

Genome-wide association study identifies human genetic variants associated with fatal outcome from Lassa fever

Received: 30 September 2022

Accepted: 14 December 2023

Published online: 07 February 2024

 Check for updates

A full list of authors and their affiliations appears at the end of the paper.

Infection with Lassa virus (LASV) can cause Lassa fever, a haemorrhagic illness with an estimated fatality rate of 29.7%, but causes no or mild symptoms in many individuals. Here, to investigate whether human genetic variation underlies the heterogeneity of LASV infection, we carried out genome-wide association studies (GWAS) as well as seroprevalence surveys, human leukocyte antigen typing and high-throughput variant functional characterization assays. We analysed Lassa fever susceptibility and fatal outcomes in 533 cases of Lassa fever and 1,986 population controls recruited over a 7 year period in Nigeria and Sierra Leone. We detected genome-wide significant variant associations with Lassa fever fatal outcomes near *GRM7* and *LIF* in the Nigerian cohort. We also show that a haplotype bearing signatures of positive selection and overlapping *LARGE1*, a required LASV entry factor, is associated with decreased risk of Lassa fever in the Nigerian cohort but not in the Sierra Leone cohort. Overall, we identified variants and genes that may impact the risk of severe Lassa fever, demonstrating how GWAS can provide insight into viral pathogenesis.

Lassa fever is an illness that can result from infection with Lassa virus (LASV). Initial Lassa fever symptoms (fever, vomiting, cough, sore throat) can quickly progress to respiratory distress, mucosal bleeding, shock and multiorgan failure¹. Overall case fatality rates (CFRs) are as high as 29.7% in laboratory-confirmed patients² and more than 50% in fetuses^{3,4}. This lethality, coupled with the aerosol-based route of exposure and lack of approved therapeutics or vaccines, means that LASV is a World Health Organization risk group 4 pathogen, biosafety level 4 (BSL-4) agent and substantial threat to public health.

LASV is ubiquitous in many regions of West Africa. The main host and reservoir of LASV is *Mastomys natalensis*, a rodent that lives near houses in rural villages. Capture surveys have detected LASV in 3.2–52% of rodents^{2,5}. LASV is transmitted to humans through aerosolization of viral particles from rodent excrement. Consistent with the rodent reservoir's prevalence and virus' transmissibility, antibody surveys indicate that between 8% and 52% of residents in some regions have been exposed to LASV^{6,7}, leading to an estimated 100,000–300,000

infections of LASV annually⁸. Person-to-person transmission has been reported but usually only in nosocomial settings⁹.

Despite the prevalence of LASV, only hundreds to thousands of cases of Lassa fever are diagnosed each year¹⁰, suggesting that most infections are undocumented and mild. Why severe disease and death only occurs in a subset of LASV infections is not clear. Although old age¹¹ and pregnancy^{2,3} are associated with poor Lassa fever outcomes, they do not explain all the variability in infection outcome. Variability among LASV lineages¹² has not been linked to severity of symptoms.

Human genetic variation may contribute to variability in the outcome of LASV infection. Host genetics has been linked to symptoms caused by infection with severe acute respiratory syndrome coronavirus 2, human immunodeficiency virus (HIV), dengue and hepatitis A–C^{13–15}. The link between host genetics and LASV infection is intriguing because LASV may have been an important selective force in endemic regions, driving variants that protect against Lassa fever to higher prevalence. We previously reported a signal of positive selection in a Yoruba population from Nigeria, who live in a LASV endemic region,

✉ e-mail: dkotliar@mgb.org; happic@run.edu.ng; pardis@broadinstitute.org

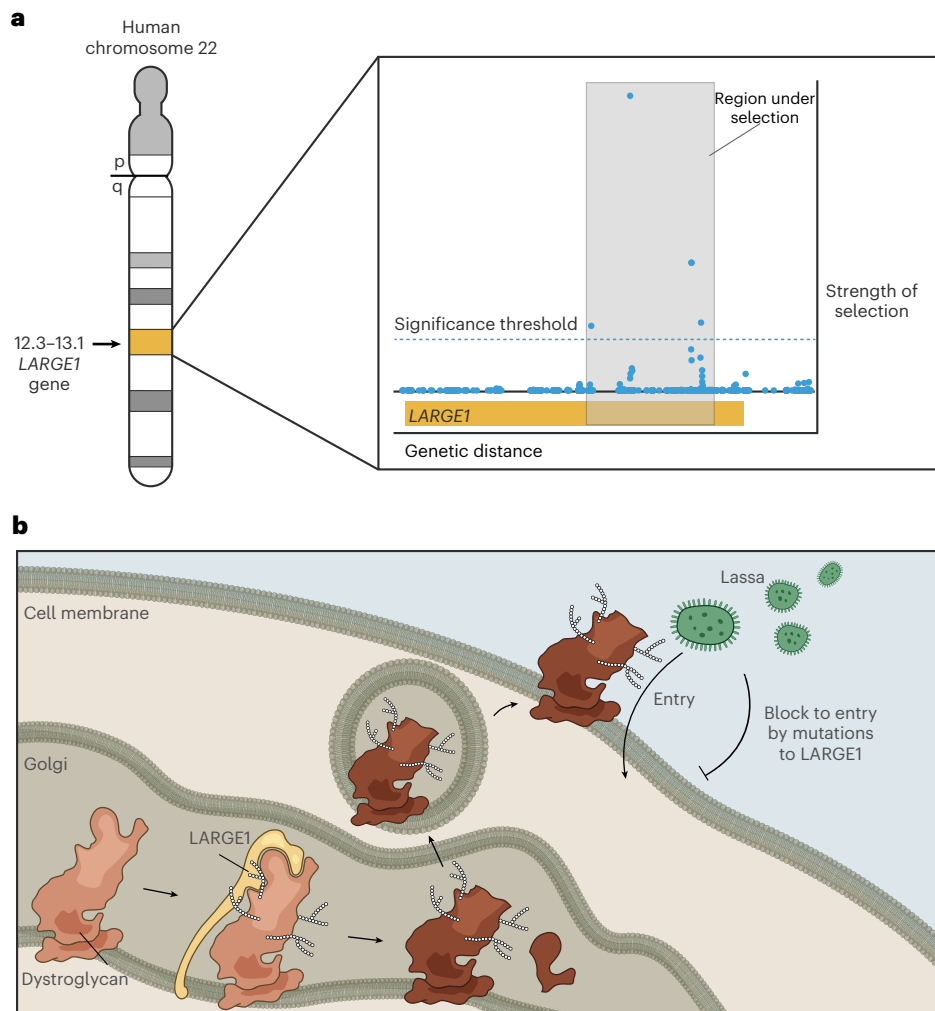


Fig. 1 | Overview of hypothesized mechanism of positive selection for resistance to Lassa fever mediated by *LARGE1*. **a**, Statistical evidence for positive selection at the *LARGE1* locus, adapted from Andersen et al.¹⁷. The y axis shows the composite likelihood score which integrates evidence of positive selection based on population differentiation (fixation index), long haplotype

(integrated haplotype score, delta integrated haplotype score, cross-population extended haplotype homozygosity) and derived allele frequency. On the figure, p refers to the short arm of the chromosome, while q refers to the long arm. See Andersen et al.¹⁷ for details. **b**, Hypothesized mechanism by which decreased activity of *LARGE1* increases resistance to LASV infection and Lassa fever.

at a locus overlapping the gene *LARGE1* (refs. 16,17) (Fig. 1a). *LARGE1* encodes a protein that glycosylates α -dystroglycan, the primary cellular receptor for LASV^{18,19}. LASV infectivity in vitro depends on the level of *LARGE1* expression¹⁹. Therefore, a variant in the putative region under positive selection may have been driven to high allele frequencies by impacting expression levels of *LARGE1*, thereby reducing the risk of severe Lassa fever (Fig. 1b). Given Lassa fever's lethality among diagnosed cases and the high seroprevalence to LASV, it is plausible that host variants providing resistance might have an impact on reproductive fitness. In addition, phylogenetic dating indicates that LASV has been present for over 1,000 years in Nigeria¹², making it feasible that the virus might have exerted evolutionary pressure on humans. However, no previous studies have systematically assessed the impact of host variation in LASV infection.

Despite the clinical importance of Lassa fever, there are practical obstacles to studying it in human patients. First, LASV is a BSL-4 pathogen endemic in countries that have only recently obtained infrastructure for safe virus handling. Second, medical infrastructure is lacking in the villages where Lassa fever is most common, so most symptomatic Lassa fever cases are undocumented. Finally, genetic diversity of LASV isolates means that diagnostics based on nucleic acid amplification or

immunoassays can have low sensitivity. As there are no US Food and Drug Administration-approved LASV diagnostics²⁰, proven diagnoses require viral culture, which is generally not feasible. We anticipated that it would be challenging to obtain a sizable enough cohort to carry out a Lassa fever genome-wide association study (GWAS) but hypothesized that increased power would arise if natural selection for resistance to Lassa fever was present. This is because natural selection would increase the prevalence of advantageous alleles, over time generating common resistance alleles. Such highly protective variants might be detectable in genetic association studies of modest sample size. For instance, the sickle cell allele in haemoglobin is one of the most robust signals of genetic resistance to infectious disease and can be detected in small samples^{21,22}. We hypothesized that if this was the case, a Lassa fever GWAS could elucidate the biological basis of Lassa fever resistance.

Beginning in 2008, we established public health and research capabilities for Lassa fever in two countries in West Africa. To obtain an adequate cohort size, we recruited and genotyped patients with Lassa fever and geographically matched individuals who do not have LASV symptoms (population controls) during a 7 year period from LASV endemic regions of Nigeria and Sierra Leone using an array of

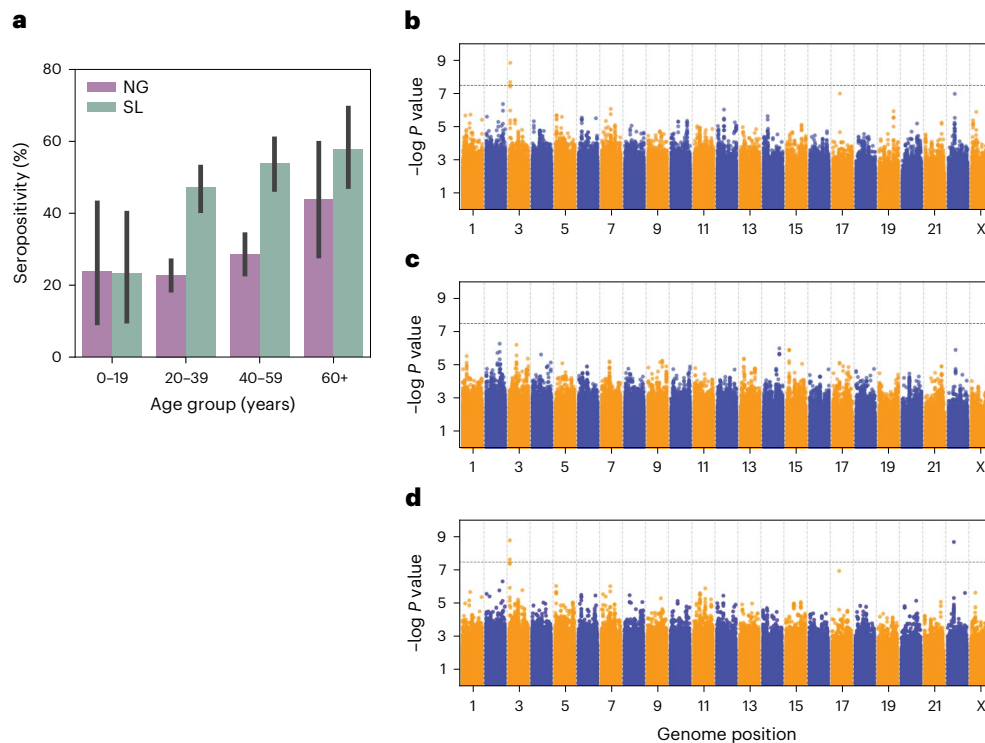


Fig. 2 | GWAS of Lassa fever clinical outcome. a, Immunoglobulin G seropositivity rate in Nigerian (NG) and Sierra Leonean (SL) controls stratified by age. Error bars represent 95% bootstrap confidence intervals. NG: *N* of 24 in 0–19 years, 424 in 20–39 years, 269 in 40–59 years and 34 in 60+ years. SL: *N* of 33 in 0–19 years, 282 in 20–39 years, 191 in 40–59 years and 83 in 60+ years.

b–d, Manhattan plots showing the $-\log P$ value for each genomic variant for the Lassa fever outcome association for Nigeria (**b**), Sierra Leone (**c**) and meta-analysis (**d**). *P* values for **b** and **c** are based on SAIGE, while *P* values for **d** are derived from meta-analysis (METAL) of *P* values shown in **b** and **c**.

diagnostic tests to capture the broadest possible set of cases while minimizing false positives. We tested for genome-wide association with Lassa fever susceptibility and fatal outcomes, with sub-analyses specifically considering variation at *LARGE1* and the human leukocyte antigen (HLA) loci.

GWAS recruitment and clinical characterization

We recruited and genotyped 411 people with LASV and 1,187 controls from Nigeria and 122 people with LASV and 799 controls from Sierra Leone (Extended Data Table 1 and Extended Data Fig. 1).

We used the standard-of-care assays for case definition at each recruitment site and also used next-generation sequencing to detect additional people with LASV missed by traditional diagnostics (Supplementary Note and Extended Data Table 2).

All sequenced LASV genomes from Nigeria were clade II or III, and those from Sierra Leone were clade IV, matching the expected distributions²³. Furthermore, all but one of the Nigeria genomes matched the expected phylogeographic distribution of clade III samples deriving from northern Nigeria and clade II samples deriving from southern Nigeria²⁴.

As we recruited population controls from Lassa fever endemic villages, we suspected that many controls were exposed to LASV in their lifetimes but never developed clinically relevant Lassa fever, thus increasing their likelihood of harbouring protective genetic variation. We used enzyme-linked immunosorbent assays (ELISAs) to measure immunoglobulin G antibodies against LASV for 751 and 589 of the controls from Nigeria and Sierra Leone, respectively (Supplementary Note). We found that 25.9% and 49.6% of the Nigeria and Sierra Leone controls were seropositive, respectively (compared to 0/117 of United States-based controls²⁵), consistent with the upper end of previous seroprevalence surveys in these countries⁶. Furthermore, we found that seropositivity was associated with older age

(rank-sum test $P = 0.0022$ for Nigeria and 0.00053 for Sierra Leone) and increased gradually with age (Fig. 2a), suggesting continuous lifetime exposure to LASV.

We tested whether demographic variables were associated with Lassa fever susceptibility and fatal outcomes. Previous studies reported higher proportions of women and girls with Lassa fever^{26–32}, suggesting increased susceptibility to LASV or exposure to LASV among women^{32,33}. Consistent with this, we found that women and girls are significantly overrepresented within our Nigeria cases (242/411 or 58.9%, binomial test $P = 0.0003$). However, we did not find significant sex differences in the Sierra Leone cases (50/122 or 41.0%, $P = 0.057$). We found that people with LASV were younger than controls in both Nigeria and Sierra Leone (rank-sum test $P = 0.0010$ and 2.15×10^{-17} , respectively) (Extended Data Fig. 2a). CFR was estimated to be 35.3% and 64.8% in our Nigeria and Sierra Leone cases, respectively, consistent with previous estimates in these countries² (Extended Data Table 1).

We tested the association between symptoms and age (Extended Data Table 3) and found that younger patients in both Nigeria and Sierra Leone were more likely to present with vomiting ($P = 0.016$ and 0.012 , respectively) and cough ($P = 0.08$ and 0.001 , respectively) than older patients. We also observed a trend toward higher probability of fatal outcome in older people with LASV, but this was not significant ($P = 0.11$ and 0.17 , respectively, in Nigeria and Sierra Leone).

GWAS of Lassa fever susceptibility and clinical outcome

Owing to the prolonged, interrupted recruitment over 7 years and changes in genotyping platforms over the time frame of recruitment, samples were genotyped on three different arrays: H3Africa, Omni 2.5M and Omni 5M (Extended Data Table 2). We corrected for array-derived batch effects before joint imputation across all arrays (Supplementary Note). This yielded a pre-imputation set of 1,453,101 genotyped variants

Table 1 | Description of lead variants for the susceptibility GWAS analysis

Lead SNP	Chromosome	Position (hg19)	Nearest gene	Lead SNP								
				Nigeria OR	Nigeria 95% CI	Nigeria <i>P</i> value	Nigeria MAF (%)	Sierra Leone OR	Sierra Leone 95% CI	Sierra Leone <i>P</i> value	Sierra Leone MAF (%)	Meta-analysis <i>P</i> value
rs114992845	7	146356694	<i>CNTNAP2</i>	9.19	[3.5, 23.9]	2.7×10 ⁻⁶	1.21	4.77	[1.3, 17.8]	0.010	1.86	1.2×10 ⁻⁷
rs143130878	17	33192408	<i>CCT6B</i>	1.20	[0.6, 2.6]	0.64	3.38	6.87	[3.3, 14.2]	1.1×10 ⁻⁷	2.74	3.3×10 ⁻⁴
rs73397758	17	68745251	<i>CASC17</i>	0.84	[0.5, 1.5]	0.58	6.28	9.16	[4.0, 20.8]	5.5×10 ⁻⁸	2.42	4.8×10 ⁻³

Includes the most significant variant in the meta-analysis of both cohorts and the two most significant variants in the Sierra Leone analysis. Country-specific *P* values are based on SAIGE, while meta-analysis *P* values are derived from meta-analysis (METAL) of *P* values generated from each cohort. 95% CI, 95% confidence interval for the OR; MAF, minor allele frequency.

and a final imputed set of 12,783,971 variants in Nigeria and 12,522,562 variants in Sierra Leone.

We used generalized linear mixed models as implemented in saddlepoint-approximated score tests (SAIGE)³⁴ to account for relatedness and population stratification in our dataset (Methods). Mixed models analysis is important for this study because the dataset contained many first-degree relatives. Six hundred and sixteen (38%) and 251 (27%) individuals in the Nigerian and Sierra Leone cohorts had a first-degree relative, respectively (Extended Data Fig. 2b). In addition, principal component analysis showed evidence of stratification even after removing closely related individuals in our cohort (Extended Data Fig. 2c); we therefore included principal components (PCs) as fixed effects, which has been shown to control for confounding due to population stratification³⁵. We used a genome-wide significance threshold of 3.24×10^{-8} (previously reported to control for false positives in African populations³⁶). Quantile–quantile plots did not show any evidence of test-statistic inflation, indicating that our statistical controls accounted for dominant confounding variables (Extended Data Fig. 2d).

A GWAS of susceptibility to Lassa fever infection for all individuals in our study did not identify any variants that reached genome-wide significance in either cohort. However, two variants on chromosome 17 showed a trend toward significance in the Sierra Leone cohort (Table 1 and Extended Data Fig. 2e). rs73397758 ($P = 5.5 \times 10^{-8}$, odds ratio (OR) = 9.16) is ~350 KB (kilobase pairs) downstream of the gene *CASC17*, a long non-coding RNA named for a genetic association with prostate cancer³⁷, and 570 KB upstream of *KCNJ2*, a potassium inwardly rectifying channel³⁸. rs143130878 ($P = 1.1 \times 10^{-7}$, OR = 6.87) resides 62,472 base pairs downstream of the gene *CCT6B*³⁹, which is a member of the molecular chaperone (TRiC) family that has been shown to regulate the replication of arenaviruses, including LASV⁴⁰. Neither variant was significantly associated with susceptibility in the Nigeria cohort ($P = 0.58$ and $P = 0.64$, respectively).

The most significant variant in a meta-analysis of the two GWAS cohorts was rs114992845 in an intron of *CNTNAP2* (meta-analysis $P = 1.2 \times 10^{-7}$; Nigeria OR = 9.19, Sierra Leone OR = 4.77) (Table 1). *CNTNAP2* is a member of the neurexin family, many members of which encode proteins that bind to α -dystroglycan, the cellular receptor for LASV⁴¹. Furthermore, loss-of-function mutations in the gene *CNTNAP2* have been associated with recurrent infections⁴², although the underlying mechanism remains unknown. All three variants that were trending toward significance in the susceptibility GWAS are of low frequency (Table 1) and will require larger sample sizes for validation.

A GWAS of fatal outcomes in Lassa fever cases using the same strategy described above did identify genome-wide significant associations (Extended Data Fig. 3a). We did not observe evidence of population stratification or test statistic inflation (Supplementary Fig. 3a,b). We identified a significant association with rs9870087 in the Nigeria cohort, falling within an intron of the gene *GRM7* ($P = 1.54 \times 10^{-9}$, OR = 15.4) (Table 2 and Fig. 2b). The protein encoded by *GRM7* is a glutamate metabotropic receptor active throughout the central nervous system⁴³. While no direct role of this receptor is known in viral infection,

GRM2, another member of this family, has been previously linked to severe acute respiratory syndrome coronavirus 2⁴⁴ and rabies⁴⁵ viral entry. A recent *GRM7* knock-out mouse implicated this gene in neuroimmune signalling in anaphylaxis⁴⁶. Furthermore, *GRM7* has an important role in maintenance of hearing by inner-ear hair cells⁴⁷, and hearing loss is a symptom of Lassa fever⁴⁸. We did not identify any genome-wide significant associations in the Sierra Leone cohort (Fig. 2c).

We also carried out a meta-analysis of fatal outcomes in the Nigeria and Sierra Leone cohorts which identified a genome-wide significant association with rs73404538 (meta-analysis $P = 1.9 \times 10^{-9}$; Nigeria OR = 0.358, Sierra Leone OR = 0.389) (Fig. 2d and Extended Data Table 4). This variant falls 16,453 base pairs downstream of the 3' untranslated region of *LIF*, which encodes an interleukin 6 class cytokine⁴⁹ that has been associated with several viral infections. We further note that rs73404538 is nominally significant in the Sierra Leone susceptibility GWAS ($P = 0.039$, OR = 0.71) and in a meta-analysis of the Nigeria and Sierra Leone susceptibility GWASs ($P = 0.021$) with a concordant direction of effect (Extended Data Table 4). This suggests that in addition to increasing the lethality of Lassa fever, rs73404538 may also increase the probability of contracting clinically detected Lassa fever.

We did not include age as a covariate in our primary analysis due to missing data for many participants (2.4% of Nigeria cases and 25.5% of Sierra Leone controls), but we did so in a secondary analysis. While the *P* values for the susceptibility lead variants decrease by up to 1 order of magnitude, consistent with a loss of power from the decreased sample size, the rs73404538 variant downstream of *LIF* actually becomes genome-wide significant in the Nigeria cohort ($P = 2.2 \times 10^{-8}$, OR = 0.36) and more significant in the meta-analysis ($P = 8.0 \times 10^{-10}$) providing further support for this association (Extended Data Fig. 3c).

As each of the candidate GWAS loci described above contains multiple linked non-coding genetic variants (Extended Data Fig. 4a,b), we used a massively parallel reporter assay (MPRA) to identify which variants are most likely to be functional. MPRA⁵⁰ identifies potential regulatory variants by testing the reference and alternate alleles of thousands of variants in parallel for their ability to impact expression of a plasmid-based reporter (Supplementary Note). We carried out MPRA in K562 and HepG2 cells for loci containing the most significant variants in the susceptibility and fatal outcome GWASs (Supplementary Tables 3–5).

We identified potential regulatory variants in many of our top GWAS loci. For the *CASC17* locus, we find that the only tested variant to show regulatory activity is rs112446079 in K562 cells (\log_2 skew = -0.64, $q = 0.031$), the second most strongly associated variant in the region (Extended Data Fig. 4c, left). Similarly, for the *CNTNAP2* locus, the seventh most strongly associated variant in the region, rs150484921, showed regulatory activity by MPRA (\log_2 skew = -0.65, $q = 0.011$), but the lead variant did not (Extended Data Fig. 4c, right). Several variants were associated with the second Sierra Leone peak near *CCT6B*, the most significant of which in the GWAS was rs116948215 (\log_2 skew = -0.98, $q = 1.94 \times 10^{-6}$). This latter single-nucleotide polymorphism (SNP) is active in the MPRA in HepG2 cells as well as K562s

Table 2 | Description of lead variants for the fatal outcome GWAS analysis

Lead SNP	Chromosome	Position (hg19)	Nearest gene	Lead SNP								
				Nigeria OR	Nigeria 95% CI	Nigeria <i>P</i> value	Nigeria MAF (%)	Sierra Leone OR	Sierra Leone 95% CI	Sierra Leone <i>P</i> value	Sierra Leone MAF (%)	Meta-analysis <i>P</i> value
rs73404538	22	30619983	LIF	0.358	[0.2, 0.5]	1.1×10 ⁻⁷	47.8	0.389	[0.19, 0.79]	4.7×10 ⁻³	35.8	1.9×10 ⁻⁹
rs9870087	3	7330265	GRM7	15.4	[6.2, 37.9]	1.5×10 ⁻⁹	4.73	0.642 ^a	[0.1, 2.8] ^a	0.55 ^a	5.02 ^a	1.1×10 ^{-6a}

Includes the most significant variant per genomic locus containing at least one genome-wide significant association (including in meta-analysis). *P* values are based on SAIGE, while meta-analysis *P* values are derived from meta-analysis (METAL) of *P* values generated from each cohort. ^ars9870087 was excluded from the Sierra Leone GWAS due to low minor allele count but is included here for completeness.

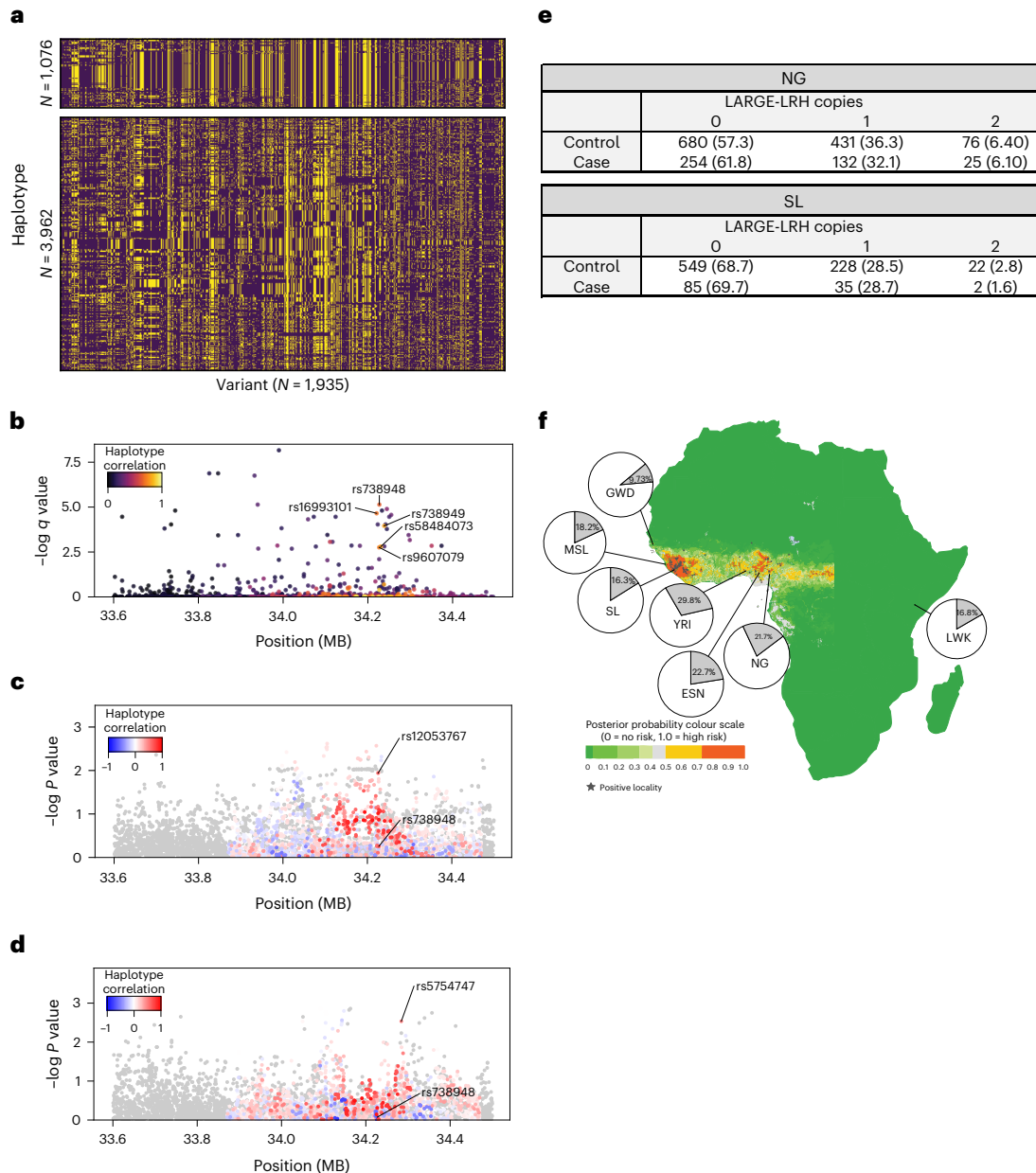


Fig. 3 | Association of the LARGE-LRH haplotype with susceptibility to Lassa fever. **a**, *K*-means clustering of haplotypes in the *LARGE1* region. Rows are phased haplotypes; columns are individual variants with reference alleles shown in purple, alternate alleles shown in yellow and *K*-means clusters separated. **b**, Scatter plot of *q* values for allelic skew in the MPRA, coloured by the absolute value of the Pearson correlation with the haplotype. **c, d**, Scatter plot of GWAS association *P* values over the *LARGE1* region for Nigeria (**c**) and Sierra Leone (**d**) coloured by Pearson correlation of the protective allele in the GWAS with

the LARGE-LRH. *P* values in **c** and **d** are based on SAIGE. **e**, Contingency table of LARGE-LRH genotype counts in cases and controls for Nigeria (NG, top) and Sierra Leone (SL, bottom). **f**, Ecologically estimated Lassa fever prevalence from Fichet-Calvet et al.⁷⁰ with pie charts indicating the frequency of the *LARGE1* haplotype in 1000 Genomes populations (YRI, Yoruba; ESN, Esan; MSL, Mende; LWK, Luhya; GWD, Gambian Mandinka)⁵¹ or our GWAS cohorts (NG, SL). Stars indicate towns, villages or hospitals that encountered outbreaks as detailed in Fichet-Calvet et al.⁷⁰.

suggesting a broader regulatory effect across cell types (Extended Data Fig. 4c, middle). For the outcome analysis, we identified one potential regulatory variant at the *GRM7* locus, rs114312118, which is active specifically in HepG2s (\log_2 skew = 0.87, $q = 0.0077$) (Extended Data Fig. 4f).

Analysis of a positive selection signal overlapping *LARGE1*

Next, we tested whether variation around the gene *LARGE1*, a required LASV entry factor, is associated with resistance to Lassa fever. Previous studies identified a long-range haplotype at this locus, that is, multiple genetic variants located up to 500 KB apart that remain in tight LD. The presence of such an extended haplotype suggests that one or more variants in the locus provides a fitness advantage, causing it to spread to high allele frequency in the population faster than genetic recombination would break down the haplotype^{16,17}.

Although no individual variants on chromosome 22 reached genome-wide significance in the GWAS, we examined the long-range haplotype overlapping the *LARGE1* locus as a single entity to further characterize its correlation with Lassa fever phenotypes. We used *K*-means clustering (with $K = 2$) of phased haplotypes and found a dominant haplotype with long-range LD (Fig. 3a and Methods). We label this haplotype ‘*LARGE1* long-range haplotype’ or LARGE-LRH, for short. LARGE-LRH was well tagged by the lead variants identified in previous positive selection scans, for example, rs5999077, rs1013337 and rs1573662, identified in ref. 16 (D' values of 0.957, 0.773 and 0.735). LARGE-LRH was present at 23.9% and 16.9% allele frequency in the Nigeria and Sierra Leone cohorts, respectively.

As LARGE-LRH comprises 96 tightly linked variants with Pearson correlation above 0.6 using the *K*-means annotation, we applied MPRA to zoom into potentially causal variants underlying the signal of positive selection. We tested a library of 5,286 oligonucleotides (of 200 base pair length) centred on different alleles of 1,674 variants in the *LARGE1* region for regulatory function using MPRA (Supplementary Note) (Fig. 3b). Fifty-four of the 1,674 tested variants (3.23%) had significant skew (false discovery rate (FDR)-adjusted $P < 0.05$) between the reference and alternate allele. Of these, five (rs738948, rs16993101, rs738949, rs58484073 and rs9607079) had an FDR-adjusted $P < 0.01$ and were linked to the haplotype with a Pearson correlation > 0.6 . This analysis shows that these variants might regulate gene expression and are candidates for positive selection effects in human populations.

We next evaluated whether any variants in linkage with LARGE-LRH were associated with susceptibility to Lassa fever (Fig. 3c,d). The haplotype-linked variant with the strongest association with Lassa fever susceptibility in the Nigeria cohort was rs12053767 ($P = 0.011$, haplotype Pearson correlation of 0.57). However, this variant was not significantly skewed by MPRA ($q = 0.998$) and was not significantly associated with Lassa fever in the Sierra Leone cohort ($P = 0.25$). The haplotype-linked variant with the strongest association to Lassa fever susceptibility in the Sierra Leone cohort was rs5754747 ($P = 0.0030$, haplotype Pearson correlation of 0.46), but this variant was also not significant in the Nigeria cohort ($P = 0.988$) or significantly skewed by MPRA ($q = 0.26$).

We reasoned that LARGE-LRH, taken together as a single allele, could yield a stronger signal than individual SNPs if the causal variant is not genotyped or if the causal mechanism involves an interaction among multiple variants on the haplotype. We tested whether LARGE-LRH is associated with Lassa fever using the same model that we used in the primary GWAS and found that LARGE-LRH was significantly associated with Lassa fever susceptibility in Nigeria ($P = 0.0492$) but not in Sierra Leone ($P = 0.412$). The overall allele frequency of LARGE-LRH was slightly higher in controls than in people with LASV (Nigeria, 24.6% allele frequency in controls versus 22.1% in people with LASV; Sierra Leone, 17.0% versus 16.0%), consistent with our hypothesized resistance model (Fig. 3e). We note that the association with LARGE-LRH is mainly driven by individuals recruited in the first cohort (Nigeria 2011–2014

recruitment $P = 0.049$, Nigeria 2016–2018 recruitment $P = 0.98$) and that there is a trend toward association in the Sierra Leone cohort during that time period (Sierra Leone 2011–2014 recruitment $P = 0.11$). As there were no controls recruited in Sierra Leone in the second cohort, we do not have a 2016–2018 comparison for it. We were surprised that people with LASV recruited in 2016–2018 did not have a lower frequency of LARGE-LRH (Extended Data Fig. 5), so further study is necessary to harmonize these conflicting observations.

To further test the link between the selection signal at *LARGE1* and Lassa fever, we used 1000 Genomes Project (1KGP) data to test whether LARGE-LRH was present at higher frequency in populations living in LASV endemic regions. We quantified the haplotype frequency of individuals from 26 populations sequenced by the 1KGP⁵¹, including several African populations in LASV endemic regions (Esan, Yoruba and Mende) (Fig. 3f). We identified tag SNPs linked to the LARGE-LRH with Pearson correlation > 0.92 . We then analysed phased 1KGP sequence data and called the LARGE-LRH if three or more of the haplotype-linked alleles were present (Methods). The 1KGP cohort contained 27 individuals homozygous for the LARGE-LRH, 198 heterozygous individuals and 2,279 carrying 0 copies. LARGE-LRH was absent from all European and Asian ancestry populations tested and was present at the highest frequency in populations in LASV endemic regions (Yoruba 30.5%, Esan 23.2% and Mende 20.0%) (Fig. 3f). It was also present in Luhya (16.7%) and Mandinka (10.2%), African populations, outside of the LASV endemic zone (Fig. 3f). Mandinka are geographically close to the Lassa fever endemic region, and the Luhya are historically tied to West Africa through the Bantu expansion, so the elevated allele frequencies could be explained by migration after the putative selective sweep or by a changing geographic distribution of LASV.

Imputation and association analysis of HLA alleles

We tested for associations between Lassa fever and genetic variation in the HLA region. HLA genes encode polymorphic proteins that present antigens to T cells and have been associated with many infectious disease phenotypes¹⁵. While we did not identify genome-wide significant associations with SNPs in the HLA genes, HLA-specific imputation approaches are frequently required to identify HLA associations⁵².

We imputed four-digit HLA alleles, which are complete amino acid sequences, and additional sequencing-based HLA typing of eight classical HLA genes to serve as ‘ground truth’ HLA calls to evaluate imputation accuracy (Methods). Sequencing-based typing of the eight classical HLA genes in 297 individuals in our Sierra Leone cohort identified 41 novel HLA alleles that were not present in the International Immunogenetics database (Extended Data Table 5). Nine of the novel alleles were from HLA class I loci, while 32 were HLA class II, with *DQB1* and *DPA1* having the most novel alleles with 11 and 9, respectively. Notably, a novel allele at 5% allele frequency, *DPA1**03:01@2, disrupts the start codon (ATG to ACG).

We compared imputation accuracy of the four-digit HLA calls with sequencing-based ground truth sets from our Sierra Leone cohort, as well as Esan and Mende individuals from 1KGP. Imputation accuracies compared to the sequencing-based calls in Sierra Leone ranged from 89.2% to 97.6% (Fig. 4a). An additional 76 and 84 Mende and Esan individuals from our Sierra Leone and Nigeria cohorts, respectively, were typed for HLA genes *A*, *B*, *C*, *DQB1* and *DRB1* as part of 1KGP⁵³. For these groups, imputation accuracy ranged from 91.4% to 99.2% (Fig. 4a). These comparisons showed adequate imputation of HLA alleles from SNP genotypes for our cohort.

We examined association of the four-digit HLA alleles with Lassa fever susceptibility phenotypes. No HLA alleles had a significant association with Lassa fever after correcting for multiple hypothesis testing (Fig. 4b). The allele with the strongest evidence of association considering both cohorts was *DRB1**15:03, which had a P value of 0.089 in the Nigeria cohort and 0.064 in the Sierra Leone cohort, resulting in

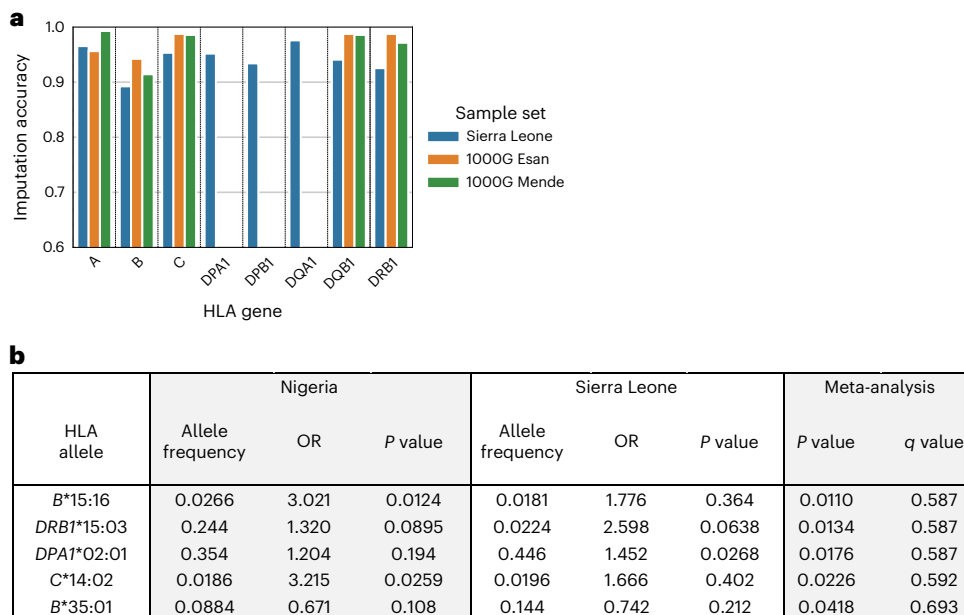


Fig. 4 | Association of HLA variation with Lassa fever susceptibility.

a, Imputation accuracy of four-digit HLA calls compared to sequencing-based ground truth sets from our Sierra Leone cohort, as well as Esan and Mende individuals from 1000 Genomes. **b**, Table of HLA alleles with the strongest

association with Lassa fever susceptibility, ordered by meta-analysis of the NG and SL cohorts. *P* values are based on SAIGE, while meta-analysis *P* values are derived from meta-analysis (METAL) of *P* values generated from each cohort. ORs are computed from Firth logistic regression.

a meta-analysis *P* value of 0.013. *B*15:16* and *C*14:02* yielded *P* values of 0.0124 and 0.0259 in the Nigeria cohort, and *DPA1*02:01* yielded a *P* value of 0.027 in the Sierra Leone cohort. After correcting for multiple hypothesis testing over all HLA tests, the most significant meta-analysis *q* value was 0.587 (Fig. 4b). Similarly, we did not find any associations for fatal outcomes after correcting for multiple hypothesis testing (*q* < 0.05). We tested the 41 novel HLA alleles that were discovered in our Sierra Leone cohort in a similar analysis (Methods), but none were significant.

Discussion

Over a 10 year period we completed the first GWAS of infection with a risk group 4 pathogen reported to date. Our cohorts were recruited in remote parts of West Africa where Lassa fever is most prevalent. They reflected the paradoxical clinical heterogeneity of Lassa fever, with high fatality rates among people with LASV and high LASV seroprevalence among population controls. We find that an intronic variant within *GRM7* and a variant downstream of *LIF* are significantly associated with Lassa fever in the Nigeria cohorts and meta-analysis of the two cohorts, respectively. We identified candidate variants that approach, but do not reach, genome-wide significance in susceptibility analyses.

Several of the loci identified in our study contain genes with potential connections to Lassa fever biology. *LIF* encodes an interleukin 6 family cytokine that was previously shown to protect against lung injury in mouse models of respiratory syncytial virus infection⁵⁴ and to be up-regulated in acute HIV infection⁵⁵ and meningococemia⁵⁶. Altered regulation of this pleiotropic cytokine due to host variation could impact Lassa fever severity, giving rise to the observed association with fatality. *GRM7* may function in viral entry akin to *GRM2* in coronavirus disease 2019 or could be involved in immune activation as was seen in a recent knock-out model of anaphylaxis⁴⁶. In addition, *GRM7* plays an important role in maintenance of hearing by inner-ear hair cells⁴⁷; interestingly, hearing loss is a notable symptom of Lassa fever⁴⁸. MPRA of the significant GWAS loci pinpointed the specific variants most likely to exert regulatory effects in the genome. None of these variants co-localized with expression quantitative trait loci in the

Genotype-Tissue Expression dataset, but this might reflect the relative lack of African ancestry individuals in this resource⁵⁷.

The variants reported here have ORs ranging from 6.87 to 9.19 for the susceptibility GWAS and as high as 15.4 for the outcome analyses (Tables 1 and 2). Intriguingly, the associated risk alleles are mostly uncommon, ranging from 1% to 5% frequency in our cohorts. Given their low frequency, they might be expected to have larger biological effects than what is typically seen for common variants⁵⁸. Furthermore, the low allele frequency may reflect strong purifying selection, with the ubiquitous virus and high CFR purifying the risk allele from the population. Alternatively, the large effect sizes might reflect ‘winner’s curse’, in which only reporting variants that pass, or approach, genome-wide significance results in systematic upward bias of reported effect sizes in GWAS⁵⁹. Larger replication studies and further biological characterization will be needed to clarify these signals.

We used our data to test a hypothesis that positive selection for genetic variation at the *LARGE1* locus provides protection from Lassa fever^{6,16,17}. We found that a haplotype with long-range LD, indicative of recent positive selection, is nominally associated with reduced likelihood of Lassa fever in the Nigeria cohort but not in the Sierra Leone cohort. We reported promising support for this hypothesis in the 2011–2014 cohort, but this did not replicate in the subsequent recruitment from 2016–2018 (Extended Data Fig. 5). The discrepancy between cohorts might represent false positives in the first, power-limited, study or underlying differences between these temporally separated cohorts. It is noteworthy that, after the Ebola outbreak from 2013 to 2016, the number of suspected cases at Irrua Specialist Teaching Hospital (ISTH) surged²⁴. Genetic epidemiology did not find evidence that a particular viral variant or extensive human-to-human transmission underpinned the surge, suggesting that it may have been driven by increased surveillance. Larger cohorts and deeper phenotypic characterization will be required to evaluate the hypothesis of *LARGE1* mediated genetic resistance to Lassa fever susceptibility.

We faced four major obstacles that will inform the design of similar studies: small sample sizes, uncertainty in case and control definitions, impact of environmental variables and insufficient characterization of genetic diversity in African populations.

Achieving large sample sizes for human studies of BSL-4 pathogens is challenging. Very few cases are documented annually, for example, less than 1,000 in Nigeria, the most populous country in the LASV endemic region¹⁰. Lassa fever is prevalent in rural areas that are far from diagnostic centres, further hampering recruitment⁶⁰. Few facilities have diagnostic capacity for LASV infection, and field-deployable LASV tests are not widely available. Therefore, only a fraction of Lassa fever cases are identified, most likely those in which extreme disease presentations motivated the patient to seek medical attention. Some practical investments that would help increase the detection and treatment of LASV infection include diagnostic centres in rural areas, field-deployable, point-of-care diagnostics, and integrated health systems.

Defining Lassa fever cases and controls remains difficult, owing to insufficient diagnostic assays and LASV's genetic diversity. These factors may result in false negatives as well as false positives that reduce power. We mitigated these limitations by using viral sequencing to supplement diagnosis at both sites. Our study also relied on population controls with unknown prior exposure to LASV. We used serology to characterize prior exposure but could not test every control in our cohort. Furthermore, interpretation of serology data is challenging as asymptomatic infections may not lead to sustained seropositivity (leading to false negatives) or could reflect the presence of undocumented Lassa fever in the past rather than asymptomatic illness. In any of these scenarios, the controls would be expected to carry the same susceptibility alleles as the people with LASV, reducing power to detect associations. Questionnaires to elicit detailed disease histories coupled with deeper serological characterization may help to distinguish individuals with previous Lassa fever from those with asymptomatic infection.

Viral genetic diversity, previous infections and co-infections, patient comorbidities and other health factors can further reduce GWAS power. LASV has up to 27% nucleotide diversity such that the specific infecting viral sequence could greatly impact outcomes. Moreover, the lineages in Nigeria and Sierra Leone are so divergent that they could potentially have different mechanisms of interaction with the host. In addition, previous infections with other endemic pathogens or co-infections with other pathogens could be a driver of observed symptoms and disease outcomes⁶¹. In future studies, metagenomic sequencing could define the genome of the infecting LASV strain while identifying the presence of co-infections, allowing these factors to be accounted for in the association model.

African populations are genetically diverse, with low levels of LD, and are under-studied, posing a challenge to GWAS of infectious diseases present mainly in Africa⁶². This issue was directly illustrated in our study; our relatively small HLA sequencing cohort of 297 individuals nevertheless identified 41 novel alleles. GWAS relies on imputing causal variants based on a relatively small number of variants included on the genotyping array. Accurate imputation requires the existence of genotyping arrays containing representative variation from the population of interest and large whole-genome sequencing reference panels, both of which are deficient for African populations. Reduced imputation accuracy can dramatically reduce power, making studies such as this one more challenging. Continuing efforts to improve our understanding of genetic variation in African populations will allow further insights into potential links between genetics and disease.

In summary, our work paves the way for follow-up studies on Lassa fever and other group 4 microbial pathogens and has contributed to an improved genetic data resource for African populations.

Methods

Institutional review board ethical review and approval

This work was approved by the following institutional review boards and local ethics committees: Nigerian National Health Research Ethics Committee and ISTH (ISTH/HREC/20170915/22), Sierra Leone Ethics and Scientific Review Committee (070716), Tulane University Human

Research Protections Office (10-191330) and Harvard University Area Committee on the Use of Human Subjects (19-0023). Enrolment procedures and sampling efforts were carried out at Irrua Specialist Teaching Hospital (ISTH), Kenema Government Hospital (KGH) (IRB 070716) and their surrounding communities with participant consent or through a waiver of consent granted by the appropriate institutional review board/local ethics committee. Some samples shared with the study collaboration include those stored at the respective hospitals as clinical excess or approved for secondary use.

Lassa fever case definition and recruitment

ISTH, Nigeria. We recruited people with Lassa fever at ISTH between 2011 and 2014 and between 2016 and 2018 with a gap from 2014 to 2016 due to the Ebola outbreak in West Africa that temporarily halted research operations. We performed molecular diagnostic testing for all individuals suspected to have LASV who met clinical diagnostic criteria for Lassa fever including fever $>38^{\circ}\text{C}$ for less than 3 weeks, absence of signs of local inflammation, absence of clinical response to anti-malarials and additional major and minor signs⁶³. Individuals suspected to have LASV who were positive by molecular diagnostic testing were recruited to the study following informed consent.

KGH, Sierra Leone. People with Lassa fever were recruited at KGH between 2011 and 2018 with a gap from 2015 to 2016 due to the Ebola outbreak in West Africa. Individuals suspected to have LASV included those who met clinical diagnostic criteria for Lassa fever⁶³ and were positive by either ELISA for a LASV antigen or immunoglobulin M antibody against LASV^{25,64}. We performed virus sequencing from a subset of enrolled people with LASV¹². We only included data from individuals suspected to have LASV who were either antigen-ELISA positive or viral sequencing positive with reads per kilobase million of >1 in the GWAS.

Population control recruitment. Study staff at ISTH and KGH recruited population controls through outreach efforts to villages with a recent history of Lassa fever cases. Village controls (Supplementary Table 2) were healthy individuals who were recruited from the same household and/or village as people with LASV, prioritizing unrelated individuals where possible. Trio controls (Supplementary Table 2) were healthy families of mother, father and child from the Esan population in Nigeria and the Mende population in Sierra Leone who were recruited jointly with phase 3 of the IKG⁵¹. The informed consent criteria for this project were developed by the Samples and Ethical, Legal and Social Implications Group of the National Human Genome Research Institute⁵¹ and extends to the analyses we carried out in this study.

See Supplementary Note for more details about real-time quantitative PCR, sequencing and ELISA assays.

DNA extraction and genotyping

For all consenting study participants, we extracted buffy coats from the diagnostic blood draw after they were spun at $1,500\text{ g}$ for 10 min. We collected the buffy coat into a 1.5 ml tube, extracted DNA using the Qiagen DNeasy kit following manufacturer's instructions and shipped DNA samples to the Broad Institute.

For samples collected between 2011 and 2014, genotyping was performed at the Broad Institute's Genomics Platform on either the Infinium Omni 2.5 M or the Omni 5 M arrays. For samples collected after 2015, genotyping was performed at Illumina in San Diego on the H3Africa array.

Variant preprocessing and genome-wide association

See Supplementary Note for detailed description of variant preprocessing, principal component analyses, GWAS analysis and meta-analysis. Briefly, we first filtered variants that showed significantly different calls across genotyping arrays. We then merged the remaining samples into a single VCF file and ran imputation using the Sanger Imputation

Service⁶⁵ and EAGLE2 v2.0.5 for phasing⁶⁶ using the African Genome Resources reference panel.

We conducted all genetic association tests using mixed models logistic regression as implemented in version 1.2.0 of SAIGE³⁴ using the leave-one-chromosome-out option. We used genotyped variants that passed quality control filters to compute PCs and the genetic relatedness matrix. We used sex, array (H3Africa versus Infinium Omni) and PCs as covariates. We used METAL (version corresponding to 25 March 2011 release)⁶⁷ to meta-analyse the results of the Nigeria and Sierra Leone cohorts using the default option of weighting each cohort by sample size.

MPRA

See Supplementary Note for details on MPRA methods.

LARGE1 haplotype analysis

To define the LARGE-LRH, we extracted phased imputed genotype data from our cohort for the region on chromosome 22 between base pairs 33,870,000 and 34,470,000 in GRCh37, which corresponds to the previously defined region of the haplotype¹⁷. We then filtered out variants with minor allele frequency below 0.05 and clustered the corresponding haplotypes using *K*-means as implemented in Scikit-learn version 0.21.3 with *K* = 2. We identified individuals who were homozygous (coded as 2), heterozygous (coded as 1) or had 0 copies of the haplotype (coded as 0) and tested for association with Lassa fever phenotypes using SAIGE as described above and in the Supplementary Note.

To tag individuals from the 1KGP dataset who were carrying the LARGE-LRH, we identified the five SNPs that were most correlated with the clustering-defined haplotype in our dataset based on Pearson correlation. These were rs59015613, rs16993014, rs4525791, rs8135517 and rs59594190, all of which had a Pearson correlation >0.92 with the LARGE-LRH. We then used the phased 1KGP data to label haplotypes as the LARGE-LRH if three or more of the linked tag SNPs were present. The results were unchanged if we required only 2 or more linked SNPs to be present, and requiring 5/5 tag SNPs to be present only decreased the number of called haplotypes called from 252 to 250.

HLA sequencing, imputation and association analysis

Sequencing-based HLA typing. We performed sequencing-based HLA typing on samples from 297 Sierra Leone study participants. We generated sequencing libraries with the TruSight HLA v2 Sequencing Panel, following manufacturer's instructions, and sequenced the samples on Illumina Miseq instruments at either the Broad Institute, Boston, MA, or Scripps Institute, La Jolla, CA. We assigned HLA calls from the raw sequencing reads using the Assign 2.0 TruSight HLA Analysis Software.

HLA imputation. We developed an HLA imputation panel from 3,608 African Americans⁶⁸. This consisted of sequencing-based HLA calls for the *HLA-A*, *HLA-B*, *HLA-C*, *HLA-DPA1*, *HLA-DPB1*, *HLA-DQA1*, *HLA-DQB1* and *HLA-DRB1* genes, as well as SNP genotyping data from either the Affymetric Genome-Wide Human SNP Array 6.0 (2259) or the Infinium Omni 2.5 M array (1349). We imputed SNPs on chromosome 6 for these individuals using the same pipeline as for our GWAS cohort (Sanger Imputation Service with Eagle2 phasing and the African Genome Resources panel). We then subsetted to the HLA region (GRCh37 position between 28191116 and 34554976) and used the HIBAG version 1.22 software `hlaParallelAttrBagging` function to create an HLA reference index consisting of seven independent classifiers that could be used to predict HLA from imputed SNP inputs⁶⁹. We then used those indices with HIBAG's `hlaPredict` function to impute HLA types for our cohort.

We evaluated imputation accuracy against the sequence-based typing ground truth sets by calculating the percentage of alleles called correctly out of $2N$ where N is the total number of individuals in the ground-truth set. We excluded novel alleles from these calculations for the Sierra Leone set. We also estimated the accuracy of our imputation for *HLA-A*, *HLA-B*, *HLA-C*, *HLA-DQB1* and *HLA-DRB1* for separate

dataset of 76 Mende and 84 Esan individuals from the 1KGP who were genotyped in our cohort and HLA-typed by Gourraud et al.³³.

HLA association analysis. We calculated dosages for each allele by summing the posterior probabilities for each genotype output by HIBAG that contained the allele. We only included alleles with minor allele frequency above 1% in a cohort for association analysis. We then used the same mixed logistic regression model as for the SNP-based GWAS to associate the HLA alleles with Lassa fever phenotypes, using the dosage for each allele as the predictor and using sex and PCs as fixed effect covariates.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Raw de-identified genetic data from this study have been submitted to the European Genome-Phenome Archive (dataset IDs [EGAD00010002510](https://ega-archive.org/datasets/EGAD00010002510) and [EGAD00010002509](https://ega-archive.org/datasets/EGAD00010002509)). The vcf file containing these data can be accessed by registering an account with EGA (<https://ega-archive.org/register/>) and making a request to the Data Access Committee, following which a download will be made available to the account holder.

Summary statistics for genetic analyses reported in this study are available in the GWAS catalogue (<https://www.ebi.ac.uk/gwas/>) under accession codes GCST90301246, GCST90301247, GCST90301248 and GCST90301249. Meta-analyses of the GWASs are available in Supplementary Tables 1 and 2. Summary statistics for the MPRA are included in Supplementary Tables 3 and 5. Data from the 1KGP are available at <https://www.internationalgenome.org/data/>. Genome assembly hg19 is available at https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_000001405.13/.

Code availability

Data analysis scripts employed in this manuscript are publicly available on GitHub at https://github.com/dylkot/lassa_fever_gwas.

References

- Merson, L. et al. Clinical characterization of Lassa fever: a systematic review of clinical reports and research to inform clinical trial design. *PLoS Negl. Trop. Dis.* **15**, e0009788 (2021).
- Kenmoe, S. et al. Systematic review and meta-analysis of the epidemiology of Lassa virus in humans, rodents and other mammals in sub-Saharan Africa. *PLoS Negl. Trop. Dis.* **14**, e0008589 (2020).
- McCormick, J. B. & Fisher-Hoch, S. P. Lassa fever. *Curr. Top. Microbiol. Immunol.* **262**, 75–109 (2002).
- Okogbenin, S. et al. Retrospective cohort study of Lassa fever in pregnancy, southern Nigeria. *Emerg. Infect. Dis.* **25**, 1494–1500 (2019).
- Happi, A. N. et al. Increased prevalence of Lassa fever virus-positive rodents and diversity of infected species found during human Lassa fever epidemics in Nigeria. *Microbiol. Spectr.* **10**, e0036622 (2022).
- Gire, S. K. et al. Epidemiology. Emerging disease or diagnosis?. *Science* **338**, 750–752 (2012).
- McCormick, J. B., Webb, P. A., Krebs, J. W., Johnson, K. M. & Smith, E. S. A prospective study of the epidemiology and ecology of Lassa fever. *J. Infect. Dis.* **155**, 437–444 (1987).
- Lassa Fever* (CDC, 2019); <https://www.cdc.gov/vhf/lassa/index.html>
- Radoshitzky, S. R. & de la Torre, J. C. Human pathogenic arenaviruses (Arenaviridae). in *Encyclopedia of Virology* 507 (Elsevier, 2019).

10. *Lassa Fever Annual Epidemiological Report for 2019* (European Centre for Disease Prevention and Control, 2021); <https://www.ecdc.europa.eu/en/publications-data/lassa-fever-annual-epidemiological-report-2019>
11. Okokhere, P. et al. Clinical and laboratory predictors of Lassa fever outcome in a dedicated treatment facility in Nigeria: a retrospective, observational cohort study. *Lancet Infect. Dis.* **18**, 684–695 (2018).
12. Andersen, K. G. et al. Clinical sequencing uncovers origins and evolution of Lassa virus. *Cell* **162**, 738–750 (2015).
13. Chapman, S. J. & Hill, A. V. S. Human genetic susceptibility to infectious disease. *Nat. Rev. Genet.* **13**, 175–188 (2012).
14. COVID-19 Host Genetics Initiative. Mapping the human genetic architecture of COVID-19. *Nature* **600**, 472–477 (2021).
15. Tian, C. et al. Genome-wide association and HLA region fine-mapping studies identify susceptibility loci for multiple common infections. *Nat. Commun.* **8**, 599 (2017).
16. Sabeti, P. C. et al. Genome-wide detection and characterization of positive selection in human populations. *Nature* **449**, 913–918 (2007).
17. Andersen, K. G. et al. Genome-wide scans provide evidence for positive selection of genes implicated in Lassa fever. *Philos. Trans. R. Soc. Lond. B* **367**, 868–877 (2012).
18. Jae, L. T. et al. Deciphering the glycosylome of dystroglycanopathies using haploid screens for Lassa virus entry. *Science* **340**, 479–483 (2013).
19. Kunz, S. et al. Posttranslational modification of alpha-dystroglycan, the cellular receptor for arenaviruses, by the glycosyltransferase LARGE is critical for virus binding. *J. Virol.* **79**, 14282–14296 (2005).
20. Raabe, V. & Koehler, J. Laboratory diagnosis of Lassa fever. *J. Clin. Microbiol.* **55**, 1629–1637 (2017).
21. Ackerman, H. et al. A comparison of case–control and family-based association methods: the example of sickle-cell and malaria. *Ann. Hum. Genet.* **69**, 559–565 (2005).
22. Hill, A. V. S. Aspects of genetic susceptibility to human infectious diseases. *Annu. Rev. Genet.* **40**, 469–486 (2006).
23. Bowen, M. D. et al. Genetic diversity among Lassa virus strains. *J. Virol.* **74**, 6992–7004 (2000).
24. Siddle, K. J. et al. Genomic analysis of Lassa virus during an increase in cases in Nigeria in 2018. *N. Engl. J. Med.* **379**, 1745–1753 (2018).
25. Boisen, M. L. et al. Field validation of recombinant antigen immunoassays for diagnosis of Lassa fever. *Sci. Rep.* **8**, 5939 (2018).
26. Johnson, K. M. et al. Clinical virology of Lassa fever in hospitalized patients. *J. Infect. Dis.* **155**, 456–464 (1987).
27. Cummins, D. et al. Acute sensorineural deafness in Lassa fever. *JAMA* **264**, 2093–2096 (1990).
28. McCormick, J. B. et al. A case–control study of the clinical diagnosis and course of Lassa fever. *J. Infect. Dis.* **155**, 445–455 (1987).
29. Monath, T. P. Lassa fever: review of epidemiology and epizootiology. *Bull. World Health Organ.* **52**, 577–592 (1975).
30. Shaffer, J. G. et al. Lassa fever in post-conflict Sierra Leone. *PLoS Negl. Trop. Dis.* **8**, e2748 (2014).
31. Klingström, J. & Ahlm, C. Sex, gender, and hemorrhagic fever viruses. in *Sex and Gender Differences in Infection and Treatments for Infectious Diseases* (eds Klein, S. L. & Roberts, C. W.) 211–230 (Springer International, 2015).
32. McCormick, J. B. Epidemiology and control of Lassa fever. *Curr. Top. Microbiol. Immunol.* **134**, 69–78 (1987).
33. Webb, P. A. et al. Lassa fever in children in Sierra Leone, West Africa. *Trans. R. Soc. Trop. Med. Hyg.* **80**, 577–582 (1986).
34. Zhou, W. et al. Efficiently controlling for case–control imbalance and sample relatedness in large-scale genetic association studies. *Nat. Genet.* **50**, 1335–1341 (2018).
35. Tucker, G., Price, A. L. & Berger, B. Improving the power of GWAS and avoiding confounding from population stratification with PC-Select. *Genetics* **197**, 1045–1049 (2014).
36. Kanai, M., Tanaka, T. & Okada, Y. Empirical estimation of genome-wide significance thresholds based on the 1000 Genomes Project data set. *J. Hum. Genet.* **61**, 861–866 (2016).
37. Eeles, R. A. et al. Multiple newly identified loci associated with prostate cancer susceptibility. *Nat. Genet.* **40**, 316–321 (2008).
38. Deeb, R., Veerapandian, A., Tawil, A. & Treidler, S. Variable penetrance of Andersen–Tawil Syndrome in a Caucasian family with a rare missense KCJN2 mutation (P3.450). *Neurology* **90**, Number 15_supplement (2018).
39. Howe, K. L. et al. Ensembl 2021. *Nucleic Acids Res.* **49**, D884–D891 (2021).
40. Sakabe, S., Witwit, H., Khafaji, R., Cubitt, B. & de la Torre, J. C. Chaperonin TRiC/CCT participates in mammarenavirus multiplication in human cells via interaction with the viral nucleoprotein. *J. Virol.* **97**, e0168822 (2023).
41. Sugita, S. et al. A stoichiometric complex of neurexins and dystroglycan in brain. *J. Cell Biol.* **154**, 435–445 (2001).
42. Mittal, R., Kumar, A., Ladda, R., Mainali, G. & Aliu, E. Pitt Hopkins-like syndrome 1 with novel CNTNAP2 mutation in siblings. *Child Neurol. Open* **8**, 2329048X211055330 (2021).
43. Song, J.-M. et al. Pathogenic GRM7 mutations associated with neurodevelopmental disorders impair axon outgrowth and presynaptic terminal development. *J. Neurosci.* **41**, 2344–2359 (2021).
44. Wang, J. et al. SARS-CoV-2 uses metabotropic glutamate receptor subtype 2 as an internalization factor to infect cells. *Cell Discov.* **7**, 119 (2021).
45. Wang, J. et al. Metabotropic glutamate receptor subtype 2 is a cellular receptor for rabies virus. *PLoS Pathog.* **14**, e1007189 (2018).
46. Rogoz, K. et al. Identification of a neuronal receptor controlling anaphylaxis. *Cell Rep.* **14**, 370–379 (2016).
47. Klotz, L. & Enz, R. MGLuR7 is a presynaptic metabotropic glutamate receptor at ribbon synapses of inner hair cells. *FASEB J.* **35**, e21855 (2021).
48. Mateer, E. J., Huang, C., Shehu, N. Y. & Paessler, S. Lassa fever-induced sensorineural hearing loss: a neglected public health and social burden. *PLoS Negl. Trop. Dis.* **12**, e0006187 (2018).
49. Christianson, J., Oxford, J. T. & Jorcyk, C. L. Emerging perspectives on leukemia inhibitory factor and its receptor in cancer. *Front. Oncol.* **11**, 693724 (2021).
50. Tewhey, R. et al. Direct identification of hundreds of expression-modulating variants using a multiplexed reporter assay. *Cell* **165**, 1519–1529 (2016).
51. 1000 Genomes Project Consortium et al. A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
52. Jia, X. et al. Imputing amino acid polymorphisms in human leukocyte antigens. *PLoS ONE* **8**, e64683 (2013).
53. Gourraud, P.-A. et al. HLA diversity in the 1000 genomes dataset. *PLoS ONE* **9**, e97282 (2014).
54. Foronjy, R. F., Dabo, A. J., Cummins, N. & Geraghty, P. Leukemia inhibitory factor protects the lung during respiratory syncytial viral infection. *BMC Immunol.* **15**, 41 (2014).
55. Tjernlund, A. et al. Early induction of leukemia inhibitor factor (LIF) in acute HIV-1 infection. *AIDS* **20**, 11–19 (2006).
56. Waring, P. M., Waring, L. J. & Metcalf, D. Circulating leukemia inhibitory factor levels correlate with disease severity in meningococemia. *J. Infect. Dis.* **170**, 1224–1228 (1994).
57. GTEx Consortium. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* **369**, 1318–1330 (2020).

58. Manolio, T. A. et al. Finding the missing heritability of complex diseases. *Nature* **461**, 747–753 (2009).
59. Palmer, C. & Pe'er, I. Statistical correction of the winner's curse explains replication variability in quantitative trait genome-wide association studies. *PLoS Genet.* **13**, e1006916 (2017).
60. Botti-Lodovico, Y. et al. The origins and future of sentinel: an early-warning system for pandemic preemption and response. *Viruses* **13**, 1605 (2021).
61. Lauck, M. et al. GB virus C coinfections in west African Ebola patients. *J. Virol.* **89**, 2425–2429 (2015).
62. Lambert, C. A. & Tishkoff, S. A. Genetic structure in African populations: implications for human demographic history. *Cold Spring Harb. Symp. Quant. Biol.* **74**, 395–402 (2009).
63. Khan, S. H. et al. New opportunities for field research on the pathogenesis and treatment of Lassa fever. *Antivir. Res.* **78**, 103–115 (2008).
64. Branco, L. M. et al. Emerging trends in Lassa fever: redefining the role of immunoglobulin M and inflammation in diagnosing acute infection. *Virol. J.* **8**, 478 (2011).
65. The Haplotype Reference Consortium. A reference panel of 64,976 haplotypes for genotype imputation. *Nat. Genet.* **48**, 1279–1283 (2016).
66. Loh, P.-R. et al. Reference-based phasing using the Haplotype Reference Consortium panel. *Nat. Genet.* **48**, 1443–1448 (2016).
67. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190–2191 (2010).
68. Ollila, H. et al. Transethnic analysis of high-resolution HLA alleles and complement 4 structural polymorphisms in schizophrenia. *Eur. Neuropsychopharmacol.* **29**, S937 (2019).
69. Zheng, X. et al. HIBAG-HLA genotype imputation with attribute bagging. *Pharmacogenomics J.* **14**, 192–200 (2014).
70. Fichet-Calvet, E. & Rogers, D. J. Risk maps of Lassa fever in West Africa. *PLoS Negl. Trop. Dis.* **3**, e388 (2009).

Acknowledgements

This work was supported by National Institutes of Health grants R01AI114855 (P.C.S.), 1DP2OD006514 (P.C.S.), HHSN272201000022C (P.C.S.), U01HG007480 (C.T.H.), German Research Foundation grants GU 883/1-1 (S.G.), GU 883/4-1 (S.G.) and GU 883/4-2 (S.G.), and the Howard Hughes Medical Institute (P.C.S.). D.K. was supported by award number T32GM007753 from the National Institute of General Medical Sciences. S.R. was supported by the FujiFilm Fellowship from Harvard Medical School. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the paper. In memoriam: three co-authors passed away before the manuscript could be published: Stefan Kunz, Michael B. A. Oldstone and S. Humarr Khan. We wish to honour their memory.

Author contributions

D.K., S.R., S.T., I.O., A.G., M.M., J.D.S., E.P., R. Tariyal, P.E.E., M.S., S.J., S.K.G., S.F.S., E.K.K., N.Y., S.H., L.M.M., K.R., S. Kunz, J.B.M., S.H.K., A. N. Honko, E.S.L., M.B.A.O., L.H., R. Tewhey, P.O.O., J.S.S., K.G.A., S.K.R.,

D.S.G., R.F.G., K.G.B., C.T.H. and P.C.S. conceived and designed the experiments. D.K., S.R., S.T., I.O., A.G., M.M., J.D.S., P.E.E., R.R.-S., K.J.S., S.J., S.K.G., S.W., M.P., L.M.B., L.K., N.B., A.G.-Y., O.O., P.-P.J., G.O.A., D.A.A., S. Kales, O.A.F., S.G., H.M.O., R.T., J.S.S., K.G.A., S.K.R., R.F.G., K.G.B. and C.T.H. performed the experiments. D.K., S.R., S.T., P.N., S.M., B.C., S.F.S., M.P., S.R.C., S.G.K., L.M.B., J.J.V., N.B., A.G.-Y. and P.-P.J. analysed the data. M.P., S.G.K., L.M.B., L.K., J.J.V., A. N. Happi, C.O.I., M.D.-I., K.I., O.O.O., A.K.O., S.A.O., S.G., H.M.O., R. Tewhey, P.O.O., J.S.S., S.K.R. and P.C.S. contributed materials and analysis tools. D.K., S.R., S.T., P.N., J.S.S., R.F.G., K.G.B., C.T.H. and P.C.S. wrote the paper.

Competing interests

P.C.S., R. Tewhey and S.K.R. are inventors on patents related to massively parallel reporter assays. P.C.S. is a co-founder of, shareholder in and consultant to Sherlock Biosciences, Inc. and Delve Bio, as well as a Board member of and shareholder in Danaher Corporation. The other authors declare no competing interests.

Additional information

Extended data is available for this paper at <https://doi.org/10.1038/s41564-023-01589-3>.

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41564-023-01589-3>.

Correspondence and requests for materials should be addressed to Dylan Kotliar, Christian T. Happi or Pardis C. Sabeti.

Peer review information *Nature Microbiology* thanks David Hinds, Michael Levin and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.












Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

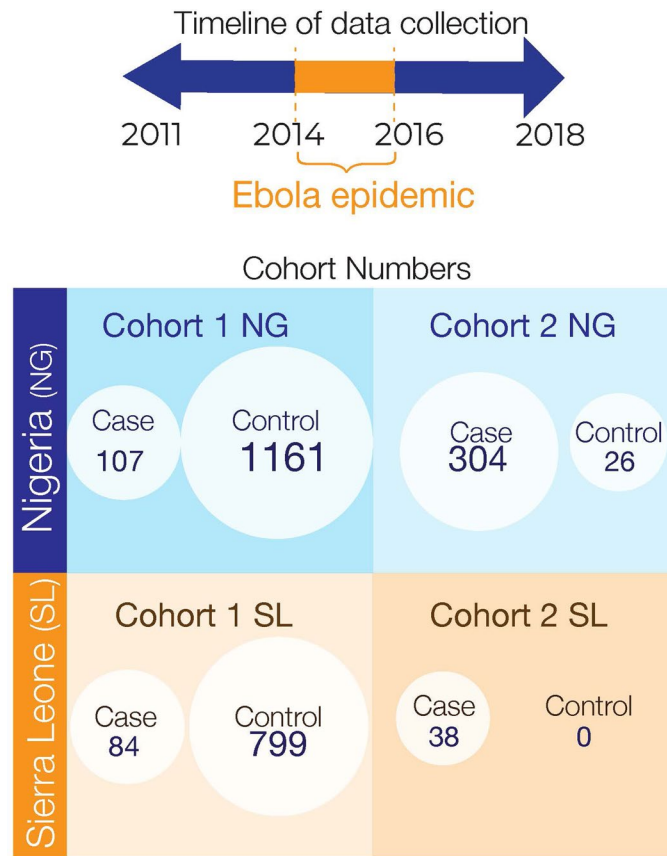
Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024

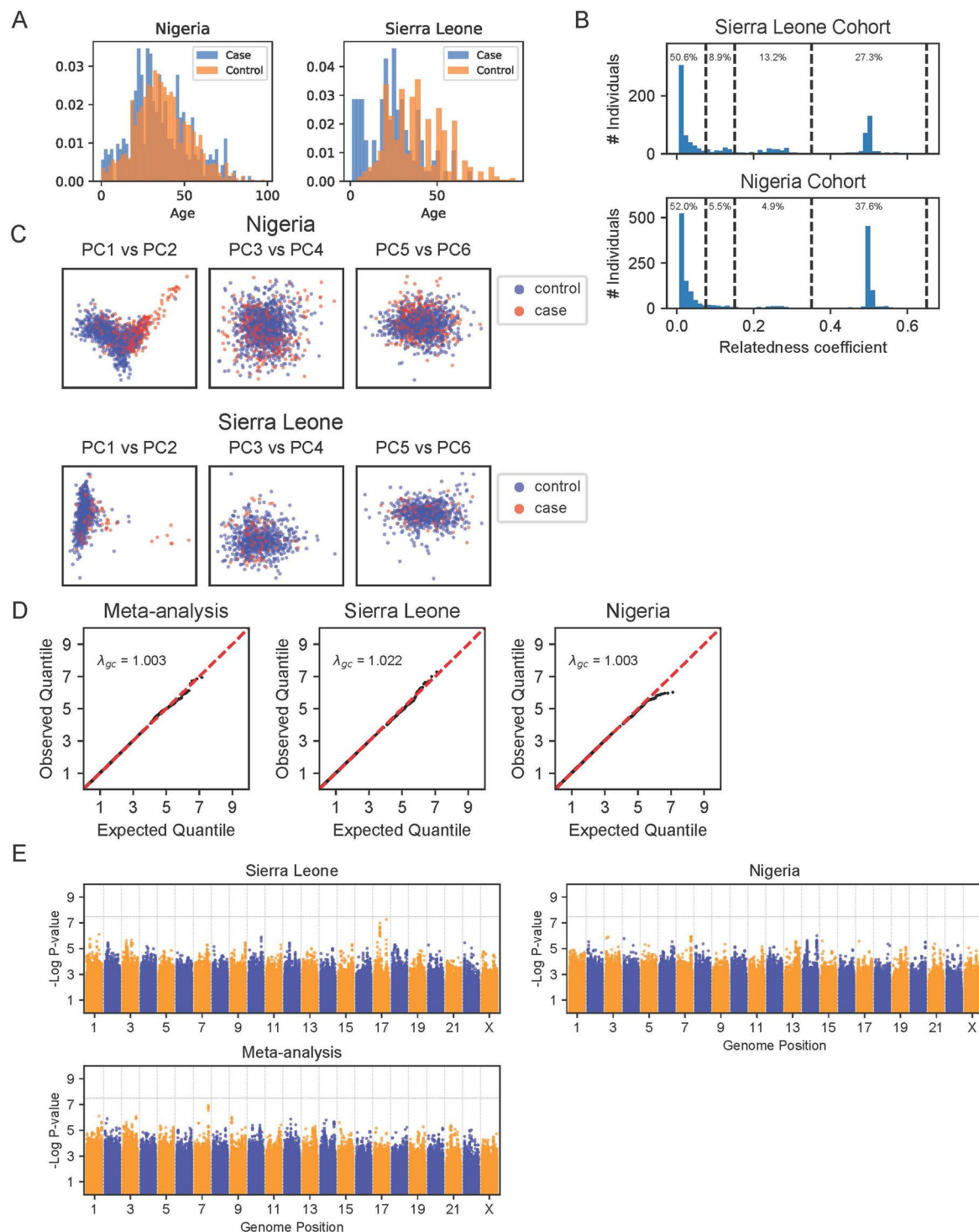
Dylan Kotliar^{1,2,3,53}✉, Siddharth Raju^{1,2,53}, Shervin Tabrizi^{1,4,5,53}, Ikponmwsa Odia^{6,53}, Augustine Goba⁷, Mambu Momoh^{7,8}, John Demby Sandi⁷, Parvathy Nair⁹, Eric Phelan¹⁰, Ridhi Tariyal¹¹, Philomena E. Eromon^{6,12}, Samar Mehta¹³, Refugio Robles-Sikisaka¹⁴, Katherine J. Sidle¹, Matt Stremelau¹⁵, Simbirie Jalloh⁷, Stephen K. Gire¹¹, Sarah Winnicki¹, Bridget Chak¹⁶, Stephen F. Schaffner^{1,17,18}, Matthias Pauthner¹⁹, Elinor K. Karlsson^{1,20,21}, Sarah R. Chapin¹, Sharon G. Kennedy^{1,22}, Luis M. Branco²³, Lansana Kanneh²⁴, Joseph J. Vitti¹, Nisha Broodie²⁵, Adrienne Gladden-Young²⁶, Omowunmi Omoniwa²⁷, Pan-Pan Jiang²⁸, Nathan Yozwiak²⁹, Shannon Heuklom³⁰, Lina M. Moses³¹, George O. Akpede^{6,32}, Danny A. Asogun³³, Kathleen Rubins³⁴, Susan Kales³⁵, Anise N. Happi¹², Christopher O. Iruolagbe³⁶, Mercy Dic-Ijiewere³⁶, Kelly Iraoyah³⁶, Omoregie O. Osazuwa³⁶, Alexander K. Okonkwo³⁶, Stefan Kunz^{37,55}, Joseph B. McCormick³⁸, S. Humarr Khan^{7,55}, Anna N. Honko³⁹, Eric S. Lander^{1,2,40}

Michael B. A. Oldstone ^{14,55}, **Lisa Hensley**⁴¹, **Onikepe A. Folarin** ^{12,42}, **Sylvanus A. Okogbenin**⁶, **Stephan Günther** ⁴³, **Hanna M. Ollila**^{1,44,45,46}, **Ryan Tewhey** ^{35,54}, **Peter O. Okokhere**^{6,32,36,54}, **John S. Schieffelin**^{47,54}, **Kristian G. Andersen** ^{14,54}, **Steven K. Reilly** ^{48,54}, **Donald S. Grant** ^{7,24,54}, **Robert F. Garry** ^{49,54}, **Kayla G. Barnes** ^{1,18,50,51,54}, **Christian T. Happi** ^{12,18,42,54} ✉ & **Pardis C. Sabeti** ^{1,9,17,18,22,52,54} ✉

¹Broad Institute of Massachusetts Institute of Technology (MIT) and Harvard, Cambridge, MA, USA. ²Department of Systems Biology, Harvard Medical School, Boston, MA, USA. ³Department of Internal Medicine, Brigham and Women's Hospital, Boston, MA, USA. ⁴Department of Radiation Oncology, Massachusetts General Hospital, Boston, MA, USA. ⁵Koch Institute for Integrative Cancer Research, Massachusetts Institute of Technology, Cambridge, MA, USA. ⁶Institute of Lassa Fever, Research and Control, Irrua Specialist Teaching Hospital, Irrua, Nigeria. ⁷College of Medicine and Allied Health Sciences, University of Sierra Leone, Freetown, Sierra Leone. ⁸Eastern Polytechnic College, Kenema, Sierra Leone. ⁹Howard Hughes Medical Institute, Chevy Chase, MD, USA. ¹⁰Prospr at Work Inc., Berlin, Germany. ¹¹NextGen Jane, Inc., Oakland, CA, USA. ¹²African Centre of Excellence for Genomics of Infectious Diseases (ACEGID), Redeemer's University, Ede, Nigeria. ¹³Department of Critical Care Medicine, University of Maryland Medical Center, Baltimore, MA, USA. ¹⁴Department of Immunology and Microbiology, The Scripps Research Institute, La Jolla, CA, USA. ¹⁵Equator Labs Incorporated, Washington, DC, USA. ¹⁶Biological Sciences Division, University of Chicago, Chicago, IL, USA. ¹⁷Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA, USA. ¹⁸Department of Immunology and Infectious Diseases, Harvard T.H. Chan School of Public Health, Harvard University, Boston, MA, USA. ¹⁹Vir Biotechnology, San Francisco, CA, USA. ²⁰Genomics and Computational Biology, UMass Chan Medical School, Worcester, MA, USA. ²¹Program in Molecular Medicine, UMass Chan Medical School, Worcester, MA, USA. ²²Department of Medicine, Massachusetts General Hospital, Boston, MA, USA. ²³Zalgen Labs, Frederick, MD, USA. ²⁴Viral Hemorrhagic Fever Program, Kenema Government Hospital, Ministry of Health and Sanitation, Kenema, Sierra Leone. ²⁵New York-Presbyterian Hospital-Columbia and Cornell, New York, NY, USA. ²⁶Molecular Microbiology, Graduate School of Biomedical Sciences, Tufts University, Boston, MA, USA. ²⁷Malaria Consortium, Abuja, Nigeria. ²⁸Google Medical Brain, Mountain View, CA, USA. ²⁹Gene and Cell Therapy Institute, Mass General Brigham, Cambridge, MA, USA. ³⁰San Francisco Community Health Center, San Francisco, CA, USA. ³¹Tulane University School of Public Health and Tropical Medicine, New Orleans, LA, USA. ³²Department of Medicine, Ambrose Alli University, Ekpoma, Nigeria. ³³Department of Community Medicine, Ambrose Alli University, Ekpoma, Nigeria. ³⁴National Aeronautics and Space Administration, Houston, TX, USA. ³⁵The Jackson Laboratory, Bar Harbor, ME, USA. ³⁶Department of Medicine, Irrua Specialist Teaching Hospital, Irrua, Nigeria. ³⁷Institute of Microbiology, University Hospital Center and University of Lausanne, Lausanne, Switzerland. ³⁸UTHealth Houston School of Public Health, Brownsville Campus, Brownsville, TX, USA. ³⁹Boston University School of Medicine, Boston, MA, USA. ⁴⁰Department of Biology, Massachusetts Institute of Technology (MIT), Cambridge, MA, USA. ⁴¹National Institutes of Health Integrated Research Facility, Frederick, MA, USA. ⁴²Department of Biological Sciences, Redeemer's University, Ede, Nigeria. ⁴³Bernhard Nocht Institute for Tropical Medicine, Hamburg, Germany. ⁴⁴Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki, Finland. ⁴⁵Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA, USA. ⁴⁶Anesthesia, Critical Care, and Pain Medicine, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA. ⁴⁷Section of Infectious Disease, Department of Pediatrics, Tulane University School of Medicine, New Orleans, LA, USA. ⁴⁸Department of Genetics, Yale School of Medicine, New Haven, CT, USA. ⁴⁹Tulane University School of Medicine, New Orleans, LA, USA. ⁵⁰Malawi-Liverpool-Wellcome Trust Clinical Research Programme, Kamuzu University of Health Sciences, Blantyre, Malawi. ⁵¹Department of Vector Biology and Tropical Disease Biology, Liverpool School of Tropical Medicine, Liverpool, UK. ⁵²Massachusetts Consortium on Pathogen Readiness, Boston, MA, USA. ⁵³These authors contributed equally: Dylan Kotliar, Siddharth Raju, Shervin Tabrizi, Ikponmwoa Odia. ⁵⁴These authors jointly supervised this work: Ryan Tewhey, Peter O. Okokhere, John S. Schieffelin, Kristian G. Andersen, Steven K. Reilly, Donald S. Grant, Robert F. Garry, Kayla G. Barnes, Christian T. Happi, Pardis C. Sabeti. ⁵⁵Deceased: Stefan Kunz, Michael B. A. Oldstone, S. Humarr Khan. ✉e-mail: dkotliar@mgb.org; happic@run.edu.ng; pardis@broadinstitute.org



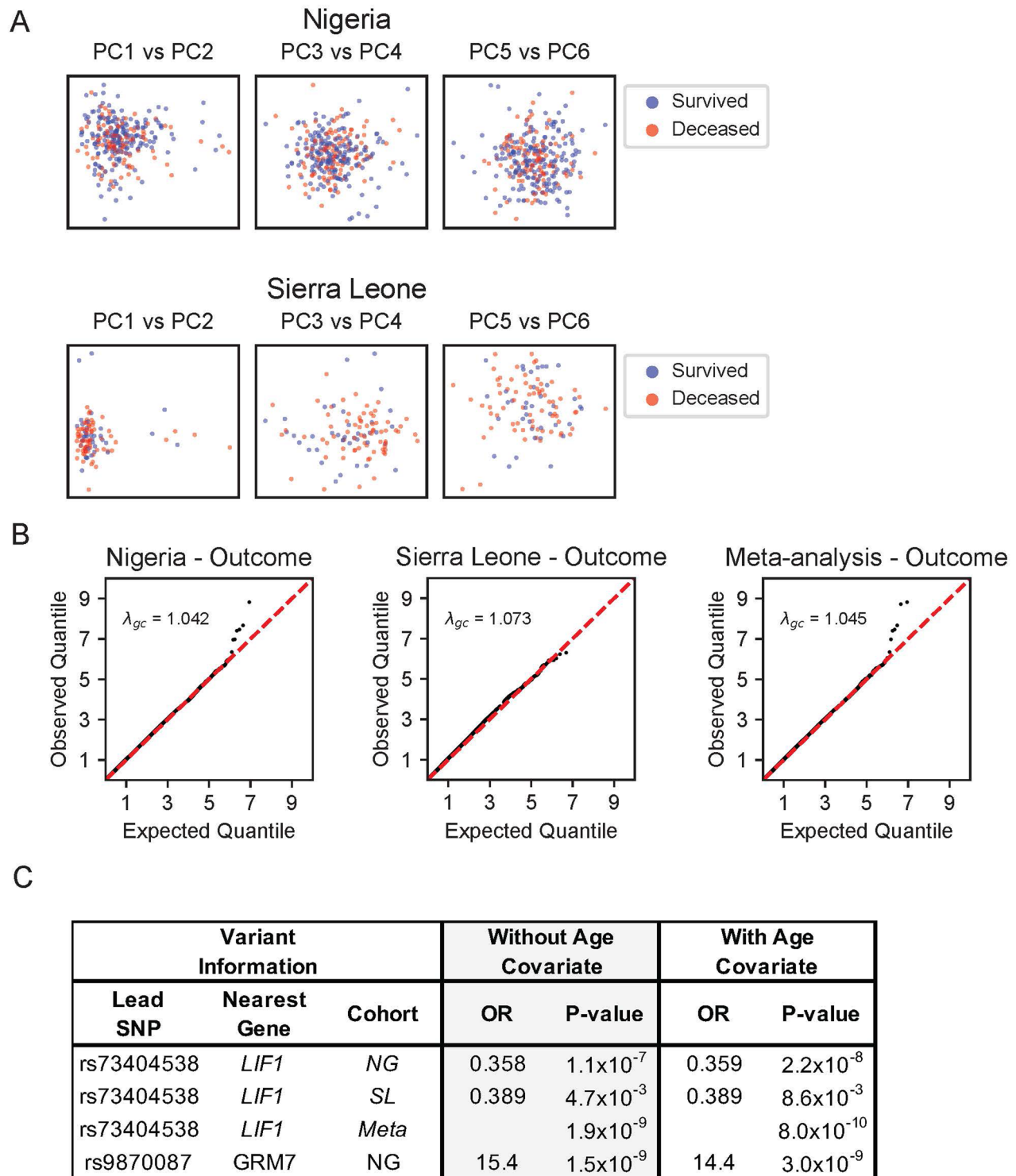
Extended Data Fig. 1 | Timeline of cohort recruitment in each country. Breakdown of enrolled patients by country, cohort, and disease status.



Extended Data Fig. 2 | Quality control analyses for the susceptibility GWAS.

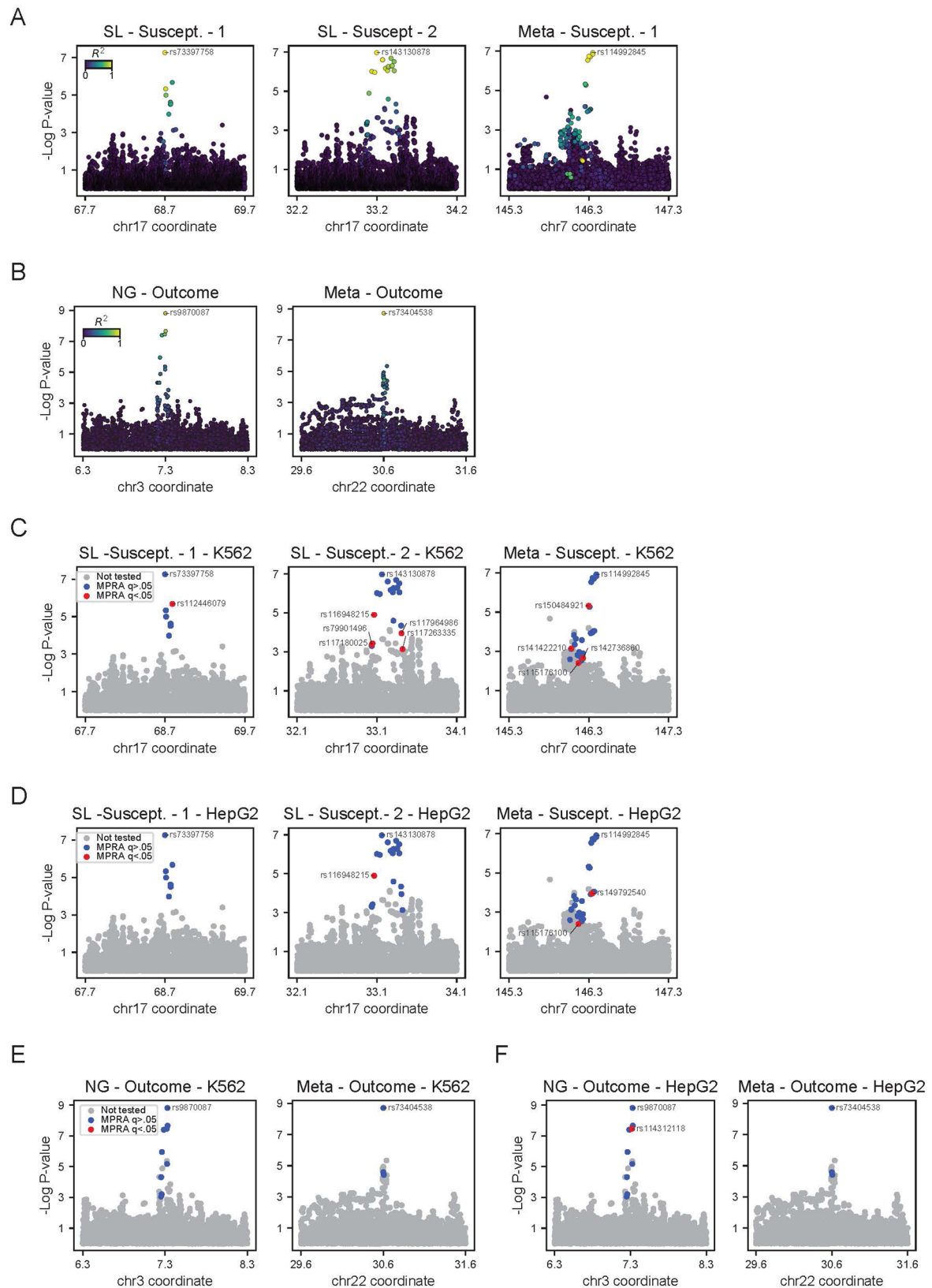
(A) Histogram of ages in the Nigeria and Sierra Leone cohorts, separated by case/control status. (B) Histogram of the maximum relatedness coefficient between each individual and all other individuals in the Nigerian (NG) and Sierra Leonean (SL) cohorts. (C) Principal component analysis (PCA) of the NG and SL cohorts, colored by case-control status. PCs were computed on unrelated individuals and then all individuals were projected onto those components (Methods).

(D) Quantile-quantile plots of $-\log_{10}$ P-values from the susceptibility GWAS against expected quantiles. (E) Manhattan plots showing the $-\log_{10}$ P-value for each genomic variant for the LF susceptibility associations. P-values in D and E are based on saddlepoint-approximated score tests (SAIGE), while meta-analysis P-values are derived from meta-analysis (METAL) of P-values generated from each cohort.



Extended Data Fig. 3 | Quality control analyses for the GWAS of LF clinical outcome. (A) Principal component analysis (PCA) of the NG and SL cohorts, colored by clinical outcome. PCs were computed on unrelated individuals, and then all individuals were projected onto those components. (B) Quantile-quantile plots of $-\log_{10}$ P-values from the outcome GWAS against expected

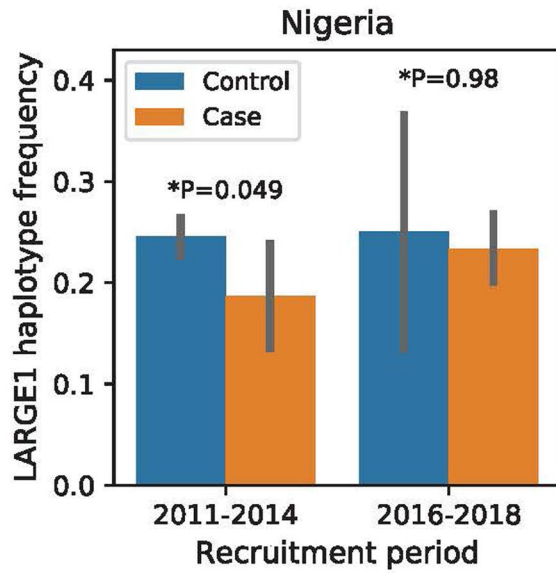
quantiles. (C) Comparison of the outcome GWAS lead variants with and without inclusion of age as a covariate. P-values in B and C are based on saddlepoint-approximated score tests (SAIGE), while meta-analysis P-values are derived from meta-analysis (METAL) of P-values generated from each cohort. Odds ratios are computed from Firth logistic regression.



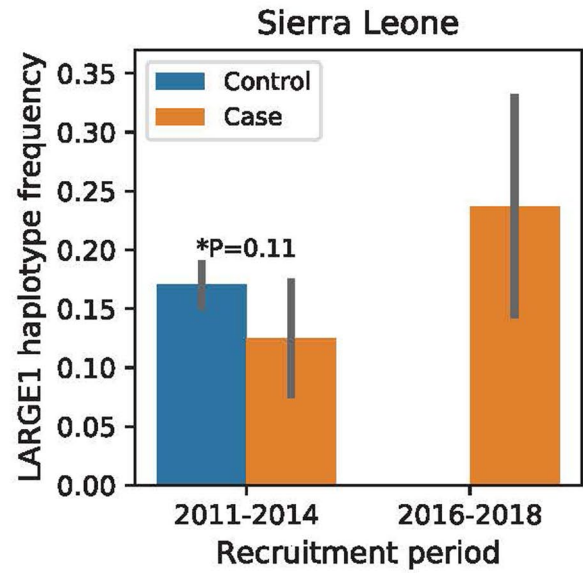
Extended Data Fig. 4 | MPRA analyses of the susceptibility and outcome GWAS peaks. (A) Scatter plot of lead susceptibility GWAS loci described in the main text showing chromosomal position against $-\log_{10}$ association P-value. Variants are colored by the linkage disequilibrium (LD) coefficient of determination R^2 between each variant and the most significant 'lead' variant in the locus. (B) Same as A but for the lead variants in the fatal outcome GWAS.

(C-F) Same as A and B but colored by whether the variant showed statistically significant skew (q -value < 0.05) in the massively parallel reporter assay in the K562 cell line (C and E) or HepG2 cell line (D and F). P-values are based on saddlepoint-approximated score tests (SAIGE), while meta-analysis P-values are derived from meta-analysis (METAL) of P-values generated from each cohort.

A



B



Extended Data Fig. 5 | LARGE1 haplotype association by recruitment period. (A, B) Frequencies of the long-range LARGE1 haplotype by the period of recruitment as well as by case-control status for Nigeria (A) and Sierra Leone (B).

P-values are from mixed logistic models association testing within the indicated recruitment period. Error bars represent 95% bootstrap confidence intervals for allele frequency. N for each cohort within each country is defined in Table S2.

Extended Data Table 1 | Summary of GWAS collections for the Nigerian and Sierra Leonean cohorts

Country	Phenotype	Total	Female (%)	Male (%)	Age Mean	Age SD	Deceased (%)	Survived (%)	Unknown (%)
Nigeria	Control	1187	497 (41.9)	690 (58.1)	37.2	16.1			
	Case	411	242 (58.9)	169 (41.1)	35.2	17.9	107 (26.0)	196 (47.7)	108 (26.3)
Sierra Leone	Control	799	324 (40.6)	475 (59.4)	38.8	16.6			
	Case	122	50 (41.0)	72 (59.0)	24.1	14.9	70 (57.4)	38 (31.1)	14 (11.5)

Includes breakdown of samples by sex, includes age mean and SD for each sample set, and breakdown of clinical outcome for cases.

Extended Data Table 2 | Detailed summary of GWAS collections

Cohort		Count	Collection Period		Genotyping Array			Control Breakdown		Cases Diagnostics				
Country	Phenotype	Total	2011-2015 (%)	2016-2018 (%)	H3 (%)	Omni2.5M (%)	Omni5M (%)	Village Controls	Trio Controls	qPCR+ (%)	Antigen+ (%)	Seq+ (%)	Both+ (%)	
Nigeria	Control	1187	1161 (97.8)	26 (2.2)	44 (3.7)	756 (63.7)	387 (32.6)	638 (53.7)	549 (46.3)	119 (29.0)		114 (27.7)	178 (43.3)	
	Case	411	107 (26.0)	304 (74.0)	317 (77.1)	91 (22.1)	3 (0.7)							
Sierra Leone	Control	799	799 (100.0)	0 (0.0)	46 (5.8)	753 (94.2)	0 (0.0)	604 (75.6)	195 (24.4)			80 (65.6)	11 (9.0)	31 (25.4)
	Case	122	84 (68.9)	38 (31.1)	38 (31.1)	84 (68.9)	0 (0.0)							

Includes breakdown of samples by collection time period, genotyping array, split of controls into village and trio recruitments, and diagnostic categories of cases. For the case diagnostic category, Nigerian cases were positive by RT-qPCR (qPCR+) and/or sequencing (Seq+), whereas Sierra Leonean cases were positive by antigen ELISA (Antigen+) and/or sequencing. The last column Both+ specifies the number of cases who were positive by both sequencing and RT-qPCR or ELISA.

Extended Data Table 3 | Overview of clinical symptoms

Symptom/Sign	Sierra Leone Cohort							Nigeria Cohort								
	Symptom Frequency By Age (%)						Statistics		Symptom Frequency By Age (%)						Statistics	
	0-9 (N=28)	10-19 (N=25)	20-29 (N=36)	30-39 (N=17)	40+ (N=16)	Total (N=122)	Z	P	0-9 (N=6)	10-20 (N=8)	20-30 (N=26)	30-40 (N=31)	40+ (N=38)	Total (N=109)	Z	P
Weakness	96.3	82.6	75	80	73.3	82.1	-2.01	0.045	16.7	25	53.8	45.2	42.1	43.1	0.78	0.435
Cough	96.3	78.3	75	53.3	66.7	76.8	-3.21	0.001	33.3	37.5	34.6	22.6	18.4	25.7	-1.77	0.077
Headache	63	87	75	73.3	93.3	76.8	1.87	0.062	16.7	12.5	76.9	61.3	50	55	0.6	0.546
Vomiting	81.5	73.9	56.2	66.7	46.7	66.1	-2.5	0.012	83.3	87.5	57.7	58.1	47.4	57.8	-2.4	0.016
Sore throat	55.6	69.6	62.5	60	60	61.6	-0.15	0.879	16.7	25	26.9	29	21.1	24.8	0.03	0.976
Abdominal pain	44.4	65.2	50	60	40	51.8	-0.04	0.967	50	37.5	46.2	58.1	47.4	49.5	0.29	0.769
Diarrhea	55.6	56.5	46.9	40	46.7	50	-1.47	0.141	16.7	12.5	30.8	16.1	21.1	21.1	-0.12	0.901
Fever	54.5	52.6	48	27.3	50	48.3	-1.18	0.237	33.3	87.5	42.3	58.1	47.4	51.4	-0.17	0.868
Bleeding	48.1	47.8	37.5	40	26.7	41.1	-1.35	0.177	16.7	25	30.8	29	15.8	23.9	-1.35	0.178
Swelling	40.7	34.8	37.5	33.3	6.7	33	-1.39	0.163								
Jaundice	3.7	8.7	3.1	26.7	6.7	8	1.26	0.207								
Injected conjun.									0	0	0	9.7	10.5	6.4	1.08	0.282
	0-9 (N=24)	10-19 (N=15)	20-29 (N=30)	30-39 (N=18)	40+ (N=21)	Total (N=108)	Z	P	0-9 (N=19)	10-19 (N=36)	20-29 (N=76)	30-39 (N=66)	40+ (N=98)	Total (N=295)	Z	P
Fatal outcome	45.8	73.3	66.7	72.2	71.4	64.8	1.36	0.174	26.3	19.4	36.8	36.4	42.9	35.3	1.61	0.107

Percentage of cases with a clinical sign or symptom at the time of admission, stratified by age. Below each age range is the number of individuals in that group with clinical data available. We report the large-sample approximation test statistic (Z) and P-value (P) for a Wilcoxon Ranksum test comparing the median age of subjects with and without each symptom. Conjunctival injection was recorded for the NG cohort but not the SL cohort, and lower extremity swelling or jaundice were recorded for the SL cohort but not the NG cohort. Bleeding includes any observed bleeding such as epistaxis, hematemesis, hematuria, melena, and hematochezia. Fever is defined as a temperature on admission of greater than 37.8 degrees celsius.

Extended Data Table 4 | Comparison of lead variants between the outcome and susceptibility GWAS analyses

Variant Information			Susceptibility GWAS					Outcome GWAS				
Lead SNP	Chrom	Position (hg19)	Nigeria OR	Nigeria P-value	Sierra Leone OR	Sierra Leone P-value	Meta-analysis P-value	Nigeria OR	Nigeria P-value	Sierra Leone OR	Sierra Leone P-value	Meta-analysis P-value
rs114992845	7	146356694	9.19	2.7x10 ⁻⁶	4.77	0.010	1.2x10 ⁻⁷	1.14*	0.82*	8.24*	0.098*	0.18*
rs143130878	17	33192408	1.20	0.64	6.87	1.1x10 ⁻⁷	3.3x10 ⁻⁴	1.16	0.75	0.70*	0.56*	0.85*
rs73397758	17	68745251	0.84	0.58	9.16	5.5x10 ⁻⁸	4.8x10 ⁻³	0.56	0.14	22.6*	1.8x10 ^{-3*}	0.25*
rs73404538	22	30619983	0.83	0.18	0.71	0.039	0.021	0.36	1.1x10 ⁻⁷	0.39	4.7x10 ⁻³	1.1x10 ⁻⁹
rs9870087	3	7330265	0.72	0.27	1.38*	0.46*	0.67*	15.4	1.5x10 ⁻⁹	0.64*	0.55*	1.1x10 ^{-6*}

Displays odds ratios (OR) and P-values for lead variants in either the susceptibility GWAS (top) or outcome GWAS (bottom). P-values are based on saddlepoint-approximated score tests (SAIGE), while meta-analysis P-values are derived from meta-analysis (METAL) of P-values generated from each cohort. Odds ratios are computed from Firth logistic regression. *Variants with an asterisk were excluded from the corresponding analysis due to quality control filters but are included here for completeness.

Extended Data Table 5 | Uncovered HLA alleles

Locus	Allele ID	Allele Frequency	Description
A	02:01@21	0.0019	One mismatch in exon 4; codon 245 position 3; GCG to GCA; Synonymous substitution (Ala to Ala)
B	35@1	0.0058	One mismatch in exon 5; codon 304 position 1; GCT to ACT; Nonsynonymous substitution (Ala to Thr)
	15:10@23	0.0019	One mismatch in exon 3; codon 135 position 3; GCC to GCG; Synonymous substitution (Ala to Ala)
	35@24	0.0019	One mismatch in exon 3 codon 158, position 3; GCT to GCC; Synonymous substitution (Ala to Ala)
	53@25	0.0019	One mismatch in exon 3 codon 171, position 1; CAC to TAC; Nonsynonymous substitution (His to Tyr)
	42@26	0.0019	One mismatch in exon 3 codon 138, position 3; ACC to ACG; Synonymous substitution (Thr to Thr)
C	16:01@8	0.0019	One mismatch in exon 2; codon 62 position 3; CGG to CGA; Synonymous substitution (Arg to Arg)
	07@11	0.0019	One mismatch in exon 3 codon 100 position 3; GGT to GGC; Synonymous substitution (Gly to Gly)
	17@27	0.0019	One mismatch in exon 2; codon 105 position 3; CCG to CCC; Synonymous substitution (Pro to Pro)
DPA1	03:01@2*	0.05	One mismatch in exon 1 promoter region; codon -31 position 2; ATG to ACG; Nonsynonymous substitution (Met to Thr).
	03:01@3	0.0192	One mismatch in exon 4; codon 204 position 3; GTG to GTC; Synonymous substitution (Val to Val).
	02:07@4	0.0404	One mismatch in exon 4; codon 224 position 2; CGG to CAG; Nonsynonymous substitution (Arg to Gln).
	01@10	0.0019	One mismatch in exon 2; codon 20 position 3; GGA to GGG; Synonymous substitution (Gly to Gly).
	01:03@12	0.0019	One mismatch in exon 4; codon 204 position 3; GTG to GTC; Synonymous substitution (Val to Val).
	02:02@14	0.0077	One mismatch in exon 2; codon 38 position 3 AAA to AAG; Synonymous substitution (Lys to Lys).
	02@28	0.0019	Two mismatches in exon 2; codon 31 position 1 and position 2; CAG to ATG; Nonsynonymous substitution (Gln to Met)
DPB1	414:01@17	0.0019	One mismatch in exon 4; codon 205 position 1 ATG to GTG; Nonsynonymous substitution (Met to Val).
	333@20	0.0019	One mismatch in exon 2; codon 72 position 1 GTG to TTG; Nonsynonymous substitution (Val to Leu).
	01@29	0.0019	One mismatch in exon 2; codon 43 position 3; GGG to GGA; Synonymous substitution (Gly to Gly)
	26@30	0.0019	One mismatch in exon 4; codon 194 position 2; CAG to CGG; Nonsynonymous substitution (Gln to Arg)
DQA1	01:06@22	0.0019	One mismatch in exon 2; codon 44 position 1; GCT to ACT; Nonsynonymous substitution (Ala to Thr).
DQB1	04@5	0.0019	One mismatch in exon 3; codon 123 position 2; TAT to TGT; Nonsynonymous substitution (Tyr to Cys).
	05:02@6	0.0019	One mismatch in promoter region of exon 1; Position 544; A to G, non coding region.
	06@31	0.0019	One mismatch in exon 2; codon 9 position 2; TAC to TTC; Nonsynonymous substitution (Tyr to Phe)
	06@32	0.0038	One mismatch in exon 2; codon 48 position 3; CGC to CGG; Synonymous substitution (Arg to Arg)
	06@33	0.0019	One mismatch in exon 4; codon 224 position 2; CAG to CGG; Nonsynonymous substitution (Gln to Arg)
	06@34	0.0019	One mismatch in exon 2; codon 57 position 2; GTT to GAT; Nonsynonymous substitution (Val to Asp)
	06@35	0.0019	One mismatch in exon 3; codon 125 position 2; GGC to GCC; Nonsynonymous substitution (Gly to Ala)
	06@36	0.0019	Two mismatches in exon 2; codon 38 position 3; GCG to GCA; Synonymous substitution (Ala to Ala) and codon 47 position 3 TAT to TAC Synonymous substitution (Tyr to Tyr)
DRB1	13@37	0.0019	One mismatch in exon 2; codon 6 position 1; CGT to TGT; Nonsynonymous substitution (Arg to Cys)
	01@38	0.0019	One mismatch in exon 2; codon 74 position 3; GCC to GCG; Synonymous substitution (Ala to Ala)
DRB3	01:34@16	0.0038	One mismatch in exon 2; codon 85 position 2 GTT to GCT; Non synonymous substitution (Val to Ala).
	02:02@18	0.0038	One mismatch in exon 3; codon 113 position 3 AAC to AAA; Non synonymous substitution (Asn to Lys).
	02:02@19	0.0058	One mismatch in exon 2; codon 77 position 3 AAC to AAT; Synonymous substitution (Asn to Asn).
	01@39	0.0038	One mismatch in exon 2; codon 77 position 3; AAT to AAC; Nonsynonymous substitution (Ala to Gly)
DRB4	03@40	0.0038	Two mismatches in exon 2; codon 37 positions 1 and 2; TTC to AAC; Nonsynonymous substitution (Phe to Asn)
	01@13	0.0115	One mismatch in exon 2; Codon 32 position 3; TAC to TAT; (Synonymous substitution Tyr to Tyr).
DRB5	01@41	0.0019	One mismatch in exon 2; codon 76 position 2; GAC to GGC; Nonsynonymous substitution (Asp to Gly)
	02@7	0.0365	One mismatch in exon 4; codon 203 position 1; ATC to GTC; Nonsynonymous substitution (Ile to Val).
	02@9	0.0019	One mismatch in exon 2; codon 67 position 1; TTC to ATC; Nonsynonymous substitution (Phe to Ile).
	02@15	0.0077	2 mismatches: 1) exon 3; codon 138 position 1; GAG to AAG; Nonsynonymous substitution (Glu to lys) 2) exon 4; codon 203 position 1; ATC to GTC; Nonsynonymous substitution (Ile to Val).

Novel HLA alleles identified in sequence-based HLA typing of 297 Sierra Leoneans.

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection

Data analysis

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

Raw de-identified genetic data from this study has been uploaded to the European Genome-Phenome Archive under dataset ID EGAD00010002510 and EGAD00010002509, and can be accessed at <https://ega-archive.org/register/>. Summary statistics for the genetic analyses performed in this study are included as extended data tables 1-6 and summary statistics for the massively parallel reporter assay are included as extended data tables 7-9. Data from the 1000 Genomes Project is available at <https://www.internationalgenome.org/data/>. Genome assembly hg19 is available at https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_000001405.13/.

Research involving human participants, their data, or biological material

Policy information about studies with [human participants or human data](#). See also policy information about [sex, gender \(identity/presentation\), and sexual orientation](#) and [race, ethnicity and racism](#).

Reporting on sex and gender

Self-reported gender was collected during study collection. In addition, biological sex was determined through analysis of the genetic data. Biological sex was used as a covariate in the GWAS.

Reporting on race, ethnicity, or other socially relevant groupings

We report the country of origin of all participants as being from Nigeria or Sierra Leone. In addition, for participants recruited jointly with the 1000 Genomes Project, we report the tribal affiliation of the participant as Yoruba, Esan, or Mende.

Population characteristics

- Age (years)
- Biological Sex
- SNP genotyping
- Lassa Virus IgG serostatus
- Genetic ancestry as reflected in principal component analyses
- Directly typed or imputed HLA alleles for class I and class II genes
- Lassa Fever case or population control status
- Clinical outcome from Lassa Fever (survival or fatal outcome)
- Clinical symptoms from Lassa Fever

Recruitment

Lassa Fever case definition and recruitment
Irrua Specialist Teaching Hospital (ISTH), Nigeria

We recruited LF cases at ISTH between 2011-2014 and 2016-2018 with a gap from 2014-2016 due to the Ebola outbreak in West Africa that temporarily halted research operations. We performed molecular diagnostic testing for all suspected cases who met clinical diagnostic criteria for LF including fever $>38^{\circ}\text{C}$ for less than 3 weeks, absence of signs of local inflammation, absence of clinical response to anti-malarials, and additional major and minor signs. Suspected cases who were positive by molecular diagnostic testing were recruited to the study following informed consent.

From 2011 to 2014, ISTH study staff performed RT-PCR targeting the GPC gene62 as the primary diagnostic and positive cases were recruited into the study. However, due to concerns about false positives of this initial assay, a confirmatory RT-qPCR assay was performed at the Broad Institute in Boston using primers against the LASV S segment (forward: CCCAAGCYCTHCCYACAAT, reverse: AACCCCTTATGAGAAATACTBTAYAA) and a subset of patients underwent next-generation viral sequencing¹². We only included data from recruited cases who were positive by this latter RT-qPCR or who had positive LASV sequencing with greater than 1 viral reads per kilobase (RPKM) in the GWAS analysis.

Between 2016-2018, ISTH patients who met clinical diagnostic criteria for LF were tested at ISTH with 2 RT-qPCR assays, one targeting the GPC gene (RealStar LASV RT-PCR Kit 1.0 CE, Altona Diagnostics, Hamburg, Germany) and a second targeting the LASV L segment^{63,64}. Suspected cases from this period who were positive by either RT-qPCR assay were recruited to the study following informed consent. A subset of these cases also underwent viral sequencing²⁶. We only included data from suspected LF cases who were positive by both of the RT-qPCR assays, or by viral genomic sequencing (with > 1 RPKM from the viral genome) in the GWAS.

Kenema Government Hospital (KGH), Sierra Leone

LF cases were recruited at KGH between 2011 and 2018 with a gap from 2015-2016 due to the Ebola outbreak in West Africa. Suspected cases included individuals who met clinical diagnostic criteria for LF²⁴ and were positive by either ELISA for a LASV antigen or IgM antibody against LASV^{28,65}. We additionally performed viral sequencing from a subset of enrolled cases¹². We only included data from suspected cases who were either antigen-ELISA positive or viral sequencing positive with RPKM > 1 in the GWAS.

Population control recruitment

Study staff at ISTH and KGH recruited population controls through outreach efforts to villages with a recent history of LF cases. Village controls (Table S1) were healthy individuals who were recruited from the same household and/or village as cases, prioritizing unrelated individuals where possible. Trio controls (Table S1) were healthy families of mother, father, and

child from the Esan population in NG and the Mende population in SL who were recruited jointly with Phase 3 of the 1000 Genomes Project (1KGP).

Given that very few Lassa Fever cases are documented annually and mostly occur in rural areas far from diagnostic centers, we are likely identifying only the fraction of cases in which extreme disease presentations motivated patients to seek medical attention; results are thus more likely to reflect aspects of severe cases than subclinical ones.

This work was evaluated and approved under the purview of the following institutional review boards / local ethics committees: Nigerian National Health Research Ethics Committee and Irrua Specialist Teaching Hospital (ISTH/ HREC/20170915/22), Sierra Leone Ethics and Scientific Review Committee (070716), Tulane University Human Research Protections Office (10-191330), and Harvard University Area Committee on the Use of Human Subjects (19-0023). Enrollment procedures and sampling efforts were executed at ISTH, KGH, and their surrounding communities with participant consent or through a waiver of consent granted by the appropriate institutional review board / local ethics committee. A subset of samples shared with the study collaboration include those stored as clinical excess or approved for secondary use.

Ethics oversight

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences Behavioural & social sciences Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Power calculations suggested >90% power to obtain genome-wide significance for an effect variant with minor allele frequency greater than 0.15 and genomic risk ratio greater than 2.00. Our case-control design is intended to detect causal resistance alleles that are under positive selection, thus power was determined for common variants with a larger effect-size than most variants affecting complex traits.
Data exclusions	Data was excluded for recruited individuals who did not meet our case definitions. This includes suspected Lassa Fever cases based on symptoms who did not meet the molecular diagnostic criteria for definition as a Lassa Fever case. In addition, data was excluded when the genotype-determined sex did not match the self-reported sex of the study participant as these likely reflected sample swap errors.
Replication	We recruited Lassa Fever cases and controls from two study sites, in Nigeria and Sierra Leone, to serve as replication cohorts. rs181032423 had a genome-wide significant association with susceptibility in the Nigerian cohort ($P = 2.2 \times 10^{-8}$) and was also nominally associated with susceptibility in the Sierra Leone cohort ($P = 0.026$). Other findings were significant in only one or the other cohort which could be due to viral genetic diversity between the two countries.
Randomization	This was a case-control study design and did not involve randomization.
Blinding	As this was a case-control study design, investigators were not blinded to group allocation.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data
<input type="checkbox"/>	<input checked="" type="checkbox"/> Dual use research of concern
<input checked="" type="checkbox"/>	<input type="checkbox"/> Plants

Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

Dual use research of concern

Policy information about [dual use research of concern](#)

Hazards

Could the accidental, deliberate or reckless misuse of agents or technologies generated in the work, or the application of information presented in the manuscript, pose a threat to:

- | No | Yes |
|-------------------------------------|---|
| <input type="checkbox"/> | <input checked="" type="checkbox"/> Public health |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> National security |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Crops and/or livestock |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Ecosystems |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Any other significant area |

Hazards

For examples of agents subject to oversight, see the United States Government [Policy for Institutional Oversight of Life Sciences Dual Use Research of Concern](#).

Experiments of concern

Does the work involve any of these experiments of concern:

- | No | Yes |
|-------------------------------------|--|
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Demonstrate how to render a vaccine ineffective |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Confer resistance to therapeutically useful antibiotics or antiviral agents |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Enhance the virulence of a pathogen or render a nonpathogen virulent |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Increase transmissibility of a pathogen |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Alter the host range of a pathogen |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Enable evasion of diagnostic/detection modalities |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Enable the weaponization of a biological agent or toxin |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Any other potentially harmful combination of experiments and agents |

Precautions and benefits

Biosecurity precautions

Biosecurity oversight

Benefits

Communication benefits