Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Host genetic determinants of rotavirus disease and immunity in African children

James Gilchrist
james.gilchrist@paediatrics.ox.ac.uk

University of Oxford https://orcid.org/0000-0003-2045-6788
Charles Agoti
KEMRI-Wellcome Research Programme
Adriace Chauwa
Basic Science & Immunology Unit, Centre for Infectious Disease Research in Zambia
Jacent Nassuuna
Medical Research Council/Uganda Virus Research Institute and London School of Hygiene and Tropical Medicine Uganda Research Unit
Joyce Kabagenyi
Medical Research Council/Uganda Virus Research Institute and London School of Hygiene and Tropical Medicine Uganda Research Unit
Martin Ssejjoba
Medical Research Council/Uganda Virus Research Institute and London School of Hygiene and Tropical Medicine Uganda Research Unit
Natasha Laban
Basic Science & Immunology Unit, Centre for Infectious Disease Research in Zambia
Alexandra Cardoso Pinto
Oxford Vaccine Group, Department of Paediatrics, University of Oxford.
Emily Webb
London School of Hygiene & Tropical Medicine
Alex Macharia
KEMRI-Wellcome Trust Research Programme https://orcid.org/0000-0001-5364-9610
Johnstone Makale
KEMRI-Wellcome Research Programme
Perpetual Wanjiku
KEMRI-Wellcome Research Programme
Shebe Mohammed
KEMRI-Wellcome Research Programme
Neema Mturi
KEMRI-Wellcome Trust Research Programme
Flavia Matos Santo
MRC–Weatherall Institute of Molecular Medicine, University of Oxford
Alba Verge de los Aires
MRC–Weatherall Institute of Molecular Medicine, University of Oxford
Claire Robertson
Oxford Vaccine Group, Department of Paediatrics, University of Oxford.
Seiko Makino
Centre for Human Genetics, University of Oxford
Luke Jostins-Dean
University of Oxford https://orcid.org/0000-0002-2475-3969
Julian Knight
University of Oxford https://orcid.org/0000-0002-0377-5536
Bridgious Walusimbi
Medical Research Council International Statistics and Epidemiology Group, Department of Infectious Disease Epidemiology, London School of Hygiene and Tropical Medicine
Alison Elliot
LSHTM
Gagandeep Kang
https://orcid.org/0000-0002-3656-564X
Sudhir Babji
The Wellcome Trust Research Laboratory, Christian Medical College, Vellore

Roma Chilengi
Zambia National Public Health Institute https://orcid.org/0000-0003-0221-9527
Adrian Hill
University of Oxford https://orcid.org/0000-0003-0900-9629
Benjamin Fairfax
MRC–Weatherall Institute of Molecular Medicine, University of Oxford, Oxford
Thomas Williams
Kenya Medical Research Institute https://orcid.org/0000-0003-4456-2382
Michelo Simuyandi
Centre for Infectious Disease Research in Zambia https://orcid.org/0000-0002-7348-2835
Alexander Mentzer
University of Oxford https://orcid.org/0000-0002-4502-2209
Sophie Uyoga
KEMRI-Wellcome Trust Research Programme https://orcid.org/0000-0003-1746-7873
Jessa Rop
KEMRI-Wellcome Trust Research Programme

Article

Keywords:

Posted Date: August 22nd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10565996/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Additional Declarations: There is **NO** Competing Interest.

Abstract

Rotavirus (RV) is a major cause of diarrhoeal disease and mortality in African children. The impact of RV vaccination on child health has been limited by low vaccine effectiveness in resource-constrained settings. We urgently need better vaccines to control RV disease, delivery of which will require an improved understanding of RV biology. We hypothesised that the genetics of RV disease and immunity are likely to be mutually informative of RV biology. We performed genome-wide association studies of infection-induced and vaccine-derived RV immunity, hospitalised RV gastroenteritis (RVGE) risk and all-cause diarrhoeal disease risk in > 3,000 Ugandan, Kenyan and Zambian children. In doing so, we expand on the role of human blood group antigen secretion in RV infection and immunity, resolving complex interactions between secretor status, ABO blood group, Lewis antigen and viral genotype. We further identify a separate, novel locus at chromosome 2q12, rs59241810:T, associated with increased anti-RV IgA responses and lower risk of RVGE and all-cause diarrhoea with dehydration. Variation at rs59241810 regulates monocyte *LIMS1* expression, thereby impacting IFNγ-induced hepatocyte growth factor (HGF) secretion, suggesting a role for monocyte-derived HGF in the generation of anti-RV IgA responses. Our data reveal the complex interplay between immunity, disease, pathogen variation and host genetics in RV infections, and identify LIMS1/HGF signalling as a target to design more effective anti-RV vaccines.

Main

Group A rotavirus (RV) is the most common cause of moderate-to-severe diarrhoea in children aged < 2 years-old[1]. An estimated 128,000 deaths in children under 5 years were caused by RV in 2023, over 100,000 of which occurred in sub-Saharan Africa[2]. Inclusion of RV vaccines in immunisation schedules has contributed to a reduction in RV gastroenteritis (RVGE) disease, hospitalisation and death[3–5]. Early trials of RV vaccination in high-income settings uniformly demonstrated high levels of vaccine efficacy (> 90%)[6]. However, in low-income settings, including in sub-Saharan Africa, vaccine effectiveness has been found to be as low as 38%[7,8]. Improving RV vaccine efficacy in African populations will require a robust understanding of the aetiology of disparate vaccine efficacy between high and low-income settings. Causes of this disparity are likely to be multifactorial, with force of infection, strain diversity, breastfeeding, malnutrition, environmental enteropathy and host genetics all hypothesised to contribute[9].

Rotaviruses are double-stranded RNA viruses. The RV genome consists of 11 segments that encode 6 structural proteins (VP1, VP2, VP3, VP4, VP6 and VP7) and 6 non-structural proteins (NSP1-6). Variation in one of these structural proteins, VP6 gives rise to 10 RV species (A-J), of which group A RVs (GARV) cause over 90% of human infection. Within GARVs, strain diversity is defined by sequence variation at the two segments encoding the outer layer structural proteins: VP7 encodes the glycoprotein (G), and VP4 encodes a protease-sensitive spike protein (P)[10]. Some RV strains, dependent on P genotype, have been shown to use A-type human blood group antigen (HBGA) as an attachment factor[11]. The presence of HBGA on mucosal surfaces is mediated by the alpha (1,2) fucosyltransferase enzyme encoded by *FUT2*, and homozygous carriers of a common *FUT2* nonsense mutation (rs601338:A) are termed non-secretors, and do not express HBGA on mucosal surfaces or secrete them in bodily fluids[12]. Moreover, another fucosyltransferase, encoded by *FUT3*, determines fucosylation of HBGA precursor glycans defining Lewis antigen status, which is also associated with RV binding affinity *in vitro*[13]. Identity and location of HBGA can thus determine likelihood of rotavirus attachment, invasion, disease and subsequent immunity in a P genotype-dependent manner.

Candidate gene studies have established host genetic variation as a major determinant of RV disease and immunity. These studies have focussed on the role of HBGA in RV disease risk and immune responses. In keeping with studies defining binding affinities of RV P antigen, these data have determined reproducible effects of *FUT2* genotype (secretor status) on RV disease risk in vaccinated and unvaccinated children: non-secretors are protected against P[4] and P[8] RV strains but not P[6] genotypes[14–22]. Moreover, the *FUT3* locus, determining Lewis antigen expression, is also a determinant of RV disease risk, with Lewis negative individuals protected against P[4] and P[8] genotype disease but at increased risk of P[6] disease[14–19,21]. Despite the demonstration of epistasis between the *FUT2* locus and ABO blood group in other contexts[23], ABO blood group has not been found to be consistently associated with RV disease risk[15–17,19,21,24].

Secretor status has been shown to associate with naturally-acquired anti-RV antibody responses (serum IgG, neutralising antibody titres and salivary IgA) in adults[13,15,25,26], but has been inconsistently associated with vaccine responses as have Lewis antigen expression and ABO blood group[17,27–31]. That the effects of secretor status on vaccine response are variable may reflect confounding by immune responses to natural infection, especially in settings where RV infection remains common despite immunisation. Live attenuated anti-RV vaccines contain P[8] genotype virus and are administered orally. Where an association between RV vaccine response and secretor status is observed, the direction of effect is one in which individuals protected against P[8] disease (non-secretors) have poor vaccine responses, suggesting a model where non-secretors fail to support vaccine attachment and replication preventing generation of robust immune responses.

Candidate gene studies of HBGA expression and identity have thus defined a role for host genetics in RV disease risk and immunity. They do not, however, explain the poor RV vaccine efficacy in African populations. Non-secretor allele frequencies are comparable in African and European populations (minor allele frequency – MAF-Africa = 0.50, MAF-Europe = 0.47; https://gnomad.broadinstitute.org/), and while Lewis negative individuals are more common in African populations, P[6] genotype vaccine failures among non-secretor, Lewis negative individuals make, at most, a modest contribution to observed vaccine failure[17].

Serum anti-RV IgA levels are associated with protection against RV infection in vaccinated[32] and unvaccinated[33] individuals. A protective threshold of > 20IU/ml has been proposed as a correlate of protection against RVGE in vaccinated infants[34], however, the generalizability of this threshold across settings is unclear[35]. In settings where RV immunisation lacks high effectiveness, anti-RV IgA will accrue following vaccination but also following RV infection. Naturally-acquired anti-RV immunity is thus likely to confound the interpretation of anti-RV IgA levels in settings with poor vaccine performance. In keeping with RV disease risk and vaccine take, the acquisition of anti-RV IgA following RV infection is likely to be under genetic control, however the contribution of host genetic variation to this, and its interaction with disease risk, is undefined.

Unbiased assessments of host genetics, e.g. genome-wide association studies (GWAS), are preferable to candidate gene studies, in that they do not rely on prior biological knowledge, and allow more complete estimates of the contribution of host genetics to a given trait. To date, there are no published GWAS of RV disease or immunity. We address this in this study, performing GWAS of both hospitalised RVGE in Kenyan children and anti-RV-specific IgA levels in Ugandan children, replicating our findings for RV immunity in Zambian infants and for diarrhoeal disease risk in Ugandan children (Fig. 1a). In doing so, we confirm the role of secretor status as a major genetic determinant of RV disease and immunity, define its interaction with Lewis antigen and ABO blood group, and identify genetic regulation of *LIMS1* expression and hepatocyte growth factor (HGF) secretion in monocytes as a novel determinant of both traits.

Results

RV immunity in Ugandan children

Serum anti-RV IgA quantification passed quality control thresholds in 616 Ugandan children aged 2 (n = 330) and 3 years (n = 286). At the age of 2–3 years, Ugandan children recruited to the Entebbe Mother and Baby Study (EMaBS) had median anti-RV IgA levels of 105 IU/ml, with an overall seroprevalence (anti-RV IgA ≥ 20 IU/ml) of 76% (Fig. 1b). Anti-RV IgA levels increased with age (Wilcoxon rank sum $P = 0.049$), with median levels at 2 years of 83 IU/ml, rising to 117 IU/ml at the age of 3 years (Supplementary Fig. 1, Supplementary Table 1). Similarly, the proportion of seropositive children also increased with age at sampling (72% aged 2 years, 82% aged 3 years, $P = 2.6 \times 10^{-3}$). Female participants had higher anti-RV IgA levels (male median anti-RV IgA = 82 IU/ml, female median anti-RV IgA = 124 IU/ml, Wilcoxon rank sum $P = 0.014$), although sex was not associated with seropositivity (79% male, 76% female). In keeping with the importance of rotavirus as a cause of infant diarrhoea in this population, increased episodes of diarrhoea in the first year of life are associated with raised serum anti-RV IgA at 2–3 years in Ugandan children (ordinal regression $P = 0.004$).

RV disease in Kenyan children

Of 554 children admitted with RV disease recruited to the study, 448 had complete G and P genotype data. Of these, 407 children had a single defined RV genotype (Supplementary Table 2), with 41 mixed genotype infections (9%). As we have previously reported[36], the majority of RV were genotypes G1P[8] (n = 208, 46%), G9P[8] (n = 62, 14%), G8P[4] (n = 49, 11%), G8P[6] (n = 23, 5%) and G2P[4] (n = 18, 4%). In total, there were 333 P[8] genotype infections, 47 P[6] and 92 P[4]. Children hospitalised with RV disease had a median age of 10 months (range 0-133) and 225 (41%) were female (Fig. 1d). Age and sex were not predictive of P genotype of RV disease. The majority of children presented with diarrhoea (n = 514, 97%), vomiting (n = 463, 87%) and fever (n = 408, 77%). Children presented with clinical features of hypovolaemic shock in 135 cases (25%) and 8 children died (2%). The median duration of hospital stay was 4 days (range 0–40). Among children hospitalised with RV disease in this population, P genotype, age or sex were not significant modifiers of clinical presentation or course (Supplementary Table 3), although diarrhoea was less prominent as a presenting symptom in older children.

GWAS of RV immunity & disease

Following SNP and sample QC, genotypes at 16,394,501 autosomal SNPs from 611 samples were included in the GWAS of anti-RV-specific serum IgA levels in Ugandan children. The genomic inflation factor ($\lambda = 1.00$) and inspection of the QQ-plot (Supplementary Fig. 2a) are not suggestive of uncontrolled confounding variation. Genome-wide association analysis revealed two loci associated with RV immunity at genome-wide significance (Fig. 1c, Table 1). The first of these loci, rs601338:A, located in *FUT2* on chromosome 19 ($P = 1.46 \times 10^{-15}$, β=-0.42), determines secretor status, which is well described as a major genetic determinant of RV infection traits[14–22]. The second associated locus, rs59241810:T, located on chromosome 2q12 ($P = 1.65 \times 10^{-12}$, $\beta = 0.56$), has not previously been identified as a determinant of RV disease or immunity.

Table 1
Association statistics for secretor status, ABO, Lewis & EDAR genotypes with serum anti-RV IgA levels and RV seropositivity in Ugandan children.

				Serum anti-RV IgA levels			Serostatus		
			Number (total n = 616)	Median IU/ml (range)	beta (95% CI)	*P*-value	Seropositive (%)	OR (95% CI)	*P*-value
Secretor status		Secretors*	465	139 (0–2,100)	-0.77 (-0.93, -0.61)	2.54E-19	396 (85%)	0.18 (0.12–0.27)	4.91E-16
		Non-secretors	151	18 (0–1,197)			75 (50%)		
ABO blood group	Overall	A	143	75 (0–2,100)	-0.16 (-0.34, 0.02)	0.08	106 (74%)	0.85 (0.53–1.35)	0.49
ABO blood group		B	120	94 (0–2,100)	-0.10 (-0.29, 0.09)	0.3	93 (78%)	1.00 (0.60–1.66)	0.99
		AB	22	129 (0–2,100)	-0.06 (-0.47, 0.34)	0.75	17 (77%)	1.02 (0.36–2.90)	0.97
		O*	331	116 (0–2,100)			255 (77%)		
	Secretors	A	108	121 (0–2,100)	-0.18 (-0.38, 0.02)	0.09	91 (84%)	0.91 (0.48–1.72)	0.77
		B	87	116 (0–2,100)	-0.14 (-0.36, 0.08)	0.2	75 (86%)	1.04 (0.51–2.12)	0.92
		AB	18	153 (0–2,100)	0.00 (-0.43, 0.42)	0.99	15 (83%)	0.96 (0.26–3.55)	0.95
		O*	252	156 (0–2,100)			215 (85%)		
	Non-secretors	A	35	11 (0-909)	-0.12 (-0.46, 0.21	0.47	15 (42%)	0.63 (0.27–1.48)	0.19
		B	33	36 (0–1,197)	0.16 (-0.18, 0.50)	0.36	18 (55%)	1.48 (0.62–3.51)	0.3
		AB	4	11 (0–27)	-0.52 (-1.35, 0.32)	0.23	2 (50%)	1.20 (0.15–9.36)	0.38
		O*	79	21 (0–1,056)			40 (51%)		
Lewis antigen phenotypes	Overall	Le (+)*	381	83 (0–2,100)	0.10 (-0.05, 0.25)	0.19	277 (73%)	1.74 (1.15–2.64)	0.008
		Le (-)	235	114 (0–2,100)			194 (83%)		
	Secretors	Le (+)*	278	154 (0–2,100)	-0.15 (-0.32, 0.01)	0.07	241 (87%)	0.71 (0.42–1.21)	0.21
		Le (-)	287	116 (0–2,100)			155 (83%)		
	Non-secretors	Le (+)*	103	0 (0-909)	0.77 (0.52, 1.03)	1.59E-08	36 (35%)	8.83 (3.67–21.24)	1.12E-06
		Le (-)	48	93 (0–1,197)			39 (81%)		
EDAR rs59241810 genotype		CC	452	74 (0–2,100)	0.58 (0.41, 0.71)	1.77E-12	328 (73%)	2.64 (1.59–4.39)	1.67E-04
		CT	156	185 (0–2,100)			135 (87%)		

Association statistics are calculated by linear (normalised IgA levels) and logistic (seropositivity) regression. Secretors, O blood group and Lewis (+) individuals are used as baseline for regression analysis (highlighted*). Association statistics for rs59241810 are calculated under an additive model with T as the effect allele. All regression models are corrected for age at sampling, sex and 4 genetic principal components.

	TT	8	482 (41 – 2,100)	8 (100%)

Association statistics are calculated by linear (normalised IgA levels) and logistic (seropositivity) regression. Secretors, O blood group and Lewis (+) individuals are used as baseline for regression analysis (highlighted*). Association statistics for rs59241810 are calculated under an additive model with T as the effect allele. All regression models are corrected for age at sampling, sex and 4 genetic principal components.

Genotypes at 15,415,726 autosomal SNPs from 554 cases and 1,556 controls were included in the GWAS of hospitalised RVGE in Kenyan children. The genotypes of causative RV isolates, and case demographics, clinical features and outcome are outlined in Supplementary Table 2. There is no evidence of uncontrolled confounding variation (Supplementary Fig. 2b). The GWAS reveals a single locus at *FUT2* (peak SNP rs602662:A, $P = 3.07x10^{-8}$, OR = 0.63; rs601338:A/secretor status, $P = 5.72x10^{-8}$, OR = 0.63) associated with hospitalised RVGE at genome-wide significance (Fig. 1e).

Secretor status is associated with RV immunity & disease

Fine-mapping of the RV association at *FUT2* for both serum anti-RV IgA and RVGE disease risk (Fig. 2a) demonstrates a shared genetic architecture, with the rs601338:A nonsense mutation being among the most significantly associated variants for both traits. Our discovery analyses mapped association in an additive model, however non-secretors are rs601338:AA homozygotes, and in keeping with this, the rs601338 genotype-dependent reduction of serum anti-RV IgA concentration is seen only in the context of rs601338:AA homozygosity (Fig. 2b).

It has previously been reported that the protection afforded by non-secretor status against RVGE is dependent on RV-type. To explore the interaction between RV P-type, secretor status and disease risk, we used multinomial logistic regression to estimate the protective effect of non-secretor status on disease caused by P[4] (n = 92), P[6] (n = 47) and P[8] (n = 333) disease (Fig. 2c). In that analysis, non-secretors (rs601338:AA homozygotes) are significantly protected against P[4] ($P = 1.73x10^{-3}$, OR = 0.18) and P[8] ($P = 2.55x10^{-9}$, OR = 0.17) RVGE, but not P[6] RVGE ($P = 0.93$, OR = 0.97).

Secretor status & ABO blood group

An interaction between RV disease, secretor status and ABO blood group has been inconsistently reported[17,29]. To explore whether the identity of the secreted blood group antigen was a determinant of RVGE disease, we imputed ABO blood group in the study samples using phased genotype data at rs8176719 (determines O/non-O blood group) and rs8176747 (determines A/B type) as previously described[37] (Supplementary Table 4). We observed a significant association (Table 2, Fig. 2d) of non-O blood group with increased risk of hospitalised RVGE in Kenyan children (O as baseline); A ($P = 0.003$, OR = 1.45), B ($P = 0.016$, OR = 1.37), and AB ($P = 0.005$, OR = 2.03). This increased risk was only observed for P[8] genotype disease and among secretors (Fig. 2d). There was no evidence that ABO blood group determines anti-RV immunity in Ugandan children, irrespective of secretor status.

Table 2
Secretor status, ABO, Lewis & EDAR genotypes and risk of hospitalised RV disease risk in Kenyan children.

			Counts (%)					Risk of disease						
			Controls	Cases				Total		P[4]		P[6]		P[
				Total	P[4]	P[6]	P[8]	OR (95% CI)	*P*-value	OR (95% CI)	*P*-value	OR (95% CI)	*P*-value	OR (9 CI
Secretor status		Secretors*	1,273 (82%)	518 (94%)	89 (97%)	38 (81%)	316 (95%)	0.24 (0.16–0.37)	2.52E-11	0.18 (0.06–0.53)	1.83E-03	0.97 (0.42–2.23)	0.93	0. (0 0.
		Non-secretors	283 (18%)	36 (6%)	3 (%)	9 (19%)	17 (5%)							
ABO blood group	Overall	A	365 (23%)	156 (28%)	22 (24%)	16 (34%)	92 (28%)	1.45 (1.14–1.84)	2.66E-03	0.99 (0.59–1.67)	0.97	1.95 (0.99–3.85)	0.05	1. (1 1.
		B	300 (19%)	123 (22%)	17 (18%)	10 (21%)	81 (24%)	1.37 (1.06–1.78)	0.016	0.98 (0.68–3.24)	0.93	1.49 (0.68–3.24)	0.32	1. (1 2.
		AB	48 (3%)	29 (5%)	4 (4%)	2 (4%)	19 (6%)	2.03(1.24–3.32)	4.84E-03	1.49 (0.51–4.32)	0.47	1.79 (0.40–7.95)	0.44	2. (1 4.
		O*	843 (54%)	246 (44%)	49 (53%)	19 (40%)	141 (42%)							
	Secretors	A	304 (24%)	147 (28%)	21 (24%)	13 (34%)	90 (28%)	1.46 (1.13–1.88)	3.59E-03	0.95 (0.56–1.63)	0.86	1.97 (0.92–4.23)	0.08	1. (1 2.
		B	251 (20%)	117 (23%)	17 (19%)	8 (21%)	78 (25%)	1.39 (1.06–1.83)	0.017	0.98 (0.55–1.75)	0.96	1.48 (0.62–3.53)	0.38	1. (1 2.
		AB	40 (3%)	29 (6%)	4 (4%)	2 (5%)	19 (6%)	2.19 (1.31–3.66)	2.69E-03	1.52 (0.52–4.47)	0.44	2.19 (0.48–9.97)	0.31	2. (1 4.
		O*	678 (53%)	225 (43%)	47 (53%)	15 (39%)	129 (41%)							
	Non-secretors	A	61 (22%)	9 (25%)	1 (33%)	3 (33%)	2 (12%)	1.07 (0.45–2.51)	0.88	1.13 (0.09-14.00)	0.92	1.76 (0.37–8.41)	0.48	0. (0 1.
		B	49 (17%)	6 (17%)	0 (0%)	2 (22%)	3 (18%)	0.87 (0.33–2.32)	0.78	NA	NA	1.52 (0.26–8.82)	0.63	0. (0 2.
		AB	8 (3%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	NA	NA	NA	NA	NA	NA	NA
		O*	165 (58%)	21 (58%)	2 (67%)	4 (44%)	12 (71%)							
Lewis antigen phenotypes	Overall	Le (+)*	1,257 (81%)	483 (87%)	79 (86%)	32 (68%)	302 (91%)	0.65 (0.49–0.87)	3.20E-03	0.74 (0.40–1.35)	0.33	2.02 (1.07–3.79)	0.03	0. (0 0.
		Le (-)	299 (19%)	71 (13%)	13 (14%)	15 (32%)	31 (9%)							
	Secretors	Le (+)*	1,042 (82%)	460 (89%)	78 (88%)	27 (71%)	289 (91%)	0.61 (0.48–0.84)	2.25E-03	0.70 (0.36–1.34)	0.27	1.86 (0.90–3.84)	0.09	0. (0 0.
		Le (-)	231 (18%)	58 (11%)	11 (12%)	11 (29%)	27 (9%)							
	Non-secretors	Le (+)*	215 (76%)	23 (64%)	1 (33%)	5 (56%)	13 (76%)	1.69 (0.80–3.59)	0.17	4.47 (0.38–52.33)	0.23	2.64 (0.66–10.56)	0.17	0. (0 2.
		Le (-)	68 (24%)	13 (36%)	2 (67%)	4 (44%)	4 (24%)							
EDAR rs59241810 genotype		CC	1,241 (80%)	447 (81%)	69 (75%)	38 (81%)	279 (84%)	0.88 (0.69–1.12)	0.3	1.15 (0.71–1.86)	0.56	0.87 (0.43–1.79)	0.71	0. (0 0.
		CT	295 (19%)	104 (19%)	23 (25%)	9 (18%)	53 (16%)							

Association statistics are calculated by logistic (overall) and multinomial logistic (P-type specific estimates) regression. Secretors, O blood group and Lewis individuals are used as baseline for regression analysis (highlighted*). Association statistics for rs59241810 are calculated under an additive model with T a allele. All regression models are corrected for sex and 4 genetic principal components.

	TT	20 (1%)	3 (1%)	0 (0%)	0 (0%)	1 (0%)

Association statistics are calculated by logistic (overall) and multinomial logistic (P-type specific estimates) regression. Secretors, O blood group and Lewis individuals are used as baseline for regression analysis (highlighted*). Association statistics for rs59241810 are calculated under an additive model with T a allele. All regression models are corrected for sex and 4 genetic principal components.

Secretor status & Lewis phenotype

A fucosyltransferase, encoded by *FUT3*, determines Lewis antigen status, and has been demonstrated to be associated with RV disease in an RV genotype-dependent manner[14–19,21]. Lewis antigen status has not been reported to be associated with naturally-acquired anti-RV immunity[15,26]. To explore this, we imputed Lewis phenotypes in our data. To do so, we identified common, well-imputed exonic variants in *FUT3* in our dataset (n = 16, Supplementary Table 5). To impute Lewis status we restricted our analysis to variants previously shown to result in Lewis negative phenotypes, and used phased data to define individuals as being Lewis negative if both *FUT3* copies were affected by at least one deleterious variant.

Lewis negative children in Kenya are protected against RV disease overall (Table 2, OR = 0.65 95% CI 0.49–0.87, $P = 3.2x10^{-3}$). That protection is mediated by reduced risk of P[8] genotype disease (OR = 0.45 95% CI 0.31–0.67, $P = 9.0x10^{-5}$), and, in contrast, Lewis negative children are at increased risk of P[6] genotype disease (OR = 2.02 95% CI 1.07–3.80, $P = 0.03$). We found no significant effect of Lewis status on risk of P[4] genotype disease (Fig. 2e). We further explored whether Lewis phenotypes interact with secretor status in determining RV disease risk. There is evidence of significant interaction between Lewis and secretor status for RV disease risk overall (interaction $P = 8.6x10^{-4}$), with the protective effect of Lewis negative phenotypes against P[8] genotype disease only observed among secretors (Fig. 2e). By contrast, the effect estimates for increased risk of P[6] disease among secretors (OR = 1.86) and non-secretors (OR = 2.64) are similar, suggesting Lewis antigens modify P[6] disease risk independent of secretor status (Fig. 2e).

Lewis antigen phenotype does not determine naturally-acquired anti-RV IgA responses in Ugandan children (Fig. 2f, β = 0.10, 95%CI -0.05-0.25, $P = 0.19$). However, there is strong evidence for interaction (interaction $P = 1.65x10^{-8}$) between secretor status and Lewis phenotype, with Lewis negative children having increased anti-RV IgA titres among non-secretors (Fig. 2f, β = 0.78, 95%CI 0.53–1.03, $P = 8.5x10^{-9}$). This association is not observed among secretors, indeed there is weak evidence that the effect of the Lewis negative phenotype among secretors may be to decrease naturally acquired anti-RV IgA titres (β=-0.15, 95%CI -0.32-0.01, $P = 0.07$).

These data are consistent with a hypothesis that the raised risk of P[6] RV disease among Lewis negative children does not result in altered anti-RV IgA levels among secretors as this is balanced by a reduced risk of P[4] and P[8] disease. However, among non-secretors, who are resistant to P[4] and P[8] disease regardless of Lewis status, increased risk of P[6] disease is the major determinant of RV disease risk and thus immune responses. The absence of effect of Lewis status on anti-RV immune responses in other settings[15,26] may reflect that P[6] genotypes of RV are more common in African settings[38].

Variation at chromosome 2q12 modifies RV disease & immunity

As above, our GWAS of anti-RV IgA in Ugandan children identified a novel determinant of anti-RV immunity at chromosome 2q12 (Fig. 3a). The lead SNP, rs59241810:T ($P = 1.65x10^{-12}$, β = 0.56), is located within the 5th intron of *EDAR*, with the minor, ancestral (T) allele being associated with increased anti-RV IgA at 2–3 years of age in unvaccinated Ugandan children (Fig. 3b). We further sought to replicate this association between genetic variation at rs59241810 and anti-RV immunity in vaccinated Zambian children. Of 159 Zambian children recruited to the ROVAS-2 trial[39] (a randomized controlled trial of two versus three doses of Rotarix vaccine for boosting and longevity of vaccine immune responses in Zambia) with an anti-RV IgA measurement at 12 months of age, we generated genotyping data for 144, with 141 children passing genotyping QC filters. Of these children, 65 (46%) were female, 72 (51%) were randomised to a third dose of Rotarix, and 10 (7%) had mixed formula and breast milk feeding (the remainder were exclusively breastfed). At 12 months of age, Zambian children had median anti-RV IgA levels of 19 IU/ml (Fig. 3c), with an overall seroprevalence (anti-RV IgA ≥ 20 IU/ml) of 49%. As we have reported previously[39], a 3rd dose of anti-RV vaccine at 9 months was not associated with anti-RV IgA levels at 12 months of age. Non-secretor children in Zambia had significantly lower anti-RV IgA levels at 12 months than did their secretor counterparts (Fig. 3d, $P = 2.0x10^{-4}$, β=-0.79). Consistent with our data in unvaccinated Ugandan children, among vaccinated Zambian infants, rs59241810:T is associated with increased anti-RV IgA levels (Fig. 3e) at 12 months of age ($P = 0.048$, β = 0.33). Among our study participants rs59241810:T is common: Uganda MAF = 0.14, Kenya MAF = 0.10, Zambia MAF = 0.13. More broadly, rs59241810:T is common across African populations (MAF = 0.08) but is rare in European and Asian populations (MAF ≤ 0.01; https://gnomad.broadinstitute.org/).

We further explored whether rs59241810:T, as a mediator of RV immune responses, has a detectable effect on risk of hospitalised RVGE in Kenyan children. Carriage of the rs59241810:T allele does not modify overall risk of hospitalised RVGE ($P = 0.30$; OR = 0.88), however, in an analysis stratified by P-type, rs59241810:T is associated with reduced risk of hospitalised RVGE secondary to P[8] RV strains (Fig. 3f; $P = 0.04$; OR = 0.72). Children carrying the rs59241810:T allele have increased serum anti-RV IgA levels at 2–3 years of age but reduced risk of hospitalised P[8] RVGE. This suggests a hypothesis in which genetic variation at rs59241810 enhances RV immune responses, thereby reducing risk of severe RV disease. Given our study samples in Kenya were collected prior to RV vaccine rollout, acquired immunity to RV will be mediated by natural infection, and is expected to accrue with age. Thus, the protective effect of rs59241810:T carriage would be expected to be more pronounced in older children. We tested this by stratifying our analysis of RVGE disease risk into cases under (n = 342) and over (n = 211) 12 months of age. In that analysis, in keeping with our hypothesis, rs59241810:T carriage protects against hospitalised RVGE but only among children over 12 months of age (Fig. 3g; $P = 0.032$, OR = 0.65). Again, this effect is mediated by protection against P[8] ($P = 0.015$, OR = 0.50), with no detectable effect of rs59241810:T on risk of P[4] and P[6] disease.

Finally, we investigated whether genetic variation at chromosome 2q12 was associated with all-cause diarrhoea in Ugandan children enrolled in the Entebbe Mother and Baby Study (EMaBS) over the first 10 years of life. Of 1,876 children with complete follow-up to 10 years of age, 1,230 children had genotype data available passing QC filters. Among those children, the median number of clinic attendances for diarrhoea in the first 10 years of life was 3 (range 1–18 attendances). There were 218 children (18%) who experienced at least one episode of dysentery, 32 (3%) had at least one episode of prolonged diarrhoea, and 152 (12%) children had at least one episode of diarrhoea with dehydration. Genotype at rs59241810:T had no effect on the frequency of all-cause diarrhoea in the first 10 years of life in Ugandan children (Fig. 3h, P = 0.79, β = 0.02). This lack of effect on all-cause diarrhoea may be consistent with genotype at rs59241810:T exerting a RV-specific effect on diarrhoeal disease risk. The attributable fraction of RV for moderate-severe diarrhoea is highest in infants < 1 year, however, even among infants of that age, the attributable fraction in low-income countries prior to vaccine introduction was 0.16–0.28[1]. RV as a causative agent of diarrhoea is likely to be particularly enriched among children with diarrhoea and dehydration, but not so among children with dysentery or prolonged diarrhoea[40]. In keeping with this, we observe a protective effect of rs59241810:T genotype on all-cause diarrhoea with dehydration (Fig. 3i, P = 0.022, OR = 0.62, 95% CI 0.41–0.93), but no effect of rs59241810:T genotype on risk of dysentery or all-cause diarrhoea.

RV-associated genetic variation at chromosome 2q12 has regulatory function

Trait-associated genetic variation identified by GWAS are highly enriched for regulatory variation[41]. To explore whether rs59241810 has regulatory function, we explored the effects of genetic variation at this locus on local gene expression in immune cells. As above, rs59241810 is located in 5th intron of *EDAR*, and the ENCODE-rE2G (enhancer to gene) model[42], integrating genomic location, epigenomic and chromatin state data, predicts that this variant may act to regulate *EDAR* expression, in particular in T cells (maximum rE2G score in immune cells = 0.84 in Th2 cells). However, using expression quantitative trait locus (eQTL) mapping data in whole blood, peripheral blood mononuclear cells (PBMCs) and T cells as well as protein QTL (pQTL) mapping data in whole blood from European and African ancestry individuals[43–47] we were unable to identify robust evidence for a regulatory effect of rs59241810 on *EDAR* expression in these tissues (Supplementary Table 6, maximum posterior probability of colocalization, PP4 = 0.29). In light of this, we interrogated previously-published immune cell eQTL datasets to understand whether rs59241810 may have *cis* regulatory activity for genes other than *EDAR*. To do so, we utilised previously-published eQTL mapping data in PBMCs from African American participants in the Jackson Heart Study[48] (n = 1,012) and primary immune cells from healthy European-ancestry adults: monocytes[49] (n = 414), NK cells[50] (n = 245), B cells[51] (n = 281), CD4 + T cells[45] (n = 290) and CD8 + T cells[45] (n = 277). In primary monocytes, rs59241810:T carriage is associated with significantly reduced *LIMS1* expression ($P = 4.25 \times 10^{-14}$, Fig. 4a), an effect we replicate in an independent set of monocytes from healthy European-ancestry adults (n = 194, $P = 3.42 \times 10^{-4}$, Fig. 4b). Colocalization analysis best-supports a model in which *LIMS1* expression in monocytes shares a causal genetic determinant with anti-RV IgA levels in Ugandan children (posterior probability of colocalization, PP4 = 0.97). This relationship appears to be specific to *LIMS1* expression in monocytes: while there is weak evidence of association between genotype at rs59241810 and *LIMS1* RNA expression in bulk PBMCs, NK cells, B cells and $CD4^+$ T cells, there is no evidence of colocalization of eQTL signals in these cell types with anti-RV IgA genetic determinants (Fig. 4c, Supplementary Table 7).

LIMS1, LIM zinc finger domain containing 1, encodes an adaptor protein localised at focal adhesion plaques, which interacts with ILK (integrin-linked kinase) contributing to integrin-mediated signalling, and NCK2 (NCK adaptor protein 2) thereby also contributing to growth factor signalling[52]. The *LIMS1* eQTL in monocytes operates in naïve cells, as well as following IFNγ stimulation and 2 hours of LPS stimulation (Fig. 4d, Supplementary Table 7), but not following 24 hours of LPS stimulation. Integrins are transmembrane receptors mediating cellular interactions between cells and with the extracellular matrix, regulating multiple leukocyte functions including migration, immunological synapse formation and cytokine secretion[53]. The LIMS1 binding partner, ILK, contributes to proinflammatory cytokine production[54] and mediates TNF production following LPS stimulation in myeloid cells[55]. To expand on the observation that a *LIMS1* eQTL in naïve and stimulated monocytes is a determinant of RV immunity and disease, we explored the effect of *LIMS1* expression on cytokine secretion in monocytes following LPS (2 and 24 hours) and IFNγ (24 hours) stimulation using previously-published data[56]. In that analysis, *LIMS1* expression does not predict cytokine secretion following LPS stimulation (Supplementary Table 8). However, *LIMS1* RNA expression is inversely correlated with hepatocyte growth factor (HGF) production in IFNγ-stimulated monocytes ($P = 2.0 \times 10^{-5}$, Fig. 4e, Supplementary Table 8). HGF secretion following IFNγ stimulation is reduced in comparison to unstimulated monocytes ($P = 7.3 \times 10^{-4}$, Supplementary Fig. 3, Supplementary Table 9). *LIMS1* RNA expression is not associated with HGF secretion in unstimulated monocytes, but rather with the degree of the IFNγ-induced downregulation of HGF secretion (Supplementary Fig. 3). In keeping with *LIMS1* expression being a modifier of IFNγ-induced HGF production by monocytes, rs59241810:T allele carriage is also associated with increased HGF secretion in IFNγ-stimulated monocytes (Fig. 4f, P = 0.02).

Discussion

In this study we performed complementary GWAS of RV disease and immunity. In doing so, we replicate evidence defining secretor status as a key mediator of RV disease risk and acquisition of anti-RV immunity through infection in early childhood[13,14,16–19,21,27–29]. We further provide evidence that Lewis antigen status is an important determinant of RV disease and immunity. We find that Lewis negative children, specifically children who are also secretors, are at reduced risk of P[8] RV disease but at increased risk of P[6] disease, in a manner that is independent of secretor status. We further demonstrate that Lewis negative children have increased anti-RV IgA levels, but only among non-secretors, an effect that we hypothesize we are able to observe due to comparatively high rates of P[6] disease in our study populations. Among secretors, we further show that the ABO blood group is an important secondary determinant of disease risk, with non-O blood group increasing the risk of disease. In addition, we identify a novel genetic determinant of RV immunity in unvaccinated Ugandan children at chromosome 2q12. We further validate this locus as a determinant of RV immunity and disease, demonstrating significant effects on anti-RV immunity in vaccinated Zambian infants, as well as hospitalised RV disease and all-cause diarrhoeal disease with dehydration in Kenyan and Ugandan children respectively. A key feature of our data is that all four of the genetic mediators of RV disease and immunity identified here, namely, secretor

status, Lewis, ABO and chromosome 2q12, show evidence of effects dependent on RV P-type. As such, the effect of these variants on disease risk and immunity will critically depend upon the prevailing burden of RV strains in a given population.

The way in which the RV-associated genetic loci interact with RV disease and immunity is informative. Non-secretors lack an attachment factor for P[8] and P[4] RV, and are therefore at reduced risk of severe disease, but also have reduced anti-RV immunity at age 2–3 years, presumably secondary to a reduced frequency of RV infections in infancy/childhood. This is important as it highlights that any assessment of immunity relying on anti-RV IgA levels needs to consider secretor status, Lewis and ABO blood group, as a proportion of children with low serum anti-RV IgA levels will be at reduced (rather than increased) risk of disease. By contrast, genetic variation at chromosome 2q12 increased anti-RV IgA levels in unvaccinated Ugandan children and vaccinated Zambian infants, but reduced risk of hospitalised RVGE disease in Kenya and all-cause diarrhoea with dehydration in Uganda. In keeping with this effect on disease risk being secondary to acquisition of anti-RV immunity, this protective effect on hospitalised disease is restricted to older children in Kenya. The demonstration of a shared genetic locus mediating enhanced anti-RV IgA responses and reduced risk of RV disease, provides support for a model in which anti-RV IgA is causal of protection against RV disease in young children.

The RV-associated variant at chromosome 2q12 is located is located in the 5th intron of the *EDAR* gene. Interestingly, the *EDAR* locus has been long-recognised as a site of recent positive selection in humans[57]. The mediators of this selection pressure remain unclear, although selected *EDAR* alleles modulate hair thickness and sweat gland density[58], as well as increased ductal branching in breast tissue[59]. Infectious diseases are common determinants of survival to reproductive age, and as such have been frequently identified as agents driving selection pressure[60]. A role for genetic variation at *EDAR* in host defence against RV disease provides an alternative hypothesis for the recent selection seen at this locus. Using eQTL mapping data in immune cells, we demonstrate that, rather than modifying *EDAR* expression or function, the RV-associated variation at chromosome 2q12 is an eQTL for *LIMS1* expression in naïve and stimulated monocytes. Exploring the effect of *LIMS1*, and its eQTL, on cytokine secretion in stimulated monocytes, we identify that *LIMS1* RNA expression is associated with HGF secretion in IFNγ-stimulated monocytes. Taken together, our data suggests a model in which a common genetic variant, rs59241810:T, increases anti-RV IgA responses following RV infection and vaccination, thereby protecting children against subsequent episodes of rotavirus diarrhoea. This effect is mediated through the regulation of *LIMS1* expression in monocytes, with carriers of the rs59241810:T allele having reduced *LIMS1* RNA expression, which in turn limits the downregulation of HGF secretion following IFNγ stimulation. HGF is a polyfunctional cytokine[61], which in addition to its roles in growth, motility and differentiation in hepatocytes and other cells, has well-defined roles in B cell proliferation and antibody production. The HGF receptor, MET, is expressed on B cells[62] and follicular dendritic cell-derived HGF acts to enhance antibody production on B cell activation in germinal centres[63,64]. In addition, transforming growth factor-β (TGFβ) drives IgA class-switching in B cells[65,66], and there is evidence of cross-talk between TGFβ and HGF signalling pathways[67], suggesting a potential role for HGF in IgA production specifically.

A key motivation for improving our understanding of the genetic determinants of RV disease and immunity is to allow the design of improved control strategies for RV disease, and to understand the mechanisms underlying differential vaccine efficacy across settings. The study samples investigated here define disease risk and immunity in populations with significant ongoing force of RV infection, either prior to, or despite, vaccine implementation. In defining the genetics of RV disease risk and immune acquisition in these populations, we advance our understanding of the ways that host genetics will shape risk of RV disease and anti-RV immunity in populations with significant residual RV disease despite immunisation. Our data have important implications for how we interpret vaccine failure, and in particular the use of anti-RV IgA as a correlate of protection. Critically, they highlight that the interpretation of anti-RV IgA levels, and how they relate to risk of RV disease, will depend on human genetic variation and the prevalent RV strains in the population studied. Integrating these data with future GWAS of RV vaccine failure and vaccine-derived anti-RV immunity will be central to constructing more complete models of the contribution of human genetics of RV disease and its prevention.

Our study has some limitations. We have defined genetics of RV disease largely in unvaccinated children. It is plausible, however, that the genetics of vaccine failure are distinct from that of unvaccinated disease risk, and it will be important to explore whether this is the case in future studies. African-ancestry populations are critically underrepresented in genomic studies, including eQTL studies, and regulatory variation mapped by eQTL studies are commonly context dependent. The RV-associated genetic variation that we identify on chromosome 2 is infrequent in non-African ancestry populations, and while we identify that this locus regulates *LIMS1* expression in monocytes, richer eQTL mapping data from diverse populations would likely allow a more complete characterisation of the tissue and biological contexts in which that regulation may operate and thus modify RV immunity and disease.

In summary, we present the first genome-wide association studies of RV disease and immunity. Our data reveal a complex interplay between common human genetic variation, viral genotype, naturally-acquired immunity and disease risk. We highlight that interpretation of serum anti-RV IgA requires consideration of human genetic variation in the study population and identify genetic regulation of *LIMS1* expression and downstream monocyte HGF secretion as a novel mediator of RV immunity and disease. These insights are of fundamental importance in advancing our understanding of current anti-RV vaccine performance and provide new avenues to inform the design of more effective anti-RV vaccines in the future.

Methods

GWAS sample recruitment & ethics

To investigate the genetic determinants of naturally-acquired immunity to RV infection, we leveraged stored samples from healthy Ugandan children recruited to the Entebbe Mother and Baby Study (EMaBS) birth cohort in 2003–2006, prior to RV vaccine rollout in Uganda. EMaBS was established as a randomized, double-blind, placebo-controlled trial of anthelmintic treatment during pregnancy and early childhood, designed to investigate the impact of helminth infection on response to immunization and susceptibility to infection and other diseases in Ugandan children[68]. Episodes of diarrhoea and dysentery in children recruited to the study were recorded by study physicians when the child presented to the study clinic with an illness. Diarrhoeal episodes were

defined as per maternal report, with dysentery defined as diarrhoea with stools containing blood. Episodes of prolonged diarrhoea were defined as loose, frequent stools persisting for more than 14 days. Dehydration was defined as at least 2 of the following clinical signs; restlessness, lethargy, increased thirst, drinking poorly, sunken eyes and reduced skin turgor. In addition to clinical and demographic data, high-density genome-wide genotyping data has been generated for 1,390 EMaBS participants[69]. From these samples, we randomly selected serum samples collected at 2 or 3 years of age for anti-RV IgA quantification (n = 745).

To investigate the genetic determinants of hospitalised RVGE, we recruited children with RVGE requiring hospitalisation at Kilifi County Hospital (KCH) and community-based controls in Kilifi, Kenya. Prospective RVGE surveillance has been carried out in children < 13 years admitted to KCH, Kenya, since 2003[70]. Children (n = 580) admitted to KCH with RVGE between 14/01/2002 and 31/05/2014 prior to Kenyan RV vaccine introduction (July 2014) were included as cases. Healthy community controls (n = 1,615) were recruited at 3–12 months of age from the same population (2006–2008), for whom high-density genome-wide genotyping data has been generated previously[71].

To validate novel genetic determinants of anti-RV IgA in a vaccinated population, we leveraged samples from Zambian children recruited to a randomised controlled trial (ROVAS-2) of a booster dose of anti-RV vaccine (Rotarix) at 9 months[39]. Infants aged 6–12 weeks were recruited to the trial between 13th September 2018 and 15th November 2018. All infants received two doses of Rotarix from 6 weeks of age according to the routine Zambian expanded immunisation schedule. Infants were then randomised in a 1:1 ratio to receive an additional dose of Rotarix at 9 months of age or no additional vaccine. As part of this trial, plasma anti-RV IgA was quantified in participants at 12 months of age.

Participants were recruited to the original studies following written, informed consent from the child's parent/guardian. The study protocol for the genetic association study in Kenya was reviewed by the Kenya Medical Research Institute Scientific Ethics and Research Review Unit. The study protocol for the genetic association study in Uganda was reviewed and approved by the Research and Ethics Committee of the Uganda Virus Research Institute and the Uganda National Council for Science and Technology and we notified the London School of Hygiene & Tropical Medicine Observational/Interventions Research Ethics Committee. The study protocol for the genetic association study in Zambia was reviewed by the University of Zambia Biomedical Research Ethics Committee. The study as a whole was reviewed by the Oxford Tropical Research Ethics Committee.

Anti-RV IgA ELISA

As previously described[72], serum and plasma anti-RV IgA concentrations were determined in stored serum samples by sandwich enzyme-linked immunosorbent assay (ELISA). In brief, 96 well plates were coated with hyperimmune rabbit serum, prior to washing and coating with cell lysates from MA104 cells infected with bovine G6P[5] (strain WC3). Following incubation with study serum samples and a standardised pool of human serum, bound anti-RV-specific IgA was detected with *o*-phenylenediamine oxidation using peroxidase-conjugated rabbit anti-human IgA.

Study sample anti-RV IgA concentrations were derived by comparison to a standard curve of serial dilutions from the human pooled serum, and are reported in arbitrary units (IU/ml). Samples were run in duplicate, and in triplicate if the sample coefficient of variance (CV) exceeded 20%. If, after running in triplicate, the CV remained > 20% then the sample was excluded from downstream analysis. Study samples consistently below the assay limit of quantification were assigned a value of 0 IU/ml. We defined seropositivity as an anti-RV IgA ≥ 20 IU/ml[34].

Rotavirus identification and diversity

Cases of RVGE among admissions to KCH were identified among children presenting with a history of diarrhoea, using an enzyme immunoassay (IDEIA Rotavirus, Dako) to detect rotavirus-specific antigen in stool. Rotavirus G and P genotypes were determined with nested, genotype-specific reverse transcription PCRs[73] or by di-deoxy Sanger nucleotide sequencing[36] and analysis of the VP7 and VP4 genes.

Genotyping & imputation

Kenyan RVGE case samples and Zambian study samples were genome-wide genotyped using the Illumina H3Africa consortium array. Ugandan study samples and Kenyan control samples were genotyped with the Illumina Omni2.5M array. Following genotyping, we excluded SNPs failing to meet the following quality control (QC) metrics: MAF > 1%, Hardy-Weinberg equilibrium (HWE) $P>1x10^{-20}$, SNP call rate > 0.97, plate effect $P>1x10^{-6}$. We further excluded one sample of pairs with relatedness coefficients > 0.2, and samples with outlier call rates, heterozygosity and population structure as estimated by principal components (PC) of genome-wide genotyping data (Supplementary Fig. 4). Calculation of QC metrics and sample/SNP exclusion was performed in QCTOOL (https://www.well.ox.ac.uk/~gav/qctool_v2/index.html) and PLINK[74].

For the Kenyan RVGE GWAS, where cases and controls were genotyped on different platforms, following initial QC we restricted downstream imputation to SNPs genotyped on both platforms. To facilitate QC for these samples, 16 Kenyan control samples were genotyped on both arrays, and SNPs with genotyping calls not concordant on both platforms were excluded. Genotype PCs capture self-reported ethnicity in both populations and are non-differential with respect to case-control status in the Kenyan samples (Supplementary Fig. 4).

Following QC, 2,013,220 and 1,017,849 autosomal SNPs were taken forward for genome-wide imputation in the Ugandan and Kenyan samples respectively. Typed genotypes were phased using Eagle2[75] and missing genotype information imputed using minimac[76], using 1000 Genomes Project Phase 3 as a reference panel[77], as implemented in the Michigan Imputation Server[78]. Following imputation, we retained common (MAF > 1%), well-imputed ($r^2 > 0.7$) SNPs, without evidence of departure from HWE ($P>1x10^{-10}$).

Genetic association analysis

SNPs and samples passing QC were taken forward for genetic association analysis using SNPTEST2[79]. For the Kenyan RVGE case-control GWAS, association analysis was performed by logistic regression, including sex and the major four genotype PCs to control for confounding variation. For the GWAS of anti-RV-specific serum IgA acquisition in Ugandan children, association analysis was performed by linear regression with rank-based inverse normal-transformed anti-RV-specific serum IgA concentrations. Sex, age and the major two genotype PCs were included to control for confounding variation. In both studies, analysis was performed using an additive model of association. For validation of the rs59241810 locus in the Zambian samples, we correlated genotype with rank-based inverse normal-transformed anti-RV-specific serum IgA concentrations at 12 months of age using linear regression. Study arm, sex, formula feeding and secretor status were included in the model as covariates. Inspection of genotype PCs in Zambia was not suggestive of confounding variation secondary to population structure and these were not included in the model. Downstream analysis and visualisation were performed in R (version 4.2.2).

eQTL and cytokine secretion samples

The samples used for eQTL mapping in monocytes and cytokine secretion assays have been described previously[49,56,80]. In brief, we isolated PBMCs from healthy European-ancestry donors with density gradient centrifugation (Ficoll), followed by isolation of primary monocytes using magnetic $CD14^+$ isolation kits (miltenyi). Isolated monocytes were cultured at 500,000 cells/ml in RPMI supplemented with 5% FCS, Penicillin/Streptomycin and L-Glutamine. Monocytes were left untreated (n = 414) or incubated with LPS (20ng/ml, Ultrapure LPS, Invivogen) for 2 (n = 261) and 24 (n = 322) hours or IFNγ (20ng/ml, R&D Systems) for 24 hours (n = 367). Following incubation, total RNA was extracted (RNeasy Mini Kit, Qiagen) and culture supernatants frozen at -80°C for subsequent cytokine quantification[49]. In addition, we performed eQTL mapping in untreated monocytes from a second donor collection of healthy European-ancestry adults (n = 192) using the same monocyte isolation and culture protocols.

RNA quantification and genotyping

For the first donor collection (n = 414 untreated monocytes and LPS/IFNγ-stimulated samples), RNA was quantified using Illumina HumanHT-12 v4 BeadChip gene expression arrays as described previously[49]. For the second donor collection (n = 192 untreated monocytes), RNA was quantified with paired-end 100bp sequencing on Illumina Hiseq-4000 machines, with read alignment using HISAT2[81] and gene expression quantification with HTSeq[82]. Genome-wide genotyping was performed with Illumina HumanOmniExpress-12v1.0 (first donor collection) and Affymetrix UK Biobank Axiom (second donor collection) arrays, with imputation of unassayed variants using SHAPEIT2/IMPUTE2[83,84] with 1000 Genomes Project Phase 3 as a reference panel[77].

Cytokine quantification

Cytokine quantification methods have been described previously[56]. In brief, we used the ProcartaPlex Immunoassay (Affymetrix / Thermo-Fisher, Massachusetts, USA) to quantify the concentrations of 50 cytokines (28 cytokines included post QC, Supplementary Table 9) in monocyte culture supernatants. The assay was performed in the Thermo-Fisher laboratories (Bender MedSystems GmbH, Vienna, Austria) using recommended conditions and controls for all 50 cytokines provided by Affymetrix. We assayed cytokine secretion in 1,000 samples with the assay handlers blinded to sample treatment status, and samples randomised across 96-welled plates: 77 untreated monocytes, 265 LPS (2 hours), 316 LPS (24 hours) and 342 IFNg (24 hours). We excluded cytokines from downstream analysis where the cytokine was not detectable in ≥ 10% of samples in ≥ 1 condition (n = 10), and assayed cytokines in duplicate for 36 samples, excluding a further 12 cytokines from downstream analysis with poor concordance between replicates.

eQTL mapping & colocalization

We mapped *cis* eQTLs, correlating SNP genotypes within 1Mb of a gene's transcription start site with normalised gene expression estimates. Correlation was calculated by linear regression, including 25 principal components of gene expression data in each condition to account for confounding variation.

In addition, we downloaded eQTL summary statistics[85] from blood/immune cell eQTL studies: whole blood[86] (n = 614), PBMCs[48] (n = 1,012), NK cells[50] (n = 245), B cells[51] (n = 281), T cells[46] (n = 167), CD4 + T cells[44,45,87] (3 studies; n = 290, 167, 280) and CD8 + T cells[45,87] (2 studies; n = 277, 269). We also downloaded summary statistics for studies describing pQTL mapping of plasma EDAR abundance among individuals of African ancestry within the UK biobank[43] (n = 889). To test if the mapped eQTLs share a causal variant with the anti-RV IgA GWAS locus, we used a Bayesian approach implemented in Coloc v5.1.0.1[88], using the coloc.abf() command with default settings. We considered a posterior probability of colocalization (PP4) > 0.8 to be significant.

Cytokine association analysis

Following quality control, we normalised cytokine concentrations with inverse normal transformations, normalising within each batch and condition. Prior to association analysis, we further regressed out the effects of age and sex on each cytokine, and for IFNγ-stimulated samples the measured level of IFNγ. We correlated cytokine concentrations with *LIMS1* expression by linear regression, correcting for 25 principal components of gene expression data as above. We corrected for multiple testing using the Benjamini & Hochberg method, and considered an FDR < 0.05 to be significant.

Declarations

Acknowledgements

We thank Francesca Liu and Markus Miholts from Thermo-Fisher for their help planning, performing and analysing the ProcartaPlex assay. JJG is supported by a Wellcome Career Development Award (311117). JJG and AJM were supported by NIHR Academic Clinical Lectureships. The work was supported by an Academy of Medical Sciences Starter Grant (SGL021\1036) and a Wellcome Trust Clinical PhD Fellowship (102342) to JJG, a Wellcome Trust Career Development Award (226535) to BPF and the CONACYT-Newton Fund (FONCICYT/50/2016). JR was supported by Wellcome International Master's Fellowship (200070). TNW was supported by a Wellcome Trust Senior Research Fellowship (202800). During this work, AVSH was supported by a Wellcome Trust Senior Investigator Award (HCUZZ0) and by a European Research Council advanced grant (294557). JCK is supported by a Wellcome Trust Investigator

Award (204969), the National Institute for Health Research (NIHR) Oxford Biomedical Research Centre and the Chinese Academy of Medical Sciences (CAMS) Innovation Fund for Medical Science (grant 2018-I2M-2-002). The research was supported by the Wellcome Trust Core Award Grant Number 203141/Z/16/Z with additional support from the NIHR Oxford BRC. The EMaBS was supported by the Wellcome Trust grant numbers 064693 and 079110, and conducted at the MRC/UVRI and LSHTM Uganda Research Unit which is jointly funded by the UK Medical Research Council (MRC) part of UK Research and Innovation (UKRI) and the UK Foreign, Commonwealth and Development Office (FCDO) under the MRC/FCDO Concordat agreement and is also part of the EDCTP2 programme supported by the European Union. This publication uses genotyping data from the MalariaGEN consortial project, as described in Malaria Genomic Epidemiology Network, et al. Nature Communications, 2019 (https://doi.org/10.1038/s41467-019-13480-z). The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health. This paper was published with the permission of the Director of KEMRI.

References

1. Kotloff KL et al (2013) Burden and aetiology of diarrhoeal disease in infants and young children in developing countries (the Global Enteric Multicenter Study, GEMS): a prospective, case-control study. Lancet 382:209–222
2. GBD 2023 Diarrhoeal Disease and Enteric Infectious Diseases Collaborators (2026) Global burden of enteric infectious diseases, diarrhoeal diseases, and corresponding aetiologies, 1990–2023: a systematic analysis for the Global Burden of Disease Study 2023. Lancet Infect Dis. 10.1016/s1473-3099(26)00194-5
3. Richardson V et al (2010) Effect of rotavirus vaccination on death from childhood diarrhea in Mexico. N Engl J Med 362:299–305
4. Carmo GMI, do et al (2011) Decline in diarrhea mortality and admissions after routine childhood rotavirus immunization in Brazil: a time-series analysis. PLoS Med 8:e1001024
5. Patel MM, Glass R, Desai R, Tate JE, Parashar UD (2012) Fulfilling the promise of rotavirus vaccines: how far have we come since licensure? Lancet Infect Dis 12:561–570
6. Vesikari T et al (2007) Efficacy of human rotavirus vaccine against rotavirus gastroenteritis during the first 2 years of life in European infants: randomised, double-blind controlled study. Lancet 370:1757–1763
7. Prunas O et al (2025) Global estimates of rotavirus vaccine efficacy and effectiveness: a rapid review and meta-regression analysis. eClinicalMedicine 81:103122
8. Bonkoungou IJO et al (2018) Impact and effectiveness of pentavalent rotavirus vaccine in children < 5 years of age in Burkina Faso. Vaccine 36:7170–7178
9. Parker EP et al (2018) Causes of impaired oral vaccine efficacy in developing countries. Future Microbiol 13:97–118
10. Desselberger U, Rotaviruses (2014) Virus Res 190:75–96
11. Hu L et al (2012) Cell attachment protein VP8* of a human rotavirus specifically interacts with A-type histo-blood group antigen. Nature 485:256–259
12. Kelly RJ, Rouquier S, Giorgi D, Lennon GG, Lowe JB (1995) Sequence and Expression of a Candidate for the Human Secretor Blood Group α(1,2)Fucosyltransferase Gene (FUT2) HOMOZYGOSITY FOR AN ENZYME-INACTIVATING NONSENSE MUTATION COMMONLY CORRELATES WITH THE NON-SECRETOR PHENOTYPE (⁎). J Biol Chem 270:4640–4649
13. Barbé L et al (2018) Histo-blood group antigen-binding specificities of human rotaviruses are associated with gastroenteritis but not with in vitro infection. Sci Rep 8:12961
14. Nordgren J et al (2014) Both Lewis and secretor status mediate susceptibility to rotavirus infections in a rotavirus genotype-dependent manner. Clin Infect Dis 59:1567–1573
15. Zhang X-F et al (2016) P[8] and P[4] Rotavirus Infection Associated with Secretor Phenotypes Among Children in South China. Sci Rep 6:34591
16. Sun X et al (2016) Rotavirus infection and histo-blood group antigens in the children hospitalized with diarrhoea in China. Clin Microbiol Infect 22:740e1–740e3
17. Pollock L et al (2018) Non-secretor histo-blood group antigen phenotype is associated with reduced risk of clinical rotavirus vaccine failure in Malawian infants. Clin Infect Dis. 10.1093/cid/ciy1067
18. Lee B et al (2018) Histo-Blood Group Antigen Phenotype Determines Susceptibility to Genotype-Specific Rotavirus Infections and Impacts Measures of Rotavirus Vaccine Efficacy. J Infect Dis 217:1399–1407
19. Trang NV et al (2014) Association between norovirus and rotavirus infection and histo-blood group antigen types in Vietnamese children. J Clin Microbiol 52:1366–1374
20. Payne DC et al (2015) Epidemiologic Association Between FUT2 Secretor Status and Severe Rotavirus Gastroenteritis in Children in the United States. JAMA Pediatr 169:1040–1045
21. Pérez-Ortín R et al (2019) Histo-Blood Group Antigens in Children with Symptomatic Rotavirus Infection. Viruses 11:339
22. Imbert-Marcille B-M et al (2014) A FUT2 gene common polymorphism determines resistance to rotavirus A of the P[8] genotype. J Infect Dis 209:1227–1230
23. Ahluwalia TS et al (2020) FUT2–ABO epistasis increases the risk of early childhood asthma and Streptococcus pneumoniae respiratory illnesses. Nat Commun 11:6398
24. Ayouni S et al (2015) Rotavirus P[8] Infections in Persons with Secretor and Nonsecretor Phenotypes, Tunisia. Emerg Infect Dis 21:2055–2058

25. Rodríguez-Díaz J et al (2017) Relevance of secretor status genotype and microbiota composition in susceptibility to rotavirus and norovirus infections in humans. Sci Rep 7:45559
26. Günaydın G, Nordgren J, Sharma S, Hammarström L (2016) Association of elevated rotavirus-specific antibody titers with HBGA secretor status in Swedish individuals: The FUT2 gene as a putative susceptibility determinant for infection. Virus Res 211:64–68
27. Boniface K, Byars SG, Cowley D, Kirkwood CD, Bines JE (2020) Human Neonatal Rotavirus Vaccine (RV3-BB) Produces Vaccine Take Irrespective of Histo-Blood Group Antigen Status. J Infect Dis 221:1070–1078
28. Armah GE et al (2018) Rotavirus Vaccine Take in Infants Is Associated With Secretor Status. J Infect Dis 86:4833
29. Kazi AM et al (2017) Secretor and Salivary ABO Blood Group Antigen Status Predict Rotavirus Vaccine Take in Infants. J Infect Dis 215:786–789
30. Bucardo F et al (2018) The Lewis A phenotype is a restriction factor for Rotateq and Rotarix vaccine-take in Nicaraguan children. Sci Rep 8:1502
31. Chauwa A et al (2023) Maternal and Infant Histo-Blood Group Antigen (HBGA) Profiles and Their Influence on Oral Rotavirus Vaccine (RotarixTM) Immunogenicity among Infants in Zambia. Vaccines 11:1303
32. Pollock L et al (2022) Plasma rotavirus-specific IgA and risk of rotavirus vaccine failure in infants in Malawi. Clin Infect Dis : Off Publ Infect Dis Soc Am 75:41–46
33. Lewnard JA et al (2017) Naturally Acquired Immunity Against Rotavirus Infection and Gastroenteritis in Children: Paired Reanalyses of Birth Cohort Studies. J Infect Dis 216:317–326
34. Baker JM et al (2019) Postvaccination Serum Antirotavirus Immunoglobulin A as a Correlate of Protection Against Rotavirus Gastroenteritis Across Settings. J Infect Dis 222:309–318
35. Lee B et al (2018) Rotavirus-specific IgA Responses Are Impaired and Serve as a Sub-Optimal Correlate of Protection among Infants in Bangladesh. Clin Infect Dis 382:209
36. Owor BE et al (2018) Molecular characterization of rotavirus group A strains circulating prior to vaccine introduction in rural coastal Kenya, 2002–2013. Wellcome Open Res 3:150
37. Fry AE et al (2008) Common variation in the ABO glycosyltransferase is associated with susceptibility to severe Plasmodium falciparum malaria. Hum Mol Genet 17:567–576
38. Dóró R et al (2014) Review of global rotavirus strain prevalence data from six years post vaccine licensure surveillance: Is there evidence of strain selection from vaccine pressure? Infect Genet Evol 28:446–461
39. Laban NM et al (2023) Evaluation of ROTARIX® Booster Dose Vaccination at 9 Months for Safety and Enhanced Anti-Rotavirus Immunity in Zambian Children: A Randomised Controlled Trial. Vaccines 11:346
40. Crawford SE et al (2017) Rotavirus infection. Nat Rev Dis Prim 3:17083
41. Maurano MT et al (2012) Systematic localization of common disease-associated variation in regulatory DNA. Science 337:1190–1195
42. Gschwind AR et al (2023) An encyclopedia of enhancer-gene regulatory interactions in the human genome. *bioRxiv* 2023.11.09.563812 10.1101/2023.11.09.563812
43. Sun BB et al (2023) Plasma proteomic associations with genetics and health in the UK Biobank. Nature 1–10. 10.1038/s41586-023-06592-6
44. Chen L et al (2016) Genetic Drivers of Epigenetic and Transcriptional Variation in Human Immune Cells. Cell 167:1398–1414e24
45. Momozawa Y et al (2018) IBD risk loci are enriched in multigenic regulatory modules encompassing putative causative genes. Nat Commun 9:2427
46. Gutierrez-Arcelus M et al (2013) Passive and active DNA methylation and the interplay with genetic variation in gene regulation. eLife 2:e00523
47. Wen J et al (2024) Gene expression and splicing QTL analysis of blood cells in African American participants from the Jackson Heart Study. *GENETICS* 228, iyae098
48. Wen J et al (2023) Gene Expression and Splicing QTL Analysis of Blood Cells in African American Participants from the Jackson Heart Study. *bioRxiv* 2023.04.26.538455 10.1101/2023.04.26.538455
49. Fairfax BP et al (2014) Innate immune activity conditions the effect of regulatory variants upon monocyte gene expression. Science 343:1246949–1246949
50. Gilchrist JJ et al (2022) Natural Killer cells demonstrate distinct eQTL and transcriptome-wide disease associations, highlighting their role in autoimmunity. Nat Commun 13:4073
51. Fairfax BP et al (2012) Genetics of gene expression in primary immune cells identifies cell type-specific master regulators and roles of HLA alleles. Nat Genet 44:502–510
52. Xu H, Cao H, Xiao G (2016) Signaling via PINCH: Functions, binding partners and implications in human diseases. Gene 594:10–15
53. Berton G, Lowell CA (1999) Integrin Signalling in Neutrophils and Macrophages. Cell Signal 11:621–635
54. Ahmed AU, Yim HCH, Alorro M, Ernst M, Williams BR (2017) G. Integrin-Linked Kinase Expression in Myeloid Cells Promotes Inflammatory Signaling during Experimental Colitis. J Immunol 199:2128–2139
55. Ahmed AU, Sarvestani ST, Gantier MP, Williams BRG, Hannigan GE (2014) Integrin-linked Kinase Modulates Lipopolysaccharide- and Helicobacter pylori-induced Nuclear Factor κB-activated Tumor Necrosis Factor-α Production via Regulation of p65 Serine 536 Phosphorylation*. J Biol Chem 289:27776–27793
56. Gilchrist JJ et al (2026) Genetic determinants of cytokine production in activated human monocytes. *medRxiv* 2026.05.08.26352736 10.64898/2026.05.08.26352736
57. Carlson CS et al (2005) Genomic regions exhibiting positive selection identified from dense genotype data. Genome Res 15:1553–1565

58. Kamberov YG et al (2013) Modeling Recent Human Evolution in Mice by Expression of a Selected EDAR Variant. Cell 152:691–702
59. Hlusko LJ et al (2018) Environmental selection during the last ice age on the mother-to-infant transmission of vitamin D and fatty acids through breast milk. P Natl Acad Sci Usa 115:E4426–E4432
60. Karlsson EK, Kwiatkowski DP, Sabeti PC (2014) Natural selection and infectious disease in human populations. Nat Rev Genet 15:379–393
61. Boros P, Miller CM (1995) Hepatocyte growth factor: a multifunctional cytokine. Lancet 345:293–295
62. Voort R, van der et al (1997) Paracrine Regulation of Germinal Center B Cell Adhesion through the c-Met–Hepatocyte Growth Factor/Scatter Factor Pathway. J Exp Med 185:2121–2131
63. Skibinski G, Skibinska A, James K (2001) The role of hepatocyte growth factor and its receptor c-met in interactions between lymphocytes and stromal cells in secondary human lymphoid organs. Immunology 102:506–514
64. Tjin EPM et al (2005) Follicular Dendritic Cells Catalyze Hepatocyte Growth Factor (HGF) Activation in the Germinal Center Microenvironment by Secreting the Serine Protease HGF Activator. J Immunol 175:2807–2813
65. Coffman RL, Lebman DA, Shrader B (1989) Transforming growth factor beta specifically enhances IgA production by lipopolysaccharide-stimulated murine B lymphocytes. J Exp Med 170:1039–1044
66. Sonoda E et al (1989) Transforming growth factor beta induces IgA production and acts additively with interleukin 5 for IgA production. J Exp Med 170:1415–1420
67. Mori S et al (2004) TGF-β and HGF transmit the signals through JNK-dependent Smad2/3 phosphorylation at the linker regions. Oncogene 23:7416–7429
68. Ndibazza J et al (2012) Impact of anthelminthic treatment in pregnancy and childhood on immunisations, infections and eczema in childhood: a randomised controlled trial. PLoS ONE 7:e50325
69. Lule SA et al (2019) A genome-wide association and replication study of blood pressure in Ugandan early adolescents. Mol Genet Genom Med 7:e00950
70. Nokes DJ et al (2008) Incidence and clinical characteristics of group A rotavirus infections among children admitted to hospital in Kilifi, Kenya. PLoS Med 5:e153
71. Network MGE (2019) Insights into malaria susceptibility using genome-wide data on 17,000 individuals from Africa, Asia and Oceania. Nat Commun 10:5732–5719
72. Paul A et al (2014) Human and bovine rotavirus strain antigens for evaluation of immunogenicity in a randomized, double-blind, placebo-controlled trial of a single dose live attenuated tetravalent, bovine-human-reassortant, oral rotavirus vaccine in Indian adults. Vaccine 32:3094–3100
73. Nokes DJ et al (2010) Rotavirus Genetic Diversity, Disease Association, and Temporal Change in Hospitalized Rural Kenyan Children. J Infect Dis 202:S180–S186
74. Purcell S et al (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 81:559–575
75. Loh P-R et al (2016) Reference-based phasing using the Haplotype Reference Consortium panel. Nat Genet 48:1443–1448
76. Fuchsberger C, Abecasis GR, Hinds D (2015) A. minimac2: faster genotype imputation. Bioinformatics 31:782–784
77. Auton A et al (2015) A global reference for human genetic variation. Nature 526:68–74
78. Das S et al (2016) Next-generation genotype imputation service and methods. Nat Genet 48:1284–1287
79. Marchini J, Howie B, Myers S, McVean G, Donnelly P (2007) A new multipoint method for genome-wide association studies by imputation of genotypes. Nat Genet 39:906–913
80. Gilchrist JJ et al (2024) Characterization of the genetic determinants of context-specific DNA methylation in primary monocytes. Cell Genom 4:100541
81. Kim D, Langmead B, Salzberg SL (2015) HISAT: a fast spliced aligner with low memory requirements. Nat Methods 12:357–360
82. Anders S, Pyl PT, Huber W (2015) HTSeq—a Python framework to work with high-throughput sequencing data. Bioinformatics 31:166–169
83. Delaneau O, Marchini J, Zagury J-F (2012) A linear complexity phasing method for thousands of genomes. Nat Methods 9:179–181
84. Marchini J, Howie B (2010) Genotype imputation for genome-wide association studies. Nat Rev Genet 11:499–511
85. Kerimov N et al (2021) A compendium of uniformly processed human gene expression and splicing quantitative trait loci. Nat Genet 53:1290–1299
86. Castel SE et al (2025) A map of blood regulatory variation in South Africans enables GWAS interpretation. Nat Genet 57:1628–1637
87. Kasela S et al (2017) Pathogenic implications for autoimmune mechanisms derived by comparative eQTL analysis of CD4 + versus CD8 + T cells. PLoS Genet 13:e1006643
88. Giambartolomei C et al (2014) Bayesian Test for Colocalisation between Pairs of Genetic Association Studies Using Summary Statistics. PLoS Genet 10:e1004383

Figures

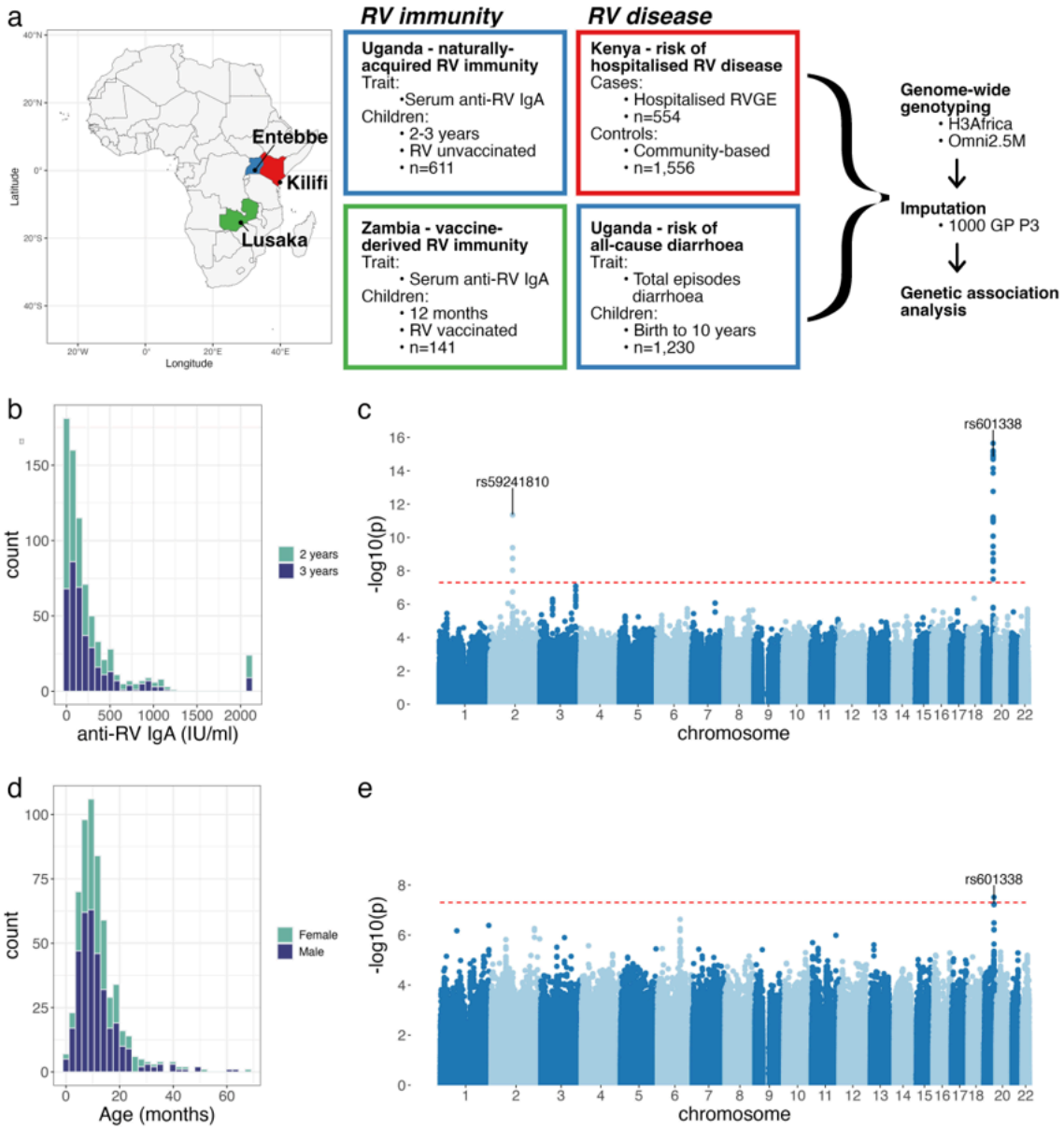


Figure 1

Genetic association studies of rotavirus immunity & disease. **a,** Study schematic for genetic association studies of rotavirus (RV) disease and immunity. RVGE, rotavirus gastroenteritis. **b,** Distribution of serum anti-RV IgA levels in Ugandan children aged 2-3 years. Samples were collected between 2004 and 2011, and none of the children had received anti-RV vaccination. Individuals are coloured according to their age at sampling. **c,** Manhattan plot of anti-RV IgA in Ugandan children aged 2 and 3 years prior to rotavirus vaccine rollout (n=611). The major causal determinant of secretor status (rs601338) and the novel anti-RV IgA-associated locus on chromosome 2 (rs59241810) are labelled. **d,** Age distribution of Kenyan children (n=554) hospitalised at Kilifi County Hospital between 2002 and 2014. Individuals are coloured by sex. **e,** Manhattan plot of RV disease risk in Kenyan children prior to rotavirus vaccine rollout (554 cases, 1,556 controls). The major causal determinant of secretor status (rs601338) is labelled. Genome-wide significance (P=5x10^{-8}) is highlighted with the red, dashed line.

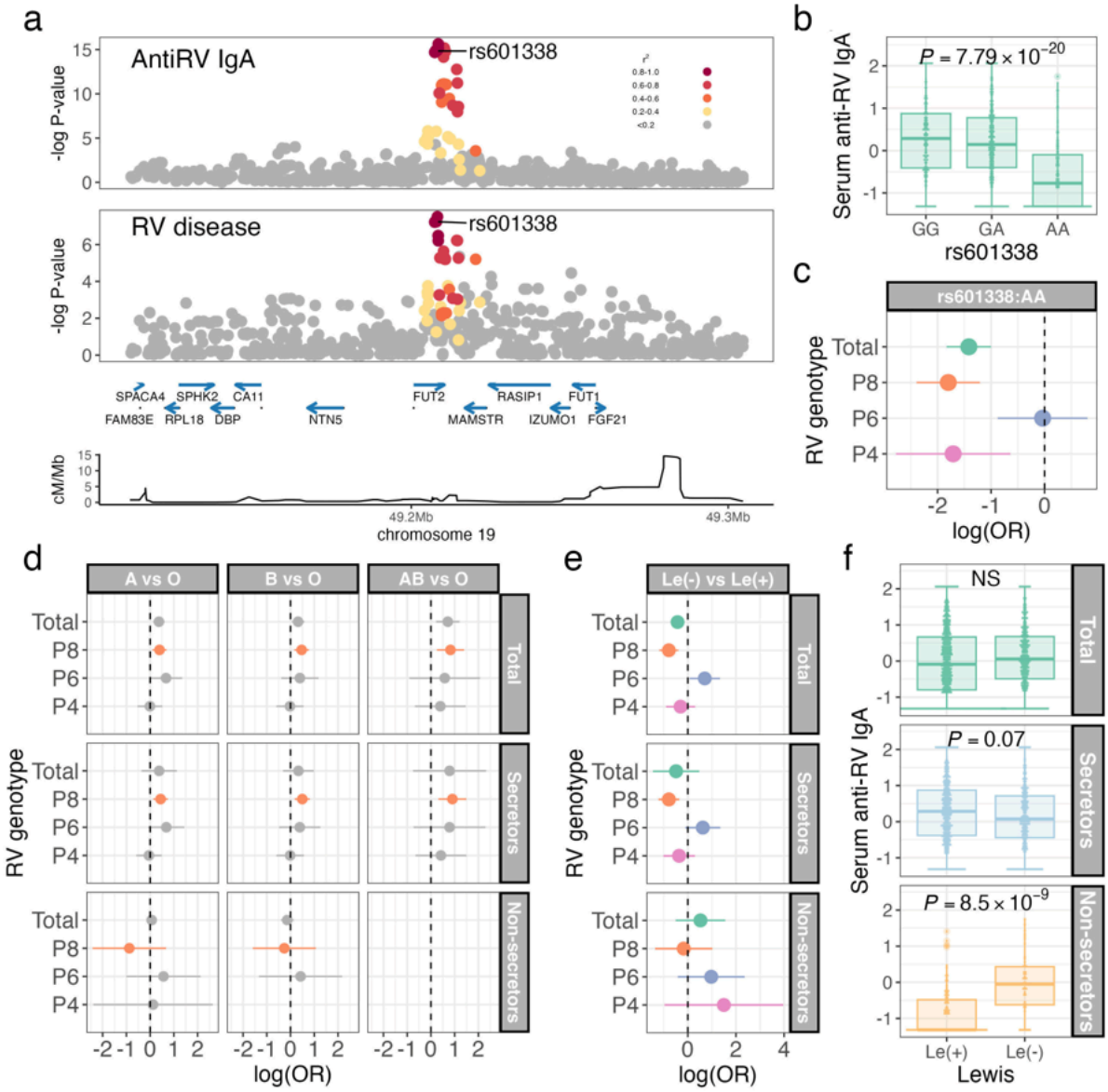


Figure 2

Secretor status, ABO and Lewis and their effects on rotavirus disease and immunity. **a,** Regional association plot of the secretor status association with serum anti-RV IgA in Ugandan children aged 2-3 years prior to rotavirus vaccine rollout (top), and hospitalised RV disease risk in Kenyan children prior to rotavirus vaccine rollout (bottom). SNPs are colour-coded according to linkage disequilibrium to the major determinant of secretor status in this population, rs601338. Protein-coding gene positions are depicted with blue arrows. Recombination rate is illustrated on the bottom track. **b,** The effect of rs601338 genotype on serum anti-RV IgA levels in Ugandan children. **c,** Effect estimates (and 95% confidence intervals) of rs601338:AA genotype (non-secretors) on hospitalised RV disease risk in Kenyan children stratified by rotavirus genotype. **d,** Effect estimates of blood group (compared to blood group O as baseline) on hospitalised RV disease risk in Kenyan children stratified by rotavirus genotype overall stratified by secretor status. A significant effect of non-O blood group on disease risk is seen in P[8] genotype disease (highlighted orange) among secretors. Where effect estimates are not plotted (e.g. for AB vs O blood group in non-secretors), this reflects inadequate sample numbers in this stratum. **e,** Effect estimates (and 95% confidence intervals) of Lewis negative phenotype on hospitalised RV disease risk in Kenyan children stratified by RV genotype in children overall and stratified by secretor status. **f,** The effect of Lewis phenotype on serum anti-RV IgA levels in Ugandan children overall and stratified by secretor status. Effects estimates for hospitalised disease associations are calculated by multinomial logistic regression adjusted for sex and 4 genetic principal components. Effects estimates for serum anti-RV IgA associations are calculated with linear regression adjusted for age at sampling, sex and 2 genetic principal components. Box and whisker plots: boxes depict the upper and lower quartiles of the data, and whiskers depict the range of the data excluding outliers (data-points >1.5x the interquartile range from the upper or lower quartiles).

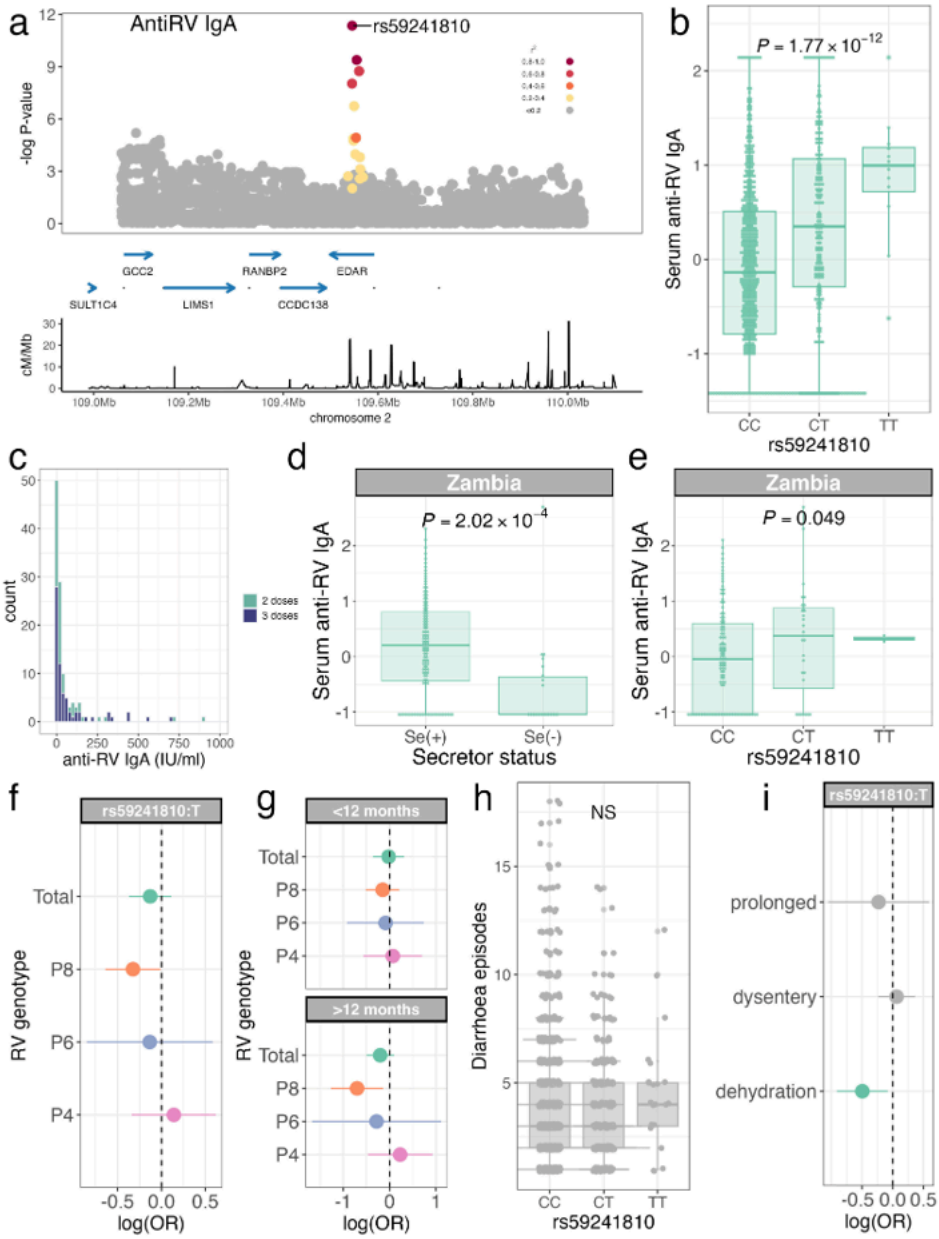


Figure 3

Genetic variation at chromosome 2q12, rotavirus disease and immunity. **a,** Regional association plot of the genetic association at the chromosome 2q12 locus with serum anti-RV IgA in Ugandan children aged 2- 3 years prior to RV vaccine rollout. **b,** The effect of rs59241810 genotype on serum anti-RV IgA levels in Ugandan children. **c,** Distribution of serum anti-RV IgA levels in Zambian children aged 12 months. All children had received anti-RV vaccination (Rotarix) at 6 and 10 weeks of age. Individuals are coloured according to whether they were randomised to a third anti-RV vaccine dose at 9 months of age. **d,** The effect of secretor status on serum anti-RV IgA levels in vaccinated Zambian children at the age of 12 months. **e,** The effect of rs59241810 genotype on serum anti-RV IgA levels in vaccinated Zambian children at the age of 12 months. **f,** Effect estimates (and 95% confidence intervals) of rs59241810:T allele carriage on hospitalised RV disease risk in Kenyan children stratified by RV genotype. **g,** Effect estimates of rs59241810:T allele carriage on hospitalised RV disease risk in Kenyan children stratified by RV genotype and age at admission. **h,** The effect of rs59241810 genotype on total diarrhoeal disease episodes in unvaccinated Ugandan children in the first 10 years of life. **i,** Effect estimates of rs59241810:T allele carriage on diarrhoeal disease episodes which were prolonged (>14 days), on episodes of dysentery and on diarrhoeal disease episodes with dehydration in unvaccinated Ugandan children in the first 10 years of life. Box and whisker plots: boxes depict the upper and lower quartiles of the data, and whiskers depict the range of the data excluding outliers (data-points >1.5x the interquartile range from the upper or lower quartiles).

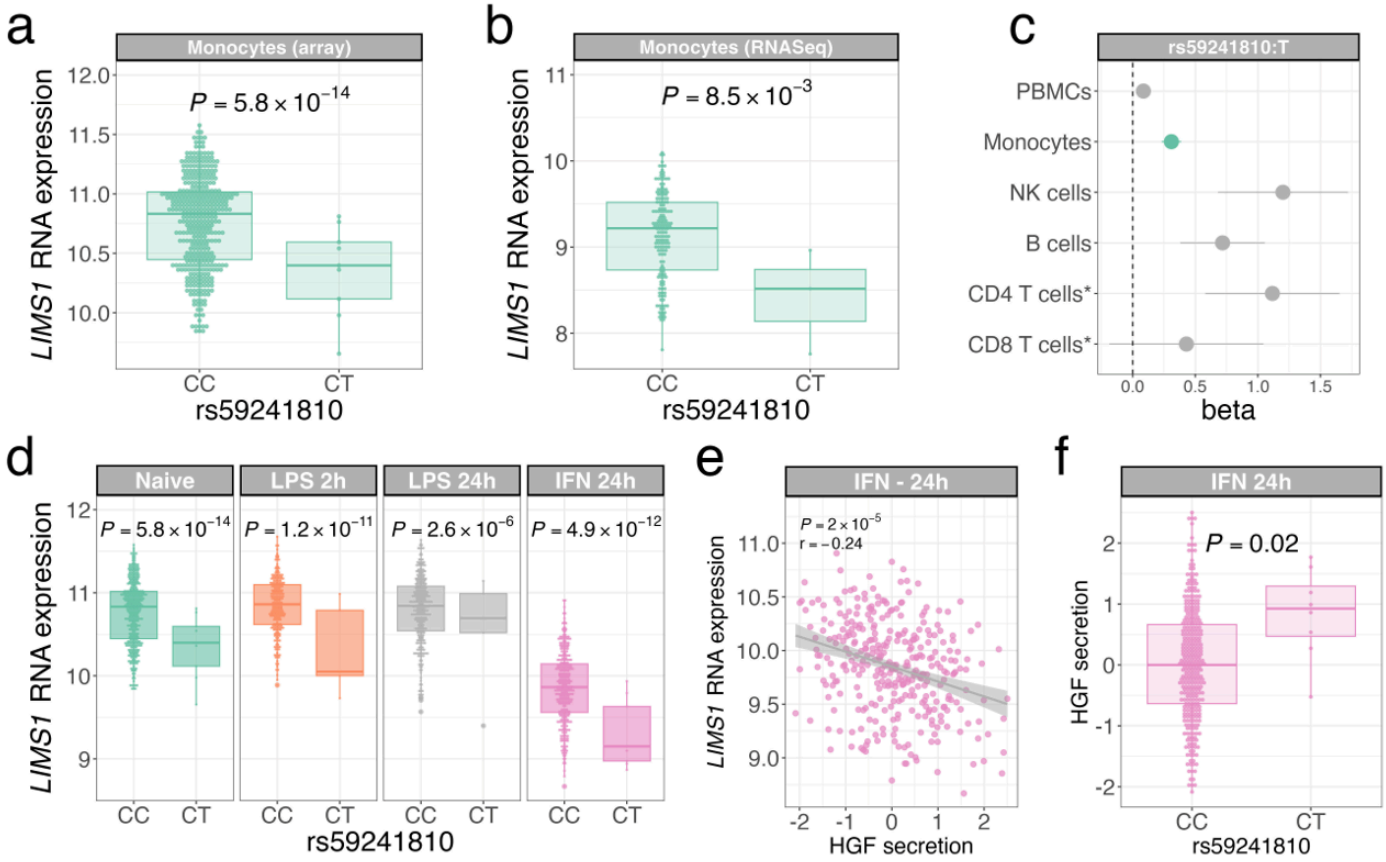


Figure 4

Regulatory function of the rotavirus-associated locus at chromosome 2q12. **a,** Effect of rs59241810 genotype on *LIMS1* RNA expression in naïve monocytes (n = 414 individuals). RNA expression quantified by microarray. **b,** Effect of rs59241810 genotype on *LIMS1* RNA expression in naïve monocytes (n = 192 individuals). RNA expression quantified by RNA-seq. **c,** The colocalization of the *LIMS1* eQTL at rs59241810 with the anti-RV IgA GWAS signal across immune cells. Significant eQTL effects colocalising (PP4>0.8) with the anti-RV IgA GWAS signal are highlighted (green), and is only observed in monocytes. **d,** Effect of rs59241810 genotype on *LIMS1* RNA expression in naïve monocytes (n = 414 individuals) and monocytes stimulated with LPS for 2 hours (n=261), LPS for 24 hours (n=322) and IFNg for 24 hours (n=367). Each eQTL signal shows evidence of colocalisation (PP4>0.8) with the anti-RV IgA GWAS signal, with the exception of that observed for monocytes stimulated with LPS for 24 hours. **e,** Correlation between LIMS1 RNA expression and hepatocyte growth factor (HGF) secretion in IFNg-stimulated monocytes (n=367). **f,** Effect of rs59241810 genotype on HGF secretion from IFNg-stimulated monocytes (n=367). Genotype phenotype associations are calculated by linear regression, and correlation coefficients represent Pearson's correlation. Box and whisker plots; boxes depict the upper and lower quartiles of the data, and whiskers depict the range of the data excluding outliers (outliers are defined as data-points > 1.5 × the inter-quartile range from the upper or lower quartiles).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- RVGWASsuppsub.docx