GWAS Identifies *LINC01184/SLC12A2* as a Risk Locus for Skin and Soft Tissue Infections

JID Open

Journal of Investigative Dermatology (2021) ■, ■–■; doi:10.1016/j.jid.2021.01.020

TO THE EDITOR

Microbial invasion of the skin and underlying soft tissues, known as skin and soft tissue infections (SSTIs), contribute to a considerable burden of disease worldwide (Kaye et al., 2019; Lozano et al., 2012). Knowledge about host factors contributing to SSTI risk is important to prevent the SSTIs. The genetics of SSTI susceptibility remain largely unknown, and the only previously published genome-wide study on SSTIs is a small family-based linkage study that did not identify significant linkage to any genes for erysipelas or cellulitis susceptibility (Hannula-Jouppi et al., 2013).

A range of cardiometabolic risk factors has been associated with SSTIs (Butler-Laporte et al., 2020; Kaye et al., 2019; Winter-Jensen et al., 2020). Few studies have used genetic variants as instrumental variables (Mendelian randomization [MR]) to assess causality, which may reduce bias owing to reverse causation and confounding (Davies et al., 2018). Increasing body mass index has been found to increase the risk of SSTIs in such a framework (Butler-Laporte et al., 2020; Winter-Jensen et al., 2020), but other cardiometabolic risk factors have, to our knowledge, not been explored.

The aims of this study were to conduct a GWAS on susceptibility to SSTIs, explore possible biological pathways through transcriptome-wide association analyses, and perform MR analyses to investigate the potential causal relationships of cardiometabolic risk factors with SSTIs.

We used two independent cohorts: UK Biobank and Trøndelag Health Study (HUNT), where the UK Biobank served as the discovery cohort in the genome-wide association analyses and the HUNT as the replication cohort. Subjects who had been hospitalized with a primary diagnosis of SSTI served as cases, whereas those who had not been hospitalized with a primary or secondary diagnosis of SSTI were considered controls (Supplementary Material and Methods).

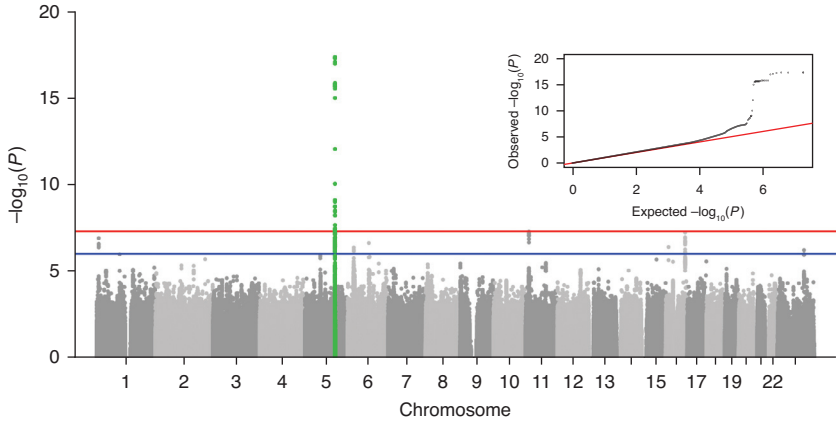


Figure 1. Manhattan plot of results for the meta-analysis. Axes display the $-\log_{10}$ transformed *P*-value by chromosomal position. The blue line indicates genome-wide–suggestive associations ($P < 1e\text{-}6$), and the red line indicates genome-wide–significant associations ($P < 5e\text{-}8$). Genome-wide–significant loci (±500 kb of lead variant) are highlighted in green. The image in the top right corner shows the quantile–quantile plot. Axes display the observed (y-axis) and expected (x-axis) $-\log_{10}$ transformed *P*-value. The black dots represent the observed *P*-values, whereas the red line represents the expected *P*-values under the null distribution. Genomic inflation factor (λ) = 1.01.

Genome-wide association analyses were conducted using scalable and accurate implementation of generalized mixed model, with age, sex, genotype chip, and ancestry-informative principal components as covariates (Zhou et al., 2018), and meta-analyses were conducted using METAL (Supplementary Materials and Methods). Associations with $P < 1e\text{-}6$ and $P < 5e\text{-}8$ were considered genome-wide suggestive and significant, respectively.

We used FUSION to perform transcriptome-wide association analyses by combining the summary statistics from the genome-wide meta-analysis with linkage disequilibrium (European ancestry in 1000 Genomes Project) and reference gene expression panels (Genotype-Tissue Expression, version 7) to estimate the gene expression patterns associated with SSTIs (Gusev et al., 2016). Sun-exposed skin (lower legs) was the tissue of interest for the transcriptome-wide analyses (8,609 genes tested), whereas all the 48 general tissues from Genotype-Tissue Expression, version 7, were analyzed for the chromosome with genome-wide–significant hits (10,518 tests). Bonferroni-corrected threshold for genome-wide significance was $P < 2.6e\text{-}6$.

Two-sample MR analyses were conducted separately for the results from the meta-analysis, UK Biobank, and HUNT. Genetic instruments for body mass index, type-2 diabetes mellitus, low-density lipoprotein cholesterol, systolic blood pressure, lifetime smoking, and sedentary lifestyle were extracted from relevant published

Abbreviations: HUNT, Trøndelag Health Study; MR, Mendelian randomization; SSTI, skin and soft tissue infection

Accepted manuscript published online XXX; corrected proof published online XXX

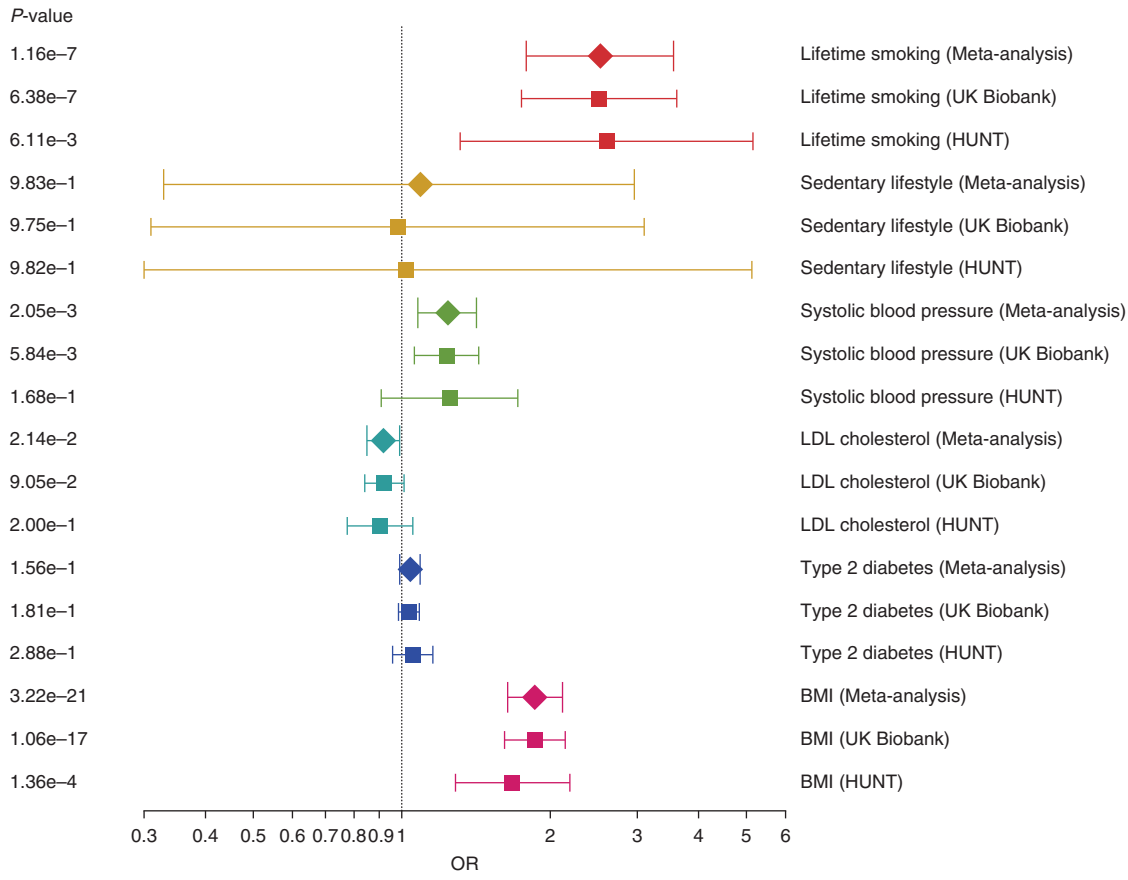


Figure 2. MR analyses of cardiometabolic risk factors on the risk of SSTI. Forest plot of the two-sample inverse-variance weighted MR analyses of cardiometabolic risk factors identified as genetically correlated with SSTI. Each risk factor was evaluated separately using the results from the meta-analysis, UK Biobank, HUNT, and the corresponding risk factors were grouped by color. The x-axis represents the increased OR per SD increase of the genetically predicted risk factor (per unit increase in log OR for genetically proxied type-2 diabetes mellitus liability). BMI, body mass index; HUNT, Trøndelag Health Study; LDL, low-density lipoprotein; MR, Mendelian randomization; SSTI, skin and soft tissue infection.

GWASs (Supplementary Table S1). The TwoSampleMR R package (version 0.5.0) (Hemani et al., 2018) was used to carry out inverse-variance weighted MR analyses (main analyses), along with statistical tests for heterogeneity, simple median, weighted median, and MR Egger (sensitivity analyses).

In both UK Biobank and HUNT, cases at baseline were older, had higher body mass index and systolic blood pressure, were more likely to be male, were more likely to be ever smoker, and were more likely to self-report as diabetic (Supplementary Table S2) than the controls.

The GWAS included 6,107 cases and 399,239 controls from UK Biobank and 1,657 cases and 67,522 controls from HUNT. UK Biobank yielded seven suggestive loci (Supplementary Table S3 and Supplementary Figure S1), of which one was replicated in HUNT: rs3749748 in the *LINC01184/SLC12A2*-gene region on chromosome 5 (Supplementary Figures S2 and S3). In the meta-analysis of 7,764 cases and 466,761 controls, only the locus in *LINC01184/SLC12A2* reached genome-wide significance (Figure 1), whereas two additional loci were close to genome-wide significance: *PSMA1* on chromosome 11 and *GAN* on chromosome 16 (Supplementary Table S3). There was no indication of genomic inflation (Figure 1 and Supplementary Figures S1 and S2).

LINC01184 is part of the long intervening noncoding RNA class of genes that do not encode for proteins but have still been found to modulate inflammation and infection risk (Atianand et al., 2016; Carpenter et al., 2013). *SLC12A2* encodes for the protein NKCC1, which regulates the transportation of chloride, potassium, and sodium across cell membranes, and is key in modulating ion movement across the epithelium, the volume of cells, and antimicrobial activity (Matthay and Su, 2007; Yang et al., 2020).

In the transcriptome-wide association analysis of the skin on the lower legs, the only gene that was statistically significantly associated with SSTIs was *LINC01184* (Supplementary Figure S4). A reduced expression of *LINC01184* was associated with an increased risk of SSTIs. The same association was observed in all tissues but less pronounced in the brain (Supplementary Figure S4).

An increase in genetically predicted body mass index, systolic blood pressure, and smoking increased the risk of SSTIs, whereas increasing low-density lipoprotein cholesterol was associated with a reduced risk of SSTIs (Figure 2). Sensitivity analyses supported the findings from the inverse-variance weighted analyses (Supplementary Table S4).

To our knowledge, this study is a GWAS published on SSTIs, with a large number of cases and controls. We were able to identify a locus—*LINC01184/SLC12A2*—robustly associated with

SSTIs in the discovery cohort and the independent replication cohort. A limitation of our study is that we did not have the power to identify more than one genome-wide–significant locus, which in part may be due to the non-differential misclassification of the outcome, and we thus encourage a replication with meta-analysis in independent cohorts. Of note, whereas the minor allele frequency of rs3749748 in the North-Western European populations is around 23%, it is only 4% in African American populations (Karczewski et al., 2020). It is therefore important to evaluate the populations of different ancestries other than the one currently considered.

In conclusion, we have identified genetic variation in *LINC01184/SLC12A2* to be strongly associated with the risk of SSTIs. Interventions to reduce smoking, hypertension, overweight, and obesity in the population will likely reduce the disease burden of SSTIs.

Data availability statement

Data from the Trøndelag Health Study and UK Biobank are available on application. Gene expression data are available through the FUSION website (http://gusevlab.org/projects/fusion/). Summary statistics are available at the GWAS Catalog (https://www.ebi.ac.uk/gwas) under identification number GCST90013411.

ORCIDs

Tormod Rogne: http://orcid.org/0000-0002-9581-7384
Kristin V. Liyanarachi: http://orcid.org/0000-0001-5499-9196
Humaira Rasheed: http://orcid.org/0000-0002-3331-5864
Laurent F. Thomas: http://orcid.org/0000-0003-0548-2486
Helene M. Flatby: http://orcid.org/0000-0002-5700-020X
Jørgen Stenvik: http://orcid.org/0000-0002-1051-9258
Mari Løset: http://orcid.org/0000-0003-3736-6551
Dipender Gill: http://orcid.org/0000-0001-7312-7078
Stephen Burgess: http://orcid.org/0000-0001-5365-8760
Cristen J. Willer: http://orcid.org/0000-0001-5645-4966
Kristian Hveem: http://orcid.org/0000-0001-8157-9744
Bjørn O. Åsvold: http://orcid.org/0000-0003-3837-2101
Ben M. Brumpton: http://orcid.org/0000-0002-3058-1059
Andrew T. DeWan: http://orcid.org/0000-0002-7679-8704
Erik Solligård: http://orcid.org/0000-0001-6173-3580
Jan K. Damås: http://orcid.org/0000-0003-4268-671X

CONFLICT OF INTEREST

DG is employed part-time by Novo Nordisk, outside of the submitted work. The remaining authors state no conflict of interest.

ACKNOWLEDGMENTS

This study was in part funded by Samarbeidsorganet Helse Midt-Norge, NTNU Norwegian University of Science and Technology (Trondheim, Norway), and The Research Council of Norway (grant 299765). The first author was funded in part by a Fulbright Scholarship by the United States-Norway Fulbright Foundation. BMB, HR, LFT, ML, KH, and BOÅ work in a research unit funded by Stiftelsen Kristian Gerhard Jebsen; Faculty of Medicine and Health Sciences, NTNU Norwegian University of Science and Technology; The Liaison Committee for education, research, and innovation in Central Norway; the Joint Research Committee between St. Olavs Hospital (Trondheim, Norway) and the Faculty of Medicine and Health Sciences, NTNU Norwegian University of Science and Technology; and the Medical Research Council Integrative Epidemiology Unit at the University of Bristol (United Kingdom), which is supported by the Medical Research Council and the University of Bristol (MC_UU_12013/1). All authors had full access to all the data in the study and can take responsibility for the integrity of the data and the accuracy of the data analysis. All authors made substantial contributions to the interpretation of the data and critically revised the manuscript. All authors have approved the submitted version and are personally accountable for the author's own contributions. The Trøndelag Health Study is a collaboration between the Trøndelag Health Study Research Centre (Faculty of Medicine and Health Sciences, Norwegian University of Science and Technology), the Trøndelag County Council, the Central Norway Regional Health Authority, and the Norwegian Institute of Public Health. This research has been conducted using the UK Biobank Resource under application number 40135.

AUTHOR CONTRIBUTIONS

Conceptualization: TR, KVL, ES, JKD, ATD, HMF; Data Curation: TR, HMF, BMB, HR, LFT, CJW, KH, BOÅ; Formal Analysis: TR, HR, LFT; Funding Acquisition: TR, ES, JKD, KH, CJW, BOÅ, JS, ATD, BMB; Investigation: TR, HR, LFT, ES, JKD, KH, CJW, BOÅ, JS, ATD, ML, BMB; Methodology: TR, HR, LFT, DG, SB, ATD, BMB; Project Administration: TR, ES, JKD, BOÅ, KH, CJW, ATD, BMB, ML; Resources: ES, JKD, BOÅ, KH, ATD, JS; Software: DG, SB, BMB, HR, LFT; Supervision: TR, ES, JKD, ATD, BMB, DG, SB, BOÅ, ML, CJW; Validation: HR, LFT, BMB, JS; Visualization: TR, HR, LFT; Writing - Original Draft Preparation: TR; Writing - Review and Editing: TR, KVL, HR, LFT, HMF, JS, ML, DG, SB, CJW, KH, BOÅ, BMB, ATD, ES, JKD

Disclaimer

The funding sources had no role in the study design; in the collection, analysis, and interpretation of data; in the writing of the report; nor in the decision to submit the article for publication. The researchers were independent of the funders.

Tormod Rogne[1,2,3,*], Kristin V. Liyanarachi[1,4,23], Humaira Rasheed[5,6,23], Laurent F. Thomas[5,7,8,9], Helene M. Flatby[1,3], Jørgen Stenvik[7,4,10], Mari Løset[5,11], Dipender Gill[12,13,14], Stephen Burgess[15,16], Cristen J. Willer[17,18,19], Kristian Hveem[5,20], Bjørn O. Åsvold[5,21], Ben M. Brumpton[5,6,22], Andrew T. DeWan[1,2], Erik Solligård[1,3,24] and Jan K. Damås[1,4,10,24]

[1]Gemini Center for Sepsis Research, Department of Circulation and Medical Imaging, NTNU Norwegian University of Science and Technology, Trondheim, Norway; [2]Center for Perinatal, Pediatric and Environmental Epidemiology, Department of Chronic Disease Epidemiology, Yale School of Public Health, New Haven, Connecticut, USA; [3]Clinic of Anaesthesia and Intensive Care, St. Olavs Hospital, Trondheim University Hospital, Trondheim, Norway; [4]Department of Infectious Diseases, St. Olavs Hospital, Trondheim University Hospital, Trondheim, Norway; [5]K.G. Jebsen Center for Genetic Epidemiology, Department of Public Health and Nursing, NTNU Norwegian University of Science and Technology, Trondheim, Norway; [6]MRC Integrative Epidemiology Unit, University of Bristol, Bristol, United Kingdom; [7]Department of Clinical and Molecular Medicine, NTNU Norwegian University of Science and Technology, Trondheim, Norway; [8]BioCore - Bioinformatics Core Facility, NTNU Norwegian University of Science and Technology, Trondheim, Norway; [9]Clinic of Laboratory Medicine, St. Olavs Hospital, Trondheim University Hospital, Trondheim, Norway; [10]Centre of Molecular Inflammation Research, Department of Clinical and Molecular Medicine, NTNU Norwegian University of Science and Technology, Trondheim, Norway; [11]Department of Dermatology, St. Olavs Hospital, Trondheim University Hospital, Trondheim, Norway; [12]Clinical Pharmacology and Therapeutics Section, Institute of Medical and Biomedical Education, St George's University of London, London, United Kingdom; [13]Institute for Infection and Immunity, St George's University of London, London, United Kingdom; [14]Clinical Pharmacology Group, Pharmacy and Medicines Directorate, St George's University Hospitals NHS Foundation Trust, London, United Kingdom; [15]MRC Biostatistics Unit, University of Cambridge, Cambridge, United Kingdom; [16]Cardiovascular Epidemiology Unit, Department of Public Health and Primary Care, University of Cambridge, Cambridge, United Kingdom; [17]Department of Internal Medicine, University of Michigan, Ann Arbor, Michigan, USA; [18]Department of Human Genetics, University of Michigan, Ann Arbor, Michigan, USA; [19]Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, Michigan, USA; [20]Department of Research, Innovation and Education, St. Olavs

Hospital, Trondheim University Hospital, Trondheim, Norway; [21]*Department of Endocrinology, Clinic of Medicine, St. Olavs Hospital, Trondheim University Hospital, Trondheim, Norway; and* [22]*Clinic of Thoracic and Occupational Medicine, St. Olavs Hospital, Trondheim University Hospital, Trondheim, Norway*
[23]*These authors contributed equally to this work.*
[24]*These authors contributed equally to this work.*
[*]*Corresponding author e-mail: tormod.rogne@ntnu.no*

SUPPLEMENTARY MATERIAL

Supplementary material is linked to the online version of the paper at www.jidonline.org, and at https://doi.org/10.1016/j.jid.2021.01.020.

REFERENCES

Atianand MK, Hu W, Satpathy AT, Shen Y, Ricci EP, Alvarez-Dominguez JR, et al. A long noncoding RNA lincRNA-EPS acts as a transcriptional brake to restrain inflammation. Cell 2016;165:1672–85.

Butler-Laporte G, Harroud A, Forgetta V, Richards JB. Elevated body mass index is associated with an increased risk of infectious disease admissions and mortality: a Mendelian randomization study [e-pub ahead of print]. Clin Microbiol Infect 2020. https://doi.org/10.1016/j.cmi.2020.06.014 (accessed 1 September 2020).

Carpenter S, Aiello D, Atianand MK, Ricci EP, Gandhi P, Hall LL, et al. A long noncoding RNA mediates both activation and repression of immune response genes. Science 2013;341:789–92.

Davies NM, Holmes MV, Davey Smith G. Reading Mendelian randomisation studies: a guide, glossary, and checklist for clinicians. BMJ 2018;362:k601.

Gusev A, Ko A, Shi H, Bhatia G, Chung W, Penninx BW, et al. Integrative approaches for large-scale transcriptome-wide association studies. Nat Genet 2016;48:245–52.

Hannula-Jouppi K, Massinen S, Siljander T, Mäkelä S, Kivinen K, Leinonen R, et al. Genetic susceptibility to non-necrotizing erysipelas/cellulitis. PLoS One 2013;8:e56225.

Hemani G, Zheng J, Elsworth B, Wade KH, Haberland V, Baird D, et al. The MR-base platform supports systematic causal inference across the human phenome. Elife 2018;7:e34408.

Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Wang Q, Collins RL, et al. The mutational constraint spectrum quantified from variation in 141,456 humans [published correction appears in Nature 2021;590:E53]. Nature 2020;581:434–43.

Kaye KS, Petty LA, Shorr AF, Zilberberg MD. Current epidemiology, etiology, and burden of acute skin infections in the United States. Clin Infect Dis 2019;68(Suppl. 3):S193–9.

Lozano R, Naghavi M, Foreman K, Lim S, Shibuya K, Aboyans V, et al. Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010. Lancet 2012;380:2095–128.

Matthay MA, Su X. Pulmonary barriers to pneumonia and sepsis. Nat Med 2007;13:780–1.

Winter-Jensen M, Afzal S, Jess T, Nordestgaard BG, Allin KH. Body mass index and risk of infections: a Mendelian randomization study of 101,447 individuals. Eur J Epidemiol 2020;35:347–54.

Yang X, Wang Q, Cao E. Structure of the human cation–chloride cotransporter NKCC1 determined by single-particle electron cryo-microscopy. Nat Commun 2020;11:1016.

Zhou W, Nielsen JB, Fritsche LG, Dey R, Gabrielsen ME, Wolford BN, et al. Efficiently controlling for case-control imbalance and sample relatedness in large-scale genetic association studies. Nat Genet 2018;50:1335–41.

SUPPLEMENTARY MATERIALS AND METHODS

Material

UK Biobank. Details about the UK Biobank have previously been described (Bycroft et al., 2018). In brief, the cohort consists of 503,325 subjects enrolled between 2006 and 2010 throughout the United Kingdom. Age at baseline was between 38 and 73 years, and 94% of the subjects self-reported being of European ancestry. At baseline, genome-wide genotyping was done on 488,377 individuals, where 84% self-reported that they were of white-British ancestry with European genetic ethnicity. Information on self-reported health and lifestyle was collected, along with measurements such as height and weight. Inpatient hospital data on all the participants were available through electronic record linkage.

Trøndelag Health Study. The Trøndelag Health Study (HUNT) study is a series of surveys conducted in the Nord-Trøndelag region in Norway (~130,000 inhabitants) between 1984 and 2019 on subjects aged ≥20 years (Krokstad et al., 2013). We used data from HUNT2 (1995–1997) and HUNT3 (2006–2008), in which 78,973 subjects representative of the adult Norwegian population participated (Krokstad et al., 2013). Baseline characteristics were collected at study enrollment, and selected measurements were carried out, including height and weight measurement. Information on all hospitalizations in the county and to the regional tertiary care hospital were linked to the study subjects. Through linkage with the Norwegian population registry, we retrieved data on the date of emigration out of the study region and date of death.

Phenotype

Cases and controls were defined similarly to the definitions in the UK Biobank and HUNT. The following International Classification of Diseases (ICD)-9 and ICD-10 codes were considered as skin and soft tissue infection (SSTI) codes: 035 (erysipelas; ICD-9), 729.4 (fasciitis, unspecified; ICD-9), A46 (erysipelas; ICD-10), L03 (cellulitis and acute lymphangitis; ICD-10), and M72.6 (necrotizing fasciitis; ICD-10). These codes are used primarily for bacterial infections, and nonbacterial infections of the skin have other specific codes not considered. Our main definition of SSTI was a case that had been hospitalized with an SSTI as the primary diagnosis. In the sensitivity analysis, we included secondary diagnoses in the definition of SSTI (i.e., SSTIs not the primary cause of hospitalization).

Those who had not been hospitalized with an SSTI (primary or secondary diagnosis) served as controls.

Genotyping

UK biobank. The Affymetrix UK BiLEVE Axiom array was used to genotype the initial 50,000 participants, and the Affymetrix UK Biobank Axiom array was used to genotype the rest of the subjects. Directly genotyped variants were prephased using SHAPEIT3 (O'Connell et al., 2016) and imputed using Impute4 and the UK10K (UK10K Consortium et al., 2015), Haplotype Reference Consortium (UK10K Consortium et al., 2015), and 1000 Genomes Phase 3 (1000 Genomes Project Consortium et al., 2015) reference panels (version 3 of the imputed data). Exclusions were made for variants with imputation score $r^2 < 0.3$. More detail is contained in a previous publication (Bycroft et al., 2018).

HUNT. As previously described, three different Illumina HumanCoreExome arrays were used to genotype the study participants (HumanCoreExome12, version 1.0; HumanCoreExome12, version1.1; and UM HUNT Biobank, version1.0) (Ferreira et al., 2017). Samples with a call rate <99%, with large chromosomal copy number variants, with contamination >2.5% as estimated with BAF Regress (Jun et al., 2012), with genotypic and phenotypic sex discordance, and that were not of European ancestry were excluded, leaving 69,422 genotyped subjects. Genetic variants of Hardy–Weinberg equilibrium ($P < 0.0001$) or with a call rate <99% were excluded. Imputation was done using Minimac3 of 2,201 whole-genome reference sequences from HUNT and Haplotype Reference Consortium, version 1.1, resulting in 24.9 million SNPs ($r^2 > 0.3$). Principal components were calculated using of TRACE (version 1.03), with 938 individuals from the Human Genome Diversity Project serving as reference (Wang et al., 2015, 2014).

Genome-wide association analyses

UK Biobank. Genome-wide association analysis was performed in scalable and accurate implementation of generalized mixed model (version 0.35.8.3) using a linear mixed model that accounts for cryptic relatedness and imbalance in the proportion of cases and controls (Zhou et al., 2018). We included birth year, sex, genotype chip, and the first six ancestry-informative principal components as covariates. We used scalable and accurate implementation of generalized mixed model with the same settings to analyze the X chromosome, coding males as diploid. Variants with minor allele frequency >0.5% were included in the analyses, and dosages were used for imputed variants.

HUNT. Genome-wide association tests were carried using scalable and accurate implementation of generalized mixed model (version 0.29.4) on autosomal chromosomes (Zhou et al., 2018), whereas BOLT-LMM (version 2.3.4) was used in the analysis of the X chromosome, coding males as diploid (Loh et al., 2015). The beta-coefficients from BOLT-LMM were transformed using the formula: log OR = $\beta/(\mu * (1 - \mu))$, where μ = case fraction. The standard errors from BOLT-LMM were transformed using the formula: $SE_{transformed} = SE_{original}/(\mu \times [1 - \mu])$. Age, sex, genotype batch, and the five first ancestry-informative principal components were included as covariates. Variants with minor allele frequency >0.5% were included in the analyses, and dosages were used for imputed variants.

Meta-analysis. We carried out a meta-analysis using METAL (version 2011-03-25), with the use of effect size estimates and standard errors as weights, adjusting for residual population stratification and relatedness through genomic control correction (Willer et al., 2010). A total of 9,211,777 SNPs that were present in both cohorts were included in the meta-analysis.

Ethical approval

The Regional Committee for Medical Research, Health Region IV, Norway approved the HUNT study, and this project is regulated in conjunction with The Norwegian Social Science Data Services. The UK Biobank study has ethical approval from the North West Multi-centre Research Ethics Committee. Approval for individual projects is covered by the Research Tissue Bank.

SUPPLEMENTARY REFERENCES

1000 Genomes Project Consortium, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, et al. A global reference for human genetic variation. Nature 2015;526:68–74.

Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, et al. The UK Biobank Resource with deep phenotyping and genomic data. Nature 2018;562:203–9.

Carter AR, Gill D, Davies NM, Taylor AE, Tillmann T, Vaucher J, et al. Understanding the consequences of education inequality on cardiovascular disease: Mendelian randomisation study. BMJ 2019;365:l1855.

Doherty A, Smith-Byrne K, Ferreira T, Holmes MV, Holmes C, Pulit SL, et al. GWAS identifies 14 loci for device-measured physical activity and sleep duration. Nat Commun 2018;9:5257.

Ferreira MA, Vonk JM, Baurecht H, Marenholz I, Tian C, Hoffman JD, et al. Shared genetic origin of asthma, hay fever and eczema elucidates allergic disease biology. Nat Genet 2017;49:1752–7.

Jun G, Flickinger M, Hetrick KN, Romm JM, Doheny KF, Abecasis GR, et al. Detecting and estimating contamination of human DNA samples in sequencing and array-based genotype data. Am J Hum Genet 2012;91:839–48.

Krokstad S, Langhammer A, Hveem K, Holmen TL, Midthjell K, Stene TR, et al. Cohort profile: the HUNT study, Norway. Int J Epidemiol 2013;42:968–77.

Loh PR, Tucker G, Bulik-Sullivan BK, Vilhjálmsson BJ, Finucane HK, Salem RM, et al. Efficient Bayesian mixed-model analysis increases association power in large cohorts. Nat Genet 2015;47:284–90.

Mahajan A, Taliun D, Thurner M, Robertson NR, Torres JM, Rayner NW, et al. Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. Nat Genet 2018;50:1505–13.

O'Connell J, Sharp K, Shrine N, Wain L, Hall I, Tobin M, et al. Haplotype estimation for biobank-scale data sets. Nat Genet 2016;48:817–20.

UK10K Consortium, Walter K, Min JL, Huang J, Crooks L, Memari Y, et al. The UK10K project identifies rare variants in health and disease. Nature 2015;526:82–90.

Wang C, Zhan X, Bragg-Gresham J, Kang HM, Stambolian D, Chew EY, et al. Ancestry estimation and control of population stratification for sequence-based association studies. Nat Genet 2014;46:409–15.

Wang C, Zhan X, Liang L, Abecasis GR, Lin X. Improved ancestry estimation for both genotyping and sequencing data using projection Procrustes analysis and genotype imputation. Am J Hum Genet 2015;96:926–37.

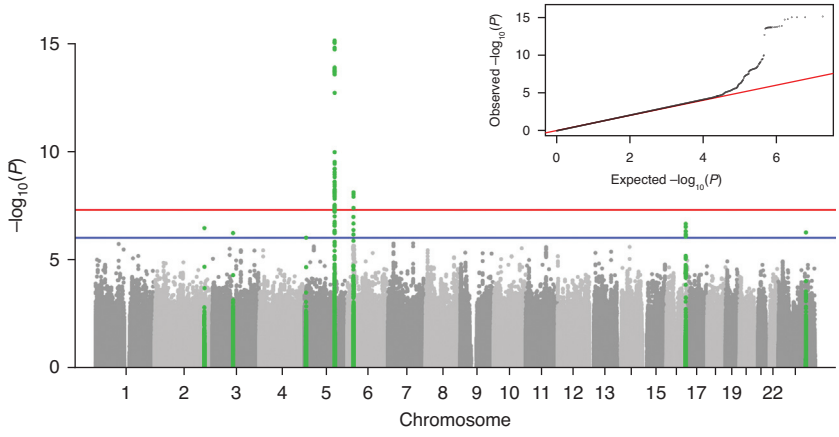
Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics 2010;26:2190–1.

Willer CJ, Schmidt EM, Sengupta S, Peloso GM, Gustafsson S, Kanoni S, et al. Discovery and refinement of loci associated with lipid levels. Nat Genet 2013;45:1274–83.

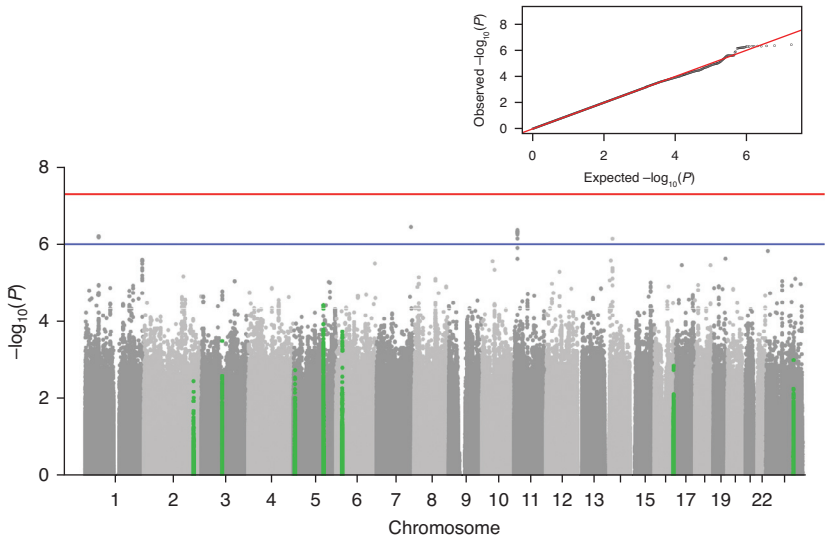
Wootton RE, Richmond RC, Stuijfzand BG, Lawn RB, Sallis HM, Taylor GMJ, et al. Evidence for causal effects of lifetime smoking on risk for depression and schizophrenia: a Mendelian randomisation study. Psychol Med 2019;50:2435–43.

Yengo L, Sidorenko J, Kemper KE, Zheng Z, Wood AR, Weedon MN, et al. Meta-analysis of genome-wide association studies for height and body mass index in ~700000 individuals of European ancestry. Hum Mol Genet 2018;27:3641–9.

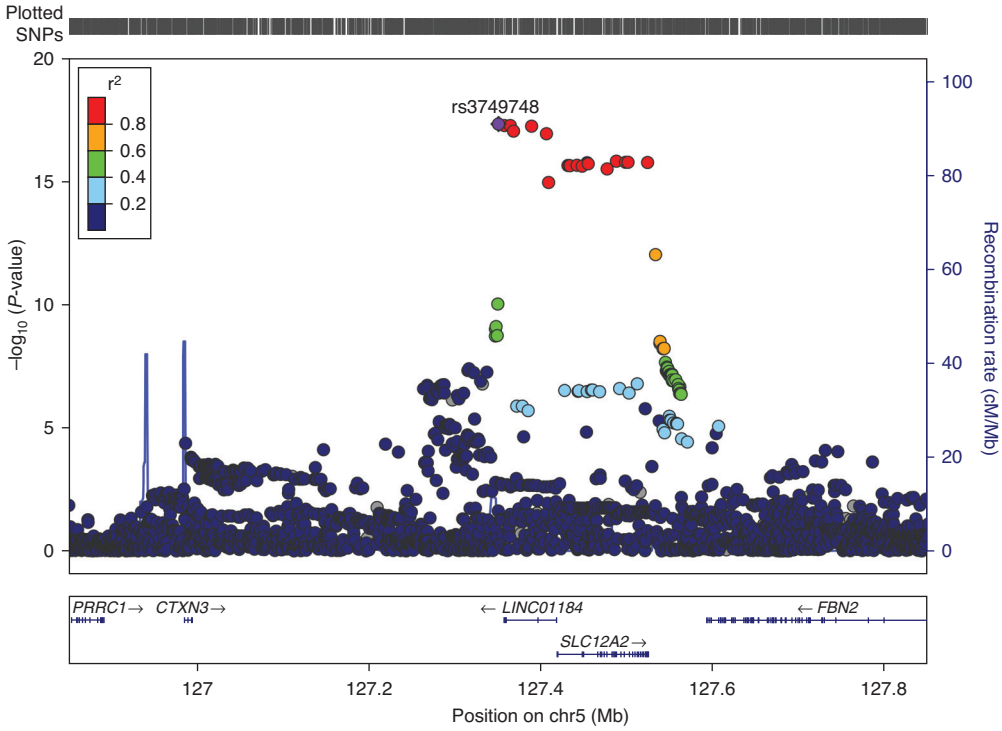
Zhou W, Nielsen JB, Fritsche LG, Dey R, Gabrielsen ME, Wolford BN, et al. Efficiently controlling for case-control imbalance and sample relatedness in large-scale genetic association studies. Nat Genet 2018;50:1335–41.



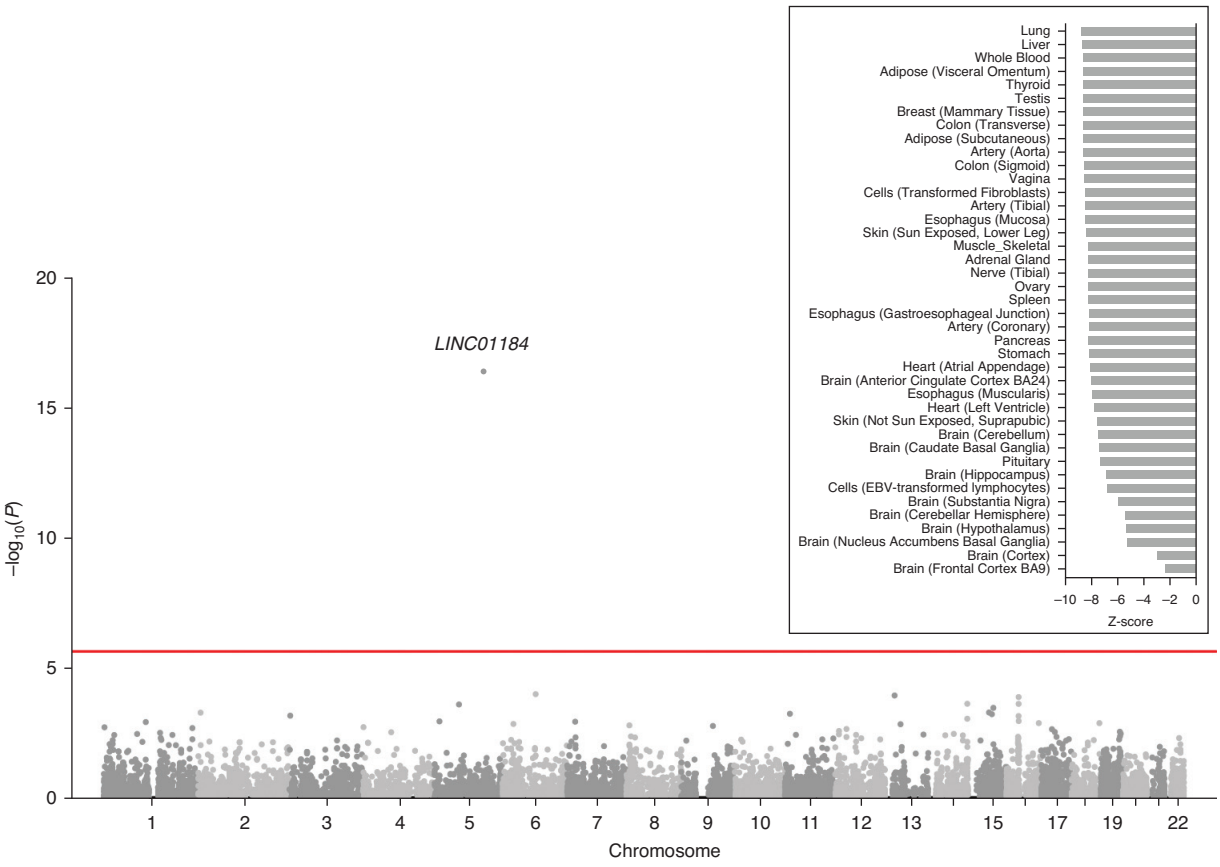
Supplementary Figure S1. Manhattan plot of the results for the discovery stage (UK Biobank). Axes display the $-\log_{10}$ transformed *P*-value by chromosomal position. The blue line indicates the genome-wide–suggestive associations ($P < 1e\text{-}6$), and the red line indicates the genome-wide–significant associations ($P < 5e\text{-}8$). Genome-wide–suggestive loci (±500 kb of lead variant) are highlighted in green. The image in the top right corner shows the quantile–quantile plot. Axes display the observed (y-axis) and expected (x-axis) $-\log_{10}$ transformed *P*-value. The black dots represent the observed *P*-values, whereas the red line represents the expected *P*-values under the null distribution. Genomic inflation factor (λ) = 1.02.kb, kilobase.



Supplementary Figure S2. Manhattan plot of the results for the replication stage (HUNT). Axes display the $-\log_{10}$ transformed *P*-value by chromosomal position. The blue line indicates the genome-wide–suggestive associations ($P < 1e\text{-}6$), and the red line indicates the genome-wide–significant associations ($P < 5e\text{-}8$). Genome-wide–suggestive loci from the discovery stage (±500 kb of lead variant) are highlighted in green. The image in the top right corner shows the quantile–quantile plot. Axes display the observed (y-axis) and expected (x-axis) $-\log_{10}$ transformed *P*-value. The black dots represent the observed *P*-values, whereas the red line represents the expected *P*-values under the null distribution. Genomic inflation factor (λ) = 1.00. HUNT, Trøndelag Health Study; kb, kilobase.



Supplementary Figure S3. Regional plot of the association results of the discovery stage genome-wide–significant locus that was replicated. Associations between genetic variants and SSTI from the meta-analysis are plotted by position (x-axis) and $-\log_{10}$ transformed *P*-values (left y-axis). rs3749748 served as sentinel variant, whereas the remaining variants are color coded in terms of the linkage disequilibrium (r^2) to the sentinel variant. Estimated recombination rates are plotted as light blue lines (right y-axis). The European population from 1000 Genomes Project, November 2014 release, was used as the reference on genome build hg19. chr, chromosome; Mb, megabase; SSTI, skin and soft tissue infection.



Supplementary Figure S4. Manhattan plot of transcriptome-wide association analysis. Each dot represents the association between the predicted gene expressions in the skin on the lower legs with the risk of SSTIs. The red line indicates the statistically significant associations ($P < 2.6e-6$). The image in the top right corner shows the transcriptome association statistic for *LINC01184* in all the 48 tissues from GTEx v7. BA, Brodmann area; EBV, Epstein-Barr virus; GTEx v7, Genotype-Tissue Expression, version 7; SSTI, skin and soft tissue infection.

Supplementary Table S1. Genetic Instruments for Cardiometabolic Exposures

Trait	Sample Size	Population Ancestry	Number of Variants	Variance Explained, %	Reference
Body mass index	681,275	European	595	6.0	(Yengo et al., 2018)
Type-2 diabetes mellitus	74,124 cases and 824,006 controls	European	202	16.3	(Mahajan et al., 2018)
Low-density lipoprotein cholesterol	188,577	European	80	7.9	(Willer et al., 2013)
Systolic blood pressure	318,417	European	192	2.9	(Carter et al., 2019)
Lifetime smoking index	462,690	European	126	0.4	(Wootton et al., 2019)
Sedentary lifestyle	91,105	European	4	0.08	(Doherty et al., 2018)

Only independent SNPs ($r^2 < 0.001$) with $P < 5e-8$ in these GWASs were included.

Supplementary Table S2. Background Characteristics at Entry in the UK Biobank and the HUNT Study

Characteristic	UK Biobank			HUNT		
	Cases (n = 6,107)	Controls (n = 399,239)	All (N = 405,346)	Cases (n = 1,657)	Controls (n = 67,522)	All (N = 69,179)
Female sex	2,535 (41.5)	216,956 (54.3)	219,491 (54.1)	825 (49.8)	35,829 (53.1)	36,654 (53.0)
Age, y	60 (53–65)	58 (51–63)	58 (51–63)	55 (43–68)	46 (34–60)	46 (34–60)
Ever smoker	3,895 (63.8)	240,412 (60.2)	244,307 (60.3)	923 (57.4)	37,518 (56.6)	38,441 (56.6)
Sedentary lifestyle[1]	—	—	(7.1)	192 (13.4)	4,180 (7.0)	4,372 (7.1)
Diabetes (self-reported)	115 (1.9)	2,860 (0.7)	2,975 (0.7)	102 (6.2)	2,003 (3.0)	2,105 (3.1)
Body mass index, kg/m^2	30.6 (6.6)	27.3 (4.7)	27.4 (4.7)	28.8 (5.2)	26.3 (4.1)	26.4 (4.2)
LDL cholesterol, mmol/l	3.4 (0.9)	3.6 (0.9)	3.6 (0.9)	3.8 (1.1)	3.6 (1.1)	3.6 (1.1)
Systolic blood pressure, mmHg	141.1 (19.1)	138.2 (18.6)	138.2 (18.6)	142.1 (22.7)	134.9 (20.9)	135.0 (21.0)

Abbreviation: HUNT, Trøndelag Health Study; LDL, low-density lipoprotein.

Data are presented as mean (SD), median (25th and 75th centile), or number (%).

[1]Sedentary lifestyle: the proportion with sedentary lifestyle among all subjects in UK Biobank was estimated from none of the above from data field 6,164 (types of physical activity in the last 4 weeks) because individual-level data were unavailable; in HUNT, sedentary lifestyle was defined as a self-reported average of 0 hours of low or vigorous physical activity per week in the last year.

Supplementary Table S3. Genetic Variants with $P < 1e\text{-}6$ in the Discovery Cohort or $P < 1e\text{-}7$ in the Meta-Analysis on the Risk of SSTIs

					Discovery (UK Biobank)			Replication (HUNT)			Meta-Analysis	
Variant Name	Chr	Pos (hg19)	Closest Gene	EA/OA	EAF	OR (95% CI)	*P*-Value	EAF	OR (95% CI)	*P*-Value	OR (95% CI)	*P*-Value
rs72989928	2	210,196,618	*MAP2*	G/T	0.017	0.69 (0.60–0.79)	3.5e-7	0.014	0.95 (0.68–1.33)	7.7e-1	0.72 (0.63–0.83)	2.0e-6
rs62267025	3	87,726,132	*AC108749.1*	C/T	0.012	1.60 (1.33–1.92)	6.0e-7	0.010	0.92 (0.63–1.35)	6.6e-1	1.44 (1.22–1.70)	2.0e-5
rs150468829	5	7,081,850	*LINC02196*	A/G	0.009	1.67 (1.36–2.05)	9.7e-7	0.009	0.98 (0.67–1.42)	9.0e-1	1.47 (1.23–1.77)	2.7e-5
rs3749748[1]	5	127,350,549	*LINC01184*	T/C	0.248	1.19 (1.14–1.24)	7.6e-16	0.231	1.15 (1.06–1.25)	6.3e-4	1.18 (1.14–1.23)	4.4e-18
rs115740542	6	26,123,502	*H2BC4*	C/T	0.075	1.23 (1.14–1.31)	7.8e-9	0.091	1.01 (0.90–1.14)	8.4e-1	1.17 (1.10–1.24)	4.2e-7
rs2007361	11	14,662,722	*PSMA1*	G/A	0.342	0.93 (0.90–0.97)	4.0e-4	0.365	0.83 (0.77–0.89)	4.7e-7	0.91 (0.88–0.94)	5.1e-8
rs78625038	16	81,402,279	*GAN*	CT/C	0.006	1.98 (1.53–2.56)	2.2e-7	0.006	1.56 (1.00–2.41)	4.9e-2	1.86 (1.48–2.32)	5.9e-8
rs5910356	X	117,606,177	*WDR44*	T/C	0.058	0.84 (0.79–0.90)	5.6e-7	0.055	1.04 (0.91–1.17)	5.9e-1	0.88 (0.83–0.94)	8.1e-5

Abbreviations: Chr, chromosome; CI, confidence interval; EA, effect allele; EAF, effect allele frequency; HUNT, Trøndelag Health Study; OA, other allele; Pos, chromosome position.
[1]Suggestive variants ($P < 1e\text{-}6$) in the discovery cohort that replicate in the HUNT cohort ($P < 7.1e\text{-}3$ and β coefficient in the same direction) are presented.

Supplementary Table S4. MR Sensitivity Analyses of Cardiometabolic Risk Factors on the Risk of SSTI

Trait	UK Biobank			HUNT			Meta-Analysis		
	OR (95% CI) or Q	*P*-Value	Number of SNPs	OR (95% CI) or Q	*P*-Value	Number of SNPs	OR (95% CI) or Q	*P*-Value	Number of SNPs
Lifetime smoking									
IVW	2.51 (1.75–3.61)	6.38e-7	126	2.61 (1.31–5.17)	6.11e-3	125	2.53 (1.79–3.56)	1.16e-7	125
Heterogeneity IVW	135.53	2.45e-1	126	125.35	4.49e-1	125	148.49	6.62e-2	125
Simple median	2.45 (1.46–4.12)	7.31e-4	126	2.92 (1.03–8.28)	4.44e-2	125	2.67 (1.67–4.28)	4.03e-5	125
Weighted median	2.36 (1.38–4.03)	1.69e-3	126	3.16 (1.18–8.42)	2.17e-2	125	2.17 (1.34–3.52)	1.71e-3	125
MR Egger	1.52 (0.36–6.44)	5.71e-1	126	7.17 (0.45–113.72)	1.65e-1	125	2.06 (0.52–8.06)	3.04e-1	125
MR Egger intercept	1.01 (0.99–1.02)	4.81e-1	126	0.99 (0.97–1.02)	4.60e-1	125	1.00 (0.99–1.02)	7.61e-1	125
Sedentary lifestyle									
IVW	0.98 (0.31–3.11)	9.75e-1	4	1.02 (0.20–5.13)	9.82e-1	4	1.09 (0.33–2.96)	9.83e-1	4
Heterogeneity IVW	9.30	2.55e-2	4	4.89	1.80e-1	4			4
Simple median	0.67 (0.29–1.52)	3.34e-1	4	1.00 (0.21–4.81)	9.99e-1	4	0.86 (0.41–1.80)	6.93e-1	4
Weighted median	0.65 (0.27–1.54)	3.29e-1	4	1.01 (0.22–4.66)	9.89e-1	4	0.85 (0.41–1.78)	6.72e-1	4
MR Egger	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
MR Egger intercept	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Systolic blood pressure									
IVW	1.23 (1.06–1.43)	5.84e-3	192	1.25 (0.91–1.72)	1.68e-1	187	1.24 (1.08–1.42)	2.05e-3	187
Heterogeneity IVW	182.96	6.49e-1	192	217.98	5.42e-2	187	185.37	4.99e-1	187
Simple median	1.43 (1.14–1.79)	1.70e-3	192	1.14 (0.74–1.76)	5.61e-1	187	1.21 (1.00–1.47)	4.78e-2	187
Weighted median	1.27 (1.01–1.60)	3.82e-2	192	1.31 (0.82–2.09)	2.60e-1	187	1.10 (0.90–1.35)	3.34e-1	187
MR Egger	0.76 (0.47–1.21)	2.45e-1	192	2.52 (0.93–6.87)	7.19e-2	187	0.99 (0.65–1.52)	9.77e-1	187
MR Egger intercept	1.01 (1.00–1.02)	3.23e-2	192	0.99 (0.97–1.01)	1.50e-1	187	1.00 (1.00–1.01)	2.87e-1	187
Low-density lipoprotein cholesterol									
IVW	0.92 (0.84–1.01)	9.05e-2	80	0.90 (0.78–1.05)	2.00e-1	78	0.92 (0.85–0.99)	2.14e-2	78
Heterogeneity IVW	112.71	7.65e-3	80	48.79	9.95e-1	78	83.58	2.85e-1	78
Simple median	0.89 (0.77–1.03)	1.17e-1	80	0.99 (0.77–1.28)	9.46e-1	78	0.87 (0.77–0.99)	3.05e-2	78
Weighted median	0.90 (0.79–1.01)	7.64e-2	80	0.98 (0.78–1.25)	8.95e-1	78	0.91 (0.82–1.01)	8.14e-2	78
MR Egger	0.89 (0.78–1.02)	1.01e-1	80	0.89 (0.71–1.12)	3.13e-1	78	0.89 (0.80–0.99)	3.81e.2	78
MR Egger intercept	1.00 (0.99–1.01)	4.88e-1	80	1.00 (0.99–1.02)	8.38e-1	78	1.00 (1.00–1.01)	4.58e-1	78
Type-2 diabetes mellitus									
IVW	1.03 (0.98–1.09)	1.81e-1	199	1.05 (0.96–1.16)	2.88e-1	195	1.04 (0.99–1.09)	1.56e-1	195
Heterogeneity IVW	243.51	1.93e-2	199	216.12	1.32e-1	195	263.37	6.75e-4	195
Simple median	1.05 (0.97–1.14)	1.99e-1	199	1.07 (0.92–1.23)	3.85e-1	195	1.09 (1.02–1.17)	1.47e-2	195
Weighted median	0.96 (0.89–1.04)	3.43e-1	199	0.97 (0.81–1.16)	7.39e-1	195	0.97 (0.90–1.04)	3.35e-1	195
MR Egger	0.90 (0.81–1.00)	4.85e-2	199	1.05 (0.85–1.29)	6.54e-1	195	0.92 (0.83–1.02)	1.26e-1	195
MR Egger intercept	1.01 (1.00–1.02)	3.61e-3	199	1.00 (0.99–1.01)	9.64e-1	195	1.01 (1.00–1.02)	1.38e-2	195

(*continued*)

Supplementary Table S4. Continued

Trait	UK Biobank			HUNT			Meta-Analysis		
	OR (95% CI) or Q	*P*-Value	Number of SNPs	OR (95% CI) or Q	*P*-Value	Number of SNPs	OR (95% CI) or Q	*P*-Value	Number of SNPs
Body mass index									
IVW	1.86 (1.62−2.15)	1.06e-17	594	1.68 (1.29−2.19)	1.36e-4	580	1.86 (1.64−2.12)	3.22e-21	580
Heterogeneity IVW	658.06	3.26e-2	594	532.31	9.18e-1	580	641.16	3.72e-2	580
Simple median	1.91 (1.56−2.34)	6.17e-10	594	1.62 (1.11−2.37)	1.28e-2	580	1.92 (1.60−2.31)	2.29e-12	580
Weighted median	1.63 (1.33−2.00)	2.06e-6	594	1.53 (1.02−2.30)	4.03e-2	580	1.83 (1.51−2.21)	7.05e-10	580
MR Egger	1.70 (0.95−3.04)	7.38e-2	594	1.02 (0.34−3.05)	9.78e-1	580	1.41 (0.83−2.41)	2.03e-1	580
MR Egger intercept	1.00 (0.99−1.01)	7.50e-1	594	1.01 (0.99−1.02)	3.55e-1	580	1.00 (1.00−1.01)	2.96e-1	580

Abbreviations: CI, confidence interval; HUNT, Trøndelag Health Study; IVW, inverse-variance weighted; MR, Mendelian randomization; N/A, not applicable; SSTI, SSTI, skin and soft tissue infection.
The effect estimates are presented as OR per SD increase of the genetically predicted risk factor (per unit increase in log OR for genetically proxied type-2 diabetes mellitus liability). For the heterogeneity test of the IVW analysis, the Q-statistic along with its *P*-value is presented.