Request for Applications (RFA): Submission of Longitudinal Samples for Whole Genome Sequencing in the South African 110,000 Human Genome Pilot Programme

[RFA-COHORTS-SA-PILOT-HGP]

Issued by: South African Medical Research Council (SAMRC) in partnership with the Department of Science, Technology, and Innovation (DSTI)

RFA Release Date: 26 May 2025
Submission Deadline: 26 June 2025

Contact Email: 10K-Genome_Pilot@mrc.ac.za

1. Introduction

The SAMRC, in collaboration with the DSTI, invites applications for the submission of existing longitudinal patient cohort samples to be considered for inclusion in the pilot phase of the South African 110,000 Human Genome Programme (SA110K-HGP).

This pilot phase will aim to sequence 10,000 human genomes over a two-year period using Next Generation Sequencing (NGS) technologies. The initiative forms part of a broader national strategy to build sustainable infrastructure for precision medicine, genomics research, and healthcare innovation. Therefore, this pilot programme does not just aim to build sequencing capacity but rather to develop a clinical genomics ecosystem. This can be achieved by embedding WGS into trials, the programme will test how genomic data can be used to improve care and research in African populations, with real health needs within the region. The selected samples will be foundational in validating protocols, establishing cost structures, optimising workflows, and demonstrating feasibility for nationwide scale-up toward a national population genome programme.

Applications are welcome from research institutions, hospitals, and clinical trial networks for the submission of existing longitudinal patient cohorts to conduct Whole Genome Sequencing (WGS) at 30x coverage. This pilot initiative aims to sequence 10,000 samples using Next Generation Sequencing (NGS) technology to establish a coordinated federated ecosystem within South Africa capable of supporting precision medicine, improving clinical outcomes, and advancing understanding of disease mechanisms and therapeutic responses. Applicants should clearly demonstrate how genomic insights from the proposed cohorts will support

improved diagnostics, therapeutic stratification, or elucidation of disease etiology in existing or planned clinical studies.

We are not simply looking for patient samples; rather, we aim to build a national system for precision medicine. Your cohort must add scientific value, help test infrastructure, contribute data to a national archive, and show how it can improve patient outcomes. Successful applicants must be ready to explain how data generated on their patient samples will be used!

2. Objectives

This RFA aims to address the following scope and objectives:

2.1. To identify diverse patient cohorts for WGS at 30x coverage.
2.2. To utilize genomic data to address critical gaps in disease understanding and therapeutic response.
2.3. To integrate genomic insights into clinical trial studies for improved patient outcomes.
2.4. To build national genomic resources that inform the scalable workflow of a larger scale National Genome Programme.
2.5. To contribute data to a National Archive for enhanced analytics and broader population-based impact.
2.6. To promote tertiary analyses that extract actionable insights from genomic data and strengthen the scientific contributions of this pilot phase of the SA-110K-HGP.

3. How will these selected cohorts integrate into the South African 110, 000 Human Genome Programme Workflow

Selected cohorts will be included in the pilot phase to enable a centralised whole genome sequencing (WGS) workflow. South African laboratories with the necessary infrastructure are being positioned to utilize their core sequencing platforms for this pilot, contributing to the development of a harmonized workflow across all sites. This will ensure the generation of high-quality data from both Illumina and MGI Next Generation Sequencing (NGS) platforms.

Data generated at these central laboratories will be processed using standardised pipelines and archived in a central National Genome Archive. Additionally, a separate phenotype data registry will be established to collect and harmonize metadata across various cohorts.

This approach will support the pilot phase and lay the groundwork for a large-scale population genome programme in South Africa.

4. Scope of Required Samples

Applicants must propose a minimum of 500 samples per cohort and max 2500 for submission. Each sample must meet the following:

- Be blood-derived, with sufficient volume for DNA extraction and sequencing.
- Accompanied by a minimum set of de-identified phenotype/clinical metadata linked to each sample. Pertinent aspects such as (age, sex, ethnicity and main disease aspect), this type of information is necessary to provide some level of sample differentiation within a genome Archive.
- Have traceability and documentation from the point of collection to current storage.
- Be of high integrity and stored under appropriate conditions for genomic analysis.

5. Eligibility Criteria

The following eligibility criteria sets the minimum requirement for participation:

- Be a recognised academic institution, clinical research centre, hospital, or clinical trial institute or network based in South Africa.
- Submit an existing or planned cohort that is part of an ethically approved clinical or research study, with valid participant consent for genetic studies such as Whole Genome Sequencing (WGS) and data sharing.
- Ethical approval must be from a registered South African Human Research Ethics Committee.
- Applicants must agree to deposit resulting genomic and associated metadata into the National Genomics Archive for integration, analysis, and future reuse, subject to appropriate governance and access controls.
- Commitment to comply with standardized sample preparation and DNA extraction protocols.

6. Selection Criteria for Patient Cohorts

6.1 Scientific & Clinical Relevance

All proposed cohorts should focus on diseases or clinical conditions where WGS is expected to produce scientifically or clinically meaningful insights. While South African cohorts are broadly underrepresented in global genomic databases, proposals will be assessed based on:

- The strength of the scientific rationale for sequencing (e.g., potential to identify disease mechanisms, novel variants, or biomarkers).
- The clinical relevance of the cohort, including potential for WGS to improve diagnostics, treatment stratification, or care.
- The readiness of the study to produce results that inform downstream research, clinical trials, or public health interventions
- Proposals demonstrating potential to inform policy or drive scientific breakthroughs are encouraged.

6.2. Integration with Clinical Trials

Clinical trials often have structured, high-quality data, demographics, lab results, treatment response, adverse events, etc. This enhances the power of genomic analyses, enabling (i)Genotype-phenotype associations; (ii)Biomarker discovery; (iii) Validation of disease mechanisms.

- Cohorts should be embedded in active or planned clinical trials. Applicants must define having WGS tied to real-world patient outcomes with the potential for translation of genomic findings into actionable insights to improve trial outcomes, refine interventions or perhaps support regulatory submissions.
- WGS use must be justified in terms of improving clinical study outcomes with already collected metadata such as treatment response, biomarker discovery, or patient management.

6.3 Diversity & Representativeness

The pilot phase has a main focus on technical feasibility, cohort maturity, and ethical compliance. Genomics research has historically underrepresented African populations. Therefore, preference will be given to cohorts that contribute to the geographic, ethnic, and clinical diversity of the South African population. While not mandatory, such diversity will strengthen the programme's ability to generate findings that are broadly applicable and equitable. Applicants are encouraged to describe how their cohort enhances the overall representativeness of the pilot data set.

6.4 Data Quality & Feasibility

Applicants must demonstrate that their cohort has organized, high-quality clinical data already collected and available. Provide evidence of well-curated clinical trial datasets with phenotype metadata. This may include:

- Patient demographics (age, sex, ethnicity, etc).
- Clinical characteristics (diagnoses, treatment regimens, outcomes).
- Lab results, imaging, or other phenotype indicators.
- Ensure biospecimen protocols support high-quality DNA extraction for WGS. Whole blood is preferred for good quality DNA.

6.5 Ethical Considerations & Data Governance

All selected cohorts will be required to deposit de-identified genomic data (FASTQ, BAM/CRAM, and VCF formats) along with associated metadata into the South African National Genomics Archive and a separate Phenotype registry. This archive will serve as a centralized, secure, and governed repository that facilitates data reuse, supports secondary research, and enables integration across cohorts. Applicants must confirm that appropriate participant consent has been obtained to allow genomic data sharing for research purposes. Data must comply with FAIR (Findable, Accessible, Interoperable, Reusable) principles. Controlled access mechanisms will be implemented to ensure ethical, legal, and secure use of

the data. Cohort PIs will retain primary access rights and contribute to downstream use governance.

All applications must convey:

- Clear plans for informed consent and ethical oversight.
- Compliance with national/international genomic data sharing standards.
- Adherence to FAIR data principles (Findable, Accessible, Interoperable, Reusable).
- Data access committee members for every application.

7. Funding & Support

- The programme will centrally procure sequencing services to ensure competitive pricing on Illumina and MGI platforms, within South Africa.
- Therefore, funding allocations will support sample transport, minimum personnel, and tertiary bioinformatics analysis.
- Proposals must include a plan for genotype-phenotype association studies with clearly defined scientific objectives.
- Funding amounts will be determined based on cohort size, disease relevance, and scientific merit.
- Sequencing will take place at local sequencing centres; data processing will follow a centralised pipeline within the central programme workflow and data will be processed and stored within a National Archive.

Therefore, successful applicants will work with the programme infrastructure team to refine budgets and finalise their applications.

8. Application Process

1. Cohort description (disease focus, number of participants, demographics).
2. Justification for WGS (clinical utility, trial integration, outcomes).
3. Patient inclusion and consent strategy.
4. Clinical and genomic data management plan.
5. Minimum budget outline and resource needs.

Submission Deadline: **26 June 2025**

Review Process: Applications will be evaluated by an independent panel of genomic and clinical experts.

7. Contact Information

For inquiries regarding this funding call, please contact:

Rizwana Mia (Senior Programme Manager – South African 110K Human Genome Programme)
Email: 10K-Genome_Pilot@mrc.ac.za