Impact of maternal antibodies and microbiota development on the immunogenicity of oral rotavirus vaccine in African, Indian, and European infants: a prospective cohort study

Edward P. K. Parker[1*], Christina Bronowski[2*], Kulandaipalayam Natarajan C. Sindhu[3*], Sudhir Babji[3], Blossom Benny[3], Noelia Carmona-Vicente[2], Nedson Chasweka[4], End Chinyama[4], Nigel A. Cunliffe[2], Queen Dube[4], Sidhartha Giri[3], Nicholas C. Grassly[6], Annai Gunasekaran[3], Deborah Howarth[2], Sushil Immanuel[3], Khuzwayo C. Jere[2,4,5], Beate Kampmann[1], Jenna Lowe[2], Jonathan Mandolo[4], Ira Praharaj[3], Bakthavatsalam Sandya Rani[3], Sophia Silas[3], Vivek Kumar Srinivasan[3], Mark Turner[7], Srinivasan Venugopal[3], Valsan Philip Verghese[8], Alistair C. Darby[9†], Gagandeep Kang[3†], Miren Iturriza-Gómara[2†§]

* Joint first authors; † Joint senior authors; § Lead contact

[1] *The Vaccine Centre, Department of Clinical Research, London School of Hygiene and Tropical Medicine, London, WC1E 7HT, UK;*

[2] *Institute of Infection and Global Health, University of Liverpool, Liverpool, L69 7BE, UK;*

[3] *Division of Gastrointestinal Sciences, Christian Medical College, Vellore, Tamil Nadu, 632004, India;*

[4] *Malawi-Liverpool-Wellcome Trust Clinical Research Programme, University of Malawi, Blantyre, PO Box 30096, Malawi;*

[5] *Department of Medical Laboratory Sciences, College of Medicine, University of Malawi, Private Bag 360, Chichiri, Blantyre 3, Malawi;*

[6] *Department of Infectious Disease Epidemiology, Imperial College London, London, W2 1PG, UK;*

[7] *Institute of Translational Medicine, University of Liverpool, Liverpool, L8 7SS, UK;*

[8] *Department of Child Health, Christian Medical College, Vellore, Tamil Nadu, 632004, India;*

[9] *Institute of Integrative Biology, University of Liverpool, Liverpool, L69 7ZB, UK*

Corresponding authors:

Edward Parker (edward.parker@lshtm.ac.uk)

Miren Iturriza-Gómara (M.Iturriza-Gomara@liverpool.ac.uk)

Word count: 3,999

Abstract

Identifying risk factors for impaired oral rotavirus vaccine (ORV) efficacy in low-income countries may lead to improvements in vaccine design and delivery. We measured maternal rotavirus antibodies, environmental enteric dysfunction (EED), and bacterial gut microbiota development among infants receiving two doses of Rotarix in India (n = 307), Malawi (n = 119), and the UK (n = 60), using standardised methods across cohorts. ORV shedding and seroconversion rates were significantly lower in Malawi and India than the UK. Maternal rotavirus-specific antibodies in serum and breastmilk were negatively correlated with ORV response in India and Malawi, and this was mediated partly by a reduction in ORV replication. In the UK, ORV replication was not inhibited despite comparable maternal antibody levels. In both India and Malawi, pre-vaccination microbiota diversity was negatively correlated with ORV immunogenicity, suggesting that high early-life microbial exposure may contribute to impaired vaccine efficacy.

Key words

Oral vaccines, environmental enteric dysfunction, intestinal microbiota, maternal antibodies, rotavirus, vaccination

Introduction

The roll-out of oral rotavirus vaccine (ORV) has now reached over 100 countries. This initiative – in parallel with advances in sanitation infrastructure and increased use of oral rehydration therapy – has led to substantial declines in global diarrhoeal mortality[1]. Whereas rotavirus was linked with over 500,000 infant deaths annually at the turn of the century[2], this number currently stands at approximately 130,000[3]. Yet the potential impact of ORV is constrained by the impaired performance of current vaccines in low- and middle-income countries (LMICs). The 1-year protective efficacy of Rotarix against severe rotavirus-associated gastroenteritis is >95% in Europe[4] but may fall below 50% in sub-Saharan Africa[5]. Moreover, while a variety of interventions to boost ORV performance have been tested (e.g. temporary withholding of breastfeeding), these have generally proven either ineffective or of modest benefit[6], highlighting the need for new strategies informed by a deeper understanding of the mechanisms underlying the vaccine efficacy gap.

A variety of risk factors have been linked with impaired oral vaccine performance in LMICs[7]. Passively acquired maternal antibodies appear to interfere with ORV immunogenicity [8], as documented for a variety of parenteral vaccines [9]. However, it remains unclear whether rotavirus-specific antibody concentrations and their inhibitory effect are higher in LMICs. Environmental enteric dysfunction (EED) – a subclinical condition associated with intestinal inflammation and increased permeability – is common in LMICs and has been cited as a possible cause of oral vaccine failure[10]. However, EED is generally measured using faecal or plasma biomarkers as a proxy for mucosal immune status, and studies of these biomarkers at the time of oral vaccine delivery have yielded mixed results[11]. The bacterial gut microbiota shapes and is in turn shaped by the developing infant immune system. Among infants in Ghana, Rotarix immunogenicity was positively correlated with the relative abundance of *Streptococcus bovis* and negatively correlated with Bacteroidetes abundance at the time of the first vaccine dose[12]. However, no significant discrepancies in microbiota composition were apparent when comparing Rotarix responders with non-responders in India[13]. Consistent predictors of ORV failure within the bacterial gut microbiota thus remain elusive.

Much of what we know about the mechanisms shaping oral vaccine performance comes from single-population studies focusing on individual risk factors. Here, we present a multicentre cohort study exploring the effect of maternal antibodies, EED biomarkers, and bacterial gut microbiota development on Rotarix response among infants in Malawi, India, and the UK.

Results

We enrolled pregnant women during the third trimester in Blantyre (Malawi; n = 119), Vellore (India; n = 306), and Liverpool (UK; n = 60; **Table 1**). After delivery, infants received routine vaccines including two doses of Rotarix according to the national immunisation schedule (6 and 10 weeks in Malawi and India, 8 and 12 weeks in the UK). We measured rotavirus-specific IgA (RV-IgA) in maternal blood, cord blood, and breastmilk samples collected during or in the week after delivery, and in infant blood samples collected pre- and post-vaccination. In Indian participants, all serum samples were also assayed for rotavirus-specific IgG (RV-IgG). Six longitudinal stool samples were collected from each infant and assayed for rotavirus shedding, bacterial microbiota composition (16S rRNA V3–V4 sequencing), and/or EED biomarkers (**Figure 1A** and **Table S1**).

ORV replication and immunogenicity. In the UK, we observed near-ubiquitous shedding of the first dose of ORV, with 55/60 (92%) shedding 1 week after vaccination and 24/54 (44%) continuing to shed after 1 month (**Figure S1A**). By contrast, dose 1 shedding was detected in 82/305 (27%) infants in India and 56/101 (55%) in Malawi (**Figure 1B**), and continued shedding after 1 month was much rarer in these cohorts (**Figure S1A**). ORV shedding after the second dose was also more common in the UK than India or Malawi, although discrepancies were less marked for this dose (**Figure 1B**). Shedding following at least one dose was observed in 54/56 (96%) infants in the UK, 151/304 (50%) in India, and 50/72 (69%) in Malawi (Fisher's p values <0.005 for all between-country comparisons after false discovery rate [FDR] adjustment; **Figure 1B**).

Seroconversion was observed in 85/305 (28%) infants in India, 24/103 (23%) infants in Malawi, and 27/51 (53%) infants in the UK (**Figure 1C**). Geometric mean concentrations (GMCs) of RV-IgA (IU/ml) after vaccination were 20 (95% CI 16–25) in India, 9 (6–12) in Malawi, and 27 (17–45) in the UK (Tukey's post-hoc p values <0.005 for comparisons between Malawi and other cohorts and 0.489 for India vs UK). Overall, at least one indicator of ORV response (seroconversion or shedding after at least one dose) was observed in 181/302 (60%) infants in India, 43/61 (70%) in Malawi, and 46/48 (96%) in the UK (**Figure 1D**).

Neonatal rotavirus infection. A unique feature among Indian infants was the high rate of wild-type rotavirus shedding in the first week of life (136/306 [44%] based on qPCR, compared with 6/102 [6%] in Malawi and 1/58 [2%] in the UK; **Figure 1B**). We successfully characterised rotavirus genotype in 104 samples, of which 103 were positive for the strain G10P[11]. Overall, 166/304 (55%) infants in India

had some indication of pre-vaccination rotavirus infection (week 1 shedding or baseline RV-IgA ≥20 IU/ml), whereas the corresponding rates were 10/90 (11%) in Malawi and 2/54 (4%) in the UK. Pre-vaccination infection was more common among infants born in tertiary care facilities in India (relative risk [RR] 1.98 [95% CI 1.72–2.19]; **Table S2)** and among infants delivered by caesarean section versus vaginal delivery (RR 1.31 [1.05–1.53]). All neonatal rotavirus infections were asymptomatic.

Consistent with the high prevalence of neonatal rotavirus infection, we observed significantly higher pre-vaccination RV-IgA concentrations in India compared with Malawi and the UK (**Figure 1C**). Prior infection was associated with a reduced likelihood of shedding ORV after at least one dose (RR 0.65 [0.47–0.83]; **Figure S1B** and **Table S3**). By contrast, pre-vaccination infection did not influence the likelihood of seroconversion (**Figure S1B** and **Table S4**), and where ORV replication was observed among previously infected infants, this significantly boosted post-vaccination RV-IgA concentrations (**Figure 1E**). Overall, our findings point to the high frequency and immunogenic nature of neonatal rotavirus infection in India. Indeed, despite the lower seroconversion rates in India compared with the UK, post-vaccination GMCs did not differ significantly between these countries (**Figure 1C**).

Breastfeeding, growth, and sanitation. We examined several baseline health and demographic variables for their potential correlation with ORV response (**Tables S3** to **S5**). In India, seroconversion was positively correlated with exclusive breastfeeding (RR 2.04 [1.07–3.68]). Recruitment in this cohort spanned the global switch from trivalent OPV (tOPV) to bivalent OPV (bOPV) in April 2016. Seroconversion was more common in infants who received bOPV-only than tOPV-containing schedules (**Table S4**), although infants receiving tOPV were also more likely to be immunised during the dry season (January–May) given the timing of the tOPV–bOPV switch. Similar associations were apparent when considering post-vaccination RV-IgA concentration as an endpoint (**Table S5**). Moreover, RV-IgA was positively correlated with height-for-age Z score at the time of the first ORV dose, and was higher among infants living in houses built from permanent versus temporary or mixed materials (beta 2.03 [95% CI 1.26–3.26], where beta represents the estimated ratio of GMCs) or with access to treated water (beta 1.66 [1.04–2.67]). Baseline health and demographic variables were not strongly associated with ORV response in Malawi or the UK, although fewer covariates were measured in these cohorts.

Maternal antibodies. Serum RV-IgA concentrations were significantly higher among mothers in Malawi than in India or the UK (**Figure 2A**; GMCs of 134 [120–150], 340 [256–452], and 186 [123–281] in

143 India, Malawi, and the UK, respectively). Interestingly, maternal RV-IgA levels did not differ
144 significantly between India and the UK, while breastmilk RV-IgA concentrations were significantly
145 higher in Malawi and the UK than in India (GMCs of 25 [22–29], 97 [84–112], and 100 [68–146] in India,
146 Malawi, and the UK, respectively). Thus, while the degree of maternal antibody exposure was greatest
147 in Malawi, infants in the UK were by no means exempt from this potential ORV inhibitor.

148 Our analysis of the association between maternal antibodies and ORV outcome yielded several
149 notable findings. First, maternal serum RV-IgA concentrations were negatively correlated with infant
150 post-vaccination RV-IgA levels in each cohort (Pearson coefficient [*r*] of -0.29, -0.13, and -0.12 in Malawi,
151 India, and the UK, respectively; **Figure 2B**). Although this correlation was not significant in the UK,
152 this may reflect a lack of statistical power owing to the smaller size of this cohort, and we observed no
153 significant evidence of effect modification between cohorts (p values of >0.05 for interaction terms
154 between maternal RV-IgA and country). Serum IgA is not transmitted efficiently across the placenta
155 and is therefore unlikely to directly influence ORV. However, we observed a strong correlation between
156 maternal RV-IgA and cord blood RV-IgG in India (*r* 0.41, p <0.001), suggesting that maternal RV-IgA
157 may be a valid (if imperfect) proxy for the degree of transplacental RV-IgG transfer (**Figure 2C**). Second,
158 maternal serum RV-IgA levels were negatively correlated with ORV shedding in both India and
159 Malawi (**Figure 2C** and **Table S3**). By contrast, ORV shedding in the UK was ubiquitous despite the
160 fact that maternal antibody levels were equivalent to those in India. Finally, we observed no significant
161 difference in maternal antibody levels according to week 1 rotavirus shedding status in Indian infants
162 (**Table S2**), suggesting that neonatal G10P[11] viruses may not be subject to the same replication
163 inhibition observed for ORV. However, among infants with pre-vaccination infection, we observed a
164 strong negative correlation between maternal RV-IgG levels and infant RV-IgA at 6 weeks (*r* -0.42, p
165 <0.001; **Figure 2B**).

166 Breastmilk RV-IgA was negatively correlated with ORV immunogenicity and/or shedding in
167 India and Malawi (**Figure 2B** and **Tables S3** to **S5**). Disentangling the relative influence of breastmilk
168 versus transplacental antibodies is challenging given the correlation between the two (**Figure 2C**).
169 However, as noted above, ORV immunogenicity was higher in conjunction with exclusive
170 breastfeeding in India – a finding at odds with a significant inhibitory effect of breastmilk antibodies
171 on ORV response.

172

173 **Inflammatory biomarkers.** We observed strong geographic discrepancies in EED biomarkers. Whereas
174 α1-antitrypsin (α1AT; a marker of protein-losing enteropathy) was highest in Malawi and lowest in the

175 UK, both myeloperoxidase (MPO; a marker of neutrophil activity) and α1 acid glycoprotein (a marker
176 of systemic inflammation) were highest in Indian infants (**Figure S2A**). Despite their marked variation
177 within and between cohorts, none of the biomarkers were significantly associated with seroconversion,
178 ORV shedding, or post-vaccination RV-IgA (**Figure S2B** and **Tables S3** to **S5**).
179
180 **Geographic differences in microbiota development.** We obtained high-quality faecal microbiota
181 profiles in 2,086 samples (142,880 ± 136,113 [mean ± s.d.] reads per sample). Microbiota profiles were
182 highly consistent across sequencing runs and facilities (explored by independently re-sequencing 10%
183 of samples; **Figure S3**). Infant samples were dominated by a small number of taxa (**Figure S4**). The
184 trajectory of microbiota development was highly distinct to each cohort, with significant deviations
185 apparent as early as the first week of life. In particular, microbiota diversity and richness were
186 significantly higher in Malawi than both other cohorts (**Figure 3A**), and numerous genera were
187 enriched in Malawi compared with both India (21 genera with FDR $p < 0.1$ in longitudinal mixed-effects
188 models) and the UK (22 genera with FDR $p < 0.1$; **Figure S5**). On the other hand, *Bifidobacterium*,
189 *Enterococcus*, *Staphylococcus*, and *Streptococcus* were enriched in Indian infants compared with both
190 other cohorts, while *Bacteroides*, *Citrobacter*, *Enterobacter*, *Klebsiella*, *Haemophilus*, *Lachnospiraceae*, and
191 *Veillonella* were enriched in the UK (**Figures 3B** and **S5**). Several of these taxonomic discrepancies were
192 also evident in maternal stool samples collected in the week after delivery (**Figure S4C**). Clustering of
193 samples by country at the time of the first ORV dose was evident based on Bray–Curtis distances
194 (**Figure 3C**). Moreover, using the machine-learning method Random Forests, samples could be
195 accurately distinguished by country at all time-points (mean cross-validation accuracies of 83–98%;
196 baseline accuracy 50%; **Figure 3D**). The genera most important to the predictive accuracy of these
197 models differed between cohorts in both their prevalence and abundance (**Figures 3E** and **S6**). Indeed,
198 simple cross-sectional comparisons of taxon prevalence were adept at selecting differential microbiota
199 colonisation patterns between cohorts (**Figure S5C** and **Table S6**).
200
201 **Microbiota composition versus ORV response.** In both India and Malawi, we observed a significant
202 negative correlation between microbiota diversity and ORV immunogenicity. In India, these
203 discrepancies were apparent in longitudinal models (**Figure 4A**) and cross-sectional comparisons at the
204 time of the first ORV dose (**Figure 4B** and **Table S4**). Interestingly, stratified analyses revealed the
205 discrepancies to be specific to infants with no rotavirus exposure prior to receiving ORV (**Figure 4A**).
206 A similar trend was evident in Malawian infants, albeit specifically at the time of the first dose of ORV

(**Figure 4B** and **Table S4**). By contrast, microbiota diversity was not strongly associated with ORV immunogenicity in the UK. Neither Shannon index nor richness differed according to ORV shedding status in any cohort (**Figure S7** and **Table S3**).

In spite of the significant differences in microbiota diversity, we did not observe clear differences in taxonomic composition according to ORV response. Cross-sectional Random Forests models based on genus abundances failed to accurately predict immunogenicity (**Figures 4C** and **4D**) or shedding status (**Figures S7C** and **S7D**). While numerous genera either increased or decreased in abundance with age, very few exhibited differential abundance according to ORV outcome when assessed via longitudinal models (**Figures 4E** and **S7E**). Exceptions include *Sutterella*, which was negatively correlated with dose 1 ORV shedding in Malawian infants, while *Lactobacillus*, *Corynebacterium*, and *Dolosigranulum* were negatively correlated with ORV shedding and/or final RV-IgA concentration in India (**Table S7**). Overall, while microbial exposure and colonisation appears to be significantly higher in infants who fail to respond to ORV, we did not observe consistent discrepancies in specific bacterial taxa according to vaccine outcome.

Multivariate analysis. To delineate the contribution of different risk factors in shaping ORV response, we performed an integrated analysis based on three modules of input data: (i) demographic and baseline health variables (n = 18; see **Table S4** for complete list); (ii) exposure/antibody data (n = 12), including EED biomarkers and maternal antibody concentrations; and (iii) pre-vaccination microbiota composition, comprising genus abundances at the time of the first ORV dose (n = 126). We used the machine-learning algorithm Random Forests to predict ORV outcome in Indian infants (the largest cohort in terms of number of infants and number of recorded variables) using each module in turn and all modules combined. The prediction of seroconversion and ORV shedding was poor for all modules (**Figure 5A**). Post-vaccination RV-IgA concentration was predicted with modest accuracy by exposure/antibody data (cross-validation R^2 of 0.36 ± 0.11 [mean ± s.d.]), and this module alone offered predictive accuracy commensurate with the integrated analysis combining all variables (R^2 of 0.38 ± 0.10; **Figure 5A**). Consistent with the univariate analyses described above, the most important variables underlying the prediction of post-vaccination RV-IgA levels were pre-vaccination rotavirus infection and maternal RV-IgG concentration (**Figure 5B**).

We complemented the machine-learning models with multivariate regressions. Covariates with a p value of <0.05 in univariate analyses were selected for inclusion in these models (**Tables S3** to **S5**). Notably, this approach revealed breastfeeding practice, linear growth, pre-vaccination exposure,

239 and polio vaccine schedule to be significantly correlated with post-vaccination RV-IgA (**Figure 5C**).
240 Associations varied according to pre-vaccination exposure status – whereas breastfeeding practice and
241 maternal antibody levels were strong predictors of RV-IgA among infants with neonatal rotavirus
242 exposure, microbiota diversity, house type, and access to treated water were significantly associated
243 with vaccine immunogenicity in those without prior exposure. Overall, the multivariate regressions
244 offered a predictive accuracy commensurate with or exceeding that of Random Forests (R^2 of 0.43 in the
245 complete cohort and 0.11–0.19 in the stratified analyses for RV-IgA regressions).
246

247 **Discussion**

248 ORV is unique among oral vaccines in being routinely administered in both high-income
249 countries and LMICs. As such, the vaccine provides a valuable tool for probing the divergent
250 trajectories of the developing immune system in different settings. Here, we report on the extent to
251 which several key determinants of infant immune development differ among three disparate
252 populations, and the degree to which these factors shape ORV response.

253 As expected, Rotarix response was significantly impaired among infants in Malawi and India.
254 While the later vaccination schedule in the UK (8/12 weeks vs 6/10 weeks) may contribute to these
255 differences by providing more time for passively acquired maternal antibodies to wane, overall
256 shedding rates after both doses in Malawi and India still fell short of the near-ubiquitous dose 1
257 shedding observed in the UK, highlighting the barrier to ORV replication that emerges within the first
258 months of life in LMICs. Seroconversion rates in India and Malawi were broadly consistent with those
259 previously reported in LMICs[14,15], but lower than expected in the UK[16]. These findings highlight the
260 suboptimal nature of RV-IgA as a correlate of protection. While higher RV-IgA concentrations correlate
261 with a lower risk of subsequent rotavirus infection, these associations often fail to hold when
262 considering protection against rotavirus-associated gastroenteritis[17]. We advocate the further use of
263 vaccine virus replication as an adjunctive measure of ORV response in future trials.

264 For a risk factor to account for broad geographic trends in ORV efficacy, it is reasonable to
265 expect it to be more common or observed at higher levels in LMICs than high-income countries. This
266 was not the case for several of the factors examined in this study. Maternal serum RV-IgA levels were
267 comparable in the UK and India (although significantly higher in Malawi), while breastmilk RV-IgA
268 levels were higher in the UK than in India. These results are somewhat surprising given that recent
269 rotavirus exposure and consequent boosting of antibody levels might be expected to occur more
270 frequently among mothers in LMICs with a high rotavirus disease burden. Moreover, in previous

studies, RV-IgA concentrations were significantly higher in India than the USA, with intermediate levels in South Africa[18,19]. Our findings caution against the broader extrapolation of these trends. The high maternal RV-IgA levels in the UK may reflect the immunogenic nature of early rotavirus exposures and the subsequent boosting of antibodies by mild or asymptomatic re-infection [20].

Nonetheless, our findings confirm the potential inhibitory effect of maternal antibodies on ORV among infants in LMICs. Whereas previous studies of this phenomenon have typically focused on immunogenicity endpoints[8,21], we observed a reduction in both shedding and immunogenicity, suggesting that the effect of maternal antibodies is mediated in part through a reduction in ORV replication efficiency. This could potentially involve the transudation of transplacental RV-IgG antibodies across the mucosal epithelium or the direct neutralisation of vaccine viruses by RV-IgA in breastmilk. In neonatal macaques, passively acquired monoclonal IgG antibodies have been shown to confer protection against lentivirus infection at mucosal surfaces[22], and a similar mechanism may influence ORV. The presence of a leaky mucosal barrier following early-life enteropathogen exposure could feasibly enhance the transfer of maternal RV-IgG to the gut lumen among infants in LMICs[23] and thereby enhance their inhibitory effect, although this possibility remains untested. If breastmilk RV-IgA is involved in the inhibition of ORV replication, the effect does not appear to be mitigated by the temporary withholding of breastfeeding, which failed to boost ORV response in controlled trials[24,25]. Our findings are also consistent with the inhibition of downstream antigen processing by transplacental antibodies, as evidenced by the strong inverse correlation between maternal RV-IgG and infant RV-IgA formation in Indian infants with neonatal rotavirus infection. Overall, one could reasonably envision a combination of upstream inhibition of ORV replication (via RV-IgA and/or transudated RV-IgG) and downstream inhibition of antigen processing (via circulating RV-IgG) that synergistically inhibit ORV outcome. However, these mechanisms alone are clearly insufficient to prevent a robust ORV response given that the vaccine was ubiquitously shed and was highly immunogenic in the UK, where maternal antibody levels were high and >75% of infants were breastfed.

We considered several biomarkers as a proxy for EED. Although α1AT concentrations were significantly lower in the UK than both other cohorts, MPO levels were similar in the UK and Malawi, calling into question the extent to which this marker accurately captures the onset of EED. We did not observe a significant relationship between any of the measured EED biomarkers and ORV outcome, in keeping with the growing literature in this field[11,26]. Together, these findings imply either that EED is not a primary driver of oral vaccine failure or that the condition is not accurately captured by the current suite of faecal and systemic biomarkers (quite possibly both).

We observed marked geographic discrepancies in composition of the bacterial microbiota. Microbiota diversity was significantly higher among Malawian infants, while Indian infants could be distinguished by their high *Bifidobacterium* abundance. The latter is interesting given that *Bifidobacterium* has previously been suggested to enhance immunological memory responses to a range of vaccine targets in Bangladesh[27], yet this genus was least abundant among infants in the UK that were most responsive to ORV. Similarly, we failed to recapitulate other associations that have between documented between bacterial microbiota composition and ORV outcome, such as the higher abundance of Bacteroidetes among non-responders in Ghana[12]. However, we observed a significant negative correlation between microbiota diversity and ORV immunogenicity – apparent among Indian infants throughout infancy and in Malawian infants specifically prior to the first vaccine dose. Although this is the first study to find such an association for ORV, we previously reported a negative correlation between microbiota diversity and monovalent type 3 OPV response among 6–11 month-old Indian infants[28]. Several indirect explanations for this trend are plausible. For example, microbiota richness may act as a proxy for other early-life exposures that shape ORV outcome, such as non-polio enteroviruses[29]. The fact that ORV immunogenicity was impaired among Indian infants in households without access to treated water would support this notion. It is also possible that exposure to greater microbial diversity may foster a state of hyporesponsiveness at the mucosal epithelium that impairs oral vaccine outcome. When viewed in this way, impaired vaccine response could be seen as a counterpart to hyperresponsive immune states associated with low early-life microbiota diversity, such as atopic disease[30]. Notably, low microbiota diversity is often used as a marker of 'dysbiosis'[31], albeit largely based on studies in high-income settings. A clear understanding of the signatures of 'healthy' microbiota development in different settings is lacking. Our findings allude to a more complex narrative whereby such signatures vary not only by geographic setting and age, but by the health outcome being considered.

Our study has several limitations. As noted above, RV-IgA is a suboptimal correlate of protection against rotavirus-associated gastroenteritis. We compensated for this by considering multiple endpoints, including seroconversion and post-vaccination shedding. Whilst the potential contribution of batch effects to the observed differences between populations cannot be completely ruled out, we went to considerable lengths to ensure reproducibility and comparability across sites, using the same reagent lots and standards. In the case of microbiota composition, we independently re-sequenced 10% of samples at a separate facility and recapitulated the original microbiota composition with high accuracy. Finally, profiling the virome, proteome, transcriptome, and

335 metabolome can now be achieved using small sample volumes and would provide valuable insight
336 into systemic immune status at the time of vaccination in future studies[32].
337 Notwithstanding these caveats, this study advances our understanding of the potential
338 mechanisms influencing ORV response in several ways. By considering how several risk factors of ORV
339 failure differ within and between populations, we have been able to place these mechanisms within a
340 broader global context. Our study confirms the inhibitory effect of maternal antibodies on ORV
341 immunogenicity in LMICs, and suggests this is mediated partly by a reduction in ORV replication. We
342 observed that asymptomatic neonatal rotavirus exposure was strongly associated with RV-IgA
343 formation, providing further support for the potential of neonatal vaccination as a pragmatic approach
344 to achieve greater rotavirus vaccine impact. Finally, while specific bacterial taxa do not appear to drive
345 within-population differences in ORV response, high microbial exposure in early life may contribute
346 to the impaired efficacy of this vaccine in LMICs.
347
348 **Acknowledgements**
349 We thank all members of the clinical study teams in Vellore, Blantyre, and Liverpool, including
350 Falak Diab, Siobhan Holt, and the research midwives at the Liverpool Women's Hospital; Dawn
351 Redman and the team of research nurses at Alder Hey Children's Hospital; Uma Raman, Charlet,
352 Margaret, Jacklin, and the field research assistants at Christian Medical College, Vellore; and James
353 Tamani, Anna Ainani, Amisa Chisale, Bertha Masamba, Carlo Gondwe, and Evelyn Gondwe in
354 Blantyre, Malawi. The UK and Malawi sites were funded by the UK Medical Research Council and the
355 UK Department for International Development (Newton Fund MR/N006259/1). M.I.G. is partly
356 supported by the NIHR HPRU in Gastrointestinal Infections. K.C.J. is funded by a Wellcome
357 International Training Fellowship (number 201945/Z/16/Z). The site in India was funded by the
358 Government of India's Department of Biotechnology. Richard Eccles, Anita Lucaci, Richard Gregory,
359 John Kenny, and other staff at the Centre for Genomic Research (University of Liverpool) provided
360 valuable support for the 16S microbiota sequencing work, as did Laurence Game at the Imperial BRC
361 Genomics Laboratory. Above all, we are grateful to the families involved in the study.
362
363 **Author contributions**
364 Conceptualisation, M.I.G., N.A.C., A.C.D., N.C.G., and G.K.; Methodology, C.B., E.P.K.P.,
365 A.C.D., M.I.G., I.P., and S.B.; Software, E.P.K.P.; Validation, E.P.K.P.; Formal Analysis, E.P.K.P.;
366 Investigation, C.B., E.P.K.P., N.C.-V., D.H., J.L., S.B., I.P., S.G., S.I., V.K.S., B.B., S.S., B.S.R., J.M., and

E.C.; Resources Provision, K.N.S., A.G., V.P.V., Q.D., N.C., M.T., and N.A.C.; Data Curation, C.B., E.P.K.P., K.N.S., S.V., J.M., and M.I.G.; Writing – Original Draft, E.P.K.P., C.B., A.C.D., S.B., I.P., A.D., M.I.G., and N.C.G.; Writing – Review & Editing, B.K., Visualisation, E.P.K.P; Supervision, N.A.C., A.C.D., K.C.J., N.C.G., B.K.; Project Administration, C.B., M.I-G., I.P., S.B., K.N.S., K.C.J.; Funding Acquisition, M.I.G., G.K.

Declaration of Interests

M.I.G. has received research grants from GSK and Merck, and has provided expert advice to GSK. K.C.J. and N.A.C. have received investigator-initiated research grant support from GSK. N.A.C. has received research grant support and honoraria for participation in rotavirus vaccine data safety monitoring committee meetings from GSK.

Figure titles and legends

Figure 1. Study Design and Oral Rotavirus Vaccine Response. (A) Study design. (B to D) Geographic differences in (B) rotavirus shedding, (C) immunogenicity, and (D) cumulative vaccine response (seroconversion or shedding after either dose). Error bars represent Clopper–Pearson 95% confidence intervals. (E) Association between rotavirus shedding and post-vaccination antibody concentration. Gene targets for shedding assays are indicated in brackets. Groups were compared by Fisher's exact tests with FDR correction (binary outcomes) or ANOVAs with post-hoc Tukey tests (continuous outcomes). The dotted lines at 20 IU/ml indicate the standard cut-off for RV-IgA seropositivity. ns, not significant; ORV, oral rotavirus vaccine; RV, rotavirus; * $p < 0.05$; ** $p < 0.005$; *** $p < 0.0005$.

Figure 2. Association between Maternal Antibodies and Oral Rotavirus Vaccine Response. (A) Maternal antibody concentrations by cohort. Groups were compared by ANOVAs with post-hoc Tukey tests. The dotted lines at 20 IU/ml indicate the standard cut-off for RV-IgA seropositivity. (B) Association between maternal antibodies and RV-IgA formation. Log-transformed concentrations were compared using Pearson's correlation coefficient (*r*). (C) Correlation between antibody concentrations and rotavirus shedding quantity (1/Ct) among Indian infants. Variables were compared using Spearman's rank correlation coefficient (rho). neo+, infected with rotavirus pre-vaccination; neo-, uninfected with rotavirus pre-vaccination; ORV, oral rotavirus vaccine; RV, rotavirus; * $p < 0.05$; ** $p < 0.005$; *** $p < 0.0005$.

Figure 3. Geographic Differences in Microbiota Development. (A) Longitudinal analysis of alpha diversity. Points represent mean values. Groups were compared using longitudinal mixed-effects models including week as a covariate and study ID as a random effect. Pairwise comparisons between countries were FDR corrected. (B) Longitudinal plot of mean genus abundances. Genera were included if they were present with a mean relative abundance of ≥5% at one or more timepoints. (C) Principal coordinates analysis plot of weighted Bray-Curtis distances for samples collected before the first dose of ORV (week 6 in India and Malawi; week 8 in the UK). (D) Cross-validation accuracy of Random Forests models (mean ± s.d.). (E) The 20 most important genera selected by Random Forests for discriminating infants by country at the time of the first dose of ORV. Mean cross-validation importance scores based on Gini index are depicted alongside the prevalence (≥0.1% abundance threshold) and mean abundance of the corresponding genera in each country. PC, principal coordinate; ns, not significant; * $p < 0.05$; *** $p < 0.0005$.

410

411 **Figure 4. Association between Microbiota Development and Oral Rotavirus Vaccine Response.** (A)
412 Longitudinal analysis of microbiota diversity. Points represent mean values. Groups were compared
413 using longitudinal mixed-effects models including week as a covariate and study ID as a random effect.
414 (B) Association between microbiota diversity and RV-IgA formation. Variables were compared using
415 Pearson's correlation coefficient (*r*). Microbiota richness values (number of genera detectable at ≥0.1%
416 abundance) were log-transformed. (C and D) Random Forests cross-validation accuracy for (C)
417 seroconversion and (D) RV-IgA concentrations based on genus abundance data (mean ± s.d.). (E)
418 Longitudinal zero-inflated negative binomial models of genus abundance. Genera with FDR-adjusted
419 p values of <0.1 are highlighted. neo+, infected with rotavirus pre-vaccination; neo-, uninfected with
420 rotavirus pre-vaccination; nf, not fitted as <10 responders; ORV, oral rotavirus vaccine; RV, rotavirus;
421 ZINB, zero-inflated negative binomial; * p <0.05; ** p <0.005; *** p <0.0005.

422

423 **Figure 5. Integrated Analysis for Prediction of Oral Rotavirus Vaccine Response.** (A) Random Forests
424 models cross-validation accuracy for prediction of vaccine outcome based on demographic,
425 exposure/antibody data, and pre-vaccination (week 6) microbiota data (mean ± s.d.). (B) The top 10
426 variables selected by Random Forests for prediction of RV-IgA concentration based on mean cross-
427 validation importance score. (C) Multivariate regression models for post-vaccination RV-IgA. Variables
428 were included if significantly associated with RV-IgA concentration (p <0.05) during univariate
429 analyses. See **Table S5** for full univariate and multivariate outputs. neo+, infected with rotavirus pre-
430 vaccination; neo-, uninfected with rotavirus pre-vaccination; MSE, mean squared error; ORV, oral
431 rotavirus vaccine; RV, rotavirus.

Table 1. Baseline characteristics of study cohorts.

	India	Malawi	UK
N	307	119	60
Date of birth			
First	03/12/2015	02/11/2016	29/09/2016
Last	15/11/2016	25/04/2018	30/03/2018
Caesarean delivery	70 (22.8)	0 (0.0)	1 (1.7)*
Birthweight (kg)	2.96 (0.42)	3.02 (0.42)	3.68 (0.48)
Female	152 (49.5)	60 (50.4)	24 (40.0)
Polio vaccine schedule			
tOPV	57 (18.6)	–	–
Mixed tOPV/bOPV	56 (18.2)	–	–
bOPV	94 (30.6)	119 (100.0)	–
IPV	100 (32.6)	–	60 (100.0)
Breastfeeding†			
Exclusive	265 (86.3)	108 (90.7)	26 (43.3)
Partial	32 (10.4)	11 (9.2)	20 (33.3)
None	10 (3.3)	0 (0.0)	14 (23.3)
Exposed to antibiotics	84 (27.4)	33 (27.7)	6 (10.0)
HIV status			
Exposed	0 (0.0)	27 (22.7)	-
Unexposed	299 (97.4)	88 (73.9)	-
Unknown	8 (2.6)	4 (3.4)	60 (100.0)

Data are n (%) or mean (s. d.).

* Elective caesarean was an exclusion criterion for this cohort (with 1 delivery by emergency caesarean).

† Data from birth to week 11 in Malawi and India, and birth to week 13 in the UK.

456 **Supplemental figure titles and legends**

457 **Figure S1. Association between pre-vaccination rotavirus exposure and oral rotavirus vaccine**
458 **response.** (A) Rotavirus shedding across all timepoints based on VP6-specific PCR. Groups were
459 compared by Fisher's exact tests with FDR correction. (B) Association between pre-vaccination
460 rotavirus exposure and ORV outcome in Indian infants. Prior exposure was defined by week 1 rotavirus
461 shedding or detection of RV-IgA at baseline (≥20 IU/ml). Exposed and unexposed groups were
462 compared by ANOVA (log antibody concentrations) or Fisher's exact tests (shedding/seroconversion).
463 The cohorts from Malawi and the UK are displayed for reference; statistical comparisons among cohorts
464 are shown in **Figure 1**. Error bars represent Clopper–Pearson 95% confidence intervals. ns, not
465 significant; ORV, oral rotavirus vaccine; RV, rotavirus; *, $p < 0.05$; **, $p < 0.005$; ***, $p < 0.0005$.

466

467 **Figure S2. Association between environmental enteric dysfunction and oral rotavirus vaccine**
468 **response.** (A) Environmental enteric dysfunction biomarkers by cohort. Groups were compared by
469 ANOVAs with post-hoc Tukey tests. (B) Association between inflammatory biomarkers and RV-IgA
470 formation. Log-transformed concentrations were compared using Pearson's correlation coefficient (*r*).
471 α1AT, α1-antitrypsin; α1AG, α1 acid glycoprotein; MPO, myeloperoxidase; neo+, infected with
472 rotavirus pre-vaccination; neo-, uninfected with rotavirus pre-vaccination; ORV, oral rotavirus vaccine;
473 RV, rotavirus.

474

475 **Figure S3. Replicability of microbiota sequencing across runs and facilities.** Positive controls
476 including an infant stool (BSctrl), maternal stool (MSctrl), and a mock bacterial community (MCctrl)
477 were included in every plate, with up to up to four plates per run. Alpha diversity (A) and beta diversity
478 (B) for these positive controls were highly replicable from run to run, with sample type accounting for
479 the majority of variation among controls based on linear regression (for alpha diversity) and
480 PERMANOVA (for beta diversity). A minimum detection threshold of 0.1% abundance was used for
481 richness and unweighted Bray–Curtis metrics. (C) Proportion of variation associated with run for each
482 sample group. R^2 and statistical significance were determined separately for each sample group by
483 PERMANOVA. Mean R^2 values are indicated by dotted lines. (D–F) For 10% of samples (90 week-1
484 infant samples, 90 week-10/12 infant samples, and 90 maternal samples evenly distributed across the
485 study sites), we validated the microbiota pipeline at a separate sequencing facility in London. PCR,
486 quantification, and pooling all involved the same methods and reagents (ordered separately), although
487 sequencing was performed via Illumina MiSeq rather than HiSeq (used for the majority of runs in

488 Liverpool). The order of samples during the 10% validation was randomised across three plates.
489 Microbiota composition was highly replicable based on alpha diversity (D), beta diversity (E), and
490 relative abundance of major genera (F), with sample ID accounting for >97% of variation based on
491 linear regression or PERMANOVA. In (B), replicates are linked by a line, although often the overlap is
492 so precise that these lines are indistinguishable. The proportion of variation associated with sample
493 code was equivalent when using unweighted distances ($R^2 = 0.996$). PC, principal coordinate; * $p<0.05$.
494
495 **Figure S4. Genus profiles of infant and maternal samples.** (A) Hierarchical clustering of stool samples
496 (columns) based on presence/absence of common genera (rows). Genera detected in 1% of samples
497 were included. Rows and columns were clustered using Ward's minimum variance hierarchical
498 clustering method. (B) Genus abundance versus prevalence for infant samples. Samples in each cohort
499 were generally dominated by a small number of taxa with high prevalence and abundance. All infant
500 samples from each cohort were included in the prevalence and abundance calculations (week 1, week
501 4, week 6/8, and week 10/12). Margins display density plots. (C) Geographic discrepancies in maternal
502 microbiota composition, including comparisons of alpha diversity (left panel) and genus composition
503 (right panel). Diversity metrics were compared using ANOVAs with post-hoc Tukey tests. Only genera
504 from **Figure 4B** are shown. ns, not significant; **, $p <0.05$; **, $p <0.005$; ***, $p <0.0005$.
505
506 **Figure S5. Discriminant genera by country.** (A) Longitudinal relative abundance plots for major
507 genera (mean abundance ≥5% in at least one country for one or more timepoints). Lines show local
508 weighted regression (loess) fits with 95% confidence intervals. (B) Mixed-effects zero-inflated negative
509 binomial models of genus abundance. Genera were included if present in at least 20% of samples from
510 at least one country being compared. Study ID was included as a random effect. Genera with an FDR-
511 adjusted p values of <0.1 are highlighted. (C) Cross-sectional comparisons of genus prevalence. A
512 minimum detection threshold of 0.1% abundance was used to define the presence/absence of a taxon.
513 Genera with an FDR-adjusted p values of <0.1 are highlighted.
514
515 **Figure S6. Top 20 genera for prediction of country using Random Forests.** See **Figures 4E** and **4F** for
516 additional details.
517
518 **Figure S7. Association between microbiota development and oral rotavirus vaccine response.** (A and
519 B) Longitudinal analysis of microbiota diversity in relation to (A) dose 1 shedding status and (B)

shedding after either dose. Points represent mean values. Groups were compared using longitudinal mixed-effects models including week as a covariate and study ID as a random effect. (C and D) Random Forests cross-validation accuracy for (C) dose 1 shedding and (D) shedding after either dose based on genus abundance data (mean ± s.d.). (E) Longitudinal zero-inflated negative binomial models of genus abundance. Genera with FDR-adjusted p values of <0.1 are highlighted, with further details provided in **Table S7**. (D) Number of samples per cross-validation fold during cross-sectional Random Forests models. All available samples were included in models for post-vaccination RV-IgA concentration. For binary outcomes, we randomly selected 50 infants per group in each iteration of cross-validation, or the number of samples in the minority group if this was <50. neo+, infected with rotavirus pre-vaccination; neo-, uninfected with rotavirus pre-vaccination; nf, not fitted as <10 samples in minority group; ORV, oral rotavirus vaccine; RV, rotavirus; ZINB, zero-inflated negative binomial.

Figure S8. Removal of extraction batches with potential cross-contamination. (A) Read counts in no-template and extraction controls. All no-template controls were clear of substantial amplification (left panel), as were pooled extraction controls from Malawi and the UK (middle panel). Four pools of extraction controls from India yielded significant amplification (>10,000 sequences). When the individual extraction controls from these pools were sequenced (right panel), two gave significant amplification (corresponding to extraction batches 14 and 27). (B) Beta diversity plots and (C) genus composition of these extraction controls revealed two clusters consistent with contamination of extraction batches 14 and 27. We therefore removed the samples associated with these extraction batches from the final analysis (n = 8). NTC, no-template PCR control; PC, principal coordinate.

541 **Methods**

542 **Study design**

543 Pregnant women were enrolled during the third trimester at three study sites: Blantyre
544 (Malawi), Vellore (India), and Liverpool (UK). Women were included if they provided informed
545 written consent to participate in the study and were willing to stay in the study area for 4 months
546 following delivery. Exclusion criteria included: congenital immune deficiency; chronic renal or liver
547 failure; other chronic illnesses which may affect immune function; non-singleton pregnancy; low
548 birthweight or pre-term birth (<34 weeks gestation); congenital anomalies and other neonatal
549 complications requiring prolonged hospitalisation; and delivery by elective caesarean (UK only).
550 Mothers provided a venous blood sample as well as a cord blood sample during birth and a stool
551 sample during the week after delivery. Infants provided two blood samples and four stool samples
552 over the course of the study (**Figure 1A**). The study was approved by the Institutional Review Board at
553 the Christian Medical College (CMC) in Vellore, the College of Medicine Research and Ethics
554 Committee in Blantyre, and the North West – Liverpool East Research Ethics Committee in Liverpool.
555 The trial is registered with the Clinical Trials Registry of India (CTRI/2015/11/006354) and the study
556 protocol has previously been published[33].

557 Routine vaccines were administered according to the routine schedule at each site, including
558 OPV at 0, 6, and 10 weeks of age in Malawi and India. In Malawi, all infants were born after April 2016
559 – the date of the global switch from tOPV to bOPV. In India, 98/307 (32%) infants were born before
560 April 2016. Infants therefore received tOPV (57/307 [19%]), a mixed schedule of tOPV and bOPV
561 (56/307 [18%]), or bOPV (94/307 [31%]). To explore the potential inhibitory effect of OPV on ORV[34],
562 the 6 and 10 week doses of OPV were replaced with inactivated poliovirus vaccine (IPV) in a
563 sequentially recruited cohort in India (100/307 [33%]). The comparison of OPV and IPV arms is
564 described elsewhere (Babji et al, in preparation). Briefly, while shedding after the first dose of ORV was
565 less common in OPV than IPV recipients, cumulative shedding after both doses was comparable in the
566 two arms, as were seroconversion rates and post-vaccination RV-IgA levels. Given the lack of
567 association between study arm and ORV immunogenicity, we pooled the IPV and OPV arms in the
568 present study.

569

570 **Sample processing and storage**

571 Whole blood was collected in anti-coagulation EDTA-tubes (BD) and stored at 4°C for up to 12
572 hours until collection by laboratory staff. Fractions were separated by centrifugation at 1,500–2,000 g in

573 a benchtop centrifuge (Eppendorf). A disposable plastic transfer pipette was used to aspirate the
574 plasma down to ~1 mm from the red blood cells. The plasma was aliquoted into two screw cap cryo-
575 tubes (Starlab). The buffy coat and red blood cells were subsequently collected into separate cryovials
576 and all samples were stored at -70°C.

577 Stool and breastmilk samples were collected in sterile sample pots by participants and were
578 shipped to the respective laboratory by courier within 24 hours in the UK and within 4 hours in India
579 and Malawi (to account for ambient temperatures). Samples were kept at 4°C for a maximum of 8 hours
580 until processing. Breastmilk samples were stored in 2 ml aliquots in SuperLock tubes (Starlab) at -70°C
581 for maternal RV-IgA analysis. Two 10% stool suspensions were prepared in sterile phosphate-buffered
582 saline (PBS) for assessing rotavirus shedding. Further aliquots of neat stool were stored in 2 ml
583 SuperLock tubes (Starlab) at -70°C for microbiota analysis and inflammatory biomarker measurement.
584 Stool samples were stored at -70°C for a maximum of 2 weeks prior to DNA extraction for microbiota
585 analysis.

586

587 **Rotavirus-specific antibodies**

588 RV-IgA and RV-IgG laboratory assays were conducted at CMC, Vellore, for samples from all
589 three study sites using a custom antibody-sandwich ELISA [35]. Briefly, 96-well plates (Costar) coated
590 with rabbit hyperimmune serum to rotavirus were incubated with purified cell culture lysates (WC3)
591 or mock-infected MA104 cells. Serial dilutions of standard and test sera were added followed by
592 biotinylated rabbit anti-human IgA (Jackson ImmunoResearch Laboratories) for detection of RV-IgA
593 and biotinylated rabbit anti-human IgG (Vector Laboratories). Absorbance was subsequently read at
594 492 nm. Background-corrected optical density values from sample wells were compared with the
595 standard curve and IgA or IgG concentration was determined based on derived units of IgA or IgG
596 arbitrarily assigned to the respective standard curve, with a minimum detection limit of 1 IU/ml.
597 Seroconversion was defined as detection of RV-IgA at ≥20 IU/ml in previously seronegative infants or
598 a 4-fold increase in RV-IgA concentration among infants who were seropositive at baseline.

599

600 **Rotavirus shedding**

601 Quantification of rotavirus shedding was performed at each study site. RNA was extracted
602 using the RNeasy mini kit (Qiagen) following the manufacturer's instructions. Briefly, 10% PBS stool
603 suspensions were defrosted on ice, vortexed, and briefly centrifuged (10,000 g for 10 seconds) to pellet
604 larger debris. 250 μl of β-ME in buffer RTL was added to 250 μl of supernatant. The mixture was then

605 vortexed before the addition of 250 µl of 70% molecular-grade ethanol. The RNeasy Mini Handbook
606 (5th Edition, October 2013) was followed for subsequent steps, and RNA was eluted in 50 µl RNase-free
607 water. cDNA was generated immediately after RNA extraction. The master mix for cDNA conversion
608 comprised: 7 µl of 10x Taq DNA PCR buffer (Invitrogen), 7 µl of 50 mM $MgCl_2$ (BIOLINE), 1 µl of
609 Random Primers (Invitrogen), 2 µl of 10 mM dNTP (Invitrogen), 1 µl of DTT (Invitrogen), 2 µl of
610 SuperScript III reverse transcriptase (Invitrogen), and 10 µl of RNase-free water (Invitrogen). 40 µl of
611 eluted RNA was denatured at 95°C for 5 minutes and subsequently cooled on ice for 2 minutes. 30 µl
612 of master mix was added to each RNA sample, mixed gently by pipetting, and centrifuged briefly. The
613 tube was placed in a thermal cycler and incubated at 25°C for 10 minutes, 37°C for 1 hour, then 95°C
614 for 5 minutes. The cDNA was stored at -20°C.

615 Previously validated primers and probes were used for the qPCR-based detection of rotavirus
616 NSP2 [36] and VP6 [37] in two separate reactions. cDNA and reagents were thawed on ice. The master mix
617 for qPCR comprised: 12.5 µl of Platinum qPCR supermix-UDG 2x (Invitrogen), 2 µl of 10 µM forward
618 primer (5′-GAACTTCCTTGAA-TATAAGATCACACTGA-3′ [NSP2] or 5′-
619 GACGGVGCRACTACATGGT-3′ [VP6]), 2 µl of 10 µM reverse primer (5′-
620 TTGAAGACGTAAATGCATACCAATTC-3′ [NSP2] or 5′-GTCCAATTCAT-NCCTGGTG-3′ [VP6]),
621 0.25 µl of 10 µM probe (FAM5′-TCCAATAGATTGAAGTCAGTAACGT-TTCCA-3′MGB [NSP2] or
622 FAM5′-CCACCRAAYATGACRCCAGCNGTA-3′MGB [VP6]), and 6.25 µl of nuclease-free water
623 (ThermoFisher Scientific). In a 96-well plate, 23 µl of master mix was combined with 2 µl of cDNA. Six
624 standards for the target gene as well as a no-template control and a positive control were added to each
625 PCR plate. The plate was sealed with a Microseal 'C' Film (Bio-Rad) and subjected to the following
626 cycles in a Bio-Rad CFX qPCR Instrument (Bio-Rad): initial UDG-incubation 50°C for 2 minutes;
627 denaturation for 95°C for 2 minutes; then 40 cycles of 95°C for 15 seconds and 60°C for 1 minute. Data
628 were considered valid if amplification was absent in the negative control and present in the positive
629 control. qPCR standards consisting of TOPO-TA plasmid constructs containing either NPS2 or VP6
630 gene amplicons (genotype 1) were produced at the University of Liverpool and distributed to the other
631 study sites.

632

633 **Inflammatory biomarkers**

634 For infant stools collected before each dose of ORV, MPO and α1AT concentrations were
635 measured by ELISA at each site. ELISA kits with the same lot number were distributed across the study
636 sites. Stool samples were defrosted on ice and 100 mg then aliquoted into a polypropylene tube. ELISAs

637 were carried out following the manufacturer's instructions (IDK 2018 instructions for MPO and
638 BioVendor RIC6200 for α1AT). Paired samples from a given infant were run on the same plate, and
639 MPO and α1AT assays were carried out in tandem to avoid freeze-thaw cycles. For week 6 serum
640 samples of Indian and Malawian infants, α1AG was measured at CMC, Vellore, using a commercial
641 ELISA kit (abcam) according to the manufacturer's instructions.
642
643 **DNA extraction from stool**
644 DNA extraction from stool was carried out using the QIAamp DNA stool mini kit (Qiagen)
645 according to the manufacturer's instructions (QIAamp DNA Stool Handbook 06/2012), with several
646 modifications to increase DNA yield. Briefly, the vortexed suspension of ASL buffer and stool was
647 added to 370 mg of 0.1 mm zirconia-silicate beads (BioSpec) with 1.67 µl of lyzozyme (30 mg/ml)
648 (Sigma-Aldrich). Samples were then incubated at 37°C on a shaker (Eppendorf Thermo Mixer Comfort)
649 at 250-300 rpm for 10 minutes. 10 µl of proteinase K (Qiagen), 50 µl of 10% sodium dodecyl sulphate
650 (Sigma-Aldrich), and 20 µl of RNase A (1mg/ml; Sigma-Aldrich) was added before incubation on a
651 heating block (Eppendorf Thermo Mixer Comfort) at 70°C for 10 minutes. After allowing the samples
652 to cool for 3 minutes, bead beating was performed for 5 minutes using a Tissue Lyser II (Qiagen) at 25
653 Hz. Samples were microcentrifuged at 16,100 g for 1 minute and the supernatant was then transferred
654 into a new microcentrifuge tube containing an InhibitEX tablet and vortexed for 1 minute. The
655 suspension was incubated at room temperature for 1 minute, centrifuged at 16,100 g for 5 minutes, and
656 the supernatant (approximately 600 µl) then transferred into a new 2 ml tube. The supernatant was
657 centrifuged at 16,100 g for 3 minutes, then 400 µl added to a 2 ml tube containing 400 µl of buffer AL.
658 400 µl of ethanol (96–100%) was added to the lysate and mixed by vortexing. The manufacturer's
659 instructions were then resumed until elution in 50 µl of AE Buffer. An extraction control was carried
660 out with each batch of DNA extraction at all sites. Extracted DNA was shipped from India and Malawi
661 to the University of Liverpool on dry ice for 16S rRNA sequencing.
662
663 **Amplicon generation for 16S rRNA sequencing**
664 Amplicons spanning 16S rRNA gene variable regions 3 and 4 (primers 309F 5'-overhang-
665 ACTCCTACGGGAGGCAGCAG-3' and 819R 5'-overhang-GGACTACHVGGGTWTCTAAT-3') were
666 produced following the established Illumina protocol (Illumina, 16S Metagenomic Sequencing Library
667 Preparation Protocol Part # 15044223 Rev. B) with the following amendments: 1 µl of DNA was used
668 as a starting template; NEBNext Q5 Hot Start HiFi PCR Master Mix (NEB) was used for amplicon and

669 indexing PCRs; PCR reaction volumes were halved to 25 µl; primers were used at a concentration of 10
670 µM; and 20 µg of molecular-grade bovine serum albumin (NEB) was added to the amplicon PCR master
671 mix to mitigate the effect of PCR inhibitors. Cycling conditions for amplicon PCR involved:
672 denaturation at 98°C for 30 seconds; 10 cycles of 98°C for 10 seconds, 55°C for 15 seconds, and 72°C for
673 40 seconds; and final extension at 72°C for 60 seconds. Cycling conditions for indexing PCR involved:
674 denaturation at 98°C for 3 minutes; 15 cycles of 98°C for 10 seconds, 55°C for 15 seconds, and 72°C for
675 40 seconds; and final extension at 72°C for 5 minutes. AMPure XP beads were substituted with a custom
676 preparation of Sera-Mag SpeedBeads Protein A/G particles (MERCK/GE Healthcare) throughout the
677 protocol. To determine volumes required for equimolar pooling, amplicon concentrations were
678 determined by Quanti-it (ThermoFisher Scientific) and size distributions by fragment analysis on a 5300
679 Fragment Analyzer System (Agilent). Amplicons were pooled using a mosquito X1 (TTP Labtech)
680 liquid handling robot. The final library underwent size selection to remove potentially contaminating
681 primer-dimers and genomic DNA using the Pippin Prep (Sage Science) 1.5% Agarose Gel Cassette
682 (Labtech).

683 Overall, we sequenced amplicons from 2,138 samples across 14 Illumina HiSeq2500 lanes (v2
684 chemistry, 600 cycles in rapid run mode) and one Illumina MiSeq lane (v3 chemistry, 600 cycles). Since
685 enrolment and sample collection across the three sites were not synchronised, sequencing was batched
686 by geographic origin according to sample availability. Samples from a mother–infant pair were
687 processed on the same PCR plate. Each PCR plate contained a no-template PCR control, stool controls
688 provided by a mother and infant in the UK who were not enrolled in the study, DNA from a mock
689 community (Zymo Research D6306), and a pool of extraction controls corresponding to the samples
690 contained on the PCR plate. Final libraries consisted of up to four PCR plates (384 amplicons). Eight
691 samples were excluded from the analysis owing to the presence of significant amplification from their
692 corresponding extraction controls (**Figure S8**).

693

694 **Illumina sequencing of 16S rRNA libraries**

695 The quantity and quality of each amplicon pool was assessed by Qubit HS DNA kit
696 (ThermoFisher Scientific), Bioanalyzer High Sensitivity DNA Kit (Agilent), and qPCR using the KAPA
697 Library Quantification Kit (Roche) on a Roche Light Cycler LC480II (Roche) according to the
698 manufacturer's instructions. qPCR data were used to calculate sample molarity. 5 µl of the final pool
699 was denatured for 5 minutes at room temperature using 5 µl of freshly diluted 0.1 N sodium hydroxide
700 (Illumina). The reaction was subsequently terminated by the addition of HT1 hybridization buffer

701 (Illumina) and the library diluted post-denaturation to a final loading concentration of 7–8.5 pM.
702 Libraries were sequenced with 10–20% PhiX (Illumina) using the 2x300 bp paired-end protocol. The
703 raw sequence data generated during this study have been deposited in the European Nucleotide
704 Archive (accession number PRJEB38948).
705
706 **Independent validation of 16S rRNA sequencing**
707 To validate the robustness of the microbiota protocol, 10% of DNA samples were shipped to
708 Imperial College London and sequenced according to the methods described above. We included 30
709 mother–infant pairs per study site in the validation subset and sequenced two infant stools (from weeks
710 1 and 10/12) and the maternal stool from each pair. Protocol deviations included the use of AMPure
711 XP beads for PCR product purification, fragment analysis via TapeStataion D1000 ScreenTape (Agilent
712 Technologies), concentration quantification via Qubit dsDNA High Sensitivity Assay Kit
713 (ThermoFisher Scientific), and manual pooling of equimolar quantities. The order of samples was
714 randomised across three PCR plates and the pooled libraries were sequenced with 10–20% PhiX
715 (Illumina) on two 2x300 bp paired-end Illumina MiSeq runs (v3 chemistry).
716
717 **Bioinformatic processing of sequence data**
718 Adapters were trimmed from raw sequences using cutadapt version 1.18 [38]. Subsequent steps
719 for quality filtering, denoising, merging of paired reads, and chimera removal followed the DADA2
720 pipeline [39] implemented in QIIME 2 with default parameters [40]. Forward reads were truncated to 270
721 bp and reverse reads to 200 bp to account for the fall in sequencing quality scores towards the end of
722 each read. Sequence data were processed separately for each run, and the feature tables and
723 corresponding sequences then merged. All subsequent steps were performed in the programming
724 language R. Taxonomy assignment was performed using the RDP naïve Bayesian classifier trained on
725 the Silva rRNA database (version 132). Sequence variants were included in the analysis if they were
726 390–440 bp in length, bacterial, detectable at an abundance of ≥0.1% in at least two samples, and passed
727 frequency-based contamination filtering using the *decontam* package (which screens sequences that are
728 more abundant in low-concentration samples) [41]. Nanodrop concentrations (ng/μl) of extracted DNA
729 served as the basis for contaminant filtering. Samples with at least 25,000 quality-filtered sequences
730 were retained in the final analysis.
731
732

733 **Statistical analysis**

734 Processed data and analysis code generated during this study are available on github
735 (https://github.com/eparker12/RoVI). All analyses were performed in the programming language R.
736 Seroconversion and shedding proportions were compared by country using pairwise Fisher's exact
737 tests with Benjamini–Hochberg FDR correction. Antibody and inflammatory biomarker concentrations
738 were log transformed and compared by country and shedding subgroup (see **Figure 1E**) using ANOVA
739 with post-hoc Tukey tests. To explore covariates associated with ORV outcome, we used either logistic
740 regression (for seroconversion and shedding status) or linear regression and Pearson's *r* (for log-
741 transformed post-vaccination RV-IgA). In Indian mother–infant dyads, Spearman's rank correlation
742 coefficients were calculated between all measured antibody concentrations (RV-IgG and RV-IgA) and
743 rotavirus shedding quantities (reciprocal of qPCR Ct value). A hierarchical clustering analysis of all
744 infant and maternal gut microbiota samples was conducted using Ward's minimum variance
745 hierarchical clustering method based on the presence or absence of common genera (≥1% prevalence
746 at ≥0.1% abundance). Longitudinal mixed-effects models were used to compare microbiota diversity
747 (Shannon index, genus-level richness) by country (pairwise comparisons with FDR correction) and
748 ORV outcome (seroconversion and shedding status), including week as a covariate and study ID as a
749 random effect. We used longitudinal zero-inflated negative binomial models of genus counts to identify
750 taxa that discriminated infants according to country, seroconversion, and shedding status, adjusting
751 for age (by including week as a fixed effect) and including study ID as a random effect. For taxa with a
752 prevalence of >95%, negative binomial models without zero inflation were used. Fisher's exact test was
753 also used for cross-sectional comparisons of genus prevalence (≥0.1% abundance threshold) by country,
754 excluding taxa that were ubiquitous (>95% in both comparison groups) or scarce (<5% in both groups).
755 We report on genera with FDR-adjusted p values of <0.1.

756 We applied the Random Forests algorithm in a series of cross-sectional analyses to predict
757 country, seroconversion, or shedding status (classification approach), or post-vaccination log-
758 transformed RV-IgA concentration (regression approach) based on genus relative abundances. For each
759 analysis, we performed 20 iterations of 5-fold cross-validation. For classification models (of binary
760 outcomes), we standardised the baseline accuracy at 50% by fitting each iteration of 5-fold cross-
761 validation on a random subset of 50 samples per group (or the number of samples in the minority group
762 if this was <50). Models were excluded if there were <10 samples in the minority group. Mean cross-
763 validation variable importance was determined based on the increase in Gini index (classification) or
764 mean squared error (regression) of out-of-bag sample prediction following random permutation. For

765 Indian infants, we assessed additional models based on demography/baseline health measurements
766 (18 variables), exposure/antibody data (12 variables, encompassing inflammatory biomarkers, wild-
767 type rotavirus shedding, and maternal antibodies), and genus relative abundances (126 variables), and
768 all variables combined.

769 For technical replicates (including positive controls and samples selected for independent
770 validation), we quantified the proportion of variance explained by sample ID using linear regression
771 (alpha diversity, genus abundances) or PERMANOVA with 999 permutations (weighted and
772 unweighted Bray–Curtis distances). We also used PERMANOVA to quantify the proportion of
773 variation in microbiota composition associated with sequencing run, stratified by age and country.
774

775 **Supplemental table titles and legends**

776 **Table S1. Sample counts for primary outcomes.** α1AG, α1 acid glycoprotein; α1AT, α1-antitrypsin;
777 MPO, myeloperoxidase; neo+, infected with rotavirus pre-vaccination; neo-, uninfected with rotavirus
778 pre-vaccination; RV, rotavirus.

779

780 **Table S2. Cofactors associated with neonatal rotavirus infection.** α1AG, α1 acid glycoprotein; α1AT,
781 α1-antitrypsin; bOPV, bivalent oral poliovirus vaccine; GMC, geometric mean concentration; IPV,
782 inactivated poliovirus vaccine; LRT, likelihood ratio test for fit of model with vs without polio vaccine
783 schedule; MPO, myeloperoxidase; ORV, oral rotavirus vaccine; RR, relative risk; RV, rotavirus; tOPV,
784 trivalent oral poliovirus vaccine; *log-transformed to approximate normality in statistical models.

785

786 **Table S3A. Cofactors associated with rotavirus vaccine shedding.** Data are presented for (A) dose 1
787 shedding and (B) shedding after either dose. Measurements obtained after the first dose of ORV are
788 excluded to focus on potential predictors of ORV shedding rather than shedding–induced changes.
789 α1AG, α1 acid glycoprotein; α1AT, α1–antitrypsin; bOPV, bivalent oral poliovirus vaccine; GMC,
790 geometric mean concentration; IPV, inactivated poliovirus vaccine; LRT, likelihood ratio test for fit of
791 model with vs without polio vaccine schedule; MPO, myeloperoxidase; ORV, oral rotavirus vaccine;
792 RR, relative risk; RV, rotavirus; tOPV, trivalent oral poliovirus vaccine; *log–transformed to
793 approximate normality in statistical models; † excluded to minimise multicollinearity.

794

795 **Table S4. Cofactors associated with rotavirus seroconversion.** α1AG, α1 acid glycoprotein; α1AT, α1-
796 antitrypsin; bOPV, bivalent oral poliovirus vaccine; GMC, geometric mean concentration; IPV,
797 inactivated poliovirus vaccine; LRT, likelihood ratio test for fit of model with vs without polio vaccine
798 schedule; MPO, myeloperoxidase; ORV, oral rotavirus vaccine; RR, relative risk; RV, rotavirus; tOPV,
799 trivalent oral poliovirus vaccine; *log-transformed to approximate normality in statistical models; †
800 excluded to minimise multicollinearity.

801

802 **Table S5. Cofactors associated with post-vaccination RV-IgA concentration.** α1AG, α1 acid
803 glycoprotein; α1AT, α1-antitrypsin; bOPV, bivalent oral poliovirus vaccine; GMC, geometric mean
804 concentration; IPV, inactivated poliovirus vaccine; LRT, likelihood ratio test for fit of model with vs
805 without polio vaccine schedule; MPO, myeloperoxidase; ORV, oral rotavirus vaccine; RV, rotavirus;

806 tOPV, trivalent oral poliovirus vaccine; *log-transformed to approximate normality in statistical
807 models; † excluded to minimise multicollinearity.
808
809 **Table S6. Geographic discrepancies in genus prevalence.** Data are presented for (A) Malawi vs India,
810 (B) India vs UK, and (C) Malawi vs UK. Genus prevalences were compared using Fisher's exact test,
811 with a 0.1% abundance threshold to define presence/absence. Taxa were excluded if ubiquitous (>95%
812 in both groups) or scarce (<5% in both groups).
813
814 **Table S7. Taxa correlated with ORV response in longitudinal zero-inflated negative binomial**
815 **models.** FDR, false discovery rate; ORV, oral rotavirus vaccine; RV, rotavirus.

816 **References**

817 1 Mokomane, M., Kasvosve, I., de Melo, E., Pernica, J. M. & Goldfarb, D. M. The global problem
818 of childhood diarrhoeal diseases: emerging strategies in prevention and management. *Ther Adv*
819 *Infect Dis* **5**, 29-43 (2018).

820 2 Tate, J. E., Burton, A. H., Boschi-Pinto, C., Parashar, U. D. & World Health Organization-
821 Coordinated Global Rotavirus Surveillance, N. Global, regional, and national estimates of
822 rotavirus mortality in children <5 years of age, 2000-2013. *Clin Infect Dis* **62 Suppl 2**, S96-S105
823 (2016).

824 3 Troeger, C. *et al.* Estimates of the global, regional, and national morbidity, mortality, and
825 aetiologies of diarrhoea in 195 countries: a systematic analysis for the Global Burden of Disease
826 Study 2016. *Lancet Infect Dis* **18**, 1211-1228 (2018).

827 4 Vesikari, T. *et al.* Efficacy of human rotavirus vaccine against rotavirus gastroenteritis during
828 the first 2 years of life in European infants: randomised, double-blind controlled study. *Lancet*
829 **370**, 1757-1763 (2007).

830 5 Madhi, S. A. *et al.* Effect of human rotavirus vaccine on severe diarrhea in African infants. *N*
831 *Engl J Med* **362**, 289-298 (2010).

832 6 Church, J. A., Parker, E. P., Kirkpatrick, B. D., Grassly, N. C. & Prendergast, A. J. Interventions
833 to improve oral vaccine performance: a systematic review and meta-analysis. *Lancet Infect Dis*
834 **19**, 203-214 (2019).

835 7 Parker, E. P. *et al.* Causes of impaired oral vaccine efficacy in developing countries. *Future*
836 *Microbiol* **13**, 97-118 (2018).

837 8 Moon, S. S. *et al.* Prevaccination rotavirus serum IgG and IgA are associated with lower
838 immunogenicity of live, oral human rotavirus vaccine in South African infants. *Clin Infect Dis*
839 **62**, 157-165 (2016).

840 9 Voysey, M. *et al.* The Influence of Maternally Derived Antibody and Infant Age at Vaccination
841 on Infant Vaccine Responses : An Individual Participant Meta-analysis. *JAMA Pediatr* **171**, 637-
842 646 (2017).

843 10 Naylor, C. *et al.* Environmental Enteropathy, Oral Vaccine Failure and Growth Faltering in
844 Infants in Bangladesh. *EBioMedicine* **2**, 1759-1766 (2015).

845 11 Church, J. A. *et al.* Exploring the relationship between environmental enteric dysfunction and
846 oral vaccine responses. *Future Microbiol* **13**, 1055-1070 (2018).

847 12 Harris, V. C. *et al.* Significant Correlation Between the Infant Gut Microbiome and Rotavirus
848 Vaccine Response in Rural Ghana. *J Infect Dis* **215**, 34-41 (2017).
849 13 Parker, E. P. K. *et al.* Influence of the intestinal microbiota on the immunogenicity of oral
850 rotavirus vaccine given to infants in south India. *Vaccine* **36**, 264-272 (2018).
851 14 Bennett, A. *et al.* Infrequent Transmission of Monovalent Human Rotavirus Vaccine Virus to
852 Household Contacts of Vaccinated Infants in Malawi. *J Infect Dis* **219**, 1730-1734 (2019).
853 15 Lazarus, R. P. *et al.* The effect of probiotics and zinc supplementation on the immune response
854 to oral rotavirus vaccine: A randomized, factorial design, placebo-controlled study among
855 Indian infants. *Vaccine* **36**, 273-279 (2018).
856 16 Patel, M. *et al.* A systematic review of anti-rotavirus serum IgA antibody titer as a potential
857 correlate of rotavirus vaccine efficacy. *J Infect Dis* **208**, 284-294 (2013).
858 17 Lewnard, J. A. *et al.* Naturally Acquired Immunity Against Rotavirus Infection and
859 Gastroenteritis in Children: Paired Reanalyses of Birth Cohort Studies. *J Infect Dis* **216**, 317-326
860 (2017).
861 18 Moon, S. S. *et al.* Inhibitory effect of breast milk on infectivity of live oral rotavirus vaccines.
862 *Pediatr Infect Dis J* **29**, 919-923 (2010).
863 19 Moon, S. S. *et al.* Differential profiles and inhibitory effect on rotavirus vaccines of nonantibody
864 components in breast milk from mothers in developing and developed countries. *Pediatr Infect*
865 *Dis J* **32**, 863-870 (2013).
866 20 Phillips, G., Lopman, B., Rodrigues, L. C. & Tam, C. C. Asymptomatic rotavirus infections in
867 England: prevalence, characteristics, and risk factors. *Am J Epidemiol* **171**, 1023-1030 (2010).
868 21 Becker-Dreps, S. *et al.* Rotavirus-specific IgG antibodies from mothers' serum may inhibit infant
869 immune responses to the pentavalent rotavirus vaccine. *Pediatr Infect Dis J* **34**, 115-116 (2015).
870 22 Baba, T. W. *et al.* Human neutralizing monoclonal antibodies of the IgG1 subtype protect
871 against mucosal simian-human immunodeficiency virus infection. *Nat Med* **6**, 200-206 (2000).
872 23 Cohen, O. R. *et al.* Systemically administered IgG anti-toxin antibodies protect the colonic
873 mucosa during infection with Clostridium difficile in the piglet model. *PLoS One* **9**, e111075
874 (2014).
875 24 Ali, A. *et al.* Impact of withholding breastfeeding at the time of vaccination on the
876 immunogenicity of oral rotavirus vaccine--a randomized trial. *PLoS One* **10**, e0127622 (2015).
877 25 Rongsen-Chandola, T. *et al.* Effect of withholding breastfeeding on the immune response to a
878 live oral rotavirus vaccine in North Indian infants. *Vaccine* **32 Suppl 1**, A134-139 (2014).

879 26 Bucardo, F. *et al.* The Lewis A phenotype is a restriction factor for Rotateq and Rotarix vaccine-
880 take in Nicaraguan children. *Sci Rep* **8**, 1502 (2018).
881 27 Huda, M. N. *et al.* Bifidobacterium Abundance in Early Infancy and Vaccine Response at 2
882 Years of Age. *Pediatrics* **143** (2019).
883 28 Praharaj, I. *et al.* Influence of Nonpolio Enteroviruses and the Bacterial Gut Microbiota on Oral
884 Poliovirus Vaccine Response: A Study from South India. *J Infect Dis* **219**, 1178-1186 (2019).
885 29 Taniuchi, M. *et al.* Impact of enterovirus and other enteric pathogens on oral polio and rotavirus
886 vaccine performance in Bangladeshi infants. *Vaccine* **34**, 3068-3075 (2016).
887 30 Abrahamsson, T. R. *et al.* Low diversity of the gut microbiota in infants with atopic eczema. *J*
888 *Allergy Clin Immunol* **129**, 434-440, 440 e431-432 (2012).
889 31 Valdes, A. M., Walter, J., Segal, E. & Spector, T. D. Role of the gut microbiota in nutrition and
890 health. *BMJ* **361**, k2179 (2018).
891 32 Lee, A. H. *et al.* Dynamic molecular changes during the first week of human life follow a robust
892 developmental trajectory. *Nat Commun* **10**, 1092 (2019).
893 33 Sindhu, K. N. *et al.* Impact of maternal antibodies and infant gut microbiota on the
894 immunogenicity of rotavirus vaccines in African, Indian and European infants: protocol for a
895 prospective cohort study. *BMJ Open* **7**, e016577 (2017).
896 34 Church, J. A. *et al.* Predictors of oral rotavirus vaccine immunogenicity in rural Zimbabwean
897 infants. *Vaccine* **38**, 2870-2878 (2020).
898 35 Paul, A., Gladstone, B. P., Mukhopadhya, I. & Kang, G. Rotavirus infections in a community
899 based cohort in Vellore, India. *Vaccine* **32 Suppl 1**, A49-54 (2014).
900 36 Gautam, R. *et al.* One-step multiplex real-time RT-PCR assay for detecting and genotyping
901 wild-type group A rotavirus strains and vaccine strains (Rotarix(R) and RotaTeq(R)) in stool
902 samples. *PeerJ* **4**, e1560 (2016).
903 37 Iturriza Gomara, M., Wong, C., Blome, S., Desselberger, U. & Gray, J. Molecular
904 characterization of VP6 genes of human rotavirus isolates: correlation of genogroups with
905 subgroups and evidence of independent segregation. *J Virol* **76**, 6596-6601 (2002).
906 38 Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads.
907 *EMBnet.journal* **17(1)**, 10-12 (2011).
908 39 Callahan, B. J. *et al.* DADA2: High-resolution sample inference from Illumina amplicon data.
909 *Nat Methods* **13**, 581-583 (2016).

40 Bolyen, E. *et al*. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat Biotechnol* **37**, 852-857 (2019).

41 Davis, N. M., Proctor, D. M., Holmes, S. P., Relman, D. A. & Callahan, B. J. Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data. *Microbiome* **6** (2018).

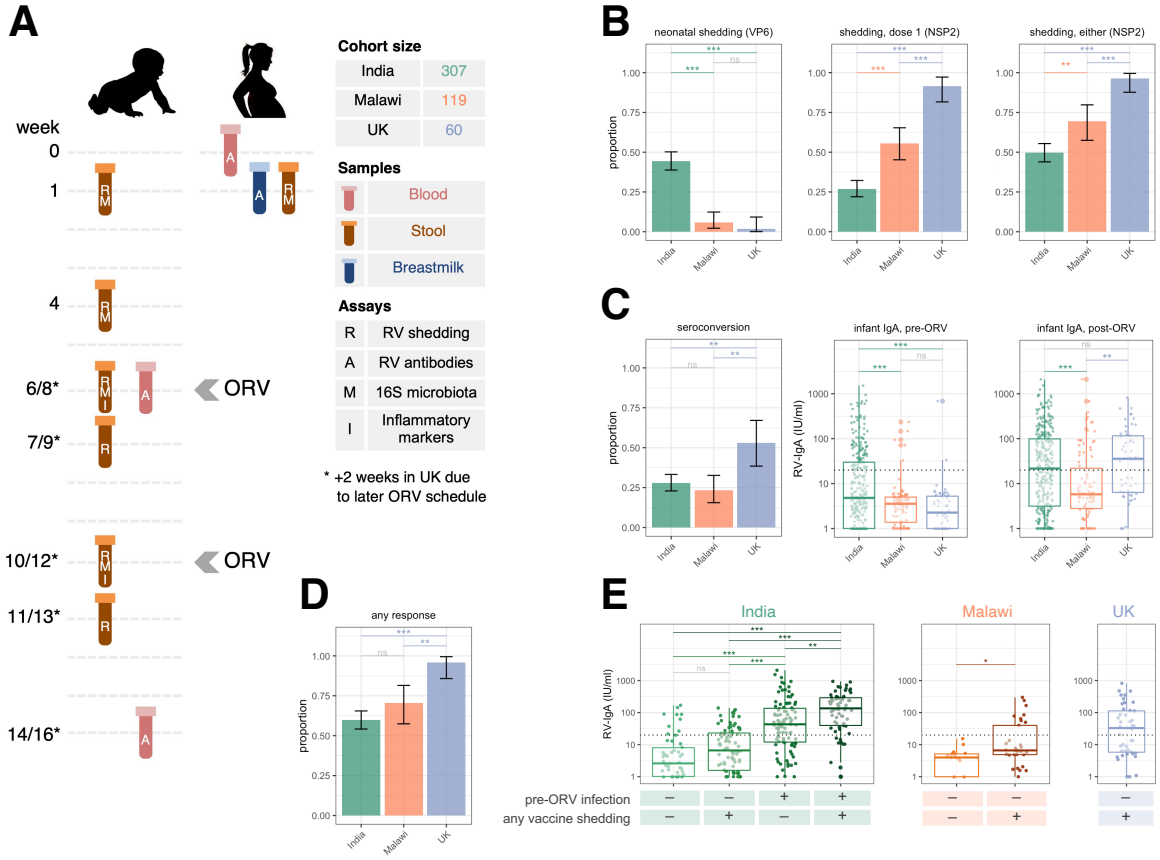

A
week
0
1
4
6/8*
7/9*
10/12*
11/13*
14/16*
ORV
ORV
Cohort size
India 307
Malawi 119
UK 60
Samples
Blood
Stool
Breastmilk
Assays
R RV shedding
A RV antibodies
M 16S microbiota
I Inflammatory markers
* +2 weeks in UK due to later ORV schedule
B
neonatal shedding (VP6)
shedding, dose 1 (NSP2)
shedding, either (NSP2)
proportion
India
Malawi
UK
C
seroconversion
infant IgA, pre-ORV
infant IgA, post-ORV
RV-IgA (IU/ml)
D
any response
E
India
Malawi
UK
pre-ORV infection
any vaccine shedding

Figure 1

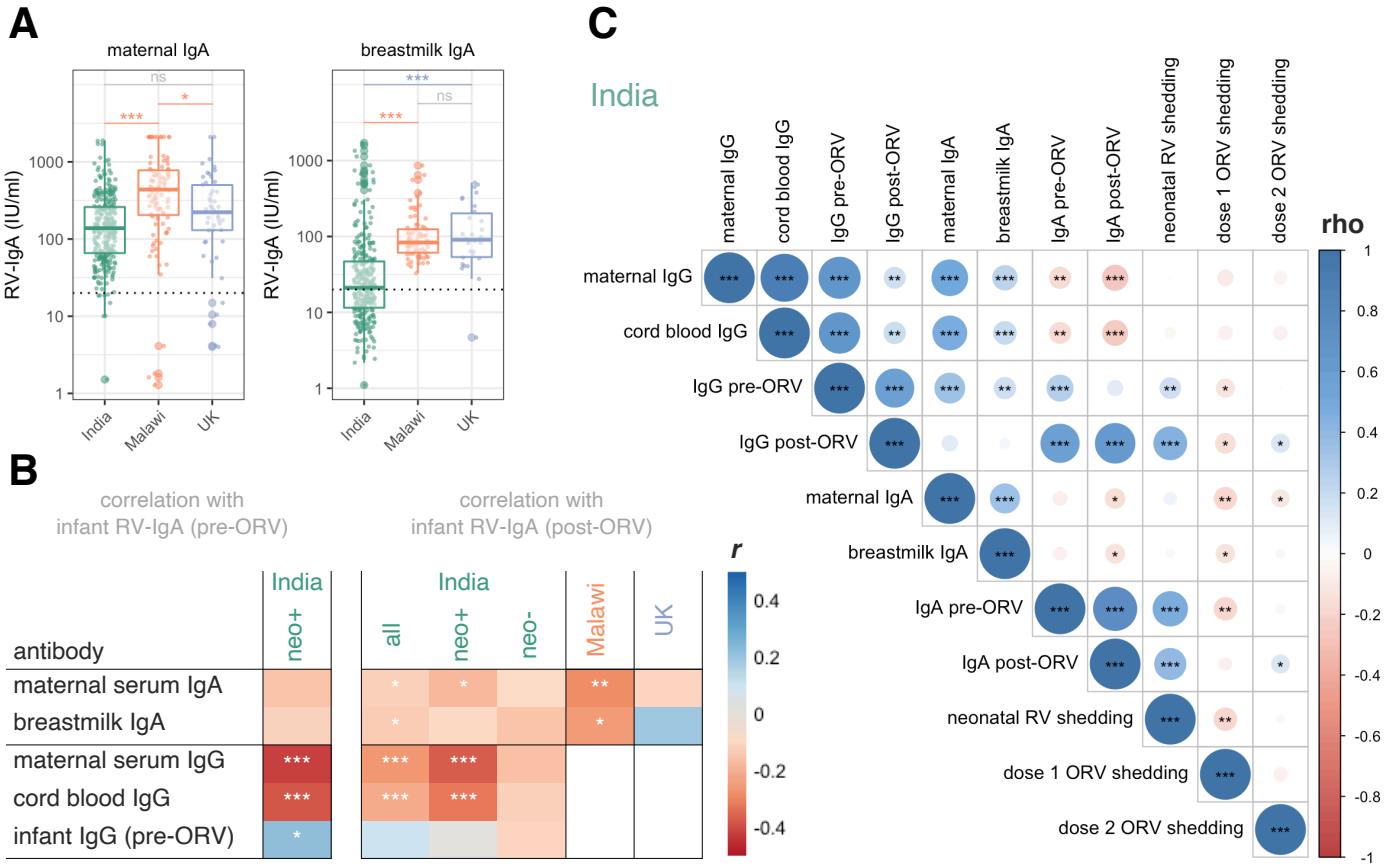

A
maternal IgA
breastmilk IgA
RV-IgA (IU/ml)
1000
100
10
1
ns
***
*
India
Malawi
UK
B
correlation with
infant RV-IgA (pre-ORV)
correlation with
infant RV-IgA (post-ORV)
India
neo+
India
all
neo+
neo-
Malawi
UK
antibody
maternal serum IgA
breastmilk IgA
maternal serum IgG
cord blood IgG
infant IgG (pre-ORV)
r
0.4
0.2
0
-0.2
-0.4
C
India
maternal IgG
cord blood IgG
IgG pre-ORV
IgG post-ORV
maternal IgA
breastmilk IgA
IgA pre-ORV
IgA post-ORV
neonatal RV shedding
dose 1 ORV shedding
dose 2 ORV shedding
rho
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1

Figure 2

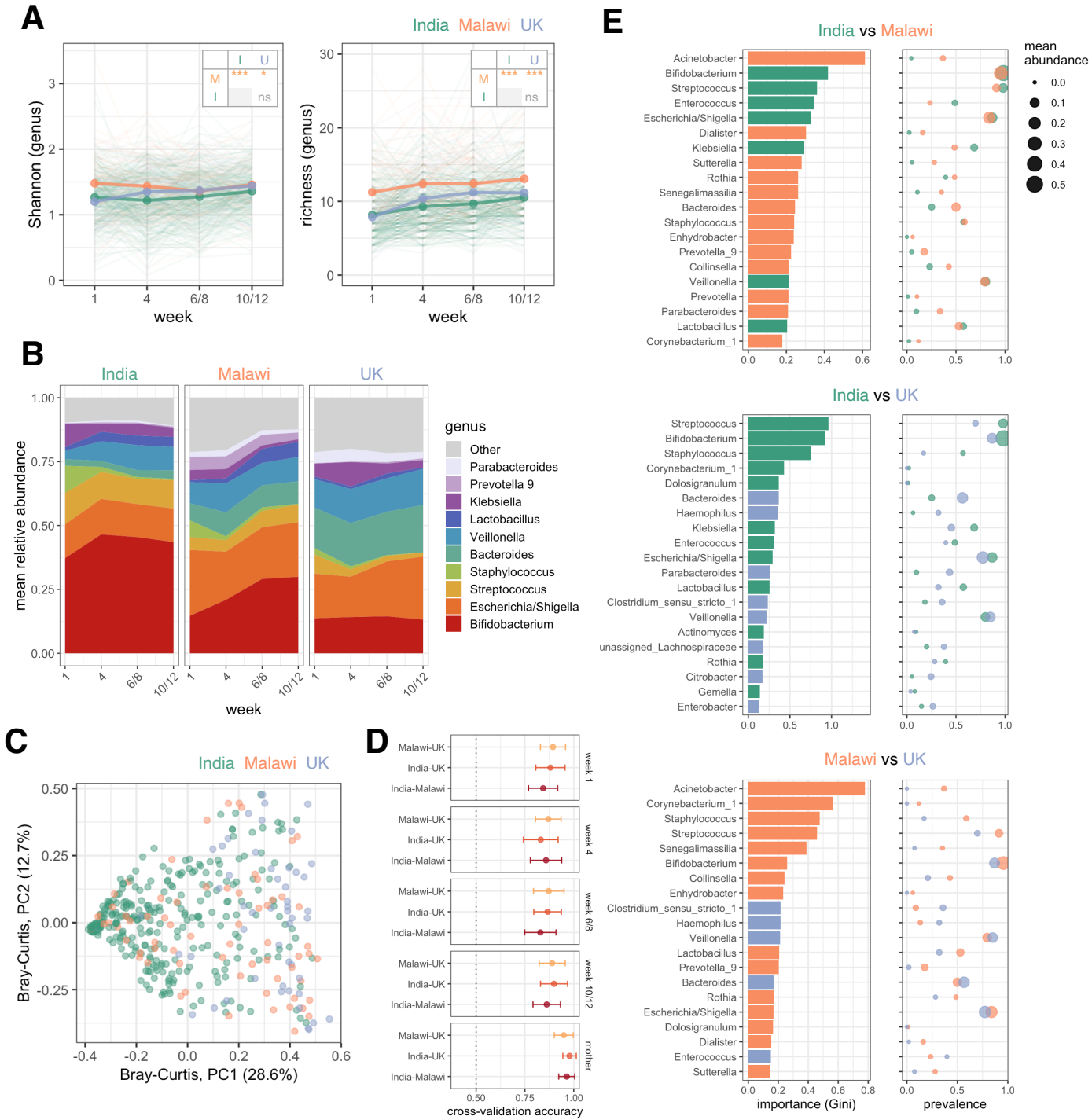
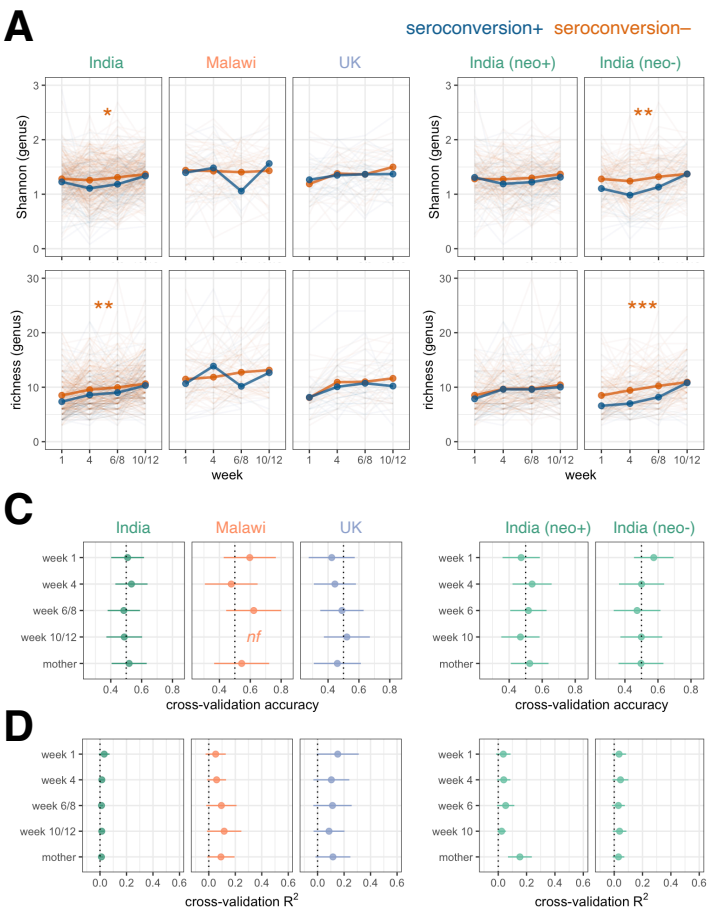
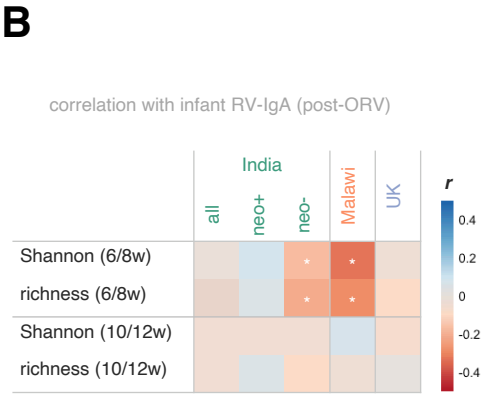

A
India Malawi UK
Shannon (genus)
richness (genus)
week
B
India
Malawi
UK
mean relative abundance
genus
Other
Parabacteroides
Prevotella 9
Klebsiella
Lactobacillus
Veillonella
Bacteroides
Staphylococcus
Streptococcus
Escherichia/Shigella
Bifidobacterium
C
Bray-Curtis, PC2 (12.7%)
Bray-Curtis, PC1 (28.6%)
D
Malawi-UK
India-UK
India-Malawi
week 1
week 4
week 6/8
week 10/12
mother
cross-validation accuracy
E
India vs Malawi
India vs UK
Malawi vs UK
mean abundance
importance (Gini)
prevalence

Figure 3





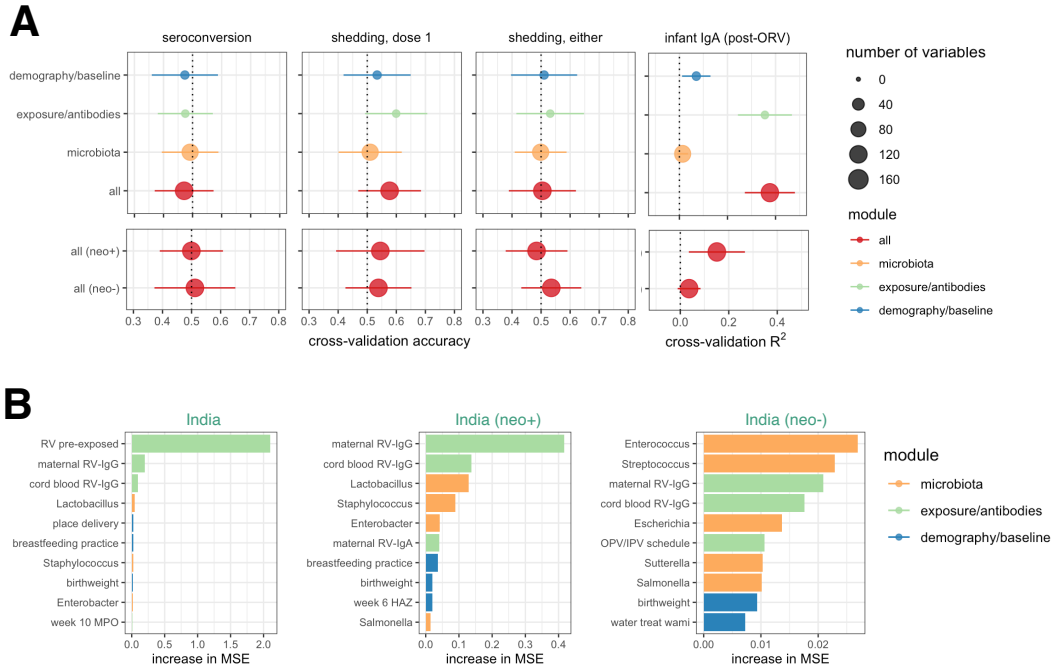
E

longitudinal ZINB models of taxon abundance

	India			Malawi	UK
	all	neo+	neo-		
N genera tested*	25	26	24	42	25
Increases with age	7	8	8	16	4
Decreases with age	6	7	6	13	13
+ correlation with seroconversion	0	0	0	0	0
– correlation with seroconversion	0	0	0	0	0
+ correlation with RV-IgA post-ORV	0	0	0	0	0
– correlation with RV-IgA post-ORV	0	1	0	0	0

** Present in ≥20% of samples*

Figure 4



C

outcome: infant RV-IgA (post-ORV)	India (full, n = 286)		India (neo+, n = 166)		India (neo-, n = 124)	
variable	adjusted beta (95% CI)	p	adjusted beta (95% CI)	p	adjusted beta (95% CI)	p
delivery mode (caesarean vs vaginal)	1.45 (0.93–2.28)	0.102	–	–	1.74 (0.81–3.73)	0.153
delivery place (tertiary vs non-tertiary)	1.08 (0.72–1.64)	0.701	–	–	1.17 (0.60–2.28)	0.637
breastfeeding practice (exclusive vs non-exclusive)	**1.89 (1.11–3.22)**	**0.019**	**3.45 (1.62–7.33)**	**0.001**	–	–
height-for-age Z score (6 weeks)	**1.31 (1.10–1.55)**	**0.002**	1.25 (0.99–1.58)	0.061	–	–
exposed to rotavirus pre-vaccination	**8.48 (5.67–12.67)**	**<0.001**	–	–	–	–
polio vaccine schedule	–	–	–	–	–	–
bOPV only	ref	–	–	–	–	–
bOPV/tOPV mixed	0.48 (0.28–0.85)	0.011	–	–	–	–
tOPV only	**0.43 (0.25–0.75)**	**0.003**	–	–	–	–
IPV	**0.77 (0.49–1.22)**	**0.265**	–	–	–	–
house made of permanent materials (yes vs no)	1.28 (0.86–1.91)	0.228	–	–	**1.72 (1.02–2.92)**	**0.043**
access to treated water (yes vs no)	1.35 (0.93–1.96)	0.120	–	–	**1.82 (1.08–3.08)**	**0.026**
maternal RV-IgG (IU/ml), log-transformed	**0.59 (0.49–0.71)**	**<0.001**	**0.54 (0.41–0.70)**	**<0.001**	–	–
Shannon index (6 weeks)	–	–	–	–	**0.48 (0.26–0.90)**	**0.021**
adjusted R^2	0.434		0.194		0.108	

Figure 5