

South Africa

Frampton, Dan, Division of Infection and Immunity, UCL, London
Division of Infection and Immunity, UCL, London UCL, London
McKendry, Rachel, London Center for Nanotechnology, UCL, London
Pillay, Deenan, Africa Health Research Institute

**Visualisation of sequence and demographic data
to assist HIV surveillance in Northern KwaZulu-
Natal: extending the TasP/iSense dashboard to
include markers of HIV drug resistance mutations**

Study Documentation

June 3, 2019

Metadata Production

Metadata Producer(s)	Africa Health Research Institute (AHRI)
Identification	DDI.AHRI.TasP.DataEveryWhere.Dashboard.2016.v1

Table of Contents

Overview.....	4
Scope & Coverage.....	4
Producers & Sponsors.....	5
Sampling.....	5
Data Collection.....	5
Data Processing & Appraisal.....	5
Accessibility.....	5
Files Description.....	7
AHRI.TasP.DataEveryWhere.Hexagons.2016.v1.....	7
AHRI.TasP.DataEveryWhere.SequnceMappedToHexagons.2016.v1.....	7
Variables List.....	8
AHRI.TasP.DataEveryWhere.Hexagons.2016.v1.....	8
AHRI.TasP.DataEveryWhere.SequnceMappedToHexagons.2016.v1.....	8
Variables Description.....	9
AHRI.TasP.DataEveryWhere.Hexagons.2016.v1.....	10
AHRI.TasP.DataEveryWhere.SequnceMappedToHexagons.2016.v1.....	11

Visualisation of sequence and demographic data to assist HIV surveillance in Northern KwaZulu-Natal: extending the TasP/iSense dashboard to include markers of HIV drug resistance mutations

Overview	
Identification	AHRI.TasP.DataEveryWhere.Dashboard.2016.v1
Version	V1.0.0
<p>Abstract</p> <p>This proposal aims to extend an existing collaboration between AHRI and UCL. As part of the iSense project, teams from AHRI and UCL have successfully developed a dashboard that integrates information from mobile computers used for TasP field visits with data from the clinics to display spatial coverage of homestead visits, highlighting those that require follow-up visits to ensure linkage to care. The dashboard provides a broad snapshot of the state of the study, spatially aggregating geographical zones in order to preserve the privacy of trial participants.</p> <p>The aim of this proposal is to extend this framework to visualise presence and prevalence of drug resistance mutations (DRMs) within the study area. A higher prevalence of DRMs than expected may be linked to several factors, e.g. poor drug adherence, and thus of value to clinicians and healthcare workers in terms of focusing efforts and resource allocation.</p>	
Kind of Data	The existing dashboard is built on pseudonymised demographic data grouped into geographic hexagons. Within the PANGAEA project, we have independently assembled HIV genomes from patient clinical samples and identified DRMs for each sample. To map the DRMs to each geographical hexagon, we require a lookup table to link the TasP/PANGAEA IDs to hexagons.
Unit of Analysis	Each sample ID should be linked to a single hexagon where possible.

Scope & Coverage	
Keywords	Mapping, visualisation, data linkage, drug resistance, DRM, HIV-1
Topics	HIV-1; Incidence; Phylogeny; Epidemics; Population Surveillance; Rural Population; HIV Infections; Africa
Time Period(s)	2012-2016
Countries	South Africa
<p>Geographic Coverage</p> <p>South Africa</p>	
<p>Universe</p> <p>Clusters: the trial area consists of 150 local areas (neighbourhoods). These were aggregated into 34 clusters of between one and six contiguous neighbourhoods, each cluster comprising an average of 1 000 individuals >15 years of age. Clusters were designed to encompass social networks based on earlier studies. TasP phase 1 has been implemented in 10 geographic clusters (5 control and 5 intervention). Twelve additional clusters have been opened in June 2014.</p> <p>Locations: corresponds to physical locations. There are two types of locations: homesteads and TasP clinics.</p> <p>Homesteads: the population lives in scattered homesteads that are not concentrated into villages or compounds. All usable and occupied homesteads were eligible for trial participation.</p> <p>Households: each homestead could be composed of one or several households. An household remains always attached to the same homestead.</p> <p>Eligible individuals : all 16 years or older and resident household members. Each individuals is attached to an household. In case of internal migration, an individual could move to another household.</p>	

TasP clinics: dedicated trial clinics implemented in each survey cluster.

Producers & Sponsors

Primary Investigator(s)	Frampton, Dan, Division of Infection and Immunity, UCL, London Division of Infection and Immunity, UCL, London McKendry, Rachel, London Center for Nanotechnology, UCL, London Pillay, Deenan, Africa Health Research Institute
Other Producer(s)	Africa Health Research Institute (AHRI)
Funding Agency/ies	South African Medical Research Council (SAMRC) , Genotyping funding source Engineering and Physical Sciences Research Council (EPSRC) , Design and implementation of dashboard within the iSense project
Other Acknowledgment(s)	Jaco Dreyer , Linking sample and demographic data design , Africa Health Research Institute Ed Manley , Implementation of iSense dashboard design , UCL Centre for Advanced Spatial Analysis (CASA), London Dave Concannon , Implementation of iSense dashboard , UCL Centre for Advanced Spatial Analysis (CASA), London

Sampling

Sampling Procedure

HIV positive individuals within the TasP surveillance area of the Africa Health Research Institute from 2012 to 2016.

Data Collection

Data Collection Dates	start 2012-01-01 end 2016-12-04
------------------------------	------------------------------------

Data Processing & Appraisal

Data Editing

Samples were sequenced at the Durban based laboratory of AHRI; genome assembly and downstream sequence analysis was performed at UCL.

Accessibility

Access Conditions

The representative of the Receiving Organization agrees to comply with the following conditions:

1. Access to the restricted data will be limited to the Lead Researcher and other members of the research team listed in this request.
2. Copies of the restricted data or any data created on the basis of the original data will not be copied or made available to anyone other than those mentioned in this Data Access Agreement, unless formally authorized by the Data Archive.
3. The data will only be processed for the stated statistical and research purpose. They will be used for solely for reporting of aggregated information, and not for investigation of specific individuals or organizations. Data will not in any way be used for any administrative, proprietary or law enforcement purposes.
4. The Lead Researcher must state if it is their intention to match the restricted microdata with any other micro-dataset. If any matching is to take place, details must be provided of the datasets to be matched and of the reasons for the matching. Any datasets created as a result of matching will be considered to be restricted and must comply with the terms of this Data Access Agreement.

5. The Lead Researcher undertakes that no attempt will be made to identify any individual person, family, business, enterprise or organization. If such a unique disclosure is made inadvertently, no use will be made of the identity of any person or establishment discovered and full details will be reported to the Data Archive. The identification will not be revealed to any other person not included in the Data Access Agreement.
6. The Lead Researcher will implement security measures to prevent unauthorized access to licensed microdata acquired from the Data Archive. The microdata must be destroyed upon the completion of this research, unless the Data Archive obtains satisfactory guarantee that the data can be secured and provides written authorization to the Receiving Organization to retain them. Destruction of the microdata will be confirmed in writing by the Lead Researcher to the Data Archive.
7. Any books, articles, conference papers, theses, dissertations, reports, or other publications that employ data obtained from the Data Archive will cite the source of data in accordance with the citation requirement provided with the dataset.
8. An electronic copy of all reports and publications based on the requested data will be sent to the Data Archive.
9. The original collector of the data, the Data Archive, and the relevant funding agencies bear no responsibility for use of the data or for interpretations or inferences based upon such uses.
10. This agreement will come into force on the date that approval is given for access to the restricted dataset and remain in force until the completion date of the project or an earlier date if the project is completed ahead of time.
11. If there are any changes to the project specification, security arrangements, personnel or organization detailed in this application form, it is the responsibility of the Lead Researcher to seek the agreement of the Data Archive to these changes. Where there is a change to the employer organization of the Lead Researcher this will involve a new application being made and termination of the original project.
12. Breaches of the agreement will be taken seriously and the Data Archive will take action against those responsible for the lapse if willful or accidental. Failure to comply with the directions of the Data Archive will be deemed to be a major breach of the agreement and may involve recourse to legal proceedings. The Data Archive will maintain and share with partner data archives a register of those individuals and organizations which are responsible for breaching the terms of the Data Access Agreement and will impose sanctions on release of future data to these parties.

Citation Requirements

<https://doi.org/10.23664/AHRI.TasP.DataEveryWhere.Dashboard.2016.v1>

Files Description

Dataset contains 2 file(s)

AHRI.TasP.DataEveryWhere.Hexagons.2016.v1	
# Cases	1021
# Variable(s)	5

AHRI.TasP.DataEveryWhere.SequnceMappedToHexagons.2016.v1	
# Cases	1750
# Variable(s)	6

Variables List

Dataset contains 11 variable(s)

File AHRI.TasP.DataEveryWhere.Hexagons.2016.v1							
#	Name	Label	Type	Format	Valid	Invalid	Question
1	Id	Id	continuous	numeric-12.0	1021	0	-
2	Uid	Uid	discrete	character-36	1021	0	-
3	Perimeter	Perimeter	discrete	character-244	1021	-	-
4	Centroid	Centroid	discrete	character-244	1021	-	-
5	t	t	discrete	character-244	1021	-	-

File AHRI.TasP.DataEveryWhere.SequnceMappedToHexagons.2016.v1							
#	Name	Label	Type	Format	Valid	Invalid	Question
1	PangeaId	PANGEA Unique Sample Identifier	discrete	character-244	1750	-	-
2	Cohort	Cohort	discrete	character-244	1750	-	-
3	SampleId	Unique Sample Identifier	discrete	character-244	1750	-	-
4	Sequence ..	Unique Sequence Identifier in the Lab	discrete	character-244	1750	-	-
5	Individu ..	Unique Individual Identifier	continuous	numeric-12.0	1750	0	-
6	HexagonId	Unique Hexagon Identifier	continuous	numeric-12.0	1749	1	-

Variables Description

Dataset contains 11 variable(s)

File : AHRI.TasP.DataEveryWhere.Hexagons.2016.v1

Id: Id

Information	[Type= continuous] [Format=numeric] [Range= 1-1021] [Missing=*]
Statistics [NW/ W]	[Valid=1021 /-] [Invalid=0 /-] [Mean=511 /-] [StdDev=294.882 /-]

Uid: Uid

Information	[Type= discrete] [Format=character] [Missing=*]
Statistics [NW/ W]	[Valid=1021 /-] [Invalid=0 /-]

Perimeter: Perimeter

Information	[Type= discrete] [Format=character] [Missing=*]
Statistics [NW/ W]	[Valid=1021 /-]

Centroid: Centroid

Information	[Type= discrete] [Format=character] [Missing=*]
Statistics [NW/ W]	[Valid=1021 /-]

t: t

Information	[Type= discrete] [Format=character] [Missing=*]
Statistics [NW/ W]	[Valid=1021 /-]

File : AHRI.TasP.DataEveryWhere.SequenceMappedToHexagons.2016.v1

PangeaId: PANGEA Unique Sample Identifier

Information

[Type= discrete] [Format=character] [Missing=*]

Statistics [NW/ W]

[Valid=1750 /-]

Cohort: Cohort

Information

[Type= discrete] [Format=character] [Missing=*]

Statistics [NW/ W]

[Valid=1750 /-]

Value	Label	Cases	Percentage
TASP		935	<div></div> 53.4%
TasP		815	<div></div> 46.6%

Warning: these figures indicate the number of cases found in the data file. They cannot be interpreted as summary statistics of the population of interest.

SampleId: Unique Sample Identifier

Information

[Type= discrete] [Format=character] [Missing=*]

Statistics [NW/ W]

[Valid=1750 /-]

SequenceLabId: Unique Sequence Identifier in the Lab

Information

[Type= discrete] [Format=character] [Missing=*]

Statistics [NW/ W]

[Valid=1750 /-]

IndividualId: Unique Individual Identifier

Information

[Type= continuous] [Format=numeric] [Range= 9-100102] [Missing=*]

Statistics [NW/ W]

[Valid=1750 /-] [Invalid=0 /-] [Mean=12724.27 /-] [StdDev=8856.461 /-]

HexagonId: Unique Hexagon Identifier

Information

[Type= continuous] [Format=numeric] [Range= 412-1013] [Missing=*]

Statistics [NW/ W]

[Valid=1749 /-] [Invalid=1 /-] [Mean=850.456 /-] [StdDev=104.355 /-]