

SOUTH AFRICAN 110,000 Human Genome Programme

SA110K-HGP

The Pilot Phase

Building South Africa's clinical genomics and big-data-in-health ecosystem –
A model for Africa

10

Deeply phenotyped
cohorts

10 000

Patients being
evaluated

30×

Whole genome
sequencing

4

federated sequencing
laboratories

2

Staggered phases
to scale to 10K



TABLE OF CONTENTS

ABOUT THE PROGRAMME 3

WHY SOUTH AFRICA NEEDS THE SA110K-HGP 4

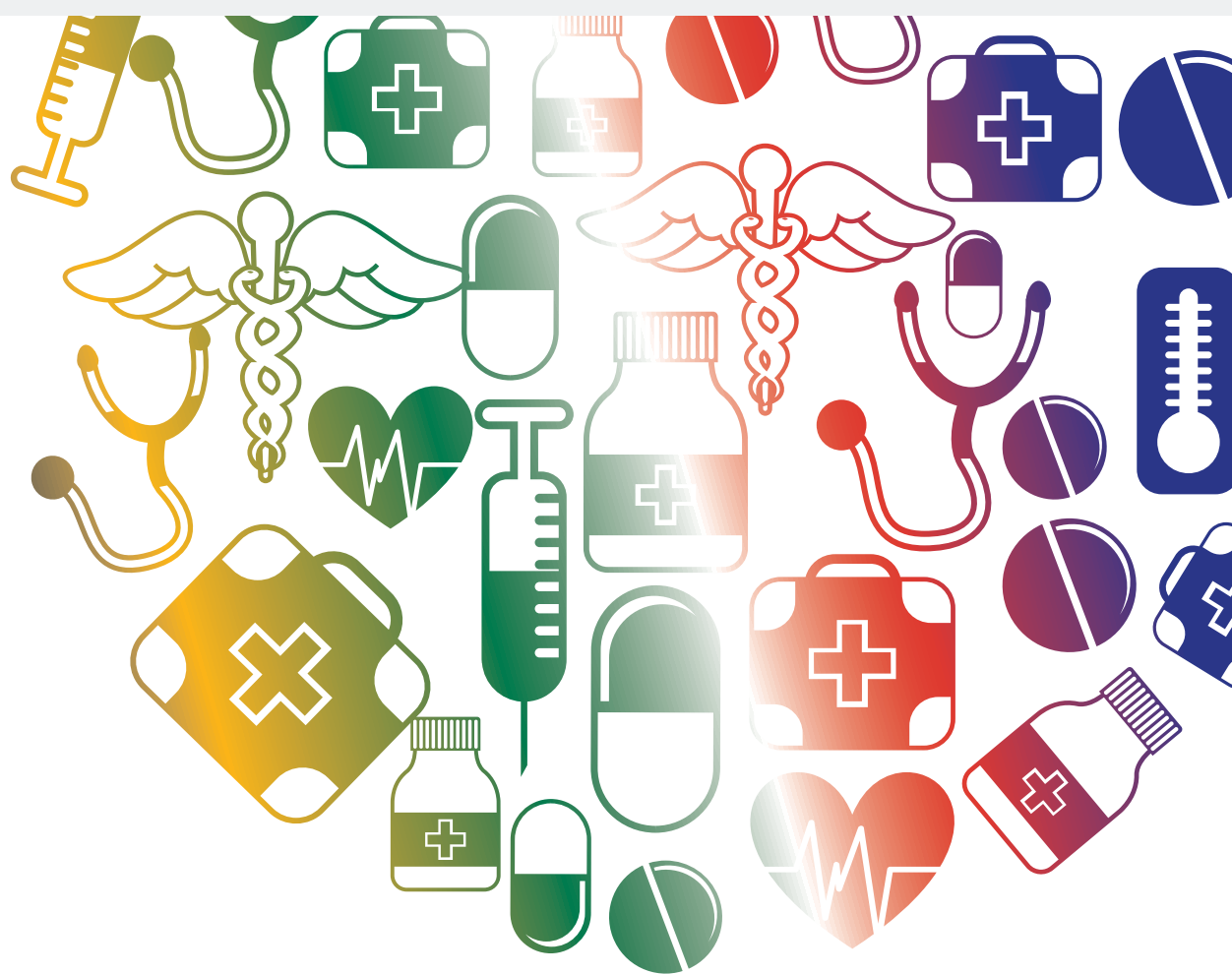
A CONNECTED NATIONAL ECOSYSTEM..... 5

HOW THE COHORTS WERE SELECTED 6

THE TEN COHORTS 7

CONCLUSION: A BIG-DATA-IN-HEALTH ECOSYSTEM FOR AFRICA 10

ABOUT US..... 11



ABOUT THE PROGRAMME

The South African Medical Research Council (SAMRC), in partnership with the Department of Science, Technology and Innovation (DSTI), is implementing the pilot phase of the South African 110,000 Human Genome Programme (SA110K-HGP). The pilot, known as the SA110K-HGP Pilot Phase, aims to sequence 10,000 human genomes using next-generation whole genome sequencing at 30× coverage.

The SA110K-HGP Pilot Phase is the first stage of a two-phase national strategy to build sustainable infrastructure for precision medicine, genomics research and healthcare innovation. Its purpose is to prove, at national scale, that South Africa can run genomics end to end: sound governance, ethics and regulation, harmonised laboratory and data workflows, the feasibility and evidence needed to move to scaled sequencing on established national platforms.

Programme Objectives

- Sequence the whole genomes (30×) of participants in diverse, well-characterised South African patient cohorts.
- Use genomic data to close critical gaps in disease understanding and therapeutic response for conditions of high relevance to South African and African populations.
- Integrate genomic insights into clinical trials and clinical services to improve patient outcomes.
- Build national genomic resources and scalable workflows that inform the larger 100K phase of the National Genome Programme.
- Contribute data to a National Genome–Phenome Archive for enhanced analytics and broader population impact.
- Promote tertiary analyses that extract actionable insights and strengthen the scientific contribution of the pilot.

THE ESSENCE OF THE SA110K-HGP PILOT PHASE

The SA110K-HGP Pilot Phase does not start from scratch. It sequences participants from existing, longitudinal, deeply phenotyped clinical cohorts, so every genome arrives already linked to years of clinical, laboratory and outcome data. Each genome therefore does double duty: it answers a specific disease question for its cohort, and it builds a shared national genome–phenome resource for South Africa. It is a disease-focused clinical-cohort resource, not a population-representative reference genome.

WHY SOUTH AFRICA NEEDS THE SA110K-HGP

For most of modern medicine, healthcare has been built around averages. Patients with the same diagnosis are offered the same treatment, even though individual responses vary significantly: some people experience substantial benefits, others little to none, and some may have adverse drug reactions. Precision medicine recognises that health and disease are shaped by a unique combination of genetic, environmental, behavioural and social factors.

The right intervention, for the right person, at the right time.



EVIDENCE BUILT ELSEWHERE

Much of the science behind genomic tools, diagnostic tests and prediction algorithms comes from populations outside Africa. Clinicians here often work with evidence generated elsewhere, which limits variant interpretation and the accuracy of risk models for African patients. Without local data, the benefits of precision medicine will not be shared equally.



UNMATCHED GENETIC DIVERSITY

Africa holds the greatest human genetic diversity on earth. South Africa is a melting pot of that diversity, home to people from across sub-Saharan Africa as well as communities of European, Middle Eastern and Asian heritage, with strong founder effects. A resource built here can have impact far beyond South Africa's borders.



A COMPLEX DISEASE BURDEN

South Africa carries a quadruple burden of disease – HIV and TB, rising non-communicable disease, maternal and child health challenges, and injury and violence – increasingly complicated by multimorbidity. Overlapping conditions delay diagnosis, and adverse drug reactions add cost and harm. Precision medicine offers more accurate diagnoses and better-targeted therapies.



GENES, ENVIRONMENT AND LIVED EXPERIENCE

Health outcomes emerge from the interplay of biology with poverty, nutrition, housing, stress and access to care. Integrating genomic data with clinical, environmental and population-level data reveals distinct biological subtypes hidden within single diagnoses such as diabetes, hypertension and cancer – the basis for earlier prevention and better-matched care.

Built on a decade of investment

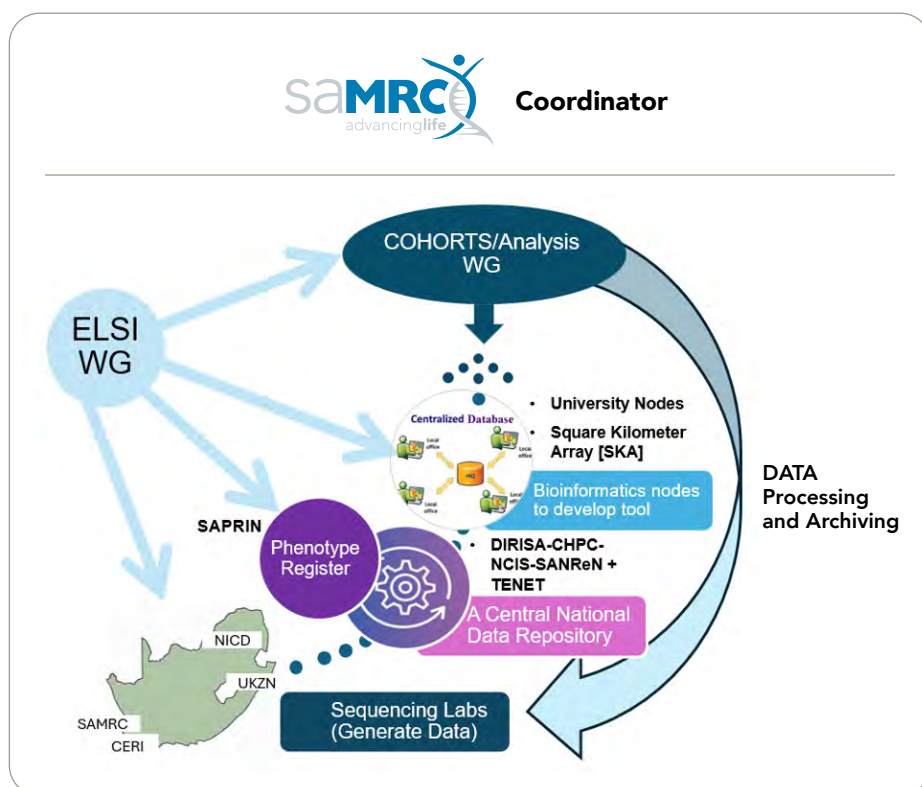
South Africa is not starting from scratch. Over ten years, government, funders and scientific institutions have invested in genomics, bioinformatics, pharmacogenomics, public health genomics and international collaboration. Milestones include the DSTI Precision Medicine Platform; collaborative funding partnerships with the United Kingdom, Europe, China and Japan; growing pharmacogenomics research programmes; and local sequencing capability at the SAMRC Genomics Centre. During COVID-19, the Network for Genomic Surveillance in South Africa showed that genomics is not only a research tool but a real-time public health capability.

The SA110K-HGP also leverages the national big-data infrastructure that government built for data-intensive programmes such as the Square Kilometre Array (SKA). The lesson of the past decade is that precision medicine cannot be built through isolated projects. The SA110K-HGP is the natural next step: connecting these components into one functioning, nationally governed system – harmonising ten multi-disease, deeply phenotyped cohorts onto a central phenotype registry and establishing a National Genome Archive for genomic data generated on South African populations.

The goal is not merely to generate genomic data. The goal is to ensure that genomic information becomes clinically useful and contributes to better healthcare.

A CONNECTED NATIONAL ECOSYSTEM

SAMRC coordinates a single ecosystem that links cohorts, ethics oversight, sequencing laboratories, national cyber-infrastructure and bioinformatics teams. Rather than building new infrastructure, the programme deliberately leverages South Africa's National Integrated Cyber Infrastructure System (NICIS) – the Centre for High Performance Computing (CHPC), the Data Intensive Research Initiative of South Africa (DIRISA), and the SANReN and TENET research networks – together with a shared phenotype registry developed with the South African Population Research Infrastructure Network (SAPRIN). Lab infrastructure was substantially upgraded for scalability over the last two years to facilitate the programme's needs.



1

COHORT READINESS

Samples,
consent and
phenotype meta-
data confirmed

2

SEQUENCING

WGS at NICD,
CERI, SAMRC
Genomics Centre
and UKZN

3

PROCESSING

Agreed bioinformatics pipelines and quality control

4

**NATIONAL
ARCHIVE**

Secure storage
via NICIS (CHPC
/ DIRISA), moved
over SANReN with
Globus Endpoint

5

ANALYSIS

Governed strictly
at cohort-level
and cross-cohort
analysis

Governance Structure

Scientific Advisory Board	Independent scientific oversight, strategic guidance and international benchmarking.
Operational Management Committee	Operational management, coordination and technical oversight of the programme.
Ethics Working Group	Ethics validation, benchmarking and governance readiness; a shared ethical framework for cohort inclusion and data sharing.
Data Working Group	Central data workflow and archive: DIRISA storage, CHPC compute, SANReN connectivity, pipelines and phenotype harmonisation.
Laboratory Working Group	SOPs, workflow harmonisation, external quality assessment and sequencing validation across the four federated laboratories.
Cohort PIs / Analysis Working Group	Cohort and cross-cohort analysis, tertiary data use and joint publication.

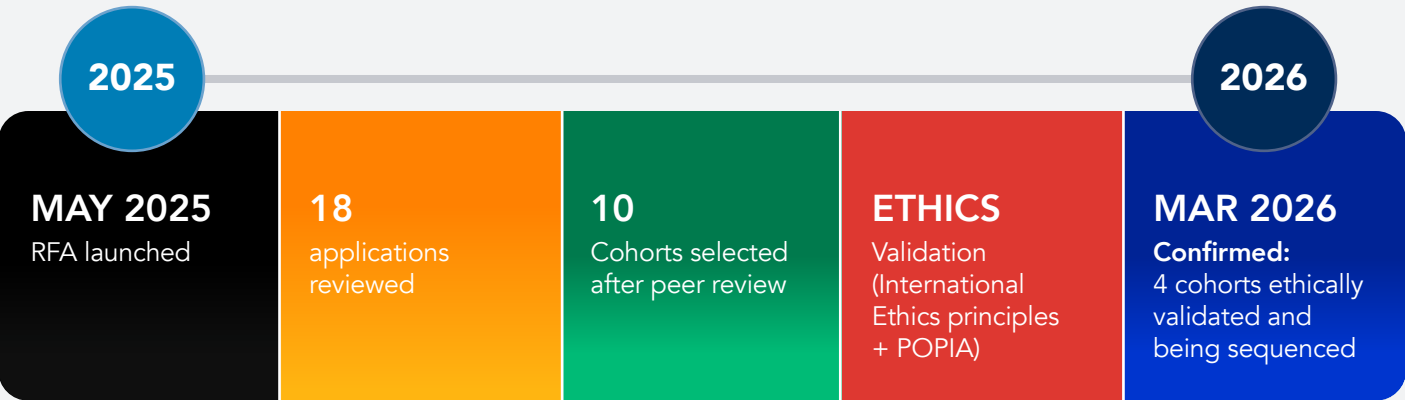
Each participating cohort nominates representatives to the Ethics, Data and Laboratory Working Groups, so cohorts help shape the programme rather than simply supply samples to it.

HOW THE COHORTS WERE SELECTED

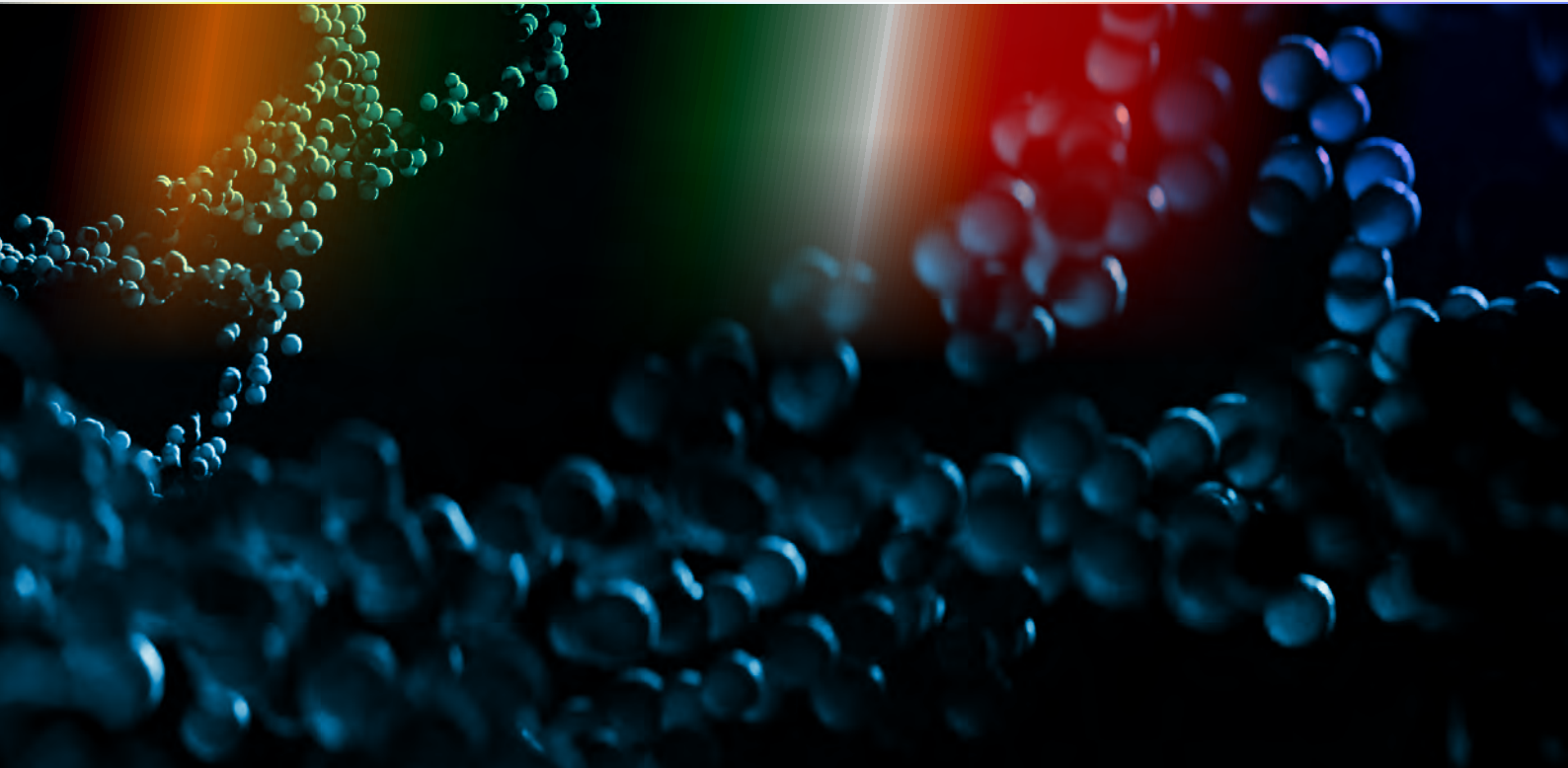
Following a public call in May 2025, 18 applications were assessed by independent local and international experts for scientific and clinical relevance, ethics and governance, and readiness for data harmonisation. Shortlisted cohorts also underwent a structured ethics review aligned with international principles and South Africa’s POPIA Act. Ten ethically ready cohorts were selected for phased inclusion: Four cohorts are proceeding to sequencing now, while six are provisionally approved, subject to ethics approval

and funding. All ten are participating in national phenotype data harmonisation.

Shortlisted cohorts then passed through a dedicated three-step ethics validation — a theoretical framework, a benchmarking tool, and a scoring and classification system — benchmarked against International Ethics principles and POPIA. In March 2026 the SAMRC Executive Management Committee used the combined scores to rank the ten selected cohorts and set their phased inclusion.



Phase 1 cohorts proceed to sequencing now. The remaining six are provisionally approved for Phase 2, subject to ethics approvals and funding. All ten cohorts are taking part in national phenotype data harmonisation.



THE TEN COHORTS

Each cohort was chosen because it brings a question that only African genomes can answer, and the clinical depth to answer it. Together they span South Africa's major burdens of communicable and non-communicable disease, across the life course.

GENARK-SA

Chronic kidney disease

Dr June Fabian • University of the Witwatersrand

WHO IS IN THE COHORT

Rural Black South African adults aged 20–80 in the Agincourt Health and Demographic Surveillance System, Mpumalanga. More than 90% speak Xitsonga and about a third have Mozambican migrant heritage. Existing GWAS array, APOL1 genotyping, and metabolomic and proteomic data on a deeply phenotyped subset.

WHY IT MATTERS

CKD is highly prevalent in this community, yet risk pathways defined in other populations may miss African-specific variants. GENARK-SA will map the genetic contributors to CKD and its links with hypertension, diabetes, cardiovascular risk and HIV, informing population-appropriate screening and the relevance of emerging APOL1-targeted therapies.

Undiagnosed Disease Programme (UDP)

Rare and undiagnosed disease in children

Prof Shahida Moosa • Stellenbosch University

WHO IS IN THE COHORT

Patients from Tygerberg Hospital and outreach centres, 80% under 17 years, largely Afrikaans- and isiXhosa-speaking, from the Western, Eastern and Northern Cape and beyond. All remain undiagnosed after exome or panel testing and have HPO-coded phenotypes and broad data-sharing consent.

WHY IT MATTERS

For families on a long diagnostic odyssey, a genome may be the last test available. WGS detects variant types that exomes miss, is expected to lift the current ~40% post-exome yield, and will uncover novel disease genes in one of sub-Saharan Africa's first deeply phenotyped rare-disease cohorts. Actionable results are returned to families.

AFRI-CODE

Hypertension and pre-eclampsia pharmacogenomics

Prof Collet Dandara • University of Cape Town

WHO IS IN THE COHORT

African and mixed-ancestry adults from Groote Schuur and Mbombela hospitals, spanning resistant and controlled hypertension, pre-eclampsia and normotensive controls, with up to five years of follow-up on blood pressure, treatment and pregnancy outcomes.

WHY IT MATTERS

Hypertension is one of Africa's largest disease burdens, but its genetic basis in African populations is poorly understood. AFRI-CODE seeks the factors that separate women who develop pre-eclampsia from those who do not, and pharmacogenomic markers of treatment response – towards biomarkers for risk prediction and better-targeted therapy.

THRIVE-SA

Drug-resistant TB and HIV pharmacogenomics

A/Prof Phumla Sinxadi • University of Cape Town

WHO IS IN THE COHORT

Participants from the SHIFT-TB cohort and the BEAT-TB phase 3 trial (650 enrolled) at Groote Schuur, East London, Khayelitsha and Durban; predominantly Black African, with or without HIV, followed for 18 months with detailed drug-level, efficacy and adverse-reaction data.

WHY IT MATTERS

Rifampicin-resistant TB is treated with long, toxic regimens still dosed one-size-fits-all. By linking genomes to drug exposure and toxicity for bedaquiline, linezolid, clofazimine and delamanid — including cardiac QT prolongation — THRIVE-SA aims to predict who is at risk before treatment starts and to inform African pharmacogenetic dosing guidelines.

GENE-HT (CAPRISA / FRESH / DREAMS)

Host genomics of HIV and TB

Prof Veron Ramsuran • University of KwaZulu-Natal

WHO IS IN THE COHORT

Young Black African women in high HIV-burden districts of KwaZulu-Natal and Gauteng, drawn from the long-running CAPRISA studies and the FRESH and DREAMS cohorts. FRESH captured HIV infection at its very earliest stages through intensive testing.

WHY IT MATTERS

A genome-wide study in DREAMS found a previously unseen chromosome 1 variant associated with strong HIV control in the Zulu population — using only ~0.1% of the genome. Whole genomes can reveal far more: protective variants, determinants of acquisition and progression, ART response, and leads for HIV cure and vaccine research.

HeLTI-SA

Maternal, child and life-course health

Prof Shane Norris • University of the Witwatersrand

WHO IS IN THE COHORT

Women aged 18–28 in Soweto from the Healthy Life Trajectories Initiative trial (7,735 randomised from ~38,600 households approached), with data from preconception, through pregnancy, into childhood. A WHO-coordinated trial with harmonised data dictionaries.

WHY IT MATTERS

By age 22, half of these women already live with at least one chronic condition. HeLTI-SA will test whether ancestry-informed polygenic risk scores improve prediction of maternal cardiometabolic and mental health, pregnancy and early-childhood outcomes, and how genes and environment interact across generations.

DIAMOND

Paediatric monogenic disorders

Prof Zané Lombard • University of the Witwatersrand

WHO IS IN THE COHORT

Four sub-studies — DDD-Africa developmental disorders, Wits and UKZN clinical services, CHAMPS stillbirth and neonatal deaths, and Wits PeCan paediatric cancer — with unsolved probands and parents. Over 90% of African descent; HPO-coded, harmonised phenotypes.

WHY IT MATTERS

DIAMOND is generating the evidence for WGS as a first-line diagnostic in South Africa's public health system. Outputs include a curated catalogue of disease-causing variants for global databases, locally built interpretation pipelines, and evidence to guide future newborn screening and carrier testing.

RICE

Mental health (PTSD) after trauma

Dr Jani Nöthling • SAMRC Gender and Health Research Unit

WHO IS IN THE COHORT

isiZulu-speaking women in KwaZulu-Natal — rape survivors and a comparison group — followed for 24 months with neuropsychiatric assessments, cardiometabolic biomarkers and stored DNA, serum and plasma.

WHY IT MATTERS

African women are disproportionately affected by trauma yet almost absent from psychiatric genomics. RICE will map genomic predictors of PTSD and depression trajectories and their links to physical health, supporting culturally relevant risk prediction and targeted mental-health interventions for local clinics.

ACEI-Angioedema & IMARI-Africa

Severe adverse drug reactions

Prof Jonny Peter • UCT Lung Institute, University of Cape Town

WHO IS IN THE COHORT

Two case-control cohorts from the Western and Eastern Cape: ~250 ACE-inhibitor angioedema cases and ~300 severe cutaneous drug-reaction cases (mostly people living with HIV on TB drugs, ARVs or co-trimoxazole), each matched about 2:1 with drug-tolerant controls. Multi-omic proteomic data available.

WHY IT MATTERS

Some HLA risk markers for drug reactions carry very large effects, so even modest cohorts can find them. Markers that tell clinicians which HIV/TB patients can be safely re-challenged with a drug would save lives, time and scarce health-system resources.

Johannesburg Heart Failure Cohort

Heart failure and cardiomyopathy

Prof Nqoba Tsabedze • University of the Witwatersrand

WHO IS IN THE COHORT

Predominantly Black African patients from Charlotte Maxeke Johannesburg Academic Hospital, drawn from across South Africa and neighbouring countries, with reduced-EF, acute, dilated and peripartum heart failure. Patients are roughly a decade younger than in Western cohorts; median follow-up 36 months.

WHY IT MATTERS

Screening against European-derived dilated cardiomyopathy gene panels gave a very low diagnostic yield in this population. WGS will define African-specific and structural variants, pharmacogenomic responses to therapy and genotype-guided risk of sudden cardiac death — enabling family cascade screening and an African heart-failure reference.

CONCLUSION: A BIG-DATA-IN-HEALTH ECOSYSTEM FOR AFRICA

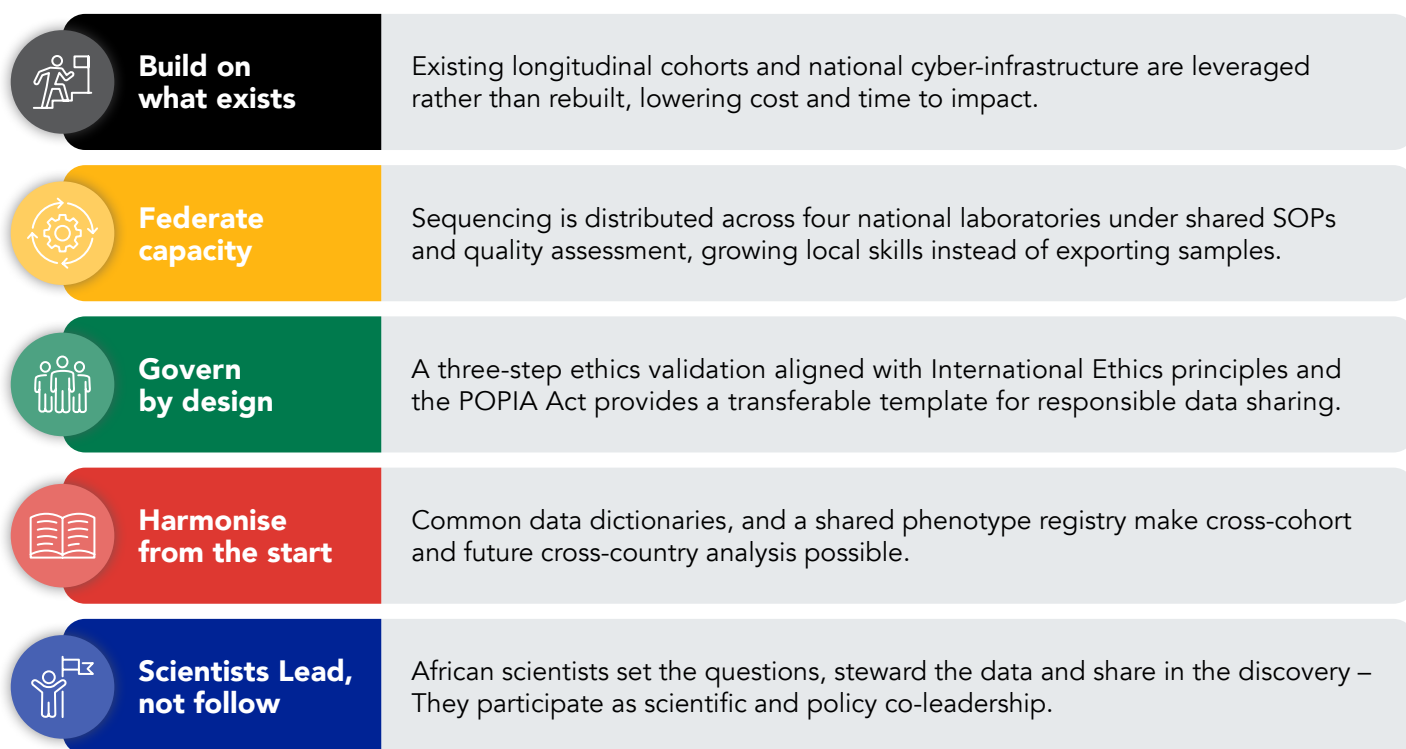
A model for low- and middle-income countries in Africa

The SA110K-HGP Pilot Phase is more than a 10,000-genome sequencing effort. It is using South Africa's existing research infrastructure as the foundation for a coordinated, scalable genomics workflow. Existing high-bandwidth research networks and national high-performance computing and data-storage capacity are being adapted to support secure data transfer, processing and archiving, and to establish the national phenotype registry and National Genome Archive.

The pilot is connecting these capabilities with a federated network of sequencing laboratories and a

shared ethics and governance framework. Together, these components will create a governed pathway from participant consent and deep phenotyping through sequencing, data linkage and analysis.

By linking genomic data with longitudinal clinical, laboratory, pharmacological and social data, the pilot will lay the groundwork for a national genome-phenome resource. This resource can support research and, over time, inform diagnosis, drug safety, risk prediction and health planning. The workflow is designed to scale towards the full 110,000-genome programme.



SA110K-HGP is set to establish a big-data-in-health ecosystem that other African low- and middle-income countries can adopt and adapt – demonstrating that the continent can generate, govern and use its own genomic data to deliver precision medicine and health equity for its people.

ABOUT US

The South African Medical Research Council (SAMRC)

Established in 1969, the SAMRC is dedicated to improving the health of people in South Africa through research, innovation, development and technology transfer. Its research spans laboratory investigations, clinical research and public health studies, addressing South Africa's quadruple burden of disease: maternal, newborn and child health; HIV/AIDS and TB; non-communicable diseases; and interpersonal violence. Its work generates evidence to inform health policy and practice and improve the health status of South Africans.

The SAMRC is the largest local funder of health research, medical diagnostics, medical devices and therapeutics, and a pioneer of cutting-edge medical innovation in genomic research, novel treatment regimens, vaccines, diagnostic tools, and new drugs and devices. Transformation is integral to its mission: through Self-Initiated Research grants, the Mid-Career Scientist Programme, the Bongani Mayosi National Health Scholars Programme and other platforms, the SAMRC continues to address gender, racial, institutional and geographic parity in health research capacity.



VISION

Building a healthy nation through research, innovation and transformation.



MISSION

To advance the nation's health and quality of life and address inequity by conducting and funding relevant and responsive health research, capacity development, innovation and research translation.



FOCUS

The top ten causes of death and disability and their risk factors, and how health systems function – giving policymakers the tools for informed healthcare decisions.

STRATEGIC HEALTH INNOVATION PARTNERSHIPS (SHIP) PROGRAM

SHIP is a partnership between the SAMRC and the Department of Science, Technology and Innovation (DSTI). It funds health innovation projects focused on the development of new drugs, treatments, vaccines, medical devices and prevention strategies, delivering life-saving innovations to the health industry and the South African public through new and improved treatment options.



PROGRAMME LEAD

Rizwana Mia

Programme Lead: South African 110,000 Human Genome Programme (SA110K-HGP)

Senior Programme Manager:

South African Precision Medicine Programme
Grants, Innovation and Product Development (GIPD) Unit,
South African Medical Research Council

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

Web: www.samrc.ac.za

