



# Genetic predictors of lifelong medication-use patterns in cardiometabolic diseases

Received: 19 October 2021

Accepted: 8 November 2022

Published online: 18 January 2023

Check for updates

**Tuomo Kiiskinen** <sup>1,2,3,17</sup>, **Pyry Helkkula** <sup>1,17</sup>, **Kristi Krebs** <sup>4,17</sup>, **Juha Karjalainen** <sup>1,3,5</sup>, **Elmo Saarentaus** <sup>1,6</sup>, **Nina Mars** <sup>1</sup>, **Arto Lehisto** <sup>1</sup>, **Wei Zhou** <sup>3</sup>, **Mattia Cordioli** <sup>1</sup>, **Sakari Jukarainen** <sup>1</sup>, **Joel T. Rämö** <sup>1</sup>, **Juha Mehtonen** <sup>1</sup>, **Kumar Veerapen** <sup>3</sup>, **Markus Räsänen** <sup>7</sup>, **Sanni Ruotsalainen** <sup>1</sup>, **Mutaamba Maasha** <sup>3</sup>, **FinnGen\***, **Teemu Niiranen** <sup>2,8,9</sup>, **Tiinamaija Tuomi** <sup>1,10,11,12,13</sup>, **Veikko Salomaa** <sup>2</sup>, **Mitja Kurki** <sup>1,3,5</sup>, **Matti Pirinen** <sup>1,14,15</sup>, **Aarno Palotie** <sup>1,5,16</sup>, **Mark Daly** <sup>1,5,16</sup>, **Andrea Ganna** <sup>1,3</sup>, **Aki S. Havulinna** <sup>1,2</sup>, **Lili Milani** <sup>4</sup> & **Samuli Ripatti** <sup>1,3,11</sup>

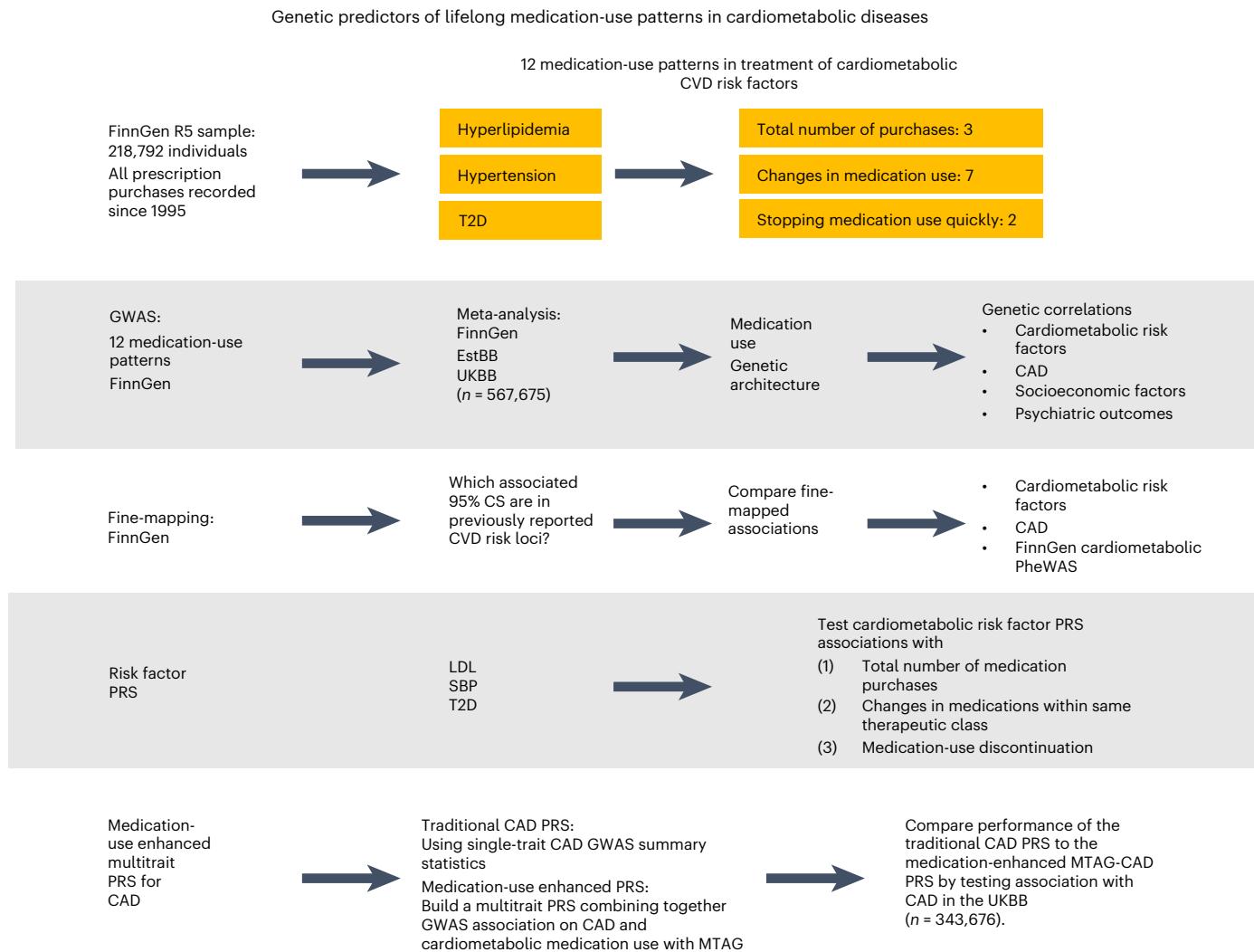
Little is known about the genetic determinants of medication use in preventing cardiometabolic diseases. Using the Finnish nationwide drug purchase registry with follow-up since 1995, we performed genome-wide association analyses of longitudinal patterns of medication use in hyperlipidemia, hypertension and type 2 diabetes in up to 193,933 individuals (55% women) in the FinnGen study. In meta-analyses of up to 567,671 individuals combining FinnGen with the Estonian Biobank and the UK Biobank, we discovered 333 independent loci ( $P < 5 \times 10^{-9}$ ) associated with medication use. Fine-mapping revealed 494 95% credible sets associated with the total number of medication purchases, changes in medication combinations or treatment discontinuation, including 46 credible sets in 40 loci not associated with the underlying treatment targets. The polygenic risk scores (PRS) for cardiometabolic risk factors were strongly associated with the medication-use behavior. A medication-use enhanced multitrait PRS for coronary artery disease matched the performance of a risk factor-based multitrait coronary artery disease PRS in an independent sample (UK Biobank,  $n = 343,676$ ). In summary, we demonstrate medication-based strategies for identifying cardiometabolic risk loci and provide genome-wide tools for preventing cardiovascular diseases.

Cardiovascular disease (CVD) is the leading cause of excess mortality in the developed countries<sup>1</sup>, and although approximately half of the variability in cardiometabolic diseases is heritable<sup>2</sup>, most related harm is preventable<sup>3,4</sup>. Pharmacotherapies targeting cardiometabolic risk factors—type 2 diabetes (T2D), hyperlipidemia and hypertension—remain at the core of CVD prevention<sup>5,6</sup>.

Challenges in pharmacological prevention of CVD involve identifying patients in need of therapy, setting the targets of the treatment and selecting therapies of adequate efficacy and acceptable risk profiles. In addition to socioeconomic factors, both the set and dose

of medicines that patients start their treatment with and continue to use depends on factors such as cardiovascular risk profiles, disease etiology, drug responsiveness and adverse effects<sup>5,6</sup>. Abandoning or inadequately adhering to therapies worsens outcomes<sup>7–10</sup>. With limited tools to predict treatment suboptimality, pharmacotherapy is traditionally optimized in a reactive trial-and-error manner when patients experience side effects, miss their treatment targets or experience events such as myocardial infarction or stroke<sup>5,6</sup>. Real-world data from electronic health records and registries provide massive datasets with sufficient statistical power to explore long-term medication use.

A full list of affiliations appears at the end of the paper. e-mail: [samuli.ripatti@helsinki.fi](mailto:samuli.ripatti@helsinki.fi)



**Fig. 1 | Workflow of the study.** GWAS were performed for 12 phenotypes of medication-use patterns in treating hyperlipidemia, hypertension and T2D (three continuous analyses of the total number of medication purchases and nine binary analyses of medication changing and fast discontinuation) in FinnGen, with data capturing all medication purchases since 1 January 1995. Meta-analyses of up to 567,671 participants combined data in FinnGen ( $n = 29,990$ –193,933), EstBB ( $n = 5,110$ –184,892) and UKBB ( $n = 188,846$ ). Fine-mapping was performed in FinnGen for all associated ( $P < 5 \times 10^{-8}$ ) regions. Genetic architectures between

medication-use traits and the underlying cardiometabolic risk were juxtaposed by comparing genome-wide significant associations, calculating LDSC regression genetic correlations and testing associations between PRS for LDL, SBP and T2D and the medication-use phenotypes. A medication-use enhanced multitrait PRS for CAD was built using MTAG method and its performance was compared to a traditional CAD PRS by testing associations with CAD in an independent sample (UKBB,  $n = 343,676$ ).

Behavioral patterns derived from prescription data can be used as a proxy to detect suboptimal or harmful prescriptions<sup>11,12</sup>. Such studies are essential to identify factors that influence interindividual variability in treatment response and can inform clinicians on initial treatments that are different from first-line therapies.

Genetic information has been proposed as a tool to optimize pharmacotherapy<sup>13,14</sup> and advance drug development<sup>15</sup>. Despite the progress in disease genetics, we have relatively limited knowledge on the role of genetic factors driving the variation in lifelong patterns of medication use in cardiometabolic diseases. Meanwhile, large-scale genome-wide association studies (GWAS) have identified a complex polygenic architecture comprising hundreds of associated loci for lipid levels<sup>16–18</sup>, blood pressure<sup>19,20</sup> and T2D<sup>21,22</sup>, and using genetic information in clinical risk prediction shows promise of clinical relevance<sup>23,24</sup>. A study of self-reported medication use in the UK Biobank (UKBB) revealed variants associated with medication use, but the self-reported nature and ambiguous names of medications may have limitations in

the accuracy of the investigated phenotypes<sup>25</sup>. In contrast, pharmacogenetic associations, such as the *SLCO1B1* polymorphism (rs4149056) increasing risk for simvastatin-induced myopathy<sup>26</sup>, have been identified in smaller pharmacogenetic studies<sup>27–29</sup>, mostly having a relatively narrow focus on drug response as the primary outcome.

Our work is a systematic study of genetic effects on medication-use patterns using cardiometabolic medications as a model. We test three main hypotheses: (1) medication data can be used to identify genetic factors underlying cardiometabolic diseases, (2) genome-wide data inform us about the likelihood of medication switching and stopping, and (3) medication-use-associated genetic variation allows for enhancing polygenic prediction. We study three fundamental medication-use patterns: the cumulative medication use, medication switching and treatment discontinuation. We develop polygenic risk scores (PRS) to predict medication patterns and use medication data to identify risk factors for cardiometabolic diseases. We conduct a genetic population-based biobank study comprising up to 193,933

**Table 1 | Medication patterns in hyperlipidemia, hypertension and T2D derived from the nationwide drug purchase registry in FinnGen**

	ATC codes	n	Patients	Controls
Treatment of hyperlipidemia				
(1) Drugs targeting hyperlipidemia, number of purchases	C10	193,933		
Participants with at least one purchase		77,439		
(2) Changing simvastatin to another statin	C10AA01	45,134	19,228	25,906
(3) Fast discontinuation of statin use	C10AA	76,499	7,459	68,990
Treatment of hypertension				
(4) Drugs targeting hypertension, number of purchases	C02, C03, C07, C08, C09	193,933		
Participants with at least one purchase		121,491		
(5) More than one antihypertensive group		125,586	82,551	43,035
(6) More than two antihypertensive groups		125,586	50,482	75,104
(7) More than three antihypertensive groups		125,586	22,476	103,110
(8) All five antihypertensive groups		125,586	2,900	122,686
(9) Fast discontinuation of hypertension medication use		125,212	14,661	110,551
Treatment of T2D				
(10) Drugs targeting T2D, number of purchases	A10B	193,933		
Participants with at least one purchase		31,223		
(11) Use of second-line agents in T2D	A10BD, A10BH, A10BJ, A10BK	31,665	14,292	17,373
(12) Use of insulin	A10A	29,990	6,861	23,129

The total number of participants included in the analyses, and the number of cases and controls for the binary traits, are presented.

Finnish participants of the FinnGen study<sup>30</sup> linked to the nationwide drug purchase registry covering every prescription drug purchase in Finland since 1995. We perform GWAS for patterns of medication use on (1) the total number of drug purchases during the follow-up, (2) switching medications within the same therapeutic class and (3) discontinuation of the use, and fine-map our findings to a single-variant resolution. We meta-analyze our GWAS results using 184,892 participants in the Estonian Biobank (EstBB)<sup>31</sup> and 188,846 participants in the UKBB<sup>32</sup>. We evaluate the effect of genome-wide PRS for low-density lipoprotein (LDL), systolic blood pressure (SBP) and T2D on medication use. Finally, we build a medication-use-enhanced multitrait PRS for coronary artery disease (CAD) (Fig. 1).

## Results

### Study population and medication-use patterns

FinnGen release 5 (ref. <sup>30</sup>) ([www.finngen.fi/en](http://www.finngen.fi/en)) consisted of 218,792 genotyped individuals of Finnish ancestry (Supplementary Figs. 1–4) with 5,118,565 years of drug-registry-based follow-up. In total, 56.5% were women and the mean age at the end of the follow-up was 59.8 years. Overall, 44,343,661 drug purchases were recorded and 3,650,495 (8.2%) of these were drugs used in treating hyperlipidemia, hypertension or T2D.

We identified all participants' complete purchase histories of drugs targeting cardiometabolic risk factors: hyperlipidemia, hypertension and T2D. For each disorder, we used the sum of the total number of purchases of the drugs to quantify the magnitude of the pharmacological intervention (Extended Data Fig. 1) and every purchase of any package of any drug coded with an appropriate Anatomical Therapeutic Chemical Classification (ATC) code during the study period (from the start of the drug purchase registry, 1 January 1995, until death or the end of follow-up, 31 December 2018) increased the sum by one. Multiple factors, including the age of disease onset, cardiometabolic disease severity, drug resistance, perceived cardiometabolic risk and adherence, all affect this proxy of overall cumulative medication use. We included 193,933 participants (55.3% women), being at least 10 years

old and alive on 1 January 1995 (Supplementary Table 1a). In total, 77,439, 121,491 and 31,223 participants recorded at least one purchase of drugs used in treating hyperlipidemia, hypertension and T2D, respectively. The longest treatment durations (mean = 11.7 years, standard deviation, s.d. = 7.97) and most purchases were seen in hypertension (mean = 69.1, excluding participants without purchases, s.d. = 67.6, median = 50). The shortest treatments were seen in T2D (mean = 8.05 years, s.d. = 6.35) and the fewest purchases in hyperlipidemia (mean = 30.7, s.d. = 27.1, median = 25) (Supplementary Table 1b).

We identified common binary patterns in medication use (Extended Data Fig. 2). Although simvastatin was the most common first-choice statin, with 45,134 participants (59.0%) starting their treatment with it, 19,228 (42.6%) switched to another statin. Early discontinuation of statin use was frequent: 7,459 participants (9.86%) stopped after one or two purchases, comparable with a previous estimate<sup>33</sup>. In hypertension, 82,551 (65.7%) had purchases in more than one and 2,900 (2.3%) in all five ATC-based medication subgroups. In T2D, 14,292 (45.1%) recorded purchases of second-line treatments and 6,861 (22.9%) progressed to use insulin (Table 1; for EstBB and UKBB, see Supplementary Table 2). Total purchases and medication changes were positively correlated ( $r = 0.01$ – $0.60$ ) with one another, and each of these was negatively correlated with medication discontinuation ( $r = -0.46$  to  $-0.03$ ) (Extended Data Fig. 3).

### Genome-wide association analyses in FinnGen

We conducted GWAS for the 12 medication-use phenotypes (Table 1) of total purchases, medication switching and discontinuation of up to 24 years of data (mean = 23.4, s.d. = 2.68) on hyperlipidemia, hypertension and T2D medications in FinnGen. Of  $16.4 \times 10^6$  variants ( $10.3 \times 10^3$  with minor allele frequency (MAF)  $> 0.005$  included in quantitative analyses), 23,577 were genome-wide significant (GWS,  $P < 5 \times 10^{-8}$ ) in at least one analysis, with 6211 Mb windows of genome-wide significant associations to medication-use phenotypes (Manhattan plots for FinnGen GWAS: [https://meds.finngen.fi](http://meds.finngen.fi)). Combining all GWS variants from all analyses within 1.5 Mb windows resulted in 303 independent loci: 91 with leading

associations (smallest  $P$  values) for hyperlipidemia, 145 for hypertension and 67 for T2D medication patterns. In total, 298 loci had leading associations for the quantitative and five for the binary medication phenotypes (Extended Data Figs. 4–6 and Supplementary Table 3).

Secondary sensitivity GWAS displayed overlapping GWS loci (case-control analyses of ever having purchased the studied drugs: 132 overlapping/2 additional GWS loci, annual purchases: 225/7, total purchases, excluding participants without any purchases: 73/5; Supplementary Tables 4–7). All linkage disequilibrium score regression (LDSC)<sup>34,35</sup> genetic correlations (RGs) between the initial and the sensitivity analyses were close to 1 (Supplementary Table 8). Sex-stratified GWAS of total purchases ( $n_{\text{women}} = 107,231$ ,  $n_{\text{men}} = 86,702$ ) supported shared genetic architectures between the sexes as RGs between sexes were approximately 1 for all medications (hyperlipidemia, hypertension, T2D) and effect sizes for the 298 GWS lead variants were highly correlated (correlation coefficient for lead variant effect sizes ( $r_{\beta}$ ) = 0.95) between sexes (Supplementary Table 9 and Extended Data Fig. 7). Finally, we performed interaction analyses for the 303 lead variants, introducing lead variant allele (G)  $\times$  age or G  $\times$  follow-up time interaction terms in the models. Interactions with age were significant ( $P < 0.05/303$  GWS lead variants) for 37 variants (Supplementary Table 10), whereas there were no significant interactions with follow-up time.

### Meta-analyses in FinnGen, the EstBB and the UKBB

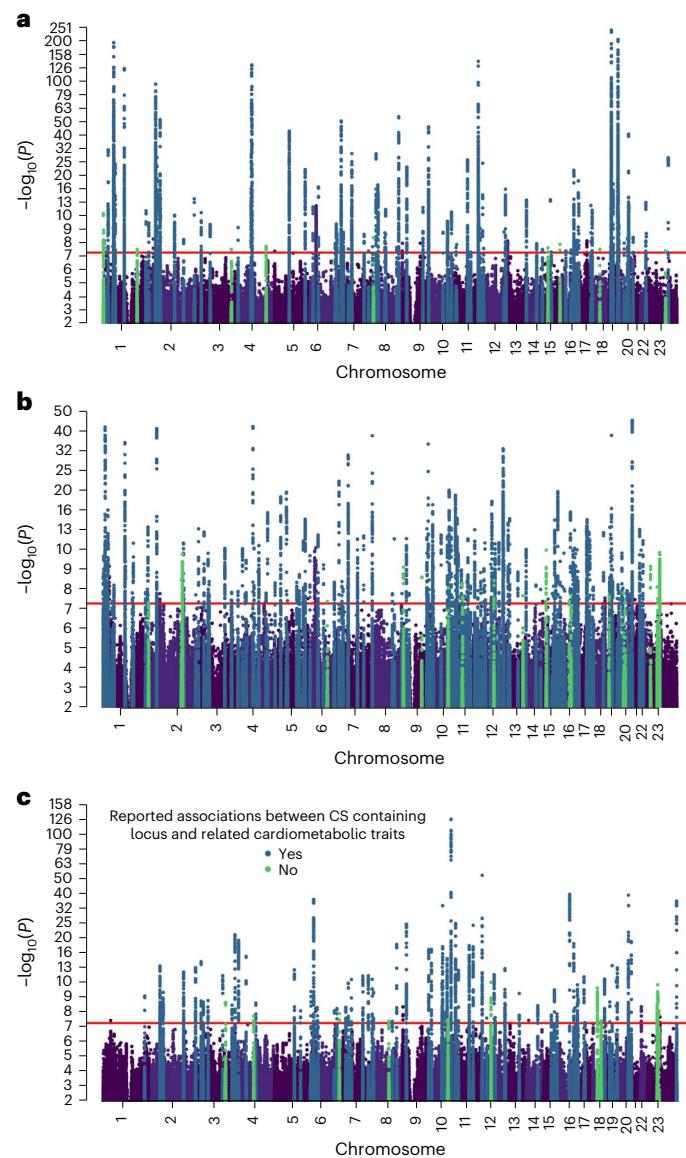
In GWAS for all 12 medication-use patterns in EstBB with medication purchase data ( $n_{\text{EstBB}} = 184,892$ , binary phenotypes  $n = 5,067$ –74,699) and three medication quantity analyses using prescription data in UKBB ( $n_{\text{UKBB}} = 188,846$ ), we saw highly concordant effect directions (251 out of 288 lead variants in FinnGen were concordant in all samples,  $P$  value from binomial test ( $P_{\text{binom}}$ ) =  $6.0 \times 10^{-111}$ ; Supplementary Table 3). We performed meta-analyses of all three samples for the number of purchases ( $n_{\text{total}} = 567,671$ ), of FinnGen and EstBB for the binary phenotypes ( $n_{\text{total}} = 42,332$ –200,285) for all loci with a suspected association in FinnGen ( $P < 5 \times 10^{-6}$ ). In total, 333 independent loci were associated with at least one medication phenotype (74 with leading association for hyperlipidemia, 181 for hypertension and 78 for T2D), with a stricter criterion for genome-wide significance ( $P < 5 \times 10^{-9}$ ), including 94 loci not significant ( $P \geq 5 \times 10^{-8}$ ) in FinnGen (Supplementary Table 11).

### Fine-mapping in FinnGen

Fine-mapping the significant ( $P < 5 \times 10^{-8}$ ) 480 locus–medication associations (a locus may be associated with multiple phenotypes) in FinnGen using the sum of single effects method<sup>36</sup> resulted in 494 95% credible sets (CS) in 347 locus–medication associations. In 73 CS the variant with the highest posterior probability for causality was at least two-fold Finnish-enriched compared with non-Finnish-Swedish-Estonian European populations. In total, 81 CS included one or more functional missense or predicted loss-of-function variants (Supplementary Tables 12–18). Overall, 448 CS overlapped with associations reported previously for the underlying trait (lipids, blood pressure and fasting glucose levels), whereas 40 distinct loci (with 46 CS) had no previous GWS associations in external summary statistics and the GWAS catalog.

### CS for total number of medication purchases

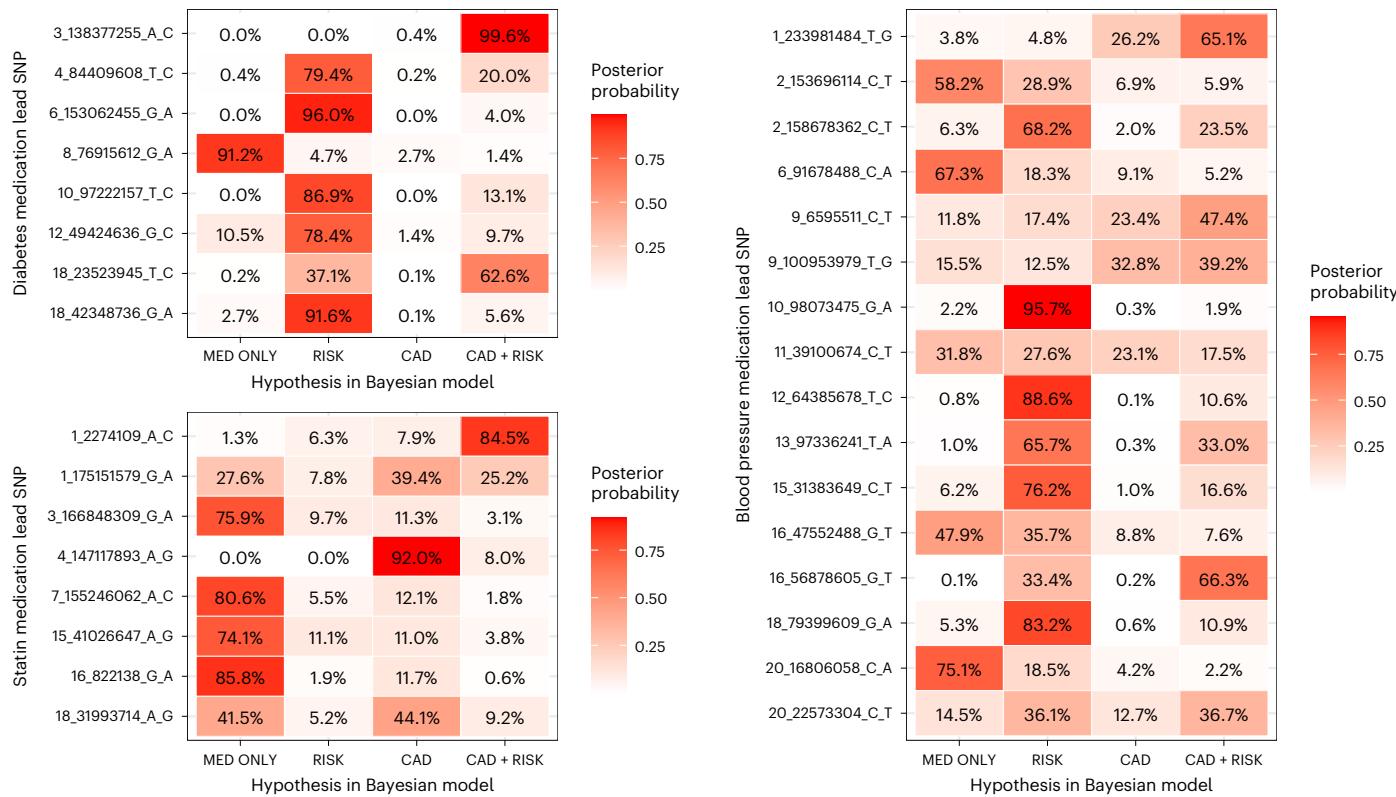
In total, 450 CS were associated with the overall number of purchases of drugs used in the treatment of hyperlipidemia (83 loci, 137 CS), hypertension (145 loci, 221 CS) and T2D (77 loci, 92 CS) (Fig. 2, Extended Data Fig. 4 and Supplementary Tables 14–16). Of loci, 255 were GWS in the meta-analysis of FinnGen, UKBB and EstBB ( $P < 5 \times 10^{-9}$ ) (Supplementary Table 12). Most CS (85.2%) overlap with previously associated loci for cardiometabolic risk factors: variants elevating risk factor levels were associated with more purchases and risk factors that lowered variants with fewer purchases. There were 38 loci containing 44 CS (including six missense variants) that had no related previous GWS associations (Supplementary Tables 14–16).



**Fig. 2 | GWAS association analysis results for the total number of medication purchases in FinnGen. a–c.** Manhattan plots with two-sided  $P$  values of quantitative SAIGE mixed-model GWAS ( $n = 193,933$ ) for the number of recorded purchases of drugs used in the treatment of hyperlipidemia (a), hypertension (b) and T2D (c). All loci containing one or more 95% CS are highlighted according to their previously reported related cardiometabolic associations; loci that have been previously associated with lipid-related traits (a), blood pressure-related traits (b) or blood glucose-related traits (c). The horizontal line signifies genome-wide significance ( $P < 5 \times 10^{-8}$ ) without additional multiple testing correction.

The strongest associations ( $P_{\text{lead variant}} < 10^{-100}$ , Bayes factor (BF)<sub>CS</sub>  $> 10^{100}$ ) for hyperlipidemia medications were in established lipid loci (*APOE*, *LDLR*, *PCSK9*, *APOC1*, *CBL* and *ANKRD17*). The 137 CS included 29 functional variants (28 missense, one predicted loss of function). Nine loci had no previous lipid-related associations (Supplementary Table 14).

In total, 221 CS associated with hypertension medication purchases contained 49 functional variants. There were 20 loci that had no associations with blood pressure (Supplementary Table 15). Overall, 92 CS associated with T2D medication purchases contained 21 functional variants. Of 11 loci that had not been previously associated with glucose-related traits, four contained a Finnish-enriched CS lead variant (more than threefold) (Supplementary Table 16).



**Fig. 3 | Shared effect between medically treated cardiometabolic risk factor, CAD and lead SNPs from GWAS of the total number of purchases of medications for hyperlipidemia, hypertension and T2D.** Bayesian posterior probabilities of hypothetical models are displayed on the x axis for the number of purchases of T2D (upper left), hyperlipidemia (lower left) and hypertension (right) medications lead SNPs (y axis). MED ONLY, SNP affects medication use only model; RISK, correlated effect model across medication use and risk factor

but not CAD; CAD, correlated effect model across medication use and CAD but not the risk factor; CAD + RISK, correlated effect model across medication use, risk factor and CAD. The variant association statistics used to compute the posterior probabilities for LDL, SBP and T2D and came from previous studies<sup>20,21,23</sup> with sample sizes of 340,951, 757,601 and 898,130, respectively. For CAD, a meta-analysis including FinnGen, UKBB and CARDIoGRAMplusC4D<sup>37</sup> ( $n = 811,555$ )<sup>40</sup> was used.

## CS for changes in medication use

CS in four established lipid loci (*PCSK9*, *CELSR2*, *APOE* and *CHD4*) were associated with switching simvastatin to another statin with shared variant effect directions for LDL. A locus ( $P$  value in meta-analysis ( $P_{\text{meta}}$ ) =  $2.6 \times 10^{-9}$ ) without previous cardiometabolic associations (nearest gene: *AP000472.2*) contained a 6.4-fold Finnish-enriched lead variant (21:22082072:G:A, MAF =  $8.0 \times 10^{-4}$ ) (Extended Data Fig. 5a and Supplementary Tables 12 and 17).

In total, 32 CS were associated with purchasing hypertension medications from different ATC-based medication groups (Table 1). One locus (*WNT2B*) was associated with purchases from all four analyzed thresholds of the number of antihypertensive drug groups (more than one, two, three or four groups). Three loci (*HOXA13*, *CASZ1*, *KCNK3*) were associated with three thresholds (more than one, two and three groups) (Extended Data Fig. 5c–f and Supplementary Tables 12 and 17).

A previously reported T2D locus (*TCF7L2*) was associated with both using second-line T2D treatments ( $P_{\text{meta}} = 2.1 \times 10^{-13}$ ) and insulin ( $P_{\text{meta}} = 2.3 \times 10^{-11}$ ) (one CS per analysis) (Extended Data Fig. 5h,i and Supplementary Tables 12 and 17).

Out of 37 CS containing loci–phenotype associations for medication changing, 24 were significant in meta-analyses ( $P < 5 \times 10^{-9}$ ) and 36 had concordant effect directions in FinnGen and EstBB ( $P_{\text{binom}} = 7.3 \times 10^{-12}$ ) (Supplementary Table 12).

## CS for discontinuation of medication use

CS in known lipid- (*PCSK9*, *LDLR* and *APOE*) and blood pressure-related (*WNT2B* and *HOXA13*) loci were associated with discontinuation of medication use in hyperlipidemia (Extended Data Fig. 5b and

Supplementary Table 18) and hypertension (Extended Data Fig. 5g and Supplementary Table 18), respectively. The lead variants had the opposite effect to the underlying related risk factors, number of medication purchases and medication changing.

## Medication-use-specific associations and cardiometabolic risk

We analyzed whether the 40 medication-use-associated loci (with 46 CS) without previous GWAS associations for related cardiometabolic traits would indicate cardiometabolic or medication-specific links. Of these loci (20 CS), 18 were associated ( $P < 5 \times 10^{-9}$ ) in the meta-analyses of FinnGen, EstBB and UKBB and the remaining 22 loci had seven CS with at least a twofold Finnish-enriched lead variant (Supplementary Table 19). We compared hyperlipidemia, hypertension and T2D medication associations with LDL<sup>18</sup>, SBP<sup>19</sup> and T2D<sup>21</sup> in previous GWAS, respectively. In total, 16 of the 35 lead variants present in external disease-related summary statistics were associated with the risk factors (multiple comparison adjusted  $P$  ( $P_{\text{adj}}$ )  $< 0.05$  at false discovery rate (FDR) 5%), whereas 29 had shared effect directions ( $P_{\text{binom}} = 7.0 \times 10^{-7}$ ) (Supplementary Table 20). Eight were associated with CAD ( $P_{\text{adj}} < 0.05$  at FDR 5%) in a meta-analysis of FinnGen, UKBB and CARDIoGRAMplusC4D<sup>37</sup> ( $n_{\text{cases}} = 113,168$ ,  $n_{\text{total}} = 811,555$ ) and 35 had concordant effect directions ( $P_{\text{binom}} = 9.3 \times 10^{-8}$ ) (Supplementary Table 21). Overall, 16 of the CS lead variants were associated ( $P < 0.05$  out of 231 endpoints) with at least one cardiometabolic disease in a list of more detailed endpoints in FinnGen (Supplementary Table 22).

Finally, we used a Bayesian framework to assess shared effect with the underlying risk factor (LDL, SBP, T2D) and CAD in the 32 novel autosomal associations for medication purchases. We found a likely

CAD association (posterior inclusion probability (PIP) > 60%) in seven loci, including a locus (*chr4:147117893:A:G/TT2C29*) associated with CAD and hyperlipidemia medication, but not with the underlying risk factor (LDL). Seven loci did not associate with either the considered risk factor or CAD (PIP > 60% for association with medication use only) (Fig. 3).

### Genetic correlations between medication use and cardiometabolic risk

Using LDSC<sup>34,35</sup>, we report ( $P < 0.05$  at 0.05 FDR) genetic correlation (RG) between the medication use and the underlying cardiometabolic risk (RG approximately 0.9–1.0). Medication use was correlated with CAD (Fig. 4). This is in line with 158 of the 333 medication-use-associated loci ( $P < 5 \times 10^{-9}$ ) being associated with CAD ( $P_{adj} < 0.05$  at FDR < 0.05, shared direction of beta; Supplementary Table 23).

Of the non-cardiometabolic conditions, ADHD, depression, educational attainment, loneliness, personality trait neuroticism and self-related health were all correlated with all three medication-use phenotypes (medication GWAS summary statistics were from FinnGen, others from external sources; Supplementary Table 24 and Extended Data Fig. 8). In a sensitivity GWAS in EstBB ( $n = 132,982$ –132,987) adjusting for educational attainment (a proxy for socioeconomic status), we did not observe any change in effect sizes or statistical significance levels of the GWS lead variants (Supplementary Table 25 and Extended Data Fig. 9).

### Polygenic cardiometabolic risk associates with medicine use

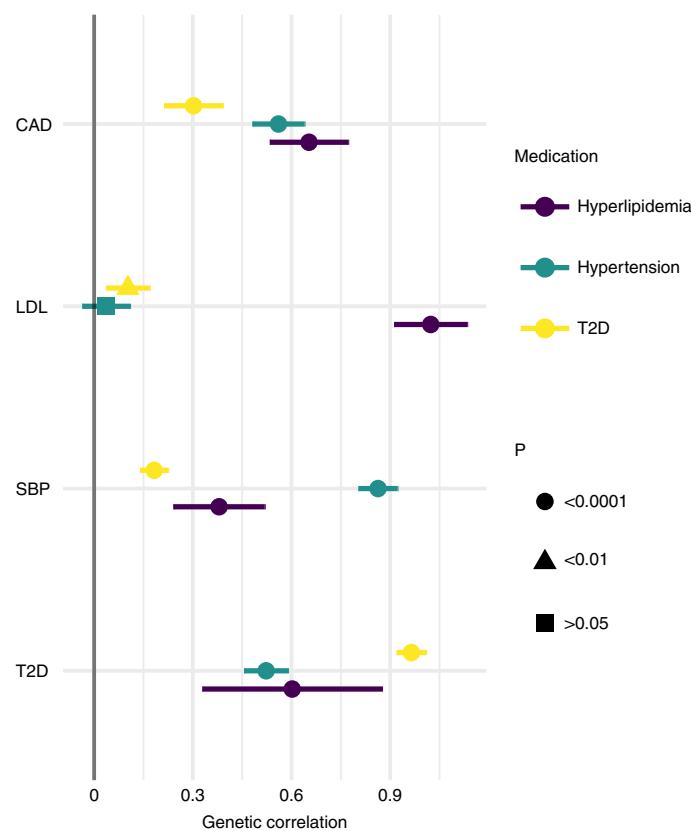
The PRS of LDL, SBP and T2D were strongly associated with the related medication-use patterns (Fig. 5 and Supplementary Table 26). Participants in the highest decile recorded, on average, 15 (linear regression coefficient (beta) = 14.6; 95% confidence interval (95% CI) = 14.2–15.0), 42 (beta = 42.0; 95% CI = 40.9–43.1) and 12 (beta = 11.9; 95% CI = 11.5–12.3) more hyperlipidemia (mean number of purchases = 12.2; s.d. = 22.8), hypertension (mean = 40.1; s.d. = 61.6) and T2D purchases (mean = 6.54; s.d. = 21.9) compared with the bottom decile, respectively. In the top decile of the LDL PRS, the odds of switching simvastatin to another statin were 103% higher (odds ratio (OR) = 2.03; 95% CI = 1.85–2.23) than in the bottom decile. In contrast, the odds of discontinuing statin use were lower in the top decile (OR = 0.421; 95% CI = 0.376–0.473). For the top decile of the SBP PRS, the odds of using more than one group of antihypertensives were 210% higher (OR = 3.11; 95% CI = 2.93–3.31) than in the bottom decile and the odds of having used all five groups of antihypertensives were 320% higher (OR = 4.21; 95% CI = 3.42–5.17). In contrast, the odds of discontinuing hypertension medication were 70% lower (OR = 0.305; 95% CI = 0.280–0.333) in the top decile. Finally, individuals in the top decile of the T2D PRS had 160% higher odds of using insulin (OR = 2.62; 95% CI = 2.21–3.10) and 120% higher odds to use second-line treatments (OR = 2.23; 95% CI = 1.96–2.54).

Participants in the top deciles of the PRS started medication use 3.9–6.3 years earlier than in the bottom deciles. All associations remained significant in analyses adjusting for the age of treatment onset (Supplementary Table 26).

A missense variant **rs4149056** in *SLCO1B1* has previously been associated with statin-induced myopathy<sup>26</sup>. We did not observe effect to statin discontinuation (OR, homozygous carriers versus non-carriers = 1.08; 95% CI = 0.957–1.220;  $P = 0.22$ ) or switching simvastatin (OR = 1.08; 95% CI = 0.982–1.190;  $P = 0.11$ ). There was a small additive effect for total lipid-lowering drug purchases (OR = 0.988; 95% CI = 0.982–0.995;  $P = 0.0004$ ) and changing (OR = 1.04; 95% CI = 1.00–1.07;  $P = 0.03$ ) but not for discontinuation (OR = 1.02; 95% CI = 0.978–1.020;  $P = 0.33$ ). We summarize results for previously reported pharmacogenomic variants in Supplementary Table 27.

### Medication-use-enhanced polygenic risk score for CAD

We used multitrait analysis of GWAS (MTAG)<sup>38</sup> to perform a joint analysis of CAD and the number of drug purchases for



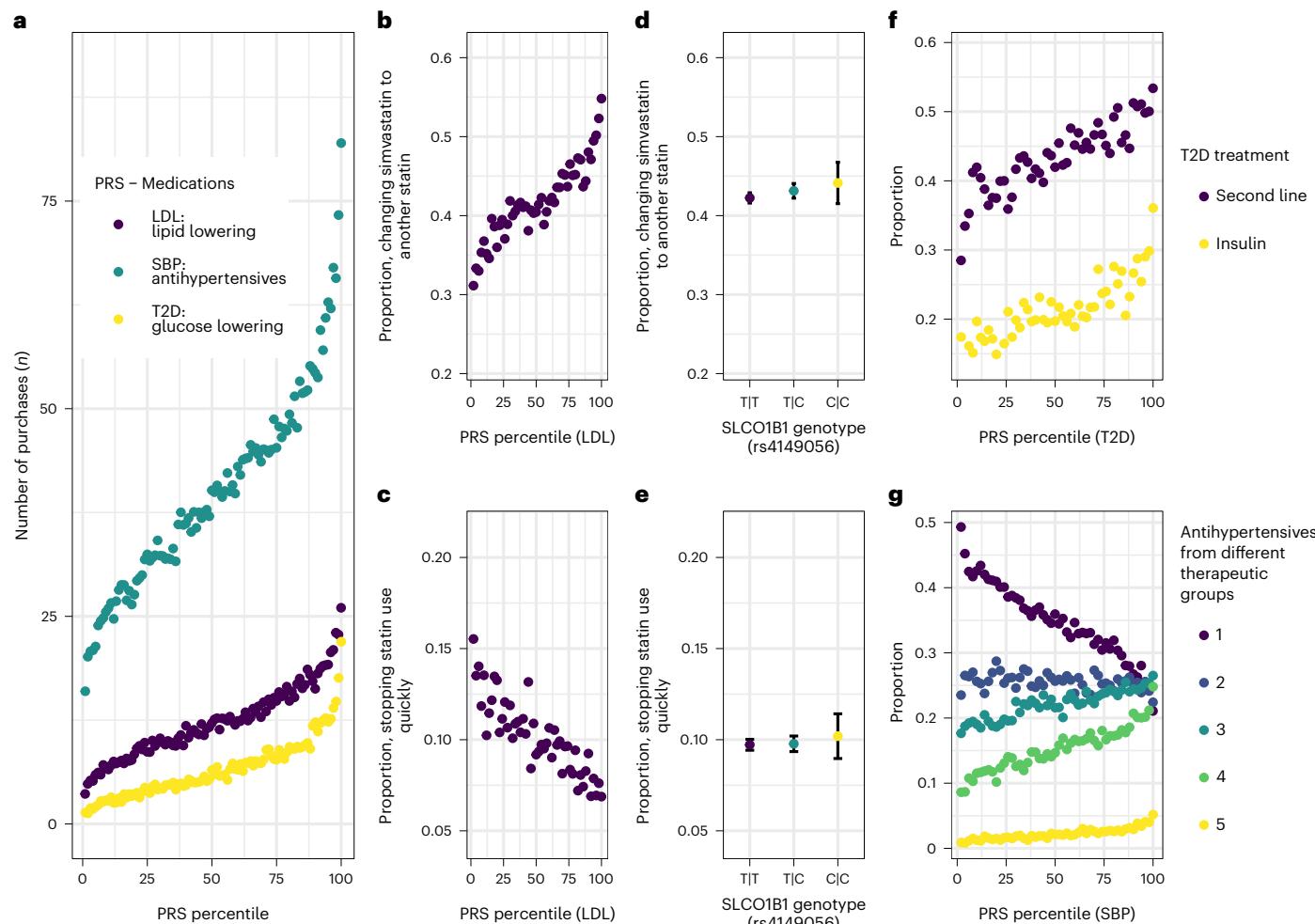
**Fig. 4 | Genetic correlations between total numbers of drugs purchased for cardiometabolic indications and the underlying cardiometabolic risk factors and CAD.**

Genetic correlations between the total number of hyperlipidemia, hypertension and T2D medication purchases ( $n = 193,933$ ) and CAD (CARDIoGRAMplusC4D<sup>37</sup>,  $n = 194,427$ ), LDL<sup>18</sup> ( $n = 340,951$ ), SBP<sup>19</sup> ( $n = 757,601$ ) and T2D<sup>21</sup> ( $n = 898,130$ ) estimated with LDSC. The association statistics for LDL, SBP, T2D and CAD used to compute genetic correlations came from previous studies. Point estimates of genetic correlation and their 95% CI, indicated using error bars, are presented. The genetic correlation between LDL and the total number of hypertension medications was not significant, whereas all other genetic correlations were significant (two-sided  $P < 0.05$  at 5% FDR).

hyperlipidemia, hypertension and T2D, restricting analyses to HapMap 3 single-nucleotide polymorphisms (SNPs)<sup>39</sup>. We built two PRS (CAD PRS and MTAG-CAD PRS) from the two GWAS summary statistics. Both PRS were associated with CAD in the UKBB (17,986 cases,  $n = 343,676$ ). Comparing the top 1% to the middle quintile, OR = 4.40 for CAD PRS (95% CI = 3.96–4.89) and 4.78 for MTAG-CAD PRS (95% CI = 4.31–5.30). The MTAG PRS had a 22% higher pseudo- $R^2$  for CAD ( $R^2_{MTAG-PRS} = 0.0424$ ,  $R^2_{PRS} = 0.0348$ , Davidson–Mackinnon  $J$ -test  $P(P_{J-test}) = 1.0 \times 10^{-208}$ ) and a 0.6 percentage points higher area under the receiver operating characteristic curve (AUC) ( $AUC_{MTAG-PRS} = 0.790$ ,  $AUC_{PRS} = 0.784$ ). A similar risk factor MTAG PRS combining risks for CAD, hyperlipidemia, hypertension and T2D resulted in an OR estimate between CAD PRS and medication-enhanced MTAG-CAD PRS (OR 99% versus 1% = 4.68, 95% CI = 4.22–5.19, AUC = 0.788,  $R^2 = 0.040$ ) (Extended Data Table 1). Assessing effect of rare variants with MAF < 1% excluded by MTAG, a CAD PRS together with a traditional weighted-sum PRS based on effect sizes of the lead SNPs of the medication-use GWAS loci performed equally, whether the rare variants were included ( $R^2 = 0.0378$ ) or not ( $R^2 = 0.0374$ ).

### Discussion

In this study, we demonstrate the highly polygenic nature of lifelong medication use in cardiometabolic conditions. We discovered hundreds



**Fig. 5 | Associations between cardiovascular risk-factor PRS and medication-use patterns.** Associations between PRS for LDL, SBP and T2D and medication-use patterns of hyperlipidemia, hypertension and T2D are presented, respectively. **a**, Associations between PRS percentiles and the number of medication purchases (LDL PRS and hyperlipidemia medications, SBP PRS and hypertension medications, T2D PRS and T2D medications,  $n = 193,933$ ). **b–e**, LDL PRS and prevalence of changing simvastatin to another statin ( $n = 45,134$ ) (**b**) and discontinuation of statin use ( $n = 76,499$ ) (**c**), in comparison with differences in

statin-associated myopathy-related SLCO1B1 genotypes ( $n = 45,134$ ,  $P = 0.11$  (**d**);  $n = 76,499$ ,  $P = 0.22$  (**e**)). **f**, Associations between T2D PRS and use of second-line T2D treatments ( $n = 31,665$ ) and insulin ( $n = 29,990$ ). **g**, Association between SBP PRS and use of hypertension medications from different numbers of distinct medication groups ( $n = 125,586$ ). PRS are split into bins of 1% (**a**) and 2% (**b,c,f,g**). The LDL, SBP and T2D PRS were computed from the GWAS association statistics from previous studies<sup>18,19,21</sup> with sample sizes of 340,951, 757,601 and 898,130, respectively. The error bars signify 95% CI.

of genetic predictors for temporal medication-use patterns: (1) the total quantity of lifelong purchases, (2) changing medication and (3) discontinuation. We report 333 independent loci associated with medication use and 495 CS with  $\geq 95\%$  posterior probability for causality. Whereas most associations have been previously linked to lipid, blood pressure or glucose-related traits, we discovered loci without previous cardiometabolic link. We developed polygenic predictors for medication use and showed how genetic association results for medication use can refine polygenic prediction.

Most genetic factors driving the differences in medication use in cardiometabolic conditions are shared with the underlying risk factors. Most loci associated with total purchases and medication changes (hyperlipidemia: *APOE*, *PCSK9* and *LDLR*; hypertension: *WNT3B*, *HOXA13*, *CASZ1* and *KCNK3*; T2D: *TCF7L2*) have been associated with cardiometabolic risk factors with shared effect directions<sup>16–22</sup>. In contrast, the associations for discontinuation (statins: *PCSK9*, *LDLR* and *APOE*; antihypertensives: *WNT3B*, *HOXA13*, *CASZ1* and *KCNK3*) had opposite effect directions. The PRS for LDL, SBP and T2D were strongly associated with medication-use patterns, supported by a reported association between T2D PRS and T2D patients' progression to use

insulin<sup>40</sup>. Genetic correlations were highest between total purchases and the treatment targets. In contrast, known pharmacogenetic variants associated with medication-related adverse events<sup>26–28</sup>, such as *SLCO1B1* increasing statin-related myopathy, were not among the leading associations. Together, these results demonstrate the central role of the underlying genetic liability in medication use.

There are several explanations for these findings. First, the total number of purchases is an aggregate phenotype combining both the length of the treatment and the number of different medications used within a certain time window. It is driven by the underlying polygenic risk for the risk factor itself associated with the age of onset<sup>24,41</sup>, levels of the measured risk factor<sup>18,42,43</sup> and hard cardiovascular events<sup>18,22,41,42</sup>. Second, these factors are all driving earlier initiation of the medication, a greater number of different therapeutic agents used and the total number of medications purchased. Third, the polygenic risk for the underlying risk factor was inversely associated with treatment discontinuation.

We propose reasons for the negative association between cardiometabolic risk and treatment discontinuation. Patients with higher genetic cardiometabolic risk are at higher risk for severe

cardiometabolic diseases, cardiovascular complications and a family history of CVD, all motivating the treatment, even if side effects are experienced. Medical professionals might try to find tolerated medication combinations if patients wanted to abandon the medications or side effects are experienced. Also, nonpharmacological lifestyle-related interventions are more likely sufficient<sup>5,6</sup> if the genetic risk is low as higher genetic risk correlates with higher measures of cardiovascular risk factors<sup>18,42</sup>. Finally, we know that higher genetic risk for individual risk factors is associated with hard CVD outcomes<sup>18,22,41,42</sup>. In secondary prevention of CVD, the treatment is more aggressive and the stricter treatment goals are unlikely to be met without ongoing pharmacotherapy. Individuals with high inherited risk have a harder time meeting the goals and need continued treatment, whereas individuals with lower genetic risk meet the goals more easily, sometimes leading to treatment discontinuation.

In total, 40 associated loci were not previously linked with the underlying risk factors. The link between medication use and cardiovascular risk, partially independent of the risk factors, may explain why these loci had not been reported in well-powered risk factor studies<sup>18,19,21</sup>, because high CVD risk indicates pharmacological treatment of CVD risk, despite normal risk factor measures<sup>5,6,44</sup>. For example, lipid-lowering treatment is initiated earlier and targets are stricter in individuals with CVD or diabetes<sup>6</sup>. Whether these associations indicate medication-specific pathways for action, including drug efficacy and side effects or are more behavioral by nature, warrants further study.

A medication-enhanced multitrait PRS for CAD showed better risk discrimination compared with a traditional single-trait CAD PRS, although the improvement was not clinically meaningful. This is in line with results for combining risk factor GWAS results with CAD for polygenic prediction<sup>45</sup>. Rare variants may provide further predictive power to the PRS.

The study has strengths and limitations. FinnGen consisted of individuals linked to the Finnish drug purchase registry covering all drug purchases since 1995, being superior to our other samples. The Finnish bottleneck population provides for genetic discovery<sup>46</sup> and 73 CS had a Finnish-enriched lead variant. Replication in datasets with enough carriers of Finnish-specific variants is therefore challenging. By combining drugs with varying mechanisms for action to maximize statistical power, the possible molecule-specific pharmacogenomic aspects might be overlooked. The medication dosage could have an effect, but dosage information was unavailable. For example, we may underestimate the effect of *SLCO1B1 rs414956* polymorphisms on medication use as the related statin-induced myopathy is dependent on simvastatin doses of at least 40 mg daily<sup>47</sup>. However, our quantitative phenotypes grouped together different medications and our binary trajectory design was dose independent, and we show the robust association between the underlying cardiovascular risk and the medication patterns even without adjustment for dosage. In addition, because of widespread off-label use and additional indications for many of the studied medications<sup>48</sup>, not all purchases were to treat cardiometabolic conditions, but our key findings display the association between the cardiometabolic risk and the studied medication patterns.

The long follow-up time in our study highlights time-dependent progression in medication-use practices, seen in the overrepresentation of simvastatin, the first widely used statin<sup>49</sup>, as the first-choice statin in FinnGen. However, the wide use of simvastatin allowed us to test associations on switching to other statins. Although the medication registry covers the whole nation of Finland, the biobank ascertainment may cause biases, including survival bias, because participants need to be alive until they are recruited (Supplementary Fig. 5). Given the replicability of associations, it is unlikely that these biases largely affect our main findings. Also, due to possible spectrum bias, the results from our quantile-based PRS analyses should be generalized to other populations with caution<sup>50</sup>. In addition, given the self-reporting of psychological and socioeconomic factors in our study, the effect of

these traits may be underestimated. Finally, although the study was conducted in three independent biobanks, all our samples were of European ancestry, limiting the generalizability of our results<sup>51,52</sup>.

In summary, we demonstrate the highly polygenic genetic architecture of lifelong medication-use patterns in hyperlipidemia, hypertension and T2D largely shared with cardiometabolic traits. Our findings highlight the possible utility of using risk-factor-related genetic information for optimizing pharmacological treatment with medicine-use-related genetic information to improve the prediction and prevention of cardiometabolic diseases.

## Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41591-022-02122-5>.

## References

1. James, S. L. et al. Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet* **392**, 1789–1858 (2018).
2. Marenberg, M. E., Risch, N., Berkman, L. F., Floderus, B. & de Faire, U. Genetic susceptibility to death from coronary heart disease in a study of twins. *N. Engl. J. Med.* **330**, 1041–1046 (1994).
3. Yusuf, S. et al. Effect of potentially modifiable risk factors associated with myocardial infarction in 52 countries (the INTERHEART study): case-control study. *Lancet* **364**, 937–952 (2004).
4. Virani, S. S. et al. Heart disease and stroke statistics—2020 update: a report from the American Heart Association. *Circulation* **141**, e139–e596 (2020).
5. Arnett, D. K. et al. 2019 ACC/AHA guideline on the primary prevention of cardiovascular disease: a report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. *J. Am. Coll. Cardiol.* **74**, e177–e232 (2019).
6. Piepoli, M. F. et al. 2016 European Guidelines on cardiovascular disease prevention in clinical practice: The Sixth Joint Task Force of the European Society of Cardiology and Other Societies on Cardiovascular Disease Prevention in Clinical Practice (constituted by representatives of 10 societies and by invited experts) Developed with the special contribution of the European Association for Cardiovascular Prevention & Rehabilitation (EACPR). *Eur. J. Preventive Cardiol.* **23**, NP1–NP96 (2016).
7. Allonen, J. et al. Mortality rate increases steeply with nonadherence to statin therapy in patients with acute coronary syndrome. *Clin. Cardiol.* **35**, E22–E27 (2012).
8. Allonen, J., Nieminen, M. S. & Sinisalo, J. Poor adherence to beta-blockers is associated with increased long-term mortality even beyond the first year after an acute coronary syndrome event. *Ann. Med.* **52**, 74–84 (2020).
9. Vrijens, B., Vincze, G., Kristanto, P., Urquhart, J. & Burnier, M. Adherence to prescribed antihypertensive drug treatments: longitudinal study of electronically compiled dosing histories. *Brit. Med. J.* **336**, 1114–1117 (2008).
10. Khunti, K., Seidu, S., Kunutsor, S. & Davies, M. Association between adherence to pharmacotherapy and outcomes in type 2 diabetes: a meta-analysis. *Diabetes Care* **40**, 1588–1596 (2017).
11. Aguayo-Orozco, A. et al. Optimizing drug selection from a prescription trajectory of one patient. *npj Digit. Med.* **4**, 150 (2021).

12. Sen, A. et al. Systematic assessment of prescribed medications and short-term risk of myocardial infarction – a pharmacopeia-wide association study from Norway and Sweden. *Sci. Rep.* **9**, 8257 (2019).

13. Leopold, J. A. & Loscalzo, J. Emerging role of precision medicine in cardiovascular disease. *Circ. Res.* **122**, 1302–1315 (2018).

14. Hamburg, M. A. & Collins, F. S. The path to personalized medicine. *N. Engl. J. Med.* **363**, 301–304 (2010).

15. King, E. A., Davis, J. W. & Degner, J. F. Are drug targets with genetic support twice as likely to be approved? Revised estimates of the impact of genetic support for drug mechanisms on the probability of drug approval. *PLoS Genet.* **15**, e1008489 (2019).

16. Sinnott-Armstrong, N. et al. Genetics of 35 blood and urine biomarkers in the UK Biobank. *Nat. Genet.* **53**, 185–194 (2021).

17. Tabassum, R. et al. Genetic architecture of human plasma lipidome and its link to cardiovascular disease. *Nat. Commun.* **10**, 4329 (2019).

18. Ripatti, P. et al. Polygenic hyperlipidemias and coronary artery disease risk. *Circ. Genom. Precis. Med.* **13**, e002725 (2020).

19. Evangelou, E. et al. Genetic analysis of over 1 million people identifies 535 new loci associated with blood pressure traits. *Nat. Genet.* **50**, 1412–1425 (2018).

20. Surendran, P. et al. Discovery of rare variants associated with blood pressure regulation through meta-analysis of 1.3 million individuals. *Nat. Genet.* **52**, 1314–1332 (2020).

21. Mahajan, A. et al. Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. *Nat. Genet.* **50**, 1505–1513 (2018).

22. Vujkovic, M. et al. Discovery of 318 new risk loci for type 2 diabetes and related vascular outcomes among 1.4 million participants in a multi-ancestry meta-analysis. *Nat. Genet.* **52**, 680–691 (2020).

23. Khera, A. V. et al. Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nat. Genet.* **50**, 1219–1224 (2018).

24. Mars, N. et al. Polygenic and clinical risk scores and their impact on age at onset and prediction of cardiometabolic diseases and common cancers. *Nat. Med.* **26**, 549–557 (2020).

25. Wu, Y. et al. Genome-wide association study of medication-use and associated disease in the UK Biobank. *Nat. Commun.* **10**, 1891 (2019).

26. SEARCH Collaborative Group. SLCO1B1 variants and statin-induced myopathy—a genomewide study. *N. Engl. J. Med.* **359**, 789–799 (2008).

27. Ahangari, N. et al. Personalised medicine in hypercholesterolemia: the role of pharmacogenetics in statin therapy. *Ann. Med.* **52**, 462–470 (2020).

28. Oliveira-Paula, G. H., Pereira, S. C., Tanus-Santos, J. E. & Lacchini, R. Pharmacogenomics and hypertension: current insights. *Pharmacogenomics Pers. Med.* **12**, 341–359 (2019).

29. Mannino, G. C., Andreozzi, F. & Sesti, G. Pharmacogenetics of type 2 diabetes mellitus, the route toward tailored medicine. *Diabetes Metab. Res. Rev.* **35**, e3109 (2019).

30. Kurki, M. I. et al. FinnGen: Unique genetic insights from combining isolated population and national health register data. Preprint at *medRxiv* <https://doi.org/10.1101/2022.03.03.22271360> (2022).

31. Leitsalu, L. et al. Cohort profile: Estonian biobank of the Estonian genome center, university of Tartu. *Int. J. Epidemiol.* **44**, 1137–1147 (2015).

32. Bycroft, C. et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203–209 (2018).

33. Halava, H. et al. Predictors of first-year statin medication discontinuation: a cohort study. *J. Clin. Lipidol.* **10**, 987–995 (2016).

34. Bulik-Sullivan, B. K. et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).

35. Bulik-Sullivan, B. et al. An atlas of genetic correlations across human diseases and traits. *Nat. Genet.* **47**, 1236–1241 (2015).

36. Wang, G., Sarkar, A., Carbonetto, P. & Stephens, M. A simple new approach to variable selection in regression, with application to genetic fine mapping. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **82**, 1273–1300 (2020).

37. Nikpay, M. et al. A comprehensive 1000 Genomes-based genome-wide association meta-analysis of coronary artery disease. *Nat. Genet.* **47**, 1121–1130 (2015).

38. Turley, P. et al. Multi-trait analysis of genome-wide association summary statistics using MTAG. *Nat. Genet.* **50**, 229–237 (2018).

39. International HapMap 3 Consortium. Integrating common and rare genetic variation in diverse human populations. *Nature* **467**, 52–58 (2010).

40. Jiang, G. et al. Obesity, clinical, and genetic predictors for glycemic progression in Chinese patients with type 2 diabetes: a cohort study using the Hong Kong Diabetes Register and Hong Kong Diabetes Biobank. *PLoS Med.* **17**, e1003209 (2020).

41. Pazoki, R. et al. Genetic predisposition to high blood pressure and lifestyle factors: associations with midlife blood pressure levels and cardiovascular events. *Circulation* **137**, 653–661 (2018).

42. Vaura, F. et al. Polygenic risk scores predict hypertension onset and cardiovascular risk. *Hypertension* **77**, 1119–1127 (2021).

43. Fava, C. et al. Prediction of blood pressure changes over time and incidence of hypertension by a genetic risk score in Swedes. *Hypertension* **61**, 319–326 (2013).

44. Rabar, S., Harker, M., O'Flynn, N. & Wierzbicki, A. S. Lipid modification and cardiovascular risk assessment for the primary and secondary prevention of cardiovascular disease: summary of updated NICE guidance. *Brit. Med. J.* **349**, g4356 (2014).

45. Inouye, M. et al. Genomic risk prediction of coronary artery disease in 480,000 adults: implications for primary prevention. *J. Am. Coll. Cardiol.* **72**, 1883–1893 (2018).

46. Locke, A. E. et al. Exome sequencing of Finnish isolates enhances rare-variant association power. *Nature* **572**, 323–328 (2019).

47. Ramsey, L. B. et al. The clinical pharmacogenetics implementation consortium guideline for SLCO1B1 and simvastatin-induced myopathy: 2014 update. *Clin. Pharmacol. Ther.* **96**, 423–428 (2014).

48. Butt, A. K. et al. Beneficial extracardiac effects of cardiovascular medications. *Curr. Cardiol. Rev.* **18**, 47–63 (2022).

49. Plosker, G. L. & McTavish, D. Simvastatin. *Drugs* **50**, 334–363 (1995).

50. Goehring, C., Perrier, A. & Morabia, A. Spectrum bias: a quantitative and graphical analysis of the variability of medical diagnostic test performance. *Stat. Med.* **23**, 125–135 (2004).

51. Martin, A. R. et al. Clinical use of current polygenic risk scores may exacerbate health disparities. *Nat. Genet.* **51**, 584–591 (2019).

52. Peterson, R. E. et al. Genome-wide association studies in ancestrally diverse populations: opportunities, methods, pitfalls, and recommendations. *Cell* **179**, 589–603 (2019).

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

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate

if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted

use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023

<sup>1</sup>Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki, Finland. <sup>2</sup>Finnish Institute for Health and Welfare, Helsinki, Finland.

<sup>3</sup>Broad Institute of the Massachusetts Institute of Technology and Harvard University, Cambridge, MA, USA. <sup>4</sup>Estonian Genome Center, Institute of Genomics, University of Tartu, Tartu, Estonia. <sup>5</sup>Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA, USA. <sup>6</sup>Department of Otorhinolaryngology – Head and Neck Surgery, University of Helsinki and Helsinki University Hospital, Helsinki, Finland. <sup>7</sup>Wilhuri Research Institute, University of Helsinki, Helsinki, Finland. <sup>8</sup>Department of Internal Medicine, University of Turku, Turku, Finland. <sup>9</sup>Division of Medicine, Turku University Hospital, Turku, Finland. <sup>10</sup>Research Programs Unit, Clinical and Molecular Metabolism, University of Helsinki, Helsinki, Finland. <sup>11</sup>Lund University Diabetes Center, Malmö, Sweden. <sup>12</sup>Folkhälsan Research Centre, Helsinki, Finland. <sup>13</sup>Department of Endocrinology, Helsinki University Hospital, Helsinki, Finland.

<sup>14</sup>Department of Public Health, Clinicum, Faculty of Medicine, University of Helsinki, Helsinki, Finland. <sup>15</sup>Department of Mathematics and Statistics, University of Helsinki, Helsinki, Finland. <sup>16</sup>Program in Medical and Population Genetics and Stanley Center for Psychiatric Research, Broad Institute of Harvard and MIT, Cambridge, MA, USA. <sup>17</sup>These authors contributed equally: Tuomo Kiiskinen, Pyry Helkkula, Kristi Krebs. \*A list of authors and their

affiliations appears at the end of the paper.  e-mail: [samuli.ripatti@helsinki.fi](mailto:samuli.ripatti@helsinki.fi)

## FinnGen

**Tuomo Kiiskinen<sup>1,2,3,17</sup>, Juha Karjalainen<sup>1,3,5</sup>, Elmo Saarentaus<sup>1,6</sup>, Nina Mars<sup>1</sup>, Arto Lehisto<sup>1</sup>, Wei Zhou<sup>3</sup>, Joel T. Rämö<sup>1</sup>, Juha Mehtonen<sup>1</sup>, Sanni Ruotsalainen<sup>1</sup>, Mutaamba Maasha<sup>3</sup>, Teemu Niiranen<sup>2,8,9</sup>, Tiinamaija Tuomi<sup>1,10,11,12,13</sup>, Veikko Salomaa<sup>2</sup>, Mitja Kurki<sup>1,3,5</sup>, Aarno Palotie<sup>1,5,16</sup>, Mark Daly<sup>1,5,16</sup>, Andrea Ganna<sup>1,3</sup>, Aki S. Havulinna<sup>1,2</sup> & Samuli Ripatti<sup>1,3,11</sup>**

## Methods

### Study sample

The data comprised 218,792 Finnish individuals from FinnGen Data Freeze 5, which includes prospective epidemiological and disease-based cohorts and hospital biobank samples (Supplementary Table 28). The data were linked by the unique national personal identification numbers to national hospital discharge (available from 1968), death (from 1969), cancer (from 1953) and drug purchase (from 1995) registries. The registry data were available until 31 December 2018.

### Drug purchase registry

The Finnish drug purchase registry managed by The Social Insurance Institution of Finland (Kela) contains all prescription drug purchases starting from 1 January 1995. All prescription drug purchases are dated and coded with the national version of the World Health Organization ATC, thus enabling molecular and higher drug-group-level classifications.

### Medication-use phenotypes

We identified the analyzed drug-use phenotypes using the drug purchase registry and ATC codes (Table 1). First, we identified all purchases of drugs used in the management of three common risk factors of CVD, hyperlipidemias, hypertension and T2D, lipid-modifying agents (ATC code starts with C10), agents used in the treatment of hypertension (five different ATC groups: C02\*, C03\*, C07\*, C08\* and C09\*) and blood glucose-lowering drugs, excluding insulin (ATC = A10B\*).

For the analysis of quantitative phenotypes of drugs used in the treatment of hyperlipidemia, hypertension and T2D, we counted all single drug purchases in these categories for all the study participants. All the nonusers were included in these phenotypes with zero purchases (Supplementary Table 1a). To exclude age-related medicine-free follow-up from the analyses, we excluded all the participants who were dead or younger than 10 years old at the beginning of the follow-up (1 January 1995), resulting in the final sample sizes of 193,933 participants.

For the analysis of binary phenotypes, we identified risk-factor-specific common drug-use patterns from the purchase registries for hyperlipidemias, hypertension and T2D from the whole dataset of 218,588 individuals alive at the beginning of the follow-up.

Among the hyperlipidemia drug users, we identified as patients (1) those who started with simvastatin (ATC: C10AA01) but then changed to another statin (ATC: C10AA\*, not C10AA01) and (2) those who started using statins but stopped the use quickly (only one to two purchases of statins, ATC = C10AA\*, last purchase more than 1 year before the end of the follow-up, excluding those discontinuing later from the analysis) (Supplementary Fig. 6).

For medications used in the treatment of hypertension, we analyzed four thresholds of the number of drugs purchased from different hypertension medicine subgroups, requiring at least one purchase during the follow-up: (3) more than one group, (4) more than two groups, (5) more than three groups and (6) all five groups used. Controls were individuals with records of subgroup purchases that are less than in the case-defining groups (for example, controls for the first analysis were individuals with purchases of drugs from one hypertension drug subgroup, for the second analysis individuals with purchases from one or two hypertension drug subgroups, and so on). Discontinuation was defined by having only one or two purchases of hypertension medication, with at least 1 year of purchase free follow-up after the last purchase (Supplementary Fig. 6).

For T2D, we identified users of (7) second-line treatments (combinations: ATC = A10BD\*, DPP4: ATC = A10BH\*, GLP-1: ATC = A10BJ\* and SGLT2: ATC = A10JK\*) and (8) those starting insulin (ATC = A10A\*) after initially starting with non-insulin T2D drugs.

For these drug-use patterns, only non-case users of hyperlipidemia (1 and 2), hypertension (3–6) and T2D (7 and 8) drugs were considered

as controls, and those participants who never purchased the related drugs were excluded.

### Genotyping and imputation

FinnGen samples were genotyped with Illumina (Illumina Inc.) and Affymetrix arrays (Thermo Fisher Scientific). For imputation, a population-specific SISu v.3 imputation reference panel (Pärn et al., manuscript in preparation) comprising 3,775 high-coverage (25–30x) whole genomes was used (public protocol: <https://www.protocols.io/view/genotype-imputation-workflow-v3-0-xbgfjw>). Postimputation quality control involved nonreference concordance analyses, checking expected conformity of the imputation information (INFO)-values distribution, MAF differences between the target dataset and the imputation reference panel, and checking chromosomal continuity of the imputed genotype calls. After these steps, variants with imputation INFO score < 0.6 or MAF < 0.0001 were excluded. For details about the genotype calling, quality controls and imputation in FinnGen, see Kurki<sup>30</sup>.

Genotyping of DNA samples from the EstBB<sup>31</sup> was done at the Core Genotyping Lab of the Institute of Genomics, University of Tartu using the Illumina Global Screening Arrays (GSA v.1.0, GSA v.2.0 and GSA v.2.0\_EST). Altogether 206,448 samples were genotyped and PLINK format files were created using Illumina GenomeStudio v.2.0.4. During the quality control all individuals with call-rate < 95% or mismatching sex, which was defined based on the heterozygosity of X chromosome and sex in the phenotype data, were excluded from the analysis. Variants were filtered by call-rate over 95% and Hardy–Weinberg equilibrium  $P$  value  $< 1 \times 10^{-4}$  (autosomal variants only). Variant positions were updated to GenomeReference Consortium Human Build 37 and all variants were changed to be from TOP strand using reference information provided by Dr. W. Rayner from the University of Oxford (<https://www.well.ox.ac.uk/~wrayner/strand/>). After quality control the dataset contained 202,910 samples. Before imputation, variants with MAF < 1% and indels were removed. Prephasing was done using the Eagle v.2.3 software 1 (number of conditioning haplotypes Eagle2 uses when phasing each sample was set to:  $-Kpbwt = 20,000$ ) and imputation was carried out using Beagle v.28Sep18.793 (refs. <sup>33,54</sup>) with an effective population size  $n = 20,000$ . As a reference, an Estonian population-specific imputation reference of 2,297 whole-genome sequencing samples was used<sup>55</sup>.

The genotyping in the UKBB was performed using the Applied Biosystems UK BiLEVE Axiom Array or the Applied Biosystems UKB Axiom Array. Imputation was done using combined Haplotype Reference Consortium, UK10K and 1000 Genomes Project phase 3 reference panels with IMPUTE4 software. For quality control, variants with INFO score  $\leq 0.8$ , MAF  $\leq 0.01$  and Hardy–Weinberg equilibrium  $P$  value  $\leq 1 \times 10^{-10}$  were excluded. For a detailed description on genotyping, imputation and quality control in the UKBB, see Bycroft<sup>32</sup>.

### GWAS

For GWAS analyses, we used scalable and accurate implementation of generalized mixed model (SAIGE), which controls for unbalanced case-control ratios of binary phenotypes and sample relatedness by a two-step approach: (1) fitting the null model with the covariates age, sex, the first ten principal components and genotyping batch (for batches with at least ten cases and controls) as fixed effects and assuming the random effects are distributed as  $N(0, \tau \psi)$ , where  $\psi$  is the genetic relationship matrix and  $\tau$  is the additive genetic variance, and (2) testing for association between each genetic variant and phenotypes by applying saddle-point approximation to the score test statistics, which obtains more accurate  $P$  values than the normal approximation in the presence of the unbalanced case-control ratios of binary phenotypes<sup>56</sup>.

For the analysis of quantitative phenotypes, we used SAIGE mixed-modell linear regression. Before the final association analyses, we transformed the count data to continuous phenotypes (Supplementary

Fig. 7) with the rank-based inverse normal transformation (INT) using the ‘rankNorm’ function of the R-package rNORM<sup>57</sup>, as required (<https://saigegit.github.io/SAIGE-doc/>). Age, age<sup>2</sup>, follow-up time (end of the follow-up—1 January 1995), sex (imputed with PLINK, not included in the sex-stratified analyses), the first ten first principal components of ancestry and genotyping batch (for batches with at least ten cases and controls) were used as fixed-effect covariates. To account for the possible type I error of rare variants using INT-normalized phenotypes, we excluded variants with MAF < 0.005. Also, we excluded participants dead or under 10 years old at the beginning of the follow-up.

### Genome-wide association analyses in EstBB and UKBB

We identified the drug-use phenotypes at the EstBB in the same way as described above for the FinnGen cohort by using the drug purchasing data of biobank participants and purchased drug ATC codes. Sample sizes for genome-wide analyses were as follows. Quantitative drug purchases phenotypes: hyperlipidemia  $n = 184,880$  ( $n_{\text{purchases}>0} = 28,659$ ), hypertension  $n = 184,892$  ( $n_{\text{purchases}>0} = 74,534$ ) and T2D  $n = 184,880$  ( $n_{\text{purchases}>0} = 12,795$ ) (there were 12 new participants in the hypertension analysis because their inclusion in the EstBB was confirmed by a recorded hypertension prescription purchase); binary phenotypes switching from simvastatin: 3,013 patients and 2,054 controls; discontinuation of statins: 5,005 patients and 21,679 controls; the four thresholds of the number of drugs purchased from different hypertension medicine subgroups: (1) more than one group, 45,365 patients and 29,334 controls, (2) more than two groups, 26,490 patients and 48,209 controls, (3) more than three groups, 12,542 patients and 62,157 controls, (4) more than four groups, 3,103 patients and 71,596 controls; users of traditionally second-line T2D drugs: 3,501 patients and 9,268 controls; starting insulin after other T2D drugs: 1,728 patients and 10,614 controls.

We conducted GWAS using the SAIGE software. For the analysis of quantitative phenotypes we normalized the phenotypes with the rank-based INT using the ‘rankNorm’ function of the R-package rNORM and used SAIGE mixed-model linear regression for genome-wide analysis, adjusting for covariates age at the end of follow-up, age at the end of follow-up squared, follow-up time (end of the follow-up to 1 January 2004), sex and the first ten principal components of the genotype matrix. All individuals who were less than 10 years old at the beginning of follow-up were excluded from the analysis. For the analysis of binary phenotypes we used SAIGE mixed-model logistic regression and adjusted for the covariates age at the end of follow-up, sex and ten principal components. For the first steps of SAIGE models we used only the genotyped data and set the parameters trace CV cutoff to 0.002, ratio CV cutoff to 0.001 and LOCO to TRUE. For the second steps of SAIGE models, we used imputed data and the parameters minMAF was set to 0.0001, minMAC to 1 and LOCO to TRUE.

In the UKBB we identified proxy drug-use phenotypes for the total number of medication purchases in the UKBB general practitioner prescription data (data field 42039) by counting all prescriptions for hyperlipidemia (British National Formulary code (BNF) = 02120\*), hypertension (BNF = 0202\*, 0204\*, 0205\* and 020602\*) and T2D (BNF = 060102\*) and normalizing the phenotypes. We did not consider the binary phenotypes in the UKBB because the data were restricted to prescriptions by general practitioners and there are no data available for whether the prescriptions were dispensed. The final sample was restricted to 188,846 participants, with at least one prescription of any kind in the records. We conducted GWAS using the SAIGE software. In the UKBB, instead of the lacking follow-up time covariate, we adjusted for birth year.

### Meta-analyses

GWAS analyses for (1) the total number of purchases in hyperlipidemia, hypertension and T2D in FinnGen, EstBB and UKBB, (2) binary medication purchase phenotypes in FinnGen and EstBB and (3) CAD

summary statistics from FinnGen, UKBB<sup>56</sup> (<https://pheweb.org/UKB-SAIGE/pheno/411>) and CARDIOGRAM4+<sup>37</sup> were combined using an inverse-variance weighted fixed-effect meta-analysis with GWAMA v.2.2.2 (ref. <sup>58</sup>). For the medication-use pattern meta-analyses, all loci ( $\pm 1.5$  Mb) with a potential association ( $P < 5 \times 10^{-6}$  for the lead variant) in FinnGen were included. A stricter criterion for genome-wide significance was used ( $P < 5 \times 10^{-9}$ )<sup>16</sup>. For chromosomes 1–22, we included all variants that were present in FinnGen and at least one additional cohort (EstBB, UKBB or both). Chromosome X variants were only analyzed in FinnGen and interpreted genome-wide significance only if meeting the stricter meta-analysis threshold of significance ( $P < 5 \times 10^{-9}$ ).

### Bayesian fine-mapping

Using GWAS summary statistics from the medication pattern analyses, we fine-mapped all regions where the lead variant reached the standard genome-wide significant  $P$  value of  $\sim 5 \times 10^{-8}$  in FinnGen using the sum of single effects method<sup>36</sup> to allow discovery of associations driven by Finnish-enriched variants, possibly missed in other populations. Fine-mapping regions were formed using a 3 Mb window ( $\pm 1.5$  Mb) around each lead variant and combining overlapping regions into one. Linkage disequilibrium (LD) between each variant was calculated from individual-level FinnGen data and a minimum pairwise correlation value of 0.5 ( $r^2 = 0.25$ ) for variants in a CS was required. We report all 95% CS and for CS with over ten variants, we report those variants with a  $\geq 0.01$  probability of being causal. We report which CS are in loci with a genome-wide significant association for the given medication-use phenotype in the meta-analysis ( $P < 5 \times 10^{-9}$ ) and which CS lead variants are Finnish enriched, with a minor allele over twofold enriched in Finns as compared to non-Finnish-Swedish-Estonian European in gnomAD v.2.0.1<sup>59</sup>.

### Automatized novelty reporting

To evaluate the possible novelty of the found CS associations, we compared our results to those in the GWAS catalog<sup>60</sup> and the largest available GWAS summary statistics with non-overlapping samples (lipid-lowering agent patterns against statin-use adjusted LDL<sup>18</sup>, blood pressure-related medication patterns against hypertension medication adjusted SBP summary statistics<sup>19</sup>, and glucose-lowering medication patterns against T2D<sup>21</sup>), using a 3 Mb window ( $\pm 1.5$  Mb) around each 95% CS top variant. For loci with no corresponding genome-wide significant ( $P < 5 \times 10^{-8}$ ) variants in the respective GWAS summary statistics, we checked all the previous associations in the GWAS catalog. If no associations related to the indication of the drugs used (lipid-modifying agents: lipid levels, lipid-level measurements and morbidities, and use of lipid-modifying agents; glucose-lowering agents: glucose-related measurements, morbidities and use of glucose altering agents; blood pressure medications: blood pressure measurements and morbidities, and use of blood pressure modifying agents) were found, we considered the association to be of potential novelty. We also report whether these possibly novel loci have been previously associated with any CVDs or any other known CVD risk factors. (For a more detailed description of the automated novelty reporting method, see <https://aaltodoc.aalto.fi/handle/123456789/41629>.)

### Cardiometabolic phenotype-wide association study in FinnGen

For the lead variants of the CS in the possibly novel loci, we performed a cardiometabolic phenotype-wide association study in FinnGen, testing the lead variants associations among 231 cardiometabolic endpoints: T1D, T2D, all endpoints from the ICD-10 main category IX ‘Diseases of the circulatory system’ with a name starting with I9\_\* and hyperlipidemia endpoints (E4\_HYPERCHOL, E4\_HYPERGLYCER, E4\_HYPERLIPMIX, E4\_HYPERCHYLO, E4\_HYPERLIPNAS, E4\_LIPODEF, E4\_LIPOPROTNAS). A  $P$  value of  $< 0.05/231$  (the number of the considered cardiometabolic endpoints) ( $= 2.16 \times 10^{-4}$ ) was considered statistically significant. The full list of FinnGen endpoints with detailed

descriptions of their definitions ([https://www.finngen.fi/en/researchers/clinical\\_endpoints](https://www.finngen.fi/en/researchers/clinical_endpoints)) and thorough documentation (<https://finngen.gitbook.io/documentation/methods/endpoints>) are available online. The publicly available results for FinnGen R5 genome-wide association study are available online (<https://r5.finngen.fi/>).

### LD score regression

To estimate the genetic correlation between the medication phenotypes and the underlying risk factors we used LDSC software<sup>34,35</sup>. LDSC uses the LD score regression method, which quantifies the contribution of each variant by examining the relationship between test statistics and LD. We used LD scores readily calculated from the 1000 Genomes Project (EUR)<sup>61</sup>. To restrict to a set of common, well-imputed variants, we retained only those SNPs in the HapMap 3 reference panel<sup>39</sup>. The medication-use summary statistics from GWAS conducted in FinnGen were compared with summary statistics from external sources (Supplementary Table 24).

### PRS

We calculated genome-wide PRS for LDL, SBP and T2D using the PRS-continuous shrinkage priors (CS) method, a Bayesian method to infer posterior effect sizes for variants using summary statistics from GWAS and an external LD reference panel<sup>62</sup>. For the input weights, we used available summary statistics from external GWAS for LDL<sup>18</sup>, SBP<sup>19</sup>, T2D<sup>21</sup> and a selected set of other health-related traits (Supplementary Table 28) and limited our variants to those in the HapMap 3 (ref. <sup>39</sup>) reference panel. The 1000 Genomes Project (EUR)<sup>61</sup> was used as an LD reference panel. Traditional PRS for medication use were calculated as effect-size weighted sums of the lead variants in the GWS loci in FinnGen (Supplementary Table 3), after pruning the summary statistics to variants in the UKBB. Two traditional weighted-sum PRS were calculated, one excluding and the other including rare variants (MAF < 1%).

### Bayesian meta-analysis

We analyzed the hypothesized shared effect of specific variants using a previously reported method of Bayesian meta-analysis<sup>63</sup>. This approach uses GWAS summary statistics to infer an approximate BF (ABF) for different assumptions about heterogeneity between studies, and the probability of a model is derived from its ABF relative to the sum of all model ABFs for a particular variant. We used this Bayesian framework to compare the effect between medication phenotypes in FinnGen and the respective cardiometabolic risk factors (RISK) in studies in non-Finns when the overlap between cases was negligible<sup>18,19,21</sup>. For CAD effect, we used the GWAS summary statistics from the meta-analysis (CAD). The models used in all tests were the null model, where the effect on RISK and CAD is zero (MED ONLY); endpoint-specific models, where the effect is assumed to be non-zero for only one endpoint (RISK or CAD); and the shared effect model, where the effect is fixed or highly correlated between RISK and CAD (CAD + RISK). In model comparison, a model was considered supported if the posterior probability of the model was  $\geq 60\%$ . Prior probability distribution was uniform. Analyses were performed for loci found in all summary statistics (medication use, risk factor and CAD).

### Traditional and MTAG-PRS for CAD in the UKBB

For the medication-enhanced multitrait PRS of CAD, we jointly analyzed the summary statistics for the total number of purchased medications in the treatment of hyperlipidemia, hypertension and T2D (in FinnGen) and CAD<sup>37</sup> using MTAG<sup>38</sup> with variants found in HapMap 3 (ref. <sup>39</sup>) and all the summary statistics with a MAF of  $>0.01$ , resulting in 1,168,733 variants. The PRS were calculated using PRS-CS as described above, the traditional PRS using the CAD summary statistics and the MTAG-PRS using the MTAG-CAD summary statistics, both with the same 1,168,733 variants. For the risk factor enhanced multitrait PRS, we used a similar approach with FinnGen R5 GWAS summary statistics for hypertension

([https://r5.finngen.fi/pheno/I9\\_HYPHTENS](https://r5.finngen.fi/pheno/I9_HYPHTENS)), T2D (<https://r5.finngen.fi/pheno/T2D>) and a custom hyperlipidemia GWAS (cases: cases of FinnGen R5 endpoints E4\_HYPERCHOL, E4\_HYPERGLYCER, E4\_HYPERLIPMIX, E4\_HYPERLIPNAS or International Classification of Diseases 10th revision (ICD-10) diagnosis code E78.0–5 in the Register of Primary Health Care Visits/Avohilmo',  $n_{\text{cases}} = 29,837$ ,  $n_{\text{controls}} = 188,955$ ).

### CAD definition in the UKBB

In the analyses testing the associations between the CAD PRS and CAD in the UKBB, CAD was defined as (1) any of I20–I25, I46 or R96 (ICD-10) as the primary or secondary cause of death (from data fields 40,001 and 40,002, age from data field 40,007), (2) any of I20.0, I21–I22 (ICD-10) or 410, 4110 (ICD-9) in the hospital inpatient records (from data fields 41,270 and 41,271), or (3) any coronary revascularization procedure (OPCS-4, variable 41,272, codes K40, K41, K42, K43, K44, K45, K46, K49, K50 and K75, and age defined based on data field 41,282; OPSC-3, data field 41,273, code 3,043, self-reported operations, data field 20,004, codes 1,070 and 1,095, with age being defined based on data field 20,010). Diagnoses dating both before and after the study recruitment were considered. The analyses were restricted to participants of White British ethnicity (self-reported, data field 21,000).

### Statistical models

R (v.3.6.0, 4.0.3, 4.1.2 and 4.1.3) was used for the analyses of age or time dependencies of the significant associations. We performed interaction analyses for the FinnGen GWS loci lead variants by introducing G  $\times$  age or G  $\times$  follow-up time interaction terms in the (linear or logistic) regression models, where age and follow-up time in the interaction terms were dummy variables split to be 0 or 1 by their sample medians. In addition, the same covariates were used as in the original GWAS analyses (including continuous age and follow-up time). The analyses were restricted to 134,695 non-related independent FinnGen R5 samples. For the PRS calculation, we used PLINK 2.0. R (v.4.0.0–4.1.3) was used in the analysis of PRS and drug-use patterns. Separate models were created for continuous normalized PRS and decile stratified PRS. In the analysis of associations between the PRS and the medication-use patterns, linear regression was used for the continuous traits and logistic regression for the binary traits. Age, sex, follow-up time, the first ten principal components and genotyping batch were included as covariates, and the sensitivity analyses were restricted to medication users only and adjusted for age of onset for the medication use. For the analyses studying the relationship between the PRS and the age of onset we used linear regression and included, age, sex, the first ten principal components and genotyping batch as covariates. In the analyses in the UKBB testing associations between PRS and CAD diagnosis, we used a logistic regression model with age, sex and the first ten principal components as covariates. Davidson–MacKinnon *J*-test was used to test the specification of the non-nested models<sup>64</sup> and the Nagelkerke pseudo- $R^2$  was used as a measure of model fit<sup>65</sup>. AUCs were calculated by splitting the dataset by random sampling into training (75%) and validation datasets (25%), repeated for a total of 100 rounds.

### Ethics

Patients and control subjects in FinnGen provided informed consent for biobank research, based on the Finnish Biobank Act. Alternatively, older research cohorts, collected before the start of FinnGen (in August 2017), were based on study-specific consents and later transferred to the Finnish Biobanks after approval by Valvira, the National Supervisory Authority for Welfare and Health. Recruitment protocols followed the Biobank protocols approved by Valvira. The Coordinating Ethics Committee of the Hospital District of Helsinki and Uusimaa approved the FinnGen study protocol HUS/990/2017. The FinnGen study is approved by the Finnish Institute for Health and Welfare, approval number THL/2031/6.02.00/2017, amendments THL/1101/5.05.00/2017,

THL/341/6.02.00/2018, THL/2222/6.02.00/2018, THL/283/6.02.00/2019, THL/1721/5.05.00/2019, digital and population data service agency VRK43431/2017-3, VRK/6909/2018-3, VRK/4415/2019-3, the Social Insurance Institution (KELA) KELA 58/522/2017, KELA 131/522/2018, KELA 70/522/2019, KELA 98/522/2019, and Statistics Finland TK-53-1041-17. The Biobank Access Decisions for FinnGen samples and data utilized in FinnGen Data Freeze 5 include: THL Biobank BB2017\_55, BB2017\_111, BB2018\_19, BB\_2018\_34, BB\_2018\_67, BB2018\_71, BB2019\_7, BB2019\_8, BB2019\_26, Finnish Red Cross Blood Service Biobank 7.12.2017, Helsinki Biobank HUS/359/2017, Auria Biobank AB17-5154, Biobank Borealis of Northern Finland\_2017\_1013, Biobank of Eastern Finland 1186/2018, Finnish Clinical Biobank Tampere MH0004, Central Finland Biobank 1-2017 and Terveystalo Biobank STB 2018001. The EstBB study was approved by the Ethics Review Committee on Human Research of the University of Tartu. At recruitment, participants signed an informed consent to allow follow-up linkage of their electronic health records. UKBB has approval from the North West Multi-center Research Ethics Committee as a Research Tissue Bank approval. All participants provided an informed consent for their participation in UKBB.

## Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

## Data availability

The FinnGen GWAS associations for medication-use patterns can be explored with the PheWeb portal (<https://med.finngen.fi>). The FinnGen release 5 GWAS results for the clinical endpoints can be browsed with the FinnGen PheWeb portal (<https://r5.finngen.fi/>). Registered researchers can access the medication-use pattern GWAS summary statistics, which have been added to be part of FinnGen R5 public release ([https://www.finngen.fi/en/access\\_results](https://www.finngen.fi/en/access_results)). The FinnGen data may be accessed by approved researchers through Finnish Biobanks' FinBB portal ([www.finbb.fi](http://www.finbb.fi); email: info.fingenious@finbb.fi). Access to EstBB (<https://genomics.ut.ee/en/content/estonian-biobank>) and access to UKBB (<http://www.ukbiobank.ac.uk/resources/>) is restricted for approved researchers and can be requested. Previously reported GWAS associations can be accessed at the NHGRI-EBI GWAS Catalog (<https://www.ebi.ac.uk/gwas/>) and gnomAD can be accessed at <https://gnomad.broadinstitute.org/>.

## Code availability

The full genotyping and imputation protocol for FinnGen is described at <https://doi.org/10.17504/protocols.io.xbgfjw>. Codes for FinnGen pipelines can be accessed at <https://github.com/FINNGEN>. Codes for ATC- and BNF-based medication phenotypers for the quantitative phenotypes and code (based on the qqman package: <https://cran.r-project.org/web/packages/qqman/index.html>) to draw a Manhattan plot with added functionality to highlight two sets of SNPs with two different colors can be accessed at <https://github.com/yuj1r0/Medication-Use-Patterns>.

## References

53. Browning, B. L., Zhou, Y. & Browning, S. R. A one-penny imputed genome from next-generation reference panels. *Am. J. Hum. Genet.* **103**, 338–348 (2018).
54. Browning, S. R. & Browning, B. L. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *Am. J. Hum. Genet.* **81**, 1084–1097 (2007).
55. Mitt, M. et al. Improved imputation accuracy of rare and low-frequency variants using population-specific high-coverage WGS-based imputation reference panel. *Eur. J. Hum. Genet.* **25**, 869–876 (2017).
56. Zhou, W. et al. Efficiently controlling for case-control imbalance and sample relatedness in large-scale genetic association studies. *Nat. Genet.* **50**, 1335–1341 (2018).
57. McCaw, Z. R., Lane, J. M., Saxena, R., Redline, S. & Lin, X. Operating characteristics of the rank-based inverse normal transformation for quantitative trait analysis in genome-wide association studies. *Biometrics* **76**, 1262–1272 (2020).
58. Mägi, R. & Morris, A. P. GWAMA: software for genome-wide association meta-analysis. *BMC Bioinform.* **11**, 288 (2010).
59. Karczewski, K. J. et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **581**, 434–443 (2020).
60. Buniello, A. et al. The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res.* **47**, D1005–D1012 (2019).
61. 1000 Genomes Project Consortium. A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
62. Ge, T. et al. Polygenic prediction via Bayesian regression and continuous shrinkage priors. *Nat. Commun.* **10**, 1776 (2019).
63. Trochet, H. et al. Bayesian meta-analysis across genome-wide association studies of diverse phenotypes. *Genet. Epidemiol.* **43**, 532–547 (2019).
64. Davidson, R. & MacKinnon, J. G. Several tests for model specification in the presence of alternative hypotheses. *Econometrica* **49**, 781–793 (1981).
65. Nagelkerke, N. J. A note on a general definition of the coefficient of determination. *Biometrika* **78**, 691–692 (1991).

## Acknowledgements

We would like to thank S. Kivikko, H.-Y. Shen and U. Tuomainen for management assistance. We would like to thank all participants of the FinnGen study for their generous participation. S.Ri was supported by the Academy of Finland Center of Excellence in Complex Disease Genetics (grant 312062), Academy of Finland (grant 285380), the Finnish Foundation for Cardiovascular Research, the Sigrid Juselius Foundation and the University of Helsinki HiLIFE Fellow grant. The funding agencies had no role in the design and conduct of the study, the collection, analysis and interpretation of data, the writing of the manuscript or the decision to submit it for publication. The FinnGen project is funded by two grants from Business Finland (HUS4685/31/2016 and UH4386/31/2016) and nine industry partners (AbbVie, AstraZeneca, Biogen, Celgene, Genentech, GlaxoSmithKline, MSD, Pfizer and Sanofi). The following biobanks are acknowledged for collecting the FinnGen project samples: Auria Biobank (<https://www.auria.fi/biopankki/en>), THL Biobank (<https://thl.fi/fi/web/thl-biopankki>), Helsinki Biobank (<https://www.terveytskyla.fi/helsinginbiopankki/en>), Northern Finland Biobank Borealis (<https://www.pppshp.fi/Tutkimus-ja-opetus/Biopankki>), Finnish Clinical Biobank Tampere ([https://www.tays.fi/en-US/Research\\_and\\_development/Finnish\\_Clinical\\_Biobank\\_Tampere](https://www.tays.fi/en-US/Research_and_development/Finnish_Clinical_Biobank_Tampere)), Biobank of Eastern Finland (<https://ita-suomenbiopankki.fi/>), Central Finland Biobank (<https://www.ksshp.fi/fi-FI/Potilaalle/Biopankki>) and the Finnish Red Cross Blood Service Biobank (<https://www.bloodservice.fi/Research%20Projects/biobanking>). A.S.H. was supported by the Academy of Finland (grant 321356). T.T. was supported by grants from Folkhälsan Research Foundation, the Academy of Finland Center of Excellence in Complex Disease Genetics (grants 312072 and 336826). T.K. and P.H. were supported by the Doctoral Programme in Population Health, University of Helsinki. P.H. was also supported by a stipend from Nylands Nation, University of Helsinki. EstBB would like to thank all participants and staff of the EstBB for their contribution to this research and the analytical work of EstBB was carried out in part in the High Performance Computing Center of the University of Tartu. L.M. and K.K. were supported by grants from the Estonian Research Council PRG184 and by the European Union through the European Regional

Development Fund (Project 2014-2020.4.01.15-0012). S.Ri and L.M. were supported by EU H2020 funded project Safepolymed. We thank all the participants of the UKBB. This research was conducted in the UKBB Resource under application numbers 31063 (GWAS) and 22627 (PRS analyses).

## Author contributions

S.Ri., L.M., A.H., T.K., P.H., K.K. and J.K. conceptualized the study. S.Ri., T.K., L.M., A.H., P.H., K.K., J.K., N.M., E.S., A.L., W.Z., M.K., M.P., A.G., J.M. and M.C. designed the analyses. T.K., K.K. and W.Z. performed genome-wide association analyses. T.K., P.H., K.K., N.M., E.S., W.Z., A.L., M.C. and J.M. performed statistical analyses. N.M. provided statistical and methodological support for the PRS analyses. S.J., J.T., S.Ru., K.V. and M.R. provided methodological support. M.M. built the interactive PheWeb portal for the results. T.K., P.H., K.K., A.G., L.M., A.H. and S.R. mainly drafted the manuscript. All authors participated in writing and reviewing the manuscript.

## Funding

Open Access funding provided by University of Helsinki including Helsinki University Central Hospital.

## Competing interests

The authors declare no competing interests.

## Additional information

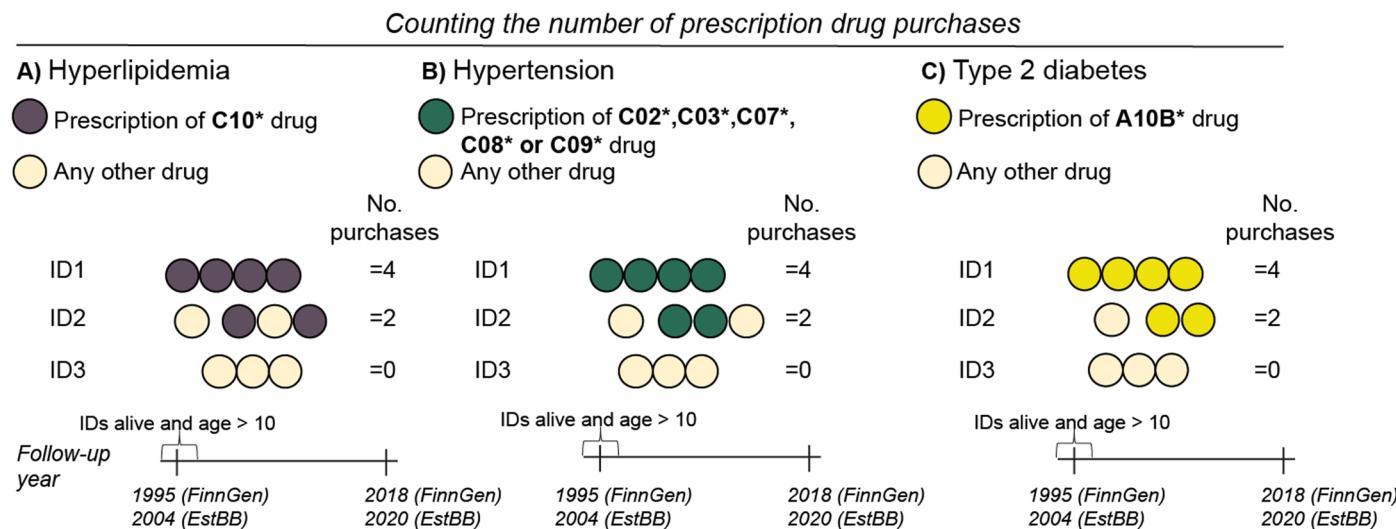
**Extended data** is available for this paper at <https://doi.org/10.1038/s41591-022-02122-5>.

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41591-022-02122-5>.

**Correspondence and requests for materials** should be addressed to Samuli Ripatti.

**Peer review information** *Nature Medicine* thanks P. Franks and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Primary Handling Editor: A. M. Ranzoni, in collaboration with the *Nature Medicine* team.

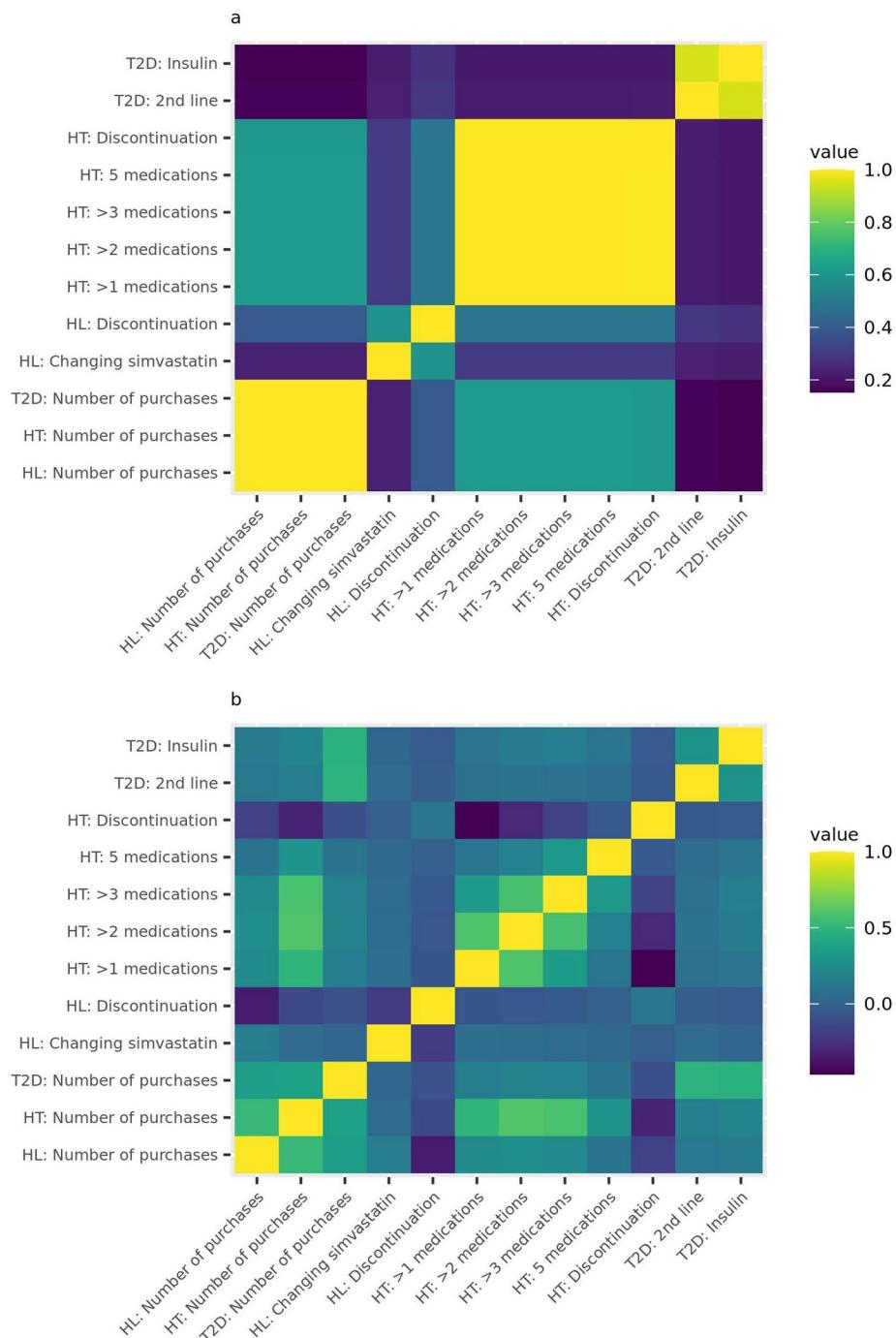
**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).



**Extended Data Fig. 1 | Drug use quantitative phenotypes.** Drugs used in the treatment of a) Hyperlipidemia b) Hypertension and c) Type 2 diabetes. Dots indicate the prescriptions and text in bold indicates the drug ATC code for the specific phenotype. We added all the single purchases in underlying drug groups,

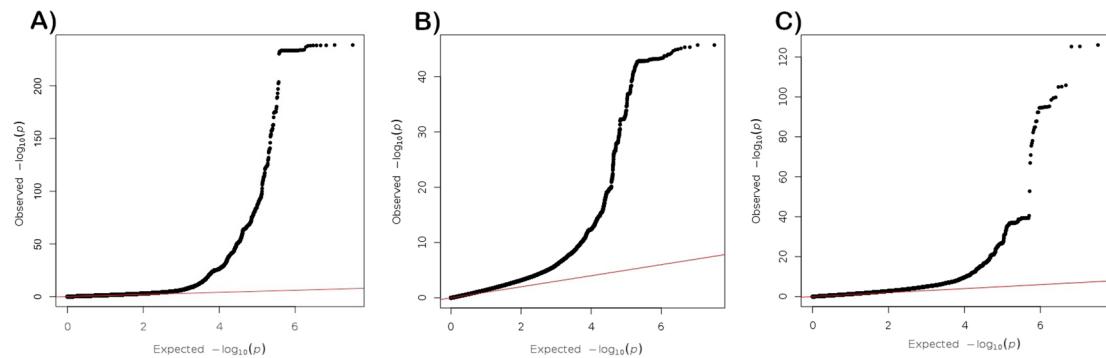
the non-users were included with 0 purchases of the specific drug. Participants dead before or less than 10 years old at the start of follow-up (Jan 1st, 1995 for FinnGen and Jan 1st, 2004 for EstBB) were excluded from the analysis.





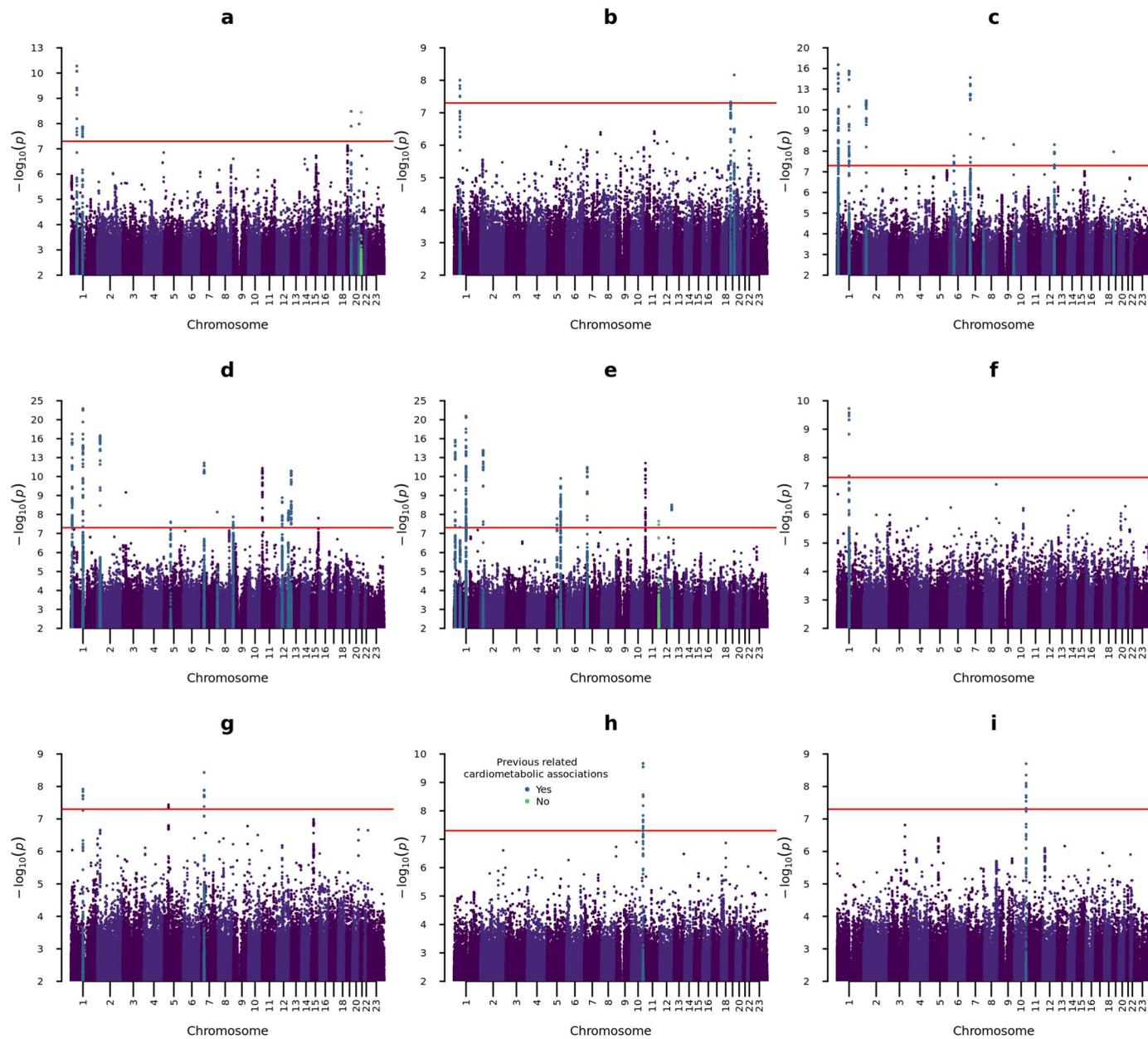
**Extended Data Fig. 3 | Correlation metrics between the medication use pattern phenotypes.** HL = Hyperlipidemia, HT = Hypertension, T2D = Type 2 Diabetes. **a**) Sample overlap displayed by the Jaccard Index (the intersection of samples divided by the union of samples) for each phenotype. **b**) Correlation

between each phenotype combination calculated in the overlapping sample intersection (Pearson's correlation coefficient for continuous-continuous phenotype pairs, Point-biserial correlation coefficient for continuous-binary pairs, and Phi-coefficient for binary-binary pairs).



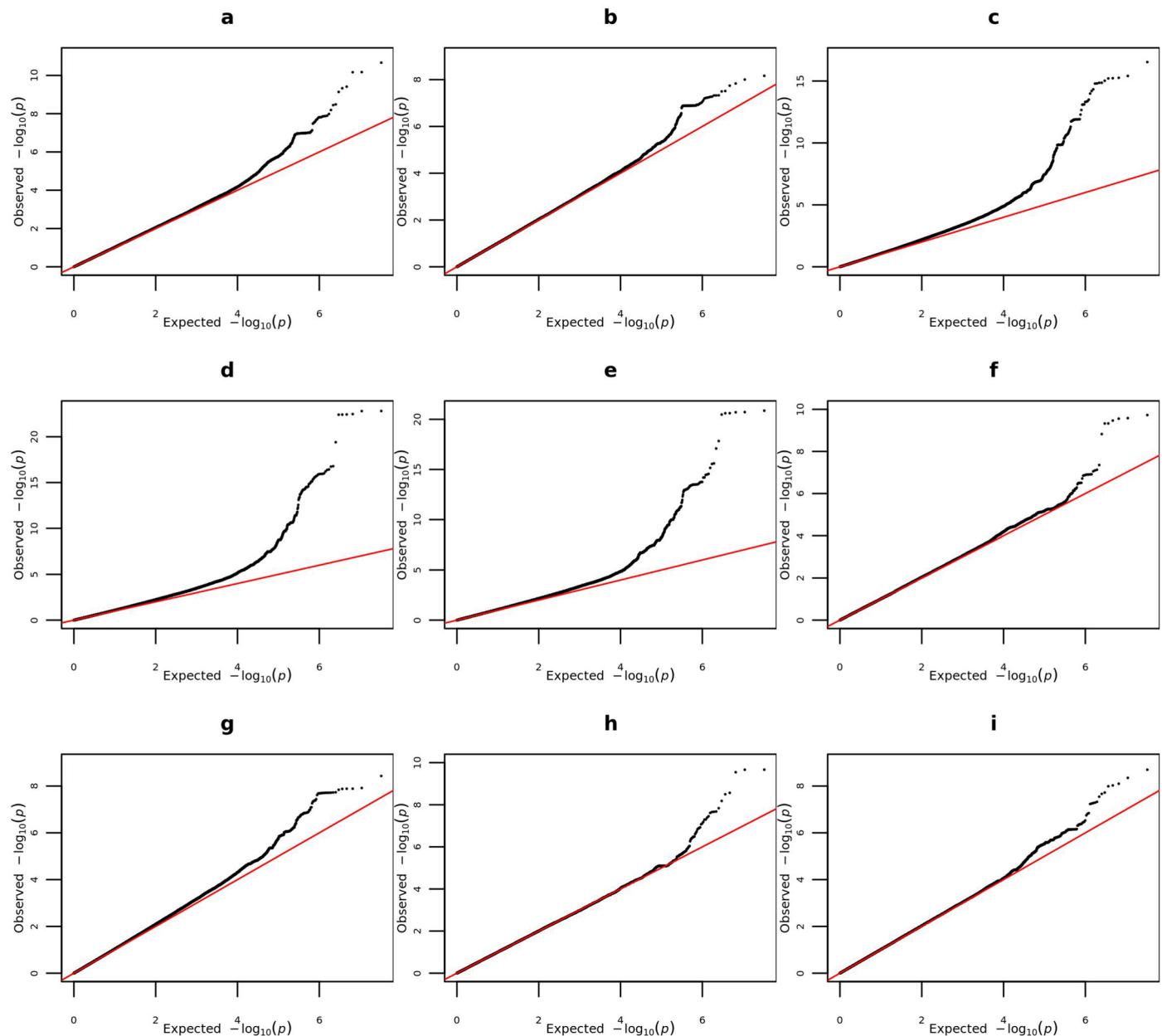
**Extended Data Fig. 4 | QQ-plots, number of purchases (FinnGen).** Two-sided  $-\log_{10}$ -transformed P values of quantitative SAIGE mixed-model GWAS analyses for the total number of prescriptions of **a**) Drugs used in the treatment of

hyperlipidemia **b**) Drugs used in the treatment of hypertension **c**) Drugs used in the treatment of type 2 diabetes (excluding insulins) in 193,933 FinnGen R5 participants.


**Extended Data Fig. 5 | Manhattan plots, binary analyses (FinnGen).**

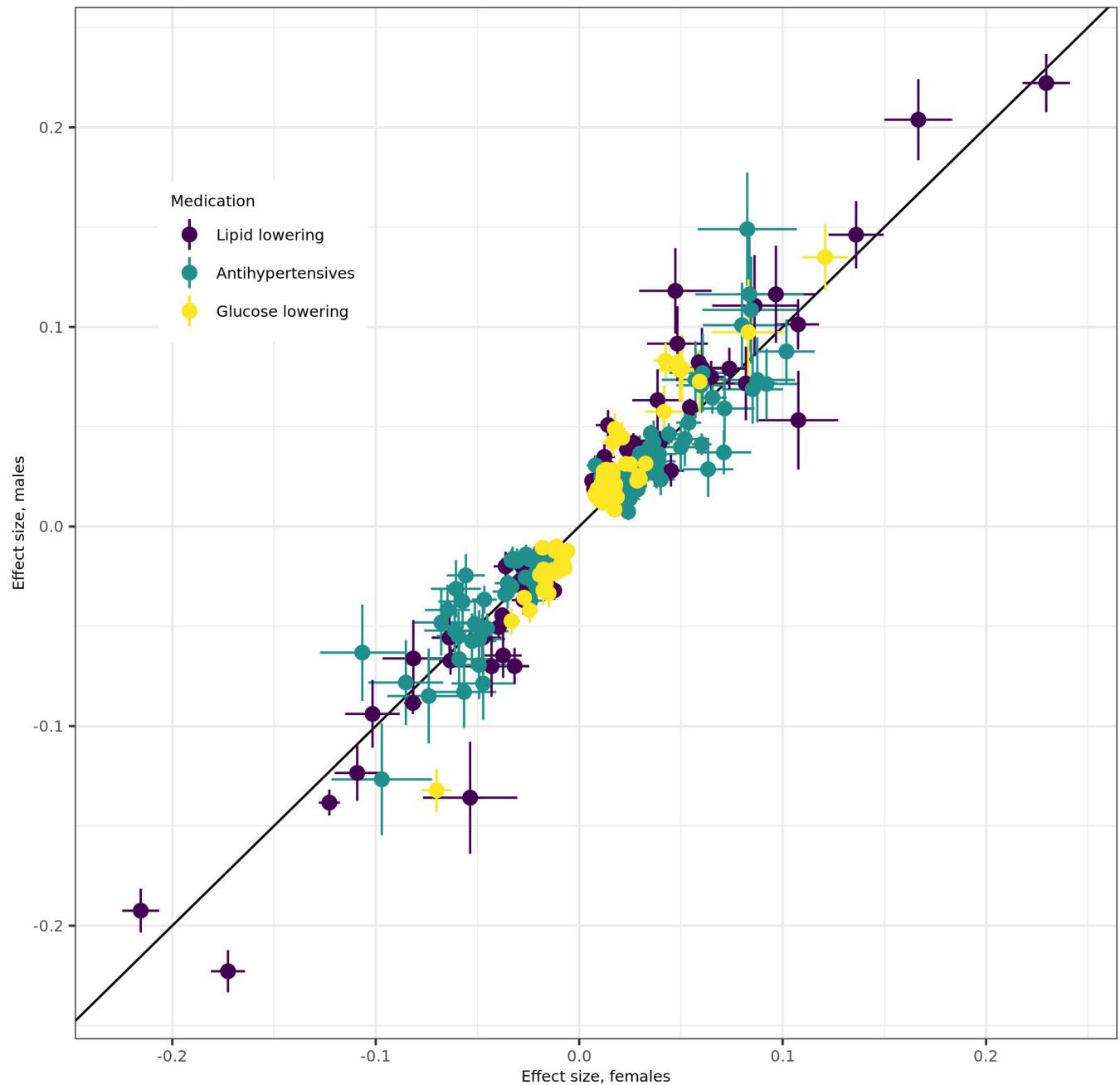
Manhattan plots, binary analyses (FinnGen). Two-sided  $-\log_{10}$ -transformed P values of binary SAIGE mixed-model GWASs for **a**) Changing simvastatin to another statin **b**) Discontinuation of statin use **c**) Hypertension medications from more than 1 different therapeutic group **d**) Hypertension medications from more than 2 different therapeutic groups **e**) Hypertension medications from more than 3 different therapeutic groups **f**) Hypertension medications from all 5 therapeutic groups **g**) Discontinuation of hypertension medication use **h**)

Using 2nd line type 2 diabetes medications **i**) Using insulin in type 2 diabetes. All loci containing  $\geq 195\%$  credible set(s) are highlighted related to their previously reported related cardiometabolic associations; loci that have been previously associated with lipid-related traits (**a-b**), blood pressure-related traits (**c-g**), or blood glucose-related traits (**h-i**) in blue and potentially novel associations in green. The red horizontal line signifies genome-wide significance ( $p < 5 \times 10^{-8}$ ) without multiple testing corrections.



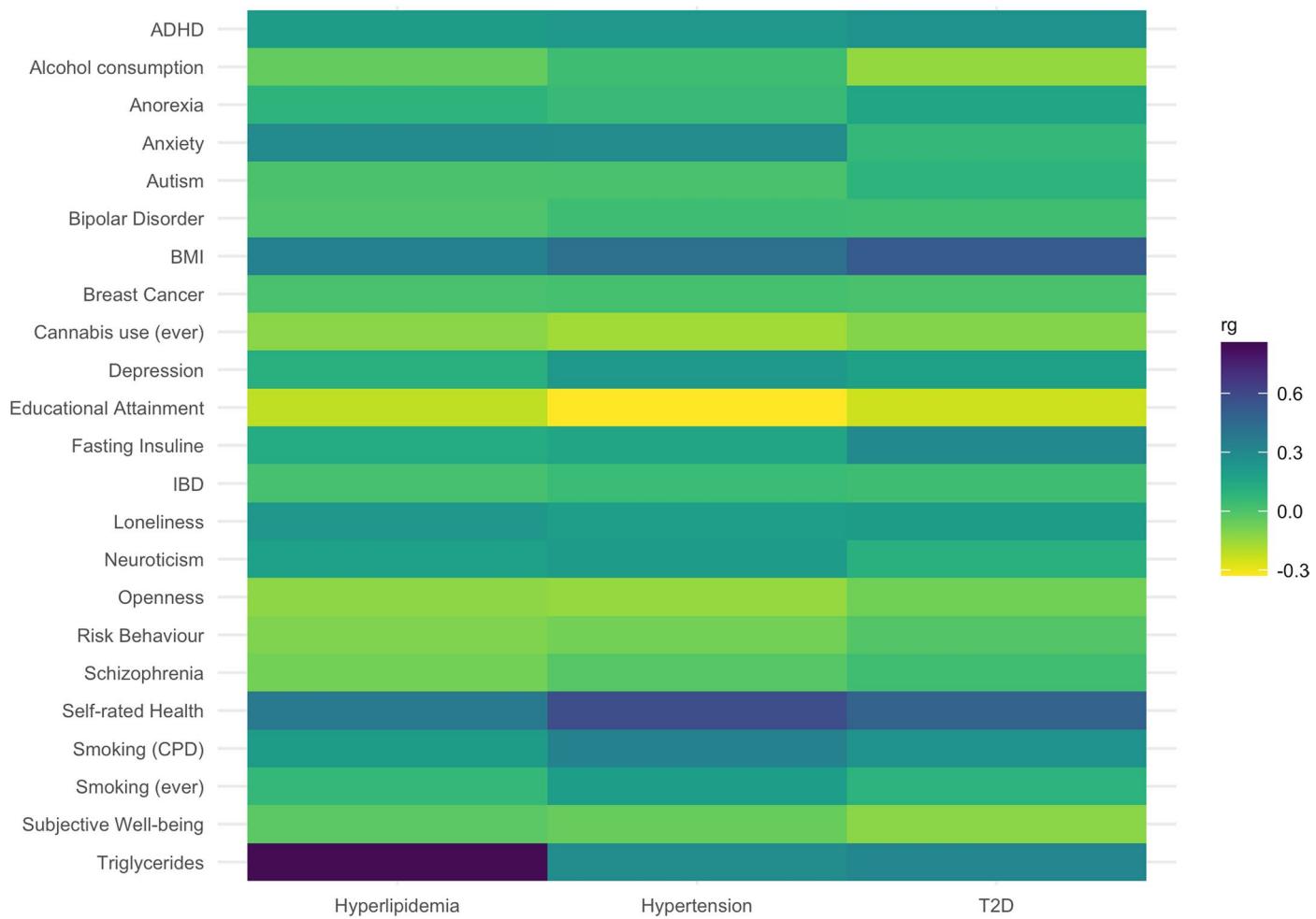
**Extended Data Fig. 6 | QQ-plots, binary analyses (FinnGen).** Two-sided  $-\log_{10}$ -transformed P values of binary SAIGE mixed-model GWASs for **a**) Changing simvastatin to another statin **b**) Discontinuation of statin use **c**) Discontinuation of hypertension medication use **d**) Hypertension medications from more than 1 different therapeutic group **e**) Hypertension medications from more than

different therapeutic groups **f**) Hypertension medications from more than 3 different therapeutic groups **g**) Hypertension medications from all 4 therapeutic groups **h**) Using 2nd line type 2 diabetes medications **i**) Using insulin in type 2 diabetes. The error bands indicate 95% confidence intervals for the quantiles of a normal distribution.

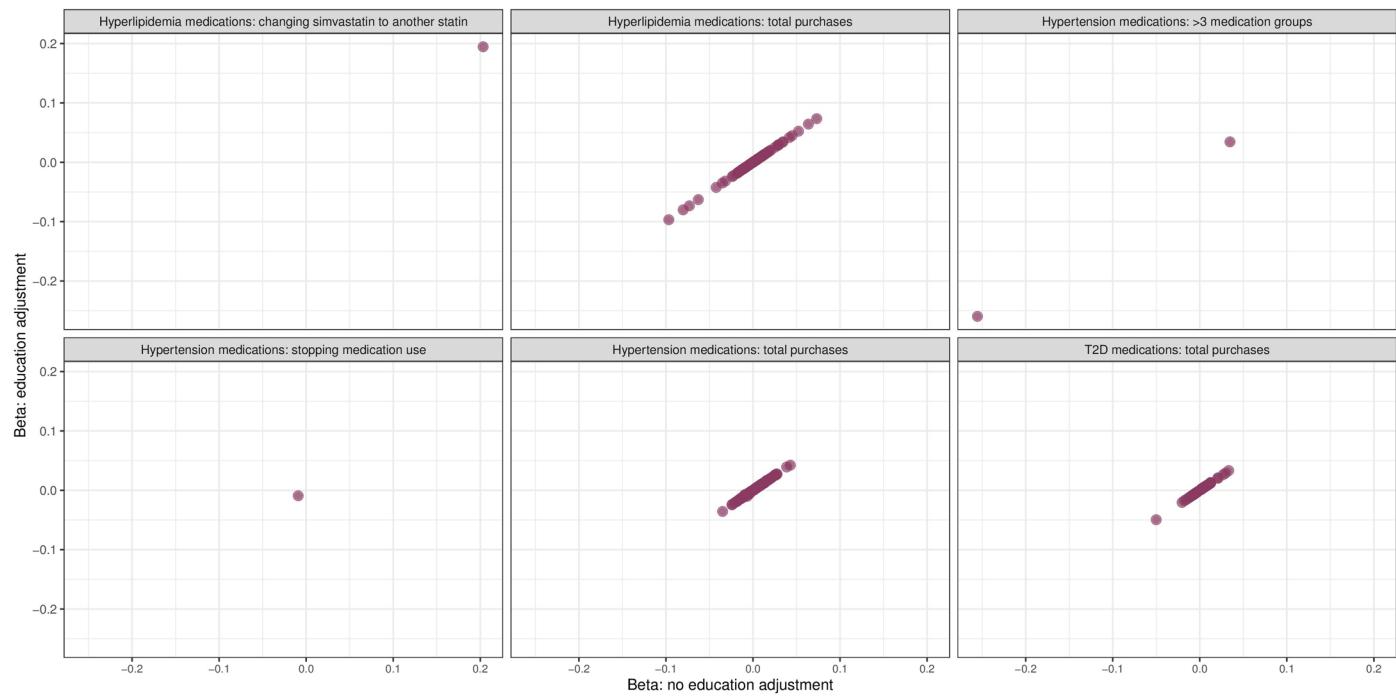


**Extended Data Fig. 7 | Effect sizes in sex-stratified genome-wide association analyses.** Effect sizes (regression coefficient betas) with 95% confidence intervals (error bars) of the lead variants in genome-wide significant loci (in the analyses of all participants) for females ( $n = 107,231$ ) and males ( $n = 86,702$ ) in sex-

stratified genome-wide analyses for the total number of prescriptions of drugs used in the treatment of hyperlipidemia, and type 2 diabetes (excluding insulins) in 193,933 FinnGen R5 participants.



**Extended Data Fig. 8 | LDSC genetic correlations.** Linkage disequilibrium score regression (LDSC) genetic correlation estimates: the total number of medication purchases (hyperlipidemia, hypertension, and type 2 diabetes) and selected cardiometabolic and socioeconomic, psychological, and psychiatric traits.



**Extended Data Fig. 9 | Betas in EstBB without and with educational attainment.** Effects of the independent 289 medication use lead association (in FinnGen) found in EstBB without (x-axis) and with (y-axis) the inclusion of educational attainment as a covariate in the GWAS models. Including educational attainment had no effect on either effect directions or sizes.

**Extended Data Table 1 | Comparison of associations between the traditional and MTAG-based coronary artery disease (CAD) polygenic risk scores and CAD in the UKBB. The UKBB contained 343,676 samples (17,986 CAD cases). The middle quintile of the PRS (40th to 60th percentiles) was used as the reference group. The AUCs and Nagelkerke pseudo-R<sup>2</sup> were calculated using PRSs as continuous variables in the models including the PRS and all covariates or only the PRS. 75% of the full sample was used for training and 25% for validation**

PRS	OR: >99% PRS	OR: >95% PRS	OR: ≤5% PRS	OR: ≤1% PRS	AUC – covariates, full sample	AUC – full model, training sample	AUC – full model, validation sample	AUC – PRS only, validation sample	R <sup>2</sup> – PRS only, full sample
CAD	4.10 (3.69-4.56)	2.06 (1.98-2.15)	0.51 (0.46-0.58)	0.27 (0.20-0.36)	0.755 (0.749-0.760)	0.785 (0.783-0.787)	0.784 (0.779-0.789)	0.635 (0.628-0.643)	0.0348
MTAG-CAD (medications)	4.78 (4.31-5.30)	2.24 (2.15-2.34)	0.45 (0.39-0.51)	0.24 (0.17-0.33)	0.755 (0.749-0.760)	0.791 (0.789-0.793)	0.790 (0.785-0.795)	0.650 (0.642-0.657)	0.0424
MTAG-CAD (risk factors)	4.68 (4.22-5.19)	2.22 (2.13-2.32)	0.51 (0.45-0.57)	0.26 (0.19-0.35)	0.755 (0.749-0.760)	0.789 (0.787-0.790)	0.788 (0.783-0.793)	0.644 (0.637-0.652)	0.0396

## Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

### Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size ( $n$ ) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided  
*Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g.  $F$ ,  $t$ ,  $r$ ) with confidence intervals, effect sizes, degrees of freedom and  $P$  value noted  
*Give  $P$  values as exact values whenever suitable.*
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's  $d$ , Pearson's  $r$ ), indicating how they were calculated

*Our web collection on [statistics for biologists](#) contains articles on many of the points above.*

### Software and code

Policy information about [availability of computer code](#)

Data collection For this study, no software was used for data collection.

Data analysis Software: R versions v3.6.0, v4.0.2, v4.0.3 and v4.1.2 (<https://www.r-project.org/>), PLINK v2.0, Eagle v2.3, AxiomGT1 algorithm for Affymetrix data (Thermo Fisher Scientific, Santa Clara, CA, USA), Beagle v.28Sep18.793, PRS-CS (<https://github.com/getian107/PRSs>), IMPUTE4, SISu v3, SAIGE (version 0.35.8.8, 0.43.1, <https://github.com/weizhouUMICH/SAIGE>), LDSC v1.0.1 (<https://github.com/bulik/ldsc>), GWAMA v2.2.2

Codes for FinnGen pipelines can be accessed at <https://github.com/FINNGEN>. Analysis pipelines: FineMapping pipeline (<https://github.com/FINNGEN/finemapping-pipeline>), genetic ancestry and PCA pipeline ([https://github.com/FINNGEN/pca\\_kinship](https://github.com/FINNGEN/pca_kinship)), PRS-CS pipeline (<https://github.com/FINNGEN/CS-PRS-pipeline>). The full genotyping and imputation protocol for FinnGen is described at <https://doi.org/10.17504/protocols.io.xbgfijw>.

Codes for ATC- and BNF-based medication phenotypers for the quantitative phenotypes and code (based on the qqman package: <https://cran.r-project.org/web/packages/qqman/index.html>) to draw a Manhattan plot with added functionality to highlight two sets of SNPs with two different colors can be accessed at <https://github.com/yuj1r0/Medication-Use-Patterns>.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

## Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

The FinnGen GWAS associations for medication use patterns can be explored with the PheWeb portal (<https://med.finngen.fi>). The summary statistics have been added to be part of FinnGen R5 public release ([https://www.finngen.fi/en/access\\_results](https://www.finngen.fi/en/access_results)). The FinnGen release 5 GWAS results for the clinical endpoints can be browsed with the FinnGen PheWeb portal (<https://r5.finngen.fi/>). The FinnGen data may be accessed through Finnish Biobanks' FinBB portal (web link: [www.finbb.fi](http://www.finbb.fi), email: [info.fingenious@finbb.fi](mailto:info.fingenious@finbb.fi)). Access to Estonian Biobank (<https://genomics.ut.ee/en/content/estonian-biobank>) and access to UK Biobank (<http://www.ukbiobank.ac.uk/resources/>) can be requested. Previously reported GWAS associations can be accessed at The NHGRI-EBI GWAS Catalog (<https://www.ebi.ac.uk/gwas/>) and gnomAD can be accessed (<https://gnomad.broadinstitute.org/>).

## Human research participants

Policy information about [studies involving human research participants](#) and [Sex and Gender in Research](#).

### Reporting on sex and gender

Sex was imputed with PLINK. All participants were included in the analyses regardless of sex, and imputed sex was included as a covariate in all analyses, excluding sex-stratified analyses. Effect sizes of genome-wide significant loci lead variants were compared in sex-stratified analyses.

### Population characteristics

Our main dataset, FinnGen release 5, consisted of 218,792 genotyped individuals of Finnish ancestry with a total of 5,118,565 years of drug-registry-based follow-up. Of the cohort 56.5% were female and the mean age at the end of the follow-up was 59.8 years. A total of 44,343,661 drug purchases were recorded and 3,650,495 (8.2%) of these were drugs used in the treatment of hyperlipidemia, hypertension, or T2D. The Estonian Biobank (EstBB) consisted of 184,892 individuals (females: 65.5%, mean age: 44.2 years) and the UK Biobank (UKBB) consisted of 343,676 individuals (females: 53.7%, mean age: 56.9 years).

### Recruitment

Data in FinnGen Data Freeze 5 are administered by regional biobanks (Auria Biobank, Biobank of Central Finland, Biobank of Eastern Finland, Biobank of Eastern Finland, Borealis Biobank, Helsinki Biobank, Tampere Biobank), the Blood Service Biobank, the Terveystalo Biobank, and biobanks administered by the Finnish Institute for Health and Welfare. EstBB is a population-based cohort of 200,000 participants. At recruitment, participants have signed a consent to allow follow-up linkage of their electronic health records. UKBB comprises phenotype data from 500,000 volunteer individuals from the UK population aged between 40 and 69 during recruitment in 2006–2010. Data for all participants have been linked with national Hospital Episode statistics.

### Ethics oversight

The FinnGen Study protocol UHUS/990/2017 was approved by the Ethics Review Board of the Hospital District of Helsinki and Uusimaa. EstBB study was approved by the Ethics Review Committee on Human Research of the University of Tartu. UKBB has approval from the North West Multi-centre Research Ethics Committee (MREC) as a Research Tissue Bank (RTB) approval.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

## Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences       Behavioural & social sciences       Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://nature.com/documents/nr-reporting-summary-flat.pdf)

## Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

### Sample size

Sample sizes were determined by the current biobank sample sizes of FinnGen, EstBB and UKBB. The sample sizes were considered sufficient as performing genome-wide association studies in biobanks of these sizes (hundreds of thousands), including meta-analysis, is the state-of-the-art for genomic discovery for polygenic traits.

### Data exclusions

In FinnGen and EstBB we excluded participants dead or less than 10 years at the beginning of the follow-up from the quantitative analyses (not expected to have any prescription medication purchases). In UKBB the analyses were restricted to "white British" ethnicity (self-reported, UKBB Data-field 21000), and participants without any prescriptions were excluded from the GWAS analyses.

## Replication

For the GWAS analyses, quantitative phenotypes were analyzed in 3 biobanks: FinnGen, EstBB and UKBB. Binary GWAS analyses were performed in FinnGen and EstBB. A high concordance of effect directions (87.2%) was checked between the biobanks for all 303 significant ( $p < 5e-8$ ) loci lead variants in FinnGen. A meta-analysis was performed for all loci with a suggestive association ( $p < 5e-6$ ) in FinnGen: quantitative traits (FinnGen, EstBB, UKBB) and binary traits (FinnGen + EstBB). 333 loci were significant ( $p < 5e-9$ ), including 94 loci not significant ( $p \geq 5e-8$ ) in the initial FinnGen analyses. Of the all 347 CS containing fine-mapped loci-phenotype associations (loci lead variant  $p < 5e-8$  and  $\geq 1$  95% CS) in FinnGen, 282 (81.3%) were genome-wide significant in the meta-analysis ( $p < 5e-9$ ).

## Randomization

Analyses were performed in population based biobanks. Participants were allocated into groups by their medication use behavior (quantitative and binary phenotypes). Age, Age2, follow-up time (end of the follow-up – 1.1.1995), sex (imputed with PLINK, not included in the sex-stratified analyses), the first 10 first principal components of ancestry, and genotyping batch (for batches with at least 10 cases and controls) were used as covariates.

## Blinding

Blinding was not relevant as no intervention occurred. The data points were collected from the biobank participants medical records and blood samples (genotype data), and all data were collected before any of the analyzes of this study were planned (so this study protocol could not have had any effect on any data points). Also, blinding of the studied predictors (genomes), was not applicable. The data was pseudonymized, and the participants could not be identified nor interacted with.

## Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

### Materials & experimental systems

n/a	Involved in the study
<input checked="" type="checkbox"/>	Antibodies
<input checked="" type="checkbox"/>	Eukaryotic cell lines
<input checked="" type="checkbox"/>	Palaeontology and archaeology
<input checked="" type="checkbox"/>	Animals and other organisms
<input checked="" type="checkbox"/>	Clinical data
<input checked="" type="checkbox"/>	Dual use research of concern

### Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	ChIP-seq
<input checked="" type="checkbox"/>	Flow cytometry
<input checked="" type="checkbox"/>	MRI-based neuroimaging