AWI-Gen
Wits-INDEPTH Partnership
Genomic and environmental risk factors for cardiometabolic disease in Africans
Collaborative Centre

Members at this meeting:
Marianne Alberts
Nadia Carstens
Nigel Crowther
Zane Lombard
Michele Ramsay
Himla Soodyall
Ernest Tambo
Alisha Wade
Ananyo Choudhury
Daniel Achinko

Co-PI
Osman Sankoh
INDEPTH - International Network for the Demographic Evaluation of Populations and their Health in low and middle-income Countries

H3Africa NIH funded project
Top 5 leading **risk factors** for burden of disease (DALYs) in South Africa

- High BMI as a risk factor
  - Ghana 7th
  - Kenya 14th
  - Burkina Faso not in top 15

---

H3Africa – Johannesburg - October 2013
Genomic and environmental contributions to complex traits in African populations

• High genetic diversity and low LD
• Population structure
• Inter-ethnic variation
  – Global risk variants
  – Population specific risk variants
• Environments and diets
• Few studies in African populations
  – African Americans reflect predominantly west African ancestry
AWI-Gen Collaborative Center

• Wits-INDEPTH Partnership
• Project Aims
• Phenotyping & sample collection
• Data Management
• Important outcomes
• Time line
AWI-Gen overview

**AWI-Gen**
(African Wits-INDEPTH Partnership for the Genomic study of body composition and cardiometabolic disease risk)
The INDEPTH Network of Health and Demographic Surveillance Systems

*Founded 1998, constituted 2002*

Osman Sankoh
Executive Director of INDEPTH

Informing global efforts to improve the health and wellbeing of low and middle-income populations

Stephen Tollman
Principal Scientist

Kathleen Kahn
Board Member

H3Africa – Johannesburg - October 2013
Low- and Middle-Income Countries with INDEPTH member centres
Running Health and Demographic Surveillance Systems (HDSSs)

Currently 43 HDSSs in 20 countries
30 HDSSs in Africa
12 HDSSs in Asia
1 HDSS in Oceania

Over 3,200,000 people under surveillance
AWI-Gen study sites in Africa:

- Ghana, Navrongo (Rural)  
  Abraham Oduro

- Burkina Faso, Nanoro (Rural)  
  Halidou Tinto

- Kenya, Nairobi (Urban)  
  Catherine Kyobutungi

- South Africa, Soweto (Urban)  
  Shane Norris

- South Africa, Agincourt (Rural)  
  Stephen Tollman

- South Africa, Dikgale (Rural)  
  Marianne Alberts
1. Pilot Project – Soweto (~2000 individuals)
2. Population structure and genome architecture
3. Genetic and environmental contributions to body composition across six Centres in Africa (~12 000 individuals)
Urban Soweto study

• Study design
  – Population sample
  – Age 40 to 60 yrs
  – Male & Female
  – Body composition phenotype

• Platform
  – Candidate gene/region assessment
  – Metabochip

• Analysis
  – Correlations with quantitative traits related to body composition

• Progress
  – ~1000 females
  – 40 to 60 years
  – Phenotyped
  – Genotyped

• Next steps
  – Prepare DNA from next 1000 individuals for genotyping
  – Data analysis
Advantages:
- Cost effective
- Fine mapping (previous associations)
- Replication study
- Rapid results
- Good training opportunity

Disadvantages:
- SNP choice predominantly Eurocentric
- Previous associations not in African populations
- Limits novel discovery
Role of pilot project in capacity development

• PhD student – Venesa Pillay
• Soon to join: PhD student and postdoctoral fellow
• Complexity of the data - genotyping and phenotype (including biochemical markers) lends itself to multiple enquiry
Aim 2: Population structure and genomic architecture

- **Study design**
  - 30 unrelated trios
  - 40 unrelated individuals

- **Genotyping Platform**
  - uncertain

- **Outcome**
  - HapMap equivalent for each population
  - Common variant allele frequencies

- **Challenge**
  - Which populations to test
# Complexity of population structure

Africa
2,146 languages spoken (30.2% of all living languages)
789,138,977 people (12.7% of all people)

<table>
<thead>
<tr>
<th>Country</th>
<th>No. Living languages</th>
<th>Indigenous languages</th>
<th>Immigrant languages</th>
<th>Population size</th>
<th>Diversity Index</th>
</tr>
</thead>
<tbody>
<tr>
<td>Burkina Faso</td>
<td>70</td>
<td>68</td>
<td>2</td>
<td>10.9 M</td>
<td>0.768</td>
</tr>
<tr>
<td>Ghana</td>
<td>86</td>
<td>81</td>
<td>5</td>
<td>25.1M</td>
<td>0.835</td>
</tr>
<tr>
<td>Kenya</td>
<td>72</td>
<td>67</td>
<td>5</td>
<td>37.6M</td>
<td>0.928</td>
</tr>
<tr>
<td>South Africa</td>
<td>44</td>
<td>28</td>
<td>16</td>
<td>44.6M</td>
<td>0.874</td>
</tr>
</tbody>
</table>

### Language divisions

<table>
<thead>
<tr>
<th>Language divisions</th>
<th>NUHDSS</th>
<th>Kenya</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bantu</td>
<td>78.5</td>
<td>52.0</td>
</tr>
<tr>
<td>Nilotic</td>
<td>13.8</td>
<td>28.6</td>
</tr>
<tr>
<td>Cushites</td>
<td>6.4</td>
<td>8.0</td>
</tr>
<tr>
<td>Other</td>
<td>1.3</td>
<td>11.4</td>
</tr>
</tbody>
</table>

### Ethnic Groups

<table>
<thead>
<tr>
<th>Ethnic Group</th>
<th>% NUHDSS</th>
<th>% Kenya</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kikuyu</td>
<td>29.9</td>
<td>17</td>
</tr>
<tr>
<td>Luhya</td>
<td>15.7</td>
<td>14</td>
</tr>
<tr>
<td>Kamba</td>
<td>24.6</td>
<td>10</td>
</tr>
<tr>
<td>Luo</td>
<td>12.9</td>
<td>10</td>
</tr>
<tr>
<td>Other</td>
<td>16.9</td>
<td>49</td>
</tr>
</tbody>
</table>

Catherine Kyobutungi: Nairobi Health and Demographic Surveillance System
Aim 3: Genetic and environmental contributions to body composition

- Standardised phenotype questionnaire
- Instructions & SOPs
- Equipment purchase
  - Stadiometers
  - Scales
  - Ultrasound machines
- Training

- Field roll out
  - Staggered to ensure QA
  - Years 2, 3 & 4
  - Phenotyping
  - Blood sampling
- Data (demography & phenotype)
  - Collection
  - Data Management
Phenotype & Sample collection

**Funded**
- Demographic information
  - Home language & self-reported ethnicity
  - Medical & health histories
  - Living conditions (SES)
- Body composition
  - Height & Weight
  - Blood pressure
  - Waist & hip circumference
  - Ultrasound subcutaneous & visceral fat

**Not funded**
- Glucose
- Lipids

**Blood samples (fasting):**
- EDTA (DNA)
- Clotted (serum - lipids)
- NaF (plasma - glucose)

**Added sampling:**
- Spot urine collections

H3Africa – Johannesburg - October 2013
## RedCAP for AWI-Gen

<table>
<thead>
<tr>
<th>Demographic Collection Data (89)</th>
<th>Phenotypic Collection Data (231)</th>
<th>Sample Data Collection (25)</th>
<th>Blood Collection Data (21)</th>
<th>Checklist (11)</th>
</tr>
</thead>
<tbody>
<tr>
<td>General information</td>
<td>Marital Status</td>
<td>Anthropometric measurements</td>
<td>Blood collection</td>
<td>Questionnaire</td>
</tr>
<tr>
<td>Demographic information</td>
<td>Education</td>
<td>Blood pressure</td>
<td>HIV</td>
<td>Anthropometric measurements</td>
</tr>
<tr>
<td>• Age</td>
<td>Employment</td>
<td>Pulse</td>
<td>Test results</td>
<td>Blood pressure</td>
</tr>
<tr>
<td>• Country</td>
<td></td>
<td>Ultrasound measurements</td>
<td>Urine collection</td>
<td>Pulse</td>
</tr>
<tr>
<td>• Home language*</td>
<td></td>
<td></td>
<td></td>
<td>Blood Samples</td>
</tr>
<tr>
<td>• Ethnicity*</td>
<td></td>
<td></td>
<td></td>
<td>Ultrasound</td>
</tr>
<tr>
<td>• Family Ethnicity*</td>
<td></td>
<td></td>
<td></td>
<td>cIMT</td>
</tr>
<tr>
<td>Family Composition</td>
<td>Household attributes*</td>
<td></td>
<td></td>
<td>HIV test</td>
</tr>
<tr>
<td></td>
<td>Substance use</td>
<td></td>
<td></td>
<td>Urine sample</td>
</tr>
<tr>
<td></td>
<td>• Tobacco use</td>
<td></td>
<td></td>
<td>Travel reimbursement</td>
</tr>
<tr>
<td></td>
<td>• Alcohol use</td>
<td></td>
<td></td>
<td>Quality Controller ID</td>
</tr>
<tr>
<td></td>
<td>• Drug use</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>General health</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Infection history</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Malaria</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• TB</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Cardiometabolic risk factors</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Diabetes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Stroke</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Thyroid disease</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Kidney disease</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Physical activity</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Sleep</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
In a population sample of 2000 individuals.....

<table>
<thead>
<tr>
<th></th>
<th>Expected number HIV infected individuals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agincourt</td>
<td>462</td>
</tr>
<tr>
<td>Dikgale</td>
<td>274</td>
</tr>
<tr>
<td>Nairobi</td>
<td>248</td>
</tr>
<tr>
<td>Nanoro</td>
<td>22</td>
</tr>
<tr>
<td>Navrongo</td>
<td>30</td>
</tr>
<tr>
<td>Soweto</td>
<td>304</td>
</tr>
</tbody>
</table>

Based on regional averages

Based on country average
Data Management

MODULE 1
Repository
Demographic & Phenotype data

MODULE 2
LIMS
Biorepository & Laboratory

MODULE 3
Genetic & Genomic Data Results Management

MODULE 4
Shared data sets
Repository
collaborators & users

Unique Individual identifier

Centre 1  Centre 2  Centre 3  Centre 4  Centre 5  Centre 6

Repository
Demographic & Phenotype data

LIMS
Biorepository & Laboratory

Shared data sets
Repository
collaborators & users

Genetic & Genomic Data Results Management
AWI-Gen Data Management Workshop
July 2013
Summary of Connectivity and Storage Capacity

<table>
<thead>
<tr>
<th>Site</th>
<th>Speed*</th>
<th>Reliability</th>
<th>Storage capacity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nairobi</td>
<td>8MBps</td>
<td>99%</td>
<td>1.4 TB</td>
</tr>
<tr>
<td>Nanoro</td>
<td>1MBps</td>
<td>Back up system</td>
<td>300 GB +</td>
</tr>
<tr>
<td>Navrongo</td>
<td>1MBps</td>
<td>98%</td>
<td>730 GB</td>
</tr>
<tr>
<td>Agincourt</td>
<td>4MBps</td>
<td>96%</td>
<td>5 TB</td>
</tr>
<tr>
<td>Dikgale</td>
<td>45MBps</td>
<td>95%</td>
<td>30 TB</td>
</tr>
<tr>
<td>Soweto</td>
<td>94MBps</td>
<td>99%</td>
<td>300 GB +</td>
</tr>
</tbody>
</table>
Important outcomes

• Capacity development
  – PhD students, postdocs, scientists
  – Epidemiology, population genetics, genomics, bioinformatics

• Phenotype and blood profiles
  – Means and ranges for African populations

• New knowledge
  – African population diversity
    • African variation enhanced chip (cost effective)
    • African population structure
  – Pilot study
    • Replication data, Longitudinal analysis, Training
  – Main study
### Timeline (Aug 2012 – July 2017)

<table>
<thead>
<tr>
<th>ACTIVITY</th>
<th>YEAR 1</th>
<th>YEAR 2</th>
<th>YEAR 3</th>
<th>YEAR 4</th>
<th>YEAR 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training and capacity development</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>African genome structure</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phenotyping and sampling for Cohorts</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Obesity and body composition pilot study – urban South Africa</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Genome association study – west, east and south Africa</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Acknowledgements

• Wits
  Michele Ramsay
  Himla Soodyall
  Shane Norris
  Stephen Tollman
  Alisha Wade
  Nadia Carstens
  Nigel Crowther
  Zane Lombard
  Kathleen Khan
  Cassandra Soo
  Ananyo Choudhury
  Venesa Pillay

• INDEPTH
  Osman Sankoh
  Kathleen Kahn
  Stephen Tollman
  Abraham Oduro
  Godfred Agongo
  Halidou Tinto
  Hermann Sorgho
  Marianne Alberts
  Catherine Kyobutungi
  Kate Theron