nature genetics

BRIEF COMMUNICATION

https://doi.org/10.1038/s41588-021-00986-w

Check for updates

The *UGT2A1/UGT2A2* locus is associated with COVID-19-related loss of smell or taste

Janie F. Shelton, Anjali J. Shastri, Kipper Fletez-Brant, The 23andMe COVID-19 Team*, Stella Aslibekyan and Adam Auton ✉

Using online surveys, we collected data regarding COVID-19-related loss of smell or taste from 69,841 individuals. We performed a multi-ancestry genome-wide association study and identified a genome-wide significant locus in the vicinity of the *UGT2A1* and *UGT2A2* genes. Both genes are expressed in the olfactory epithelium and play a role in metabolizing odorants. These findings provide a genetic link to the biological mechanisms underlying COVID-19-related loss of smell or taste.

Loss of sense of smell (anosmia) or taste (ageusia) are distinctive symptoms of COVID-19 and are among the earliest and most often reported indicators of the acute phase of SARS-CoV-2 infection. It is notable from other viral symptoms in its sudden onset and the absence of mucosal blockage[1]. While a large fraction of COVID-19 patients report loss of smell or taste, the underlying mechanism is unclear[2]. In this study, we conducted a genome-wide association study (GWAS) of COVID-19-related loss of smell or taste, having collected self-reported data from over 1 million 23andMe research participants as described previously[3]. By asking study participants to report the symptoms they encountered during their COVID-19 experience, we identified SARS-CoV-2 test-positive individuals who reported a loss of smell or taste and contrasted them with test-positive individuals who did not report a loss of smell or taste.

Of the individuals who self-reported having received a SARS-CoV-2 positive test, 68% reported loss of smell or taste as a symptom (47,298 out of a total of 69,841 individuals). Female respondents were more likely than male respondents to report this symptom (72% versus 61%; chi-squared test, $P = 5.7 \times 10^{-178}$) and those with this symptom were typically younger than those without this symptom (mean age of 41 years for those with loss of smell or taste versus 45 years for those without; $P = 2.34 \times 10^{-199}$, Welch's *t*-test). Among genetically determined ancestral groups, rates of loss of smell or taste ranged between 63% and 70% (Table 1). As expected, compared to other symptoms surveyed, loss of smell or taste was much more common among those with a SARS-CoV-2 positive test compared to those who self-reported other cold or flu-like symptoms but who tested negative for SARS-CoV-2 (Extended Data Fig. 1). In a logistic regression model predicting loss of smell or taste as a function of age, sex and genetic ancestry, individuals of East Asian or African American ancestry were significantly less likely to report loss of smell or taste (odds ratio (OR) = 0.8 and 0.88, respectively) relative to individuals of European ancestry (Supplementary Table 1).

For unrelated individuals with complete data, we conducted GWAS within each ancestry group separately (total sample size = 56,373) before performing a multi-ancestry meta-analysis using a fixed effects model. Each input GWAS was adjusted for inflation via genomic control ($\lambda = 1.029$, 1.037, 1.024, 1.042 and 1.071 within the European, Latino, African American, East Asian and South Asian ancestry GWAS, respectively), as was the subsequent meta-analysis ($\lambda = 1.001$). Within the multi-ancestry meta-analysis, we identified a single associated locus at chr4q13.3 (Fig. 1). No other locus achieved genome-wide significance in the multi-ancestry meta-analysis or in any of the input populations. The index SNP at this locus was rs7688383 (C/T, with T being the risk allele, $P = 1.4 \times 10^{-14}$, OR = 1.11). While most of the support for this genetic association within the multi-ancestry analysis comes from the European population (for which we have the largest sample size), the estimated effect sizes are consistent across populations

Table 1 | Sample sizes and percentages comparing self-reported loss of smell or taste versus no loss of smell or taste among those with a positive SARS-CoV-2 test result

		Positive SARS-CoV-2 test result				
		Loss of smell or taste, *n*	%	No loss of smell or taste, *n*	%	Total, *n*
	Total	47,298	68	22,543	32	69,841
Sex	Female	31,608	72	12,562	28	44,170
	Male	15,690	61	9,981	39	25,671
Age	≤25	6,276	71	2,620	29	8,896
	26–35	13,855	73	5,134	27	18,989
	36–45	10,539	70	4,552	30	15,091
	46–55	8,321	67	4,080	33	12,401
	56–65	5,673	62	3,522	38	9,195
	66–75	2,059	51	1,945	49	4,004
	>75	574	45	689	55	1,263
Ancestry	European	33,336	67	16,257	33	49,593
	Latino	9,233	70	3,944	30	13,177
	African American or Black	1,860	66	949	34	2,809
	East Asian	626	65	338	35	964
	South Asian	284	63	167	37	451
	Other	1,959	69	888	31	2,847

[1]23andMe Inc., Sunnyvale, CA, USA. *A list of authors and affiliations appears at the end of the paper. ✉e-mail: aauton@23andme.com

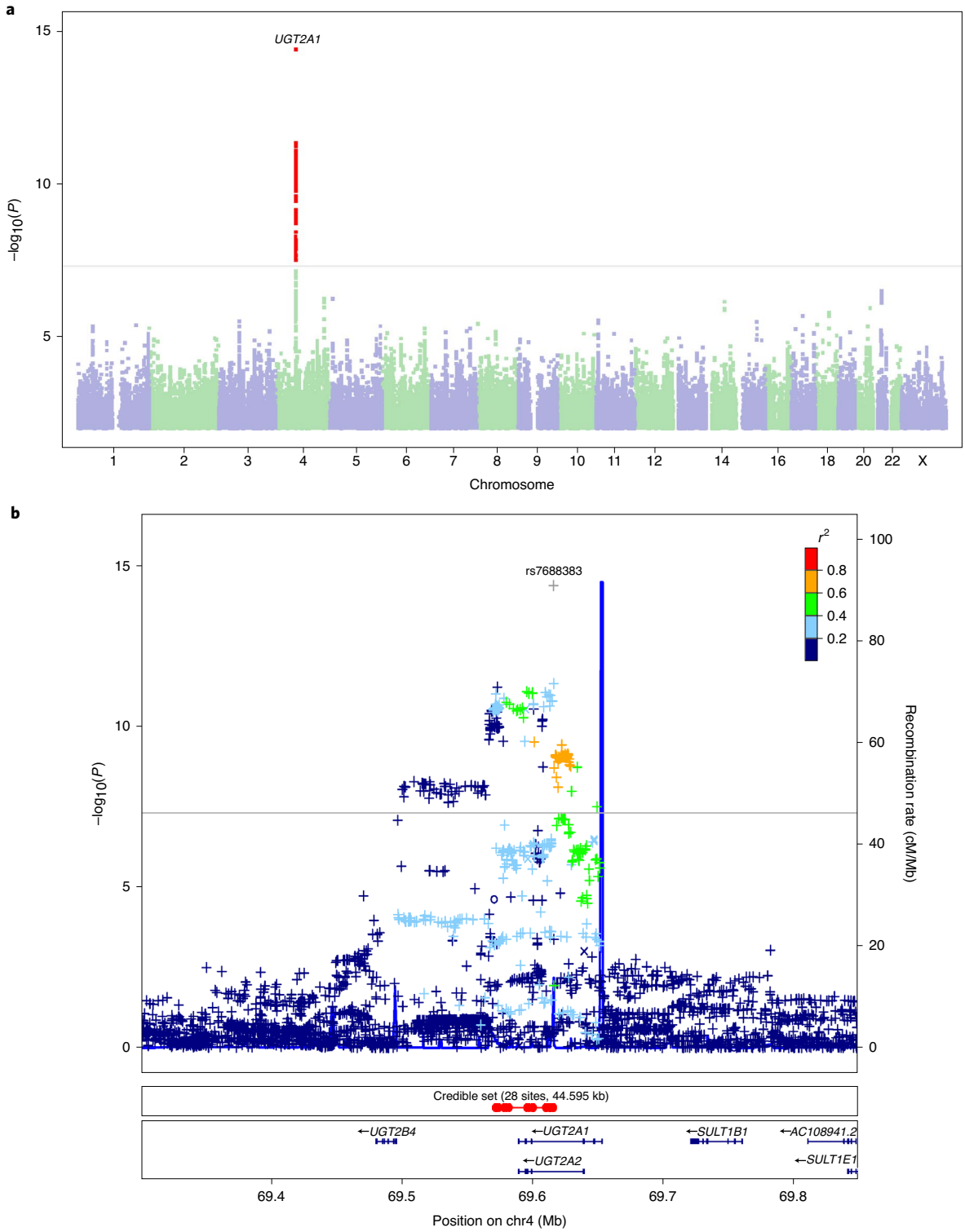


Fig. 1 | Manhattan plot and regional plot for the 'loss of taste or smell' phenotype comparing SARS-CoV-2-positive individuals with and without this symptom. a, Manhattan plot. SNPs achieving genome-wide significance are highlighted in red. The nearest gene to the index SNP is indicated above the relevant association peak. **b**, Regional plot around the *UGT2A1/UGT2A2* locus. The colors indicate the strength of linkage disequilibrium (r^2) relative to the index SNP (rs7688383). Imputed variants are indicated with '+' symbols; coding variants are indicated with 'x' symbols. Where imputed variants were not available, directly genotyped variants are indicated by 'o' symbols; coding variants are indicated by diamond symbols.

(Supplementary Table 2). The credible set from the multi-ancestry analysis contained 28 variants covering a 44.6-kilobase (kb) region (chr4:69.57–69.62 megabases (Mb); Supplementary Table 3).

We performed a phenome-wide association study on the index SNP across approximately 1,300 phenotypes defined in the 23andMe database. We identified four additional associated phenotypes with

$P < 1 \times 10^{-6}$, of which two are related to the ability to smell, one is related to ice cream taste preference and one is related to tobacco use (Supplementary Table 4). We detected no other associations with COVID-19 symptoms, susceptibility or severity.

There are four genes within 150 kb of the association (*UGT2A1*, *UGT2A2*, *UGT2B4*, *SULT1B1*), with the index variant itself being within an intron of the overlapping *UGT2A1* and *UGT2A2* genes. While the GWAS index variant appears to be physically proximal to expression quantitative trait loci (eQTLs) for *UGT2A1*, evidence for colocalization between the GWAS and eQTLs for any of the nearby genes is generally weak (Supplementary Note and Extended Data Figs. 2 and 3). Nonetheless, of the four genes in the vicinity, *UGT2A1* and *UGT2A2* are not only the most proximal but also the most biologically plausible causal gene candidates. *UGT2A1* and *UGT2A2* are part of a family of uridine diphosphate glycosyltransferases, enzymes that metabolize lipophilic substrates through conjugation with glucoronic acid. During olfaction, animal studies show that these enzymes, which are expressed in the olfactory epithelium, are involved in the elimination of the odorants that enter the nasal cavity and bind to olfactory receptors. For example, glucuronidation of odorants fails to stimulate the olfactory bulb, which prevents the odor from being detected by the brain, functionally demonstrating the effect of the enzyme produced by these genes on the odorant[4]. This results in the clearance of the odorant to facilitate the transient experience of olfaction, once the stimuli are no longer present in the environment[5]. *UGT2A2* is a splice variant of *UGT2A1*, with identical C-terminal residues but different N termini[6]. Conversely, *UGT2B4* and *SULT1B1* appear less plausible from a biological perspective, with neither having a clear link to olfactory or gustatory function.

While mechanistic explanations have been proposed[7] for COVID-19-related loss of smell, experimental studies suggest that loss of smell is related to damage to the cilia and olfactory epithelium but not infection of the olfactory neurons. For example, in an experiment where hamsters were nasally infected with SARS-CoV-2, the olfactory epithelium and cilia became very damaged, which can completely inhibit the ability to smell, but no infection was observed in the olfactory neurons[8]. Recent evidence suggests that SARS-CoV-2 enters and accumulates in olfactory support cells, specifically, sustentacular cells, which unlike olfactory neurons abundantly express the viral entry proteins angiotensin-converting enzyme 2 (ACE2) and transmembrane protease serine 2 (TMPRSS2; refs. [9,10]). These support cells are metabolically and functionally associated with olfactory neurons and with odorant signal transduction (processing odorants by endocytosing the odorant-binding protein complex, detoxifying, maintaining the cilia of mature olfactory receptor neurons and maintaining epithelial integrity). It has been proposed that olfactory sensation is impaired when these essential functions are disrupted, causing ciliary impairment[7]. How *UGT2A1* and *UGT2A2* are involved in this process is unclear but given their localization and essential function in metabolizing and detoxifying such compounds, these genes may play a role in the physiology of infected cells and the resulting functional impairment that contributes to loss of ability to smell. Notably, the variant identified in this study also appears to be associated with general ability to smell, which may suggest that those with heightened smell or taste sensitivity may be more prone to notice a loss of these senses resulting from a SARS-CoV-2 infection.

Our study has several limitations. First, while our study was large in scale, it was biased toward individuals of European ancestry and lacked a replication cohort. Second, we relied on self-reported case and symptom status; replication within a cohort with clinical ascertainment could be beneficial. Third, given that loss of smell or taste were combined in a single survey question, we cannot further disentangle these two symptoms. Loss of smell without loss of taste may be distinct from loss of both or loss of taste without loss of smell. Given this, it is unclear if our findings relate more strongly to one symptom or the other.

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/s41588-021-00986-w.

Received: 14 June 2021; Accepted: 12 November 2021;
Published online: 17 January 2022

References

1. Parma, V. et al. More than smell—COVID-19 is associated with severe impairment of smell, taste, and chemesthesis. *Chem. Senses* **45**, 609–622 (2020).
2. Mutiawati, E. et al. Anosmia and dysgeusia in SARS-CoV-2 infection: incidence and effects on COVID-19 severity and mortality, and the possible pathobiology mechanisms—a systematic review and meta-analysis. *F1000Res.* **10**, 40 (2021).
3. Shelton, J. F. et al. Trans-ancestry analysis reveals genetic and nongenetic associations with COVID-19 susceptibility and severity. *Nat. Genet.* **53**, 801–808 (2021).
4. Neiers, F., Jarriault, D., Menetrier, F., Briand, L. & Heydel, J.-M. The odorant metabolizing enzyme UGT2A1: immunolocalization and impact of the modulation of its activity on the olfactory response. *PLoS ONE* **16**, e0249029 (2021).
5. Lazard, D. et al. Odorant signal termination by olfactory UDP glucuronosyl transferase. *Nature* **349**, 790–793 (1991).
6. Mackenzie, P. I. et al. Nomenclature update for the mammalian UDP glycosyltransferase (UGT) gene superfamily. *Pharmacogenet. Genomics* **15**, 677–685 (2005).
7. Butowt, R. & von Bartheld, C. S. Anosmia in COVID-19: underlying mechanisms and assessment of an olfactory route to brain infection. *Neuroscientist* **27**, 582–603 (2021).
8. Bryche, B. et al. Massive transient damage of the olfactory epithelium associated with infection of sustentacular cells by SARS-CoV-2 in golden Syrian hamsters. *Brain Behav. Immun.* **89**, 579–586 (2020).
9. Brann, D. H. et al. Non-neuronal expression of SARS-CoV-2 entry genes in the olfactory system suggests mechanisms underlying COVID-19-associated anosmia. *Sci. Adv.* **6**, eabc5801 (2020).
10. Bilinska, K., Jakubowska, P., Von Bartheld, C. S. & Butowt, R. Expression of the SARS-CoV-2 entry proteins, ACE2 and TMPRSS2, in cells of the olfactory epithelium: identification of cell types and trends with age. *ACS Chem. Neurosci.* **11**, 1555–1562 (2020).

The 23andMe COVID-19 Team

Adam Auton, Adrian Chubb, Alison Fitch, Alison Kung, Amanda Altman, Andy Kill, Anjali J. Shastri, Antony Symons, Catherine Weldon, Daniella Coker, Janie F. Shelton, Jason Tan, Jeff Pollard, Jey McCreight, Jess Bielenberg, John Matthews, Johnny Lee, Lindsey Tran, Maya Lowe, Monica Royce, Nate Tang, Pooja Gandhi, Raffaello d'Amore, Ruth Tennen, Scott Dvorak, Scott Hadly, Stella Aslibekyan, Sungmin Park, Taylor Morrow, Teresa Filshtein Sonmez, Trung Le and Yiwen Zheng

Methods

Overview of study recruitment and data collection. Participants in this study were recruited from the customer base of 23andMe, a personal genetics company. Participants provided informed consent and participated in the research online, under a protocol approved by the external Association for the Accreditation of Human Research Protection Programs-accredited institutional review board, Ethical and Independent Review Services. Participants were included in the analysis based on consent status as checked at the time data analyses were initiated.

Full details of the data collection paradigm for this study have been described previously[3]. In brief, primary recruitment was carried out by email to approximately 6.7 million 23andMe research participants over 18 years of age and living in the USA or UK. Additionally, pre-existing customers were invited to participate in the study through promotional materials on the 23andMe website, the 23andMe mobile application and via social media. Study participation consisted solely of web-based surveys, including an initial baseline survey and three follow-up surveys fielded each month after completion of the baseline survey. The surveys collected information regarding individuals' experiences with COVID-19 and included questions regarding recently experienced symptoms with or without a SARS-CoV-2 positive test. Enrollment continued after the initial recruitment efforts until a data freeze was taken for this study in March 2021, when 1.3 million participants had completed the baseline survey.

Phenotype definition for GWAS. Using the information derived from the surveys, we defined a phenotype to contrast SARS-CoV-2 positive individuals that experienced COVID-19-related loss of smell or taste from those who did not. Specifically, participants were asked to respond to the question 'Have you been tested for COVID-19 (coronavirus)?', with possible responses 'Yes, it was positive/Yes, it was negative/No/My results are pending/I'm not sure'. Of those who responded 'Yes, it was positive', we further considered the question 'During your illness, did you experience any of the following symptoms?', to which participants could select as many as needed from the following list of responses: 'Muscle or body aches/Fatigue/Dry cough/Sore throat/Coughing up of sputum or phlegm (productive cough)/Loss of smell or taste/Chills/Difficulty breathing or shortness of breath/Pressure or tightness in upper chest/Diarrhea/Nausea or vomiting/Sneezing/Loss of appetite/Runny nose/Headache/Intensely red or watery eyes'. We defined cases as SARS-CoV-2 test-positive individuals who also reported 'Loss of smell or taste', and controls as SARS-CoV-2 test-positive individuals who did not report 'Loss of smell or taste'. While some participants reported a COVID-19 diagnosis absent a confirmed positive test for SARS-CoV-2, we did not include such individuals within this analysis.

Descriptive statistics. Sample sizes and proportions were calculated by age, sex and ancestry. Differences in loss of smell or taste by sex were statistically evaluated with a chi-squared statistics and mean differences in age were evaluated with a *t*-test. A logistic regression model was constructed to evaluate loss of smell or taste as a function of ancestry, age (categorical) and sex. All analyses were conducted in R v.3.6.3.

Genotyping and SNP imputation. DNA extraction and genotyping were performed on saliva samples by Clinical Laboratory Improvement Amendments-certified and College of American Pathologists-accredited clinical laboratories of Laboratory Corporation of America. Samples were genotyped on one of five genotyping platforms. The V1 and V2 platforms were variants of the Illumina HumanHap550 BeadChip and contained a total of about 560,000 SNPs, including about 25,000 custom SNPs selected by 23andMe. The V3 platform was based on the Illumina OmniExpress BeadChip and contained a total of about 950,000 SNPs and custom content to improve the overlap with our V2 array. The V4 platform was a fully custom array of about 950,000 SNPs and included a lower redundancy subset of V2 and V3 SNPs with additional coverage of lower-frequency coding variation. The V5 platform was based on the Illumina Global Screening Array, consisting of approximately 654,000 preselected SNPs and approximately 50,000 custom content variants. Samples that failed to reach 98.5% call rate were reanalyzed. Individuals whose analyses failed repeatedly were recontacted by the 23andMe customer service to provide additional samples as done for all 23andMe customers.

Participant genotype data were imputed using the Haplotype Reference Consortium (HRC) panel[11], augmented by the phase 3 1000 Genomes Project panel[12] for variants not present in the HRC. We phased and imputed data for each genotyping platform separately. For the non-pseudoautosomal region of the X chromosome, males and females were phased together in segments, treating males as already phased; the pseudoautosomal regions were phased separately. We then imputed males and females together, treating males as homozygous pseudo-diploids for the non-pseudoautosomal region.

GWAS. Genotyped participants were included in the GWAS analyses on the basis of ancestry as determined by a genetic ancestry classification algorithm[13]. We selected a set of unrelated individuals so that no 2 individuals shared more than 700 cM of DNA identical by descent (IBD). If a case and a control were identified as having at least 700 cM of DNA IBD, we preferentially discarded the control from the sample. This filtering paradigm resulted in approximately 1.76% of the sample being excluded.

We tested for association using logistic regression, assuming additive allelic effects. For tests using imputed data, we used the imputed dosages rather than best-guess genotypes. We included covariates for age, age squared, sex, a sex:age interaction, the top ten principal components to account for residual population structure and dummy variables to account for the genotyping platform. The association test *P* value was computed using a likelihood ratio test, which in our experience is better behaved than a Wald test on the regression coefficient. Results for the X chromosome were computed similarly, with men coded as if they were homozygous diploid for the observed allele.

We combined the GWAS summary statistics from both genotyped and imputed data. When choosing between imputed and genotyped GWAS results, we favored the imputed result, unless the imputed variant was unavailable or failed quality control. For imputed variants, we removed variants with low imputation quality ($r^2 < 0.5$ averaged across batches or a minimum $r^2 < 0.3$) or with evidence of batch effects (analysis of variance (ANOVA) *F*-test across batches, $P < 10^{-50}$). For genotyped variants, we removed variants only present on our V1 or V2 arrays (due to small sample size) that failed a Mendelian transmission test in trios ($P < 10^{-20}$), failed a Hardy–Weinberg test in individuals of European ancestry ($P < 10^{-20}$), failed a batch effect test (ANOVA $P < 10^{-50}$) or had a call rate <90%.

We repeated the GWAS analysis separately in each population cohort for which we had sufficient data (European, Latino, African American, East Asian and South Asian ancestry); the resulting summary statistics were adjusted for inflation using genomic control when the inflation factor was estimated to be greater than 1. We then performed multi-ancestry meta-analysis using a fixed effects model (inverse variance method[14]), restricting to variants of at least 1% minor allele frequency in the pooled sample and minor allele count > 30 within each subpopulation. Both the input GWAS and resulting meta-analysis were adjusted for inflation using genomic control where necessary.

We identified regions with genome-wide significant associations. We defined the region boundaries by identifying all SNPs with $P < 10^{-5}$ within the vicinity of a genome-wide significance association and then grouping these regions into intervals so that no 2 regions were separated by less than 250 kb. We considered the SNP with the smallest *P* value within each interval to be the index SNP. Within each region, we calculated a credible set using the method of Maller et al.[15].

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

Data availability

The full set of de-identified summary statistics can be made available to qualified investigators who enter into an agreement with 23andMe that protects participant confidentiality. Interested investigators should visit https://research.23andme.com/covid19-dataset-access/.

References

11. McCarthy, S. et al. A reference panel of 64,976 haplotypes for genotype imputation. *Nat. Genet.* **48**, 1279–1283 (2016).
12. Auton, A. et al. A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
13. Durand, E. Y., Do, C. B., Mountain, J. L. & Macpherson, J. M. Ancestry Composition: a novel, efficient pipeline for ancestry deconvolution. Preprint at *bioRxiv* https://doi.org/10.1101/010512 (2014).
14. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190–2191 (2010).
15. Maller, J. B. et al. Bayesian refinement of association signals for 14 loci in 3 common diseases. *Nat. Genet.* **44**, 1294–1301 (2012).

Acknowledgements

We thank the 23andMe research participants and employees who made this study possible. The 23andMe Research Team is: S. Aslibekyan, A. Auton, E. Babalola, R. K. Bell, J. Bielenberg, K. Bryc, E. Bullis, D. Coker, G. Cuellar Partida, D. Dhamija, S. Das, S. L. Elson, T. Filshtein, K. Fletez-Brant, P. Fontanillas, W. Freyman, P. M. Gandhi, K. Heilbron, B. Hicks, D. A. Hinds, E. M. Jewett, Y. Jiang, K. Kukar, K.-H. Lin, M. Lowe, J. McCreight, M. H. McIntyre, S. J. Micheletti, M. E. Moreno, J. L. Mountain, P. Nandakumar, E. S. Noblin, J. O'Connell, A. A. Petrakovitz, G. D. Poznik, M. Schumacher, A. J. Shastri, J. F. Shelton, J. Shi, S. Shringarpure, V. Tran, J. Y. Tung, X. Wang, W. Wang, C. H. Weldon, P. Wilton, A. Hernandez, C. Wong and C. Toukam Tchakouté.

Author contributions

J.F.S., A.J.S., S.A. and A.A. designed this study. The 23andMe COVID-19 Team developed the recruitment and participant engagement strategy and acquired and processed the data. J.F.S., K.F.-B. and A.A. analyzed the data. J.F.S., A.J.S., K.F.-B. and A.A. interpreted the data. J.F.S., A.J.S. and A.A. wrote the manuscript. All authors participated in the preparation of the manuscript by reading and commenting on the drafts before submission.

Competing interests

J.F.S., A.J.S., K.F.-B., S.A. and A.A. are current employees of 23andMe and hold stock or stock options in 23andMe.

Additional information

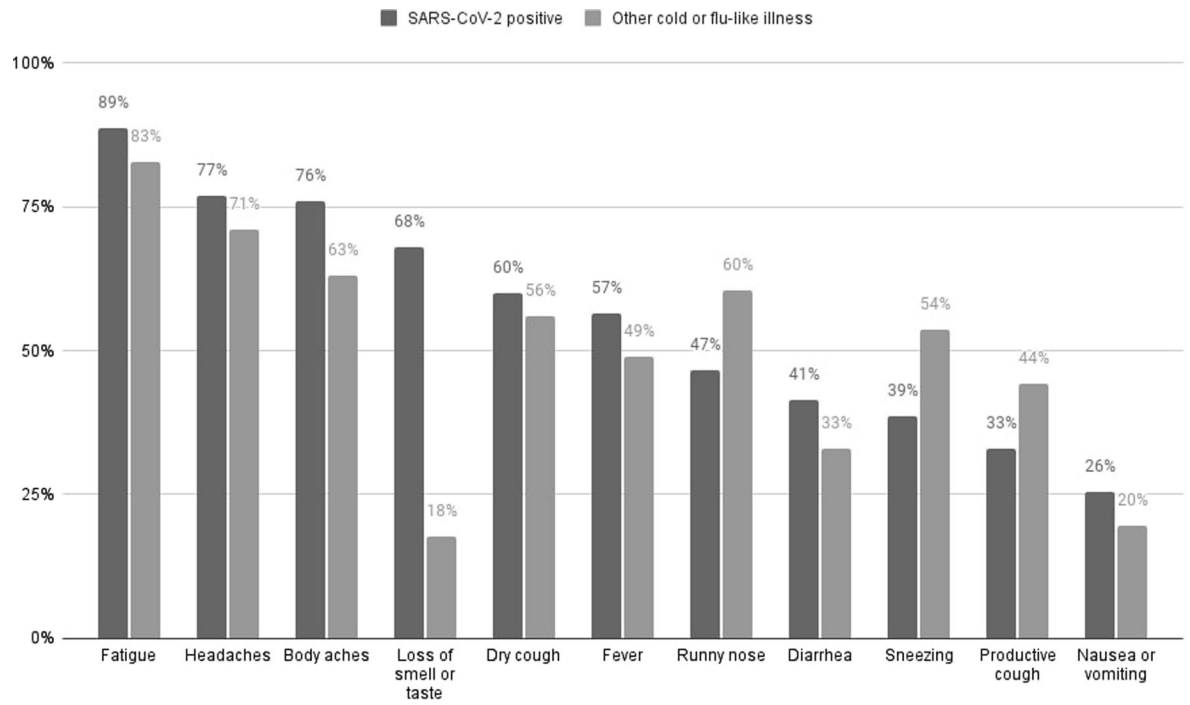
Extended data is available for this paper at https://doi.org/10.1038/s41588-021-00986-w.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41588-021-00986-w.

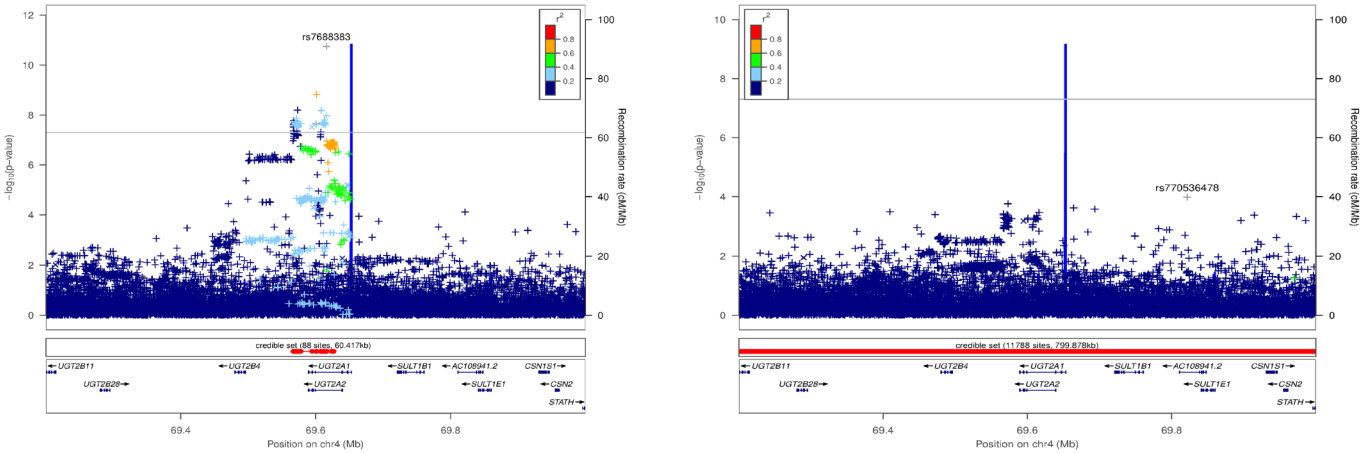
Correspondence and requests for materials should be addressed to Adam Auton.

Peer review information *Nature Genetics* thanks Patrick Sulem and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

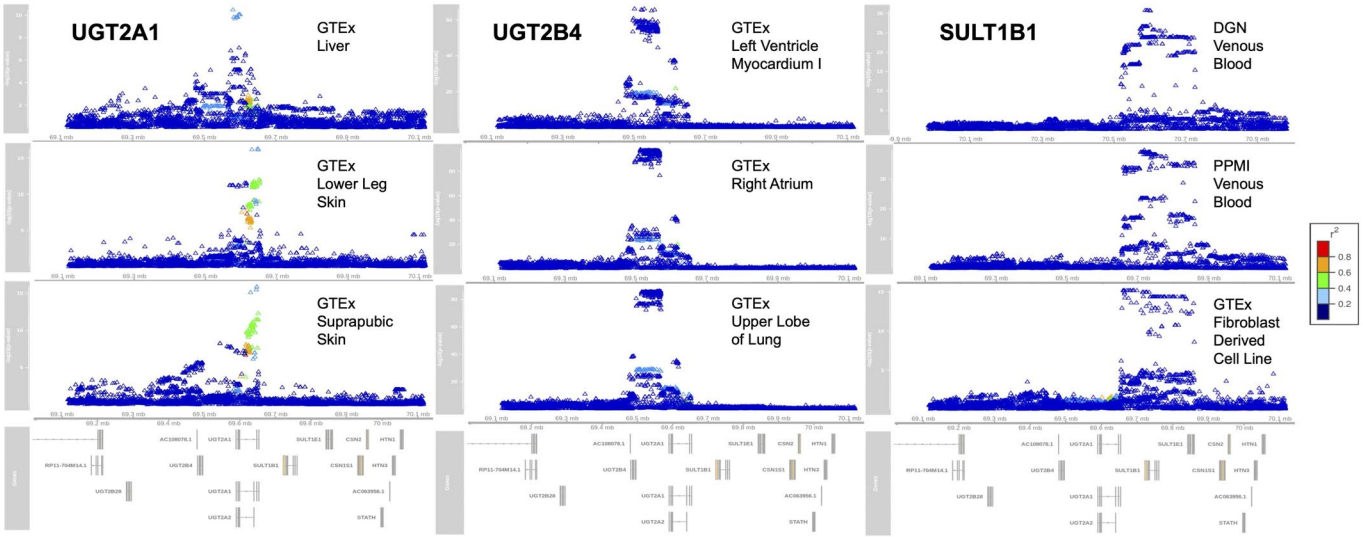
Reprints and permissions information is available at www.nature.com/reprints.



Extended Data Fig. 1 | Self-reported symptoms experienced during SARS-CoV-2 infection with a positive test ($n = 69,841$) as compared to individuals self-reporting cold or flu-like illness but with a negative SARS-CoV-2 test ($n = 314,441$). Loss of smell or taste was reported by 68% of individuals with a positive test for SARS-CoV-2 infection.



Extended Data Fig. 2 | Conditional association LocusZoom plots for rs768838. Lack of evidence for conditional associations. Left, LocusZoom plot of primary association in the European population prior to conditional analysis. Right, LocusZoom plot of the same region having included rs7688383 in the regression model.



Extended Data Fig. 3 | Examples of eQTL associations for *UTG2A1*, *UGT2B4*, and *SULT1B1*. eQTL association plots for *UGT2A1* (left), *UGT2B4* (middle), and *SULT1B1* (right). No eQTL associations were observed for *UGT2A2*. For each gene, the three tissues with the strongest eQTL associations are shown. Colors represent the linkage disequilibrium with the GWAS index SNP (rs7688383).

nature portfolio

Corresponding author(s): Adam Auton

Last updated by author(s): Sep 22, 2021

Reporting Summary

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

Statistics

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

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

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

Software and code

Policy information about availability of computer code

Data collection	Phenotype data were collected via online surveys using the 23andMe Research platform. No additional software was used.
Data analysis	Standard GWAS analyses were performed using the 23andMe Research platform that has been described in numerous peer-reviewed publications (https://research.23andme.com/publications/). Additional analyses were conducted in R version 3.6.3.

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

Data

Policy information about availability of data

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

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

Provide your data availability statement here.

The full set of de-identified GWAS summary statistics can be made available to qualified investigators who enter an agreement with 23andMe that protects participants confidentiality. Interested investigators should visit the following: https://research.23andme.com/covid19-dataset-access/.

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.

[X] Life sciences [ ] Behavioural & social sciences [ ] Ecological, evolutionary & environmental sciences

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

Life sciences study design

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

Sample size	Sample size was determined based on the number of respondents to online surveys and was judged appropriate adequately power GWAS analyses.
Data exclusions	Only 23andMe Research participants that consented to participate in research and completed the on line survey were included. No other exclusion criteria were used.
Replication	The primary association replicated across populations contained within this study, but has not been replicated in external data.
Randomization	This is an observational population study for which randomization was not applicable.
Blinding	This is an observational population study for which blinding was not applicable.

Behavioural & social sciences study design

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

Study description	*Briefly describe the study type including whether data are quantitative, qualitative, or mixed-methods (e.g. qualitative cross-sectional, quantitative experimental, mixed-methods case study).*
Research sample	*State the research sample (e.g. Harvard university undergraduates, villagers in rural India) and provide relevant demographic information (e.g. age, sex) and indicate whether the sample is representative. Provide a rationale for the study sample chosen. For studies involving existing datasets, please describe the dataset and source.*
Sampling strategy	*Describe the sampling procedure (e.g. random, snowball, stratified, convenience). Describe the statistical methods that were used to predetermine sample size OR if no sample-size calculation was performed, describe how sample sizes were chosen and provide a rationale for why these sample sizes are sufficient. For qualitative data, please indicate whether data saturation was considered, and what criteria were used to decide that no further sampling was needed.*
Data collection	*Provide details about the data collection procedure, including the instruments or devices used to record the data (e.g. pen and paper, computer, eye tracker, video or audio equipment) whether anyone was present besides the participant(s) and the researcher, and whether the researcher was blind to experimental condition and/or the study hypothesis during data collection.*
Timing	*Indicate the start and stop dates of data collection. If there is a gap between collection periods, state the dates for each sample cohort.*
Data exclusions	*If no data were excluded from the analyses, state so OR if data were excluded, provide the exact number of exclusions and the rationale behind them, indicating whether exclusion criteria were pre-established.*
Non-participation	*State how many participants dropped out/declined participation and the reason(s) given OR provide response rate OR state that no participants dropped out/declined participation.*
Randomization	*If participants were not allocated into experimental groups, state so OR describe how participants were allocated to groups, and if allocation was not random, describe how covariates were controlled.*

Ecological, evolutionary & environmental sciences study design

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

Study description	*Briefly describe the study. For quantitative data include treatment factors and interactions, design structure (e.g. factorial, nested, hierarchical), nature and number of experimental units and replicates.*
Research sample	*Describe the research sample (e.g. a group of tagged Passer domesticus, all Stenocereus thurberi within Organ Pipe Cactus National*

Research sample	*Monument), and provide a rationale for the sample choice. When relevant, describe the organism taxa, source, sex, age range and any manipulations. State what population the sample is meant to represent when applicable. For studies involving existing datasets, describe the data and its source.*
Sampling strategy	*Note the sampling procedure. Describe the statistical methods that were used to predetermine sample size OR if no sample-size calculation was performed, describe how sample sizes were chosen and provide a rationale for why these sample sizes are sufficient.*
Data collection	*Describe the data collection procedure, including who recorded the data and how.*
Timing and spatial scale	*Indicate the start and stop dates of data collection, noting the frequency and periodicity of sampling and providing a rationale for these choices. If there is a gap between collection periods, state the dates for each sample cohort. Specify the spatial scale from which the data are taken*
Data exclusions	*If no data were excluded from the analyses, state so OR if data were excluded, describe the exclusions and the rationale behind them, indicating whether exclusion criteria were pre-established.*
Reproducibility	*Describe the measures taken to verify the reproducibility of experimental findings. For each experiment, note whether any attempts to repeat the experiment failed OR state that all attempts to repeat the experiment were successful.*
Randomization	*Describe how samples/organisms/participants were allocated into groups. If allocation was not random, describe how covariates were controlled. If this is not relevant to your study, explain why.*
Blinding	*Describe the extent of blinding used during data acquisition and analysis. If blinding was not possible, describe why OR explain why blinding was not relevant to your study.*

Did the study involve field work? ☐ Yes ☐ No

Field work, collection and transport

Field conditions	*Describe the study conditions for field work, providing relevant parameters (e.g. temperature, rainfall).*
Location	*State the location of the sampling or experiment, providing relevant parameters (e.g. latitude and longitude, elevation, water depth).*
Access & import/export	*Describe the efforts you have made to access habitats and to collect and import/export your samples in a responsible manner and in compliance with local, national and international laws, noting any permits that were obtained (give the name of the issuing authority, the date of issue, and any identifying information).*
Disturbance	*Describe any disturbance caused by the study and how it was minimized.*

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
X	☐ Antibodies
X	☐ Eukaryotic cell lines
X	☐ Palaeontology and archaeology
X	☐ Animals and other organisms
☐	X Human research participants
X	☐ Clinical data
X	☐ Dual use research of concern

Methods

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

Antibodies

Antibodies used	*Describe all antibodies used in the study; as applicable, provide supplier name, catalog number, clone name, and lot number.*
Validation	*Describe the validation of each primary antibody for the species and application, noting any validation statements on the manufacturer's website, relevant citations, antibody profiles in online databases, or data provided in the manuscript.*

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	*State the source of each cell line used.*

Authentication	*Describe the authentication procedures for each cell line used OR declare that none of the cell lines used were authenticated.*
Mycoplasma contamination	*Confirm that all cell lines tested negative for mycoplasma contamination OR describe the results of the testing for mycoplasma contamination OR declare that the cell lines were not tested for mycoplasma contamination.*
Commonly misidentified lines (See ICLAC register)	*Name any commonly misidentified cell lines used in the study and provide a rationale for their use.*

Palaeontology and Archaeology

Specimen provenance	*Provide provenance information for specimens and describe permits that were obtained for the work (including the name of the issuing authority, the date of issue, and any identifying information). Permits should encompass collection and, where applicable, export.*
Specimen deposition	*Indicate where the specimens have been deposited to permit free access by other researchers.*
Dating methods	*If new dates are provided, describe how they were obtained (e.g. collection, storage, sample pretreatment and measurement), where they were obtained (i.e. lab name), the calibration program and the protocol for quality assurance OR state that no new dates are provided.*

☐ Tick this box to confirm that the raw and calibrated dates are available in the paper or in Supplementary Information.

Ethics oversight	*Identify the organization(s) that approved or provided guidance on the study protocol, OR state that no ethical approval or guidance was required and explain why not.*

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

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals	*For laboratory animals, report species, strain, sex and age OR state that the study did not involve laboratory animals.*
Wild animals	*Provide details on animals observed in or captured in the field; report species, sex and age where possible. Describe how animals were caught and transported and what happened to captive animals after the study (if killed, explain why and describe method; if released, say where and when) OR state that the study did not involve wild animals.*
Field-collected samples	*For laboratory work with field-collected samples, describe all relevant parameters such as housing, maintenance, temperature, photoperiod and end-of-experiment protocol OR state that the study did not involve samples collected from the field.*
Ethics oversight	*Identify the organization(s) that approved or provided guidance on the study protocol, OR state that no ethical approval or guidance was required and explain why not.*

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

Human research participants

Policy information about studies involving human research participants

Population characteristics	Participants in this study were recruited from the customer base of 23andMe, Inc., a personal genetics company. All individuals included in the analyses provided informed consent and answered surveys on-line according to our human subjects research protocol, which was reviewed and approved by Ethical and Independent Review Services, a private institutional review board (http://www.eandireview.com). The demographic characteristics of the cohort are fully detailed within Table 1 of the manuscript.
Recruitment	Participants in this study were self-selected by response to an email requesting participation and thereby may not be reflective of the general population. This type of bias to the study sample may have resulted in a population of cases who had less severe course of disease by virtue of being healthy enough to participate. In the event that participation is related to loss of smell or taste, the percent of the population with this symptom should not be interpreted as a population estimate.
Ethics oversight	Ethical and Independent Review Services

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

Clinical data

Policy information about clinical studies

All manuscripts should comply with the ICMJE guidelines for publication of clinical research and a completed CONSORT checklist must be included with all submissions.

Clinical trial registration	*Provide the trial registration number from ClinicalTrials.gov or an equivalent agency.*
Study protocol	*Note where the full trial protocol can be accessed OR if not available, explain why.*
Data collection	*Describe the settings and locales of data collection, noting the time periods of recruitment and data collection.*
Outcomes	*Describe how you pre-defined primary and secondary outcome measures and how you assessed these measures.*

Dual use research of concern

Policy information about dual use research of concern

Hazards

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

No	Yes	
☐	☐	Public health
☐	☐	National security
☐	☐	Crops and/or livestock
☐	☐	Ecosystems
☐	☐	Any other significant area

Experiments of concern

Does the work involve any of these experiments of concern:

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

ChIP-seq

Data deposition

☐ Confirm that both raw and final processed data have been deposited in a public database such as GEO.

☐ Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

Data access links *May remain private before publication.*	*For "Initial submission" or "Revised version" documents, provide reviewer access links. For your "Final submission" document, provide a link to the deposited data.*
Files in database submission	*Provide a list of all files available in the database submission.*
Genome browser session (e.g. UCSC)	*Provide a link to an anonymized genome browser session for "Initial submission" and "Revised version" documents only, to enable peer review. Write "no longer applicable" for "Final submission" documents.*

Methodology

Replicates	*Describe the experimental replicates, specifying number, type and replicate agreement.*
Sequencing depth	*Describe the sequencing depth for each experiment, providing the total number of reads, uniquely mapped reads, length of reads and whether they were paired- or single-end.*
Antibodies	*Describe the antibodies used for the ChIP-seq experiments; as applicable, provide supplier name, catalog number, clone name, and lot number.*
Peak calling parameters	*Specify the command line program and parameters used for read mapping and peak calling, including the ChIP, control and index files used.*
Data quality	*Describe the methods used to ensure data quality in full detail, including how many peaks are at FDR 5% and above 5-fold enrichment.*
Software	*Describe the software used to collect and analyze the ChIP-seq data. For custom code that has been deposited into a community repository, provide accession details.*

Flow Cytometry

Plots

Confirm that:

☐ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☐ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☐ All plots are contour plots with outliers or pseudocolor plots.

☐ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	*Describe the sample preparation, detailing the biological source of the cells and any tissue processing steps used.*
Instrument	*Identify the instrument used for data collection, specifying make and model number.*
Software	*Describe the software used to collect and analyze the flow cytometry data. For custom code that has been deposited into a community repository, provide accession details.*
Cell population abundance	*Describe the abundance of the relevant cell populations within post-sort fractions, providing details on the purity of the samples and how it was determined.*
Gating strategy	*Describe the gating strategy used for all relevant experiments, specifying the preliminary FSC/SSC gates of the starting cell population, indicating where boundaries between "positive" and "negative" staining cell populations are defined.*

☐ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.

Magnetic resonance imaging

Experimental design

Design type	*Indicate task or resting state; event-related or block design.*
Design specifications	*Specify the number of blocks, trials or experimental units per session and/or subject, and specify the length of each trial or block (if trials are blocked) and interval between trials.*
Behavioral performance measures	*State number and/or type of variables recorded (e.g. correct button press, response time) and what statistics were used to establish that the subjects were performing the task as expected (e.g. mean, range, and/or standard deviation across subjects).*

Acquisition

Imaging type(s)	*Specify: functional, structural, diffusion, perfusion.*
Field strength	*Specify in Tesla*
Sequence & imaging parameters	*Specify the pulse sequence type (gradient echo, spin echo, etc.), imaging type (EPI, spiral, etc.), field of view, matrix size, slice thickness, orientation and TE/TR/flip angle.*
Area of acquisition	*State whether a whole brain scan was used OR define the area of acquisition, describing how the region was determined.*

Diffusion MRI ☐ Used ☐ Not used

Preprocessing

Preprocessing software	*Provide detail on software version and revision number and on specific parameters (model/functions, brain extraction, segmentation, smoothing kernel size, etc.).*
Normalization	*If data were normalized/standardized, describe the approach(es): specify linear or non-linear and define image types used for transformation OR indicate that data were not normalized and explain rationale for lack of normalization.*
Normalization template	*Describe the template used for normalization/transformation, specifying subject space or group standardized space (e.g. original Talairach, MNI305, ICBM152) OR indicate that the data were not normalized.*
Noise and artifact removal	*Describe your procedure(s) for artifact and structured noise removal, specifying motion parameters, tissue signals and physiological signals (heart rate, respiration).*

Volume censoring

Define your software and/or method and criteria for volume censoring, and state the extent of such censoring.

Statistical modeling & inference

Model type and settings

Specify type (mass univariate, multivariate, RSA, predictive, etc.) and describe essential details of the model at the first and second levels (e.g. fixed, random or mixed effects; drift or auto-correlation).

Effect(s) tested

Define precise effect in terms of the task or stimulus conditions instead of psychological concepts and indicate whether ANOVA or factorial designs were used.

Specify type of analysis: ☐ Whole brain ☐ ROI-based ☐ Both

Statistic type for inference
(See Eklund et al. 2016)

Specify voxel-wise or cluster-wise and report all relevant parameters for cluster-wise methods.

Correction

Describe the type of correction and how it is obtained for multiple comparisons (e.g. FWE, FDR, permutation or Monte Carlo).

Models & analysis

n/a	Involved in the study
☐	☐ Functional and/or effective connectivity
☐	☐ Graph analysis
☐	☐ Multivariate modeling or predictive analysis

Functional and/or effective connectivity

Report the measures of dependence used and the model details (e.g. Pearson correlation, partial correlation, mutual information).

Graph analysis

Report the dependent variable and connectivity measure, specifying weighted graph or binarized graph, subject- or group-level, and the global and/or node summaries used (e.g. clustering coefficient, efficiency, etc.).

Multivariate modeling and predictive analysis

Specify independent variables, features extraction and dimension reduction, model, training and evaluation metrics.