

SUPPLEMENTARY METHODS

Supplementary Methods A: Sleep Trait Questionnaires in UK Biobank

At baseline, participants completed a touchscreen questionnaire, which included questions about their sleep behaviours.

Chronotype (morning or evening preference) was assessed in the question “Do you consider yourself to be?” with one of six possible answers: “Definitely a ‘morning’ person,” “More a ‘morning’ than ‘evening’ person,” “More an ‘evening’ than a ‘morning’ person,” “Definitely an ‘evening’ person,” “Do not know,” or “Prefer not to answer”, from which we derived a five category variable for chronotype where “Definitely a ‘morning’ person” and “Definitely an ‘evening’ person” define either extreme.

Sleep duration was assessed by asking: “About how many hours sleep do you get in every 24 hours? (Including naps).” The answer could only contain integer values.

To assess insomnia symptoms, participants were asked: “Do you have trouble falling asleep at night or do you wake up in the middle of the night?” where responses “Never/rarely”, “Sometimes” or “Usually” were coded as a three-category variable.

Frequency of napping was assessed based on the question: “Do you have a nap during the day?”, with possible responses “Never/rarely”, “Sometimes” and “Usually” coded as a three-category variable.

The frequency of daytime-sleepiness using the question: “How likely are you to doze off or fall asleep during the daytime when you don’t mean to? (e.g.: when working, reading, or driving)”, with the answers coded as a four-category variable “never”, “sometimes”, “often”, or “all of the time”.

Ease of getting up in the morning was assessed using the question: “On an average day, how easy do you find getting up in the morning?”, with the answers coded as a four-category variable “not at all easy”, “not very easy”, “fairly easy”, or “very easy”.

For all UKB-derived variables, answers “Prefer not to answer” were coded as missing.

Supplementary Methods B: Genotyping and Imputation (UKB)

Individuals were genotyped using the UK BiLEVE array (N=49,979) and the UK Biobank axion array (N=438,398), with the full data release comprising of N=488,377 successfully genotyped samples. Multiallelic SNPs or those with a MAF of $\leq 1\%$ were removed prior to phasing, which was then performed using a modified version of SHAPEIT2 [2]. Genotype imputation was performed using IMPUTE2 algorithms [3,4]. Further details for pre-imputation quality control, phasing and imputation are described elsewhere [5].

Individuals with mismatch between genetic sex and reported sex or with sex-chromosome aneuploidy were excluded from the analysis (N=814). Remaining samples were restricted to individuals of 'European' ancestry, defined using four principal components provided by UKB (N=464,708) [6].

GWAS in this study were conducted using the linear mixed model (LMM) association method as implemented in BOLT-LMM (v2.3) [7]. Population structure was modelled using directly genotyped SNPs, obtained after filtering on MAF > 0.01; Hardy-Weinberg equilibrium p-value < 0.0001 and LD pruning to an r^2 threshold of 0.001 using PLINKv2.00.

Supplementary Methods C: Sleep Trait Questionnaires in 23andMe

23andMe is a personal genetics company[1], in which participants are given the option to consent to their data being used in research. In 23andme, an insomnia phenotype was determined through completion of questionnaires through online surveys.

Participants were asked to answer one or more questions related to seven sleep-related traits.

Participants with positive response to any of the following questions were considered as cases: 1) 'Have you ever been diagnosed with, or treated for, insomnia?', 2) 'Were you diagnosed with insomnia?', 3) 'Have you ever been diagnosed by a doctor with any of the following neurological conditions?' (Sleep disturbance), 4) 'Do you routinely have trouble getting to sleep at night?', 5) 'What sleep disorders have you been diagnosed with? Please select all that apply.' (Insomnia, trouble falling or staying asleep), 6) 'Have you ever taken these medications?' (Prescription sleep aids) and 7) 'In the last 2 years, have you taken any of these medications?' (Prescription sleep aids). Participants who did not provide either positive or uncertain answer ('I don't know' or 'I am not sure') in any of the 7 questions nor the followings are considered as controls: 1) 'Have you ever been diagnosed with, or treated for, any of the following conditions?' (Insomnia; Narcolepsy; Sleep apnoea; Restless leg syndrome), 2) 'In the past 12 months, have you been newly diagnosed with any

of the following conditions by a medical professional?’ (Insomnia; Sleep apnoea; Migraines), 3) ‘Have you ever been diagnosed with or treated for any of the following conditions?’ (Posttraumatic stress disorder; Autism; Asperger’s; Sleep disorder), 4) ‘Have you ever been diagnosed with or treated for a sleep disorder?’, and 5) ‘Have you ever been diagnosed with or treated for any of the following conditions?’ (A sleep disorder).

Supplementary Methods D: Genotyping and Imputation (BCAC)

Patients were genotyped with iCOGS and OncoArray arrays [8,9], and were then restricted to variants with a minor allele frequency (MAF) > 0.01 and to participants of ‘European’ ancestry, based on genotype data. Non-genotyped variants were either imputed based on the 1000 Genomes Project Phase 3 (October 2014) release as reference panel, or using a reference panel from the Haplotype Reference Consortium (HRC) [10], to improve imputation quality for rarer variants.

Further details of methods related to genotyping and genotype imputation have been described previously [11–13].

REFERENCES

1. 23andMe [Internet]. Available from: <https://www.23andme.com/en-gb/>
2. 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(7):817–20.
3. Huang J, Howie B, McCarthy S, Memari Y, Walter K, Min JL, et al. Improved imputation of low-frequency and rare variants using the UK10K haplotype reference panel. *Nat Commun.* 2015;6(1):8111.
4. Howie B, Marchini J, Stephens M. Genotype imputation with thousands of genomes. *G3 (Bethesda).* 2011/11/01. 2011 Nov;1(6):457–70.
5. 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(7726):203–9.
6. Mitchell RE, Hemani G, Dudding T, Corbin L, Harrison S, Paternoster L. UK Biobank Genetic Data : MRC-IEU Quality Control , version 2 , 18 / 01 / 2019. 2019;
7. 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(3):284–90.
8. Amos CI, Dennis J, Wang Z, Byun J, Schumacher FR, Gayther SA, et al. The OncoArray Consortium: A Network for Understanding the Genetic Architecture of Common Cancers. *Cancer Epidemiol biomarkers Prev a Publ Am Assoc Cancer Res cosponsored by Am Soc Prev Oncol.* 2017 Jan;26(1):126–35.
9. Michailidou K, Hall P, Gonzalez-Neira A, Ghoussaini M, Dennis J, Milne RL, et al. Large-scale genotyping identifies 41 new loci associated with breast cancer risk. *Nat Genet.* 2013 Apr;45(4):353–61, 361e1-2.
10. McCarthy S, Das S, Kretzschmar W, Delaneau O, Wood AR, Teumer A, et al. A reference panel of 64,976 haplotypes for genotype imputation. *Nat Genet.* 2016;48(10):1279–83.
11. Amos CI, Dennis J, Wang Z, Byun J, Schumacher FR, Gayther SA, et al. The oncoarray consortium: A network for understanding the genetic architecture of common cancers. *Cancer Epidemiol Biomarkers Prev.* 2017;26(1):126–35.
12. Guo Q, Schmidt MK, Kraft P, Canisius S, Chen C, Khan S, et al. Identification of novel genetic

markers of breast cancer survival. *J Natl Cancer Inst.* 2015;107(5):1–9.

13. Escala-Garcia M, Guo Q, Dörk T, Canisius S, Keeman R, Dennis J, et al. Genome-wide association study of germline variants and breast cancer-specific mortality. *Br J Cancer.* 2019 Mar;120(6):647–57.