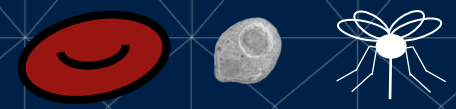
The genetics of susceptibility to malaria

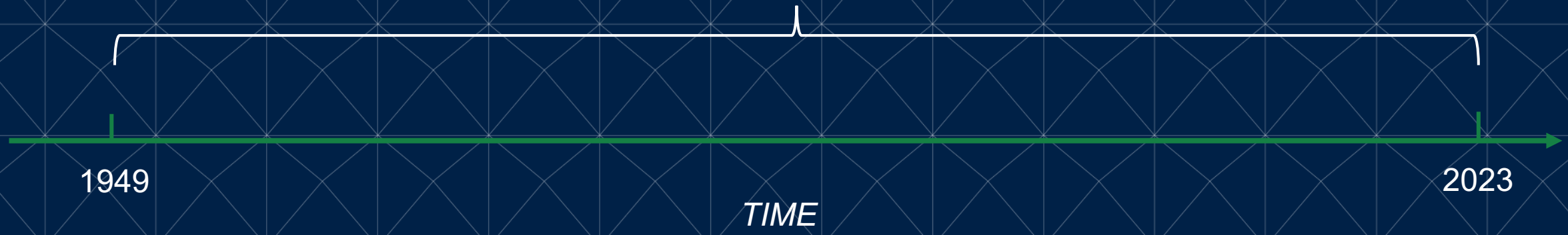
Gavin Band

Adelphi Genetics Forum

Wednesday 18th October 2023

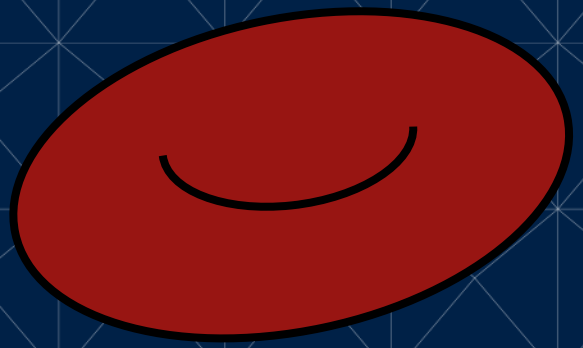


75 years
in
30 minutes

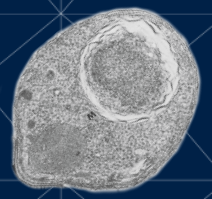


A tale of three genomes

Anopheles mosquitos



Human host



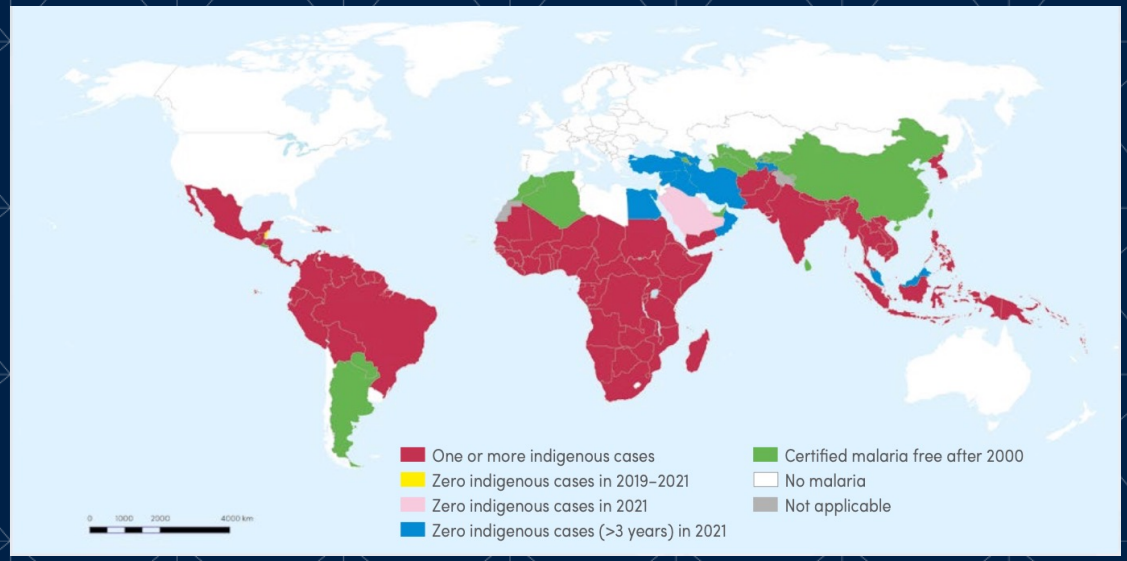
Plasmodium falciparum
(malaria parasite)

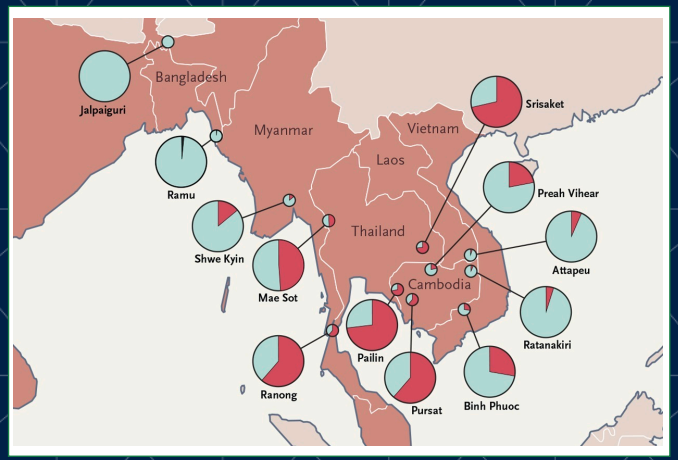
also *P.vivax*, *P. malariae*, *P.ovale*

Reasons to be optimistic

Several countries have eliminated malaria, several more on course to do so by 2025

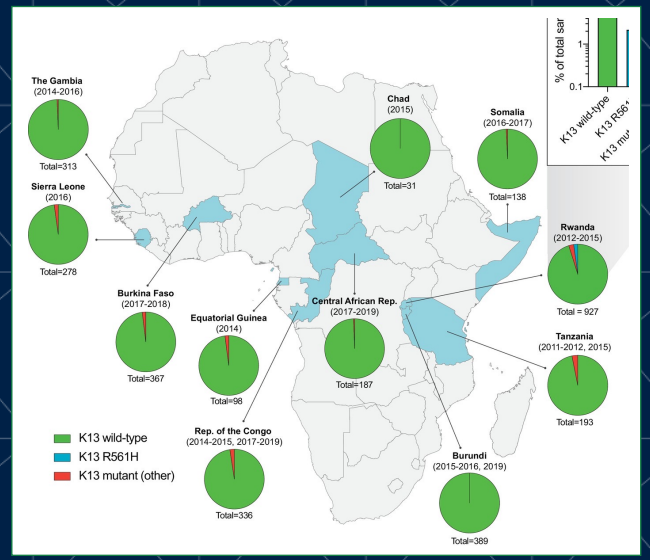
WHO World Malaria Report 2022





Reasons to be concerned

Rising antimalarial and insecticide resistance, including in Africa

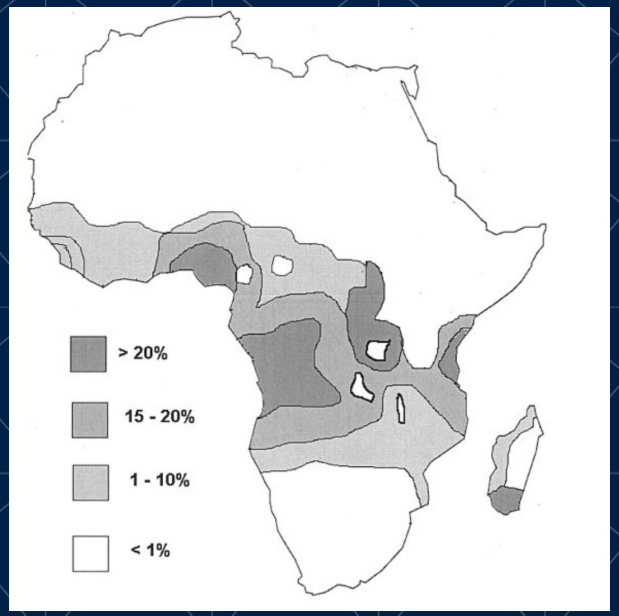


Ashley et al, N. Engl. J. Med. 2014, Stokes et al, eLife 2021

Talk outline

The malaria hypothesis

Distribution of *P.falciparum* malaria in Africa



Distribution of sickle heterozygotes in Africa

"Montalenti [recalled] a gene, lethal in the homozygous state, which occurs in the heterozygous state with such frequency in some populations (more than 10%) that one must admit that it represents [..] an advantage for individuals that carry it"

- transcript of 1949 meeting between J.B.S Haldane and G. Montalenti

"1949 was an annus mirabilis when the modern phase of research on haemoglobinopathies was initiated. […] Montalenti observed high frequencies of thalassemia heterozygotes in formerly malarious parts of Italy, and I observed high frequencies of sickle-cell heterozygotes in formerly malarious parts of Kenya.

- A.C.Allison

P.W. Hedrick, Heredity (2011)
A. C. Allison - Biochem molecular biology education (2006)

TABLE I

	With Parasitaemia	Without Parasitaemia	Total
Sicklers	12 (27·9%)	31 (72·1%)	43
Non-sicklers ..	113 (45·7%)	134 (53·3%)	247

Allison A. C. BMJ (1954)

The discovery of the protective effect of sickle haemoglobin.

Infection observed at ~2-fold lower rates in A/S individuals in this sample of children from Uganda (N = 290)

TABLE I

	With Parasitaemia	Without Parasitaemia	Total
Sicklers	12 (27·9%)	31 (72·1%)	43
Non-sicklers ..	113 (45·7%)	134 (53·3%)	247

Allison A. C. BMJ (1954)

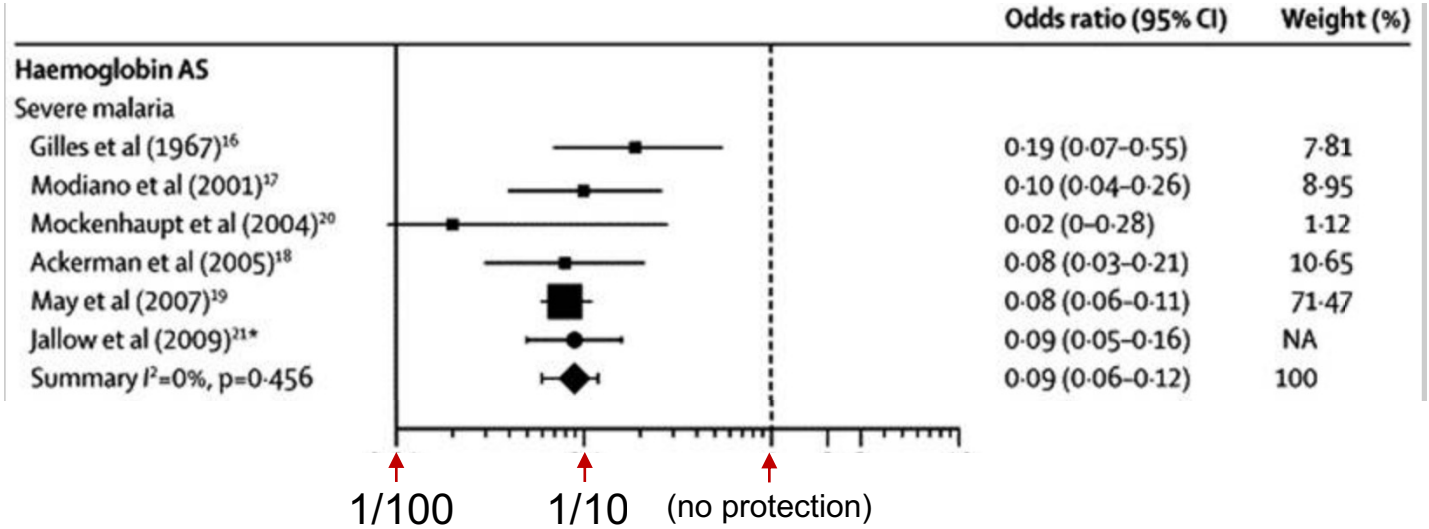
The discovery of the protective effect of sickle haemoglobin.

Infection observed at ~2-fold lower rates in this sample of children from Uganda (N = 290)

Effect size against *severe malaria* is stronger – a 10-fold protective effect .

Now replicated across many populations and studies.

In severe malaria cases:



Taylor et al Lancet Infect. Dis 2013

The malaria hypothesis

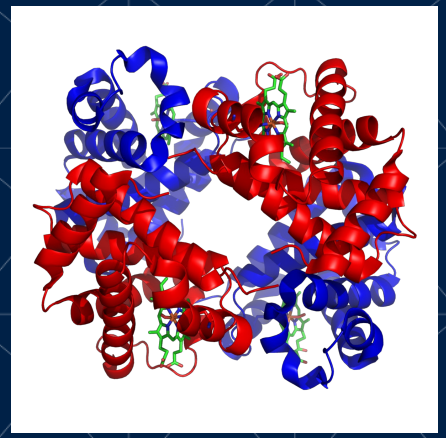
"In the less malarious parts of Accra about 5% of children died from malaria before they were 5 years old [..] fairly evenly spread over the first five years of life. In the more malarious suburban areas, there were many more deaths from malaria in the first two years of life"

- Colbourne M.J. and Edington G.M., BMJ 1956

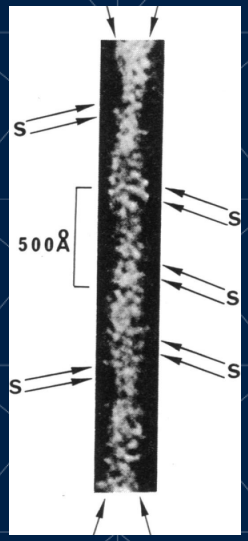
Malaria provides a *positive selective force* on these human genetic alleles – pushing them to higher frequency.

Sickle homozygotes and heterozygots

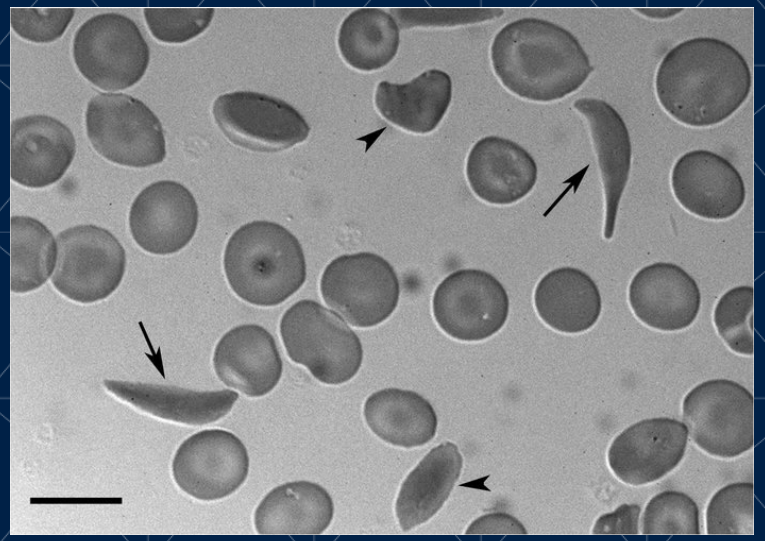
Sickle cell anaemia aetiology is well understood…



Haemoglobin



Edelstein 1973



P.S. Frenette and G.F. Atweh, JCI (2007)

…but it's still not really understood how sickle heterozygotes are protected
although many mechanisms have been proposed

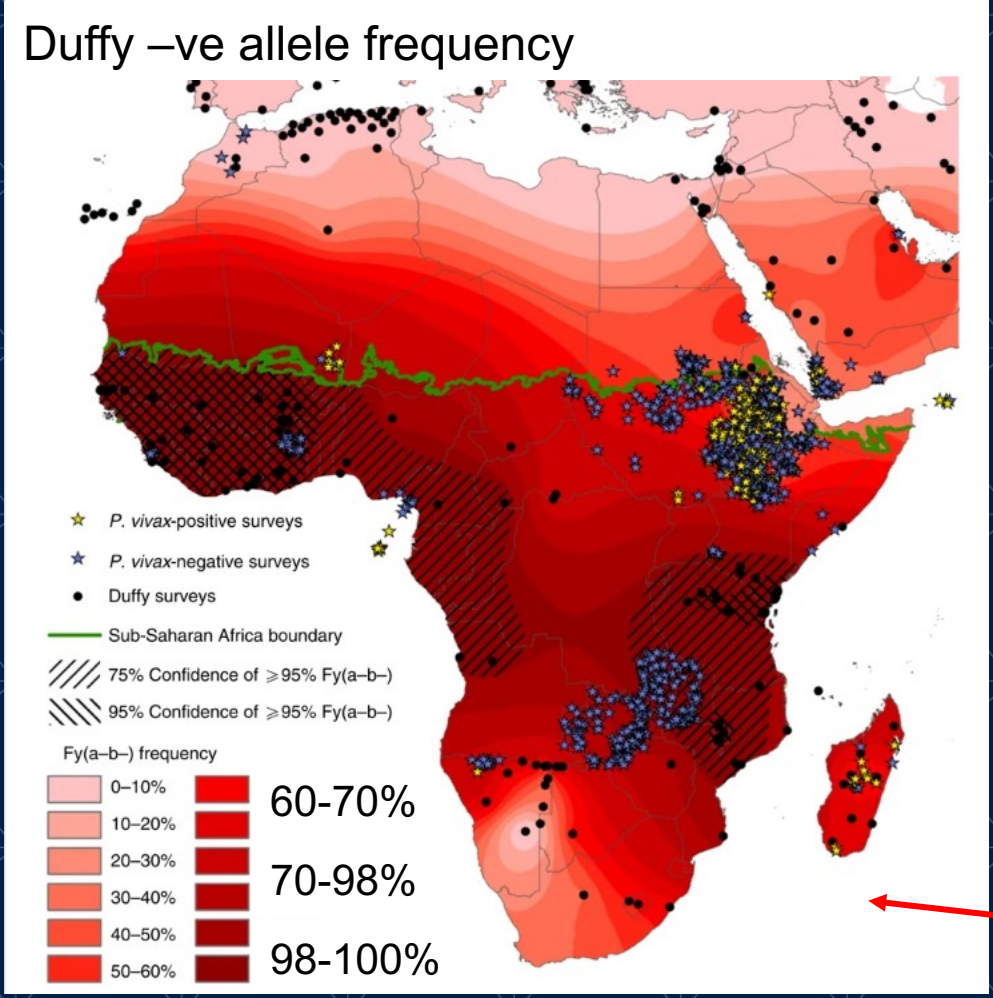
The malaria hypothesis

Malaria provides a *positive selective force* on these human genetic alleles – pushing them to higher frequency.

Anaemia causes a *negative* selective force.

What would happen if this balancing force wasn't there?

The sweep of protection against *P.vivax*



…CCAA**GATA**AGAGC…

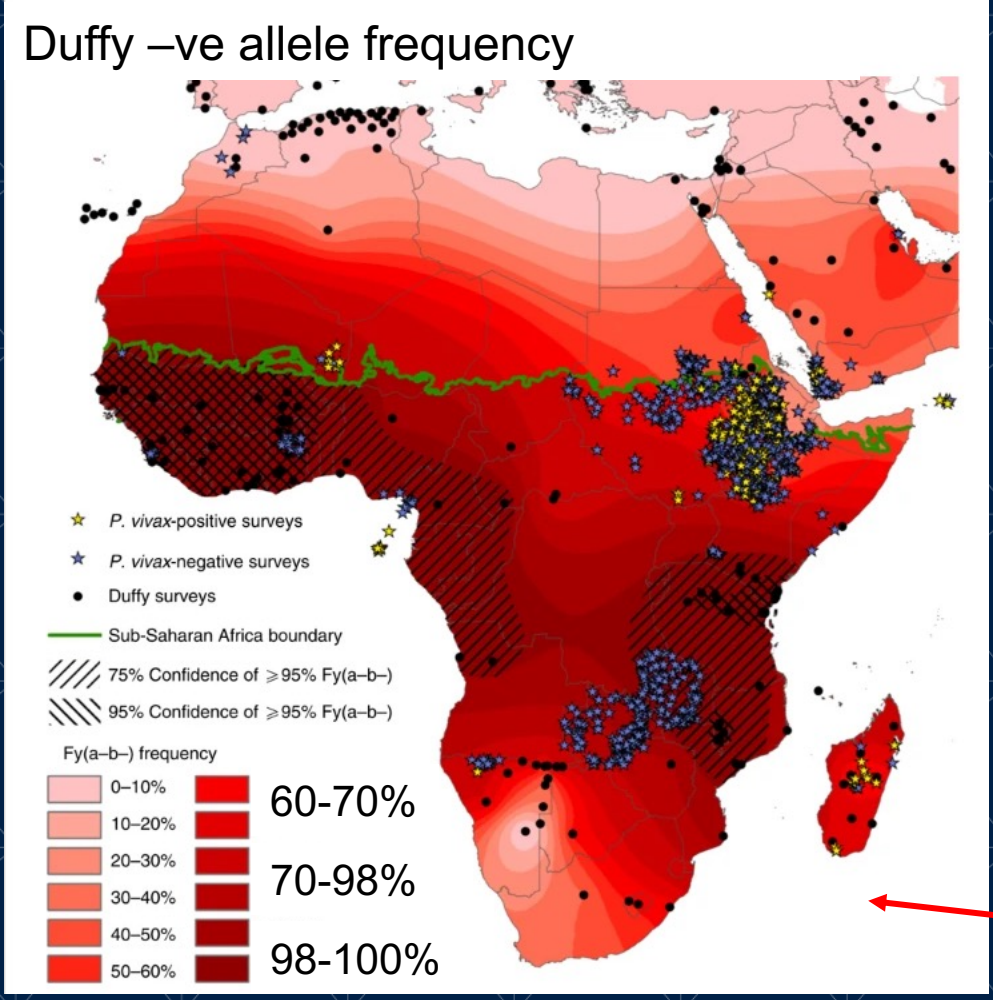
Gene expressed

The GATA transcription factor binds to a DNA motif upstream of the *DARC* gene, causing gene expression

One of the few examples of a selective sweep in the human genome

Rosalind R. Howes et al, Nat. Comm. 2011

The sweep of protection against *P.vivax*



…CCAA**GATA**AGAGC…

DARC

Gene expressed

…CCAAGGTAAGAGC…

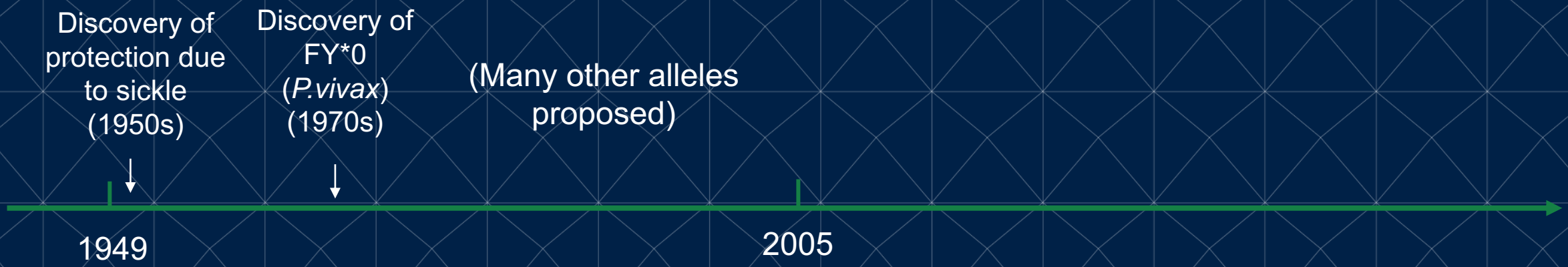
DARC

Gene not expressed

The ‘FY*0’ mutation ‘turns off’ expression by disrupting the GATA site.

One of the few examples of a selective sweep in the human genome

Rosalind R. Howes et al, Nat. Comm. 2011



"How Malaria Has Affected the Human Genome and What Human Genetics Can Teach Us about Malaria"
– Dominic Kwiatkowski Am J Hum Genet 2005

Genetic variants proposed to provide protection circa 2005

: *Beta-globin* HbS, HbC, HbE β - thalassaemia; *Alpha-globin* α - thalassaemia; Glycophorin A and B deletions; *G6PD* deficiency; "Band 3" ovalocytosis; *(DARC For P.vivax)*

Haptoglobin; O blood group; *PIEZO1*

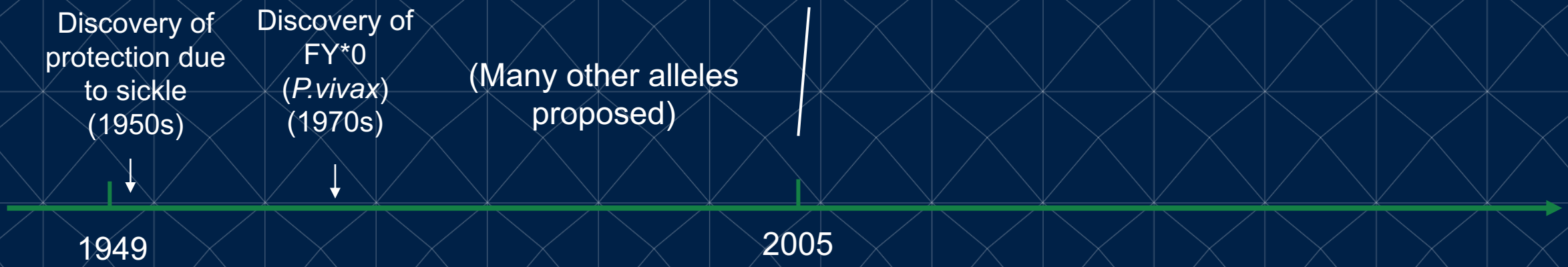
Cytoadhesion: *CD36*; *CR1*; *ICAM1*; *PECAM1*

Immunity: *FCGR2A*; *CD40* ligand Immune activation; HLA antigen presentation; Interferon receptor components; Interleukins

MBL2; *NOS2A*; Tumer Necrosis Factor; Interferon Gamma

"How Malaria Has Affected the Human Genome and What Human Genetics Can Teach Us about Malaria"
– Dominic Kwiatkowski Am J Hum Genet 2005

"When the same polymorphism has been tested in different geographical locations, the results have been variable - not simply failure to replicate but, in some cases, the association of the same polymorphism with susceptibility to severe malaria in one study and with resistance in another."



"How Malaria Has Affected the Human Genome and What Human Genetics Can Teach Us about Malaria"
– Dominic Kwiatkowski Am J Hum Genet 2005

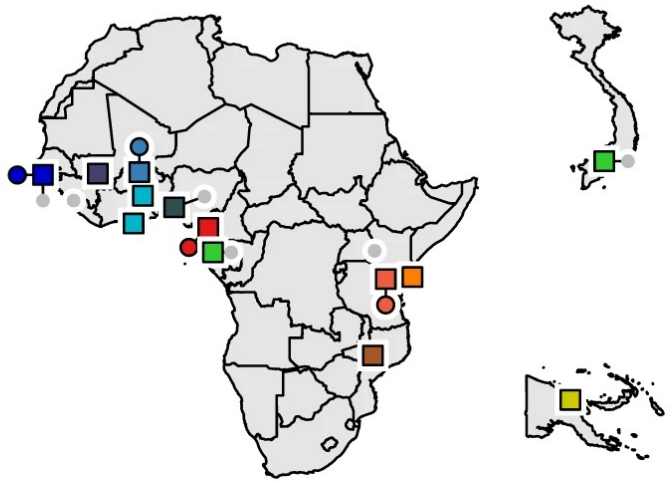
Talk outline

MalariaGEN collected over 30,000 cases and controls

a

Whole–genome sequences

Group	Trios	Duos	Other	TOTAL
Gambia				
FULA	31	1	5	100
JOLA	32	1	2	100
MANDINKA	33	0	1	100
WOLLOF	32	1	3	98
Burkina Faso				
MOSSI	0	0	57	57
Cameroon				
BANTU	5	3	11	31
SEMIBANTU	8	0	7	32
Tanzania				
CHAGGA	21	2	13	80
PARE	22	2	7	77
WASAAMBA	23	6	9	90

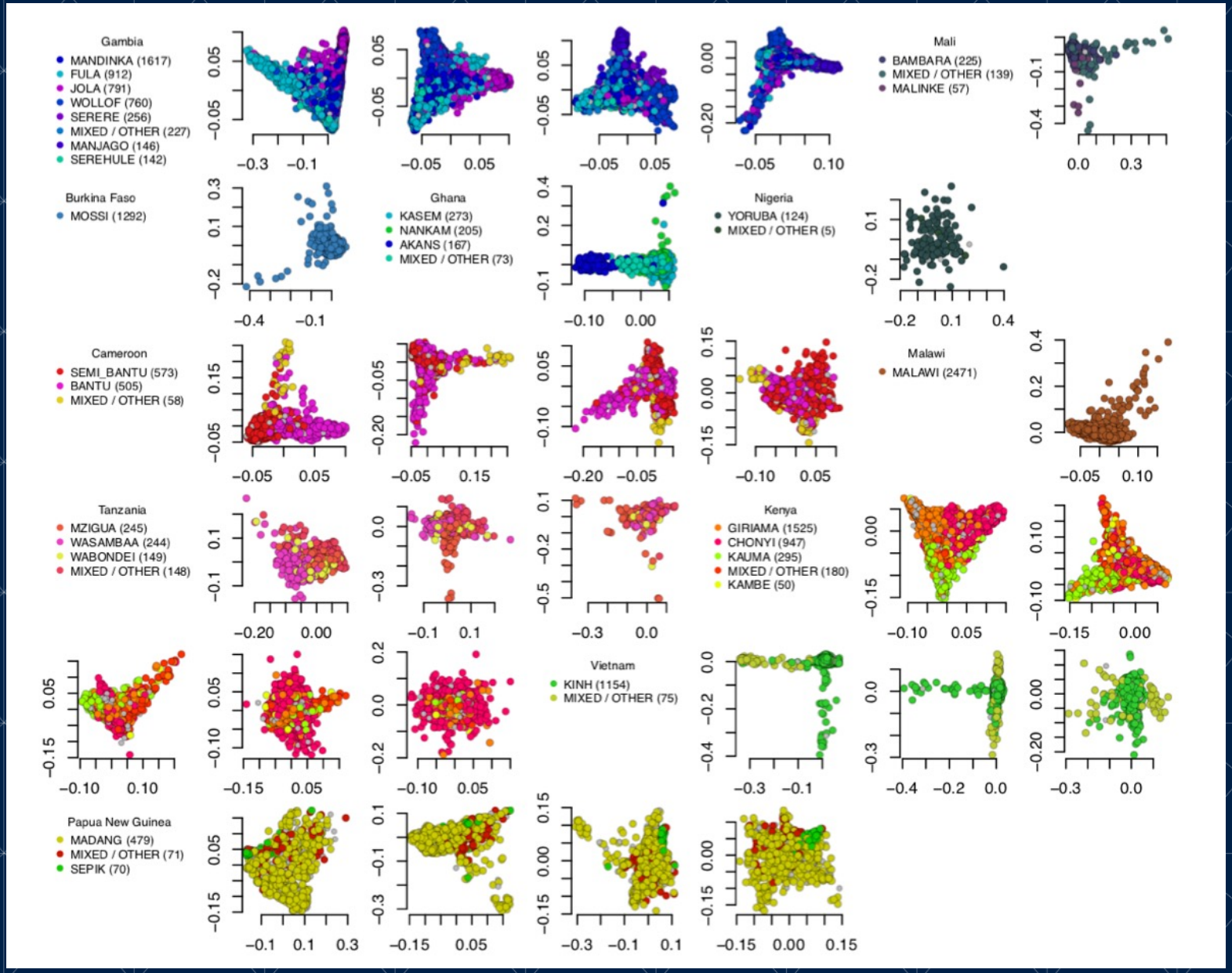


Study samples

Group	Cases	Controls	TOTAL
Africa			
Gambia	2567	2605	5172
Mali	274	183	457
Burkina Faso	733	596	1329
Ghana	399	320	719
Nigeria	113	22	135
Cameroon	592	685	1277
Malawi	1182	1317	2499
Tanzania	416	403	819
Kenya	1681	1615	3296
Asia			
Vietnam	718	546	1264
Oceania			
PNG	402	374	776

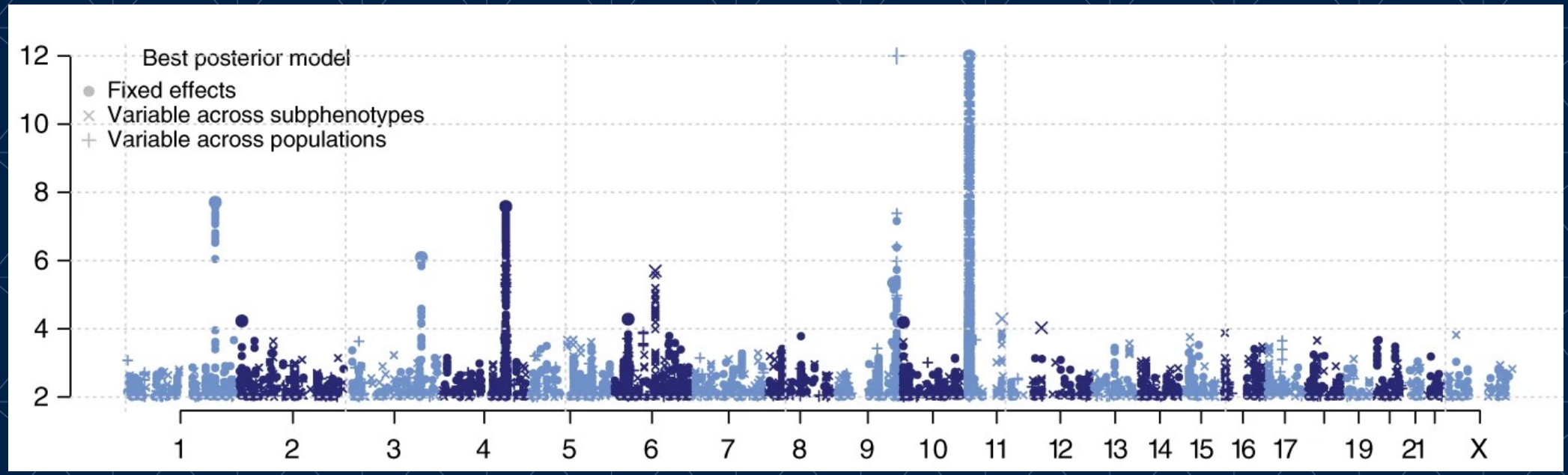
~400 population individuals to add to available reference genetic data

Over 17,000 severe cases and controls with human genome-wide data


Gambia
MANDINKA (1617)
FULA (912)
JOLA (791)
WOLLOF (760)
SERERE (256)
MIXED / OTHER (227)
MANJAGO (146)
SEREHULE (142)
Mali
BAMBARA (225)
MIXED / OTHER (139)
MALINKE (57)
Burkina Faso
MOSSI (1292)
Ghana
KASEM (273)
NANKAM (205)
AKANS (167)
MIXED / OTHER (73)
Nigeria
YORUBA (124)
MIXED / OTHER (5)
Cameroon
SEMI_BANTU (573)
BANTU (505)
MIXED / OTHER (58)
Malawi
MALAWI (2471)
Tanzania
MZIGUA (245)
WASAMBAA (244)
WABONDEI (149)
MIXED / OTHER (148)
Kenya
GIRIAMA (1525)
CHONYI (947)
KAUMA (295)
MIXED / OTHER (180)
KAMBE (50)
Vietnam
KINH (1154)
MIXED / OTHER (75)
Papua New Guinea
MADANG (479)
MIXED / OTHER (71)
SEPIK (70)

Only a handful of variants show strong associations

Evidence for association $(\log_{10} BF)$



Position in human genome

The human genome is very big, with millions of variants.
You need very strong evidence (say > 10^6) to get convinced in this type of study.

cf also Timmann et al Nature 2012

Genetic variants proposed to provide protection circa 2005

: *Beta-globin* HbS, HbC, HbE β - thalassaemia; *Alpha-globin* α - thalassaemia; Glycophorin A and B deletions; *G6PD* deficiency; "Band 3" ovalocytosis

Haptoglobin; O blood group; *PIEZO1*

Cytoadhesion: *CD36*; *CR1*; *ICAM1*; *PECAM1*

Immunity: *FCGR2A*; *CD40* ligand Immune activation; HLA antigen presentation; Interferon receptor components; Interleukins

MBL2; *NOS2A*; Tumer Necrosis Factor; Interferon Gamma

Genetic variants proposed to provide protection circa 2005

: *Beta-globin* HbS, HbC, HbE β - thalassaemia; *Alpha-globin* α - thalassaemia; Glycophorin A and B ~~deletions~~; *G6PD* deficiency; "Band 3" ovalocytosis

~~Haptoglobin~~; O blood group; ~~*PIEZO1*~~

Cytoadhesion: ~~*CD36*~~; ~~*CR1*~~; ~~*ICAM1*~~; ~~*PECAM1*~~

Immunity: ~~*FCGR2A*~~; ~~*CD40* ligand~~ Immune activation; ~~HLA~~ antigen presentation; ~~Interferon receptor~~ components; ~~Interleukins~~

~~*MBL2*~~; ~~*NOS2A*~~; Tumer ~~Necrosis Factor~~; ~~Interferon~~ Gamma

Genetic variants proposed to provide protection circa 2005

:

Beta-globin
HbS, HbC, HbE
β - thalassaemia

Alpha-globin
α - thalassaemia

Glycophorin A and B
~~deletions~~
+ duplications!

G6PD
deficiency

"Band 3"
ovalocytosis

~~Haptoglobin~~

O blood group

~~*PIEZO1*~~

ATP2B4
RBC calcium pump!

Cytoadhesion:

~~*CD36*~~

~~*CR1*~~

~~*ICAM1*~~

~~*PECAM1*~~

Immunity:

~~*FCGR2A*~~

~~*CD40* ligand~~
Immune activation

~~HLA~~
antigen presentation

~~Interferon receptor components~~

~~Interleukins~~

~~*MBL2*~~

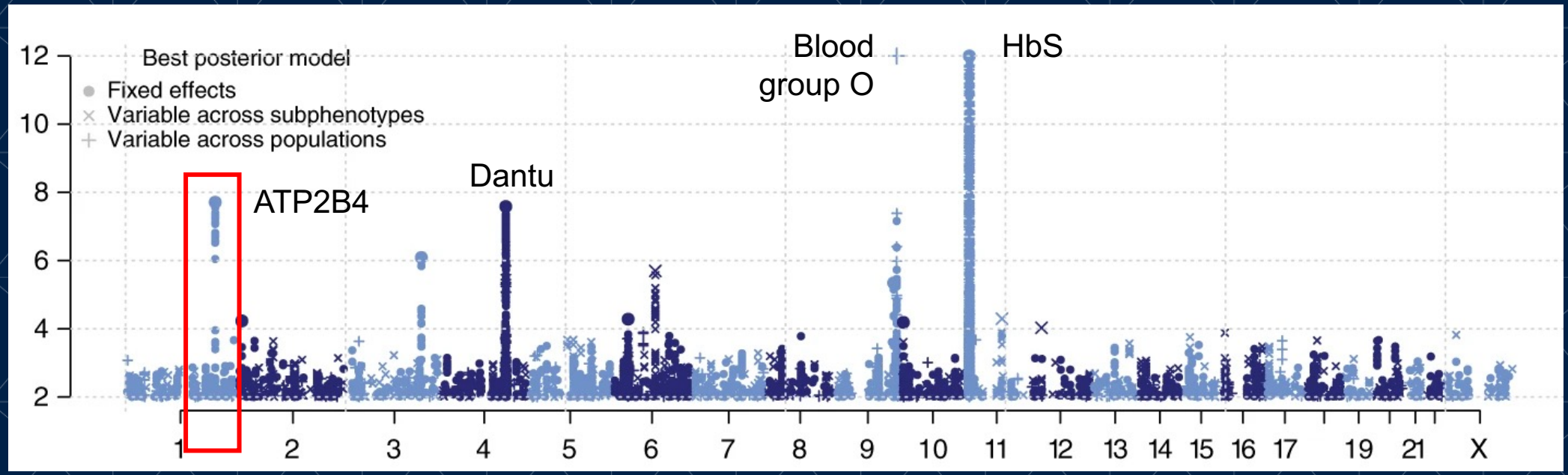
~~*NOS2A*~~

~~Tumer Necrosis Factor~~

~~Interferon Gamma~~

Only a handful of variants show strong associations

Evidence for association $(\log_{10} BF)$



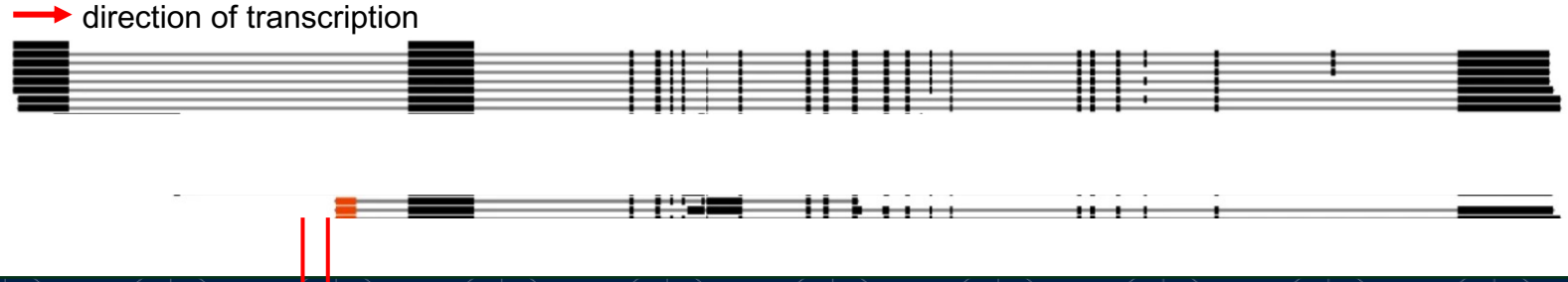
Position in human genome

The human genome is very big, with millions of variants.
You need very strong evidence (say > 10^6) to get convinced in this type of study.

Erythrocyte-specific calcium control at *ATP2B4*

ATP2B4 is expressed in all cells in this form

But in red cells in this form



The **risk allele** has a GATA site, and expressed the gene

…GAGCGATAAGAT…

…GAGCGGTAAGAT…

The **protective allele** disrupts the GATA site and does not express the gene

Erythrocyte-specific calcium control at *ATP2B4*

ATP2B4 is expressed in all cells in this form

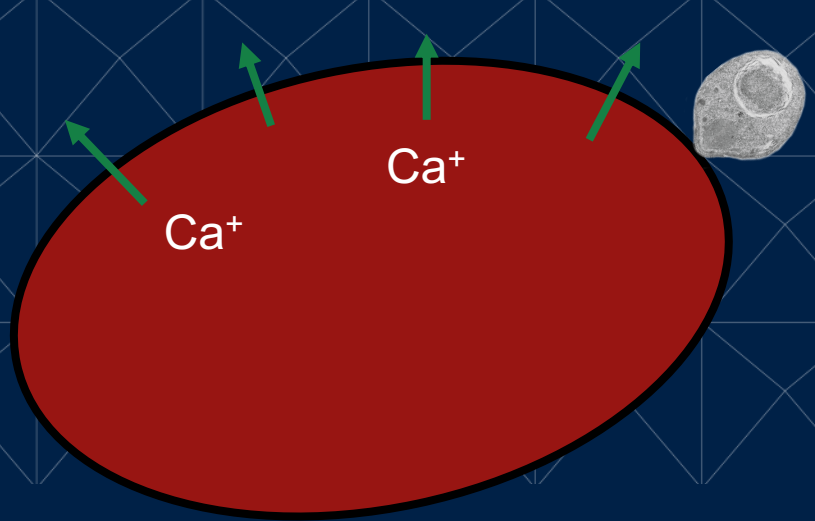
But in red cells in this form

The **risk allele** has a GATA site, and expressed the gene

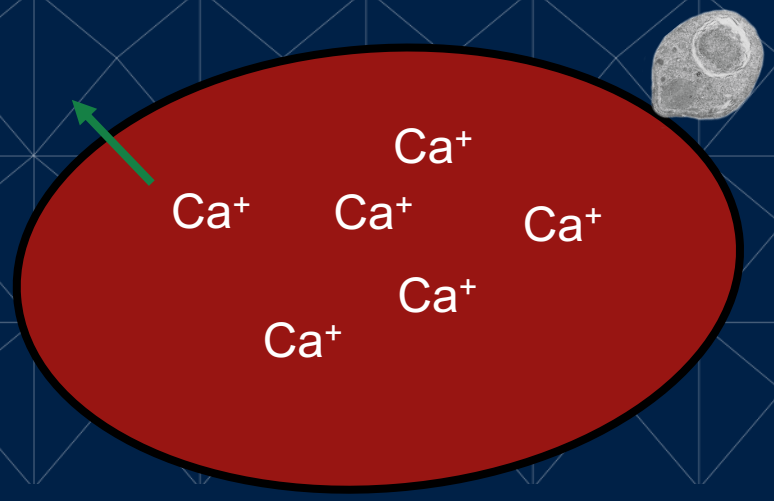
…GAGC**GATA**AGAT…

…GAGCGGTAAGAT…

The **protective allele** disrupts the GATA site and does not express the gene



ATP2B4 is a calcium pump.
It removes Ca^+ from the cell.



Erythrocyte-specific calcium control at *ATP2B4*

ATP2B4 is expressed in all cells in this form

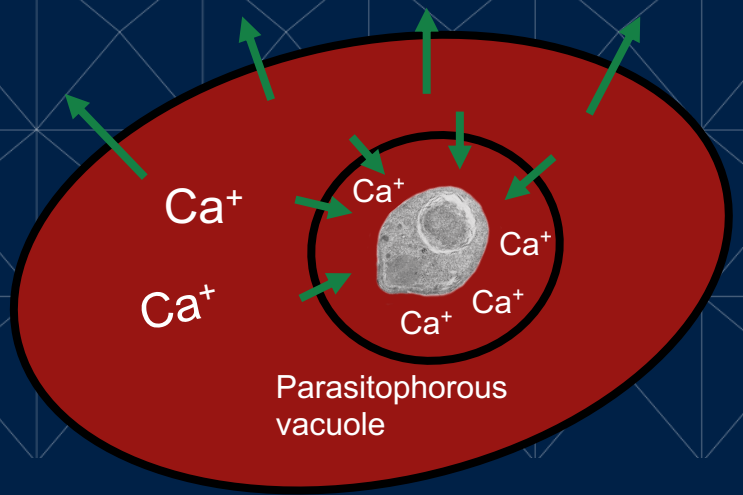
But in red cells in this form

The **risk allele** has a GATA site, and expressed the gene

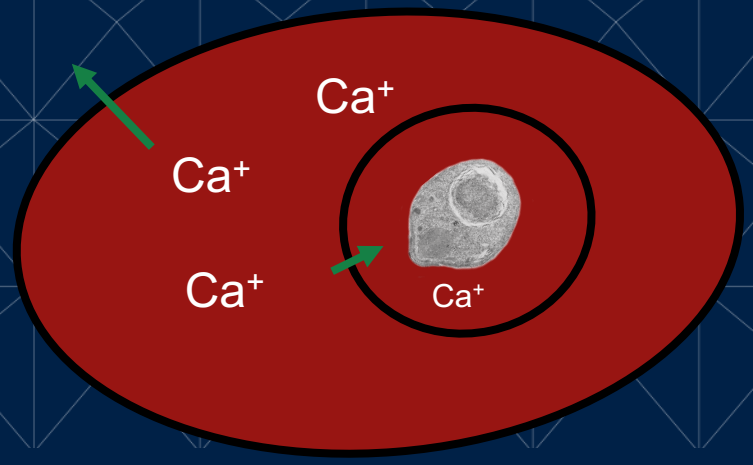
…GAGC**GATA**AGAT…

…GAGCGGTAAGAT…

The **protective allele** disrupts the GATA site and does not express the gene



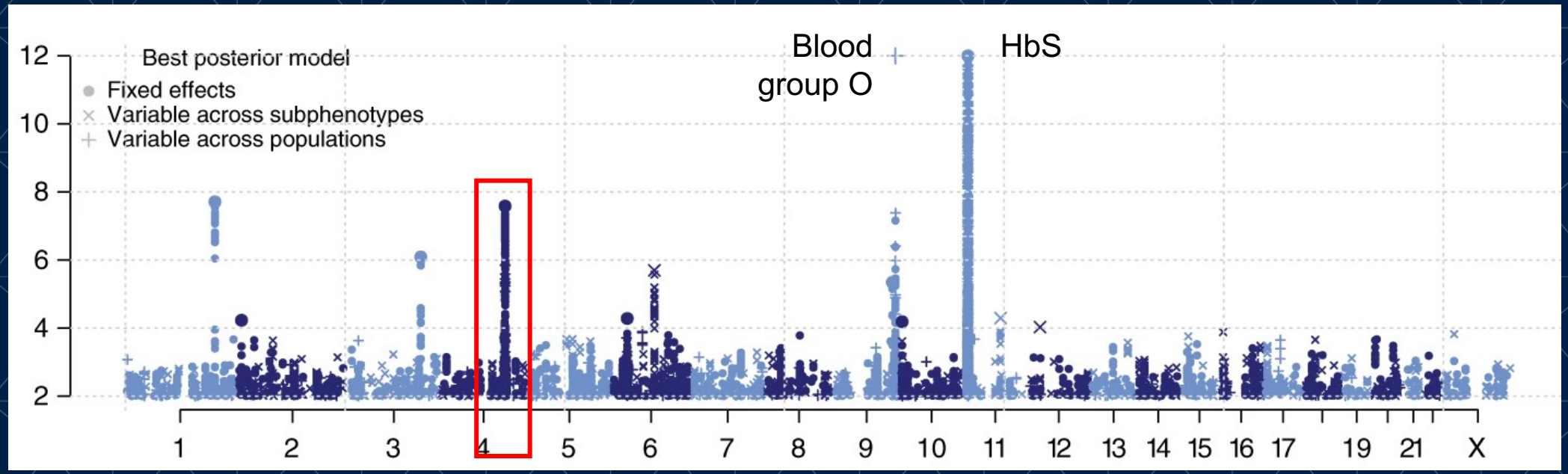
ATP2B4 is a calcium pump.
It removes Ca^+ from the cell.



Lessard et al 2017, Zambo et al 2017

Only a handful of variants show strong associations

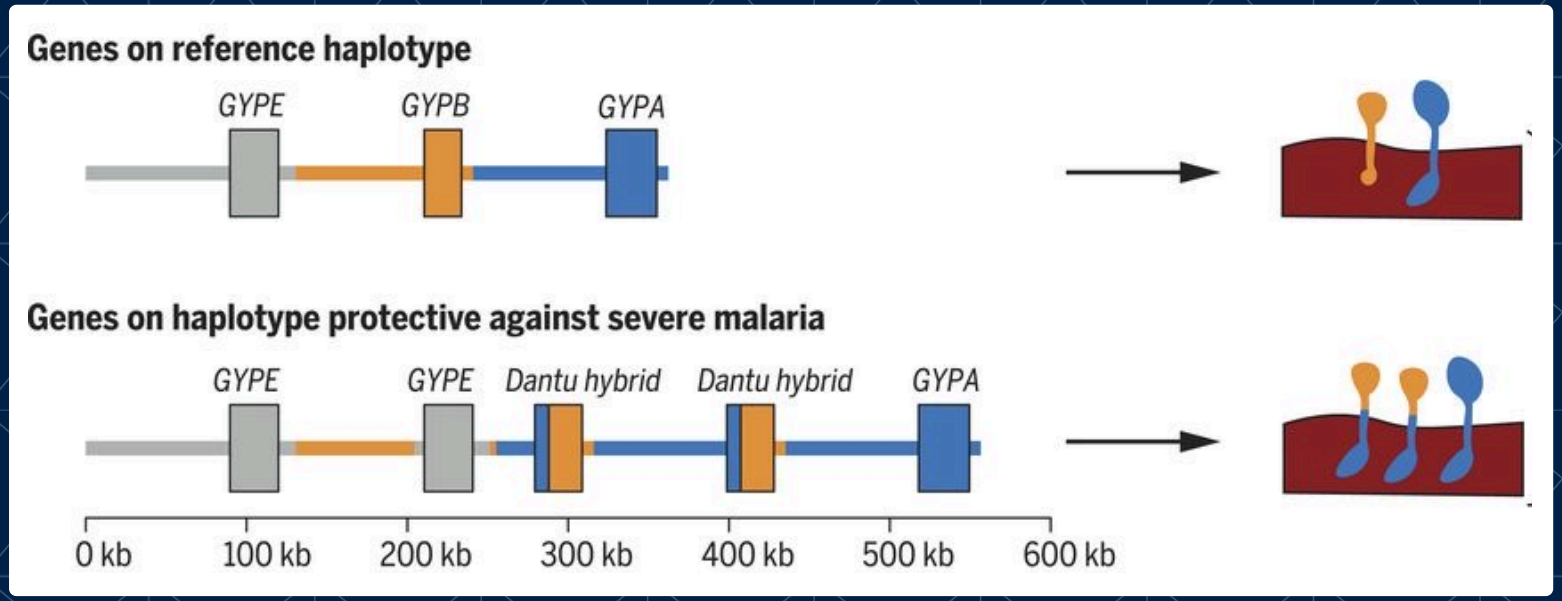
Evidence for association $(\log_{10} BF)$



Position in human genome

The human genome is very big, with millions of variants.
You need very strong evidence (say > 10^6) to get convinced in this type of study.

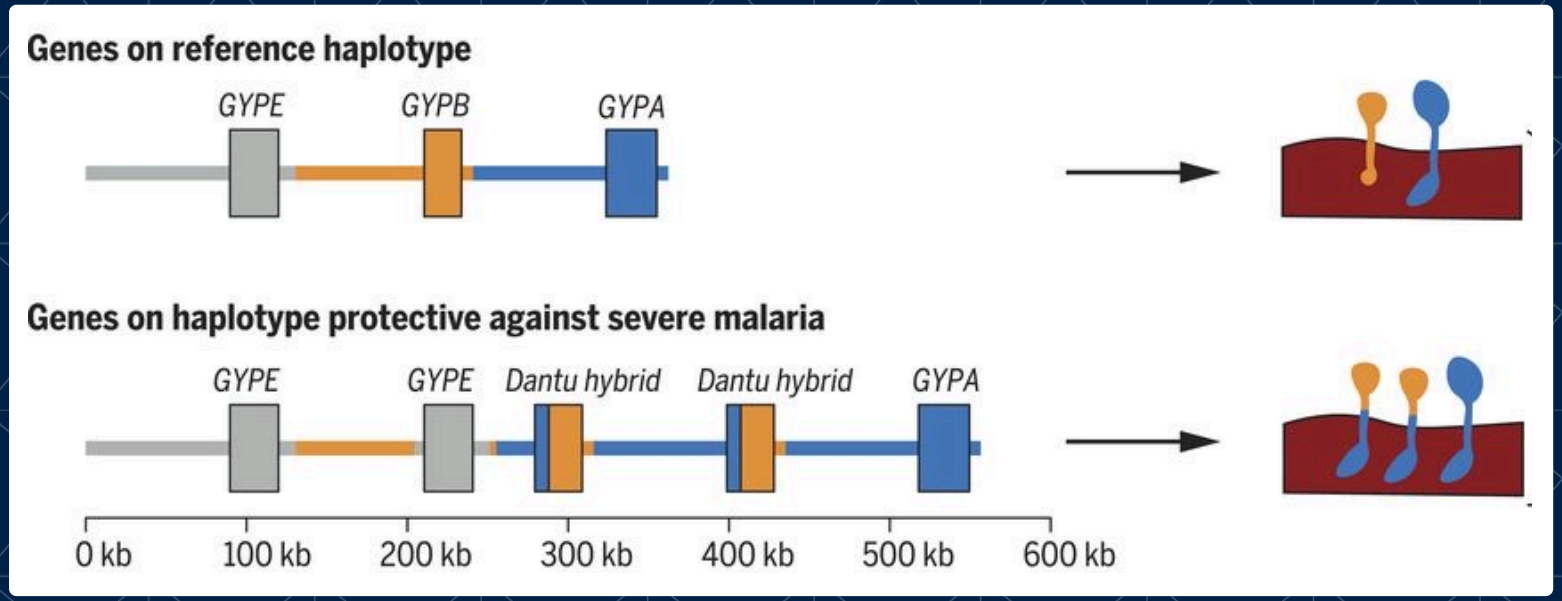
Structural variation of glycophorins is associated with protection



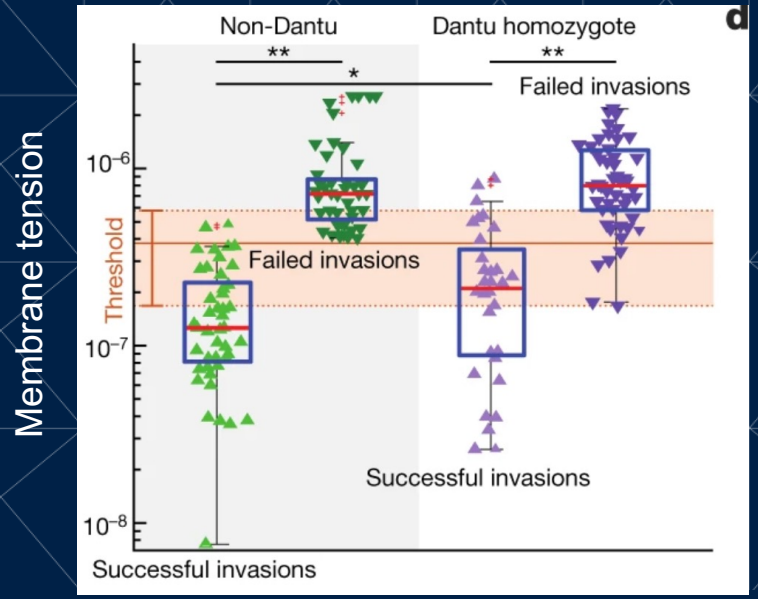
"Dantu NE" blood group variant, only found in parts of east Africa

E. Leffler et al Science 2017

Structural variation of glycophorins is associated with protection



"Dantu NE" blood group variant, only found in parts of east Africa



E. Leffler et al Science 2017

S.Kariuki et al Nature 2020

Sickle haemoglobin

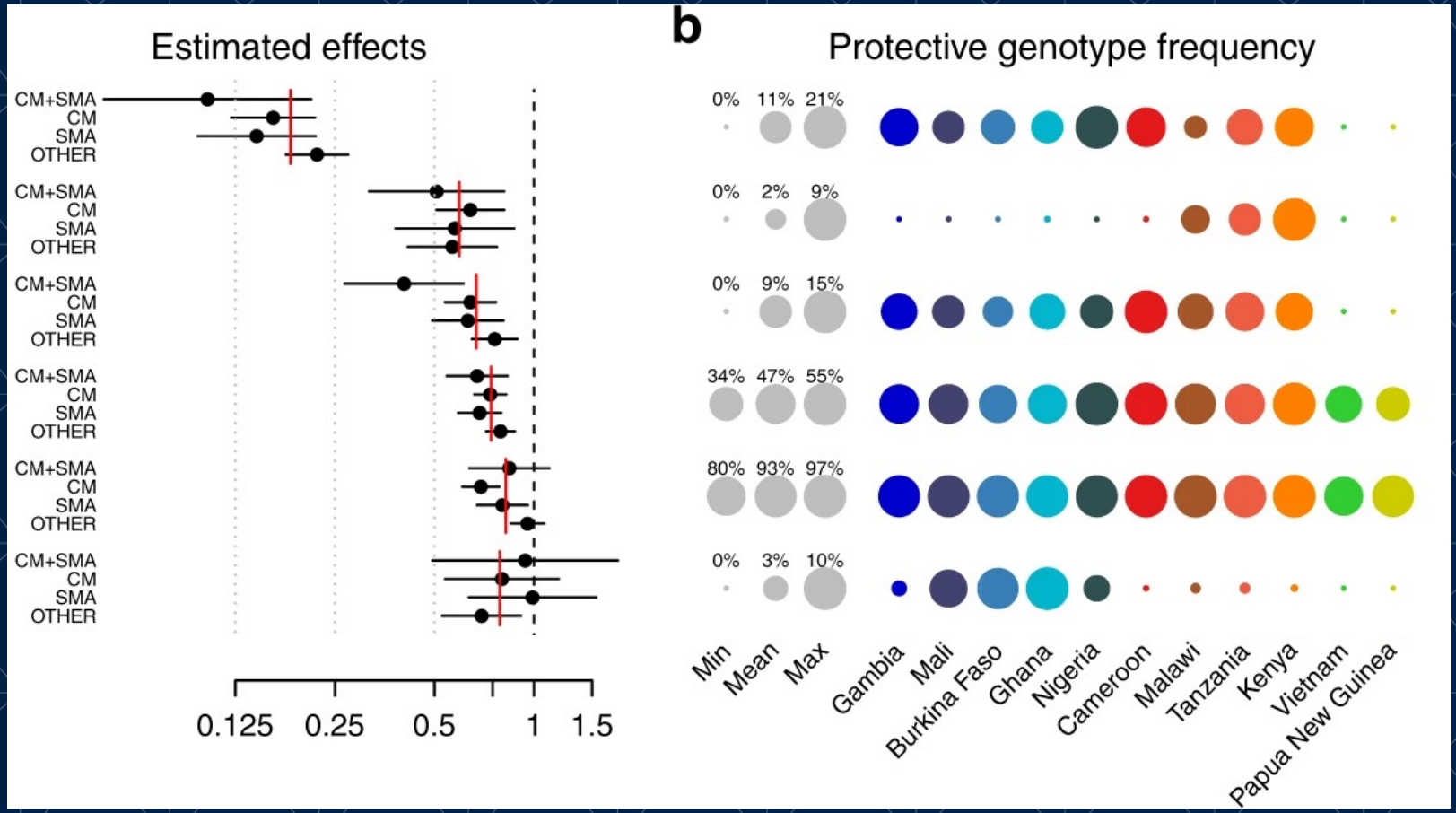
Glycophorin A/B duplications!

ATP2B4 (RBC calcium channel)

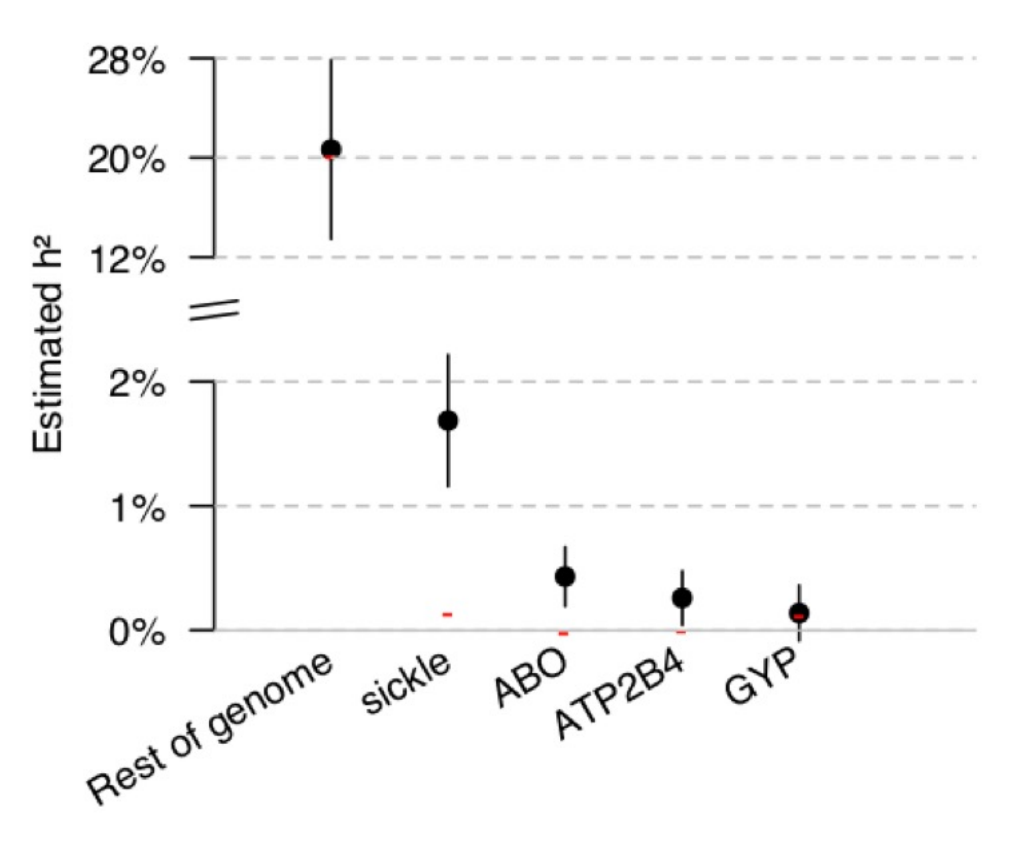
Blood group O

(Intragenic SNP)

HbC



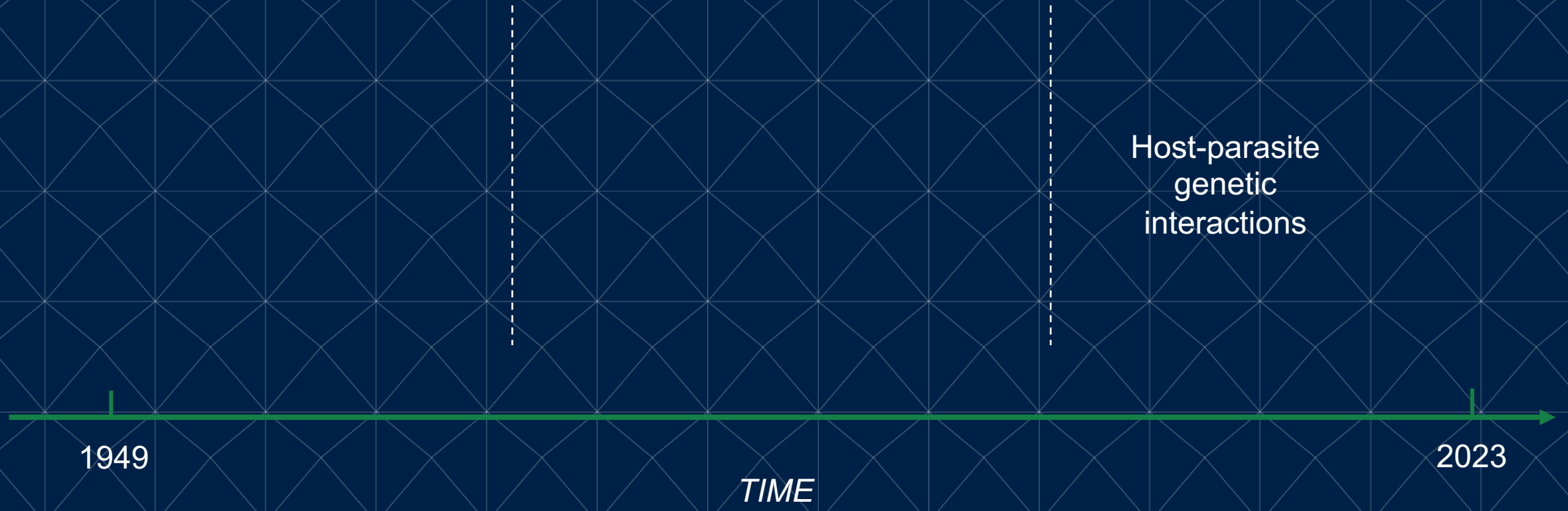
West ⟶ East



A substantial proportion of the heritability **remains unexplained**

That is – there is probably more to find, if only we had large enough samples.

Talk outline



O blood type

frequency ~ 50%

RR ~ 0.75 (recessive)

Dantu blood type

frequency ~ 0-10%

RR ~ 0.6

(additive)

ATP2B4 calcium pump variation

frequency ~ 50%

RR ~ 0.66

(recessive)

Sickle haemoglobin (HbS)

Frequency ~ 2-20%

RR ~ 0.1-0.2

(heterozygote)

Allison Br Med. J. (1954)

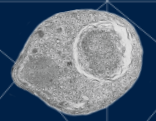
O blood type

Dantu blood type

ATP2B4 calcium pump variation

Sickle haemoglobin (HbS)

?



P.falciparum
23Mb genome
>5,000 genes
millions of variants

Have parasite populations adapted?

(And is this detectable in current populations?)

Human – *P.falciparum* genetic association study

"*Malaria protection due to sickle haemoglobin depends on parasite genotype*", 2021

https://doi.org/10.1038/s41586-021-04288-3

Plan of our analysis:

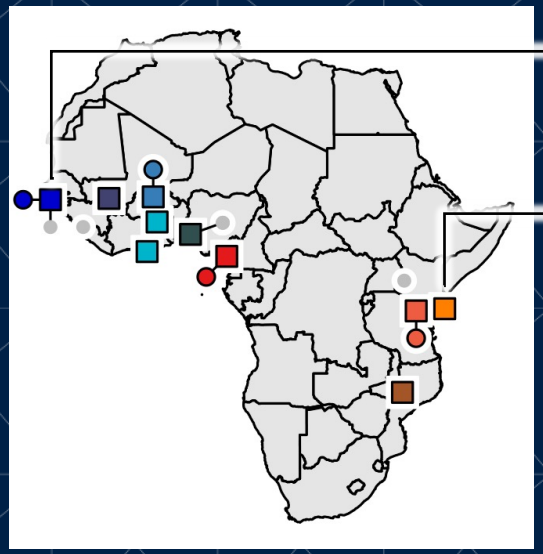
1. Sequence the *P.falciparum* genome in severe malaria cases
   selected from our previously published human GWAS*

2. Test for association between human and parasite genetic variants

*MalariaGEN human GWAS, Nature Communications 2019

Investigating human-parasite genetic interaction in severe malaria cases

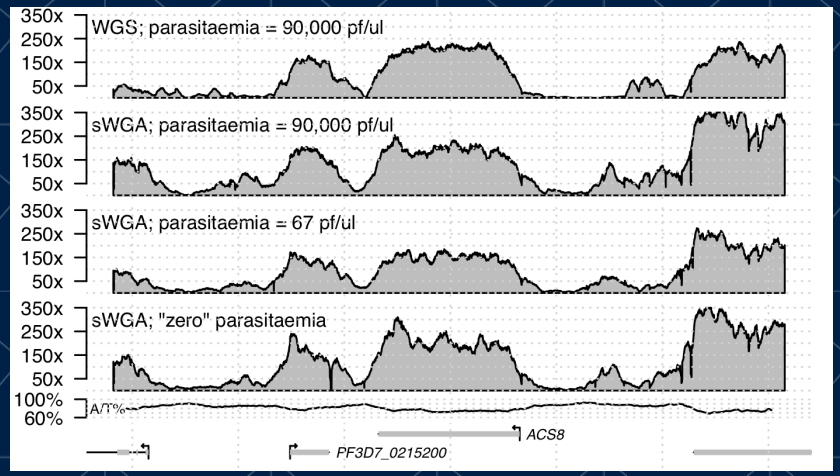
1. Sequence the *P.falciparum* genome in severe malaria cases from the published human GWAS



Banjul, The Gambia
N = 2,721

Kilifi, Kenya
N = 2,375

Collected in 1995-2009



Variant calling and quality control

Overlap with human data
N = 3,346 samples

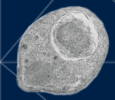
Previously generated human genome-wide genotypes and imputation

WELLCOME CENTRE *for* HUMAN GENETICS

2. Test for association pairwise between human and *Pf* variants using a simple logistic regression framework:

$$g_{Pf} \sim g_{\text{human}} + \text{country}$$

Software at: www.well.ox.ac.uk/~gav/hptest



Focus on candidates:

- Known protective mutations
- Further putative associations
- Blood group gene variants
- HLA alleles

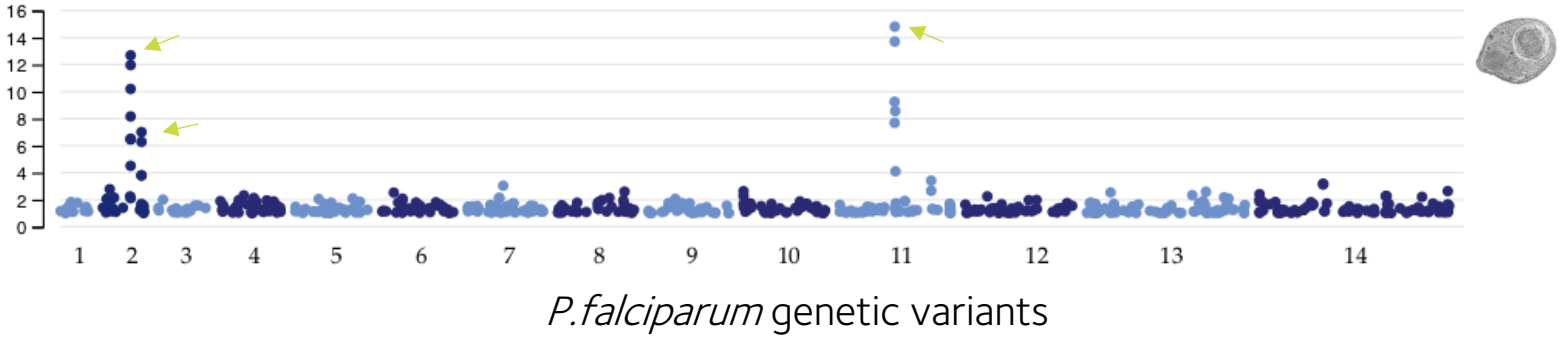
Focus on 'easy' parts:

- Biallelic variants in core genome
- Seen in at least 25 infections across the sample.
- 51,552 variants in total

(...excludes multiallelics and complex regions)

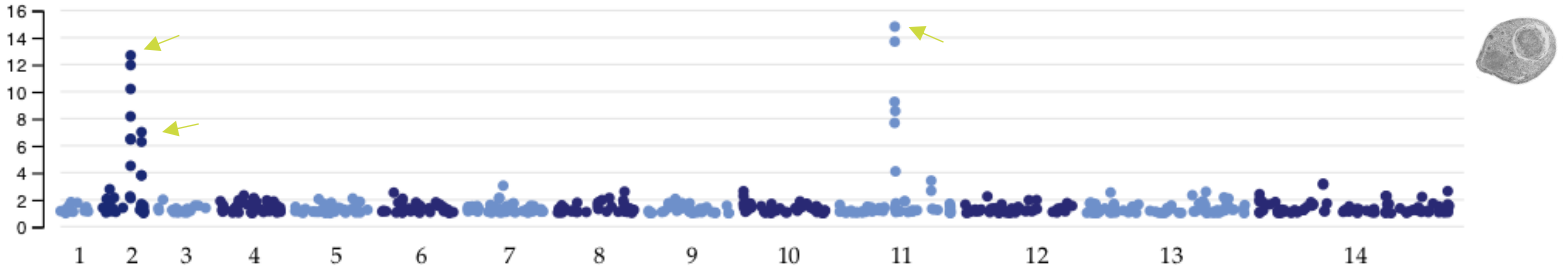
Three regions of the Pf genome are associated…

Evidence for association
for *P.falciparum* variants
(averaged over human variants)

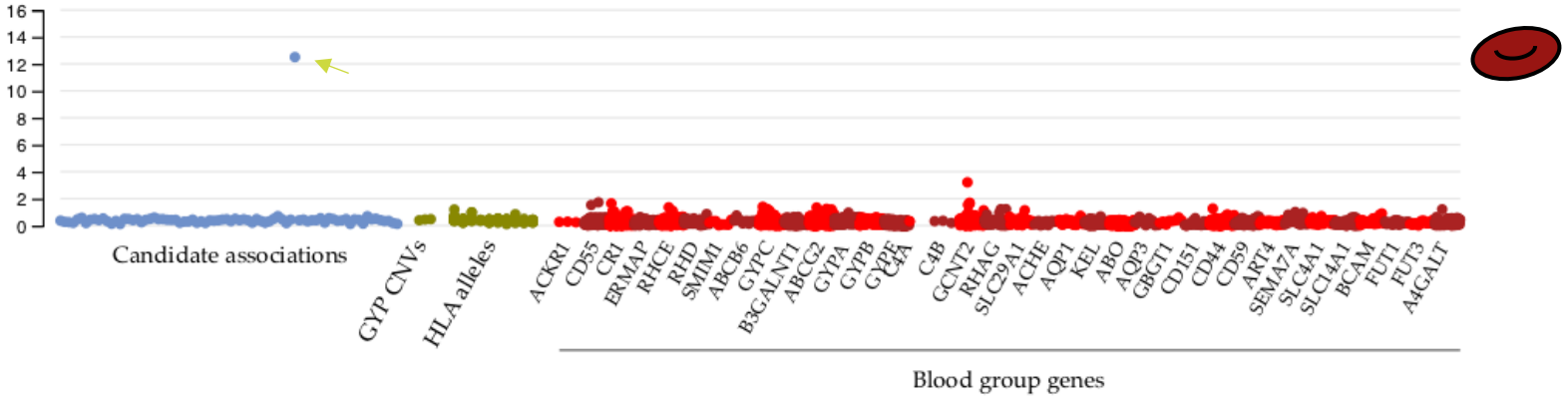


Three regions of the Pf genome are associated with…

Evidence for association
for *P.falciparum* variants
(averaged over human variants)



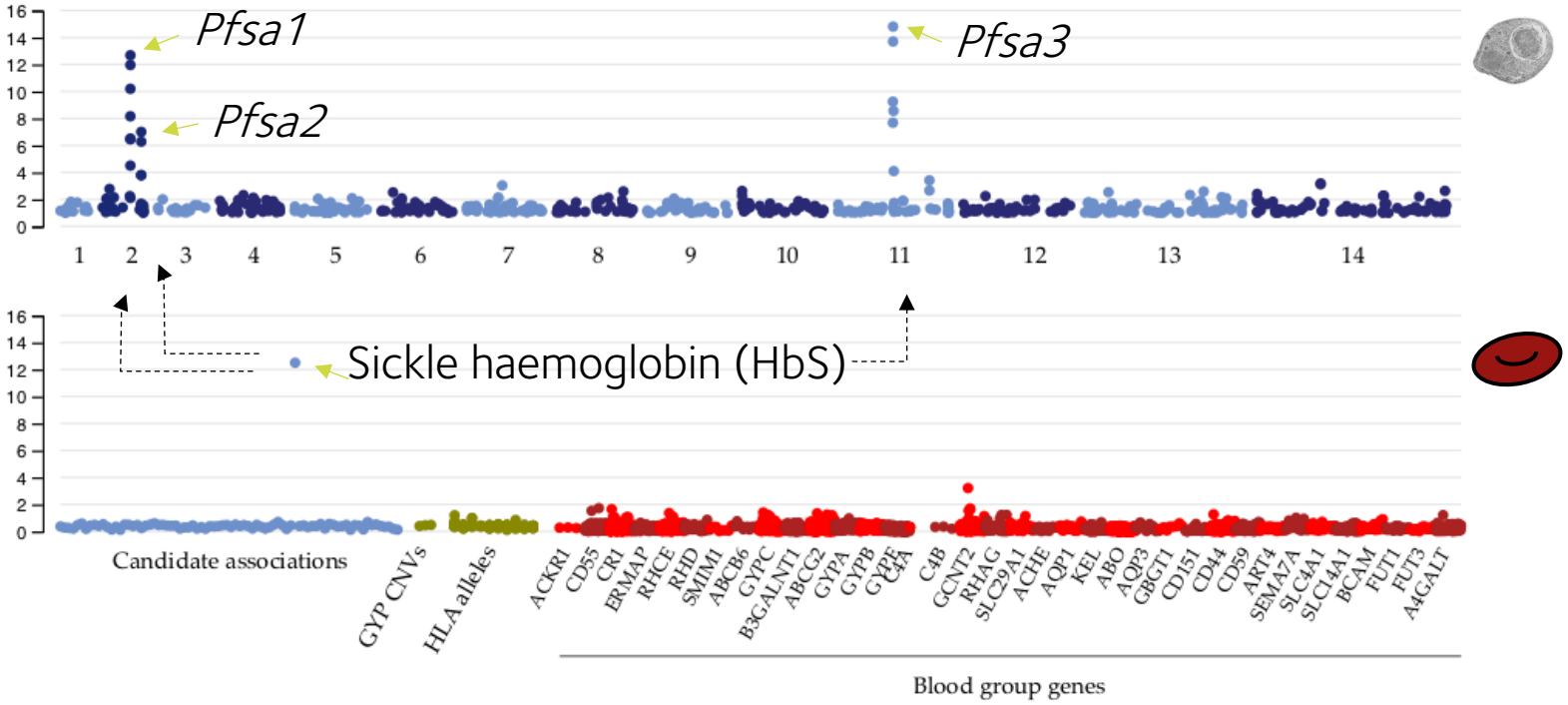
Evidence for association
for human variants
(averaged over *Pf* variants)



Three regions of the Pf genome are associated with HbS

Evidence for association
for *P.falciparum* variants
(averaged over human variants)

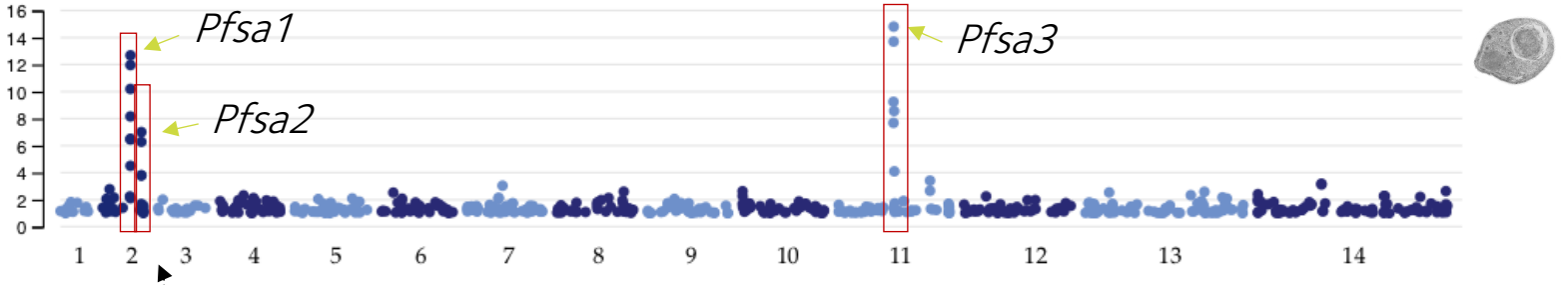
Evidence for association
for human variants
(averaged over *Pf* variants)



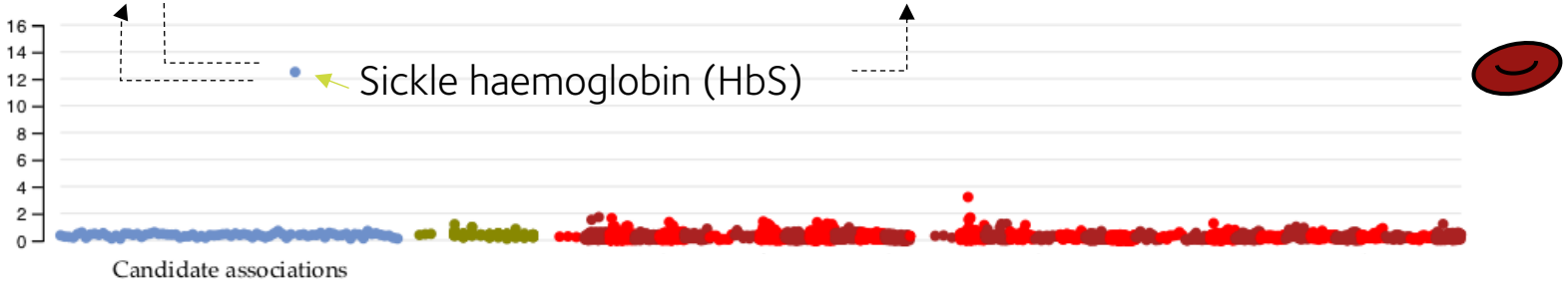
Pfsa = "Plasmodium falciparum sickle-associated"

Three regions of the Pf genome are associated with HbS

Evidence for association
for *P.falciparum* variants
(averaged over human variants)

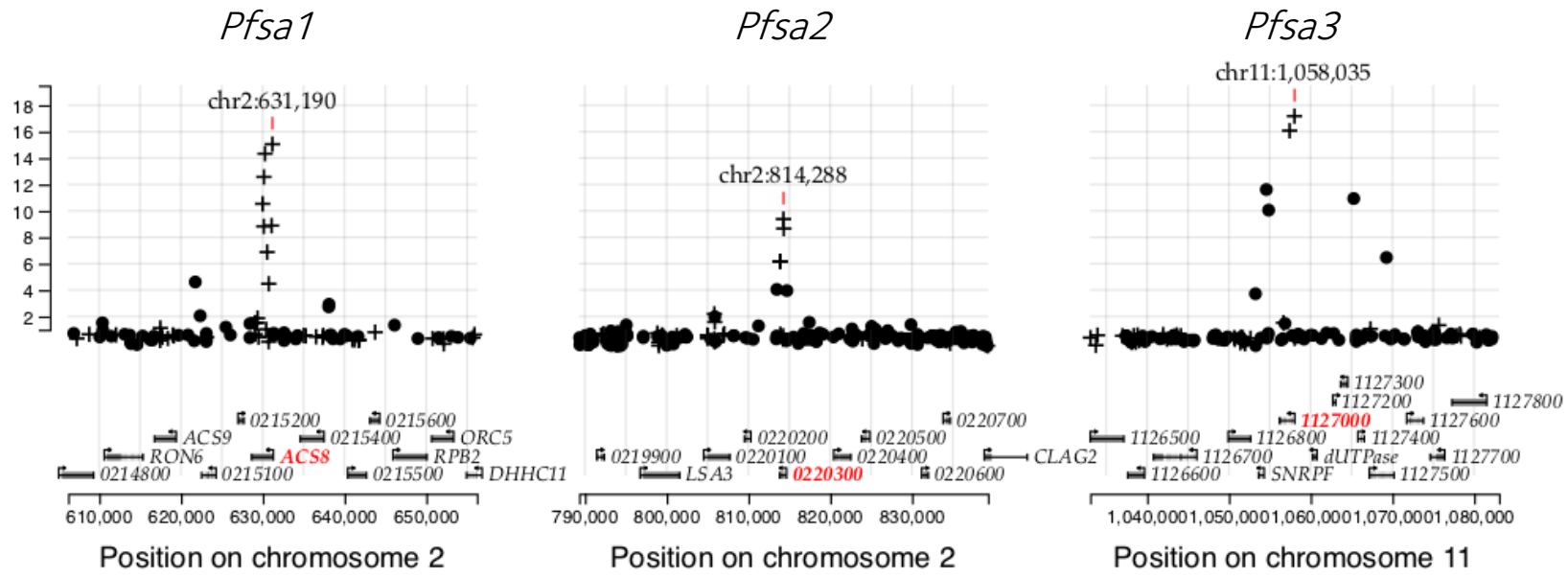


Evidence for association
for human variants
(averaged over *Pf* variants)



Zoom in to
Pf genome:

Evidence for association
for *P.falciparum* variants
with HbS

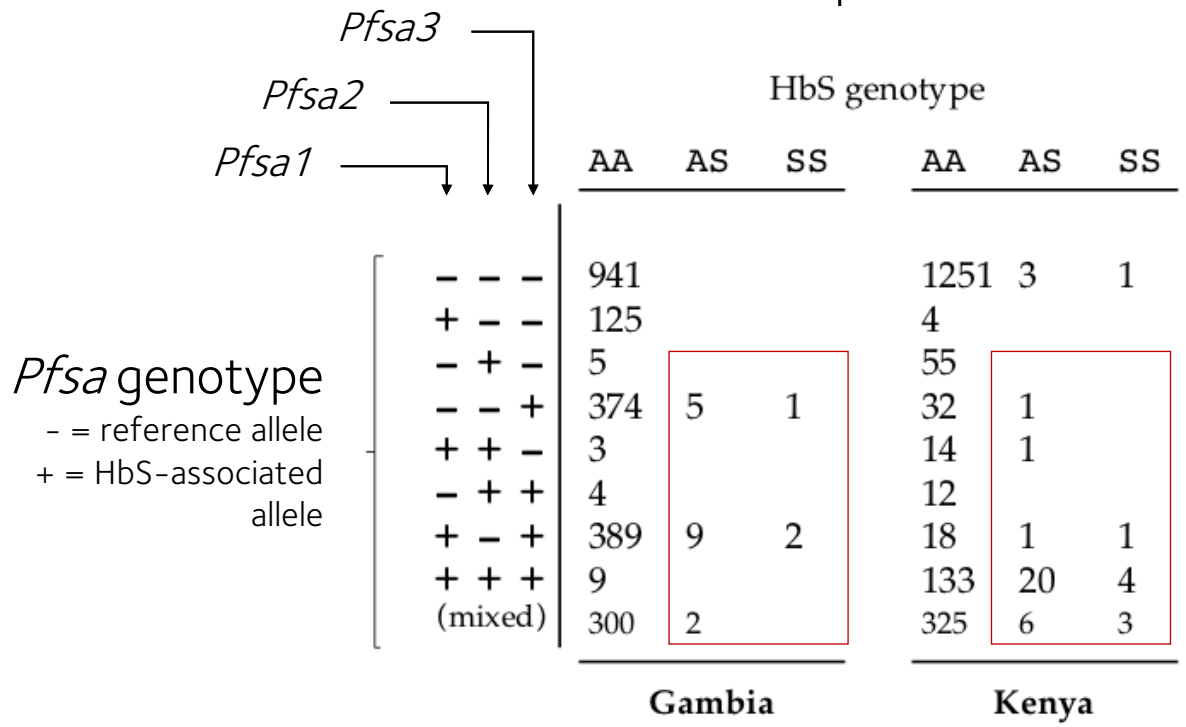


WELLCOM

The protective effect of HbS varies with *Pfsa* genotype

$N = 4{,}071$ severe malaria cases

Sample counts



Pfsa genotype	Gambia AA	Gambia AS	Gambia SS	Kenya AA	Kenya AS	Kenya SS
– – –	941			1251	3	1
+ – –	125			4		
– + –	5			55		
– – +	374	5	1	32	1	
+ + –	3			14	1	
– + +	4			12		
+ – +	389	9	2	18	1	1
+ + +	9			133	20	4
(mixed)	300	2		325	6	3

45 of 49 severe infections of individuals with HbS genotypes were with *Pfsa+* parasites

The protective effect of HbS varies with *Pfsa* genotype

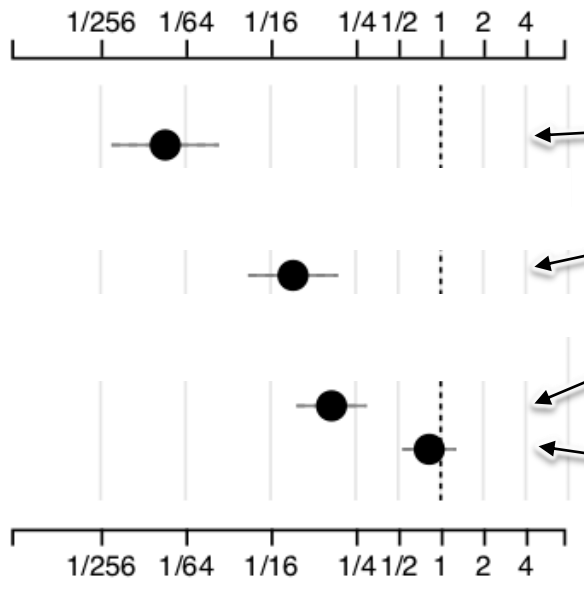
Sample counts

Relative risk of HbS by Pfsa genotype

Pfsa3
Pfsa2
Pfsa1
Population controls

Pfsa genotype
- = reference allele
+ = HbS-associated allele

Pfsa genotype	HbS genotype: Gambia AA	Gambia AS	Gambia SS	Kenya AA	Kenya AS	Kenya SS
	2284	396	15	3365	600	30
– – –	941			1251	3	1
+ – –	125			4		
– + –	5			55		
– – +	374	5	1	32	1	
+ + –	3			14	1	
– + +	4			12		
+ – +	389	9	2	18	1	1
+ + +	9			133	20	4
(mixed)	300	2		325	6	3



$RR = 0.012$

$RR = 0.09$

$RR = 0.17$

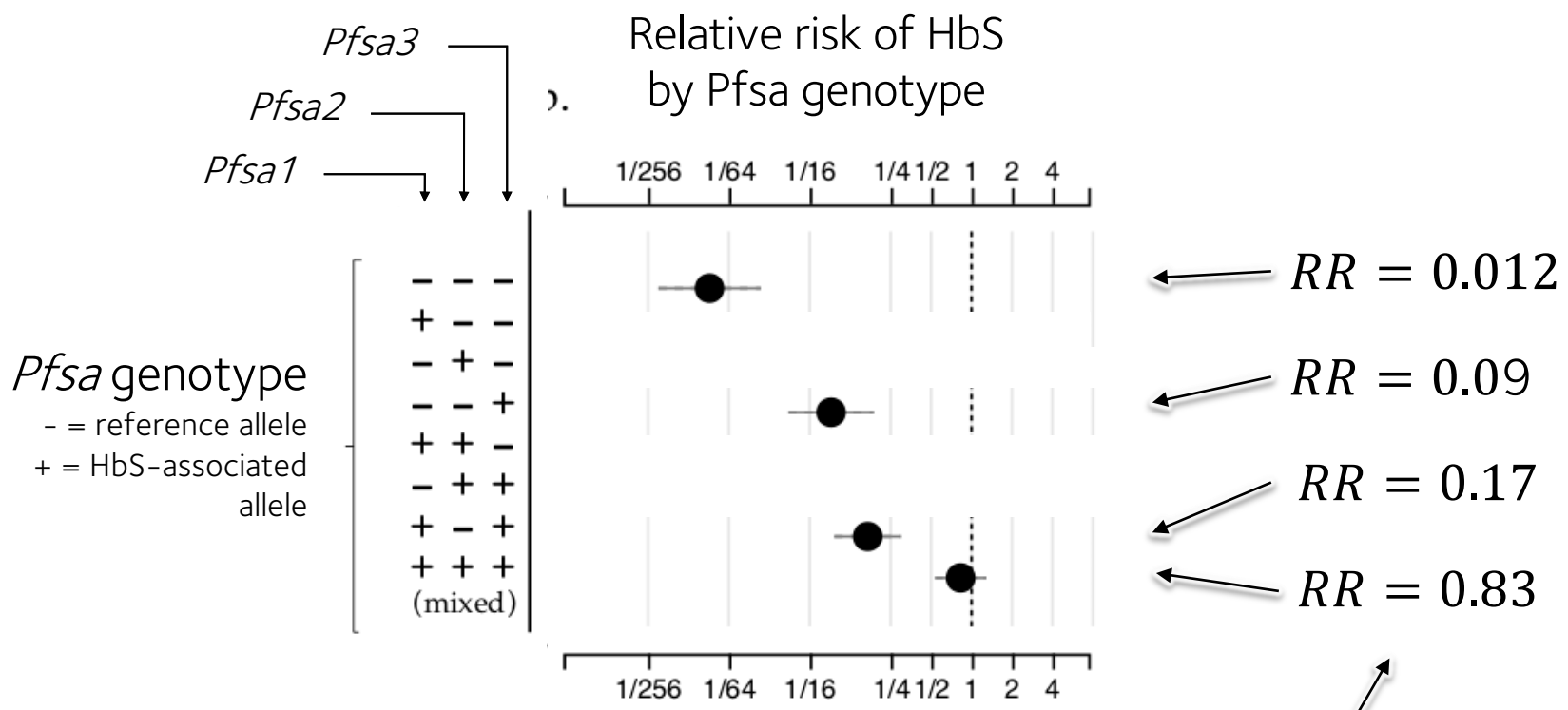
$RR = 0.83$

Pfsa+ parasites may have overcome HbS protection

https://doi.org/10.1038/s41586-021-04288-3

+ caveats!

The protective effect of HbS varies with *Pfsa* genotype



Pfsa+ parasites may be able to overcome HbS protection

https://doi.org/10.1038/s41586-021-04288-3

+ caveats!

Pfsa frequencies vary widely within and between populations

Wellcome Open Research

Wellcome Open Research 2021, 6:42 Last updated: 29 MAR 2021

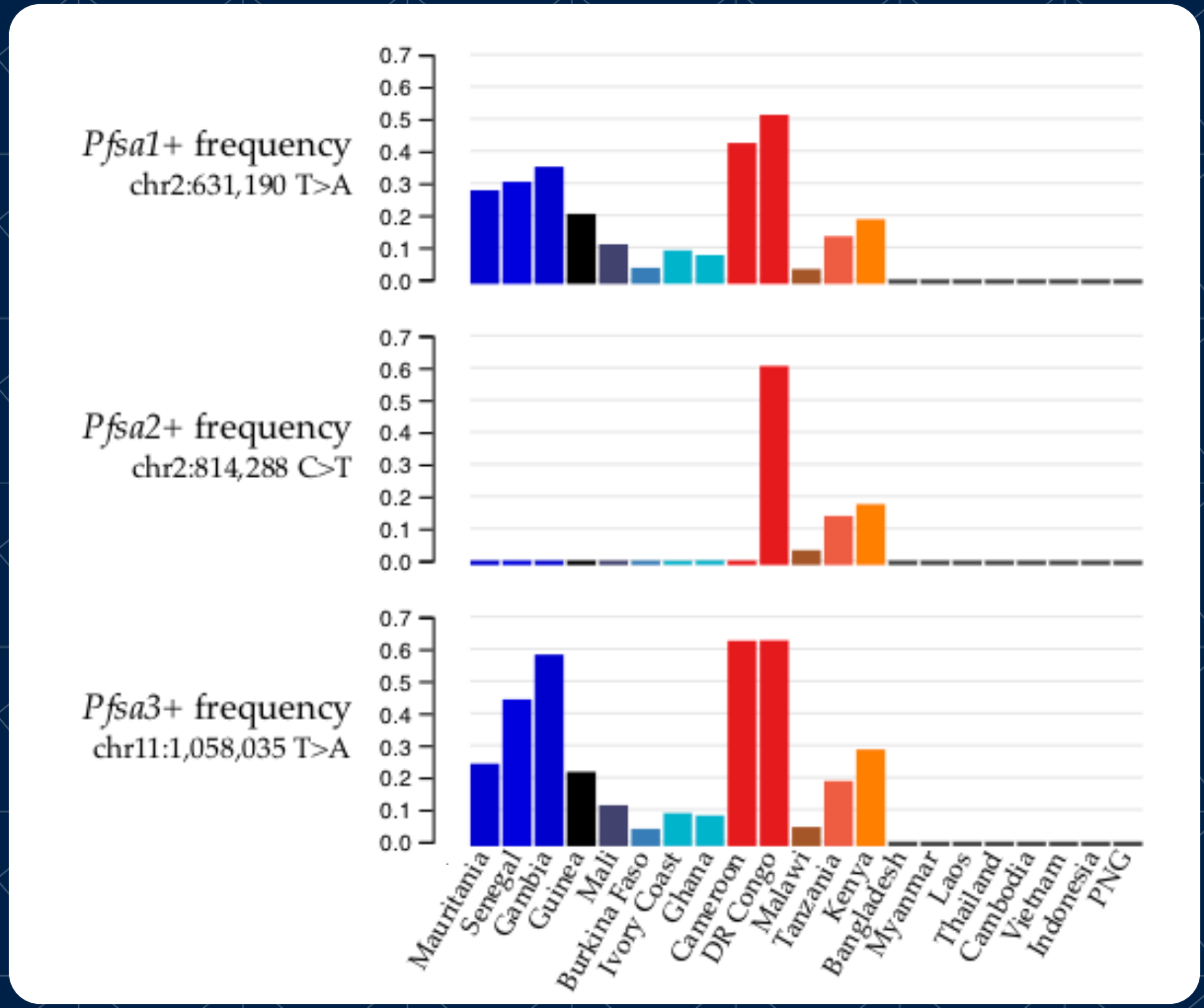
Check for updates

RESEARCH ARTICLE

An open dataset of *Plasmodium falciparum* genome variation in 7,000 worldwide samples

MalariaGEN Pf6

Wellcome Open Research 2021



Pfsa frequencies vary widely within and between populations

Wellcome Open Research

Wellcome Open Research 2021, 6:42 Last updated: 29 MAR 2021

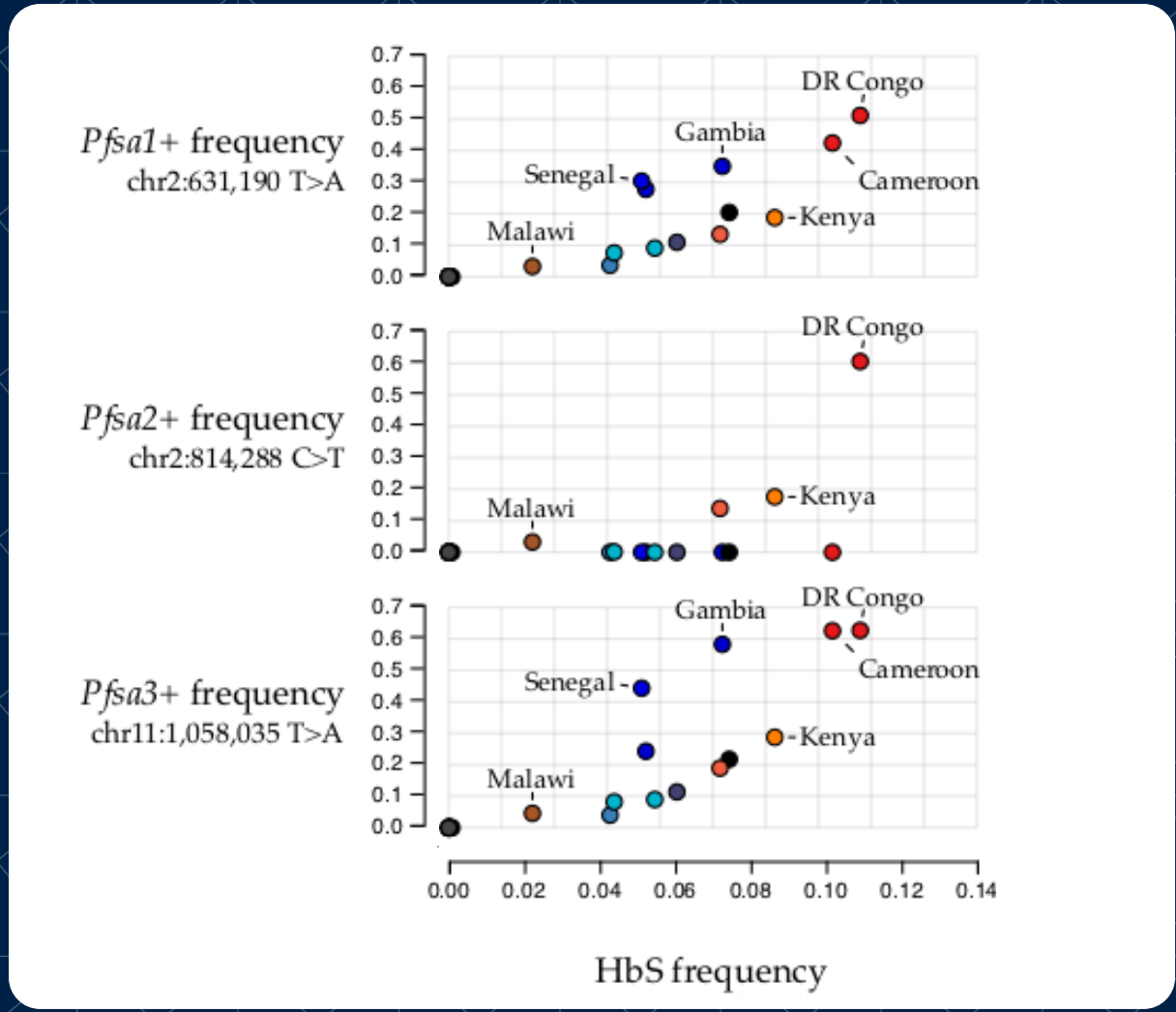
Check for updates

RESEARCH ARTICLE

An open dataset of *Plasmodium falciparum* genome variation in 7,000 worldwide samples

MalariaGEN Pf6

Wellcome Open Research 2021



(Malaria Atlas Project - Piel et al Lancet 2013)

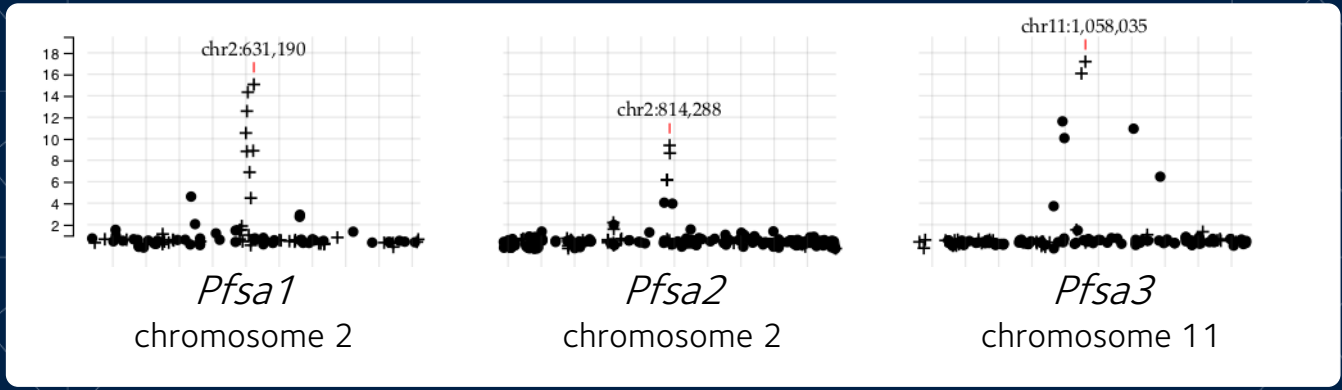
The *Pfsa* alleles are in strong linkage disequilibrium i.e. they co-occur

r=0.80

r=0.75

r=0.66

Kenya:



Pfsa1 chromosome 2

Pfsa2 chromosome 2

Pfsa3 chromosome 11

Gambia:

r=0.43

Correlation between *Pfsa+* alleles in severe malaria cases...

After excluding HbS individuals

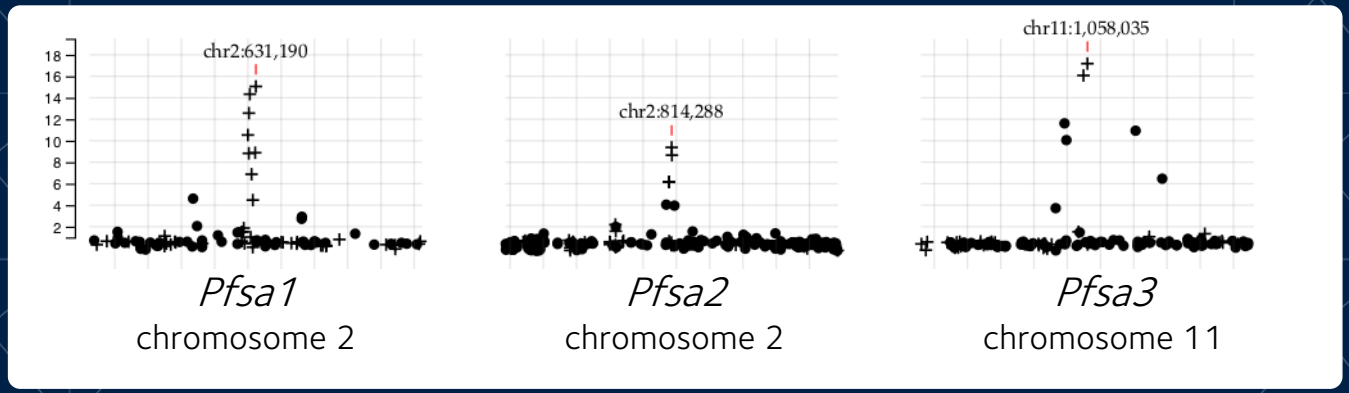
The *Pfsa* alleles are in strong linkage disequilibrium i.e. they co-occur

r=0.80

Kenya:

r=0.75

r=0.66



Pfsa1 chromosome 2 | *Pfsa2* chromosome 2 | *Pfsa3* chromosome 11

Gambia:

r=0.43

Correlation between *Pfsa+* alleles in severe malaria cases...

After excluding HbS individuals

...or in milder infections:

Pfsa1+ vs *Pfsa3+*

Country	N	r
Gambia	169	0.20
Guinea	133	0.79
Mali	379	0.84
Ghana	807	0.86
Cameroon	174	0.52
Congo	241	0.64
Malawi	239	0.79
Tanzania	282	0.59
Kenya	89	0.71

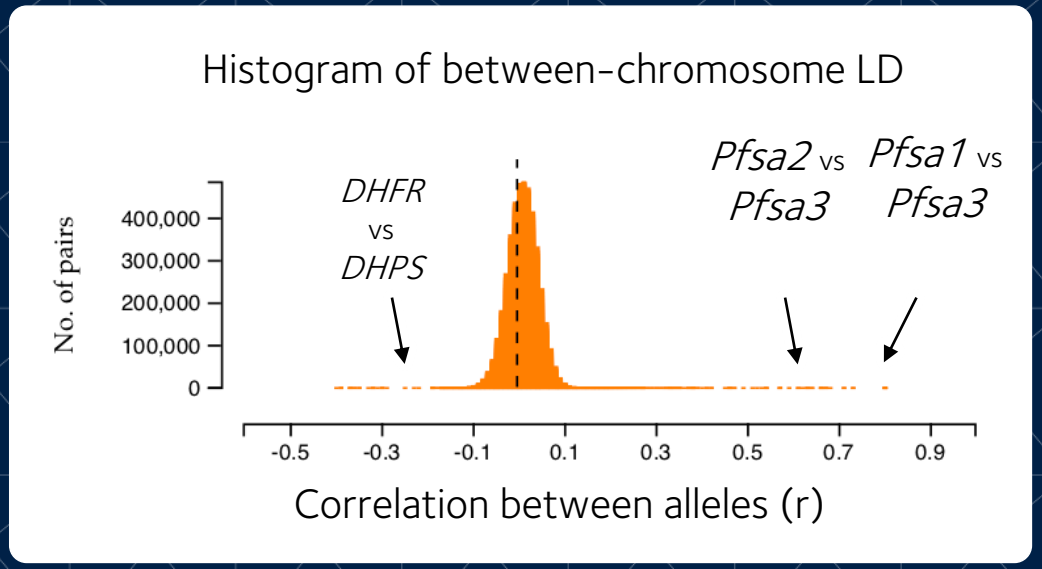
MalariaGEN Pf6

The *Pfsa* alleles are in strong linkage disequilibrium i.e. they co-occur

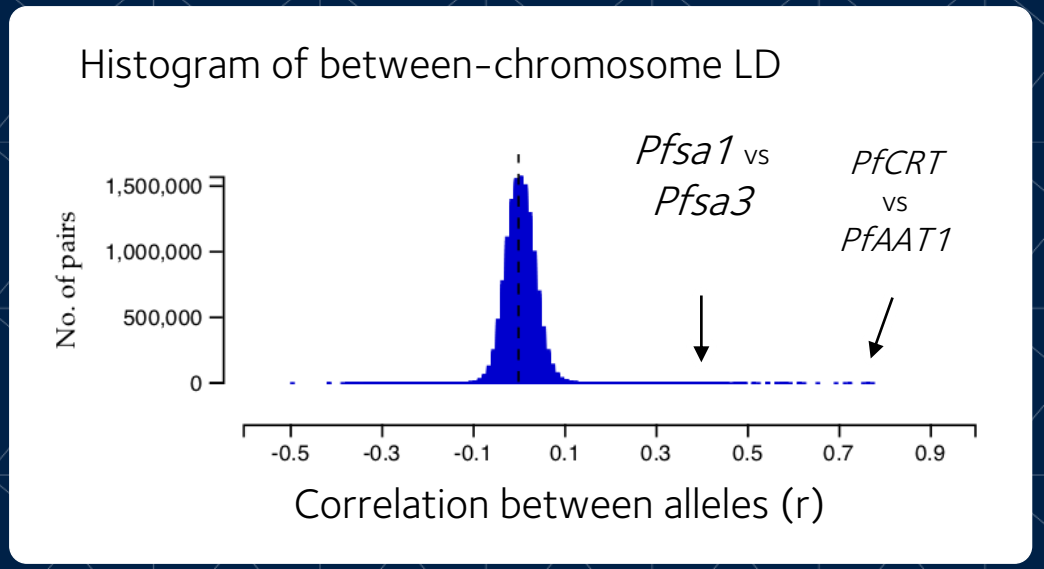
Parasites undergo sexual reproduction (meiosis) in mosquitos

This breaks down LD.

Kenya:



Gambia:



?

”Reverse malaria hypothesis”

- *Pfsa*-carrying parasites are able to infect and cause disease in HbS-carrying individuals…

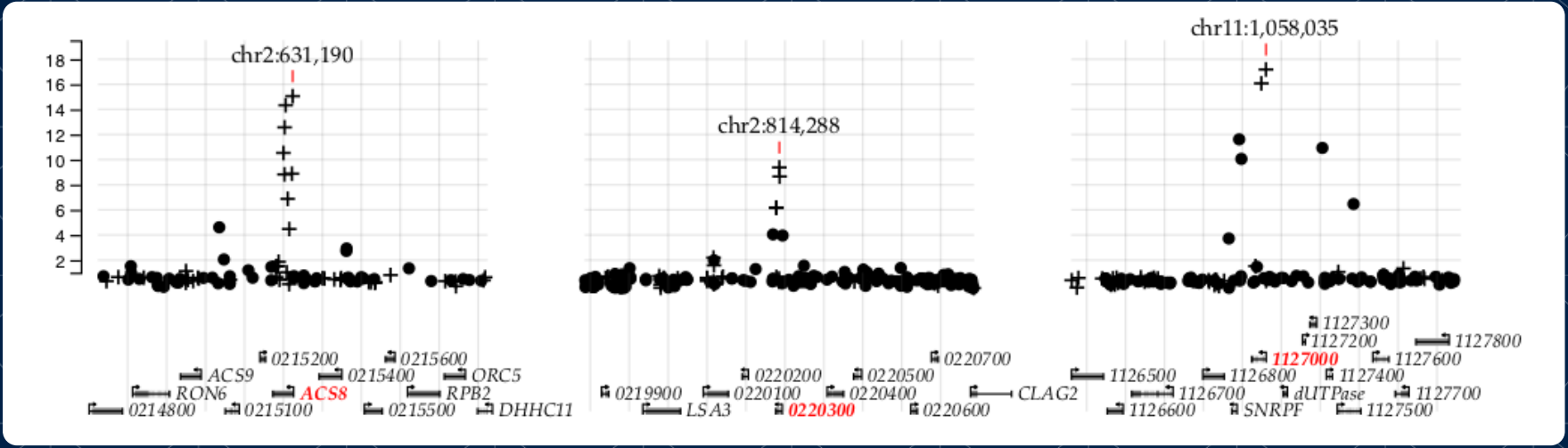
- …if so, they are presumably positively selected in individuals carrying HbS – a kind of reverse malaria hypothesis.

- This must be one of the selective forces generating the unusual population-genetic features - but any competing forces are currently unknown.

- Raises many questions…

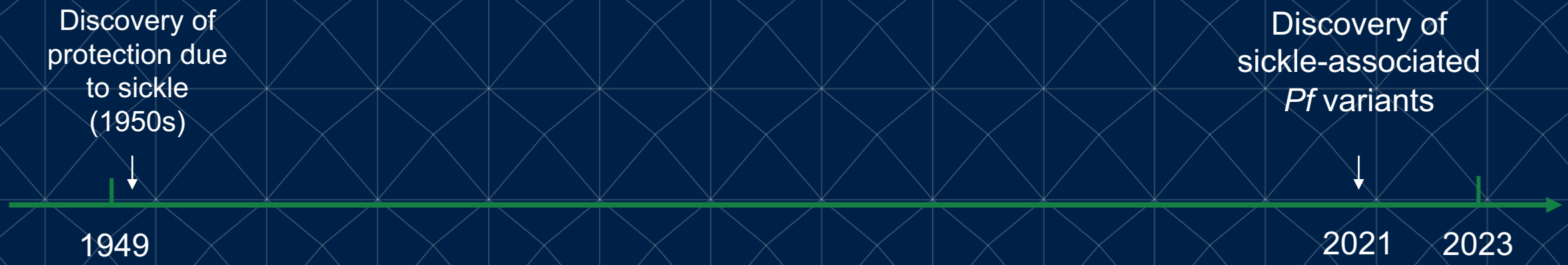
Puzzles and questions

- What is the underlying biology?



The function of the *Pfsa* genes is not known

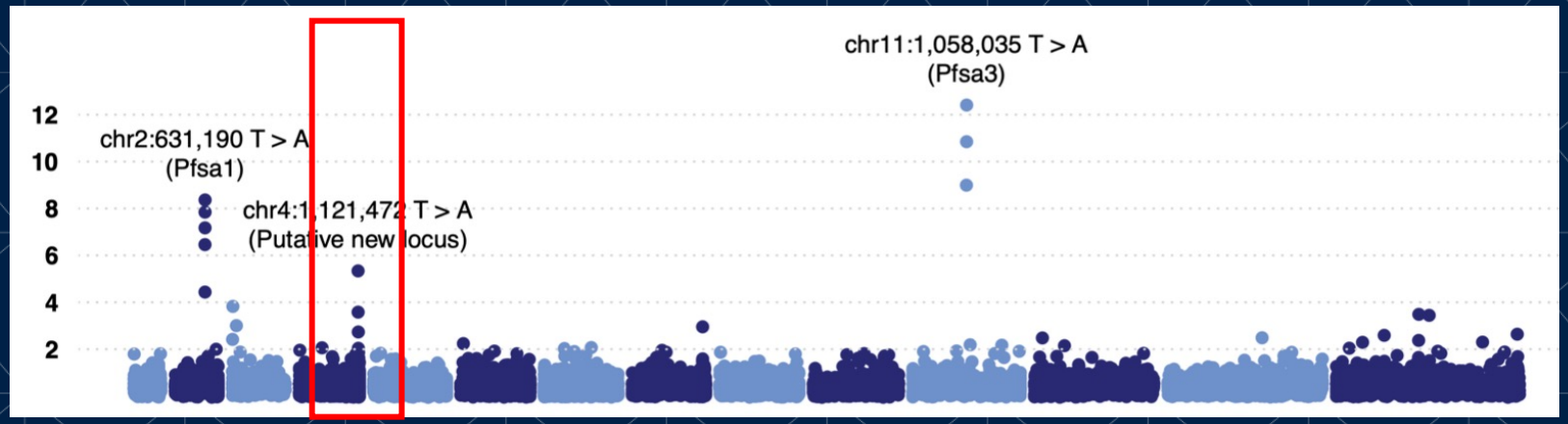
- How is the long-range LD maintained?
- What happens in other populations – or in milder cases?
- How does sickle provide protection anyway?


Discovery of protection due to sickle (1950s)
Discovery of sickle-associated *Pf* variants
1949
2021
2023

Stop press

Hamilton W. , Amenga-Etego L.
"A fourth locus in the Plasmodium falciparum genome associated with sickle haemoglobin"
bioRxiv (2023)

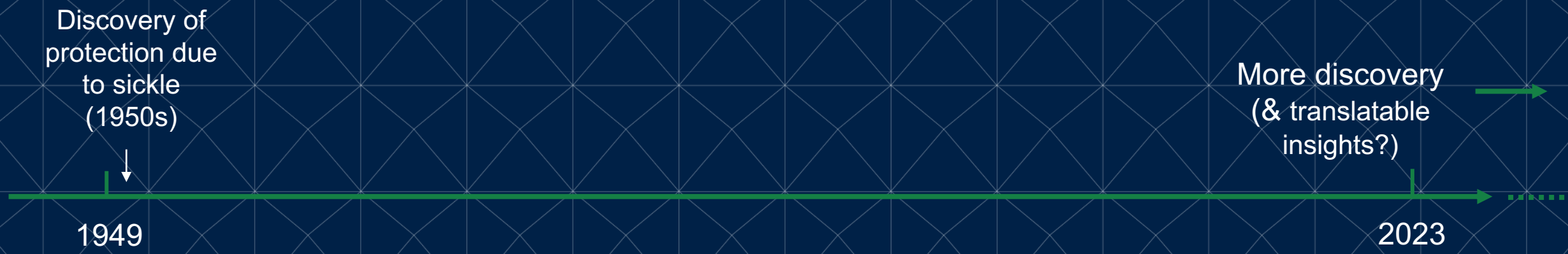
Evidence



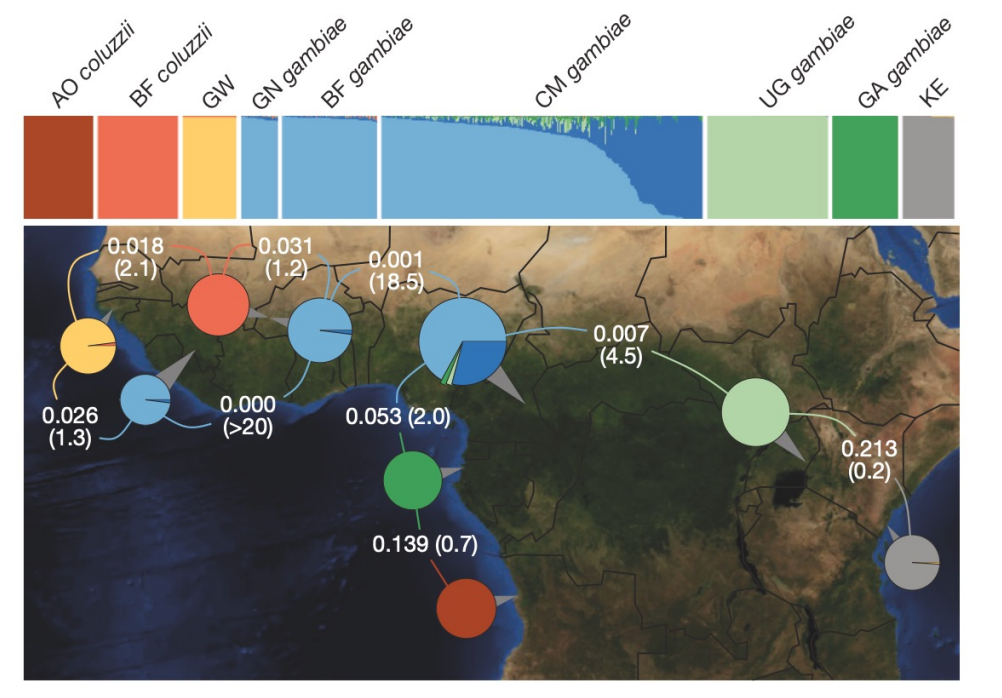
Parasite genome

There appears to be a 4th *Pfsa* locus in west Africa.

Discovered in non-severe cases from Ghana


Discovery of protection due to sickle (1950s)
More discovery (& translatable insights?)
1949
2023

The missing part of my talk: mosquito genomics



The *Anopheles gambiae* 1000 Genomes Consortium, Nature 2017

Describing population structure.

Tracking insecticide resistance and gene flow.

Parasite-vector interactions?

wellcome centre human genetics

wellcome sanger institute

KEMRI | Wellcome Trust

MRC Unit The Gambia

London School of Hygiene & Tropical Medicine

Dominic Kwiatkowski 1953-2023

Kirk Rockett

Ellen Leffler

Muminatou Jallow

Tom Williams

MRC The Gambia @ LSHTM
Muminatou Jallow
Fatoumatta Sisay-Joof
Kalifa A. Bojang
David J. Conway
Giorgio Sirugo
Umberto d'Alessandro

KEMRI-Wellcome, Kilifi, Kenya
Carolyne M. Ndila
Alexandar W. Macharia
Norbert Peshu
Kevin Marsh
Thomas N. Williams

University of Oxford
Kirk A. Rockett
Anna E. Jeffreys
Christina Hubbart
Kate Rowlands
John Todd

Wellcome Sanger Institute
Thuy Nguyen
Sonia M. Goncalves
Cristina V. Ariani
Jim Stalker
Richard D.Pearson
Roberto Amato
Eleanor Drury
Dominic P. Kwiatkowski

Duke University
Steve M. Taylor
Joseph P. Saeelens

USTTB Mali:
Mahamadou Diakite

Thanks also:
Annie Forster
Jia-Yuan Zhang
Andre Python

www.malariagen.net

"We thank the patients and staff at the Paediatric Department of the Royal Victoria Hospital in Banjul, Gambia, and at Kilifi County Hospital and the KEMRI-Wellcome Trust Research Programme, Kilifi, Kenya for their help with this study."

Publication at:
doi.org/10.1038/s41586-021-04288-3

wellcome

MRC Medical Research Council

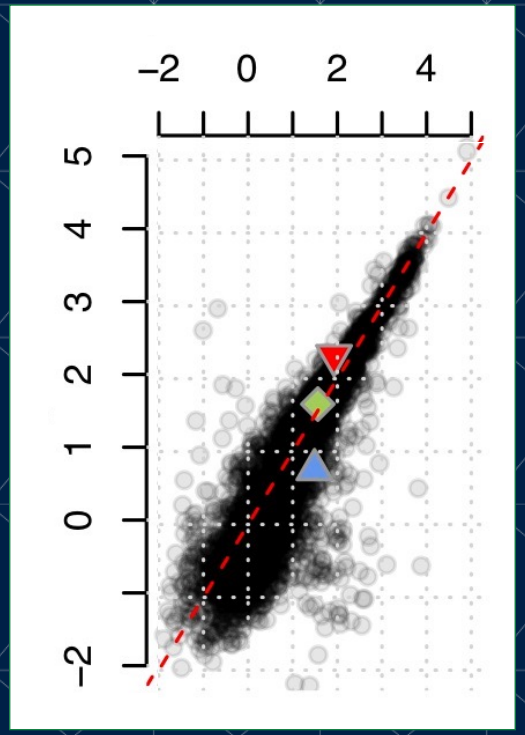
FNIH Foundation for the National Institutes of Health

Bill & Melinda Gates foundation

MalariaGEN
GENOMIC EPIDEMIOLOGY NETWORK

Parasite gene expression (RNA levels)

sickle-associated (*Pfsa+*) parasite



Pfsa− parasite

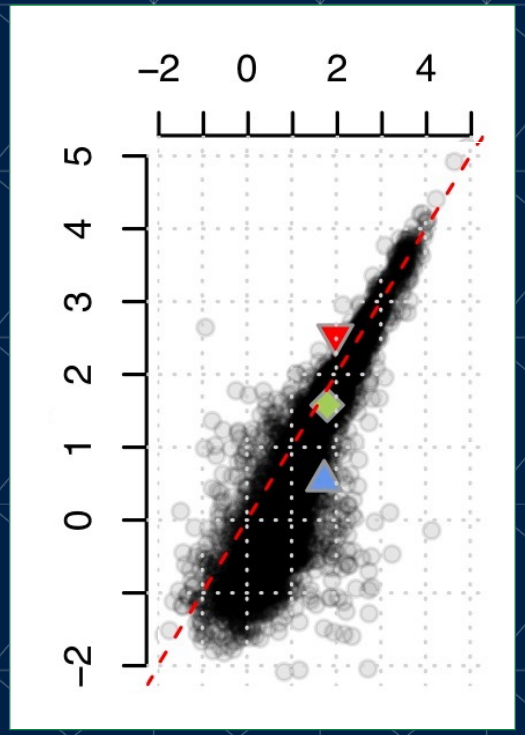
T=9

Hours post-invasion

▼ PF3D7_1127000 ● other genes

The parasite takes about 48 hours to replicate within red cells (then they burst and the parasites re-invade). What does gene expression look like across this cycle?

sickle-associated (*Pfsa+*) parasite



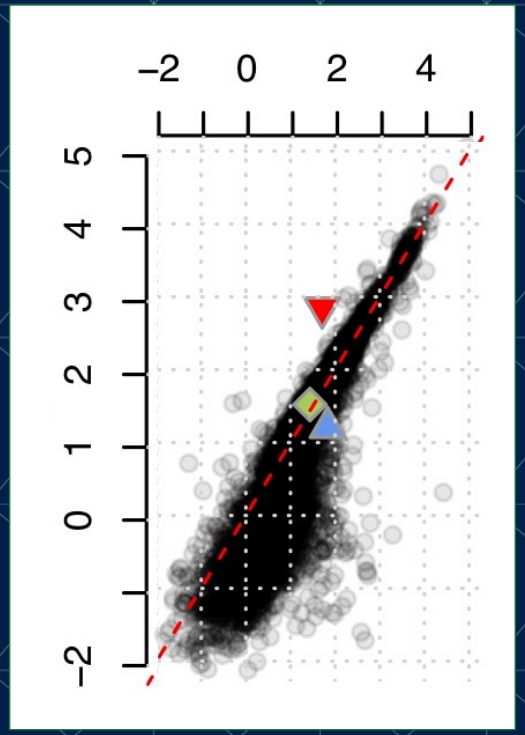
Pfsa− parasite

T=15

Hours post-invasion

PF3D7_1127000 other genes

sickle-associated (*Pfsa+*) parasite



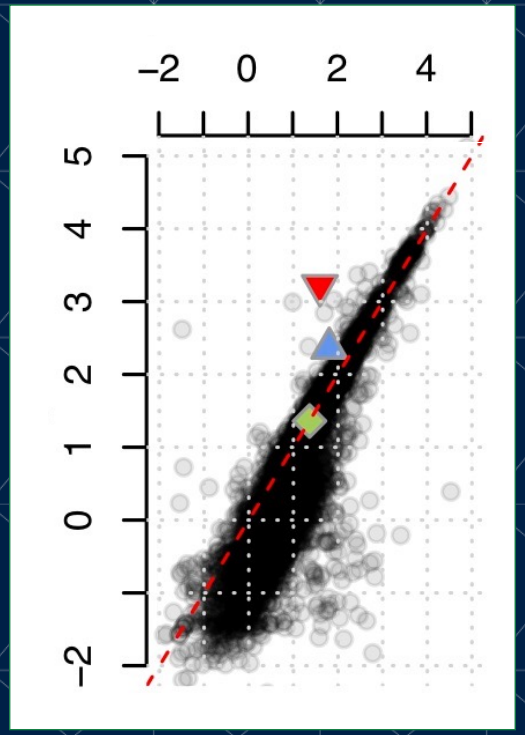
Pfsa− parasite

T=21

Hours post-invasion

▼ PF3D7_1127000 ● other genes

sickle-associated (*Pfsa+*) parasite



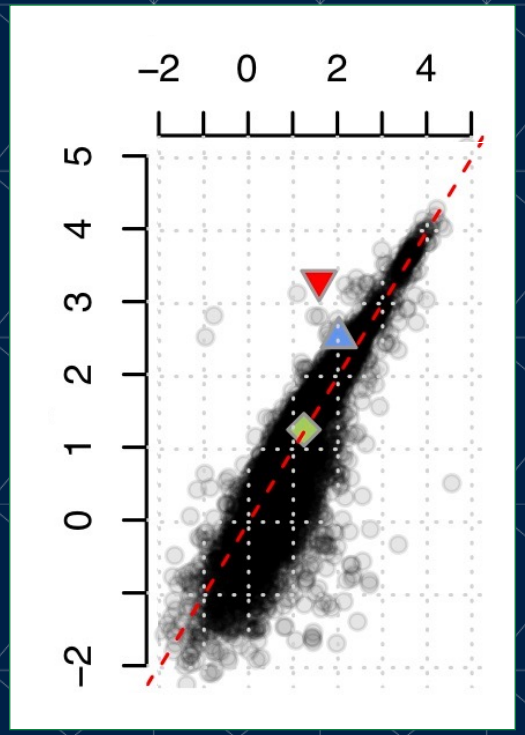
Pfsa− parasite

T=24

Hours post-invasion

PF3D7_1127000 other genes

sickle-associated (*Pfsa+*) parasite



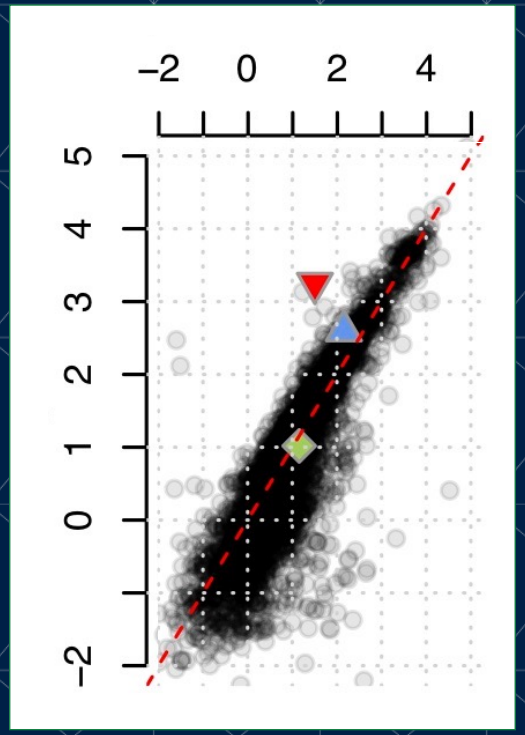
Pfsa− parasite

T=27

Hours post-invasion

PF3D7_1127000 other genes

sickle-associated (*Pfsa+*) parasite



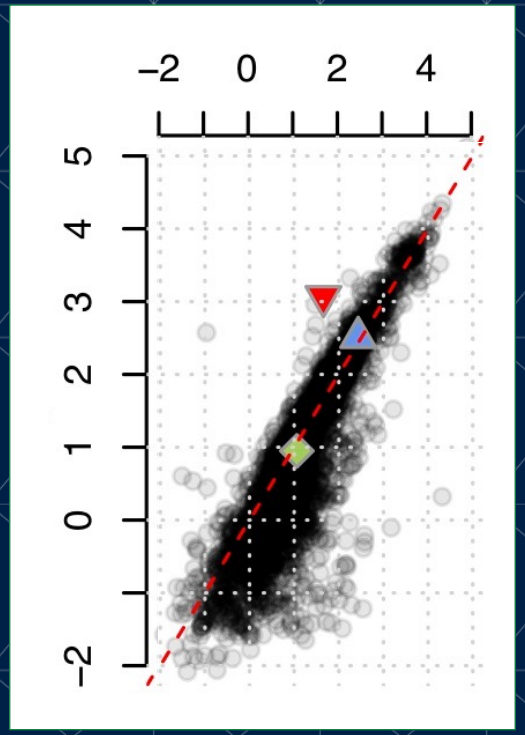
Pfsa− parasite

T=30

Hours post-invasion

PF3D7_1127000 other genes

sickle-associated
(*Pfsa+*) parasite



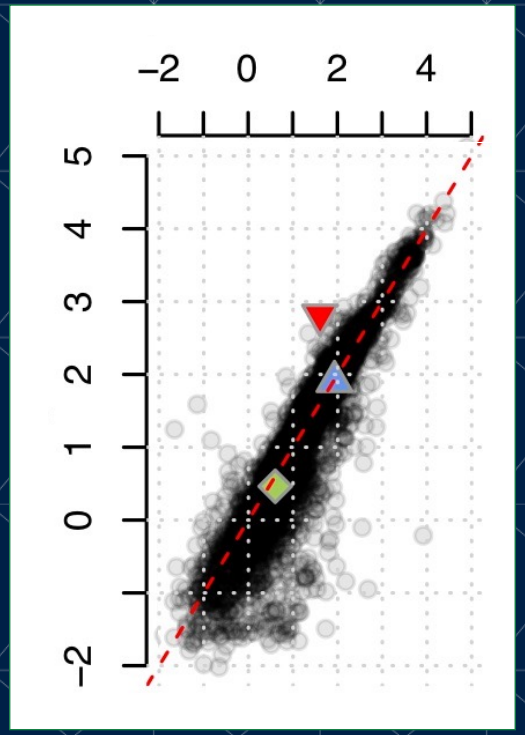
Pfsa− parasite

T=33

Hours post-invasion

PF3D7_1127000 other genes

sickle-associated
(*Pfsa+*) parasite



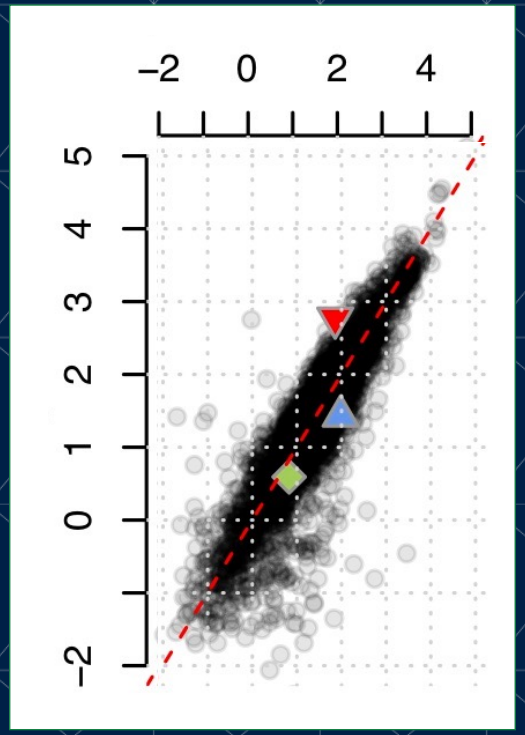
Pfsa− parasite

T=39

Hours post-invasion

PF3D7_1127000 other genes

sickle-associated (*Pfsa+*) parasite

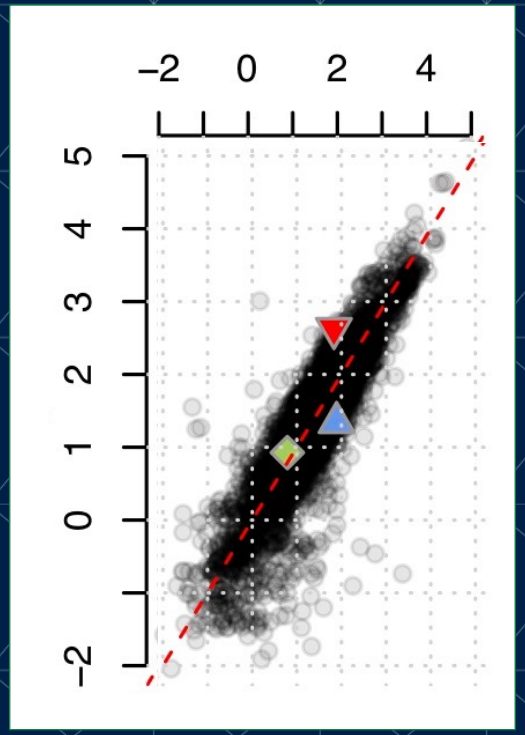


Pfsa− parasite

T=45

Hours post-invasion

PF3D7_1127000 other genes



T=48

Hours post-invasion

▼ PF3D7_1127000 ● other genes