



AALBORG UNIVERSITY
DENMARK

Aalborg Universitet

The genetic architecture of medication-use

Rohde, Palle Duun; Sørensen, Peter; Nyegaard, Mette

DOI (link to publication from Publisher):
[10.1101/2020.09.30.20204438](https://doi.org/10.1101/2020.09.30.20204438)

Creative Commons License
CC BY-ND 4.0

Publication date:
2020

Document Version
Publisher's PDF, also known as Version of record

[Link to publication from Aalborg University](#)

Citation for published version (APA):
Rohde, P. D., Sørensen, P., & Nyegaard, M. (2020). *The genetic architecture of medication-use*. medRxiv. <https://doi.org/10.1101/2020.09.30.20204438>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal -

Take down policy

If you believe that this document breaches copyright please contact us at vbn@aub.aau.dk providing details, and we will remove access to the work immediately and investigate your claim.

1 **The genetic architecture of medication-use**

2
3 **Palle Duun Rohde^{1,*}, Peter Sørensen² & Mette Nyegaard³**

4
5 ¹ Department of Chemistry and Bioscience, Aalborg University, 9220 Aalborg, Denmark

6 ² Centre for Quantitative Genetics and Genomics, Aarhus University, 8830 Tjele, Denmark

7 ³ Department of Biomedicine, Aarhus University, 8000 Aarhus, Denmark

8 * Corresponding author: palleldr@bio.aau.dk

9 **Abstract**

10 Genomics has been forecasted to revolutionise human health by improving medical treatment through a better
11 understanding of the molecular mechanisms of human diseases. Despite great successes of the last decade's
12 genome-wide association studies (GWAS), the results have to a limited extent been translated to genomic medicine.
13 We propose, that one route to get closer to improved medical treatment is by understanding the genetics of
14 medication-use. Here we obtained entire medication profiles from 335,744 individuals from the UK Biobank and
15 performed a GWAS to identify which common genetic variants are major drivers of medication-use. We analysed 9
16 million imputed genetic variants, estimated SNP heritability, partitioned the genomic variance across functional
17 categories, and constructed genetic scores for medication-use. In total, 59 independent loci were identified for
18 medication-use and approximately 18% of the total variation was attributable to common genetic (minor allele
19 frequency >0.01) variants. The largest fraction of variance was captured by variants with low to medium minor
20 allele frequency. In particular coding and conserved regions, as well as transcription start sites, displayed
21 significantly enrichment of heritability. The average correlation between medication-use and predicted genetic
22 scores was 0.14. These results demonstrate that medication-use *per se* is a highly polygenic complex trait and that
23 individuals with higher genetic liability are on average more diseased and have a higher risk for adverse drug
24 reactions. These results provide an insight into the genetic architecture of medication use and pave the way for
25 developments of multicomponent genetic risk models that includes the genetically informed medication-use.

26

27 Introduction

28 Understanding the relationship between DNA sequence variation and the predisposition to common diseases has
29 interested researchers for decades. In particular after the initial release of the human genome (1), the number of
30 polymorphic genetic variants associated with disease predisposition has grown exponentially to more than 60,000
31 associations (2–4). Genome-wide association studies (GWAS) have provided new insight into the biology and
32 genetic epidemiology of many human complex diseases, which is essential for innovative developments within
33 genomic medicine.

34 The traditional *one-size-fits-all* approach for disease diagnosis and treatment has been proven to be inefficient,
35 expensive and sometimes with adverse clinical reactions. Personalised medicine or genomic medicine, which often
36 are used interchangeably (5), is foreseen to change the way we prevent, diagnose and treat medical conditions
37 (6,7). The goal of genomic medicine is to develop targeted preventive or treatment approaches based on the
38 individual's genetic makeup, environmental exposures and lifestyle parameters (6,7). Fundamental for
39 development of genomic medicine is accurate knowledge about the disease pathogenesis. Equally important is the
40 realisation of the genetic contributions to drug response *per se* (8–10). Genetic variation among patients modulates
41 the drug efficiency and can impose toxic effects (adverse drug reactions) (11,12). Therefore, understanding the
42 genetic architecture of drug response is absolutely essential for the development of genomic medicine.

43 A major challenge and hindrance in studying genetic factors influencing drug response variability is the lack
44 of accessible data. Despite the emerge of large biobanks such as the United Kingdom Biobank (13), Japan Biobank
45 (14) and Estonia Biobank (15), that contain genetic and deep phenotypic information of the participants,
46 information on response to medical treatment is absent. The accessibility of electronic health records and self-
47 reported health status may provide means for alternative approaches for studying the genetic basis of traits of
48 relevance for medication-use.

49 Previously, Wu *et al.* (2019)(16) performed a genetic analysis of medication-use in the United Kingdom
50 Biobank (UKBB). They categorised medications based on the drugs active substances according to the organ or
51 system they act on and their pharmacological properties. They performed a genetic analysis on 23 isolated groups
52 of medications and identified a very large number of independent loci associated with the different drug categories.
53 Instead of grouping drugs based on the active substances and analysing those individually, we investigated the
54 entire medication profile of individuals from the UKBB. The aim of the current study was to investigate the genetic
55 basis of self-reported medication-use in the UKBB. We defined medication-use as the total number of different
56 prescription and over-the-counter medications individuals were taking at the time of the verbal interview with the

57 UKBB assessment centres. Medication-use – as defined above – has the clear advantage that it is easily quantifiable
58 compared to for example drug efficiency. Assuming current medication-use as a quantitative trait phenotype we
59 conducted a genetic analysis of 335,744 unrelated individuals from the UKBB. We hypothesised that across medical
60 conditions, medication-use has a detectable genetic component, and expected medication-use to be genetically
61 correlated with common diseases as commonly prescribed drugs is likely to be a proxy for major disease groups.
62 Extensive medication-use has in individuals older than 65 years been shown to be associated with ill health and
63 morbidity (17,18); hence, we further hypothesised that medication-use was genetically correlated with health-
64 related outcomes. If medication-use has a genetic component, understanding the genetic basis is important because
65 many medications has side effects, and increased drug usage might be associated with higher risk of toxic effects.
66 Hence, genetic predisposition to high medication-use could be used as guidance in treatment plans aiming to reduce
67 the total number of medications.

68

69 Results

70 We performed a genetic analysis of current medication-use within the white British cohort of the UK Biobank
71 ($n=335,744$). We defined current medication-use as the number of different prescription and over-the-counter
72 medicines the participants were taking regularly at the time of the verbal interview (short term medication, like
73 antibiotics or analgesics was not included). The average number of medications taken by males were 2.34 (standard
74 deviation (sd) 2.7) and 2.67 (sd 2.7) for females, with a linear increase in the number of medications taken with
75 increasing age ([Supplementary Figure S1](#)). Interestingly, the mean number of drugs taken for individuals with an
76 ICD10-code for adverse drug reaction ([Supplementary Table S1](#))(19) was significant larger (mean=4.09, sd=3.57)
77 than individuals without such diagnoses (mean=2.23, sd=2.43; $t\text{-test}=-101.77$, $df=46662$, $P<2.2\times 10^{-16}$), suggesting
78 that individuals taking a larger number of medications are more likely to encounter an adverse drug reaction or that
79 individuals experiencing adverse drug reactions are harder to treat, requiring more medication. . The participants
80 have reported a total of 3,247 different medications, where the most frequently used drugs were paracetamol
81 ($n=61,604$), aspirin ($n=44,894$), ibuprofen ($n=41,756$) and simvastatin ($n=38,379$) ([Supplementary Table S2](#)).

82 After SNP quality control (see Methods) there were 9,804,629 autosomal SNPs left for GWAS analysis. In total,
83 we identified 59 independent quantitative trait loci for current medication-use ([Fig. 1, Supplementary Table S3](#)).
84 The strongest associated locus was within the human leucocyte antigens (HLA) complex ($rs35248896$, $P=1.52\times 10^{-}$
85 46). Because of the complexity of the HLA region (20,21) we only allowed one significant locus at this genomic region

(additional three loci passed the significance threshold within the HLA region but were excluded, [Supplementary Table S3](#)). Among the 59 genome-wide significant loci 14 of them were located in intergenic regions.

Using the medication-use GWAS summary statistics (excluding the HLA region) we estimated the proportion of variation in medication-use explained by the SNPs (h_{SNP}^2) to 0.18 ± 0.005 . Next, we partitioned the total heritability to heritability captured by individual autosomal chromosomes and found a linear association between the proportion of heritability captured by each autosomal chromosome and the number of SNPs per chromosome ($R^2=0.9$, [Fig. 2A](#)) suggesting that medication-use is a highly polygenic trait. The genomic variance explained per variant within minor allele frequency bins indicate that low frequent variants capture about 3 times more genetic variance than higher frequent variants does ([Fig. 2B](#)). Also, we estimated the enrichment score across 24 functional categories (obtained from Finucane et al (2015) (22)), which is the estimated share of h_{SNP}^2 divided by its expected share under the assumed heritability model (22,23) ([Fig 2C](#)). In particular, the conserved genomic region and transcription start sites (TSS) were highly enriched accounting for 4.5% and 1.3% of h_{SNP}^2 , respectively.

We constructed genetic scores for medication-use by re-estimating the SNP-effects using a five-fold cross validation scheme. The scoring was performed on five levels of LD pruning (r^2) and across eleven P -value thresholds. The maximum prediction accuracy (correlation between genetic scores and medication-use ~ 0.14) was obtained when removing markers with $r^2 > 0.5$ ([Supplementary Fig. S2](#)) at a P -value of 0.9 ([Fig. 3A](#)), which includes about 1.5 million genetic markers ([Supplementary Fig. S3](#)). Stratifying individuals based on their genetic score we see that individuals within the top 5% highest genetic score has increased medication-use compared with individuals with the 5% lowest genetic score ([Fig. 3B](#), [Supplementary Fig. S4](#)). Moreover, individuals with the 5% highest genetic scores has significant more ICD10 diagnoses than those individuals with the 5% lowest genetic scores (10.8 diagnoses and 6.4 diagnoses, respectively, [Supplementary Fig. S5](#)). There is, however, no visual difference in which diseases they have been diagnosed with ([Supplementary Fig. S6](#)); individuals with the highest genetic score for medication-use simply has more diagnoses than those with low genetic scores ([Supplementary Fig. S5B and Fig. S6](#)). Using the classification of adverse drug reactions from Hohl *et al.* (2014) (19) ([Supplementary Table S1](#)), we further see, that individuals with predicted high risk and an experience of adverse drug reactions on average have a 1.6 fold higher medication-use compared with low risk individuals ([Supplementary Fig. S7 and Table S5](#)).

Finally, we computed the genetic correlation between medication-use and 257 quantitative traits and complex diseases using LD Hub (24). We found a significant genetic correlation with 115 traits (Bonferroni adjusted P value < 0.01) across 26 categories, except bone traits, where no genetic correlation with medication use was found ([Fig. 4](#), [Supplementary Table S4](#)). Expectedly, medication-use was positively genetically correlated with major common

117 complex diseases in particularly coronary artery disease, type 2 diabetes, asthma, lung cancer and major depressive
118 disorder (Fig. 5). Parents age at death was the trait that was most negatively genetically correlated with medication-
119 use indicating that higher medication use correlate with lower age at death (higher mortality) of parents (Fig. 5).
120 We also found that the number of years in school and if completed college were negatively correlated with
121 medication-use. Number of cigarettes smoked per day and medication-use was positively genetically correlated,
122 and medication-use was also genetically correlated with sleep traits; insomnia and sleep duration (Fig. 5). Finally,
123 we estimated genetic correlations between medication-use and the medication categories previously published by
124 Wu *et al.* (2019) (16) (Supplementary Figure S8). The average genetic correlation between our definition of
125 medication-use and the 23 medication categories was 0.52 (sd 0.22; Supplementary Table S6), and the category
126 'drugs affecting bone structure and mineralization' was the only insignificant result, which is in agreement with the
127 observation that medication-use was not genetically correlated with any bone traits (Fig. 4).

129 Discussion

130 In this study, we used data from the UK Biobank to perform a genetic analysis of current medication-use, defined
131 as the total number of different prescription and over-the-counter drugs UKBB participants were taking at the time
132 of the initial assessment. Our aim was to investigate the genetic basis of self-reported medication-use. We identified
133 59 linkage disequilibrium independent SNPs ($P < 5 \cdot 10^{-8}$) associated with current medication-use (Fig. 1). The
134 strongest signal was located within the major histocompatibility complex (MHC); *HLA-DQA1* (lead SNP
135 rs35248896, P value = $1.52 \cdot 10^{-46}$), which belongs to the MHC class II gene. The MHC region is a large genomic
136 region on chromosome 6 that is associated with more diseases than any other region of the genome (20,25).
137 Additional three loci within MCH reached LD-independently genome-wide significance (Supplementary Table S3),
138 however, the complexity and extreme variant polymorphism, combined with strong LD, within MHC complicates
139 interpretation and disentanglement of individual loci (26). Given the biological involvement of MHC in immune
140 response, it was unsurprisingly that exactly this region was the strongest associated loci.

141 The associated loci have previously been linked to a large number of different quantitative traits and complex
142 diseases. Since the genetic correlations express the extent to which two measurements reflect what is genetically
143 the same character (27), it is not too surprisingly that we see a good correspondence between the identified loci
144 (Fig. 1) and their previous associations and the significant genetic correlations (Fig. 5, Supplementary Table S4).
145 For example, among our candidate genes (Supplementary Table S3) we have susceptibility loci for diabetes
146 (*PTPN22*, *CEP68*, *RREB1*, *TCF7L2* (28–30)), coronary artery disease (*PSRC1*, *UNC5C*, *LPLA* (31)), depression

147 (*MAD1L1*, *YLPM1* (32,33)), and insomnia (*NMT1* (34)). But also for non-disease traits including BMI (*RABGAP1L*,
148 *HEYL* (35,36)), smoking (*NLGN1* (37)), and age at menarche (*RBM6* (38)).

149 The polygenic nature of medication-use was, aside from the large number of identified quantitative trait loci,
150 further supported by the linear association between proportion of genetic variance captured by each autosomal
151 chromosome and the proportion of genetic variants (Fig. 2A), which is similar to what is seen for other complex,
152 polygenic traits (39–43). We observed that low frequent alleles capture more genetic variance than common
153 variants (Fig. 2B), which is similar to what is seen in for example type 2 diabetes (44) and coronary artery disease
154 (42,45). Neurodevelopmental and -degenerative disorders, like schizophrenia, Tourette syndrome and Alzheimer
155 disease, do, however, show the opposite pattern (39,40,46). Therefore, it is not surprisingly that our findings
156 correspond with what is seen for common diseases since the disease prevalence of common diseases in the UKBB
157 follows the population prevalence, and further that the prevalence of mental disorders is too low compared to the
158 population frequency.

159 Our statistical genetic analysis was performed across any medical condition the included UKBB participants
160 may be diagnosed with as our interest was to study the genetic contributions to variation in medication-use. Thus,
161 our findings could be biased towards the common diseases with the highest disease prevalence, for example
162 through partially shared genetic aetiology. However, this would inevitable imply that some disease groups require
163 numerically more drugs for treatment than other disease groups. Our results clearly demonstrated that individuals
164 with high genetic score for medication-use had an increased medication usage (Fig. 3B), and that this was
165 associated with higher probability of have more disease diagnoses (Supplementary Fig. S5) and having experienced
166 an adverse drug reaction (Supplementary Fig. S7). Thus, it appears, that the individuals with large medication-use
167 has a genetic predisposition towards increased medication demand likely because of poor health from being
168 predisposed to a larger number of different diseases, leading to increased risk of responding negatively to the
169 prescribed medical treatment. Supportive for this, is that we see medication-use to be genetically correlated with
170 known indicators of poor health (Fig. 5). We found, for example, overweight and high body mass index, known to
171 be under strong genetic influence (47), to be positively correlated with medication-use, and these are strongly
172 associated with poor health (48–50). Moreover, the behavioural characteristics smoking and sleep patterns, are
173 also known factors for bad state of health (51,52), which also displayed significant genetic correlations with
174 medication-use (Fig. 5).

175 This study has a number of limitations. First, despite the information on medication-use was obtained by
176 trained nurses during interviews, the same drug may have been reported under different names, which may limit
177 the accuracy of the analysis. Second, our definition of drug usage also includes vitamins. However, for many

178 diseases, dietary supplements are used in the pharmacological intervention, thus, to prune out vitamins might not
179 more accurately capture the individual's medication profile. Third, the lack of information on medication duration,
180 dosage and response entails that true pharmacogenomic analysis cannot be performed. Fourth, the self-reported
181 nature of the phenotype may limit the precision; however, studies have shown good concordance between self-
182 reported and electronic health records, both for diseases (53) and for medication-use (54). Fifth, the findings
183 presented here are specific to the UK Biobank participants, which is not representative of the general UK population
184 (55,56), and may not translate to other populations and other health systems.

185 In conclusion, we have shown that the genetic basis of current medication-use in the UK Biobank among
186 335,744 individuals appears genetically heterogeneous. We identified 59 independent quantitative trait loci for
187 medication-use and found that 18% of the observed variation could be ascribed to common genetic variants. The
188 genetic heterogeneous nature of medication-use was further supported as the genetic variance was spread across
189 the genome, and that the highest accuracy of prediction was observed when including 1.5 million genetic markers.
190 Our findings that medication-use was strongly correlated with smoking behaviour, parents age of death,
191 educational level and insomnia, suggest that the genetic architecture is not biased towards common diseases.
192 Understanding the genetic aetiology of complex diseases has been suggested as a route for improving medical
193 treatment (57,58). Medication-use – as defined in the current study – is an easy quantifiable trait, and because of
194 its genetically correlated nature with many complex traits and diseases, incorporating such information into new
195 multicomponent genetic risk scores (59) could further increase the accuracy of predicting disease predispositions
196 and disease trajectories. From genetic data we can identify those individuals with high medication-use, which
197 concurrently are the most diseased individuals, and they are at an increased risk for adverse drug reaction. Thus,
198 individual medication profiles are likely to be yet another puzzle piece in understanding complex for human
199 diseases, and for providing better medical treatment for future generation.

200

201 **Material and Methods**

202 **Genotype and phenotype data**

203 Genetic and phenotypic data were obtained from the United Kingdom Biobank (UKB) (13,60). Data has been
204 collected for more than 500,000 individuals aged 37-73 years. Genotyping details has been described previously
205 (13). To obtain a genetic homogeneous study population we restricted our analyses to unrelated Caucasians, and
206 excluded individuals with more than 5,000 missing genotypes or individuals with autosomal aneuploidy
207 ($n=335,744$). We first converted the imputed genotype probabilities to hard-call genotypes using PLINK2 (--hard-

call 0.1) (61), and variants with MAF < 0.01, missing genotype rate >0.05, Hardy-Weinberg equilibrium test P value < 1×10^{-6} , or imputation info score <0.3 were excluded ($m = 9.804.629$).

We defined current medication-use as the number of different prescription and over-the-counter medicines (Data Field: 20003) the participants were taking regularly at the time of the verbal interview. Any short-term medications, like antibiotics or analgesics were not registered at the interview. From Hohl *et al.* (2013)(19) we obtained a list of ICD10 codes commonly used to describe adverse drug reactions ([Supplementary Table S1](#)).

Genome-wide association study (GWAS) of medication-use

Using PLINK2 (61) we performed a GWAS on medication-use for the entire White-British cohort with sex, age, UKB assessment centre, and the first ten genetic principal components as covariates. To identify high confidence independent associated loci, we performed LD-based clumping in window size of 1000 kb with $r^2 < 0.01$, and for the major histocompatibility complex region we only allowed one significant locus. Lead SNPs within each independent genome-wide significant locus was annotated to nearest gene (genome build GRCh37, hg19) within 2000 kb using Variant Effect Predictor (62).

Estimation of heritability and genetic correlations

We estimated the proportion of variation in medication-use explained by SNPs (h_{SNP}^2) using SumHer (23). We excluded variants within the HLA region (as suggested by the authors of SumHer), and the estimation was performed assuming the LDK heritability model. In addition, we estimated enrichment for 24 functional categories (obtained from Finucane *et al.* (2015)(22)), and further partitioned h_{SNP}^2 to autosomal chromosomes and minor allele frequency bins.

As it has been shown that SumHer and LD Score regression (63) has similar accuracy in estimation of genetic correlations (23), we used LD Hub to estimate the genetic correlations between medication-use (excluding the HLA region) and 257 quantitative and disease traits (24). To account for multiple testing, we adjusted all P -values using Bonferroni ($P_{adj} < 0.01$). Using LD Score regression (63) we also computed the genetic correlation between our definition of medication-use and the previously published genetic analysis of categories of medication traits (16). The univariate LD scores were computed using the 1000 Genomes European data.

Constructing genetic scores for medication-use

To construct within-cohort genetic scores we reran the GWAS in PLINK2 on five subdivision of the White-British cohort (i.e., a five-fold cross validation scheme). For each set of summary statistics we performed LD clumping (for

239 $r^2 < \{0.1, 0.3, 0.5, 0.7, 0.9\}$) implemented in the R package qgg (41) using a range of P value thresholds (0.001, 0.01,
240 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.7, 0.9, 0.999), and computed the genetic scores as $GS = \sum_{i=1}^m \mathbf{w}_i \hat{\mathbf{s}}_i$, where \mathbf{w}_i is the i -th
241 genotype (allelic counts), $\hat{\mathbf{s}}_i$ is the estimated SNP effect, and m is the number of SNPs left after LD pruning and P -
242 value thresholding.

243 The accuracy of the genetic scores were obtained as the correlation between the number of medications taken
244 of the individuals in the training set and the predicted genetic score. The genetic scores were divided into
245 percentiles, and regression coefficients (β) were estimated by linear regression of the number of medications taken
246 on the genetic score percentile relative to the 50th genetic score percentile, adjusted for sex, age, UKB assessment
247 centre, and the first ten genetic principal components.

248

249 **Conflict of interest statement**

250 The authors declare no competing interests.

251

252 **Funding**

253 This work was supported by a Lundbeck Foundation grant to PDR (R287-2018-735).

254

255 **Abbreviations**

256 GWAS, genome-wide association study; ICD10, International Classification of Diseases, version10; SNP, single
257 nucleotide polymorphisms; UKBB, United Kingdom Biobank

258

259 **Acknowledgements**

260 All of the computing for this project was performed on the GenomeDK cluster. We would like to thank GenomeDK
261 and Aarhus University for providing computational resources and support that contributed to these research
262 results. The genetic and phenotypic data were obtained from the UK Biobank Resource (ID 31269).

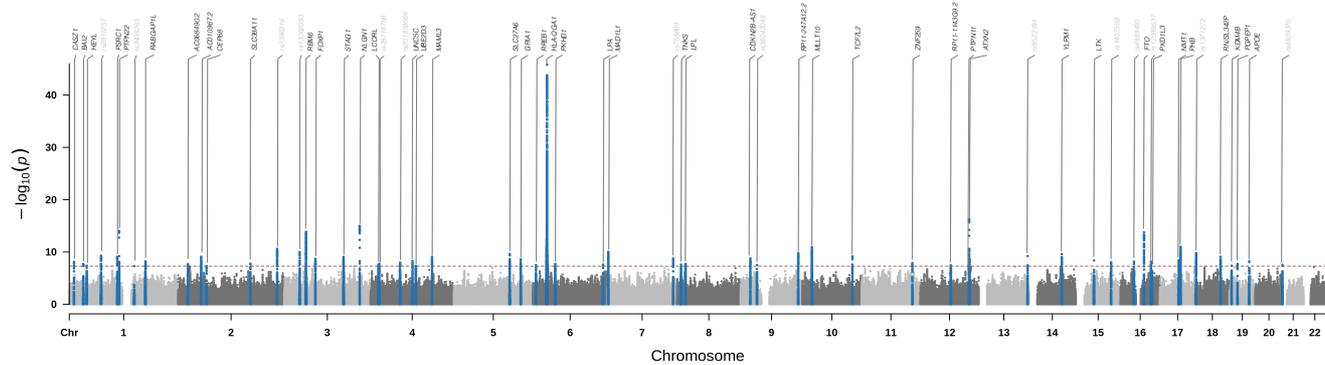
263 References

- 264 1. Lander, E. S., Linton, L. M., Birren, B., et al. (2001) Initial sequencing and analysis of the human genome. *Nature*,
265 **409**, 860–921.
- 266 2. Buniello, A., MacArthur, J. A. L., Cerezo, M., et al. (2019) The NHGRI-EBI GWAS Catalog of published genome-wide
267 association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res.*, **47**, D1005–D1012.
- 268 3. Claussnitzer, M., Cho, J. H., Collins, R., et al. (2020) A brief history of human disease genetics. *Nature*, **577**, 179–
269 189.
- 270 4. Visscher, P. M., Wray, N. R., Zhang, Q., et al. (2017) 10 Years of GWAS Discovery: Biology, Function, and
271 Translation. *Am. J. Hum. Genet.*, **101**, 5–22.
- 272 5. Ciardiello, F., Arnold, D., Casali, P. G., et al. (2014) Delivering precision medicine in oncology today and in future–
273 the promise and challenges of personalised cancer medicine: A position paper by the European Society for
274 Medical Oncology (ESMO). *Ann. Oncol.*, **25**, 1673–1678.
- 275 6. Ginsburg, G. S. and McCarthy, J. J. (2001) Personalized medicine: Revolutionizing drug discovery and patient
276 care. *Trends Biotechnol.*, **19**, 491–496.
- 277 7. Ashley, E. A. (2016) Towards precision medicine. *Nat. Rev. Genet.*, **17**, 507–522.
- 278 8. Schork, N. J. (2015) Personalized medicine: Time for one-person trials. *Nature*, **520**, 609–611.
- 279 9. Schärfe, C. P. I., Tremmel, R., Schwab, M., et al. (2017) Genetic variation in human drug-related genes. *Genome*
280 *Med.*, **9**, 1–15.
- 281 10. Madian, A. G., Wheeler, H. E., Jones, R. B., et al. (2012) Relating human genetic variation to variation in drug
282 responses. *Trends Genet.*, **28**, 487–495.
- 283 11. Pirmohamed, M. and Park, B. K. (2001) Genetic susceptibility to adverse drug reactions. *Trends Pharmacol. Sci.*,
284 **22**, 298–305.
- 285 12. Roden, D. M., Wilke, R. A., Kroemer, H. K., et al. (2011) Pharmacogenomics: The genetics of variable drug
286 responses. *Circulation*, **123**, 1661–1670.
- 287 13. Bycroft, C., Elliott, L. T., Young, A., et al. (2018) The UK Biobank resource with deep phenotyping and genomic
288 data. *Nature*, **562**, 203–209.
- 289 14. Nagai, A., Hirata, M., Kamatani, Y., et al. (2017) Overview of the BioBank Japan Project: Study design and profile.
290 *J. Epidemiol.*, **27**, S2–S8.
- 291 15. Leitsalu, L., Haller, T., Esko, T., et al. (2015) Cohort profile: Estonian biobank of the Estonian genome center,
292 university of Tartu. *Int. J. Epidemiol.*, **44**, 1137–1147.
- 293 16. Wu, Y., Byrne, E. M., Zheng, Z., et al. (2019) Genome-wide association study of medication-use and associated
294 disease in the UK Biobank. *Nat. Commun.*, **10**, 1–10.

- 295 17. Patel, R. B. (2003) Polypharmacy and the Elderly. *J. Infus. Nurs.*, **26**, 166–169.
- 296 18. Ramage-Morin, P. L. (2009) Medication use among senior Canadians. *Heal. Reports, Stat. Canada*, **20**, 37–44.
- 297 19. Hohl, C. M., Karpov, A., Reddekopp, L., et al. (2014) ICD-10 codes used to identify adverse drug events in
298 administrative data: A systematic review. *J. Am. Med. Informatics Assoc.*, **21**, 547–557.
- 299 20. Kennedy, A. E., Ozbek, U. and Dorak, M. T. (2017) What has GWAS done for HLA and disease associations? *Int. J.*
300 *Immunogenet.*, **44**, 195–211.
- 301 21. Altshuler, D. L., Durbin, R. M., Abecasis, G. R., et al. (2010) A map of human genome variation from population-
302 scale sequencing. *Nature*, **467**, 1061–1073.
- 303 22. Finucane, H. K., Bulik-Sullivan, B., Gusev, A., et al. (2015) Partitioning heritability by functional annotation using
304 genome-wide association summary statistics. *Nat. Genet.*, **47**, 1228–1235.
- 305 23. Speed, D. and Balding, D. J. (2019) SumHer better estimates the SNP heritability of complex traits from summary
306 statistics. *Nat. Genet.*, **51**, 277–284.
- 307 24. Zheng, J., Erzurumluoglu, A. M., Elsworth, B. L., et al. (2017) LD Hub: A centralized database and web interface to
308 perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability and
309 genetic correlation analysis. *Bioinformatics*, **33**, 272–279.
- 310 25. Dendrou, C. A., Petersen, J., Rossjohn, J., et al. (2018) HLA variation and disease. *Nat. Rev. Immunol.*, **18**, 325–339.
- 311 26. Trowsdale, J. and Knight, J. C. (2013) Major histocompatibility complex genomics and human disease. *Annu. Rev.*
312 *Genomics Hum. Genet.*, **14**, 301–323.
- 313 27. Falconer, D. S. and Mackay, T. F. C. (1996) Introduction to Quantitative Genetics. *Introduction to Quantitative*
314 *Genetics*; Fourth edi.; Pearson, (1996) .
- 315 28. Todd, J. A., Walker, N. M., Cooper, J. D., et al. (2007) Robust associations of four new chromosome regions from
316 genome-wide analyses of type 1 diabetes. *Nat. Genet.*, **39**, 857–864.
- 317 29. Mahajan, A., Go, M. J., Zhang, W., et al. (2014) Genome-wide trans-ancestry meta-analysis provides insight into
318 the genetic architecture of type 2 diabetes susceptibility. *Nat. Genet.*, **46**, 234–244.
- 319 30. Mahajan, A., Taliun, D., Thurner, M., et al. (2018) Fine-mapping type 2 diabetes loci to single-variant resolution
320 using high-density imputation and islet-specific epigenome maps. *Nat. Genet.*, **50**, 1505–1513.
- 321 31. Van Der Harst, P. and Verweij, N. (2018) Identification of 64 novel genetic loci provides an expanded view on
322 the genetic architecture of coronary artery disease. *Circ. Res.*, **122**, 433–443.
- 323 32. Howard, D. M., Adams, M. J., Clarke, T. K., et al. (2019) Genome-wide meta-analysis of depression identifies 102
324 independent variants and highlights the importance of the prefrontal brain regions. *Nat. Neurosci.*, **22**, 343–352.
- 325 33. Nagel, M., Jansen, P. R., Stringer, S., et al. (2018) Meta-analysis of genome-wide association studies for
326 neuroticism in 449,484 individuals identifies novel genetic loci and pathways. *Nat. Genet.*, **50**, 920–927.

- 327 34. Jansen, P. R., Watanabe, K., Stringer, S., et al. (2019) Genome-wide analysis of insomnia in 1,331,010 individuals
328 identifies new risk loci and functional pathways. *Nat. Genet.*, **51**, 394–403.
- 329 35. Kichaev, G., Bhatia, G., Loh, P. R., et al. (2019) Leveraging Polygenic Functional Enrichment to Improve GWAS
330 Power. *Am. J. Hum. Genet.*, **104**, 65–75.
- 331 36. Zhu, Z., Guo, Y., Shi, H., et al. (2020) Shared genetic and experimental links between obesity-related traits and
332 asthma subtypes in UK Biobank. *J. Allergy Clin. Immunol.*, **145**, 537–549.
- 333 37. Liu, M., Jiang, Y., Wedow, R., et al. (2019) Association studies of up to 1.2 million individuals yield new insights
334 into the genetic etiology of tobacco and alcohol use. *Nat. Genet.*, **51**, 237–244.
- 335 38. Perry, J. R. B., Day, F., Elks, C. E., et al. (2014) Parent-of-origin-specific allelic associations among 106 genomic
336 loci for age at menarche. *Nature*, **514**, 92–97.
- 337 39. Lee, S. H., Decandia, T. R., Ripke, S., et al. (2012) Estimating the proportion of variation in susceptibility to
338 schizophrenia captured by common SNPs. *Nat. Genet.*, **44**, 247–250.
- 339 40. Davis, L. K., Yu, D., Keenan, C. L., et al. (2013) Partitioning the Heritability of Tourette Syndrome and Obsessive
340 Compulsive Disorder Reveals Differences in Genetic Architecture. *PLoS Genet.*, **9**.
- 341 41. Rohde, P. D., Fourie Sørensen, I. and Sørensen, P. (2020) qgg: an R package for large-scale quantitative genetic
342 analyses. *Bioinformatics*, **36**, 2614–2615.
- 343 42. Nikpay, M., Stewart, A. F. R., McPherson, R., et al. (2017) Partitioning the heritability of coronary artery disease
344 highlights the importance of immune-mediated processes and epigenetic sites associated with transcriptional
345 activity. *Cardiovasc. Res.*, **113**, 973–983.
- 346 43. Yang, J., Manolio, T. A., Pasquale, L. R., et al. (2011) Genome partitioning of genetic variation for complex traits
347 using common SNPs. *Nat. Genet.*, **43**, 519–525.
- 348 44. Xue, A., Wu, Y., Zhu, Z., et al. (2018) Genome-wide association analyses identify 143 risk variants and putative
349 regulatory mechanisms for type 2 diabetes. *Nat. Commun.*, **9**.
- 350 45. Nikpay, M., Turner, A. W. and McPherson, R. (2018) Partitioning the Pleiotropy Between Coronary Artery
351 Disease and Body Mass Index Reveals the Importance of Low Frequency Variants and Central Nervous System-
352 Specific Functional Elements. *Circ. Genomic Precis. Med.*, **11**, e002050.
- 353 46. Lee, S. H., Harold, D., Nyholt, D. R., et al. (2013) Estimation and partitioning of polygenic variation captured by
354 common snps for alzheimer's disease, multiple sclerosis and endometriosis. *Hum. Mol. Genet.*, **22**, 832–841.
- 355 47. Locke, A. E., Kahali, B., Berndt, S. I., et al. (2015) Genetic studies of body mass index yield new insights for
356 obesity biology. *Nature*, **518**, 197–206.
- 357 48. Al-Goblan, A. S., Al-Alfi, M. A. and Khan, M. Z. (2014) Mechanism linking diabetes mellitus and obesity. *Diabetes,*
358 *Metab. Syndr. Obes. Targets Ther.*, **7**, 587–591.

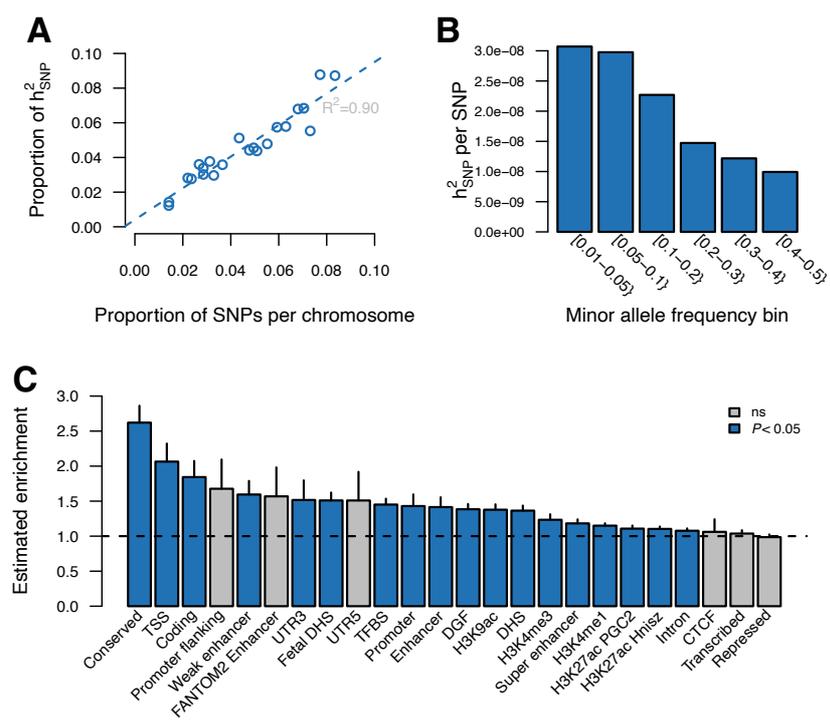
- 359 49. Hruby, A., Manson, J. A. E., Qi, L., et al. (2016) Determinants and consequences of obesity. *Am. J. Public Health*,
360 **106**, 1656–1662.
- 361 50. U.S. Department of Health and Human Services (2013) Managing Overweight and Obesity in Adults: Systematic
362 Evidence Review from the Obesity Expert Panel. *Managing Overweight and Obesity in Adults: Systematic
363 Evidence Review from the Obesity Expert Panel*; (2013) .
- 364 51. Paruthi, S., Brooks, L. J., D'Ambrosio, C., et al. (2016) Consensus Statement of the American Academy of Sleep
365 Medicine on the Recommended Amount of Sleep for Healthy Children: Methodology and Discussion. *J. Clin. Sleep
366 Med.*, **12**, 1549–1561.
- 367 52. (2004) The 2004 United States Surgeon General's Report: The Health Consequences of Smoking. *N. S. W. Public
368 Health Bull.*, **15**, 107.
- 369 53. DeBoever, C., Tanigawa, Y., Aguirre, M., et al. (2020) Assessing Digital Phenotyping to Enhance Genetic Studies
370 of Human Diseases. *Am. J. Hum. Genet.*
- 371 54. Hafferty, J. D., Campbell, A. I., Navrady, L. B., et al. (2018) Self-reported medication use validated through record
372 linkage to national prescribing data. *J. Clin. Epidemiol.*, **94**, 132–142.
- 373 55. Fry, A., Littlejohns, T. J., Sudlow, C., et al. (2017) Comparison of Sociodemographic and Health-Related
374 Characteristics of UK Biobank Participants with Those of the General Population. *Am. J. Epidemiol.*, **186**, 1026–
375 1034.
- 376 56. Keyes, K. M. and Westreich, D. (2019) UK Biobank, big data, and the consequences of non-representativeness.
377 *Lancet*, **393**, 1297.
- 378 57. Plenge, R. M., Scolnick, E. M. and Altshuler, D. (2013) Validating therapeutic targets through human genetics.
379 *Nat. Rev. Drug Discov.*, **12**, 581–594.
- 380 58. Nelson, M. R., Tipney, H., Painter, J. L., et al. (2015) The support of human genetic evidence for approved drug
381 indications. *Nat. Genet.*, **47**, 856–860.
- 382 59. Krapohl, E., Patel, H., Newhouse, S., et al. (2018) Multi-polygenic score approach to trait prediction. *Mol.
383 Psychiatry*, **23**, 1368–1374.
- 384 60. Sudlow, C., Gallacher, J., Allen, N., et al. (2015) UK Biobank: An Open Access Resource for Identifying the Causes
385 of a Wide Range of Complex Diseases of Middle and Old Age. *PLoS Med.*, **12**, 1–10.
- 386 61. Chang, C. C., Chow, C. C., Tellier, L. C. A. M., et al. (2015) Second-generation PLINK: Rising to the challenge of
387 larger and richer datasets. *Gigascience*, **4**, 1–16.
- 388 62. McLaren, W., Gil, L., Hunt, S. E., et al. (2016) The Ensembl Variant Effect Predictor. *Genome Biol.*, **17**, 1–14.
- 389 63. Bulik-Sullivan, B., Finucane, H. K., Anttila, V., et al. (2015) An atlas of genetic correlations across human diseases
390 and traits. *Nat. Genet.*, **47**, 1236–1241.



391

392 **Figure 1 | Manhattan plot for medication-use in the UKBB (n=335,744).** The x-axis is chromosomal position, and the y-axis
393 shows the negative logarithm base-10 to the P values from regression of current medication-use on 9,804,629 SNPs. The
394 horizontal red line shows the genome-wide significance level (5×10^{-8}). Independent genome-wide significant loci (within 1000
395 kb and $r^2 < 0.01$) are depicted in blue. For each significant locus the gene within 2000kb is shown (for intergenic loci the lead
396 SNP is shown).

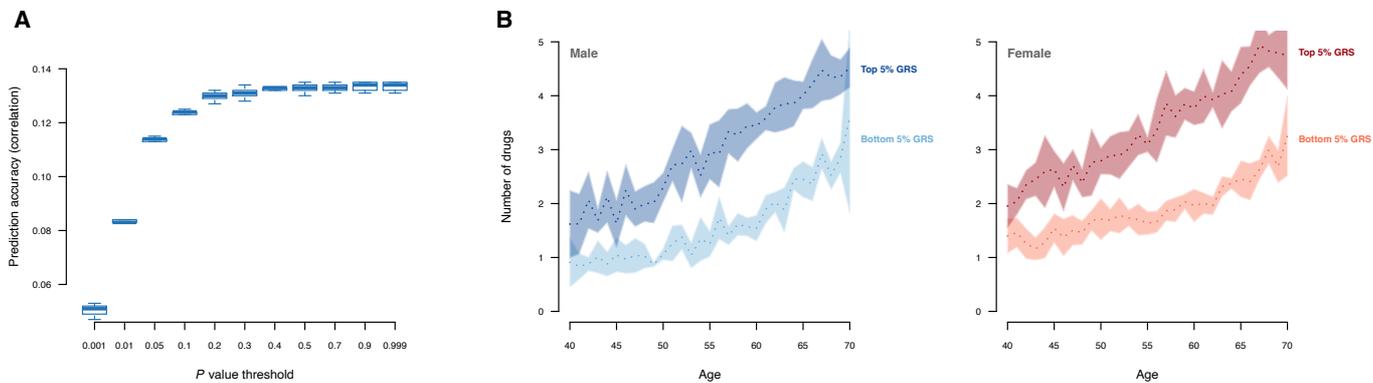
397



398

399 **Figure 2 | Partitioning of genomic variance for medication-use.** **A:** Proportion of genomic variance captured per autosomal
400 chromosome as function of the proportion of SNPs per chromosome. **B:** Proportion genomic variance, scaled by the number of
401 SNPs, captured by minor allele frequency. **C:** Estimated enrichment score for functional categories. Vertical line segments mark
402 the standard deviation on enrichment score. Horizontal dashed line marks an enrichment score of no enrichment. CTCF: a highly
403 conserved multifunctional DNA-binding protein, DGF: digital genomic footprint, DHS: DNase I hypersensitivity sites, TFBS:
404 transcription factor binding site, TSS: transcription start site.

405



406

407

408

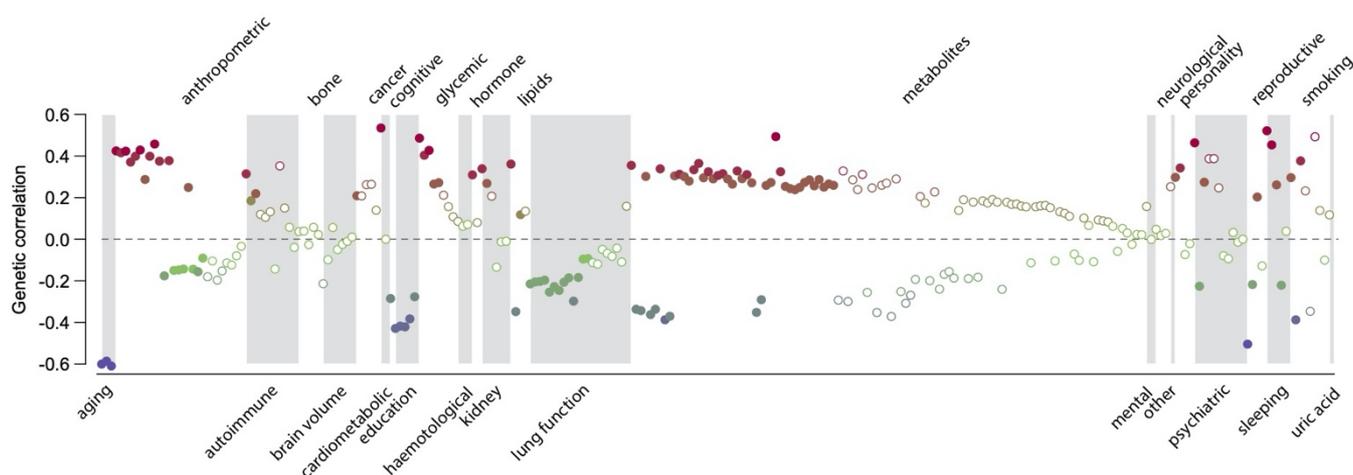
409

410

411

412

Figure 3 | Prediction of number of medications used. **A:** Prediction accuracy (measured by the correlation between observed and predicted values) for medication-use across the range of P values. Results shown is for $r^2 < 0.5$ as it gave the highest accuracy across different r^2 -values (see [Supplementary Figure S2](#) for all r^2 values). **B:** Averaged number of drugs (over the five training sets) used by males and females stratified by top 5% and bottom 5% of genetic scores. Shadings corresponds to the standard error over the five training sets.



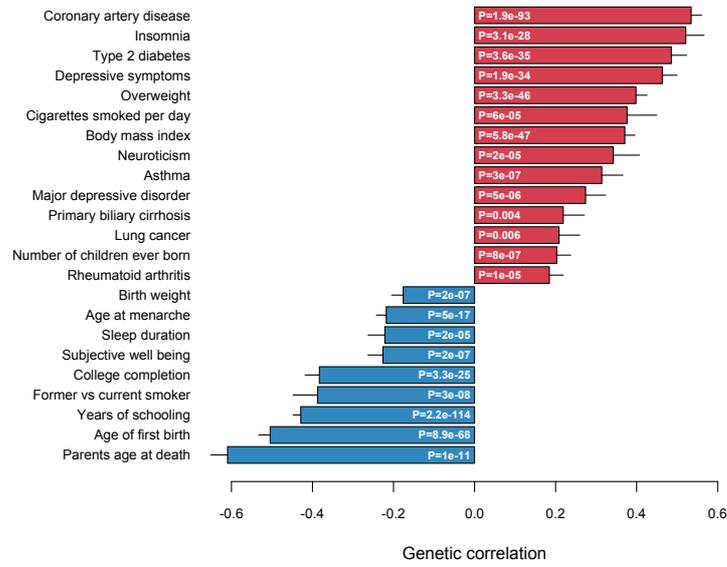
413

414

415

416

Figure 4 | Estimated genetic correlations between medication-use and 257 traits and diseases. Traits displaying significant genetic correlations are displayed as filled symbols. Details can be found in [Supplementary Table S3](#).



417

418 **Figure 5 | Genetic correlations between medication-use and selected top categories.**