8 19H risk allele in cholesterol absorption and gallstone disease. *BMC Gastroenterol.* (2013) 13:30. doi: 10.1186/1471-230X-13-30
26. Nicolas A, Fatima S, Lamri A, Bellili-Muñoz N, Halimi JM, Saulnier PJ, et al. ABCG8 polymorphisms and renal disease in type 2 diabetic patients. *Metabolism.* (2015) 64:713–9. doi: 10.1016/j.metabol.2015.03.005
27. Koeijvoets KCMC, van der Net JB, Dallinga-Thie GM, Steyerberg EW, Mensink RP, Kastelein JJP, et al. ABCG8 gene polymorphisms, plasma cholesterol concentrations, and risk of cardiovascular disease in familial hypercholesterolemia. *Atherosclerosis.* (2009) 204:453–8. doi: 10.1016/j.atherosclerosis.2008.09.018
28. Tada H, Okada H, Nomura A, Yashiro S, Nohara A, Ishigaki Y, et al. Rare and deleterious mutations in ABCG5/ABCG8 genes contribute to mimicking and worsening of familial hypercholesterolemia phenotype. *Circ J.* (2019) 83:1917–24. doi: 10.1253/circj.CJ-19-0317
29. Hong TT, Cogswell R, James CA, Kang G, Pullinger CR, Malloy MJ, et al. Plasma BIN1 correlates with heart failure and predicts arrhythmia in patients with arrhythmogenic right ventricular cardiomyopathy. *Hear Rhythm.* (2012) 9:961–7. doi: 10.1016/j.hrthm.2012.01.024
30. Jiang XX, Zhu YR, Liu HM, Chen SL, Zhang DM. Effect of BIN1 on cardiac dysfunction and malignant arrhythmias. *Acta Physiol.* (2020) 228:e13429 doi: 10.1111/apha.13429
31. Ohno S. The genetic background of arrhythmogenic right ventricular cardiomyopathy. *J Arrhythmia.* (2016) 32:398–403. doi: 10.1016/j.joa.2016.01.006
32. Iskandar K, Dwianingsih EK, Pratiwi L, Kalim AS, Mardhiah H, Putranti AH, et al. The analysis of DMD gene deletions by multiplex PCR in Indonesian DMD/BMD patients: The era of personalized medicine. *BMC Res Notes.* (2019) 12:704. doi: 10.1186/s13104-019-4730-1
33. Mohammed F, Elshafey A, Al-balool H, Alaboud H, Ali MA Ben, Baqer A, et al. Mutation spectrum analysis of Duchenne/ Becker muscular dystrophy in 68 families in Kuwait: The era of personalized medicine. *PLoS ONE.* (2018) 13:e0197205. doi: 10.1371/journal.pone.0197205
34. Huang H, Chen Y-Q, Fan L-L, Guo S, Li J-J, Jin J-Y, et al. Whole-exome sequencing identifies a novel mutation of *GPD1L* (R189X) associated with familial conduction disease and sudden death. *J Cell Mol Med.* (2017) 22:1350–4. doi: 10.1111/jcmm.13409
35. Liu M, Sanyal S, Gao G, Gurung IS, Zhu X, Gaconnet G, et al. Cardiac Na+ current regulation by pyridine nucleotides. *Circ Res.* (2009) 105:737–45. doi: 10.1161/CIRCRESAHA.109.197277
36. Valdivia CR, Ueda K, Ackerman MJ, Makielski JC. GPD1L links redox state to cardiac excitability by PKC-dependent phosphorylation of the sodium channel SCN5A. *Am J Physiol–Hear Circ Physiol.* (2009) 297:1446–52. doi: 10.1152/ajpheart.00513.2009
37. Hosseini SM, Kim R, Udupa S, Costain G, Jobling R, Liston E, et al. Reappraisal of Reported Genes for Sudden Arrhythmic Death: Evidence-Based Evaluation of Gene Validity for Brugada Syndrome. *Circulation.* (2018) 138:1195–205. doi: 10.1161/circulationaha.118.035070
38. Fan J, Ji CC, Cheng YJ, Yao H, Chen XM, Zheng ZH, et al. A novel mutation in GPD1-L associated with early repolarization syndrome via modulation of cardiomyocyte fast sodium currents. *Int J Mol Med.* (2020) 45:947–955. doi: 10.3892/ijmm.2020.4454
39. Boyden SE, Mahoney LJ, Kawahara G, Myers JA, Mitsuhashi S, Estrella EA, et al. Mutations in the satellite cell gene MEGF10 cause a recessive congenital myopathy with minicores. *Neurogenetics.* (2012) 13:115–24. doi: 10.1007/s10048-012-0315-z
40. Chen X, Wang X, Chen Q, Williamson V, van den Oord E, Maher BS, et al. MEGF10 Association with Schizophrenia. *Biol Psychiatry.* (2008) 63:441–8. doi: 10.1016/j.biopsych.2007.11.003
41. Logan CV, Lucke B, Pottinger C, Abdelhamed ZA, Parry DA, Szymanska K, et al. Mutations in MEGF10, a regulator of satellite cell myogenesis, cause early onset myopathy, areflexia, respiratory distress and dysphagia (EMARDD). *Nat Genet.* (2011) 43:1189–93. doi: 10.1038/ng.995
42. Shang Y, Smith S, Hu X. Role of Notch signaling in regulating innate immunity and inflammation in health and disease. *Protein Cell.* (2016) 7:159–74. doi: 10.1007/s13238-016-0250-0
43. Irtyuga O, Malashicheva A, Zhiduleva E, Freylikhman O, Rotar O, Bäck M, et al. NOTCH1 mutations in aortic stenosis: association with osteoprotegerin/RANK/RANKL. *Biomed Res Int.* (2017) 2017:6917907. doi: 10.1155/2017/6917907
44. Hopkins PN, Defesche J, Fouchier SW, Bruckert E, Luc G, Cariou B, et al. Characterization of Autosomal Dominant Hypercholesterolemia Caused by PCSK9 Gain of Function Mutations and Its Specific Treatment with Alirocumab, a PCSK9 Monoclonal Antibody. *Circ Cardiovasc Genet.* (2015) 8:823–31. doi: 10.1161/CIRCGENETICS.115.001129
45. Tarugi P, Averna M. Hypobetalipoproteinemia: genetics, biochemistry, and clinical spectrum. *Adv Clin Chem.* (2011) 54:81–107. doi: 10.1016/B978-0-12-387025-4.00004-2
46. Halmo SM, Singh D, Patel S, Wang S, Edlin M, Boons GJ, et al. Protein O-linked mannose β-1,4-N-acetylglucosaminyltransferase 2 (POMGNT2) is a gatekeeper enzyme for functional glycosylation of α-dystroglycan. *J Biol Chem.* (2017) 292:2101–09. doi: 10.1074/jbc.M116.764712

47. Endo Y, Dong M, Noguchi S, Ogawa M, Hayashi YK, Kuru S, et al. Milder forms of muscular dystrophy associated with POMGNT2 mutations. *Neurol Genet.* (2015) 1:e33. doi: 10.1212/NXG.0000000000000033
48. Cai R, Liu F, Hua C, Yu Z, Ramien M, Malic C. et al. A novel RASA1 mutation causing capillary malformation-arteriovenous malformation (CM-AVM): the first genetic clinical report in East Asia. *Hereditas.* (2018) 155:24. doi: 10.1186/s41065-018-0062-8
49. Koh HR, Lee YK, Ko SY, Shin SM, Han B-H. RASA1-related parkes weber syndrome in a neonate. *Neonatal Med.* (2018) 25:126–30. doi: 10.5385/nm.2018.25.3.126
50. Mahlke N, Dittmann S, Schulze-Bahr E, Ritz-Timme S, Hartung B. Sudden unexpected cardiac death and postmortem identification of a novel RYR2 gene mutation. *Int J Legal Med.* (2019) 133:1835–8. doi: 10.1007/s00414-019-02117-x
51. Aizawa Y, Fujisawa T, Katsumata Y, Kohsaka S, Kunitomi A, Ohno S, et al. Sex-dependent phenotypic variability of an SCN5a mutation: Brugada syndrome and sick sinus syndrome. *J Am Heart Assoc.* (2018) 7:e009387. doi: 10.1161/JAHA.118.009387
52. Brugada R, Campuzano O, Sarquella-Brugada G, Brugada J, Brugada P. Brugada syndrome. *Methodist Debakey Cardiovasc J.* (2014) 10:25–8. doi: 10.14797/mdcj-10-1-25
53. Ito S, Taketani T, Sugamori T, Okada T, Sato H, Adachi T, et al. A case of long QT syndrome having compound mutations of KCNH2 and SCN5A. *J Cardiol Cases.* (2012) 6:e170–e172. doi: 10.1016/j.jccase.2012.07.004
54. Riuró H, Beltran-Alvarez P, Tarradas A, Selga E, Campuzano O, Vergés M, et al. A missense mutation in the sodium channel β2 subunit reveals SCN2B as a new candidate gene for brugada syndrome. *Hum Mutat.* (2013) 34:961–6. doi: 10.1002/humu.22328
55. Faiella W, Bessoudo R. Cardiac manifestations in Emery–Dreifuss muscular dystrophy. *CMAJ.* (2018) 190:E1414–E1417. doi: 10.1503/cmaj.180410
56. Zhou C, Li C, Zhou B, Sun H, Koullourou V, Holt I, et al. Novel nesprin-1 mutations associated with dilated cardiomyopathy cause nuclear envelope disruption and defects in myogenesis. *Hum Mol Genet.* (2017) 26:2258–76. doi: 10.1093/hmg/ddx116
57. Fellmann F, van El CG, Charron P, Michaud K, Howard HC, Boers SN, et al. European recommendations integrating genetic testing into multidisciplinary management of sudden cardiac death. *Eur J Hum Genet.* (2019) 27:1763–73. doi: 10.1038/s41431-019-0445-y
58. Gollob MH, Blier L, Brugada R, Champagne J, ChauhanV, Connors S, et al. Recommendations for the use of genetic testing in the clinical evaluation of inherited cardiac arrhythmias associated with sudden cardiac death: Canadian cardiovascular society/Canadian heart rhythm society joint position paper. *Can J Cardiol.* (2011) 27:232–45. doi: 10.1016/j.cjca.2010.12.078
59. Moccia F, Negri S, Shekha M, Faris P, Guerra G. Endothelial Ca2+ signaling, angiogenesis and vasculogenesis: just what it takes to make a blood vessel. *Int J Mol Sci.* (2019) 20:3962. doi: 10.3390/ijms20163962
60. London B, Michalec M, Mehdi H, Zhu X, Kerchner L, Sanyal S, et al. Mutation in glycerol-3-phosphate dehydrogenase 1-like gene (GPD1-L) decreases cardiac Na+ current and causes inherited arrhythmias. *Circulation.* (2007) 116 2260–8. doi: 10.1161/CIRCULATIONAHA.107.703330
61. Snir AD, Raju H. Current controversies and challenges in brugada syndrome. *Eur Cardiol Rev.* (2019) 14:169–74. doi: 10.15420/ecr.2019.12.2
62. Darbar D, Yang T, Churchwell K, Wilde AAM, Roden DM. Unmasking of Brugada syndrome by lithium. *Circulation.* (2005) 112:1527–31. doi: 10.1161/CIRCULATIONAHA.105.548487
63. Masrur S, Memon S, Thompson PD. Brugada syndrome, exercise, and exercise testing. *Clin Cardiol.* (2015) 38:323–6. doi: 10.1002/clc.22386
64. Roomi SS, Ullah W, Abbas H, Abdullah H, Talib U, Figueredo V. Brugada syndrome unmasked by fever: a comprehensive review of literature. *J Community Hosp Intern Med Perspect.* (2020) 10:224–8. doi: 10.1080/20009666.2020.1767278
65. Campuzano O, Fernandez-Falgueras A, Sarquella-Brugada G, Cesar S, Arbelo E, García-Álvarez A, et al. Personalized interpretation and clinical translation of genetic variants associated with cardiomyopathies. *Front Genet.* (2019) 10:450. doi: 10.3389/fgene.2019.00450
66. Verkerk AO, Amin AS, Remme CA. Disease modifiers of inherited SCN5A channelopathy. *Front Cardiovasc Med.* (2018) 5:137. doi: 10.3389/fcvm.2018.00137
67. Lee SR, Han J. Mitochondrial mutations in cardiac disorders. *Adv Exp Med Biol.* (2017) 982:81–111. doi: 10.1007/978-3-319-55330-6_5
68. Brinkmann B. Harmonization of medico-legal autopsy rules. Committee of Ministers. Council of Europe. *Int J Legal Med.* (1999) 113:1–14. doi: 10.1007/s004140050271 PMID: 10654232.
69. Lahrouchi N, Raju H, Lodder EM, Papatheodorou E, Ware JS, Papadakis M, et al. Utility of post-mortem genetic testing in cases of sudden arrhythmic death syndrome. *J Am Coll Cardiol.* (2017) 2 69:2134–45. doi: 10.1016/j.jacc.2017.02.046
70. Papadakis M, Papatheodorou E, Mellor G, Raju H, Bastiaenen R, Wijeyeratne Y, et al. The diagnostic yield of brugada syndrome after sudden death with normal autopsy. *J Am Coll Cardiol.* (2018) 71:1204–14. doi: 10.1016/j.jacc.2018.01.031
71. Medeiros Domingo A, Bolliger S, Gräni C, Rieubland C, Hersch D, Asatryan B, et al. Swiss Working Group on Sudden Cardiac Death. Recommendations for genetic testing and counselling after sudden cardiac death: practical aspects for Swiss practice. *Swiss Med Wkly.* (2018) 148:w14638. doi: 10.4414/smw.2018.14638
72. Valencia OM, Samuel SE, Viscusi RK, Riall TS, Neumayer LA, Aziz H. The role of genetic testing in patients with breast cancer: a review. *JAMA Surg.* (2017) 152:589–94. doi: 10.1001/jamasurg.2017.0552

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 in Genetics

REVIEW
published: 23 March 2021
doi: 10.3389/fgene.2021.638730

The History of Gene Hunting in Hereditary Spinocerebellar Degeneration: Lessons From the Past and Future Perspectives

Ashraf Yahia[1,2,3,4] *and Giovanni Stevanin*[3,4]*

[1]Department of Biochemistry, Faculty of Medicine, University of Khartoum, Khartoum, Sudan, [2]Department of Biochemistry, Faculty of Medicine, National University, Khartoum, Sudan, [3]Institut du Cerveau, INSERM U1127, CNRS UMR7225, Sorbonne Université, Paris, France, [4]Ecole Pratique des Hautes Etudes, EPHE, PSL Research University, Paris, France

Hereditary spinocerebellar degeneration (SCD) encompasses an expanding list of rare diseases with a broad clinical and genetic heterogeneity, complicating their diagnosis and management in daily clinical practice. Correct diagnosis is a pillar for precision medicine, a branch of medicine that promises to flourish with the progressive improvements in studying the human genome. Discovering the genes causing novel Mendelian phenotypes contributes to precision medicine by diagnosing subsets of patients with previously undiagnosed conditions, guiding the management of these patients and their families, and enabling the discovery of more causes of Mendelian diseases. This new knowledge provides insight into the biological processes involved in health and disease, including the more common complex disorders. This review discusses the evolution of the clinical and genetic approaches used to diagnose hereditary SCD and the potential of new tools for future discoveries.

Keywords: hereditary spastic paraplegia, hereditary cerebellar ataxia, spinocerebellar ataxia, spinocerebellar degeneration, gene discovery, diagnosis, neurogenetics

OPEN ACCESS

Edited by:
Ayman Hussein, University of Khartoum, Sudan

Reviewed by:
Zhipeng Liu, Purdue University, United States
Bolin Chen, Northwestern Polytechnical University, China

****Correspondence:***
Giovanni Stevanin
giovanni-b.stevanin@inserm.fr

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

Received: *07 December 2020*
Accepted: *02 March 2021*
Published: *23 March 2021*

Citation:
Yahia A and Stevanin G (2021) The History of Gene Hunting in Hereditary Spinocerebellar Degeneration: Lessons From the Past and Future Perspectives. Front. Genet. 12:638730. doi: 10.3389/fgene.2021.638730

INTRODUCTION

Hereditary forms of spastic paraplegia (SPG), cerebellar ataxia (CA), spastic ataxia, and spinocerebellar ataxia (SCA) are distinct clinical entities caused by related mechanisms and encompassing a continuum of phenotypes known as hereditary spinocerebellar degenerations (SCDs; Parodi et al., 2018). To date, they have been shown to be caused by pathogenic variants in more than 200 genes and loci inherited through X-linked, autosomal recessive, autosomal dominant, and mitochondrial patterns, and occasionally due to *de novo* events (Synofzik and Schüle, 2017; Parodi et al., 2018; Boutry et al., 2019; Manto et al., 2020; see **Supplementary Table 1**). SCDs are characterized clinically by cerebellar ataxia and/or spastic limbs, often complicated by other neurological or extra-neurological features (Parodi et al., 2018). The genetic and phenotypic heterogeneity of SCDs complicates their diagnosis (Elsayed et al., 2019). Moreover, the overlap between the clinical presentations of SCD and those of other genetic and non-genetic neurological conditions further complicates diagnosing these diseases (Elsayed et al., 2019). These diagnostic hurdles require the use of constantly evolving tools and approaches.

Spinocerebellar degenerations are estimated to affect 1:10,000 individuals worldwide, and thus fall within the category of rare diseases (Ruano et al., 2014; Nguengang Wakap et al., 2020). More than 70% of rare diseases are genetic, the majority of which affect the nervous system (Lee et al., 2020; Nguengang Wakap et al., 2020). From a biological perspective, Mendelian diseases can be viewed as natural "knockouts" or gains of function of a gene, a pathway, or a cellular process that can provide insight into complex biological systems and the more common complex diseases (Lee et al., 2020). From epidemiological and clinical perspectives, Mendelian diseases affect >2.5% of the human population (Nguengang Wakap et al., 2020), and therefore reaching a clinical diagnosis is valuable for the patients affected by these conditions and their families (Clark et al., 2018; Wright et al., 2018). The discovery of the cause of a Mendelian disease has, on some occasions, had a snowball effect and led to the discovery of more diseases (van der Knaap et al., 2002; Abou Jamra et al., 2011). Finding more causative genes helps to complete the puzzle and increases the probability of identifying biomarkers and therapeutic targets, hopefully of interest for multiple clinical-genetic entities.

The identification of SCD genes has passed several milestones, and multiple tools have been used in this quest (**Figure 1**; **Supplementary Table 1**). The evolution of the number of SCD entities has greatly benefited from improvements in terms of the tools used, their cost, the clinical phenotyping, and the availability of multiple samples in affected families. Several reviews have addressed the best practices to diagnose SCD subtypes at different periods (Harding, 1993; Fink, 1997; Brusse et al., 2007; De Silva et al., 2019; Shribman et al., 2019). Here, we review the evolution of the approaches used to identify novel SCD genes over time and discuss the possibility of combining some of these approaches with novel tools to increase the current SCD diagnostic rates and answer the current challenges in these diseases.

TOOLS USED TO DISCOVER SCD GENES IN THE PAST

Linkage Analysis and Homozygosity Mapping

Whole genome linkage analyses, and homozygosity mapping in consanguineous cases, were the primary approaches for identifying genes and loci associated with Mendelian diseases (Lipner and Greenberg, 2018). These methods were developed in the 1980s thanks to improvements in detection of genetic markers and their assignment in chromosomal maps. These strategies were and are still, suitable for family-based approaches and realizable only in large pedigrees, although single nuclear sibships can be used in autosomal recessive consanguineous families (Lipner and Greenberg, 2018).

The linkage analysis concept relies on genetic variant co-segregation with nearby markers at the disease locus, as recombination is less likely to separate adjacent loci (Lipner and Greenberg, 2018). Microsatellite markers and array-based single nucleotide polymorphism (SNP) markers have been extensively used in SCD gene discovery. These markers replaced the restriction-fragment-length polymorphisms (RFLP) and the variable numbers of tandem repeats (VNTR) used in the 1980s. SNP markers are now commonly used either with DNA microarray or extracted from exome/genome sequencing data (Guergueltcheva et al., 2012; Novarino et al., 2014).

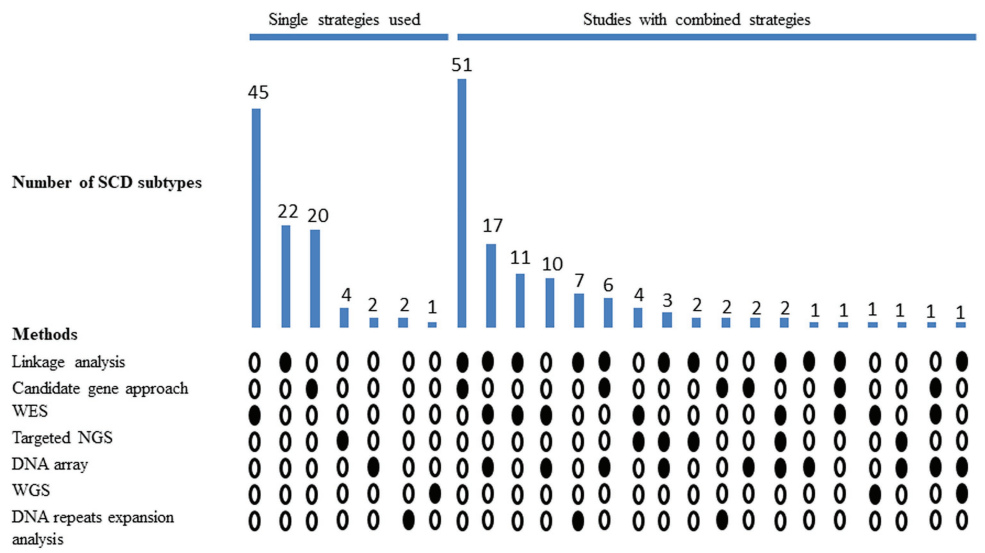


FIGURE 1 | Methods used in diagnosing novel subtypes of hereditary spinocerebellar degeneration (SCD). Filled circles represent the methods used in identifying SCD subtypes, and the bar chart shows the number of SCD subtypes identified by each method or combination of methods. Empty circles represent the unused methods. SCD, hereditary spinocerebellar degeneration; WES, whole-exome sequencing; NGS, next-generation sequencing; and WGS, whole-genome sequencing.

Among the 56% (123/220) of known SCD subtypes identified by approaches that involved linkage analysis, a good example of its diagnostic utility is the study by Novarino et al. (2014) that reported 15 novel causative genes in consanguineous families with various forms of SPG; they used SNP marker-based homozygosity to filter and prioritize the exome data. Indeed, most of the linkage-based SCD studies identified the responsible genes by coupling linkage analysis with candidate genes screening (51/220) or, more recently through biased (2/218) or unbiased next-generation sequencing (NGS; 31/220). For example, Guergueltcheva et al. (2012) pinpointed the mutation responsible for autosomal recessive spinocerebellar ataxia (SCAR) type 13 (SCAR13; OMIM # 614831) using linkage studies and mutational screening. There are still 22 SCD loci for which the causative gene has not been found and may have been missed if the causative variant is in non-coding regions, often remaining unexplored. Indeed, chromosomal location through linkage studies was critical for the identification of several of the nucleotide repeat expansion disorders such as the recently reported "cerebellar ataxia neuropathy and vestibular areflexia" syndrome (CANVAS) due to intronic repeat expansions in *RFC1* (Cortese et al., 2019).

The use of microsatellite or SNP markers is not confined to identifying disease loci but can extend to establishing the founder effects of disease-causing variants (Engert et al., 2000; Zivony-Elboum et al., 2012). This is clearly illustrated in studies performed to determine the ancestral origin of the pathological expansions in SCA3/Machado-Joseph disease (OMIM # 109150; Martins et al., 2007; Sharony et al., 2019).

Repeat Expansion Detection Methods

Expansion of simple DNA sequence repeats causes more than 40 diseases, most with phenotypes affecting the central nervous system (Paulson, 2018; Rodriguez and Todd, 2019). They were first identified in a series of SCA loci in the 1990s following linkage studies and DNA/cDNA cloning strategies. These included the studies performed by Koide et al. (1994), David et al. (1997), and Koob et al. (1999) that directly screened patients with an autosomal dominant pattern of inheritance for repeat expansions and identified the genes responsible for dentatorubral-pallidoluysian atrophy (DRPLA; OMIM # 125370), SCA7 (OMIM # 164500), and SCA8 (OMIM # 608768). These discoveries benefited from the development of repeat expansion detection methods that used oligonucleotide probes containing triplet repeats and antibodies against polyglutamine tracks (Li et al., 1993; Trottier et al., 1995; Bahlo et al., 2018; Mitsuhashi and Matsumoto, 2020). The 1990s also witnessed the identification of the cause of Friedreich ataxia (OMIM # 229300), the most frequent autosomal recessive cerebellar ataxia commonly caused by intronic repeat expansions in the *FXN* gene (Campuzano et al., 1996).

There has been renewed interest in repeat expansion in recent years, with the development of diagnostic applications of PCR-based approaches (Cagnoli et al., 2018) and the improvement of algorithms for short-read sequence analysis. Recently, a combination of linkage analysis and short-read whole-genome sequencing (WGS) identified repeat expansions in *DAB1* and *RFC1* causing SCA37 (OMIM # 615945) and CANVAS (OMIM # 614575), respectively (Seixas et al., 2017; Cortese et al., 2019). In all likelihood, long-read sequencing methods will soon further increase the power of detecting repeat expansions (Bahlo et al., 2018; De Roeck et al., 2019; Mitsuhashi and Matsumoto, 2020). Up to now, methods for detecting DNA repeat expansions were applied in discovering 5% (11/220) of the SCD subtypes, mostly (9/11) when coupled with whole genome linkage analysis.

Candidate Gene Approaches

Prioritizing genes is a complementary but crucial step in all the approaches taken to identify SCD genes. This prioritization usually depends on the gene expression, the gene's function or its product's function, and the phenotype of patients and/or animal models. Interestingly, a straightforward candidate gene approach was the main methodology in identifying 9.1% (20/220) of SCD subtypes.

There are multiple mouse models of neurodegeneration, particularly with cerebellar features, and they can be very informative in guiding candidate gene selection. Escayg et al. (2000) screened patients with neurological phenotypes for mutations in the *CACNB4* gene after they had identified bi-allelic mutations in *CACNB4* in a mouse model presenting with ataxia and epilepsy, and they identified *CACNB4* as the cause of type 5 episodic ataxia (EA5; OMIM # 613855; Burgess et al., 1997). Loss of function variants in the *NHE1* gene coding the sodium-hydrogen exchanger 1 (NHE1) are associated with Lichtenstein-Knorr syndrome (OMIM# 616291), a cerebellar syndrome complicated by sensorineural hearing loss (Guissart et al., 2015). Interestingly, the Lichtenstein-Knorr syndrome's phenotype is partially observed in the mouse knockout for *NHE1* and for *CHP1* as well, coding an NHE1 interactor protein (Bell et al., 1999; Liu et al., 2013). These observations prompted Mendoza-Ferreira et al. (2018) to prioritize a frame-shift variant in *CHP1* observed on whole-exome sequencing (WES) data of Moroccan patients with complex cerebellar ataxia and intellectual disability, the first cases of spastic ataxia type 9 (OMIM # 618438). Human genetics can also benefit animal genetics. This is shown by the identification of a *KIFC* variant in Charolais cattle developing a spastic ataxia phenotype by prioritization of the WGS data on this gene after it was involved in a similar human disorder (Duchesne et al., 2018).

The clinical and biological features of the patients have also helped to point to specific candidate genes. The patients' clinical presentation and enzymatic assays significant for ubiquinone deficiency led to the implication of pathogenic mutations in *ADCK3* in SCAR type 9 (SCAR9; OMIM # 612016; Mollet et al., 2008). The patients' phenotype and hormonal assays implying thyroid hormone resistance led to the implication of *SLC16A2* mutations in X-linked SPG22 (OMIM # 300523; Friesema et al., 2004). Similarly, the clinical phenotype led to the discovery of some SCD subtypes caused by mitochondrial alterations (Holt et al., 1990; Winterthun et al., 2005; Sarzi et al., 2007).

Many attempts to identify the SCD genes' interactors have been performed using yeast two-hybrid screening (Lim et al., 2006) and *in silico* networks (Novarino et al., 2014). The latter study identified three causative genes in HSP

families using an HSP *in silico* functional network (Novarino et al., 2014). Based on this "HSP network," the authors selected three candidate genes and tested them by direct sequencing in a large series of HSP cases, a strategy that led to *REEP2*, *MAG*, and *BICD2* being identified as novel HSP genes (Novarino et al., 2014). Along the same lines, Słabicki et al. (2010) adopted another approach in identifying the gene responsible for SPG48; they first performed RNAi genome-wide screening in HeLa cells for genes involved in DNA double-strand breaks and identified the *AP5Z1* gene among the modifier genes. *AP5Z1* was shown to interact with spastascin and spastizin, the proteins involved in SPG11 and SPG15, respectively. Słabicki et al. (2010) then sequenced the *AP5Z1* gene in 166 patients with unexplained HSP and identified the first SPG48 case. These studies demonstrated the usefulness of knowing the interactors of the SCD proteins as they are likely the source of potential variants in patients.

Knowledge of the physiopathology at the cell level can also guide the candidate gene analysis. Jen et al. (2005) screened a cohort of patients with unexplained episodic ataxia and hemiplegic migraine for mutations in *SLC1A3*, which encodes a glutamate transporter essential for removing glutamate from the synaptic cleft. Glutamate is the most abundant amino acid neurotransmitter in the central nervous system. In their study, Jen et al. (2005) identified a heterozygous mutation in *SLC1A3* in a patient with episodic ataxia called later episodic ataxia (EA) type 6 (EA6; OMIM # 612656).

Other studies have implicated genes known for their association with specific neurological phenotypes in developing new SCD forms based on phenotypic similarities between the old and the new phenotypes. These include studies that implicated *SCN2A* in EA9 (OMIM # 618924; Liao et al., 2010), *GJC2* in SPG44 (OMIM# 613206; Orthmann-Murphy et al., 2009), *PGN* in the autosomal dominant SPG7 (Sánchez-Ferrero et al., 2013), *FMR1* in fragile X tremor/ataxia syndrome (FXTAS, OMIM # 300623; Hagerman et al., 2001), and *DARS2* in hereditary SPG (Lan et al., 2017).

DNA Microarrays

DNA microarrays are used in clinical settings to investigate copy number variations (CNVs) and can also help determine SNP genotypes for linkage analysis or homozygosity mapping strategies (Levy and Burnside, 2019). DNA microarrays used for CNV detection are the first-tier tests for diagnosing neurodevelopmental disorders, disease entities that include intellectual disabilities, developmental delay disorders, autism spectrum disorders, and disorders with multiple congenital anomalies (Jang et al., 2019). However, a recent meta-analysis suggested that WES outperforms microarrays in diagnosing neurodevelopmental disorders (Srivastava et al., 2019). Furthermore, in a recent study performed over a 10-year period, Ciaccio et al. (2020) discouraged the use of chromosomal microarrays as the first-tier test for diagnosing neurodevelopmental disorders with ataxia.

DNA arrays were the sole genetic screening tool used in discovering two SCD genes. Moreno-De-Luca et al. (2011) used a custom-designed oligonucleotide array in discovering *AP4E1*, the gene responsible for SPG51 (OMIM# 613744). Moreno-De-Luca et al. (2011) used a genome-wide microarray in their study to identify a region of interest, and then further analyzed the identified region with a higher-resolution microarray. In another study, Utine et al. (2013) used an SNP array in discovering the gene causing SCAR18 (OMIM # 616204). Guided by array-based homozygosity mapping, Utine et al. (2013) identified deletions in the third and fourth exons of *GRID2* in three Turkish siblings. They confirmed the absence of these exons in the patients using real-time PCR (RT-PCR).

Coupling arrays to other genetic screening tools has led to the identification of 20% (44/220) of all currently known SCD genes. More than half of the genes thus identified (27/44) were discovered using approaches that involved microarrays (mainly for linkage studies) and WES.

Next-Generation Sequencing Approaches

The advent of NGS has revolutionized the field of medical genetics as it enabled parallel sequencing of massive targets in a short duration of time and for a plummeting cost (Boycott et al., 2017; Mazzarotto et al., 2020). NGS is classified according to the targeted sequences. The targeted sequence in WGS is the whole genome, in WES it is the coding part of the genome, and in targeted NGS (TS) it is a custom-ordered set of genes and sequences (Gao and Smith, 2020). NGS is also classified as long-read sequencing and short-read sequencing according to the length of the sequenced fragments, also known as the sequence reads. The length of the sequence read in short-read sequencing is ~75–300 base-pairs, while in long-read sequencing; it is equal to the length of the template sequence, at least in theory (Caspar et al., 2018).

Next-generation sequencing is extensively and routinely used in diagnosing Mendelian diseases, including those with neurological phenotypes (Bhatia et al., 2020; Rouleau et al., 2020). To date, 47.3% (104/220) of SCD genes have been discovered using NGS-based approaches, and the slope of the curve in SCD gene discovery drastically changed from 2009, when this technique was made available (**Figure 2**).

Targeted NGS

Targeted NGS is the most commonly used NGS tool in clinical practice (Gao and Smith, 2020). In the clinical context, TS investigates the roles of a predetermined set of genes in a clinical phenotype (Gao and Smith, 2020). In the setting of research and gene discovery, researchers use TS to focus on specific regions and sequences of interest determined by the clinical data or other diagnostic modalities.

Targeted NGS was involved in the discovery of 7.3% (16/220) of known SCD subtypes, mostly (73%) when coupled with other diagnostic modalities. TS was the primary genetic test used in four studies that identified four new SCD subtypes. In the first, Zanni et al. (2012) sequenced all the exons in the X chromosome in a family with X-linked cerebellar ataxia and identified a missense mutation in the *ATP2B3* gene. Pathogenic mutations in *ATP2B3* cause X-linked SCA type 1 (SCAX1; OMIM # 302500; Zanni et al., 2012). In the other three studies, TS identified mutations in *SPTBN2*, *KIF1A*, and

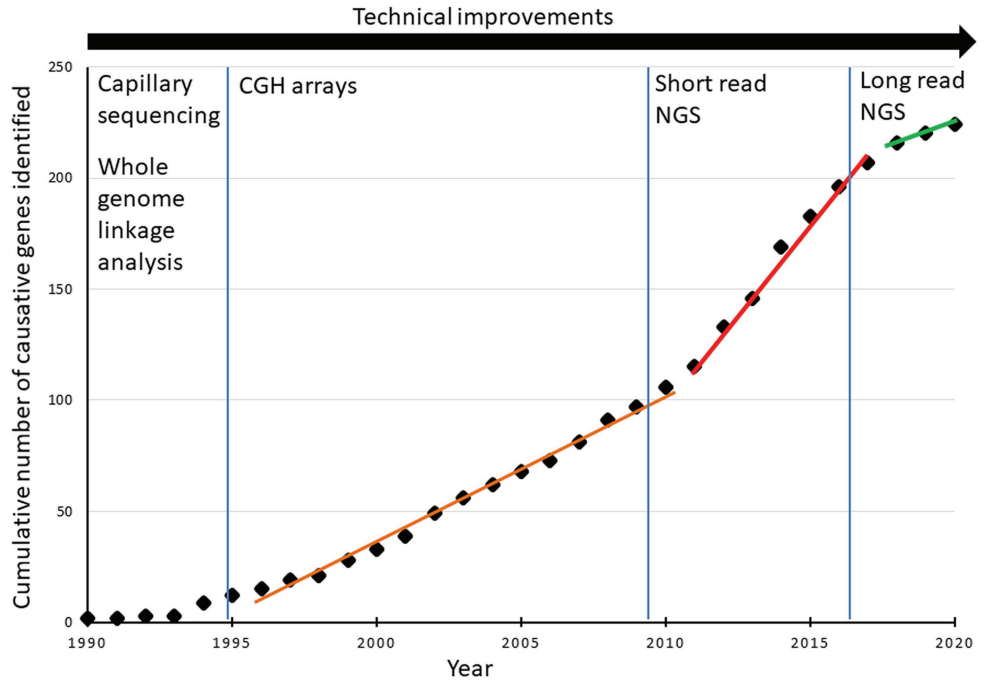


FIGURE 2 | Cumulative evolution of the number of subtypes of hereditary SCD identified per year.

TUBB4A as the causes of SCAR14 (OMIM # 615386), the autosomal dominant form of SPG30 (OMIM # 610357), and *TUBB4A*-associated SPG, respectively (Lise et al., 2012; Ylikallio et al., 2015; Sagnelli et al., 2016). Pathogenic mutations in *SPTBN2*, *KIF1A*, and *TUBB4A* were already known to cause SCA5 (OMIM # 600224), autosomal recessive SPG30 (OMIM # 610357), and both autosomal dominant type 4 dystonia (OMIM # 128101) and type 6 hypomyelination leukodystrophy (OMIM # 612438), respectively (Ikeda et al., 2006; Erlich et al., 2011; Hersheson et al., 2013; Simons et al., 2013). Thus, the three subsequent studies added SCD to the clinical phenotypes associated with genes previously known to cause diseases other than SCD, a situation becoming more frequent in recent years than real gene identification in SCD as shown by the change in the slope of the curve in **Figure 2** since 2017.

Targeted NGS is mainly used in genetic diagnosis in clinical practice as it allows focused analysis, thereby limiting incidental findings (Elsayed et al., 2016; Morais et al., 2017). In addition, given that it is optimized for a specific set of genes, sequence capture is usually homogeneous, allowing CNV detection through coverage analysis (Moreno-Cabrera et al., 2020).

Whole-Exome Sequencing

Whole-exome sequencing covers ~1–2% of the human genome and is extensively used for diagnosing Mendelian diseases (Hansen et al., 2020; Rouleau et al., 2020). The 3' and 5' untranslated regions (UTRs) are also included in most WES kits, at a variable level according to the company; however, variants in these regions are often difficult to analyze and interpret (Devanna et al., 2018). To date, mutations in ~20% of the protein-coding genes are known associated with a disease trait (Posey, 2019).

There are two views on whether TS or WES is the best first-tier test for diagnosing patients with SCDs (Galatolo et al., 2018; Shribman et al., 2019). TS has a lower cost, higher depth, lower analysis time, and fewer incidental findings (Gao and Smith, 2020; Mazzarotto et al., 2020; Platt et al., 2020). WES has higher coverage and outperforms TS in discovering new Mendelian disease-causing genes (Gao and Smith, 2020; Mazzarotto et al., 2020; Platt et al., 2020). WES was the only genetic screening tool used in discovering 20.5% (45/220) of known SCD subtypes and an additional 21.4% (47/220) when coupled with other technics. Focused exome sequencing (FES) is a WES-based approach that possesses some of the advantages of TS and WES (Pengelly et al., 2020). FES has higher coverage than TS, as it currently targets ~5,000 disease-associated genes, ~20% of the sequences targeted by WES (Pengelly et al., 2020). FES has greater depth and produces fewer incidental findings compared to WES but has a lower potential for discovering new SCD genes (Pengelly et al., 2020).

The diagnostic yield of WES is higher in familial cases compared to sporadic cases (Coutelier et al., 2018; Ngo et al., 2020). However, this generalization has some exceptions. For example, Fogel et al. (2014) reported a success rate of 73% (22/30) in diagnosing patients with sporadic cerebellar ataxia. Reanalysis of the undiagnosed exomes and coupling WES with other genetic approaches, such as mRNA sequencing and CNV detection, also increase WES's diagnostic yield (Deelen et al., 2019; Jalkh et al., 2019; Fung et al., 2020; Matalonga et al., 2020). CNVs are common in some SCD subtypes, e.g., SPG4 (OMIM

182601) and SCA15 (OMIM # 606658; Depienne et al., 2007; Hsiao et al., 2017). They can be looked for in short-reads sequencing data, but with variable rates according to the used algorithms and the library preparation, making comparative genomic hybridization (CGH) arrays or multiplex-ligation-probe-amplification (MLPA) favored to detect CNV. One study reported a rise of 7% in the diagnostic success rate upon reanalyzing the WES data of a cohort of SCD patients (Ngo et al., 2020). Recent reports advocate for a periodic reanalysis of WES data (Jalkh et al., 2019; Liu et al., 2019; Fung et al., 2020; Ngo et al., 2020). The 5-year cumulative increase in the diagnostic success rate upon WES data reanalysis ranges from 12 to 22% (Liu et al., 2019).

Whole-Genome Sequencing

Whole-genome sequencing is the most comprehensive tool to study the genome (Gao and Smith, 2020). A significant advantage of WGS is its ability to interrogate intergenic and intronic regions that constitute most of the human genome. This ability is a double-edged sword as the interpretation of intergenic and intronic variants is hard compared to the interpretation of exonic variants (Gao and Smith, 2020). For that reason, most of the time, WGS is interrogated for exonic sequences first, and often, limited to them (Rexach et al., 2019). WGS has a lower but more uniform coverage compared to WES (Posey, 2019). This uniformity in coverage of exonic sequences has resulted in better detection of GC-rich regions, e.g., at the first exons, and CNVs, compared to WES (Posey, 2019; Rexach et al., 2019).

To date, WGS has been involved in the discovery of three SCD genes. Melo et al. (2015) identified a homozygous 216-bp deletion in the non-coding sequence upstream to the *KLC2* gene in patients with SPG, optic atrophy, and neuropathy (SPOAN). Melo et al. (2015) identified this 216-bp deletion in screening a series of 275 patients (273 from Brazil and two from Egypt) with SPOAN and validated its pathogenicity in a zebrafish model. The candidate region containing *KLC2* was refined using linkage analysis and WES (Macedo-Souza et al., 2009; Melo et al., 2015). Similarly, Cortese et al. (2019) identified bi-allelic intronic expansions in *RFC1* in 56 patients from the United Kingdom, Italy, and Brazil with the CANVAS phenotype using a WGS-based approach focused to the linked-chromosomal region. In another study, Wagner et al. (2019) using WGS, identified a biallelic splice-site variant in the *RNF170* gene in a family with early-onset HSP complicated with peripheral neuropathy. Wagner et al. (2019) corroborated their finding by identifying additional patients with pathogenic *RNF170* variants on the GENESIS platform[1] of shared sequencing data (Gonzalez et al., 2015).

Studies that follow the WGS-based approach to identify novel genes generally use other techniques, such as whole-genome linkage analysis, to focus the search for candidate variants. Indeed, the bottleneck of the WGS-based approach is the large number of variants that can be identified. Multiomics integration will ameliorate the interpretation of the functional and regulatory effects of variants identified by NGS, especially WGS (Agrawal et al., 2014; Posey, 2019). Furthermore, long-read WGS is probably a promising new tool for diagnosing genetic diseases for the near future with great hope on detection of nucleotide repeats as well (Posey, 2019; Wenger et al., 2019).

[1]https://www.tgp-foundation.org/genesis-log-in

DISCUSSION AND FUTURE DIRECTIONS

The Need for Combined Approaches

Genome-wide approaches have proved the most successful in identifying new SCD subtypes, having identified 86.4% (190/220) of the known SCD subtypes. These approaches have made use of linkage analysis, WES, WGS, and DNA microarrays. Linkage analysis with candidate gene screening dominated the discovery of SCD genes in the past, often thanks to prior knowledge of these genes' functions and expression profiles, their products, the pathways they serve, or pertinent animal models. WES is now dominating SCD gene discovery and we can expect that WGS will dominate the discovery of the genes causing Mendelian diseases in the future, probably within integrated multi-omics strategies to reduce the number of candidate variants (Posey, 2019; Kerr et al., 2020).

Indeed, approaches that utilized more than one genetic screening tool identified 55.9% (123/220) of SCD subtypes compared to the 43.2% (95/220) identified by single tools. Historically, the most successful approaches in identifying SCD genes were those involving linkage analysis and exome sequencing. However, the use of NGS to diagnose rare diseases has been accompanied by a shift away from studying single large multiplex pedigrees with a restricted phenotype toward studying phenotypically diverse multiple small pedigrees and sporadic cases as well (Wright et al., 2018; Claussnitzer et al., 2020). In our opinion, WES and SNP arrays are the best combination to accelerate the discovery of SCD genes, since SNP arrays can be used in detecting CNVs, chromosomal aberrations, and linkage/homozygosity regions (Levy and Burnside, 2019). SNP arrays are also the most widely used tools for studying complex traits by genome-wide association studies (Tam et al., 2019). WGS could achieve all these goals, in an even better way, and eliminate the need to couple WES and SNP arrays (**Figure 3**). However, at the present time, the use of WGS on a broad scale is prohibited by its relatively high costs and the technical difficulties in storing, analyzing, and sharing WGS data (Suwinski et al., 2019; Tam et al., 2019).

Emerging Approaches

Several tools and developments in the genome-wide approaches have emerged to have inherent potentials for augmenting the discovery of new Mendelian diseases. These advancements include, but not limited to, the optical mapping methods for detecting genomic rearrangements (e.g., Bionano optical mapping), long-read genomic sequencing valuable for detecting copy number variants and nucleotide expansions (e.g., Oxford Nanopore or PacBio technologies), and mosaicism detection tools (e.g., MosaicForecast; Lam et al., 2012; Jain et al., 2018;

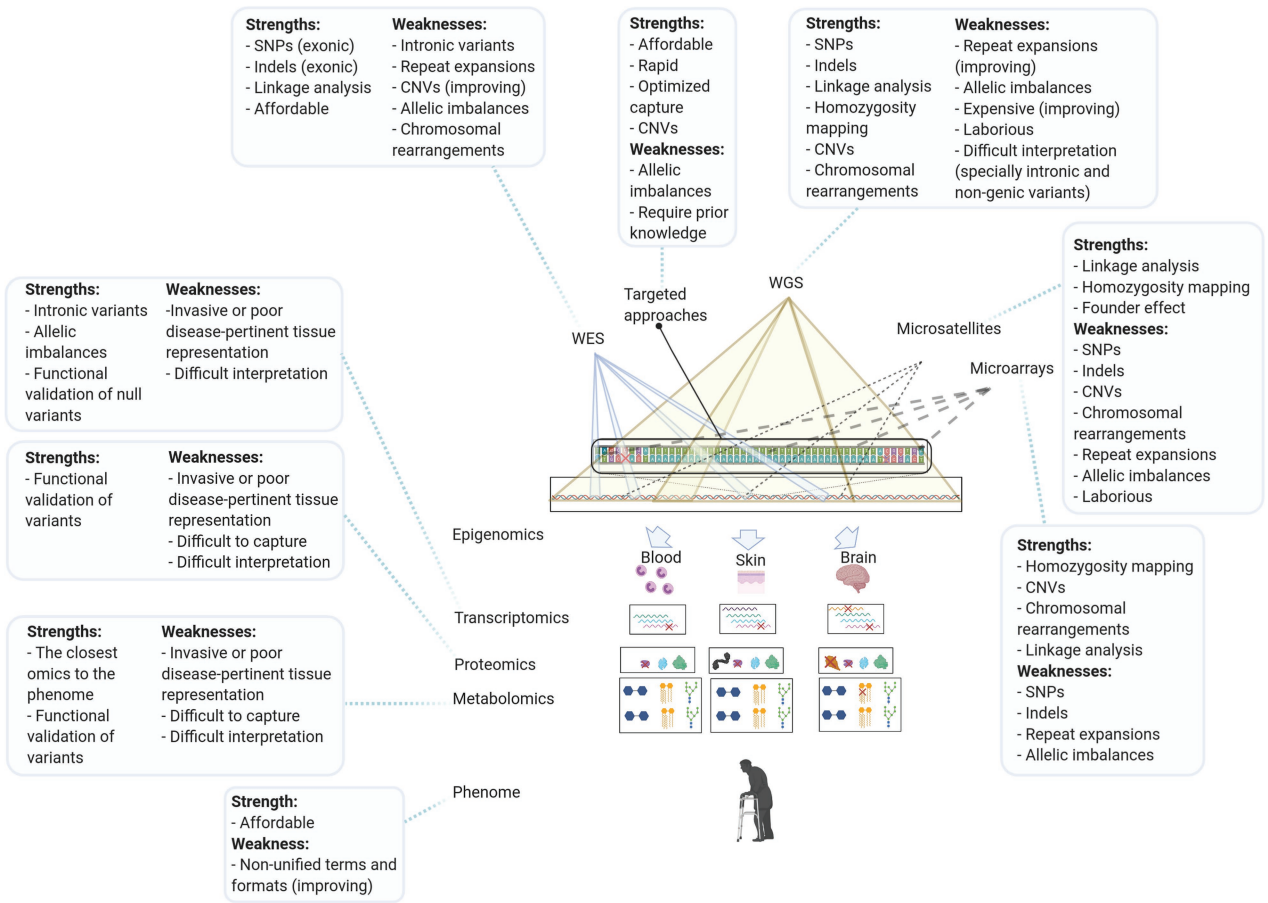


FIGURE 3 | The strengths and weaknesses of the methods used in diagnosing subtypes of hereditary SCD. WES, whole-exome sequencing; WGS, whole-genome sequencing; Targeted approaches, targeted next-generation sequencing, candidate gene approaches, and repeats expansion detection methods; SNPs, single nucleotide polymorphisms; and CNVs, copy number variations.

Dou et al., 2020). Structural rearrangements, CNVs, sequence expansions, and somatic mutations account for a significant proportion of the missing heritability in Mendelian diseases (Maroilley and Tarailo-Graovac, 2019).

The comprehensiveness of WES and WGS should not lead us to underestimate the role of proper clinical phenotyping in diagnosing genetic diseases (Pena et al., 2018). Tools that incorporate phenotypic data in NGS analysis pipelines are under development and have been used with success (Thuriot et al., 2018; Zhao et al., 2020). These include tools that empower the clinicians to have a central role in prioritizing variants, e.g., tools that apply gene-pertinence metrics (Segal et al., 2020). Proper clinical phenotyping is of value to increase the diagnostic rate of rare diseases (Bhatia et al., 2020). The high degree of overlap between various neurodegenerative disorders, including SCD, calls for better documentation, and follow-up for patients (Klebe et al., 2015; Parodi et al., 2018). This is clearly seen by the constant identification of more genes involved in other diseases and found mutated in SCD as well. The reverse is also true and illustrates the clinical overlap with various diseases. International registries, such as SCA Global and ARCA Global[2] and SPATAX,[3] have been launched to meet the need for better documentation of SCD cases. A prerequisite for international patient registries, besides appropriate research ethics policies, is the development of unified common disability scales to facilitate analysis and comparison (Trouillas et al., 1997; Lin et al., 2020). Many disability scales have been developed for SCD, the most commonly used scales being the scale for the assessment and rating of ataxia (SARA) and the International Cooperative Ataxia Rating Scale (ICARS; Perez-Lloret et al., 2020). Disability scales do not only measure disability; unlike what the name implies, but they can also measure the response to therapeutic interventions including attempts to preserve motor capacities of patients through alternative treatment options, e.g., dance, video game-based coordinative training (exergames), etc. (Ilg et al., 2012).

On the other hand, clinical phenotypes will, in many cases, evolve over time, and this needs to be considered when analyzing NGS data (Liu et al., 2019; Bhatia et al., 2020).

[2]http://ataxia-global-initiatives.net/
[3]https://spatax.wordpress.com/

Artificial intelligence tools have recently been employed in phenotyping, known as next-generation phenotyping, with good results, especially in neurological diseases (Elmas and Gogus, 2020; Pode-Shakked et al., 2020).

Experimental forward genetics has identified a few SCD genes. Forward genetics investigates the genetic basis of a particular phenotype, in contrast to reverse genetics, which investigates the phenotypic consequences of altering a particular gene (Moresco et al., 2013). The current development in forward genetics and its equipment with NGS tools could further enhance the application of forward genetics for discovering new genes implicated in Mendelian diseases (Singh, 2020). Furthermore, gene essentiality screens on model organisms and cell lines of different origins are gaining momentum and are progressively integrated into gene-prioritization steps in NGS analysis pipelines (Cacheiro et al., 2020).

Remaining Challenges

There is still a pressing need for biomarkers to complement the advances in SCD genomics and phenomics. The role of biomarkers is not limited to diagnosis, as biomarkers can objectively measure disease evolution, prognosis, and response to therapy (Gülbakan et al., 2016). Currently, the means for validating the biological impact of candidate pathogenic SCD variants in the clinical setting are not available except for a few SCD subtypes, e.g., SPG5A caused by mutations in the *CYP7B1* gene (27-OH cholesterol), SPG9A (OMIM # 601162), and SPG9B (OMIM # 616586) caused by mutations in the *ALDH18A1* gene (amino-acid levels in blood), and ataxia with isolated vitamin E deficiency (OMIM # 277460) caused by mutations in the *TTPA* gene (Ouahchi et al., 1995; Coutelier et al., 2015; Marelli et al., 2018).

The rapidly evolving field of metabolomics is expected to contribute to the SCD biomarkers list (Wishart, 2016; Yang et al., 2019). Metabolomics is analyzing the body metabolites in a comprehensive, yet systematic approach (Gülbakan et al., 2016). Metabolic disturbances are core pathological mechanisms in many SCD subtypes (Blackstone, 2018; Darios et al., 2020). Consequently, metabolomics will not only provide biomarkers but also will shed light on the SCD biochemistry and provide an objective means for classifying SCD subtypes (Wishart, 2016; del Mar Amador et al., 2018; Yang et al., 2019). It can also unravel the overlapping phenotypes caused by mutations in a single gene, a pattern common in neurological diseases, including SCD (Esterhuizen et al., 2021). Developing an objective SCD nosology is a prerequisite for advancing SCD research and therapeutics (Elsayed et al., 2019).

Metabolomics is the closest omics layer to the phenotype and could have central roles in validating the newly discovered Mendelian genes (Feussner and Polle, 2015; Graham et al., 2018; Almontashiri et al., 2020). Metabolomics studies are conducted using two main approaches: a hypothesis-directed approach that studies a predetermined set of metabolites (targeted metabolomics) and a hypothesis-free comprehensive approach that aims to analyze the whole set of metabolites within a system (untargeted or global metabolomics; Almontashiri et al., 2020; Chen et al., 2020). Global metabolomics is gaining momentum over targeted approaches in clinical settings (Almontashiri et al., 2020). However, neither approach is suitable yet for the *ab initio* discovery of new genes; rather, global metabolomics can be integrated with exome and genome sequencing to enhance new genes discovery (Graham et al., 2018; Almontashiri et al., 2020).

Multiple hurdles are facing the application of metabolomics in Mendelian genes discovery. Firstly, the human metabolome is made of thousands of compounds influenced by the genome but also by a plethora of non-genomic determinants, including the microbiome, lifestyle, environment, and others (Bar et al., 2020). Secondly, a proportion of the human metabolome is not identified yet and there is no single tool currently capable of capturing the whole metabolome (Graham et al., 2018). Furthermore, there is no congruence between the concentration of certain metabolites in the blood and the brain (Yang et al., 2019). Lastly, the limited number of patients with rare diseases, particularly those in whom new phenotypes are discovered, complicates the analysis of metabolomics data and the differentiation between genuine signals and background noise (Graham et al., 2018; Yang et al., 2019). However, recently, some promising solutions have emerged to overcome those hurdles.

Multiple repositories were launched to enhance metabolomics data sharing, and, in parallel, tools that integrate the data from different repositories were recently developed (Krassowski et al., 2020; Palermo et al., 2020). Examples for such repositories include MetaboLights, GNPS-MASSIVE, and Metabolomics Workbench (Haug et al., 2013; Sud et al., 2016; Wang et al., 2016). Animal models can be used to alleviate the need for human tissues to study the metabolome (Kilk, 2019). Even more promising are the human organoids, which are more representative substitutes for the human brain tissues (Kim et al., 2020). To a large extent, organoids can also solve the scarcity of patients with rare diseases (Kim et al., 2020).

Other omics layers, particularly the transcriptome and the proteome, can be exploited in the quest for identifying new SCD genes with the limitation of the access to the brain tissue in most cases. A substantial proportion of the undiagnosed Mendelian diseases are caused by variants impacting RNA expression and/or regulation, including deep intronic variants (Gonorazky et al., 2019). Transcriptome sequencing (RNA-seq) is deemed to enhance identifying Mendelian variants by ~10–30% (Kremer et al., 2017; Gonorazky et al., 2019). Allelic imbalances in transcriptomics data can identify imprinting, uni-parental disomy, and X chromosome inactivation (Gonorazky et al., 2019; Yépez et al., 2021). However, the tissue transcriptome does not necessarily reflect the proteome, metabolome, and phenome (Wang et al., 2019). In the abnormal tissues, the discrepancy between the transcriptome and the proteome can be more evident depending on the underlying disease's pathophysiology and the type of the disease-causing variants (Roos et al., 2018). Although the proteome is closer to the phenome than the transcriptome, it is more easier to recover data by RNA-seq than proteome-capturing approaches due to the higher depth of RNA-seq and the much wider proteins' dynamic range of expression (Feussner and Polle, 2015; Wang et al., 2019). Moreover, ~13% of the human proteome is still missing and not captured (Omenn et al., 2017; Wang et al., 2019).

A holistic approach that integrates data obtained from multiple omics layers (multi-omics approach) has a higher biological reliability and can substantially enhance new Mendelian genes' discovery than the single-omics approaches (Labory et al., 2020). The multi-omics have been successfully used in many diseases, including neurological diseases, and is thus a promising approach to enhance the discovery of new SCD genes (Crowther et al., 2018; Kerr et al., 2020; Labory et al., 2020).

In the current big data era, tools that integrate the data generated at different biological scales, namely multi-omic tools, are rapidly emerging (Anderson et al., 2019; Kerr et al., 2020; Subramanian et al., 2020). This promotes collaborative studies. In some instances, establishing associations between variants, especially non-exonic variants, and diseases, required a large number of patients and controls. Involving patients from different genetic backgrounds has aided the refinement of the loci associated with some SCD genes. An increasing number of consortia and collaborative platforms have been launched to enhance the discovery of new genes by enabling the sharing of sequence data, e.g., GENESIS platform; variants and phenotypes, e.g., GeneMatcher; and connecting geneticists and biologists to enhance validating new candidate genes, e.g., the Canadian Rare Diseases Models and Mechanisms Network (Gonzalez et al., 2015; Sobreira et al., 2015; Boycott et al., 2020). Larger-scale collaborations are essential to explore a very challenging aspect of all neurogenetic conditions, the clinical variability. Modifier genes can explain part of the phenotypic variability commonly seen between patients, as shown in autosomal dominant SCAs (Du Montcel et al., 2014).

Sporadic cases remain a challenge in these diseases. They are often the most frequent in clinical practice and their analysis possibility is limited. Collaborations, trio analysis, and multi-omic approaches can assist in solving them. In our experience, 10% of sporadic cases can be explained by *de novo* variants or inherited variants due to incomplete penetrance, adoption, family censure, etc. We expect multigenic alterations or somatic variants to explain as well part of the missing heritability in some of the patients with sporadic forms.

Finally, one of the bottlenecks in diagnosing SCD is the functional validation of the variants of unknown significance (VUS). Three-dimensional modeling and functional enzymatic assays can be precious, as shown for variants in the *CYP2U1* gene, the gene associated with SPG56 (OMIM # 615030; Durand et al., 2018). Small animal models such as *C. elegans* and zebrafish can also be used (Martin et al., 2013; Novarino et al., 2014; Chiba et al., 2019). Moreover, organoids emerged as promising models to establish pathogenicity and circumvent the difficulty in obtaining brain tissues (Kim et al., 2020). Without such systematic studies, we can reasonably think that a significant part of variants are not really causative which may limit the finding of phenotype-genotype correlations. In that respect, the report of second mutated families is required for multiple clinico genetic entities, including SPG59, 60, 65, 67, 68, 69, 70, and 71, found in single families yet (Novarino et al., 2014).

CONCLUSION

In conclusion, the toolbox for diagnosing Mendelian diseases is continuously expanding, including multiple "out of the box" approaches. These valuable utensils and approaches (**Figure 3**) have solved and deemed to continue solving some of the Mendelian diseases diagnostic odysseys. This review highlighted some of the approaches successfully used to discover SCD genes and shed light on some of the promising resources and approaches expected to play central roles in diagnosing SCD soon. Discovering novel genes associated with SCD is moving toward integrated multi-omic research using innovative technologies. We hypothesize that many neurogenetic entities will join the spectrum of SCD, necessitating newer and more comprehensive nosologies at the SCD, and likely at the neurogenetic level as well.

AUTHOR CONTRIBUTIONS

AY and GS designed, wrote, and critically revised this review article. All authors contributed to the article and approved the submitted version.

FUNDING

GS is financially supported by a European Union grant (grant 779257 Solve-RD from the Horizon 2020 Research and Innovation Program) and by Grants from the Spastic Paraplegia Foundation (US), the HSP-ASL-France and the association Connaitre les Syndromes Cérébelleux. AY is supported by a grant from the Ministry of Higher Education in Sudan and the French embassy in Sudan.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.638730/full#supplementary-material

REFERENCES

Abou Jamra, R., Philippe, O., Raas-Rothschild, A., Eck, S. H., Graf, E., Buchert, R., et al. (2011). Adaptor protein complex 4 deficiency causes severe autosomal-recessive intellectual disability, progressive spastic paraplegia, shy character, and short stature. *Am. J. Hum. Genet.* 88, 788–795. doi: 10.1016/j.ajhg.2011.04.019

Agrawal, P. B., Joshi, M., Marinakis, N. S., Schmitz-Abe, K., Ciarlini, P. D. S. C., Sargent, J. C., et al. (2014). Expanding the phenotype associated with the NEFL mutation neuromuscular disease in a family with overlapping myopathic and neurogenic findings. *JAMA Neurol.* 71, 1413–1420. doi: 10.1001/jamaneurol.2014.1432

Almontashiri, N. A. M., Zha, L., Young, K., Law, T., Kellogg, M. D., Bodamer, O. A., et al. (2020). Clinical validation of targeted and untargeted metabolomics

testing for genetic disorders: a 3 year comparative study. *Sci. Rep.* 10, 1–8. doi: 10.1038/s41598-020-66401-2

Anderson, D., Baynam, G., Blackwell, J. M., and Lassmann, T. (2019). Personalised analytics for rare disease diagnostics. *Nat. Commun.* 10, 1–8. doi: 10.1038/s41467-019-13345-5

Bahlo, M., Bennett, M. F., Degorski, P., Tankard, R. M., Delatycki, M. B., and Lockhart, P. J. (2018). Recent advances in the detection of repeat expansions with short-read next-generation sequencing. *F1000Research* 7:736. doi: 10.12688/f1000research.13980.1

Bar, N., Korem, T., Weissbrod, O., Zeevi, D., Rothschild, D., Leviatan, S., et al. (2020). A reference map of potential determinants for the human serum metabolome. *Nature* 588, 135–140. doi: 10.1038/s41586-020-2896-2

Bell, S. M., Schreiner, C. M., Schultheis, P. J., Miller, M. L., Evans, R. L., Vorhees, C. V., et al. (1999). Targeted disruption of the murine Nhe1 locus induces ataxia, growth retardation, and seizures. *Am. J. Phys.* 276, C788–C795. doi: 10.1152/ajpcell.1999.276.4.c788

Bhatia, N. S., Lim, J. Y., Bonnard, C., Kuan, J. L., Brett, M., Wei, H., et al. (2020). Singapore undiagnosed disease program: genomic analysis aids diagnosis and clinical management. *Arch. Dis. Child.* 106, 31–37. doi: 10.1136/archdischild-2020-319180

Blackstone, C. (2018). Converging cellular themes for the hereditary spastic paraplegias. *Curr. Opin. Neurobiol.* 51, 139–146. doi: 10.1016/j.conb.2018.04.025

Boutry, M., Morais, S., and Stevanin, G. (2019). Update on the genetics of spastic paraplegias. *Curr. Neurol. Neurosci. Rep.* 19:18. doi: 10.1007/s11910-019-0930-2

Boycott, K. M., Campeau, P. M., Howley, H. E., Pavlidis, P., Rogic, S., Oriel, C., et al. (2020). The Canadian rare diseases models and mechanisms (RDMM) network: connecting understudied genes to model organisms. *Am. J. Hum. Genet.* 106, 143–152. doi: 10.1016/j.ajhg.2020.01.009

Boycott, K. M., Rath, A., Chong, J. X., Hartley, T., Alkuraya, F. S., Baynam, G., et al. (2017). International cooperation to enable the diagnosis of all rare genetic diseases. *Am. J. Hum. Genet.* 100, 695–705. doi: 10.1016/j.ajhg.2017.04.003

Brusse, E., Maat-Kievit, J. A., and van Swieten, J. C. (2007). Diagnosis and management of early- and late-onset cerebellar ataxia. *Clin. Genet.* 71, 12–24. doi: 10.1111/j.1399-0004.2006.00722.x

Burgess, D. L., Jones, J. M., Meisler, M. H., and Noebels, J. L. (1997). Mutation of the Ca^{2+} channel β subunit gene Cchb4 is associated with ataxia and seizures in the lethargic (lh) mouse. *Cell* 88, 385–392. doi: 10.1016/S0092-8674(00)81877-2

Cacheiro, P., Muñoz-Fuentes, V., Murray, S. A., Dickinson, M. E., Bucan, M., Nutter, L. M. J., et al. (2020). Human and mouse essentiality screens as a resource for disease gene discovery. *Nat. Commun.* 11:655. doi: 10.1038/s41467-020-14284-2

Cagnoli, C., Brussino, A., Mancini, C., Ferrone, M., Orsi, L., Salmin, P., et al. (2018). Spinocerebellar ataxia tethering PCR: a rapid genetic test for the diagnosis of spinocerebellar ataxia types 1, 2, 3, 6, and 7 by PCR and capillary electrophoresis. *J. Mol. Diagn.* 20, 289–297. doi: 10.1016/j.jmoldx.2017.12.006

Campuzano, V., Montermini, L., Moltò, M. D., Pianese, L., Cossée, M., Cavalcanti, F., et al. (1996). Friedreich's ataxia: autosomal recessive disease caused by an intronic GAA triplet repeat expansion. *Science* 271, 1423–1427. doi: 10.1126/science.271.5254.1423

Caspar, S. M., Dubacher, N., Kopps, A. M., Meienberg, J., Henggeler, C., and Matyas, G. (2018). Clinical sequencing: from raw data to diagnosis with lifetime value. *Clin. Genet.* 93, 508–519. doi: 10.1111/cge.13190

Chen, L., Zhong, F., and Zhu, J. (2020). Bridging targeted and untargeted mass spectrometry-based metabolomics via hybrid approaches. *Meta* 10, 1–19. doi: 10.3390/metabo10090348

Chiba, K., Takahashi, H., Chen, M., Obinata, H., Arai, S., Hashimoto, K., et al. (2019). Disease-associated mutations hyperactivate KIF1A motility and anterograde axonal transport of synaptic vesicle precursors. *Proc. Natl. Acad. Sci. U. S. A.* 116, 18429–18434. doi: 10.1073/pnas.1905690116

Ciaccio, C., Pantaleoni, C., Bulgheroni, S., Sciacca, F., and D'Arrigo, S. (2020). Chromosomal microarray analysis has a poor diagnostic yield in children with developmental delay/intellectual disability when concurrent cerebellar anomalies are present. *Cerebellum* 19, 629–635. doi: 10.1007/s12311-020-01145-3

Clark, M. M., Stark, Z., Farnaes, L., Tan, T. Y., White, S. M., Dimmock, D., et al. (2018). Meta-analysis of the diagnostic and clinical utility of genome and exome sequencing and chromosomal microarray in children with suspected genetic diseases. *NPJ Genomic Med.* 3:16. doi: 10.1038/s41525-018-0053-8

Claussnitzer, M., Cho, J. H., Collins, R., Cox, N. J., Dermitzakis, E. T., Hurles, M. E., et al. (2020). A brief history of human disease genetics. *Nature* 577, 179–189. doi: 10.1038/s41586-019-1879-7

Cortese, A., Simone, R., Sullivan, R., Vandrovcova, J., Tariq, H., Yan, Y. W., et al. (2019). Biallelic expansion of an intronic repeat in RFC1 is a common cause of late-onset ataxia. *Nat. Genet.* 51, 649–658. doi: 10.1038/s41588-019-0372-4

Coutelier, M., Goizet, C., Durr, A., Habarou, F., Morais, S., Dionne-Laporte, A., et al. (2015). Alteration of ornithine metabolism leads to dominant and recessive hereditary spastic paraplegia. *Brain* 138, 2191–2205. doi: 10.1093/brain/awv143

Coutelier, M., Hammer, M. B., Stevanin, G., Monin, M. L., Davoine, C. S., Mochel, F., et al. (2018). Efficacy of exome-targeted capture sequencing to detect mutations in known cerebellar ataxia genes. *JAMA Neurol.* 75, 591–599. doi: 10.1001/jamaneurol.2017.5121

Crowther, L. M., Poms, M., and Plecko, B. (2018). Multiomics tools for the diagnosis and treatment of rare neurological disease. *J. Inherit. Metab. Dis.* 41, 425–434. doi: 10.1007/s10545-018-0154-7

Darios, F., Mochel, F., and Stevanin, G. (2020). Lipids in the physiopathology of hereditary spastic paraplegias. *Front. Neurosci.* 14:74. doi: 10.3389/fnins.2020.00074

David, G., Abbas, N., Stevanin, G., Dürr, A., Yvert, G., Cancel, G., et al. (1997). Cloning of the SCA7 gene reveals a highly unstable CAG repeat expansion. *Nat. Genet.* 17, 65–70. doi: 10.1038/ng0997-65

Deelen, P., van Dam, S., Herkert, J. C., Karjalainen, J. M., Brugge, H., Abbott, K. M., et al. (2019). Improving the diagnostic yield of exome- sequencing by predicting gene–phenotype associations using large-scale gene expression analysis. *Nat. Commun.* 10, 1–13. doi: 10.1038/s41467-019-10649-4

del Mar Amador, M., Colsch, B., Lamari, F., Jardel, C., Ichou, F., Rastetter, A., et al. (2018). Targeted versus untargeted omics—the CAFSA story. *J. Inherit. Metab. Dis.* 41, 447–456. doi: 10.1007/s10545-017-0134-3

Depienne, C., Fedirko, E., Forlani, S., Cazeneuve, C., Ribaï, P., Feki, I., et al. (2007). Exon deletions of SPG4 are a frequent cause of hereditary spastic paraplegia. *J. Med. Genet.* 44, 281–284. doi: 10.1136/jmg.2006.046425

De Roeck, A., De Coster, W., Bossaerts, L., Cacace, R., De Pooter, T., Van Dongen, J., et al. (2019). NanoSatellite: accurate characterization of expanded tandem repeat length and sequence through whole genome long-read sequencing on PromethION. *Genome Biol.* 20:239. doi: 10.1186/s13059-019-1856-3

De Silva, R., Greenfield, J., Cook, A., Bonney, H., Vallortigara, J., Hunt, B., et al. (2019). Guidelines on the diagnosis and management of the progressive ataxias. *Orphanet J. Rare Dis.* 14:51. doi: 10.1186/s13023-019-1013-9

Devanna, P., Chen, X. S., Ho, J., Gajewski, D., Smith, S. D., Gialluisi, A., et al. (2018). Next-gen sequencing identifies non-coding variation disrupting miRNA-binding sites in neurological disorders. *Mol. Psychiatry* 23, 1375–1384. doi: 10.1038/mp.2017.30

Dou, Y., Kwon, M., Rodin, R. E., Cortés-Ciriano, I., Doan, R., Luquette, L. J., et al. (2020). Accurate detection of mosaic variants in sequencing data without matched controls. *Nat. Biotechnol.* 38, 314–319. doi: 10.1038/s41587-019-0368-8

Duchesne, A., Vaiman, A., Frah, M., Floriot, S., Legoueix-Rodriguez, S., Desmazières, A., et al. (2018). Progressive ataxia of Charolais cattle highlights a role of KIF1C in sustainable myelination. *PLoS Genet.* 14:e1007550. doi: 10.1371/journal.pgen.1007550

Du Montcel, S. T., Durr, A., Bauer, P., Figueroa, K. P., Ichikawa, Y., Brussino, A., et al. (2014). Modulation of the age at onset in spinocerebellar ataxia by CAG tracts in various genes. *Brain* 137, 2444–2455. doi: 10.1093/brain/awu174

Durand, C. M., Dhers, L., Tesson, C., Tessa, A., Fouillen, L., Jacqueré, S., et al. (2018). CYP2U1 activity is altered by missense mutations in hereditary spastic paraplegia 56. *Hum. Mutat.* 39, 140–151. doi: 10.1002/humu.23359

Elmas, M., and Gogus, B. (2020). Success of face analysis technology in rare genetic diseases diagnosed by whole-exome sequencing: a single-center experience. *Mol. Syndromol.* 11, 4–14. doi: 10.1159/000505800

Elsayed, L. E. O., Eltazi, I. Z. M., Ahmed, A. E. M., and Stevanin, G. (2019). Hereditary spastic paraplegias: time for an objective case definition and a

new nosology for neurogenetic disorders to facilitate biomarker/therapeutic studies. *Expert. Rev. Neurother.* 19, 409–415. doi: 10.1080/14737175.2019.1608824

Elsayed, L. E. O., Mohammed, I. N., Hamed, A. A. A., Elseed, M. A., Johnson, A., Mairey, M., et al. (2016). Hereditary spastic paraplegias: identification of a novel SPG57 variant affecting TFG oligomerization and description of HSP subtypes in Sudan. *Eur. J. Hum. Genet.* 25, 100–110. doi: 10.1038/ejhg.2016.108

Engert, J. C., Bérubé, P., Mercier, J., Doré, C., Lepage, P., Ge, B., et al. (2000). ARSACS, a spastic ataxia common in northeastern Québec, is caused by mutations in a new gene encoding an 11.5-kb ORF. *Nat. Genet.* 24, 120–125. doi: 10.1038/72769

Erlich, Y., Edvardson, S., Hodges, E., Zenvirt, S., Thekkat, P., Shaag, A., et al. (2011). Exome sequencing and disease-network analysis of a single family implicate a mutation in KIF1A in hereditary spastic paraparesis. *Genome Res.* 21, 658–664. doi: 10.1101/gr.117143.110

Escayg, A., De Waard, M., Lee, D. D., Bichet, D., Wolf, P., Mayer, T., et al. (2000). Coding and noncoding variation of the human calcium-channel β4-subunit gene CACNB4 patients with idiopathic generalized epilepsy and episodic ataxia. *Am. J. Hum. Genet.* 66, 1531–1539. doi: 10.1086/302909

Esterhuizen, K., Lindeque, J. Z., Mason, S., van der Westhuizen, F. H., Rodenburg, R. J., de Laat, P., et al. (2021). One mutation, three phenotypes: novel metabolic insights on MELAS, MIDD and myopathy caused by the m.3243A > G mutation. *Metabolomics* 17, 1–16. doi: 10.1007/s11306-020-01769-w

Feussner, I., and Polle, A. (2015). What the transcriptome does not tell—proteomics and metabolomics are closer to the plants' patho-phenotype. *Curr. Opin. Plant Biol.* 26, 26–31. doi: 10.1016/j.pbi.2015.05.023

Fink, J. K. (1997). Advances in hereditary spastic paraplegia. *Curr. Opin. Neurol.* 10, 313–318. doi: 10.1097/00019052-199708000-00006

Fogel, B. L., Lee, H., Deignan, J. L., Strom, S. P., Kantarci, S., Wang, X., et al. (2014). Exome sequencing in the clinical diagnosis of sporadic or familial cerebellar ataxia. *JAMA Neurol.* 71, 1237–1246. doi: 10.1001/jamaneurol.2014.1944

Friesema, E. C. H., Grueters, P. A., Biebermann, H., Krude, H., von Moers, A., Reeser, M., et al. (2004). Association between mutations in a thyroid hormone transporter and severe X-linked psychomotor retardation. *Lancet* 364, 1435–1437. doi: 10.1016/S0140-6736(04)17226-7

Fung, J. L. F., Yu, M. H. C., Huang, S., Chung, C. C. Y., Chan, M. C. Y., Pajusalu, S., et al. (2020). A three-year follow-up study evaluating clinical utility of exome sequencing and diagnostic potential of reanalysis. *NPJ Genomic Med.* 5:37. doi: 10.1038/s41525-020-00144-x

Galatolo, D., Tessa, A., Filla, A., and Santorelli, F. M. (2018). Clinical application of next generation sequencing in hereditary spinocerebellar ataxia: increasing the diagnostic yield and broadening the ataxia-spasticity spectrum. A retrospective analysis. *Neurogenetics* 19, 1–18. doi: 10.1007/s10048-017-0532-6

Gao, G., and Smith, D. I. (2020). Clinical massively parallel sequencing. *Clin. Chem.* 66, 77–88. doi: 10.1373/clinchem.2019.303305

Gonorazky, H. D., Naumenko, S., Ramani, A. K., Nelakuditi, V., Mashouri, P., Wang, P., et al. (2019). Expanding the boundaries of RNA sequencing as a diagnostic tool for rare mendelian disease. *Am. J. Hum. Genet.* 104, 466–483. doi: 10.1016/j.ajhg.2019.01.012

Gonzalez, M., Falk, M. J., Gai, X., Postrel, R., Schüle, R., and Zuchner, S. (2015). Innovative genomic collaboration using the genesis (GEM.App) platform. *Hum. Mutat.* 36, 950–956. doi: 10.1002/humu.22836

Graham, E., Lee, J., Price, M., Tarailo-Graovac, M., Matthews, A., Engelke, U., et al. (2018). Integration of genomics and metabolomics for prioritization of rare disease variants: a 2018 literature review. *J. Inherit. Metab. Dis.* 41, 435–445. doi: 10.1007/s10545-018-0139-6

Guergueltcheva, V., Azmanov, D. N., Angelicheva, D., Smith, K. R., Chamova, T., Florez, L., et al. (2012). Autosomal-recessive congenital cerebellar ataxia is caused by mutations in metabotropic glutamate receptor 1. *Am. J. Hum. Genet.* 91, 553–564. doi: 10.1016/j.ajhg.2012.07.019

Guissart, C., Li, X., Leheup, B., Drouot, N., Montaut-Verient, B., Raffo, E., et al. (2015). Mutation of SLC9A1, encoding the major Na^+/H^+ exchanger, causes ataxia-deafness Lichtenstein-Knorr syndrome. *Hum. Mol. Genet.* 24, 463–470. doi: 10.1093/hmg/ddu461

Gülbakan, B., Özgül, R. K., Yüzbaşioglu, A., Kohl, M., Deigner, H. P., and Özgüç, M. (2016). Discovery of biomarkers in rare diseases: innovative approaches by predictive and personalized medicine. *EPMA J.* 7:24. doi: 10.1186/s13167-016-0074-2

Hagerman, R. J., Leehey, M., Heinrichs, W., Tassone, F., Wilson, R., Hills, J., et al. (2001). Intention tremor, parkinsonism, and generalized brain atrophy in male carriers of fragile X. *Neurology* 57, 127–130. doi: 10.1212/WNL.57.1.127

Hansen, M. C., Haferlach, T., and Nyvold, C. G. (2020). A decade with whole exome sequencing in haematology. *Br. J. Haematol.* 188, 367–382. doi: 10.1111/bjh.16249

Harding, A. E. (1993). Hereditary spastic paraplegias. *Semin. Neurol.* 13, 333–336. doi: 10.1055/s-2008-1041143

Haug, K., Salek, R. M., Conesa, P., Hastings, J., De Matos, P., Rijnbeek, M., et al. (2013). MetaboLights: an open-access general-purpose repository for metabolomics studies and associated meta-data. *Nucleic Acids Res.* 41, D781–D786. doi: 10.1093/nar/gks1004

Hersheson, J., Mencacci, N. E., Davis, M., MacDonald, N., Trabzuni, D., Ryten, M., et al. (2013). Mutations in the autoregulatory domain of β-tubulin 4a cause hereditary dystonia. *Ann. Neurol.* 73, 546–553. doi: 10.1002/ana.23832

Holt, I. J., Harding, A. E., Petty, R. K. H., and Morgan-Hughes, J. A. (1990). A new mitochondrial disease associated with mitochondrial DNA heteroplasmy. *Am. J. Hum. Genet.* 46, 428–433.

Hsiao, C. T., Liu, Y. T., Liao, Y. C., Hsu, T. Y., Lee, Y. C., and Soong, B. W. (2017). Mutational analysis of ITPR1 in a Taiwanese cohort with cerebellar ataxias. *PLoS One* 12:e0187503. doi: 10.1371/journal.pone.0187503

Ikeda, Y., Dick, K. A., Weatherspoon, M. R., Gincel, D., Armbrust, K. R., Dalton, J. C., et al. (2006). Spectrin mutations cause spinocerebellar ataxia type 5. *Nat. Genet.* 38, 184–190. doi: 10.1038/ng1728

Ilg, W., Schatton, C., Schicks, J., Giese, M. A., Schöls, L., and Synofzik, M. (2012). Video game-based coordinative training improves ataxia in children with degenerative ataxia. *Neurology* 79, 2056–2060. doi: 10.1212/WNL.0b013e3182749e67

Jain, M., Koren, S., Miga, K. H., Quick, J., Rand, A. C., Sasani, T. A., et al. (2018). Nanopore sequencing and assembly of a human genome with ultra-long reads. *Nat. Biotechnol.* 36, 338–345. doi: 10.1038/nbt.4060

Jalkh, N., Corbani, S., Haidar, Z., Hamdan, N., Farah, E., Abou Ghoch, J., et al. (2019). The added value of WES reanalysis in the field of genetic diagnosis: lessons learned from 200 exomes in the Lebanese population. *BMC Med. Genet.* 12:11. doi: 10.1186/s12920-019-0474-y

Jang, W., Kim, Y., Han, E., Park, J., Chae, H., Kwon, A., et al. (2019). Chromosomal microarray analysis as a first-tier clinical diagnostic test in patients with developmental delay/intellectual disability, autism spectrum disorders, and multiple congenital anomalies: a prospective multicenter study in Korea. *Ann. Lab. Med.* 39, 299–310. doi: 10.3343/alm.2019.39.3.299

Jen, J. C., Wan, J., Palos, T. P., Howard, B. D., and Baloh, R. W. (2005). Mutation in the glutamate transporter EAAT1 causes episodic ataxia, hemiplegia, and seizures. *Neurology* 65, 529–534. doi: 10.1212/01.WNL.0000172638.58172.5a

Kerr, K., McAneney, H., Smyth, L. J., Bailie, C., McKee, S., and McKnight, A. J. (2020). A scoping review and proposed workflow for multi-omic rare disease research. *Orphanet J. Rare Dis.* 15:107. doi: 10.1186/s13023-020-01376-x

Kilk, K. (2019). Metabolomics for animal models of rare human diseases: an expert review and lessons learned. *Omi. A J. Integr. Biol.* 23, 300–307. doi: 10.1089/omi.2019.0065

Kim, J., Koo, B. K., and Knoblich, J. A. (2020). Human organoids: model systems for human biology and medicine. *Nat. Rev. Mol. Cell Biol.* 21, 571–584. doi: 10.1038/s41580-020-0259-3

Klebe, S., Stevanin, G., and Depienne, C. (2015). Clinical and genetic heterogeneity in hereditary spastic paraplegias: from SPG1 to SPG72 and still counting. *Rev. Neurol.* 171, 505–530. doi: 10.1016/j.neurol.2015.02.017

Koide, R., Ikeuchi, T., Onodera, O., Tanaka, H., Igarashi, S., Endo, K., et al. (1994). Unstable expansion of CAG repeat in hereditary dentatorubral-pallidoluysian atrophy (DRPLA). *Nat. Genet.* 6, 9–13. doi: 10.1038/ng0194-9

Koob, M. D., Moseley, M. L., Schut, L. J., Benzow, K. A., Bird, T. D., Day, J. W., et al. (1999). An untranslated CTG expansion causes a novel form of spinocerebellar ataxia (SCA8). *Nat. Genet.* 21, 379–384. doi: 10.1038/7710

Krassowski, M., Das, V., Sahu, S. K., and Misra, B. B. (2020). State of the field in multi-omics research: from computational needs to data mining and sharing. *Front. Genet.* 11:610798. doi: 10.3389/fgene.2020.610798

Kremer, L. S., Bader, D. M., Mertes, C., Kopajtich, R., Pichler, G., Iuso, A., et al. (2017). Genetic diagnosis of Mendelian disorders via RNA sequencing. *Nat. Commun.* 8:15824. doi: 10.1038/ncomms15824

Labory, J., Fierville, M., Ait-El-Mkadem, S., Bannwarth, S., Paquis-Flucklinger, V., and Bottini, S. (2020). Multi-omics approaches to improve mitochondrial

disease diagnosis: challenges, advances, and perspectives. *Front. Mol. Biosci.* 7:327. doi: 10.3389/fmolb.2020.590842

Lam, E. T., Hastie, A., Lin, C., Ehrlich, D., Das, S. K., Austin, M. D., et al. (2012). Genome mapping on nanochannel arrays for structural variation analysis and sequence assembly. *Nat. Biotechnol.* 30, 771–776. doi: 10.1038/nbt.2303

Lan, M. Y., Chang, Y. Y., Yeh, T. H., Lin, T. K., and Lu, C. S. (2017). Leukoencephalopathy with brainstem and spinal cord involvement and lactate elevation (LBSL) with a novel DARS2 mutation and isolated progressive spastic paraparesis. *J. Neurol. Sci.* 372, 229–231. doi: 10.1016/j.jns.2016.11.058

Lee, C. E., Singleton, K. S., Wallin, M., and Faundez, V. (2020). Rare genetic diseases: nature's experiments on human development. *iScience* 23:101123. doi: 10.1016/j.isci.2020.101123

Levy, B., and Burnside, R. D. (2019). Are all chromosome microarrays the same? What clinicians need to know. *Prenat. Diagn.* 39, 157–164. doi: 10.1002/pd.5422

Li, S. H., McInnis, M. G., Margolis, R. L., Antonarakis, S. E., and Ross, C. A. (1993). Novel triplet repeat containing genes in human brain: cloning, expression, and length polymorphisms. *Genomics* 16, 572–579. doi: 10.1006/geno.1993.1232

Liao, Y., Anttonen, A. K., Liukkonen, E., Gaily, E., Maljevic, S., Schubert, S., et al. (2010). SCN2A mutation associated with neonatal epilepsy, late-onset episodic ataxia, myoclonus, and pain. *Neurology* 75, 1454–1458. doi: 10.1212/WNL.0b013e3181f8812e

Lim, J., Hao, T., Shaw, C., Patel, A. J., Szabó, G., Rual, J. F., et al. (2006). A protein-protein interaction network for human inherited ataxias and disorders of purkinje cell degeneration. *Cell* 125, 801–814. doi: 10.1016/j.cell.2006.03.032

Lin, C.-C., Ashizawa, T., and Kuo, S.-H. (2020). Collaborative efforts for spinocerebellar ataxia research in the United States: CRC-SCA and READISCA. *Front. Neurol.* 11:902. doi: 10.3389/fneur.2020.00902

Lipner, E. M., and Greenberg, D. A. (2018). "The rise and fall and rise of linkage analysis as a technique for finding and characterizing inherited influences on disease expression" in *Disease gene identification*. ed. J. DiStefano (Humana Press Inc.), 381–397.

Lise, S., Clarkson, Y., Perkins, E., Kwasniewska, A., Sadighi Akha, E., Parolin Schnekenberg, R., et al. (2012). Recessive mutations in SPTBN2 implicate β-III spectrin in both cognitive and motor development. *PLoS Genet.* 8:e1003074. doi: 10.1371/journal.pgen.1003074

Liu, P., Meng, L., Normand, E. A., Xia, F., Song, X., Ghazi, A., et al. (2019). Reanalysis of clinical exome sequencing data. *N. Engl. J. Med.* 380, 2478–2480. doi: 10.1056/NEJMc1812033

Liu, Y., Zaun, H. C., Orlowski, J., and Ackerman, S. L. (2013). CHP1-mediated NHE1 biosynthetic maturation is required for purkinje cell axon homeostasis. *J. Neurosci.* 33, 12656–12669. doi: 10.1523/JNEUROSCI.0406-13.2013

Macedo-Souza, L. I., Kok, F., Santos, S., Licinio, L., Lezirovitz, K., Cavaçana, N., et al. (2009). Spastic paraplegia, optic atrophy, and neuropathy: new observations, locus refinement, and exclusion of candidate genes. *Ann. Hum. Genet.* 73, 382–387. doi: 10.1111/j.1469-1809.2009.00507.x

Manto, M., Gandini, J., Feil, K., and Strupp, M. (2020). Cerebellar ataxias: an update. *Curr. Opin. Neurol.* 33, 150–160. doi: 10.1097/WCO.0000000000000774

Marelli, C., Lamari, F., Rainteau, D., Lafourcade, A., Banneau, G., Humbert, L., et al. (2018). Plasma oxysterols: biomarkers for diagnosis and treatment in spastic paraplegia type 5. *Brain* 141, 72–84. doi: 10.1093/brain/awx297

Maroilley, T., and Tarailo-Graovac, M. (2019). Uncovering missing heritability in rare diseases. *Gene* 10:275. doi: 10.3390/genes10040275

Martin, E., Schüle, R., Smets, K., Rastetter, A., Boukhris, A., Loureiro, J. L., et al. (2013). Loss of function of glucocerebrosidase GBA2 is responsible for motor neuron defects in hereditary spastic paraplegia. *Am. J. Hum. Genet.* 92, 238–244. doi: 10.1016/j.ajhg.2012.11.021

Martins, S., Calafell, F., Gaspar, C., Wong, V. C. N., Silveira, I., Nicholson, G. A., et al. (2007). Asian origin for the worldwide-spread mutational event in Machado-Joseph disease. *Arch. Neurol.* 64, 1502–1508. doi: 10.1001/archneur.64.10.1502

Matalonga, L., Laurie, S., Papakonstantinou, A., Piscia, D., Mereu, E., Bullich, G., et al. (2020). Improved diagnosis of rare disease patients through systematic detection of runs of homozygosity. *J. Mol. Diagn.* 22, 1205–1215. doi: 10.1016/j.jmoldx.2020.06.008

Mazzarotto, F., Olivotto, I., and Walsh, R. (2020). Advantages and perils of clinical whole-exome and whole-genome sequencing in cardiomyopathy. *Cardiovasc. Drugs Ther.* 34, 241–253. doi: 10.1007/s10557-020-06948-4

Melo, U. S., Macedo-Souza, L. I., Figueiredo, T., Muotri, A. R., Gleeson, J. G., Coux, G., et al. (2015). Overexpression of KLC2 due to a homozygous deletion in the non-coding region causes SPOAN syndrome. *Hum. Mol. Genet.* 24, 6877–6885. doi: 10.1093/hmg/ddv388

Mendoza-Ferreira, N., Coutelier, M., Janzen, E., Hosseinibarkooie, S., Löhr, H., Schneider, S., et al. (2018). Biallelic CHP1 mutation causes human autosomal recessive ataxia by impairing NHE1 function. *Neurol. Genet.* 4:e209. doi: 10.1212/NXG.0000000000000209

Mitsuhashi, S., and Matsumoto, N. (2020). Long-read sequencing for rare human genetic diseases. *J. Hum. Genet.* 65, 11–19. doi: 10.1038/s10038-019-0671-8

Mollet, J., Delahodde, A., Serre, V., Chretien, D., Schlemmer, D., Lombes, A., et al. (2008). CABC1 gene mutations cause ubiquinone deficiency with cerebellar ataxia and seizures. *Am. J. Hum. Genet.* 82, 623–630. doi: 10.1016/j.ajhg.2007.12.022

Morais, S., Raymond, L., Mairey, M., Coutinho, P., Brandão, E., Ribeiro, P., et al. (2017). Massive sequencing of 70 genes reveals a myriad of missing genes or mechanisms to be uncovered in hereditary spastic paraplegias. *Eur. J. Hum. Genet.* 25, 1217–1228. doi: 10.1038/ejhg.2017.124

Moreno-Cabrera, J. M., del Valle, J., Castellanos, E., Feliubadaló, L., Pineda, M., Brunet, J., et al. (2020). Evaluation of CNV detection tools for NGS panel data in genetic diagnostics. *Eur. J. Hum. Genet.* 28, 1645–1655. doi: 10.1038/s41431-020-0675-z

Moreno-De-Luca, A., Helmers, S. L., Mao, H., Burns, T. G., Melton, A. M. A., Schmidt, K. R., et al. (2011). Adaptor protein complex-4 (AP-4) deficiency causes a novel autosomal recessive cerebral palsy syndrome with microcephaly and intellectual disability. *J. Med. Genet.* 48, 141–144. doi: 10.1136/jmg.2010.082263

Moresco, E. M. Y., Li, X., and Beutler, B. (2013). Going forward with genetics: recent technological advances and forward genetics in mice. *Am. J. Pathol.* 182, 1462–1473. doi: 10.1016/j.ajpath.2013.02.002

Ngo, K. J., Rexach, J. E., Lee, H., Petty, L. E., Perlman, S., Valera, J. M., et al. (2020). A diagnostic ceiling for exome sequencing in cerebellar ataxia and related neurological disorders. *Hum. Mutat.* 41, 487–501. doi: 10.1002/humu.23946

Nguengang Wakap, S., Lambert, D. M., Olry, A., Rodwell, C., Gueydan, C., Lanneau, V., et al. (2020). Estimating cumulative point prevalence of rare diseases: analysis of the Orphanet database. *Eur. J. Hum. Genet.* 28, 165–173. doi: 10.1038/s41431-019-0508-0

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, 506–511. doi: 10.1126/science.1247363

Omenn, G. S., Lane, L., Lundberg, E. K., Overall, C. M., and Deutsch, E. W. (2017). Progress on the HUPO draft human proteome: 2017 metrics of the human proteome project. *J. Proteome Res.* 16, 4281–4287. doi: 10.1021/acs.jproteome.7b00375

Orthmann-Murphy, J. L., Salsano, E., Abrams, C. K., Bizzi, A., Uziel, G., Freidin, M. M., et al. (2009). Hereditary spastic paraplegia is a novel phenotype for GJA12/GJC2 mutations. *Brain* 132, 426–438. doi: 10.1093/brain/awn328

Ouahchi, K., Arita, M., Kayden, H., Hentati, F., Hamida, M. B., Sokol, R., et al. (1995). Ataxia with isolated vitamin E deficiency is caused by mutations in the α–tocopherol transfer protein. *Nat. Genet.* 9, 141–145. doi: 10.1038/ng0295-141

Palermo, A., Huan, T., Rinehart, D., Rinschen, M. M., Li, S., O'Donnell, V. B., et al. (2020). Cloud-based archived metabolomics data: a resource for in-source fragmentation/annotation, meta-analysis and systems biology. *Anal. Sci. Adv.* 1, 70–80. doi: 10.1002/ansa.202000042

Parodi, L., Coarelli, G., Stevanin, G., Brice, A., and Durr, A. (2018). Hereditary ataxias and paraparesias: clinical and genetic update. *Curr. Opin. Neurol.* 31, 462–471. doi: 10.1097/WCO.0000000000000585

Paulson, H. (2018). "Repeat expansion diseases" in *Handbook of clinical neurology*. eds. D. H. Geschwind, H. L. Paulson and C. Klein (Elsevier B.V.), 105–123.

Pena, L. D. M., Jiang, Y. H., Schoch, K., Spillmann, R. C., Walley, N., Stong, N., et al. (2018). Looking beyond the exome: a phenotype-first approach to molecular diagnostic resolution in rare and undiagnosed diseases. *Genet. Med.* 20, 464–469. doi: 10.1038/gim.2017.128

Pengelly, R. J., Ward, D., Hunt, D., Mattocks, C., and Ennis, S. (2020). Comparison of Mendeliome exome capture kits for use in clinical diagnostics. *Sci. Rep.* 10:3235. doi: 10.1038/s41598-020-60215-y

Perez-Lloret, S., van de Warrenburg, B., Rossi, M., Rodríguez-Blázquez, C., Zesiewicz, T., Saute, J. A. M., et al. (2020). Assessment of ataxia rating scales and cerebellar functional tests: critique and recommendations. *Mov. Disord.* 36, 283–297. doi: 10.1002/mds.28313

Platt, C. D., Zaman, F., Bainter, W., Stafstrom, K., Almutairi, A., Reigle, M., et al. (2020). Efficacy and economics of targeted panel versus whole-exome sequencing in 878 patients with suspected primary immunodeficiency. *J. Allergy Clin. Immunol.* 147, 723–726. doi: 10.1016/j.jaci.2020.08.022

Pode-Shakked, B., Finezilber, Y., Levi, Y., Liber, S., Fleischer, N., Greenbaum, L., et al. (2020). Shared facial phenotype of patients with mucolipidosis type IV: a clinical observation reaffirmed by next generation phenotyping. *Eur. J. Med. Genet.* 63:103927. doi: 10.1016/j.ejmg.2020.103927

Posey, J. E. (2019). Genome sequencing and implications for rare disorders. *Orphanet J. Rare Dis.* 14:153. doi: 10.1186/s13023-019-1127-0

Rexach, J., Lee, H., Martinez-Agosto, J. A., Németh, A. H., and Fogel, B. L. (2019). Clinical application of next-generation sequencing to the practice of neurology. *Lancet Neurol.* 18, 492–503. doi: 10.1016/S1474-4422(19)30033-X

Rodriguez, C. M., and Todd, P. K. (2019). New pathologic mechanisms in nucleotide repeat expansion disorders. *Neurobiol. Dis.* 130:104515. doi: 10.1016/j.nbd.2019.104515

Roos, A., Thompson, R., Horvath, R., Lochmüller, H., and Sickmann, A. (2018). Intersection of proteomics and genomics to "solve the unsolved" in rare disorders such as neurodegenerative and neuromuscular diseases. *Proteomics Clin. Appl.* 12:1700073. doi: 10.1002/prca.201700073

Rouleau, G. A., Ross, J. P., and Dion, P. A. (2020). Exome sequencing in genetic disease: recent advances and considerations. *F1000Research* 9:336. doi: 10.12688/f1000research.19444.1

Ruano, L., Melo, C., Silva, M. C., and Coutinho, P. (2014). The global epidemiology of hereditary ataxia and spastic paraplegia: a systematic review of prevalence studies. *Neuroepidemiology* 42, 174–183. doi: 10.1159/000358801

Sagnelli, A., Magri, S., Farina, L., Chiapparini, L., Marotta, G., Tonduti, D., et al. (2016). Early-onset progressive spastic paraplegia caused by a novel TUBB4A mutation: brain MRI and FDG-PET findings. *J. Neurol.* 263, 591–593. doi: 10.1007/s00415-016-8020-8

Sánchez-Ferrero, E., Coto, E., Beetz, C., Gámez, J., Corao, A., Díaz, M., et al. (2013). SPG7 mutational screening in spastic paraplegia patients supports a dominant effect for some mutations and a pathogenic role for p.A510V. *Clin. Genet.* 83, 257–262. doi: 10.1111/j.1399-0004.2012.01896.x

Sarzi, E., Goffart, S., Serre, V., Chrétien, D., Slama, A., Munnich, A., et al. (2007). Twinkle helicase (PEO1) gene mutation causes mitochondrial DNA depletion. *Ann. Neurol.* 62, 579–587. doi: 10.1002/ana.21207

Segal, M. M., George, R., Waltman, P., El-Hattab, A. W., James, K. N., Stanley, V., et al. (2020). Clinician-centric diagnosis of rare genetic diseases: performance of a gene pertinence metric in decision support for clinicians. *Orphanet J. Rare Dis.* 15:191. doi: 10.1186/s13023-020-01461-1

Seixas, A. I., Loureiro, J. R., Costa, C., Ordóñez-Ugalde, A., Marcelino, H., Oliveira, C. L., et al. (2017). A pentanucleotide ATTTC repeat insertion in the non-coding region of DAB1, mapping to SCA37, causes spinocerebellar ataxia. *Am. J. Hum. Genet.* 101, 87–103. doi: 10.1016/j.ajhg.2017.06.007

Sharony, R., Martins, S., Costa, I. P. D., Zaltzman, R., Amorim, A., Sequeiros, J., et al. (2019). Yemenite-Jewish families with Machado-Joseph disease (MJD/SCA3) share a recent common ancestor. *Eur. J. Hum. Genet.* 27, 1731–1737. doi: 10.1038/s41431-019-0449-7

Shribman, S., Reid, E., Crosby, A. H., Houlden, H., and Warner, T. T. (2019). Hereditary spastic paraplegia: from diagnosis to emerging therapeutic approaches. *Lancet Neurol.* 18, 1136–1146. doi: 10.1016/S1474-4422(19)30235-2

Simons, C., Wolf, N. I., McNeil, N., Caldovic, L., Devaney, J. M., Takanohashi, A., et al. (2013). A de novo mutation in the β-tubulin gene TUBB4A results in the leukoencephalopathy hypomyelination with atrophy of the basal ganglia and cerebellum. *Am. J. Hum. Genet.* 92, 767–773. doi: 10.1016/j.ajhg.2013.03.018

Singh, J. (2020). Harnessing the power of genetics: fast forward genetics in *Caenorhabditis elegans*. *Mol. Gen. Genomics.* 296, 1–20. doi: 10.1007/s00438-020-01721-6

Słabicki, M., Theis, M., Krastev, D. B., Samsonov, S., Mundwiller, E., Junqueira, M., et al. (2010). A genome-scale DNA repair RNAi screen identifies SPG48 as a novel gene associated with hereditary spastic paraplegia. *PLoS Biol.* 8:e1000408. doi: 10.1371/journal.pbio.1000408

Sobreira, N., Schiettecatte, F., Valle, D., and Hamosh, A. (2015). GeneMatcher: a matching tool for connecting investigators with an interest in the same gene. *Hum. Mutat.* 36, 928–930. doi: 10.1002/humu.22844

Srivastava, S., Love-Nichols, J. A., Dies, K. A., Ledbetter, D. H., Martin, C. L., Chung, W. K., et al. (2019). Meta-analysis and multidisciplinary consensus statement: exome sequencing is a first-tier clinical diagnostic test for individuals with neurodevelopmental disorders. *Genet. Med.* 21, 2413–2421. doi: 10.1038/s41436-019-0554-6

Subramanian, I., Verma, S., Kumar, S., Jere, A., and Anamika, K. (2020). Multi-omics data integration, interpretation, and its application. *Bioinform. Biol. Insights* 14:1177932219899051. doi: 10.1177/1177932219899051

Sud, M., Fahy, E., Cotter, D., Azam, K., Vadivelu, I., Burant, C., et al. (2016). Metabolomics workbench: an international repository for metabolomics data and metadata, metabolite standards, protocols, tutorials and training, and analysis tools. *Nucleic Acids Res.* 44, D463–D470. doi: 10.1093/nar/gkv1042

Suwinski, P., Ong, C. K., Ling, M. H. T., Poh, Y. M., Khan, A. M., and Ong, H. S. (2019). Advancing personalized medicine through the application of whole exome sequencing and big data analytics. *Front. Genet.* 10:49. doi: 10.3389/fgene.2019.00049

Synofzik, M., and Schüle, R. (2017). Overcoming the divide between ataxias and spastic paraplegias: shared phenotypes, genes, and pathways. *Mov. Disord.* 32, 332–345. doi: 10.1002/mds.26944

Tam, V., Patel, N., Turcotte, M., Bossé, Y., Paré, G., and Meyre, D. (2019). Benefits and limitations of genome-wide association studies. *Nat. Rev. Genet.* 20, 467–484. doi: 10.1038/s41576-019-0127-1

Thuriot, F., Buote, C., Gravel, E., Chénier, S., Désilets, V., Maranda, B., et al. (2018). Clinical validity of phenotype-driven analysis software PhenoVar as a diagnostic aid for clinical geneticists in the interpretation of whole-exome sequencing data. *Genet. Med.* 20, 942–949. doi: 10.1038/gim.2017.239

Trottier, Y., Lutz, Y., Stevanin, G., Imbert, G., Devys, D., Cancel, G., et al. (1995). Polyglutamine expansion as a pathological epitope in huntington's disease and four dominant cerebellar ataxias. *Nature* 378, 403–406. doi: 10.1038/378403a0

Trouillas, P., Takayanagi, T., Hallett, M., Currier, R. D., Subramony, S. H., Wessel, K., et al. (1997). International cooperative ataxia rating scale for pharmacological assessment of the cerebellar syndrome. *J. Neurol. Sci.* 145, 205–211. doi: 10.1016/S0022-510X(96)00231-6

Utine, G. E., Haliloğlu, G., Salanci, B., Çetinkaya, A., Kiper, P. Ö., Alanay, Y., et al. (2013). A homozygous deletion in GRID2 causes a human phenotype with cerebellar ataxia and atrophy. *J. Child Neurol.* 28, 926–932. doi: 10.1177/0883073813484967

van der Knaap, M. S., Leegwater, P. A. J., Könst, A. A. M., Visser, A., Naidu, S., Oudejans, C. B. M., et al. (2002). Mutations in each of the five subunits of translation initiation factor eIF2B can cause leukoencephalopathy with vanishing white matter. *Ann. Neurol.* 51, 264–270. doi: 10.1002/ana.10112

Wagner, M., Osborn, D. P. S., Gehweiler, I., Nagel, M., Ulmer, U., Bakhtiari, S., et al. (2019). Bi-allelic variants in RNF170 are associated with hereditary spastic paraplegia. *Nat. Commun.* 10:4790. doi: 10.1038/s41467-019-12620-9

Wang, M., Carver, J. J., Phelan, V. V., Sanchez, L. M., Garg, N., Peng, Y., et al. (2016). Sharing and community curation of mass spectrometry data with global natural products social molecular networking. *Nat. Biotechnol.* 34, 828–837. doi: 10.1038/nbt.3597

Wang, D., Eraslan, B., Wieland, T., Hallström, B., Hopf, T., Zolg, D. P., et al. (2019). A deep proteome and transcriptome abundance atlas of 29 healthy human tissues. *Mol. Syst. Biol.* 15:e8503. doi: 10.15252/msb.20188503

Wenger, A. M., Peluso, P., Rowell, W. J., Chang, P. C., Hall, R. J., Concepcion, G. T., et al. (2019). Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome. *Nat. Biotechnol.* 37, 1155–1162. doi: 10.1038/s41587-019-0217-9

Winterthun, S., Ferrari, G., He, L., Taylor, R. W., Zeviani, M., Turnbull, D. M., et al. (2005). Autosomal recessive mitochondrial ataxic syndrome due to mitochondrial polymerase γ mutations. *Neurology* 64, 1204–1208. doi: 10.1212/01.WNL.0000156516.77696.5A

Wishart, D. S. (2016). Emerging applications of metabolomics in drug discovery and precision medicine. *Nat. Rev. Drug Discov.* 15, 473–484. doi: 10.1038/nrd.2016.32

Wright, C. F., FitzPatrick, D. R., and Firth, H. V. (2018). Paediatric genomics: diagnosing rare disease in children. *Nat. Rev. Genet.* 19, 253–268. doi: 10.1038/nrg.2017.116

Yang, Z., Shi, C., Zhou, L., Li, Y., Yang, J., Liu, Y., et al. (2019). Metabolic profiling reveals biochemical pathways and potential biomarkers of spinocerebellar ataxia 3. *Front. Mol. Neurosci.* 12:159. doi: 10.3389/fnmol.2019.00159

Yépez, V. A., Mertes, C., Müller, M. F., Klaproth-Andrade, D., Wachutka, L., Frésard, L., et al. (2021). Detection of aberrant gene expression events in RNA sequencing data. *Nat. Protoc.* 16, 1276–1296. doi: 10.1038/s41596-020-00462-5

Ylikallio, E., Kim, D., Isohanni, P., Auranen, M., Kim, E., Lönnqvist, T., et al. (2015). Dominant transmission of de novo KIF1A motor domain variant underlying pure spastic paraplegia. *Eur. J. Hum. Genet.* 23, 1427–1430. doi: 10.1038/ejhg.2014.297

Zanni, G., Calì, T., Kalscheuer, V. M., Ottolini, D., Barresi, S., Lebrun, N., et al. (2012). Mutation of plasma membrane Ca^{2+} ATPase isoform 3 in a family with X-linked congenital cerebellar ataxia impairs Ca^{2+} homeostasis. *Proc. Natl. Acad. Sci. U. S. A.* 109, 14514–14519. doi: 10.1073/pnas.1207488109

Zhao, M., Havrilla, J. M., Fang, L., Chen, Y., Peng, J., Liu, C., et al. (2020). Phen2Gene: rapid phenotype-driven gene prioritization for rare diseases. *NAR Genomics Bioinform.* 2:lqaa032. doi: 10.1093/nargab/lqaa032

Zivony-Elboum, Y., Westbroek, W., Kfir, N., Savitzki, D., Shoval, Y., Bloom, A., et al. (2012). A founder mutation in Vps37A causes autosomal recessive complex hereditary spastic paraparesis. *J. Med. Genet.* 49, 462–472. doi: 10.1136/jmedgenet-2012-100742

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 in Genetics

ORIGINAL RESEARCH
published: 02 July 2021
doi: 10.3389/fgene.2021.637945

Novel Reclassification of Adult Diabetes Is Useful to Distinguish Stages of β-Cell Function Linked to the Risk of Vascular Complications: The DOLCE Study From Northern Ukraine

*Olena Fedotkina[1], Oksana Sulaieva[2,3], Turkuler Ozgumus[1], Liubov Cherviakova[4], Nadiya Khalimon[5], Tetiana Svietleisha[6], Tetiana Buldenko[7], Emma Ahlqvist[2], Olof Asplund[2], Leif Groop[2], Peter M. Nilsson[2] and Valeriya Lyssenko[1,2]**

[1] Department of Clinical Science, Center for Diabetes Research, University of Bergen, Bergen, Norway, [2] Lund University Diabetes Center, Department of Clinical Sciences, Lund University, Skåne University Hospital, Malmö, Sweden, [3] Medical Laboratory CSD, Kyiv, Ukraine, [4] Chernihiv Regional Hospital, Chernihiv, Ukraine, [5] City Hospital Mo. 2, Chernihiv, Ukraine, [6] City Hospital No., Chernihiv, Ukraine, [7] Department of Health Care of Chernihiv Regional State Administration, Chernihiv, Ukraine

OPEN ACCESS

Edited by:
Ayman Hussein,
University of Khartoum, Sudan

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

***Correspondence:**
Valeriya Lyssenko
Valeriya.Lyssenko@uib.no

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

Received: 04 December 2020
Accepted: 03 June 2021
Published: 02 July 2021

Citation:
Fedotkina O, Sulaieva O, Ozgumus T, Cherviakova L, Khalimon N, Svietleisha T, Buldenko T, Ahlqvist E, Asplund O, Groop L, Nilsson PM and Lyssenko V (2021) Novel Reclassification of Adult Diabetes Is Useful to Distinguish Stages of β-Cell Function Linked to the Risk of Vascular Complications: The DOLCE Study From Northern Ukraine. Front. Genet. 12:637945. doi: 10.3389/fgene.2021.637945

Background: Presently, persons with diabetes are classified as having type 1 (T1D) or type 2 diabetes (T2D) based on clinical diagnosis. However, adult patients exhibit diverse clinical representations and this makes treatment approaches challenging to personalize. A recent Scandinavian study proposed a novel classification of adult diabetes into five clusters based on disease pathophysiology and risk of vascular complications. The current study aimed to characterize new subgroups of adult diabetes using this strategy in a defined population from northern Ukraine.

Methods: We analyzed 2,140 patients with established diabetes from the DOLCE study (n = 887 with new-onset diabetes and n = 1,253 with long duration). We used the k-means approach to perform clustering analyses using BMI, age at onset of diabetes, HbA_{1c}, insulin secretion (HOMA2-B), and insulin resistance (HOMA2-IR) indices and glutamic acid decarboxylase antibodies (GADA) levels. Risks of macro- (myocardial infarction or stroke) and microvascular [retinopathy, chronic kidney disease (CKD) and neuropathy] complications and associations of genetic variants with specific clusters were studied using logistic regression adjusted for age, sex, and diabetes duration.

Results: Severe autoimmune diabetes (SAID, 11 and 6%) and severe insulin-deficient diabetes (SIDD, 25 and 14%) clusters were twice as prevalent in patients with long-term as compared to those with new-onset diabetes. Patients with long duration in both SAID and SIDD clusters had highest risks of proliferative retinopathy, and elevated risks of CKD. Long-term insulin-resistant obese diabetes 1 (IROD1) subgroup had elevated risks of CKD, while insulin-resistant obese diabetes 2 (IROD2) cluster exhibited the highest HOMA2-B, lowest HbA_{1c}, and lower prevalence of all microvascular complications as compared to all other clusters. Genetic analyses of IROD2 subgroup identified reduced

frequency of the risk alleles in the *TCF7L2* gene as compared to all other clusters, cumulatively and individually (p = 0.0001).

Conclusion: The novel reclassification algorithm of patients with adult diabetes was reproducible in this population from northern Ukraine. It may be beneficial for the patients in the SIDD subgroup to initiate earlier insulin treatment or other anti-diabetic modalities to preserve β-cell function. Long-term diabetes cases with preserved β-cell function and lower risk for microvascular complications represent an interesting subgroup of patients for further investigations of protective mechanisms.

Keywords: clustering, β-cell function, diabetes complications, genetics, adult diabetes

INTRODUCTION

Diabetes represents a global health problem, affecting today more than 400 million people worldwide, which is estimated to increase up to 600 million by 2030 (IDF, 2019). Diabetes in adults, of which type 2 diabetes (T2D) being the most common type, comprises around 90% of all diabetes cases. The incidence and prevalence of adult diabetes are more rapidly increasing in the low- and middle-income countries (IDF, 2019). The official prevalence of diabetes in Ukraine was reported to rise by about 25% from 2007 to 2019, reaching 8.4% in 2019 (Khalangot and Tronko, 2007; IDF, 2019). This situation is particularly alarming because nearly an equal number of people have undiagnosed diabetes in Ukraine meaning that the prevalence rate is doubled in real life (IDF, 2019; Mankovsky, 2020). Diabetes is one of the leading causes of blindness, end-stage renal disease, limb amputation, heart disease, stroke, liver cirrhosis, and premature death in working age adults (IDF, 2019). Numerous clinical and genetic studies demonstrated that adult diabetes is a highly heterogeneous metabolic disease with diverse underlying mechanisms (Tuomi et al., 2014; McCarthy, 2017). Lack of pathophysiology-based classification hampers personalized therapy targeting specific complications of diabetes (Fitipaldi et al., 2018). This shortcoming was recently addressed in a Scandinavian cohort of patients with newly diagnosed adult diabetes, by applying an unbiased unsupervised clustering approach to propose a reclassification of the disease (Ahlqvist et al., 2018). Based on six clinically affordable parameters including glycated hemoglobin (HbA_{1c}), body-mass index (BMI), age at diabetes onset, insulin resistance, insulin secretion calculated using homeostasis model assessment (HOMA2), and glutamic acid decarboxylase antibody (GADA) levels, a new five-cluster classification scheme for adult diabetes was proposed. These clusters reflected essential pathophysiological mechanisms in the disease and were named accordingly as severe autoimmune diabetes (SAID), severe insulin-deficient diabetes (SIDD), severe insulin-resistant diabetes (SIRD), mild obesity-related diabetes (MOD), and mild age-related diabetes (MARD). This novel classification has now been validated in a number of studies across Europe and Asia, emphasizing the robustness of the approach (Dennis et al., 2019; Zaharia et al., 2019; Zou et al., 2019) even though there was some ethnic heterogeneity reported in an Asian-Indian study (Anjana et al., 2020). The suggested clustering approach is attractive for clinical practice because of the opportunity to predict the risks of diabetic complications at diagnosis and tailor patients' therapy accordingly (Ahlqvist et al., 2018; Zaharia et al., 2019).

The aim of the present study was to assess the prognostic value of the new reclassification approach by Ahlqvist et al. (Ahlqvist et al., 2018) for the first time in an East-European population from Ukraine including patients with newly diagnosed and long-term adult diabetes.

PARTICIPANTS AND METHODS

Study Cohort

The Diagnostic Optimization and Treatment of Diabetes and its Complications in the Chernihiv Region, Ukraine (DOLCE) is a hospital- and primary health care-based study of individuals with diabetes and their healthy relatives, including total of 6,095 participants. The cohort consists of 785 persons with T1D, 4,297 with T2D, 62 with unspecified diabetes, and 951 healthy first- or second-degree relatives. All participants completed a questionnaire supervised by an endocrinologist and a trained diabetes nurse, which covered the person's medical history and included information of family history of diabetes, anthropometric measurements (weight, height, and blood pressure), alcohol intake, smoking, diabetes medication, antihypertensive and lipid-lowering treatments. Information of prevalent cardiovascular events, neuropathy, chronic kidney disease, and stages of retinopathy was reported by primary care physicians using patients' hospital discharge records as primary source and was used as data entry into the DOLCE database at the screening visit. Fasting blood samples were drawn for plasma glucose and HbA_{1c} measurements. Additional plasma and serum samples were stored at –80°C for C-peptide, insulin, lipids, and glutamic acid decarboxylase antibodies (GADA) measurements, which were performed at the Department of Clinical Chemistry, Scania University Hospital, Malmö, Sweden. A written informed consent form was obtained from every participant. The DOLCE study was approved by the local ethics committees (approval number for Ukraine Dnr17/2011–09–14; for Norway 2019/28968).

Measurements and Calculations

C-peptide concentrations were measured with an electro-chemiluminescence immunoassay on Cobas e411 (Roche Diagnostics, Mannheim, Germany) or a radioimmunoassay (Human C-peptide RIA; Linco, St Charles, MO, United States; or Peninsula Laboratories, Belmont, CA, United States). GADA were measured with an ELISA from the samples collected at the screening visit. Test results greater than or equal to 5 U/ml were considered as positive. The radio binding assays had 62–88% sensitivity and 91–99% specificity, and the ELISA assay had 72% sensitivity and 99% specificity (Combinatorial Autoantibody or Diabetes/Islet Autoantibody Standardization Programs 1998–2013). β-cell function (HOMA2-B) and insulin resistance (HOMA2-IR) were assessed with Homoeostasis Model Assessment 2 (HOMA2) and were calculated with the HOMA2 calculator using C-peptide and fasting glucose measurements (Levy et al., 1998; HOMACalculator, 2020). Data values for BMI, HOMA2-B, and HOMA2-IR with more than three standard deviations were excluded. Only individuals with age at diabetes diagnosis older than 18 years and complete information on BMI, age at onset, HbA_{1c}, HOMA2-B, HOMA2-IR, sex, duration of diabetes, and complication status were included in the final analysis (n = 2,140).

Definition of Diabetic Complications

Proliferative diabetic retinopathy (PDR) was defined as having one of the following conditions: (a) proliferative retinopathy, (b) laser treatment, or (c) blindness of either one or both of the eyes. Stages of PDR were based on fundus photographs and were evaluated by ophthalmologists. Chronic kidney disease (CKD) was defined as having at least one of the following conditions: (a) estimated glomerular filtration rate less than 60 mL/min/1.73 m^2 (eGFR $<$ 60), (b) clinically documented diagnosis of nephropathy, dialysis, or end-stage renal disease (ESRD). eGFR was calculated using Modification of Diet in Renal Disease (MDRD) formula as: $186 \times (\text{Creatinine}/88.4)^{-1.154} \times (\text{age})^{-0.203} \times 0.742$ for female participants; $186 \times (\text{Creatinine}/88.4)^{-1.154} \times (\text{age})^{-0.203}$ for male participants (Mula-Abed et al., 2012). Neuropathy was defined as clinically diagnosed peripheral neuropathy. Cardiovascular diseases were defined as either presence of myocardial infarction or stroke using International Classification of Diseases (ICD)-10 codes I21, I24, and I61–I64, respectively.

Statistical Analysis

In SAID cluster we included individuals with positive GADA. SIDD, SIRD, MOD, and MARD cluster analysis was performed on individuals who were negative for GADA antibodies (where GADA measurements were available) using previously defined parameters by Ahlqvist et al. (Ahlqvist et al., 2018): HbA_{1c}, BMI, age at onset of diabetes, HOMA2-B, and HOMA2-IR. The analyses were done in two groups of patients: (i) long-term diabetes with more than 3 years of disease duration and (ii) new-onset diabetes with less than 3 years of disease duration to harmonize with the Swedish ANDIS cohort employed as the reference (Ahlqvist et al., 2018). The analyses were performed using k-means clustering on the centered and scaled values. We used "kmeansruns" function from R package fpc v2.2–8 with 4 as the assigned number of clusters and default parameters: krange = 4, criterion = "asw," iter.max = 100, runs = 100, alpha = 0.001, critout = FALSE (Hennig, 2020). Male and female patients were analyzed separately to avoid sex bias, and results were merged afterward. Risks for diabetic complications in each cluster were calculated using logistic regression adjusted for age, sex, and diabetes duration with the MARD cluster used as reference. To adjust for multiple phenotype testing, we calculated a false discovery rate (FDR). FDR was calculated using p.adjust function implemented in the "stats" package in R (R Core Team, 2021). FDR $<$ 0.05 was considered statistically significant.

Genetics

Genotyping of DNA samples was available only in patients with clinically defined T2D using Infinium Core Exome Chip InfiniumCoreExome-24v1-1 (https://www.illumina.com). Imputation was done using Michigan Imputation Server and Reference Panel HRC r1.1 2016. Standard quality control steps for Genome-Wide Association Studies (GWAS) were applied (Marees et al., 2018). We analyzed SNPs associated with T2D, which were previously reported in the DIAGRAM GWAS meta-analysis (Morris et al., 2012). Logistic regression adjusted for sex and age was used to study associations between the insulin resistant obese diabetes 2 cluster and the genetic variants as compared with all other clusters in patients with long-term diabetes. SAID cluster was not included in the genetic analysis due to small sample size of individuals with genotypes in this group (n = 26). Bonferroni correction was used to adjust for multiple testing in genetic association tests. $P < 0.05$ was considered statistically significant.

RESULTS

The clinical characteristics and prevalence of macro- and microvascular complications in adult patients with (a) new-onset and with (b) long-term diabetes are shown in **Table 1**.

New-Onset Adult Diabetes

There were 887 (36% men) with new-onset diabetes (years, 1.1 ± 1.1). Oral glucose-lowering treatment was reported for 42.6% of the patients, while insulin treatment was initiated in 8.8% of the patients. The prevalence of PDR was 0.7%, and CKD was 12.4%. Peripheral neuropathy was reported in 34.3%, while non-fatal CVD occurred in 8.6% of the individuals (**Table 1**).

Long-Term Adult Diabetes

There were 1,253 persons (34% men) with long diabetes duration (mean ± SD, years, 11.0 ± 6.9). BMI in this group was similar to the new-onset diabetes group (kg/m^2, 30.9 ± 5.3 and 30.7 ± 5.6). Oral glucose-lowering treatment was reported for 51.6% of the patients, while insulin treatment was initiated in 27.7% of the patients. Glycemic control was observed to be worse in patients with long-term diabetes (HbA_{1c}%, 9.0 ± 2.0 and 7.9 ± 2.1) than in those with new-onset disease. As expected, insulin secretion

TABLE 1 | Clinical characteristics of patients with adult diabetes in the DOLCE study.

Phenotype	New-onset	Long-term
N (men,%)	887 (36%)	1,253 (34%)
Age at visit, years	56.9 (12.5)	61.2 (10.0)
Age at onset of diabetes, years	55.8 (12.4)	50.1 (10.8)
Diabetes duration, years	1.1 (1.1)	11.0 (6.9)
HbA_{1c},%	7.9 (2.1)	9.0 (2.0)
HbA_{1c}, mmol/mol	62.9 (23.4)	74.4 (21.3)
BMI, kg/m^2	30.7 (5.6)	30.9 (5.3)
Waist, cm	97.4 (14.3)	99.1 (12.7)
HOMA2-B	81.1 (44.6)	63.9 (45.9)
HOMA2-IR	2.4 (1.2)	2.2 (1.3)
C-peptide, nmol/l	0.9 (0.5)	0.8 (0.5)
Without treatment,%	46.9%	12.1%
Tablets,%	42.6%	51.6%
Insulin,%	8.8%	27.7%
Tablets and insulin,%	1.7%	8.5%
Sulfonylurea,%	27%	45%
PDR,%	0.7%	5.5%
CKD,%	12.4%	26.7%
Neuropathy,%	34.3%	85.6%
CVD,%	8.6%	14.9%

Data are represented as mean ± (SD).
HOMA2-B, homoeostatic model assessment 2 estimates of β-cell function; HOMA2-IR, homoeostatic model assessment 2 estimates of insulin resistance; PDR, proliferative diabetic retinopathy; CKD, chronic kidney disease; CVD, cardiovascular disease.

estimated with HOMA2-B was lower in the persons with long-term diabetes (%, 63.9 ± 45.9 and 81.1 ± 44.6) as compared to those with new-onset (**Table 1**). The prevalence of PDR was 5.5%, CKD was 26.7%, peripheral neuropathy was 85.6%, and non-fatal CVD was 14.9% (**Table 1**).

Characteristics of Different Clusters in Patients With Adult Diabetes

Descriptive characteristics and frequency of the clusters in the patients with long-term and new-onset adult diabetes are presented in **Figures 1**, **2** and **Supplementary Table 1**.

Severe autoimmune diabetes (SAID) was twice prevalent in patients with long-term (11 and 6%) as compared to those with new-onset diabetes (**Figure 1**). Of these, in total 74 (37 and 43%) had clinical diagnosis as T2D. Individuals in this cluster had low BMI, younger age at diagnosis, and the lowest HOMA2-B and HOMA2-IR as compared to other clusters (**Figure 2** and **Supplementary Table 1**). Individuals with SAID most frequently had insulin treatment, particularly among those with the long-term diabetes (84.1 and 58.8%), while oral antidiabetic treatment was more frequent in the new-onset group (9.4 and 27.5%) (**Supplementary Table 1**).

Severe insulin-deficient diabetes (SIDD) was almost twice as prevalent in the patients with long diabetes duration than in those with new-onset diabetes (25 and 14%) (**Figure 1**). This cluster was characterized by low insulin secretion as shown by HOMA2-B, relatively low BMI and the high HbA_{1c} than the other clusters (**Figure 2** and **Supplementary Table 1**). The frequency of insulin therapy alone (42.6 and 25.6%) or in combination with oral anti glycemic medications (17.9 and 4.1%) was higher than in all other clusters apart from SAID for the patients with long-term and new-onset disease (**Supplementary Table 1**).

Severe insulin-resistant diabetes (SIRD) cluster occurred with 27% frequency in patients with new-onset diabetes (**Figure 1**). This cluster in the new-onset group was characterized by the highest insulin-resistance index HOMA2-IR (%, 3.4 ± 1.1) and HOMA2-B (%, 121.7 ± 42.2), elevated BMI and lower HbA_{1c} (**Figure 2** and **Supplementary Table 1**). The insulin resistant obese patients with long-term diabetes differed from the original SIRD individuals. They exhibit similarly elevated HOMA2-IR (%, 3.5 ± 1.1) while insulin secretion index HOMA2-B was lower (%, 65.7 ± 27.8) and therefore we named this cluster insulin resistant obese diabetes 1 (IROD1) (**Figure 2** and **Supplementary Table 1**). IROD1 cluster also showed the highest waist circumference among all clusters, reflecting the presence of an abdominal adiposity (**Supplementary Table 1**). Oral anti-glycemic treatment was common in long-term and new-onset groups (72.2 and 48.3%), however, long-term IROD1 reported more frequent use of sulfonylurea (61%). In addition, half of the individuals with new-onset diabetes (49.1%) and only 8.6% with long-term did not receive any anti-diabetic medications.

Mild obesity-related diabetes (MOD) cluster in new-onset group occurred with the frequency of 7% and was the smallest after SAID cluster (**Figure 1**). This group in the new-onset group was characterized by elevated BMI (kg/m^2, 34.0 ± 4.9), HOMA2-IR (%, 3.3 ± 1.2), and moderately elevated HOMA2-B index (%, 66.9 ± 23.8). Patients with long-term diabetes in this cluster showed elevated BMI (kg/m^2, 32.0 ± 4.5) and HOMA2-IR (%, 3.1 ± 1.1), but did not match their original MOD counterparts in the new-onset group by markedly elevated HOMA2-B index (%, 139.4 ± 38.8), and therefore we named them insulin resistant obese diabetes 2 (IROD2) (**Figure 2** and **Supplementary Table 1**). Patients in both groups received often oral anti-hyperglycemic treatment (64.7 and 70.8%), or were controlled with diet and/or lifestyle intervention (27.9 and 23.6%) of long-term and new-onset diabetes, respectively (**Supplementary Table 1**).

Finally, mild age-related diabetes (MARD) was the largest cluster in both long-term and new-onset adult diabetes (27 and 46%) (**Figure 1**). This cluster was by definition characterized by the highest age at disease diagnosis (years, 59.7 ± 11.7 and 59.9 ± 11.7) and had similar characteristics across the two patients' groups (**Figure 2** and **Supplementary Table 1**).

Risk of Diabetes Complications in Different Clusters

We assessed the risk of macro- and microvascular complications of diabetes in each cluster using MARD as reference group, and adjusted for age, sex, and diabetes duration. In this cross-sectional study, patients with new-onset adult diabetes had few PDR events, and therefore analysis was not performed to calculate PDR risk in these patients. Only few cases (4.2%) had CKD

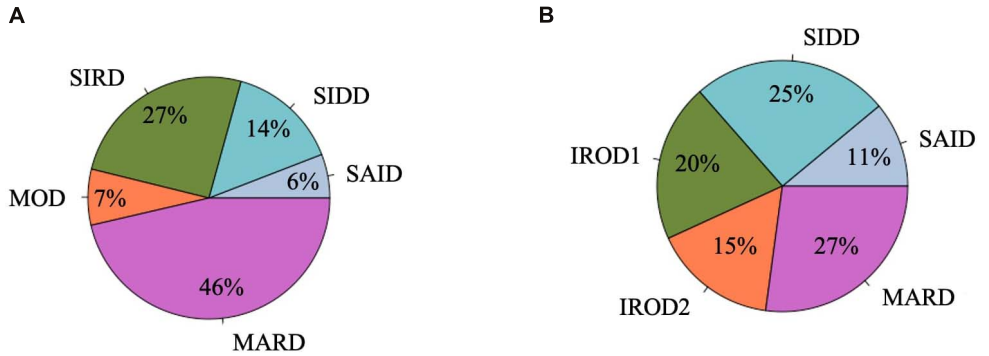


FIGURE 1 | Frequency distribution of clusters in the DOLCE study. **(A)** Individuals with new-onset adult diabetes with less than 3 years of disease duration (n = 887). **(B)** Individuals with long-term adult diabetes with more than 3 years of disease duration (n = 1,253).

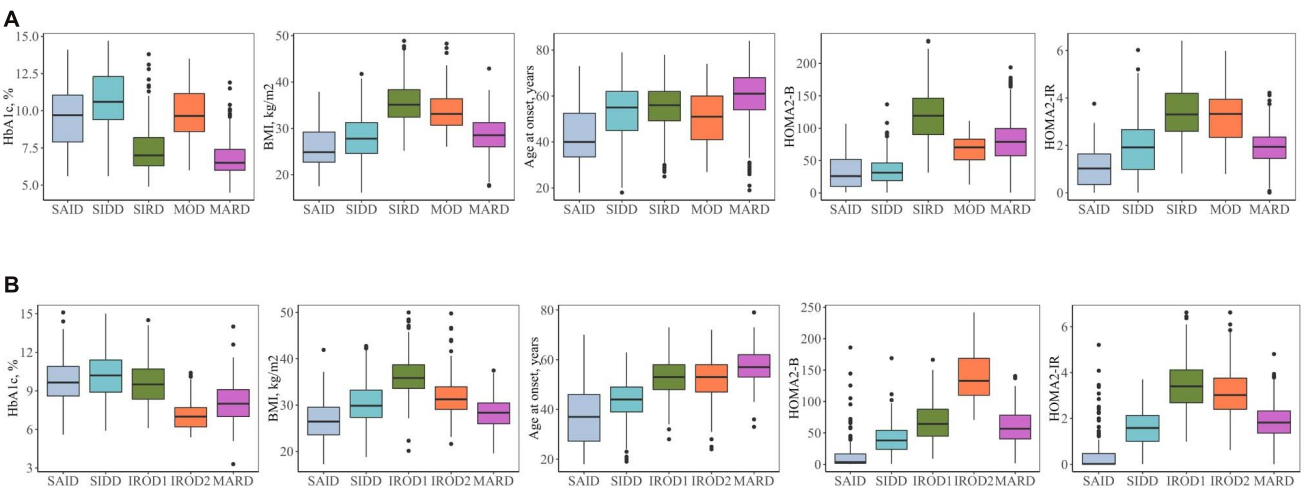


FIGURE 2 | Clinical characteristics of the clusters. **(A)** Individuals with new-onset adult diabetes with less than 3 years of disease duration (n = 887). **(B)** Individuals with long-term adult diabetes with more than 3 years of disease duration (n = 1,253). SAID, severe autoimmune diabetes**;** SIDD, severe insulin-deficient diabetes; SIRD, severe insulin-resistant diabetes; MOD, mild obesity-related diabetes; MARD, mild age-related diabetes; IROD1 and 2, insulin-resistant obese diabetes 1 and 2; HOMA2-B, homoeostatic model assessment two estimates of β-cell function; HOMA2-IR, homoeostatic model assessment two estimates of insulin resistance.

in MOD cluster with the new onset diabetes (**Figure 3** and **Supplementary Table 1**). Similar to the reference ANDIS study (Ahlqvist et al., 2018), in long-term group the SAID cluster was characterized with high prevalence of PDR (10.9%) with OR of 9.32-fold (95% CI, 2.15–40.46, p = 0.003) relative to MARD cluster. Prevalence of CVD in this cluster was detected to be lower than in the other clusters (5.8%). The SIDD cluster had similarly to SAID high prevalence and increased risk of PDR (11%, OR 2.42, 95% CI, 0.96– 6.07, p = 0.06) relative to MARD cluster (**Figure 3**, **Table 2**, and **Supplementary Table 1**). Prevalence of CKD was found to be elevated in all severe clusters, i.e., SAID (29.7%), SIDD (32%), and insulin-resistant obese diabetes 1 (IROD1) (30.2%) conferring an increased risk of 2.59-fold (1.34–5.00, p = 0.005), 2.03-fold (1.21– 3.40, $p = 7.1 \times 10^{-3}$), and 1.63-fold (1.08 –2.46, p = 0.02) relative to MARD in patients with long diabetes duration. In contrast to ANDIS, neuropathy prevalence was high across all clusters, but the relative risk was highest in SIDD cluster (OR 13.60, 95% CI 5.20–35.50, $p = 1.10 \times 10^{-7}$). In general, the IROD2 cluster exhibited the lowest prevalence of all microvascular complications, particularly CKD and neuropathy.

Genetic Analyses

We analyzed variants reported in the DIAGRAM GWAS meta-analyses (Morris et al., 2012) to provide insights on the key T2D SNPs associated with protective phenotype of insulin resistant obese (IROD2) cluster in patients with long-term diabetes relative to all other clusters (excluding SAID due to the lack of genetic information in this cluster). The top SNP was rs7903146 in *TCF7L2* (OR, 95%CI; 0.54, 0.39–0.74, p = 0.0001), which showed significantly lower frequency of the risk T-allele in IROD2 than in IROD1 (p = 0.008), SIDD (p = 0.001) and MARD (p = 0.0003) (**Table 3**). The same directionality for lower risk allele frequencies in IROD2 was also found for *KCNQ1* locus (rs163184, 0.67, 0.52–0.88, p = 0.003), *NOTCH2* locus (rs1493694, 0.49, 0.30– 0.80, p = 0.004), locus in the gene related to cell-matrix interplay *ADAMTS9* (rs6795735: 0.72, 0.56– 0.93,

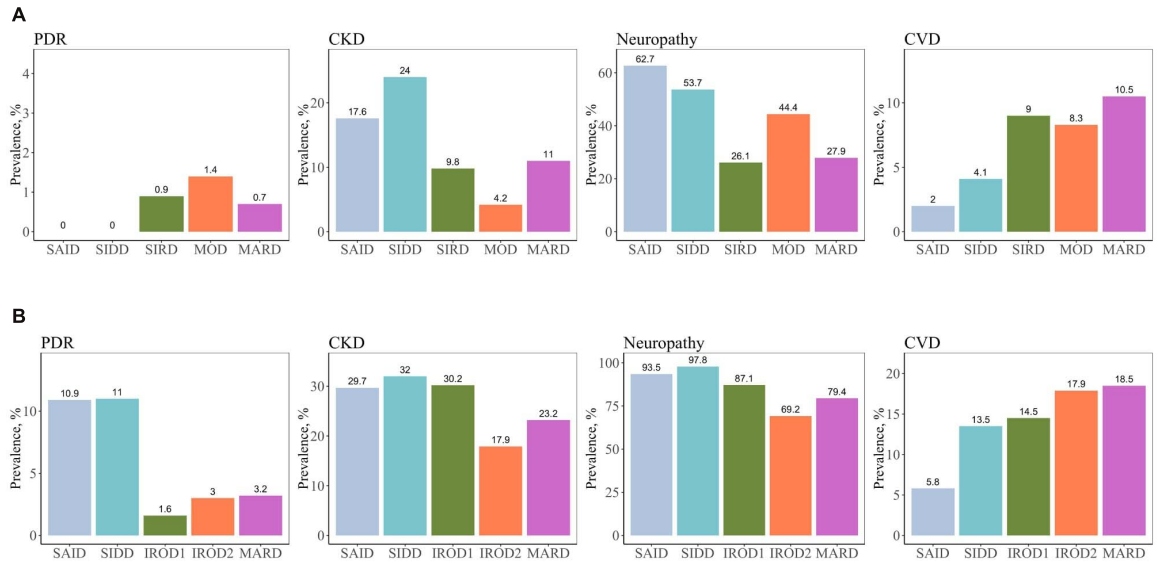


FIGURE 3 | Prevalence of macro- and microvascular complications in different clusters of adult patients with the new-onset and long-term diabetes. **(A)** Individuals with new-onset adult diabetes with less than 3 years of disease duration (n = 887). **(B)** Individuals with long-term adult diabetes with more than 3 years of disease duration (n = 1,253). PDR, proliferative diabetic retinopathy; CKD, chronic kidney disease; CVD, cardiovascular disease.

TABLE 2 | Risk of macro- and microvascular complications in different clusters in long-term adult diabetes relative to MARD.

Complications	SAID		SIDD		IROD1		IROD2	
	OR (95% CI)	*P*-value	OR (95% CI)	*P*-value	OR (95% CI)	*P*-value	OR (95% CI)	*P*-value
PDR	9.32 (2.15, 40.46)	0.003*	2.42 (0.96, 6.07)	0.06	0.53 (0.16, 1.79)	0.31	0.73 (0.23, 2.32)	0.59
CKD	2.59 (1.34, 5.00)	0.005*	2.03 (1.21, 3.40)	7.1×10^{-3}*	1.63 (1.08, 2.46)	0.02*	0.85 (0.53, 1.39)	0.52
Neuropathy	3.30 (1.29, 8.43)	0.01*	13.60 (5.20, 35.50)	1.0×10^{-7}*	2.08 (1.28, 3.37)	3.0×10^{-3}*	0.70 (0.45, 1.09)	0.12
CVD	0.61 (0.25, 1.51)	0.3	1.13 (0.60, 2.11)	0.71	1.11 (0.68, 1.8)	0.68	1.44 (0.87, 2.36)	0.15

PDR, proliferative diabetic retinopathy; CKD, chronic kidney disease; CVD, cardiovascular disease; OR, Odds Ratios; 95% CI, 95% confidence intervals; and p-values were calculated using logistic regression adjusted for sex, age, and diabetes duration.
**Significant after adjustment for the multiple testing using FDR (<0.05).*

p = 0.012), loci in genes *ZFAND6* (rs11634397: 0.73, 0.56–0.94, p = 0.016), and *ZFAND3* (rs4299828: 0.73, 0.54– 0.99, p = 0.042). On the contrary, higher frequencies were observed for the risk alleles of SNPs in the gene related to cell signaling *PTPRD* (rs16927668: 1.40, 1.03–1.91, p = 0.033). However, after adjustment for multiple testing using Bonferroni correction, only variant rs7903146 in *TCF7L2* remained statistically significant (p-value $_{IROD2vs.allclusters}$ = 0.008, p-value $_{IROD2vs.SIDD}$ = 0.048, p-value $_{IROD2vs.MARD}$ = 0.002) (**Table 3**).

DISCUSSION

The findings from this observational study demonstrate that the novel approach using SAID, SIDD, SIRD, MOD, and MARD clustering of diabetes subgroups (Ahlqvist et al., 2018) is replicated in the patients with adult diabetes from northern Ukraine. In accordance with the published studies, the SIDD cluster had the highest prevalence of retinopathy and neuropathy, and the insulin resistant subgroups were linked to high risk of CKD. In contrast to Scandinavian and German populations (Ahlqvist et al., 2018; Zaharia et al., 2019) and similar to the insulin-deficient insulin-resistant subgroup in a recently published Asian-Indian cohort (Anjana et al., 2020), SIDD cluster in this cohort also showed high risk of CKD. With longer duration of diabetes, the clusters might change, and insulin resistant obese cases could be challenging to match to the original SIRD and MOD cluster of corresponding new-onset diabetes. In general, patients with long-term diabetes and preserved β-cell function demonstrated better glycemic control measured by HbA_{1c} and lower risk of all microvascular complications than expected.

An important observation was lower insulin secretion in patients from northern Ukraine with new-onset adult diabetes compared to the Swedish ANDIS cohort. The history of Ukraine during the first half of the 20^{th} century could contribute to this difference. It has been previously reported that children born to parents exposed to the Ukrainian Holodomor famine (1932–1933) showed increased risk of developing T2D later in life (Lumey et al., 2015). Early life exposure to starvation might

TABLE 3 | Top T2D SNPs nominally associated with IROD2 cluster in individuals with long-term diabetes.

SNP	Gene	Chr	BP	Risk allele	RAF	IROD2 vs. all clusters		IROD2 vs. IROD1		IROD2 vs. SIDD		IROD2 vs. MARD	
						OR (95% CI)	*P*-value	OR (95% CI)	*P*-value	OR (95% CI)	*P*-value	OR (95% CI)	*P*-value
rs7903146	*TCF7L2*	10	114758349	T	0.28	0.54 (0.39, 0.74)	0.0001*	0.61 (0.42, 0.88)	0.008	0.54 (0.37, 0.77)	0.001*	0.45 (0.31, 0.65)	0.0003*
rs163184	*KCNQ1*	11	2847069	G	0.50	0.67 (0.52, 0.88)	0.003	0.60 (0.44, 0.82)	0.001	0.74 (0.54, 1.01)	0.059	0.71 (0.52, 0.98)	0.036
rs10923931	*NOTCH2*	1	120517959	T	0.13	0.49 (0.30, 0.80)	0.004	0.52 (0.29, 0.90)	0.020	0.50 (0.29, 0.87)	0.014	0.45 (0.26, 0.78)	0.004
rs6795735	*ADAMTS9*	3	64705365	C	0.59	0.72 (0.56, 0.93)	0.012	0.68 (0.50, 0.93)	0.015	0.78 (0.58, 1.06)	0.113	0.72 (0.53, 0.97)	0.033
rs11634397	*ZFAND6*	15	80432222	G	0.66	0.73 (0.56, 0.94)	0.016	0.71 (0.51, 0.99)	0.041	0.70 (0.51, 0.96)	0.025	0.76 (0.56, 1.03)	0.081
rs16927668	*PTPRD*	9	8369533	T	0.18	1.40 (1.03, 1.91)	0.033	1.45 (1.00, 2.10)	0.047	1.53 (1.04, 2.24)	0.032	1.25 (0.85, 1.82)	0.255
rs4299828	*ZFAND3*	6	38177667	A	0.81	0.73 (0.54, 0.99)	0.042	0.65 (0.46, 0.93)	0.018	0.80 (0.55, 1.15)	0.227	0.79 (0.55, 1.13)	0.195
rs459193	*ANKRD55*	5	55806751	G	0.72	1.36 (1.01, 1.84)	0.044	1.30 (0.90, 1.86)	0.159	1.32 (0.92, 1.88)	0.128	1.44 (1.01, 2.07)	0.045

Association of T2D SNPs with IROD2 cluster as case group. OR and 95% CI were calculated using logistic regression adjusted for sex and age.
Chr, chromosome; BP, base pair; RAF, risk allele frequency.
Number of individuals in the analyses IROD2 vs. all clusters: n = 837, IROD2 vs. IROD1: n = 357, IROD2 vs. SIDD: n = 355, IROD2 vs. MARD: n = 415.
**Significant after adjustment for the multiple testing using Bonferroni correction ($p < 0.05$).*

exhibit detrimental effects, which might result in malformation of the pancreas, thereby reducing the number of β-cells (β-cell mass) and/or affecting the secretory activity of β-cells (β-cell function) (Chen et al., 2017). The decrease in insulin output as a consequence of the reduced β-cell mass could be caused by a progressive β-cell death through established starvation-induced processes on apoptosis and disrupted autophagy (Marrif and Al-Sunousi, 2016). Thus, intrauterine programming may be related to the restricted insulin secretory capacity of pancreatic islets as a response to the demands imposed by increased insulin resistance linked to obesity in adult life. This could contribute to more severe manifestation of diabetes in the Ukrainian population. Lower frequency of insulin therapy in Ukrainian patients with new-onset diabetes in adults despite more profound insulin secretory defects compared to the ANDIS cohort could indicate that many more patients might benefit from early intensified treatments targeting preservation of insulin secretion. Naturally, discovering new ways to maintain or refine β-cell mass via enhancing β-cell survival and/or reducing apoptosis is of specific interest to investigate.

SIDD was the most severe cluster in this study, which supported consistently documented high prevalence of retinopathy across all published studies, and also showed elevated prevalence of CKD. These results are opposite to ANDIS and other cohorts (Ahlqvist et al., 2018; Zaharia et al., 2019; Zou et al., 2019) showing increased risk of CKD conferred to SIRD cluster, but somewhat similar to the recently reported findings in Asian-Indian population (Anjana et al., 2020) demonstrating increased risk of CKD in the combined insulin-deficient insulin-resistant cluster. One potential explanation for this observation could reflect and support the above-mentioned effects and malformations of organs as a consequence of abnormal intrauterine programming, which may lead to increased susceptibility to vascular complications in patients with adult diabetes later in life. In support of this, studies from the Dutch famine birth cohort have reported increased risk of microalbuminuria in adults after prenatal exposure to famine (Painter et al., 2005). Researchers suggested that fetal undernutrition may lead to lower nephron endowment, reduced number of glomeruli and consequently hyperfiltration, which may cause glomerular damage and lead to a reduction in renal function.

In line with study by Zaharia et.al of German patients with adult diabetes (Zaharia et al., 2019), the highest waist circumference was observed in the IROD1 cluster of long-term diabetes. From this perspective, the data suggest that IROD1 is associated with abdominal (visceral) type of adiposity as compared to IROD2 in which subcutaneous adipose tissue is prevailing fat deposition. Abdominal adiposity is considered to be much unhealthier than subcutaneous fat accumulation (Kahn et al., 2001) and associated with increased insulin resistance, inflammation, atherosclerosis, and vascular complications (Reijrink et al., 2019) as well as increased mortality (Christakoudi et al., 2020). Similarly, in the German Diabetes Study, the patients with newly diagnosed adult diabetes in the SIRD cluster demonstrated the highest hepatocellular lipid content and the highest prevalence of hepatic fibrosis at 5-year follow-up (Zaharia et al., 2019). This supports the idea of visceral obesity contributing to elevated risks of metabolic disorders and vascular complications in SIRD.

An interesting finding of the clustering approach in long-term adult diabetes was related to the IROD2 cluster with preserved β-function. It is tempting to speculate that people in the IROD2 cluster could have escaped starvation during the

historical famines and thereby have had their β-cell function preserved relative to the people who experienced undernutrition during times of exposure. However, we do not have information on caloric intake during famine periods; hence, this hypothesis could only be tested in a population where such information exists (e.g., Dutch famine birth cohort).

To shed light on possible underlying genetic factors of associated with beta-cell function in IROD2 cluster, we compared the frequency of established T2D variants (Morris et al., 2012). The genetic results identified consistently lower frequency of risk alleles at the *TCF7L2* locus in IROD2 cluster compared to each other cluster in this group of patients. Since 2007, polymorphisms of *TCF7L2*, encoding for transcription factor-7–like 2, are considered to be guilty of β-cells dysfunction and increased risk of diabetes in different ethnic populations (Scott et al., 2007; Dimas et al., 2014). *TCF7L2* is a member of the T-cell–specific high-mobility group box-containing family of transcription factors, that acts through Wnt-β-catenin dependent and independent pathways (Karve et al., 2020) and coordinates expression of various genes regulating cell cycle and fate determination. In pancreas *TCF7L2* was shown to play a crucial role in regulation of β-cell survival and proliferation rate. Moreover, *TCF7L2* controls the expression of genes involved in insulin granule fusion at the plasma membrane through syntaxin repression, affecting insulin secretion levels (da Silva Xavier et al., 2009). According to our results, *TCF7L2* rs7903146, which is a lead among T2D susceptibility loci (Hattersley, 2007), associated with more severe T2D phenotypes, and the lower frequencies of the risk allele appear to be associated with the protection against progression toward several vascular complications. As rs7903146 has been shown to reside in islet-selective open chromatin (Gaulton et al., 2010), this clearly motivates further metabolic studies of this group to identify epigenetic factors that play multifaced roles.

Second top signal in IROD2 group of long-term diabetes suggested reduced frequency of the risk allele in the imprinting gene *KCNQ1* (rs 163,184) that has also been shown to be expressed in the pancreatic β-cells and to act through impaired islet function on the risk of future T2D (Jonsson et al., 2009). *KCNQ1* locus was first discovered as a top signal in the two GWAS for T2D from Japan (Unoki et al., 2008; Yasuda et al., 2008) and identified as a GWAS locus for parent-of-origin effects in a large family-based study from Iceland (Kong et al., 2009). Functional and analyses of imprinting status of this genomic region suggested that metabolic effects conferred by the risk alleles at the *KCNQ1* locus target the cyclin-dependent kinase inhibitor *CDKN1C* playing a key role in regulating pancreatic β-cells proliferation and development (Kassem et al., 2001). Expression of both *KCNQ1* and *CDKN1C* demonstrated to exhibit temporal effects in fetal and adult human pancreas and islets emphasizing that the diabetes risk may be mediated in early development (Travers et al., 2013). In line with this notion, unbalanced placental expression of *CDKN1C* has been associated with intrauterine growth retardation (McMinn et al., 2006). This further supports the idea of possible contributing role of intrauterine programming of reduced pancreatic β-cell function in this population.

Several other genetic loci might deserve attention such as consistently lower frequency of risk variants in *NOTCH2* rs1493694 in IROD2 cluster. *NOTCH2*, which encodes for neurogenic locus notch homolog protein 2, is known to be responsible for regulating interactions between adjacent cells. It was demonstrated that *NOTCH2* is involved in insulin secretion and sensitivity as well as growth and development of the pancreas (Jonsson et al., 2013). The protein's extracellular domain consists of multiple EGF-like repeats while intracellular domain is involved in cell signaling affecting a variety of developmental processes controlling cell fate determination. Altered *NOTCH2* expression was found to be related to diabetic complications (Rasheed et al., 2017).

It is worth discussing that comparison of cluster characteristics in patients with longer duration of diabetes demonstrated that we for sure could not tell which of the two insulin resistant obese clusters in long-term would match the original SIRD and MOD. The original SIRD cluster in the new-onset group was characterized by high HOMA2-IR and HOMA2-B. In long-term diabetes, we could not observe the original phenomenon of simultaneously elevated both indices. To avoid confusion with the original clusters, we named the groups IROD1 in which for the given high HOMA2-IR a reduced HOMA2-B was observed, and IROD2 in which for the relatively lower HOMA2-IR higher HOMA2-B was observed. One possible explanation for the long-term HOMA-B changes in IROD1 group could be related to more frequent use of sulfonylurea drugs (61 vs. 48%, $p = 0.007$), which increase insulin secretion in short term but are considered to lead to lower insulin secretion in long term (Maedler et al., 2005; Shin et al., 2012). Additionally, these clusters differed in respect to the risk of complications with IROD2 having lowest prevalence of CKD. The IROD2 cluster was characterized by better insulin secretion and reduced frequency of the risk allele in the *TCF7L2* gene in line with genetic data in the SIRD cluster from ANDIS (Aly et al., 2020). These findings emphasize that in long-term cases changes in HOMA2-IR and HOMA2-B might occur and rather a general fit to the cluster might be considered as opposed to the given preference to one of these measures, and genetic information could be beneficial to assign people to the original SIRD cluster.

Limitations

The analyses in the present study were conducted on the patients with adult diabetes from a defined population of northern Ukraine (Chernihiv and Kyiv regions) with a history of Holodomor famine. Therefore, the findings might not be generalized to the other regions of Ukraine. The cohort comprised of adult patients with established diabetes, and the blood sampling was conducted at the study visit instead of the time of diagnosis, which limits the prognostic assessment complications risk in this cross-sectional study. Follow-up data would be required to determine if the difference between the SIRD and MOD groups in new onset diabetes and the IROD1 and IROD2 groups are due to phenotype changes over time, and in that case which of the two indices of insulin resistance and

β-cell function (HOMA2-IR or HOMA2-B) in patients with long-term would have a better fit to the original insulin resistant SIRD and MOD clusters coined for ANDIS. An alternative explanation is poor performance of the clustering algorithm due to the relatively small cohort size. It has been seen also in ANDIS that the MOD and SIRD clusters are the least stable but that a larger sample size can improve stability and reproducibility of the clusters. There were more women than men in this cohort, which could potentially give gender-specific effects. However, all analyses were separately performed in men and women, and gave similar results. The K_{mean} clustering algorithm used in the current approach presumes that all the clustering variables have the same weight. Nevertheless, giving variables different weights or prioritizing importance or using another clustering algorithm might improve the approach. All patients were recruited at the primary health care centers or outpatient clinic, which minimized the bias related to recruitment of severe diabetes patients admitted to the hospital or being on ward. Although the sample size is limited for genetic analyses, the power calculations showed that significant effects ($p < 0.00014$) would be reached for variants with the Genotype Relative Risk (disease probability for individuals with 1 risk allele divided by disease probability for individuals with 0 risk alleles) above 1,8 (Skol et al., 2006; Goncalo Abecasis et al., 2017).

CONCLUSION

In conclusion, pathophysiology-based clustering is undoubtedly beneficial for diagnosing different subtypes of adult diabetes related to risk of micro- and macrovascular complications. Assessment of GADA is prerequisite to correctly re-classify SAID patients with adult diabetes, which in the clinical practice can be misclassified as T2D. It can be a clear advantage for the patients belonging to the SIDD cluster to start treatment with insulin or other therapeutic modalities at an earlier stage in order to preserve and maintain β-cell function. The persons with long-term diabetes assigned to IROD2 cluster exhibited preserved insulin secretion and lower risk for microvascular complications. Thus, this cluster represents an interesting subgroup of patients for further investigations of protective mechanisms. The current diabetes cluster approach could be further refined and optimized by including other new biomarkers derived from ongoing omics studies.

DATA AVAILABILITY STATEMENT

The datasets presented in this article are not readily available because of GDPR and ethical restrictions. Requests to access the datasets should be directed to valeriya.lyssenko@uib.no.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Ethics Committee at "Regional Hospital," Chernihiv, Ukraine (approval number Dnr17/2011-09-14) and Regional Committee for Medical and Health Research (Ethics, South-east, Panel A, approval number 2019/28968). The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

OF did the statistical analyses, data interpretation, and wrote the manuscript. TO, EA, and OA assisted in statistical analyses. LC, NK, TS, and TB performed recruitment of patients, data acquisition, and clinical interpretation. OS, TO, EA, LG, and PN contributed to editing of the manuscript and data interpretation. VL conceived and designed the study, planned the analyses, supervised all parts of the study, interpreted the data, and wrote the manuscript. All authors contributed to the interpretation of the data, and approved the final version of the manuscript.

FUNDING

This work was supported by the Swedish Research Council (Dnr2015-03574, Dnr349-2006-237, 2017-02688, and 2020-02191), Strategic Research Area Exodiab (Dnr2009-1039), the Novonordisk Foundation (NNF12OC1016467 and NNF18OC0034408), Swedish Foundation for Strategic Research (DnrIRC15–0067), the Swedish Heart-Lung Foundation, the Steno Diabetes Center Copenhagen, Bergen Research Foundation (BFS811294), and the University of Bergen.

ACKNOWLEDGMENTS

We thank the patients for their diligent and active participation and laboratory assistant Valentina Burkhanova from the Chernihiv Regional Blood Transfusion Station for the excellent technical assistance.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.637945/full#supplementary-material

REFERENCES

R Core Team. (2021). *R: A Language and Environment for Statistical Computing (Version 3.6.2)*. R Foundation for Statistical Computing. Available online at: https://www.R-project.org/

Ahlqvist, E., Storm, P., Käräjämäki, A., Martinell, M., Dorkhan, M., Carlsson, A., et al. (2018). Novel subgroups of adult-onset diabetes and their association with outcomes: a data-driven cluster analysis of six variables. *Lancet Diabet. Endocrinol.* 6, 361–369. doi: 10.1016/s2213-8587(18)30051-2

Anjana, R. M., Nair, A. T. N., Jebarani, S., Siddiqui, M. K., Pradeepa, R., et al. (2020). Novel subgroups of type 2 diabetes and their association with microvascular outcomes in an Asian Indian population: a data-driven cluster analysis: the INSPIRED study. *BMJ Open Diabet. Res. Care* 8:e001506. doi: 10.1136/bmjdrc-2020-001506

Chen, C., Cohrs, C. M., Stertmann, J., Bozsak, R., and Speier, S. (2017). Human beta cell mass and function in diabetes: Recent advances in knowledge and technologies to understand disease pathogenesis. *Mol. Metab.* 6, 943–957. doi: 10.1016/j.molmet.2017.06.019

Christakoudi, S., Muller, D. C., Freisling, H., Weiderpass, E., Overvad, K., et al. (2020). A Body Shape Index (ABSI) achieves better mortality risk stratification than alternative indices of abdominal obesity: results from a large European cohort. *Sci. Rep.* 10:14541. doi: 14510.11038/s41598-14020-71302-14545

da Silva Xavier, G., Loder, M. K., McDonald, A., Tarasov, A. I., Carzaniga, R., et al. (2009). TCF7L2 regulates late events in insulin secretion from pancreatic islet beta-cells. *Diabetes* 58, 894–905. doi: 10.2337/db08-1187

Dennis, J. M., Shields, B. M., Henley, W. E., Jones, A. G., and Hattersley, A. T. (2019). Disease progression and treatment response in data-driven subgroups of type 2 diabetes compared with models based on simple clinical features: an analysis using clinical trial data. *Lancet Diabet. Endocrinol.* 7, 442–451. doi: 10.1016/s2213-8587(19)30087-7

IDF (2019). *IDF DIABETES ATLAS*, 9th Edn. Brussels: International Diabetes Federation. Available online at: https://www.diabetesatlas.org/en/

Dimas, A. S., Lagou, V., Barker, A., Knowles, J. W., Mägi, R., Hivert, M. F., et al. (2014). Impact of type 2 diabetes susceptibility variants on quantitative glycemic traits reveals mechanistic heterogeneity. *Diabetes* 63, 2158–2171.

Aly, D. M., Dwivedi, O. P., Prasad, R. B., Käräjämäki, A., Hjort, R., Åkerlund, M., et al. (2020). Aetiological differences between novel subtypes of diabetes derived from genetic associations. [Preprint].

Goncalo Abecasis, J. L. J., Clark, C., Li, K. W., and Caron, S. (2017). *GAS Power Calculator*. Michigan: University of Michigan.

Fitipaldi, H., McCarthy, M. I., Florez, J. C., and Franks, P. W. (2018). A Global Overview of Precision Medicine in Type 2 Diabetes. *Diabetes* 67, 1911–1922. doi: 10.2337/dbi17-0045

Gaulton, K. J., Nammo, T., Pasquali, L., Simon, J. M., Giresi, P. G., and Fogarty, M. P et al. (2010). A map of open chromatin in human pancreatic islets. *Nat. Genet.* 42, 255–259. doi: 10.1038/ng.530

Hattersley, A. T. (2007). Prime suspect: the TCF7L2 gene and type 2 diabetes risk. *J. Clin. Invest.* 117, 2077–2079. doi: 10.1172/jci33077

Hennig, C. (2020). *R-package 'fpc'*. Vienna: R Core Team. Available Online at: https://cran.r-project.org/web/packages/fpc/fpc.pdf.

HOMACalculator (2020). *HOMA Calculator*. Oxford: University of Oxford.

Jonsson, A., Isomaa, B., Tuomi, T., Taneera, J., Salehi, A., Nilsson, P., et al. (2009). A variant in the KCNQ1 gene predicts future type 2 diabetes and mediates impaired insulin secretion. *Diabetes* 58, 2409–2413. doi: 10.2337/db09-0246

Jonsson, A., Ladenvall, C., Ahluwalia, T. S., Kravic, J., Krus, U., Taneera, J., et al. (2013). Effects of common genetic variants associated with type 2 diabetes and glycemic traits on alpha- and beta-cell function and insulin action in humans. *Diabetes* 62, 2978–2983. doi: 10.2337/db12-1627

Kahn, S. E., Schwartz, R. S., Fujimoto, W. Y., Knopp, R. H., Brunzell, J. D., and Porte, D. Jr. (2001). Obesity, body fat distribution, insulin sensitivity and Islet beta-cell function as explanations for metabolic diversity. *J. Nutr.* 131, 354S–160S. doi: 110.1093/jn/1091

Karve, K., Netherton, S., Deng, L., Bonni, A., and Bonni, S. (2020). Regulation of epithelial-mesenchymal transition and organoid morphogenesis by a novel TGFbeta-TCF7L2 isoform-specific signaling pathway. *Cell Death Dis.* 11:704.

Kassem, S. A., Ariel, I., Thornton, P. S., Hussain, K., Smith, V., Lindley, K. J., et al. (2001). p57(KIP2) expression in normal islet cells and in hyperinsulinism of infancy. *Diabetes* 50, 2763–2769. doi: 10.2337/diabetes.50.12.2763

Khalangot, M., and Tronko, M. (2007). Primary care diabetes in Ukraine. *Prim. Care Diabet.* 1, 203–205. doi: 10.1016/j.pcd.2007.10.041

Kong, A., Steinthorsdottir, V., Masson, G., Thorleifsson, G., Sulem, P., Besenbacher, S., et al. (2009). Parental origin of sequence variants associated with complex diseases. *Nature* 462, 868–874. doi: 10.1038/nature08625

Levy, J. C., Matthews, D. R., and Hermans, M. P. (1998). Correct homeostasis model assessment (HOMA) evaluation uses the computer program. *Diabet. Care* 21:2191. doi: 10.2337/diacare.21.12.2191

Lumey, L. H., Khalangot, M. D., and Vaiserman, A. M. (2015). Association between type 2 diabetes and prenatal exposure to the Ukraine famine of 1932-33: a retrospective cohort study. *Lancet Diabet. Endocrinol.* 3, 787–794. doi: 10.1016/s2213-8587(15)00279-x

Maedler, K., Carr, R. D., Bosco, D., Zuellig, R. A., Berney, T., and Donath, M. Y. (2005). Sulfonylurea induced beta-cell apoptosis in cultured human islets. *J. Clin. Endocrinol. Metab.* 90, 501–506. doi: 10.1210/jc.2004-0699

Mankovsky, B. (2020). Diabetes Care at the Times of Transition and COVID-19 Pandemics (Ukrainian Experience). *J. Diabet. Sci. Technol.* 14, 754–755. doi: 10.1177/1932296820930031

Marees, A. T., de Kluiver, H., Stringer, S., Vorspan, F., Curis, E., Marie-Claire, C., et al. (2018). A tutorial on conducting genome-wide association studies: Quality control and statistical analysis. *Int. J. Methods Psychiatr. Res.* 27:e1608. doi: 10.1002/mpr.1608

Marrif, H. I., and Al-Sunousi, S. I. (2016). Pancreatic β Cell Mass Death. *Front. Pharmacol.* 7:83. doi: 10.3389/fphar.2016.00083

McCarthy, M. I. (2017). Painting a new picture of personalised medicine for diabetes. *Diabetologia* 60, 793–799. doi: 10.1007/s00125-017-4210-x

McMinn, J., Wei, M., Schupf, N., Cusmai, J., Johnson, E. B., Smith, A. C., et al. (2006). Unbalanced placental expression of imprinted genes in human intrauterine growth restriction. *Placenta* 27, 540–549. doi: 10.1016/j.placenta.2005.07.004

Morris, A. P., Voight, B. F., Teslovich, T. M., Ferreira, T., Segrè, A. V., Steinthorsdottir, V., et al. (2012). Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nat. Genet.* 44:981. doi: 10.1038/ng.2383

Mula-Abed, W. A., Al Rasadi, K., and Al-Riyami, D. (2012). Estimated Glomerular Filtration Rate (eGFR): A Serum Creatinine-Based Test for the Detection of Chronic Kidney Disease and its Impact on Clinical Practice. *Oman Med. J.* 27, 108–113. doi: 10.5001/omj.2012.23

Painter, R. C., Roseboom, T. J., van Montfrans, G. A., Bossuyt, P. M., Krediet, R. T., Osmond, C., et al. (2005). Microalbuminuria in adults after prenatal exposure to the Dutch famine. *J. Am. Soc. Nephrol.* 16, 189–194. doi: 10.1681/asn.2004060474

Rasheed, M. A., Kantoush, N., Abd El-Ghaffar, N., Farouk, H., Kamel, S., Ibrahim, A. A., et al. (2017). Expression of JAZF1, ABCC8, KCNJ11and Notch2 genes and vitamin D receptor polymorphisms in type 2 diabetes, and their association with microvascular complications. *Ther. Adv. Endocrinol. Metab.* 8, 97–108. doi: 10.1177/2042018817708910

Reijrink, M., dBS, Spoor, D. S., Lefrandt, J. D., Lambers Heerspink, H. J., Boellaard, R., et al. (2019). Visceral adipose tissue volume is associated with premature atherosclerosis in early type 2 diabetes mellitus independent of traditional risk factors. *Atherosclerosis* 290, 287–293. doi: 210.1016/j.atherosclerosis.2019.1009.1016

Scott, L. J., Mohlke, K. L., Bonnycastle, L. L., Willer, C. J., Li, Y., Duren, W. L., et al. (2007). A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. *Science* 316, 1341–1345.

Shin, M. S., Yu, J. H., Jung, C. H., Hwang, J. Y., Lee, W. J., Kim, M. S., et al. (2012). The duration of sulfonylurea treatment is associated with beta-cell dysfunction in patients with type 2 diabetes mellitus. *Diabet. Technol. Ther.* 14, 1033–1042. doi: 10.1089/dia.2012.0144

Skol, A. D., Scott, L. J., Abecasis, G. R., and Boehnke, M. (2006). Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. *Nat. Genet.* 38, 209–213. doi: 10.1038/ng1706

Travers, M. E., Mackay, D. J., Dekker Nitert, M., Morris, A. P., Lindgren, C. M., Berry, A., et al. (2013). Insights into the molecular mechanism for type 2 diabetes susceptibility at the KCNQ1 locus from temporal changes in imprinting status in human islets. *Diabetes* 62, 987–992. doi: 10.2337/db12-0819

Tuomi, T., Santoro, N., Caprio, S., Cai, M., Weng, J., and Groop, L. (2014). The many faces of diabetes: a disease with increasing heterogeneity. *Lancet* 383, 1084–1094. doi: 10.1016/s0140-6736(13)62219-9

Unoki, H., Takahashi, A., Kawaguchi, T., Hara, K., Horikoshi, M., Andersen, G., et al. (2008). SNPs in KCNQ1 are associated with susceptibility to type 2

diabetes in East Asian and European populations. *Nat. Genet.* 40, 1098–1102. doi: 10.1038/ng.208

Yasuda, K., Miyake, K., Horikawa, Y., Hara, K., Osawa, H., Furuta, H., et al. (2008). Variants in KCNQ1 are associated with susceptibility to type 2 diabetes mellitus. *Nat. Genet.* 40, 1092–1097.

Zaharia, O. P., Strassburger, K., Strom, A., Bönhof, G. J., Karusheva, Y., Antoniou, S., et al. (2019). Risk of diabetes-associated diseases in subgroups of patients with recent-onset diabetes: a 5-year follow-up study. *Lancet Diabet. Endocrinol.* 7, 684–694.

Zou, X., Zhou, X., Zhu, Z., and Ji, L. (2019). Novel subgroups of patients with adult-onset diabetes in Chinese and US populations. *Lancet Diabet. Endocrinol.* 7, 9–11. doi: 10.1016/s2213-8587(18)30316-4

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 in Medicine

PERSPECTIVE
published: 26 July 2021
doi: 10.3389/fmed.2021.663014

Legal Challenges in Precision Medicine: What Duties Arising From Genetic and Genomic Testing Does a Physician Owe to Patients?

Scott P. McGrath[1]*, *Arthur E. Peabody Jr.*[2], *Derek Walton*[3] and *Nephi Walton*[4]

[1] CITRIS and the Banatao Institute, University of California, Berkeley, Berkeley, CA, United States, [2] Hooper, Lundy & Bookman, Professional Corporation, Washington, DC, United States, [3] Walton Legal Professional Limited Liability Company, Salt Lake City, UT, United States, [4] Intermountain Healthcare, Salt Lake City, UT, United States

Precision medicine is increasingly incorporated into clinical practice *via* three primary data conduits: environmental, lifestyle, and genetic data. In this manuscript we take a closer look at the genetic tier of precision medicine. The volume and variety of data provides a more robust picture of health for individual patients and patient populations. However, this increased data may also have an adverse effect by muddling our understanding without the proper pedagogical tools. Patient genomic data can be challenging to work with. Physicians may encounter genetic results which are not fully understood. Genetic tests may also lead to the quandary of linking patients with diseases or disorders where there are no known treatments. Thus, physicians face a unique challenge of establishing the proper scope of their duty to patients when dealing with genomic data. Some of those scope of practice boundaries have been established as a result of litigation, while others remain an open question. In this paper, we map out some of the legal challenges facing the genomic component of precision medicine, both established and some questions requiring additional guidance. If physicians begin to perceive genomic data as falling short in overall benefit to their patients, it may detrimentally impact precision medicine as a whole. Helping to develop guidance for physicians working with patient genomic data can help avoid this fate of faltering confidence.

Keywords: precision medicine, scope of practice, genetic testing, genomics, physicians

OPEN ACCESS

Edited by:
Muntaser Eltayeb Ibrahim, University of Khartoum, Sudan

Reviewed by:
Nurlan Dauletbayev, McGill University, Canada
David Kisor, Manchester University, United States

***Correspondence:**
Scott P. McGrath
smcgrath@berkeley.edu

Specialty section:
This article was submitted to Precision Medicine, a section of the journal Frontiers in Medicine

Received: *02 February 2021*
Accepted: *28 June 2021*
Published: *26 July 2021*

Citation:
McGrath SP, Peabody AE Jr, Walton D and Walton N (2021) Legal Challenges in Precision Medicine: What Duties Arising From Genetic and Genomic Testing Does a Physician Owe to Patients? Front. Med. 8:663014. doi: 10.3389/fmed.2021.663014

INTRODUCTION

As precision medicine proliferates in the provision of medical care (1, 2), a critical component is the use of an individual's genomic information in diagnosis and individualized treatments (3–5). Yet physicians face unique challenges. Physician knowledge of genetics largely remains incomplete (6) and there is a gap in their ability to interpret genetic results in comparison to genetic specialists (7). Genomics is rapidly changing, both at the focused level of genetic tests in the wider scope of genomic tests, making it difficult to keep abreast of the latest developments (8, 9). Studies have shown that labs can yield both inconsistent results and more information than the typical physician can reasonably digest (10–12). Clinical applications can be confusing and not readily available in many cases. Polygenic risk scores, e.g., may be adding complexity to the corpus of knowledge a health provider must be aware of, and also may risk exacerbating health disparities by catering

certain treatments to different racial/ethnic groups (13, 14). There are few physicians that possess the education, skills, and experience to take full advantage of genetic testing, including what test is optimal to select (15–18). Numerous ethical and privacy concerns lurk over the shoulder of every physician choosing to use genetic information (19–23). This article seeks to analyze the current state of physician liability of using genetic information. In the field of genetic testing, there are more questions than answers. In many areas, there is a need to analyze the specific facts at issue and develop carefully crafted solutions to avoid the ever-looming specter of legal liability. There is currently a lack of case law in the United States, Europe, or in the United Kingdom that directly addresses these issues (24–26), therefore, we focus on legal principles established in American jurisprudence and those few reported cases in the courts of the United States to provide a framework to explore this topic.

LEGAL STANDARDS

The legal liability of physicians is generally governed by the common law of negligence. In general, a physician is negligent or liable for challenged conduct when they fail to exercise reasonable judgment and departs from generally acceptable standards of practice. In other words, a physician is negligent when he or she fails to follow generally accepted practice in any particular medical domain. State law is especially important because state courts set the parameters of what kind of claims can be sought and on what basis. Some genetic legal cases have required explicit statutes (27), while others have relied on common law notions of negligence or personal injury law absent any statute (28). Yet, in precision medicine, the question may be far more complicated. When the interpretation of genetic information is at issue, there may be no generally accepted practice or standard. Physicians are constantly challenged by what to do with genetic knowledge that is available but may not be fully understood or for diseases for which there is no known treatment.

At what specific point does clinical genetic knowledge become a standard of practice? This issue is being tested in both federal and state courts. In *Williams v. Quest/Athena* (29), the plaintiff sued the laboratory that conducted genetic testing of the plaintiff's son, who subsequently died. The plaintiff has argued that two genetic studies conducted prior to the lab's allegedly erroneous reporting that the variant was of "unknown significance." This misdiagnosed genetic mutation led directly to the child's inappropriate treatment and death. Are two studies sufficient to establish a standard of practice? Do two studies establish that the variant in question is "significant" and must be properly identified by the lab and addressed by the physician? The federal court ultimately found for the Defendants holding, among other things, that there was insufficient evidence that the plaintiff's son's variant was pathogenic and, as a result, no reasonable jury could find negligence (30). This case points out how unclear the standard for legal liability can be under the current state of the law.

Some have suggested that evidence-based medicine should determine when there is sufficient evidence to find a variant to be pathogenic and, as a result, provide direction to physicians navigating unfamiliar genetic or genomic results. However, consensus acquired through evidence-based medicine relating to genomic data is still maturing. As a result, this approach does not always provide medical practitioners guidance on what steps to take. Nonetheless, efforts are being made to determine if the principles of evidence-based medicine can be effectively applied to genomic data (31) and how exactly genomics should be integrated into healthcare (32–34).

PHYSICIAN CHALLENGES

Yet the challenges do not end with the question of a legal standard. How does a physician keep up or remain current with the evolving body of genetic knowledge? There are five categorizations of genetic variant: pathogenetic, likely pathogenic, variant of uncertain significance (VUS), likely benign, and benign. The usage of "likely" is defined as "a >90% certainty that a specific variant is pathogenic or benign" (35). A 25-gene cancer susceptibility panel will report at least one VUS about 33% of the time (36). These VUS can obfuscate a physician's duty and generate several questions.

- Are physicians responsible for tracking reported VUS, in case they become classified as "significant" for both current and previous patients?
- Does the nature of the duty change when the doctor is the patient's "primary physician?"
- Do general practitioners or genetic specialists owe the same or a different duty based on their expertise in genetics?

A further confounding factor is the gap, sometimes indefinite, between the capacity to diagnose using genetic tests and the capacity to treat due to the lack of effective therapies for any number of genetically linked conditions (37, 38).

But the physicians' quandary does not stop here. As the ability to interpret genetic variations grows, the physician is faced with the fact that known variants that have the potential for disease may never lead to the disease, including patients with a genetic disease for which there is no known treatment (39). What duty does a physician owe to these patients even when it is understood that known variants may never lead to the disease in particular, individual patients?

The challenge of genetic knowledge is knowing what to do with most genetic information. For example, women who test positive for BRCA-1 or BRCA-2 pathogenic variants, may undergo prophylactic mastectomy because of their fear of developing an aggressive cancer. However, some of these patients may have been either misdiagnosed or do not fully appreciate the uncertain nature of the genetic mutations (40, 41). Some have undergone unnecessary surgery; others may have escaped breast cancer. How do we balance the benefits and potential risks of

Abbreviations: VUS, Variant of Uncertain Significance; BRCA-1, Breast Cancer Gene 1; BRCA-2, Breast Cancer Gene 2; HIPAA, Health Insurance Portability and Accountability Act; AMA, American Medical Association; ASHG, American Society of Human Genetics; NIH, National Institutes of Health.

harm of genetic knowledge? This is a question society, in general, and physicians, in particular, have only begun to address.

These alleged unnecessary mastectomies have led to litigation. In *Moore v. Curry County Health* (42), the plaintiff alleged that the physician misread her genetic tests and even if the variant was interpreted correctly, the variant did not lead to breast cancer. The facts as alleged, if true, may form a plausible claim for damages as the physician may not have followed the standard of practice in interpreting the plaintiff's genetic test results, but like all cases in this field to date, there is always a wrinkle. The plaintiff had a family history of breast cancer. Would the family history of breast cancer have justified the mastectomy, notwithstanding whatever the genetic testing may have revealed? In *Moore v. Curry*, the plaintiff settled for $600 k+ in damages without disclosing the terms of the settlement (43).

ADVANCED GENOMIC TECHNIQUES AND PHYSICIAN DUTY

Whole exome sequencing (WES) is becoming popular for diagnosing patients with complex disease. However, tests often return results outside of the condition for which it was originally ordered (44). Does the physician have an obligation to review all of the findings even though they were not ordered and may lack relevancy to their diagnosis? Pharmacogenomic testing results can guide prescribing. However, physicians are unlikely to look at a WES test report to assess for pharmacogenomic findings prior to prescribing a medication (45). Is the physician liable for not reviewing WES data in a patient's chart, if the failure to do so leads to harm of the patient? This issue is exacerbated by the volume of information that may be provided and the numerous findings of "variant of unknown significance" that may be contained in the report. At least one court has found liability for failing to diagnose medical issues that may be disclosed during the course of tests and other measures a physician undertakes to resolve a patient's other issues. The courts have labeled the medical issues discovered in the course of other tests as "incidental findings." For example, in *Lo v. Burke* (46), a court found a radiologist liable for not detecting a pancreatic tumor when searching for a tumor in the liver.

In addition to the aforementioned challenges, WES may not adequately cover all genetic regions (promotors or intronic areas, e.g.,) that are important to a certain pharmacogenomic result or secondary finding (47). What happens if the physician does review the pharmacogenomic data from WES but is unaware of the limitations of the regions covered and, as a result, fails to order the more comprehensive test that would have more appropriately guided the prescribing? Legal precedent has not yet addressed this and other issues relating to the extent of physician liability that may arise from the use of WES in diagnosing and planning a patient's course of treatment.

In confronting test results the physician must decide both what to disclose and who else to advise. For example, if a patient's test reveals a significant variant known to be linked to a disease that is potentially hereditary in nature, is the physician under any obligation to advise the patient's children or any other family member, e.g., the patient's siblings, of the variant, and their potential exposure to the disease? Here, the courts have reached different results.

In *Pate v. Threlkel* (48), the court ultimately held that the physician had the duty to advise the patient to warn her children of their increased risk of disease due to the genetic makeup of their mother – but no duty of the physician to advise the children directly. This holding was based on Florida's "HIPAA statute" barring the disclosure of protected health information absent consent of the patient. HIPAA protects most private health information except for certain explicit reasons, including treatment. This exception, however, applies only to the patient. HIPAA does authorize a patient to provide "informal" consent to a physician to disclose genetic information to potentially affected family members, but state law, as in Threlkel, may impose more formal requirements or bar the disclosures.

In *Safer v. Estate of Pack* (49), the court held that a physician has a duty to warn all members of a patient's immediate family of the patient's potentially genetically transferrable disease – in spite of HIPAA. The American Medical Association provides some guidance to physicians who need to counsel patients about sharing genetic test results to family members (50). The American Society of Human Genetics' statement suggests that a physician may be justified in warning family members directly if the patient declines to cooperate in circumstances posing serious risks to family members (51). The alleged duty to warn may have serious unintended consequences. For example, in diagnosing potential birth defects in an unborn child, both parents may be subjected to genetic testing. The result may call into question the paternity of the child. The question of scope of physician duty is again brought into focus when dealing with families.

- What bearing does infrequent application of genetic knowledge impact a physician's duty to warn a patient's children and family members?
- What guidance should be provided to physicians in order to make such decisions when faced with a multitude of "variants of unknown significance" results?
- When does a physician have a duty to recommend genetic testing in face of a patient's history of disease generally associated with a genetic mutation?

While questions surrounding family consultations are numerous, the duty to recommend genetic testing is not a novel issue for courts to address. Case precedent exits extending as far back as the early 1980's holding that the failure to recommend genetic testing in appropriate circumstances, e.g., a family history of a genetically linked disease, constitutes a departure from the standard of care (52). State courts have found a similar duty. In *Downs v. Trias* (53), the court found a physician liable for negligence when he failed to recommend genetic testing to a woman who died of ovarian cancer where her family had a history of cancer.

CONCLUSIONS

In summary, physicians face many challenges in the interpretation of genetic testing results given the current state of knowledge. The issue is only heightened because the field

is changing rapidly with ever-increasing information. Physicians carry enormous responsibility when they enter the arena of genetic testing. The law is only beginning to articulate the duties each must fulfill. Even though, there are legal challenges that are presented with these new capabilities. As legal challenges associated with precision medicine are assessed, it is critical to not overlook the risk associated with avoiding new technology which may incur even larger liability. A recent review of genomic malpractice cases in the United States through the end of 2016 showed that most cases were not based on misinterpretation of genetic variants but were related to failure to perform genetic testing or failure to act on the results of genetic testing. In fact, it suggested that 57% of the genomic medical malpractice cases could have been avoided if genetic testing had been performed when the patient first presented with the condition (54). There is an expanding list of more 7,000 rare diseases, a majority of which are believed to have a genetic cause (55). Many of these diseases are actionable in that they have specific treatments catered to the genetic cause; for others appropriate management is just beginning to be developed or understood. Failure to test for and treat these diseases appropriately can lead to poor outcomes for patients and significant liability for providers. It is important that new technologies are embraced for the benefits they can offer and used cautiously, for fear of these technologies must not become a barrier to providing the best possible patient care.

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/s.

AUTHOR CONTRIBUTIONS

NW and AP performed the primary research on the cases and constructed the initial draft manuscript. DW assisted with the legal reviews and contributed to the writing about legal cases. SM contributed to both the writing and research and led the revision and preparation of the final manuscript. All authors contributed to the article and approved the submitted version.

ACKNOWLEDGMENTS

The authors would like to thank Melanie McGrath (PhD, LAT, ATC) for aid in the preparation of this manuscript.

REFERENCES

1. Fröhlich H, Balling R, Beerenwinkel N, Kohlbacher O, Kumar S, Lengauer T, et al. From hype to reality: data science enabling personalized medicine. *BMC Med*. (2018) 16:150. doi: 10.1186/s12916-018-1122-7
2. Kattan MW, Hess KR, Amin MB, Lu Y, Moons KGM, Gershenwald JE, et al. American Joint Committee on Cancer acceptance criteria for inclusion of risk models for individualized prognosis in the practice of precision medicine. *CA Cancer J Clin*. (2016) 66:370–4. doi: 10.3322/caac.21339
3. Abul-Husn NS, Owusu Obeng A, Sanderson SC, Gottesman O, Scott SA. Implementation and utilization of genetic testing in personalized medicine. *Pharmgenomics Pers Med*. (2014) 7:227–40. doi: 10.2147/PGPM.S48887
4. Bray MS, Loos RJF, McCaffery JM, Ling C, Franks PW, Weinstock GM, et al. NIH working group report-using genomic information to guide weight management: from universal to precision treatment. *Obesity*. (2016) 24:14–22. doi: 10.1002/oby.21381
5. Hyman DM, Taylor BS, Baselga J. Implementing genome-driven oncology. *Cell*. (2017) 168:584–99. doi: 10.1016/j.cell.2016.12.015
6. Owusu Obeng, Fei K, Levy K, Elsey A, Pollin T, Ramirez A, et al. Physician-reported benefits and barriers to clinical implementation of genomic medicine: a multi-site IGNITE-network survey. *J Pers Med*. (2018) 8:24. doi: 10.3390/jpm8030024
7. McGrath SP, Walton N, Williams MS, Kim KK, Bastola K. Are providers prepared for genomic medicine: interpretation of Direct-to-Consumer genetic testing (DTC-GT) results and genetic self-efficacy by medical professionals. *BMC Health Serv Res*. (2019) 19:844. doi: 10.1186/s12913-019-4679-8
8. Plunkett-Rondeau J, Hyland K, Dasgupta S. Training future physicians in the era of genomic medicine: trends in undergraduate medical genetics education. *Genet Med*. (2015) 17:927–34. doi: 10.1038/gim.2014.208
9. Stoll K, Kubendran S, Cohen SA. The past, present and future of service delivery in genetic counseling: keeping up in the era of precision medicine. *Am J Med Genet C Semin Med Genet*. (2018) 178:24–37. doi: 10.1002/ajmg.c.31602
10. Manrai AK, Funke BH, Rehm HL, Olesen MS, Maron BA, Szolovits P, et al. Genetic misdiagnoses and the potential for health disparities. *N Engl J Med*. (2016) 375:655–65. doi: 10.1056/NEJMsa1507092
11. Rubanovich CK, Cheung C, Mandel J, Bloss CS. Physician preparedness for big genomic data: a review of genomic medicine education initiatives in the United States. *Hum Mol Genet*. (2018) 27:R250–8. doi: 10.1093/hmg/ddy170
12. Tandy-Connor S, Guiltinan J, Krempely K, LaDuca H, Reineke P, Gutierrez S, et al. False-positive results released by direct-to-consumer genetic tests highlight the importance of clinical confirmation testing for appropriate patient care. *Genet Med*. (2018) 20:1515–21. doi: 10.1038/gim.2018.38
13. Martin AR, Kanai M, Kamatani Y, Okada Y, Neale BM, Daly MJ. Clinical use of current polygenic risk scores may exacerbate health disparities. *Nat Genet*. (2019) 51:584–91. doi: 10.1038/s41588-019-0379-x
14. Torkamani, Wineinger NE, Topol EJ. The personal and clinical utility of polygenic risk scores. *Nat Rev Genet*. (2018) 19:581–90. doi: 10.1038/s41576-018-0018-x
15. Arora NS, Davis JK, Kirby C, McGuire AL, Green RC, Blumenthal-Barby JS, et al. Communication challenges for non-geneticist physicians relaying clinical genomic results. *Per Med*. (2017) 14:423–31. doi: 10.2217/pme-2017-0008
16. Christensen KD, Vassy JL, Jamal L, Lehmann LS, Slashinski MJ, Perry DL, et al. Are physicians prepared for whole genome sequencing? A qualitative analysis. *Clin Genet*. (2016) 89:228–34. doi: 10.1111/cge.12626
17. Mikat-Stevens NA, Larson IA, Tarini BA. Primary-care providers' perceived barriers to integration of genetics services: a systematic review of the literature. *Genet Med*. (2015) 17:169–76. doi: 10.1038/gim.2014.101
18. Reed EK, Johansen Taber KA, Ingram Nissen T, Schott S, Dowling LO, O'Leary JC, et al. What works in genomics education: outcomes of an evidenced-based instructional model for community-based physicians. *Genet Med*. (2016) 18:737–45. doi: 10.1038/gim.2015.144
19. Borry P, Shabani M, Howard HC. Is there a right time to know?: the right not to know and genetic testing in children. *J Law Med Ethics*. (2014) 42:19–27. doi: 10.1111/jlme.12115
20. Fulda KG, Lykens K. Ethical issues in predictive genetic testing: a public health perspective. *J Med Ethics*. (2006) 32:143–7. doi: 10.1136/jme.2004.010272
21. Green MJ, Botkin JR. "Genetic exceptionalism" in medicine: clarifying the differences between genetic and non-genetic tests. *Ann Intern Med*. (2003) 138:571–5. doi: 10.7326/0003-4819-138-7-200304010-00013
22. Klitzman R, Chung W, Marder K, Shanmugham A, Chin LJ, Stark M, et al. Attitudes and practices among internists concerning genetic testing. *J Genet Counsel*. (2013) 22:90–100. doi: 10.1007/s10897-012-9504-z

23. Rotz SJ, Kodish E. Ethical conundrums in pediatric genomics. *Hematology*. (2018) 2018:301–6. doi: 10.1182/asheducation-2018.1.301
24. Clayton EW, Haga S, Kuszler P, Bane E, Shutske K, Burke W. Managing incidental genomic findings: legal obligations of clinicians. *Genet Med*. (2013) 15:624–9. doi: 10.1038/gim.2013.7
25. Mitchell C, Ploem C, Chico V, Ormondroyd E, Hall A, Wallace S, et al. Exploring the potential duty of care in clinical genomics under UK law. *Med Law Int*. (2017) 17:158–82. doi: 10.1177/0968533217721966
26. Thorogood, Zawati M, Knoppers BM. Oversight, governance, and policy for making decisions about return of individual genomic findings. In: *Secondary Findings in Genomic Research*. Elsevier (2020). p. 29–41. doi: 10.1016/B978-0-12-816549-2.00002-3
27. Atlanta Ob. & Gyn. v. Abelson. *Supreme Court of Georgia*. Atlanta, GA (1990). Available online at: https://law.justia.com/cases/georgia/supreme-court/1990/s90g1000-1.html (accessed June 21, 2021).
28. Plowman v. FT. *Madison Community Hosp*. Iowa: Supreme Court (2017). Available online at: https://law.justia.com/cases/iowa/supreme-court/2017/150974.html (accessed June 21, 2021).
29. Williams v. Quest Diagnostics, Inc. *United States District Cort for South Carolina District*. (2020). Available online at: https://crain-platform-genomeweb-prod.s3.amazonaws.com/s3fs-public/files_copied/williams_v_quest_summary_judgment_11052020.pdf (accessed June 21, 2021).
30. Ray T. *Quest Diagnostics Win in Wrongful Death Case Reveals Ongoing Challenges for Variant Classification. GenomeWeb*. (2020). Available online at: https://www.genomeweb.com/molecular-diagnostics/quest-diagnostics-win-wrongful-death-case-reveals-ongoingchallenges-variant (accessed June 21, 2021).
31. Ioannidis JPA, Khoury MJ. Evidence-based medicine and big genomic data. *Hum Mol Genet*. (2018) 27:R2–7. doi: 10.1093/hmg/ddy065
32. Manolio TA, Abramowicz M, Al-Mulla F, Anderson W, Balling R, Berger AC, et al. Global implementation of genomic medicine: we are not alone. *Sci Transl Med*. (2015) 7:290ps13. doi: 10.1126/scitranslmed.aab0194
33. Stark Z, Boughtwood T, Phillips P, Christodoulou J, Hansen DP, Braithwaite J, et al. Australian genomics: a federated model for integrating genomics into healthcare. *Am J Hum Genet*. (2019) 105:7–14. doi: 10.1016/j.ajhg.2019.06.003
34. Stark Z, Dolman L, Manolio TA, Ozenberger B, Hill SL, Caulfied MJ, et al. Integrating genomics into healthcare: a global responsibility. *Am J Hum Genet*. (2019) 104:13–20. doi: 10.1016/j.ajhg.2018.11.014
35. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, et al. 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*. (2015) 17:405–24. doi: 10.1038/gim.2015.30
36. Tung N, Lin NU, Kidd J, Allen BA, Singh N, Wenstrup RJ, et al. Frequency of germline mutations in 25 cancer susceptibility genes in a sequential series of patients with breast cancer. *J Clin Oncol*. (2016) 34:1460–8. doi: 10.1200/JCO.2015.65.0747
37. Gross T, Daniel J. Overview of pharmacogenomic testing in clinical practice. *Ment Health Clin*. (2018) 8:235–41. doi: 10.9740/mhc.2018.09.235
38. Oliveri S, Ferrari F, Manfrinati A, Pravettoni G. A systematic review of the psychological implications of genetic testing: a comparative analysis among cardiovascular, neurodegenerative and cancer diseases. *Front Genet*. (2018) 9:624. doi: 10.3389/fgene.2018.00624
39. National Institutes of Health. *What Does It Mean to Have a Genetic Predisposition to a Disease? Lister Hill National Center for Biomedical Communications*. (2020). Available online at: https://medlineplus.gov/genetics/understanding/mutationsanddisorders/predisposition/ (accessed July 4, 2021).
40. Dean M. "It's not if I get cancer, it's when I get cancer": BRCA-positive patients' (un)certain health experiences regarding hereditary breast and ovarian cancer risk. *Soc Sci Med*. (2016) 163:21–7. doi: 10.1016/j.socscimed.2016.06.039
41. Silva FC, Torrezan GT, Brianese RC, Stabellini R, Carraro DM. Pitfalls in genetic testing: a case of a SNP in primer-annealing region leading to allele dropout in *BRCA1*. *Mol Genet Genom Med*. (2017) 5:443–7. doi: 10.1002/mgg3.295
42. Moore v. *Curry County Health. Circuit Court of the State of Oregon for the County of Curry*. Gold Beach, OR (2017). Available online at: http://media.oregonlive.com/pacific-northwest-news/other/mastectomy.suit.pdf (accessed July 4, 2021).
43. Green. *Woman Claims she Underwent Mastectomy based on Mistaken Diagnosis: Settles for $600k+*. The Oregonian (2018). Available online at: https://www.oregonlive.com/pacific-northwest-news/2018/08/woman_claims_she_underwent_mas.html (accessed June 21, 2021).
44. Smith ED, Blanco K, Sajan SA, Hunter JM, Shinde DN, Wayburn B, et al. A retrospective review of multiple findings in diagnostic exome sequencing: half are distinct and half are overlapping diagnoses. *Genet Med*. (2019) 21:2199–207. doi: 10.1038/s41436-019-0477-2
45. Sauver JLS, Bielinski SJ, Olson JE, Bell EJ, Mc Gree ME, Jacobson DJ, et al. Integrating pharmacogenomics into clinical practice: promise vs reality. *Am J Med*. (2016) 129:1093–9. doi: 10.1016/j.amjmed.2016.04.009
46. Lo v. Burke. *Supreme Court of Virginia*. Richmond, VA (1995). Available online at: http://www.ecases.us/case/va/c1219127/lo-v-burke (accessed July 4, 2021).
47. Londin ER, Clark P, Sponziello M, Kricka LJ, Fortina P, Park JY. Performance of exome sequencing for pharmacogenomics. *Pers Med*. (2015) 12:109–15. doi: 10.2217/pme.14.77
48. Pate v. Threlkel. *Supreme Court of Florida*. Tallahassee, FL (1995).
49. Safer v. Estate of Pack. New Jersey Superior Court. Camden, NJ (1995).
50. AMA. Opinion 2.131 - disclosure of familial risk in genetic testing. *Virtual Mentor*. (2012) 14:627. doi: 10.1001/virtualmentor.2012.14.8.coet1-1208
51. ASHG. Professional disclosure of familial genetic information. The American Society of human genetics social issues subcommittee on familial disclosure. *Am J Hum Genet*. (1998) 62:474–83. doi: 10.1086/301707
52. Phillips v. United States. *US District Court for the District of South Carolina - 575 F. Supp. 1309*. Charleston, SC (1981). Available online at: https://law.justia.com/cases/federal/district-courts/FSupp/575/1309/1462039/ (accessed July 4, 2021).
53. Downs v. Trias. *Supreme Court of Connecticut*. Hartford, CT (2012). Available online at: https://casetext.com/case/downs-v-trias-nox10-uwy-cv-07-5009295s-aug (accessed July 4, 2021).
54. Marchant GE, Lindor RA. *Genomic Malpractice: An Emerging Tide or Gentle Ripple?* Food and Drug Law Institute (2018). p. 1–37. Available online at: https://www.fdli.org/2018/02/genomic-malpractice-emerging-tide-gentle-ripple/ (accessed January 16, 2020).
55. NIH. *FAQs About Rare Diseases*. (2017). Available online at: https://rarediseases.info.nih.gov/diseases/pages/31/faqs-about-rare-diseases (Accessed June 21, 2021).

Conflict of Interest: AP was employed by the company Hooper, Lundy & Bookman, and DW was employed by the company Walton Legal PLLC.

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.

frontiers in Genetics

SYSTEMATIC REVIEW
published: 09 August 2021
doi: 10.3389/fgene.2021.610728

Emerging Trends and Research Foci in Cataract Genes: A Bibliometric and Visualized Study

Hongli Zhu[1] *and Zhichang Zhang*[2]*

[1] *Department of Ophthalmology, The 4th People's Hospital of Shenyang, Shenyang, China,* [2] *Department of Computer, School of Intelligent Medicine, China Medical University, Shenyang, China*

Background: Approximately 50% of cataracts are associated with genetic factors. Genetic etiology and molecular mechanisms based on gene research increase the understanding of cataracts and provide direction for diagnosis and intervention. In the present study, SCIE papers related to the modeling of cataract gene research from 2010–2019 were evaluated and qualitative and quantitative analyses with modeling performed.

Methods: The SCIE database was searched on July 6, 2021 for cataract gene publications and relevant papers published since 2010 were considered for review. Subsequently, 1,904 SCIE papers associated with cataract genes from 2010–2019 were analyzed using a bibliometric method. The publication, country, institution, journal, references, knowledgebase, keywords, and research hotspots of the papers were analyzed using an online analysis platform of literature metrology, bibliographic item co-occurrence matrix builder (BICOMB), CiteSpace V, and VOS viewer analysis tool.

Results: 78 countries published the related articles, and the United States ranks of America had the most publications. Two thousand seven hundred and eighty three institutions contributed to the related publications. Fudan University had the most publications. The reference clusters of SCI papers were clustered into six categories, namely, causing congenital cataract-microcornea syndrome, functional snp, cataractous lenses, a1 mutation, foxe3 mutation, cell adhesion gene pvrl3, nid1 gene. The key words representing the research frontiers were cerebrotendinous xanthomatosis (2017-2019), oxidative stress (2017–2019).

Conclusion: This study provided a systematic, objective and comprehensive analysis of the literature related to gene research of cataract. Moreover, this study demonstrated the current hotspots and the future trends in the field of gene research of cataract. This review will help ophthalmologist to discern the dynamic evolution of cataract gene research, as well as highlight areas for future research.

Keywords: cataract, gene, data visualization, bibliometric, citespace, VOSviewer

OPEN ACCESS

Edited by:
Maria G. Stathopoulou, Institut National de la Santé et de la Recherche Médicale (INSERM), France

Reviewed by:
Mohammed Abdallah, University of Cape Town, South Africa
Sergio Cuevas, National Autonomous University of Mexico, Mexico

***Correspondence:**
Zhichang Zhang
zczhang@cmu.edu.cn

Specialty section:
This article was submitted to Genomic Medicine, a section of the journal Frontiers in Genetics

Received: *27 September 2020*
Accepted: *15 July 2021*
Published: *09 August 2021*

Citation:
Zhu H and Zhang Z (2021) Emerging Trends and Research Foci in Cataract Genes: A Bibliometric and Visualized Study. Front. Genet. 12:610728. doi: 10.3389/fgene.2021.610728

BACKGROUND

Cataract is not only a multifactorial but also a monogenic disease. In addition to genetic factor, its pathogenesis and development are also related to age, gender, radiation, oxidation, physical injury, diet, and medication. These factors can lead to abnormal gene expression and affect the transparency of the lens, and eventually form cataract. Despite the in-depth study of eye's genes, It is still hard to find many causes that lead to abnormal lens (Gillespie et al., 2014). Hereditary cataract is a kind of clinical and genetic heterogeneity disease (Berry et al., 1999). Isolated cataract and syndromic congenital cataract are a heterogeneous developmental defect. The identification of the related genes is challenging (Anand et al., 2018b). After overcoming the bottleneck of genetic heterogeneity, more and more people realize that the explanation of genetic variation and the relationship between new genes and specific phenotypes are still challenging. However, further understanding of the heredity and variation basis of lens and anterior segment abnormalities will be of great value to our understanding of eye diseases.

Genetic studies have identified mutations in over 30 causative genes for congenital or other early-onset forms of cataract as well as several gene variants associated with age-related cataract (Shiels and Hejtmancik, 2017). Cataract is a major cause of blindness worldwide. It is characterized by lens opacification and is accompanied by extensive post-translational modifications (PTMs) in various proteins (Zhang et al., 2018). PTMs play an essential role in lens opacification. And post-translational modification (PTM) of lens proteins is believed to play various roles in age-related lens function and development. Several PTMs have been described in proteins isolated from relatively old human lenses, including phosphorylation, deamidation, racemization, truncation, acetylation, and methylation. An overwhelming majority of previous cataract proteomic studies have exclusively focused on crystallin proteins, which are the most abundant proteome components of the lens. Elucidating the role of these modifications in cataract formation has been a challenging task because they are among the most difficult PTMs to study analytically (Huang et al., 2011). The proteomic status of some amides presents similar properties in normal aged and cataractous lenses, whereas some may undergo greater PTMs in cataract.

In this study, conducted from Jan 1, 2010 to Dec 31, 2019, we analyzed the SCIE papers for studies related to cataract gene research using bibliometric methods, and included articles on cataract research published in various countries, regions, and by different research institutions. We further analyzed journals that published papers on cataract research, and we analyzed the "top 10 cited references," and we calculated the number of times popular references were cited. By clustering the reference network of co-cited references, we also analyzed the knowledge base of this topic. The research hotspots of this topic were detected by burst keywords, which could provide some reference for future relevant research (Small, 1973; Chen, 2006). These analyses afford ophthalmologist with both a macroscopically understanding and a microscopically characterization of the knowledge domain as a whole. Compared with traditional systematic reviews written by experts, this bibliometric analysis provides a timely, visual, and unbiased approach to track the development and explore the specific knowledge domains.

Abbreviations: BICOMB, bibliographic item co-occurrence matrix builder; CTX: cerebrotendinou xanthomatosis; WoSCC, Web of Science Core Collection; TRPM3, transient receptor potential cation channel subfamily M member-3; LOD, logarithm of odds; EPHA2, Eph-receptor type-A2; ADCC, autosomal dominant congenital cataract.

METHODS

All data were downloaded from Web of Science Core Collection (WoSCC) on July 6, 2021, and were verified by two authors (ZH and ZZ) independently. The literature research was performed for publications from 2010 to 2019, using the following search terms (TS = cataract* and TS =gene), and The literature types was all document types. We collected the following basic information for each article: title, abstract, authors, institution, country/region, journal, keywords, and references. Articles that met the following criteria were included: (1) those indexed in the Web of Science Core Collection and (2) The following articles were excluded: (1) irrelevant meeting abstracts, irrelevant proceedings paper, book chapter, data paper, editorial material, and repeated articles and (2) unpublished documents without enough information for further analysis. A total of 83 papers with duplicates were excluded. The detailed search processes and analysis procedures were shown in **Figure 1**.

DATA ANALYSIS

We tried to describe all publication characteristics, including countries, institutes, journals, keywords, and so on. We inquired the H-index, which was regarded as an important indicator to measure the scientific value of research (Eyre-Walker and Stoletzki, 2013). In this study, the Online Analysis Platform of Literature Metrology (http://bibliometric.com/), CiteSpace V (Drexel University, Philadelphia, PA, USA) and VOSviewer (Leiden University, Leiden, the Netherlands) were used to perform co-occurrence analysis and visualize the collaborative networks of the countries/institutes/journals/keywords. Through CiteSpace, reference co-citation analysis was performed, and related knowledge maps were constructed, and burst keyword detection was also performed to investigate the recurrent new keywords (Chen, 2006).

RESULTS

Distribution of Articles by Publication Years

A total of 1904 papers from 2010 to 2019 were published. **Figure 2** shows the trend in the number of cataract gene related publications. Since 2017, the activity in cataract gene research reached a peak.

Countries/Regions and Institutes

A total of 78 countries / regions published the related articles, Collaborations among these countries were shown

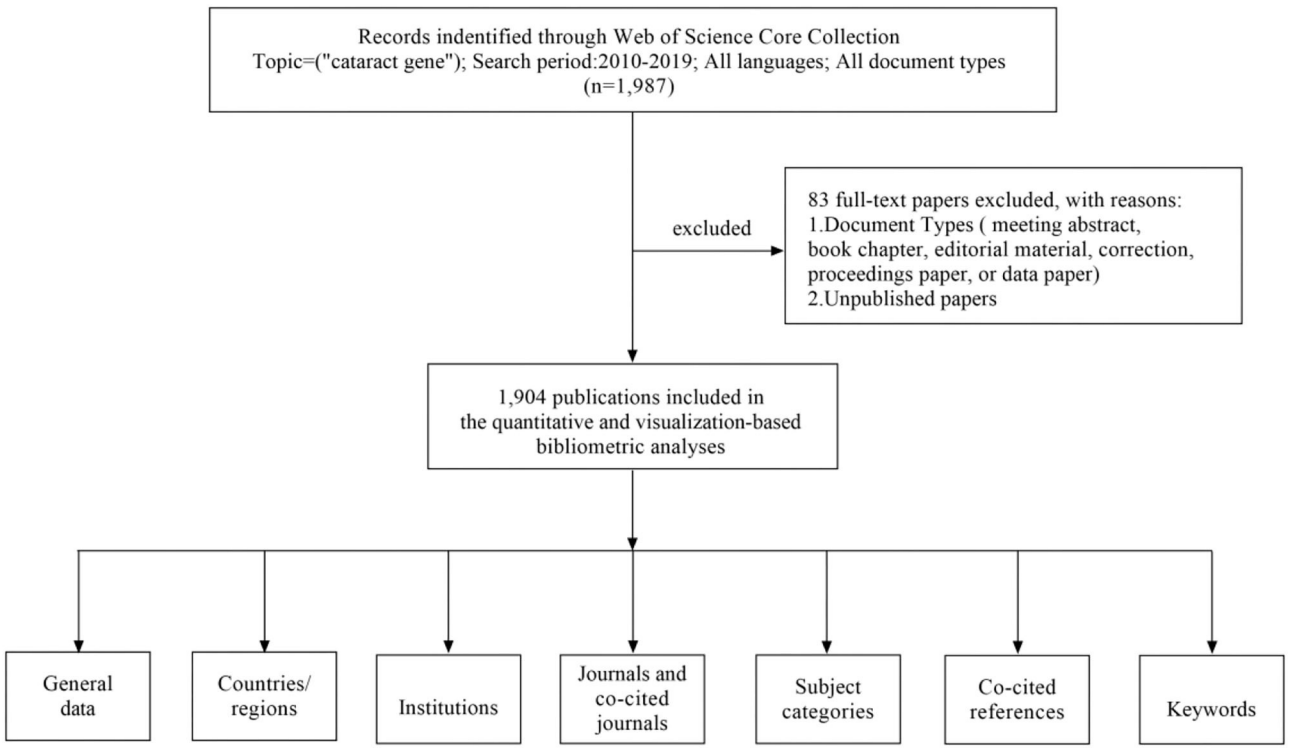


FIGURE 1 | A frame flow diagram. The diagram showed details selection criteria for Cataract gene publications from WoSCC database and the steps of bibliometric analysis.

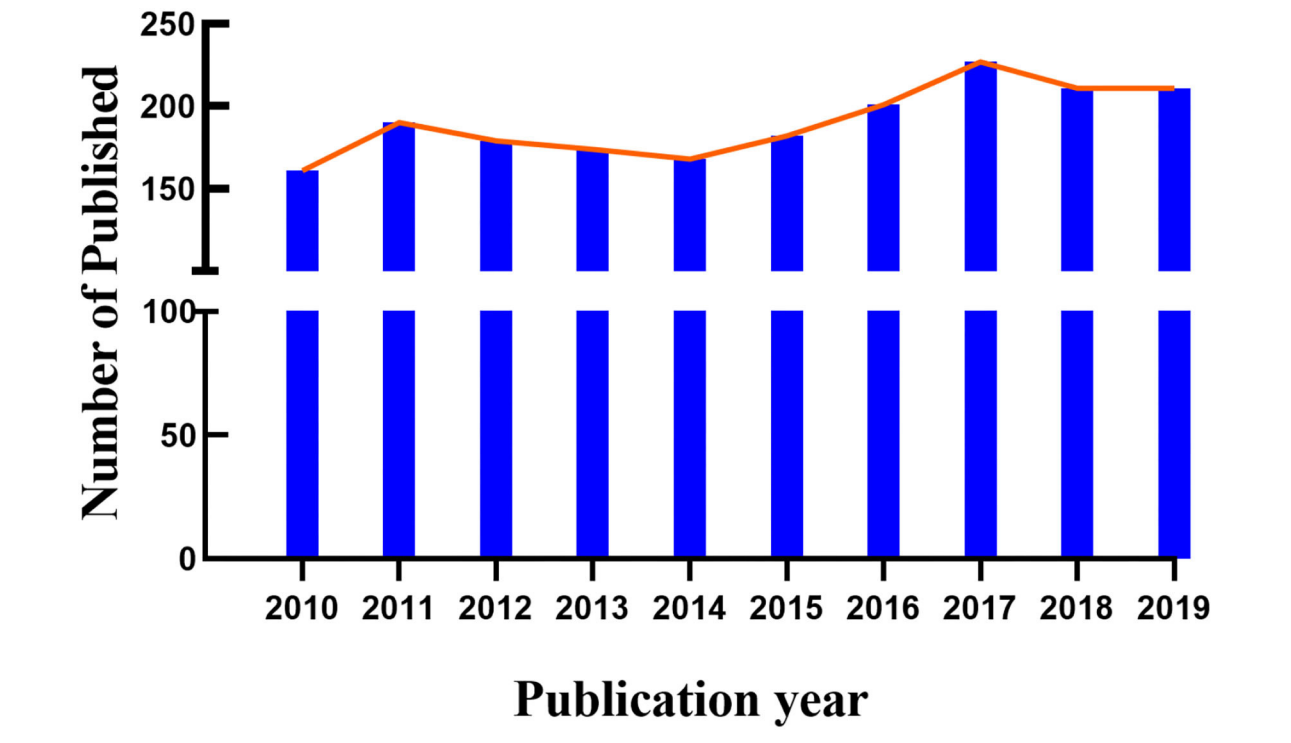


FIGURE 2 | Trends in the number of publications on Cataract-gene from 2010 to 2019.

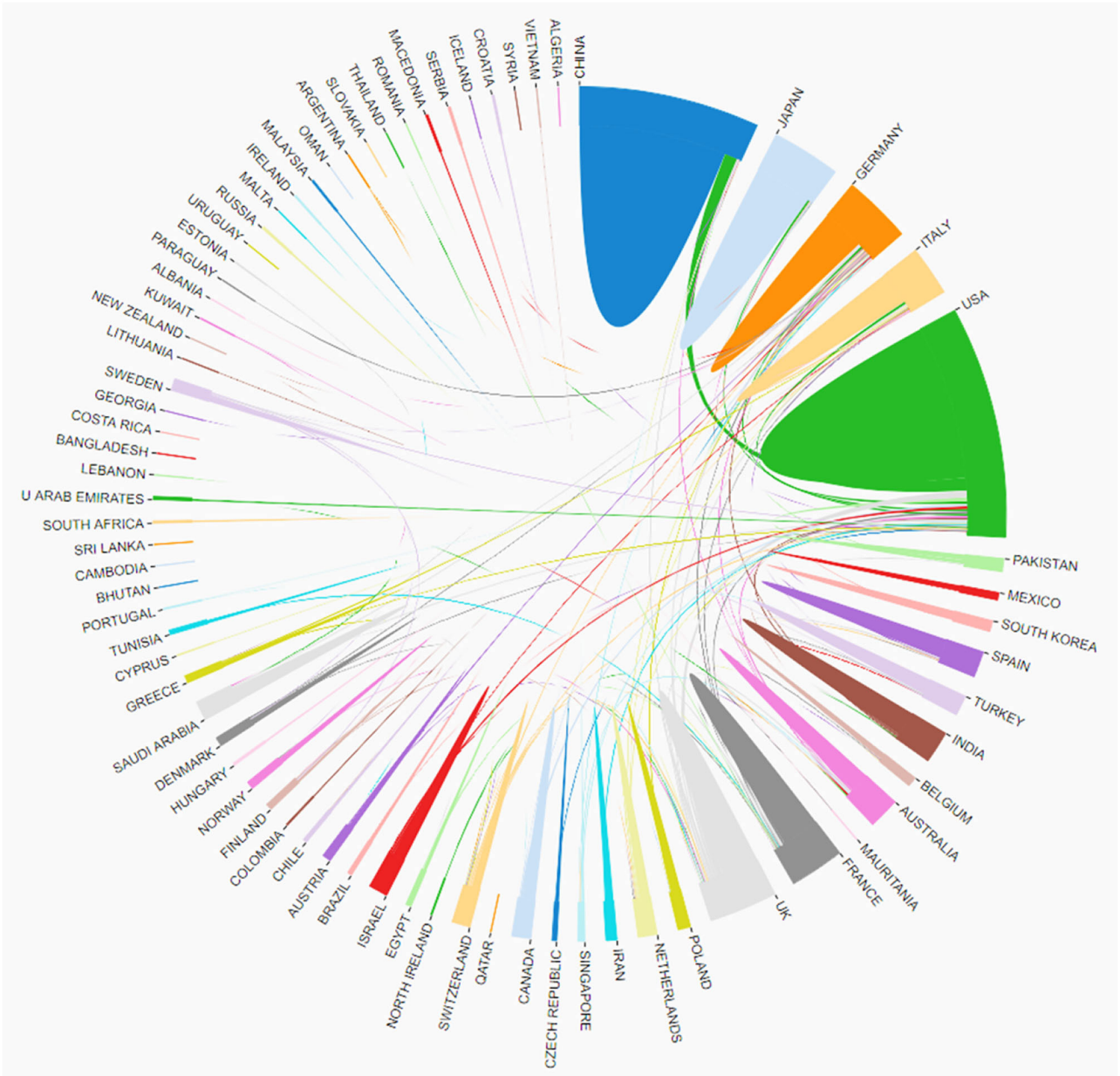


FIGURE 3 | The cooperation of countries/regions contributed to publications on Cataract-gene from 2010 to 2019.

in **Figure 3**. The top 10 countries were listed in **Table 1**. The United States of America (USA) had the most publications (581), followed by China (487), Germany (160) and the United Kingdom (146).

A total of 2,783 institutes contributed to the related publications, the top 10 institutes listed in **Table 1**. Collaborations among these institutes were shown in **Figure 4**. Fudan University had the most publications (52), followed by Sun Yat Sen University (45), the National Eye Institute (42), and Capital Med University (33).

Journals

The referential relationship of academic journals represents the situation of knowledge exchange in the research field in question, where the citing papers form the frontier of knowledge, and the cited papers the knowledge basis. The top 10 journals were presented in **Table 2**. Collaborations among these journals were shown in **Figure 5**. The first was MOL VIS (151), followed by INVEST OPHTH VIS SCI (92).

The dual-map overlay of journals is shown in **Figure 6**, with the citing journals on the left side, cited journals on

TABLE 1 | The top 10 Countries/Regions and Institutions to publications on Cataract-gene from 2010 to 2019.

Rank	Countries/regions	Count	H-index	Institutions	Count	H-index
1	USA	581	46	FUDAN UNIV	52	14
2	PEOPLES R CHINA	487	28	SUN YAT SEN UNIV	45	11
3	GERMANY	160	31	NEI	42	16
4	UK	146	28	CAPITAL MED UNIV	33	12
5	JAPAN	138	23	WASHINGTON UNIV	32	11
6	INDIA	104	18	ZHEJIANG UNIV	31	14
7	FRANCE	95	24	CENT S UNIV	30	15
8	ITALY	103	23	JOHNS HOPKINS UNIV	30	7
9	CANADA	70	21	CHINESE ACAD MED SCI	27	14
10	AUSTRALIA	67	22	HARVARD UNIV	26	12

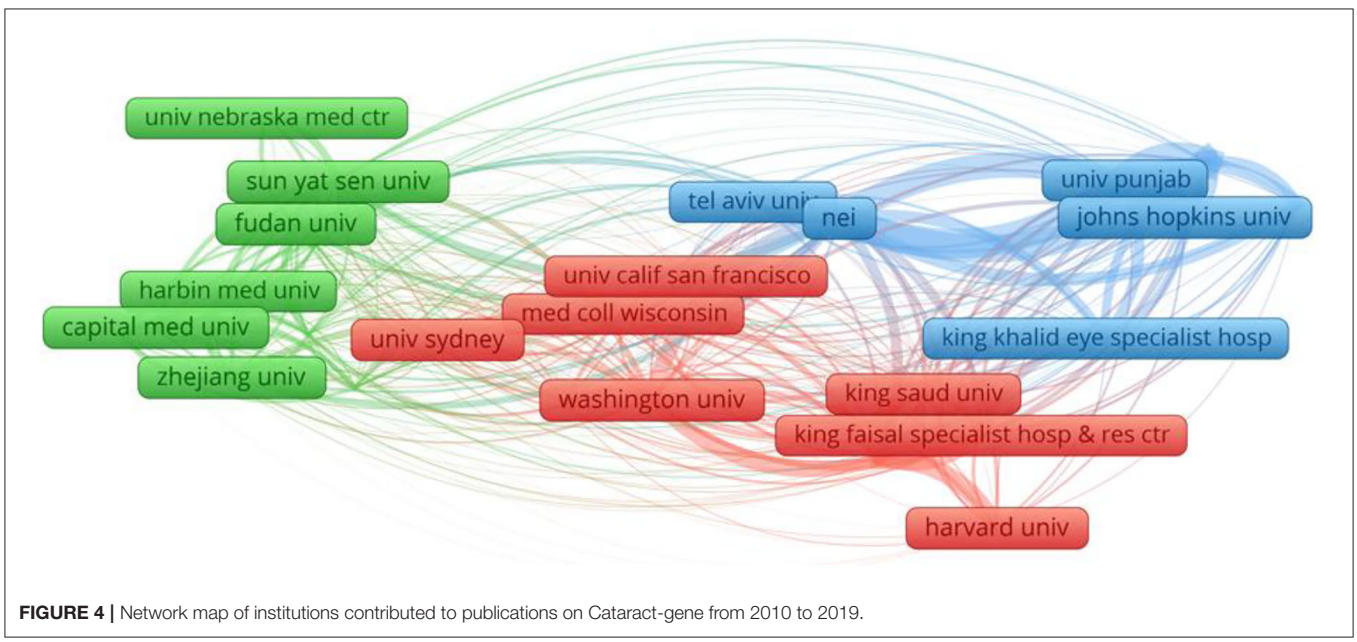


FIGURE 4 | Network map of institutions contributed to publications on Cataract-gene from 2010 to 2019.

the right side, and the colored paths indicate the citation relationships. The color paths indicated that studies, published in Molecular/Biology/Genetics journals, are usually cited in the studies, published in Molecular/Biology/Immunology and Neurology/Sports/Ophthalmology journals.

References

Analysis of references is one of the most important indicators of bibliometric. Frequently cited documents are generally of great influence in their respective research field. On this topic, a co-cited documents-based clustering analysis may present subfields and connecting nodes of the research in question.

A network of co-cited references was constructed to assay the scientific relevance of the related publications (**Figure 7**). The cluster setting parameters were as follows: # Years Per Slice = 2, Top N% = 0.3, pruning algorithm was adopted. The Modularity Q score was 0.7347, >0.5, showing the network was reasonably divided into loosely coupled clusters. The Weighted Mean silhouette score was 0.9383, more than 0.5, indicating that the homogeneity of these clusters was acceptable. Index items extracted from literature were used as cluster markers. The largest cluster #0 was marked as "causing congenital cataract-microcornea syndrome" (Huang and He, 2010; Hu et al., 2010; Jiang, 2010; Li et al., 2010) the next cluster #1 was marked as "functional snp"(Gu et al., 2016; Wang et al., 2016; Anand et al., 2018a), the second cluster #2 was marked as "cataractous lenses" (Sousounis and Tsonis, 2012; Weisschuh et al., 2012; Deng and Yuan, 2014), the third cluster #3 was marked as "a1 mutation" (Zhu et al., 2010; Yu et al., 2012; Wu et al., 2017), the fourth cluster #4 was marked as "foxe3 mutation" (Wada et al., 2011; Anand et al., 2018a; Plaisancie et al., 2018), cluster #5 was marked as "cell adhesion gene pvrl3" (Lachke et al., 2012; Weatherbee et al., 2019), cluster #6 was marked as "nid1 gene" (Murgiano et al., 2014; Osinchuk et al., 2017; Braun et al., 2019).

TABLE 2 | The top 10 journals and references of publications on Cataract-gene from 2010 to 2019.

Rank	Source titles	Count	Title of co-cited reference	Pmid	Count	Interpretation of the findings
1	MOL VIS	151	Congenital cataracts and their molecular genetics (Hejtmancik, 2008)	18035564	91	This study reported the molecular genetics of congenital cataracts.
2	INVEST OPHTH VIS SCI	92	Cat-Map: Putting Cataract on the Map (Shiels et al., 2010)	21042563	77	This paper summarized the genetic complexity of Mendelian and age-related cataract by Cat-Map.
3	PLOS ONE	89	Genetics of Human Cataract (Shiels and Hejtmancik, 2013)	23647473	42	This study showed the genetics associated with cataract.
4	OPHTHALMIC GENET	47	A Method and Server for Predicting Damaging Missense Mutations (Adzhubei et al., 2010)	20354512	37	This article presented PolyPhen-2 software to analyze the effects of non-synonymous mutations on proteins in humans
5	CURR EYE RES	42	Molecular Characteristics of Inherited Congenital Cataracts (Huang and He, 2010)	20624502	33	This paper showed the characteristics of inherited congenital cataracts.
6	EXP EYE RES	42	Comprehensive mutational screening in a cohort of Danish families with hereditary congenital cataract (Hansen et al., 2009)	19182255	32	This study analyzed family cohorts with inherited congenital cataract in a comprehensive mutation screening strategy and demonstrated the effectiveness of the strategy.
7	AM J MED GENET A	41	Crystallin gene mutations in Indian families with inherited pediatric cataract (Devi et al., 2008)	18587492	32	This document identified the frequency and frequency of crystalloprotein gene mutations in a population of Indian cataract patients
8	INT J OPHTHALMOL-CHI	28	The EPHA2 Gene Is Associated With Cataracts Linked to Chromosome 1p (Shiels et al., 2008)	19005574	29	This article provided the EPHA2 gene is associated with cataracts Linked to chromosome 1p.
9	SCI REP-UK	27	Mutation analysis of CRYAA, CRYGC, and CRYGD associated with autosomal dominant congenital cataract in Brazilian families (Santana et al., 2009)	19390652	29	This study reported the results of an analysis of mutations in CRYAA, CRYGC, and CRYGD associated with autosomal dominant congenital cataracts.
10	AM J HUM GENET	23	Mutations in the RNA granule component TDRD7 cause cataract and glaucoma (Lachke et al., 2011)	21436445	28	This article identified that TDRD7-RGs play an essential role in the regulation of specific genes that are critical for lens development.

Keywords

Keywords in the related publications were extracted and analyzed. The top 20 keywords were listed in **Table 3**. In addition to cataract, congenital cataract and mutation occurred more than 50 times. Keywords analysis of the 1,904 articles identified 100 keywords with a minimum of 20 occurrences and divided them into five clusters (cataract, gene, mutation, age-related cataract, differentiation) (**Figure 8**). We analyzed the temporal trend of hotspot shift according to the top 11 keywords with the strongest citation bursts, such as cerebrotendinous xanthomatosis (2017–2019), and oxidative stress (2017–2019) (**Figure 9**).

DISCUSSION

General Data

One thousand nine hundred and four SCI papers related to the gene research of cataract and published from 2010 to 2019 were analyzed in this study. The United States of America had the most publications (581), accounting for 30.9%. China had the second most publications (487), accounting for 27.1%. The top 10 institutions included three in the United States of America, six in Asia, and one in Australia. MOL VIS was the most published journal. These observations showed that MOL VIS predominantly contributed to the research in this field. In addition, the top 10 cited publications were also investigated. The first article was published in SEMINARS IN CELL & DEVELOPMENTAL BIOLOGY by HEJTMANCIK JF, and was cited 91 times. The second was published in MOLECULAR VISION by SHIELS A, which was cited 77 times.

Knowledge Base

According to previous studies, many genetic types of cataract have been investigated in animal models. Substantial advances have taken place regarding the mapping of genes and their variations involved in congenital cataract formation, and the genetic causes of age-related cataract have been discovered. As shown in **Figure 6**, after clustering the co-cited references, the key nodes in the clustering resulted reveal the knowledge bases in this research field. Namely: #0 "causing congenital cataract-microcornea syndrome," #1 "functional snp," #2 "cataractous

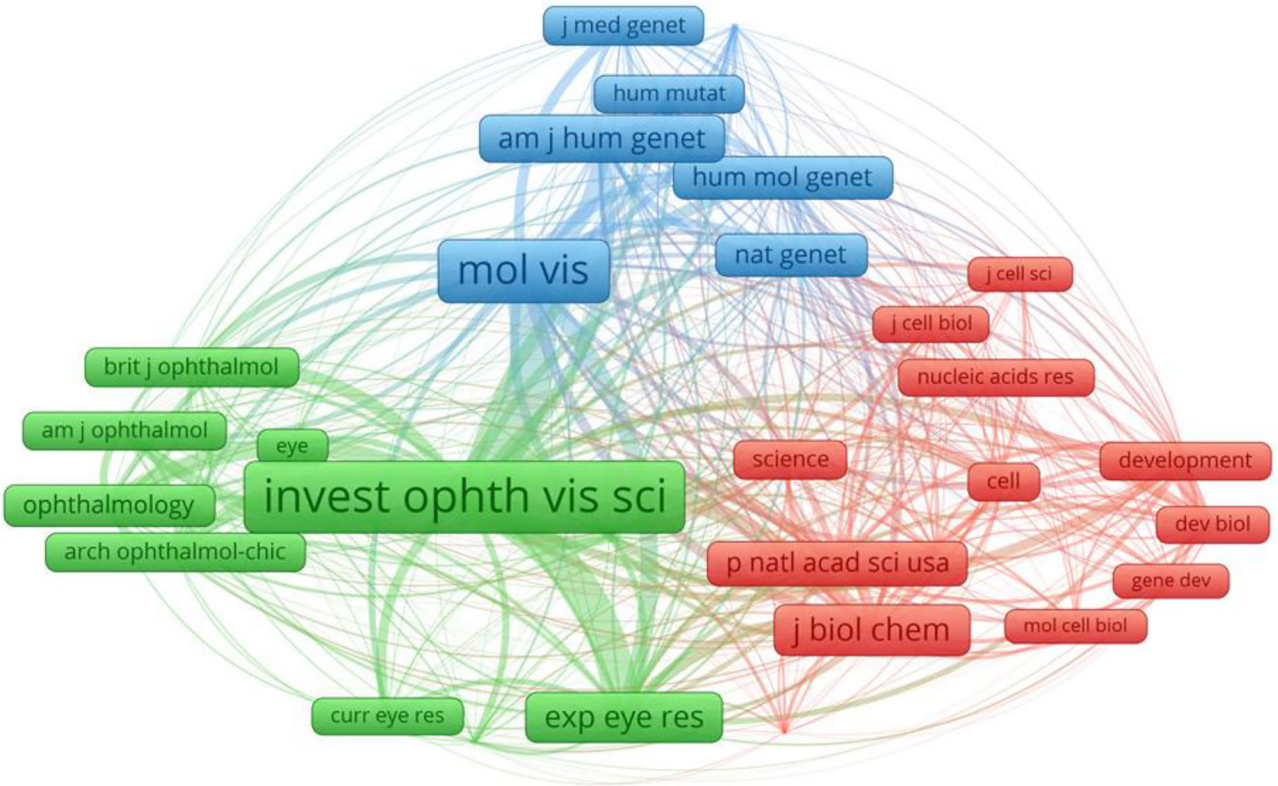


FIGURE 5 | The network map of cited journals contributed to publications on Cataract-gene from 2010 to 2019.

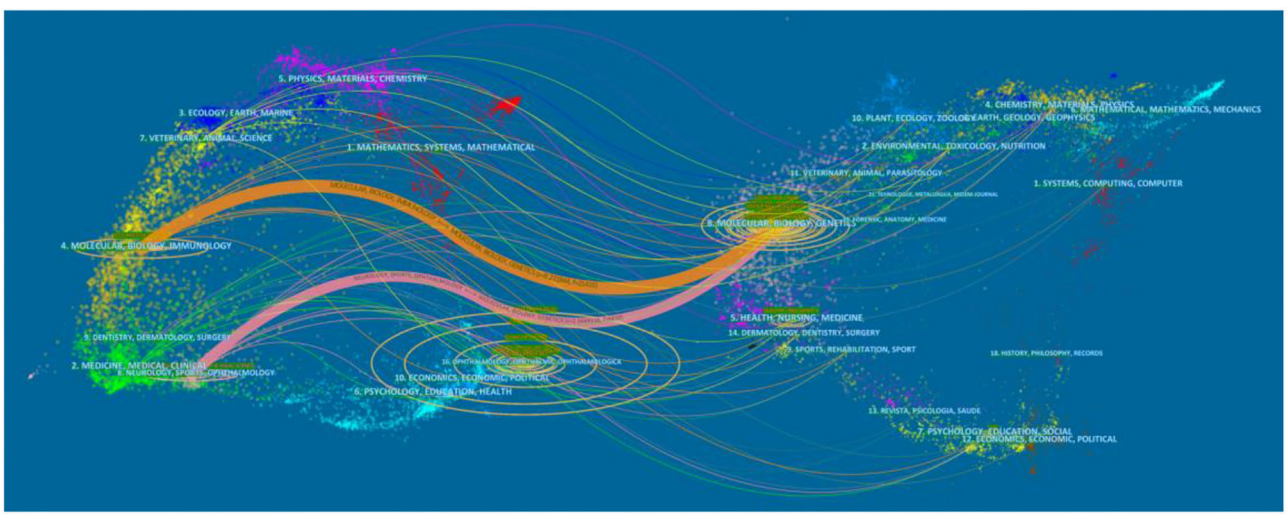


FIGURE 6 | The dual-map overlay of journals contributed to publications on Cataract-gene from 2010 to 2019.

lenses," #3 "a1 mutation," #4 "foxe3 mutation," #5 "cell adhesion gene pvrl3," #6 "nid1 gene." This paper described the knowledge base of cataract gene research according to different clusters.

In #0 "causing congenital cataract-microcornea syndrome," Shanshan Hu et al. identified the underlying genetic defect in a four-generation family of Chinese origin with autosomal dominant congenital cataract-microcornea syndrome (CCMC). They direct sequencing of the encoding regions of the candidate genes revealed a heterozygous mutation c.592CT in exon 2 of the gap junction protein, alpha8 (GJA8) gene. This mutation was

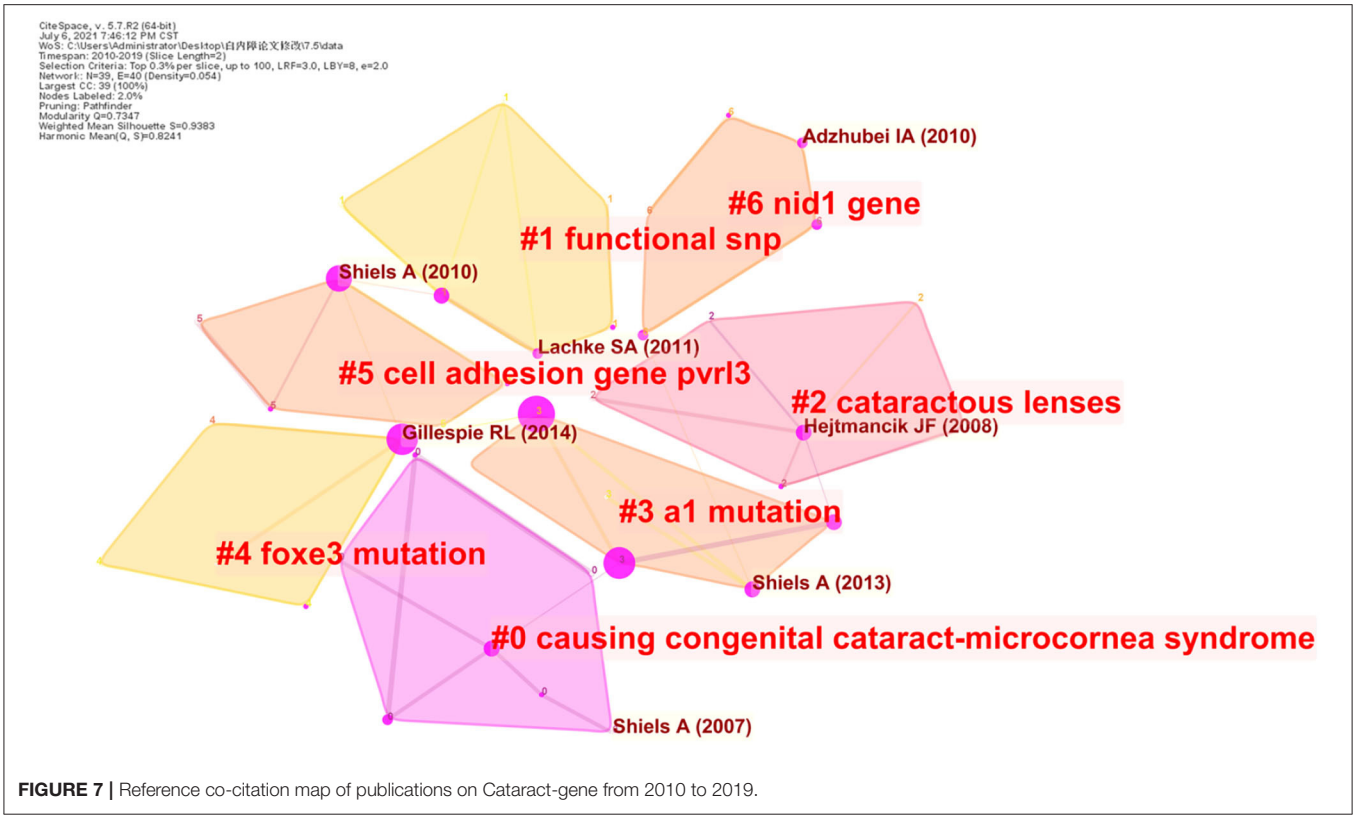


FIGURE 7 | Reference co-citation map of publications on Cataract-gene from 2010 to 2019.

TABLE 3 | The top 20 Keywords on Cataract-gene from 2010 to 2019.

Rank	Keyword	Count	Rank	Keyword	Count
1	Cataract	147	11	Gene expression	12
2	Congenital cataract	72	12	Apoptosis	12
3	Mutation	50	13	Lowe syndrome	11
4	Len	34	14	pax6	11
5	Age-related cataract	24	15	gja8	9
6	Exome sequencing	16	16	Microphthalmia	8
7	Genetics	15	17	Next-generation sequencing	8
8	Oxidative stress	15	18	Linkage	7
9	Aniridia	15	19	Anophthalmia	7
10	Crystallin	14	20	Retinitis pigmentosa	6

responsible for the familial disorder through the substitution of a highly conserved arginine to tryptophan at coden198 (p.R198W). That report is the first to relate p.R198W mutation in GJA8 with CCMC. The result expanded the mutation spectrum of GJA8 in associated with congenital cataract and microcornea and implied that this gene had direct involvement with the development of the lens as well as the other anterior segment of the eye.

In #1 "functional snp," SNPs is single-nucleotide polymorphisms. As DNA repair is implicated in ARC pathogenesis and SNPs in the 3′-terminal untranslated region (3′-UTR) targeted by micro RNA(miRNAs) can alter the gene function.

In #2 "cataractous lenses," Konstantinos Sousounis et al. had examined the patterns of gene expression in cataractous lenses. The purpose was to evaluate unique and common patterns of gene expression during development, aging and cataracts.

In #3 "a1 mutation," Yanan Zhu et al. first report of a phenotype of progressive nuclear and cortical cataracts related to the βA1/A3-crystallin gene (CRYBA3/A1) mutation IVS3+1 G>A. This finding expands the spectrum of cataract phenotypes caused by the IVS3+1 G>A mutation of CRYBA3/A1, confirms the phenotypic heterogeneity of this mutation and suggests the mechanism that influences the cataractogenesis in different ethnic backgrounds. C.30-2 A>G mutation of CRYBA3/A1 gene is a novel mutation and broadens the genetic spectrum of ADCC.

In #4 "foxe3 mutation," FOXE3 gene, which was initially described in individuals with dominantly inherited anterior segment dysgenesis and, subsequently, associated with recessively inherited primary aphakia, sclerocornea and microphthalmia. Mutations in the transcription factor genes FOXE3 cause congenital lens defects including cataracts that may be accompanied by defects in other components of the eye or in nonocular tissues. All individuals with ocular abnormalities described in the literature for which a FOXE3 mutation was identified and demonstrated that correlations exist between the mutation type, mode of inheritance and the phenotype severity. a mutation was located in the regulatory regions of the Foxe3 gene. This gene is responsible for cataracts in humans and

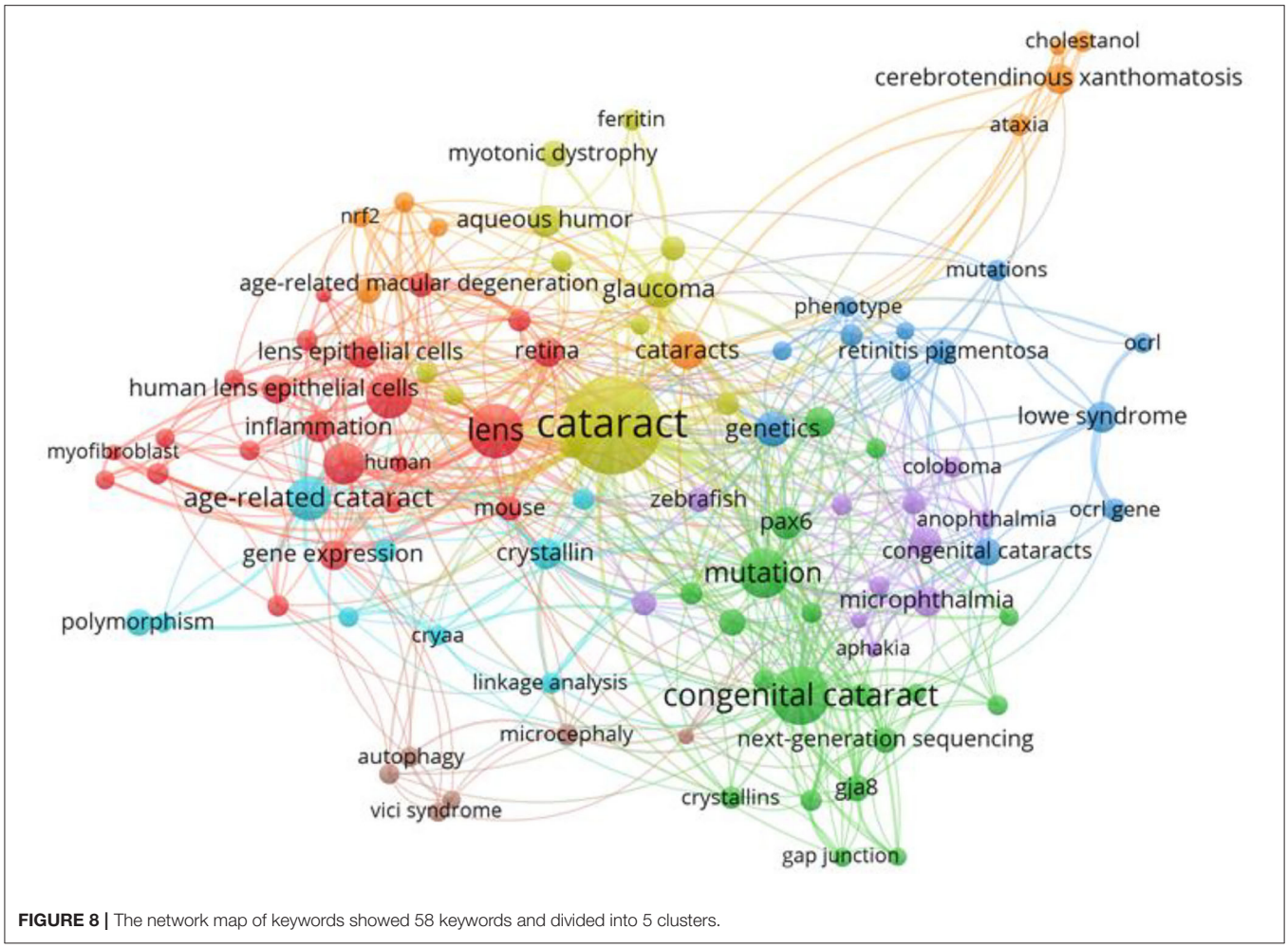


FIGURE 8 | The network map of keywords showed 58 keywords and divided into 5 clusters.

mice, and it plays a crucial role in the development of the lens. Furthermore, mutation of Foxe3 causes various ocular defects. Kenta Wada et al. suggested that cataracts in rct mice were caused by reduced Foxe3 expression in the lens and that this decreased expression was a result of a deletion in a cis-acting regulatory element. Deepti Anand et al. comprehensively describe here all the variants in FOXE3 genes linked to human developmental defects. A total of 52 variants for FOXE3, the effort revealed FOXE3 had 33 unique causal mutations. Finally, they made the detailed FOXE3 variant information available in the Leiden Online Variation Database (LOVD) platform at https://www.LOVD.nl/FOXE3.

In #5 "cell adhesion gene pvrl3," the expression of PVRL3, which encodes the cell adhesion protein Nectin 3, is significantly reduced in patient DGAP113 lymphoblastoid cells, likely due to a position effect caused by the chromosomal translocation. Moreover, Pvrl3 knockout mice as well as a spontaneous mouse mutant ari (anterior retinal inversion), that maps to the Pvrl3 locus, exhibit lens and other ocular defects involving the ciliary body. Collectively, these data identify PVRL3 as a critical gene involved in a Nectin-mediated cell-cell adhesion mechanism in human ocular development.

In #6 "nid1 gene," nidogen 1 (NID1) gene (c.3579_3604+829del) deletion leads to the skipping of exon 19 during transcription and is therefore predicted to cause a frameshift and premature stop codon (p.1164fs27X). Nidogen 1 deficient mice show neurological abnormalities and highly irregular crystal lens alterations. This study adds NID1 to the list of candidate genes for inherited cataract in humans and is the first report of a naturally occurring mutation leading to non-syndromic cataract in cattle provides a potential large animal model for human cataract.

Research Hotspots and Frontiers

Keywords concentrate expression of current research issues or concepts. Burst keywords stand for emerging trends and research frontiers. In the present study, We further used CiteSpace to capture the burst keywords. Two frontiers of related research were found as follows: cerebrotendinous xanthomatosis (2017–2019) (Tibrewal et al., 2017; Freedman et al., 2019), oxidative stress (2017–2019) (Zoric et al., 2008; Petrou and Terzidaki, 2017), and these key words cover the research frontier of the current topic.

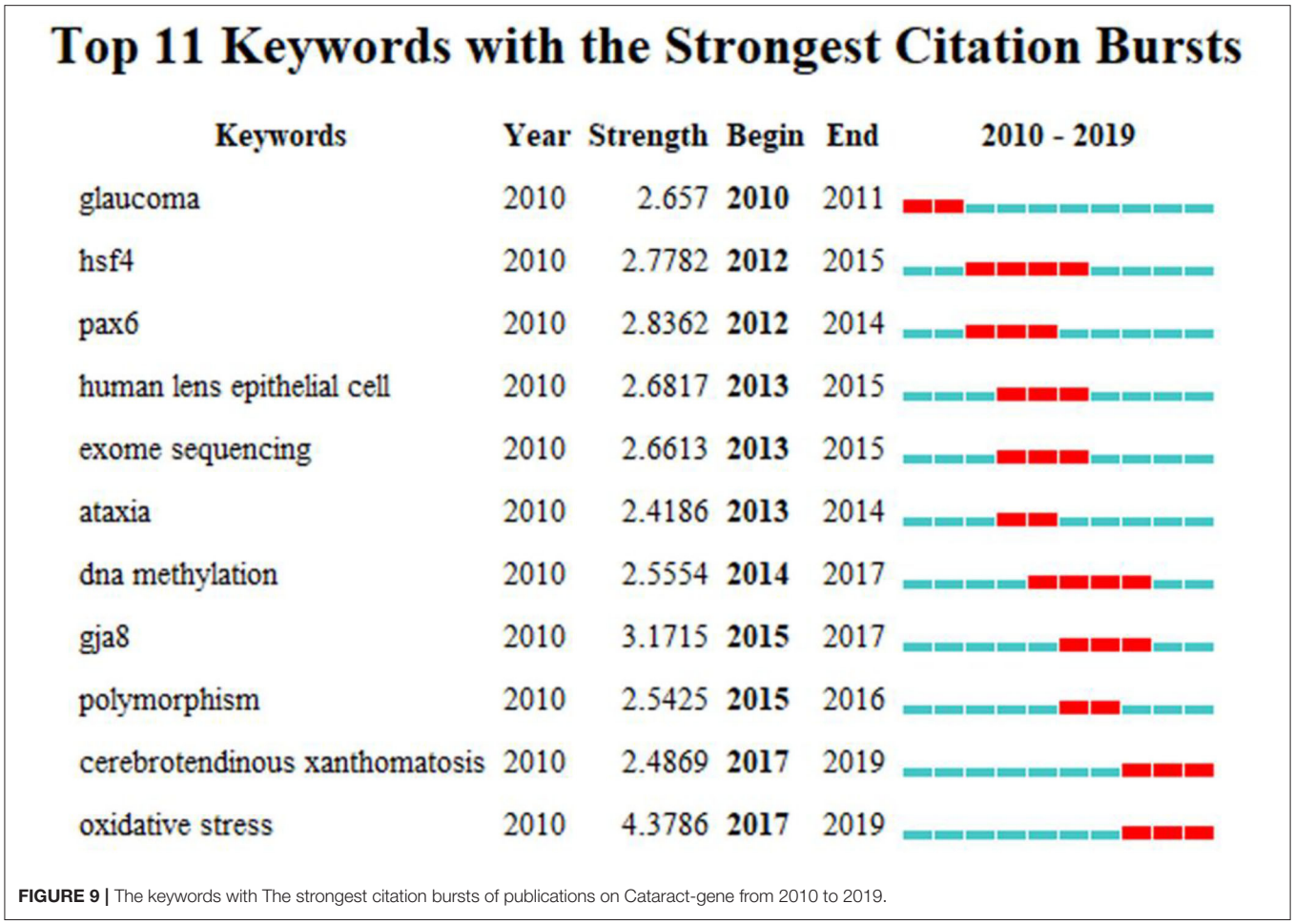


FIGURE 9 | The keywords with The strongest citation bursts of publications on Cataract-gene from 2010 to 2019.

Cerebrotendinous Xanthomatosis

CTX is a uncommon autosomal recessive metabolic condition, which is characterized by multiple system damage caused by lipid metabolism disorders, and is often manifested as tenoxanthoma. In cataract in adolescence and arteriosclerosis in early years, cholestanol and other metabolites are often deposited in lipid-rich nerve tissue, which can cause neurotoxicity and lead to damage of central and peripheral nerves. The pathogenic gene of CTX is located at 2q33-qter. The variation of the CTX gene causes the deficiency of CYP27, resulting in cholesterol metabolism disturbance and the accumulation of cholesterol, cholestanol and other neurotoxic substances in various tissues and organs, particularly nervous tissue. Freedman et al. studied the prevalence of the CTX gene in a study population that was diagnosed with idiopathic, early-onset, and bilateral cataracts. The study included patients aged 2–21 years; 1.8% of these patients were diagnosed as having the CTX gene variation. Within this study population, the CTX gene was about 500-fold the presently estimated its prevalence in human (3 to 5/100 000). The data suggested that idiopathic, early-onset, and bilateral cataracts could be a screening tool for CTX for early identification (Freedman et al., 2019). Tibrewal et al. previously reported the case of a child who presented with bilateral cataracts, which led to a diagnosis of CTX. They described the cataracts' morphologic characteristics and the outcome of systemic treatment on cataract. The authors believed CTX to be an uncommon autosomal recessive disease that results in lipid storage abnormalities and presents in a wide range of clinical manifestations, including juvenile bilateral cataracts. Left untreated, CTX can lead to irreversible progressive neurologic devastation and early death. Frequently, juvenile bilateral cataracts occur in early childhood; this allows ophthalmologists the chance to make an early diagnosis and initiate the sysmetic treatment (Tibrewal et al., 2017).

Oxidative Stress

In vivo, oxidative stress is a state of imbalance between oxidation and antioxidation; this state results in inflammatory infiltration of neutrophils, an increase in protease secretion, and a production of numerous oxidation intermediates. Oxidative stress has a negative effect on the body as a result of free radicals and is considered to be an important factor leading to aging and disease. Zoric et al. reviewed a retrospective cross-sectional study that evaluated 80 samples of aqueous humor and corticonuclear lens blocks. The authors believed that cataract type and pigmentation may be dictated by the form and intensity

of oxidative stress. If true, this would make efforts in cataract prevention challenging and more complex. Zoric et al. proposed that the role of oxidative stress in cataract formation was not the same for all cataract types. For example, high levels of lipid peroxides may result in certain pigmented cataracts whereas lipid peroxidation and consumption of SH groups (as seen in the development of cortical cataracts) might be of less importance (Zoric et al., 2008). Petrou et al. calculated the thermodynamic parameters for nucleation, elongation, fibrillization, and other processes of proteinaceous diseases that were related to β-amyloid protein (Alzheimer disease), tau protein (Alzheimer and Pick disease), α-synuclein (Parkinson disease), prion, amylin (type 2 diabetes), and α-crystallin (cataract). From kinetic data (k, T), it can be concluded that $\Delta G\neq$ is equal to the energy needed for ground state oxygen excitation of the singlet oxygen state ($1\Delta g$, first excited). The similarity of $\Delta G\neq$ values is an indication that there may be a common mechanism in the previously mentioned disorders. Petrou and colleagues attributed this common mechanism to oxidative stress and specifically to the singlet oxygen molecule ($1\Delta g$) (Petrou and Terzidaki, 2017).

CONCLUSION

Using the bibliometric analysis this study provided a systematic analysis of the literature related to cataract gene. The analysis was also objective and comprehensive. Moreover, this study demonstrated the research basis, the current hotspots and the future trends in the field of cataract gene. The knowledge bases in this research field were causing congenital cataract-microcornea syndrome, functional snp, cataractous lenses, a1 mutation, foxe3 mutation, cell adhesion gene pvrl3, nid1 gene. The emerging trends and research frontiers of current research theme were cerebrotendinous xanthomatosis and oxidative stress. Papers published at different stages were collected for this study, some of which were not comprehensive and might have publication bias, which may affect the results of this systematic review.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

HZ and ZZ designed and conceived the general idea and context of this review, wrote and integrated all sections, and contributed to the relevant references of this manuscript. All the authors read and approved the final manuscript.

FUNDING

This study was supported by grants from Association of Fundamental Computing Education in Chinese Universities (AFCEC) to Zhichang Zhang (2019-AFCEC-199).

REFERENCES

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

Anand, D., Agrawal, S. A., Slavotinek, A., and Lachke, S. A. (2018a). Mutation update of transcription factor genes FOXE3, HSF4, MAF, and PITX3 causing cataracts and other developmental ocular defects. *Hum. Mutat.* 39, 471–494. doi: 10.1002/humu.23395

Anand, D., Kakrana, A., Siddam, A. D., Huang, H., Saadi, I., and Lachke, S. A. (2018b). RNA sequencing-based transcriptomic profiles of embryonic lens development for cataract gene discovery. *Hum. Genet.* 137, 941–954. doi: 10.1007/s00439-018-1958-0

Berry, V., Mackay, D., Khaliq, S., Francis, P. J., Hameed, A., Anwar, K., et al. (1999). Connexin 50 mutation in a family with congenital "zonular nuclear" pulverulent cataract of Pakistani origin. *Hum. Genet.* 105, 168–170. doi: 10.1007/s004399900094

Braun, M., Struck, A. K., Reinartz, S., Heppelmann, M., Rehage, J., Eule, J. C., et al. (2019). Study of congenital *Morgagnian cataracts* in Holstein calves. *PLoS ONE* 14:e0226823. doi: 10.1371/journal.pone.0226823

Chen, C. (2006). CiteSpace II: detecting and visualizing emerging trends and transient patterns in scientific literature. *J. Am. Soc. Inf. Sci. Technol.* 57, 359–377. doi: 10.1002/asi.20317

Deng, H., and Yuan, L. (2014). Molecular genetics of congenital nuclear cataract. *Eur. J. Med. Genet.* 57, 113–122. doi: 10.1016/j.ejmg.2013.12.006

Devi, R. R., Yao, W., Vijayalakshmi, P., Sergeev, Y. V., Sundaresan, P., and Hejtmancik, J. F. (2008). Crystallin gene mutations in Indian families with inherited pediatric cataract. *Mol. Vis.* 14, 1157–1170.

Eyre-Walker, A., and Stoletzki, N. (2013). The assessment of science: the relative merits of post-publication review, the impact factor, and the number of citations. *PLoS Biol.* 11:e1001675. doi: 10.1371/journal.pbio.1001675

Freedman, S. F., Brennand, C., Chiang, J., Debarber, A., Del Monte, M. A., Duell, P. B., et al. (2019). Prevalence of cerebrotendinous xanthomatosis among patients diagnosed with acquired juvenile-onset idiopathic bilateral cataracts. *JAMA Ophthalmol.* 137, 1312–1316. doi: 10.1001/jamaophthalmol.2019.3639

Gillespie, R. L., Lloyd, I. C., and Black, G. C. (2014). The use of autozygosity mapping and next-generation sequencing in understanding anterior segment defects caused by an abnormal development of the lens. *Hum. Hered.* 77, 118–137. doi: 10.1159/000362599

Gu, S., Rong, H., Zhang, G., Kang, L., Yang, M., and Guan, H. (2016). Functional SNP in 3'-UTR MicroRNA-binding site of ZNF350 confers risk for age-related cataract. *Hum. Mutat.* 37, 1223–1230. doi: 10.1002/humu.23073

Hansen, L., Mikkelsen, A., Nurnberg, P., Nurnberg, G., Anjum, I., Eiberg, H., et al. (2009). Comprehensive mutational screening in a cohort of Danish families with hereditary congenital cataract. *Invest. Ophthalmol. Vis. Sci.* 50, 3291–3303. doi: 10.1167/iovs.08-3149

Hejtmancik, J. F. (2008). Congenital cataracts and their molecular genetics. *Sem. Cell Dev. Biol.* 19, 134–149. doi: 10.1016/j.semcdb.2007.10.003

Hu, S., Wang, B., Zhou, Z., Zhou, G., Wang, J., Ma, X., et al. (2010). A novel mutation in GJA8 causing congenital cataract-microcornea syndrome in a Chinese pedigree. *Mol. Vis.* 16, 1585–1592.

Huang, B. Y., and He, W. G. (2010). Molecular characteristics of inherited congenital cataracts. *Eur. J. Med. Genet.* 53, 347–357. doi: 10.1016/j.ejmg.2010.07.001

Huang, C. H., Wang, Y. T., Tsai, C. F., Chen, Y. J., Lee, J. S., and Chiou, S. H. (2011). Phosphoproteomics characterization of novel phosphorylated sites of lens proteins from normal and cataractous human eye lenses. *Mol. Vis.* 17, 186–198.

Jiang, J. X. (2010). Gap junctions or hemichannel-dependent and independent roles of connexins in cataractogenesis and lens development. *Curr. Mol. Med.* 10, 851–863. doi: 10.2174/1566524107939 37750

Lachke, S. A., Alkuraya, F. S., Kneeland, S. C., Ohn, T., Aboukhalil, A., Howell, G. R., et al. (2011). Mutations in the RNA granule component TDRD7 cause cataract and glaucoma. *Science* 331, 1571–1576. doi: 10.1126/science.1195970

Lachke, S. A., Higgins, A. W., Inagaki, M., Saadi, I., Xi, Q., Long, M., et al. (2012). The cell adhesion gene PVRL3 is associated with congenital ocular defects. *Hum. Genet.* 131, 235–250. doi: 10.1007/s00439-011-1064-z

Li, F. F., Yang, M., Ma, X., Zhang, Q., Zhang, M., Wang, S. Z., et al. (2010). Autosomal dominant congenital nuclear cataracts caused by a CRYAA gene mutation. *Curr. Eye Res.* 35, 492–498. doi: 10.3109/02713681003624901

Murgiano, L., Jagannathan, V., Calderoni, V., Joechler, M., Gentile, A., and Drogemuller, C. (2014). Looking the cow in the eye: deletion in the NID1 gene is associated with recessive inherited cataract in Romagnola cattle. *PLoS ONE* 9:e110628. doi: 10.1371/journal.pone.0110628

Osinchuk, S., Petrie, L., Leis, M., Schumann, F., Bauer, B., Sandmeyer, L., et al. (2017). Congenital nuclear cataracts in a Holstein dairy herd. *Can. Vet. J.* 58, 488–492.

Petrou, A. L., and Terzidaki, A. (2017). A meta-analysis and review examining a possible role for oxidative stress and singlet oxygen in diverse diseases. *Biochem. J.* 474, 2713–2731. doi: 10.1042/BCJ20161058

Plaisancie, J., Ragge, N. K., Dollfus, H., Kaplan, J., Lehalle, D., Francannet, C., et al. (2018). FOXE3 mutations: genotype-phenotype correlations. *Clin. Genet.* 93, 837–845. doi: 10.1111/cge.13177

Santana, A., Waiswol, M., Arcieri, E. S., Cabral De Vasconcellos, J. P., and Barbosa De Melo, M. (2009). Mutation analysis of CRYAA, CRYGC, and CRYGD associated with autosomal dominant congenital cataract in Brazilian families. *Mol. Vis.* 15, 793–800.

Shiels, A., Bennett, T. M., and Hejtmancik, J. F. (2010). Cat-Map: putting cataract on the map. *Mol. Vis.* 16, 2007–2015.

Shiels, A., Bennett, T. M., Knopf, H. L., Maraini, G., Li, A., Jiao, X., et al. (2008). The EPHA2 gene is associated with cataracts linked to chromosome 1p. *Mol. Vis.* 14, 2042–2055.

Shiels, A., and Hejtmancik, J. F. (2013). Genetics of human cataract. *Clin. Genet.* 84, 120–127. doi: 10.1111/cge.12182

Shiels, A., and Hejtmancik, J. F. (2017). Mutations and mechanisms in congenital and age-related cataracts. *Exp. Eye Res.* 156, 95–102. doi: 10.1016/j.exer.2016.06.011

Small, H. (1973). Co-citation in the scientific literature: a new measure of the relationship between two documents. *J. Am. Soc. Inf. Sci.* 24, 265–269. doi: 10.1002/asi.4630240406

Sousounis, K., and Tsonis, P. A. (2012). Patterns of gene expression in microarrays and expressed sequence tags from normal and cataractous lenses. *Hum. Genomics* 6:14. doi: 10.1186/1479-7364-6-14

Tibrewal, S., Duell, P. B., Debarber, A. E., and Loh, A. R. (2017). Cerebrotendinous xanthomatosis: early diagnosis on the basis of juvenile cataracts. *J. AAPOS* 21, 505–507. doi: 10.1016/j.jaapos.2017.07.211

Wada, K., Maeda, Y. Y., Watanabe, K., Oshio, T., Ueda, T., Takahashi, G., et al. (2011). A deletion in a cis element of Foxe3causes cataracts and microphthalmia in rct mice. *Mamm. Genome* 22, 693–702. doi: 10.1007/s00335-011-9358-y

Wang, Y., Li, F., Zhang, G., Kang, L., and Guan, H. (2016). Ultraviolet-B induces ERCC6 repression in lens epithelium cells of age-related nuclear cataract through coordinated DNA hypermethylation and histone deacetylation. *Clin. Epigenetics* 8:62. doi: 10.1186/s13148-016-0229-y

Weatherbee, B. A. T., Barton, J. R., Siddam, A. D., Anand, D., and Lachke, S. A. (2019). Molecular characterization of the human lens epithelium-derived cell line SRA01/04. *Exp. Eye Res.* 188:107787. doi: 10.1016/j.exer.2019.107787

Weisschuh, N., Aisenbrey, S., Wissinger, B., and Riess, A. (2012). Identification of a novel CRYBB2 missense mutation causing congenital autosomal dominant cataract. *Mol. Vis.* 18, 174–180.

Wu, M. H., Yu, Y. H., Hao, Q. L., Gong, X. H., and Yao, K. (2017). A novel splice site mutation of CRYBA3/A1 gene associated with congenital cataract in a Chinese family. *Int. J. Ophthalmol.* 10, 1–5. doi: 10.18240/ijo.2017.01.01

Yu, Y., Li, J., Xu, J., Wang, Q., Yu, Y., and Yao, K. (2012). Congenital polymorphic cataract associated with a G to A splice site mutation in the human beta-crystallin gene CRYbetaA3/A1. *Mol. Vis.* 18, 2213–2220.

Zhang, K., Zhu, X., and Lu, Y. (2018). The proteome of cataract markers: focus on crystallins. *Adv. Clin. Chem.* 86, 179–210. doi: 10.1016/bs.acc.2018.05.005

Zhu, Y., Shentu, X., Wang, W., Li, J., Jin, C., and Yao, K. (2010). A Chinese family with progressive childhood cataracts and IVS3+1G>A CRYBA3/A1 mutations. *Mol. Vis.* 16, 2347–2353.

Zoric, L., Elek-Vlajic, S., Jovanovic, M., Kisic, B., Djokic, O., Canadanovic, V., et al. (2008). Oxidative stress intensity in lens and aqueous depending on age-related cataract type and brunescense. *Eur. J. Ophthalmol.* 18, 669–674. doi: 10.1177/112067210801800501

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.2022.853969

Individualized Medicine in Africa: Bringing the Practice Into the Realms of Population Heterogeneity

Ayman A. Hussein[1], *Reem Hamad*[1], *Melanie J. Newport*[2] *and Muntaser E. Ibrahim*[1]*

[1]Unit of Diseases and Diversity, Department of Molecular Biology, Institute of Endemic Diseases, University of Khartoum, Khartoum, Sudan, [2]Department of Global Health and Infection, Brighton and Sussex Medical School, University of Sussex, Brighton, United Kingdom

The declared aim of "personalized", "stratified" or "precision" approaches is to place individual variation, as ascertained through genomic and various other biomarkers, at the heart of Scientific Medicine using it to predict risk of disease or response to therapy and to tailor interventions and target therapies so as to maximize benefit and minimize risk for individual patients and efficiency for the health care system overall. It is often contrasted to current practices for which the scientific base is rooted in concepts of a "universal biology" and a "typical" or "average patient" and in which variation is ignored. Yet both approaches equally overlook the hierarchical nature of human variation and the critical importance of differences between populations. Impact of genetic heterogeneity has to be seen within that context to be meaningful and subsequently useful. In Africa such complexity is compounded by the high effective size of its populations, their diverse histories and the diversity of the environmental terrains they occupy, rendering analysis of gene environment interactions including the establishment of phenotype genotype correlations even more cumbersome. Henceforth "Individualized" methods and approaches can only magnify the shortcomings of universal approaches if adopted without due regard to these complexities. In the current perspective we review examples of potential hurdles that may confront biomedical scientists and analysts in genomic medicine in clinical and public health genomics in Africa citing specific examples from the current SARS-COV2 pandemic and the challenges of establishing reference biobanks and pharmacogenomics reference values.

Keywords: individualized medicine, Africa, population heterogeneity, effective size, genetics

OPEN ACCESS

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

Reviewed by:
Dipali Dhawan, Baylor Genetics, United States

***Correspondence:**
Muntaser E. Ibrahim
mibrahim@iend.org

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

Received: *13 January 2022*
Accepted: *28 March 2022*
Published: *14 April 2022*

Citation:
Hussein AA, Hamad R, Newport MJ and Ibrahim ME (2022) Individualized Medicine in Africa: Bringing the Practice Into the Realms of Population Heterogeneity. Front. Genet. 13:853969. doi: 10.3389/fgene.2022.853969

BACKGROUND

The concept of individualized medicine invokes some fundamental paradigms relating to human categorization, the attribution to the individual of membership in a group, and to the characterization of variation both within and between groups. The work of the pioneering biometricians, such as Francis Galton, who studied variation between individuals, was often inextricably tied to the beliefs and practices of eugenics which had the expressed aim of "improvement of stock", "cultivation of race" or some similar formulation of the goal of improving the "genetic quality" of a population. In the period leading up to the Second World War, the misguided adoption of such aims and practices by states in relation to their citizens marred not only eugenics but also to varying degrees its founders and their scientific contributions. In a

similar vein, the study of variation between populations and the linked attempts at categorization had descended through scientifically aberrant theories of race to morally aberrant practices of racism and eventually genocide. In contrast, during roughly the same period, the extensive global usage of antibiotics, based in the universal scientific concepts of biology, gave us a lifesaving tool reinforcing the unity of humankind and altering forever our relationship with microbes. The triumphs of antimicrobial drugs, consolidated thereafter by the eradication of smallpox and malaria in the tropics, and the abuse of the concepts of race and eugenics both promoted a holistic humanistic approach to health.

However, the mainstay of the practice of medicine, in the sense of the conventional relationship between the physician and patient, remained individualized. Ironically, individualized and universal approaches have a common thread in that both overlook the stark realities of population heterogeneity. Individuals are genetically defined not through an abstract human category, but through bonds to defined social units like families and populations formed through, among other things, mating and admixture. These units are defined biologically by genetic structure and allele frequency metrics. One of the earliest demonstrations of the relevance of such units to identification and mapping of diseases susceptibility loci is Paul Mckeigue's work on admixture mapping (Mckeigue, 1998), a concept based on the tenet that population with differential disease phenotype frequencies and recent admixture can help map disease susceptibility loci that are eventually and by nature individualized. Another example is the utility in association studies of alleles imputed from ancestry-adjusted population reference panels, which has been shown to be sensitive to population stratification (Jallow et al., 2009; Huang et al., 2011). Whether stratified, individualized or precise, often used alternately or synonymously, these terms suffer the inherent limitations of definition, for each may describe in different ways the kinds of future medicine and healthcare that these new methods promise to bring forth (Erikainen and Chan, 2018).

It is not only the concepts that await the development of a working paradigm, the practice itself is searching to define its functional domains, including the development of robust and efficient systems for estimation of personalized intervention recommendations as well as the generation of new knowledge about a particular decision process under study (Hicks et al., 2016; Kosorok et al., 2021).

In this perspective we discuss how these paradigms and concepts may translate within the peculiar African setting to describe situations of relevance to its future practice both semantically and functionally.

AFRICA'S BURDEN OF EXTREMES: ENVIRONMENT INTERACTIONS AND SELECTION

With sequencing of the human genome and the eruption of massive genomic data and knowledge, biological understanding of inheritance progressed from simplistic single gene models into dictums of complex inheritance. The spectrum of interaction from a systems point of view is pivotal. There are conditions where a limited number of variants decide the ultimate phenotype vis-a-vis others that involve multiple variants. It is not only the complexity in the spectrum of gene/gene and gene/protein interactions in a two-dimensional scale as in **Figure 1** below; matters may become more complex and multidimensional when bearing in mind that our understanding of genome regulatory mechanisms and the basis of complex inheritance is still in its infancy.

On average, and excluding environmental evolutionary pressure in favor of heterosis, the normal distribution of traits holds. That said, in most situations, population variation remains under the influence of genes with modest or low effects in relation to environmental determinants which constitute the core and essence of complex genetic inheritance. Unfortunately, the prevailing mind-set in the realm of inheritance is largely molded onto single genes with large effects and has little consideration of the compound world of complex and polygenic inheritance i.e., real-life genetics. Ramification of such scope on individualized medicine is consequentially significant. Diagnosticians and therapists should be informed when seeking successful intervention in ailments underpinned by gene-environmental interaction, which form the majority of diseases particularly in our modern lifestyle, characterized by vast and novel psychological and environmental onslaughts on our body and mind. Grappling with such situations is a major undertaking, and genomic science might not be prepared to handle, computationally and conceptually, complex situations of this sort. In Africa, given the large effective population size, its vast and diverse environmental terrains these challenges are even far greater. Ibrahim and Bekele (2019); A Framework for the Implementation of Genomic Medicine for Public Health in Africa: A report by the African Academy of Sciences, 2021).

THE CAVEAT OF AFRICAN EFFECTIVE POPULATION SIZE

Effective Population size (Ne) is a mathematical term denoting the amount of variation in a group, which could be a population, a group of people or even the entire human species. The estimated Ne of Homo sapiens, which describes the number of heterozygous contributing unique genomes at the naissance of our species, is between 1,000 and 10,000 genomes (Elhassan et al., 2014). The majority of this variation is retained in Africa as demonstrated in the spectacular diversity among and within African populations. In fitness terms this may constitute a huge advantage in terms of adaptedness and adaptation, and may have contributed in a historical sense to the survival of our species. However, when it comes to developing therapies and investigations, whether individualized or universal, such heterogeneity can become a liability. In economic and practical terms, the rules of aggregate demand and economies of scale favor large homogeneous populations.

The practice of individualized genomics, though fashionable and tempting, and perhaps within reach in the

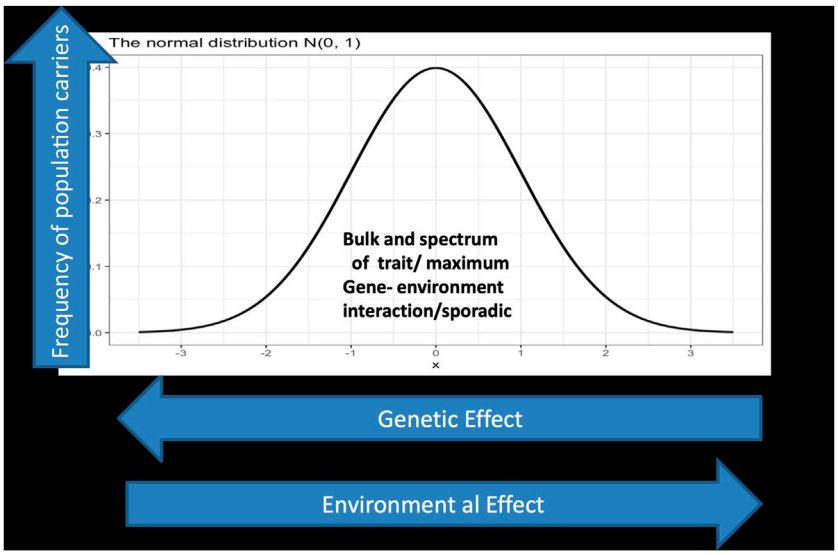


FIGURE 1 | displays a spectrum of gene environment interaction in a two dimensional landscape. The *Y* axis represents the phenotypic outcome of interactions which is normally distributed, while the *x* axis depicts the genetic/environmental effects which is a continuum starting on the left side with large genetic effect as in the case of monogenic and Mendelian inheritance and ends at the right side with low genetic effect but larger environmental effect, in the sense that genetic architecture matters less. Individuals carrying deleterious mutations in their genome with large effects fall under this category which represents in normal situations a tiny fraction of the population. For example a gene that exert strong effect in a Mendelian fashion between 0 and 10 in the *X* axis but has much lower effect in the scale between 90 and 100 in the *Y* axis. In the far right one assumes all the lucky ones who indulge in reckless lifestyle but ends up exceeding the average life span well into their eighties and nineties.

foreseeable future, may not become commonly applicable in Africa for years to come. Reasons may obviously pertain to Africa's under-resourced health systems but also to several other innate obstacles, including Africa's known lower linkage disequilibrium and high genetic diversity (Reich et al., 2001). A major concern questions the argument of a putative ancestral genome that may be reconstructed from existing genomes. If such a premise is flawed progress in translational genomics may not be transferrable in a globalized manner. Two arguments were mentioned along this line: firstly, human populations in and out of Africa were influenced and shaped by drift and serial founder effects (Chiaroni et al., 2009), and secondly, functional variants are multiple, formed through an extended evolutionary history in Africa and retained within its extant populations (Ibrahim and Musa, 2015). It is essentially the same reservation that surrounds the use of genome wide association studies (GWAS) and SNPChips in Africa. There are sobering examples to guard against similar undertakings. Malaria is the most common infectious diseases in Africa, and its phenotypes both as infection or diseases could be readily ascertained. In a GWAS markers close to the sickle cell allele, a locus with a markedly strong effect in protecting against malaria, failed to appear in the initial analysis (minimum $p = 3.9 \times 10^{-7}$). However, when the sickle allele (rs334) was directly genotyped an impressive signal of $p = 1.3 \times 10^{-28}$ was recovered. A reference panel of Gambian control DNAs were sequenced across this HHB region and imputation using data derived from this gave a p value of $p = 4.5 \times 10^{-14}$ for the sickle allele (Jallow et al., 2009).

Even if technologies availed a panel of genetic markers with persuasive power for detecting disease association at a reasonable effect size, such markers cannot guarantee satisfactory outcomes in prediction/intervention without accounting for potential epistasis and network effects (Maron et al., 2021), as well as the excess of minor/rare/private pathogenic alleles in Africa (Koko et al., 2018).

POPULATION STRUCTURE IN PUBLIC HEALTH GENOMICS, PHARMACOGENOMICS AND BIOBANKING

Africa's diverse and complex population structure is a vivid reminder to the importance of human genome variation to genomic medicine. It also exposes the feebleness of universal approaches and shows the extent to which management can often fall into uninformed and guess-like practice. As the road towards individualized medicine is still long in Africa, population health maps may be a valuable aid to physicians and diagnosticians. These are now cropping up following genomics mega projects like H3Africa (H3Africa Consortium et al., 2014) and countries/ population maps like the recent copy number variation in Tunisian populations (Romdhane et al., 2021).

In many countries the distribution of monogenic diseases is related to ethnic origin or population subgroup and such knowledge is of great assistance. This may similarly apply to polygenic diseases, infectious or non-infectious alike. For monogenic non-infectious diseases, ethnic and geographic clusters are reported at the global, regional and country levels.

Similarly, infectious diseases, including the current coronavirus pandemic, can have clear ethnic and geographic distributions (Ibrahim and Salih, 2021). Such clusters may be interrogated for allele frequency patterns and ancestry markers that may foster identification of functional loci and suggest treatments.

Genomic data on such clusters and foci combined with well-documented clinical phenotypic data is expected to improve our understanding of concepts of risk and their measures whether at the level of population, family or individual, and provide clues to why some individuals remain healthy while others are more susceptible to disease. Such knowledge may be used to make predictions about the likelihood that a person will contract a disease and his/her expected clinical phenotype (Molster et al., 2018). It could also be applied to develop new tools for risk prediction or predictive testing in relation to the onset or recurrence of disease, in the context of public health genomics (PHG). PHG is a domain that is most suited to relate populations to both individualized and "universalized" approaches.

Genomic knowledge can offer new ways of stratifying individuals and sub-groups within populations, taking public health beyond the traditional correlates of disease risk factors such as gender, age, and socio-economic status (Boccia et al., 2014). PHG promises wide prospects in informing preventive health measures for all. The P4 (Preventive Proactive, Personalized and Predictive) project, is one such exciting prospect with immense possibilities for harnessing the genomics revolution to achieve the example of health for all (Flores et al., 2013). Pharmacogenomics and pharmacogenetics are largely seen as the domains where PHG could readily inform individualized approaches to be put into practice. Already substantial progress has been made in antihypertensive drugs, by targeted therapy based on renin/aldosterone phenotyping (Akintunde et al., 2017). Among other success stories where the logic was made emphatically on the necessity that African populations begin capturing their own pharmacogenetic SNPs for precision medicine, is the Warfarin case (Ndadza et al., 2019; Ndadza et al., 2021; Muyambo et al., 2022) with several private and population specific polymorphisms identified.

However, apart from these gains, the promise of therapies tailored to individual patients, which is the ultimate aspiration of individualized and precision medicine, again faces the obstacles associated with Africa's large effective population size, and the potential excess of rare pathogenic variants. It is not only a challenge to patients and doctors; it may also diminish the interest from industry. Potential gaps between discovery and translation and industry are not easy to bridge and companies will be less eager if the market is limited. Alternatively, and until cheaper approaches for translating genomics data into individualized interventions for individual patients is within reach, health systems can make substantial gains by employing population stratification and identification of less rare pathogenic variants that could be typed in patients. Although this is short of ideal where all possible biological interactions are considered, it is still superior to a non-informed administration of drugs in a population. Baker et al. (2017), rightly propose a more systematic characterization of the genetic diversity of African ancestry populations, which might help customize the intervention thus reducing cost and effort.

Biobanks are platforms to obtain, store, manage and distribute large collections of biological samples of defined phenotypes and of marked interest for researchers in human, animal or plant specimens. Given Africa's reported scale of genomic variation, a major question arises on which representative sample that should be analyzed to capture the true amount of genomic variation present in a country, population or community. Hence, in order to form a perspective of alleles frequency spectrum; the scientific community should invest not only on clinical sequencing efforts, which is becoming the norm for biomedical enterprise, but on population data as well. Such data may be obtained from 1) control samples in association studies, 2) active population surveys and 3) biological samples including tissues from tertiary facilities. Biobanks are by virtue a combination of all the above with the added value of including time series data and data from rare diseases that are seldom encountered in cross sectional surveys and sampling.

INDIVIDUALIZED MEDICINE IN LIGHT OF THE PANDEMIC

The current pandemic caused by SARS-CoV-2, came as an anticlimax to public perception of human health following decades of successful eradication of infectious diseases. It embodied the gravity of challenges facing the human species in a globalized and environmentally frail world. Unlike the HIV/AIDS where parts of the continent bore the brunt of its burden, in the current SARS-CoV-2 pandemic a different hitherto unexplained situation prevails, characterized by low disease burden and some of the lowest mortality indices (see Corona world meter). Conventional attributes and explanations are lack of transparency, ailing health systems, young populations (Sofonias and Nkengasong, 2021) and again an oversight of the critical elements of genetic susceptibility and population structure (Pairo-Castineira et al., 2021).

Disparities in the population infection and mortality rates in the COVID-19 pandemic (Ibrahim and Salih, 2021), underscores the need for adaptive strategies where personalized and communal intervention becomes directly proportional to the degree of complexity of its etiological basis including the effect of the environment. Environmental triggers are themselves subject to compatibility with molecular targets which are in turn under influence of allele frequency changes which vary between individuals and population to create a situation of unusual patterns of transmission and epidemicity, where even herd immunity, a central tenet in the control of pandemic, becomes contested (Buss et al., 2021). Vaccination as a measure of quelling pandemics falters between astounding successes like in smallpox eradication, measles and polio vaccines and the modest outcomes in the influenza or dengue virus. Understanding the dynamics of COVID-19 immune responses is not critical only for mass vaccination and attainment of herd immunity but also at the individual

level for those who are likely to develop serious adverse side effects from the vaccine. Most of the poor compliance reported in some of the COVID-19 vaccines is related to media coverage of rare serious side effects. Africa's poor vaccination coverage, however, may be attributable not only to anti-vaccination sentiments and vaccine availability/affordability but possibly also to poor compliance emanating from a communal feeling of lack of threat. Yazdan and Robineau argue In all cases, prediction of such adverse effects and sequel of infection requires a move from "population-based policies" to "individual-centred" measures, at both a public health and patient's management given the heterogeneity of the clinical presentations of Sars-cov2. https://www.icpermed.eu/en/covid-19_personalised-medicine.php. According to Muyambo et al. (2022), given the "current historical moment of COVID-19 pandemic" with systemic and cardiovascular effects of the virus such move might prove mandatory considering the several private and population specific polymorphisms identified in genes potentially influencing interactions with drugs like warfarin.

POPULATION DIFFERENCES AND STRATIFICATION FROM THE GROUP TO THE PERSON

Although the term population is an appealing one, it can often be quite misleading in actual situations given how complex human history is. We have seen above that ethnic allele frequency maps facilitated the identification of diseases susceptibility loci, in conjunction with imputation (Jallow et al., 2009; Huang et al., 2011).

The challenge of population heterogeneity and hence potential disease propensities has been highlighted in the comprehensive report on genomic medicine by the African Academy of Sciences. That report adopted a widely scientifically acclaimed classification of human populations according to linguistic groups (Chiaroni et al., 2009), with Africa host to four major linguistic families. A panel of 1,327 microsatellite markers further identified 14 ancestral population clusters in Africa that also correlated with self-described ethnicity and shared cultural and/or linguistic properties (Tishkoff et al., 2009).

Such clusters should become the prospective aim and focus of Africa's genomic endeavors as we zoom in to fill the gap between the populations and individual, the latter being the ultimate goal of precision and individualized medicine.

A vivid example of the pivotally of both obvious and cryptic population structures is visceral leishmaniasis (VL). Ethnic differences were intuitively observed in visceral leishmaniasis (VL) and during the quest to understand the genetic basis of susceptibility to VL, stratification of a target ethnically homogenous population was carried out based on Y chromosome haplogroups. Interestingly, this approach increased the likelihood of linkage to an impressive LOD score of 5.656 defining a susceptibility locus associated with carriers of haplogroup A1b1b2b, indicating the presence of hidden population structure and village-specific genetic lineage possibly due to a founder effect (Miller et al., 2007).

The identification, for purposes of transitioning to personalized risk assessment and intervention, of individualized markers that underlie group and population-based risk is hence not a straight forward path in either stratum, and a marker related to disease risk in one population is not necessarily universal either. Two examples of differences between populations illustrate this point. Firstly, Prdx5 a molecule shown to pose considerable risk for various cancers in Asians, displayed different mode of gene expression and reduced risk in Africans (Sudanese) stressing the need for caution in applying findings about risk association from one population to another (Elamin et al., 2013). The other example concerns apparent differences in the biological etiology of breast cancer between Eritrea and neighboring Sudan as seen in the presence of a marked viral etiology in Sudan and a minor contribution in Eritrea, a contrast that may be related to environmental or genetic causes (Fessahaye et al., 2017).

In Sudan, The field of cancer is rife with similar examples where cancers have shown to present with infrequent mutations in TP53 the notorious tumor suppressor (Masri et al., 2002). However, in many cases, genome and epigenome changes in cancers presented with mutations in other p53 family members like P63 and P73 (Suleiman et al., 2015; Abdallah et al., 2018). BRCA1, BRCA2 and p53 mutations are infrequent in Sudanese breast cancer patients. Epigenetic changes are suggested as alternative mechanisms to account for the minor contribution of genetic alterations in these genes in sporadic and familial breast cancer cases in Sudan, and evidence for a strong association with EBV was reported (Yahia et al., 2014). These findings were subsequently corroborated by whole methylome analysis which pointed out the dysregulation of specific developmental and viral pathways (Abdallah et al., 2018). In the progression of hereditary colorectal cancer, emphasis is usually placed on the pathogenic role of inactivating mutations in tumor-suppressor genes; however, a study of a multi-case family revealed a different pathway of similar complexity with the identification of a pivotal role for the RNA binding protein ELAV1 and other partners (Suleiman et al., 2015). The presence or absence of a variant is not necessarily a direct reflection of the absence or presence of a trait. An interesting example is lactase persistence. In Europe a single mutation associated with lactase persistence dominates, whereas at least 4 mutations associated with the persistence of the enzyme, mainly in milk-consuming pastoralists communities, were identified in Africa (Ranciaro et al., 2014). However, more than one population that has a history of dairy consumption has no obvious mutations in lactase gene to explain the trait. Apparently milk digestion and lactose tolerance may not be solely an outcome of genetic adaptation. Speculation points to lactobacilli possibly underlying the trait. If this is the case either directly or epigenetically, another layer of variation has to be taken in to account which is the role of the microbiome.

Beyond conventional paradigms of gene-gene interaction, where many gray areas still exists, defining gene-environment interaction is also exceedingly challenging in Africa for the

same reasons: These complex interactions between environmental determinants (viruses, food etc.), genomic sequences and epigenetic machinery is what renders the pursuits of individualized approaches in Africa a daunting task but all more interesting. Amid the mist of these disparate and multitude of contributors one should possess a methodological compass to point the right direction where the phenotypic outcome are plausibly explained by genetic variants shown in population differences, while environmental and epigenetic determinants need to be worked out in a frame work of system science. (Kosorok et al., 2021; Maron et al., 2021). As the aim of precision medicine is to design and improve diagnosis, therapeutics and prognostication through the use of large complex datasets that incorporate individual gene, functional, and environmental variations. The implementation of high-performance computing (HPC) and artificial intelligence (AI) can predict risks with greater accuracy based on available multidimensional clinical and biological datasets(Hicks et al., 2016; Subramanian et al., 2020; Kosorok et al., 2021).

In conclusion the road towards an effective, equitable and cost-efficient personalized health practices in Africa still has a long way to go. Major obstacles include the complex phenotypic outcomes of gene-environment interactions, the difficulty of applying customized approaches in an environment of remarkable genetic diversity, the low translational value of genetic and genomic research, prohibitive cost of technologies and applications, and availability of these applications to the bulk of Africans and finally the disinterest of state and business. For Africa to enter the epoch of individualized medicine will not be possible without addressing these challenges and carrying out fundamental reform to integrate genomics into the educational and health system.

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

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

ACKNOWLEDGMENTS

Gratitude goes to Safaa AbuShora and Lamees Ammar for assistance with the graphics.

REFERENCES

Abdallah, M. O. E., Algizouli, U. K., Suliman, M. A., Abdulrahman, R. A., Koko, M., Fessahaye, G., et al. (2018). EBV Associated Breast Cancer Whole Methylome Analysis Reveals Viral and Developmental Enriched Pathways. *Front. Oncol.* 8 (8). doi:10.3389/fonc.2018.00316

Akintunde, A., Nondi, J., Gogo, K., Jones, E. S. W., Rayner, B. L., Hackam, D. G., et al. (2017). Physiological Phenotyping for Personalized Therapy of Uncontrolled Hypertension in Africa. *Am. J. Hypertens.* 30 (9), 923–930. doi:10.1093/ajh/hpx066

Baker, J. L., Shriner, D., Bentley, A. R., and Rotimi, C. N. (2017). Pharmacogenomic Implications of the Evolutionary History of Infectious Diseases in Africa. *Pharmacogenomics J.* 17 (2), 112–120. doi:10.1038/tpj.2016.78

Boccia, S., Mc Kee, M., Adany, R., Boffetta, P., Burton, H., Cambon-Thomsen, A., et al. (2014). Beyond Public Health Genomics: Proposals from an International Working Group. *Eur. J. Public Health* 24 (6), 877–879. doi:10.1093/eurpub/cku142

Buss, L. F., Prete, C. A., Abrahim, C. M. M., Mendrone, A., Jr, Salomon, T., de Almeida-Neto, C., et al. (2021). Three-quarters Attack Rate of SARS-CoV-2 in the Brazilian Amazon during a Largely Unmitigated Epidemic. *Science* 371, 288–292. doi:10.1126/science.abe9728

Chiaroni, J., Underhill, P. A., and Cavalli-Sforza, L. L. (2009). Y Chromosome Diversity, Human Expansion, Drift, and Cultural Evolution. *Proc. Natl. Acad. Sci. U S A.* 106 (48), 20174–20179. doi:10.1073/pnas.0910803106

Elamin, A., Zhu, H., Hassan, A. M. E., Xu, N., and Ibrahim, M. E. (2013). Peroxiredoxin V: A Candidate Breast Tumor Marker of Population Specificity. *Mol. Clin. Oncol.* 1 (3), 541–549. doi:10.3892/mco.2013.91

Elhassan, N., Gebremeskel, E. I., Elnour, M. A., Isabirye, D., Okello, J., Hussien, A., et al. (2014). The Episode of Genetic Drift Defining the Migration of Humans Out of Africa Is Derived from a Large East African Population Size. *PLoS ONE* 9 (5), e97674. doi:10.1371/journal.pone.0097674

Erikainen, S., and Chan, S. (2018). Contested Futures: Envisioning "Personalized," "Stratified," and "Precision" Medicine. *New Genet. Soc. Crit. Stud. Contemp. Biosciences* 38, 308–330.

Fessahaye, G., Elhassan, A. M., Elamin, E. M., Adam, A. A. M., Ghebremedhin, A., and Ibrahim, M. E. (2017). Association of Epstein - Barr Virus and Breast Cancer in Eritrea. *Infect. Agents Cancer* 12, 62. doi:10.1186/s13027-017-0173-2

Flores, M., Glusman, G., Brogaard, K., Price, N. D., and Hood, L. (2013). P4 Medicine: How Systems Medicine Will Transform the Healthcare Sector and Society. *Personalized Med.* 10 (6), 565–576. doi:10.2217/pme.13.57

Hicks, J. K., Dunnenberger, H. M., Gumpper, K. F., Haidar, C. E., and Hoffman, J. M. (2016). Integrating Pharmacogenomics into Electronic Health Records with Clinical Decision Support. *Am. J. Health Syst. Pharm.* 73 (23), 1967–1976. doi:10.2146/ajhp160030

Huang, L., Jakobsson, M., Pemberton, T. J., Ibrahim, M., Nyambo, T., Omar, S., et al. (2011). Haplotype Variation and Genotype Imputation in African Populations. *Genet. Epidemiol.* 35 (8), 766–780. doi:10.1002/gepi.20626

Ibrahim, M. E., and Bekele, E. (2019). "Disease, Selection, and Evolution in the African Landscape," in *(edits) Ibrahim M and Rotimi C* (Cambridge University Press), 50–70. doi:10.1017/9781139680295.003

Ibrahim, M., and Musa, M. K. (2015). Effective Size and Effectiveness: Next Generation Sequencing and the Practice of Genomics in Africa. *J. Next Generation Seq. Appl.* 01 (S1), 2469–9853. doi:10.4172/2469-9853.s1-008

Ibrahim, M., and Salih, A. (2021). The Y Chromosome Ancestry Marker R1b1b2: a Surrogate of the SARS-CoV-2 Population Affinity. *Hum. Genome* 8 (1), 11. doi:10.1038/s41439-021-00141-1

Jallow, M., Teo, Y. Y., Teo, Y. Y., Small, K. S., Rockett, K. A., Deloukas, P., et al. (2009). Genome-wide and fine-resolution Association Analysis of Malaria in West Africa. *Nat. Genet.* 41 (6), 657–665. doi:10.1038/ng.388

Koko, M., Abdallah, M. O. E., Amin, M., and Ibrahim, M. (2018). Challenges Imposed by Minor Reference Alleles on the Identification and Reporting of Clinical Variants from Exome Data. *BMC Genomics* 19 (1), 46. doi:10.1186/s12864-018-4433-3

Kosorok, M. R., Laber, E. B., Small, D. S., and Zeng, D. (2021). Introduction to the Theory and Methods Special Issue on Precision Medicine and Individualized Policy Discovery. *J. Am. Stat. Assoc.* 116 (533), 159–161. doi:10.1080/01621459.2020.1863224

Maron, B. A., Wang, R.-S., Shevtsov, S., Drakos, S. G., Arons, E., Wever-Pinzon, O., et al. (2021). Individualized Interactomes for Network-Based Precision Medicine in Hypertrophic Cardiomyopathy with Implications for Other Clinical Pathophenotypes. *Nat. Commun.* 12, 873. doi:10.1038/s41467-021-21146-y

Masri, M. A., Abdel Seed, N. M., Fahal, A. H., Romano, M., Baralle, F., El Hassan, A. M., et al. (2002). Minor Role for BRCA2 (Exon11) and P53 (Exon 5-9) Among Sudanese Breast Cancer Patients. *Breast Cancer Res. Treat.* 71 (2), 145–147. doi:10.1023/a:1013807830329

Mckeigue, P. M. (1998). Mapping Genes that Underlie Ethnic Differences in Disease Risk: Methods for Detecting Linkage in Admixed Populations, by Conditioning on Parental Admixture. *Am. J. Hum. Genet.* 63, 241–251. doi:10.1086/301908

Miller, E. N., Fadl, M., Mohamed, H. S., Elzein, A., Jamieson, S. E., Cordell, H. J., et al. (2007). Y Chromosome Lineage- and Village-specific Genes on Chromosomes 1p22 and 6q27 Control Visceral Leishmaniasis in Sudan. *Plos Genet.* 3 (5), e71. doi:10.1371/journal.pgen.0030071

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 FuturePersonalized Medicine for Patients with MDR-TB. *Front. Public Healthjournal Antimicrob. Chemother.* 671 (4), 247852–247855. doi:10.3389/fpubh.2018.00247

Muyambo, S., Ndadza, A., Soko, N. D., Kruger, B., Kadzirange, G., Chimusa, E., et al. (2022). Warfarin Pharmacogenomics for Precision Medicine in Real-Life Clinical Practice in Southern Africa: Harnessing 73 Variants in 29 Pharmacogenes. *OMICS: A J. Integr. Biol.* 26 (1), 35–50. doi:10.1089/omi.2021.0199

Ndadza, A., Cindi, Z., Makambwa, E., Chimusa, E., Wonkam, A., Kengne, A. P., et al. (2019). Warfarin Dose and CYP2C Gene Cluster: An African Ancestral-specific Variant Is a Strong Predictor of Dose in Black South African Patients. *OMICS: A J. Integr. Biol.* 23 (1), 36–44. doi:10.1089/omi.2018.0174

Ndadza, A., Muyambo, S., Mntla, P., Wonkam, A., Chimusa, E., Kengne, A. P., et al. (2021). Profiling of Warfarin Pharmacokinetics-associated Genetic Variants: Black Africans Portray Unique Genetic Markers Important for an African Specific Warfarin Pharmacogenetics-dosing Algorithm. *J. Thromb. Haemost.* 19 (12), 2957–2973. doi:10.1111/jth.15494

Pairo-Castineira, E., Clohisey, S., Clohisey, S., Klaric, L., Bretherick, A. D., Rawlik, K., et al. (2021). Genetic Mechanisms of Critical Illness in COVID-19. *Nature* 591 (7848), 92–98. doi:10.1038/s41586-020-03065-y

Ranciaro, A., Campbell, M. C., Hirbo, J. B., Ko, W.-Y., Froment, A., Anagnostou, P., et al. (2014). Genetic Origins of Lactase Persistence and the Spread of Pastoralism in Africa. *Am. J. Hum. Genet.* 94 (4), 496–510. doi:10.1016/j.ajhg.2014.02.009

Reich, D. E., Cargill, M., Bolk, S., Ireland, J., Sabeti, P. C., Richter, D. J., et al. (2001). Linkage Disequilibrium in the Human Genome. *Nature* 411, 199–204. doi:10.1038/35075590

Romdhane, L., Mezzi, N., Dallali, H., Messaoud, O., Shan, J., Fakhro, K. A., et al. (2021). A Map of Copy Number Variations in the Tunisian Population: a Valuable Tool for Medical Genomics in North Africa. *Npj Genom. Med.* 6 (1), 3. doi:10.1038/s41525-020-00166-5

H3Africa ConsortiumRotimi, C., Rotimi, C., Abayomi, A., Abimiku, A., Adabayeri, V. M., Adebamowo, C., et al. (2014). Research Capacity. Enabling the Genomic Revolution in Africa. *Science* 344 (6190), 1346–1348. doi:10.1126/science.1251546

Sofonias, T. K., and Nkengasong, J. N. (2021). Understanding COVID-19 in Africa. *Nat. Rev Immunol* 21, 469–470.

Subramanian, M., Wojtusciszyn, A., Favre, L., Boughorbel, S., Shan, J., Letaief, K. B., et al. (2020). Precision Medicine in the Era of Artificial Intelligence: Implications in Chronic Disease Management. *J. Transl Med.* 18, 472. doi:10.1186/s12967-020-02658-5

Suleiman, S. H., Koko, M. E., Nasir, W. H., Elfateh, O., Elgizouli, U. K., Abdallah, M. O. E., et al. (2015). Exome Sequencing of a Colorectal Cancer Family Reveals Shared Mutation Pattern and Predisposition Circuitry along Tumor Pathways. *Front. Genet.* 6 (6), 288. doi:10.3389/fgene.2015.00288

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

Yahia, Z. A., Adam, A. A., Elgizouli, M., Hussein, A., Masri, M. A., Kamal, M., et al. (2014). Epstein Barr Virus: A Prime Candidate of Breast Cancer Aetiology in Sudanese Patients. *Infect. Agents Cancer* 9 (1), 9. doi:10.1186/1750-9378-9-9

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 Medicine

METHODS
published: 23 May 2022
doi: 10.3389/fmed.2022.852529

Development of a Simple Risk Model to Predict Mortality in Patients With Osteosarcoma of the Extremity

Yu Chen[1,2], *Chao Li*[3], *Xin Wang*[3], *Zhiyong Liu*[3] *and Zhigang Ren*[1,4]*

[1] Gene Hospital of Henan Province, Precision Medicine Center, The First Affiliated Hospital of Zhengzhou University, Zhengzhou, China, [2] Academy of Medical Science, Zhengzhou University, Zhengzhou, China, [3] Department of Orthopaedic Surgery, Affiliated Cancer Hospital of Zhengzhou University, Zhengzhou, China, [4] Department of Infectious Diseases, The First Affiliated Hospital of Zhengzhou University, Zhengzhou, China

Background: Osteosarcoma (OS) is the most prevalent primary malignant bone cancer with poor prognosis. The aim of this study was to explore the prognostic factors that influence survival, and build up and validate a simple risk model to predict mortality in OS patients.

Materials and Methods: This was a single-center retrospective cohort study. A total of 153 patients with newly diagnosed OS were enrolled as the training group. We analyzed the clinical data and outcomes of the OS patients. Prognostic risk factors were identified and evaluated by a logistic regression model with Markov Chain Monte Carlo simulation. The risk score was constructed based on the training group and was further validated using each patient.

Results: Among the 153 patients, the mean (standard deviation) age was 21.6 (14.2) years, and 62 (40.5%) patients were females. The rate of in-hospital mortality of patients was 41.2% (95% CI, 31.6–50.7%). The candidate prognostic factors were selected and evaluated in relation to patient age, sex, tumor site (lower/upper extremity), tumor volume, intramedullary length of lesion, serum levels of alkaline phosphatase (ALP) and primary metastasis. However, only tumor size and primary metastasis were identified as independent prognostic indicators for patients with osteosarcoma. The risk model had a C-statistic of 0.7308 with a predictive range of 21.05–68.42%. Based on the distribution of the risk score, 24.8, 49.7 and 25.5% of patients were stratified into the high-, average- and low-risk groups for in-hospital mortality, with corresponding probabilities of 0.684, 0.329, and 0.210, respectively.

Conclusion: A simple risk model was developed and validated to predict the prognosis for patients with osteosarcoma of the extremity at primary diagnosis. The simple risk score system could be used to stratify patients into different risk groups of in-hospital mortality and may help clinicians judge the outcomes of prognosis and establish appropriate surveillance strategies.

Keywords: osteosarcoma, clinical data, outcome, prognostic factors, risk model

OPEN ACCESS

Edited by:
Muntaser Eltayeb Ibrahim, University of Khartoum, Sudan

Reviewed by:
Chao Zhao, Fudan University, China
Fei Xiao, Beijing Hospital, China

***Correspondence:**
Zhigang Ren
fccrenzg@zzu.edu.cn

Specialty section:
This article was submitted to Precision Medicine, a section of the journal Frontiers in Medicine

Received: *11 January 2022*
Accepted: *21 April 2022*
Published: *23 May 2022*

Citation:
Chen Y, Li C, Wang X, Liu Z and Ren Z (2022) Development of a Simple Risk Model to Predict Mortality in Patients With Osteosarcoma of the Extremity. Front. Med. 9:852529. doi: 10.3389/fmed.2022.852529

INTRODUCTION

Osteosarcoma (OS) is one of the most common primary malignant bone tumors in children and adolescents, accounting for approximately two-thirds of the bone cancers diagnosed in the second decade of life (1–3). The incidence of OS is approximately one to three cases annually per one million people worldwide (1–3). OS is characterized by the presence of an osteoid matrix or immature bone, mainly frequent sites in the metaphysis and diaphysis of long bones (femur, tibia) (2), and can be broadly classified into three histologic subtypes (intramedullary, surface, and extraskeletal) (4). The pathogenesis and etiology of OS are still unclear. In the past few decades, the 5-year survival of patients with localized OS has considerably improved to 78% with the development of neoadjuvant chemotherapy and surgical techniques. However, the 5-year survival rate drops to 25% in cases with metastasis at diagnosis or relapse (5). Moreover, another study showed that the 5-year survival rate showed no significant improvement in patients with localized disease and no improvement in metastatic patients over the past four decades (6). Thus, improving the survival of OS patients has proved challenging, although the therapy for osteosarcoma is on the precipice of advancement. The historical invariability of survival outcomes and the limited number of predictive risk factors known to be active in the development of this disease facilitate clinical trials designed to identify efficacious prognostic factors in patients with osteosarcoma.

Currently, some clinical studies have been reported regarding identifying the prognostic factors that influence survival in osteosarcoma, including patient age, sex, tumor site and size, histologic subtype, presence and location of metastases, histologic response to chemotherapy and type of surgery and surgical margins (7–14). However, there is no standardized system for evaluating the prognostic factors correlated with survival among these studies. Specifically, there were variations in the statistical methods and the study population. Consequently, the results are somewhat inconsistent and even contradictory in the published importance of some variables (such as patient age, tumor site, and tumor size) (8–10, 13). However, predictive models are important tools to provide estimates of patient outcome (15). Moreover, it is imperative to explore effective prognostic models to predict the mortality of patients with OS.

Herein, we mined the relevant clinical data of patients with osteosarcoma (OS) of the extremities spanning 3 years from 2013 to 2015. We subsequently identified prognostic risk factors related to poor outcomes and then developed and validated a risk model to stratify patients into different risk groups of in-hospital mortality and help clinicians provide patients with appropriate surveillance strategies.

MATERIALS AND METHODS

Patients and Study Design

Medical records of patients with OS of the extremities who were admitted to the Affiliated Cancer Hospital of Zhengzhou University between 1 January 2013 and 31 December 2015 with a minimum follow-up of 5 years were reviewed. The eligible patients for enrolment fulfilled the following criteria: typical radiographic and histologic features of primary, high-grade central osteosarcoma of the extremity. Exclusion criteria included non-extremity locations, low- or intermediate-grade osteosarcoma, treatment regimens that did not follow National Comprehensive Cancer Network (NCCN) guidelines, concomitant with previous history of cancer, and incomplete medical records. In total, we identified 153 patients who satisfied the prespecified study inclusion and exclusion criteria. The original 153 samples were used as the training group, and the bootstrap method was used as an internal test of the performance of the model (**Figure 1**). The ethics committee of the Affiliated Cancer Hospital of Zhengzhou University approved this study (No. 2017407).

Potential Risk Factors and Outcome

The 7 candidate risk factors were easily collected, reliable, clinically important, and potentially associated with the outcome, including patient age, sex, tumor site (lower/upper limbs), tumor volume, intramedullary length of lesion, ALP and primary metastasis. The outcome was in-hospital mortality.

STATISTICAL ANALYSIS

Risk Factor Selection and Evaluation

Using the original sample with all candidate risk factors, we employ Markov Chain Monte Carlo (MCMC) simulation to select the primitive risk factors, which have a positive coefficient in more than 90% or <10% of the simulations and are regarded with a stable association with the outcome. The final risk model to predict the outcome was constructed by fitting a logistic regression model to the original sample using the selected risk factors by the MCMC simulation.

Due to the small sample size, the bootstrap method (16) was used to test model performance, and a total of 1,000 bootstrap samples were drawn with replacement of the same sample size as the original sample. Models were developed in the bootstrap samples and tested in the original sample.

We calculated the four indicators to evaluate the risk model performance. Discrimination was assessed with an internally validated c-statistic, and the distribution of the c-statistic for the bootstrap samples and the original sample (**Figure 2A**) are also presented. The internally validated calibration slope was used to measure calibration, and the observed in-hospital mortality in strata defined by quantiles of the predictive probabilities is presented (**Figure 2B**). We divided patients in the original sample into 3 mutually exclusive risk classes based on the quantiles of the predicted probability of in-hospital mortality, i.e., the lowest risk (class 1, <25% quantile), moderate risk (class 2, between 25 and 75% quantile), and highest risk (class 3, >75% quantile) for evaluation. We also calculated the internally validated Brier score to assess the overall fit of the model

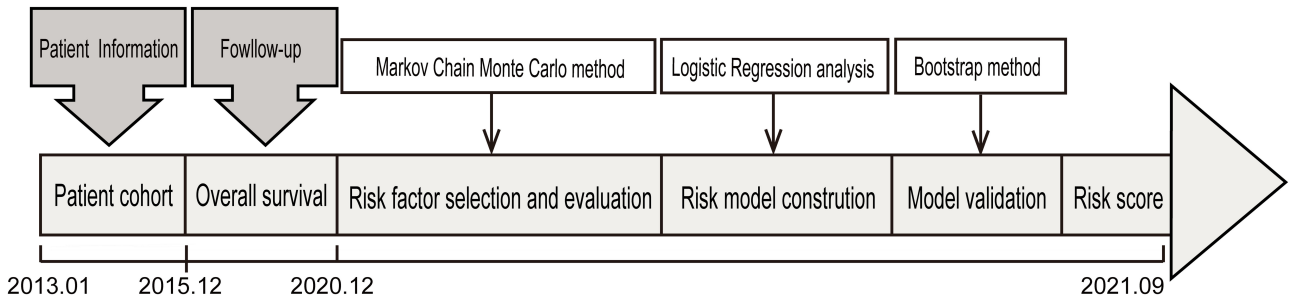


FIGURE 1 | The design of the study. The patient cohort including 153 patients with newly diagnosed OS after a strict pathological diagnosis and exclusion process was used to identify prognostic factors for in-hospital mortality and was used for further evaluation and validation analysis.

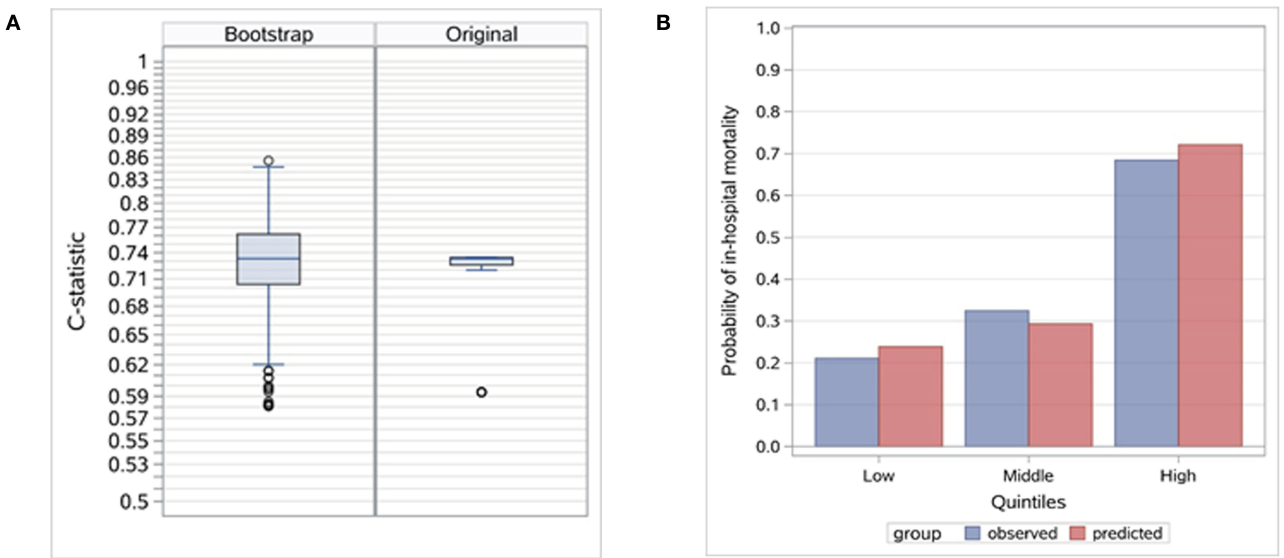


FIGURE 2 | Evaluation of the risk model. **(A)** Distribution of c-statistic in bootstrap and original samples. Internally validated c-statistic is 0.7308. **(B)** Distribution of observed in-hospital mortality by stratified predictive probability.

and internally validated Nagelkerke's R^2 to measure the explained variation.

Risk Score

To facilitate the use of the selected risk factors and the risk model, we constructed a simple risk score for each patient based on the regression coefficients estimated from the risk model with the original sample. Points for each risk factor were calculated by dividing the risk factor's coefficient by summing the absolute value of coefficients in the model, multiplying by 100, and rounding to the nearest integer. We stratified patients into three risk groups based on the distribution of the risk score: low (<25th percentile), average (25th−75th percentile), and high (>75th percentile).

Analysis was conducted using SAS statistical software version 9.4 (SAS Institute Inc.). The study followed the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD) reporting guideline. Each of the 22 items of the TRIPOD statement was addressed.

RESULTS

Patient Characteristics

A total of 153 eligible patients were enrolled. The general characteristics of the 153 patients are shown in the **Table 1**. The mean age at diagnosis was 21.6 years (standard deviation 14.2) and 62 (40.5%) patients were female. Fifty-nine patients died and 94 remained continuously survival during follow-up phase. The rate of in-hospital mortality was 41.2% (95% CI, 31.6–50.7%) for the original sample. According to the outcome of patients with OS, these patients were divided into two groups named the death and survival groups.

Of the extremity tumors, 141 (92.2%) were situated in the lower extremities (femur, fibula, tibia), and 12 (7.8%) were situated in the upper extremities (humerus, radius). The primary tumors involved lower limbs tumors for the death and survival groups were 57 (96.6%) patients and 84 patients (89.4%), respectively. As to tumor lesions, the median length of lesion was 10.1 centimeter [interquartile range (IQR) 7–13.2] in the death

TABLE 1 | General characteristics of the 153 patients.

Characteristics	Aggregate	Death	Survival	*p*-values
Total	153	59	94	
Age, mean age (SD) (years)	21.6 (14.2)	19.7 (12.3)	22.8 (15.2)	0.3002
Median (IQR) (years)	16 (13, 24)	16 (13, 22)	16.5 (13, 27)	
Gender, female, *n* (%)	62 (40.5)	24 (40.7)	38 (40.4)	0.9753
Tumor site, lower limbs, *n* (%)	141 (92.2)	57 (96.6)	84 (89.4)	0.1045
Tumor length of lesion, mean (SD) (cm)	10.1 (4.5)	11 (5.4)	9.5 (3.8)	0.1653
Median (IQR) (cm)	9 (6.7, 12)	10.1 (7, 13.2)	9 (6.7, 12)	
Tumor volume, mean (SD) (cm^3)	545.6 (661.7)	710 (725.1)	442.5 (599.9)	0.0009
Median (IQR) (cm^3)	291.2 (142.1, 680.6)	523.7 (201.1, 923.4)	250.4 (119.2, 448.5)	
Primary metastasis, *n* (%)	34 (22.2)	25 (42.4)	9 (9.6)	<0.0001
Alkaline phosphatase, mean (SD) (U/L)	246 (281.2)	297.6 (321.3)	213.6 (249.1)	0.003
Median (IQR) (U/L)	147 (104, 251)	196 (136, 346)	126 (94, 223)	
In-hospital mortality rate, % (95% CI)		41.2% (95%CI, 31.6–50.7%)		

SD, standard deviation; IQR, interquartile range; cm, centimeter; CI, confidence interval.

group and 9.0 centimeter (IQR 6.7–12) in the survival group. There were no significant differences in age, sex, tumor site, and tumor lesion between the two groups, while tumor volume and serum levels of alkaline phosphatase were markedly increased in OS patients with death vs. survival and primary metastasis was significantly related to the overall survival of patients (**Table 1**).

Risk Factor Selection and Validation

The MCMC method selected three primitive factors with a posterior probability of at least 0.90, including lower limbs, tumor volume and primary metastasis. The final model was developed with a logistic model, and primary metastasis and tumor volume were selected.

The risk model based on the two risk factors demonstrated good discrimination, calibration, overall fit, and explained variation. The internally validated c-statistic was 0.7308. For calibration, the validated calibration slope was 0.9660. The mean observed in-hospital mortality rate ranged from 21.05% in the lowest predicted quantile to 68.42% in the highest predicted quantile, a range of 47.4%. The validated Brier score was 0.2021, and the validated explained variation was 0.2006.

Risk Score

The risk stratification of in-hospital mortality is demonstrated in **Table 2**. Based on the distribution of the risk score, 24.8, 49.7, and 25.5% of patients were stratified into the high- (risk score $\geq$ 67.5+), average- (risk score 4.7–67.5) and low-risk (risk score 0–4.7) groups for in-hospital mortality, with corresponding probabilities of 0.684, 0.329, and 0.210, respectively (**Figure 3**).

TABLE 2 | Patients risk stratification based on risk score.

Risk groups	Patients, *n* (%)	In-hospital mortality, mean (%)
High (risk score 67.5+)	38 (24.8)	68.4
Average (risk score 4.7–67.5)	76 (49.7)	32.9
Low (0–4.7)	39 (25.5)	21.0

DISCUSSION

Osteosarcoma is the most common primary malignant bone tumor with higher rates of invasion and metastasis. Although these therapeutic regimens for osteosarcoma have been unprecedentedly advanced, the prognosis is still poor in patients with OS (17). The 5-year overall survival rate of OS patients has remained virtually unchanged in recent decades, especially for metastatic osteosarcomas (6, 17). This may be due to the rarity and heterogeneity of the tumor, together with the lack of pathognomonic mutations identified and the limited targeted treatments thus far (3, 7, 17, 18). Thus, a novel and innovative risk model predicting mortality is urgently needed to increase the understanding of factors identified to exert prognostic effects in patients with OS.

Several clinical trials have been performed to identify the prognostic indicators relevant to the mortality of OS patients, which have been mostly conducted in developed countries. However, the outcomes vary among these studies. The common clinically detected tumor size is widely evaluated in prognostic modeling, whereas the predictive performance of tumor size in predicting outcome was inconsistent among different studies. Some studies indicated that tumor size was one of the significant prognostic factors to predict in-hospital mortality in OS patients (9, 19–22), while tumor size lost its significance in another study; it did not appear to be a significant prognostic factor at all (13). Moreover, worldwide, there is still no consensus as to the standardized risk model that can be used to evaluate the prognosis of OS. Thus, in the absence of availability of these clinical data (such as histologic response to chemotherapy and type of surgery and surgical margins) before the treatment of OS cases, it is difficult to determine whether the prognostic factor evaluated is the true prognostic effect or not and to

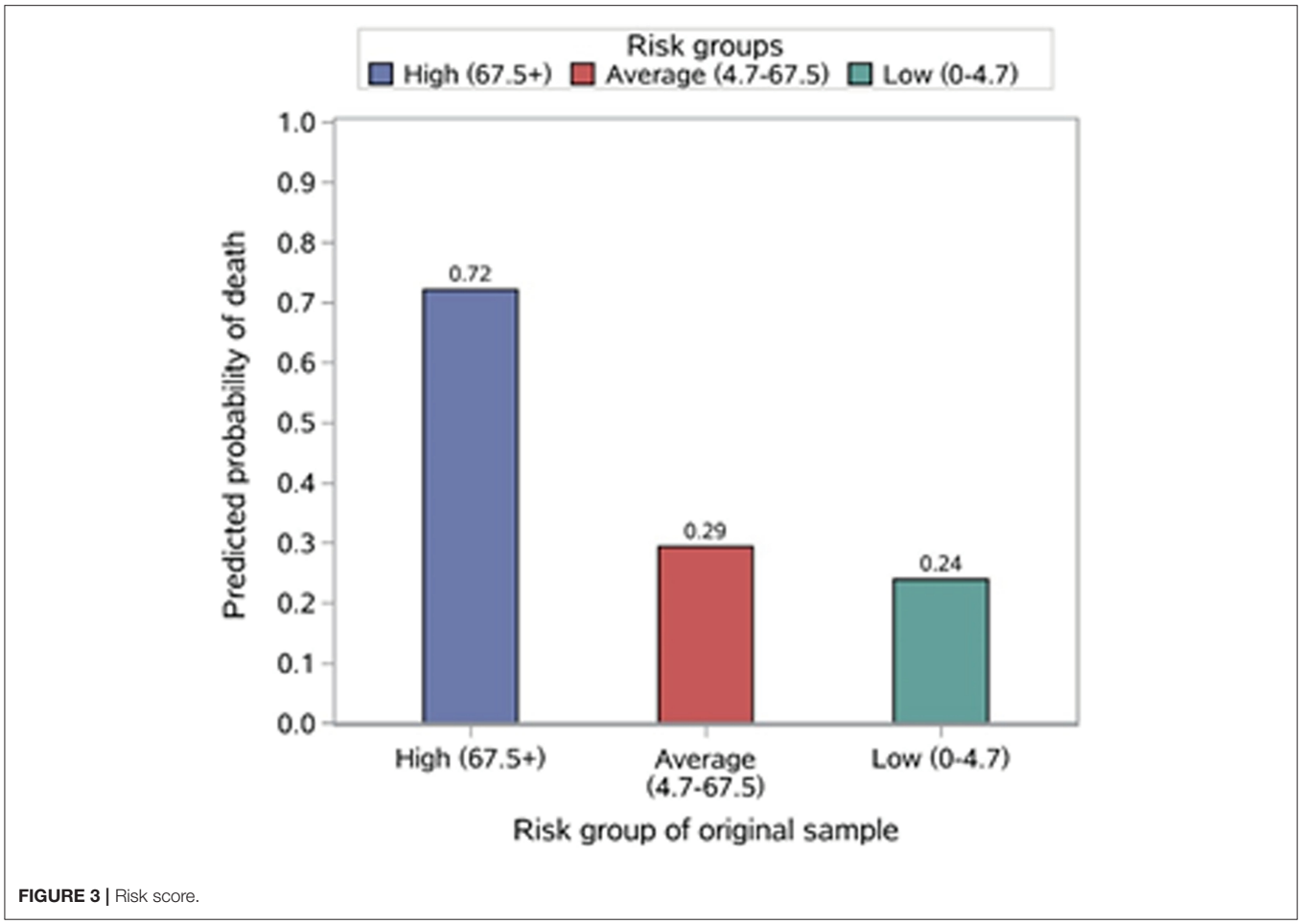


FIGURE 3 | Risk score.

apply the proposed risk factors to stratify patients at diagnosis before treatment.

We conducted this study based on the patient cohort with OS. First, we analyzed the clinical data of OS cases and selected these potential prognostic variables associated with the outcome. These clinical data, as predictive factors, are conveniently collected, widely reliable, and clinically important during hospitalization and could be applied to predict the outcomes of OS cases by statistical algorithms. Then, we employed MCMC simulation to explore the strength of the correlation between these factors and the prognosis of patients with OS. Two risk factors (metastasis, tumor volume) were identified with the aggressive regimens, which were independent predictive indicators of 5-year survival in OS cases. Additionally, their prognostic value has been recognized in many studies (8–10, 23, 24). Finally, a simple risk model was constructed and evaluated based on the two factors, which reflected good discrimination, calibration, overall fit, and explained variation. We further constructed a simple risk score to stratify patients into three risk groups of in-hospital mortality. Through the risk stratification, we found that 24.8% of the patients were at high risk of in-hospital mortality, which emphasized the importance of identifying these patients to provide them with targeted and systemic treatment and establish appropriate surveillance strategies. Moreover, on the foundation of the risk model, we are able to offer OS cases useful prognostic information and predict survival at diagnosis. Thus, the results of this study demonstrated that not only this risk model but also the risk score had an important potential application in clinical work.

Despite the advantages outlined above, our study has several limitations. First, the study did not have sufficient OS patients at new diagnosis. The main reason is the rarity of the disease and the difficulty in accumulating adequate cases. Second, this was a single-center retrospective study in China, and the performance of the risk model lacked validation in more independent regions and different races. Third, our results were based on the foundation of the existing medical records database and lacked the independent and external validity of this study. Finally, further studies are needed to estimate and confirm the generality of our results.

CONCLUSION

In summary, in this study, we established a novel prognostic risk model based on clinical data from OS patients at primary

diagnosis in China. It may help clinicians stratify patients into different risk groups of in-hospital mortality, provide them with targeted and systemic treatment and establish appropriate surveillance strategies. Hence, these findings offer a direction to predict the prognosis of OS cases.

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/s.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Ethics Committee of Affiliated Cancer Hospital of Zhengzhou University. Written informed consent for participation was not provided by the participants' legal guardians/next of kin because: The clinical study was retrospective involving human participants. Only the clinical data (gender, age, tumor site, etc.) of hospitalized patients with osteosarcoma are collected, it does not involve intervention in the treatment and prognosis of patients, and does not bring risks to the physiology of patients. Meanwhile, we will do our best to protect patient information and not disclose patient privacy. The Ethics Committee waived the requirement for written informed consent for participation.

AUTHOR CONTRIBUTIONS

YC and ZR designed the study. YC, CL, XW, and ZL retrieved references and analyzed data. YC and CL wrote the manuscript. ZR revised the manuscript. All authors reviewed and approved the manuscript.

FUNDING

This study was sponsored by grants from National Key Research and Development Program of China (2018YFC2000501), National Natural Science Foundation of China (U2004121), and China Postdoctoral Science Foundation (2020T130609).

ACKNOWLEDGMENTS

We thank the colleagues from Department of Orthopedic Surgery and Medical Record Department of Affiliated Cancer Hospital of Zhengzhou University. We thank all the participants who have involved in this study.

REFERENCES

1. Ottaviani G, Jaffe N. The epidemiology of osteosarcoma. *Cancer Treat Res.* (2009) 152:3–13. doi: 10.1007/978-1-4419-0284-9_1
2. Mirabello L, Troisi R, Savage SA. Osteosarcoma incidence and survival rates from 1973 to 2004: data from the surveillance, epidemiology, and end results program. *Cancer.* (2009) 115:1531–43. doi: 10.1002/cncr.24121
3. Kansara M, Teng M, Smyth M, Thomas DJ. Translational biology of osteosarcoma. *Nat Rev Cancer.* (2014) 14:722–35. doi: 10.1038/nrc3838
4. Klein M, Siegal GP. Osteosarcoma: anatomic and histologic variants. *Am J Clin Pathol.* (2006) 125:555–81. doi: 10.1309/UC6KQHLD9LV2KENN
5. Gaspar N, Occean B, Pacquement H, Bompas E, Bouvier C, Brisse H, et al. Results of methotrexate-etoposide-ifosfamide based regimen (M-Ei) in osteosarcoma patients included in the French Os2006/Sarcome-09 Study. *Eur J Cancer.* (2018) 88:57–66. doi: 10.1016/j.ejca.2017.09.036
6. Allison D, Carney S, Ahlmann E, Hendifar A, Chawla S, Fedenko A, et al. A meta-analysis of osteosarcoma outcomes in the modern medical era. *Sarcoma.* (2012) 2012:704872. doi: 10.1155/2012/704872
7. Davis A, Bell R, Goodwin PJ. Prognostic factors in osteosarcoma: a critical review. *J Clin Oncol.* (1994) 12:423–31. doi: 10.1200/JCO.1994.12.2.423
8. Ferrari S, Bertoni F, Mercuri M, Picci P, Giacomini S, Longhi A, et al. Predictive factors of disease-free survival for non-metastatic osteosarcoma of the extremity: an analysis of 300 patients treated at the Rizzoli Institute. *Ann Oncol.* (2001) 12:1145–50. doi: 10.1023/A:1011636912674
9. Bielack S, Kempf-Bielack B, Delling G, Exner G, Flege S, Helmke K, et al. Prognostic factors in high-grade osteosarcoma of the extremities or trunk: an analysis of 1,702 patients treated on neoadjuvant cooperative osteosarcoma study group protocols. *J Clin Oncol.* (2002) 20:776–90. doi: 10.1200/JCO.2002.20.3.776
10. Bacci G, Longhi A, Ferrari S, Briccoli A, Donati D, De Paolis M, et al. Prognostic significance of serum lactate dehydrogenase in osteosarcoma of the extremity: experience at Rizzoli on 1421 patients treated over the last 30 years. *Tumori.* (2004) 90:478–84. doi: 10.1177/030089160409000507
11. Bacci G, Longhi A, Versari M, Mercuri M, Briccoli A, Picci P. Prognostic factors for osteosarcoma of the extremity treated with neoadjuvant chemotherapy: 15-year experience in 789 patients treated at a single institution. *Cancer.* (2006) 106:1154–61. doi: 10.1002/cncr.21724
12. Whelan J, Jinks R, McTiernan A, Sydes M, Hook J, Trani L, et al. Survival from high-grade localised extremity osteosarcoma: combined results and prognostic factors from three european osteosarcoma intergroup randomised controlled trials. *Ann Oncol.* (2012) 23:1607–16. doi: 10.1093/annonc/mdr491
13. Ogura K, Fujiwara T, Yasunaga H, Matsui H, Jeon D, Cho W, et al. Development and external validation of nomograms predicting distant metastases and overall survival after neoadjuvant chemotherapy and surgery for patients with nonmetastatic osteosarcoma: a multi-institutional study. *Cancer.* (2015) 121:3844–52. doi: 10.1002/cncr.29575
14. Bertrand T, Cruz A, Binitie O, Cheong D, Letson GD. Do surgical margins affect local recurrence and survival in extremity, nonmetastatic, high-grade osteosarcoma? *Clin Orthop Relat Res.* (2016) 474:677–83. doi: 10.1007/s11999-015-4359-x
15. Harrell FE Jr, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med.* (1996) 15:361–87. doi: 10.1002/(SICI)1097-0258(19960229)15:4<361::AID-SIM168>3.0.CO;2-4
16. Steyerberg E, Harrell F, Borsboom G, Eijkemans M, Vergouwe Y, Habbema JD. Internal validation of predictive models: efficiency of some procedures for logistic regression analysis. *J Clin Epidemiol.* (2001) 54:774–81. doi: 10.1016/S0895-4356(01)00341-9
17. Gill J, Gorlick G. Advancing therapy for osteosarcoma. *Nat Rev Clin Oncol.* (2021) 18:609–24. doi: 10.1038/s41571-021-00519-8
18. Corre I, Verrecchia F, Crenn V, Redini F, Trichet VJC. The osteosarcoma microenvironment: a complex but targetable ecosystem. *Cells.* (2020) 9. doi: 10.3390/cells9040976
19. Wuisman P, Enneking W, Roesner A. Local growth and the prognosis of osteosarcoma. *Int Orthop.* (1992) 16:55–8. doi: 10.1007/BF00182986
20. Spanier S, Shuster J, Vander Griend RA. The effect of local extent of the tumor on prognosis in osteosarcoma. *J Bone Joint Surg.* (1990) 72:643–53. doi: 10.2106/00004623-199072050-00002

21. Mankin H, Hornicek F, Rosenberg A, Harmon D, Gebhardt MC. Survival data for 648 patients with osteosarcoma treated at one institution. *Clin Orthopaed Relat Res.* (2004) 429:286–91. doi: 10.1097/01.blo.0000145991.65770.e6
22. Petrilli A, Gentil F, Epelman S, Lopes L, Bianchi A, Lopes A, et al. Increased survival, limb preservation, and prognostic factors for osteosarcoma. *Cancer.* (1991) 68:733–7. doi: 10.1002/1097-0142(19910815)68:4<733::AID-CNCR2820680412>3.0.CO;2-0
23. Kager L, Zoubek A, Pötschger U, Kastner U, Flege S, Kempf-Bielack B, et al. Primary metastatic osteosarcoma: presentation and outcome of patients treated on neoadjuvant cooperative osteosarcoma study group protocols. *J Clin Oncol.* (2003) 21:2011–8. doi: 10.1200/JCO.2003.08.132
24. Petrilli AS, de Camargo B, Filho VO, Bruniera P, Brunetto AL, Jesus-Garcia R, et al. Results of the Brazilian osteosarcoma treatment group studies III and IV: prognostic factors and impact on survival. *J Clin Oncol.* (2006) 24:1161–8. doi: 10.1200/JCO.2005.03.5352

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.

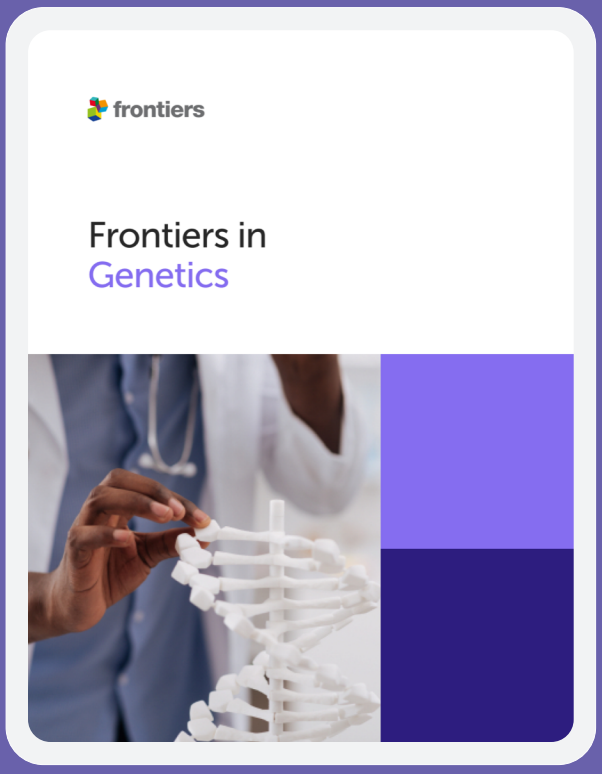
Frontiers in Genetics

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