NATIONAL HUMAN GENOME RESEARCH INSTITUTE Division of Intramural Research

U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES | NATIONAL INSTITUTES OF HEALTH | genome.gov/DIR

Center for Research on Genomics and Global Health

NATIONAL HUMAN GENOME RESEARCH INSTITUTE Division of Intramural Research

Wrap Up

Adebowale Adeyemo, MD
Deputy Director, Center for Research on Genomics & Global Health
NHGRI/NIH

WT Genome Epidemiology Course, Durban, RSA – June 2015

Genome epidemiology

- Tools and Technology: high throughput genotyping and sequencing platforms, high performance computers, clusters and storage…
- Data and Databases: reference databases, genome browsers, central repositories of study data,…
- Analytic and Visualization Paradigms

In this course:

- Genetic association studies
- Genome wide association studies (GWAS)
- Population genetics
- Meta-analysis of GWAS
- Sequencing
- Public resources

- Practicals: R, RStudio, plink, Linux

- Emphasis on process

Published Genome-Wide Associations through 12/2012
Published GWA at $p \leq 5 \times 10^{-8}$ for 17 trait categories

NHGRI GWA Catalog

www.genome.gov/GWAStudies

www.ebi.ac.uk/fgpt/gwas/

National Human Genome Research Institute

NATIONAL HUMAN GENOME RE

Divisio

EMBL-EBI

GWAS in Africa*: July 2013

First Author (Date)	**Disease/Trait**	**Initial sample size**	**Replication sample size**	**Platform**	**Location(s)**	**Comment**
Jallow et al (2009)	Severe malaria	2500	3400	Affymetrix 500K	Gambia	First GWAS in Africa
Timman et al (2012)	Severe malaria	2153	3542	Affymetrix 6.0	Ghana	
Band et al (2013)	Severe malaria	~12000	NA	Illumina (various)	Malawi; Kenya; Gambia	*meta-analysis
Petrovski et al (2010)	HIV-1	1532	NA	Illumina 1M/1M-Duo	Malawi	
Lingappa et al (2011)	HIV-1	798	NA	Illumina 1M-Duo	Multiple sites in & Southern Africa	
Luo et al (2012)	HIV resistance	108	NA	Affymetrix 50k Xba240	Kenya	*Low resolution ?GWAS
Thye et al (2010)	Tuberculosis	3117	4384	Affymetrix 500K/ Affymetrix 6.0	Ghana; Gambia	
Thye et al (2012)	Tuberculosis	3176	~13000	Affymetrix 6.0	Ghana; Gambia-Indonesia-Russia	
Kang et al (2010)	Anthropometric traits	1188 Nigerians (+ 743 African-Americans)	2728	Affymetrix 6.0	Nigeria (USA)	
Ayele et al (2012)	Podoconiosis	397	606	Illumina 610	Ethiopia	First GWAS in Africa of an NCD and of an NTD

*Does not include papers about selection, e.g. Alkorta-Aranburu et al PLoS Genetics 2012

More recent GWAS from Africa

- Mtatiro et al: **Genome wide association study of fetal hemoglobin in sickle cell anemia in Tanzania**. PLoS One. 2014 Nov 5;9(11):e111464. PubMed PMID: 25372704; PubMed Central PMCID: PMC4221031.

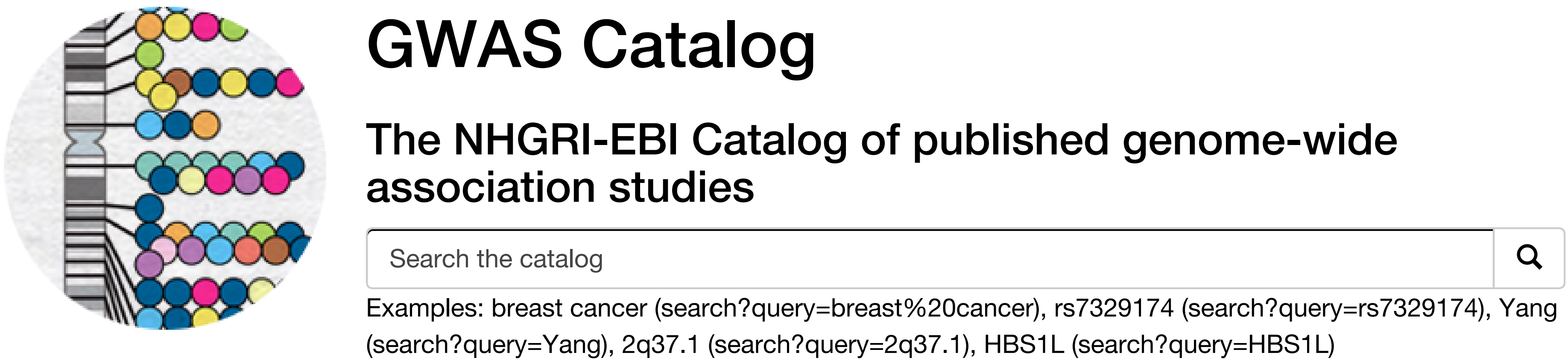
- Cook et al: **A genome-wide association study of prostate cancer in West African men**. Hum Genet. 2014 May;133(5):509-21. PubMed PMID: 24185611; PubMed Central PMCID: PMC3988225.

- Chimusa et al: **Genome-wide association study of ancestry-specific TB risk in the South African Coloured population**. Hum Mol Genet. 2014 Feb 1;23(3):796-809. doi: 10.1093/hmg/ddt462. PubMed PMID: 24057671; PubMed Central PMCID: PMC3888262.

GWAS Catalog

NHGRI GWAS Catalog is now the NHGRI-EBI GWAS of published genome wide association studies

Now at http://www.ebi.ac.uk/gwas/



GWAS Catalog

The NHGRI-EBI Catalog of published genome-wide association studies

Examples: breast cancer (search?query=breast%20cancer), rs7329174 (search?query=rs7329174), Yang (search?query=Yang), 2q37.1 (search?query=2q37.1), HBS1L (search?query=HBS1L)

H3Africa

National Institutes of Health
Wellcome Trust H3Africa Research Network

● NIH primary award institution
● WT primary award institution
● Collaborating institution

Other collaborating institutions are in Belgium, Canada, France, the United Kingdom and the United States of America

Diseases/disorders from RHD to stroke

ELSI

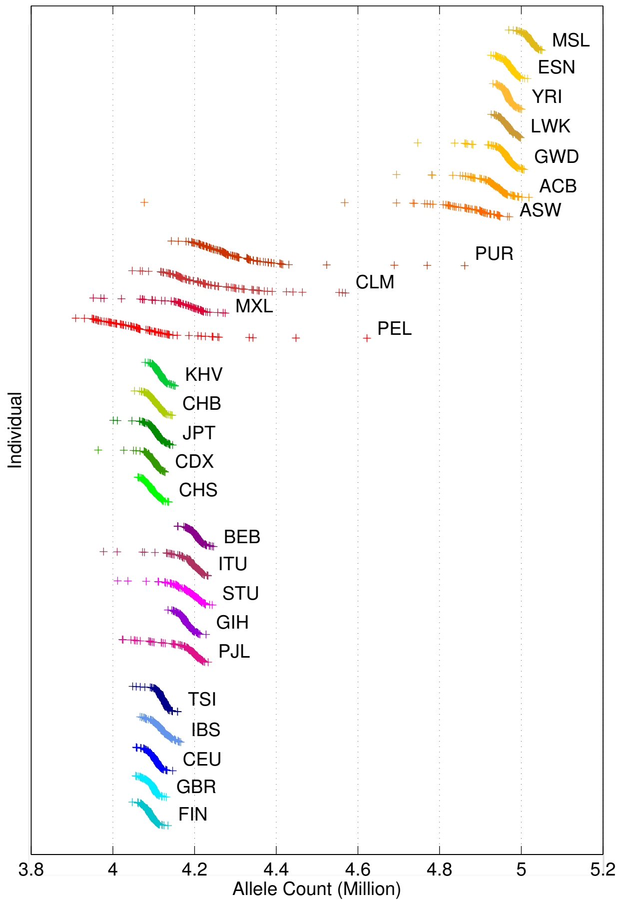
Biorepositories

Bioinformatics network

Why an African GWAS array?

African-derived populations have greater genetic diversity and lower levels of LD, requiring a greater density of SNPs to provide genome-wide coverage of common variation

Arrays optimized for African populations derive content from HapMap/1000 Genomes African populations and African-Americans/ Afro-Caribbeans/other African diaspora populations



Variants per genome (1KG phase 3)

Why an African GWAS array?

H3Africa projects cover many African ethnic groups, few of which have dense genotype/sequence data and for whom coverage of existing products is often suboptimal

National Institutes of Health
Wellcome Trust H3Africa Research Network

● NIH primary award institution
● WT primary award institution
● Collaborating institution

Other collaborating institutions are in Belgium, Canada, France, the United Kingdom and the United States of America

Draft parameters for African custom array

- A GWAS chip that adequately captures common variation in African populations, especially those in H3A projects

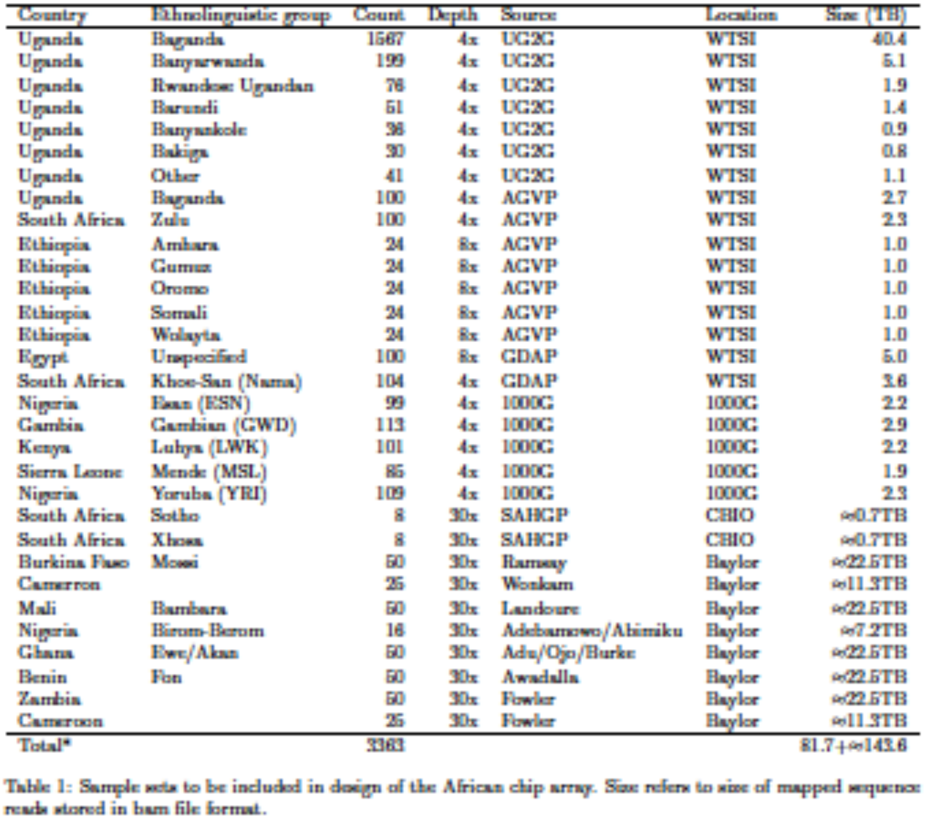
- Desired features:
  - “good” coverage of common variation in H3A populations;
  - good scaffold for imputation;
  - useful specific content - e.g. reported GWAS hits, PGx variants, HLA, fingerprint SNPs,…
  - ? option to add custom content

- Timeline: ready before H3A projects samples ready for genotyping; possibility of more than one version

Data to be used

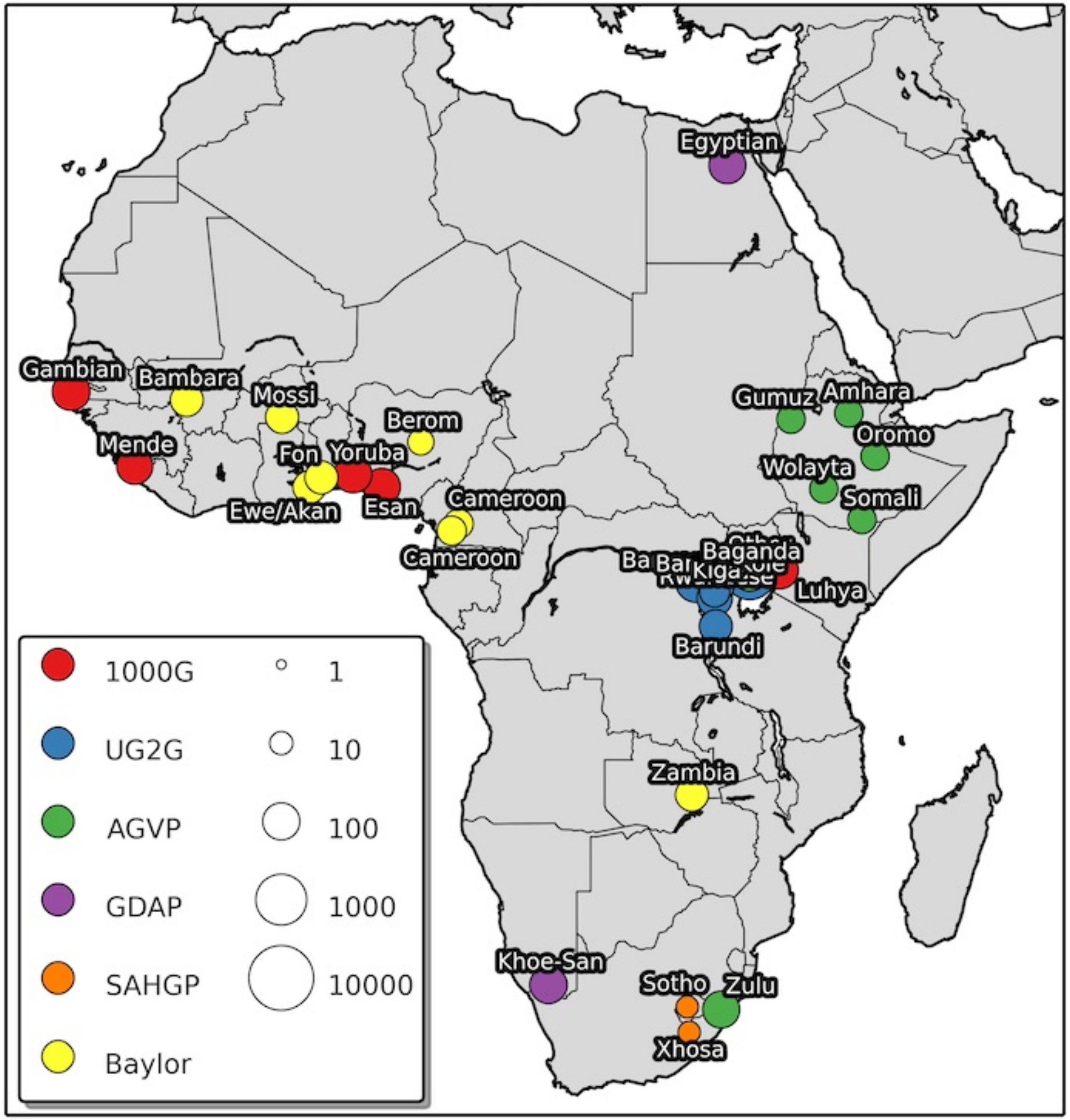
WGS from African populations

- Previous studies
- Recent/ongoing studies
- NIH/NHGRI supplement for WGS



Country	Ethnolinguistic group	Count	Depth	Source	Location	Size (TB)
Uganda	Baganda	1567	4x	UG2G	WTSI	40.4
Uganda	Banyarwanda	199	4x	UG2G	WTSI	5.1
Uganda	Rwandese Ugandan	76	4x	UG2G	WTSI	1.9
Uganda	Barundi	51	4x	UG2G	WTSI	1.4
Uganda	Banyankole	38	4x	UG2G	WTSI	0.9
Uganda	Bakiga	30	4x	UG2G	WTSI	0.8
Uganda	Other	41	4x	UG2G	WTSI	1.1
Uganda	Baganda	100	4x	AGVP	WTSI	2.7
South Africa	Zulu	100	4x	AGVP	WTSI	2.3
Ethiopia	Amhara	24	8x	AGVP	WTSI	1.0
Ethiopia	Gumuz	24	8x	AGVP	WTSI	1.0
Ethiopia	Oromo	24	8x	AGVP	WTSI	1.0
Ethiopia	Somali	24	8x	AGVP	WTSI	1.0
Ethiopia	Wolayta	24	8x	AGVP	WTSI	1.0
Egypt	Unspecified	100	8x	GDAP	WTSI	5.0
South Africa	Khoe-San (Nama)	104	4x	GDAP	WTSI	3.6
Nigeria	Esan (ESN)	99	4x	1000G	1000G	2.2
Gambia	Gambian (GWD)	113	4x	1000G	1000G	2.9
Kenya	Luhya (LWK)	101	4x	1000G	1000G	2.2
Sierra Leone	Mende (MSL)	85	4x	1000G	1000G	1.9
Nigeria	Yoruba (YRI)	109	4x	1000G	1000G	2.3
South Africa	Sotho	8	30x	SAHGP	CBIO	≈0.7TB
South Africa	Xhosa	8	30x	SAHGP	CBIO	≈0.7TB
Burkina Faso	Mossi	50	30x	Ramsay	Baylor	≈22.5TB
Camerron		25	30x	Wonkam	Baylor	≈11.3TB
Mali	Bambara	50	30x	Landoure	Baylor	≈22.5TB
Nigeria	Birom-Berom	16	30x	Adebamowo/Abimiku	Baylor	≈7.2TB
Ghana	Ewe/Akan	50	30x	Adu/Ojo/Burke	Baylor	≈22.5TB
Benin	Fon	50	30x	Awadalla	Baylor	≈22.5TB
Zambia		50	30x	Fowler	Baylor	≈22.5TB
Cameroon		25	30x	Fowler	Baylor	≈11.3TB
Total*		3363				81.7+≈143.6

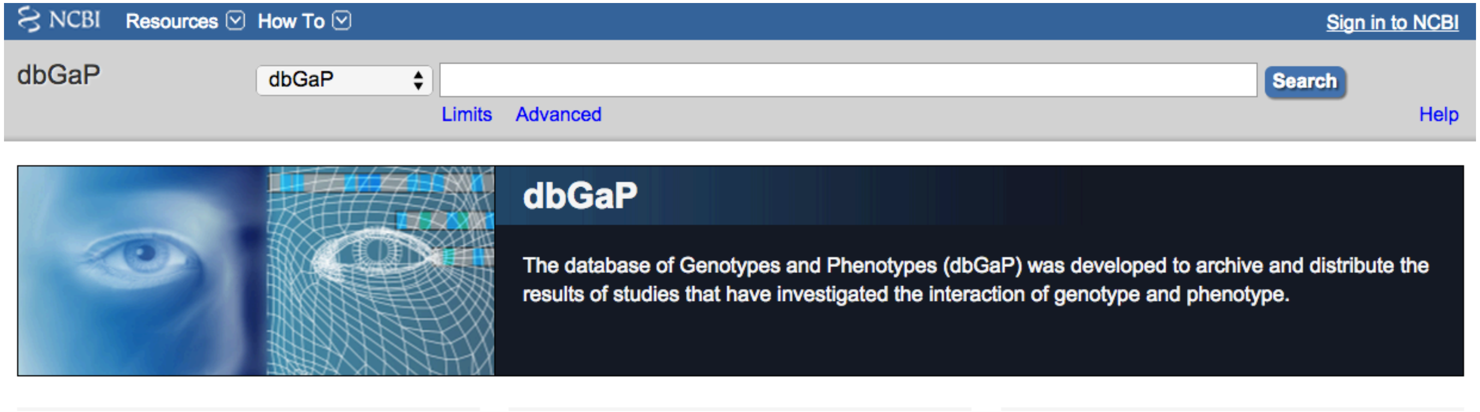
Table 1: Sample sets to be included in design of the African chip array. Size refers to size of mapped sequence reads stored in bam file format.


Egyptian
Gambian
Bambara
Mossi
Berom
Mende
Fon
Yoruba
Ewe/Akan
Esan
Cameroon
Cameroon
Gumuz
Amhara
Oromo
Wolayta
Somali
Baganda
Ba
Kiga
Luhya
Barundi
Zambia
Khoe-San
Sotho
Zulu
Xhosa
1000G
UG2G
AGVP
GDAP
SAHGP
Baylor
1
10
100
1000
10000

Resources with both genotype & phenotype data

dbGAP: The database of Genotypes and Phenotypes

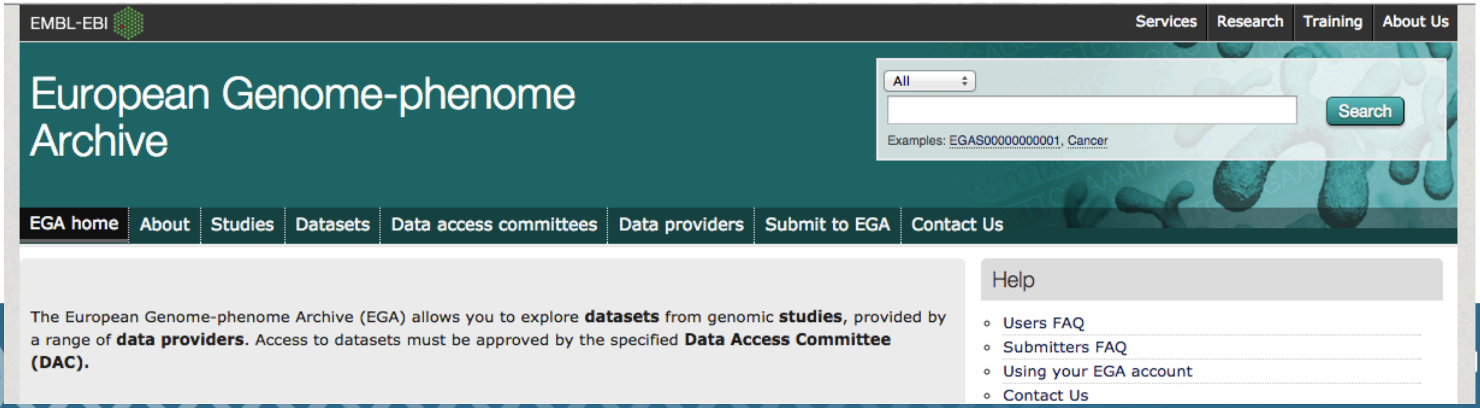
http://www.ncbi.nlm.nih.gov/gap



The database of Genotypes and Phenotypes (dbGaP) was developed to archive and distribute the results of studies that have investigated the interaction of genotype and phenotype.

EGA: European Genome-phenome Archive

https://www.ebi.ac.uk/ega/home



The European Genome-phenome Archive (EGA) allows you to explore **datasets** from genomic **studies**, provided by a range of **data providers**. Access to datasets must be approved by the specified **Data Access Committee (DAC).**

In this course:

- Genetic association studies
- Genome wide association studies (GWAS)
- Population genetics
- Meta-analysis of GWAS
- Sequencing
- Public resources

- Practicals: R, Rstudio, plink, Linux

What next?

Review course materials

Practice with the aid of the practicals

Apply what you learnt to your own/other datasets

Keep in touch with your colleagues

What next?

Indicate to your PI/Group Leader that you now have these skills; interest in applying the skills

Participate in the analysis of consortium and similar projects

Keep abreast of developments in the field

Look for opportunities for career development

What next?

Safe travels!