UNIVERSITY *of the*
WESTERN CAPE

RESPICE PROSPICE

SANBI
South African National
Bioinformatics Institute

SOUTH AFRICAN NATIONAL BIOINFORMATICS INSTITUTE

ANNUAL REPORT 2024

CONTENTS

ABOUT SANBI

Bioinformatics is a specialist discipline straddling the fields of biology, mathematics and computer sciences and it is integral to modern biological research. The South African National Bioinformatics Institute (SANBI) is situated at the University of the Western Cape (UWC) in Bellville, Cape Town.

WHO WE ARE

Our primary focus is the development and implementation of computational methodologies that allow biomedical researchers to accelerate their genomics data analyses. SANBI aims to heighten awareness of bioinformatics in South Africa and to assist the country in making optimal use of bioinformatics tools. As the leading bioinformatics entity in Africa, we continue to foster local and regional collaborations on health-related topics that cover both communicable and non- communicable diseases.

SANBI provides a focus for biological research located in Africa and as such, is dedicated to:

- the development of online specialised resources for genomics and genome informatics;
- capacity development in genomics and bioinformatics in Africa; and the development and implementation of genome annotation methods.

OUR VISION

- To be a global leader in computational biology, achieving the highest level in biomedical research and education in the global, African and South African context.

OUR MISSION

- To conduct cutting edge bioinformatics and computational biology research relevant to South African, African and global populations.
- To develop human resources in bioinformatics and computational biology by educating and mentoring scientists.
- To increase awareness of, and access to, bioinformatics and computational biology resources.

OUR GOALS

- To generate and publish high quality, relevant biomedical research.
- To train and graduate competent and productive researchers.
- To add value to the academic programme of UWC.
- To enhance other research fields through collaborative projects.
- To establish sources of renewable funding to pursue the mission of the institute.

POLICY MANDATES

5th National Strategic Plan (NSP) for HIV, TB and STIs (2023 - 2028)

The vision and mission of SANBI align with the NSP 2023 - 2028 that outlines how the country will respond to the prevention and treatment of HIV, TB and Sexually transmitted infections (STIs). The NSP aims to *"...build resilient systems for HIV, TB and STIs that are integrated into systems for health, social protection and pandemic response...".*

National Research Foundation (NRF) Strategy 2025

The vision and mission of SANBI aligns with the NRF's Strategy 2025, specifically *"promoting global and competitive research and innovation".*

Research and Development aligns with the Department of Science and Innovation White Paper on Science, Technology and Innovation 2018 - 2028 by contributing to *"the development of human capabilities, knowledge expansion and innovation performance...".*

The SA Medical Research Council (SAMRC) Act (Act 58 of 1991)

As an extramural unit of the SAMRC, SANBI falls under the legislative mandates of the SAMRC. At Section 3, this Act states that the Legislative Mandate of the SAMRC is: *"...through research, development and technology transfer, to promote the improvement of the health and quality of life of the population of the Republic, and to perform such functions as may be assigned to the SAMRC by or under this Act..."*

DIRECTOR'S MESSAGE

Professor Alan Christoffels
PhD, M.ASSAf
Director & DSI/NRF Research Chair in Bioinformatics and Health Genomics
Director of SAMRC Bioinformatics Unit

South African National Bioinformatics Institute
University of the Western Cape

The research and innovation space continues to be filled with opportunities despite the financial constraints placed on universities and public health entities. SANBI-UWC has navigated this space with key decisions in 2024 to embark on strategic projects that strengthen our partnerships with the public health sector, industry and other academic institutions.

During 2024, we have extended our projects related to non-communicable and communicable diseases as reflected in the research updates and outputs. Excitingly we have expanded our partnerships in South America through a diabetes data modeling project. And our role in infectious disease bioinformatics continues to grow.

Funding through the US federal government continue to be at risk as we await clarity on those projects. Nevertheless, resilience needs to shape our thinking as we traverse the demands of 2025.

YEAR IN REVIEW

2024 HIGHLIGHTS

Peer reviewed publications include 19 journal publications and 1 book chapter. Two female PhD students graduated in 2024 and and they are working as postdocs or as faculty at the university.

We have supported international exposure to many of our students and staff who have presented the following:

- 19 conference talks
- 5 media articles
- 13 training courses
- 2 mentorships

RESEARCH PROJECT HIGHLIGHTS:

- Development of a virtual genotyped cohort to study drivers of multimorbidity in WCGHW healthcare clients (VCAMM study, PI N Tiffin)

- Development of bioinformatics for public health solutions to support pandemic preparedness (PHA4GE, PI A Christoffels)

- Modeling information flow in public health labs (PI P van Heusden)

- Development of data governance and ethical tools and guidelines (PI N Tiffin)

- Development of databases for drug discovery targets (PI R Cloete)

- Data standards for Cholera (PI D Anderson)

- Establishment of an Africa data modelling network to better understand the key drivers of gestational diabetes and pre-eclampsia (B&MGF, PI N Tiffin)
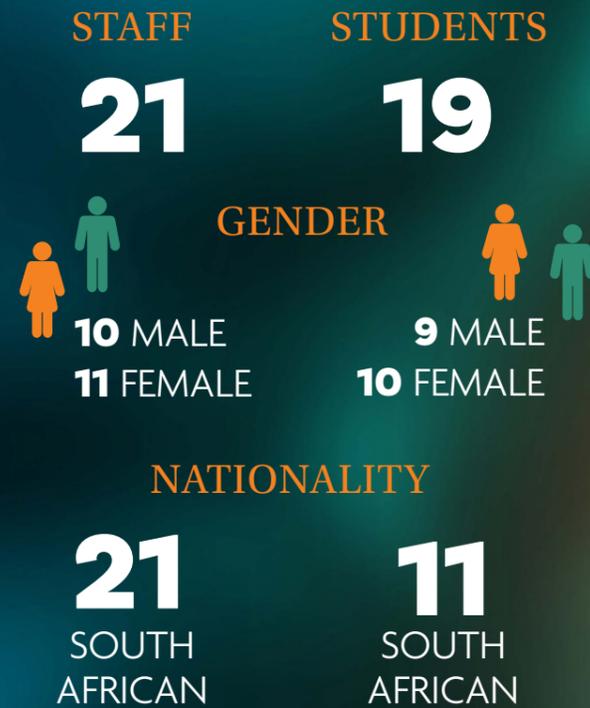
2024 IN SUMMARY

INCOME

ZAR 51 433 340

EXPENDITURE

ZAR 40 286 266

RESEARCH OUTPUTS

JOURNALS
19

BOOK CHAPTER
1

GRADUATIONS
2

AWARDS/HONOURS
5

MENTORSHIPS
2

KEYNOTES/INVITED TALKS
10

TEACHING & LEARNING
16

CONFERENCE PRESENTATIONS
16

EXPERT PANEL/ COMMITTEE MEMBERSHIP
33

44 COLLABS IN **18** COUNTRIES

RESEARCH PROJECTS **24**

80 ALUMNI IN **20** COUNTRIES

DIVERSITY PROFILE

STAFF
21

STUDENTS
19

GENDER

10 MALE
11 FEMALE

9 MALE
10 FEMALE

NATIONALITY

21 SOUTH AFRICAN

11 SOUTH AFRICAN

EDUCATION LEVEL

POSTDOCTORAL
6

POSTDOCTORAL
3

MSC
5

PHD
7

BSC
1

MSC
9

DIP
9

STAFF

SANBI comprises a highly skilled, diverse group of research scientists, technical and administrative staff who all contribute to a dynamic productive working environment. As SANBI is a research institute, the Director reports through the faculty of Natural Sciences at UWC.

RESEARCH STAFF

SANBI research staff focus on the development of analytical tools to address health challenges facing South Africa and the African continent. Through a series of national and international collaborations we have developed methods and used these methods to gain insight into disease mechanisms or to reduce the barrier to entry for researchers in resource-limited environments.

Alan Christoffels, Prof
SANBI Director
SARChI Bioinformatics
SAMRC Unit Director
*DSI/NRF Research Chairs Programme

Nicki Tiffin, Prof
SANBI Deputy Director

*UWC

Gordon Harkins, Prof
Associate Professor

*UWC

Dominique Anderson, Dr
Senior Researcher

*UWC

Anja Bedeker
Research Associate

*B&MGF

Ruben Cloete, Dr
Senior Lecturer

*UWC

Judit Kumuthini, Dr
Senior Researcher

*H3ABioNetwork

* Funder

TECHNICAL STAFF

SANBI Technical staff develop and maintain the computing infrastructure.

Quinton Coert
Software Developer

*Baobab LIMS

Campbell Rae
Web Developer

*DSI/NRF Research Chairs Programme

Peter van Heusden
Senior Systems Developer

*DSI/NRF Research Chairs Programme

ADMINISTRATIVE STAFF

SANBI administrators ensure the smooth running of daily operations.

Saajidah Beghardien
Administrator

*SAMRC

Fungiwe Mpithi
PA/Administrator

*SAMRC

Ferial Mullins
Finance Administrator

*UWC

Junita Williams
Marketing Administrator (part/time)
*DSI/NRF Research Chairs Programme

PHA4GE SECRETARIAT

The PHA4GE Secretariat hosted at SANBI provide administrative and project management support to the PHA4GE working groups.

Michael Bridger
Business Development
(PHA4GE)

*B&MGF

Rangarirai Matima
Communication Specialist
(PHA4GE)

*B&MGF

Tracey Calvert-Joshua, Dr
Technical Manager
(PHA4GE)

*B&MGF

Jamie Southgate
Communications Officer
(PHA4GE)

*B&MGF

Nawaal Nacerodien-Weitz
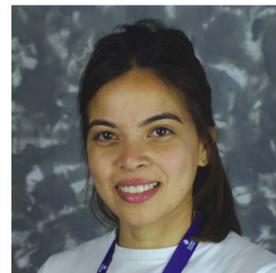Project Manager
(PHA4GE)

*B&MGF

Gabrielle Arendse
Project Administrator
(PHA4GE)

*B&MGF

Keaghan Brown
Training Co-ordinator

*B&MGF

Farzaana Diedericks
Training Co-ordinator

*B&MGF

STAFF AWARDS AND HONOURS

Alan Christoffels won the prestigious Gold Scientific Achievement Award at the SAMRC Merit Awards ceremony held on 7 March 2024. The awards recognise individuals who have demonstrated exceptional scientific acumen and have made innovative strides in addressing public health challenges, potentially influencing policy and enhancing the well-being of the South African population.

SERVICE TO THE SCIENTIFIC COMMUNITY

SANBI academic staff are actively involved in translating research into policy. In 2024, all of the academic staff were appointed on national and international councils, committees and expert panels that develop research frameworks to accelerate data science and biomedical research. Nearly all of SANBI faculty participated as editorial board members or as journal article reviewers. SANBI staff are also actively involved in supporting other institutions of learning through thesis examination and external moderation.

MEMBERSHIP OF EXPERT PANELS AND COMMITTEES

NAME	INSTITUTE	ROLE AND PURPOSE
Alan Christoffels	African Society for Bioinformatics and Computational Biology	President of Governing Council
	Academy of Science of South Africa	Member - Standing Committee on Biosafety and Biosecurity and Standing Committee on POPIA
	Sydney Brenner Institute for Molecular Bioscience, Wits	Research Advisory Committee
	Global Emerging Pathogens Consortium	Member - host an annual conference on emerging pathogens as part of our mandate to create awareness of biosecurity threats across Africa, and to empower African scientists to respond to disease outbreaks.
	Africa CDC	Senior advisor on pathogen genomics
Nicki Tiffin	University of Cape Town, Computational Biology Division	Honorary Professor
	Stellenbosch University, Centre for Bioinformatics and Computational Biology	Member of Governing Board
	International Health Cohorts Consortium	Member - Executive Committee and Steering Committee
	African Population Cohorts Consortium	Independent Advisory Group
	PHA4GE Consortium	Member - Steering Committee; Chair - Ethics and Data Sharing Working Group
	Global Health EDCTP3 Joint Undertaking	Member of the Scientific Committee
	G2MC Policy and Advocacy Working Group/Core Committee	Co-Lead of working group
	WHO Guiding Principles for Pathogen Genome Data Sharing	Panel of experts
	Wellcome Pathogen Genomics Data Sharing: Technical Solutions Workshops	Workshop Participant and Expert Reviewer
	WHO Attributes and Principles for pathogen genomic data sharing platforms	Panel of experts
	UKRI MRC African Research Leaders Programme	Reviewer
	UKRI Funding Applications	Reviewer

Dominique Anderson	Academy of Science of South Africa	Standing Committee on POPIA - to guide the development of the POPIA Code of Conduct for Research
	National Scientific R Collections Platform	Medical Biobanks Working Group Co-Chair
	PHA4GE Consortium	Data Curator
	UWC Technology Transfer Office	Seed Fund Committee
	African Society for Bioinformatics and Computational biology	Conference program committee member
	UWC Science Teaching and Learning	Committee member
	Seq-Africa technical and strategic committee	Committee member
Anja Bedeker	PHA4GE Consortium	Member of the Steering Committee and Co-Chair of the Ethics and Data Sharing Working Group
	Global Alliance for Genomics and Health (GA4GH)	Member - Regulatory & Ethics Work Stream Product lead: Ethical Preparedness for Pandemics and Epidemics (EPPE) Framework
	Public Policy Projects	Panelist - Global Genomics 2024 Roundtable Series
Ruben Cloete	South African Society for Bioinformatics	Member
	African Society for Bioinformatics and Computational biology	Program Committee
	Biophysical Society	Member
Gordon Harkins	Centre for the AIDS Programme of Research in South Africa (CAPRISA)	Appointed Director of the Centre of Excellence in HIV Prevention at UWC
Peter van Heusden	PHA4GE Consortium	Infrastructure Working Group
	Galaxy Project	Intergalactic Utilities Commission

FURTHER DEVELOPMENT OF STAFF

Throughout the year, SANBI staff are encouraged to update their skills by attending relevant training interventions. In April 2024, Anja Bedeker graduated with a Masters' degree in Psychology at UWC.

NAME	INSTITUTION	COURSE NAME AND PURPOSE OF COURSE
Alan Christoffels	UCT	Masters' degree in Public Health
Nicki Tiffin	University of Washington	Course on developing policy and advocacy for global health. Online

Dominique Anderson	IBM	AI and Cybersecurity short courses Online
	UWC	Oxford Nanopore Technology (ONT) NGS hands-on training
	SAMRC/NRF	Human Frontiers Science Programme Grantsmanship course

JOURNAL EDITING AND REVIEWING

NAME	REVIEWER	EDITORIAL BOARD MEMBER	JOURNAL ASSOCIATE EDITOR
Alan Christoffels	BMC Genomics	Data Journal	
	BMC Bioinformatics	Bioinformatics Advances	
	Bioinformatics		
	PLoS ONE		
Nicki Tiffin	BMJ Global Health		Nucleic Acids Research Genomics and Bioinformatics Population Studies
Anja Bedeker	BMC Medical Ethics		
Ruben Cloete	Heliyon		
	MDPI Viruses		
Peter van Heusden	Bioinformatics Advances		

SANBI IN THE MEDIA

The ground breaking research at SANBI regularly features in print and online media. SANBI academics are also frequently requested to provide commentary on pressing issues affecting bioinformatics researchers. Below are some articles which appeared during 2024.

NAME	TITLE	PUBLISHED BY URL
Alan Christoffels	Increasing the presence of BIPOC researchers in computational science	nature.com https://www.nature.com/articles/s43588-024-00693-6
Ruben Cloete	Sarepta Secondary School Empowers Students with Inspiring Career Day	TygerBurger https://www.news24.com/news24/community-newspaper/tygerburger/careers-day-for-sarepta-gr-12s-20240402-2
Nicki Tiffin	Strengthening Health and Disease Modelling for Public Health Decision Making in Africa	Grand Challenges Champions https://gcgh.grandchallenges.org/champion/nicki-tiffin
	Genomic Medicine: The Value of Genes for Healthcare	Health Tech Africa Podcast https://podcasts.apple.com/us/podcast/genomic-medicine-the-value-of-genes-for-healthcare/id1670395120?i=1000675741623
Peter van Heusden	Mpox in SA: stigma may be keeping people from coming forward, getting tested	In Health E-News, Journalism for Public Health by Yoliswa Sobuwa https://health-e.org.za/2024/06/07/mpox-in-sa-stigma-may-be-keeping-people-from-coming-forward-getting-tested/

COMMUNITY OUTREACH

Career Day

Together with the principal and Life Orientation teacher at Sarepta Secondary High School, alumni Ruben Cloete and Ashley Hendricks arranged a Career Day for 185 matric learners in March 2024. The aim of this initiative was to motivate and inspire learners to look beyond the school for their ongoing development and advancement. Learners at the school were exposed to professionals in their field and were given guidance on how to go about making career choices. As a result of Career Day, Ruben Cloete hosted two high school students to job shadow him during the month of September.

Mandela Day

On 25 July 2024, to mark Mandela Day, PHA4GE secretariat hosted learners from Woodlands Primary in Heideveld at the SANBI offices. Tracey Calvert-Joshua and other PHA4GE staff introduced the learners to concepts such as the PHA4GE pathogens and the human immune system, and to the production of medicines. Calvert-Joshua says. "I felt that we shouldn't really underestimate kids, and that we could guide them through these concepts and ideas."

Left: TraceyCalvert-Joshua with the primary school learners.

Right: Talk at Sarepta High School.

In addition to its links to Mandela Day, the visit formed part of PHA4GE's 'Youth 4 Science' initiative, which looks to teach young people about possible career pathways in science and public health, regardless of their backgrounds. "We wanted to show them that science can be fun while highlighting the important roles scientists have in today's world," says Rangarirai Matima, communications lead with PHA4GE.

The visit was arranged by Prof Alan Christoffels, PHA4GE's Secretariat and UWC Institutional Advancement PR and Events Unit, in collaboration with Great Commission United (GCU), a not-for-profit organisation operating in Heideveld. In keeping with the GCU's feeding initiative Mother City Kitchen and their ambition to feed not just minds but also bodies, PHA4GE provided their young visitors with a nutritious lunch, including snacks.

Alan Christoffels and the PHA4GE Team with Woodlands Primary School learners.

CAPACITY DEVELOPMENT

SANBI offers training programmes which are in keeping with its vision of becoming a centre of excellence in biomedical research and education in the global, African and South African context.

UNDERGRADUATE TRAINING PROGRAMME

Students who are interested in Bioinformatics as a career path are encouraged to take a combination of relevant subjects in Life or Health Sciences, Statistics, Computer Science and Mathematics during their undergraduate degree.

Bioinformatics Module (BTN 315)

Each year the UWC undergraduate Bioinformatics Module is taught to third-year Biotechnology students by the SANBI faculty. 78 students were taught a range of topics which include: Introduction to Bioinformatics and Databases, Protein structure prediction, Sequence alignments and Phylogenetics.

POSTGRADUATE TRAINING PROGRAMME

Postgraduate training at SANBI is well-established and alumni are now working all over South Africa and at bioinformatics research sites around the world.

Honours Programme

Although SANBI does not have an Honours programme, students who attain a pass rate of >60% can include a bioinformatics component to their Honours thesis project, branched off the BSc Honours Biotechnology programme.

Masters' Programme

SANBI offers a research MSc in Bioinformatics by research thesis. Candidates with a BSc Honours degree or equivalent in a related scientific subject area may apply. The MSc degree is usually completed within two years.

Doctoral Programme

Candidates should be in possession of a MSc degree in Bioinformatics or in a related scientific field subject areas such as Computer Science, Mathematics, Biochemistry and Engineering. The PhD degree must be completed within five years.

Postdoctoral Programme

Postdoctoral fellows are admitted to the research programme after consultation with a potential SANBI supervisor. Fellowships may last for a maximum of three years.

INTERNSHIPS

When opportunities arise, SANBI will employ interns as part of their initiatives to recruit students. Peter van Heusden hosted a college student, Zach Moss who is interested in Computer Systems. He learned about Linux server administration and helped out with IT tasks. Felicia Khoza, an MSc Computational Health Informatics student from UCT was hosted by Dominique Anderson for an 8-week internship during September and October.

SANBI STUDENTS

The 2024 SANBI student complement comprises a cohort of diverse and multi-talented researchers. In total there are 3 Postdoctoral Fellows, 7 PhD and 9 MSc candidates who are supervised by the SANBI academic staff.

REGISTERED POSTDOCTORAL FELLOWS

NAME	SURNAME	GROUP	COUNTRY	PRIMARY SUPERVISOR
Eddie	Lulamba	Postdoctoral research fellow with the health data integration group (HDIG)	DRC	Nicki Tiffin
		Postdoctoral research fellow with the PHA4GE group	DRC	Alan Christoffels
Catherine	Rossouw	Postdoctoral research fellow with Christoffels Lab	South Africa	Alan Christoffels
Tsaone	Tamuhla	Postdoctoral research fellow with the health data integration group (HDIG)	Botswana	Nicki Tiffin

Eddie Lulamba

Catherine Rossouw

Tsaone Tamuhla

REGISTERED DOCTORAL STUDENTS

NAME	SURNAME	COUNTRY	PRIMARY SUPERVISOR
Adetutu	Akinnuwesi	Nigeria	Ruben Cloete
Keaghan	Brown	South Africa	Ruben Cloete
Jeremy	Burgess	South Africa	Ruben Cloete
Mohammed	Hassan	Ethiopia	Ruben Cloete
Adeshina	Odugbemi	Nigeria	Alan Christoffels
Josephine	Peka	South Africa	Gordon Harkins
Sohail	Simon	South Africa	Nicki Tiffin

Adetutu Akinnuwesi

Keaghan Brown

Jeremy Burgess

Mohammed Hassan

Adeshina Odugbemi

Josephine Peka

Sohail Simon

REGISTERED MASTERS' STUDENTS

NAME	SURNAME	COUNTRY	PRIMARY SUPERVISOR
Jeremy	Burgess	South Africa	Ruben Cloete
Farzana	Diedricks	South Africa	Alan Christoffels
Ridaa	Diedricks	South Africa	Alan Christoffels
Kriheska	Francis	South Africa	Gordon Harkins
Tatenda	Majuru	Zimbabwe	Dominique Anderson
Lunathi	Ndlondlo	South Africa	Ruben Cloete
Nkosazana	Nyembezi	South Africa	Gordon Harkins
Chance	Hwenude	Benin	Alan Christoffels
Veronica	Wokibula	Uganda	Dominique Anderson

Jeremy Burgess **Farzana Diedricks** **Ridaa Diedricks** **Kriheska Francis**
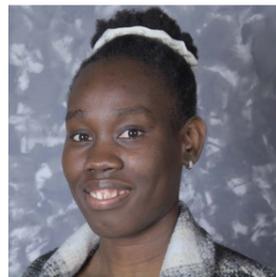
Tatenda Majuru **Lunathi Ndlondlo** **Nkosazana Nyembezi**
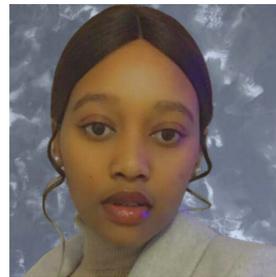
Chance Hwenude **Veronica Wokibula**

POSTGRADUATE STUDENT DIVERSITY EXPRESSED BY COUNTRY AND GENDER

	South Africa	Zimbabwe	Benin	Uganda	Nigeria	Ethiopia	Botswana	DRC
	11	1	1	1	2	1	1	1

— Male — Female
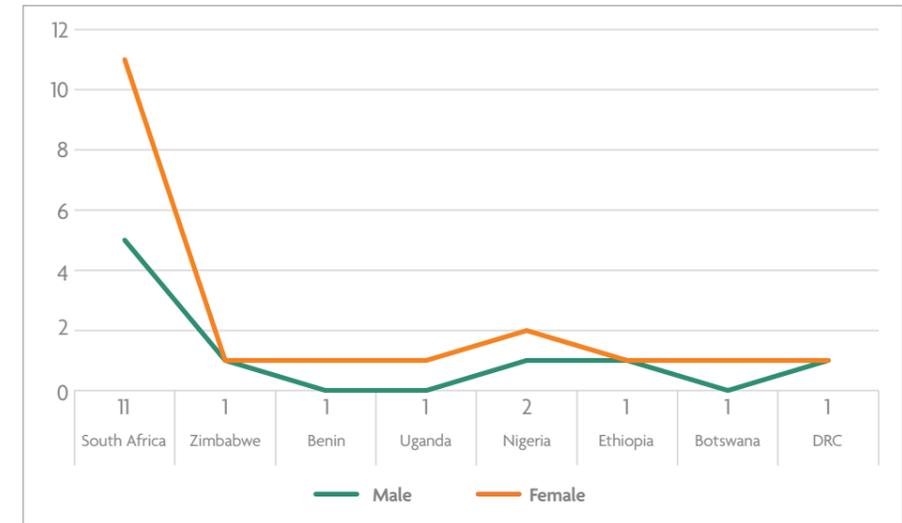
POSTGRADUATE STUDENT REGISTRATIONS 2015 – 2024

— MSc — PhD — Postdoctoral
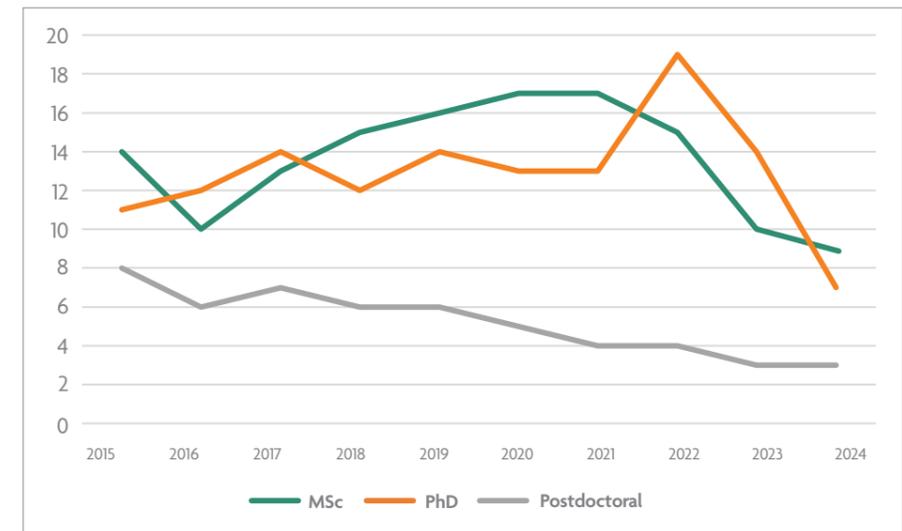
2024 SANBI GRADUATIONS

NAME	DEGREE	THESIS TITLE	PRIMARY SUPERVISOR	SECONDARY SUPERVISOR
Bridget Langa	PhD Bioinformatics	Discovering cancer subtypes by tracking cancer progression with transcriptomic data through the multi-stage process of cancer development.	Junaid Gamieldien	Burtram Fielding
Michelle Livesey	PhD Bioinformatics	Semantic discovery and computational filtering to identify potentially novel breast cancer genes and signatures in omics data.	Hocine Bendou	Alan Christoffels

Michelle Livesey.
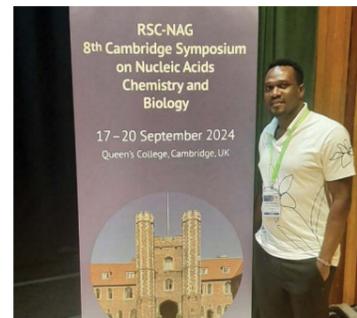
Bridget Langa.

STUDENT AWARDS

Above: Eddie Lulamba
Oral: The African Data and Biospecimen Exchange (ADBEx) – a platform to facilitate equitable and ethical sharing of African data and biospecimens.
WON: 1ST PLACE SHORT TALK PRIZE

Above centre: Darius Martin
Poster: Computational approaches to identify aptamers and their application in the rapid detection of infectious agents.
WON: BIOCHEMICAL JOURNAL POSTER PRIZE/AWARD

Above right: Adeshina Odugbemi
Poster: Cyberinfrastructure Collaboration: Towards Accelerated Impact.
WON 1ST PRIZE FOR PhD POSTER

TEACHING AND LEARNING BY SANBI

SANBI's commitment to capacity development in the field of bioinformatics is apparent in all their activities to organise or conduct training and workshop opportunities for a variety of participants throughout Africa.

Bioinformatics for Genomic Epidemiology

This introductory course on Pathogen Bioinformatics for Genomic Epidemiology ran over the first two weeks of March 2024. 18 Trainees from public health laboratories and research facilities from around the continent attended this course which had an urgent focus on cholera.

Peter van Heusden, Tracey Calvert-Joshua, Keaghan Brown, George Githinji, John Juma and Nasr Eshibona contributed in-person instruction combining theoretical with practical instruction and opportunity for hands-on practice of the skills presented. This course was organised by SANBI at UWC in collaboration with Africa CDC, part of the African Union (AU) and the Computational Biology Department at UCT.

Introduction to Bioinformatics

In March 2024, Dominique Anderson taught this component of the Pathology and molecular biology module to 15 Masters' in Medicine students at the University of Stellenbosch.

Consultative Meeting on Training Curriculum and Competency Framework for Public Health Genomics and Bioinformatics in Africa

At the beginning of April, Africa CDC in collaboration with the African Society for Laboratory Medicine (ASLM) and the NGS Academy, organised a 3-day consultative meeting on training curriculum and competency framework for public health genomics and bioinformatics in Africa. The meeting, held at the Africa CDC HQ in Addis Ababa, was attended by Peter van Heusden of SANBI. Also present were 30 people drawn from expert groups either in pathogen genomics or in bioinformatics with several key public health institutes represented by universities, research institutes, industry players and the ASLM.

This curriculum will help standardise training provision on the continent and improve multi-pathogen genomic surveillance in the AU member states.

Attendees at the 3-day meeting.

Introduction to Bioinformatics (IBT)

This course aims to use a combination of theoretical and practical sessions in order for participants to gain practical experience in using various bioinformatics tools and resources. 30 participants attended in person at SANBI and the rest via virtual classrooms. The in-person classrooms were co-ordinated by Nicki Tiffin and Catherine Rossouw. The course ran from April - July 2024. Funding was provided by the H3ABioNet.

Introduction to Study Design and R
In May 2024 Nicki Tiffin taught this course to 6 postgraduate students over 3 weeks at SANBI. This course is aimed at life science postgraduates who lack the basic study design knowledge and basic statistical skills for different study types. It is also designed to get students coding in R.

Data Curation Workshops
Building on the launch of Africa's first pathogen data-sharing platform (AGARI) in 2023, the Africa CDC conducted the first 3-day workshop in data curation from 18 - 20 June 2024 in Addis Ababa, as part of its programme to expand capacity in and support for pathogen genomics around Africa. The workshop attended by participants from public health laboratories in 20 Member States, aimed to accelerate data verification and validation before sharing for public health use. This continental platform is intended for use by national public health institutions, national reference laboratories, research and academic institutions from around Africa to upload, manage, and share pathogen sequence and associated metadata to effectively respond to public health threats across Member States in a co-ordinated manner.

Second data curation workshop held in Cape Town.

The 2nd Data Curation Workshop at the Africa CDC was taught by Dominique Anderson along with non-SANBI trainers in December 2024. The course aimed to introduce participants to data curation, data management and database utilisation in public health. 30 participants from various African countries attended.

Sysadmin Curriculum Planning Workshop
Alan Christoffels secured funding from the World Bank to develop a sysadmin curriculum. In August and December 2024 workshops were held at SANBI to design the curriculum on filling the sysadmin skills gap to support public health bioinformatics and genomic epidemiology in Africa. Participants from across Africa and beyond have been drawing on insights from public health labs, research institutions and the work in the Galaxy Project, UK MRC CLIMB-BIG DATA, RSSE Africa, Public Health Alliance for Genomic Epidemiology (PHA4GE), H3ABioNet and others.

Sysadmin Curriculum Planning attendees at SANBI.

Advanced Bioinformatics and Genomic Epidemiology Training for AMR Pathogens
During September 2024 an Advanced Bioinformatics and Genomic Epidemiology Training for AMR pathogens was organised by Africa CDC in collaboration with SANBI. This training focused on strengthening national genomic surveillance of Multi-Drug Resistance (MDR) as it continues to increase globally.

Masters' in Data Analytics
Nicki Tiffin ran one module: Omics in Health for this programme in the Faculty of Community Health to 15 online participants during September 2024.

Introductory Course on Klebsiella Pneumoniae Bioinformatics
Tracey Calvert-Joshua, Keaghan Brown, Farzaana Diedericks from SANBI in collaboration with Africa CDC, the APHF, ASLM, LSHTM and Oxford University ran this 2 week in-person instruction to introduce analysis of bacterial sequence data with a focus on Klebsiella. 17 Participants attended from Cameroon, Chad, DRC, Ethiopia, Malawi, Morocco, Mozambique, Nigeria, Zambia and Zimbabwe during September 2024.

Mpox Virus Sequence Analysis
During October 2024, Peter van Heusden along with trainers from the NGS Academy and Africa CDC ran this course which was organised at short notice to help grow Mpox virus sequence analysis capacity in public health labs in Mpox clade-I affected countries. Participants from Central African Republic, Republic of Congo, Uganda, Rwanda, Cameroon and Burundi and trainers from South Africa, Côte d'Ivoire, Kenya and Tanzania attended. This training was taught in both English and French.

African Climate and Health Capacity Accelerator Network (CAN)
From August to October 2024 Nicki Tiffin contributed three sessions as a guest trainer to this online programme which was run by OpenUp - Mapping geospatial data in public health; Healthcare data ecosystems; and a brief overview of Genomics in Health. Approximately 50 participants attended online.

Information Systems for Biobanking with Hands-on Practical on BAOBAB LIMS
Dominique Anderson taught a group of 11 professionals, academics, and students this module at the 1st Medical Biorepositories of South Africa (MBirSA) Summer School: Fundamentals of Biobanking in October 2024.

MENTORSHIPS

2024 1st Africa, Law & Emerging Technologies Policy Hackathon
The policy hackathon took place on 31 July which was aimed at registered postgraduate SA students to brainstorm and pitch creative policy solutions that address the intersection between law and data (access, use, reuse and management) in various fields of AI and data science applications (health, culture and languages, climate, etc) in South Africa. Dominique Anderson and Peter van Heusden were judges. The Data Science Law Lab in collaboration with the Data Science for Social Impact research group, The Centre for Child Law and the Expression, Information and Digital Rights Unit, Centre for Human Rights (all at the University of Pretoria) as well as SANBI provided mentorship, collaboration and support as Challenge Partners.

Student Cluster Competition Mentoring
SANBI once again assisted UWC Computer Science students in preparing for the CHPC's annual Student Cluster Competition. Two teams from UWC were selected to articipate in the National Round of the competition with the "HPC Wizards" Team coming third overall, with one of the team members, Lisa Pitsi, chosen to take part in the national team that will participate in the Student Cluster Competition at the ISC Conference in Germany in 2025.

VISITING SPEAKER AT SANBI

On 20 August SANBI hosted Dr Idrissa Dien, a researcher at the Virology Department from the Institut Pasteur de Dakar in Senegal who presented a seminar on "Genomic Surveillance of Dengue virus in Senegal and West Africa".

COMPUTATIONAL RESOURCES

SANBI's IT and bioinformatics software development team supports the work of the institute through software development and engagement with other computing providers to ensure that SANBI researchers are able to access appropriate computing resources.

Our major computing partners are the Centre for High Performance Computing (CHPC) whose HPC cluster supports the Ilifu cluster and cloud resources. Dr Ruben Cloete's research group makes extensive use of both the CHPC and Ilifu for molecular dynamic simulations and docking simulation between candidate drugs and target proteins. Profs Alan Christoffels and Nicki Tiffin use virtual machines hosted on the Ilifu cloud to host database resources.

RESEARCH AND DEVELOPMENT DRIVING SANBI'S COMPUTATIONAL INFRASTRUCTURE DEPLOYMENT

PRINCIPAL INVESTIGATOR	THEME	PROJECT
Alan Christoffels	Pathogen Archive Prototype/AGARI	The African pathogen archive prototype was built to curate Priority Pathogens.
Nicki Tiffin	Hosting a public platform	The African Data and Biospecimen Exchange - Ilifu. Development of this programme has been completed during 2024, and it will go live in 2025.
Peter van Heusden	Research Computing Infrastructure	Studying the deployment of research computing infrastructure and development of associated skills on the African continent through RSSE Africa.

RESEARCH SOFTWARE AND SYSTEMS ENGINEERING AFRICA (RSSE AFRICA)

RSSE Africa is an initiative started by SANBI's Peter van Heusden and Eugene de Beste in 2017 to stimulate the growth of an African community of practice around digital research infrastructure development and research software and system engineering practices. With support from Talarify (a startup founded by SANBI alumnus Anelda van der Walt), RSSE Africa collaborated with RSE Asia to run a series of events on Enabling Open Science Through Research Code.

Previous events hosted by RSSE Africa highlighted the need for more resources to assist those developing research software, many of whom have in fact only recently heard the term "research software engineer" despite the centrality of software development to their research work. The RSSE Africa/RSE Asia collaboration aimed to provide an onramp for those working in RSE, especially in lower and middle income countries (LMICs). Three online events were held in 2024 with three more planned for 2025.

Summaries, recordings and resource guides are available on the Community Meetups section of the RSSE website (**https://rsse.africa**).

EVENT TITLE	SUMMARY
A Conversation With Researchers Who Code 10 October	An introduction into what the life of a researcher who codes looks like with four researchers from South Africa, Egypt and the UK talking about their favourite programming languages, career journeys and advice for fellow research software developers.
Enabling Reproducibility Through Research Code 14 November	What steps can you take to ensure that your code enables reproducibility in your research project? From carefully planned directory structures to the use of automated workflows and containers, researchers from Japan, Norway and South Africa presented their experiences.
Opening Up Research Code 12 December	Why make your code "open"? What are the steps involved and how can you benefit others - and even your future self - by making research software easier to install and embracing "open source" development practices.

RESEARCH OUTPUTS

SANBI's profile of excellence is reflected in its ability to publish collaborative high-impact scientific articles in international publications.

BOOK CHAPTER

Understanding the COVID-19 Syndemic in South Africa: Concrete Responses and a Call to Action
van Heusden P, Lewins K, Reynolds L.
in
COVID-19 Syndemics and the Global South: A World Divided
Ali I, Singer M, Bulled (pp. 104-137)
Routledge Studies in Health and Medical Anthropology July 2024
eBook ISBN9781003365358

"This book focuses on syndemics in the Global South and uses COVID-19 as a window to understand clusters of disparities and disease comorbidities.

JOURNAL PUBLICATIONS

PUBLICATION DETAILS	DATE
Molecular modeling and simulation studies of SELEX-derived high affinity DNA aptamers to the Ebola virus nucleoprotein. Martin DR, Mutombwera AT, Madiehe AM, Onani MO, Meyer, Cloete R. Journal of Biomolecular Structure and Dynamics 2024 **Doi: 10.1080/07391102.2024.2302922**	13 January
Compliance with research ethics in epidemiological studies targeted to conflict-affected areas in Western Ethiopia: validity of informed consent (VIC) by information comprehension and voluntariness (ICV). Tiruneh G, Yilma M, Wakuma B, Abdisa E, Bayisa L, Nichols M, Bedeker A, Tiffin N. BMC Medical Ethics 25, 9. **https://doi.org/10.1186/s12910-024-01003-5**	18 January
Translating the consent form is the tip of the iceberg: using cognitive interviews to assess the barriers to informed consent in South African health facilities. Pillay N, Ncube N, Moopelo K, Mothoagae G, Welte O, Shogole M, Gwiji N, Scott L, Moshani N, Tiffin N, Boulle A, Griffiths F, Fairlie L, Mehta U, LeFevre A, Scott K. Sexual and Reproductive Health Matters, 31:4 **Doi: 10.1080/26410397.2024.2302553**	26 January
The role and potential of computer-aided drug discovery strategies in the discovery of novel antimicrobials. Oselusi SO, Dube P, Odugbemi AI, et al. Computers in Biology and Medicine 169:107927. **Doi: 10.1016/j.compbiomed.2024.107927**	29 February
WCN24-736 incidence and characteristics of healthcare clients with acute-on-chronic kidney disease in the City of Cape Town, South Africa. Caskey F, Phelanyane F, Birnie K, Ben-Shlomo Y, Aylward R, Rayner B, Tiffin N. Kidney International Reports, 2024 9(4): S24-S25.	1 April
Kidney function in healthcare clients in Khayelitsha, South Africa: Routine laboratory testing and results reflect distinct healthcare experiences by age for healthcare clients with and without HIV. Osei-Yeboah R, Ngwenya O, Tiffin N. PLOS Global Public Health 4(5): e0002526 **Doi: 10.1371/journal.pgph.0002526**	1 May
The workshops on computational applications in secondary metabolite discovery (CAiSMD). Ntie-Kang F, Eni DB, Telukunta KK, Chukwudi Osamor V, Egieyeh SA, et al. Physical Sciences Reviews **Doi: 10.1515/psr-2024-0015**	8 May
Impact of subtype C-specific amino acid variants on HIV-1 Tat-TAR interaction: insights from molecular modelling and dynamics. Gotora PT, Brown K, Martin DR, van der Sluis R, Cloete R, Williams ME. Virology Journal 21(1) **Doi: 10.1186/s12985-024-02419-6**	25 June
Artificial Intelligence in antidiabetic drug discovery: The advances in QSAR and the prediction of α-glucosidase inhibitors. Odugbemi AI, Nyirenda C, Christoffels A, Egieyeh SA. Computational and Structural Biotechnology Journal, Volume 23, 2964 – 2977.	6 July

Diagnostic Prediction Model for Tuberculous Meningitis: An Individual Participant Data Meta-Analysis
16 July
Stadelman-Behar AM, **Tiffin N,** Ellis J, *et al.*
The American Journal of Tropical Medicine and Hygiene Volume 111:3(546-553)
Doi: https://doi.org/10.4269/ajtmh.23-0789

Interdisciplinary perspectives on multimorbidity in Africa: Developing an expanded conceptual model.
30 July
Dixon J, Morton B, Nkhata MJ, Silman A, Simiyu IG, Spencer SA, Van Pinxteren M, Bunn C, Calderwood C, Chandler CIR, Chikumbu E, Crampin AC, Hurst JR, Jobe M, Kengne AP, Bickton F, Bilungula AM, Bosire E, Chawani MS, Chinoko B, Chisala M, Chiwanda J, Drew S, Farrant L, Ferrand RA, Gondwe M, Gregson CL, Harding R, Kajungu D, Kasenda S, Katagira W, Kwaitana D, Mendenhall E, Mensah ABB, Mnenula M, Mupaza L, Mwakasungula M, Nakanga W, Ndhlovu C, Nkhoma K, Nkoka O, Opare-Lokko EA, Phulusa J, Price A, Rylance J, Salima C, Salimu S, Sturmberg J, Vale E, Limbani F. Levitt NS, Moshabela M, Owolabi M, Peer N, Phiri N, Singh SJ, **Tamuhla T**, Tembo M, **Tiffin N**, Worrall E, Yongolo NM, Banda GT,
PLOS Global Public Health. 2024 Jul 30;4(7): e0003434.
Doi: 10.1371/journal.pgph.0003434.

Progress in Pharmacometrics Implementation and Regulatory Integration in Africa: A Systematic Review.
20 August
Ndzamba BS, **Egieyeh S**, Fasinu P.
Clinical Pharmacology and Therapeutics (0009-9236)
Doi: 10.1002/cpt.3415

Consistency of alerts generated by, and implementation of, the NHS England acute kidney injury detection algorithm in English laboratories.
4 August
Aylward R, Casula A, **Tiffin N**, Ben-Shlomo Y, Rayner B, *et al.*
Journal of Nephrology ID: 001283253800001

Increasing the presence of BIPOC researchers in computational science.
1 September
Chen CY, **Christoffels A**, Dube R.
Nature Computational Science 4(9): 646-653
Doi: 10.1038/s43588-024-00693-6

Africa in the era of pathogen genomics: Unlocking data barriers.
19 September
Mboowa G, Tessema SK, **Christoffels A**, Ndembi N, Kebede Tebeje Y, *et al.*
Cell: 5146-5150, 2024-09-19, ID:85204513060

Data governance for ethical usage of linked routine health data in South Africa: balancing privacy and data sharing.
30 September
Mutemaringa T, Boulle A, **Tiffin N**.
International Journal of Population Data Science, 9(5).
Doi: 10.23889/ijpds.v9i5.2721.

Implementation of a genotyped African population cohort, with virtual follow-up: A feasibility study in the Western Cape Province, South Africa
22 October
Tamuhla T, Coussens AK, Abrahams M, *et al.*
Wellcome Open Research 2024, 9:620
Doi.org/10.12688/wellcomeopenres.23009.1

The PHA4GE Microbial Data-Sharing Accord: establishing baseline consensus microbial data-sharing norms to facilitate cross-sectoral collaboration.
30 October
Griffiths EJ, **van Heusden P, Tamuhla T**, *et al.*
BMJ Global Health 2024;9: e016474.
Doi:10.1136/ bmjgh-2024-016474

Antimicrobial activity of Petivera alliacea L. root and its constituents: in vitro and in silico studies
1 November
Adesanya EO, Adesanya OO, Ogunlakin AD, AjayiOdoko OA, Ojo OA,
Odugbemi AI, Egieyeh SA, Oyesiku OO.
International Journal of Plant Research.
Doi: 10.1007/s42535-024-01085-x.

NUMBER OF OUTPUTS PER YEAR 2015-2024

NON-PEER REVIEWED DOCUMENT

24 March marks a significant date on the global health calendar as World TB Day. PHA4GE is actively working to combat TB, with researchers from around the world coming together to collaborate on various challenges and resources. Peter van Heusden and Tracey Calvert-Joshua of SANBI contributed to the TB guidance document *"Mycobacterium tuberculosis Bioinformatics Guidance Document"*, which was designed to offer direction for those engaged in bioinformatics analyses, pointing them towards the pertinent knowledge and tools needed.

KEYNOTES/INVITED TALKS PRESENTED

PRESENTER NAME	CONFERENCE DETAILS	PARTICIPATION AND TITLE
Alan Christoffels	Launch of GLOWACON: the Global consortium for wastewater & environmental surveillance for public health. Brussels, Belgium. March 2024.	Invited Panelist: Waste Water Surveillance Data management
	10th Federation of infectious diseases Societies of Southern Africa. Johannesburg, South Africa. May 2024	Plenary: The role of metagenomics in South Africa
	Accelerating the use of pathogen genomics & metagenomics in public health workshop. National Academies Sciences, Engineering, Medicine. USA. June 2024	Invited talk: PHA4GE data standards system and structure
	GET 10th African Conference on One Health and Biosecurity, Lagos Continental Hotel, Nigeria. November 2024	Speaker: From Ebola to COVID-19 and Beyond: Strengthening Africa's Capacity to Address Emerging Global Health Security Threats
Nicki Tiffin	Wellcome Connecting Science: Genomics for Antimicrobial Resistance Surveillance in OneHealth Symposium, Johannesburg, South Africa. March 2024	Workshop presenter: Challenges and opportunities in genomic data sharing
	PRECISE-IHCC Conference 2024, Singapore. August 2024	Speaker: Strategic Cohort Design for the Global South
	8th Annual Western Cape Research Ethics Workshop, UWC, South Africa. September 2024	Speaker: Making sense of secondary data use with an ethical and legal lens
	African Microbiomes in Health and Disease Symposium, UCT, South Africa. October 2024	Session Chair: Creating Robust Ecosystems for Microbiome Research in Africa
	Nature Forum on Health Equity and Nature Awards for Inclusive Health Research, Stellenbosch University, South Africa. October 2024	Speaker and Panelist: Investing in genomics in Africa.
	Asia PGI Webinar Series, Online. November 2024	Presenter: Data Sharing in Pathogen Genomics, What's the way forward?
Tsaone Tamuhla	CIDRI-Africa Annual Scientific Meeting, Somerset West, South Africa. July 2024	Presenter: Implementation of a genotyped African population cohort with virtual follow-up: A feasibility study in the Western Cape Province, South Africa
	Africa CDC Pathogen Genomic Data Sharing Workshop, Addis Ababa, Ethiopia. December 2024	Virtual Talk: From Sharing to Secondary Use: PHA4GE Accord and Tools for Equitable Microbial Data Access in Africa
Peter van Heusden	Pathogen Genomics Initiative meeting at the ASLM Special Convention on Diagnostics, Abidjan, Ivory Coast. November 2024	Speaker: Quality Management and Standardisation for Bioinformatics Workflows
	WHO IPSN Global Partners Forum., Bangkok, Thailand. November 2024	Panelist: Equitable access to pathogen genomic surveillance: the analytical barriers

At the GET Conference in June 2024 Alan Christoffels, a founding member of the Global Emerging Pathogens Treatment Consortium, was honoured for his contributions to academia, public health, and community engagement over the past 10 years.

Nicki Tiffin presenting at the Precision Health Research, Singapore (Precise) International Health Cohorts Consortium (IHCC) Conference.

CONFERENCE PARTICIPATION

Additional to the postgraduate programmes, SANBI offers opportunities for staff and students alike to participate in relevant scientific conferences in the form of organiser, facilitator, attendance or presentations.

PRESENTER NAME	CONFERENCE DETAILS	PARTICIPATION AND TITLE
Peter van Heusden and Dominique Anderson	ASLM Workshop on Development of an NGS Quality Management System framework, Nairobi, Kenya. July 2024	Participants/Consultants
Dominique Anderson	IMBM Symposium, UWC, SA. August 2024	Attendee
	ISBER regional meeting, St Petersburg, Florida, USA. November 2024	Poster: Establishment of a Medical Biorepositories of South Africa (MBirSA) Network
	Data Advocacy and Sharing Workshop, Africa CDC, Ethiopia. December 2024	Attendee and Session Facilitator
Eddie Lulamba	Genomics for Antimicrobial Resistance Surveillance in One Health Symposium 2024, Wellcome Connecting Science. National Institute for Communicable Diseases (NICD), Johannesburg, SA. March 2024	Oral: The African Data and Biospecimen Exchange (ADBEx) – a platform to facilitate equitable and ethical sharing of African data and biospecimen **WON: 1ST PLACE SHORT TALK PRIZE**
	South African Society for Bioinformatics and South African Genetics Society (SASBi/SAGS) BIO2024 Conference, Future Africa, Pretoria, SA. September 2024	Oral: The African Data and Biospecimen Exchange (ADBEx) – a platform to facilitate equitable and ethical sharing of African data and biospecimens
	Genomic Epidemiology of Malaria conference, Wellcome Genome Campus, Cambridge, UK. September 2024	Online Attendee
Darius Martin	8th Cambridge Symposium on Nucleic Acids Chemistry and Biology in Cambridge, UK. 17-20 September 2024	Poster: Computational approaches to identify aptamers and their application in the rapid detection of infectious agents **WON: BIOCHEMICAL JOURNAL POSTER PRIZE/AWARD**
Irene Muchada	Genomics of Rare Disease 2024, Wellcome Connecting Science online. March 2024	Attendee
	Beyond the Sequence: Ethical, Legal, and Social Contexts in Genomics, Wellcome Trust offices, Cambridge, UK. March 2024	Attendee
	Beyond the Sequence: Ethical, Legal, and Social Contexts in Genomics, Radisson Blu Hotel, Cape Town, South Africa. April 2024	Attendee
	Human Immunology - Genes and Environment 2024, Wellcome Connecting Science online. May 2024	Attendee
	Southern African Society of Human Genetics 2024 conference, Sun City, South Africa. October 2024	Poster: A Researcher's Reflections on Participants' Experiences During Implementation of Tiered Informed Consent for Genomic Health Research: A Qualitative Iterative Observation
	Young and Early Career Geneticists Group Symposium, Witwatersrand, Johannesburg, South Africa. October 2024	Oral: A Researcher's Reflections on Participants' Experiences During Implementation of Tiered Informed Consent for Genomic Health Research: A Qualitative Iterative Observation
Themba Mutemaringa	International Population Data Linkage Network, Chicago - Illinois, USA. September 2024	Poster: Data governance for ethical usage of linked routine health data in South Africa: balancing privacy and data sharing
Adeshina Odugbemi	18th Centre for High Performance Computing (CHPC) National Conference at the Boardwalk International Convention Centre, Gqeberha, SA. December 2024	Poster: Cyberinfrastructure Collaboration: Towards Accelerated Impact **WON 1ST PRIZE FOR PhD POSTER**
Tsaone Tamuhla	PRECISE-IHCC Conference 2024, Singapore. August 2024	Poster: A modular approach to compiling data- and biospecimen-sharing agreements to promote more equitable terms for resource sharing
Sohail Simon	World Congress of Epidemiology 2024 (WCE2024), CTICC, Cape Town, South Africa. September 2024	Attendee

RESEARCH LABORATORIES

PROF ALAN CHRISTOFFELS

HIGHLIGHTS OF 2024

One PhD student, Michelle Livesey, graduated during 2024.

RESEARCH PROJECT THEMES

Pathogen genomics
- In collaboration with the Tygerberg SAMRC Unit, we are developing methods to analyse high throughput sequencing data for microbial genomes.
- Identification of novel drug targets in pathways known to contain drug resistant genes.
- Identification of operons and non-coding RNAs to understand gene regulation in pathogen genomes.

Pathogen surveillance
- Developing systems for pathogen data management.
- Development of global standards for data sharing and integration.
- Public Health Alliance for Genomic Epidemiology (PHA4GE).

Drug discovery
- In partnership with the School of Pharmacy at UWC, delineate the pharmacokinetics of drugs in response to host genetic factors.

RESEARCH PROJECTS

The projects below underpin our translational work:

Tuberculosis
A comprehensive research programme is underway that includes investigating genetics determinants in both host (Human) and pathogen (*Mycobacterium tuberculosis*) to understand drug resistance, and protein structure determination to inform patient-centric drug design. These findings inform the development of a scientific workflow management system to support reproducible high throughput computational experiments.
One of the projects focus on predicting operons. Operons may be important drug targets for the development of effective anti-microbials to combat the emerging, global drug resistance challenge. However, there is a shortage of known *Mycobacterium tuberculosis* operons. We developed COSMO, an algorithm that removes the limitations imposed by using the constraints of one specific organism's genome and exploits raw RNA-seq data instead. The code and examples of input and output files for testing and validation, are available at the project home page at **https://github.com/SANBI-SA/COSMO**

Biobanking
In collaboration with Dominique Anderson, we continue to strengthen our biobank software capabilities so that genetics data and biospecimens can be curated and stored.

RESEARCH COLLABORATIONS

1. Develop a Biobank Informatics Management System
Collaborating Parties:
Dr Dominique Anderson - SANBI, UWC and Africa CDC
Nature and Purpose:
Development of a laboratory management system for use by Biorepositories.
Output in the last 12 months:
Baobab LIMS Lite core module has been developed.
Future Direction:
Building functionality to support biodiversity biorepositories.

2. Analysing Genetic Networks in M.tuberculosis
Collaborating Parties:
Prof Rob Warren - University of Stellenbosch
Nature and Purpose:
Identification of Operons: applying Operon finder, COSMO, to clostridium difficile.
Output in the last 12 months:
Refining COSMO for other pathogens.

3. Computational Bacterial Analytical Toolkit for Tuberculosis (COMBAT-TB)
Collaborating Parties:
Profs Samantha Sampson and Rob Warren - University of Stellenbosch
Peter van Heusden - SANBI, UWC
Nature and Purpose:
Develop a computational platform to store Tuberculosis -omic data and to provide a visualisation tool.

4. Biosecurity Preparedness in Africa
Collaborating Parties:
Dr Dominique Anderson - SANBI, UWC
Global Consortium on emerging infectious diseases **(www.getafrica.org).**
Nature and Purpose:
We aim to establish infrastructure and research capacity to respond to highly infectious emerging pathogens.
Future Direction:
Advocating for Biosecurity Best Practice in Africa.

5. Public Health Alliance for Genomic Epidemiology (PHA4GE)
Collaborating Parties:
63 Organisations including - Africa CDC; Oxford University; Washington University; University of Melbourne; USA CDC; University of Birmingham UK; Simon Fraser University, Vancouver; Canada; Broad Institute in Boston and H3Africa.
Nature and Purpose:
The PHA4GE consortium was established to bring bioinformatics closer to public health - to build data standards.

6. African Genomics Archive
Collaborating Parties:
Africa CDC
Nature and Purpose:
To establish a data platform in Africa to manage disease surveillance data with a view to strengthen public health institutes.
Output in the last 12 months:
Prototype of the archive - AGARI.
Future Direction:
Develop a production ready server to host 5 priority pathogens for African countries.

PROF NICKI TIFFIN

TEAM MEMBERS

Postdocs: Tsaone Tamuhla, Eddie Lulamba
PhD students: Sohail Simon, Florence Phelanyane and Themba Mutemaringa
Researcher: Irene Muchada
Administration: Saajidah Beghardien

HIGHLIGHTS OF 2024

Students

During 2024, Themba Mutemaringa and Florence Phelanyane, who both work at the Provincial Health Centre at Western Cape Government Health and Wellness have made good progress in their part-time PhD degrees under my supervision at UCT. We are also pleased to see Sohail Simon launch his PhD work during 2024, funded by NRF, and make excellent progress. He is being supervised by Tsaone Tamuhla and myself. These students have a variety of projects working with large routine health datasets as well as conducting pilot studies to link genomic and routine health data.

During 2024, Dr Eddie Lulamba has continued working as a postdoctoral fellow on the ADBEx project, and we congratulate him on his successful transition to the PHA4GE project at the end of 2024. We wish him every success in his new role and thank him for all his hard work on the ADBEx programme to date.

Tsaone has continued as a postdoc, working on the ADBEx programme, and she also has been the co-lead of the VCAMM project and has been primarily responsible for overseeing the successful completion in December 2024 of the pilot phase of the project funded by UKRI/MRC. She has been co-ordinating participant recruitment together with Irene Muchada, who has continued to work as a genetic counsellor on the VCAMM project, as well as running the DNA prep pipeline and building the bioinformatics and epidemiological pipelines for analysis of the data generated to date.

Ms Tiffany Fredericks also joined our team at the end of 2024, in preparation of her MSc (Bioinformatics) degree with us starting in 2025 - also working with data from the VCAMM programme. She used the interim between her successful completion of her BSc (Hons) degree at UWC and the MSc programme to very efficiently help us get our VCAMM stakeholder meeting successfully organised.

Thank you to the HDIG team for a very successful and enjoyable year.

Funding

Our research group has been awarded a new grant from the Bill & Melinda Gates Foundation to develop advanced data modelling approaches for analysing risk factors and maternal and child outcomes for gestational diabetes and pre-eclampsia, using large health datasets. Led by Principal Investigator Nicki Tiffin and co-PI Tsaone Tamuhla, the project launched in November 2024 and brings together African collaborators from South Africa, Kenya and Uganda, as well as global South collaborators in India and Brazil. By leveraging innovative data-driven modelling methodologies, this initiative aims to generate insights that can contribute to improving maternal and child health outcomes in diverse healthcare settings.

In 2024 we completed our third year of funding with the Bill & Melinda Gates Foundation provided through the Calestous Juma Science Leadership Fellowship awarded to myself. This fellowship provides five years of funding for building the African Data and Biospecimen Exchange (ADBEx) as well as opportunities for networking and training in non-scientific expertise for the fellows.

We also successfully completed the pilot phase of our VCAMM pilot programme to build a virtual genotyped multi-morbidity population cohort through the linkage of routine health data to genotype data for consenting participants in the Western Cape, South Africa. This first stage of the cohort programme has been supported by UKRI/MRC.

Finally, we participated as co-investigators on a collaborative programme funded by the Bill & Melinda Gates Foundation together with partners from Centro de Integração de Dados e Conhecimentos para Saúde (CIDACS) in Brazil and the Provincial Health Data Centre (PHDC) at the Western Cape Department of Health and Wellness. I attended a workshop at CIDACS in Salvador, Brazil, together with PHDC colleagues (April 2024), and together with Tsaone and Themba also attended the reciprocal workshop held in Cape Town (September 2024) for this project to develop a common data model, and undertake analysis of drivers of disease outcomes in pregnant women with syphilis and TB.

Themba Mutemaringa and Nicki Tiffin with CIDACS, Brazil visitors on a technical exchange (First Leg).

Themba Mutemaringa and Nicki Tiffin on the second leg of the technical exchange in Cape Town.

HDIG Team presenting at conferences during 2024. From left: Eddie Lulamba, Irene Muchada, Themba Mutemaringa and Tsaone Tamuhla.

RESEARCH PROJECTS AND COLLABORATIONS

1. Multimorbidity in the Western Cape

The focus of this project is the management, integration and analysis of routine health data from health care clients attending government health facilities in the Western Cape.

1.1 Establishing a virtual population cohort using routine health data for epidemiological analyses

Nature and Purpose: We are analysing anonymised and perturbed routine health data from a virtual cohort of all health care clients who attended any government health facility in the Khayelitsha sub-district in Cape Town during 2016 and 2017. Analysis of this data provides insights into multimorbidity in this population group that suffers from a high burden of both infectious and non-communicable diseases. In addition, we have continued to work in collaboration with the Provincial Health Data Centre (PHDC) at the Western Cape Department of Health in epidemiological and service-delivery projects working with routine health data from the Western Cape.

Researchers: Tsaone Tamuhla, Sohail Simon, Florence Phelanyane.

Outputs in the past 12 months:

Dixon J, Morton B, Nkhata MJ, Silman A, Simiyu IG, Spencer SA, et al. including Tiffin N. (2024) Interdisciplinary perspectives on multimorbidity in Africa: Developing an expanded conceptual model. PLoS Glob Public Health 4(7): e0003434. **https://doi.org/10.1371/journal.pgph.0003434**

Osei-Yeboah R, Ngwenya O, Tiffin N. Kidney function in healthcare clients in Khayelitsha, South Africa: Routine laboratory testing and results reflect distinct healthcare experiences by age for healthcare clients with and without HIV (2024). PLoS Global Public Health, 2024 May, 4(5): e0002526. **https://doi.org/10.1371/journal.pgph.0002526**

Future Direction: In this project, we will continue to build on appropriate methodologies for analysis of complex routine health data in order to inform evidence-based care. Sohail Simon will continue working within this project to further explore analysis using latent factor analysis as well as machine learning approaches to analyse risk factors for kidney disease in this population group. In addition, we will be using this dataset in our preliminary data modelling and epidemiological analyses for our new programme investigating drivers of gestational diabetes and pre-eclampsia in African populations.

1.2 Piloting a virtual genotyped population cohort linking genotype and routine health data:

Nature and Purpose: With informed consent from health care clients, it is possible to link individuals' genotype data with their routine health data in order to establish a genotyped virtual cohort with complex clinical phenotype data that can be updated into the future using routine health records. We have piloted this approach by establishing the informed consent process, sample and data workflows to create, in the first instance, a virtual genotyped cohort for patients with diabetes in Cape Town. UKRI/MRC has funded a pilot study for proof-of-principle establishment of a virtual genotyped population cohort using this approach, as well as establishment of the infrastructure to return actionable findings from the analysis to clinicians at the Western Cape Department of Health. In addition to myself and Tsaone, Ms Irene Muchada has been working as a genetic counsellor on this programme throughout 2024, successfully establishing excellent working relationships with our collaborators at the Nephrology Unit at Groote Schuur Hospital, our ongoing collaborators at the Diabetes Clinic at Groote Schuur Hospital, and our new collaborators at the UWC Dentistry Faculty clinics. We successfully completed the pilot project in December 2024 and will continue to work with collaborators across a variety of projects as new approaches to genomic data analyses are applied to genomic data from African populations. Ms Tiffany Fredericks is also joining this project to work with the VCAMM data for her MSc (Bioinformatics) which will start in January 2025.

Researchers: Tsaone Tamuhla, Nicki Tiffin, Irene Muchada, Tiffany Fredericks.

Outputs in the past 12 months:

Tamuhla T, Coussens AK, Abrahams M, Blumenthal MJ, Lakay F, Wilkinson RJ, Riou C, Raubenheimer P, Dave JA, Tiffin N (2024) Implementation of a genotyped African population cohort, with virtual follow-up: A feasibility study in the Western Cape Province, South Africa. **https://wellcomeopenresearch.org/articles/9-620/v1.**

Future Direction: A proof-of-concept pilot project has now been completed, with development of the analysis pipeline with an appropriate informed consent process, collection of buccal swabs and preparation of high-quality DNA sample, generation and analysis of genotype data, and linkage to routine health data for phenotype definition. We are currently completing the building of our data infrastructure to ensure that actionable genetic findings from these studies are appropriately returned to the Western Cape Department of Health to inform evidence-based care. We have a stakeholder meeting taking place in early 2025 and are actively seeking new funding resources to take this work forward.

1.3 Optimising data linkage of African health client data

Nature and Purpose: When integrating individualised data from a variety of sources, different linkage algorithms can be used to ensure the most accurate linkage and de-duplication of records that originate from the same individual. Most of these algorithms, however, have been developed using Euro-centric datasets and do not necessarily perform as well in African contexts, and we are working on a systematic approach to improve data linkage for routine health data. Themba Mutemaringa has continued to take this work forward by successfully developing a curated test dataset, and ongoing benchmarking of existing linkage algorithms using African data.

Researcher: Themba Mutemaringa

Collaborators: Andrew Boulle, Alexa Heekes, Western Cape Department of Health.

Outputs in the past 12 months:

Mutemaringa T, Boulle A and Tiffin N. (2024) Data governance for ethical usage of linked routine health data in South Africa: balancing privacy and data sharing. Conference Proceedings: International Journal of Population Data Science, 9(5). doi:10.23889/ijpds.v9i5.2721.

Future Direction: In this project, Themba Mutemaringa will continue to lead the validation and benchmarking of existing linkage and de-duplication algorithms, and will develop and test new methodologies and algorithms to best achieve de-duplication and data linkage.

2. Facilitating ethical and equitable sharing of African data and biospecimens: the African Data and Biospecimen Exchange - ADBEx

Nature and Purpose: Towards the end of 2021 I established a programme, funded by the Calestous Juma fellowship from the Bill & Melinda Gates Foundation, to build an online platform to facilitate ethical and equitable secondary sharing of African data and biospecimen resources. The query-able platform will store meta-data about sample and data collections, submitted by researchers who hold those resources, and will enable direct sharing partnerships to be established between resource providers and resource consumers without centralisation of samples or data. It will also provide online infrastructure for drawing up and recording data-sharing, benefit-sharing and collaborative agreements as well as building a repository of data and biospecimen guidelines and governance resources.

The development of the ADBEx platform has reached completion at the end of 2024, thanks to the highly competent and professional work of our partners in User Interface - Hominum (**https://www.hominum.digital/**) and software developers MethodLab.
We have developed two related projects that are linked to ADBEx: Tsaone created a stand-alone, browser-based application that can be used to build fit-for-purpose

data- and biospecimen-sharing agreements. Whilst this functionality will be incorporated into the ADBEx platform we were keen to also test this concept and introduce it to a wider audience. The app can be found here **https://adbex-template-mou-builder.streamlit.app/** (just wake up the app if it is sleeping!).

We completed and published a collaborative project with PHA4GE to create a Data Sharing Accord - a document that outlines the basic "common sense" principles for sharing microbial data, that end users can adopt as a common ground for general accepted principles for sharing data for secondary use. We presented this Accord to various stakeholders, at the PHA4GE Conference in Cape Town and also at the Wellcome AMR Workshop in Johannesburg, and were able to incorporate feedback from the community in the version of the Accord that we published.

We are still in the process of building a Data Sharing license. Drawing inspiration from the Creative Commons licensing structure, we have developed a prototype for a similar type of license that can be used to effectively and unambiguously communicate the nuanced sharing requirements for individual datasets. In 2025, this will be operationalised as an online browser-based license generator.

Researchers: Nicki Tiffin, Eddie Lulamba, Tsaone Tamuhla.

Collaborators: Hominum Global, **www.hominum.global** - User Interface and Software Specification and Development; MethodLab, **https://methodlab.io/** - Software development.

PHA4GE collaborators: Peter van Heusden, Emma Griffiths, Michelle Nichols, Anja Bedeker.

Outputs in the past 12 months:

Story of the UI Development by Hominum: **https://www.hominum.digital/blog/adbex**
MOU Builder for data- and biospecimen-sharing agreements: **https://adbex-template-mou-builder.streamlit.app/**

Griffiths EJ, van Heusden P, Tamuhla T, Lulamba ET, Bedeker A, Nichols M, Christoffels A, Tiffin N (2024). The PHA4GE Microbial Data-Sharing Accord: establishing baseline consensus microbial data-sharing norms to facilitate cross-sectoral collaboration. BMJ Glob Health. 2024 Oct 30;9(10): e016474. doi: 10.1136/bmjgh-2024-016474.

Future Direction: We are excited to test and launch the ADBEx platform in 2025 and will be undertaking a series of roadshows to introduce the platform to African researchers across the Continent, and to help individuals to start uploading the metadata for their sample collections and datasets.
We will also be building an online tool for creating data-sharing licenses for end-users wishing to stipulate what kinds of data use are acceptable for the datasets that they are sharing.

The Calestous Juma Fellowship
https://gcgh.grandchallenges.org/article/nicki-tiffin

In November 2021 the Bill & Melinda Gates Foundation launched the Calestous Juma Fellowship program for scientists permanently located at African research organisations. I was one of fourteen recipients of these fellowships and I have been participating in a series of leadership trainings, networking opportunities and skills development as part of this program.

In April 2024 I travelled to the Bill & Melinda Gates Foundation in Seattle together with the other CJ Fellows for a week of networking and engagement with the Foundation. In addition, I was able to contribute to the programme of Calestous Juma Fellow Prof Jo-Ann Passmore at the conference "African Microbiomes in Health and Disease" in October 2024 in Cape Town. Furthermore, two of our collaborators in our new B&MGF funding to study gestational diabetes and pre-eclampsia in African populations are Calestous Juma Fellows Prof Annettee Nakimuli (Uganda) and Prof Moses Obimbo (Kenya). As always, the cohort of CJ Fellows are extremely inspirational, and it is an honour to work together with this amazing group of African scientists.

3. Ethics and governance to promote ethical and equitable health research in Africa

Nature and Purpose: In order to undertake ethical research in Africa and other regions of the global South, there are multiple considerations around risks, benefits, informed consent and equitable research. We explore different ways to better do research and to ensure that health research in Africa is ethical and equitable through the development of tools and frameworks. I also participate in multiple ethics and governance working groups of international consortia including PHA4GE (Chair: Ethics and Data Sharing Working Group), G2MC/IHCC (Co-chair: Policy and Advocacy working group). We have also some exciting outputs from the PHA4GE Ethics and Data Sharing sub-grant awardees who were previously funded to conduct research projects in ethics and data sharing.

Outputs in the past 12 months:

Tiruneh G, Yilma M, Wakuma B, Abdisa E, Bayisa L, Nichols M, Bedeker A, Tiffin N. Compliance with research ethics in epidemiological studies targeted to conflict-affected areas in Western Ethiopia: validity of informed consent (VIC) by information comprehension and voluntariness (ICV). BMC Med Ethics. 2024 Jan 18;25(1):9. doi:10.1186/s12910-024-01003-5.

Pillay N, Ncube N, Moopelo K, Mothoagae G, Welte O, Shogole M, Gwiji N, Scott L, Moshani N, Tiffin N, Boulle A, Griffiths F, Fairlie L, Mehta U, LeFevre A, Scott K. Translating the consent form is the tip of the iceberg: using cognitive interviews to assess the barriers to informed consent in South African health facilities. Sex Reprod Health Matters. 2023 Dec;31(4):2302553. doi: 10.1080/26410397.2024.2302553. Epub 2024 Jan 26.

Griffiths EJ, van Heusden P, Tamuhla T, Lulamba ET, Bedeker A, Nichols M, Christoffels A, Tiffin N (2024). The PHA4GE Microbial Data-Sharing Accord: establishing baseline consensus microbial data-sharing norms to facilitate cross-sectoral collaboration. BMJ Glob Health. 2024 Oct 30;9(10): e016474. doi: 10.1136/bmjgh-2024-016474.

Future Direction: We are currently developing a data sharing license to capture and communicate nuances of data sharing permissions, and to facilitate data sharing that respects conditions of sharing.

DR DOMINIQUE ANDERSON

HIGHLIGHTS OF 2024

- Supervisor of 2 MSc students
- Co-supervisor of 1 PhD student and 1 MSc student
- Lead in Data curation capacity development and User design

The research group is investigating several areas with an interest in metagenomics, data security and privacy, biomedical data governance, informatics of biobanking and application of AI in One Health. I have an ongoing role in a team focused on informatics solutions for biobanking with continued development, enhancement, and training in the open-source Baobab laboratory information management system. Information management and quality management remains a core area of activity in my research group with hopes to expand the collaborative network of developers and researchers. A collaboration has been established with the NHLS, to evaluate data workflows and metadata standards with a view to improve data quality in the public health setting. In relation to additional interests, I am involved in community engagement and effective science communication and training and am an active committee member of the DSI Medical Biobank Cluster, the TTO seed-fund committee and the POPIA code of conduct committee.

With regards to data privacy and governance, my collaborators and I combine expertise in biobanking, bioinformatics and big data, and law to investigate how both regulations and best practices influence the biomedical data and data sharing landscape in Africa. I am also currently working towards establishing networks within the agricultural, veterinary, medical, and environmental sphere in order to intensify research in the area of One Health and machine learning and investigating field-able molecular biology and '-omics' sequencing for pathogen genomic surveillance, as well as investigating venom bioinformatics and the associated governance of indigenous knowledge resources.

SOUTH AFRICAN NATIONAL BIOINFORMATICS INSTITUTE

RESEARCH PROJECT THEMES

Informatics Tools
- Biobank informatics - multiple collaborations for biocollection sample quality management and e-infrastructure with a focus on Baobab LIMS.
- Pathogen genome databases for impact in public health and pathogen surveillance.

Biomedical Data Governance
- Collaboration with biobank and law experts to evaluate data governance of biological diversity and human biomedical collections.
- Cybersecurity and data management in Africa.
- Metadata standardisation and impacts of data management in public health.

Venom Bioinformatics
- Bioinformatic evaluation of potential therapeutics and targets from venomous animals

RESEARCH COLLABORATIONS

1. Data Governance and the POPI Act
Collaborating Parties:
ASSAf
Nature and Purpose: To investigate the impact of local and international regulation on the sharing of biomedical research data in Africa. Drafting committee member for the POPIA code of conduct for research.
Output in the last 12 months: Code of conduct submitted to the Information Regulator. A framework is in the process of being developed.
Future Direction: Ongoing engagement and amendments to the framework.

2. Venom Bioinformatics
Collaborating Parties:
Professor Bryan Maritz - UWC
Dr Ciara Staunton - EURAC
Dr Carmen Swanepoel - University of Stellenbosch
Dr Ruben Cloete - SANBI, UWC
Nature and Purpose: Investigating the therapeutic potential of venoms through bioinformatic analysis of protein-protein and protein-ligand interactions.
Output in the last 12 months: Postgraduate student currently enrolled.
Future directions: Seeking funding to support a venom biodiversity biorepository for further research.

3. Biobank Informatics and Data Quality Management/ Data Curation
Collaborating Parties:
Prof Alan Christoffels - SANBI, UWC
Africa CDC
PHA4GE Consortium
Medical Biorepositories SA
Nature and Purpose: Maintenance and enhancement of Baobab LIMS. Entrepreneurship focused on hybrid social and sustainability models for QM software. Standardisation of e-infrastructure. Evaluating the impact of data standardisation in public health and mechanisms to improve data quality across workflows. Development of data platforms for sharing and archiving pathogen data. Strengthening capacity in data curation and development of metadata standards. Facilitating the development of a network of medical biorepositories in South Africa.
Output in the last 12 months: Successful completion of Phase 1 of the African Pathogen Data Sharing and Archive platform with Phase 2 started. Version 1 of the CholGen metadata standard completed and in review. Core module for Baobab LIMS lite developed. The MBiRSA network is being established with ongoing engagements and collaboration and the first summer school training initiative has been completed.
Future Direction: New collaborations within Africa as well as training initiatives on the continent. User design thinking in software development as well as building a data curation cohort with regional representation on the continent. Phase 2 of the APA. Biodiversity module of the LIMS Lite core.

4. One Health
Collaborating Parties: Seeking collaborative partners.
Nature and Purpose: Multidisciplinary approaches to understanding infectious diseases. Development of metadata standards, novel AI based tools for surveillance and species cross-over and translational research in One Health. Examining field-forward technologies bringing molecular biology and '-omics' sequencing to the sample.

DR RUBEN CLOETE

HIGHLIGHTS OF 2024

- Currently, I supervise four PhD students, Mohammed Hassan, Jeremy Burgess, Adetutu Akinnuwesi and Keaghan Brown. I co-supervise a PhD student, Dewald Schoeman in Medical Biosciences, UWC.
- I also supervise one MSc student Lunathi Ndlondlo.
- My UWC Masters' student, Keaghan Brown, graduated in 2023.
- Grants/Funding Awarded: Co-Principal investigator on a SAMRC RCDI grant with Prof Megan Shaw for 2021-2024.

RESEARCH PROJECTS

My primary research interests focus on understanding drug resistance in *Mycobacterium tuberculosis* and Human Immunodeficiency Virus (HIV-1) integrase protein and the identification of causal variants in Parkinson's disease development. Tuberculosis and HIV-1 drug resistance is a huge problem in South Africa and calls for the identification of newer drugs to curb the spread of these diseases. The use of computational methods to speed up the process of drug discovery can reduce the cost and time spent pursuing drugs that later fail in clinical trials.

We have developed computational pipelines within my laboratory to interrogate drug targets to screen for drugs that inhibit enzyme targets that can be purchased and tested *in vitro* for activity against *Mycobacterium tuberculosis*. We also use computational methods to validate mutations and their effects on protein structure and function. This is also very useful to determine if a drug remains within an enzyme active pocket and if the mutation results in reduced binding or no binding. The result of this research may provide improve treatment regimens for South African populations to combat infectious disease and non-communicable diseases like Parkinson's. Recently, my group also embarked on identifying novel drugs to treat SARS-CoV-2 coronavirus infections using computational methods. Other collaborative projects involve studying drug transporter proteins associated with Diabetes Mellitus and investigating structural differences between human Coronavirus envelope proteins to understand pathogenicity between virulent and non-virulent strains.

RESEARCH PROJECT THEMES

Pathogen resistance
- NGS sequencing of Tuberculosis isolates and HIV plasma samples allowed the identification of novel mutations.
  To identify mutations associated with drug resistance we develop structural computational pipelines to further our understanding of HIV-1 and tuberculosis drug resistance.
- To identify drug resistance pathways in HIV-1 using Phylogenetics.

Drug discovery
- Previous pathway mapping and gene prioritisation methods allowed us to identify known and novel *Mycobacterium tuberculosis* drug targets. Here we exploit the drug targets using virtual screening methods to identify novel drugs to treat drug resistant tuberculosis. The drugs identified are tested for activity against *Mycobacterium tuberculosis* using a bioluminescent reporter assay system. *In silico* discovery of compounds with activity against the novel SARS-coronavirus and in vitro testing.

Human disease genomics
- Whole genome sequencing identified variants that co-segregate within South African families with Parkinson's disease. Here we use various software tools to prioritise these variants to identify causal variants that might be associated with Parkinson's disease.

Machine learning approaches
- Using machine learning methods to identify putative compounds with activity against *Mycobacterium tuberculosis* and HIV.

Diagnostic tool development
- Develop structural methods to aid in the identification of biomarkers useful for Tuberculosis and Ebola virus diagnosis.

RESEARCH COLLABORATIONS

1. Novel drug discovery against Mycobacterium tuberculosis and SARS-Cov-2 coronavirus
Collaborating Parties:
Prof Samuel Egieyeh - School of Pharmacy, UWC
Prof Samantha Samson and Dr Melanie Grobbelaar - University of Stellenbosch
Prof Andrej Sali - University of California San Francisco, USA
Prof John Irwin - University of California San Francisco, USA
Prof Megan Shaw - Medical Biosciences, UWC
Dr Bianca Gordon - Medical Biosciences, UWC
Nature and Purpose:
To identify drugs with a new mode of action against *Mycobacterium tuberculosis* protein.
Target proteins essential for transcription and translation in SARS-CoV-2 coronavirus.

Output in the last 12 months:
Akinnuwesi A, Cloete R and Egieyeh SA, State-of-the-art Strategies to Prioritise Mycobacterium tuberculosis Drug Targets for Drug Discovery using a Subtractive Genomics Approach. Frontiers in Drug Discovery, 3, p.1254656. (2023).
Future Direction:
Draft two manuscripts one targeting *Mycobacterium tuberculosis* drug target Rv2196 and a follow up article on Rv2421c to identify novel drug molecules to treat Tuberculosis infections.

2. Structural impact of resistance associated mutations in the South African HIV-1C integrase protein
Collaborating Parties:
The late Dr Graeme Jacobs - University of Stellenbosch (deceased)
Nature and Purpose:
Firstly, to understand genetic diversity in HIV-1 subtype C integrase gene in South African HIV-1 infected patients and recombinant subtype AG in Cameroonian patients. Secondly, to determine if second-line integrase inhibitors will be a viable option for South African and Cameroonian patients infected with HIV-1.
Output in the last 12 months:
None
Future Direction:
One manuscript submitted and under review focusing on the development of an automated pipeline to investigate the effects of mutations on HIV-1C and HIV-1 AG recombinant Integrase structure.

3. Prioritising mutations identified in South African Parkinson's disease patients using structural methods
Collaborating Parties:
Prof Soraya Bardien - University of Stellenbosch
Nature and Purpose:
To identify novel genes associated with Parkinson's disease development using Whole
Exome sequencing and using structural computational methods to understand the impact of
mutations on protein structure and function.
Output in the last 12 months:
*Cuttler K, de Swardt D, Engelbrecht L, et al. Neurexin 2 p.G849D variant, implicated in Parkinson's disease, increases reactive oxygen species, and reduces cell viability and mitochondrial membrane potential in SH-SY5Y cells. J Neural Transm 129, 1435-1446 (2022). **https://doi.org/10.1007/s00702-022-02548-8***
Future Direction:
Two manuscripts under preparation for publication following up on neurexin and neuroligin interaction associated with PD and another on a novel gene candidate identified in a South African family with PD using exome sequencing.

4. Haplotype variation within South African Xhosa population and its effect on diabetic treatment
Collaborating Parties:
Prof Mongi Benjeddou - Biotechnology Department, UWC
Nature and Purpose:
Understand the effect of haplotype variation on SLCAA2 transporter protein and the binding of diabetic drugs to SLCAA2 within a South African Xhosa population.
Output in the last 12 months:
None
Future Direction:
None to report.

5. Structural studies of more and less virulent coronavirus envelope proteins to understand human host interaction and severity of disease
Collaborating Parties:
Prof Burtram Fielding - Department of Medical Biosciences, UWC
Nature and Purpose:
Perform topology predictions, structural modelling as well as simulation studies to understand the structural differences between four coronavirus envelope proteins and their interaction with the human host PALS1 protein.
Output in the last 12 months:
None
Future Direction:
Manuscript in preparation.

6. The identification and characterisation of DNA aptamers for application in diagnosis of infectious diseases
Collaborating Parties:
Prof Mervin Meyer - Biotechnology Department, UWC
Nature and Purpose:
The aim of the research is to develop multiplex lateral flow devices (LFDs) for the detection of serum human biomarker proteins for TB and Ebola diagnosis using DNA aptamers.
Output in the last 12 months:
Martin DR, Mutombwera AT, Madiehe AM, Onani MO, Meyer M and Cloete R, 2024. Molecular modelling and simulation studies of SELEX-derived high-affinity DNA aptamers to the Ebola virus nucleoprotein. Journal of Biomolecular Structure and Dynamics, pp.1-18.
Future Direction:
To build an mysql database containing 3D structures of DNA aptamers for diagnostic purposes.

7. Ebola virus 3D database (EBOV-3D): Understanding the Ebola virus proteome and evaluating possible drug targets
Collaborating Parties:
Dr Arun Pandurangan - Department of Medicine, University of Cambridge
Prof Placide Mbala- Kingebeni - National Institute of Biomedical Research, Democratic Republic of the Congo, University of Kinshasa
Prof/Sir Tom Blundell - Department of Medicine, University of Cambridge
Dr Hocine Bendou - Division of Computational Biology, University of Cape Town
Dr Adetayo Emmanuel Obasa - Faculty of Medicine and Health Sciences, University of Stellenbosch
Nature and Purpose:
The aim of this project is to provide a user-friendly and easily accessible 3D webserver, so that end-users can navigate, inspect and download the 3D structural proteome data, visualise modelled oligomeric complexes, analyse pockets of modelled structures, and investigate Ebola virus human-protein interactions, mutations and protein–ligand docking interfaces.
Output in the last 12 months:
None
Future Direction:
Build the postgre SQL backend of the database as well as the front end and prepare a manuscript.

PROF GORDON HARKINS

My research focuses on the evolution and molecular epidemiology of DNA and RNA infectious viral pathogens where I seek to determine the evolutionary underpinnings of the emergence and spread of the numerous viral diseases that seriously threaten the health and food security of Africa and the rest of the developing world.

In early 2020, in response to the emergence and rapid early spread of COVID-19, we immediately shifted our research focus to tackle the unprecedented volume of SARS-CoV-2 genomic and spatial data that was being generated and shared with the scientific community to gain real-time insights into the virus transmission during a viral pandemic.

RESEARCH THEMES

Human immunodeficiency virus (HIV-1)
- HIV-1 latent viral reservoir dynamics

SARS-CoV-2
- The characterisation and clinical manifestation of the SARS-CoV-2 outbreak in Uruguay
- The evolutionary dynamics of SARS-CoV-2 lineages during the first epidemic wave in New York City
- The role of natural selection in the emergence and ongoing evolution of the N501Y lineages (V1-V3) of SARSCoV2
- Selection Analysis Identifies Unusual Clustered Mutational Changes in Omicron Lineage BA.1 that likely impact Spike Function
- Conserved recombination in coronavirus subgenera
- T cell immune escape mutations at HLA-binding anchor motifs found in SARS-CoV-2 epitopes from genomes isolated in South Africa and sub-Saharan Africa

RESEARCH PROJECTS AND COLLABORATIONS

1. Selection Analysis Identifies Unusual Clustered Mutational Changes in Omicron Lineage BA.1 that Likely Impact Spike Function.

Collaborating Parties:
Darren P Martin, Arne De Klerk - Department of Integrative Biomedical Sciences, UCT
Philippe Lemey - Department of Microbiology, Immunology and Transplantation, Rega Institute, Belgium
Steven Weaver, Stephen D Shank, Sergei L Kosakovsky Pond - Temple University, Pennsylvania, USA
Houryiah Tegally, Emmanuel James San, Eduan Wilkinson, Jennifer Giandhari, Richard J Lessells, Anton Nekrutenko, Tulio De Oliveira - KwaZulu-Natal Research Innovation and Sequencing Platform (KRISP), School of Laboratory Medicine & Medical Sciences, University of KwaZulu- Natal
Ben Murrell – Karolinska Institutet, Stockholm, Sweden
Carolyn Williamson - Division of Medical Virology, UCT
Spyro Lytras, Oscar S MacLean, David L Robertson, Richard J Orton – MRC-University of Glasgow
Alexander G Lucaci - Department of Biology, Temple University, Philadelphia
Wolfgang Maier, Bjorn Gruning - Department of Computer Science, University of Freiburg, Germany
Maciej F Boni - Center for Infectious Disease Dynamics, Department of Biology, Pennsylvania State University
Cathrine Scheepers, Jinal N Bhiman - National Institute for Communicable Diseases (NICD) of the National Health Laboratory Service, (NHLS), SA MRC Antibody Immunity Research Unit, School of Pathology, University of the Witwatersrand
Josie Everatt, Daniel G Amoako - National Institute for Communicable Diseases (NICD) of the National Health Laboratory Service, (NHLS)
James Emanuel San, Jennifer Giandhari - KRISP, School of Laboratory Medicine & Medical Sciences, University of KwaZulu- Natal
Alex Sigal, Ravindra K Gupta - Africa Health Research Institute
Nei-yuan Hsiao - Division of Medical Virology, UCT
Anne von Gottberg - NICD
Robert W Shafer - Division of Infectious Diseases, Stanford University
Robert J Wilkinson - Wellcome Center for Infectious Diseases Research in Africa, UCT
Brian Trevor Sewell - Institute for Infectious Diseases and Molecular Medicine, UCT
Allison J Greaney, Tyler N Starr, Jesse D Bloom - Fred Hutchinson Cancer Research Center

Nature and Purpose:
To determine what drives the mutational changes in the Omicron variant of concern (VOC).
Output in the last 12 months:
One paper published in 2022 in the Journal Molecular Biology and Evolution.
Future Direction:
This is an ongoing collaboration.

2. Conserved recombination in coronavirus subgenera

Collaborating Parties:
Darren P Martin, Arné de Klerk, Phillip Swanepoel, Mpumelelo Zondo, Isaac Abodunran - Institute of Infectious Diseases and Molecular Medicine, UCT
Rentia Lourens - Neuroscience Institute, UCT
Spyros Lytras, Oscar A MacLean, David Robertson - MRC-University of Glasgow Centre for Virus Research
Sergei L Kosakovsky Pond, Jordan D Zehr - Institute for Genomics and Evolutionary Medicine, Temple University, Philadelphia
Venkatesh Kumar, Ben Murrell - Department of Microbiology, Tumor and Cell Biology, Karolinska Institutet
Michael J. Stanhope - Department of Population and Ecosystem Health, College of Veterinary Medicine, Cornell University
Nature and Purpose:
To determine whether recombination is non-random in coronavirus subgenera.
Output in the last 12 months:
One article has been accepted in the journal Virus Evolution.
Future Direction:
This is an ongoing collaboration.

3. HIV-1 Latent Viral Reservoir Dynamics

Collaborating Parties:
Melissa-Rose Abrahams, Lynn Tyers, David Matten, Deelan Doolabh, Colin Anthony, Carolyn Williamson - Institute of Infectious Disease and Molecular Medicine, UCT
Salim Abdool Karim - CAPRISA, University of KwaZulu- Natal
Andrew Redd - Johns Hopkins University, School of Medicine
Siposethu Matzishana, Nkosazama Nyembezi - SANBI, UWC
Nature and Purpose:
To study the viral dynamics involved in HIV-1 subtype C latent reservoir formation maintenance and evolution to better understand the determinants thereof.
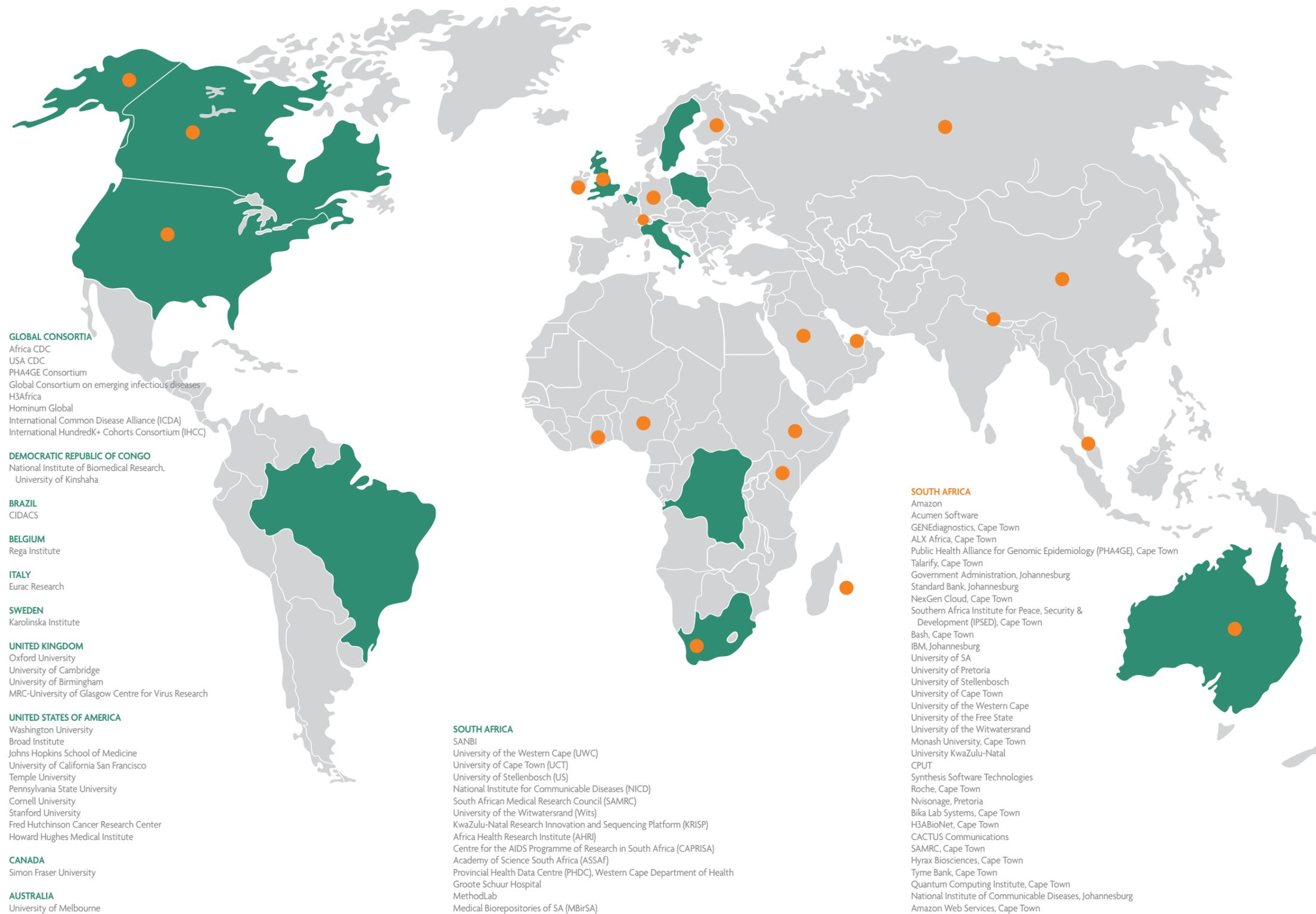**Output in the last 12 months:**
None.
Future Direction:
This project is funded by the National Institutes of Health (NIH) USA and the South African Medical Research Council and will run between 2020 and 2025.

COLLABORATIONS

SANBI researchers have established a vast network of partnerships and collaborations with peers all over the world, indicated in green on the map below.

ALUMNI

SANBI has produced many alumni over the years who have taken their research to institutions and corporates all over the world, shown in orange on the map below.

GLOBAL CONSORTIA
Africa CDC
USA CDC
PHA4GE Consortium
Global Consortium on emerging infectious diseases
H3Africa
Hominum Global
International Common Disease Alliance (ICDA)
International HundredK+ Cohorts Consortium (IHCC)

DEMOCRATIC REPUBLIC OF CONGO
National Institute of Biomedical Research,
   University of Kinshaha

BRAZIL
CIDACS

BELGIUM
Rega Institute

ITALY
Eurac Research

SWEDEN
Karolinska Institute

UNITED KINGDOM
Oxford University
University of Cambridge
University of Birmingham
MRC–University of Glasgow Centre for Virus Research

UNITED STATES OF AMERICA
Washington University
Broad Institute
Johns Hopkins School of Medicine
University of California San Francisco
Temple University
Pennsylvania State University
Cornell University
Stanford University
Fred Hutchinson Cancer Research Center
Howard Hughes Medical Institute

CANADA
Simon Fraser University

AUSTRALIA
University of Melbourne

SOUTH AFRICA
SANBI
University of the Western Cape (UWC)
University of Cape Town (UCT)
University of Stellenbosch (US)
National Institute for Communicable Diseases (NICD)
South African Medical Research Council (SAMRC)
University of the Witwatersrand (Wits)
KwaZulu-Natal Research Innovation and Sequencing Platform (KRISP)
Africa Health Research Institute (AHRI)
Centre for the AIDS Programme of Research in South Africa (CAPRISA)
Academy of Science South Africa (ASSAf)
Provincial Health Data Centre (PHDC), Western Cape Department of Health
Groote Schuur Hospital
MethodLab
Medical Biorepositories of SA (MBirSA)

SOUTH AFRICA
Amazon
Acumen Software
GENEdiagnostics, Cape Town
ALX Africa, Cape Town
Public Health Alliance for Genomic Epidemiology (PHA4GE), Cape Town
Talarify, Cape Town
Government Administration, Johannesburg
Standard Bank, Johannesburg
NexGen Cloud, Cape Town
Southern Africa Institute for Peace, Security & Development (IPSED), Cape Town
Bash, Cape Town
IBM, Johannesburg
University of SA
University of Pretoria
University of Stellenbosch
University of Cape Town
University of the Western Cape
University of the Free State
University of the Witwatersrand
Monash University, Cape Town
University KwaZulu-Natal
CPUT
Synthesis Software Technologies
Roche, Cape Town
Nvisonage, Pretoria
Bika Lab Systems, Cape Town
H3ABioNet, Cape Town
CACTUS Communications
SAMRC, Cape Town
Hyrax Biosciences, Cape Town
Tyme Bank, Cape Town
Quantum Computing Institute, Cape Town
National Institute of Communicable Diseases, Johannesburg
Amazon Web Services, Cape Town

ETHIOPIA
Africa CDC, Addis Ababa

KENYA
Deimos ML, Nairobi County
Egerton University, Njoro
International Livestock Research Institute
The Technical University of Kenya (TU-K)
University of Nairobi
Icipe, Duduville

GHANA
University of Ghana

NIGERIA
Lagos State University

MAURITIUS
Mahatma Gandhi Institute

USA
National Institutes of Health, Bethesda, Maryland
Washington University, St Louis, Missouri
Ann Arbor, Michigan
Invitae, San Francisco
Pharma US, Seattle, Washington
Mayo Clinic
The Bioinformatics CRO
Princeton University
Novocraft Technologies
Brigham and Women's Hospital: Boston
New Jersey DoH's Public Health and Environmental Laboratories
Georgetown University, Washington DC
Jamf, Minnesota
University of Connecticut

CANADA
Western University: London

UK
Illumina, Cambridge
Albemarle College
AstraZeneca, London
University of Surrey
University of Westminster, London

FINLAND
Espoo

IRELAND
National University of Ireland

SWITZERLAND
Roche
World Bank

GERMANY
Greater Leipzig Area
Evotec, Hamburg

UNITED ARAB EMIRATES
Dubai

RUSSIA
South Ural State University, Chelyabinsk

SAUDI ARABIA
KAUST

SINGAPORE
Nanyang Technology University
University of Singapore

NEPAL
Milo Logic, Kathmandu

AUSTRALIA
SpeeDx
University of Melbourne

CHINA
Shanghai World Foreign Language Primary School, Shanghai

FINANCIALS

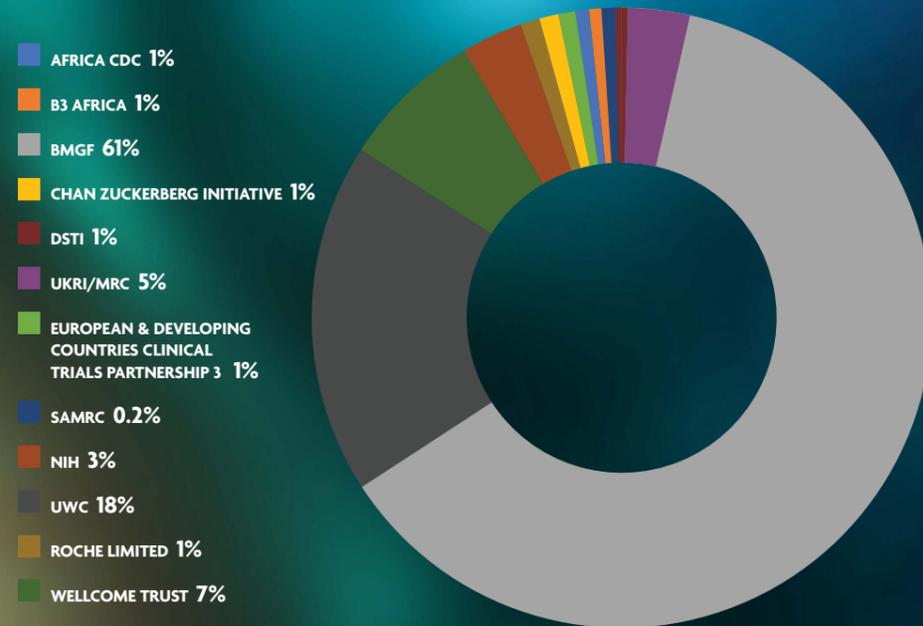SANBI's income and expenditure trends for 2024 are shown in this section.

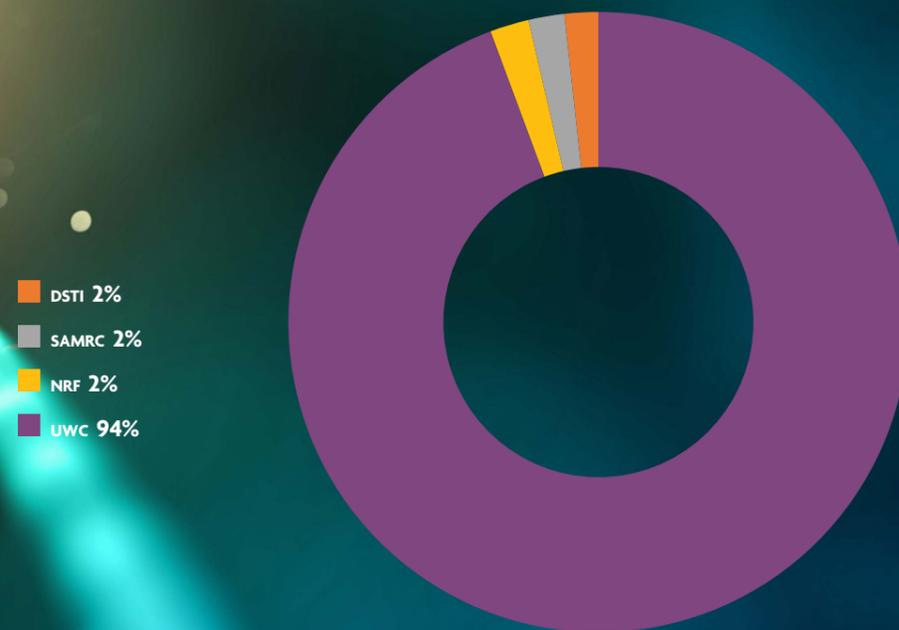**FIGURE 1. DISTRIBUTION OF INCOME FROM ALL SOURCES 2024**

Legend:
- AFRICA CDC 1%
- B3 AFRICA 1%
- BMGF 61%
- CHAN ZUCKERBERG INITIATIVE 1%
- DSTI 1%
- UKRI/MRC 5%
- EUROPEAN & DEVELOPING COUNTRIES CLINICAL TRIALS PARTNERSHIP 3 1%
- SAMRC 0.2%
- NIH 3%
- UWC 18%
- ROCHE LIMITED 1%
- WELLCOME TRUST 7%

FIGURE 2. DISTRIBUTION OF INCOME RECEIVED FROM SA SOURCES 2024

Legend:
- DSTI 2%
- SAMRC 2%
- NRF 2%
- UWC 94%
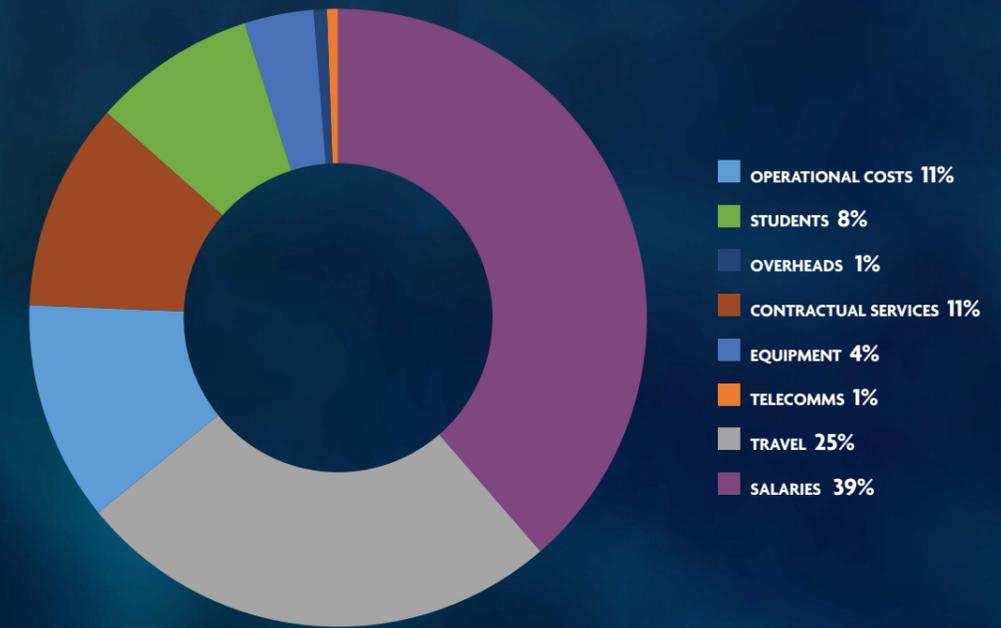
FIGURE 3. EXPENDITURE IN 2024

Legend:
- OPERATIONAL COSTS 11%
- STUDENTS 8%
- OVERHEADS 1%
- CONTRACTUAL SERVICES 11%
- EQUIPMENT 4%
- TELECOMMS 1%
- TRAVEL 25%
- SALARIES 39%

FIGURE 4. INCOME VS EXPENDITURE 2015 – 2024

YEAR	EXPENSES	INCOME
2024	40 286 266	51 433 340
2023	35 438 006	23 330 519
2022	24 849 024	37 586 181
2021	16 723 940	31 383 413
2020	10 406 504	12 045 007
2019	13 849 462	27 721 406
2018	15 981 456	13 043 930
2017	20 844 456	21 335 183
2016	15 665 001	20 495 324
2015	13 393 931	19 641 569

Note: Funding received in a calendar year often reflects projects that span multiple years.

FUNDERS

Research Institute at UWC since 1997

National Research Foundation funding since 1998

South African Medical Research Council Bioinformatics Unit since 2002

Bill & Melinda Gates Foundation funding since 2014

Technology Innovation Agency funding since 2016

UK Research Innovation/ Medical Research Council (UKRI/MRC) funding since 2023

Funding since 2021

SANBI

South African National Bioinformatics Institute