Genome Medicine

RESEARCH

Open Access

Genome Tunisia Project: paving the way for precision medicine in North Africa

Yosr Hamdi[1,2*], Mediha Trabelsi[3,4†], Kais Ghedira[5†], Maroua Boujemaa[1†], Ikhlas Ben Ayed[6,7†], Cherine Charfeddine[1,8†], Amal Souissi[7†], Imen Rejeb[9,10†], Wafa Kammoun Rebai[1], Chaima Hkimi[5], Fadoua Neifar[7], Nouha Jandoubi[1], Rahma Mkaouar[1], Melek Chaouch[5], Ayda Bennour[11,12], Selim Kamoun[5], Hend Chaker Masmoudi[1,13,14], Nabil Abid[15], Maha Mezghani Khemakhem[16], On behalf of the GTCA Consortium, Saber Masmoudi[7], Ali Saad[11,12], Lamia BenJemaa[9,10], Alia BenKahla[5], Samir Boubaker[1,2], Ridha Mrad[3,4], Hassen Kamoun[6,17], Sonia Abdelhak[1,18], Moez Gribaa[11,12†], Neila Belguith[3,17†], Najla Kharrat[7†], Dorra Hmida[11,12†] and Ahmed Rebai[7]

Abstract

Background Key discoveries and innovations in the field of human genetics have led to the foundation of molecular and personalized medicine. Here, we present the *Genome Tunisia Project, a two-phased initiative (2022–2035)* which aims to deliver the reference sequence of the Tunisian Genome and to support the implementation of personalized medicine in Tunisia, a North African country that represents a central hub of population admixture and human migration between African, European, and Asian populations.

The main goal of this initiative is to develop a healthcare system capable of incorporating *omics* data for use in routine medical practice, enabling medical doctors to better prevent, diagnose, and treat patients.

Methods A multidisciplinary partnership involving Tunisian experts from different institutions has come to discern all requirements that would be of high priority to fulfill the project's goals. One of the most urgent priorities is to determine the reference sequence of the Tunisian Genome. In addition, extensive situation analysis and revision of the education programs, community awareness, appropriate infrastructure including sequencing platforms and biobanking, as well as ethical and regulatory frameworks, have been undertaken towards building sufficient capacity to integrate personalized medicine into the Tunisian healthcare system.

Results In the framework of this project, an ecosystem with all engaged stakeholders has been implemented including healthcare providers, clinicians, researchers, pharmacists, bioinformaticians, industry, policymakers, and advocacy

†Mediha Trabelsi, Kais Ghedira and Maroua Boujemaa contributed equally to this work.

†Ikhlas Ben Ayed, Cherine Charfeddine, Amal Souissi and Imen Rejeb contributed equally to this work.

†Moez Gribaa, Neila Belguith, Najla Kharrat, and Dorra Hmida contributed equally to this work.

*Correspondence:
Yosr Hamdi
yosr.hamdi@pasteur.utm.tn
Full list of author information is available at the end of the article

© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

groups. This initiative will also help to reinforce research and innovation capacities in the field of genomics and to strengthen discoverability in the health sector.

Conclusions Genome Tunisia is the first initiative in North Africa that seeks to demonstrate the major impact that can be achieved by Human Genome Projects in low- and middle-income countries to strengthen research and to improve disease management and treatment outcomes, thereby reducing the social and economic burden on healthcare systems. Sharing this experience within the African scientific community is a chance to turn a major challenge into an opportunity for dissemination and outreach. Additional efforts are now being made to advance personalized medicine in patient care by educating consumers and providers, accelerating research and innovation, and supporting necessary changes in policy and regulation.

Keywords Tunisian reference genome, Precision medicine, Human genomics, Socio-economic impact, Data sharing, Government support, Health economics, Biobanking, Care pathway, Cost-effectiveness

Background

The first complete human genome released 20 years ago led to several achievements beyond what scientists thought possible. Currently, about 30 million people worldwide have had their genomes sequenced [1]. This remarkable progress hides large inequalities between populations. Actually, it is estimated that about 78% of available human genome data is from individuals of European ancestry, and only 2% is from African origin [2]. The UK stands out as a country that offers genome sequencing as part of routine care with a perspective of analyzing 500,000 genomes by 2023/24, while the USA ambitions, with the Precision Medicine Initiative (PMI) launched by President Obama in 2015, is to sequence over 1 million genomes by 2025 (https://allofus.nih.gov/) to move to healthcare delivery and to tailor treatment and prevention strategies. Additional initiatives from Estonia, Singapore, Qatar, Denmark, China, and France took place and provided insightful outcomes [3].

Although African populations have the highest level of ethnic diversity, with over 3 million unique genetic variants that have been recently discovered and have never been reported before [4, 5], they still remain consistently understudied [2].

Completed in 2015, the 1000 Genomes Project [6] is one of the most comprehensive genomics initiatives to date by sequencing 2054 genomes of individuals from 26 populations in Africa, East Asia, Europe, South Asia, and the Americas [7]. The African Genome Variation Project launched in 2015 also allowed the assessment of the genetic diversity among 1481 individuals from 18 ethno-linguistic groups in Sub-Saharan Africa and released whole-genome sequences (WGS) of 320 Africans [8].

Additionally, a major advance in African genomics knowledge has been made through the Human Heredity and Health in Africa (H3Africa) initiative, which was launched in 2010 and funded by the US National Institute of Health (NIH) and the UK's Wellcome Trust in partnership with the African Society of Human Genetics for a period of 10 years to support genomics-oriented research projects [9, 10]. H3Africa allowed to sequence 350 whole genomes at high coverage plus 160 at medium coverage from African individuals [4, 11]. All genomic data generated by the H3Africa projects are available in the H3Africa Data and Biospecimen Catalog (https://catalog.h3africa.org/), which currently contains 23 datasets and 23,421 biospecimens from 16 large studies, and are accessible through a request to the H3Africa Data and Biospecimen Access Committee (DBAC).

However, North African populations are among the most understudied African populations, with no data available on WGS for individuals from Tunisia, Algeria, and Mauritania, while very few are available for individuals from Morocco, Libya, and Egypt [12–16]. Thereby, there is an urgent need for genomic data generation in North Africa, not only to elucidate questions related to its population history but also to understand the involvement of the genetic variants in human diseases and to pave the way for precision medicine.

The Genome Tunisia Project was therefore launched in 2019 to set up the reference sequence of the Tunisian Genome and to implement personalized medicine in Tunisia.

The main goal of this initiative is to develop a healthcare system capable of incorporating genomic information for use in routine medical practice and in pandemic settings. Indeed, genomic analysis has already been integrated into Tunisia's healthcare system since 2005. However, this approach lacks systematic and organized implementation. In addition, access to high throughput sequencing technologies remains limited. The lack of data on the Tunisian reference genome makes the existing activity less reliable. The Genome Tunisia Project aims to address these gaps by structuring genomic analysis activities, making it more accessible, efficient, and sustainable.

With the support of the Ministry of Health, a multidisciplinary network of Tunisian experts from universities, research and clinical centers, university hospitals, industry, and civil society, called the Genome Tunisia Collaborative Alliance (GTCA), has been constituted. Extensive situation analysis has been undertaken to guide the establishment of an appropriate ecosystem towards building sufficient capacity to effectively integrate precision medicine and genetic diagnostics into the Tunisian health system [17]. The Genome Tunisia Project seeks thereby to demonstrate the major impact that can be achieved by combining population sequencing and genotyping programs to improve disease management and treatment outcomes and to reduce the socio-economic burden on healthcare systems in Tunisia and in other African populations. The Genome Tunisia Project has been recently included in the Tunisian finance law (2023) with a dedicated budget and financial support from the Tunisian Government represented by the Tunisian Ministry of Health.

Genome Tunisia Project: context and overview of the Tunisian health system

Tunisia, the northernmost country in Africa, is situated in a strategic position at the heart of the Mediterranean basin and the cross-road between Europe, the Middle East, and Africa.

The analysis of the historical and cultural context of the Tunisian population has led to conclude that the current population is composed of a mixture between an indigenous Berber majority, a relatively large Arab contribution, a less important Mediterranean contribution including a Spanish and Italian backgrounds, with a very small, relatively recent, sub-Saharan African contribution [18, 19]. This heterogeneity is mainly explained by the migratory waves of allogenic populations and various civilizations including Phoenicians, Vandals, Romans, Arabs, Ottomans, and finally French during the protectorate period from 1881 to 1956 [20, 21]. The particular structure of the Tunisian population is also due to a relatively high rate of consanguinity that has an impact on the incidence of monogenic as well as polygenic diseases.

Yet within the Middle East and North African (MENA) region, Tunisia is characterized by the improvement of its health indicators. According to the most updated World Bank's data, Tunisia has the highest life expectancy at birth with an increase of 22.25 years over the last four decades [22]. The decline of the infant mortality rate is the most important progress achieved, decreasing from 80 per thousand in 1980 to 11 per thousand in 2023 (https://data.worldbank.org/indicator/SP.DYN.LE00.IN?locations=TN). This success can be partially attributed to the relatively strong national health system. Indeed, the Tunisian healthcare system is closely monitored by the Ministry of Health and supported by several public institutions including the National Laboratory for Drug Control, the National Center of Pharmacovigilance and the Central Pharmacy of Tunisia. Additional institutions are also playing a crucial role in the health system including the National Authority for Evaluation and Accreditation in Health (INEAS) and The National Health Insurance Fund (CNAM) (http://www.santetunisie.rns.tn). Moreover, Tunisia had built a robust infrastructure for primary healthcare over the past 50 years, as a vital framework to foster care and commitment at the community level aiming to reduce health inequities [23]. In addition, remarkable progress in the biomedical research field was observed through building capacities and reinforcing infrastructure. Consequently, the comprehensive Tunisian health profile shifted by a significant decrease in the prevalence of communicable diseases and an increase of the non-communicable diseases (NCDs) burden such as genetic disorders [23]. According to the 2016 report of the World Health Organization (WHO), NCDs in Tunisia were responsible for 86% of all causes of death (49% from cardiovascular diseases, 12% from cancer, 5% from chronic respiratory diseases, and 5% from diabetes) and for 16% of premature deaths between 30 and 70 years [24]. Therefore, genetic diseases are increasingly recognized as a real public health challenge in Tunisia, explaining, in part, the number of studies published during the last three decades on Tunisian disease cohorts as well as those from neighboring countries. These studies identified genes involved in human Mendelian diseases, in particular those with a recessive mode of inheritance. The availability of next-generation sequencing (NGS) at a relatively accessible cost has accelerated the molecular and genetic investigations of these conditions and allowed to improve the diagnosis of other chronic diseases such as cancer [25], cardiac diseases [26–28] monogenic forms of diabetes [29], and several other hereditary and genetic disorders.

So far, Tunisia has four public medical genetics centers, in "Charles Nicolle Hospital (CNH)" in Tunis, in "Mongi Slim Hospital" (MSH) in La Marsa, in "Farhat Hached Hospital" (FHH) in Sousse, and in "Hedi Chaker Hospital" (HCH) in Sfax, with two private medical genetics laboratories that offer genetic diagnoses services [30]. These medical genetics centers are very well connected to two main biomedical research centers namely Institut Pasteur de Tunis (IPT) in the North and the Center of Biotechnology of Sfax (CBS) in the South. Researchers and medical

doctors from these centers are also affiliated to medicine and science faculties in the different Tunisian regions as part of their academic activities and also to keep the intersectoral connections within these institutions.

With the introduction of targeted therapies, and the development of molecular tools, genomic testing became one of the most powerful tools of cutting-edge healthcare investigations that can improve patient outcomes in Tunisia [31]. Consequently, Tunisian medical geneticists and researchers recognized the urgent need to scale up genomic research to the next level and translate research findings into tangible benefits in the clinical settings. This required a comprehensive national strategy that carefully takes into account the specific local context. These reflections led to launching a genomic-based population project called Genome Tunisia that has the potential to fulfill these main priorities.

Methods

Genome Tunisia Project: goals and vision

In 2019, a wide public debate started in order to establish a national strategy for a Tunisian Human Genome Project. It resulted in setting up a consortium called The Genome Tunisia Collaborative Alliance (GTCA). GTCA is a multidisciplinary team including highly experienced as well as young scientists, researchers, and medical doctors from different fields such as human genomics, anthropology, medical genetics, bioinformatics, biostatistics, pathology, molecular biology, ethics, and cytogenetics.

The original goal of the Genome Tunisia Project is to generate the reference sequence of the Tunisian genome and to identify genetic variations specific to the Tunisian population in order to improve genetic testing services and progressively implement personalized treatments. Subsequently, the Genome Tunisia Project exceeded its initial set of goals to reach additional aims (Fig. 1) including: (1) strengthening public health system, (2) reinforcing research system and capacity building, and (3) socio-economic development.

The Genome Tunisia Project has then been divided into two phases with a first 2-year pilot phase (Fig. 2A) structured into five milestones starting from (M1) getting appropriate ethical and regulation approvals to (M2) sample collection, (M3) sequencing of a small sample set as a pilot phase, (M4) performing data analysis, and finally (M5) sharing the results with the community. The first phase (2022–2024) will then be followed by a 10-year second phase (2025–2035) which will lead to the implementation of precision medicine in Tunisia (Fig. 2B). This second phase is formed by 12 task forces (TF) that are pivotal for the long-term

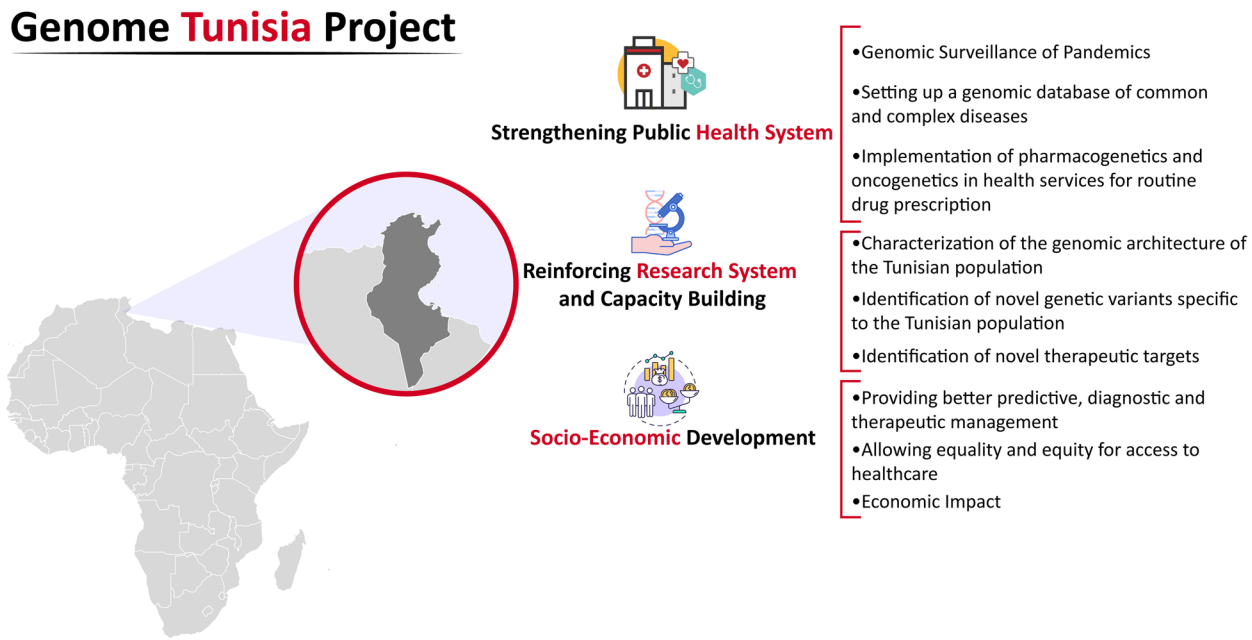

Fig. 1 Main goals of the Genome Tunisia Project (2022–2035). The Genome Tunisia Project has three main aims (1) strengthening public health system, (2) reinforcing research system and capacity building, and (3) socio-economic development. Each of these aims is divided into specific goals including identification of genetic variants specific to the Tunisian population, implementation of pharmacogenetics and oncogenetics in routine medical care, discovery of novel therapeutic targets, and enabling equality and equity for access to medical care

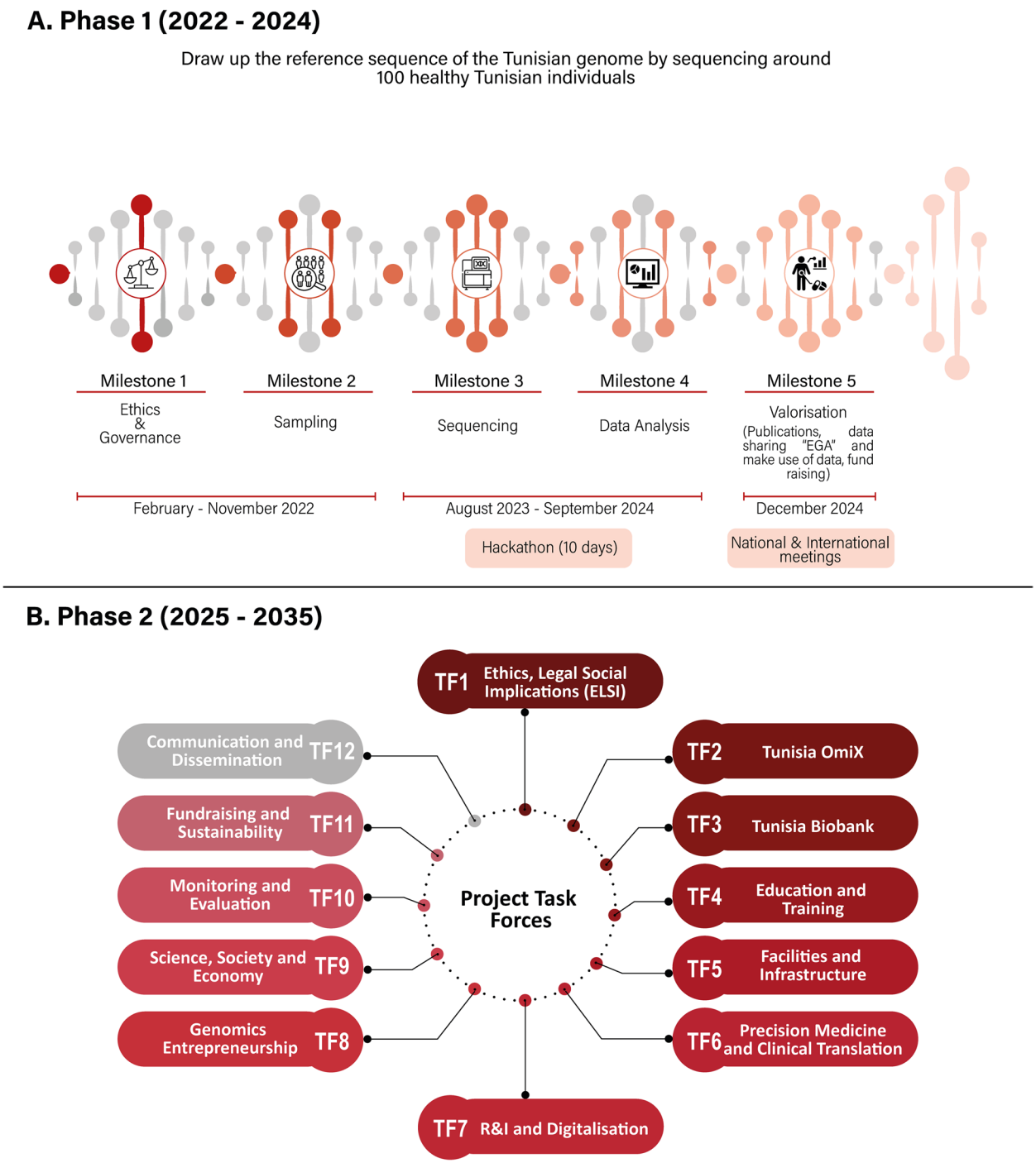

Fig. 2 Main milestones and task forces of the Genome Tunisia Project (2022–2035). **A** Phase 1—Milestones and timelines. Milestone 1: ethical and regulation approvals; milestone 2: sample collection; milestone 3: whole-genome sequencing; milestone 4: data analysis; milestone 5: results sharing. **B** Phase 2—Twelve task forces are planned for the second phase of the Genome Tunisia Project (TF1 to TF12)

sustainability of the project covering several fields including ethical legal and social implications, biobanking, training, genomic entrepreneurship, health economics, and digitalization (Additional file 1).

Ethics and regulation approvals

The research protocol describing the background, rationale, objectives, study design, methodology, and the ethical considerations relating to the study was

submitted to the Biomedical Ethics Committee of Institut Pasteur de Tunis (CEBM) for evaluation. Ethical approval according to the Declaration of Helsinki Principles was obtained after addressing the reviewers' comments (reference: 2021/31/1/V2). Moreover, the other institutional ethics committees of the research centers, hospitals, and institutions belonging to the GTCA consortium have been consulted and letters of support were also obtained to conduct the Genome Tunisia Project from the participating entities.

In order to ensure that the study is in compliance with the national data protection regulations, the project was also submitted to the Tunisian Regulatory body on Personal Data protection called "*Instance Nationale de Protection des Données Personnelles* (INPDP)" for review and approval.

After obtaining approval from both ethical and regulatory organizations (CEBM and INPDP), sample collection was planned and written informed consents were obtained from all participants before sample collection.

Sample collection

Blood samples were collected from the general population in three main regions of Tunisia: Tunis, Sousse, and Sfax, representing the Northern, Central, and Southern regions of the country. Sample collection and info days have been organized in these regions to ensure a faithful representation of the whole Tunisian population and all ethnic groups. The sampling process and the geographical distribution of the participants are illustrated in Additional file 2: Fig. S1.

A combination of stratified and purposive sampling was used. Participants were enrolled in the study on the basis of the following inclusion and exclusion criteria:

Inclusion criteria

(1) Healthy individuals
(2) Unrelated participants (to capture the diversity of the population given the high consanguinity rates)
(3) Participants representing the general Tunisian population and from different ethnic groups
(4) Signed consents
(5) Participants who provided a minimal phenotypic data set.

Exclusion criteria

(1) Non-Tunisian participants
(2) Age < 18 years old
(3) Vulnerable population

In the current study, the vulnerable population includes pregnant women, elderly individuals, and individuals with mental illness or developmental disabilities. In Tunisian national ethical rules, these vulnerable populations should be excluded in the case where their participation is not mandatory in the research studies in order to respect their physical and mental health status and their self-identification and to avoid any potential side effects.

For all participants included in the study, face-to-face questionnaires have been filled with demographic data including age, sex, language, lifestyle, ethnicity, marital status, socio-economic status, and as a detailed family medical history. The questionnaire consisted of open-ended questions in both Arabic and French languages. The questionnaires, as well as the collected samples, were anonymous and systematically stored in a coded clinical record form in order to guarantee participants' personal data protection.

Phenotypic data collection

Phenotypic data collection in genomics research encompasses social, environmental, phenotypic, epidemiological, and clinical aspects that need to be aligned with genomic data. In biomedical research projects, case report forms (CRFs) are employed for gathering primarily clinical and phenotypic data to ensure precise, efficient, and consistent data collection, leading to high-quality, dependable data, and trustworthy research outcomes.

In order to follow the recommendation of using well-established data collection standards, Genome Tunisia CRF was developed using RedCap CRFs established in the H3ABioNet Core Phenotypes Standards project (https://www.h3abionet.org/images/DataAndStandards/DataStandards/H3ABioNet_StandardCRF_Guidelines_20180907_Version_1.2.pdf). The used CRF was adapted to the specific data collection method and to the Tunisian context (Additional file 3). Experts in the field, comprising both genetic researchers and medical geneticists with prior involvement in genomic research, were consulted to review the proposed CRFs.

DNA extraction

Genomic DNA was automatically extracted from whole blood samples collected during the first phase of the project using QIAamp DNA Blood Mini Kit (Qiagen) on the QIAcube, according to the manufacturer's instructions. For genomic DNA Quality Control, concentration and purity were determined using the DeNovixTM DS-11-FX and the Thermo Scientific NanoDrop 1000 UV/VIS

spectrophotometers. Genomic DNA integrity was verified by means of agarose gel electrophoresis.

Sequencing platforms

So far, sequencing facilities have been implemented in the main public institutions of Tunisia including two medical genetics centers (CNH, FHH) as well as two research centers (IPT, CBS) belonging to the Genome Tunisia Project and affiliated to both ministries of Health and of Higher Education (Fig. 3). These facilities include a NextSeq 550, iSeq 100, NextSeq 1000, and two MiSeq platforms located in the North, Central, and South of Tunisia in order to cover the sequencing needs of the whole country. Additional sequencing facilities are installed at the participating institutions such as Hedi Chaker Hospital and Faculty of Sciences of Tunis (Fig. 3). In its first phase, Genome Tunisia aims to sequence 100 whole genomes as a pilot phase of the project with an ultimate goal to reach 10,000 WGS from the Tunisian population. To do so, the implementation of a national sequencing platform with a Novaseq X machine for large-scale whole exome and whole-genome sequencing is well underway.

Ethical legal and social implications

Genome Tunisia, as all Human Genome projects, faces challenges, with ethical, legal, and social issues. These have been and will be addressed at different levels with appropriate tools and measures. A DBAC has been established to control biospecimen and data sharing according to national ethical guidelines and respect of privacy and consent to ensure equitable benefits and access to genomic data for all stakeholders. The data collected during the project, besides their contribution to capacity building in genomics and bioinformatics,

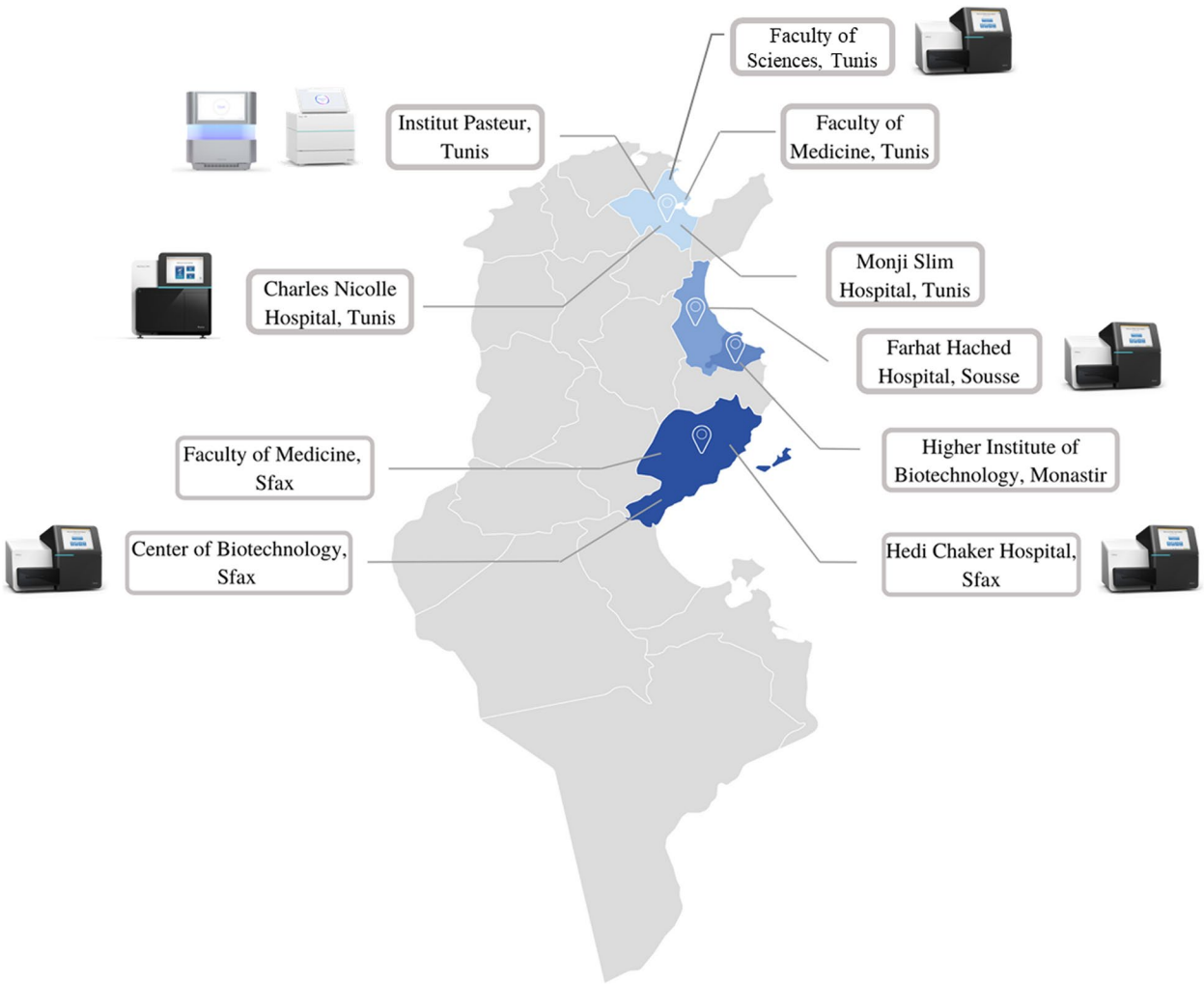

Fig. 3 NGS platforms installed in public institutions, members of the Genome Tunisia Project located in the Northern, Central, and Southern regions of the country

will facilitate the implementation of personalized medicine. The involvement of all stakeholders and decision makers (Ministry of Health and Ministry of Scientific Research and Higher Education) as partners in the project will maximize its benefit to the population without any discrimination. Moreover, drawing upon the experience of the H3Africa ELSI project (Rufus O. Akinyemi et al. 2021), several other contextual intervention tools and practices will be developed to address ethical, legal, and social issues. This includes education and training programs to strengthen the competences of researchers and health professionals, genomics, and precision medicine community awareness programs to enhance public understanding of genomics and precision medicine and also the development of informative and training videos to improve genomic and precision medicine literacy.

Results

Data infrastructure

As previously stated, one of the major goals of the Genome Tunisia Project is to reinforce the public health system through strengthening existing infrastructure and building capacities. To fulfill these goals, it was essential to evaluate the readiness level for implementing such an ambitious project and particularly to generate, manage, and analyze large-scale genomics data, including the reference sequence of the Tunisian population. An exhaustive and detailed catalog of existing wet labs, bioinformatics infrastructure, and human resources available in the different research and clinical centers involved in the project was prepared.

Indeed, information related to available bioinformatics and computational biology infrastructure for omics data analysis and data science has been gathered from each participating center. It included the available number of machines/servers and their corresponding capacities in terms of cores, central processing units (CPUs), random access memory (RAM), storage, as well as the availability of Internet connection, the bandwidth, and the backup. Among participating centers, Institut Pasteur de Tunis is equipped with an institutional high-performance computing (HPC) SLURM-managed cluster that will be used to analyze the first batch of data generated in the framework of this project. This latter comprises a centralized infrastructure designed for robust computational tasks. It includes a head/master node, managing operations across the network. Supported by five computing nodes, this cluster ensures parallel processing capabilities, enhancing efficiency and speed. Additionally, a dedicated high-memory node (750 GB, 128 GPU cores, and 272 CPU cores) facilitates memory-intensive processes. For data management, the cluster provides around 100 TB of storage alongside a backup server of equal capacity, ensuring data integrity, and accessibility. This setup is supported by two data center switches, optimizing network connectivity and data flow within the cluster. In addition, the Center of Biotechnology of Sfax includes a data infrastructure dedicated to NGS data analysis and storage and recently DRAGEN servers were installed in two hospitals participating in this project, Charles Nicolle and Farhat Hached Hospitals. Moreover, two additional data centers, Al Khwarizmi and the health data center, have been also put in place by the Tunisian Ministry of Higher Education and Scientific Research and the Tunisian Ministry of health, respectively.

For the second phase of sequencing, there are ongoing plans to establish a national platform dedicated to sequencing and data analysis equipped by high throughput sequencers and HPC clusters to effectively manage the increasing volume of data expected to be generated within the project.

In addition, experienced scientists in the different centers (engineers, system administrators, bioinformaticians, computer scientists, data scientists, etc.) were identified and are mainly involved in organizing regular training activities in the participating centers.

Data analysis team

Given that a Human Genome Project requires highly skilled human resources, a "Data Analysis Team" was created to handle genomics data that will be generated in the framework of the Genome Tunisia Project. The team members were selected based on a survey that was designed to evaluate data analysis skills of existing human resources. The survey includes questions on Linux commands, Genome assembly pipelines, Conda environment, and the use of HPC clusters. Thirty-two members responded to the survey and 20 members were selected. Moreover, as Genome Tunisia is a national project supported by the Tunisian Government, the Tunisian Ministry of Health in cooperation with the Ministry of Higher Education and Scientific Research has recently provided 15 permanent positions in different specialties including bioinformatics. The newly recruited bioinformaticians have already joined the Genome Tunisia Data Analysis Team. Additional bioinformaticians will be hired soon to reinforce other public health institutions. The first batch of data will be generated in May 2024 and will be analyzed before September 2024.

We intend to conduct population genetic studies to first investigate the genetic landscape of the Tunisian population by dressing up haplotypes and identifying the structure and admixture component of the Tunisian

population. Different bioinformatics tools (R packages, ADMIXTURE, etc.) will be employed for this analysis to make use of the generated data in order to bring new insights on the genetic diversity of our population. We will also compare the genetic architecture of the Tunisian population to that of other neighboring populations as well as European and SubSaharan populations, using suitable bioinformatics tools. In addition, we will integrate anthropological and historical data with genomics to comprehensively characterize the genetic ancestry and the diversity of the Tunisian population. This will help guide personalized medicine interventions by taking into account the unique genetic makeup of the Tunisian population.

Genome Tunisia's Data and Biospecimen Access Committee: the GTCA DBAC

Because of their sensitive nature, all the data that will be generated in the framework of the Genome Tunisia Project will be stored in a national portal and secured with controlled access that will be granted by the GTCA-DBAC. The biospecimen or data access requests (abbreviated BAR or DAR, respectively) will be evaluated by an independent committee, named data (and biospecimen if relevant) access committee (DAC or DBAC, respectively). The DACs/DBACs evaluate and then grant or decline access to data, after examination of the DAR/BAR in compliance with national and international guidelines.

The GTCA DBAC was established in December 2022, with its composition and data access guidelines adapted from the H3Africa project (H3Africa Data and Biospecimen Access Committee Guidelines, April 2020, retrieved from App-D-H3Africa-Data-and-Biospecimen-Access-Committee-Guidelines-2020-.pdf).

The GTCA DBAC is formed by eight members with relevant expertise in different fields including biobanking, data analysis, genomics, ethics, law, anthropology, and patient advocacy. The first DBAC meeting was held on March 2, 2023, to schedule the annual meeting plan and discuss specific issues on evaluation criteria of requests.

NGS analyses and care pathway

The previously mentioned achievements of the Genome Tunisia Project in its first phase in reinforcing data and sequencing infrastructure bring us to establish a "patient care pathway" for genetic testing. Indeed, genetic counseling is a complex task that involves not only patient management but also their families, by offering risk assessment, diagnosis, and predictive testing. In addition, it includes pre-analysis, analytical, and post analysis phases which involve several actors namely, medical geneticists, molecular biologists, laboratory technicians, and bioinformaticians.

Several brainstorming meetings with the different stakeholders were organized to come up with preliminary proposals of the following care pathway (Fig. 4) that goes from the test indication to the delivery of the clinical report by specifying the role of the different specialists at each phase:

The pre-analysis phase starts by the prescription of an appropriate genetic test by a medical specialist, according to the patient's clinical data and family history. Then, patients will be referred to one of the medical genetics departments in the different Tunisian hospitals. The medical geneticist will draw up a detailed pedigree and validate the prescribed genetic test (gene panel or clinical exome). When whole-exome sequencing (WES) or whole-genome sequencing (WGS) analyses are required, this decision will be made within a multidisciplinary consultation meeting, involving the molecular biologist and the medical specialist. The medical geneticist will inform the patient of the purpose of the analysis, the limitations of the technique, the risk of identifying incidental findings, and the potential consequences of a positive result. The patient must give his/her written consent before genetic examination. Finally, a blood, liquid biopsy, saliva, or a tissue sample will be collected depending on the prescribed test.

The analytical phase of the genetic testing includes the design and validation of the test, the technical validation, data analysis, and data interpretation. While the design and validation step are done under the supervision of the medical geneticist and the molecular biologist, the DNA extraction, library preparation, and sequencing steps will be performed by the laboratory technician under the supervision of a bio-engineer. The bioinformatician will ensure data analysis and processing as well as the technique validation. The medical geneticist and the molecular biologist will ensure the variants interpretation and may, if needed, review the technical validation step. Once a potential disease-causing variant has been identified, additional analysis may be performed to confirm the finding and to determine the clinical significance of the identified variation.

To harmonize data analysis and to ensure a consistent interpretation of results, FASTQ files generated from sequencing in existing facilities will be transferred to a central data center equipped with high-performance computing clusters. At this central facility, comprehensive quality control will be conducted on raw data to ensure their integrity and accuracy. A standardized bioinformatic pipeline will be used to analyze, annotate, and filter data. These data will be incorporated into a national database reporting Tunisian-specific polymorphisms

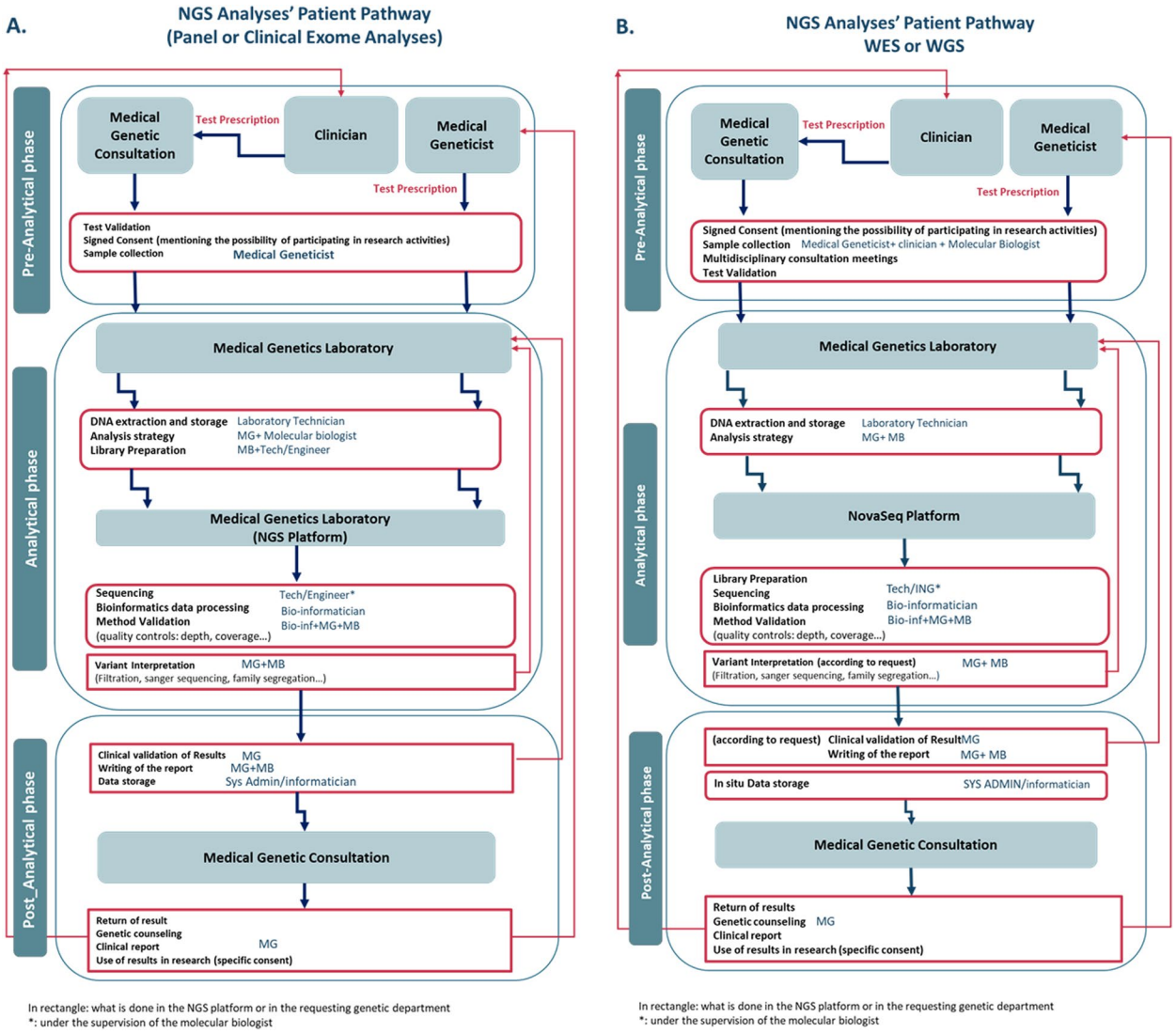

Fig. 4 Care pathway for genetic testing using NGS. **A** When using gene panels or clinical exome analysis. **B** When using WES or WGS analysis. The pathways differ in the sequencing platforms used and also the need for multidisciplinary consultation meetings to validate the need for WES or WGS

and also causal variants allowing for better clinical decision-making.

The post-analytical phase: validation of the identified causal genetic variant, if any, is essential to establish the clinical report. A phenotype–genotype correlation will be performed within a multidisciplinary board. This validation step will be followed by preparing a clinical report that will be co-signed by the medical geneticist and the molecular biologist. If any fresh or tumor biopsies have been handled by a pathologist, this later will also co-sign the report. The medical geneticist will return and explain the result to the patient who will be carefully informed on the identified genetic variants, their clinical significance, and their possible impact on the patient's health. The patient will also be informed of the importance of sharing information with the family members, who might need appropriate genetic testing in order to assess their own risk. However, it is important to respect patient's privacy and medical confidentiality. In some cases, a prenatal diagnosis may be considered for future pregnancies, particularly if the identified genetic variant is associated with a severe or life-threatening condition. Finally, a clinical report must also be sent to the medical specialist in order to ensure a better communication of the result in a clear and comprehensible format to help in the therapeutic decision-making.

Communication

The Genome Tunisia communication team prepared relevant documents to communicate about the project and to efficiently disseminate its objectives and expected findings. A cohesive corporate identity alongside the communication materials were designed including brochures, posters, and templates. Furthermore, we are in the process of developing a detailed communication plan, which will encompass the core project messages to be conveyed.

The following communication materials and tools have been used in the Genome Tunisia Project:

- A project logo was designed and is used in all project documents and publications.
- A dedicated website is under construction, which will contain a precise description of the project's objectives, partners, and a public section that describes milestones and deliverables. The website will also contain the publication list of the project partners and link to relevant projects. An exclusive collaborative portal will be accessible only to authorized members to facilitate communication between project partners. A link to the DBAC activities and data and biospecimen access rules D will also be available on this website.
- Two sets of leaflets and posters have been designed and produced. The first set will disseminate the objectives, concepts, and vision of the project in its initial phase, while the second set will additionally disseminate public results, outcomes, and findings during the second phase of the project. These materials will be used in all scientific events in which GTCA members will participate.
- Press releases, containing the most important project results, will be issued after conferences.
- Regularly updated information about the project will be provided via web and social media channels.
- Newsletters will be published annually to keep stakeholders informed about the project's progress.

In addition, the Genome Tunisia Project was presented at several national and international events, including the annual meeting of the African Society of Human Genetics, held in Rabat, Morocco in December 2022 (https://www.sm2gh.ma/) and the International Congress of Human Genetics (ICHG) that took place in Cape Town, South Africa in February 2023 (https://www.ichg2023.com/) as well as a dozen of other meetings, conferences, and workshops. These communications and presentations are extremely important in raising public awareness and disseminating information about the project's main goals and expected outcomes which provide opportunities to engage with other researchers, scientists, healthcare professionals, and other stakeholders. It also helped to expand our network, build sustainable partnerships, attract international funders, and leverage the collective knowledge of the scientific community.

Perspectives

Sequencing 10,000 whole genomes

The specific objective of the second phase of the Genome Tunisia Project is to sequence about 5 to 10% of the Tunisian adult population, reaching the optimal scale for precise population-based studies. The initial plan is to start by sequencing 10,000 healthy individuals without any familial history of diseases. Then we will progressively increase the sample size to achieve the international guidelines. However, by adhering to the inclusion and exclusion criteria of this project, we would probably need to collect much less samples than initially anticipated given that complex diseases are highly prevalent in the adult population in Tunisia, notably diabetes, hypertension, and cancer (affecting approximately 30% of the adult population). Researchers and medical personnel in specific recruitment offices continue to recruit participants throughout the country in order to ensure a better representativeness of the different ethnic groups. As detailed previously, participants will provide informed consent, blood samples, and data will be collected using the established GTCA protocols.

National sequencing and analysis platform

The implementation of a national sequencing and analysis platform will enable access to genomic medicine on a routine basis. This national platform will be linked to the regional hospitals, research centers, and universities to ensure consistency in offering sequencing testing types and clinical interpretations of the generated data. Indeed, this platform will be dedicated to (1) sequencing, (2) data analysis, and (3) data interpretation. Generated data will be centralized, trimmed, and analyzed using a standardized bioinformatic pipeline. VCFs files will then be shared with qualified medical geneticists to make appropriate annotations and interpretations using standardized tools to ensure appropriate genetic testing. Newly identified mutations will be integrated into the national portal facilitating further explorations by other investigators and to ensure knowledge transfer and sharing within the scientific community.

The implementation of this national platform will enable to address three main challenges:

– Public health: by providing equal access to sequencing facilities for genetic diagnostic, prognostic, and theranostic purposes.

– Scientific, technological, and biomedical research: by offering sequencing service to researchers allowing a better understanding of pathologies in the field of rare and common diseases, as well as by developing bioinformatics skills, data science expertise, computational capacities, and infrastructure.
– Health economy: by offering outsourcing services to neighboring countries which will promote the development of a new genomic-based industry sector and reinforce medical tourism.

National biobank

As part of this national initiative, we are aiming to create a national biorepository that will gather biological material with associated genomics and clinical and phenotypic data from Tunisian patients and citizens. The national biobank will also constitute a valuable resource for biomedical research to establish new therapeutic and preventive measures at the population level. Genome Tunisia's ambition is to create a healthcare system that can exploit the power of a centralized large cohort collection and massive data generation. Our long-term vision is to put in place a federated standards-led informatics infrastructure, spanning research, and healthcare domains to record and store biological material as well as the collected clinical and genomic data. To do so, clearly agreed standards for genotypic–phenotypic data and biospecimen collection protocols should be set up.

Authorities in Tunisia will use the biobank as part of the infrastructure they are building to accelerate the advancement of precision medicine based on national and international biobanking guidelines.

Special interest groups (SIGs)

Sequencing disease-based cohorts are under discussion to investigate the genetic risk factors associated with the most frequent diseases in the country. Six special interest groups (SIGs) on cancer, rare diseases, neurological, autoimmune, cardiovascular, and infectious diseases have been agreed on and are under construction (Fig. 5).

WGS will be performed in unsolved cases previously investigated by gene panel or WES. The number of patients to be investigated as part of the disease cohorts has been fixed to 6000, as a first stage, and this was calculated after careful examinations based on the specific research objectives and resource considerations. This investigation will facilitate the detection of new disease-causing genes and/or novel alterations, especially within regulatory regions. Additionally, WGS will enable the identification of structural variants such as copy number variations (CNVs), inversions, and translocations, with greater accuracy and sensitivity. Using SNP genotyping

Fig. 5 The special interest groups that emerged from the Genome Tunisia Project. Tunisian Society of Human Genomics (TSHG)

arrays, we intend also to perform large-scale association studies such as GWAS to identify specific genetic variants associated with these diseases within the Tunisian population. The generated data will serve for the development of polygenic risk scores (PRS) that are specific to the Tunisian population, allowing for better risk stratification. We will employ a multidisciplinary approach engaging clinicians, geneticists, and researchers to determine the most suitable investigation strategy. This assessment will consider factors such as disease complexity, genetic variability, specific research objectives, and available resources.

Tumor board meetings will be implemented for the cancer SIG as a pilot phase. These multidisciplinary meetings will then be put in place for other diseases, mainly for the neurological and rare diseases.

To strengthen the capacities of the multidisciplinary scientific community constituting the GTCA consortium, the Tunisian Society of Human Genomics (TSHG) has been recently formed (March 2023). This initiative, which emerged from the GTCA consortium, is an independent society that will work in close collaboration with the African Society of Human Genetics as well as other African societies to build strong connections between Tunisian, African, and international scientists. The executive board of TSHG is composed of 17 scientists from different Tunisian regions and institutions, with a positive gender balance (13 women) and a significant youth representation (three early-career researchers).

Upon its official creation, the TSHG will organize its first general meeting in September 2024 and plans to invite the five existing African societies, to exchange experiences, and discuss the future directions of the Human Genomics field for the benefit of the African scientific community.

Socio-economic impact

In Tunisia, the government has long prioritized a healthcare system that is accessible to all citizens, offering equitable and affordable care. Over the past decade, particularly in the post-genomics era, significant efforts have been made to raise awareness about genetic testing facilities. However, more efforts are needed to ensure that individuals from all socio-economic backgrounds are aware of and have access to these services. The Genome Tunisia Consortium is aiming to address this issue by establishing two task forces on education and training (TF4) and communication and dissemination (TF12). So far, individuals from different socio-economic backgrounds have participated in the study representing therefore the socio-economic diversity.

In addition, genetic testing reimbursement is another important issue that we are working on by involving policy makers and insurance companies on board to ensure that cost does not serve as a barrier to access genetic testing. Furthermore, as outlined in Fig. 2, we have established a task force specifically focused on cost-effectiveness and health economics (TF9) to ensure equitable access to genomic testing in Tunisia.

GTCA includes competencies and stakeholders from the private and public sectors to lay the foundation for collaborations and to maximize the potential of health economics in Tunisia. The Ministry of Health of Tunisia has recently organized an open and informative day to discuss different aspects of precision medicine in Tunisia demonstrating the interest of the government to integrate genomics in its economic reform [32].

The Genome Tunisia Project could therefore present an opportunity for Tunisia to consolidate healthcare, setting digital infrastructure, improving the connection and the interoperability of its systems, and standardizing patient medical records. Also, it is an opportunity to elaborate patient registries and undertake collaborative work to tackle public health issues by exploring cost-effective plans.

Cost-effectiveness

Micro-costing and cost-effectiveness analysis of different genomic interventions, whole exome, whole genome, and large gene panels in different indications start to be evaluated, in different countries implementing genomics [33–37].

In the Genome Tunisia Project, the health economics task force is commissioned to analyze the cost-effectiveness of genomics applications for different indications considering what has been reported from genomic initiatives in other countries and also considering the local context (epidemiology, prevalence of the disease, public health issues, and the consanguinity rate).

The three main goals of the health economics task force of Genome Tunisia are to support decision-makers for an optimal allocation of resources, to lead collaborative work toward implementing cost-effective interventions, and to follow up the continuously evolving costs of sequencing and the evolving methodology for data modeling in order to capture and value genomic testing outcomes.

Synergy with ongoing projects: AGenDA, SEED, PerMediNA, GenCoE, and other international initiatives

Genome Tunisia is typically defined as a set of interrelated activities, task forces, and milestones aiming to strengthen the public health system and scientific research in the field of Human Genomics. Although conducting original research and launching national human genome initiatives are key factors to increase competitiveness at the international level, they are not sufficient to generate major advances in the field of Genomics. For these reasons, the GTCA consortium is participating in different regional and international initiatives, such as the Assessing Genomic Diversity in Africa (AGenDA) project, where GTCA represents the North African site.

Moreover, as an innovative initiative, the Genome Tunisia Project will enable the investigation of patients who have been previously investigated in previous research projects but have remained undiagnosed. One such project is the SEED (Strengthening the Sfax university expertise for diagnosis and management of epileptic encephalopathies) project (seedtwinning.com), which focuses on exploring monogenic forms of developmental and epileptic encephalopathies (DEE). The SEED project has already collected more than 300 patients with electrophysiological, neurologic, and radiological data. One hundred of these patients have already been explored with a customized panel of 116 genes associated with DEE, achieving a diagnosis rate of 30%. The Genome Tunisia Project represents a unique opportunity to support this initiative to draw the genetic landscape of this neurodevelopmental disorder in Tunisia by exploring undiagnosed cases and to position Tunisia as a reference center for the genetic diagnosis of DEE in the MENA region.

In addition, and to start preparing the ground for precision medicine implementation, Tunisia, represented by Institut Pasteur of Tunis (IPT), is coordinating a regional

project called Personalized Medicine in North Africa (PerMediNA), funded by the European Ministry of Foreign Affairs. PerMediNA is a key project on the implementation of precision medicine in Tunisia and North Africa involving several stakeholders from different institutions and countries. The main goals of PerMediNA are strengthening collaboration within the ecosystem, generating omics data on cancer patients, and reinforcing patient care by facilitating access to genetic testing and targeted therapies.

Finally, the governmental support to the Genome Tunisia Project as well as the industry commitment and the scientific community engagement make Genome Tunisia an excellent candidate for the GenCoE initiative that was announced in the International Congress of Human Genetics (ICHG) in Cape Town 2023 (https://www.ichg2023.com/).

Precision medicine implementation

The Genome Tunisia Project endeavors to implement precision medicine in Tunisia using an evidence-based model informed by local stakeholders and based on real settings, population specificities, and national context.

The key to a successful implementation of precision medicine in Tunisia will be a bottom-up approach in which members of the project join forces with academia, healthcare authorities, policy makers, and stakeholders to build a nationally distributed infrastructure and to raise awareness on the importance and the impact of the implementation of precision medicine. During the first phase of Genome Tunisia Project, we are evaluating the readiness level at the different centers involved in the project focusing on the infrastructure and bioinformatics human resources, identifying stakeholders and funders, and working on regulation and ethical aspects to support translational and clinical research projects and adapt precision medicine national plan. The readiness level for precision medicine implementation was assessed through a multidimensional approach, which included the organization of two workshops and interviews with key opinion leaders (KOLs). These workshops served as dynamic platforms for brainstorming and idea generation. Additionally, interviews with KOLs allowed for in-depth exploration of challenges and opportunities relevant to precision medicine implementation. Moreover, further efforts are ongoing to ensure that our study is thorough and exhaustive. This includes further interviews and data collection activities, such as focus groups, surveys, and questionnaires. Regarding the findings, while some initial insights have been collected, the analysis of the gathered data is still ongoing to ensure a well-documented illustration of the current readiness level.

As illustrated in Fig. 6, a key step was to gather information on all available NGS platforms, existing biobanks,

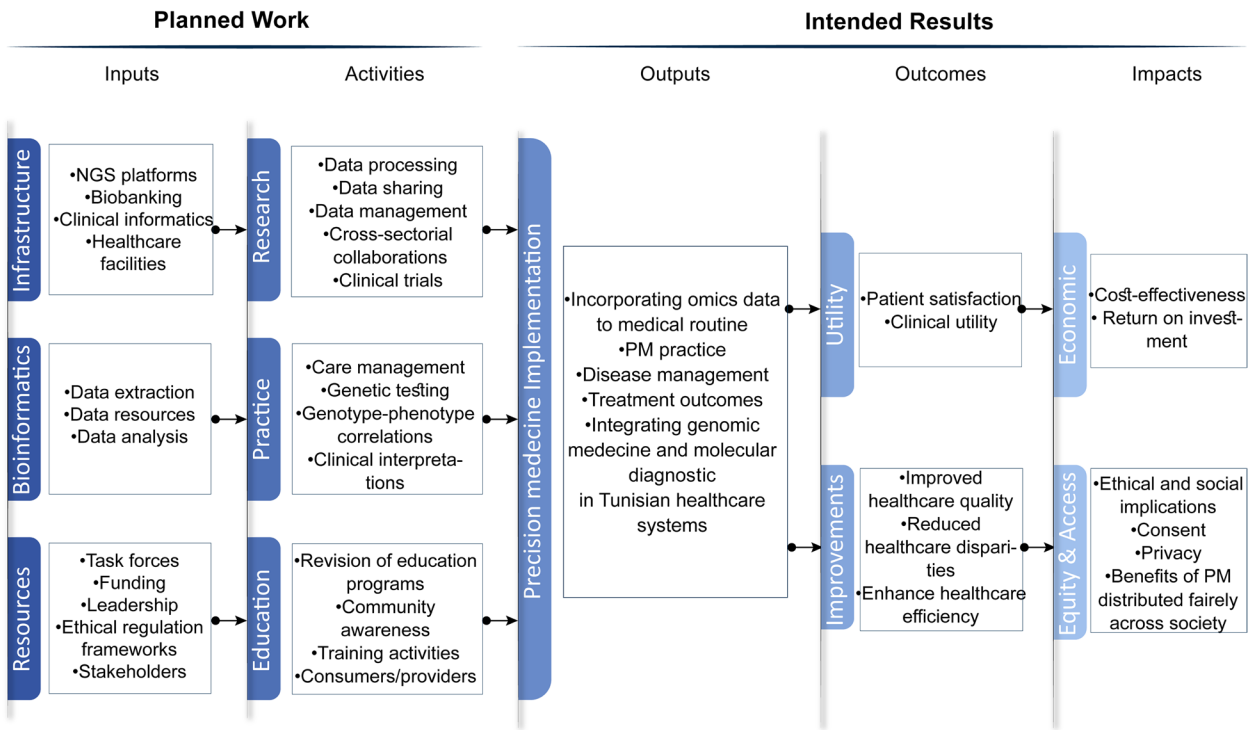

Fig. 6 Key steps for precision medicine (PM) implementation in Tunisia and the North African region (adapted from Chanfreau-Coffinier et al. 2019)

and healthcare facilities and identify gaps that need to be filled to be able to implement precision medicine in Tunisia and the North African region. Several NGS platforms are available in the different centers; however, to be able to set up an optimal environment and ecosystem, it is essential to implement a national genomics platform which will optimize the use of sequencing technologies.

Illustrating its cutting-edge activities, Tunisia takes a step forward to move toward the goal of delivering the right treatment for the right patient at the right time and at the right site of care. For a continuous development of new methods and treatment strategies, Genomic Medicine Centers will be established at all university hospitals and data generated in healthcare will be fed back to research through the Tunisian National Genomics Platform.

These Genomic Medicine Centers will be consolidated and reinforced with highly skilled bioinformaticians incorporating high throughput omics data analysis to medical routine. In this way, an innovative care pathway ensuring total or partial healthcare reimbursement and involving all national public and private stakeholders will be established to facilitate equal access to precision medicine at the national level.

Discussion

Since the publication of the first draft of a human genome in 2001, an increasing number of countries have launched their national Human Genome Projects [38–45].

The Human Genome Project (HGP) used one basic method for DNA sequencing, called Sanger DNA sequencing that has greatly advanced through a series of major technology innovations [46–49]. In parallel, sequencing costs significantly decreased over the years with the development of NGS technologies [49, 50]. Indeed, the HGP costs US$3 billion over 13 years and nowadays, a single genome costs as little as US$ 200 and takes no more than 1 week to be sequenced [51, 52]. Additionally, impressive advances in the field of bioinformatics, data analysis, and data science occurred during the last decade [53, 54].

The Genome Tunisia Project took advantage of this progress and is considered as the first Human Genome project in Africa to be supported and funded by the local government represented by the Tunisian Ministry of Health. The expected benefits of the Genome Tunisia Project include reinforced research and healthcare infrastructure, improved understanding of the genomic variation spectrum, shorter time and cost-effective genetic diagnosis, and targeted prevention, treatment, and research advances.

However, designing Human Genome projects in Arab and African populations, which have a distinctive genetic landscape due to its complex population history, should be based on the local population structure and specificities. In Tunisia, deeply rooted socio-cultural practice of strong familial links and consanguineous marriages resulted in high rates of endogamy and consanguinity. At a macro-genomic level, this has a direct impact on the frequency of genetic disorders by increasing the prevalence of rare diseases with a recessive transmission model. At a microgenetic level, this common practice of consanguinity not only results in a continuously growing list of novel variations and mutations but also impacts allele frequency and their penetrance. Thus, and in the context of small populations with high rates of consanguinity, designing a Human Genome Project does not need very large cohorts to be sequenced in order to get informative data on the reference genome sequence. Instead, sequencing a relatively small sample size representing different ethnicities and sub-populations would be the best cost-effective and efficient strategy to generate enough data that would inform on the genetic architecture of the studied population and would deliver the reference sequence of its genome.

Moreover, the relatively strong Tunisian health systems with national programs supported by the government will help to improve life expectancy resulting in a significant increase of non-communicable diseases with a remarkable decrease of communicable disease frequency in Tunisia. The current project will also enhance data sharing in public repositories for a better representation of North African genomics data in public databases. Indeed, data sharing and efficient data stewardship is a central component of successful national genomic initiatives. This will help generate more science and result in a better use of genetic findings which is especially needed in low-income settings. If done appropriately, sharing genetic data can help to reduce health inequalities. While the idea that data sharing promotes scientific progress is becoming widely accepted among research communities, funding bodies, and regulatory agencies, it raises many ethical and management issues. The Tunisian GTCA-DBAC is willing to ensure an appropriate data sharing plan of the Tunisia genomics data that will be generated in the framework of this project, opening, therefore, the doors to other African countries to share their data and to participate in flourishing public databases with more and more African datasets.

On another hand, the medical application of genome sequencing outcomes can cover the two categories of genetic disorders: monogenic and polygenic disorders. Regarding rare diseases, as an example of monogenic diseases, the Genome Tunisia Project will ensure that patients get the right diagnosis, increase awareness on rare diseases among Tunisian healthcare professionals,

improve coordination of care, and consequently enhance access to specialist care, treatments, and drugs.

For complex diseases, polygenic risk scores (PRS) are not yet routinely used by health professionals in Tunisia and in most if not all African countries because of the lack of genotyping data and because there are no local guidelines for PRS application. Nevertheless, PRS improves the assessment of disease risk for a reinforced preventive approach of frequent complex diseases. Developing and designing appropriate genotyping chips and gene panels adapted to the Tunisian and North African population is one of the goals of the Genome Tunisia Project to start targeting some specific genes and to use these specific arrays for PRS calculation.

However, before it reaches clinical application on a wide scale, PRS application and the reliable interpretation of the numerous and de novo variations revealed by WGS will require more expertise and validation. For instance, sequencing of a few thousands of African genomes has revealed millions of novel genomic variants that were mainly useful for studying African population structure [4]. The primary challenge now is to assess the pathogenicity of these novel variations and to select the pathogenic mutations to be integrated in genetic diagnostic which could orient clinical decision and patient care [55, 56]. In order to validate the functional impact and the pathogenicity of novel genetic variants, the GTCA consortium will also focus its efforts on functional genomics by implementing platforms dedicated to genome editing and other functional assays to guarantee a reliable translation of the research findings into clinical settings. Variant reclassification in ClinVar, Varsome, and other public databases is one of the urgent tasks that should be undertaken to ensure efficient genetic testing and to avoid clinical misdiagnosis. Variant interpretation guidelines specifically designed for the Tunisian population will be set up by the recently created Tunisian Society of Human Genomics on the basis of local specificities and the generated Tunisian reference genome.

Undoubtedly, lots of resources are needed to achieve all these ambitious goals. From the research perspective, the Tunisian Ministry of Higher Education and Scientific Research should optimize the available means and build on existing infrastructure by mutualizing all efforts for a more productive and insightful biomedical research. Moreover, the reality is that conventional job titles have become outdated in the current post genomics and digital era. It is then time to rethink job descriptions and profiles to be more adapted to the market demand. This requires a deep revision of the academic curriculum, the education programs, degrees as well as diplomas in higher education and even at earlier stages of the education system to create new profiles, specifically in bioinformatics, data science, data management, data protection, and molecular biology. This will also lead to a more dynamic entrepreneurship and innovation ecosystem for genomics-driven startups which needs more support to science-based innovation.

Furthermore, according to their efficient productivity, Tunisian researchers are ranked among the best researchers in the field of human genomics and precision medicine in Africa [57]; however, the country is suffering from brain drain with more than third of the human resources working in the field of biomedical research left to Europe, Middle East, and North America. By attracting the qualified and renowned Tunisian researchers working abroad and reintegrating them into the research and health system, it would be possible to raise and maintain the research quality and to improve the international standing and ranking of Tunisian universities. For instance, the GTCA consortium is not only relying on local skills but is also involving the diaspora and the Tunisian experts working on other Human Genome Projects as members of the advisory board and the DBAC of the Genome Tunisian Project.

From the healthcare system perspective, the Tunisian Ministry of Health should consolidate all efforts for an efficient precision medicine implementation. One of the main opportunities characterizing the Tunisian healthcare system is the strong national political commitment towards achieving universal health coverage and health in all policies "Health is a right for every human being" as indicated in Law no. 38 of the new Tunisian Constitution, 2022. The government in collaboration with the different stakeholders, e.g., researchers, clinicians, academic members, pharmacists, policymakers, non-governmental organizations, and civil society identified the following national health priorities: (1) ensuring equity in access to healthcare services by including genetic testing in the reimbursement system, (2) reinforcing health promotion and diseases prevention programs, (3) strengthening and bridging the research and clinical fields, and (4) setting up an integrated and disaggregated health information system based on electronic health records.

Finally, taking part of initiatives such as the H3Africa and the H3ABioNet helped the GTCA group not only to develop local competencies in genomics and bioinformatics [58] but also to deliver relevant documentations and tools that were extremely useful for the Genome Tunisia Project including the CRF questionnaires developed by the H3Africa/H3AbioNet phenotype harmonization working group and the data sharing guidelines developed by the H3Africa DBAC. Thus, establishing more South–South collaboration between Tunisia and sub-Saharan Africa is one of the priorities of the GTCA consortium. The consortium continues its involvement

in current African and international initiatives including AGenDA and PerMediNA projects.

To summarize, generating the reference sequence of the Tunisian Genome and discovering new variations will lead to more productive scientific research, improved genetic counseling practice, oriented diagnosis, reinforced prevention strategies, cost-effective, and personalized treatments as well as an improved health economic outcome. In addition, building regional and national precision medicine networks is the key to prevent rare and complex diseases, reinforce genomic surveillance, improve pandemic preparedness, and ensure national security.

Conclusions

Genome Tunisia is the first Human Genome project in Africa to be supported and funded by the local government represented by the Tunisian Ministry of Health. It represents a pioneering effort in genomics research in North Africa that aims to achieve the complete reference sequencing of the Tunisian genome. Additional national partners and stakeholders from Tunisia and from the diaspora will be involved soon in the Genome Tunisia Project for a better representation of the national efforts dedicated to this field.

More Human Genome projects might be also conducted in other African countries to overcome the huge gap of missing data in Africa. This project, whose scientific and socio-economic outcomes will impact neighboring countries, will substantially enhance our comprehension of the human genomics field and its involvement in health and pathology not only in Tunisia but in other understudied populations worldwide.

Abbreviations

AGenDA	Assessing Genomic Diversity in Africa
BAR	Biospecimen Access Requests
CBS	Centre of Biotechnology of Sfax
CEBM	Biomedical Ethics Committee of Institut Pasteur de Tunis
CPUs	Central processing units
CRFs	Case report forms
CNH	Charles Nicolle Hospital
CNAM	National Health Insurance Fund
DAC	Data access committee
DAR	Data access requests
DBAC	Data and Biospecimen Access Committee
DEE	Developmental Epileptic Encephalopathies
ELSI	Ethics, legal social implications
FHH	Farhat Hached Hospital
GDP	Gross domestic product
GenCoE	Genetic Centers Of Excellence
GTCA	Genome Tunisia Collaborative Alliance
HCH	Hedi Chaker Hospital of Sfax
HGP	Human Genome Project
H3ABioNet	H3Africa Bioinformatics Network
H3Africa	Human Heredity and Health in Africa initiative
ICHG	International Congress of Human Genetics
InGeNA	Industry Genomics Network Alliance
INPDP	Personal Data Protection Instance
INEAS	National Authority for Evaluation and Accreditation in Health
IPT	Institut Pasteur de Tunis
MENA	Middle East and North African
NCD	Non-communicable diseases
NIH	National Institute of Health
LMICs	Low- and middle-income countries
NIH	National Institute of Health
PerMediNA	Personalized Medicine in North Africa
PMI	Precision medicine initiative
PRSs	Polygenic risk scores
QALY	Quality-adjusted life year
RAM	Random access memory
R&D	Research and development
R&I	Research and innovation
ROI	Return on investment
SIGs	Specific interesting groups
SEED	Strengthening the Sfax university expertise for diagnosis and management of Epileptic Encephalopathies
TSHG	Tunisian Society of Human Genomics
UK	United Kingdom
USA	United States of America
WES	Whole-exome sequencing
WGS	Whole-genome sequencing
WHO	World Health Organization

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13073-024-01365-w.

Additional file 1. Description of the GTCA task forces. This file includes a detailed description of the twelve Task Forcesof the second phase of the GTCA project that are pivotal for the long-term sustainability of the project covering several fields including Ethical Legal and Social Implications, Biobanking, Training, Genomic Entrepreneurship, Health Economics, and digitalization.

Additional file 2: Fig. S1. Sampling process and geographical distribution of the participants in the first phase of the project. This figure illustrates the sampling process and geographical distribution of the individuals who have consented to participate in the first phase of the Genome Tunisia project.

Additional file 3. Genome Tunisia Collaborative Alliance CRF. This file represents the used CRF in Genome Tunisia project that was developed using RedCap CRFs established in the H3ABioNet Core Phenotypes Standards project and adapted to the specific data collection method and to the Tunisian context.

Acknowledgements

The authors would like to express their deepest gratitude to the Tunisian Ministry of Health for the guidance, support, and the facilitation of the process to conduct this study. The co-authors would also like to thank all those who gave generously of their time and insights for the elaboration of this project, the medical and paramedical staff that have participated in sample collection, the participants volunteering to take part in a study, and the Tunisian Ministry of Higher Education and Scientific Research for their support. The authors would also like to extend their sincere gratitude to Dr. Hamdi Mbarek for his invaluable contribution to the revision of this manuscript. His expertise has significantly enhanced the quality of our work. Thank you for your unwavering support and fruitful advice.

A heartfelt thanks to all GTCA consortium members: Ahlem Achour, Ines Ouertani, Lilia Kraoua, Rym Meddeb, Sana Skouri, Sana Gabtni, Hela Bellil, Yasmina Elaribi, Syrine Hizem, Houweyda Jilani, Sonia Kchaou, Malek Trigui, Khouloud Rjiba, Hamza Chouk, Marwa Mahdouani, Sarra Dimassi, Soumaya Mougou, Narjess Trabelsi, Yasmine El Abed, Safa Boubaker, Rihen Braham, Salma Najjar, Ahlem Msakni, Fatma Chtioui, Lamia Hammami, Hela Gargouri, Fidaa Jbeli, Riadh Ben Marzough, Mouna Choura, Meriem Ben Said, Yosra Lajmi, Souhir Guidara, and Manel Guirat. Analysis Team: Nesrine Kerkeni, Maryam Younsi, Sana Karoui, Oussema Souiai, Mariem Hannachi, Houcemeddine Othman, Haifa El Mabrouk, Feriel Bouzid, Cyrine Abid, Nihel Ammouss, Imene Boujelbene, and Oussema Khamessi. DBAC: Abdeljelal Ghram, Amade M'charek,

Hamdi Mbarek, and Thouraya Mellah. National and international experts: Noomen ElKadri, Houssein Khoujet El Khil, Hassen Chaabani, Maher Kharrat, Raja Gargouri, Lotfi Cherni, Faiza Fakhfakh, and Chahnaz Charfi Triki.

Authors' contributions

YH, SA, and AR conceived and designed the study. YH, NB, MG, MB, NJ, WKR, AS, FN, IBA, RM, CC, SK, CH, KG, AB, MC, ABK, SM, ASa, LBJ, SB, RMr, HK, DH, and MT participated in data and sample collection. YH, AR, SA, KG, HCM, NK, NB, MT, MMK, AB, MC, and DH drafted the original manuscript. YH, MB, AR, SA, HCM, WKR, AS, FN, IB, RM, SK, CC, KG, MT, MMK, AB, DH, IR, SM, ASa, LBJ, SB, RMr, HK, and NA contributed to writing—review and editing. MB, FN, and CH performed the data visualization. MB and RM formatted and submitted the manuscript. All authors read and approved the final manuscript.

Funding

This work and the Genome Tunisia Project are largely supported by the Tunisian Ministry of Health, the General Direction of Health, and the Direction of Biomedical Research. The Tunisian Ministry of Higher Education and Scientific Research has also contributed to funding sample collection.

Availability of data and materials

The datasets supporting the conclusions of this article are included within the article.

Declarations

Ethics approval and consent to participate

Ethical approval according to the Declaration of Helsinki Principles was obtained from the Biomedical Ethics Committee of Institut Pasteur de Tunis (2021/31/1/V2). All participants gave written informed consent to participate in the study.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

[1]Laboratory of Biomedical Genomics and Oncogenetics, LR16IPT05, Institut Pasteur de Tunis, University of Tunis El Manar, 13, place Pasteur, B.P. 74, Tunis, Belvédère 1002, Tunisia. [2]Laboratory of Human and Experimental Pathology, Institut Pasteur de Tunis, Tunis, Tunisia. [3]Department of Congenital and Hereditary Diseases, Charles Nicolle Hospital, Tunis, University of Tunis El Manar, Tunis, Tunisia. [4]Laboratory of Human Genetics, Faculty of Medicine of Tunis, University of Tunis El Manar, Tunis, Tunisia. [5]Laboratory of Bioinformatics, Biomathematics and Biostatistics, Institut Pasteur de Tunis, University of Tunis El Manar, Tunis, Tunisia. [6]Department of Medical Genetics, Hedi Chaker University Hospital, University of Sfax, Sfax, Tunisia. [7]Laboratory of Molecular and Cellular Screening Processes, Centre of Biotechnology of Sfax, University of Sfax, Sfax, Tunisia. [8]Higher Institute of Biotechnology of Sidi Thabet, BiotechPole of Sidi Thabet, University of Manouba, Ariana, Tunisia. [9]Department of Congenital and Hereditary Diseases, Mongi Slim University Hospital, Sidi Daoud La Marsa, Tunis, Tunisia. [10]Santé Mère-Enfant (LR22SP01), Tunis, Tunisia. [11]Faculty of Medicine, University of Sousse, Sousse, Tunisia. [12]Department of Genetics, Farhat HACHED University Hospital, Sousse, Tunisia. [13]Faculty of Pharmacy of Monastir, University of Monastir, Monastir, Tunisia. [14]Department of Histology and Cytogenetics, Institute Pasteur of Tunis, Tunis, Tunisia. [15]Laboratory of Transmissible Diseases and Biological Active Substances LR99ES27, Faculty of Pharmacy, University of Monastir, Ibn Sina Street, Monastir 5000, Tunisia. [16]Laboratory of Biochemistry and Biotechnology (LR01ES05), Faculty of Sciences of Tunis, University of Tunis El Manar, Tunis 1068, Tunisia. [17]Laboratory of Human Molecular Genetics, LR99ES33, Faculty of Medicine of Sfax, University of Sfax, Sfax, Tunisia. [18]Communication, Science and Society Support Unit (UniSS), Institut Pasteur de Tunis, Tunis, Tunisia.

Received: 2 May 2023 Accepted: 16 July 2024
Published online: 27 August 2024

References

1. Crespi S. Looking back at 20 years of human genome sequencing. Podcast Science; 2021. p. 4.
2. Pennisi E. Genomes arising. American Association for the Advancement of Science. 2021;371(6529):556–559. https://doi.org/10.1126/science.371.6529.556.
3. Kovanda A, Zimani AN, Peterlin B. How to design a national genomic project—a systematic review of active projects. Hum Genomics. 2021;15(1):1–13.
4. Choudhury A, et al. High-depth African genomes inform human migration and health. Nature. 2020;586(7831):741–8.
5. Tallman S, et al. Whole-genome sequencing of Bantu-speakers from Angola and Mozambique reveals complex dispersal patterns and interactions throughout sub-Saharan Africa. bioRxiv. 2022;2022.02. 07.478793.
6. Siva N. 1000 Genomes project. Nat Biotechnol. 2008;26(3):256–7.
7. Auton A, et al. A global reference for human genetic variation. Nature. 2015;526(7571):68–74.
8. Gurdasani D, et al. The African genome variation project shapes medical genetics in Africa. Nature. 2015;517(7534):327–32.
9. Rotimi C, et al. Research capacity. Enabling the genomic revolution in Africa. Science. 2014;344(6190):1346–8.
10. Bentley AR, Callier SL, Rotimi CN. Evaluating the promise of inclusion of African ancestry populations in genomics. npj Genomic Med. 2020;5(1):5.
11. Mulder N, et al. H3Africa: current perspectives. Pharmgenomics Pers Med. 2018;11:59–66.
12. Scott EM, et al. Characterization of greater middle eastern genetic variation for enhanced disease gene discovery. Nat Genet. 2016;48(9):1071–6.
13. Crooks L, et al. Identification of single nucleotide variants in the Moroccan population by whole-genome sequencing. BMC Genet. 2020;21(1):1–8.
14. ElHefnawi M, et al. Complete genome sequence and bioinformatics analysis of nine Egyptian females with clinical information from different geographic regions in Egypt. Gene. 2021;769:145237.
15. El-Attar EA, et al. Genomics in Egypt: current status and future aspects. Front Genet. 2022;13:797465.
16. Lorente-Galdos B, et al. Whole-genome sequence analysis of a Pan African set of samples reveals archaic gene flow from an extinct basal population of modern humans into sub-Saharan populations. Genome Biol. 2019;20(1):1–15.
17. Mezzi N, et al. Spectrum of genetic diseases in Tunisia: current situation and main milestones achieved. Genes (Basel). 2021;12(11):1820.
18. Elkamel S, et al. The orientalisation of North Africa: new hints from the study of autosomal STRs in an Arab population. Ann Hum Biol. 2017;44(2):180–90.
19. Fadhlaoui-Zid K, et al. Genome-wide and paternal diversity reveal a recent origin of human populations in North Africa. PLoS One. 2013;8(11):e80293.
20. Danver SL. Native peoples of the world: an encyclopedia of groups, cultures and contemporary issues. Routledge; 2015.
21. Kefi R, et al. Phylogeny and genetic structure of Tunisians and their position within Mediterranean populations. Mitochondrial DNA. 2015;26(4):593–604.
22. World Bank. World Development Report 2022: finance for an equitable recovery. Washington, DC: World Bank; 2022.
23. Ben Romdhane H, et al. Health system challenges of NCDs in Tunisia. Int J Public Health. 2015;60(Suppl 1):S39-46.
24. Khiari H, et al. Burden of non-communicable diseases in Tunisia, 1990–2017: results from the global burden of disease study. Pan Afr Med J. 2021;40:62.
25. Hamdi Y, et al. Identification of eleven novel BRCA mutations in Tunisia: impact on the clinical management of BRCA related cancers. Front Oncol. 2021;11:674965.
26. Jaouadi H, et al. H syndrome: clinical, histological and genetic investigation in Tunisian patients. J Dermatol. 2018;45(8):978–85.
27. Jaouadi H, et al. Multiallelic rare variants support an oligogenic origin of sudden cardiac death in the young. Herz. 2020;46(S1):94–102.
28. Jaouadi H, et al. Identification of two variants in AGRN and RPL3L genes in a patient with catecholaminergic polymorphic ventricular tachycardia suggesting new candidate disease genes and digenic inheritance. Clin Case Rep. 2022;10(2):e05339.

29. Dallali H, et al. Genetic characterization of suspected MODY patients in Tunisia by targeted next-generation sequencing. Acta Diabetol. 2019;56:515–23.
30. Elloumi-Zghal H, Bouhamed HC. Genetics and genomic medicine in Tunisia. Mol Genet Genomic Med. 2018;6(2):134.
31. Ramappa M, Verma A, Edward DP. Commentary: genomic testing is a powerful tool in diagnosing and managing anterior segment dysgenesis. Indian J Ophthalmol. 2022;70(7):2303–4.
32. Génomique et Médecine Personnalisée. Levier de modernisation du système de santé et de développement socio-économique: Quel impact pour le patient et quelle stratégie pour la Tunisie? 2023.
33. Marino P, et al. Cost of cancer diagnosis using next-generation sequencing targeted gene panels in routine practice: a nationwide French study. Eur J Hum Genet. 2018;26(3):314–23.
34. Stark Z, et al. Prospective comparison of the cost-effectiveness of clinical whole-exome sequencing with that of usual care overwhelmingly supports early use and reimbursement. Genet Med. 2017;19(8):867–74.
35. Stark Z, et al. Does genomic sequencing early in the diagnostic trajectory make a difference? A follow-up study of clinical outcomes and cost-effectiveness. Genet Med. 2019;21(1):173–80.
36. Stark Z, et al. A prospective evaluation of whole-exome sequencing as a first-tier molecular test in infants with suspected monogenic disorders. Genet Med. 2016;18(11):1090–6.
37. Tan TY, et al. Diagnostic impact and cost-effectiveness of whole-exome sequencing for ambulant children with suspected monogenic conditions. JAMA Pediatr. 2017;171(9):855–62.
38. Stark Z, et al. Australian genomics: a federated model for integrating genomics into healthcare. Am J Hum Genet. 2019;105(1):7–14.
39. Caron NR, et al. Indigenous genomic databases: pragmatic considerations and cultural contexts. Front Public Health. 2020;8:111.
40. Halbert CH. Equity in genomic medicine. Annu Rev Genomics Hum Genet. 2022;23:613–25.
41. Gibbs RA. The human genome project changed everything. Nat Rev Genet. 2020;21(10):575–6.
42. Hasanzad M, et al. Genomic medicine on the frontier of precision medicine. J Diabetes Metab Disord. 2022;21(1):853–61.
43. Armeniangenome. Unraveling our genetic background. 2023. Available from: http://armeniangenome.am/.
44. CTGA. The catalogue for transmission genetics in Arabs. Available from: https://cags.org.ae/en. Accessed 15 Apr 2023.
45. chilegenomico. Genomics of the Chilean population. Available from: http://chilegenomico.med.uchile.cl/chilegenomico1/. Accessed 15 Apr 2023.
46. Goodwin S, McPherson JD, McCombie WR. Coming of age: ten years of next-generation sequencing technologies. Nat Rev Genet. 2016;17(6):333–51.
47. Park ST, Kim J. Trends in next-generation sequencing and a new era for whole genome sequencing. Int Neurourol J. 2016;20(Suppl 2):S76.
48. Levy SE, Boone BE. Next-generation sequencing strategies. Cold Spring Harb Perspect Med. 2019;9(7):a025791.
49. Kumar KR, Cowley MJ, Davis RL. Next-generation sequencing and emerging technologies. Seminars in thrombosis and hemostasis. Thieme Medical Publishers; 2024.
50. McCombie WR, McPherson JD, Mardis ER. Next-generation sequencing technologies. Cold Spring Harb Perspect Med. 2019;9(11):a036798.
51. Dong Z, et al. Low-pass whole-genome sequencing in clinical cytogenetics: a validated approach. Genet Med. 2016;18(9):940–8.
52. The $200 genome: new tech takes on sequencing's access problem. Available from: https://www.nature.com/articles/d42473-023-00014-x. Accessed 15 Apr 2023.
53. Zhong Y, et al. Application of next generation sequencing in laboratory medicine. Ann Lab Med. 2021;41(1):25–43.
54. Uesaka K, et al. Bioinformatics in bioscience and bioengineering: recent advances, applications, and perspectives. J Biosci Bioeng. 2022;134(5):363–73.
55. Matthijs G, et al. Guidelines for diagnostic next-generation sequencing. Eur J Hum Genet. 2016;24(1):2–5.
56. Katsanis SH, Katsanis N. Molecular genetic testing and the future of clinical genomics. Nat Rev Genet. 2013;14(6):415–26.
57. euafricapermed. The EU-Africa PerMed project. 2021. Available from: https://www.euafrica-permed.eu/.
58. Fadlelmola, Faisal M., et al. H3ABioNet genomic medicine and microbiome data portals hackathon proceedings. Database. 2021;(2021):baab016.
59. Akinyemi RO, et al. Unraveling the Ethical, Legal, and Social Implications of Neurobiobanking and Stroke Genomic Research in Africa: A Study Protocol of the African Neurobiobank for Precision Stroke Medicine ELSI Project. Int J Qual Methods. 2020;19. https://doi.org/10.1177/1609406920923194.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.