Article

Assessing transmissibility of SARS-CoV-2 lineage B.1.1.7 in England

https://doi.org/10.1038/s41586-021-03470-x

Received: 31 December 2020

Accepted: 18 March 2021

Published online: 25 March 2021

Check for updates

Erik Volz[1,104]✉, Swapnil Mishra[1,104], Meera Chand[2,104], Jeffrey C. Barrett[3,104], Robert Johnson[1,104], Lily Geidelberg[1], Wes R. Hinsley[1], Daniel J. Laydon[1], Gavin Dabrera[2], Áine O'Toole[4], Robert Amato[3], Manon Ragonnet-Cronin[1], Ian Harrison[2], Ben Jackson[4], Cristina V. Ariani[3], Olivia Boyd[1], Nicholas J. Loman[2,5], John T. McCrone[4], Sónia Gonçalves[3], David Jorgensen[1], Richard Myers[2], Verity Hill[4], David K. Jackson[3], Katy Gaythorpe[1], Natalie Groves[2], John Sillitoe[3], Dominic P. Kwiatkowski[3], The COVID-19 Genomics UK (COG-UK) consortium*, Seth Flaxman[6], Oliver Ratmann[6], Samir Bhatt[1,7], Susan Hopkins[2], Axel Gandy[6,104], Andrew Rambaut[4,104] & Neil M. Ferguson[1,104]✉

The SARS-CoV-2 lineage B.1.1.7, designated variant of concern (VOC) 202012/01 by Public Health England[1], was first identified in the UK in late summer to early autumn 2020[2]. Whole-genome SARS-CoV-2 sequence data collected from community-based diagnostic testing for COVID-19 show an extremely rapid expansion of the B.1.1.7 lineage during autumn 2020, suggesting that it has a selective advantage. Here we show that changes in VOC frequency inferred from genetic data correspond closely to changes inferred by *S* gene target failures (SGTF) in community-based diagnostic PCR testing. Analysis of trends in SGTF and non-SGTF case numbers in local areas across England shows that B.1.1.7 has higher transmissibility than non-VOC lineages, even if it has a different latent period or generation time. The SGTF data indicate a transient shift in the age composition of reported cases, with cases of B.1.1.7 including a larger share of under 20-year-olds than non-VOC cases. We estimated time-varying reproduction numbers for B.1.1.7 and co-circulating lineages using SGTF and genomic data. The best-supported models did not indicate a substantial difference in VOC transmissibility among different age groups, but all analyses agreed that B.1.1.7 has a substantial transmission advantage over other lineages, with a 50% to 100% higher reproduction number.

The SARS-CoV-2 lineage B.1.1.7 spread rapidly across England between November 2020 and January 2021. This variant possesses a large number of non-synonymous substitutions of immunological importance[2]. The N501Y replacement on the spike protein has been shown to increase ACE2 binding[3,4] and cell infectivity in animal models[5], and the P618H replacement on the spike protein adjoins the furin-cleavage site[6]. B.1.1.7 also possesses a deletion at positions 69 and 70 of the spike protein (Δ69–70) that has been associated with failure of diagnostic tests using the ThermoFisher TaqPath probe, which targets the spike protein[7]. Although other variants with Δ69–70 are also circulating in the UK, the absence of detection of the *S* gene target in an otherwise positive PCR test appears to be a highly specific biomarker for the B.1.1.7 lineage. Data from national community testing in November 2020 showed a rapid increase in SGTF during PCR testing for SARS-CoV-2, coinciding with a rapid increase in the frequency of B.1.1.7 observed in genomic surveillance. The B.1.1.7 lineage was designated VOC 202012/01 by Public Health England (PHE) in December 2020.

Phylogenetic studies carried out by the UK COVID-19 Genomics Consortium (COG-UK) (https://www.cogconsortium.uk)[8] provided the first indication that B.1.1.7 has an unusual accumulation of substitutions and was growing at a higer rate than other circulating lineages. We investigated time trends in the frequency of sampling VOC genomes and the proportion of PCR tests exhibiting SGTF across the UK, which we calibrated as a biomarker of VOC infection. Using multiple approaches and both genetic and SGTF data, we conclude that B.1.1.7 is associated with a higher reproduction number (R) than previous non-VOC lineages.

We examined whole-genome SARS-CoV-2 sequences from randomly sampled residual materials obtained from community-based COVID-19 testing in England, collected between 1 October 2020 and 16 January 2021. These data included 31,390 B.1.1.7 sequences for which the time and location of sample collection were known. Over the same period, 52,795 non-VOC genomes were collected. VOC sequences were initially concentrated in London (n = 9,134), the South East (n = 5,609), and the East of England (n = 4,413), but is now widely distributed across England.

[1]MRC Centre for Global Infectious Disease Analysis, Jameel Institute for Disease and Emergency Analytics, Imperial College London, London, UK. [2]Public Health England, London, UK. [3]Wellcome Sanger Institute, Cambridge, UK. [4]Institute of Evolutionary Biology, University of Edinburgh, Edinburgh, UK. [5]Institute of Microbiology and Infection, University of Birmingham, Birmingham, UK. [6]Department of Mathematics, Imperial College London, London, UK. [7]Section of Epidemiology, Department of Public Health, University of Copenhagen, Copenhagen, Denmark. [104]These authors contributed equally: Erik Volz, Swapnil Mishra, Meera Chand, Jeffrey C. Barrett, Robert Johnson, Axel Gandy, Andrew Rambaut, Neil M. Ferguson. *A list of authors and their affiliations appears online. ✉e-mail: e.volz@imperial.ac.uk; neil.ferguson@imperial.ac.uk

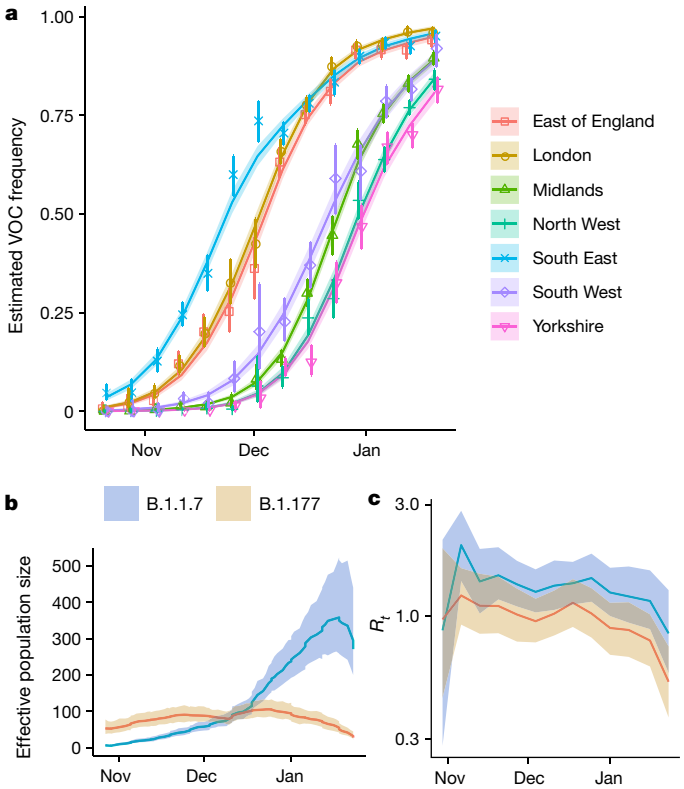


Fig. 1 | Expansion of lineage B.1.1.7 relative to co-circulating lineages in England. a, Estimated frequency of sampling the VOC (lines) over time in NHS regions (n = 84,185). Shaded regions, 95% credible region based on Bayesian regression; points, empirical proportions of the VOC in each week; error bars, 95% CI based on binomial sampling error. **b**, Effective population size over time for lineage B.1.1.7 and estimates based on a matched sample of the most abundant co-circulating lineage, B.1.177 (n = 3,000). Shaded regions, 95% bootstrap CI. **c**, The effective reproduction number inferred from growth of effective population size for both lineages in **b**.

Overall, we estimate the median posterior additive difference in growth rates between B.1.1.7 and co-circulating variants to be 0.69 per week (95% credible interval (CrI) 0.61–0.76) (Fig. 1a, Extended Data Fig. 1, Supplementary Methods section 2), and this difference was largest in November. However, in tandem with geographic expansion of the VOC and imposition of lockdown measures in 2021, this difference declined gradually to 0.43 per week (95% CrI 0.33–0.52) for the week ending 16 January.

The rate of genetic diversification of the VOC lineage over time allows epidemic growth rates to be estimated using phylodynamic modelling[9,10]. To contrast VOC and non-VOC growth patterns, we randomly sampled 3,000 VOC sequences paired with up to 3,000 non-VOC sequences and matched by week of sample collection and location (Supplementary Methods section 1). Phylodynamic modelling (Supplementary Methods section 3) of the effective population sizes of B.1.1.7 and the previously dominant non-VOC B.1.177 lineage[11] gave an estimated growth rate difference of 0.33 per week (95% confidence interval (CI) 0.09–0.62), and further indicated that the VOC overtook the B.1.177 lineage on 10 December (Fig. 1b), close to the date at which VOC sampling frequency exceeded 50% in England (3 December). Thus, we estimate that B.1.1.7 reached 50% frequency within 2.5 to 3 months after its emergence in England.

We estimated the ratio of VOC to non-VOC reproduction numbers using a renewal equation based approach (Fig. 1c, Extended Data Fig. 2, Supplementary Methods section 4). This estimator depends on the absolute growth rate of the non-VOC, estimated using the phylodynamic model. We estimate the ratio of reproduction numbers between 25 October 2020 and 16 January 2021 to be 1.89 (95% CrI 1.43–2.65), assuming a gamma-distributed generation time with mean 6.4 days and coefficient of variation of 0.66[12]. This ratio is sensitive to the assumption that the generation time distribution is identical between variants. However, even if the VOC generation time is half that of previous variants, the estimated ratio of reproduction numbers was still 1.53 (95% CrI 1.27–1.79). The ratio trended downwards over time, coinciding with the increasing frequency of the VOC. By mid-January, the ratio had fallen from 1.89 to 1.54 (95% CrI 1.34–1.82) (Extended Data Fig. 2).

Trends in SGTF attributed to the VOC

Infection with the VOC lineage results in a diagnostic failure on the *S* gene target in an otherwise positive PCR test using the ThermoFisher TaqPath assay, which is widely used for SARS-CoV-2 community PCR testing in the UK. Consequently, we gained a more detailed picture of the spatial and demographic spread of B.1.1.7 by using the much more abundant diagnostic data with SGTF than by using whole-genome sequencing only. Several SARS-CoV-2 variants can result in SGTF, but since mid-November 2020, more than 97% of PCR tests with SGTF were due to the B.1.1.7 lineage[1]. Approximately 35% of positive test results in UK community PCR testing use the TaqPath assay, and so provide *S* gene target results. Before mid-November 2020, SGTF frequency among PCR positives was a poor proxy for VOC frequency. We therefore developed a spatiotemporal model to predict the proportion of SGTF cases attributable to the VOC by area and week (Supplementary Methods section 5), here termed the true positive proportion (TPP). False positives were attributed to the *S*-gene-positive case (*S*+) category. We found that the effective population size for B.1.1.7 effective population size was highly correlated with TPP-adjusted *S* counts (Extended Data Fig. 3).

Figure 2a–c (and Supplementary Data 1, Extended Data Fig. 4) shows the spatiotemporal trends of SGTF cases (*S*−), *S*+ and total PCR-positive cases by National Health Service (NHS) England Sustainability and Transformation Plan (STP) areas (a geographical subdivision of NHS regions). Visually, it is clear that during the second England lockdown, when schools were open, *S*+ case numbers decreased but *S*− case numbers increased. However, during the third lockdown, when schools were closed, the incidence of both *S*− and *S*+ cases declined.

Using TPP-corrected SGTF frequencies applied to overall PHE case numbers, we jointly estimated weekly effective reproduction numbers (R_t) values for the VOC and non-VOC in each of the 42 STP areas using a semi-mechanistic epidemiological model[13] (Supplementary Methods section 6). The model parametrizes VOC R_t as a multiple of non-VOC R_t. The model was fitted to case numbers obtained by multiplying overall PHE case numbers by TPP-corrected SGTF frequencies. We estimated R_t for epidemiological weeks 45–55 (1 November 2020 to 16 January 2021) (Fig. 2d), as before November there were insufficient VOC cases to reliably estimate VOC reproduction numbers across England. VOC R_t was greater than non-VOC R_t for all STP–week pairs (points above the diagonal in Fig. 2e). The estimated mean ratio of R_t for the VOC and non-VOC strains was 1.79 (95% CI 1.22–2.49) over weeks 45–55. As in the phylodynamic analysis, the multiplicative advantage in R_t for the VOC declined over the time window examined, to approximately 1.5 in week 55 (Fig. 2d).

The greater R_t estimates of the VOC, even where R_t of non-VOC variants was below 1, indicates that B.1.1.7 has a transmission advantage, and that the observed frequency trends cannot be explained solely by a reduction in the mean generation time. We repeated the joint estimation of VOC and non-VOC R_t with the assumption of a 25% reduction in the mean generation time of the VOC (Extended Data Fig. 5), and this estimated the mean ratio of R_t to be 1.60 (95% CI 1.09–2.23) over weeks 45–55. Incorporating a shorter generation time for the VOC into the model reduced, but did not eliminate, the decreasing trend in transmission advantage over time.

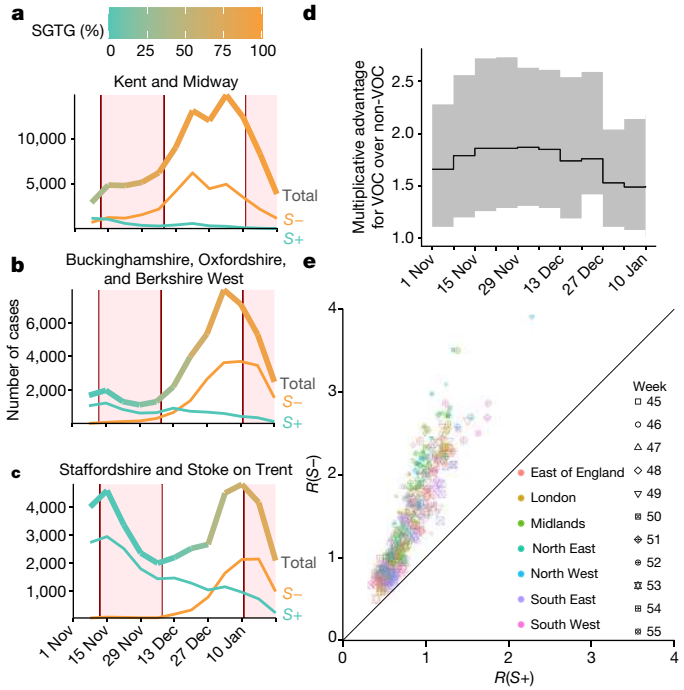


Fig. 2 | Trends of diagnosed cases and SGTF over time and between regions, and reproduction numbers of the VOC inferred from SGTF. a–**c**, The number of diagnosed cases over time for three English STP regions that represent a wide spectrum of outcomes in terms of time of VOC introduction into the region. Each line segment is shaded with the frequency of SGTF in each week (scale at top). Vertical shaded regions represent the times of the second and third UK lockdowns. **d**, The estimated (Bayesian posterior) multiplicative transmission advantage of the VOC over time inferred from STP-level SGTF count data. Shaded regions, 95% CrI. **e**, The reproduction number of *S*-gene-negative cases versus the reproduction number of *S*-gene-positive cases over time and among STP regions for epidemiological weeks 45–55 (1 November 2020 to 16 January 2021).

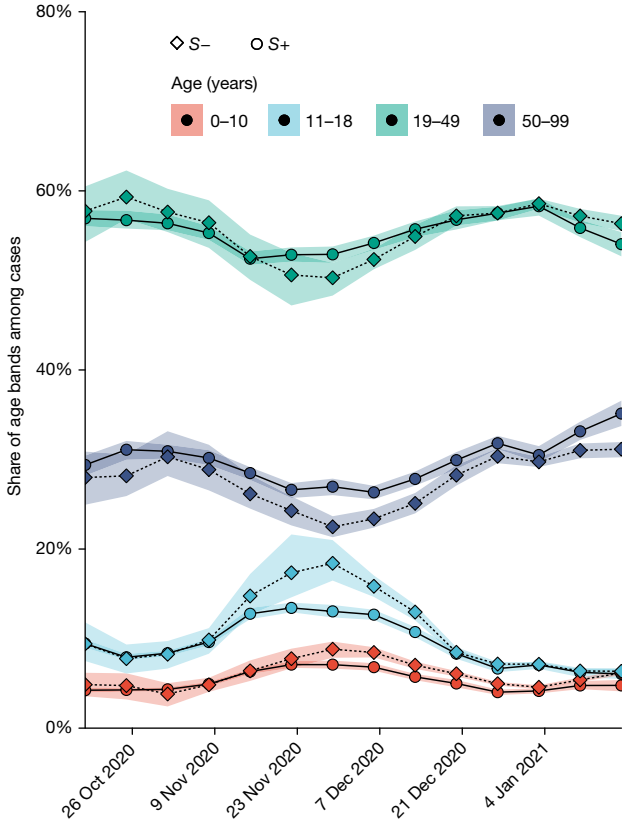


Fig. 3 | Age distribution of *S*-gene-positive and -negative cases over time in England. Observed cases were age-standardized at the level of the STP area, and age distributions were calculated for each week in STP areas and then aggregated. Shaded regions, CIs computed by bootstrapping over STP areas within NHS regions for each week.

To test whether VOC transmissibility differed by age, we first examined the age distributions of *S*+ and *S*− cases. Case numbers were age-standardized at STP area level, and then case age distributions were calculated for each STP–week (Supplementary Methods section 7). Figure 3 shows that individuals aged 19–49 years were the only age group that was consistently over-represented among observed cases relative to their share in the population (40%), with little difference between VOC and non-VOC cases. Secondary school-aged children (11–18 years) were also over-represented among observed cases relative to their share in the population (9%), and the difference between VOC and non-VOC cases was statistically significant for three weeks in November (Fig. 3, Extended Data Fig. 6). This period coincides with the second England lockdown (5 November to 2 December 2020) when schools remained open, and the differing age distributions between variants could arise from altered contact patterns when children were at greater risk of infection from all variants compared to adults.

Next, we formulated models that incorporated a difference in VOC transmission between age groups (Supplementary Methods section 7). The models were fitted variously to genome-derived and/or SGTF-derived VOC frequencies, as well as total age-specific cases in each week and region, and compared using Bayesian leave-one-out cross-validation.

Model comparison consistently favoured models that allowed the transmission advantage to vary over time and between regions, using either genomic or SGTF data. However, models that incorporated an age effect were not significantly favoured (Extended Data Table 1). Indeed, the observed fluctuations in the age distribution are equally well captured by models that do not incorporate age-specific transmission advantages (Extended Data Fig. 6). We also used these model comparisons to test the hypothesis that differences in the VOC growth rates are a consequence of a reduced generation time in B.1.1.7. In principle it is possible to statistically identify such a difference, because the data cover a period during which the overall R_t has been above and below one. Models that incorporate a change in the mean generation time were sometimes favoured (Extended Data Table 1), but the estimated ratio of mean generation times was not well identified—it varied between 0.75 and 0.96, depending on the model and data being fitted to. The mean ratio of R_t between the VOC and non-VOC ranged between 1.6 and 2.01, depending on model variant. The best fit model to both SGTF and genomic data gave an estimate of 1.74 (95% CrI 1.03–2.75), which is highly consistent with the estimates obtained from the phylodynamic analysis and the direct estimation of R_t for VOC and non-VOC described above. This model also reproduces the decline in transmission advantage over time seen in our other analyses (Extended Data Fig. 7).

Discussion

While substitutions in the B.1.1.7 lineage are associated with substantial changes in viral phenotype[3–5,14], the extent to which these substitutions lead to meaningful differences in transmission between humans is unclear, and cannot be evaluated experimentally. When randomized experimental studies are not possible, observational studies provide strong evidence if consistent patterns are seen in multiple locations and at multiple times. Increasing frequency of a new lineage is consistent with a selective advantage, but changes in frequency result from founder effects, especially for genetic variants that are repeatedly introduced from overseas[11,15]. However, in contrast to previous variants that have achieved high prevalence, we see expansion of the VOC from within the UK.

We find some evidence that the multiplicative transmission advantage of B.1.1.7 (that is, ratio of reproduction numbers) declined in late December 2020 to January 2021, coincident with stricter social distancing, school closures, and the subsequent third England lockdown (Fig. 2d, Extended Data Figs. 2, 6). A number of mechanisms could generate this effect. First, a shorter generation time of the VOC would reduce the ratio of VOC to non-VOC growth rates for small values of the non-VOC growth rate. Thus as interventions reduce both reproduction numbers, their ratio would decline, even in the absence of any underlying change in transmission advantage. Some weak support for this hypothesis is provided by our age-specific model fits to SGTF data, where model comparison generally favours models that include a change in mean generation time (Extended Data Table 1). Second, social distancing changes human contact networks, reducing the number of people contacted per day, but increasing the duration and proximity of remaining (mostly household) contacts. In such circumstances, saturation of transmission probabilities can lead to a reduction in the transmission advantage of the VOC (Extended Data Fig. 8, Supplementary Methods section 7). The observation that secondary attack rates in contacts identified through routine national contact tracing were 30–40% higher for the VOC than for non-VOC cases[16] provides some support for this hypothesis, given that the large majority of contacts identified through the UK Test and Trace system are household contacts.

The data included in this study were collected as part of routine surveillance of community testing and are not representative of SARS-CoV-2 infections in England. However, previous comparisons of community case data to random household prevalence surveys have shown very strong agreement in epidemic trends[17,18]. Furthermore, estimates of the growth advantage of B.1.1.7 obtained during earlier iterations of this study[1] have largely been predictive of its subsequent spread in January, both in the UK and internationally. Independent observations of secondary attack rates inferred from UK contact tracing data have confirmed these findings[19].

The substantial transmission advantage that we and others[20,21] have estimated has increased the challenges in controlling COVID-19. The B.1.1.7 lineage was identified quickly owing to extensive genomic surveillance in the UK, but other lineages with similar concerning features[22,23] have emerged almost concurrently, and lineages with similar features may be circulating undetected. Improving global genomic surveillance will be important for the control of COVID-19 in the presence of multiple emerging lineages with enhanced transmission or potential for immune escape.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-021-03470-x.

1. Public Health England Investigation of novel SARS-COV-2 variant 202012/01: technical briefing 1 https://www.gov.uk/government/publications/investigation-of-novel-sars-cov-2-variant-variant-of-concern-20201201 (2020).
2. Rambaut, A. et al. Preliminary genomic characterisation of an emergent SARS-CoV-2 lineage in the UK defined by a novel set of spike mutations. https://virological.org/t/preliminary-genomic-characterisation-of-an-emergent-sars-cov-2-lineage-in-the-uk-defined-by-a-novel-set-of-spike-mutations/563 (2020).
3. Chan, K. K., Tan, T. J. C., Narayanan, K. K. & Procko, E. An engineered decoy receptor for SARS-CoV-2 broadly binds protein S sequence variants. *Sci. Adv.* **7**, eabf1738 (2021).
4. Starr, T. N. et al. Deep mutational scanning of SARS-CoV-2 receptor binding domain reveals constraints on folding and ACE2 binding. *Cell* **182**, 1295–1310.e20 (2020).
5. Gu, H. et al. Adaptation of SARS-CoV-2 in BALB/c mice for testing vaccine efficacy. *Science* **369**, 1603–1607 (2020).
6. Peacock, T. P., Goldhill, D. H., Zhou, J., Baillon, L. & Frise, R. The furin cleavage site of SARS-CoV-2 spike protein is a key determinant for transmission due to enhanced replication in airway cells. Preprint at https://doi.org/10.1101/2020.09.30.318311 (2020).
7. Bal, A. et al. Screening of the H69 and V70 deletions in the SARS-CoV-2 spike protein with a RT-PCR diagnosis assay reveals low prevalence in Lyon, France. Preprint at https://doi.org/10.1101/2020.11.10.20228528 (2020).
8. The COVID-19 Genomics UK (COG-UK) consortium. An integrated national scale SARS-CoV-2 genomic surveillance network. *The Lancet Microbe* **1**, E99–E100 (2020).
9. Pybus, O. G. & Rambaut, A. Evolutionary analysis of the dynamics of viral infectious disease. *Nat. Rev. Genet.* **10**, 540–550 (2009).
10. Volz, E. M., Koelle, K. & Bedford, T. Viral phylodynamics. *PLoS Comput. Biol.* **9**, e1002947 (2013).
11. Hodcroft, E. B. et al. Emergence and spread of a SARS-CoV-2 variant through Europe in the summer of 2020. Preprint at https://doi.org/10.1101/2020.10.25.20219063 (2020).
12. Bi, Q. et al. Epidemiology and transmission of COVID-19 in 391 cases and 1286 of their close contacts in Shenzhen, China: a retrospective cohort study. *Lancet Infect. Dis.* **20**, 911–919 (2020).
13. Mishra, S. et al. A COVID-19 model for local authorities of the United Kingdom. Preprint at https://doi.org/10.1101/2020.11.24.20236661 (2020).
14. Kemp, S. et al. Recurrent emergence and transmission of a SARS-CoV-2 spike deletion ΔH69/V70. Preprint at https://doi.org/10.1101/2020.12.14.422555 (2020).
15. du Plessis, L. et al. Establishment & lineage dynamics of the SARS-CoV-2 epidemic in the UK. *Science* **371**, 708–712 (2021).
16. Public Health England. Investigation of novel SARS-COV-2 variant: Variant of Concern 202012/01: technical briefing 3. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/959360/Variant_of_Concern_VOC_202012_01_Technical_Briefing_3.pdf (2021).
17. REACT Study Investigators et al. Resurgence of SARS-CoV-2 in England: detection by community antigen surveillance. Preprint at https://doi.org/10.1101/2020.09.11.20192492 (2020).
18. Riley, S. et al. REACT-1 round 9 interim report: downward trend of SARS-CoV-2 in England in February 2021 but still at high prevalence. Preprint at https://doi.org/10.1101/2021.02.18.21251973 (2021).
19. Public Health England. Investigation of novel SARS-COV-2 variant: Variant of Concern 202012/01: technical briefing 5. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/957504/Variant_of_Concern_VOC_202012_01_Technical_Briefing_5_England.pdf (2021).
20. Leung, K., Shum, M. H., Leung, G. M., Lam, T. T. & Wu, J. T. Early transmissibility assessment of the N501Y mutant strains of SARS-CoV-2 in the United Kingdom, October to November 2020. *Euro Surveill.* **26**, 2002106 (2021).
21. Davies, N. G. et al. Estimated transmissibility and impact of SARS-CoV-2 lineage B.1.1.7 in England. Preprint at https://doi.org/10.1101/2020.12.24.20248822 (2020).
22. Tegally, H. et al. Emergence of a SARS-CoV-2 variant of concern with mutations in spike glycoprotein. *Nature* https://doi.org/10.1038/s41586-021-03402-9 (2021).
23. Sabino, E. C. et al. Resurgence of COVID-19 in Manaus, Brazil, despite high seroprevalence. *Lancet* **397**, 452–455 (2021).

The COVID-19 Genomics UK (COG-UK) consortium

Cherian Koshy[8], Amy Ash[8], Emma Wise[9], Nathan Moore[9], Matilde Mori[9], Nick Cortes[9], Jessica Lynch[9], Stephen Kidd[9], Derek J. Fairley[10], Tanya Curran[10], James P. McKenna[10], Helen Adams[11], Christophe Fraser[12], Tanya Golubchik[12], David Bonsall[12], Mohammed O. Hassan-Ibrahim[13], Cassandra S. Malone[13], Benjamin J. Cogger[13], Michelle Wantoch[14], Nicola Reynolds[14], Ben Warne[15], Joshua Maksimovic[16], Karla Spellman[16], Kathryn McCluggage[16], Michaela John[16], Robert Beer[16], Safiah Afifi[16], Sian Morgan[16], Angela Marchbank[17], Anna Price[17], Christine Kitchen[17], Huw Gulliver[17], Ian Merrick[17], Joel Southgate[17], Martyn Guest[17], Robert Munn[17], Trudy Workman[17], Thomas R. Connor[17], William Fuller[17], Catherine Bresner[17], Luke B. Snell[18], Amita Patel[18], Themoula Charalampous[19], Gaia Nebbia[19], Rahul Batra[19], Jonathan Edgeworth[19], Samuel C. Robson[20], Angela H. Beckett[20], David M. Aanensen[21], Anthony P. Underwood[21], Corin A. Yeats[21], Khalil Abudahab[21], Ben E. W. Taylor[21], Mirko Menegazzo[21], Gemma Clark[22], Wendy Smith[22], Manjinder Khakh[22], Vicki M. Fleming[22], Michelle M. Lister[22], Hannah C. Howson-Wells[22], Louise Berry[22], Tim Boswell[22], Amelia Joseph[22], Iona Willingham[22], Carl Jones[22], Christopher Holmes[23], Paul Bird[23], Thomas Helmer[23], Karlie Fallon[23], Julian Tang[23], Veena Raviprakash[24], Sharon Campbell[24], Nicola Sheriff[24], Victoria Blakey[24], Lesley-Anne Williams[24], Matthew W. Loose[25], Nadine Holmes[25], Christopher Moore[25], Matthew Carlile[25], Victoria Wright[25], Fei Sang[25], Johnny Debebe[25], Francesc Coll[26], Adrian W. Signell[27], Gilberto Betancor[27], Harry D. Wilson[27], Sahar Eldirdiri[28], Anita Kenyon[28], Thomas Davis[28], Oliver G. Pybus[29], Louis du Plessis[29], Alex E. Zarebski[29], Jayna Raghwani[29], Moritz U. G. Kraemer[29], Sarah Francois[29], Stephen W. Attwood[29], Tetyana I. Vasylyeva[29], Marina Escalera Zamudio[29], Bernardo Gutierrez[29], M. Estee Torok[30], William L. Hamilton[30], Ian G. Goodfellow[31], Grant Hall[31], Aminu S. Jahun[31], Yasmin Chaudhry[31], Myra Hosmillo[31], Malte L. Pinckert[31], Iliana Georgana[31], Samuel Moses[32], Hannah Lowe[32], Luke Bedford[33], Jonathan Moore[34], Susanne Stonehouse[34], Chloe L. Fisher[35], Ali R. Awan[35], John BoYes[36], Judith Breuer[37], Kathryn Ann Harris[37], Julianne Rose Brown[37], Divya Shah[37], Laura Atkinson[37], Jack C. D. Lee[37], Nathaniel Storey[37], Flavia Flaviani[38], Adela Alcolea-Medina[39], Rebecca Williams[40], Gabrielle Vernet[40], Michael R. Chapman[41], Lisa J. Levett[42], Judith Heaney[42], Wendy Chatterton[42], Monika Pusok[42], Li Xu-McCrae[43], Darren L. Smith[44], Matthew Bashton[44], Gregory R. Young[44], Alison Holmes[45], Paul Anthony Randell[45], Alison Cox[45], Pinglawathee Madona[45], Frances Bolt[45], James Price[45], Siddharth Mookerjee[45], Manon Ragonnet-Cronin[46], Fabricia F. Nascimento[46], David Jorgensen[46], Igor Siveroni[46], Rob Johnson[46], Olivia Boyd[46], Lily Geidelberg[46], Erik M. Volz[46], Aileen Rowan[46], Graham P. Taylor[46], Katherine L. Smollett[47], Nicholas J. Loman[5], Joshua Quick[5], Claire McMurray[5], Joanne Stockton[5], Sam Nicholls[5], Will Rowe[5], Radoslaw Poplawski[5], Alan McNally[5], Rocio T. Martinez Nunez[48], Jenifer Mason[49], Trevor I. Robinson[49], Elaine O'Toole[49], Joanne Watts[49], Cassie Breen[49], Angela Cowell[49], Graciela Sluga[50], Nicholas W. Machin[51], Shazaad S. Y. Ahmad[51], Ryan P. George[51], Fenella Halstead[52], Venkat Sivaprakasam[52], Wendy Hogsden[52], Chris J. Illingworth[53], Chris Jackson[53], Emma C. Thomson[54], James G. Shepherd[54], Patawee Asamaphan[54], Marc O. Niebel[54], Kathy K. Li[54], Rajiv N. Shah[54], Natasha G. Jesudason[54], Lily Tong[54], Alice Broos[54], Daniel Mair[54], Jenna Nichols[54], Stephen N. Carmichael[54], Kyriaki Nomikou[54], Elihu Aranday-Cortes[54], Natasha Johnson[54], Igor Starinskij[54], Ana da Silva Filipe[54], David L. Robertson[54], Richard J. Orton[54], Joseph Hughes[54], Sreenu Vattipally[54], Joshua B. Singer[54], Seema Nickbakhsh[54], Antony D. Hale[55], Louissa R. Macfarlane-Smith[55], Katherine L. Harper[55], Holli Carden[55], Yusri Taha[56], Brendan A. I. Payne[56], Shirelle Burton-Fanning[56], Sheila Waugh[56], Jennifer Collins[56], Gary Eltringham[56], Steven Rushton[57], Sarah O'Brien[57], Amanda Bradley[58], Alasdair Maclean[58], Guy Mollett[58], Rachel Blacow[58], Kate E. Templeton[59], Martin P. McHugh[59], Rebecca Dewar[59], Elizabeth Wastenge[59], Samir Dervisevic[60], Rachael Stanley[60], Emma J. Meader[60], Lindsay Coupland[60], Louise Smith[61], Clive Graham[62], Edward Barton[62], Debra Padgett[62], Garren Scott[62], Emma Swindells[63], Jane Greenaway[63], Andrew Nelson[64], Clare M. McCann[64], Wen C. Yew[64], Monique Andersson[65], Timothy Peto[65], Anita Justice[65], David Eyre[65], Derrick Crook[65], Tim J. Sloan[66], Nichola Duckworth[66], Sarah Walsh[66], Anoop J. Chauhan[67], Sharon Glaysher[67], Kelly Bicknell[67], Sarah Wyllie[67], Scott Elliott[67], Allyson Lloyd[67], Robert Impey[67], Nick Levene[68], Lynn Monaghan[68], Declan T. Bradley[69], Tim Wyatt[69], Elias Allara[2], Clare Pearson[2], Husam Osman[2], Andrew Bosworth[2], Esther Robinson[2], Peter Muir[2], Ian B. Vipond[2], Richard Hopes[2], Hannah M. Pymont[2], Stephanie Hutchings[2], Martin D. Curran[70], Surendra Parmar[70], Angie Lackenby[2], Tamyo Mbisa[2], Steven Platt[2], Shahjahan Miah[2], David Bibby[2], Carmen Manso[2], Jonathan Hubb[2], Meera Chand[2,104], Gavin Dabrera[2], Mary Ramsay[2], Daniel Bradshaw[2], Alicia Thornton[2], Richard Myers[2], Ulf Schaefer[2], Natalie Groves[2], Eileen Gallagher[2], David Lee[2], David Williams[2], Nicholas Ellaby[2], Ian Harrison[2], Hassan Hartman[2], Nikos Manesis[2], Vineet Patel[2], Chloe Bishop[2], Vicki Chalker[2], Juan Ledesma[2], Katherine A. Twohig[2], Matthew T. G. Holden[71], Sharif Shaaban[71], Alec Birchley[72], Alexander Adams[72], Alisha Davies[72], Amy Gaskin[72], Amy Plimmer[72], Bree Gatica-Wilcox[72], Caoimhe McKerr[72], Catherine Moore[72], Chris Williams[72], David Heyburn[72], Elen De Lacy[72], Ember Hilvers[72], Fatima Downing[72], Giri Shankar[72], Hannah Jones[72], Hibo Asad[72], Jason Coombes[72], Joanne Watkins[72], Johnathan M. Evans[72], Laia Fina[72], Laura Gifford[72], Lauren Gilbert[72], Lee Graham[72], Malorie Perry[72], Mari Morgan[72], Matthew Bull[72], Michelle Cronin[72], Nicole Pacchiarini[72], Noel Craine[72], Rachel Jones[72], Robin Howe[72], Sally Corden[72], Sara Rey[72], Sara Kumziene-Summerhayes[72], Sarah Taylor[72], Simon Cottrell[72], Sophie Jones[72], Sue Edwards[72], Justin O'Grady[73], Andrew J. Page[73], Alison E. Mather[73], David J. Baker[73], Steven Rudder[73], Alp Aydin[73], Gemma L. Kay[73], Alexander J. Trotter[73], Nabil-Fareed Alikhan[73], Leonardo de Oliveira Martins[73], Thanh Le-Viet[73], Lizzie Meadows[73], Anna Casey[74], Liz Ratcliffe[74], David A. Simpson[75], Zoltan Molnar[75], Thomas Thompson[75], Erwan Acheson[75], Jane A. H. Masoli[76], Bridget A. Knight[76], Sian Ellard[76], Cressida Auckland[76], Christopher R. Jones[76], Tabitha W. Mahungu[77], Dianne Irish-Tavares[77], Tanzina Haque[77], Jennifer Hart[77], Eric Witele[77], Melisa Louise Fenton[78], Ashok Dadrah[78], Amanda Symmonds[78], Tranprit Saluja[78], Yann Bourgeois[79], Garry P. Scarlett[79], Katie F. Loveson[80], Salman Goudarzi[80], Christopher Fearn[80], Kate Cook[80], Hannah Dent[80], Hannah Paul[80], David G. Partridge[81], Mohammad Raza[81], Cariad Evans[81], Kate Johnson[81], Steven Liggett[82], Paul Baker[82], Stephen Bonner[82], Sarah Essex[82], Ronan A. Lyons[83], Kordo Saeed[84], Adhyana I. K. Mahanama[84], Buddhini Samaraweera[84], Siona Silveira[84], Emanuela Pelosi[84], Eleri Wilson-Davies[84], Rachel J. Williams[85], Mark Kristiansen[85], Sunando Roy[85], Charlotte A. Williams[85], Marius Cotic[85], Nadua Bayzid[85], Adam P. Westhorpe[85], John A. Hartley[85], Riaz Jannoo[85], Helen L. Lowe[85], Angeliki Karamani[85], Leah Ensell[85], Jacqui A. Prieto[84], Sarah Jeremiah[84], Dimitris Grammatopoulos[86], Sarojini Pandey[86], Lisa Berry[86], Katie Jones[86], Alex Richter[87], Andrew Beggs[87], Angus Best[88], Benita Percival[88], Jeremy Mirza[88], Oliver Megram[88], Megan Mayhew[88], Liam Crawford[88], Fiona Ashcroft[88], Emma Moles-Garcia[88], Nicola Cumley[88], Colin P. Smith[89], Giselda Bucca[89], Andrew R. Hesketh[89], Beth Blane[90], Sophia T. Girgis[90], Danielle Leek[90], Sushmita Sridhar[90], Sally Forrest[90], Claire Cormie[90], Harmeet K. Gill[90], Joana Dias[90], Ellen E. Higginson[90], Mailis Maes[90], Jamie Young[90], Leanne M. Kermack[90], Ravi Kumar Gupta[90], Catherine Ludden[90], Sharon J. Peacock[90], Sophie Palmer[90], Carol M. Churcher[90], Nazreen F. Hadjirin[90], Alessandro M. Carabelli[90], Ellena Brooks[90], Kim S. Smith[90], Katerina Galai[90], Georgina M. McManus[90], Chris Ruis[90], Rose K. Davidson[91], Andrew Rambaut[92], Thomas Williams[92], Carlos E. Balcazar[92], Michael D. Gallagher[92], Áine O'Toole[92], Stefan Rooke[92], Verity Hill[92], Kathleen A. Williamson[92], Thomas D. Stanton[92], Stephen L. Michell[93], Claire M. Bewshea[93], Ben Temperton[93], Michelle L. Michelsen[93], Joanna Warwick-Dugdale[93], Robin Manley[93], Audrey Farbos[93], James W. Harrison[93], Christine M. Sambles[93], David J. Studholme[93], Aaron R. Jeffries[93], Alistair C. Darby[94], Julian A. Hiscox[94], Steve Paterson[94], Miren Iturriza-Gomara[94], Kathryn A. Jackson[94], Anita O. Lucaci[94], Edith E. Vamos[94], Margaret Hughes[94], Lucille Rainbow[94], Richard Eccles[94], Charlotte Nelson[94], Mark Whitehead[94], Lance Turtle[94], Sam T. Haldenby[94], Richard Gregory[94], Matthew Gemmell[94], Claudia Wierzbicki[94], Hermione J. Webster[94], Thushan I. de Silva[95], Nikki Smith[95], Adrienn Angyal[95], Benjamin B. Lindsey[95], Danielle C. Groves[95], Luke R. Green[95], Dennis Wang[95], Timothy M. Freeman[95], Matthew D. Parker[95], Alexander J. Keeley[95], Paul J. Parsons[95], Rachel M. Tucker[95], Rebecca Brown[95], Matthew Wyles[95], Max Whiteley[95], Peijun Zhang[95], Marta Gallis[95], Stavroula F. Louka[95], Chrystala Constantinidou[96], Meera Unnikrishnan[96], Sascha Ott[96], Jeffrey K. J. Cheng[96], Hannah E. Bridgewater[96], Lucy R. Frost[96], Grace Taylor-Joyce[96], Richard Stark[96], Laura Baxter[96], Mohammad T. Alam[96], Paul E. Brown[96], Dinesh Aggarwal[97], Alberto C. Cerda[98,99,100], Tammy V. Merrill[98,99,100], Rebekah E. Wilson[98,99,100], Patrick C. McClure[101], Joseph G. Chappell[101], Theocharis Tsoleridis[101], Jonathan Ball[101], David Buck[102], John A. Todd[102], Angie Green[102], Amy Trebes[102], George MacIntyre-Cockett[102], Mariateresa de Cesare[102], Alex Alderton[3], Roberto Amato[3], Cristina V. Ariani[3], Mathew A. Beale[3], Charlotte Beaver[3], Katherine L. Bellis[3], Emma Betteridge[3], James Bonfield[3], John Danesh[3], Matthew J. Dorman[3], Eleanor Drury[3], Ben W. Farr[3], Luke Foulser[3], Sonia Goncalves[3], Scott Goodwin[3], Marina Gourtovaia[3], Ewan M. Harrison[3], David K. Jackson[3], Dorota Jamrozy[3], Ian Johnston[3], Leanne Kane[3], Sally Kay[3], Jon-Paul Keatley[3], Dominic Kwiatkowski[3], Cordelia F. Langford[3], Mara Lawniczak[3], Laura Letchford[3], Rich Livett[3], Stephanie Lo[3], Inigo Martincorena[3], Samantha McGuigan[3], Rachel Nelson[3], Steve Palmer[3], Naomi R. Park[3], Minal Patel[3], Liam Prestwood[3], Christoph Puethe[3], Michael A. Quail[3], Shavanthi Rajatileka[3], Carol Scott[3], Lesley Shirley[3], John Sillitoe[3], Michael H. Spencer Chapman[3], Scott A. J. Thurston[3], Gerry Tonkin-Hill[3], Danni Weldon[3], Diana Rajan[3], Iraad F. Bronner[3], Louise Aigrain[3], Nicholas M. Redshaw[3], Stefanie V. Lensing[3], Robert Davies[3], Andrew Whitwham[3], Jennifer Liddle[3], Kevin Lewis[3], Jaime M. Tovar-Corona[3], Steven Leonard[3], Jillian Durham[3], Andrew R. Bassett[3], Shane McCarthy[3], Robin J. Moll[3], Keith James[3], Karen Oliver[3], Alex Makunin[3], Jeff Barrett[3] & Rory N. Gunson[103]

[8]Barking, Havering and Redbridge University Hospitals NHS Trust, Barking, UK. [9]Basingstoke Hospital, Basingstoke, UK. [10]Belfast Health and Social Care Trust, Belfast, UK. [11]Betsi Cadwaladr University Health Board, Betsi Cadwaladr, UK. [12]Big Data Institute, Nuffield Department of Medicine, University of Oxford, Oxford, UK. [13]Brighton and Sussex University Hospitals NHS Trust, Brighton, UK. [14]Cambridge Stem Cell Institute, University of Cambridge, Cambridge, UK. [15]Cambridge University Hospitals NHS Foundation Trust, Cambridge, UK. [16]Cardiff and Vale University Health Board, Cardiff, UK. [17]Cardiff University, Cardiff, UK. [18]Centre for Clinical Infection and Diagnostics Research, St. Thomas' Hospital and Kings College London, London, UK. [19]Centre for Clinical Infection and Diagnostics Research, Department of Infectious Diseases, Guy's and St Thomas' NHS Foundation Trust, London, UK. [20]Centre for Enzyme Innovation, University of Portsmouth (PORT), Portsmouth, UK. [21]Centre for Genomic Pathogen Surveillance, University of Oxford, Oxford, UK. [22]Clinical Microbiology Department, Queens Medical Centre, Nottingham, UK. [23]Clinical Microbiology, University Hospitals of Leicester NHS Trust, Leicester, UK. [24]County Durham and Darlington NHS Foundation Trust, Durham, UK. [25]Deep Seq, School of Life Sciences, Queens Medical Centre, University of Nottingham, Nottingham, UK. [26]Department of Infection Biology, Faculty of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, London, UK. [27]Department of Infectious Diseases, King's College London, London, UK. [28]Department of Microbiology, Kettering General Hospital, Kettering, UK. [29]Department of Zoology, University of Oxford, Oxford, UK. [30]Departments of Infectious Diseases and Microbiology, Cambridge University Hospitals NHS Foundation Trust, Cambridge, UK. [31]Division of Virology, Department of Pathology, University of Cambridge, Cambridge, UK. [32]East Kent Hospitals University NHS Foundation Trust, Canterbury, UK. [33]East Suffolk and North Essex NHS Foundation Trust, Ipswich, UK. [34]Gateshead Health NHS Foundation Trust, Gateshead, UK. [35]Genomics Innovation Unit, Guy's and St. Thomas' NHS Foundation Trust, London, UK. [36]Gloucestershire Hospitals NHS Foundation Trust, Gloucester, UK. [37]Great Ormond Street Hospital for Children NHS Foundation Trust, London, UK. [38]Guy's and St. Thomas' BRC, London, UK. [39]Guy's and St. Thomas' Hospitals, London, UK. [40]Hampshire Hospitals NHS Foundation Trust, Winchester, UK. [41]Health Data Research UK Cambridge, Cambridge, UK. [42]Health Services Laboratories, London, UK. [43]Heartlands Hospital, Birmingham, UK. [44]Hub for Biotechnology in the Built Environment, Northumbria University, Newcastle-upon-Tyne, UK. [45]Imperial College Hospitals NHS Trust, London, UK. [46]Imperial College London, London, UK. [47]Institute of Biodiversity, Animal Health and Comparative Medicine, Glasgow, UK. [48]King's College London, London, UK. [49]Liverpool Clinical Laboratories, Liverpool, UK. [50]Maidstone and Tunbridge Wells NHS Trust, Maidstone, UK. [51]Manchester University NHS Foundation Trust, Manchester, UK. [52]Microbiology Department, Wye Valley NHS Trust, Hereford, UK. [53]MRC Biostatistics Unit,

University of Cambridge, Cambridge, UK. [54]MRC-University of Glasgow Centre for Virus Research, Glasgow, UK. [55]National Infection Service, PHE and Leeds Teaching Hospitals Trust, Leeds, UK. [56]Newcastle Hospitals NHS Foundation Trust, Newcastle, UK. [57]Newcastle University, Newcastle, UK. [58]NHS Greater Glasgow and Clyde, Glasgow, UK. [59]NHS Lothian, Edinburgh, UK. [60]Norfolk and Norwich University Hospital, Norwich, UK. [61]Norfolk County Council, Norwich, UK. [62]North Cumbria Integrated Care NHS Foundation Trust, Carlisle, UK. [63]North Tees and Hartlepool NHS Foundation Trust, Stockton-on-Tees, UK. [64]Northumbria University, Newcastle-upon-Tyne, UK. [65]Oxford University Hospitals NHS Foundation Trust, Oxford, UK. [66]PathLinks, Northern Lincolnshire & Goole NHS Foundation Trust, Scunthorpe, UK. [67]Portsmouth Hospitals University NHS Trust, Portsmouth, UK. [68]Princess Alexandra Hospital Microbiology Department, Harlow, UK. [69]Public Health Agency, Belfast, UK. [70]Public Health England, Clinical Microbiology and Public Health Laboratory, Cambridge, UK. [71]Public Health Scotland, Glasgow, UK. [72]Public Health Wales NHS Trust, Cardiff, UK. [73]Quadram Institute Bioscience, Norwich, UK. [74]Queen Elizabeth Hospital, Birmingham, UK. [75]Queen's University Belfast, Belfast, UK. [76]Royal Devon and Exeter NHS Foundation Trust, Exeter, UK. [77]Royal Free NHS Trust, London, UK. [78]Sandwell and West Birmingham NHS Trust, Birmingham, UK. [79]School of Biological Sciences, University of Portsmouth (PORT), Portsmouth, UK. [80]School of Pharmacy and Biomedical Sciences, University of Portsmouth (PORT), Portsmouth, UK. [81]Sheffield Teaching Hospitals, Sheffield, UK. [82]South Tees Hospitals NHS Foundation Trust, Newcastle, UK. [83]Swansea University, Swansea, UK. [84]University Hospitals Southampton NHS Foundation Trust, Southampton, UK. [85]University College London, London, UK. [86]University Hospitals Coventry and Warwickshire, Coventry, UK. [87]University of Birmingham, Birmingham, UK. [88]University of Birmingham Turnkey Laboratory, Birmingham, UK. [89]University of Brighton, Brighton, UK. [90]University of Cambridge, Cambridge, UK. [91]University of East Anglia, Norwich, UK. [92]University of Edinburgh, Edinburgh, UK. [93]University of Exeter, Exeter, UK. [94]University of Liverpool, Liverpool, UK. [95]University of Sheffield, Sheffield, UK. [96]University of Warwick, Warwick, UK. [97]University of Cambridge, Cambridge, UK. [98]Viapath, London, UK. [99]Guy's and St Thomas' NHS Foundation Trust, London, UK. [100]King's College Hospital NHS Foundation Trust, London, UK. [101]Virology, School of Life Sciences, Queens Medical Centre, University of Nottingham, Nottingham, UK. [102]Wellcome Centre for Human Genetics, Nuffield Department of Medicine, University of Oxford, Oxford, UK. [103]West of Scotland Specialist Virology Centre, NHS Greater Glasgow and Clyde, Glasgow, UK.

Reporting summary

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

Data availability

All data used in this study, including SGTF and genome counts aggregated by region and week and multiple sequence alignments, have been deposited in Zenodo at https://doi.org/10.5281/zenodo.4593885.

Code availability

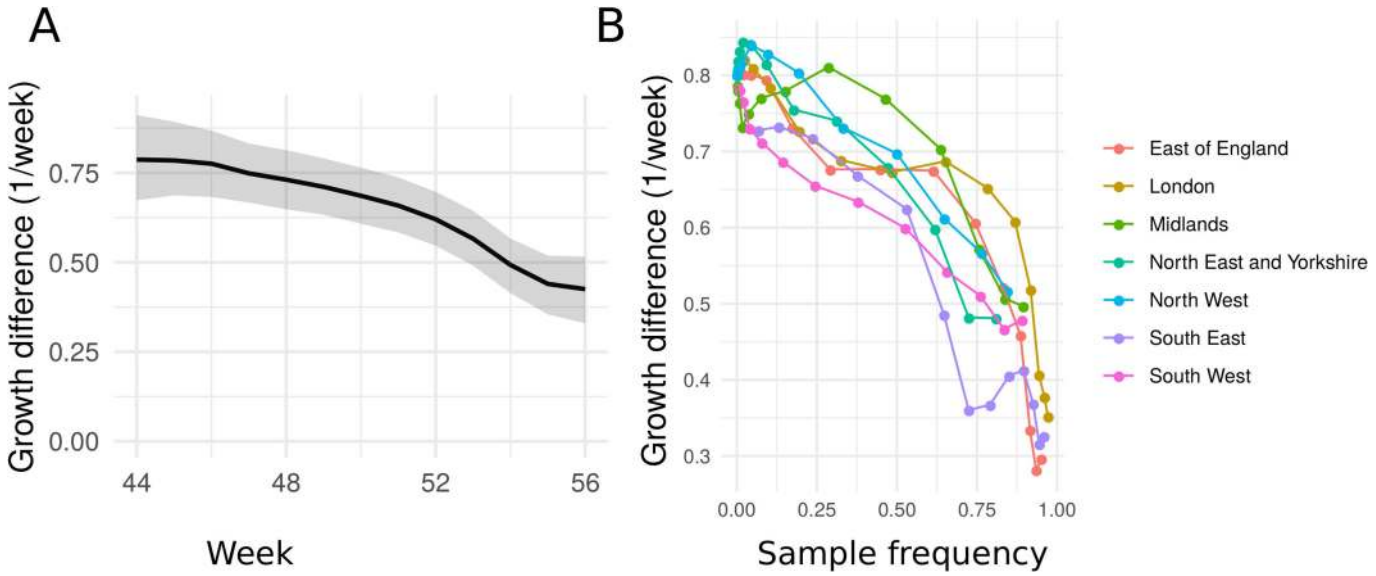
The Zenodo repository https://doi.org/10.5281/zenodo.4593885 includes code to reproduce all figures and results presented here.

Acknowledgements We thank all partners and contributors to the COG-UK consortium who are listed at https://www.cogconsortium.uk/about/. We thank the Sanger Covid Team (https://www.sanger.ac.uk/covid-team) for data preparation and feedback on the manuscript. COG-UK is supported by funding from the Medical Research Council (MRC) part of UK Research & Innovation (UKRI), the National Institute of Health Research (NIHR) and Genome Research Limited, operating as the Wellcome Sanger Institute. N.M.F., E.V., M.R. and S.B. acknowledge support from the MRC Centre for Global Infectious Disease Analysis (MR/R015600/1). R.J. and E.V. acknowledge funding from the European Commission (CoroNAb 101003653). D.J.L. and N.M.F. were supported by the NIHR VEEPED (PR-OD-1017-20002). S.B. received support from the NIHR BRC Imperial College NHS Trust Infection and COVID themes. S.B and N.M.F were supported by the UK Research and Innovation Fund MR/V038109/1. S.B. is supported by the Academy of Medical Sciences Springboard Award SBF004/1080. S.B. is supported by the Novo Nordisk Foundation Young Investigator Award NNF20OC0059309. O.R. and S.B. were supported by the Bill & Melinda Gates Foundation (OPP1175094, OPP1084362). Computational resources were provided by Amazon AWS and Microsoft AI for Health.

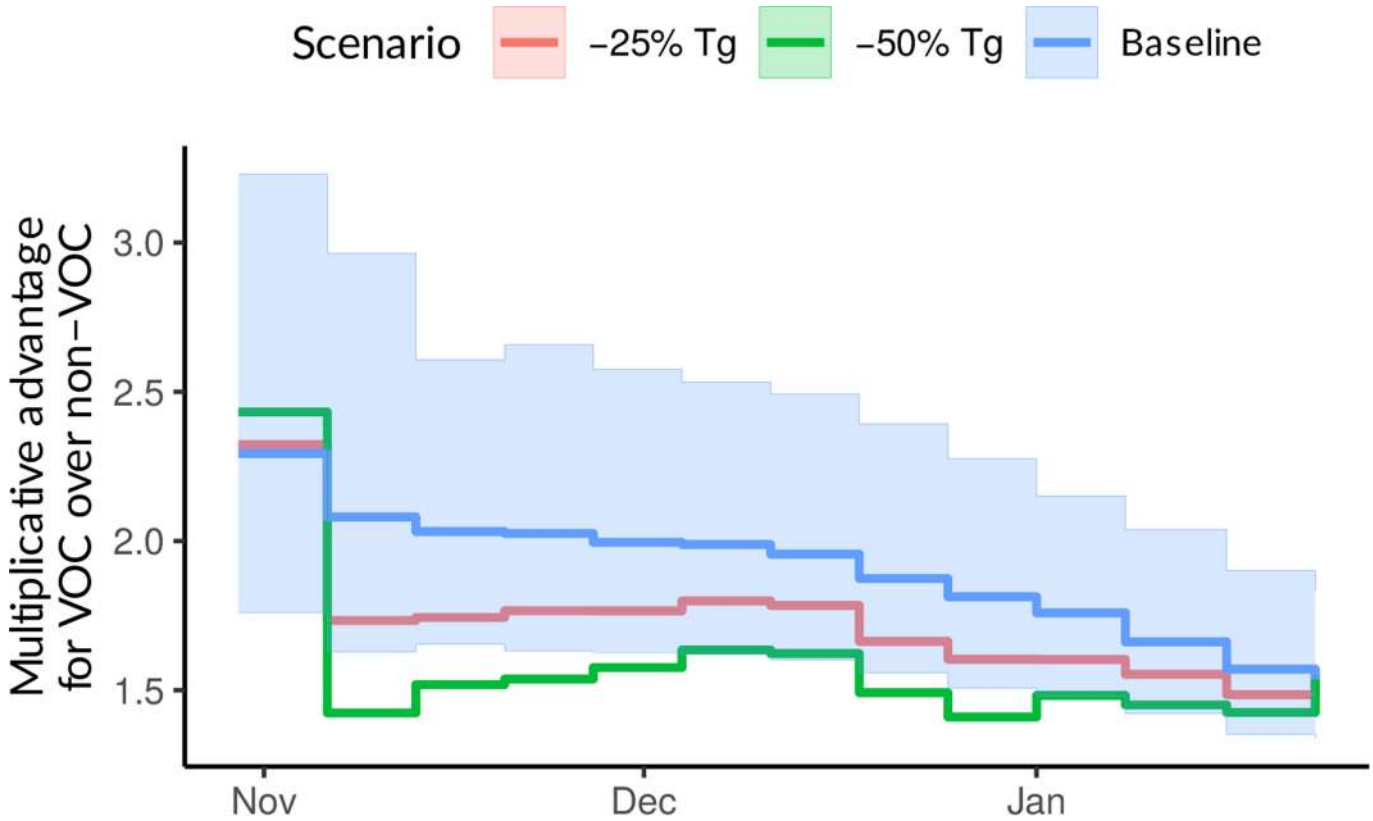
Author contributions Conceptualization: E.V., S.B., A.G., S.M., N.M.F., M.C., S.H., J.C.B., D.P.K., A.R.; writing (first draft): E.V., S.F., S.B., A.G., O.R., S.M., R.J., A.R., N.M.F.; writing (review and editing): E.V., S.F., S.B., A.G., O.R., S.M., R.J., N.M.F., M.R.-C., D.J.L., M.C., S.H., J.C.B., K.G., C.V.A., J.T.M., V.H., B.J., Á.O'T.; analysis: E.V., S.F., S.B., A.G., O.R., S.M., R.J., N.M.F., L.G., D.J., M.C., R.M., N.G., J.C.B., J.S., A.R.; methodology: E.V., S.F., S.B., A.G., O.R., S.M., R.J., N.M.F., A.R.; data generation: S.H., S.M., R.J., O.B., M.C., R.M., N.G., I.H., G.D., N.J.L., J.C.B., D.P.K., R.A., C.V.A., S.G., W.R.H.,D.K.J., J.S., O.B., K.G., J.T.M., V.H., B.J., Á.O'T., A.R.; software: E.V., S.F., A.G., S.M., R.J., N.M.F., R.A., S.G., K.G.; validation: E.V., S.B., S.M., N.M.F.; visualization: E.V., S.F., S.B., O.R., S.M., M.R.-C.; supervision: E.V., S.B., O.R., N.M.F., A.R.; funding acquisition: E.V., S.F., S.B., A.G., N.M.F., N.J.L., A.R.

Competing interests The authors declare no competing interests.

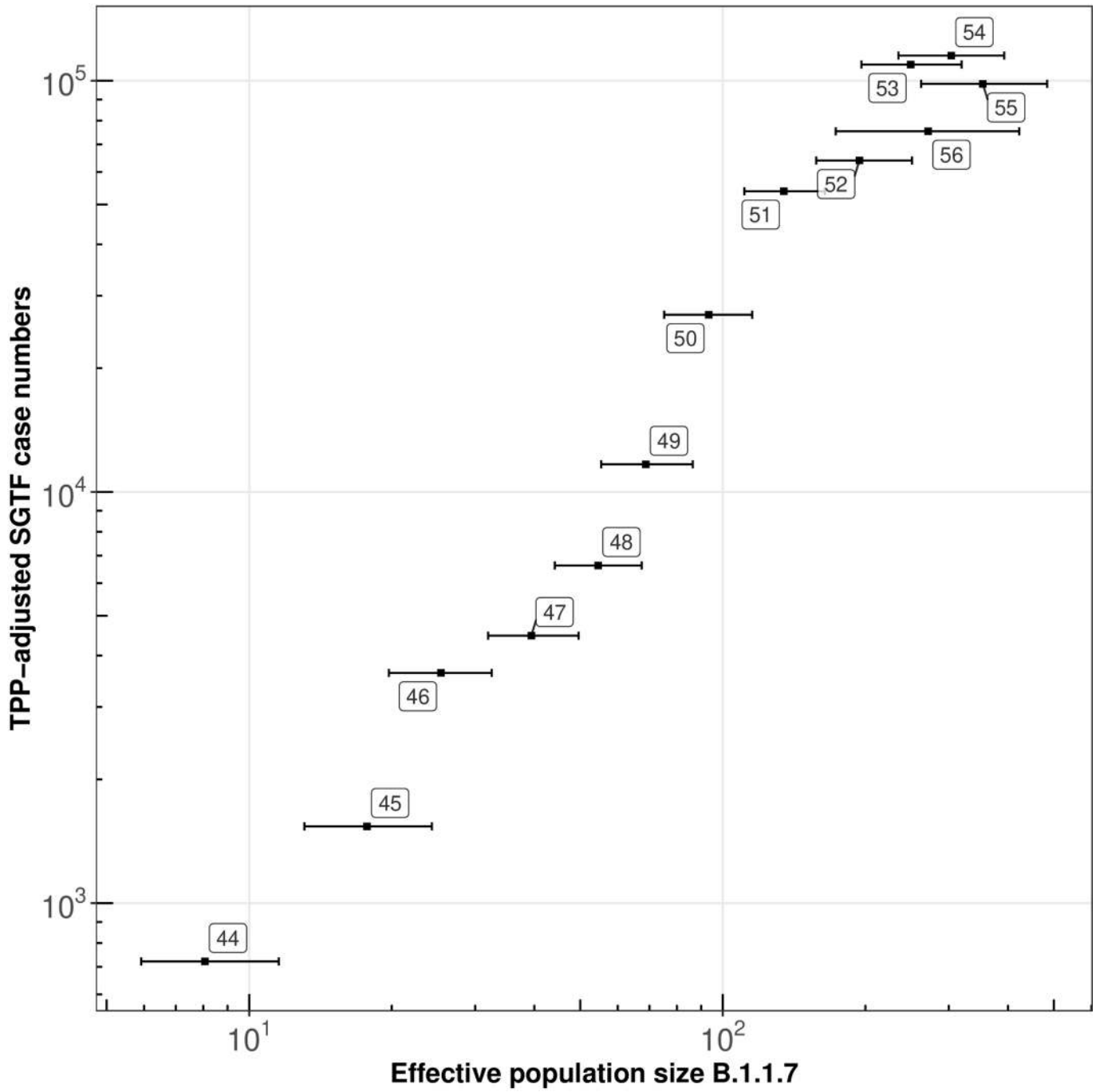
Additional information
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41586-021-03470-x.
Correspondence and requests for materials should be addressed to E.V. or N.M.F.
Peer review information *Nature* thanks Alexei Drummond and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.
Reprints and permissions information is available at http://www.nature.com/reprints.



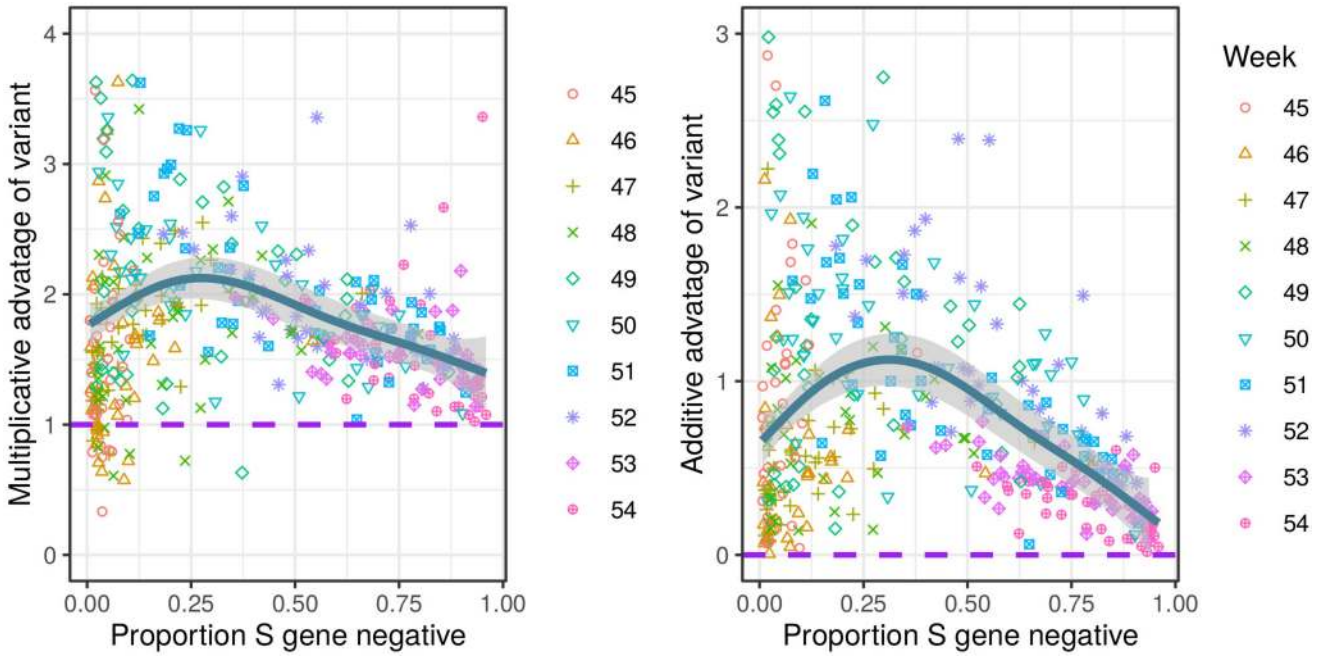
Extended Data Fig. 1 | Weekly growth rate of the VOC relative to other variants and relationship of VOC growth rate with VOC frequency. a, The additive difference in growth rate between VOC and other lineages inferred from observed frequency of VOC genomes over time (Supplementary Methods section 2). **b**, The additive difference in growth rates plotted against estimated frequency of the VOC for different NHS regions. The difference in growth rate correlates more strongly with VOC frequency than with time (data not shown). Estimates are presented for weeks 44–56 for each region.



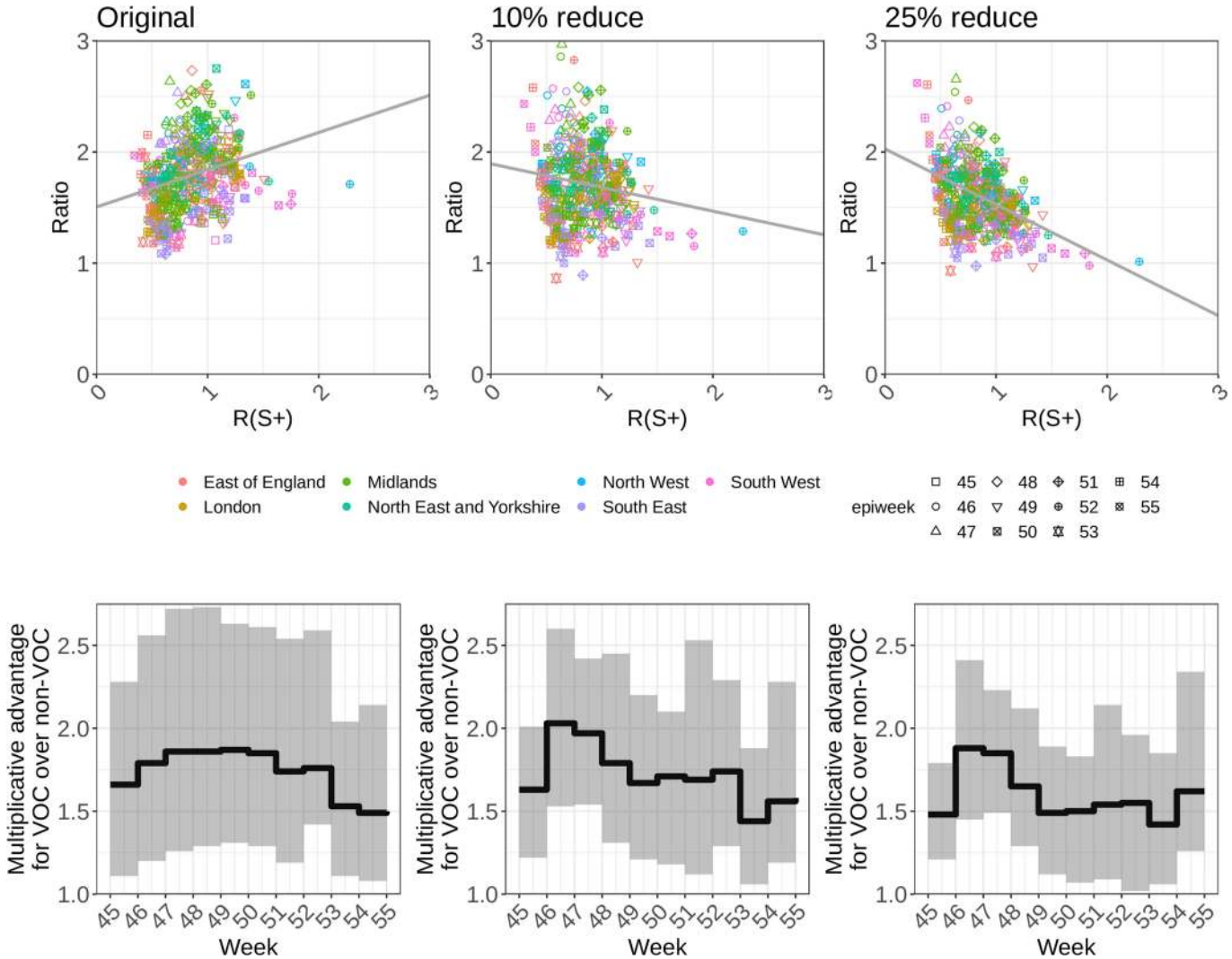
Extended Data Fig. 2 | The ratio of the reproduction number for the VOC to that for co-circulating lineages inferred from combining estimating frequencies and phylodynamic estimates of VOC and non-VOC growth rates. Shaded region, 95% CrI. Sensitivity of estimates to differences in the mean generation time is shown in red and green (25% and 50% reduced generation time in VOC, respectively).



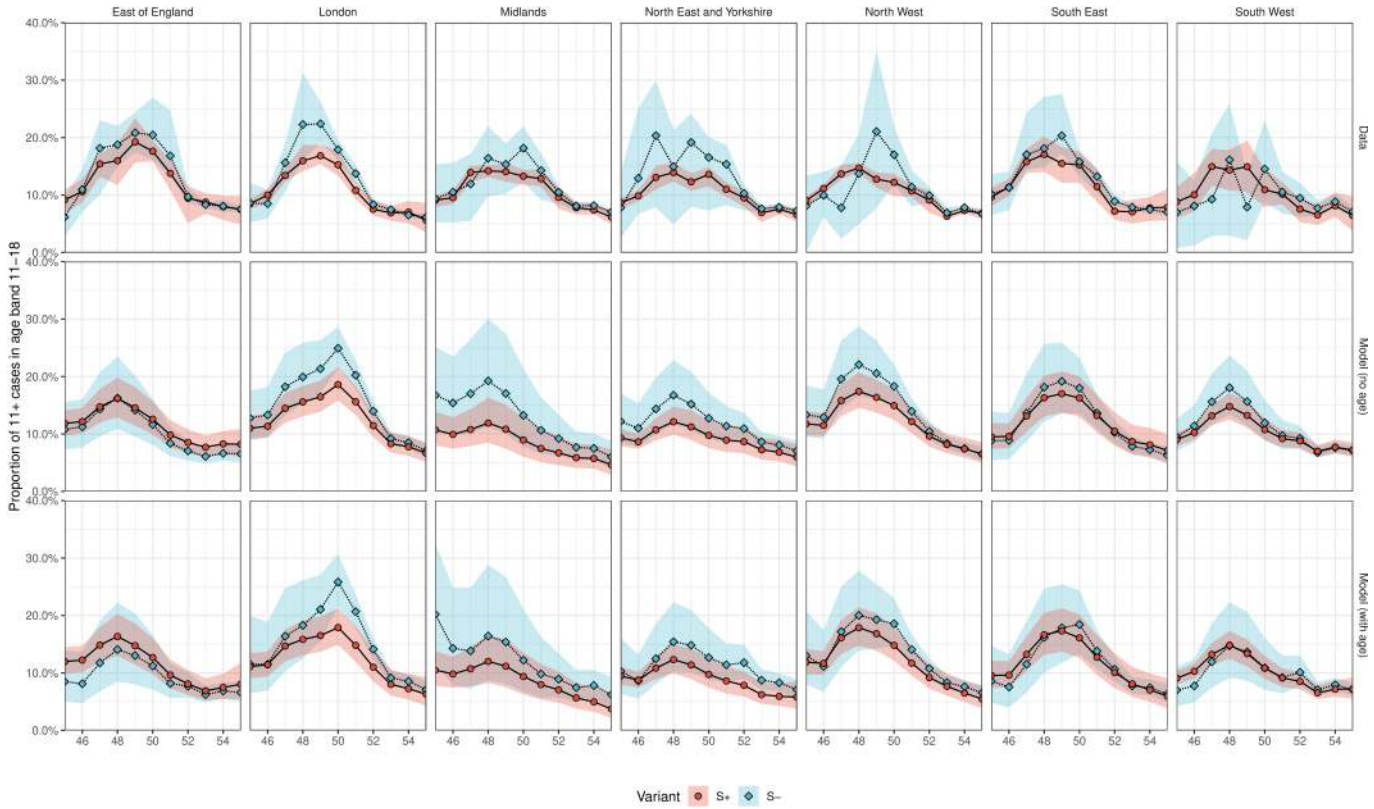
Extended Data Fig. 3 | Weekly numbers of *S*-gene-positive samples with TPP correction plotted against the effective population size of the VOC. Point labels indicate week of data collection. Effective sample size is taken on the final day of the corresponding epidemiological week. Error bars, 95% bootstrap CIs.



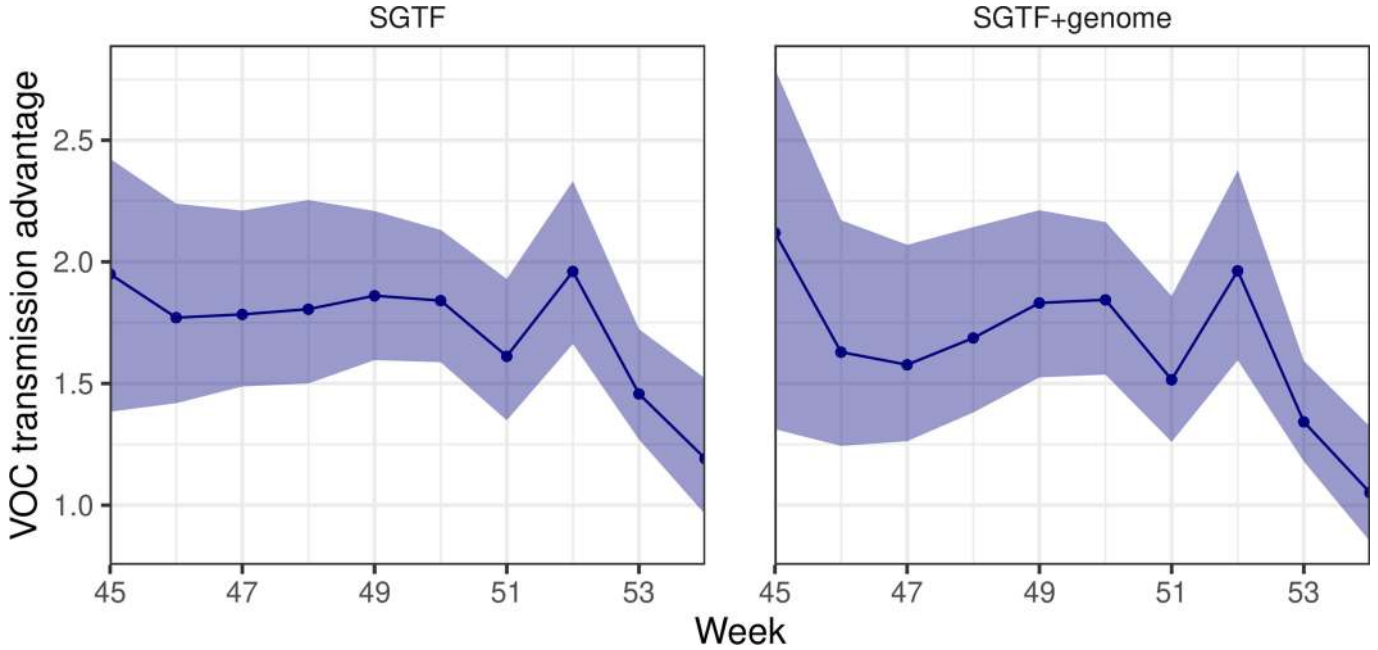
Extended Data Fig. 4 | Empirical data analysis of the advantage in weekly growth factors (cases in week *t* + 1 divided by cases in week *t*) for the VOC and non-VOC lineages (Supplementary Methods section 5). Each point represents either the ratio (left) or difference (right) of weekly growth factors for the VOC versus the non-VOC for an NHS England STP area and week, using the raw SGTF data shown in Fig. 2 (not adjusting for TPP). Colours and shapes differentiate epidemiological weeks. Numbers above 1 on the left plot and above 0 on the right plot show a transmission advantage. The blue line represents the mean advantage for a particular proportion of VOC among all cases, and the grey shading the 95% asymptotic CrI (±2σ). Scatter at low frequencies largely reflects statistical noise resulting from low counts.



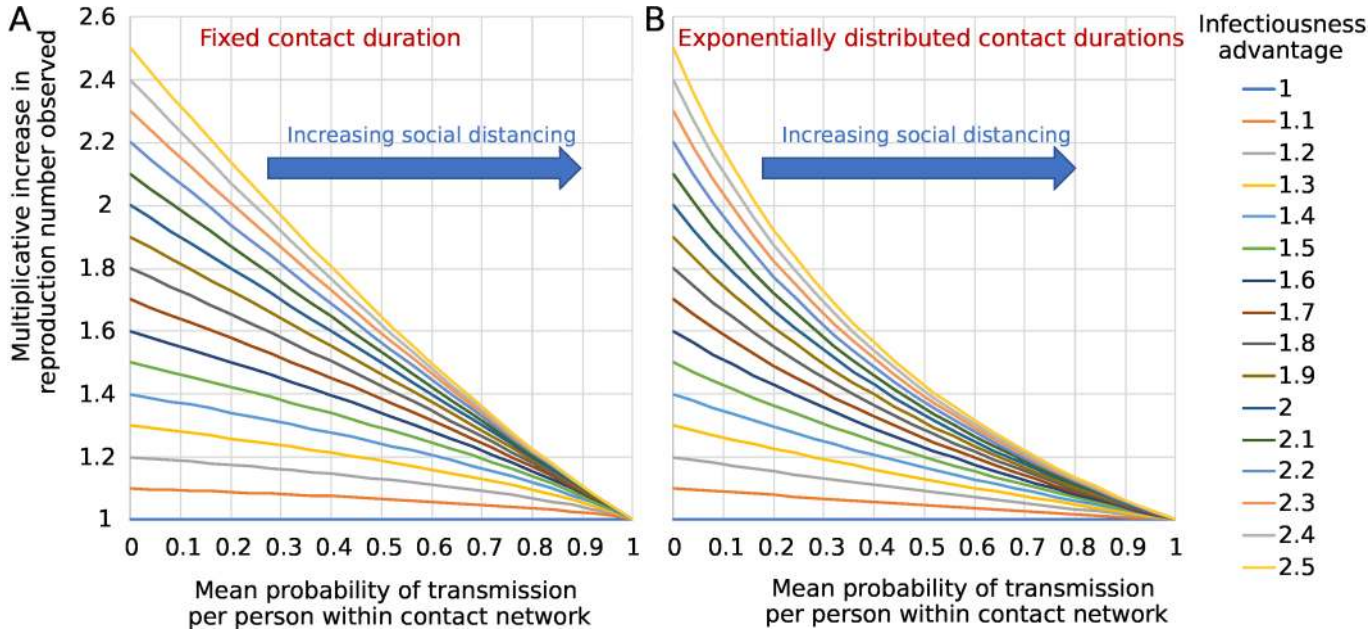
Extended Data Fig. 5 | Sensitivity of reproduction number estimates at regional level to differences in the mean generation time in the VOC (0–25% reduction in the mean, while holding coefficient of variation constant). Top, scatter plots of estimated ratio of *S*− to *S*+ reproduction numbers plotted against the reproduction number of *S*-gene-positive cases over time and among STP regions. Grey lines, linear regression—if changes in the generation time were able to completely explain temporal variation in the ratio, we would expect the slope to be zero. Bottom, ratio of reproduction number over time for each assumption about the mean generation time of the VOC. Shaded region, 95% CrI.



Extended Data Fig. 6 | Proportion of all cases in individuals over 10 years of age that were in 11–18-year-olds, stratified by SGTF status, NHS regions and week of testing. Top, data; middle, predicted trends from a semi-mechanistic model (Supplementary Methods section 7) fitting a time- and region-varying transmission advantage, but no age variation in transmission advantage; bottom, predicted trends from the model fitting a transmission that which varies by week, region and age group. Shaded region, 95% CrI.



Extended Data Fig. 7 | The overall multiplicative transmission advantage of the VOC over time estimated using a semi-mechanistic model with a VOC transmission advantage that varies by week and region, but not with age (Supplementary Methods section 7). The model fitted one transmission advantage parameter per week and per NHS region to SGTF (STP-level) counts by NHS STP area (left; finer scale than region) and SGTF counts by STP area and VOC genome counts by NHS region (right). Solid lines, medians; shading, 95% CrI pooled over all regions.



Extended Data Fig. 8 | Multiplicative increase in reproduction number seen for different intrinsic infectiousness advantages and probability of transmission per contact for the non-VOC. a, Fixed contact duration. **b**, Exponentially distributed contact durations. See Supplementary Methods section 7.

Extended Data Table 1 | Tabulation of parameter estimates and model assessment using Pareto-smoothed importance sampling leave-one-out cross validation (PSIS-LOO CV) and widely applicable information criterion (WAIC)

Covariates in reproduction number ratio	Data	Mean generation time ratio	Mean reproduction number ratio	LOO	WAIC
None	SGTF	1 (fixed)	1.92 (1.80,2.06)	-9021	-8850
None	SGTF	0.78 (0.72,0.85)	1.61 (1.48,1.74)	-9035	-8867
Age	SGTF	1 (fixed)	1.94 (1.80,2.24)	-9026	-8855
Age	SGTF	0.77 (0.72,0.84)	1.60 (1.47,1.74)	-9028	-8862
Time, region	SGTF	1 (fixed)	1.76 (1.10,2.24)	**-9013**	**-8850**
Time, region	SGTF	0.89 (0.82,0.98)	1.62 (1.12,2.04)	-9019	**-8860**
Age, time, region	SGTF	1 (fixed)	1.81 (1.19,2.30)	-9024	**-8860**
Age, time, region	SGTF	0.77 (0.70,0.85)	1.55 (1.22,1.88)	-9018	-8856
None	SGTF and genome	1 (fixed)	1.86 (1.74,2.00)	-9701	-9523
None	SGTF and genome	0.80 (0.73,0.87)	1.60 (1.48,1.75)	-9703	-9525
Age	SGTF and genome	1 (fixed)	1.85 (1.73,2.00)	-9699	-9517
Age	SGTF and genome	0.79 (0.72,0.86)	1.59 (1.46,1.74)	-9693	-9518
Time, region	SGTF and genome	1 (fixed)	1.66 (0.98,2.38)	-9654	**-9494**
Time, region	SGTF and genome	1.00 (0.90,1.11)	1.66 (0.98,2.39)	**-9650**	**-9489**
Age, time, region	SGTF and genome	1 (fixed)	1.69 (1.00,2.34)	-9676	-9514
Age, time, region	SGTF and genome	0.92 (0.82,1.02)	1.61 (1.02,2.19)	-9684	-9523
None	Genome	1	1.74 (1.61,1.89)	**-964**	**-940**
None	Genome	0.89 (0.76,1.06)	1.65 (1.50,1.84)	**-961**	**-940**
Time, region	Genome	1	1.71 (1.26,2.20)	**-929**	**-905**
Time, region	Genome	0.95 (0.82,1.11)	1.66 (1.25,2.16)	**-938**	**-915**

Bold indicates no significant difference in the estimated predictive accuracy among the best-performing models within each subset of data (SGTF counts, SGTF and genome counts, genome counts).

nature research

Corresponding author(s): Erik Volz, Neil Ferguson

Last updated by author(s): 11 March, 2021

Reporting Summary

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

Statistics

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

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

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Genomic analysesWe examined whole genome SARS-CoV-2 sequences from randomly sampled residual materials obtained from community-based COVID-19 testing in England. Sequence alignments and phylogenetic trees are available for download in the code repository. Sample weights were used to select sequences for phylodynamic modelling and to adjust counts when estimating the frequency of the variant of concern B.1.1.7 (VOC, hereafter). Weights were assigned to sequenced samples according to their local authority and their collection date. Each weight was proportional to the ratio of diagnosed cases to the number of sequence samples (reciprocal sequencing coverage rate) in each week and local authority. Case counts and sequence counts were summed over a fourteen day moving window in order to smooth over small case counts and sequence counts. Confirmed cases used for weighting were obtained using the UK government COVID-19 dashboard API (https://api.coronavirus.data.gov.uk). Code to compute sequence sample weights is available in the sequencing_coverage package (v1.0) https://git.io/Jqcve. SGTF analysesData on SGTF among national community testing was obtained from the 3 largest PCR testing laboratories and integrated into the PHE Second Generation Surveillance System (SGSS) database. We also obtained the linelist of "pillar 2" (community) PCR-positive cases from Public Health England and linked this to the SGTF data. Application of SGTF as a diagnostic for the VOC provides a large advantage over genomic sequencing in terms of cost, speed, and the sample size of available test results. We extracted 585,165 S target positive (S+) and 548,649 S target negative (S-) test results collected for weeks 43 to 56 and examined the potential to use SGTF cases (S-) as a biomarker for the VOC lineage. While the tests are not a representative sample of infections over this time period, they are a representative sample of tests within a given region and week and thus provide information about the relative abundance of the VOC versus other variants over time and between regions.
Data analysis	A large variety of open source R packages were used and described in the supporting information including version numbers. The exact code is provided in a Zenodo repository enabling reproducibility of all results and figures. Phylogenetic analysis made use of BEAST 1.10.4. Citations are provided for all packages.

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

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

All data used in this study including SGTF and genome counts aggregated by region and week and multiple sequence alignments have been deposited in Zenodo https://doi.org/10.5281/zenodo.4593885

Field-specific reporting

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

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

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

Ecological, evolutionary & environmental sciences study design

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

Study description	Genomic epidemiological analysis of the spread of a SARS-CoV-2 vaiant
Research sample	SARS CoV 2 Genomic data compiled by COG-UK consortium; community case data provided by Public Health England.
Sampling strategy	All genetic sequences and case records were utilized
Data collection	N/A
Timing and spatial scale	All of England October2020-January 2021
Data exclusions	No data were excluded
Reproducibility	The analysis is completely reproducible using code and data in the given repository
Randomization	The study involved observational data analysis that did not involve randomization. There was no original data collection. There were no experiments.
Blinding	This study involved observational data analysis and did not require blinding.

Did the study involve field work? ☐ Yes ☒ No

Reporting for specific materials, systems and methods

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

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging