Genome-wide analysis challenges for SCD

Jeffrey C. Barrett

NIH SCD Workshop, December 9, 2011

Two parallel goals in disease genetics

For a given disease, can we:

1. Explain heritability
2. Understand biology

Two parallel goals in disease genetics

For a given disease, can we:

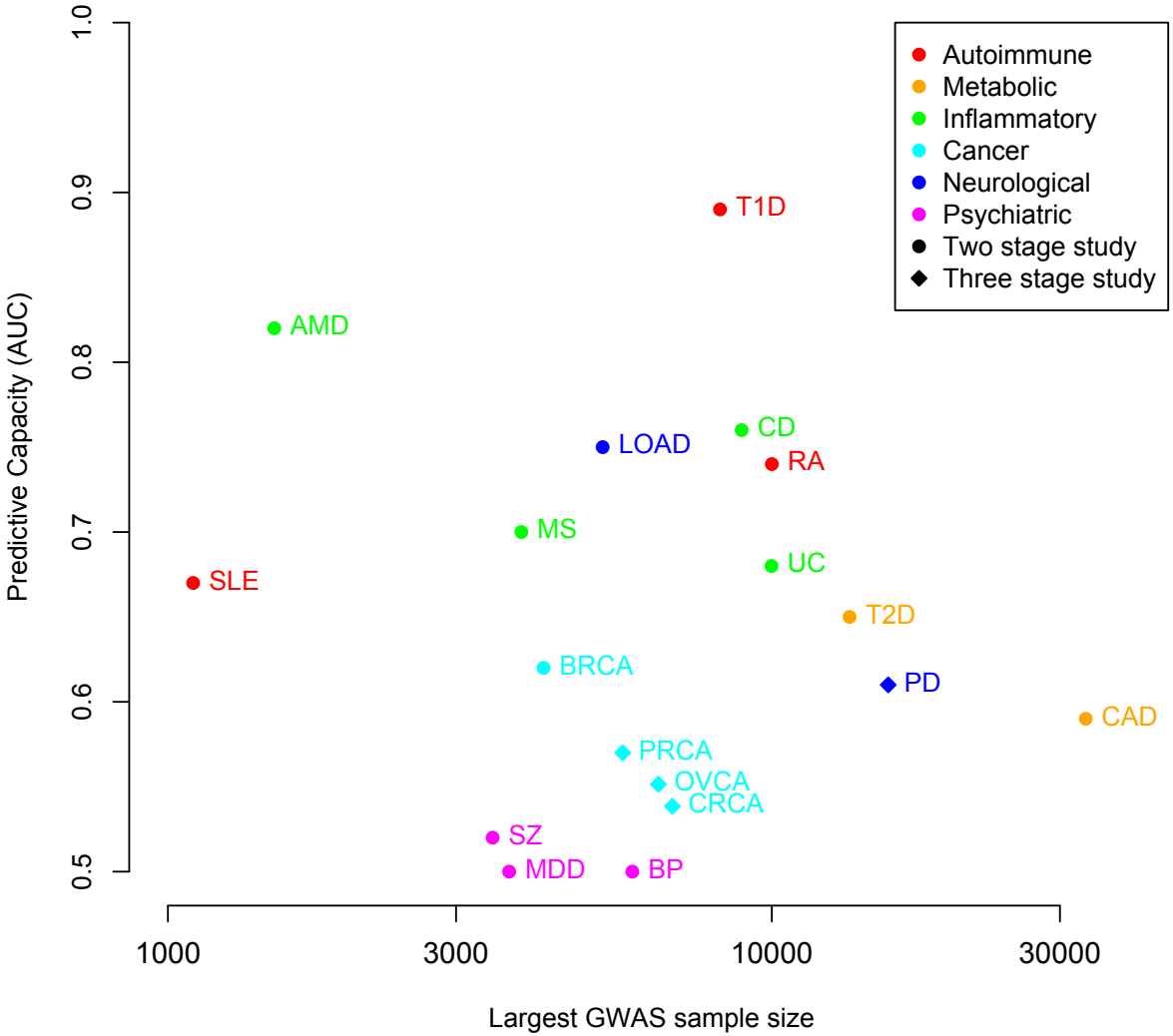
1. Explain heritability (prediction/prognosis)
2. Understand biology (prevention/treatment)

Two parallel goals in disease genetics

For a given disease, can we:

1. Explain heritability (prediction/prognosis)
2. Understand biology (prevention/treatment)

Prediction from genomes is still hard



Two parallel goals in disease genetics

For a given disease, can we:

1. Explain heritability (prediction/prognosis)
2. Understand biology (prevention/treatment)

HbF/SCD as poster child for biological insight from GWAS

nature genetics

A QTL influencing F cell production maps to a gene encoding a zinc-finger protein on chromosome 2p15

Stephan Menzel[1], Chad Garner[2], Ivo Gut[3], Fumihiko Matsuda[3], Masao Yamaguchi[3], Simon Heath[3], Mario Foglio[3], Diana Zelenika[3], Anne Boland[3], Helen Rooks[1], Steve Best[1], Tim D Spector[4], Martin Farrall[5], Mark Lathrop[3] & Swee Lay Thein[1,6]

HbF/SCD as poster child for biological insight from GWAS

A QTL influencing F cell production maps to a gene encoding a zinc-finger protein on chromosome 2p15

Stephan Menzel[1], Chad Garner[2], Ivo Gut[3], Fumihiko Matsuda[3], Masao Yamaguchi[3], Simon Heath[3], Mario Foglio[3], Diana Zelenika[3], Anne Boland[3], Helen Rooks[1], Steve Best[1], Tim D Spector[4], Martin Farrall[5], Mark Lathrop[3] & Swee Lay Thein[1,6]

Number of papers mentioning *BCL11A* and hemoglobin before this: 0

Number of papers mentioning *BCL11A* since: 47

HbF/SCD as poster child for biological insight from GWAS

nature genetics

A QTL influencing F cell production maps to a gene encoding a zinc-finger protein on chromosome 2p15

Stephan Menzel[1], Chad Garner[2], Ivo Gut[3], Fumihiko Matsuda[3], Masao Yamaguchi[3], Simon Heath[3], Mario Foglio[3], Diana Zelenika[3], Anne Boland[3], Helen Rooks[1], Steve Best[1], Tim D Spector[4], Martin Farrall[5], Mark Lathrop[3] & Swee Lay Thein[1,6]

Correction of Sickle Cell Disease in Adult Mice by Interference with Fetal Hemoglobin Silencing

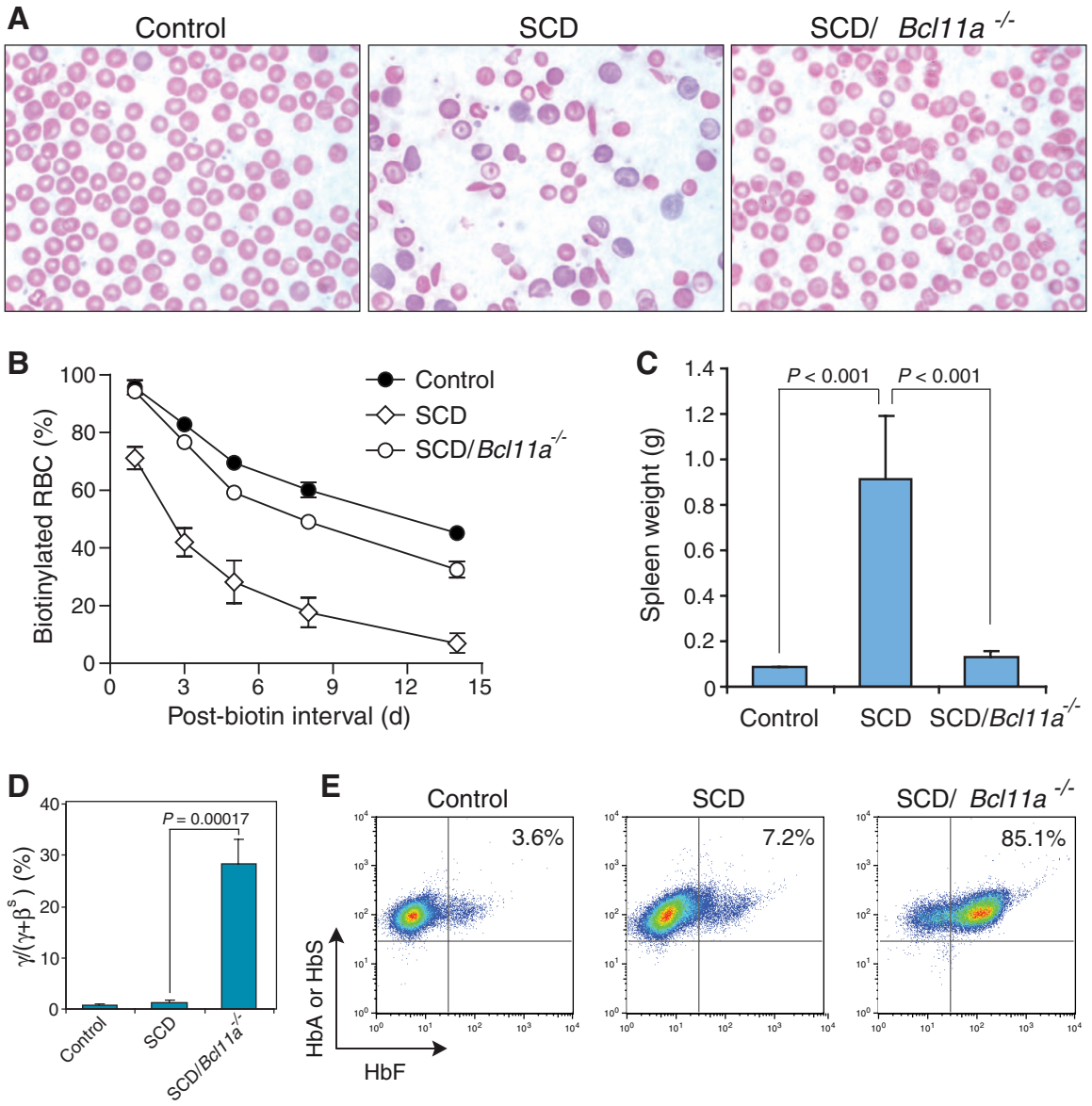
Jian Xu,[1,2] Cong Peng,[1*] Vijay G. Sankaran,[1,5*] Zhen Shao,[1] Erica B. Esrick,[1,3] Bryan G. Chong,[1] Gregory C. Ippolito,[4] Yuko Fujiwara,[1,2] Benjamin L. Ebert,[3] Philip W. Tucker,[4] Stuart H. Orkin[1,2]†

Persistence of human fetal hemoglobin (HbF, $\alpha_2\gamma_2$) in adults lessens the severity of sickle cell disease (SCD) and the β-thalassemias. Here, we show that the repressor BCL11A is required in vivo for silencing of γ-globin expression in adult animals, yet dispensable for red cell production. BCL11A serves as a barrier to HbF reactivation by known HbF inducing agents. In a proof-of-principle test of BCL11A as a potential therapeutic target, we demonstrate that inactivation of BCL11A in SCD transgenic mice corrects the hematologic and pathologic defects associated with SCD through high-level pancellular HbF induction. Thus, interference with HbF silencing by manipulation of a single target protein is sufficient to reverse SCD.

Number of papers mentioning *BCL11A* and hemoglobin before this: 0

Number of papers mentioning *BCL11A* since: 47

HbF/SCD as poster child for biological insight from GWAS



Matching biological questions to technologies

In my disease...

- What role does common variation in Europeans play? **GWAS**
- What role does low frequency variation play? **Sequencing, Targeted Sequencing, GWAS 2.0**
- What role does rare or private variation play? **Sequencing, Targeted Sequencing**
- What is the genetic architecture of disease in non-European populations? **GWAS 2.0**

MalariaGEN
GENOMIC EPIDEMIOLOGY NETWORK

- Network
  - People
  - Partner Institutions
  - Resource Centre
  - Funders
- Science
  - Consortial Projects
  - Local Projects
  - Data Access
  - Publications
- Ethics
  - Network Policies
  - Ethics Research
- Information
  - News
  - Meetings
  - Contact Information

Network
Researchers in 21 countries, working together to apply new genome research tools to the problems facing malaria-affected communities.

Science
By discovering natural mechanisms of disease resistance, we aim to develop more effective ways of preventing malaria.

Ethics
Working in partnership with communities to address the ethical and social questions posed by advances in genome research.

Wellcome Trust Advanced Course - EXTENDED DEADLINE

The Wellcome Trust is going to host an Advanced Course entitled "Malaria Experimental Genetics". The aim is to give participants a working knowledge of and practical experience in cutting edge Plasmodium experimental techniques. The goal is to facilitate the participants' own research...

More >>

Other Stories

6th Jul 2010
The aim of The Human Heredity and Health in Africa Project (H3Africa) is to support African scientists...

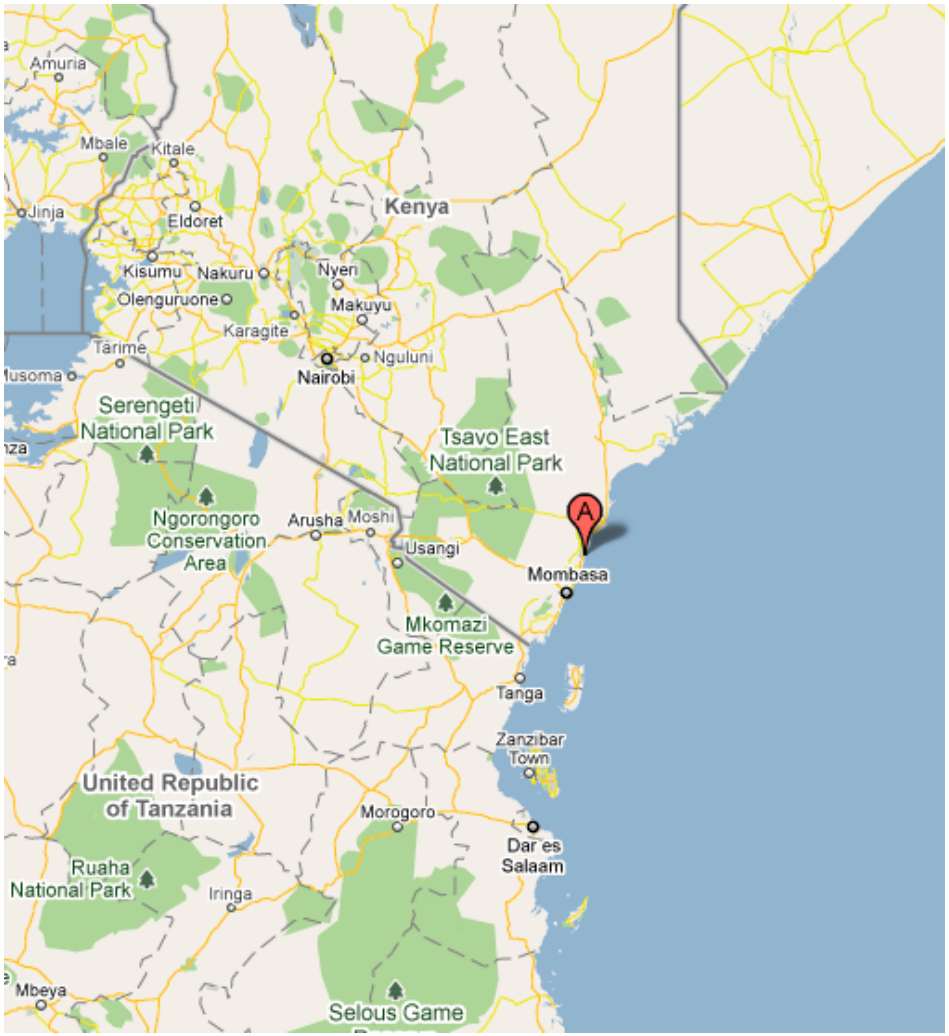
15th Jun 2010
A Wellcome Trust Advanced Course is to be held on Genomic Epidemiology in Africa at the KEMRI/Wellcome Trust Research...

20th Apr 2010
MapSeq, an interactive web-based database of genome variation in the malaria parasite, has recently been launched....

MalariaGEN is funded through the support of the following organisations:

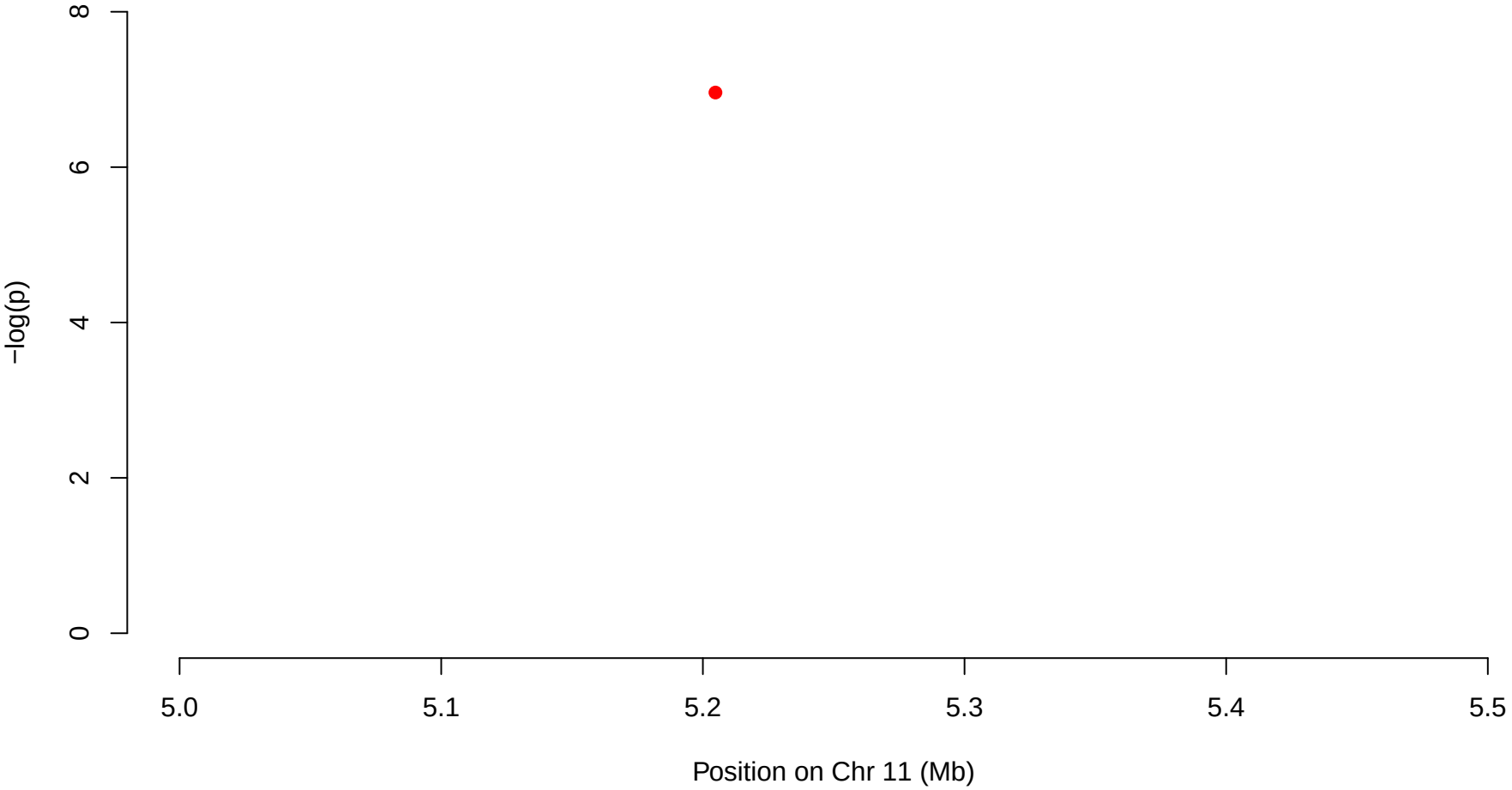
wellcome trust · Bill & Melinda Gates foundation · Foundation for the National Institutes of Health · MRC Medical Research Council

GWAS of severe malaria in Kenyan children

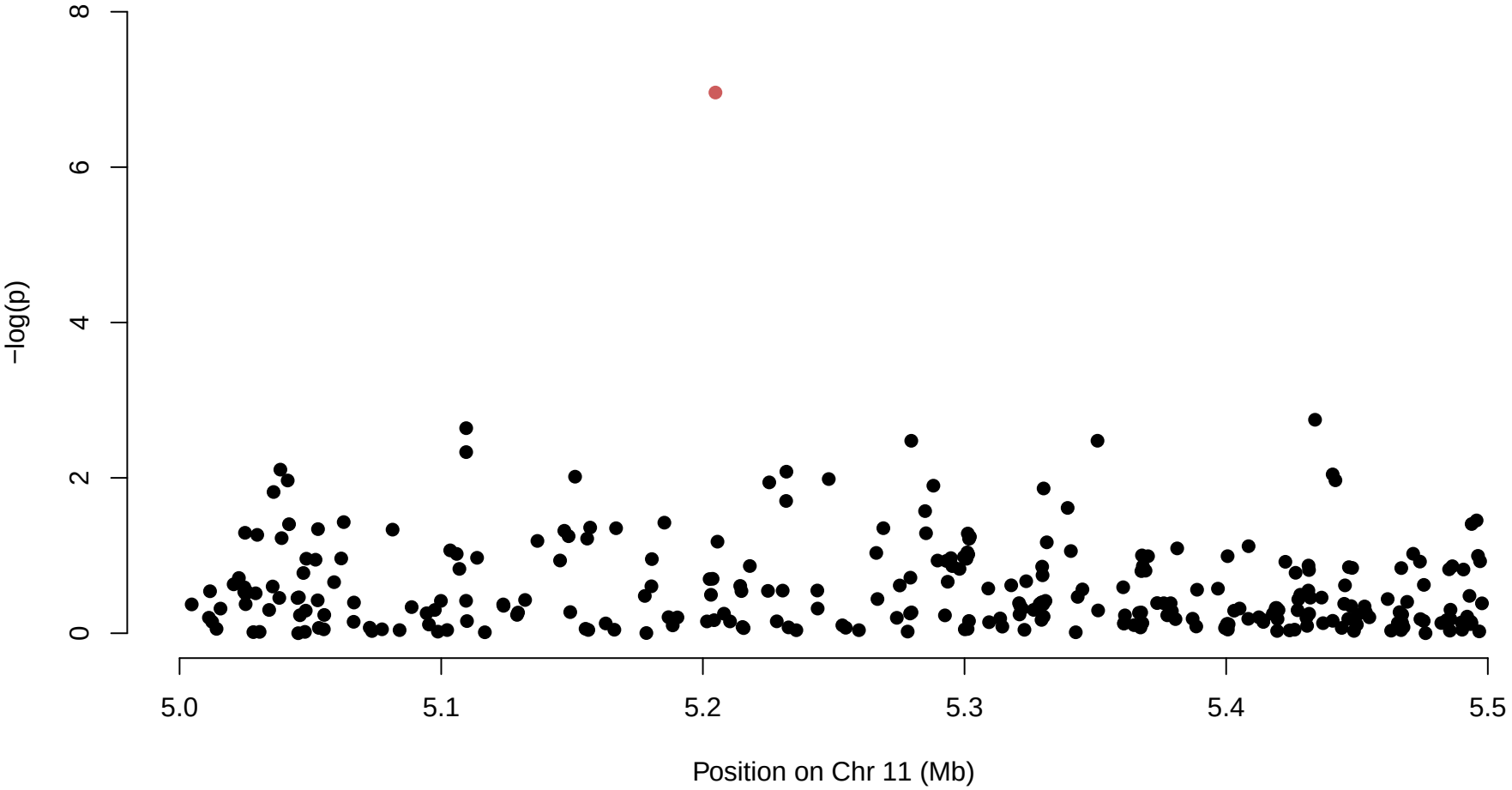


Kilifi population: 34,000

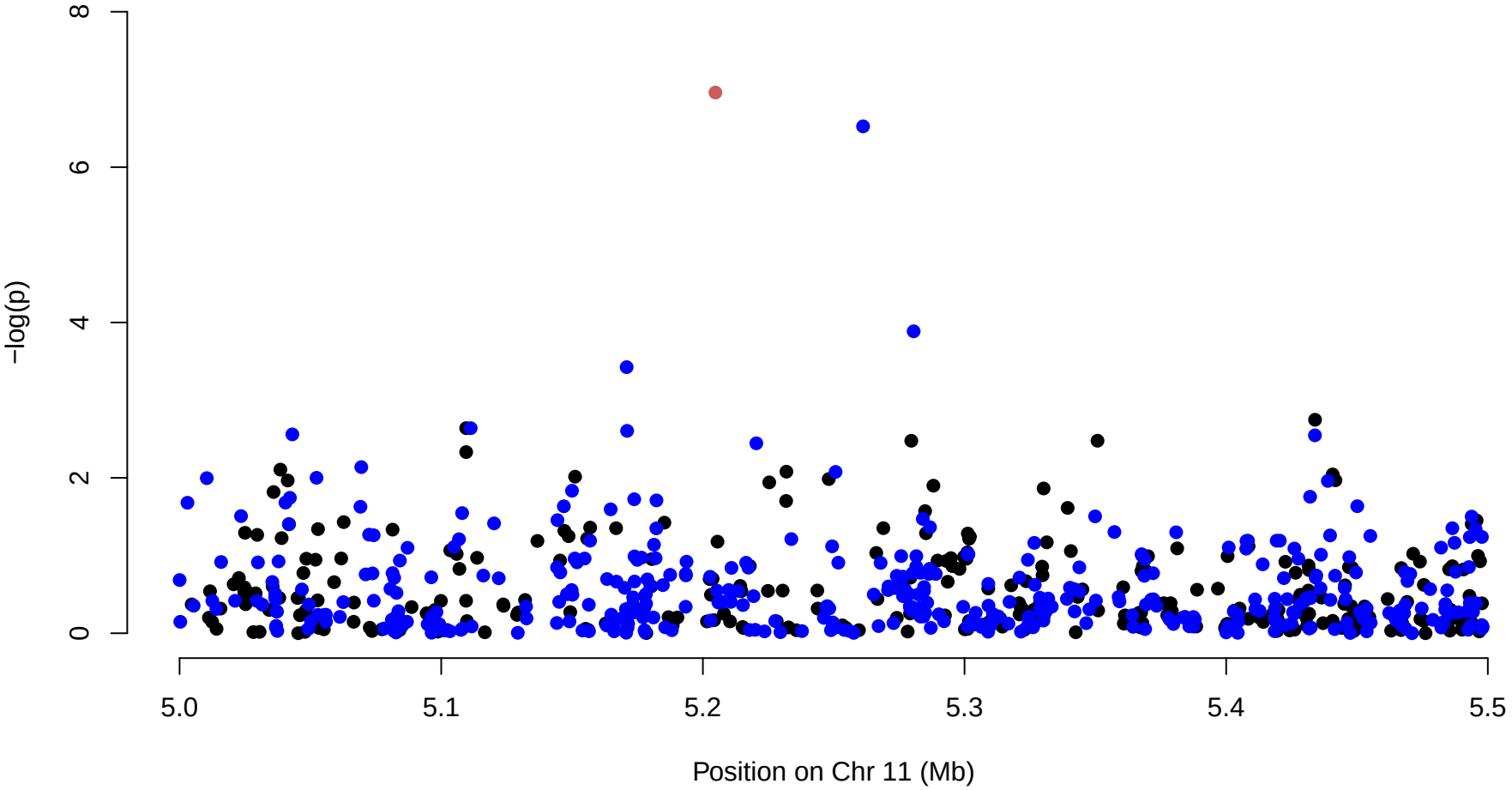
The HbS locus in Kenya: causal allele



The HbS locus in Kenya: Omni1M

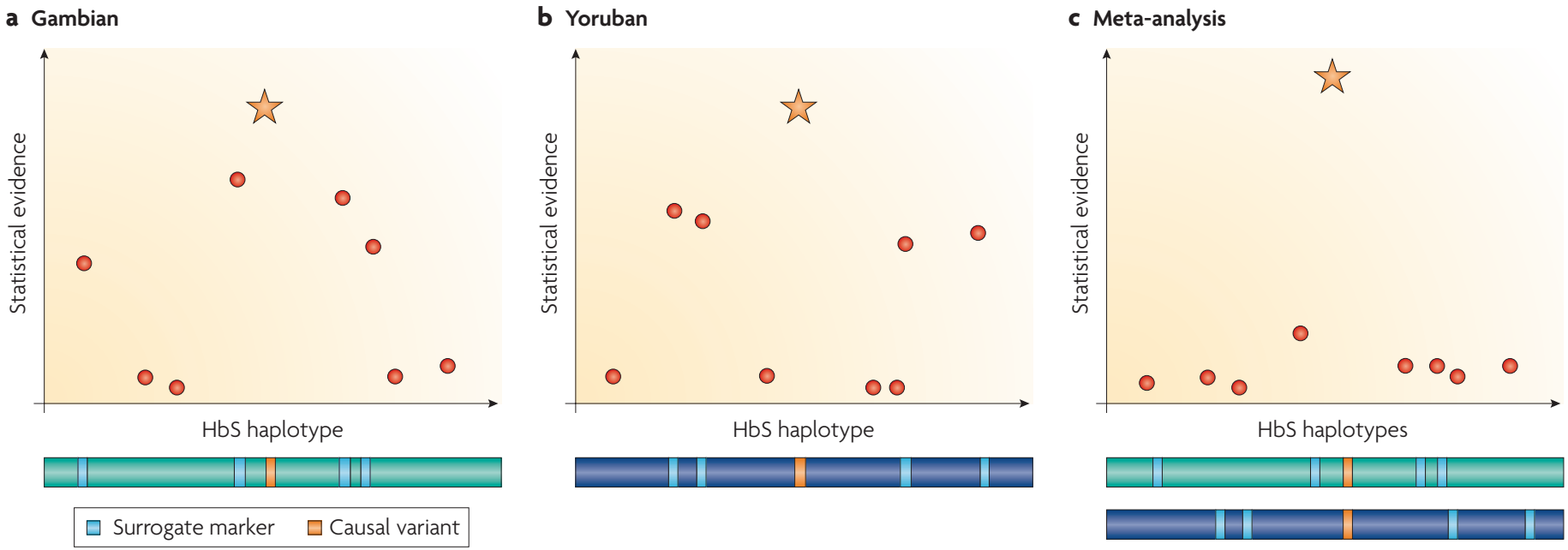


The HbS locus in Kenya: Omni2.5M



The HbS locus and malaria

HbS mutation (rs334) in *HBB*, which causes sickle cell disease when homozygous, confers protection (het OR < 0.2) from malaria.

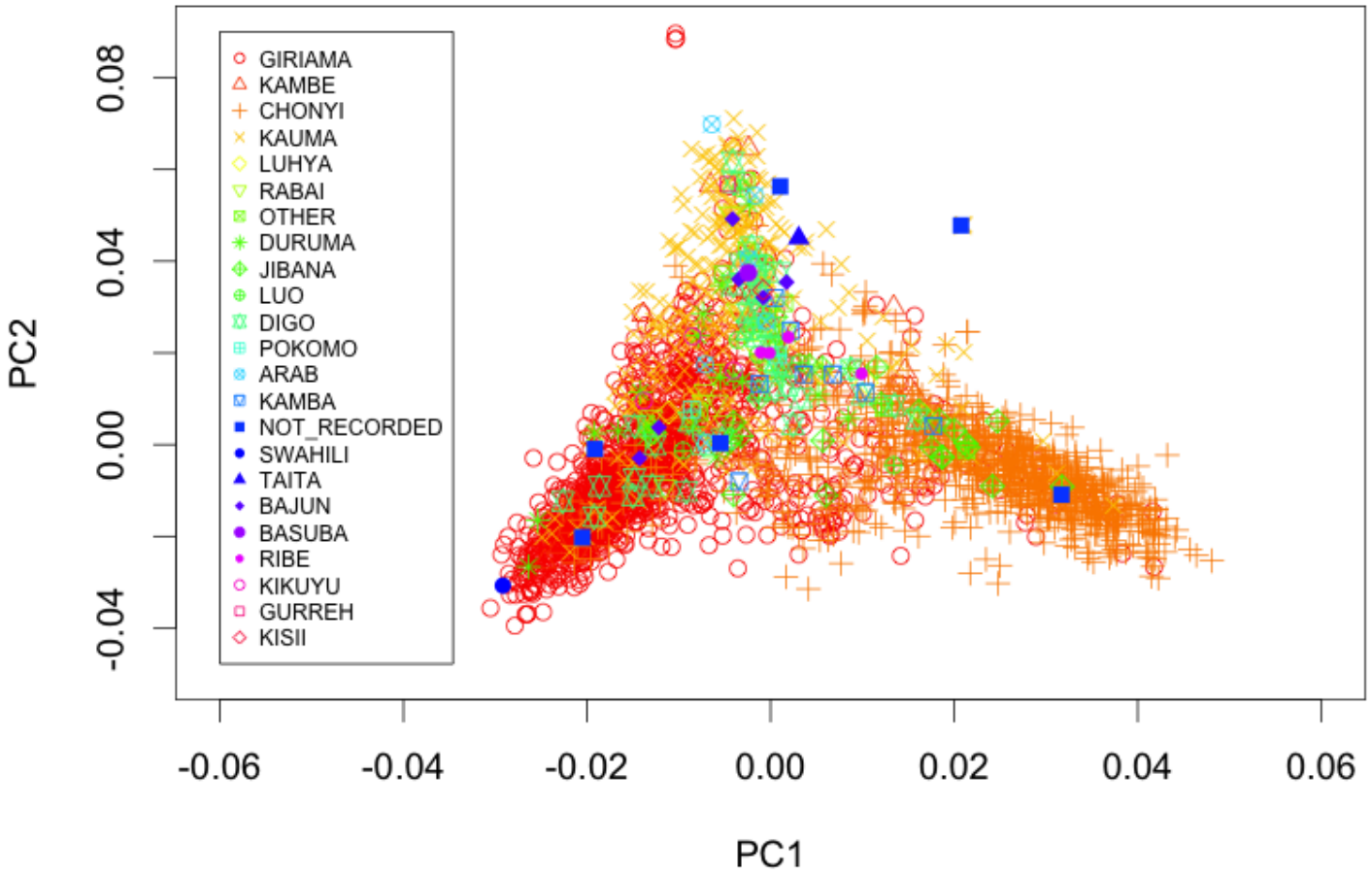


Teo, Small, Kwiatkowski. *Nat Rev Genet*. 2010

Structure & admixture within Kenya dataset

Ethnicity	N
Giriama	1838
Chonyi	1102
Kauma	313
Kambe	71
Digo	59
Jibana	32
Duruma	25
15 others	90

Population structure within Kilifi is complex



MalariaGEN African meta-analysis

- 3600 samples from Kenya
- 4000 samples from Malawi
- 3000 samples from Gambia
- 600 samples from Ghana

Mix of platforms, using targeted 2.5M genotyping and imputation for meta-analysis

"Genotyping arrays: they're just good genetic hygiene!"

GWAS of SCD clinical outcome in Tanzania

- Collaboration with Julie Makani, Muhimbili, Dar es Salaam
- 2,000 SCD patients with HbF/Hb measurements; records of heamolysis, stroke
- Genotyped on Illumina Omni2.5M chip (data generation nearly complete)

The experiment–analysis cost balance of WGS and GWAS

Consider approximately 800 Phase I samples from 1000 Genomes

	Finished storage (Gb)	**Raw storage (Gb)**	**Processing (CPU-days)**
GWAS			
WGS			

Mark Depristo, 1000 Genomes

The experiment–analysis cost balance of WGS and GWAS

Consider approximately 800 Phase I samples from 1000 Genomes

	Finished storage (Gb)	**Raw storage (Gb)**	**Processing (CPU-days)**
GWAS	50		
WGS	50		

Mark Depristo, 1000 Genomes

The experiment–analysis cost balance of WGS and GWAS

Consider approximately 800 Phase I samples from 1000 Genomes

	Finished storage (Gb)	**Raw storage (Gb)**	**Processing (CPU-days)**
GWAS	50	4	
WGS	50		

Mark Depristo, 1000 Genomes

The experiment–analysis cost balance of WGS and GWAS

Consider approximately 800 Phase I samples from 1000 Genomes

	Finished storage (Gb)	**Raw storage (Gb)**	**Processing (CPU-days)**
GWAS	50	4	
WGS	50	50,000	

Mark Depristo, 1000 Genomes

The experiment–analysis cost balance of WGS and GWAS

Consider approximately 800 Phase I samples from 1000 Genomes

	Finished storage (Gb)	**Raw storage (Gb)**	**Processing (CPU-days)**
GWAS	50	4	715
WGS	50	50,000	

Mark Depristo, 1000 Genomes

The experiment–analysis cost balance of WGS and GWAS

Consider approximately 800 Phase I samples from 1000 Genomes

	Finished storage (Gb)	**Raw storage (Gb)**	**Processing (CPU-days)**
GWAS	50	4	715
WGS	50	50,000	20,000

Mark Depristo, 1000 Genomes

The experiment–analysis cost balance of WGS and GWAS

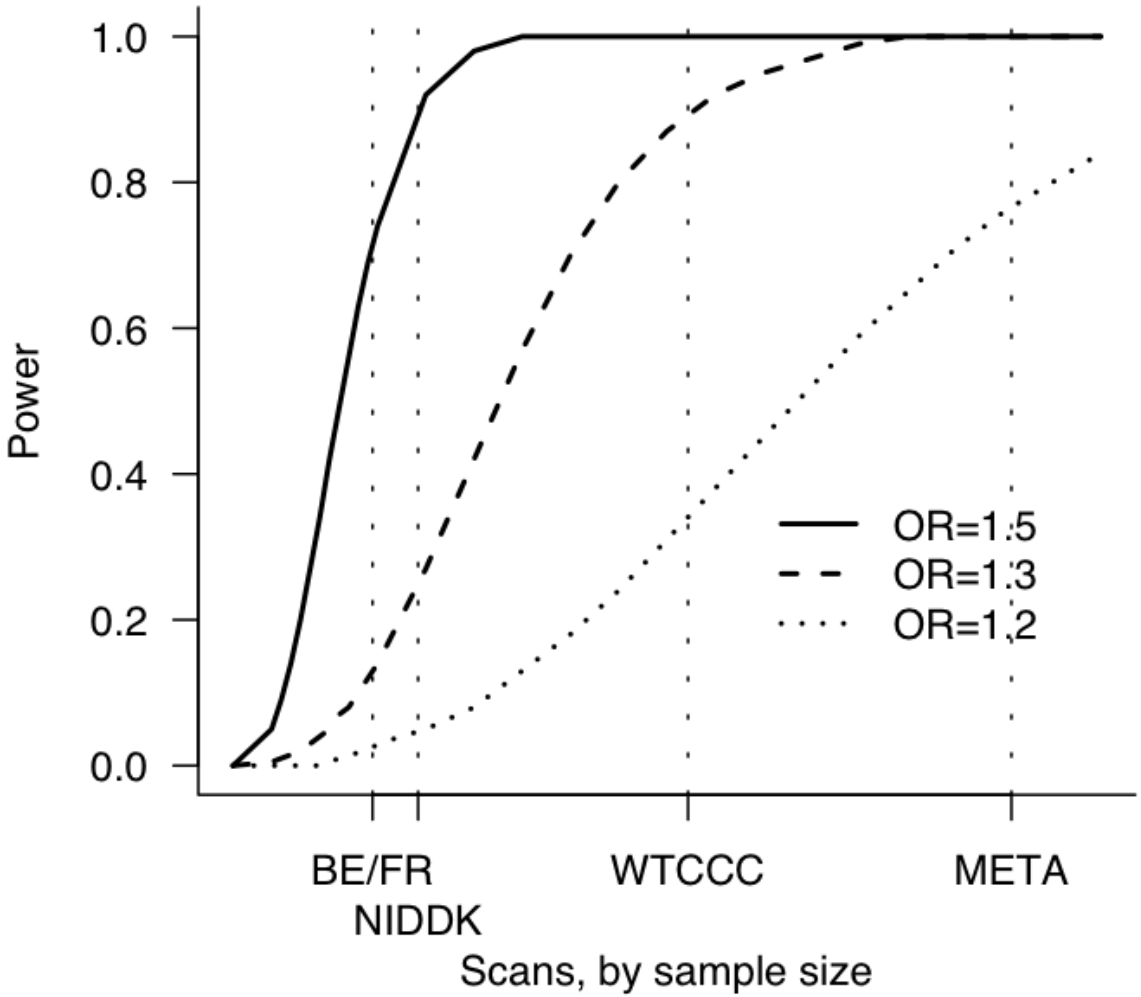
Consider approximately 800 Phase I samples from 1000 Genomes

	Finished storage (Gb)	**Raw storage (Gb)**	**Processing (CPU-days)**
GWAS	50	4	715
WGS	50	50,000	20,000

Computational bottleneck almost entirely upstream of analysis
(Moore's law since 1st GWAS: 8x)

Mark Depristo, 1000 Genomes

Success needed large sample sizes via collaboration

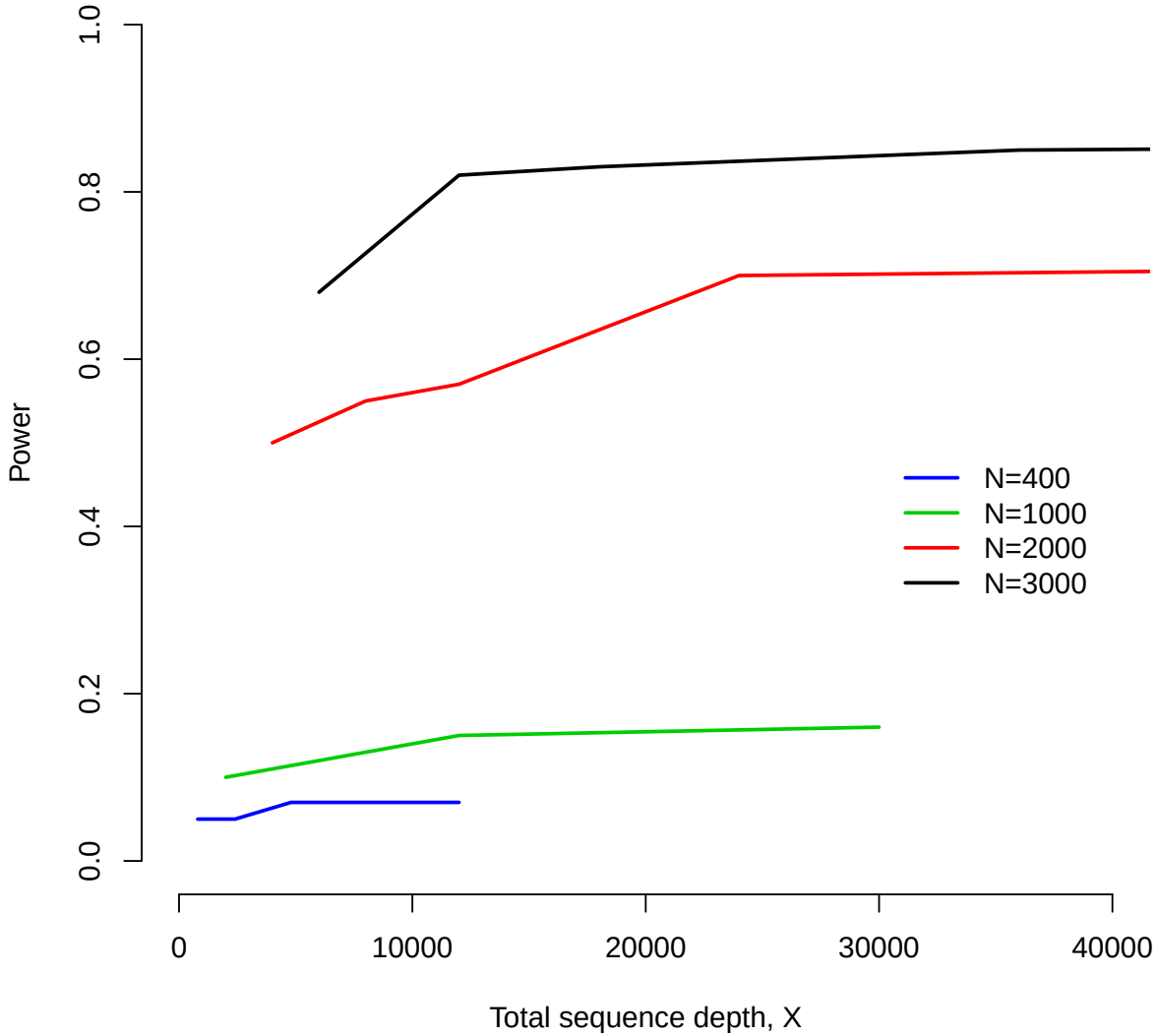


ɔristo, 1000 Genomes

Whole genome sequence association...

Adapted from Li *et al. Genome Research*, 2011

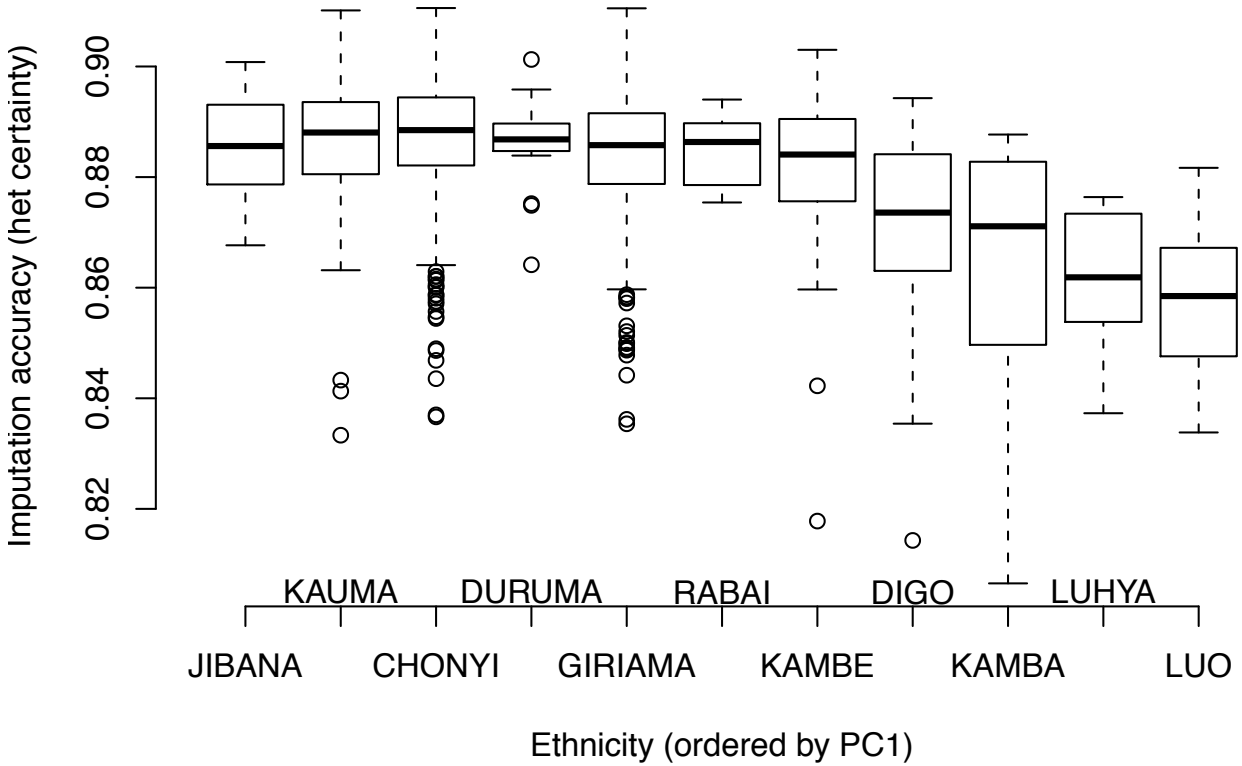
Whole genome sequence association. . . needs big samples



Adapted from Li *et al. Genome Research*, 2011

Sequencing in African populations

- 1000 Genomes (100 samples each): 2 collections sequenced, 1 being collected, 3 going through IRB
- Sanger (100 samples each): 10 collections currently being sequenced



Julie Makani
Siana Nkya
Sharon Cox
D Kwiatkowski
Chris Spencer
MalariaGEN
Kate Morley
Luke Jostins
James Morris

wellcome trust