



OPEN

Genetics of smoking and risk of clonal hematopoiesis

Michael G. Levin^{1,2,14}, Tetsushi Nakao^{3,4,5,6,14}, Seyedeh M. Zekavat^{3,7}, Satoshi Koyama³, Alexander G. Bick⁸, Abhishek Niroula^{5,9,10}, Benjamin Ebert^{5,11,13}, Scott M. Damrauer^{2,12} & Pradeep Natarajan^{3,4,13}✉

Clonal hematopoiesis of indeterminate potential (CHIP) and mosaic chromosomal alterations (mCAs) represent two forms of clonal hematopoiesis where clones bearing expanded somatic mutations have been linked to both oncologic and non-oncologic clinical outcomes including atherosclerosis and all-cause mortality. Epidemiologic studies have highlighted smoking as an important driver of somatic mutations across multiple tissues. However, establishing the causal role of smoking in clonal hematopoiesis has been limited by observational study designs, which may suffer from confounding and reverse-causality. We performed two complementary analyses to investigate the role of smoking in mCAs and CHIP. First, using an observational study design among UK Biobank participants, we confirmed strong associations between smoking and mCAs. Second, using two-sample Mendelian randomization, smoking was strongly associated with mCA but not with CHIP. Overall, these results support a causal association between smoking and mCAs and suggest smoking may variably shape the fitness of clones bearing somatic mutations.

Population-based human genetic analyses of asymptomatic individuals have shown that clonally-expanded acquired mutations in the blood are increasingly common with age^{1–5}. Clonal hematopoiesis of indeterminate potential (CHIP) is characterized by hematologic expansion of clones bearing pathogenic single nucleotide polymorphisms or small insertions/deletions, typically in *DNMT3A*, *TET2*, *ASXL1*, *JAK2*, and other oncogenic genes, detected by next-generation sequencing⁶. Mosaic chromosomal alterations (mCAs) are characterized by expanded large structural variants, 50–249 Mb, often detected by genome-wide array genotyping³. CHIP is strongly linked to myeloid malignancy and atherosclerotic cardiovascular disease^{2,7–9}, while mCAs are strongly linked to lymphoid malignancies, myeloproliferative neoplasms, and severe infections^{10–12}.

While tobacco smoking is an established strong mutagen¹³, observational associations between smoking and somatic mutations may be limited by residual confounding or reverse-causality, precluding causal inference. Smoking has been linked to mosaic loss-of-Y chromosome (LOY)¹⁴, as well as loss-of-function mutations in *ASXL1*, a CHIP-associated tumor-suppressor gene¹⁵. However, the broader relationships between smoking and variants indicative of clonal hematopoiesis, and the extent to which residual confounding from correlated lifestyle factors may explain the observed relationships is not well understood.

Genome-wide association studies (GWAS) have recently identified common germline genetic variants associated with smoking, mCA, and CHIP, enabling causal inference across these traits within the Mendelian randomization (MR) framework^{12,16–18}. Because genetic variants are randomly allocated at conception, these variants may be used as instrumental variables for MR, which under certain assumptions enables a natural experiment mimicking a randomized controlled trial¹⁹. Compared with traditional observational study designs, MR studies

¹Division of Cardiovascular Medicine, Department of Medicine, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA. ²Corporal Michael J. Crescenz VA Medical Center, Philadelphia, PA, USA. ³Program in Medical and Population Genetics and Cardiovascular Disease Initiative, Broad Institute of Harvard and MIT, Cambridge, MA, USA. ⁴Cardiovascular Research Center, Massachusetts General Hospital, 185 Cambridge Street, CPZN 3.184, Boston, MA 02114, USA. ⁵Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, MA, USA. ⁶Division of Cardiovascular Medicine, Department of Medicine, Brigham and Women's Hospital, Boston, MA, USA. ⁷Yale University School of Medicine, New Haven, CT, USA. ⁸Division of Genetic Medicine, Department of Medicine, Vanderbilt University School of Medicine, Nashville, TN, USA. ⁹Cancer Program, Broad Institute of Harvard and MIT, Cambridge, MA, USA. ¹⁰Department of Laboratory Medicine, Lund University, Lund, Sweden. ¹¹Howard Hughes Medical Institute, Boston, MA, USA. ¹²Department of Surgery, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA. ¹³Department of Medicine, Harvard Medical School, Boston, MA, USA. ¹⁴These authors contributed equally: Michael G. Levin and Tetsushi Nakao. ✉email: pnatarajan@mgh.harvard.edu

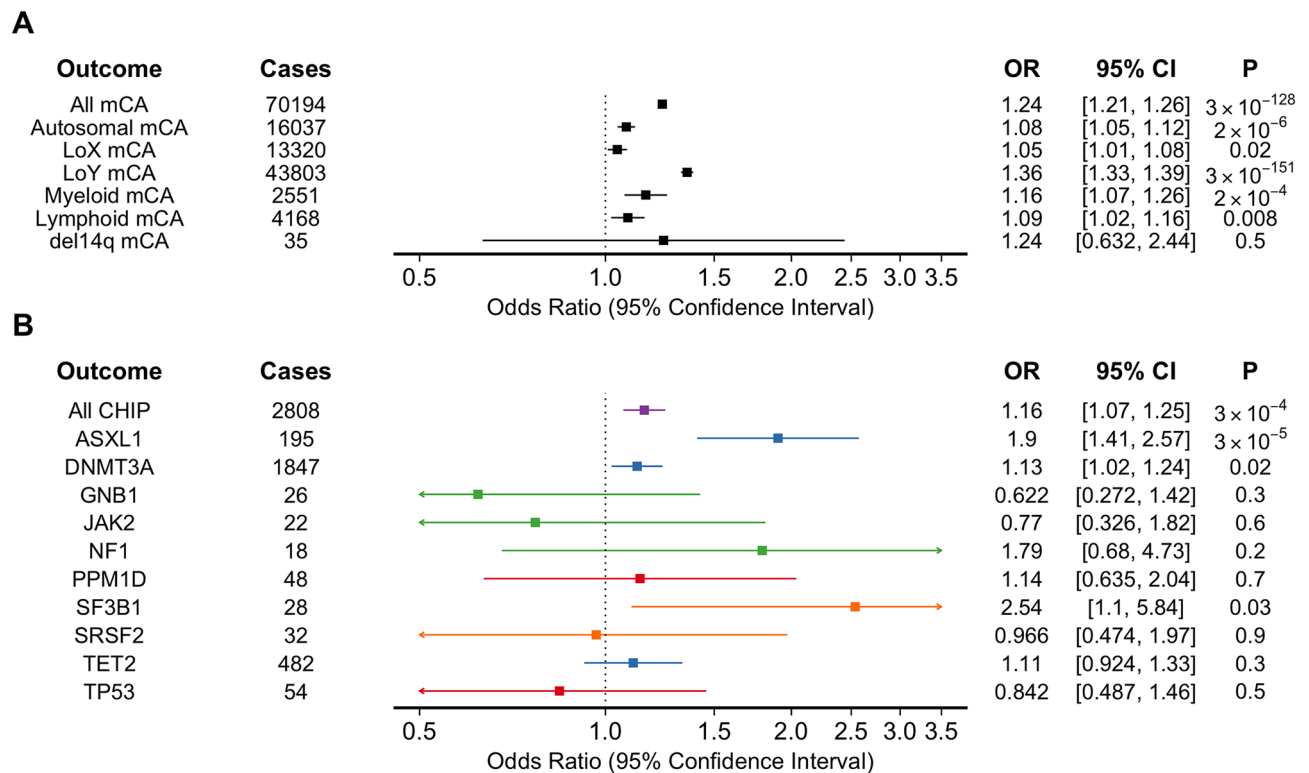


Figure 1. Observational associations between Ever smoking and somatic mutations among UK Biobank participants. Associations between ever smoking and (A) manifestations of mCA and (B) manifestations of CHIP among UK Biobank participants. Colors represent specific types of CHIP mutations (Purple = any CHIP, Blue = epigenetic, Red = DNA damage repair, Orange = splicing, Green = other). mCA = mosaic chromosomal alteration; LoX = loss-of-X chromosome; LoY = loss-of-Y chromosome; del = deletion.

may be less susceptible to confounding and reverse causality, enabling estimation of putative causal associations between exposures and outcomes¹⁹.

We aimed to (1) confirm observational associations between smoking and common manifestations of somatic mutation (mCA and CHIP), and (2) estimate putative causal associations between smoking and these outcomes within the MR framework.

In our observational analysis, mCAs were genotyped in 487,279 participants of UK Biobank who underwent genome-wide genotyping of blood DNA, and CHIP was genotyped in 49,606 participants who underwent whole-exome sequencing of blood DNA^{3,8,17}. We identified 72,176 (14.8%) individuals with any mCA, 17,108 (3.5%) with autosomal mCA, and 44,696 (20% of males) with mCA-LOY. Among the 49,606 individuals who underwent whole exome sequencing, 2888 (5.8%) had CHIP. Smoking was independently associated with any mCA (OR 1.24, 95% CI 1.21–1.26, $p = 8.55 \times 10^{-131}$), as well as subclasses: autosomal mCA (OR 1.08, 95% CI 1.05–1.12, $p = 6 \times 10^{-7}$), mCA-LOX (OR 1.05, 95% CI 1.01–1.08, $p = 0.02$), mCA-LOY (OR 1.35, 95% CI 1.32–1.39, $p = 8.16 \times 10^{-154}$), Myeloid mCA (OR 1.16, 95% CI 1.07–1.26, $p = 2 \times 10^{-4}$), and Lymphoid mCA (OR 1.09, 95% CI 1.02–1.16, $p = 0.008$) (Fig. 1). There was substantial heterogeneity in the associations between smoking and mCA subclasses ($I^2 = 97.7$, Cochran $Q = 218.5$, $p = 3.06 \times 10^{-45}$), with the association between smoking and overall mCA likely driven by mCA-LOY. Smoking was also associated with CHIP overall (OR 1.15, 95% CI 1.06–1.24, $p = 4.64 \times 10^{-4}$) (Fig. 1). In secondary analyses we tested for associations between smoking and CHIP by gene (e.g., *DNMT3A*, *TET2*, *ASXL1*, etc.). However, there was significant heterogeneity when considering associations between smoking and mutations in specific CHIP-associated genes ($I^2 = 54.4$, Cochran $Q = 19.7$, $p = 0.02$). These observational results were similar in sensitivity analyses accounting for alcohol consumption. Broadly, these findings are consistent with prior reports correlating smoking status with increased risk of somatic mutations^{4,13,14,20,21}.

Next, we performed two-sample MR using summary statistics to evaluate the causal effects of smoking on somatic mutation outcomes. As a genetic proxy for smoking, we considered up to 119 independent genetic variants associated with smoking at genome-wide significance ($p < 5 \times 10^{-8}$)¹⁶. The F-statistic for our genetic instrument ranged from 21.78 to 196 (mean 48.5), suggesting the analysis was not limited by weak-instrument bias. In the primary inverse variance-weighted MR analysis, smoking was strongly associated with mCAs (OR 1.44, 95% CI 1.16–1.79, $p = 8 \times 10^{-4}$), and mCA-LOY (OR 1.06, 95% CI 1.04–1.08, $p = 1 \times 10^{-8}$) (Fig. 2). We did not detect a significant association between smoking and CHIP (Transethnic OR 1.01, 95% CI 0.74–1.37, $p = 1$; European [EUR] OR 0.70, 95% CI 0.48–1.02, $p = 0.06$), however these confidence intervals do not exclude potentially meaningful effects (Fig. 2). Results were similar using alternative MR methods which each make different

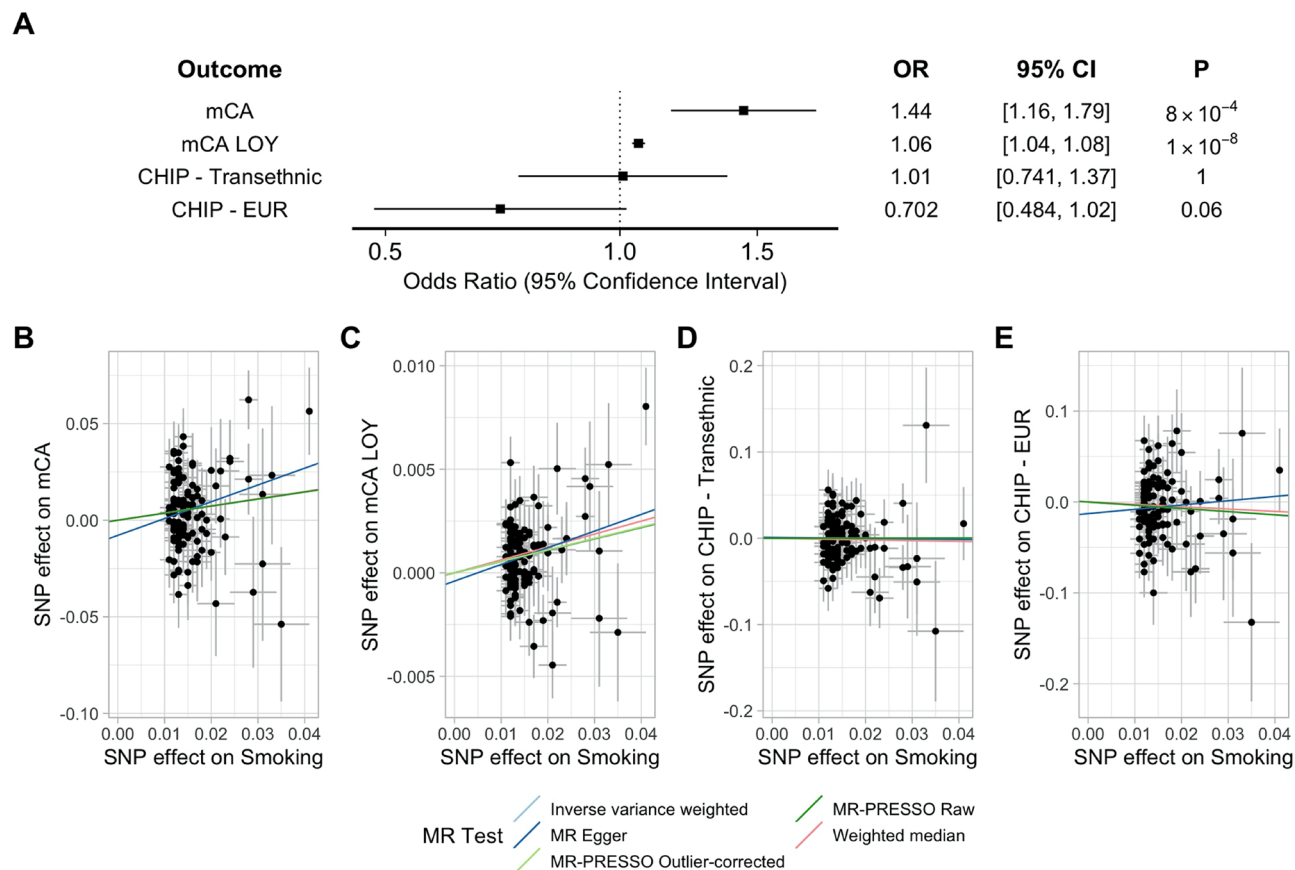


Figure 2. Mendelian Randomization associations between smoking and somatic mutations. Results of MR testing the associations between smoking and mCA and CHIP outcomes. (A) Inverse-variance weighted MR results. (B–E) MR results testing the association between smoking and each outcome using alternative MR methods. EUR = European-ancestry.

assumptions about outliers and pleiotropy (Fig. 2). The MR-Egger bias intercept test did not detect evidence of directional pleiotropy ($p > 0.05$ for all comparisons).

Overall, these results are consistent with smoking as a causal risk factor for mCA. Although smoking was associated with CHIP in our observational analysis, we did not detect an association in our MR analysis; whether smoking represents a causal risk factor for CHIP, including for specific genes, will require further study. Our findings suggest that smoking variably shapes the fitness of distinct somatic mutations, and efforts to reduce smoking would be expected to reduce the burden of the downstream consequences of somatic mutation.

This study has both strengths and limitations. In this case, the MR framework allowed us to leverage the natural randomization in the distribution of genetic variants to estimate the causal associations between smoking and manifestations of somatic mutations. By utilizing large GWAS, we were able to consider thousands of mCA and CHIP cases, which would otherwise require large, extended trials to accrue. Although we were able to identify strong associations between smoking and mCA, whether mCA mediates some of the adverse consequences of smoking will require further study. Similarly, we do not provide specific insights regarding the mechanisms by which smoking influences fitness of clones bearing somatic mutations. Given the low heritability of CHIP, and small sample size of CHIP cases limiting power, we cannot exclude meaningful associations between smoking and CHIP by gene in an MR framework. Smoking has been linked to the fitness of clones harboring mutations in particular driver mutations in the setting of lung cancer¹³, and whether similar findings extend to CHIP remains an important avenue for future study. With the growing availability of population-scale human genotype and sequence data, linking smoking to particular mutational drivers of clonal fitness should become increasingly tractable.

In conclusion, we confirm strong observational associations between smoking and somatic mutation, with MR analyses consistent with smoking as a causal risk factor for mCA. Whether smoking causes CHIP, and the specific mechanisms by which smoking influences fitness will require further study.

Methods

Observational analysis. Whole exome sequencing and array genotyping have been previously described in the UK Biobank, a population-based volunteer biobank recruited 2006–2010²². mCA was determined among 479,810 participants without hematologic malignancy who underwent genome-wide genotyping, and CHIP was

determined among up to 48,966 participants without hematologic malignancy who underwent whole-exome sequencing, as previously described^{3,8,17}. We tested for the association between ever smoking (defined by UK Biobank unique data identifier 20116-0.0) and somatic mutation outcomes (mCA, autosomal mCA, mCA-LOY [loss-of-Y chromosome], mCA-LOX [loss-of-X chromosome], lymphoid mCA, myeloid mCA, del14q, and CHIP) using logistic regression adjusted for age, age², sex, sequencing batch, and 15 genetic principal components. Myeloid and lymphoid mCAs were identified based on their association with myeloid and lymphoid malignancies²³. In a secondary analysis we tested for associations between smoking and CHIP by gene (e.g., *DNMT3A*, *TET2*, *ASXL1*, etc.). In a sensitivity analysis we included self-reported alcohol use as an additional covariate, given strong epidemiologic correlations with smoking. This work was performed using UK Biobank Application #7089. The UK Biobank obtained IRB approval from the North West Multi-centre Research Ethics Committee (approval number: 11/NW/0382), and participants provided informed consent.

Mendelian randomization analysis. We performed two-sample MR using summary statistics utilizing the *TwoSampleMR* package in R. As genetic instruments for smoking, we considered independent ($r^2 < 0.001$, distance > 10,000 kb) genetic variants associated ($p < 5 \times 10^{-8}$) with the lifetime smoking index, a previously-validated continuous measure of lifetime smoking, derived among up to 462,690 UK Biobank participants¹⁶. For each genetic variant associated with smoking, we extracted the corresponding effects for each outcome from GWAS of mCA (up to 767,891 unrelated multi-ancestry individuals without hematological cancer from UK Biobank, Biobank Japan, Mass General Brigham Biobank, and FinnGen), mCA-LOY (up to 205,011 male participants of UK Biobank), and CHIP (97,691 participants of TOPMed)^{12,17,18}. We calculated F-statistics for each exposure-outcome pair to assess for weak-instrument bias¹⁹. For the primary analysis we applied the inverse variance-weighted method, but considered weighted median, MR-Egger, and MR-PRESSO in sensitivity analyses, as these methods make different assumptions about the presence of pleiotropy and outliers²⁴.

Analyses were performed using R 4.0.3 (R Foundation for Statistical Computing, Vienna, Austria). For the primary observational and MR analyses, p values < 0.05 (after accounting for multiple comparisons using Bonferroni adjustment) were considered significant. For secondary analyses, $p < 0.05$ was considered significant. All methods were carried out in accordance with the relevant guidelines and regulations.

Received: 4 November 2021; Accepted: 25 March 2022

Published online: 04 May 2022

References

- Xie, M. *et al.* Age-related mutations associated with clonal hematopoietic expansion and malignancies. *Nat. Med.* **20**(12), 1472–1478 (2014).
- Jaiswal, S. *et al.* Clonal hematopoiesis and risk of atherosclerotic cardiovascular disease. *N. Engl. J. Med.* **377**(2), 111–121 (2017).
- Loh, P.-R. *et al.* Insights into clonal haematopoiesis from 8,342 mosaic chromosomal alterations. *Nature* **559**(7714), 350–355 (2018).
- Natarajan, P., Jaiswal, S. & Kathiresan, S. Clonal hematopoiesis. *Circ. Genom. Precis. Med.* <https://doi.org/10.1161/CIRCGEN.118.001926> (2018).
- Genovese, G. *et al.* Clonal hematopoiesis and blood-cancer risk inferred from blood DNA sequence. *N. Engl. J. Med.* **371**(26), 2477–2487 (2014).
- Steensma, D. P. *et al.* Clonal hematopoiesis of indeterminate potential and its distinction from myelodysplastic syndromes. *Blood* **126**(1), 9–16 (2015).
- Bick, A. G. *et al.* Genetic interleukin 6 signaling deficiency attenuates cardiovascular risk in clonal hematopoiesis. *Circulation* **141**(2), 124–131 (2020).
- Honigberg, M. C. *et al.* Premature menopause, clonal hematopoiesis, and coronary artery disease in postmenopausal women. *Circulation* **143**(5), 410–423 (2021).
- Fuster, J. J. *et al.* Clonal hematopoiesis associated with TET2 deficiency accelerates atherosclerosis development in mice. *Science* **355**(6327), 842–847 (2017).
- Loh, P.-R., Genovese, G. & McCarroll, S. A. Monogenic and polygenic inheritance become instruments for clonal selection. *Nature* **584**(7819), 136–141 (2020).
- Terao, C. *et al.* Chromosomal alterations among age-related haematopoietic clones in Japan. *Nature* **584**(7819), 130–135 (2020).
- Zekavat, S. M. *et al.* Hematopoietic mosaic chromosomal alterations increase the risk for diverse types of infection. *Nat. Med.* **27**, 1012–1024 (2021).
- Yoshida, K. *et al.* Tobacco smoking and somatic mutations in human bronchial epithelium. *Nature* **578**(7794), 266–272 (2020).
- Dumanski, J. P. *et al.* Smoking is associated with mosaic loss of chromosome Y. *Science* **347**(6217), 81–83 (2015).
- Bolton, K. L. *et al.* Cancer therapy shapes the fitness landscape of clonal hematopoiesis. *Nat. Genet.* **52**(11), 1219–1226 (2020).
- Wootton, R. E. *et al.* Evidence for causal effects of lifetime smoking on risk for depression and schizophrenia: A Mendelian randomisation study. *Psychol. Med.* <https://doi.org/10.1017/S0033291719002678> (2019).
- Bick, A. G. *et al.* Inherited causes of clonal haematopoiesis in 97,691 whole genomes. *Nature* **586**(7831), 763–768 (2020).
- Thompson, D. J. *et al.* Genetic predisposition to mosaic Y chromosome loss in blood. *Nature* **575**(7784), 652–657 (2019).
- Davies, N. M., Holmes, M. V. & Smith, G. D. Reading Mendelian randomisation studies: A guide, glossary, and checklist for clinicians. *BMJ (Online)* <https://doi.org/10.1136/bmj.k601> (2018).
- Dawoud, A. A. Z., Tapper, W. J. & Cross, N. C. P. Clonal myelopoiesis in the UK Biobank cohort: ASXL1 mutations are strongly associated with smoking. *Leukemia* **34**(10), 2660–2672 (2020).
- Zink, F. *et al.* Clonal hematopoiesis, with and without candidate driver mutations, is common in the elderly. *Blood* **130**(6), 742–752 (2017).
- Van Hout, C. V. *et al.* Exome sequencing and characterization of 49,960 individuals in the UK Biobank. *Nature* **586**(7831), 749–756 (2020).
- Niroula, A., Sekar, A. & Murakami, M. A. *et al.* Distinction of lymphoid and myeloid clonal hematopoiesis. *Nat. Med.* **27**(11), 1921–1927 (2021).
- Slob, E. A. W. & Burgess, S. A comparison of robust Mendelian randomization methods using summary data. *Genet. Epidemiol.* **44**(4), 313–329 (2020).

Acknowledgements

We thank the participants of the UK Biobank. SMD is supported by IK2-CX001780 from the US Department of Veterans Affairs. This publication does not represent the views of the Department of Veterans Affairs or the United States government. PN is supported by grants from the National Heart, Lung, and Blood Institute (R01HL142711, R01HL148050, R01HL151283, R01HL148565) and Fondation Leducq (TNE-18CVD04). MGL is supported by the Institute for Translational Medicine and Therapeutics of the Perelman School of Medicine at the University of Pennsylvania and the NIH/NHLBI National Research Service Award postdoctoral fellowship (T32HL007843).

Author contributions

All authors conceived of the study/experimental design and critically revised the manuscript. M.G.L. and T.N. performed the analyses.

Competing interests

PN reported receiving grants from Amgen, Apple, AstraZeneca, and Boston Scientific outside the submitted work; receiving personal fees from AstraZeneca, Blackstone Life Sciences, Novartis, Apple, Genentech/Roche, Foresite Labs; and having equity in and a spousal employment at Vertex Pharmaceuticals, all outside the present work. SMD reported receiving grants from the US Department of Veterans Affairs during the conduct of the study and receiving grants from RenalytixAI and personal fees from Calico Labs outside the submitted work. B.L.E. has received research funding from Celgene, Deerfield, Novartis, and Calico and consulting fees from GRAIL. He is a member of the scientific advisory board and shareholder for Neomorph Therapeutics, TenSixteen Bio, Skyhawk Therapeutics, and Exo Therapeutics. All other authors have no conflicts to declare.

Additional information

Correspondence and requests for materials should be addressed to P.N.

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

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



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/>.

© The Author(s) 2022