

**South Africa**

**Dr Maryam, Shahmanesh, Africa Health Research Institute**

**Genomics data collected by DREAMS  
Nested- and Multilevel HIV Prevention**

**Study Documentation**

October 20, 2020

# Metadata Production

Metadata Producer(s)	Africa Health Research Institute (AHRI)
Identification	DDI.AHRI.Fasta.Sequence.Data.For.DREAMS

# Table of Contents

<a href="#">Overview.....</a>	<a href="#">4</a>
<a href="#">Scope &amp; Coverage.....</a>	<a href="#">4</a>
<a href="#">Producers &amp; Sponsors.....</a>	<a href="#">4</a>
<a href="#">Sampling.....</a>	<a href="#">4</a>
<a href="#">Data Collection.....</a>	<a href="#">5</a>
<a href="#">Data Processing &amp; Appraisal.....</a>	<a href="#">5</a>
<a href="#">Accessibility.....</a>	<a href="#">5</a>
<a href="#">Files Description.....</a>	<a href="#">7</a>
<a href="#">AHRI.DREAMS.Fasta.Files_v1.....</a>	<a href="#">7</a>
<a href="#">Variables List.....</a>	<a href="#">8</a>
<a href="#">AHRI.DREAMS.Fasta.Files_v1.....</a>	<a href="#">8</a>
<a href="#">Variables Description.....</a>	<a href="#">9</a>
<a href="#">AHRI.DREAMS.Fasta.Files_v1.....</a>	<a href="#">10</a>

## Genomics data collected by DREAMS Nested- and Multilevel HIV Prevention

Overview	
<b>Identification</b>	AHRI.Fasta.Sequence.Data.For.DREAMS
<b>Version</b>	V1.0.0
<b>Abstract</b> The Genomics data was analysed from HIV Positive Dried Blood Spots (DBS) specimens collected during the data collection period for the DREAMS Nested- and Multilevel HIV Prevention projects.	
<b>Kind of Data</b>	Genomics Data
<b>Unit of Analysis</b>	Specimen

Scope & Coverage	
<b>Keywords</b>	DREAMS nested cohort; Genomics data; Multilevel HIV prevention
<b>Topics</b>	Genomics data
<b>Time Period(s)</b>	2017-2020
<b>Countries</b>	South Africa
<b>Geographic Coverage</b> Demographic surveillance area of the Africa Health Research Institute; KwaZulu-Natal, uMkhanyakude district.	
<b>Universe</b> DREAMS Nested cohort: A nested cohort of 800 AGYW will be followed prospectively at three time points over the two-year study period - baseline, 12 months and 24 months - at points closest aligned with periods before, during and after DREAMS implementation. In ACDIS cohorts of 400 girls aged 13-17 years and 400 young women aged 14-24 years will undergo informed consent, recruited undergo a baseline questionnaire and provide dried blood spot (DBS) sample for HSV2 and for the HIV testing, and then reviewed annually for the next two years.  Multilevel HIV Prevention: Participants were randomly selected from a southern part of AHRI surveillance area, stratified by age, sex and week-block. A target sample size was 2500 males aged 13-35 and 500 females aged 24-29 years	

Producers & Sponsors	
<b>Primary Investigator(s)</b>	Dr Maryam, Shahmanesh, Africa Health Research Institute
<b>Other Producer(s)</b>	Africa Health Research Institute (AHRI)
<b>Funding Agency/ies</b>	Bill & Melinda Gates Foundation (BMGF) , Impact Evaluation of PEPFAR funded DREAMS interventions South African Medical Research Council (SAMRC) , Genotyping funding source
<b>Other Acknowledgment(s)</b>	Jaco Dreyer , Data management , Africa Health Research Institute Nondumiso Mthiyane , Cleaning and analysis , Africa Health Research Institute

Sampling
<b>Sampling Procedure</b>

The type of sample and sample design used to select the survey respondents to represent the population. May include reference to the target sample size and the sampling fraction

DREAMS: Adolescent girls and young women aged 13-24 years who were residents in the demographic surveillance area of the Africa Health Research Institute. A total of 3013 participants were randomly selected to obtain a target sample size of 800 after 2 years of follow-up, allowing for 40% non-contact/loss-to-follow-up. Sampling was stratified by age group and area (week-blocks).

Multilevel: Participants were randomly selected from a southern part of AHRI surveillance area, stratified by age, sex and week-block. A target sample size was 2500 males aged 13-35 and 500 females aged 24-29 years.

## Data Collection

<b>Data Collection Dates</b>	start 2017-05-03 end 2020-01-29
------------------------------	------------------------------------

## Data Processing & Appraisal

### Data Editing

None

## 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**

Shahmanesh, M. (2020) “Genomics data collected by DREAMS Nested- and Multilevel HIV Prevention.” Africa Health Research Institute (AHRI). doi: 10.23664/AHRI.FASTA.SEQUENCE.DATA.FOR.DREAMS.

# Files Description

Dataset contains 1 file(s)

AHRI.DREAMS.Fasta.Files_v1	
# Cases	116
# Variable(s)	3

# Variables List

Dataset contains 3 variable(s)

File AHRI.DREAMS.Fasta.Files_v1							
#	Name	Label	Type	Format	Valid	Invalid	Question
1	<a href="#">FastaId</a>	Unique Fasta Identifier	discrete	character-14	116	0	-
2	<a href="#">SpecimenId</a>	Unique Specimen Identifier	discrete	character-13	116	0	-
3	<a href="#">Sequence</a>	Sequence	discrete	character-255	116	-	-



# Variables Description

Dataset contains 3 variable(s)

# File : AHRI.DREAMS.Fasta.Files\_v1

## # FastaId: Unique Fasta Identifier

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

**Statistics [NW/ W]** [Valid=116 /-] [Invalid=0 /-]

Value	Label	Cases	Percentage	
2017_279_00192		1		0.9%
2017_279_00196		1		0.9%
2017_279_00203		1		0.9%
2017_279_00217		1		0.9%
2017_279_00227		1		0.9%
2017_279_00312		1		0.9%
2017_279_00457		1		0.9%
2017_279_00486		1		0.9%
2017_279_00702		1		0.9%
2017_279_00715		1		0.9%
2017_279_00746		1		0.9%
2017_279_00757		1		0.9%
2017_279_00783		1		0.9%
2017_279_00855		1		0.9%
2017_279_00864		1		0.9%
2017_279_00959		1		0.9%
2017_279_01137		1		0.9%
2017_279_01260		1		0.9%
2017_279_01350		1		0.9%
2018_279_00170		1		0.9%
2018_279_00187		1		0.9%
2018_279_00364		1		0.9%
2018_279_00403		1		0.9%
2018_279_00450		1		0.9%
2018_279_00484		1		0.9%
2018_279_00629		1		0.9%
2018_279_00630		1		0.9%
2018_279_00759		1		0.9%
2018_279_00828		1		0.9%
2018_279_00960		1		0.9%
2018_279_01387		1		0.9%
2018_279_01418		1		0.9%
2018_279_01768		1		0.9%
2018_279_01829		1		0.9%
2018_279_02303		1		0.9%
2018_279_02651		1		0.9%
2018_279_02704		1		0.9%
2018_279_02733		1		0.9%
2018_279_03050		1		0.9%
2018_279_03052		1		0.9%
2018_279_03140		1		0.9%

# File : AHRI.DREAMS.Fasta.Files\_v1

# FastaId: Unique Fasta Identifier

Value	Label	Cases	Percentage
2018_279_03180		1	0.9%
2018_279_03370		1	0.9%
2018_279_03423		1	0.9%
2018_279_03520		1	0.9%
2018_279_03547		1	0.9%
2018_279_03595		1	0.9%
2018_279_03612		1	0.9%
2018_279_03664		1	0.9%
2018_279_03718		1	0.9%
2019_279_00058		1	0.9%
2019_279_00060		1	0.9%
2019_279_00078		1	0.9%
2019_279_00146		1	0.9%
2019_279_00183		1	0.9%
2019_279_00221		1	0.9%
2019_279_00230		1	0.9%
2019_279_00272		1	0.9%
2019_279_00290		1	0.9%
2019_279_00346		1	0.9%
2019_279_00361		1	0.9%
2019_279_00414		1	0.9%
2019_279_00418		1	0.9%
2019_279_00593		1	0.9%
2019_279_00681		1	0.9%
2019_279_00794		1	0.9%
2019_279_00801		1	0.9%
2019_279_00816		1	0.9%
2019_279_00838		1	0.9%
2019_279_01074		1	0.9%
2019_279_01112		1	0.9%
2019_279_01167		1	0.9%
2019_279_01229		1	0.9%
2019_279_01264		1	0.9%
2019_279_01316		1	0.9%
2019_279_01322		1	0.9%
2019_279_01326		1	0.9%
2019_279_01366		1	0.9%
2019_279_01400		1	0.9%
2019_279_01421		1	0.9%
2019_279_01545		1	0.9%
2019_279_01548		1	0.9%
2019_279_01560		1	0.9%
2019_279_01672		1	0.9%

# File : AHRI.DREAMS.Fasta.Files\_v1

## # FastaId: Unique Fasta Identifier

Value	Label	Cases	Percentage
2019_279_01837		1	0.9%
2019_279_01923		1	0.9%
2019_279_01931		1	0.9%
2019_279_02039		1	0.9%
2019_279_02077		1	0.9%
2019_279_02107		1	0.9%
2019_279_02115		1	0.9%
2019_279_02163		1	0.9%
2019_279_02217		1	0.9%
2019_279_02231		1	0.9%
2019_279_02289		1	0.9%
2019_279_02296		1	0.9%
2019_279_02297		1	0.9%
2019_279_02302		1	0.9%
2019_279_02439		1	0.9%
2019_279_02598		1	0.9%
2019_279_02863		1	0.9%
2019_279_03016		1	0.9%
2019_279_03161		1	0.9%
2019_279_03219		1	0.9%
2019_279_03234		1	0.9%
2019_279_03465		1	0.9%
2019_279_03623		1	0.9%
2019_279_03939		1	0.9%
2019_279_03958		1	0.9%
2019_279_04037		1	0.9%
2019_279_04108		1	0.9%
2019_279_04142		1	0.9%
2019_279_04386		1	0.9%
2019_279_04398		1	0.9%
2019_279_04542		1	0.9%
FastaId		1	0.9%

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.

## # SpecimenId: Unique Specimen Identifier

Information	[Type= discrete] [Format=character] [Missing=*]		
Statistics [NW/ W]	[Valid=116 /-] [Invalid=0 /-]		
Value	Label	Cases	Percentage
DBS1-27900193		1	0.9%
DBS1-27900199		1	0.9%
DBS1-27900212		1	0.9%
DBS1-27900249		1	0.9%
DBS1-27900259		1	0.9%
DBS1-27900353		1	0.9%

# File : AHRI.DREAMS.Fasta.Files\_v1

# SpecimenId: Unique Specimen Identifier

Value	Label	Cases	Percentage
DBS1-27900379		1	0.9%
DBS1-27900493		1	0.9%
DBS1-27900580		1	0.9%
DBS1-27900617		1	0.9%
DBS1-27900649		1	0.9%
DBS1-27900768		1	0.9%
DBS1-27900776		1	0.9%
DBS1-27900781		1	0.9%
DBS1-27900844		1	0.9%
DBS1-27901027		1	0.9%
DBS1-27901042		1	0.9%
DBS1-27901276		1	0.9%
DBS1-27901351		1	0.9%
DBS1-27901727		1	0.9%
DBS1-27901731		1	0.9%
DBS1-27901799		1	0.9%
DBS1-27901911		1	0.9%
DBS1-27901922		1	0.9%
DBS1-27902107		1	0.9%
DBS1-27902264		1	0.9%
DBS1-27902305		1	0.9%
DBS2-27902527		1	0.9%
DBS2-27902528		1	0.9%
DBS2-27902607		1	0.9%
DBS2-27903291		1	0.9%
DBS2-27903363		1	0.9%
DBS2-27903715		1	0.9%
DBS2-27903857		1	0.9%
DBS2-27904087		1	0.9%
DBS3-27904502		1	0.9%
DBS3-27904708		1	0.9%
DBS3-27904772		1	0.9%
DBS3-27904825		1	0.9%
DBS3-27904838		1	0.9%
DBS3-27904853		1	0.9%
DBS3-27904985		1	0.9%
DBS3-27905174		1	0.9%
DBS3-27905250		1	0.9%
DBS3-27905303		1	0.9%
DBS3-27905406		1	0.9%
DBS3-27905418		1	0.9%
DBS3-27905641		1	0.9%
DBS3-27905665		1	0.9%

# File : AHRI.DREAMS.Fasta.Files\_v1

# SpecimenId: Unique Specimen Identifier

Value	Label	Cases	Percentage
DBS3-27905800		1	0.9%
DBS3-27905905		1	0.9%
DBS3-27905937		1	0.9%
DBS3-27905963		1	0.9%
DBS3-27906045		1	0.9%
DBS3-27906051		1	0.9%
DBS3-27906058		1	0.9%
DBS3-27906096		1	0.9%
DBS3-27906120		1	0.9%
DBS3-27906189		1	0.9%
DBS3-27906238		1	0.9%
DBS3-27906287		1	0.9%
DBS3-27906298		1	0.9%
DBS3-27906529		1	0.9%
DBS3-27906677		1	0.9%
DBS3-27906775		1	0.9%
DBS3-27906795		1	0.9%
DBS3-27906840		1	0.9%
DBS3-27906884		1	0.9%
DBS3-27906925		1	0.9%
DBS3-27907009		1	0.9%
DBS3-27907094		1	0.9%
DBS3-27907222		1	0.9%
DBS4-27907339		1	0.9%
DBS4-27907379		1	0.9%
DBS4-27907382		1	0.9%
DBS4-27907399		1	0.9%
DBS4-27907407		1	0.9%
DBS4-27907439		1	0.9%
DBS4-27907456		1	0.9%
DBS4-27907555		1	0.9%
DBS4-27907626		1	0.9%
DBS4-27907748		1	0.9%
DBS4-27907779		1	0.9%
DBS4-27907781		1	0.9%
DBS4-27907838		1	0.9%
DBS4-27907869		1	0.9%
DBS4-27907885		1	0.9%
DBS4-27907990		1	0.9%
DBS4-27908158		1	0.9%
DBS4-27908186		1	0.9%
DBS4-27908187		1	0.9%
DBS4-27908285		1	0.9%

# File : AHRI.DREAMS.Fasta.Files\_v1

## # SpecimenId: Unique Specimen Identifier

Value	Label	Cases	Percentage
DBS4-27908339		1	0.9%
DBS4-27908391		1	0.9%
DBS4-27908415		1	0.9%
DBS4-27908437		1	0.9%
DBS4-27908468		1	0.9%
DBS4-27908488		1	0.9%
DBS4-27908534		1	0.9%
DBS4-27908703		1	0.9%
DBS5-27909611		1	0.9%
DBS5-27909687		1	0.9%
DBS5-27909738		1	0.9%
DBS5-27909800		1	0.9%
DBS5-27909824		1	0.9%
DBS5-27910223		1	0.9%
DBS5-27910418		1	0.9%
DBS5-27910500		1	0.9%
DBS5-27910692		1	0.9%
DBS5-27910727		1	0.9%
DBS5-27910809		1	0.9%
DBS5-27910926		1	0.9%
DBS5-27911152		1	0.9%
DBS5-27911213		1	0.9%
DBS5-27911289		1	0.9%
SpecimenID		1	0.9%

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.

## # Sequence: Sequence

Information	[Type= discrete] [Format=character] [Missing=*]
Statistics [NW/ W]	[Valid=116 /-]

Value	Label	Cases	Percentage
ACTCTTTGGCA		1	0.9%
ACTCTTTGGCA		1	0.9%
ACTCTTTGGCA		1	0.9%
ACTCTTTGGCA		1	0.9%
ATCACTCTTTG		1	0.9%
ATCACTCTTTG		1	0.9%
ATCACTCTTTG		1	0.9%
ATCACTCTTTG		1	0.9%
ATCACTCTTTG		1	0.9%
ATCACTCTTTG		1	0.9%
ATCACTCTTTG		1	0.9%
ATCACTCTTTG		1	0.9%
ATCACTCTTTG		1	0.9%
ATCACTCTTTG		1	0.9%
ATCACTCTTTG		1	0.9%
ATCACTCTTTG		1	0.9%

## File : AHRI.DREAMS.Fasta.Files\_v1

# Sequence: Sequence

[illegible]



## File : AHRI.DREAMS.Fasta.Files\_v1

# Sequence: Sequence

[illegible]

## File : AHRI.DREAMS.Fasta.Files\_v1

### # Sequence: Sequence

Value	Label	Cases	Percentage	
CCTCAGATCAC		1	0.9%	CAGTATTAGAA
CCTCAGATCAC		1	0.9%	CAGTATTAGAA
CCTCAGATCAC		1	0.9%	CAGTATTAGAA
CCTCAGATCAC		1	0.9%	CAGTATTAGAA
CCTCAGATCAC		1	0.9%	CAGTGTTAGAA
CGACCCCTTGT		1	0.9%	TACCAGGAAA
CTCAAATCACT		1	0.9%	AGTATTAGAA
CTTTGGCAGCC		1	0.9%	AAATAAATTT
CTTTGGCAGCC		1	0.9%	ACATAAATTT
GAGGGCCAAA'		1	0.9%	ATGATAGGAG
GGGGGCCAAA'		1	0.9%	ATGATAGGGG
Sequence		1	0.9%	
TGGCAGCGACG		1	0.9%	ATAAGTTTGCC
TGGCAGCGACG		1	0.9%	TAGATTTGCC
TGGCAGCGACG		1	0.9%	TAGATTTACC
TTCCCTCAGAT		1	0.9%	TACAGTACTA

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.