

CORRESPONDENCE

Open Access

No evidence that *ACE2* or *TMPRSS2* drive population disparity in COVID risks



Nathaniel M. Pearson^{1*}  and John Novembre²

Abstract

Early in the SARS-CoV2 pandemic, in this journal, Hou et al. (*BMC Med* 18:216, 2020) interpreted public genotype data, run through functional prediction tools, as suggesting that members of particular human populations carry potentially COVID-risk-increasing variants in genes *ACE2* and *TMPRSS2* far more often than do members of other populations. Beyond resting on predictions rather than clinical outcomes, and focusing on variants too rare to typify population members even jointly, their claim mistook a well known artifact (that large samples reveal more of a population's variants than do small samples) as if showing real and congruent population differences for the two genes, rather than lopsided population sampling in their shared source data. We explain that artifact, and contrast it with empirical findings, now ample, that other loci shape personal COVID risks far more significantly than do *ACE2* and *TMPRSS2*—and that variation in *ACE2* and *TMPRSS2* per se unlikely exacerbates any net population disparity in the effects of such more risk-informative loci.

Keywords *ACE2*, COVID, COVID19, Functional prediction, GWAS, Host genetics, Human genes, Immunity, Infection, Polygenic risk, Population genetics, Population structure, Rare variants, Sample design, Sample size, Sampling, SARS-CoV2, *TMPRSS2*

Background

In mid-2020, concurrent with early empirical inquiry into roles of host genomic variation in SARS-CoV2 infection [1–5], Hou et al. set out to offer predictive guidance for such efforts, by assessing pre-pandemic public DNA data from two human genes, *ACE2* and *TMPRSS2*, whose protein products were known to interact with other coronaviruses [6].

Pooling public genotype data sampled from various human groups, without phenotypes, they shortlisted *ACE2* and *TMPRSS2* variants that some computational heuristics predicted likely to alter protein function, and found that most such variants (each, typically, very rare)

came from subsets of data labeled “African/African-American” or “Non-Finnish European” versus labeled otherwise (e.g., “East Asian”).¹ This, they held, suggested real-world population disparities in *ACE2* and *TMPRSS2* functional variant load, similar for both genes, that might in turn drive population differences in COVID outcomes.

Errant interpretation of genotype data

Alas, Hou et al. had neglected a basic feature of the public data they used—lopsided population sample sizes—that made their summary findings artifactually likely even with no difference between real populations.

*Correspondence:

Nathaniel M. Pearson
nathaniel.pearson@gmail.com

¹ Root Deep Insight, Boston, MA, USA

² Department of Human Genetics, University of Chicago, Chicago, IL, USA

¹ Denoted by Hou et al. as, respectively, “African/African-American” and “Non-Finnish [and implicitly non-Amish, non-Ashkenazi] European” in Hou et al., these classifications group genetically and culturally disparate peoples, across wide modern and ancestral geographic ranges, at different scales, under single umbrella terms.



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

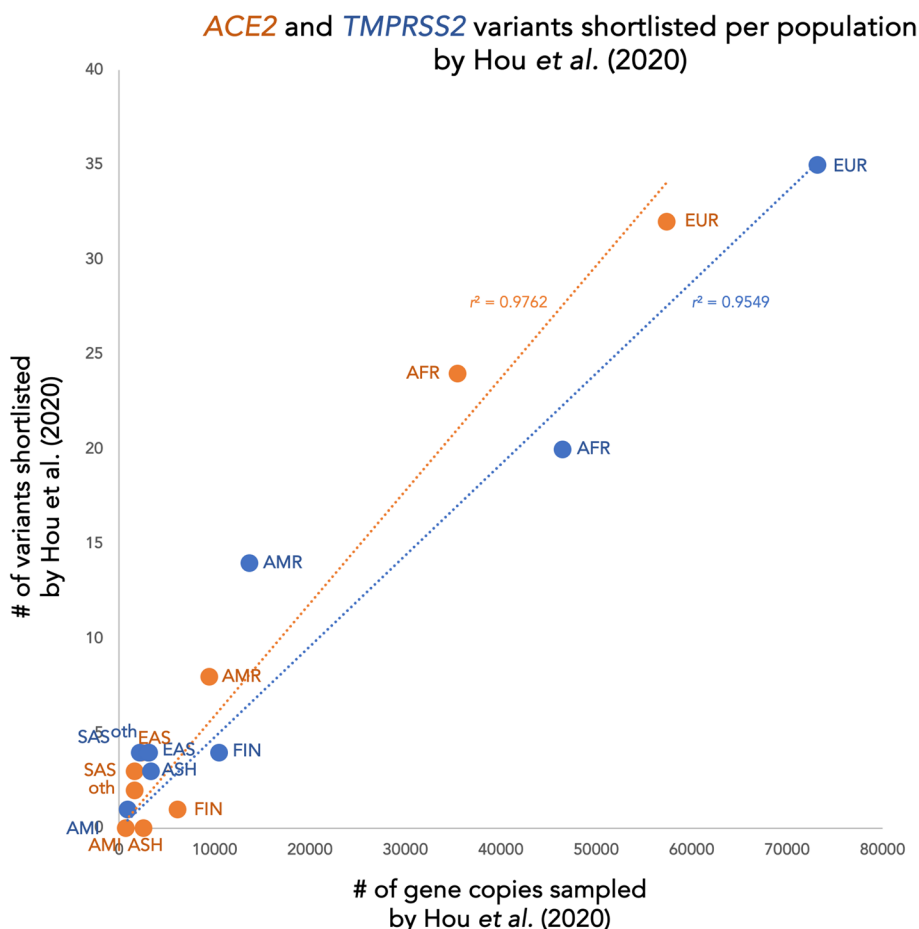


Fig. 1 Population-specific variant tallies in Hou et al. [6] reflect lopsided sampling. Scatterplots of population-specific tallies (y-axis) of shortlisted variants in *ACE2* (orange) or *TMPRSS2* (blue), by sample size (x-axis; values denote maximum sampled alleles among shortlisted variant-position genotypes for that gene in gnomAD (v.3.0) + NHLBI-GO ESP6500 genotypes, as pooled by Hou et al. [6]). Datapoints mark values for African/African-American (AFR; *TMPRSS2* count excludes 1 variant (p.Pro444Leu) reported by Hou et al., but not in public data, and not consistent with reference variant at given protein residue); Amish (AMI); Ashkenazi (ASH); east Asian (EAS); south Asian (SAS); non-Finnish, non-Amish, non-Ashkenazi European (EUR; *TMPRSS2* count excludes 1 variant (p.Gly6Arg) reported by Hou et al., but not in public data, and not consistent with reference variant at given protein residue); Finnish (FIN in Hou et al. [6]); Latino/Admixed American (AMR; *ACE2* count includes 2 variants wrongly omitted from this population by Hou et al.); or other (oth; PNA in Hou et al. [6]; *ACE2* count excludes 2 variants wrongly tallied in this population by Hou et al.). Best-fit trends (dashed) mark origin-rooted linear regression, conservatively proxying independent (versus cumulative) discovery of potentially selection-constrained (versus selectively neutral) variants in samples from variably sized, mutually diverged populations (versus one steady-sized randomly mating population). We note that even in the contrasting case of cumulative discovery in a steady-sized population, variants under selective constraint (as Hou et al. sought to tally) tend to accrue quasi-linearly, rather than strictly logarithmically, with increasing overall sample size [8–11]

Specifically, they had pooled genotypes from >36,000 “non-Finnish European” and >23,000 “African/African-American” people, but far fewer “Amish” (450), “Ashkenazi” (1662), “East Asian” (1567), or other (<15,000) people.² As such, even if variants were uniformly

distributed across real populations, Hou et al. would likelier find a given rare variant in either of their big samples (“African/African-American” or “non-Finnish European”) than in any of their much smaller samples of other groups.

Consistent with such artifact, the number of *ACE2* or *TMPRSS2* variants Hou et al. shortlisted for a given population scales well with how many genotypes they sampled from that population (Fig. 1, origin-rooted linear

² Compounding their lopsided sampling, Hou et al. also reported adding, to the foregoing totals, duplicate copies of many non-Amish, non-Ashkenazi genotypes in particular (from the 1000 Genomes Project dataset, which their main gnomAD v3.0 data already included), further confounding any frequency estimation of individual variants.

$r^2 > 0.95$ for both genes).^{3,4} Long known in theoretical and empirical population genetics [7, 11], the sampling effect apparent in Hou et al.'s summary findings reflects a simple fact: much as counting more of a forest's birds can help document rare taxa living there, sequencing more of a population's gene copies helps document rare variants among them. While other factors shape the emergence and fate of such variants, and their rate of discovery with increasing sample size [7–10, 12, 13], reliably finding and quantifying them entails sampling from many individuals.

As such, current public genotype data inherently document more of the variants actually circulating in some populations than of those actually circulating in other populations—and allow more precise estimation, in the same best-sampled (if at all) groups, of each variant's actual frequency (as may figure in functional prediction) or absence. For example, to be 90% confident that even the least rare shortlisted *ACE2* variant in non-Finnish, non-Amish, non-Ashkenazi European sample data (p.R219C) is not actually more common in Amish, Ashkenazi, and/or Finnish populations, despite its absence in those groups' small sample data, would require sampling > 22,700 additional Amish, Ashkenazi, and Finnish copies of *ACE2*.⁵

Beyond summary tallies, none of Hou et al.'s shortlisted variants reliably proxies any population to begin with. One (*TMPRSS2* p.V160M) appears in all studied populations—and in many individuals in each—so offers scant ground to guess which population(s) a carrier comes from. All 130 other shortlisted variants appear too rare in every population to typify members of any of them (even in aggregate, their data suggest that > 96% of people in every studied population likely carry none of those rare variants (Fig. 2)).⁶ And as real populations also harbor

unsampled but functionally relevant variants, whose effects on basic protein function (let alone response to a particular virus) current heuristics cannot reliably predict [14–16], the tallies and predictions of Hou et al. do not warrant positing that *ACE2* or *TMPRSS2* (let alone both) functions worse with respect to SARS-CoV2 in any human population (let alone particular ones) than in others.

In principle, real populations may indeed differ, if modestly, in functionally relevant patterns of variation in a gene (or even genome-wide), pending their histories. All else equal, for example, big populations tend to accrue and keep more genetic variation (especially if non-harmful) than do small populations [12, 18, 19]; fast-growing populations to accrue more new rare variation in particular [9, 20]; and genome segments under strong directional selection, in a given population's environment, may (along with flanking segments coupled to them without recombination) tend to lose such rare variation in particular [10].

But as human genes thus vary in summary patterns of variation, via potentially population-distinctive histories, meaningfully comparing such patterns requires sampling well, and assessing not just which variants appear at all in data sampled from a population, but the summary distributions of their estimated frequencies (respective proportions of sampled gene copies harboring them).

To that end, frequency-sensitive summary metrics⁷ show less variation in human *ACE2*, both within and between most human populations,⁸ than for most other X-borne⁹ or autosomal [21, 22, 28, 29] human genes, limiting the extent to which populations' distinctive histories may yield disparate patterns of variation. By comparison, such well grounded summary metrics show more overall variation in human *TMPRSS2* [21]—much of it shared across populations, in varied patterns that reflect the cross-regional spread of variants old (and generally non-harmful) enough to have become common.

³ While cumulative discovery of selectively neutral variants, in a steady-sized randomly mating population, tends to grow logarithmically with sample size [7], Hou et al. instead tallied selection-labile (protein-altering) variants, non-cumulatively (i.e., via independent single samplings) from mutually diverged, variably sized, and potentially variably growing populations. The foregoing factors more conservatively suit linear than logarithmic (or other presumptive higher-order) regression [8–10].

⁴ Further, each *ACE2* variant tally, and all but two *TMPRSS2* variant tallies (those for the highly recently admixed “Admixed Latino” and “Population Not Assigned” samples), accords the 95%-confident expectation for a random sample of its respective size, without replacement, from total genotypes studied for that gene, given total shortlisted variant incidences.

⁵ Confidence bound presumes accurate estimate of variant frequency in non-Finnish, non-Amish, non-Ashkenazi Europeans, and large randomly mating populations such that sampling probabilities approximate binomial expectation.

⁶ Estimates (products of binomial probabilities) presume that variants assort randomly, independently, at sampled population-specific frequencies, in half-XX/half-XY populaces.

⁷ For example, copy-pairwise per-site mismatch probability (nucleotide diversity), Tajima's *D*, and Wright's *F_{ST}*.

⁸ Save for in mutually neighboring Baka, Bakola, and Bedzan peoples, among whom three protein-coding variants rare or absent in other central African and broader populations have become moderately common (perhaps via shared linkage to regulatory variants favored by local selection over many past generations) [21]. A smaller survey [22] that reported some metrics in European-Americans suggested excess intermediate-frequency variation; but those metrics tend to spuriously infer such skew in small samples [23]. By contrast, in the same study [21], metrics robust to sample size [23] affirmed less overall variation in *ACE2* than in other X-linked genes.

⁹ Overall, X-specific genes such as *ACE2* tend, theoretically and empirically, to vary less among people overall or within a particular population, but more (relative to such scant overall variation) between randomly chosen members of different populations, than do autosomal genes [24–27].

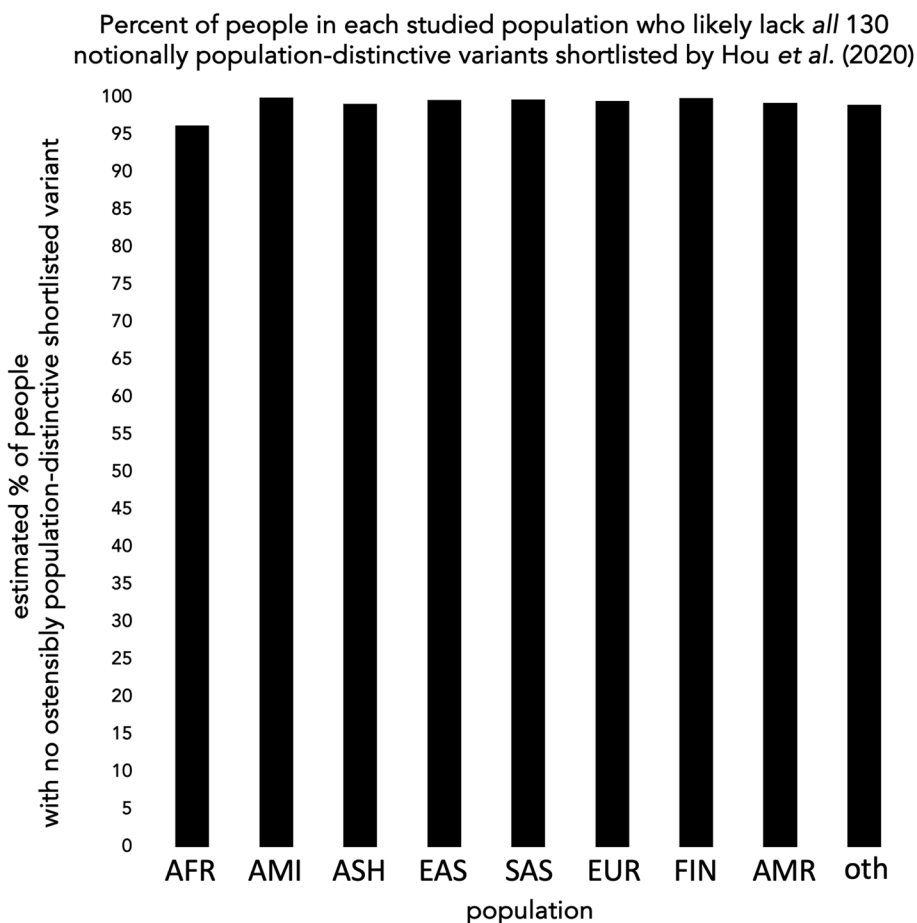


Fig. 2 Nearly everyone, in all studied populations, likely lacks all ostensibly population-distinctive variants shortlisted by Hou et al. [6]. Bar plot of estimated percentage of people in each studied population who likely have none of the 130 notionally population-distinctive (i.e., absent in sample data from at least one studied population) *ACE2* and *TMPRSS2* variants shortlisted (without empirical evidence for any effect on protein function or other physiology, and omitting many other potentially functionally relevant variants in all populations) by Hou et al. [6]. Estimates (product of binomial probabilities) presume variants assort randomly, independently, at sampled population-specific frequencies, in half-XX/half-XY populaces. AFR = African/African-American; AMI = Amish; ASH = Ashkenazi; EAS = east Asian; SAS = south Asian; EUR = non-Finnish, non-Amish, non-Ashkenazi European; FIN = Finnish; AMR = Latino/Admixed American; oth = other. Values may underestimate true minimum region-wide percentage, as (i) the least rare such variant (*ACE2* p.L731F), which most strongly suppresses the AFR estimate, appears mainly in data from over-proportionately sampled west Africa, more so than in data yet sampled from likewise populous peoples elsewhere in Africa and diaspora [17]; and (ii) any pairwise linkage among shortlisted variants would increase the proportion of people inheriting neither variant in such pairs

Importantly, even beyond the two genes’ contrasting patterns of variation, pandemic-long cohort outcomes have not shown variation in either *ACE2* or *TMPRSS2* to shape personal COVID risks nearly as significantly as variation elsewhere in our genomes—including the most strongly and significantly risk-shaping locus, on the short arm of chromosome 3; the *ABO* blood group locus on chromosome 9; and other autosomal loci [1–5, 17, 30]. Some non-protein-altering variants in *ACE2* and *TMPRSS2* have met multiple-test-stringent significance criteria for association with risks of SARS-CoV2 infection (an *ACE2* regulatory variant cluster) or severity (*TMPRSS2* intronic variant), but their significance falls

short of that evident for other loci. And among variants shortlisted by Hou et al., only one (the relatively common *TMPRSS2* p.V160M) has shown even suggestive (not multiple-test-stringent) evidence for association with any COVID risk [30–33]—while broader tests, tuned and powered specifically to detect rare variant association per se in clinically characterized population cohorts, have not implicated shortlisted or other rare protein-coding variation in either gene in COVID risks [34, 35].

Moreover, across all COVID-implicated loci, risk-informative variants differ in their population distributions and inferred effects, in many cases in partial counter-balance to one another. The *ACE2*-regulatory

rs190509934C variant, for example, associates significantly with below-average risk of SARS-CoV2 infection (and suggestively with below-average risk of severe COVID), but appears least rare in a studied broad population (south Asian) in which other loci harbor variant loads most strongly associated with above-average risk of infection (and severe disease) [36, 37].

Altogether, such mixes of above- and below-average risk-associated variants in human genomes worldwide leave a broad range of risk-relevant personal variant load within every studied human population [38–41]. And those loads, in turn, explain < 10% of personal variability in COVID risks evident, so far, in clinically characterized cohorts [39], while other factors, such as age, background health, and immune exposures, show far stronger effects.¹⁰ As such, now-ample data suggest that COVID incidence and severity likely differ among human populations far less by genetics than by factors such as age structure, past and ongoing immune exposures, comorbidity prevalences, and access to effective health interventions [4, 17, 30, 39, 40, 45–52]—and pointedly, do not support speculation that variation in *ACE2* or *TMPRSS2* drives net population disparity in genetically attributable (or overall) COVID risks.¹¹

Conclusions

Speculating early in the pandemic, on potential COVID-relevance of variation in two of humanity's many genes, Hou et al. understandably settled for predictive heuristics in lieu of clinical data. But in tallying shortlisted *ACE2* and *TMPRSS2* variants among populations, they mistook

¹⁰ Even for severe COVID (Hou et al.'s focus), sensitive liability estimates [42] at conservatively estimated population prevalences [43] were < 10% [30]. One non-genetic (pedigree-based) survey suggested possibly greater heritability, still surpassed by inferred effects of residence environment alone, but in lacking genetic data, could not control for close-kin or/and family size biases in exposures within, between, and beyond residences (e.g., in room-sharing, social/care visitation, and/or similarity of work), or for family-wise differences in age or non-residence environmental exposures [44]. And on the specific question of population (versus personal or family) disparity in inborn severe COVID risk, early speculation from sparsely typed small samples [38] has not borne out in well powered cohort data [39, 45].

¹¹ Specifically, empirically estimated variance in population-mean genomically attributable risk (the summed product of significantly risk-associated variants' sampling-inferred genotype frequencies and empirically estimated additive risk contributions [36]), for the focal outcome of Hou et al. (severe COVID) [6], in the same populations they studied, is slightly greater (~0.02089) when omitting *ACE2* variation than when including it (~0.02004) or when also including *TMPRSS2* variation (0.01822) [37]. These estimates warrant much caution, as variant-attributable effects themselves vary among populations [30] (plausibly via population-stratified genome-wide variation, local haplotype structure, environmental modifiers [53–56], and other factors), but suggest that, all else equal, *ACE2* and (if relevant) *TMPRSS2* variation unlikely exacerbates any net disparity in such additively genomically attributable COVID risk among studied human populations.

a sample size artifact as if evidence of population differences—and tried to proxy population-representative gene function by tallying shortlisted variants found at all, instead of summing empirically estimated, genotype frequency-scaled effects. Though their shortlist offered well intended, if unvalidated, candidates for early-pandemic study, genome-wide empirical insights have eclipsed its utility—while leaving, unaddressed, their artifactual summary claims. And those claims, in turn, have drawn credulous citation in public discourse, hindering understanding of COVID risks and of human genetic diversity itself [57–59].

While pitfalls of methodology and interpretation have long plagued basic and clinical research [60–67], public discourse invoking Hou et al. [6] highlights how platforms to usefully share and discuss such research can also virally spread faulty inferences missed by authors and reviewers, misleading not just researchers and clinicians, but also lay-people who may rely on published science in personal, professional, family, and civic decisions [64]. As such, we hope that correcting basic errors that misled Hou et al., and those citing their work, helps right the record on human genomic variation in *ACE2*, *TMPRSS2*, and loci more informative of COVID risks—and, further, encourages critical stringency in the interpretation of population genetic and other data, amid efforts [68, 69] to more promptly and soundly validate research published, and subsequently invoked, in and between societal crises.

Abbreviations

<i>ACE2</i>	Angiotensin converting enzyme 2 [gene]
AMI	Amish
AMR	[Latina/o] Admixed American
ASH	Ashkenazi
COVID	Coronavirus disease
DNA	Deoxyribonucleic acid
EAS	East Asian
EUR	[Non-Amish, non-Ashkenazi, non-Finnish] European
FIN	Finnish
oth	Other [population(s)]
SARS-CoV2	Severe acute respiratory syndrome coronavirus 2
SAS	South Asian
<i>TMPRSS2</i>	Transmembrane serine protease 2 [gene]

Acknowledgements

We thank Arjun Biddanda, Charleston Chiang, Mark Iles, Ann Morgan, Richard Sever, reviewers, and preprint commenters for helpful insights and suggestions; and Konrad Karczewski for guidance in accessing the gnomAD v3.0 data used by Hou et al.

Authors' contributions

NMP analyzed data, and conceived and drafted the manuscript; JN critically refined analysis and manuscript with NMP. Both authors read and approved the final manuscript.

Authors' Twitter handles

@genomenathan (NMP), @jnovembre (JN).

Funding

NIH NIGMS grant R35 GM149521 (JN).

Availability of data and materials

Datasets analyzed in this study are available at:

- <https://cloud.google.com/life-sciences/docs/resources/public-datasets/gnomad> (gnomAD v3.0 data includes 1000Genomes data, as analyzed by Hou et al.)
- https://esp.gs.washington.edu/drupal/dbGaP_Releases
- <https://www.covid19hg.org/>

Declarations**Ethics approval and consent to participate**

N/A.

Consent for publication

N/A.

Competing interests

The authors declare that they have no competing interests.

Received: 5 December 2023 Accepted: 22 July 2024

Published online: 26 August 2024

References

1. Severe Covid-19 GWAS Group, Ellinghaus D, Degenhardt F, Bujanda L, Buti M, Alballos A, et al. Genomewide association study of severe COVID-19 with respiratory failure. *N Engl J Med*. 2020;383:1522–34.
2. Zhao J, Yang Y, Huang H, Li D, Gu D, Lu X, et al. Relationship between the ABO blood group and the COVID-19 susceptibility. *Clin Infect Dis*. 2021;73(2):328–31.
3. Pairo-Castineira E, Clohisey S, Klaric L, Bretherick AD, Rawlik K, Pasko D, et al. Genetic mechanisms of critical illness in COVID-19. *Nature*. 2021;591:92–8.
4. COVID-19 Host Genetics Initiative. Mapping the human genetic architecture of COVID-19. *Nature*. 2021;600:472–7.
5. Kousathanas A, Pairo-Castineira E, Rawlik K, Stuckey A, Odhams CA, Walker S, et al. Whole-genome sequencing reveals host factors underlying critical COVID-19. *Nature*. 2022;607:97–103.
6. Hou Y, Zhao J, Martin W, Kallianpur A, Chung MK, Jehi L, et al. New insights into genetic susceptibility of COVID-19: an ACE2 and TMPRSS2 polymorphism analysis. *BMC Med*. 2020;18:216.
7. Ewens WJ. The sampling theory of selectively neutral alleles. *Theor Popul Biol*. 1972;3:87–112.
8. Desai MM, Nicolaisen LE, Walczak AM, Plotkin JB. The structure of allelic diversity in the presence of purifying selection. *Theor Popul Biol*. 2012;81:144–57.
9. Slatkin M, Hudson RR. Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations. *Genetics*. 1991;129:555–62.
10. Przeworski M, Coop G, Wall JD. The signature of positive selection on standing genetic variation. *Evolution*. 2005;59:2312–23.
11. Kiezun A, Garimella K, Do R, Stitzel NO, Neale BM, McLaren PJ, et al. Exome sequencing and the genetic basis of complex traits. *Nat Genet*. 2012;44:623–30.
12. Fisher RA. The genetical theory of natural selection. 1930.
13. Mathieson I, Reich D. Differences in the rare variant spectrum among human populations. *PLoS Genet*. 2017;13:e1006581.
14. Niroula A, Vihinen M. How good are pathogenicity predictors in detecting benign variants? *PLoS Comput Biol*. 2019;15:e1006481.
15. Gunning AC, Fryer V, Fasham J, Crosby AH, Ellard S, Baple E, et al. Assessing performance of pathogenicity predictors using clinically-relevant variant datasets. *bioRxiv*. 2020;2020.02.06.937169.
16. de Garcia FAO, de Andrade ES, Palmero EI. Insights on variant analysis in silico tools for pathogenicity prediction. *Front Genet*. 2022;13:1010327.
17. COVID-19 Host Genetics Initiative. A first update on mapping the human genetic architecture of COVID-19. *Nature*. 2022;608:E1–10.
18. Simons YB, Turchin MC, Pritchard JK, Sella G. The deleterious mutation load is insensitive to recent population history. *Nat Genet*. 2014;46:220–4.
19. Henn BM, Botigué LR, Bustamante CD, Clark AG, Gravel S. Estimating the mutation load in human genomes. *Nat Rev Genet*. 2015;16:333–43.
20. Nelson MR, Wegmann D, Ehm MG, Kessler D, St Jean P, Verzilli C, et al. An abundance of rare functional variants in 202 drug target genes sequenced in 14,002 people. *Science*. 2012;337:100–4.
21. Zhang C, Verma A, Feng Y, Dos Reis Melo MC, McQuillan M, Hansen M, et al. Impact of natural selection on global patterns of genetic variation, and association with clinical phenotypes, at genes involved in SARS-CoV-2 infection. *Res Sq*. 2021. <https://doi.org/10.21203/rs.3.rs-673011/v1>.
22. Akey JM, Eberle MA, Rieder MJ, Carlson CS, Shriver MD, Nickerson DA, et al. Population history and natural selection shape patterns of genetic variation in 132 genes. *PLoS Biol*. 2004;2:e286.
23. Subramanian S. The effects of sample size on population genomic analyses—implications for the tests of neutrality. *BMC Genomics*. 2016;17:123.
24. Vicoso B, Charlesworth B. Evolution on the X chromosome: unusual patterns and processes. *Nat Rev Genet*. 2006;7:645–53.
25. Akey JM, Zhang G, Zhang K, Jin L, Shriver MD. Interrogating a high-density SNP map for signatures of natural selection. *Genome Res*. 2002;12:1805–14.
26. Lambert CA, Connelly CF, Madeoy J, Qiu R, Olson MV, Akey JM. Highly punctuated patterns of population structure on the X chromosome and implications for African evolutionary history. *Am J Hum Genet*. 2010;86:34–44.
27. Elhaik E. Empirical distributions of F_{ST} from large-scale human polymorphism data. *PLoS One*. 2012;7:e49837.
28. Suryamohan K, Diwanji D, Stawiski EW, Gupta R, Miersch S, Liu J, et al. Human ACE2 receptor polymorphisms and altered susceptibility to SARS-CoV-2. *Commun Biol*. 2021;4:475.
29. Keinan A, Reich D. Human population differentiation is strongly correlated with local recombination rate. *PLoS Genet*. 2010;6:e1000886.
30. COVID-19 Host Genetics Initiative. A second update on mapping the human genetic architecture of COVID-19. *Nature*. 2023;621:E7–26.
31. Li W, Zhang C, Sui J, Kuhn JH, Moore MJ, Luo S, et al. Receptor and viral determinants of SARS-coronavirus adaptation to human ACE2. *EMBO J*. 2005;24(8):1634–43. <https://doi.org/10.1038/sj.emboj.7600640>.
32. David A, Parkinson N, Peacock TP, Pairo-Castineira E, Khanna T, Cobat A, et al. A common TMPRSS2 variant has a protective effect against severe COVID-19. *Curr Res Transl Med*. 2022;70:103333.
33. Yaghoobi A, Lord JS, Rezaiezhadeh JS, Yekaninejad MS, Amini M, Izadi P. TMPRSS2 polymorphism (rs12329760) and the severity of the COVID-19 in Iranian population. *PLoS One*. 2023;18:e0281750.
34. Kosmicki JA, Horowitz JE, Banerjee N, Lanche R, Marcketta A, Maxwell E, et al. Pan-ancestry exome-wide association analyses of COVID-19 outcomes in 586,157 individuals. *Am J Hum Genet*. 2021;108:1350–5.
35. Butler-Laporte G, Povysil G, Kosmicki JA, Cirulli ET, Drivas T, Furini S, et al. Exome-wide association study to identify rare variants influencing COVID-19 outcomes: results from the Host Genetics Initiative. *PLoS Genet*. 2022;18:e1010367.
36. Covid-19 HGI Browser. <https://app.covid19hg.org/variants>. Accessed 4 Aug 2023.
37. gnomAD. <https://gnomad.broadinstitute.org/>. Accessed 23 Oct 2023.
38. Marçalo R, Neto S, Pinheiro M, Rodrigues AJ, Sousa N, Santos MAS, et al. Evaluation of the genetic risk for COVID-19 outcomes in COPD and differences among worldwide populations. *PLoS One*. 2022;17:e0264009.
39. Farooqi R, Kooner JS, Zhang W. Associations between polygenic risk score and COVID-19 susceptibility and severity across ethnic groups: UK Biobank analysis. *BMC Med Genomics*. 2023;16:150.
40. Zhu D, Zhao R, Yuan H, Xie Y, Jiang Y, Xu K, et al. Host genetic factors, comorbidities and the risk of severe COVID-19. *J Epidemiol Glob Health*. 2023;13:279–91.
41. Córdova-Palomera A, Siffel C, DeBoever C, Wong E, Diogo D, Szalma S. Assessing the potential of polygenic scores to strengthen medical risk prediction models of COVID-19. *PLoS One*. 2023;18:e0285991.
42. Ojavee SE, Kutalik Z, Robinson MR. Liability-scale heritability estimation for biobank studies of low-prevalence disease. *Am J Hum Genet*. 2022;109:2009–17.
43. Anesi GL. COVID-19: epidemiology, clinical features, and prognosis of the critically ill adult. In: Manaker S, Finlay G, editors. *UpToDate*. Waltham: UpToDate; 2024.

44. Brown KL, Ramlall V, Zietz M, Gisladdottir U, Tatonetti NP. Estimating the heritability of SARS-CoV-2 susceptibility and COVID-19 severity. *Nat Commun.* 2024;15:367.
45. Crossfield SSR, Chaddock NJM, Iles MM, Pujades-Rodriguez M, Morgan AW. Interplay between demographic, clinical and polygenic risk factors for severe COVID-19. *Int J Epidemiol.* 2022;51:1384–95.
46. Romero Starke K, Reissig D, Peterleit-Haack G, Schmauder S, Nienhaus A, Seidler A. The isolated effect of age on the risk of COVID-19 severe outcomes: a systematic review with meta-analysis. *BMJ Glob Health.* 2021;6:e006434.
47. Derqui N, Koycheva A, Zhou J, Pillay TD, Crone MA, Hakki S, et al. Risk factors and vectors for SARS-CoV-2 household transmission: a prospective, longitudinal cohort study. *Lancet Microbe.* 2023;4:e397–408.
48. Antos A, Kwong ML, Balmorez T, Villanueva A, Murakami S. Unusually high risks of COVID-19 mortality with age-related comorbidities: an adjusted meta-analysis method to improve the risk assessment of mortality using the comorbid mortality data. *Infect Dis Rep.* 2021;13:700–11.
49. Gazit S, Shlezinger R, Perez G, Lotan R, Peretz A, Ben-Tov A, et al. Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) naturally acquired immunity versus vaccine-induced immunity, reinfections versus breakthrough infections: a retrospective cohort study. *Clin Infect Dis.* 2022;75:e545–51.
50. Madhi SA, Kwatra G, Myers JE, Jassat W, Dhar N, Mukendi CK, et al. Population immunity and COVID-19 severity with omicron variant in South Africa. *N Engl J Med.* 2022;386:1314–26.
51. Mongin D, Bürgisser N, Laurie G, Schimmel G, Vu D-L, Cullati S, et al. Effect of SARS-CoV-2 prior infection and mRNA vaccination on contagiousness and susceptibility to infection. *Nat Commun.* 2023;14:5452.
52. Hammond J, Leister-Tebbe H, Gardner A, Abreu P, Bao W, Wisemandle W, et al. Oral nirmatrelvir for high-risk, nonhospitalized adults with COVID-19. *N Engl J Med.* 2022;386:1397–408.
53. Martin AR, Gignoux CR, Walters RK, Wojcik GL, Neale BM, Gravel S, et al. Human demographic history impacts genetic risk prediction across diverse populations. *Am J Hum Genet.* 2017;100:635–49.
54. Martin AR, Kanai M, Kamatani Y, Okada Y, Neale BM, Daly MJ. Clinical use of current polygenic risk scores may exacerbate health disparities. *Nat Genet.* 2019;51:584–91.
55. Coop G. Reading tea leaves? Polygenic scores and differences in traits among groups. *arXiv [q-bio.GN].* 2019.
56. Polygenic Risk Score Task Force of the International Common Disease Alliance. Responsible use of polygenic risk scores in the clinic: potential benefits, risks and gaps. *Nat Med.* 2021;27:1876–84.
57. Hodgman L, Kennedy RF Jr. denies comments on 'ethnically targeted' COVID-19 were anti-Semitic. *POLITICO.* <https://www.politico.com/news/2023/07/15/rfk-jr-covid-19-ethnically-targeted-00106478>. Accessed 6 Aug 2023.
58. Kennedy RF Jr. [...]a 2021 study of the COVID-19 virus [sic] shows that COVID-19 appears to disproportionately affect certain races [...] That study is here: <https://pubmed.ncbi.nlm.nih.gov/32664879/> Twitter. 2023. <https://twitter.com/RobertKennedyJr/status/1680227322509635595>. Accessed 4 Aug 2023.
59. Owens C, Kennedy RF Jr. mentioned an actual scientific study which revealed that Chinese people & Ashkenazi Jews were the most immune to COVID-19 [...]. <https://bmcmmedicine.biomedcentral.com/articles/10.1186/s12916-020-01673-z>... Twitter. 2023. <https://twitter.com/RealCandaceO/status/1680371443446722561>. Accessed 6 Aug 2023.
60. Renschmidt C, Wichmann O, Harder T. Frequency and impact of confounding by indication and healthy vaccinee bias in observational studies assessing influenza vaccine effectiveness: a systematic review. *BMC Infect Dis.* 2015;15:429.
61. Border R, O'Rourke S, de Candia T, Goddard ME, Visscher PM, Yengo L, et al. Assortative mating biases marker-based heritability estimators. *Nat Commun.* 2022;13:660.
62. Høeg TB, Duriseti R, Prasad V. Potential, "healthy vaccinee bias" in a study of BNT162b2 vaccine against COVID-19. *N Engl J Med.* 2023;389:284–6.
63. Herrera-Perez D, Haslam A, Crain T, Gill J, Livingston C, Kaestner V, et al. A comprehensive review of randomized clinical trials in three medical journals reveals 396 medical reversals. *Elife.* 2019;8:e45183.
64. Høeg TB, Ladhani S, Prasad V. How methodological pitfalls have created widespread misunderstanding about long COVID. *BMJ Evid Based Med.* 2023. <https://doi.org/10.1136/bmjebm-2023-112338>.
65. Slutsky DJ. Statistical errors in clinical studies. *J Wrist Surg.* 2013;2:285–7.
66. Günel Karadeniz P, Uzabacı E, Atış Kuyuk S, Kaskır Kesin F, Can FE, Seçil M, et al. Statistical errors in articles published in radiology journals. *Diagn Interv Radiol.* 2019;25:102–8.
67. Dowty JG, Win AK, Buchanan DD, Lindor NM, Macrae FA, Clendenning M, et al. Cancer risks for MLH1 and MSH2 mutation carriers. *Hum Mutat.* 2013;34:490–7.
68. Simundic A-M, Nikolac N. Statistical errors in manuscripts submitted to *Biochimica Medica* journal. *Biochem Med.* 2009;19:294–300.
69. Waltman L, Kaltenbrunner W, Pinfield S, Woods HB. How to improve scientific peer review: four schools of thought. *Learn Publ.* 2023;36:334–47.

Publisher's Note

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