GWAS identifies risk locus for erectile dysfunction and implicates hypothalamic neurobiology and diabetes in etiology

Jonas Bovijn,[1,2,14] Leigh Jackson,[3,14] Jenny Censin,[1,2,14] Chia-Yen Chen,[4,5,14] Triin Laisk-Podar,[6,7,14] Samantha Laber,[1,2,14] Teresa Ferreira,[1] Sara L Pulit,[1,8,9] Craig A Glastonbury,[1] Jordan Smoller,[5] Jamie W Harrison,[10] Katherine S Ruth,[10] Robin N Beaumont,[10] Samuel E Jones,[10] Jessica Tyrrell,[10] Andrew R Wood,[10] Michael N Weedon,[10,14] Reedik Mägi,[6,14] Benjamin Neale,[4,5,14] Cecilia M Lindgren,[1,2,8,14] Anna Murray,[10,14*] Michael V Holmes,[11-13,14**]

Affiliations

1. Big Data Institute at the Li Ka Shing Centre for Health Information and Discovery, University of Oxford, Oxford, OX3 7LF, UK.
2. Wellcome Centre for Human Genetics, Nuffield Department of Medicine, University of Oxford, Oxford, OX3 7BN, UK.
3. Institute of Biomedical and Clinical Science, University of Exeter Medical School, University of Exeter, Exeter, EX2 5DW, UK.
4. Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, Massachusetts 02114 USA.
5. Psychiatric & Neurodevelopmental Genetics Unit, Massachusetts General Hospital, Boston, Massachusetts, 02114, USA.
6. Estonian Genome Center, Institute of genomics, University of Tartu, Tartu 51010, Estonia.
7. Department of Obstetrics and Gynecology, Institute of Clinical Medicine, University of Tartu, Tartu, Estonia.
8. Program in Medical and Population Genetics, Broad Institute, Cambridge, Massachusetts, 02142, USA

9. Department of Genetics, University Medical Center Utrecht, Utrecht, The Netherlands.
10. Genetics of Complex Traits, University of Exeter Medical School, University of Exeter, Exeter, EX2 5DW, UK.
11. National Institute for Health Research Oxford Biomedical Research Centre, Oxford University Hospital, Old Road, Oxford OX3 7LE, UK.
12. Clinical Trial Service Unit & Epidemiological Studies Unit (CTSU), Nuffield Department of Population Health, Big Data Institute Building, Roosevelt Drive, University of Oxford, Oxford, OX3 7LF, UK.
13. Medical Research Council Population Health Research Unit at the University of Oxford, Nuffield Department of Population Health, University of Oxford, UK.

14. These authors contributed equally to this work

Correspondence to:

****Associate Professor Michael V Holmes**, Medical Research Council Population Health Research Unit at the University of Oxford, Oxford, UK michael.holmes@ndph.ox.ac.uk

*__Professor Anna Murray__, Genetics of Complex Traits, University of Exeter Medical School, University of Exeter, Exeter, UK. A.Murray@exeter.ac.uk

Abstract

Erectile dysfunction (ED) is a common condition affecting more than 20% of men aged over 60 years, yet little is known about its genetic architecture. We performed a genome-wide association study of ED in 6,175 cases among 223,805 European men and identified one locus at 6q16.3 (lead variant rs57989773, OR 1.20 per C-allele; $p = 5.71 \times 10^{-14}$), located between *MCHR2* and *SIM1*. In-silico analysis suggests *SIM1* to confer ED risk through hypothalamic dysregulation. Mendelian randomization provides evidence that genetic risk of type 2 diabetes mellitus is a cause of ED (OR 1.11 per 1-log unit higher risk of type 2 diabetes). These findings provide insights into the biological underpinnings and the causes of ED, and may help prioritize the development of future therapies for this common disease.

Erectile dysfunction (ED) is the inability to develop or maintain a penile erection adequate for sexual intercourse.[1] ED has an age-dependent prevalence, with 20-40% men aged 60-69 years affected.[1] The genetic architecture of ED remains poorly understood, owing in part to a paucity of well-powered genetic association studies. Discovery of such genetic associations can be valuable for elucidating the etiology of ED, and can provide genetic support for potential new therapies.

We conducted a genome-wide association study (GWAS) among white European-ancestry men from the population-based UK Biobank (UKBB) and the Estonian Genome Center of the University of Tartu (EGCUT) cohorts and hospital-recruited Partners HealthCare Biobank (PHB) cohort (Supplemental Methods).

The prevalence of ED (defined as self-reported or physician-reported ED using ICD10 codes N48.4 and F52.2 or use of oral ED medication (sildenafil/Viagra, tadalafil/Cialis or vardenafil/Levitra), or a history of surgical intervention for ED (using OPCS-4 codes: L97.1 and N32.6); Supplemental Methods) in the cohorts was 1.53% (3,050/199,352) in UKBB, 7.04% (1,182/16,787) in EGCUT and 25.35% (1,943/7,666) in PHB (Table S1). Demographic characteristics of the subjects in each cohort are shown in Table S2. The reasons for the different prevalence rates in the three cohorts may include a higher median cohort age for men in PHB (65 years, compared to 59 years in UKBB and 42 years in EGCUT; Table S2), "healthy volunteer" selection bias in UKBB[2], a lack of primary care data availability in UKBB, and intercultural differences, including "social desirability" bias[3,4]. Importantly, we note that the assessment of exposure-outcome relationships remains valid, despite the prevalences likely not being representative of the general population prevalences.

GWAS in UKBB revealed a single genome-wide significant ($p < 5\times10^{-8}$) locus at 6q16.3 (lead variant rs57989773, EAF_{UKBB} (C-allele) = 0.24; OR 1.23; $p = 3.0\times10^{-11}$). Meta-analysis with

estimates from PHB (OR 1.20; $p = 9.84\times10^{-5}$) and EGCUT (OR 1.08; p = 0.16) yielded a pooled meta-analysis OR 1.20; $p = 5.72\times10^{-14}$ (heterogeneity p-value = 0.17; Figures 1A-C). Meta-analysis of all variants yielded no further genome-wide loci. Meta-analysis of our results with previously suggested ED-associated variants also did not result in any further significant loci (Supplemental Methods; Table S3), nor did X-chromosome analysis in UKBB.

The association of rs57989773 was consistent across clinically- and therapy-defined ED, as well as across different ED drug classes (Figure 1C; Figure S1). No further genome-wide significant loci were identified for ED when limited to clinically- or therapy-defined cases (2,032 and 4,142 cases, respectively).

A PheWAS of 105 predefined traits (Table S4) using the lead ED SNP rs57989773 found associations with 12 phenotypes at a p-value $< 5\times10^{-4}$ (surpassing the Bonferroni-corrected threshold of 0.05/105), including adiposity (9 traits), adult height and sleep-related traits. Sex-stratified analyses revealed sexual dimorphism for waist-hip ratio (WHR), systolic and diastolic blood pressure (Figure 1D; Table S5).

The lead variant at the 6q16.3 locus, rs57989773, lies in the intergenic region between *MCHR2* and *SIM1*, with *MCHR2* being the closest gene (distances to transcription start sites of 187kb for *MCHR2* and 284kb for *SIM1*). Conditional and joint analysis (Supplemental Methods) revealed no secondary, independent signals in the locus. Previous work has implicated the *MCHR2-SIM1* locus in sex-specific associations on age at voice-breaking and menarche.[5] The puberty timing-associated SNP in the *MCHR2-SIM1* region (rs9321659; ~500kb from rs57989773) was not in LD with our lead variant ($r^2 = 0.003$, D' = 0.095) and was not associated with ED (p = 0.32) in our meta-analysis, suggesting that the ED locus represents an independent signal.

To identify the tissue and cell types in which the causal variant(s) for ED may function, we examined chromatin states across 127 cell types[6,7] for the lead variant rs57989773 and its proxies (r^2>0.8, determined using HaploReg v4.1 (Supplemental Methods)). Enhancer marks in several tissues, including embryonic stem cells, mesenchymal stem cells and endothelial cells, indicated that the ED-associated interval lies within a regulatory locus (Figure 2A, Table S6).

To predict putative targets and causal transcripts, we assessed domains of long-range three-dimensional chromatin interactions surrounding the ED-associated interval (Figure 2B). Chromosome conformation capture (Hi-C) in human embryonic stem cells[8] showed that *MCHR2* and *SIM1* were in the same topologically associated domain (TAD) as the ED-associated variants, with high contact probabilities (referring to the relative number of times that reads in two 40-kb bins were sequenced together) between the ED-associated interval and *SIM1* (Figure 2B; Figure S2).

This observation was further confirmed in endothelial precursor cells,[9] where Capture Hi-C revealed strong connections between the *MCHR2-SIM1* intergenic region and the *SIM1* promoter (Figure 2C), pointing towards *SIM1* as a likely causal gene at this locus.

We next used the VISTA enhancer browser[10] to examine *in vivo* expression data for non-coding elements within the *MCHR2-SIM1* locus*.* A regulatory human element (hs576), located 30-kb downstream of the ED-associated interval, seems to drive *in vivo* enhancer activity specifically in the midbrain (mesencephalon) and cranial nerve in mouse embryos (Figure 2D). This long-range enhancer close to ED-associated variants recapitulated aspects of *SIM1* expression (Figure 2D), further suggesting that the ED-associated interval belongs to the regulatory landscape of *SIM1*. Taken together these data suggest that the *MCHR2-SIM1* intergenic region harbors a neuronal enhancer and that *SIM1* is functionally connected to the ED-associated region.

Single-minded homolog 1 (*SIM1*) encodes a transcription factor that is highly expressed in hypothalamic neurons.[11] Rare variants in *SIM1* have been linked to a phenotype of severe obesity and autonomic dysfunction,[12,13] including lower blood pressure. A summary of the variant-phenotype associations at the 6q16 locus in human and rodent models is shown in Table S7. Post-hoc analysis of association of rs57989773 with autonomic traits showed nominal association with syncope, orthostatic hypotension and urinary incontinence (Figure S3). The effects on blood pressure and adiposity seen in patients with rare coding variants in *SIM1* are recapitulated in individuals harbouring the common ED-risk variants at the 6q16.3 locus (Figure 1D), suggesting that *SIM1* is the causal gene at the ED-risk locus. Sim1-expressing neurons also play an important role in the central regulation of male sexual behavior as mice that lack the melanocortin receptor 4 (encoded by *MC4R*) specifically in Sim1-expressing neurons show impaired sexual performance on mounting, intromission, and ejaculation.[14] Thus, hypothalamic dysregulation of *SIM1* could present a potential mechanism for the effect of the *MCHR2-SIM1* locus on ED.

An alternative functional mechanism may be explained by proximity of the lead variant (rs57989773) to an arginase 2 processed pseudogene (LOC100129854), a long non-coding RNA (Figure 2A). RPISeq[15] predicts that the pseudogene transcript would interact with the ARG2 protein, with probabilities of 0.70-0.77. Arginine 2 is involved in nitric oxide production and has a previously established role in erectile dysfunction.[16,17] GTEx expression data[18] demonstrated highest mean expression in adipose tissue, with detectable levels in testis, fibroblasts and brain. Expression was relatively low in all tissues however, and there was no evidence that any SNPs associated with the top ED signal were eQTLs for the *ARG2* pseudogene or *ARG2* itself.

As a complementary approach, we also used the Data-driven Expression Prioritized Integration for Complex Traits and GWAS Analysis of Regulatory or Functional Information Enrichment with

LD correction (DEPICT and GARFIELD respectively; Supplemental Methods)[19,20] tools to identify gene-set, tissue-type and functional enrichments. In DEPICT, the top two prioritized gene-sets were 'regulation of cellular component size' and 'regulation of protein polymerization', whereas the top two associated tissue/cell types were 'cartilage' and 'mesenchymal stem cells'. None of the DEPICT enrichments reached an FDR threshold of 5% (Tables S8-10). GARFIELD analyses also did not yield any statistically significant enrichments, therefore limiting the utility of these approaches in this case.

ED is recognized to be observationally associated with various cardiometabolic traits and lifestyle factors[21,22], including type 2 diabetes mellitus (T2D), hypertension, smoking, and others. To further evaluate these associations, we first conducted LD score regression[23,24] to evaluate the genetic correlation of ED with a range of traits. LD score regression identified ED to share the greatest genetic correlation with T2D (r_G = 0.40, nominal p-value = 0.0008; FDR-adjusted p-value = 0.0768; Table S11). Next we performed Mendelian randomization[25] (MR) analyses to evaluate the potential causal role of 9 pre-defined cardiometabolic traits on ED risk (selected based on previous observational evidence linking such traits to ED risk[21], including T2D, insulin resistance, systolic blood pressure, LDL cholesterol, smoking heaviness, alcohol consumption, body mass index, coronary heart disease and educational attainment; Tables S12-S15). MR identified genetic risk to T2D to be causally implicated in ED: each 1-log higher genetic risk of T2D, was found to increase risk of ED by 1.11 (95% CI 1.05-1.17, p = 3.5×10^{-4}, which met our *a priori* Bonferroni-corrected significance threshold of 0.0056 (0.05/9)), with insulin resistance likely representing a mediating pathway[26] (OR 1.36 per 1 standard deviation genetically elevated insulin resistance, 95% CI 1.01-1.84, p = 0.042). Sensitivity analyses were conducted to evaluate the robustness of the T2D-ED estimate (Figure S5, Table S13), including weighted median analyses (OR 1.12, 95% CI 1.02-1.23, p = 0.0230), leave-one-out analysis for all variants (which indicated that no single SNP in the instrument unduly influenced the overall value derived from the summary

IVW estimate[27]) and a funnel plot (showing a symmetrical distribution of single-SNP IV estimates around the summary IVW causal estimate). The MR-Egger regression (intercept p = 0.35) provided no evidence to support the presence of directional pleiotropy as a potential source of confounding[28].

A potential causal effect of systolic blood pressure (SBP) was also identified, with higher SBP being linked to higher risk of ED (MR-Egger OR 2.34, 95% CI 1.26-4.36, p = 0.007, with MR-Egger intercept (p=0.007) suggesting presence of directional pleiotropy). LDL cholesterol showed minimal evidence of a causal effect (OR 1.07, 95% CI 0.98-1.17, p = 0.113), and there was no evidence to support a role for smoking heaviness or alcohol consumption. Genetic risk of coronary heart disease (CHD) showed weak effects on risk of ED, suggesting that pathways leading to CHD may be implicated in ED (OR 1.08, 95% CI 1.00-1.17, p = 0.061). Further, we identified no causal effects of BMI (using a polygenic score or a single SNP in *FTO*) or education on risk of ED.

Genetic variants may inform drug target validation by serving as a proxy for drug target modulation[29]. ED is most commonly treated using phosphodiesterase 5 (PDE5)-inhibitors such as sildenafil. To identify potential phenotypic effects of PDE5 inhibition (e.g. to predict side-effects or opportunities for repurposing), we looked for variants in or around the *PDE5A* gene, encoding PDE5, which showed association with the ED phenotype. Of all 4,670 variants within a 1Mb window of *PDE5A* (chromosome 4:119,915,550 - 121,050,146 as per GRCh37/hg19), the variant with the strongest association was rs115571325, 26Kb upstream from *PDE5A* (OR_{Meta} 1.25, nominal p-value = 8.46×10^{-4}; Bonferroni-corrected threshold (0.05/4,670) = 1.07×10^{-5}; Figure S6). We did not evaluate any further associations for this variant, given the lack of statistically significant association with the ED phenotype after correcting for the multiple testing burden.

We have gained insight into ED, a common condition with substantial morbidity, by conducting a large-scale GWAS and performing several follow-up analyses. By aggregating data from 3 cohorts, including 6,175 ED cases of European ancestry, we identified a locus associated with ED, with several lines of evidence suggesting *SIM1*, highly expressed in the hypothalamus, to be the causal gene at this locus. Our findings provide human genetic evidence in support of the key role of the hypothalamus in regulating male sexual function.[14,30–33]

LD score regression and Mendelian randomization implicated T2D as a causal risk factor for ED with suggestive evidence for insulin resistance and systolic blood pressure, corroborating well-recognized observational associations with these cardiometabolic traits.[22] Further research would be needed to explore the extent to which drugs used in the treatment of T2D might be repurposed for the treatment of ED. A non-causal effect for BMI on ED suggests that the lead SNP (rs57989773) exhibits pleiotropy, driving ED risk independent of its effect on adiposity.

In conclusion, in a large-scale GWAS of more than 6,000 ED cases, the largest to date for this phenotype, we have identified novel biology and elucidated causal effects of various risk factors. Further large-scale GWAS of ED are needed in order to provide additional clarity on its genetic architecture, etiology and shed light on potential new therapies.

Supplemental Data

Document S1. Figures S1–S7 and Supplemental Methods

Spreadsheet S2. Tables S1–S15

Acknowledgements

We thank the UK Biobank (http://www.ukbiobank.ac.uk/; application 11867), Partners HealthCare Biobank (https://biobank.partners.org/), and the Estonian Biobank of the Estonian Genome Center of the University of Tartu (https://www.geenivaramu.ee/en) and their participants.

LD Hub

We gratefully acknowledge all the studies and databases that made GWAS summary data available: ADIPOGen (Adiponectin genetics consortium), C4D (Coronary Artery Disease Genetics Consortium), CARDIoGRAM (Coronary ARtery DIsease Genome wide Replication and Meta-analysis), CKDGen (Chronic Kidney Disease Genetics consortium), dbGAP (database of Genotypes and Phenotypes), DIAGRAM (DIAbetes Genetics Replication And Meta-analysis), ENIGMA (Enhancing Neuro Imaging Genetics through Meta Analysis), EAGLE (EArly Genetics & Lifecourse Epidemiology Eczema Consortium, excluding 23andMe), EGG (Early Growth Genetics Consortium), GABRIEL (A Multidisciplinary Study to Identify the Genetic and Environmental Causes of Asthma in the European Community), GCAN (Genetic Consortium for Anorexia Nervosa), GEFOS (GEnetic Factors for OSteoporosis Consortium), GIANT (Genetic Investigation of ANthropometric Traits), GIS (Genetics of Iron Status consortium), GLGC (Global Lipids Genetics Consortium), GPC (Genetics of Personality Consortium), GUGC (Global Urate and Gout consortium), HaemGen (haemotological and platelet traits genetics consortium), HRgene (Heart Rate consortium), IIBDGC (International Inflammatory Bowel Disease Genetics Consortium), ILCCO (International Lung Cancer Consortium), IMSGC (International Multiple Sclerosis Genetic Consortium), MAGIC (Meta-Analyses of Glucose and Insulin-related traits Consortium), MESA (Multi-Ethnic Study of Atherosclerosis), PGC (Psychiatric Genomics Consortium), Project MinE consortium, ReproGen (Reproductive Genetics Consortium), SSGAC (Social Science Genetics Association Consortium) and TAG (Tobacco and Genetics Consortium),

TRICL (Transdisciplinary Research in Cancer of the Lung consortium), UK Biobank. We gratefully acknowledge the contributions of Alkes Price (the systemic lupus erythematosus GWAS and primary biliary cirrhosis GWAS) and Johannes Kettunen (lipids metabolites GWAS).

Contributions

J.B., L.J., J.C., C.Y.C., T.L.P., S.L. and T.F. collected and analysed the data. All co-authors participated in interpretation of results. J.B., L.J., J.C., S.L., M.N.W., C.M.L., A.M. and M.V.H. wrote the first draft of the manuscript. All authors commented to and edited various versions of the draft manuscript. J.B., C.M.L. and M.V.H. compiled the final manuscript. All authors approved the final manuscript. M.N.W., R.M., B.N., C.M.L., A.M. and M.V.H. supervised the study.

Data Availability

Full summary statistics of the erectile dysfunction genome-wide meta-analysis will be publicly available by the time of publication.

Declaration of Interests

M.N.W. has received speaker fees from Ipsen and Merck. B.N. is SAB of Deep Genomics and Consultant for Avanir Therapeutics. S.L. has a Postdoctoral Research Fellowship funded by Novo Nordisk. M.V.H. works in a unit that receives funding from the UK Medical Research Council and is supported by a British Heart Foundation Intermediate Clinical Research Fellowship (FS/18/23/33512). M.V.H. and C.M.L. are supported by the National Institute for Health Research Oxford Biomedical Research Centre. M.V.H. has collaborated with Boehringer Ingelheim in research, and in accordance with the policy of the Clinical Trial Service Unit and Epidemiological Studies Unit (University of Oxford), did not accept any personal payment

Web Resources

PLINK v1.9

URL: www.cog-genomics.org/plink/1.9/

BOLT-LMM v2.3

URL: https://data.broadinstitute.org/alkesgroup/BOLT-LMM/

EPACTS v3.3.0

https://github.com/statgen/EPACTS

EASYQC v9.2

http://www.uni-regensburg.de/medizin/epidemiologie-praeventivmedizin/genetische-epidemiologie/software/

METAL

http://csg.sph.umich.edu/abecasis/metal/

SNPTEST v2.5.2

https://mathgen.stats.ox.ac.uk/genetics_software/snptest/snptest.html#introduction

LD HUB v1.9.0

http://ldsc.broadinstitute.org/

DEPICT v1

https://github.com/perslab/depict

MendelianRandomization v0.2.2 (R package)

https://cran.r-project.org/web/packages/MendelianRandomization/index.html

HAPLOREG v4.1

http://archive.broadinstitute.org/mammals/haploreg/haploreg.php

GARFIELD v2

https://www.ebi.ac.uk/birney-srv/GARFIELD/

RPISeq v1.0

http://pridb.gdcb.iastate.edu/RPISeq/references.php

GCTA v1.26.0

http://cnsgenomics.com/software/gcta/#Overview

References

1. Lewis, R.W., Fugl-Meyer, K.S., Corona, G., Hayes, R.D., Laumann, E.O., Moreira, E.D., Jr, Rellini, A.H., and Segraves, T. (2010). Definitions/epidemiology/risk factors for sexual dysfunction. J. Sex. Med. 7, 1598–1607.

2. Fry, A., Littlejohns, T.J., Sudlow, C., Doherty, N., Adamska, L., Sprosen, T., Collins, R., and Allen, N.E. (2017). Comparison of Sociodemographic and Health-Related Characteristics of UK Biobank Participants With Those of the General Population. Am. J. Epidemiol. 186, 1026–1034.

3. Kelly, C.A., Soler-Hampejsek, E., Mensch, B.S., and Hewett, P.C. (2013). Social desirability bias in sexual behavior reporting: evidence from an interview mode experiment in rural Malawi. Int. Perspect. Sex. Reprod. Health 39, 14–21.

4. Catania, J.A., Gibson, D.R., Chitwood, D.D., and Coates, T.J. (1990). Methodological problems in AIDS behavioral research: influences on measurement error and participation bias in studies of sexual behavior. Psychol. Bull. 108, 339–362.

5. Day, F.R., Bulik-Sullivan, B., Hinds, D.A., Finucane, H.K., Murabito, J.M., Tung, J.Y., Ong, K.K., and Perry, J.R.B. (2015). Shared genetic aetiology of puberty timing between sexes and with health-related outcomes. Nat. Commun. 6, 8842.

6. Roadmap Epigenomics Consortium, Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A., Heravi-Moussavi, A., Kheradpour, P., Zhang, Z., Wang, J., et al. (2015). Integrative analysis of 111 reference human epigenomes. Nature 518, 317–330.

7. Ernst, J., and Kellis, M. (2015). Large-scale imputation of epigenomic datasets for systematic annotation of diverse human tissues. Nat. Biotechnol. 33, 364–376.

8. Dixon, J.R., Jung, I., Selvaraj, S., Shen, Y., Antosiewicz-Bourget, J.E., Lee, A.Y., Ye, Z., Kim, A., Rajagopal, N., Xie, W., et al. (2015). Chromatin architecture reorganization during stem cell differentiation. Nature 518, 331–336.

9. Wang, Y., Zhang, B., Zhang, L., An, L., Xu, J., Li, D., Choudhary, M.N.K., Li, Y., Hu, M., Hardison, R., et al. (2017). The 3D Genome Browser: a web-based browser for visualizing 3D genome organization and long-range chromatin interactions.

10. Visel, A., Minovitsky, S., Dubchak, I., and Pennacchio, L.A. (2007). VISTA Enhancer Browser--a database of tissue-specific human enhancers. Nucleic Acids Res. 35, D88–D92.

11. Holder, J.L., Jr, Zhang, L., Kublaoui, B.M., DiLeone, R.J., Oz, O.K., Bair, C.H., Lee, Y.-H., and Zinn, A.R. (2004). Sim1 gene dosage modulates the homeostatic feeding response to increased dietary fat in mice. Am. J. Physiol. Endocrinol. Metab. 287, E105–E113.

12. Ramachandrappa, S., Raimondo, A., Cali, A.M.G., Keogh, J.M., Henning, E., Saeed, S., Thompson, A., Garg, S., Bochukova, E.G., Brage, S., et al. (2013). Rare variants in single-minded 1 (SIM1) are associated with severe obesity. J. Clin. Invest. 123, 3042–3050.

13. Michaud, J.L., Boucher, F., Melnyk, A., Gauthier, F., Goshu, E., Lévy, E., Mitchell, G.A., Himms-Hagen, J., and Fan, C.M. (2001). Sim1 haploinsufficiency causes hyperphagia, obesity and reduction of the paraventricular nucleus of the hypothalamus. Hum. Mol. Genet. 10, 1465–1473.

14. Semple, E., and Hill, J.W. (2018). Sim1 Neurons Are Sufficient for MC4R-Mediated Sexual Function in Male Mice. Endocrinology 159, 439–449.

15. Muppirala, U.K., Honavar, V.G., and Dobbs, D. (2011). Predicting RNA-Protein Interactions Using Only Sequence Information. BMC Bioinformatics 12, 489.

16. Cox, J.D., Kim, N.N., Traish, A.M., and Christianson, D.W. (1999). Arginase-boronic acid complex highlights a physiological role in erectile function. Nat. Struct. Biol. 6, 1043–1047.

17. Lacchini, R., Muniz, J.J., Nobre, Y.T.D.A., Cologna, A.J., Martins, A.C.P., and Tanus-Santos, J.E. (2015). Relationship between Arginase 1 and Arginase 2 levels and genetic polymorphisms with erectile dysfunction. Nitric Oxide 51, 36–42.

18. GTEx Consortium (2013). The Genotype-Tissue Expression (GTEx) project. Nat. Genet. 45, 580–585.

19. Pers, T.H., Karjalainen, J.M., Chan, Y., Westra, H.-J., Wood, A.R., Yang, J., Lui, J.C., Vedantam, S., Gustafsson, S., Esko, T., et al. (2015). Biological interpretation of genome-wide association studies using predicted gene functions. Nat. Commun. 6, 5890.

20. Iotchkova, V., Ritchie, G.R.S., Geihs, M., Morganella, S., Min, J.L., Walter, K., Timpson, N.J., UK10K Consortium, Dunham, I., Birney, E., et al. (2016). GARFIELD - GWAS Analysis of Regulatory or Functional Information Enrichment with LD correction.

21. Selvin, E., Burnett, A.L., and Platz, E.A. (2007). Prevalence and risk factors for erectile dysfunction in the US. Am. J. Med. 120, 151–157.

22. Yafi, F.A., Jenkins, L., Albersen, M., Corona, G., Isidori, A.M., Goldfarb, S., Maggi, M., Nelson, C.J., Parish, S., Salonia, A., et al. (2016). Erectile dysfunction. Nat Rev Dis Primers 2, 16003.

23. Bulik-Sullivan, B.K., Loh, P.-R., Finucane, H.K., Ripke, S., Yang, J., Schizophrenia Working Group of the Psychiatric Genomics Consortium, Patterson, N., Daly, M.J., Price, A.L., and Neale, B.M. (2015). LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. Nat. Genet. 47, 291–295.

24. Bulik-Sullivan, B., Finucane, H.K., Anttila, V., Gusev, A., Day, F.R., Loh, P.-R., ReproGen Consortium, Psychiatric Genomics Consortium, Genetic Consortium for Anorexia Nervosa of the Wellcome Trust Case Control Consortium 3, Duncan, L., et al. (2015). An atlas of genetic correlations across human diseases and traits. Nat. Genet. 47, 1236–1241.

25. Davey Smith, G., and Ebrahim, S. (2003). “Mendelian randomization”: can genetic epidemiology contribute to understanding environmental determinants of disease? Int. J. Epidemiol. 32, 1–22.

26. Wang, Q., Holmes, M.V., Davey Smith, G., and Ala-Korpela, M. (2017). Genetic Support for a Causal Role of Insulin Resistance on Circulating Branched-Chain Amino Acids and Inflammation. Diabetes Care 40, 1779–1786.

27. Davies, N.M., Holmes, M.V., and Davey Smith, G. (2018). Reading Mendelian randomisation studies: a guide, glossary, and checklist for clinicians. BMJ 362, k601.

28. Holmes, M.V., Ala-Korpela, M., and Smith, G.D. (2017). Mendelian randomization in cardiometabolic disease: challenges in evaluating causality. Nat. Rev. Cardiol. 14, 577.

29. Walker, V.M., Davey Smith, G., Davies, N.M., and Martin, R.M. (2017). Mendelian randomization: a novel approach for the prediction of adverse drug events and drug repurposing opportunities. Int. J. Epidemiol. 46, 2078–2089.

30. Argiolas, A., and Melis, M.R. (2005). Central control of penile erection: role of the paraventricular nucleus of the hypothalamus. Prog. Neurobiol. 76, 1–21.

31. Argiolas, A., and Melis, M.R. (2013). Neuropeptides and central control of sexual behaviour from the past to the present: a review. Prog. Neurobiol. 108, 80–107.

405 32. Courtois, F., Carrier, S., Charvier, K., Guertin, P.A., and Journel, N.M. (2013). The control of
406 male sexual responses. Curr. Pharm. Des. 19, 4341–4356.

407 33. Giuliano, F., and Rampin, O. (2004). Neural control of erection. Physiol. Behav. 83, 189–201.

408 34. Ward, L.D., and Kellis, M. (2012). HaploReg: a resource for exploring chromatin states,
409 conservation, and regulatory motif alterations within sets of genetically linked variants. Nucleic
410 Acids Res. 40, D930–D934.

411 35. Javierre, B.M., Burren, O.S., Wilder, S.P., Kreuzhuber, R., Hill, S.M., Sewitz, S., Cairns, J.,
412 Wingett, S.W., Várnai, C., Thiecke, M.J., et al. (2016). Lineage-Specific Genome Architecture
413 Links Enhancers and Non-coding Disease Variants to Target Gene Promoters. Cell 167, 1369–
414 1384.e19.

415 36. Casper, J., Zweig, A.S., Villarreal, C., Tyner, C., Speir, M.L., Rosenbloom, K.R., Raney, B.J.,
416 Lee, C.M., Lee, B.T., Karolchik, D., et al. (2017). The UCSC Genome Browser database: 2018
417 update. Nucleic Acids Res.

Figure Legends

Figure 1. 6q16.3 (LEAD VARIANT rs57989773) IS AN ED-ASSOCIATED LOCUS AND EXHIBITS PLEIOTROPIC PHENOTYPIC EFFECTS.

(A) **Genome-wide meta-analysis revealed a single genome-wide significant locus for ED at 6q16.3**.

(B) **Six genome-wide significant variants at 6q16.3 are in high LD**.

(C) **The association of rs57989773 with ED shows a consistent direction of effect across the three cohorts and across clinically- and therapy-defined ED in UKBB**.

(D) **PheWAS reveals sex-specific associations of rs57989773 with waist-hip ratio and blood pressure**. A PheWAS of 105 predefined traits using the lead ED SNP rs57989773 found associations with 12 phenotypes at p-value < 4.8×10^{-4} (surpassing the Bonferroni-corrected threshold of 0.05/105; Table S4). Due to the nature of the ED phenotype and previously reported sex-specific effects in the *MCHR2-SIM1* locus, sex-specific analyses were performed in significant traits. Diastolic blood pressure (dbp) and systolic blood pressure (sbp) are included here (despite not meeting the Bonferroni-corrected threshold in the original analysis), due to previous reports of effects on blood-pressure in patients with rare, coding variants in *SIM1* Sexual heterogeneity was found to be significant (surpassing a Bonferroni-corrected threshold of 0.05/7 for the number of traits where sex-specific analyses were conducted) for diastolic blood pressure ($p\text{-value}_{heterogeneity} = 6.52 \times 10^{-3}$), systolic blood pressure ($p\text{-value}_{heterogeneity} = 3.73 \times 10^{-3}$), waist to hip ratio (whr; $p\text{-value}_{heterogeneity} = 2.39 \times 10^{-6}$) and waist to hip ratio adjusted for BMI ($p\text{-value}_{heterogeneity} = 1.77 \times 10^{-5}$). This plot only shows sex-specific estimates for traits showing significant sexual heterogeneity. Continuous traits were standardised prior to analysis to facilitate comparison.

Figure 2. FUNCTIONAL ANALYSIS OF 6q16.3 IMPLICATES *SIM1* IN ED PATHOGENESIS

(A) **ED-associated signal overlaps regulatory annotations in embryonic stem cells.** Chromatin state annotations for the ED-associated region across 127 reference epigenomes (rows) for cell and tissue types profiled by the Roadmap Epigenomics Project[6,7]. Grey vertical lines indicate the position of the ED-associated variant (rs57989773) and its proxies that are in LD r^2>0.8 determined using HaploReg v4.1[34] (rs17789218, rs9496567, rs78677597, rs9496614, and rs17185536). The lead variant is in proximity to 'RP3-344J20.1', an arginase 2 processed pseudogene (LOC100129854).

(B) **The ED-associated interval is functionally connected to SIM1 in embryonic stem cells.** The 3D Genome Browser[9] was used to visualize chromosome conformation capture (Hi-C) interactions contact probabilities in human embryonic stem cells[8], revealing high contact probability between the ED-associated region (highlighted in yellow) and *SIM1* at 40-kb resolution. The heat map values on a color scale correspond to the number of times that reads in two 40-kb bins were sequences together (blue - stronger interaction, white - little or no interaction).

(C) **The *MCHR2-SIM1* intergenic region forms functional connections to the *SIM1* promoter in endothelial progenitors.** The 3D Genome Browser[9] was used to visualize Capture Hi-C in endothelial precursors[35]. Light blue vertical line indicates position of the ED-associated interval.

(D) **The *MCHR2-SIM1* intergenic region harbors a neuronal enhancer.** Upper panel: Position of human element hs576 (blue vertical line) and the ED-associated variant rs57989773 and its 5 proxies in r^2>0.8 (rs17789218, rs9496567, rs78677597, rs9496614, rs17185536). hs576 is flanked by genes *MCHR2-AS1* and *SIM1.* This panel was generated using the UCSC genome browser[36]. Lower panel: Expression pattern of human element hs576 in a mouse embryo at e11.5. Expression pattern shows that hs576 drives *in vivo* enhancer activity specifically in mesencephalon (midbrain) and cranial nerve. Expression data were derived from the VISTA enhancer browser[10].