

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 Geidelberg¹, Wes R. Hinsley¹, Daniel J. Laydon¹, Gavin Dabrera², Áine O'Toole⁴, Robert Amato³, Manon Ragonnet-Cronin¹, Ian Harrison², Ben Jackson⁴, Cristina V. Ariani², Olivia Boyd¹, Nicholas J. Loman^{2,5}, John T. McCrone⁴, Sónia Gonçalves³, David Jorgensen¹, Richard Myers², Verity Hill⁴, David K. Jackson³, Katy Gaythorpe¹, Natalie Groves², John Sillitoe³, Dominic P. Kwiatkowski³, The COVID-19 Genomics UK (COG-UK) consortium*, Seth Flaxman⁶, Oliver Ratmann⁶, Samir Bhatt^{1,7}, Susan Hopkins², 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 England¹, was first identified in the UK in late summer to early autumn 2020². 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 importance². The N501Y replacement on the spike protein has been shown to increase ACE2 binding^{3,4} and cell infectivity in animal models⁵, and the P618H replacement on the spike protein adjoins the furin-cleavage site⁶. B.1.1.7 also possesses a deletion at positions 69 and 70 of the spike protein (Δ 69–70) that has been associated with failure of diagnostic tests using the ThermoFisher TaqPath probe, which targets the spike protein⁷. Although other variants with Δ 69–70 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 higher 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.

¹MRC Centre for Global Infectious Disease Analysis, Jameel Institute for Disease and Emergency Analytics, Imperial College London, London, UK. ²Public Health England, London, UK.

³Wellcome Sanger Institute, Cambridge, UK. ⁴Institute of Evolutionary Biology, University of Edinburgh, Edinburgh, UK. ⁵Institute of Microbiology and Infection, University of Birmingham, Birmingham, UK. ⁶Department of Mathematics, Imperial College London, London, UK. ⁷Section of Epidemiology, Department of Public Health, University of Copenhagen, Copenhagen, Denmark. ¹⁰⁴These 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

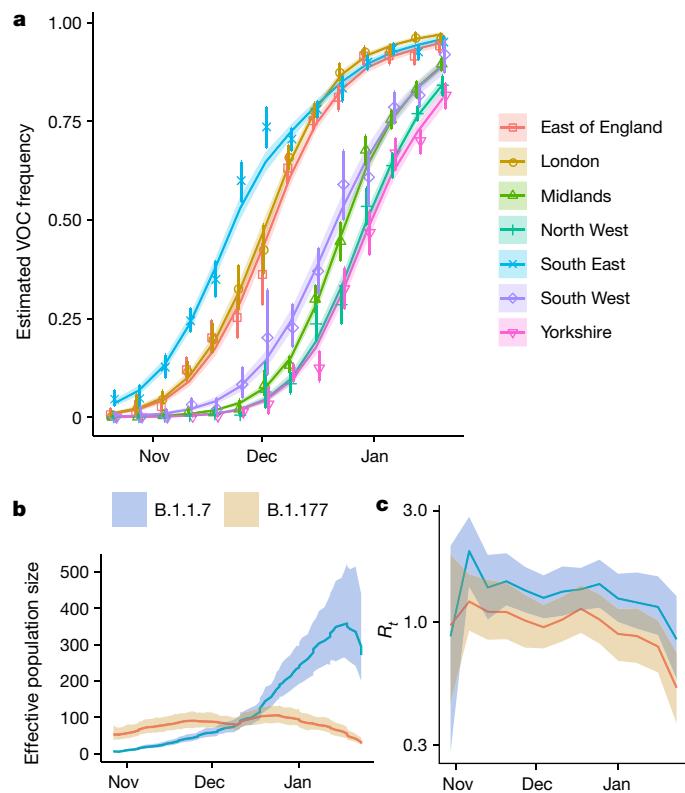


Fig. 1 | Expansion of lineage B.1.1.7 relative to co-circulating lineages in England. **a**, Estimated frequency of sampling the VOC (lines) over time in NHS regions ($n = 84,185$). Shaded regions, 95% credible region based on Bayesian regression; points, empirical proportions of the VOC in each week; 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 lineage¹¹ 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 0.66¹². 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 lineage¹. 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 model¹³ (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% CrI 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% CrI 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.

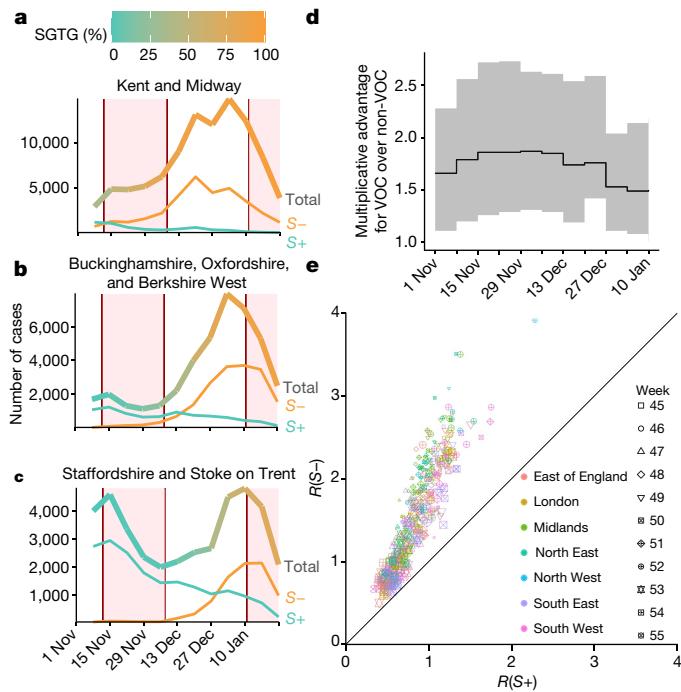


Fig. 2 | Trends of diagnosed cases and SGTG over time and between regions, and reproduction numbers of the VOC inferred from SGTG. **a–c.** The number of diagnosed cases over time for three English STP regions that represent a wide spectrum of outcomes in terms of time of VOC introduction into the region. Each line segment is shaded with the frequency of SGTG 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 SGTG 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).

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 SGTG-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 SGTG 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

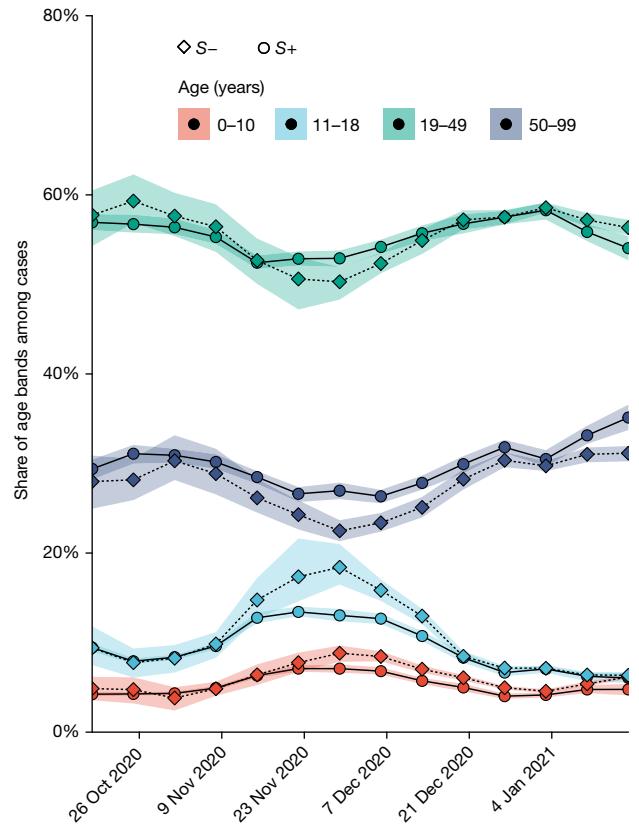


Fig. 3 | Age distribution of S-gene-positive and -negative cases over time in England. Observed cases were age-standardized at the level of the STP area, and age distributions were calculated for each week in STP areas and then aggregated. Shaded regions, CIs computed by bootstrapping over STP areas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

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 SGTG 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 phylogenetic 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.

Article

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 cases¹⁶ 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 study¹ 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 findings¹⁹.

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-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-spoke-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).

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

© The Author(s), under exclusive licence to Springer Nature Limited 2021

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

Cherian Koshy⁸, Amy Ash⁸, Emma Wise⁹, Nathan Moore⁹, Matilde Mori⁹, Nick Cortes⁹, Jessica Lynch⁹, Stephen Kidd⁹, Derek J. Fairley¹⁰, Tanya Curran¹⁰, James P. McKenna¹⁰, Helen Adams¹¹, Christophe Fraser¹², Tanya Golubchik¹², David Bonsall¹², Mohammed O. Hassan-Ibrahim¹³, Cassandra S. Malone¹³, Benjamin J. Cogger¹³, Michelle Wantoch¹⁴, Nicola Reynolds¹⁴, Ben Warne¹⁵, Joshua Maksimovic¹⁶, Karla Spellman¹⁶, Kathryn McCluggage¹⁶, Michaela John¹⁶, Robert Beer¹⁶, Safiah Afifi¹⁶, Sian Morgan¹⁶, Angela Marchbank¹⁷, Anna Price¹⁷, Christine Kitchen¹⁷, Huw Gulliver¹⁷, Ian Merrick¹⁷, Joel Southgate¹⁷, Martyn Guest¹⁷, Robert Munro¹⁷, Trudy Workman¹⁷, Thomas R. Connor¹⁷, William Fuller¹⁷, Catherine Bresner¹⁷, Luke B. Snell¹⁸, Amita Patel¹⁸, Themoula Charalampous¹⁹, Gaia Nebbia¹⁹, Rahul Batra¹⁹, Jonathan Edgeworth¹⁹, Samuel C. Robson²⁰, Angela H. Beckett²⁰, David M. Aanensen²¹, Anthony P. Underwood²¹, Corin A. Yeats²¹, Khalil Abudahab²¹, Ben W. Taylor²¹, Mirko Menegazzo²¹, Gemma Clark²², Wendy Smith²², Manjinder Khakh²², Vicki M. Fleming²², Michelle M. Lister²², Hannah C. Howson-Wells²², Louise Berry²², Tim Boswell²², Amelia Joseph²², Iona Willingham²², Carl Jones²², Christopher Holmes²³, Paul Bird²³, Thomas Helmer²³, Karlie Fallon²³, Julian Tang²³, Veena Raviprakash²⁴, Sharon Campbell²⁴, Nicola Sheriff²⁴, Victoria Blahey²⁴, Lesley-Anne Williams²⁴, Matthew W. Loose²⁵, Nadine Holmes²⁵, Christopher Moore²⁵, Matthew Carlile²⁵, Victoria Wright²⁵, Fei Sang²⁵, Johnny Debebe²⁵, Francesc Coll²⁶, Adrian W. Signell²⁷, Gilberto Betancor²⁷, Harry D. Wilson²⁷, Sahar Eldirdiri²⁸, Anita Kenyon²⁸, Thomas Davis²⁸, Oliver G. Pybus²⁹, Louis du Plessis²⁹, Alex E. Zarebski²⁹, Jayna Raghwan²⁹, Moritz U. G. Kraemer²⁹, Sarah Francois²⁹, Stephen W. Attwood²⁹, Tetyana I. Vasylyeva²⁹, Marina Escalera Zamudio²⁹, Bernardo Gutierrez²⁹, M. Estee Torok³⁰, William L. Hamilton³⁰, Ian G. Goodfellow³¹, Grant Hall³¹, Aminu S. Jahun³¹, Yasmin Chaudhry³¹, Myra Hosmillo³¹, Malte L. Pincker³¹, Iliana Georgana³¹, Samuel Moses³², Hannah Lowe³², Luke Bedford³³, Jonathan Moore³³, Susanne Stonehouse³⁴, Chloe L. Fisher³⁵, Ali R. Awan³⁵, John BoYes³⁶, Judith Breuer³⁷, Kathryn Ann Harris³⁷, Julianne Rose Brown³⁷, Divya Shah³⁷, Laura Atkinson³⁷, Jack C. D. Lee³⁷, Nathaniel Storey³⁷, Flavia Flaviani³⁸, Adela Alcolea-Medina³⁹, Rebecca Williams⁴⁰, Gabrielle Verner⁴⁰, Michael R. Chapman⁴¹, Lisa J. Levett⁴², Judith Heaney⁴², Wendy Chatterton⁴², Monika Pusok⁴², Li Xu-McCrae⁴³, Darren L. Smith⁴⁴, Matthew Bashton⁴⁴, Gregory R. Young⁴⁴, Alison Holmes⁴⁵, Paul Anthony Randell⁴⁵, Alison Cox⁴⁵, Pinglawathee Madona⁴⁵, Frances Bolt⁴⁵, James Price⁴⁵, Siddharth Mookerjee⁴⁶, Manon Ragonnet-Cronin⁴⁶, Fabricia F. Nascimento⁴⁶, David Jorgensen⁴⁶, Igor Siveroni⁴⁶, Rob Johnson⁴⁶, Olivia Boyd⁴⁶, Lily Geidelberg⁴⁶, Erik M. Volz⁴⁶, Aileen Rowan⁴⁶, Graham P. Taylor⁴⁶, Katherine L. Smollett⁴⁷, Nicholas J. Loman⁵, Joshua Quick⁵, Claire McMurray⁵, Joanne Stockton⁵, Sam Nicholls⁵, Will Rowe⁵, Radoslaw Poplawski⁵, Alan McNally⁵, Rocio T. Martinez Nunez⁵, Jenifer Mason⁴⁹, Trevor I. Robinson⁴⁹, Elaine O'Toole⁴⁹, Joanne Watts⁴⁹, Cassie Breen⁴⁹, Angela Cowell⁴⁹, Graciela Sluga⁵⁰, Nicholas W. Machin⁵¹, Shazad S. Y. Ahmad⁵¹, Ryan P. George⁵¹, Fenella Halstead⁵², Venkat Sivaprakasam⁵², Wendy Hogsden⁵², Chris J. Illingworth⁵³, Chris Jackson⁵³, Emma C. Thomson⁵⁴, James G. Shepherd⁵⁴, Patawee Asamaphan⁵⁴, Marc O. Niebel⁵⁴, Kathy K. Li⁵⁴, Rajiv N. Shah⁵⁴, Natasha G. Jesudason⁵⁴, Lily Tong⁵⁴, Alice Broos⁵⁴, Daniel Mai⁵⁴, Jenna Nichols⁵⁴, Stephen N. Carmichael⁵⁴, Kyriaki Nomikou⁵⁴, Elihu Aranday-Cortes⁵⁴, Natasha Johnson⁵⁴, Igor Starinskij⁵⁴, Ana da Silva Filipe⁵⁴, David L. Robertson⁵⁴, Richard J. Orton⁵⁴, Joseph Hughes⁵⁴, Sreenu Vattapparambil⁵⁴, Joshua B. Singer⁵⁴, Seeme Nickbaksh⁵⁴, Antony D. Hale⁵⁵, Louissa R. Macfarlane-Smith⁵⁵, Katherine L. Harper⁵⁵, Holli Carden⁵⁵, Yusri Taha⁵⁶, Brendan A. I. Payne⁵⁶, Shirelle Burton-Fanning⁵⁶, Sheila Waugh⁵⁶, Jennifer Collins⁵⁶, Gary Eltringham⁵⁶, Steven Rushton⁵⁷, Sarah O'Brien⁵⁷, Amanda Bradley⁵⁸, Alasdair Maclean⁵⁸, Guy Mollett⁵⁸, Rachel Blacock⁵⁸, Kate E. Templeton⁵⁹, Martin P. McHugh⁵⁹, Rebecca Dewar⁵⁹, Elizabeth Wastenge⁵⁹, Samir Dervisevic⁶⁰, Rachael Stanley⁶⁰, Emma J. Meader⁶⁰, Lindsay Coupland⁶⁰, Louise Smith⁶¹, Clive Graham⁶², Edward Barton⁶², Debra Padgett⁶², Garren Scott⁶², Emma Swindells⁶³, Jane Greenaway⁶³, Andrew Nelson⁶⁴, Clare M. McCann⁶⁴, Wen C. Yew⁶⁴, Monique Andersson⁶⁵, Timothy Petro⁶⁵, Anita Justice⁶⁵, David Eye⁶⁵, Derrick Crook⁶⁵, Tim J. Sloan⁶⁶, Nichola Duckworth⁶⁶, Sarah Walsh⁶⁶, Anoop J. Chauhan⁶⁷, Sharon Glashier⁶⁷, Kelly Bicknell⁶⁷, Sarah Wyllie⁶⁷, Scott Elliott⁶⁷, Allyson Lloyd⁶⁷, Robert Impey⁶⁷, Nick Levene⁶⁸, Lynn Monaghan⁶⁸, Declan T. Bradley⁶⁹, Tim Wyatt⁶⁹, Elias Allara⁷⁰, Clare Pearson⁷⁰, Husam Osman⁷⁰, Andrew Bosworth⁷⁰, Esther Robinson⁷⁰, Peter Muir⁷⁰, Ian B. Vipond⁷⁰, Richard Hopes⁷⁰, Hannah M. Pymont⁷⁰, Stephanie Hutchings⁷⁰, Martin D. Curran⁷⁰, Surendra Parmar⁷⁰, Angie Lackenby⁷², Tamyo Mbisa⁷², Steven Platt⁷², Shahjahan Miah⁷², David Bibby⁷², Carmen Manso⁷², Jonathan Hubb⁷², Meera Chand^{72,64}, Gavin Dabbera⁷², Mary Ramsay⁷², Daniel Bradshaw⁷², Alicia Thornton⁷², Richard Myers⁷², Ulf Schaefter⁷², Natalie Groves⁷², Eileen Gallagher⁷², David Lee⁷², David Williams⁷², Nicholas Ellaby⁷², Ian Harrison⁷², Hassan Hartman⁷², Nikos Manesis⁷², Vineet Patel⁷², Chloe Bishop⁷², Vicki Chalker⁷², Juan Ledesma⁷², Katherine A. Twohig⁷², Matthew T. G. Holden⁷¹, Sharif Shaaban⁷¹, Alec Birchley⁷², Alexander Adams⁷², Alisha Davies⁷², Amy Gaskin⁷², Amy Plummer⁷², Bree Gatica-Wilcox⁷², Caomhain McKerr⁷², Catherine Moore⁷², Chris Williams⁷², David Heyburn⁷², Elen De Lacy⁷², Ember Hilvers⁷², Fatima Downing⁷², Giri Shankar⁷², Hannah Jones⁷², Hibo Asad⁷², Jason Coombes⁷², Joanne Watkins⁷², Johnathan M. Evans⁷², Laia Fina⁷², Laura Gifford⁷², Lauren Gilbert⁷², Lee Graham⁷², Malorie Perry⁷², Mari Morgan⁷², Matthew Bull⁷², Michelle Cronin⁷², Nicole Pacchiarini⁷², Noel Craine⁷², Rachel Jones⁷², Robin Howe⁷², Sally Corden⁷², Sara Rey⁷², Sara Kumzienie-Summerhayes⁷², Sarah Taylor⁷², Simon Cottrell⁷², Sophie Jones⁷², Sue Edwards⁷², Justin O'Grady⁷³, Andrew J. Page⁷³, Alison E. Mather⁷³, David J. Baker⁷³, Steven Rudder⁷³, Alp Aydin⁷³, Gemma L. Kay⁷³, Alexander J. Trotter⁷³, Nabil-Fared Alikhan⁷³, Leonardo de Oliveira Martins⁷³, Thanh Le-Viet⁷³, Lizzie Meadows⁷³, Anna Casey⁷⁴, Liz Ratcliffe⁷⁴, David A. Simpson⁷⁵, Zoltan Molnar⁷⁵, Thomas Thompson⁷⁵, Erwan Acheson⁷⁵, Jane A. H. Masoli⁷⁶, Bridget A. Knight⁷⁶, Sian Ellard⁷⁶, Cressida Auckland⁷⁶, Christopher R. Jones⁷⁶, Tabitha W. Mahungu⁷⁷, Dianne Irish-Tavares⁷⁷, Tanzina Haque⁷⁷, Jennifer Hart⁷⁷, Eric Witele⁷⁷, Melisa Louise Fenton⁷⁸, Ashok Dadrah⁷⁸, Amanda Symmonds⁷⁸, Tranprit Saluja⁷⁸, Yann Bourgeois⁷⁹, Garry P. Scarlett⁷⁹, Katie F. Loveson⁸⁰, Salman Goudarzi⁸⁰, Christopher Fearn⁸⁰, Kate Cook⁸⁰, Hannah Dent⁸⁰, Hannah Paul⁸⁰, David G. Partridge⁸¹, Mohammad Raza⁸¹, Cariad Evans⁸¹, Kate Johnson⁸¹, Steven Liggett⁸², Paul Baker⁸², Stephen Bonner⁸², Sarah Essex⁸², Ronan A. Lyons⁸³, Kordo Saeed⁸⁴,

Adhyana I. K. Mahanama⁸⁴, Buddhini Samaraweera⁸⁴, Siona Silveira⁸⁴, Emanuela Pelosi⁸⁴, Eleri Wilson-Davies⁸⁴, Rachel J. Williams⁸⁵, Mark Kristiansen⁸⁵, Sunando Roy⁸⁵, Charlotte A. Williams⁸⁵, Marius Cotic⁸⁵, Nadua Bayzid⁸⁵, Adam P. Westhorpe⁸⁵, John A. Hartley⁸⁵, Riaz Janloo⁸⁵, Helen L. Lowe⁸⁵, Angeliki Karamani⁸⁵, Leah Ensell⁸⁵, Jacqui A. Prieto⁸⁴, Sarah Jeremiah⁸⁴, Dimitris Grammatopoulos⁸⁶, Sarojini Pandey⁸⁶, Lisa Berry⁸⁶, Katie Jones⁸⁶, Alex Richter⁸⁷, Andrew Beggs⁸⁷, Angus Best⁸⁸, Benita Percival⁸⁸, Jeremy Mirza⁸⁸, Oliver Megram⁸⁸, Megan Mayhew⁸⁸, Liam Crawford⁸⁸, Fiona Ashcroft⁸⁸, Emma Moles-Garcia⁸⁸, Nicola Cumley⁸⁸, Colin P. Smith⁸⁹, Giselda Bucca⁸⁹, Andrew R. Hesketh⁸⁹, Beth Blane⁹⁰, Sophia T. Girgis⁹⁰, Danielle Leek⁹⁰, Sushmita Sridhar⁹⁰, Sally Forrest⁹⁰, Claire Cormie⁹⁰, Harmeet K. Gill⁹⁰, Joana Dias⁹⁰, Ellen E. Higgins⁹⁰, Mailis Maes⁹⁰, Jamie Young⁹⁰, Leanne M. Kermack⁹⁰, Ravi Kumar Gupta⁹⁰, Catherine Ludden⁹⁰, Sharon J. Peacock⁹⁰, Sophie Palmer⁹⁰, Carol M. Churcher⁹⁰, Nazreen F. Hadjirin⁹⁰, Alessandro M. Carabelli⁹⁰, Ellena Brooks⁹⁰, Kim S. Smith⁹⁰, Katerina Galai⁹⁰, Georgina M. McManus⁹⁰, Chris Ruis⁹⁰, Rose K. Davidson⁹¹, Andrew Rambaut⁹², Thomas Williams⁹², Carlos E. Balcazar⁹², Michael D. Gallagher⁹², Aine O'Toole⁹², Stefan Rooke⁹², Verity Hill⁹², Kathleen A. Williamson⁹², Thomas D. Stanton⁹², Stephen L. Michell⁹³, Claire M. Bewshea⁹³, Ben Temperton⁹³, Michelle L. Michelsons⁹³, Joanna Warwick-Dugdale⁹³, Robin Manley⁹³, Audrey Farbos⁹³, James W. Harrison⁹³, Christine M. Sambles⁹³, David J. Studholme⁹³, Aaron R. Jeffries⁹³, Alistair C. Darby⁹⁴, Julian A. Hiscox⁹⁴, Steve Paterson⁹⁴, Miren Iturria-Gomara⁹⁴, Kathryn A. Jackson⁹⁴, Anita O. Lucaci⁹⁴, Edith E. Vamos⁹⁴, Margaret Hughes⁹⁴, Lucille Rainbow⁹⁴, Richard Eccles⁹⁴, Charlotte Nelson⁹⁴, Mark Whitehead⁹⁴, Lance Turtle⁹⁴, Sam T. Haldenby⁹⁴, Richard Gregory⁹⁴, Matthew Gemmill⁹⁴, Claudia Wierzbicki⁹⁴, Hermione J. Webster⁹⁴, Thushan I. de Silva⁹⁵, Nikki Smith⁹⁵, Adrienn Anygal⁹⁵, Benjamin B. Lindsey⁹⁵, Danielle C. Groves⁹⁵, Luke R. Green⁹⁵, Dennis Wang⁹⁵, Timothy M. Freeman⁹⁵, Matthew D. Parker⁹⁵, Alexander J. Keeley⁹⁵, Paul J. Parsons⁹⁵, Rachel M. Tucker⁹⁵, Rebecca Brown⁹⁵, Matthew Wyles⁹⁵, Max Whiteley⁹⁵, Peijun Zhang⁹⁵, Marta Gallis⁹⁵, Stavroula F. Louka⁹⁵, Chrystala Constantinidou⁹⁶, Meera Unnikrishnan⁹⁶, Sascha Ott⁹⁶, Jeffrey K. J. Cheng⁹⁶, Hannah E. Bridgewater⁹⁶, Lucy R. Frost⁹⁶, Grace Taylor-Joyce⁹⁶, Richard Stark⁹⁶, Laura Baxter⁹⁶, Mohammad T. Alam⁹⁶, Paul E. Brown⁹⁶, Dinesh Aggarwal⁹⁷, Alberto C. Cerda^{98,99,100}, Tammy L. Merrill^{98,99,100}, Rebekah E. Wilson^{98,99,100}, Patrick C. McClure¹⁰¹, Joseph G. Chappell¹⁰¹, Theocharis Tsolieridis¹⁰¹, Jonathan Ball¹⁰¹, David Buck¹⁰², John A. Todd¹⁰², Angie Green¹⁰², Amy Trebes¹⁰², George MacIntyre-Cockett¹⁰², Mariateresa de Cesare¹⁰², Alex Alderton³, Roberto Amato³, Cristina V. Ariani³, Mathew A. Beale³, Charlotte Beaver³, Katherine L. Bellis³, Emma Betteridge³, James Bonfield³, John Danesh³, Matthew J. Dorman³, Eleanor Drury³, Ben W. Farr³, Luke Foulsier³, Sonia Goncalves³, Scott Goodwin³, Marina Gourtovskaia³, Ewan M. Harrison³, David K. Jackson³, Dorota Jamrozy³, Ian Johnston³, Leanne Kane³, Sally Kay³, Jon-Paul Keatley³, Dominic Kwiatkowski³, Cordelia F. Langford³, Mara Lawniczak³, Laura Letchford³, Rich Livett³, Stephanie Lo³, Inigo Martincorena³, Samantha McGuigan³, Rachel Nelson³, Steve Palmer³, Naomi R. Park³, Minal Patel³, Liam Prestwood³, Christoph Puettel³, Michael A. Quail³, Shavanthi Rajatileka³, Carol Scott³, Lesley Shirley³, John Sillitoe³, Michael H. Spencer Chapman³, Scott A. J. Thurston³, Gerry Tonkin-Hill³, Danni Weldon³, Diana Rajan³, Iraad F. Bronner³, Louise Aigrain³, Nicholas M. Redshaw³, Stefanie V. Lensing³, Robert Davies³, Andrew Whitwham³, Jennifer Liddle³, Kevin Lewis³, Jaime M. Tovar-Corona³, Steven Leonard³, Jillian Durham³, Andrew R. Bassett³, Shane McCarthy³, Robin J. Moll³, Keith James³, Karen Oliver³, Alex Makunin³, Jeff Barrett³ & Ron N. Gunson¹⁰³

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

Article

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

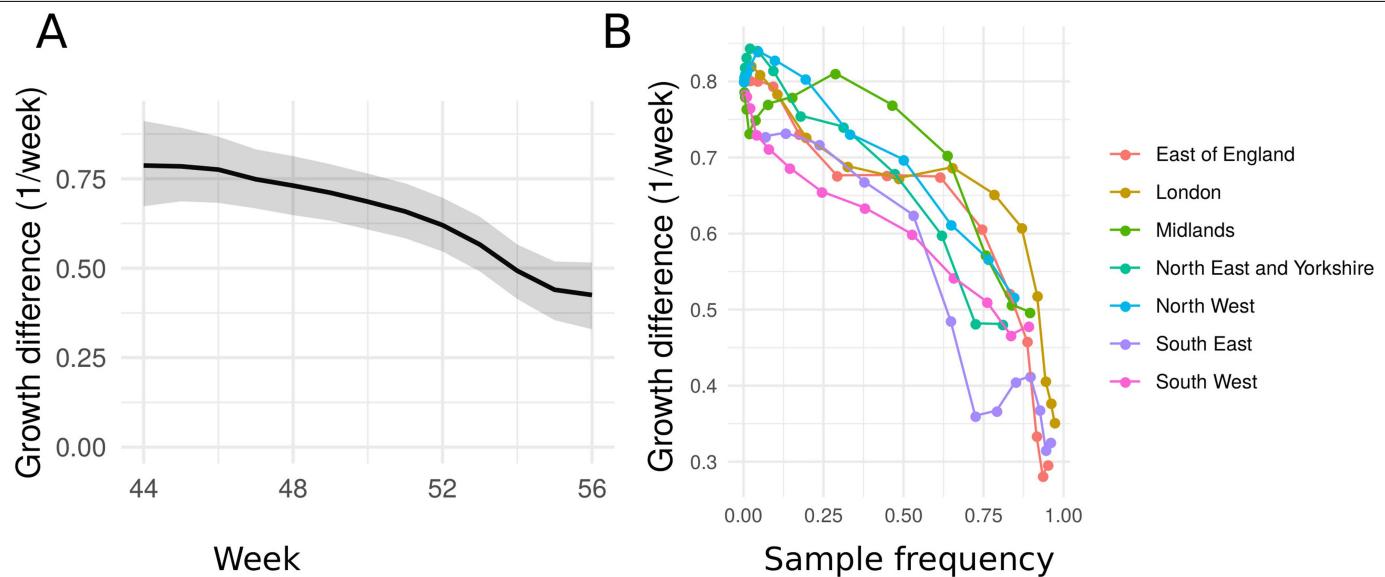
Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41586-021-03470-x>.

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.

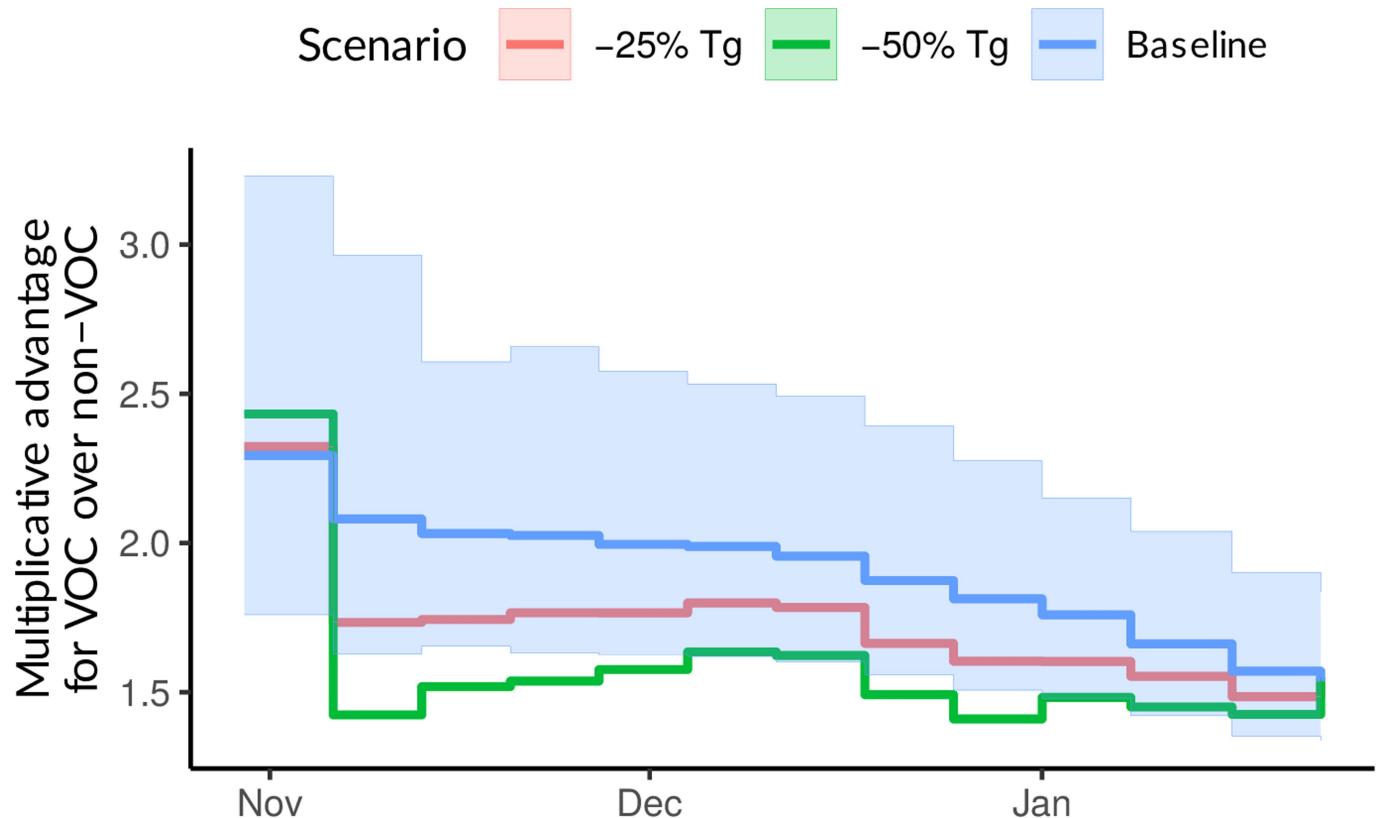
Reprints and permissions information is available at <http://www.nature.com/reprints>.

Article



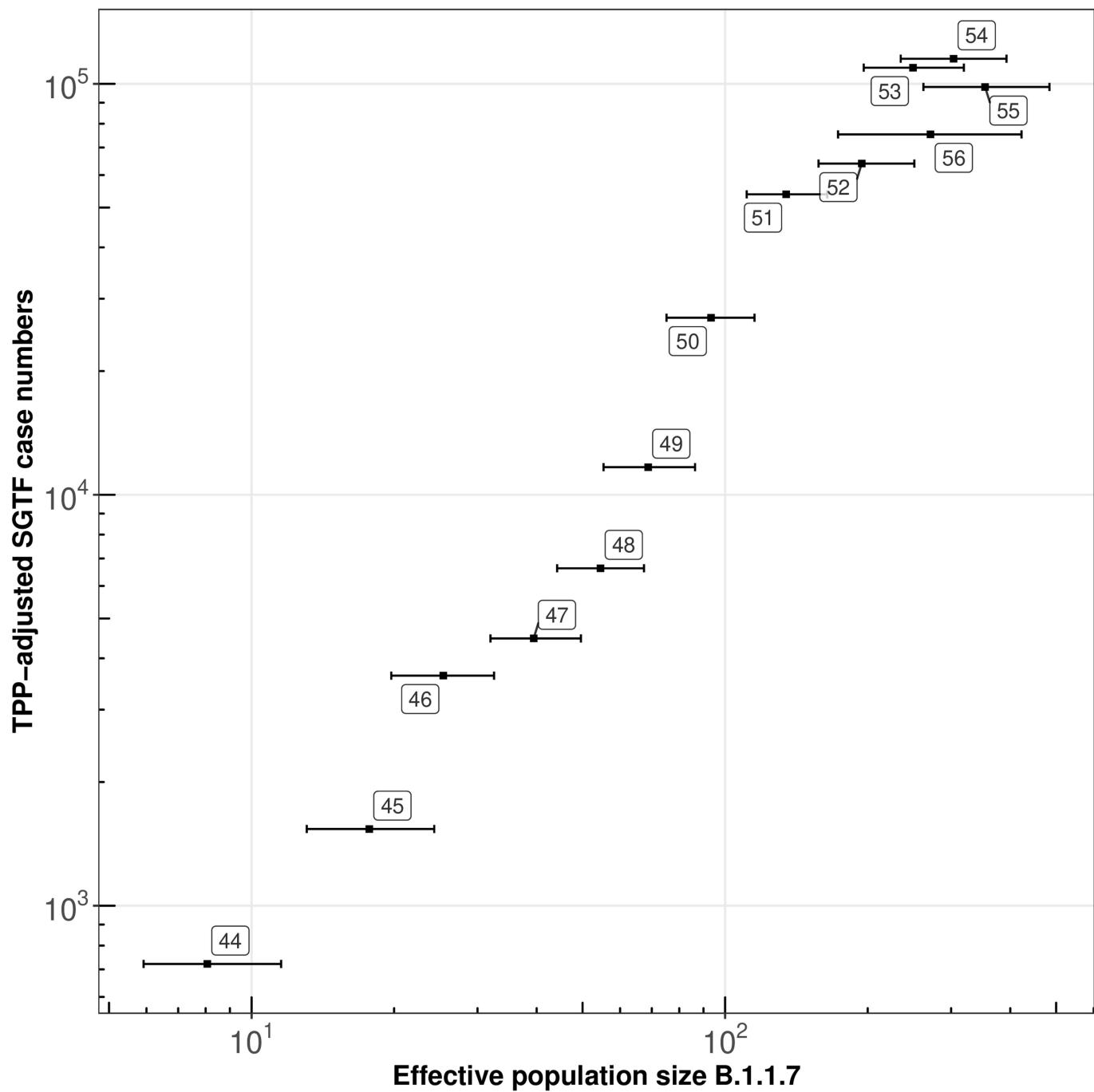
Extended Data Fig. 1 | Weekly growth rate of the VOC relative to other variants and relationship of VOC growth rate with VOC frequency. **a**, The additive difference in growth rate between VOC and other lineages inferred from observed frequency of VOC genomes over time (Supplementary Methods

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.



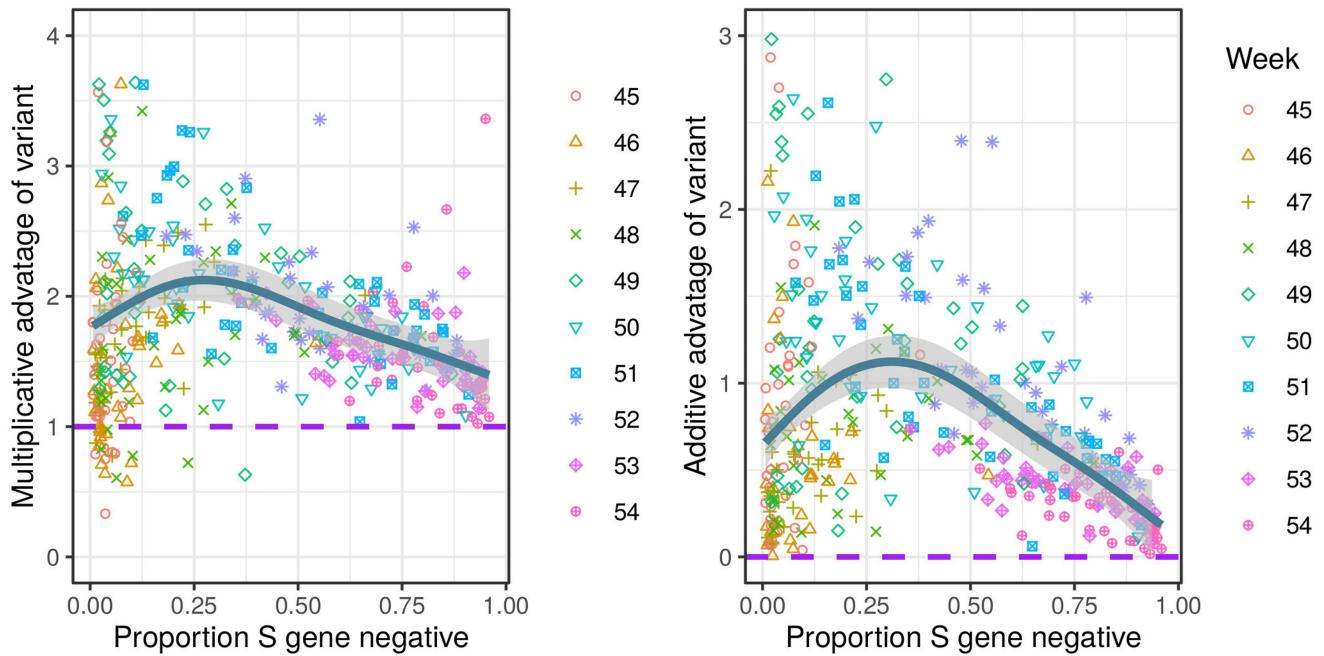
Extended Data Fig. 2 | The ratio of the reproduction number for the VOC to that for co-circulating lineages inferred from combining estimating frequencies and phylodynamic estimates of VOC and non-VOC growth

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

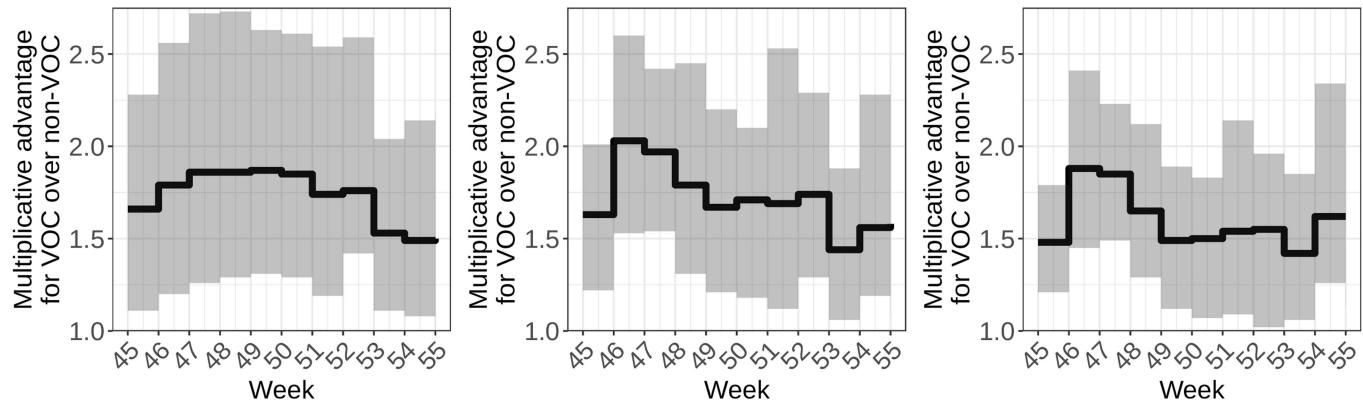
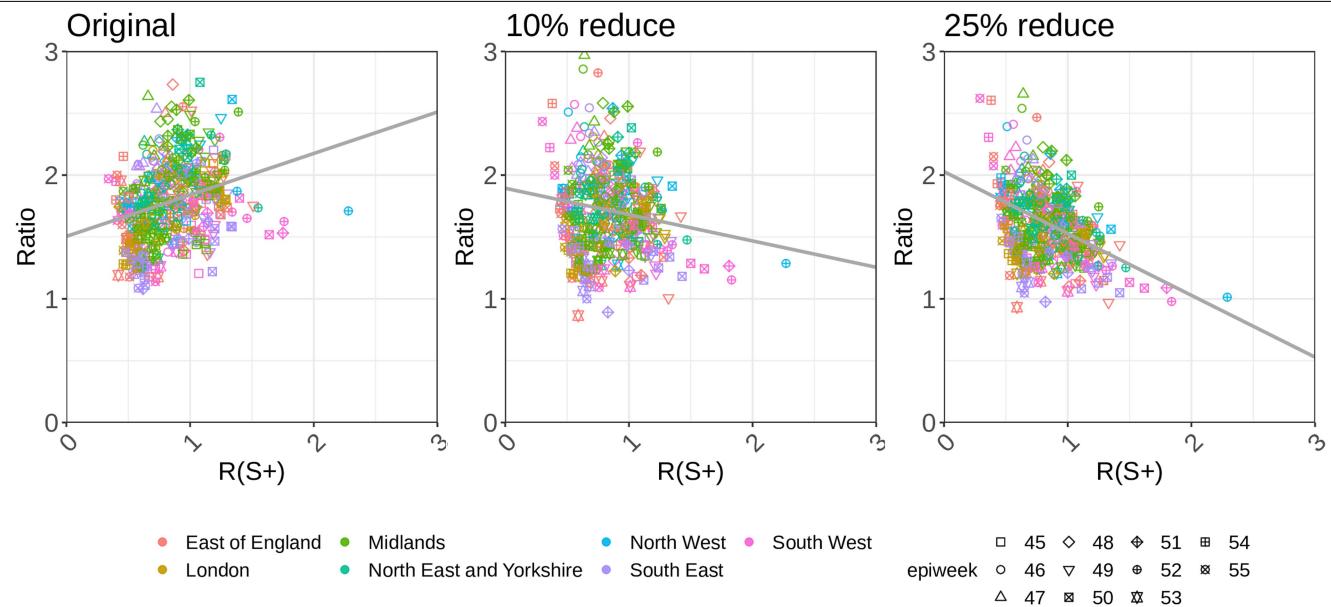
the final day of the corresponding epidemiological week. Error bars, 95% bootstrap CIs.



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

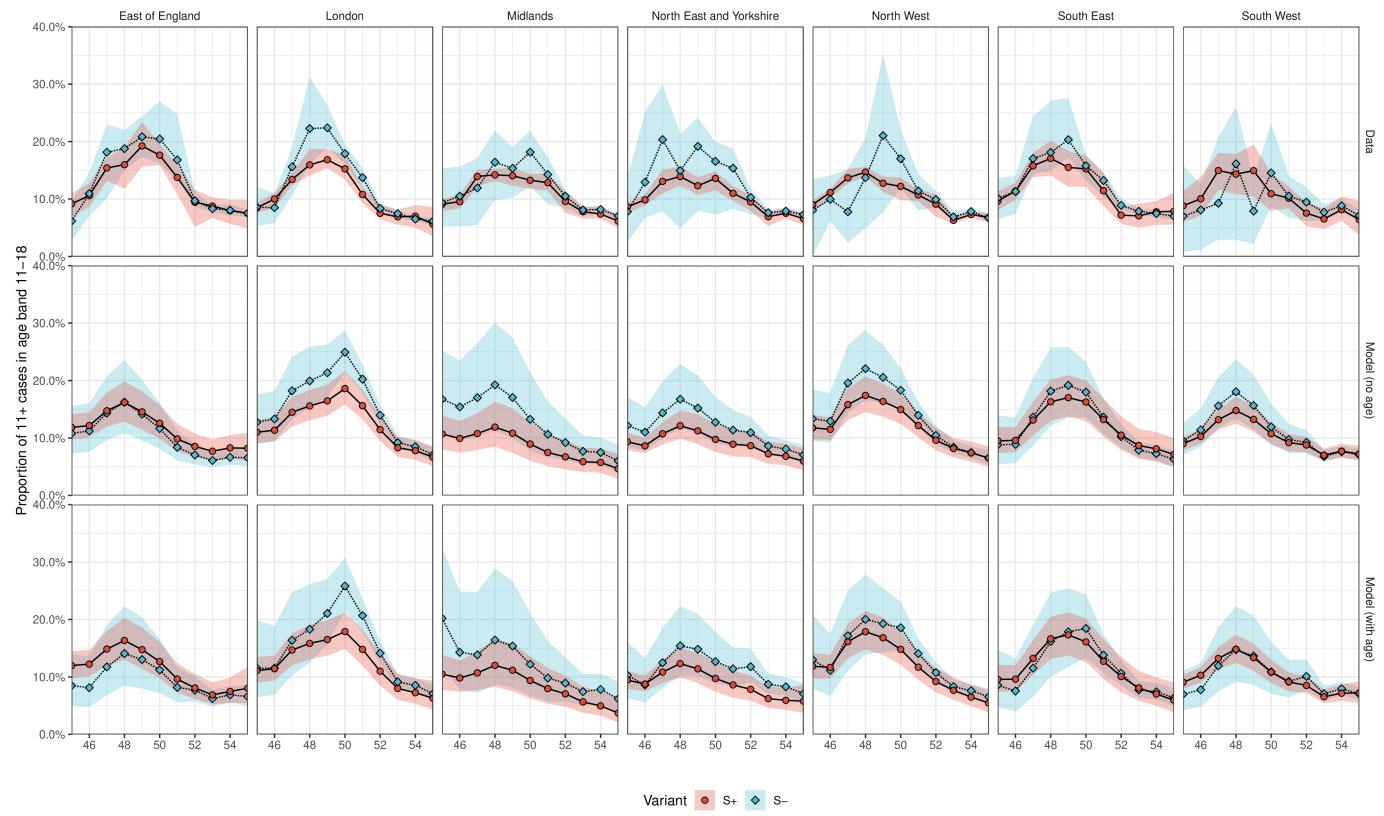
differentiate epidemiological weeks. Numbers above 1 on the left plot and above 0 on the right plot show a transmission advantage. The blue line represents the mean advantage for a particular proportion of VOC among all cases, and the grey shading the 95% asymptotic CrI ($\pm 2\sigma$). Scatter at low frequencies largely reflects statistical noise resulting from low counts.

Article



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

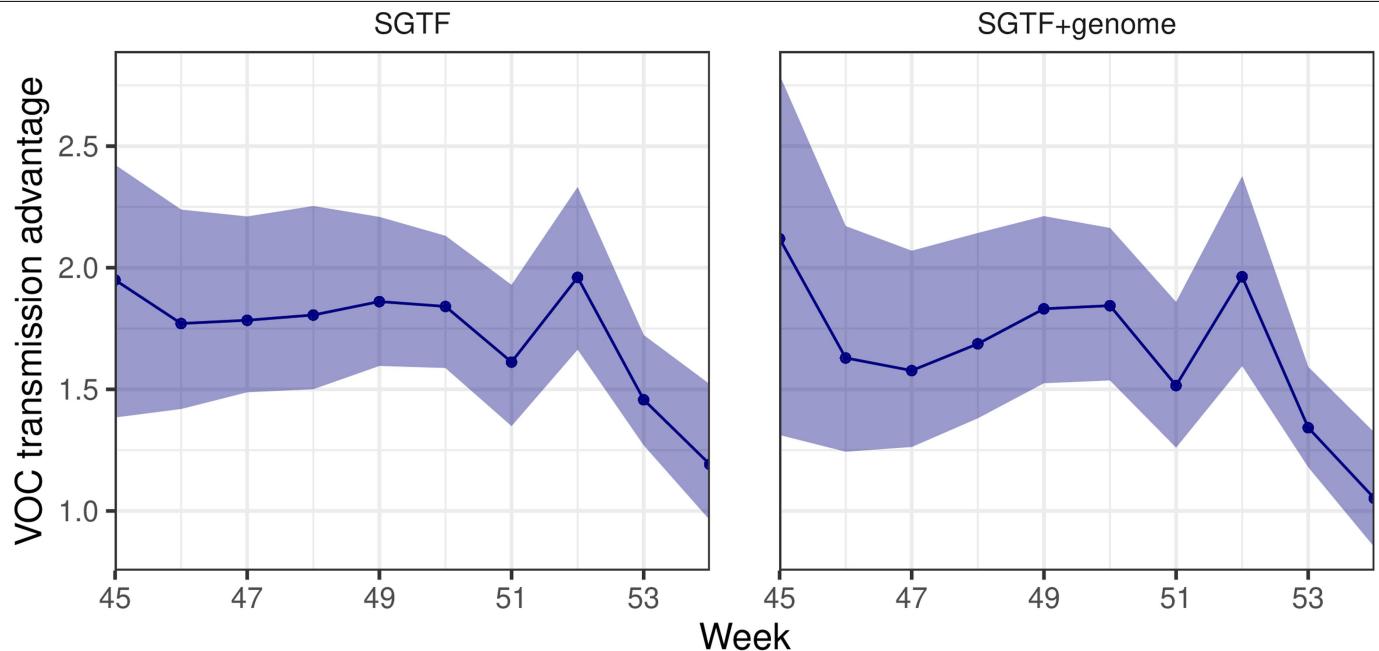
over time and among STP regions. Grey lines, linear regression—if changes in the generation time were able to completely explain temporal variation in the ratio, we would expect the slope to be zero. Bottom, ratio of reproduction number over time for each assumption about the mean generation time of the VOC. Shaded region, 95% Crl.



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

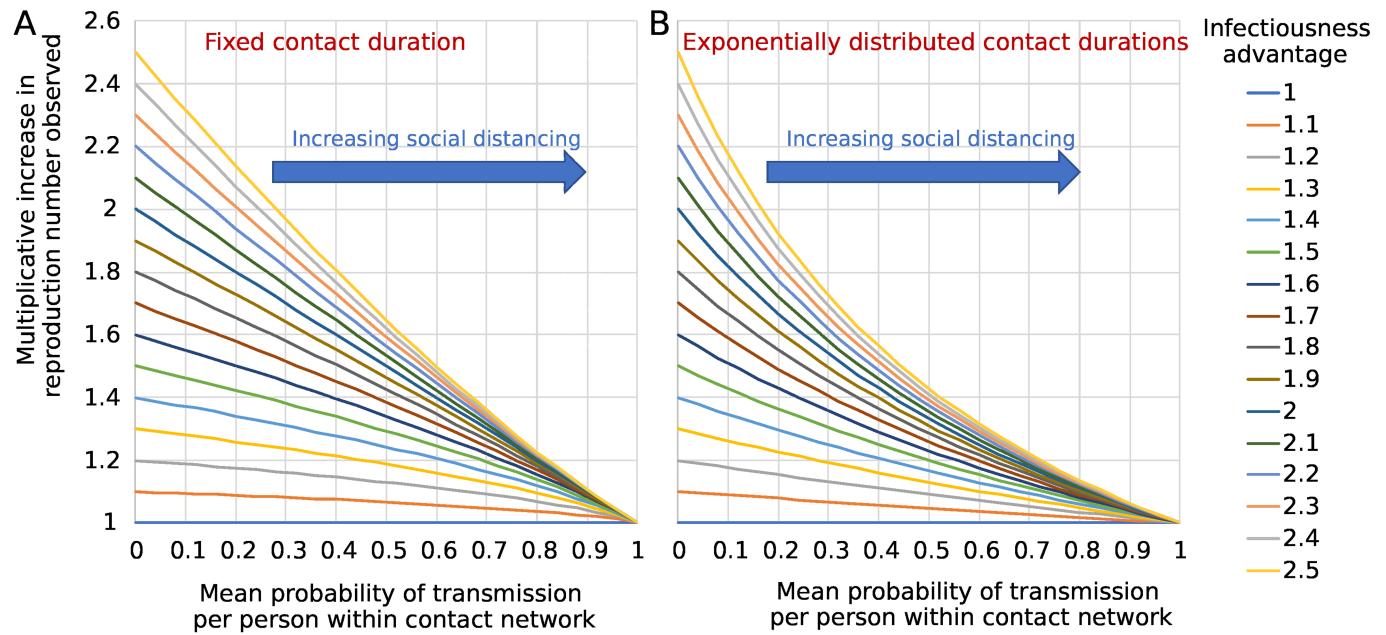
transmission advantage, but no age variation in transmission advantage; bottom, predicted trends from the model fitting a transmission that which varies by week, region and age group. Shaded region, 95% CrI.

Article



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.

Article

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).

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 analyses We 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 analyses Data 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 variant
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
<input checked="" type="checkbox"/>	Antibodies
<input checked="" type="checkbox"/>	Eukaryotic cell lines
<input checked="" type="checkbox"/>	Palaeontology and archaeology
<input checked="" type="checkbox"/>	Animals and other organisms
<input checked="" type="checkbox"/>	Human research participants
<input checked="" type="checkbox"/>	Clinical data
<input checked="" type="checkbox"/>	Dual use research of concern

Methods

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