Socio-demographic and genetic risk factors for drug adherence and persistence: a retrospective nationwide and biobank study across 5 medication classes and 1 845 665 individuals

M. Cordioli[1,#], A. Corbetta[1,2,3,#], H.M. Kariis[4,#], S. Jukarainen[1], P. Vartiainen[1], T. Kiiskinen[1], M. Ferro[1], FinnGen, Estonian Biobank Research Team, M. Niemi[5,6,7], S. Ripatti[1,8,9], K. Lehto[4], L. Milani[4], A. Ganna[1,10,11]

[1] Institute for Molecular Medicine Finland, University of Helsinki, Helsinki, Finland
[2] CHDS - Health Data Science Center, Human Technopole, Milan, Italy
[3] MOX - Laboratory for Modeling and Scientific Computing, Department of Mathematics, Politecnico di Milano, Milan, Italy
[4] Estonian Genome Centre, Institute of Genomics, University of Tartu, Tartu, Estonia
[5] Department of Clinical Pharmacology, University of Helsinki, Helsinki, Finland
[6] Individualized Drug Therapy Research Program, University of Helsinki, Helsinki, Finland
[7] Department of Clinical Pharmacology, HUS Diagnostic Center, Helsinki University Hospital, Helsinki, Finland
[8] Broad Institute of MIT and Harvard, Cambridge, MA, USA
[9] Department of Public Health, University of Helsinki, Helsinki, Finland
[10] Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA, USA
[11] Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA, USA

[#] These authors contributed equally to this work
Correspondence to Andrea Ganna (andrea.ganna@helsinki.fi)

Abstract

Low drug adherence is a major obstacle to the benefits of pharmacotherapies and it is therefore important to identify factors associated with discontinuing or being poorly adherent to a prescribed treatment regimen. Using high-quality nationwide health registry data and genome-wide genotyping, we evaluate the impact of socio-demographic and genetic risk factors on adherence and persistence for 5 common medication classes that require long-term, regular therapy (N = 1 814 591 individuals from Finnish nationwide registries, 217 005 with genetic data from Finland and Estonia).

Need for social assistance and immigration status showed a notable negative effect on persistence and adherence across the examined medications (odd ratios between 0.48 and 0.82 for persistence and between 1.1% to 4.3% decrease in adherence) while demographic and health factors showed comparably modest or inconsistent effects. A genome-wide scan did not identify genetic variants associated with the two phenotypes, while some pharmacogenes (i.e. *CYP2C9* and *SLCO1B1*) were modestly associated with persistence, but not with adherence. We observed significant genetic correlations between medication adherence and participation in research studies. Overall, our findings suggest that socio-economically disadvantaged groups would benefit from targeted interventions to improve the dispensing and uptake of pharmacological treatments.

Introduction

Drug adherence is defined as the extent to which patients follow the prescribed therapeutic regimen when taking medications [1]. High adherence is considered to be crucial for the effectiveness of pharmacological treatments [2], and improving adherence has been considered one of the leading challenges for healthcare professionals [3]. A better understanding of the risk factors driving adherence is fundamental to helping clinicians improve their clinical practice and identify patients who are at risk of poor adherence at an earlier stage.

While there are multiple dimensions of adherence, some related to the healthcare system and the mode of therapy delivery [4,5], studies have shown that individuals with low adherence in placebo arms of randomized controlled trials have higher mortality[6]. This suggests that adherence to pharmacotherapy may be a surrogate marker for overall health behavior ("healthy adherer" effect) and that risk factors for adherence might be shared across different drugs.

Accurately estimating patients drug usage and adherence is not a trivial task. Traditionally questionnaires have been used, but with the increasing availability of electronic health records, studies have started investigating determinants of drug adherence and persistence using pharmacy administrative databases or insurance claims data [7–11]. However, most studies focus on a single medication or condition and consider adherence over a limited follow-up available due to data limitations. Few studies have investigated determinants of adherence across different medications [12,13], but have only considered a limited number of demographic and socioeconomic risk factors.

Beyond socio-demographic factors, genetics might play a role in drug adherence. Pharmacogenomic studies have identified several genetic variants relevant to drug response that can increase an individual's risk for adverse drug reactions [14] and a recent randomized trial suggested that pre-emptive testing with a 12-gene pharmacogenetic panel can significantly reduce the incidence of clinically relevant adverse drug reactions [15]. Because an adverse drug reaction decreases the likelihood of a patient continuing to take a medication, genetics might impact adherence via this mechanism. Nonetheless, the impact of clinically relevant pharmacogenetic variants on drug adherence has not yet been studied in large biobank-based studies and considering multiple medications. Moreover, identifying novel genetic variants associated with adherence might help to discover new aspects of drug response. An alternative,

but not mutually exclusive hypothesis, is that genetics might capture underlying health and behavioral aspects that are associated with adherence. In this scenario, polygenic scores (PGS) for certain diseases or behavioral traits might help predict drug adherence.

Here, we conducted a comprehensive and systematic investigation of adherence and persistence to drug treatments across multiple patient-related dimensions encompassing demographic, socioeconomic, and genetic factors. We used high-quality nationwide electronic health record data on prescription medication purchases made at pharmacies to accurately estimate two established drug-taking phenotypes [16–19] (**Methods**): (a) *persistence*, defined as purchasing medications for at least one year versus early discontinuation of the medication (i.e. after one purchase) and (b) *adherence*, measured as Medication Possession Ratio (MPR), which is the "proportion (or percentage) of days supply obtained during a specified time period" (**Figure 1, Methods**). We considered five commonly prescribed medications that are typically used for long-term, continuous, regular therapy.

We first explored the role of multiple health, demographics, and socioeconomic factors in FinRegistry [20], a nationwide cohort study. Next, we considered the role of genetics in a subset of the Finnish population from the FinnGen study [21] and the Estonian Biobank [22], where genome-wide genotype information is available. We focused on well-known pharmacogenes as well as on genome-wide effects. We performed a genome-wide association study (GWAS) to identify potential novel genetic associations with adherence and persistence and we evaluated if these results can be used for risk stratification. Finally, we used genetic correlations analysis to examine the overlap between persistence and adherence and other diseases or behavioral traits, and whether PGSs capturing these traits can be used to predict the two drug-taking phenotypes of interest

77

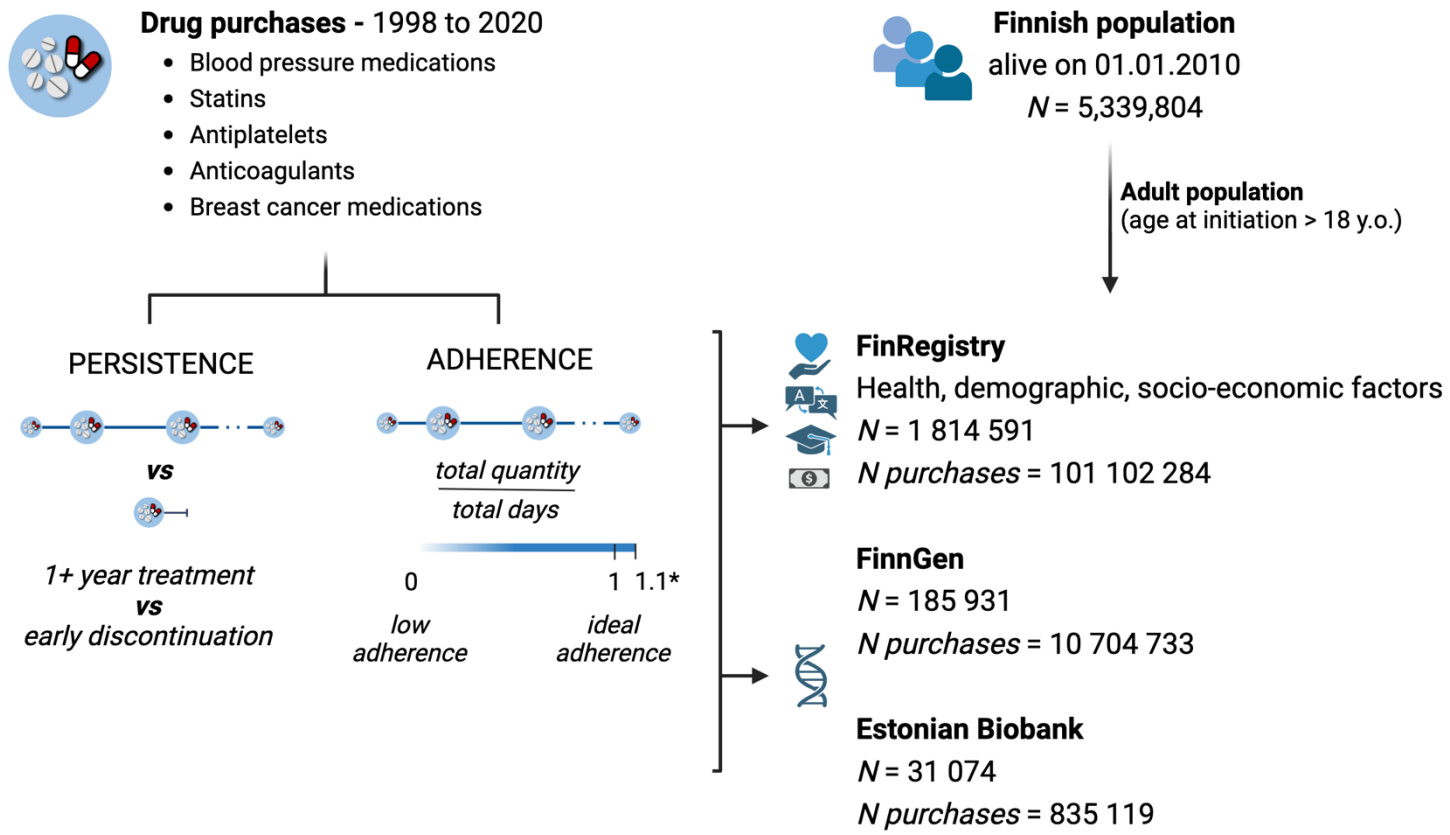


Figure 1 | Schematic overview of the study. We included all individuals resident in Finland and alive on 1.1.2010 (N=5 339 804) and used information from the Finnish drug purchase register to identify individual with at least one purchase for the following medication classes: blood pressure medications, statins, antiplatelets, direct oral anticoagulants and breast cancer medications. We considered only the adult population, starting the treatment after age 18, and reconstructed individual medication purchase trajectories as a proxy to define the medication usage phenotypes: persistence and adherence. Persistence was defined as continuing the treatment for at least one year versus discontinuing it after the first purchase. We defined adherence as the Medication Possession Ratio (ratio of the quantity purchased in the observation period to the days in the observation period), for patients continuing the treatment for at least one year. Adherence is a continuous trait ranging from 0 (complete lack of adherence) to 1 (ideal adherence). *We excluded patients with adherence above 1.1, to exclude excessive stockpiling or overbuying behaviors, but still allowing for some variability. We then studied the association of these two medication-use phenotypes with health, demographic and socio-economic factors from nationwide registries (N = 1 814 591 with complete information) and with genetic information for a subset of the population from the FinnGen study (N = 185 931) and from Estonian Biobank (N = 31 074). Figure created with Biorender.com.

78

79

Results

Adherence and persistence for 5 medications in 1 814 591 individuals in the general population and 217 005 individuals with genetic data

We considered five classes of medications: statins, blood pressure (BP) medications, antiplatelets, breast cancer medications, and direct oral anticoagulants (DOAC) (ATC codes included are described in **Table 1** and **Methods**). For breast cancer medications and DOACs, we included individuals with, respectively, a breast cancer or atrial fibrillation, pulmonary embolism, or venous thromboembolism diagnosis before treatment initiation (**Methods**). For antiplatelets, we did not analyze persistence, as this class of medications is frequently prescribed for temporary treatments of 6 to 12 months. Information about the date, quantity, and type of reimbursable medications that were purchased in pharmacies were available between 1998 and 2020 and were used to calculate adherence and persistence.

Our initial study population included every individual who was a resident in Finland on 1/1/2010 (N=5 339 804). 2 289 455 individuals (43% of the whole population) purchased at least one of the 5 medications after 18 years of age, with a total of 162 231 752 purchases.
Most individuals purchased the 5 medications for at least 12 months (ranging from 89.7% for breast cancer medications to 64.7% for anticoagulants, except for antiplatelets (44.2%), **Supplementary Table 1**) and were eligible for the adherence analysis (N = 1 876 212). We further excluded 141 452 individuals who had a medication possession ratio (MPR) > 1.1, to account for stockpiling of medications or overbuying behaviors (**Methods**), and 1 270 individuals with incomplete registry information for the 8 health and socioeconomic factors considered in the main analysis (**Supplementary Table 1**). In the persistence analysis, we considered people discontinuing the treatment after one purchase and compared them to those continuing for at least 12 months and were eligible for the adherence analysis. Our final study population thus included 1 814 591 individuals, for a total of 101 102 284 purchases.

BP medications were the most purchased (N of individuals = 1 316 889) and had a mean treatment length of over 10 years (**Table 1**). Breast cancer medications were purchased by the fewest individuals (N=55 270) and most individuals stopped after 5 years.
Persistence was above 90% for all drugs (**Figure 2**). Adherence was 0.94 averaging across all medications, highest in breast cancer medications (0.97) and lowest in statins (0.89; **Table 1**). Adherence values were similar to those reported in the literature in comparable studies or meta-

analyses [12,23] with all medication classes having at least 80% of individuals considered as “good adherers” (**Figure 2**). **Extended Figure 1** provides an overview of adherence values by showing the relationship between expected and observed consumption across the 5 medication classes.

For the subset of 185 931 individuals participating in FinnGen, persistence rates and adherence distributions were consistent with the general population. In the Estonian Biobank, a cohort with similar drug purchase data and genetic information available and 31 074 eligible medication users, we observed generally lower persistence and adherence to statins, clopidogrel, and tamoxifen (**Table 1**).

Medication class	ATC codes	Cohort	N individuals	Total purchases	Proportion of persistent individuals	Mean adherence (SD)	Mean treatment length (years)
Blood Pressure medications	C02*, C03*, C08*, C09*	FinRegistry	1 316 889	61 213 707	0.947	0.93 (0.123)	10.755
		FinnGen	134 615	6 399 494	0.937	0.932 (0.121)	11.172
Statins	C10AA*	FinRegistry	1 061 926	33 864 971	0.952	0.892 (0.123)	9.168
		FinnGen	116 439	3 828 359	0.956	0.901 (0.118)	9.454
		Estonian Biobank	27 339	792 839	0.866	0.648 (0.275)	8.815
Antiplatelets	B01AC04 B01AC30	FinRegistry	138 007	3 120 237	-	0.931 (0.114)	5.528
		FinnGen	12 618	285 902	-	0.933 (0.112)	5.572
Clopidogrel	B01AC04	Estonian Biobank	2712	33 466	-	0.663 (0.325)	5.455
Anticoagulants	B01AF01, B01AF02, B01AF03, B01AF07	FinRegistry	105 113	2 003 698	0.975	0.957 (0.094)	3.127
		FinnGen	2792	45 394	0.889	0.952 (0.101)	2.984
Breast Cancer medications	L02BA01, L02BG04, L02BG06, L02BG03	FinRegistry	55 270	899 671	0.984	0.969 (0.067)	4.556
		FinnGen	9857	145 584	0.985	0.971 (0.067)	4.128
Tamoxifen	L02BA01	Estonian Biobank	1023	8814	0.58	0.829 (0.267)	3.973

Table 1 | Descriptive statistics for the 5 medication classes included in the analysis. For each medication class, we report the ATC codes used to define the medication purchases of interest and other descriptive statistics in the general Finnish population (FinRegistry), in the subset of Finnish population with genetic data (FinnGen) and in the Estonian Biobank. Number of individuals is the total number of individuals included in the analysis, meaning the number of individuals considered for both the persistence and adherence analysis. Total purchases is the total number of purchases per each medication and each cohort considered in both analyses. Mean adherence (and its standard deviation) and the mean treatment length are reported for the individuals with at least one year long treatment (individuals considered in the adherence analysis).

146

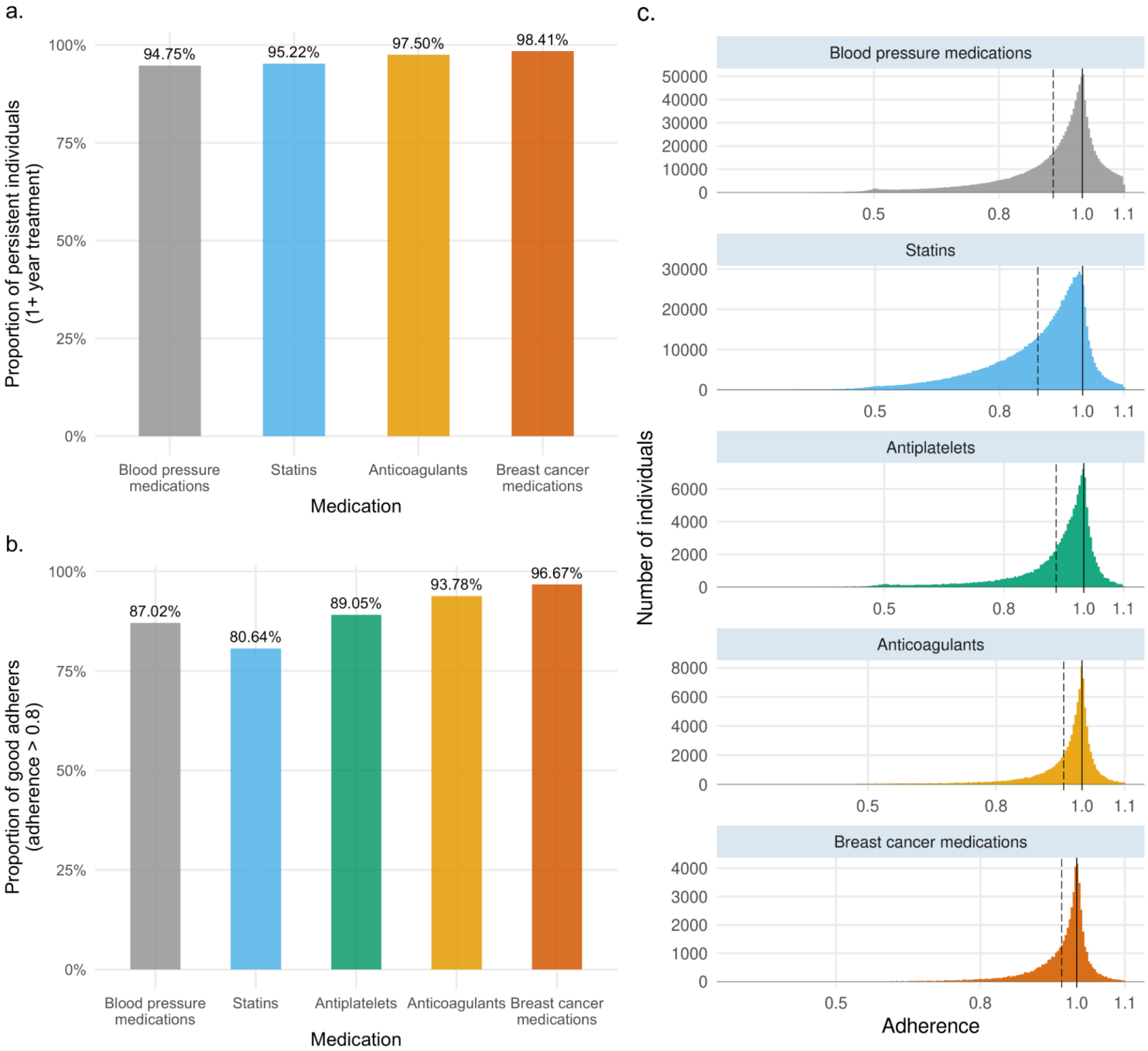


Figure 2 | Persistence and adherence to 5 medications in the general Finnish population. a., Proportion of persistent individuals for each medication class considered in the analysis. **b.**, Proportion of "good adherers" (individuals with adherence greater than 0.8) for each medication class. **c.,** Adherence distribution and mean adherence (dotted line) for the 5 medication classes considered in the analysis.

147

148 **Health and socio-demographic risk factors for persistence and adherence**

149 We studied the effects of the following risk factors, measured at the time of first purchase, on
150 persistence and adherence: age, sex, primary *vs* secondary prevention, Charlson comorbidity
151 index, years spent in education, living area, mother tongue (a proxy for first-generation
152 immigration status), and receiving any social assistance benefit in the year before starting the
153 treatment. We chose these 8 factors because they broadly capture overall health and socio-
154 demographic aspects, were previously examined in the literature [4], and were available from
155 nationwide registers. Basic descriptive statistics for each factor and each medication are
156 reported in **Supplementary Table 3**. Some socio-demographic factors were consistently
157 associated with lower persistence and adherence across all medications in a multivariable
158 analysis (**Figure 3** considers dichotomized factors while **Supplementary Tables 6 and 7** report
159 similar analysis when continuous factors are treated as such). For example, not speaking
160 Finnish or Swedish and having received social assistance benefits were associated with an
161 average of 3.2% [1.9%-4.3%] and 2% [1.1%-3.1%] lower adherence across medications,
162 respectively. Starting treatment after a related disease event (secondary prevention) was also
163 consistently associated with higher odds of persistence for statins (odds ratio (OR) = 1.53, 95%
164 confidence interval (CI) = 1.49-1.58, $P = 4.9 \times 10^{-198}$) and increased adherence to statins and
165 antiplatelets (1.7% increase, 95% CI = 1.64-1.76% $P < 2 \times 10^{-308}$ for statins; 0.47% increase, 95%
166 CI = 0.35-0.59%, $P = 7.4 \times 10^{-15}$ for antiplatelets). Age and sex showed discordant effects across
167 different medications after adjustment for the remaining factors. Being female showed a
168 consistent negative effect on persistence for all drugs but showed heterogeneous effects on
169 adherence. Age also showed heterogeneous effects across different medications
170 (**Supplementary Tables 4-7**).
171
172 Extending beyond these baseline factors, we further assessed the effect of one or more
173 concurrent treatments (i.e. polytherapy). We observed a consistent positive association between
174 having at least one or more treatment and higher adherence and increased odds of persistence
175 (**Supplementary Note**).
176
177 Overall, the considered health and socio-demographic factors showed a higher effect on
178 persistence than on adherence as indicated by the overall higher variance explained
179 (**Supplementary Table 8**).
180

181 In the Estonian Biobank, when examining variables that are present in both studies, we found
182 consistent results in the direction of effect and statistical significance for both persistence and
183 adherence to statins and age, sex, and years spent in education, but not for secondary
184 prevention (**Supplementary Table 9**).
185

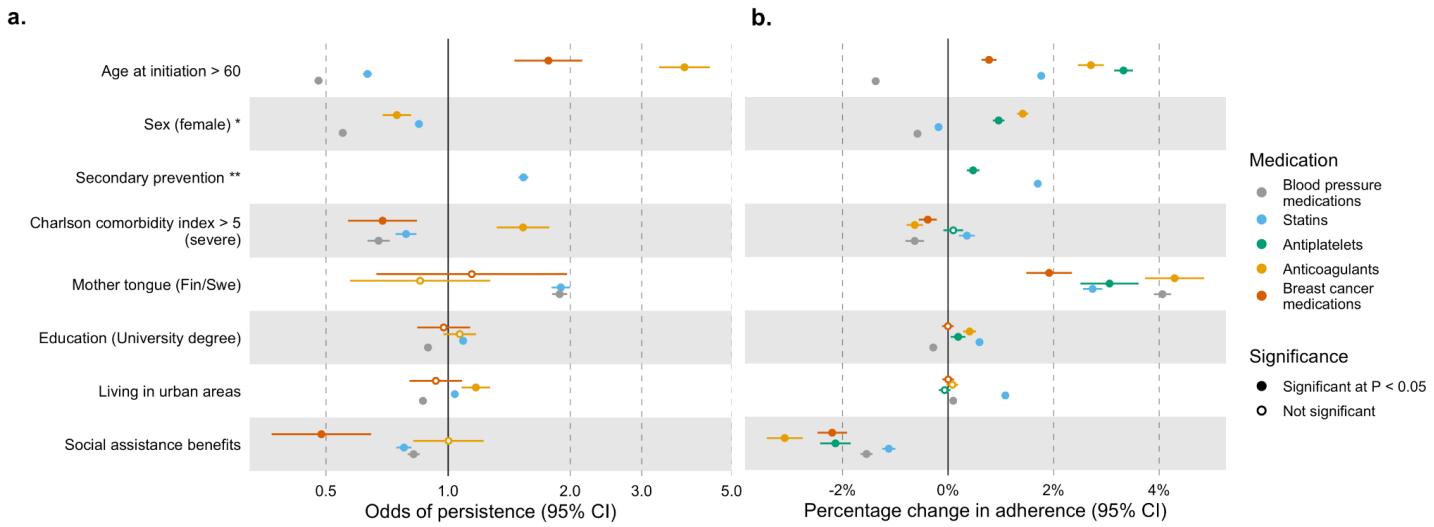


Figure 3 | Effect of 8 health, demographic and socioeconomic factors on persistence and adherence in the Finnish population. a., Odds of persistence given each factor presence. ORs are from a multivariable logistic regression model, error bars represent the 95% confidence interval for the estimates. **b.**, Coefficients are estimated from a multivariable linear regression model and were rescaled to represent the percentage change in adherence given each factor presence. Error bars represent the 95% confidence interval for the estimate. Descriptive statistics for each risk factor are reported in Supplementary Table 2. In this figure we binarized the continuous risk factors to make effect sizes comparable, however results for non-binarized risk factors are available in Supplementary Table 5-6.
*****sex not included for breast cancer medications as all individuals included are females
******secondary prevention defined only for statins and antiplatelets (**Methods**)

186

187 **Established pharmacogenes marginally associate with persistence but not adherence**

188 We tested the effects of 5 relevant pharmacogenes on persistence and adherence, in both
189 FinnGen and the Estonian Biobank. We selected the following pharmacogenes based on
190 guidelines from the Clinical Pharmacogenetics Implementation Consortium[24–27], filtering for the
191 drug/gene pairs which were reported with recommended prescribing actions (e.g. alternative
192 therapies or dosing): *ABCG2* for rosuvastatin, *CYP2C9* for fluvastatin, *SLCO1B1* for all statins,
193 *CYP2C19* for clopidogrel, and *CYP2D6* for tamoxifen. For each gene, we defined individual
194 diplotypes and grouped them into phenotypes defined based on each diplotype activity score.
195 We then examined the effect of each phenotype on persistence and adherence. The *CYP2C9*

intermediate metabolizer phenotype was nominally significantly associated with lower odds of persistence to fluvastatin in both FinnGen and the Estonian Biobank, and when meta-analyzing the results from both studies (OR $_{meta\text{-}analysis}$ = 0.6, 95% CI $_{meta\text{-}analysis}$ = 0.43-0.84, P $_{meta\text{-}analysis}$ = 0.003). Similarly, *SLCO1B1* decreased and poor function were associated with lower odds of persistence to all statins (OR $_{meta\text{-}analysis}$ = 0.92, 95% CI $_{meta\text{-}analysis}$ = 0.86-0-98, P $_{meta\text{-}analysis}$ = 0.009; OR $_{meta\text{-}analysis}$ = 0.94, 95% CI $_{meta\text{-}analysis}$ = 0.88-0.99, P $_{meta\text{-}analysis}$ = 0.04, respectively). Decreased *SLCO1B1* function was also significantly associated with lower adherence, although with a neglectable effect (-0.27% change in adherence, P = 0.033). None of the associations were significant after multiple testing correction (P < 0.002, Bonferroni correction for 19 phenotypes tested).

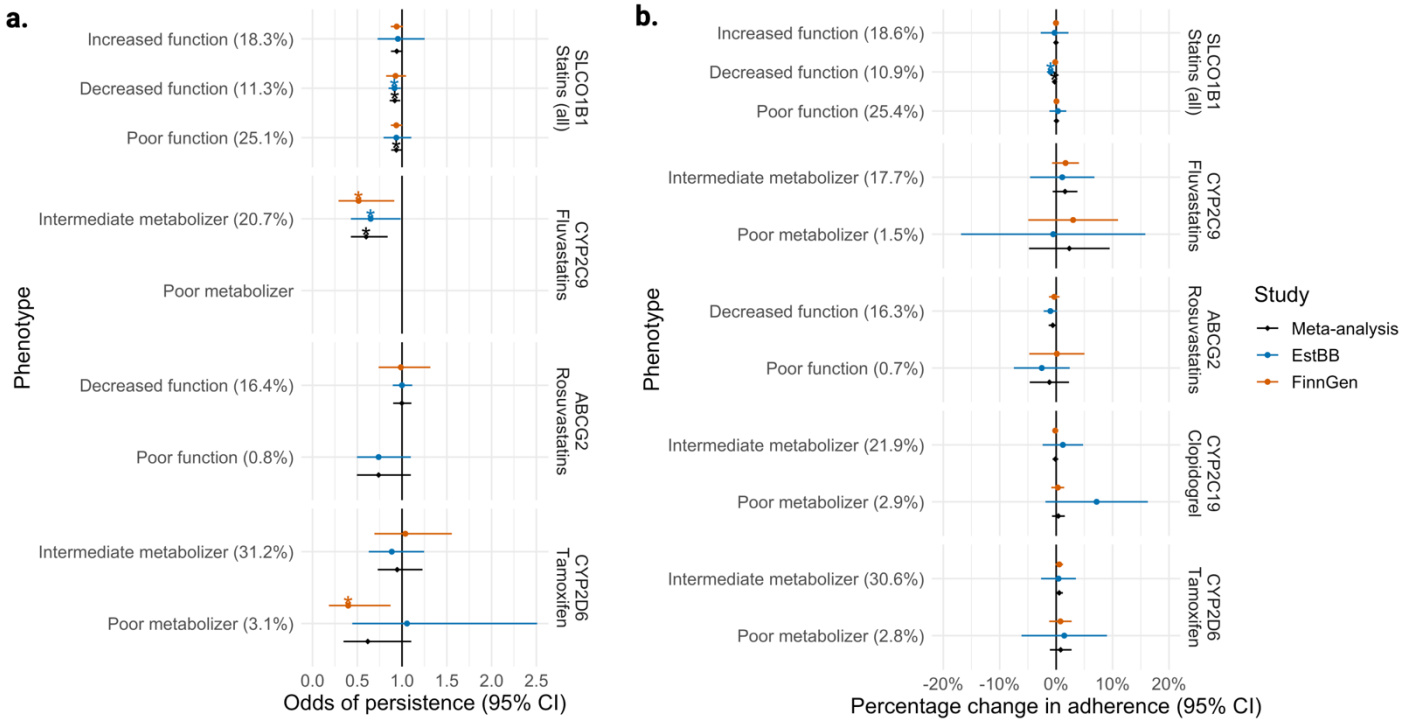


Figure 4 | Forest plot for associations between clinically relevant pharmacogene phenotypes and persistence and adherence in FinnGen and the Estonian Biobank. **a.** Odds of persistence given each phenotype level. ORs are estimated from a logistic regression model, separately for each gene/drug pair. **b**. Percentage change in adherence given each phenotype level. Coefficients are estimated from a linear regression model, separately for each gene/drug pair. For both panels, error bars represent the 95% confidence interval for the estimate. Stars represent nominal (P < 0.05) statistical significance. Estimates are not reported for phenotypes with N < 5. For each phenotype, the frequency within each drug user group is reported in parenthesis.

206 **Genome-wide association study of drug adherence and persistence**

207 We conducted a GWAS of persistence and adherence to each medication in FinnGen, adjusting
208 for age at initiation, year of birth, sex, genotyping batch, and first 10 genetic principal
209 components (**Extended Figure 2**). No variant showed significant association with either
210 phenotype at a significance level of $P < 1x10^{-8}$ (standard genome-wide significance corrected for
211 multiple testing across the 5 medications). Four SNPs, however, were genome-wide significant
212 at the standard $P < 5 \times 10^{-8}$ threshold (**Supplementary Table 12**). Notably, *rs1339882991* was
213 positively correlated with both persistence and adherence to BP medications and was
214 previously reported to be positively associated with hypertension [28]. Further characterization of
215 this and the other genome-wide significant variants is provided in the **Supplementary Note**.
216 Considering the sample size observed for statins (N = 116 439), we have at least 80% power to
217 identify genome-wide significant variants with an effect of greater than 2.7% on adherence,
218 assuming a minor allele frequency of 0.3. Such an effect is in line with what is observed for the
219 strongest epidemiological risk factor for statin adherence (**Figure 3**). As we expect individual
220 genetic effects for common variants to be smaller than socio-environmental, our results indicate
221 that at the current sample size, the power to identify genetic associations for adherence and
222 persistence is limited.

223 **Polygenic overlap between persistence, adherence and other health and behavioral traits**

224 We explored the genetic overlap between persistence and adherence and 33 major diseases,
225 disease risk factors, and behavioral traits (**Figure 5, Supplementary Tables 13,14**). Higher
226 systolic blood pressure was genetically correlated with higher persistence to BP medications (*rg*
227 = 0.77, $P = 3.1x10^{-23}$) and with higher adherence to both statins and BP medications (*rg* = 0.17,
228 $P = 3.4x10^{-06}$ and *rg* = 0.4, $P= 3.5x10^{-30}$, respectively). On the contrary, higher LDL cholesterol
229 was genetically correlated with lower statins adherence (*rg* = -0.14, P = 0.006), although when
230 considering a GWAS of LDL cholesterol adjusted by statins usage the effect was not statistically
231 significant anymore. Notably, both adherence to statins and BP medications were positively
232 genetically correlated with participating in one or more of the optional components of UK
233 Biobank [30] (*rg* between 0.11 and 0.31, $P < 8.2x10^{-3}$), indicating some shared effect between
234 drug adherence and continued participation in scientific studies.
235
236 We wondered if existing polygenic scores (PGS) could be used to predict adherence and
237 persistence and how their effect compared to that of the health and socio-demographic risk

238 factors examined before. Among the 33 PGS examined (**Extended Figure 3,4** and
239 **Supplementary Table 15,16**) the strongest effects were observed for systolic blood pressure
240 (SBP) PGS on BP medications persistence (OR = 1.64 for 1 standard deviation (SD) in PGS,
241 95% CI = 1.60-1.68, $P < 2x10^{-308}$) and adherence (0.8% increase per 1-SD change in PGS,
242 95% CI = 0.74-0.86%, $P = 2.6x10^{-126}$); LDL cholesterol (adjusted for statins usage) PGS on
243 statins persistence (OR = 1.24, 95% CI = 1.20-1.28, $P = 6.2x10^{-47}$) and SBP PGS on statins
244 adherence (0.28% increase, 95% CI = 0.22-0.34%, $P = 2x10^{-16}$); coronary artery disease PGS
245 on antiplatelets adherence (0.47% increase, 95% CI = 0.29-0.67%, $P = 4.3x10^{-7}$) and PGS for
246 schizophrenia and adherence to breast cancer medications (0.22% decrease, 95% CI = 0.1%-
247 0.34%, $P = 4.2x10^{-4}$). No significant association was observed between any of the PGS and
248 adherence or persistence to DOAC.
249
250 In the Estonian Biobank, systolic blood pressure and LDL PGS were also among the strongest
251 effects associated with increased statin adherence, and persistence, respectively (percentage
252 increase in adherence = 0.66%, $P = 1.11 \times 10^{-3}$); (OR = 1.13, 95%CI=1.093- 1.167, $P = 4.65 \times$
253 10^{-9})) **(Supplementary Tables 19-20)**.
254
255 Overall, these PGSs could explain a lower fraction of the variance in adherence than the
256 combined effect of the health and socio-demographic risk factors examined above, after
257 accounting for age and sex (**Supplementary Table 17**).
258
259 Finally, we wondered whether developing a PGS for adherence would hold clinically meaningful
260 predictive power. We used the GWAS results from FinnGen to build a PGS for statin adherence
261 and tested its association with observed adherence in the Estonian Biobank. We found the PGS
262 to be significantly associated with observed adherence (percentage change in adherence for 1
263 SD change in PGS = 0.72%, $P = 1 \times 10^{-5}$), however, the small effect size observed suggests
264 such PGS to have low predictive value and clinical relevance.
265

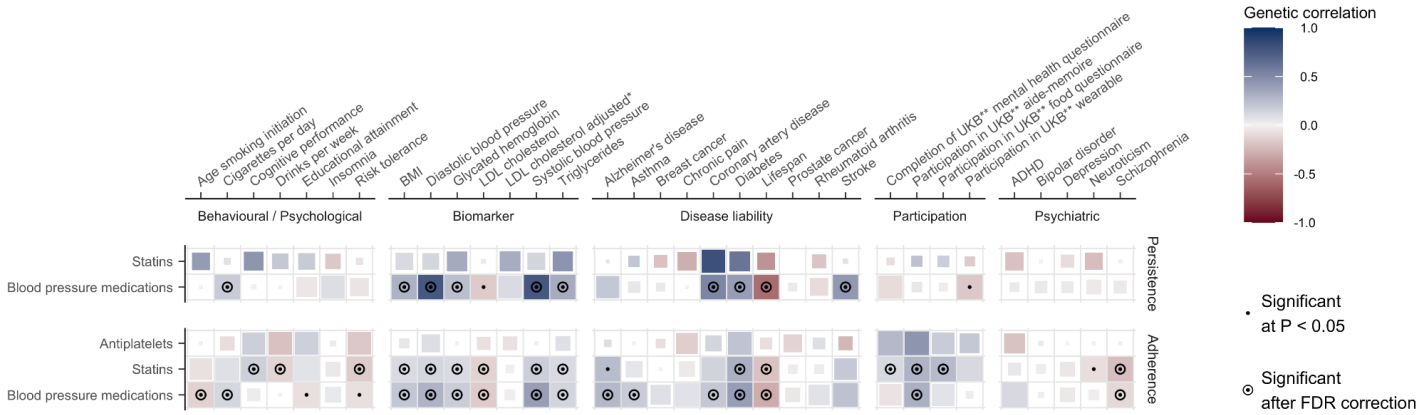


Figure 5 | Genetic correlates of persistence and adherence. Genetic correlations between 33 clinically relevant traits and persistence and adherence. Medications for which genetic correlations could not be computed are not displayed. Blue represents positive correlations, red negative. The size of each tile is proportional to the statistical significance (the bigger the tile, the lower the P-value). Correlations nominally significant at P < 0.05 are marked with a dot, while correlations which were significant after FDR correction (within each medication) are additionally marked with a circle.
***** adjusted by the use of cholesterol lowering medication
****** UKB = UK Biobank

Sensitivity analysis on persistence definition

Our definition of persistence resulted in higher proportions of persistent individuals compared to what observed in literature [31]. Thus, as sensitivity analysis, we consider a different definition of persistence, looking at the proportion of individuals that stop medication within one year among those with at least one purchase of the medication. This definition includes a broader population (i.e. does not exclude overbuying behaviors) and consider as not persistence those individuals that stop after two and three purchases, as opposed to only one. With this definition, persistence for statins was 75.6%, consistent with the literature [31]. When looking at the effect of PGS on persistence, we observed directionally consistent effects, but an overall reduction in effect sizes despite the increased number of statistically significant associations (**Supplementary Table 18**, **Supplementary Figure 1**).

Discussion

Drug adherence remains a major obstacle to the benefits of pharmacotherapies. Although many external factors (e.g. healthcare- and system-related factors) contribute to drug adherence [4,5], focusing on the patient-specific determinants might help in identifying patients at risk of low adherence. In this study we provide a comprehensive exploration of patient-specific factors affecting drug persistence and adherence across 5 different classes of medications, using nationwide high quality registry data and, for the first time, also exploring genetic factors.

Overall, our results suggest that the need for social assistance and being first-generation immigrants have a moderate to large effect on drug adherence and persistence which is consistent across medication classes. In comparison, demographic characteristics (age and sex), comorbidity burden, and genetic factors have modest or heterogeneous effects. The only health factor with a robust effect was the presence of a disease related to the medication, which substantially increased adherence among those using drugs for secondary prevention.
In line with previous reports [23], the effects of age and sex were heterogeneous across different medications. Nevertheless, females had consistently lower odds of persistence across all medications. We speculate that a higher rate of adverse drug reactions in women compared to men [32,33], or lower perceived risk for cardiometabolic diseases among women can explain these findings.

We performed three main genetic analyses. First, we evaluated if established pharmacogenes for statins, clopidogrel and tamoxifen were associated with persistence and adherence of these drugs. Our results indicate that poorer statins metabolism is associated with lower odds of persistence, with a notable effect for *CYP2C9* and fluvastatins. Since increased systemic concentration of statins is associated with higher risk of developing statin-associated muscle symptoms[25], these findings are in line with the hypothesis of patients discontinuing the treatment as they develop muscle symptoms. However, as these results were consistent in both FinnGen and the Estonia Biobank, but were not significant after multiple-testing correction, larger studies should specifically test this hypothesis. Because pharmacogenomics testing is not widespread in Finland, we do not believe these results are biased by genetically-informed drug prescription. Overall, the lack of association between most pharmacogenes and adherence suggests that widespread pharmacogenetic testing before prescribing the medications included in this study is not going to substantially improve adherence in the population, while it could be

314 valuable to tailor the prescribing dose of statins to avoid the risk of developing statin-associated
315 muscle symptoms and improve persistence.
316
317 Second, we performed a GWAS of drug adherence and persistence and used these results to
318 understand the overlap with other traits. We did not identify robust individual genetic signals but
319 the identification of four genome-wide significant variants, two of which were not previously
320 reported to be associated with traits underlying the need for medication, warrants for larger
321 genetic studies of adherence and persistence. Our power calculations indicate that even if this
322 is the largest study of this kind, power is still limited to identify genome-wide significant
323 associations of small effects. For example, at the current sample size, the variants in the
324 pharmacogenes tested, which showed marginal significance in the targeted analysis, would not
325 have been detected at a stringent genome-wide threshold.
326 A previous study using FinnGen data compared cardiometabolic-related medication user vs
327 non-user or changes in medication use and identified > 40 loci reflecting underlying
328 cardiometabolic risk [34]. The differences in the magnitude of genetic signals and genetic
329 correlation profiles in the current study compared to our previous study indicate that drug
330 adherence and persistence have lower genetic contribution and a stronger behavioral
331 component, which does not necessarily reflect the underlying disease risk. In our study,
332 interestingly, we observed genome-wide overlaps with a proxy for study participation indicating
333 some shared effect between adherence to drug therapy and participation in scientific research,
334 similarly to what we recently observed for COVID-19 vaccination uptake [35].
335
336 Third, we explored if existing PGS could predict adherence and persistence and found the
337 strongest association between systolic blood pressure PGS and adherence/persistence to BP
338 medications. This association probably reflects individuals with higher BP being more adherent
339 due to higher perceived risk [36]. Finally, while our results could be used to generate a PGS for
340 statin adherence, which significantly replicated in the Estonian Biobank, the variance explained
341 by this PGS was low.
342
343 A strength of this study is the use of Finnish registries covering all drug purchases since 1998
344 which allows an in-depth longitudinal study of drug purchasing. For example, the average
345 treatment length for BP medications was over 10 years. Moreover, by using nationwide data, we
346 could confirm that adherence and persistence measures in FinnGen were similar to what was
347 observed in the general population. Importantly, this study considers both socio-economic and

genetic risk factors, helping to better contextualize the results and bridging the divide between social science and human genetics, as advocated by others [37]. Furthermore, we derived adherence and persistence measures from 5 different medication classes, which allowed us to capture adherence-related factors beyond single disease- and drug-related factors. Finally, both epidemiological and genetic results were replicated in an external study.

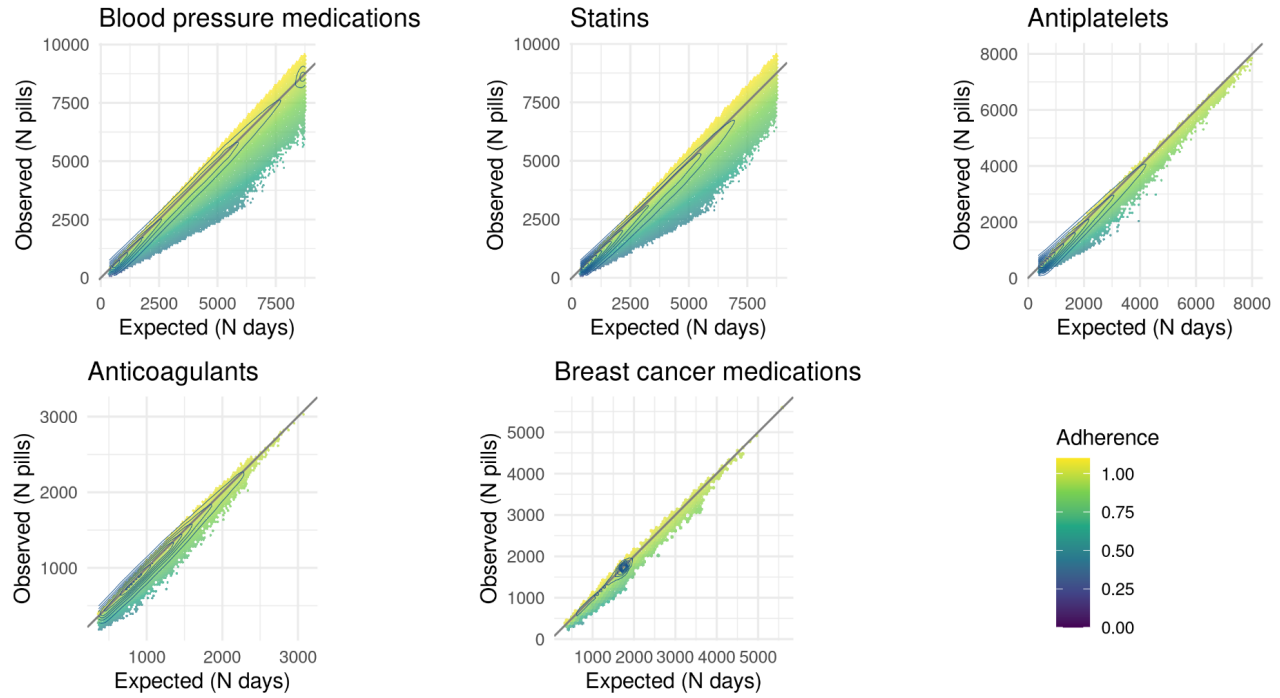
A limitation of the study is that our definitions of persistence and adherence can be biased in different ways. First, our persistence definition resulted in higher proportions of persistent individuals compared to what observed in literature[31]. Nonetheless, we chose this definition to better capture early discontinuation behaviors (i.e. after the first purchase). Moreover, we deemed it fair to compare these individuals only to those included in the adherence analysis, so as not to include overbuying behaviors. Sensitivity analyses suggest that our results are robust to a different definition of persistence and that our stricter inclusion criteria resulted in larger effect sizes at the expenses of decreased statistical power.

As for the adherence definition, first we cannot know if individuals are really taking medications and instead rely on the expected and observed consumption. Second, we don't have information on the dose prescribed by the doctor for the entire period and rely on the suggested dosage of the package. For example, individuals which are suggested by the doctor to take half-tablet of statins/day despite package indication, will result in low adherence. Two results indicate that this behavior is not common. First, we do not see enrichment in adherence values that would capture a systematic indication of non-standard dose of the medication. Second, since 2015 prescription information is available as text. Using regular expressions to extract this information (**Methods**) we found that in 2019, 96% of the prescriptions were indicating 1 statin tablet/day. Finally, the observed adherence is higher than what was reported from a previous Finnish study [38]. Considering only individuals that had taken medication for at least 12 months and excluding long inactive periods might have contributed to higher adherence values. Nonetheless, early discontinuation is captured by our persistence measure, while long breaks in drug usage might reflect hospital admission, partial emigration, or other external factors. Overall, we used a definition of adherence that is more robust to large external shocks and better suited to capture smaller and consistent variations in drug purchases over longer time periods.

In conclusion, our results suggest a limited role of genetic risk factors, including clinically relevant pharmacogenes, in explaining persistence and adherence for 5 common medication classes that require long-term, regular therapy. While larger GWASs are needed to identify smaller genetic effects, the observed genetic overlap with behavioral traits, might complicate the interpretation of the results. The robust associations with measures of social deprivation suggest that we need better understanding of the drivers of poor adherence in these socio-economically disadvantaged groups. Identifying these drivers might offer points of intervention in clinical practice, or help recognizing system-wide barriers for drug adherence.

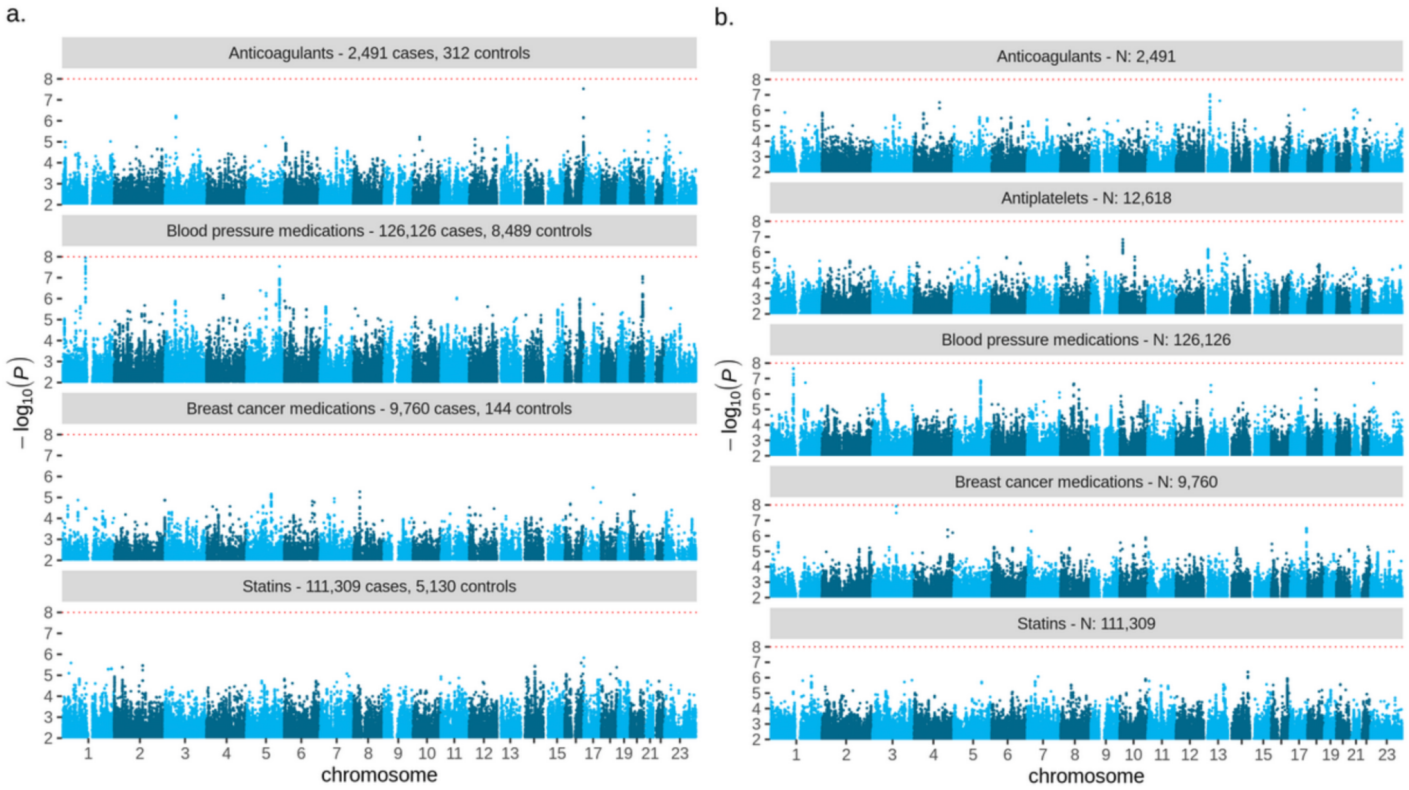
Extended Figures

Extended Figure 1



Extended Figure 1 | Adherence as function of expected and observed consumption. Hexagonal heatmap representing mean adherence levels at the varying of expected (number of days, x-axis) versus observed consumption (number of tablets, y-axis). Bins with less than 5 individuals are not displayed. The level lines represent the 2-dimensional density, each line including 10% of the individuals in the data.

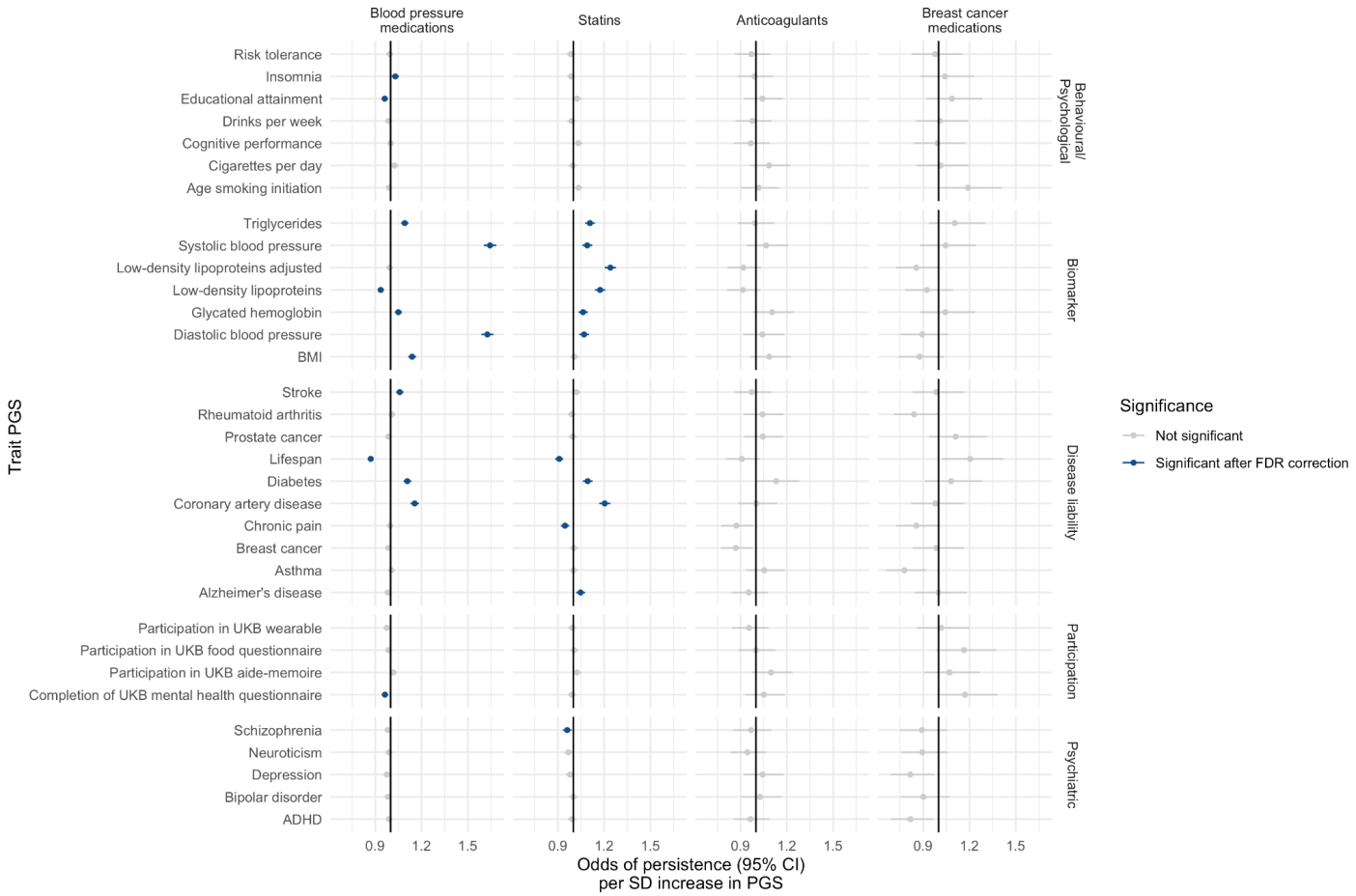
403 **Extended Figure 2**



Extended Figure 2 | Genome-wide association results for persistence and adherence. a. results of genome-wide association study of persistence to each medication. Number of cases and controls are reported next to each drug name. Red dotted line represents genome-wide significance at P-value = 1x10-8 (Bonferroni corrected for 5 medications). b. results of genome-wide association study of adherence to each medication. Total sample size is reported next to each drug name. Red dotted line represents genome-wide significance at P-value = 1x10-8 (Bonferroni corrected for 5 medications). Results for 5 variants with significance < 5x10-8 are reported in Supplementary Table 10.

404
405
406
407
408
409
410
411
412
413
414
415

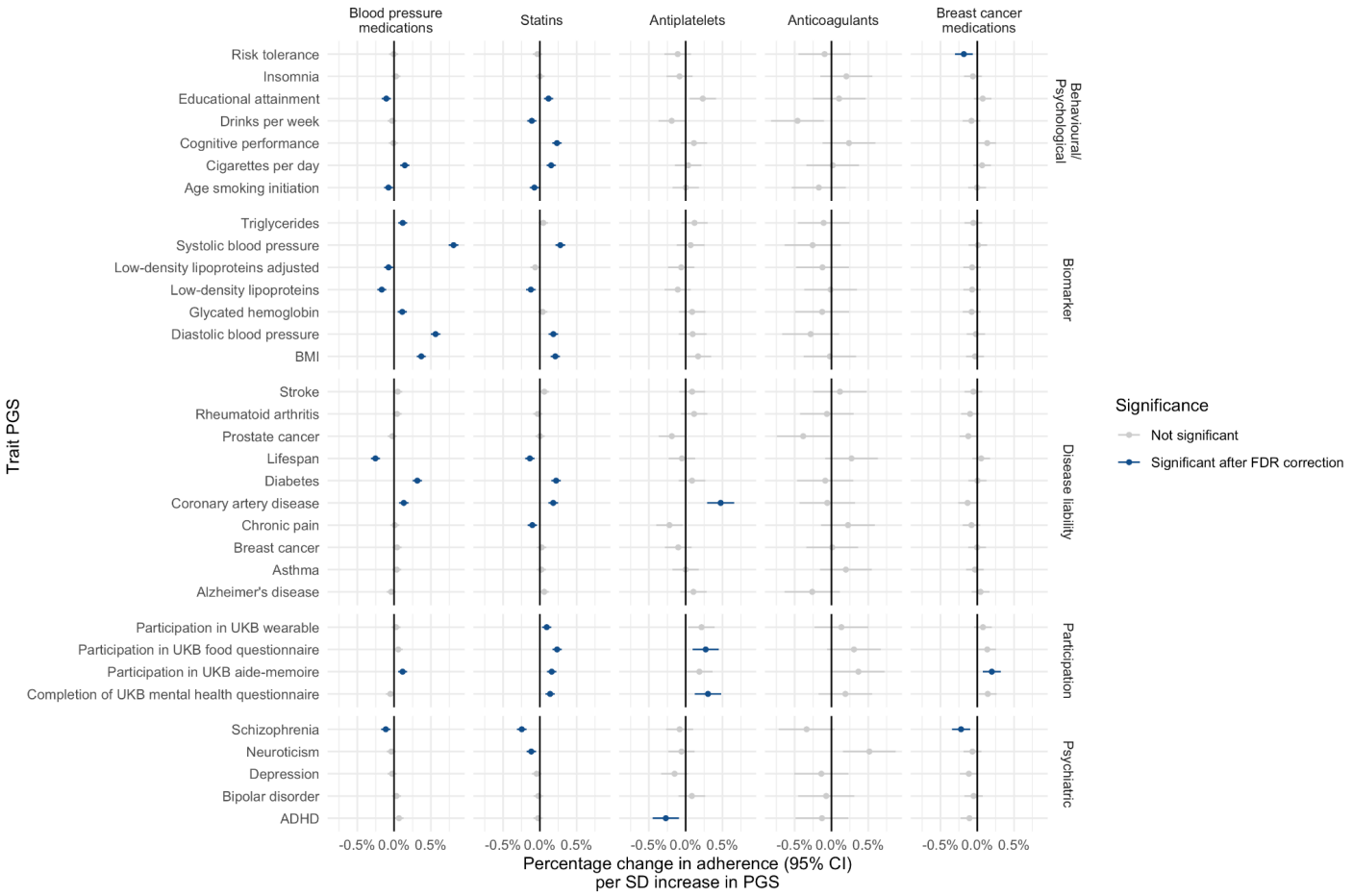
416 **Extended Figure 3**



Extended Figure 3 | Associations between persistence and PGS for 33 clinically relevant traits. Odds of persistence per 1-SD increase in trait PGS. ORs are from a logistic regression model adjusted for sex, age at initiation, first 10 genetic principal components. Error bars represent the 95% confidence interval for the estimates. Results are reported in Supplementary Table 13.

417
418
419
420
421
422
423
424
425
426
427
428

429 **Extended Figure 4**



Extended Figure 4 | Associations between adherence and PGS for 33 clinically relevant traits. Percentage change in adherence per 1-SD increase in trait PGS. Coefficients are from a linear regression model adjusted for sex, age at initiation, first 10 genetic principal components. Error bars represent the 95% confidence interval for the estimates. Results are reported in Supplementary Table 14.

430

431 **Methods**

432 **FinRegisty and FinnGen study population.**

433 The FinRegistry dataset [20] contains data from Finnish nationwide health, demographic, and
434 socioeconomic registries, for a total of 7 166 416 individuals. To ensure completeness of
435 information for the variables considered in the epidemiological analysis, we included only
436 individuals who were residents in Finland and alive on 1 January 2010 (5 339 804 individuals in
437 total). The FinnGen dataset [21] **(Supplementary Methods)** additionally contains genetic data for
438 a subset of the Finnish population (N = 430 885 for data freeze 10). We studied the two drug
439 usage phenotypes in adults only, meaning individuals needed to be at least 18 years old at
440 treatment initiation.

441 **Kela drug purchase register.**

442 The Social Insurance Institution of Finland (Kela) is a government agency that provides basic
443 economic security by financial support for everyone living in Finland. The Kela drug purchase
444 register contains information about any reimbursable prescription medication bought from
445 pharmacies in Finland between 1995 and 2020. We included only purchases from 1998 on to
446 ensure better homogeneity with the data from the other registries. Each data entry contains
447 information about the date of purchase, ATC code of the medication, number of purchased
448 packages, and the Nordic Article Number (VNR) associated with the package, which gives
449 information about the quantity and dosage of the medication. When the information was
450 available, we additionally excluded those purchases made using pharmacies dose distribution
451 services, which regularly provides patients with 2-week worth of medications at a time.
452

453 **The Estonian Biobank study population.**

454 The Estonian Biobank (EstBB) (**Supplementary Methods**) is a volunteer-based biobank that
455 has a sample size of approximately 211,000 participants, comprising over 20% of the adult
456 population of Estonia[44]. The EstBB is linked with the National Health Insurance Fund's (NHIF)
457 database which was used as the EHR source. NHIF covers over 95% of the Estonian
458 population as a result of universal health care in Estonia and thus provides information on
459 diagnoses, prescribed and dispensed drugs for up to 19 years (2004-2023) from primary and
460 specialist care. Information on prescribed and dispensed drugs is collected from all pharmacies

in Estonia. The NHIF database contains information on the prescribed and dispensed drug name, date of prescription and dispensing, International Classification of Diseases (ICD-10) diagnosis codes, Anatomical-Therapeutic-Chemical (ATC) code of the drug, number of packs dispensed, number of tablets in each package, and concentration of active ingredient in each tablet. The study population in the Estonian Biobank cohort consisted of individuals that initiated their treatment between 2007-2023.

Medications included.

We included medications commonly prescribed and purchased, that are typically used for a prolonged time and do not have a significant direct effect on wellbeing. The classes of medication we included are as follows: statins (C10AA* - all types of statins without differentiation), blood pressure medications (C02* - antihypertensives, C03* - diuretics, C08* - calcium channels blockers, C09* - agents acting on the renin-angiotensin system), antiplatelets (B01AC04 - clopidogrel and B01AC30 - acetylsalicylic acid in combination with dipyridamole), breast cancer medications (L02BA01 - tamoxifen, L02BG03 - anastrozole, L02BG04 - letrozole, L02BG06 - exemestane), direct oral anticoagulants (DOAC) (B01AF01 - rivaroxaban, B01AF02 - apixaban, B01AF03 - edoxaban, B01AE07 - dabigatran etexilate). For breast cancer medications we included only females, starting the treatment after a breast cancer diagnosis. For DOAC, only patients starting the treatment after an atrial fibrillation diagnosis were included.

The following drug classes or drugs were considered in the Estonian Biobank: statins drug class (ATC C10AA*), and separately clopidogrel (ATC B01AC04), tamoxifen (L02BA01), fluvastatin (ATC C10AA04), and rosuvastatin (ATC C10AA07). The patients were included if they were over the age of 18 at the start of treatment initiation.

Persistence.

We define persistence [18] as a binary phenotype, comparing individuals taking the medication for at least 12 months (i.e. all individuals included in the adherence analysis) versus patients early-discontinuing the drug after one purchase. To avoid right censoring, we considered as early discontinuation only purchases made at least 2 years before the end of follow-up. End of follow-up date was either death date, moving abroad date or the last date for which the data are available (01.01.2020).

492

493 In the Estonian Biobank, for the non-persistence phenotype, we included individuals
494 discontinuing the treatment after one purchase that contained less than 3 packages and 100
495 tablets. We limited the drug supply that could be bought with one purchase to exclude
496 individuals who may have stocked medications due to moving abroad or due to other reasons.
497 We considered as early discontinuation only purchases made before 01.01.2022 (last date for
498 which the data are available (28.03.2023)) and at least one year before death date.

499 **Adherence.**

500 For each individual, we considered all registered purchases for each class of medication and
501 reconstructed individual drug purchasing trajectories. We then estimated adherence using the
502 Medication Possession Ratio measure (MPR) [19]:
503

504 $$MPR = \frac{days\ of\ supply\ during\ observation\ period}{days\ in\ observation\ period}$$

505

506 where
507 - *days of supply during observation period* is the total number of tablets purchased in the
508 trajectory, normalized assuming drug-specific daily doses (e.g. statins: 1 tablet/day)
509 - *days in observation period* is the sum of the intervals (in days) between all purchases
510 considered in the trajectory
511

512 When reconstructing the trajectories, we excluded the last purchase and purchases for which
513 the interval to the next one was of at least 150 days without tablets available. We considered
514 this as a break in the treatment which should not be taken into account into the adherence
515 calculation. When patients were switching to different formulations (within the same medication
516 class) and the interval between purchases was shorter than the drug supply, we discarded the
517 leftover tablets of the previous formulation. We included only individuals with at least 1 year of
518 purchases. We further excluded individuals with adherence greater than 1.1, to exclude
519 overbuying behaviors.
520

521 In the Estonian Biobank, patients were included if they had purchased at least 150 tablets.
522 Individual purchase trajectories were constructed and adherence was estimated using the same
523 criteria specified above with the exception of excluding 150 day intervals without available

treatment. The Estonian medical system does not offer services where the treatment would be provided by the hospital without it being recorded in the drug purchase database (with the exception of inpatient care). Therefore, we considered all gaps in drug purchases as true breaks in treatment.

Sex and gender.

Sex was used as covariate in all of the analyses and no analyses were run separately in each sex. In FinRegistry, sex was defined as recorded in the population registries from the Digital and Population Data Services Agency. In FinnGen, genetically determined sex was used, after quality control removal for individuals with discrepancy between sex reported in the population registries and genetically determined sex. In the Estonian Biobank, sex was extracted from the Estonian National Identity number which is created based on the sex recorded in the Estonian Birth Registry. Quality assurance was carried out where only individuals whose sex from the National Identity number matched the sex in genotype data are included.

Health, demographic and socioeconomic risk factors for persistence and adherence.

We assessed the effect of the following 8 epidemiological risk factors on persistence and adherence:

- sex (not considered for breast cancer medications)
- age at treatment initiation: age when the first purchase was recorded
- secondary prevention: whether the treatment was started after a major event related to the medication considered. The following diagnosis (ICD-10 codes) were considered:
    - statins: major coronary heart disease (I20.0, I21, I22), cerebrovascular diseases (I60-I69), atherosclerosis (excluding cerebral, coronary and PAD) (I70)
    - antiplatelets: stroke (I60-I64), transient ischemic attack (G45), myocardial infarction (I21, I22)
    - BP medications: no specific outcome was considered, as this is a broader class of medications
    - breast cancer medications: breast cancer (C50) - only secondary prevention patients were included
    - DOAC: atrial fibrillation (I48), pulmonary embolism or venous thromboembolism (I26, I80, I87.1) - only secondary prevention patients were included

555 - Charlson comorbidity index (at treatment initiation): we considered patients comorbidity
556 status at treatment initiation by computing the Charlson comorbidity index (CCI) based
557 on all the diagnoses registered previous to the first purchase. This was computed using
558 the R package ICCI (https://github.com/dsgelab/ICCI), which uses the R package
559 comorbidity (https://cran.r-project.org/web/packages/comorbidity/index.html).
560 - years of education (at treatment initiation): years expected to complete the recorded
561 education level. When the information at treatment baseline was not available, the first
562 available record was considered.
563 - living area (at treatment initiation): urban *vs* rural, as classified by the Finnish digital and
564 population data agency (DVV).
565 - mother tongue: Finnish or Swedish (the two official languages of Finland) *vs* any other
566 language
567 - social assistance benefits: receiving (or not) any amount of social assistance benefit in
568 the year before treatment initiation
569 The effect sizes of each factor on persistence and adherence were derived by using a
570 multivariate logistic/linear regression model including all the factors, adjusting for year of birth.
571
572 For persistence, the estimated effects are reported as OR from the logistic regression model,
573 while for adherence we reported the percentage change in adherence by rescaling the log(OR)
574 from the linear model by a factor of 1.1 (maximum adherence value). To provide a better
575 interpretation of the results, in Figure 3 we report the effect sizes for a model where continuous
576 variables were dichotomized according to the following thresholds: age at initiation > 60; CCI >
577 5; education, university degree or higher. The same effect sizes keeping the continuous
578 variables as such (standardized to have mean 0 and variance equal to 1) are reported in
579 **Supplementary Table 5-6**.
580
581 Among the health, demographic and socioeconomic factors considered in FinRegistry, the
582 following were available in the Estonian Biobank: sex, age at treatment initiation, birth year, type
583 of prevention, and years of education. Age at treatment initiation, birth year and years of
584 education were standardized to have mean 0 and variance equal to 1. We assessed the
585 association between statin persistence and adherence with these factors by using multivariate
586 logistic and linear regression models, respectively, while only considering individuals with
587 complete information for all variables (N =27 339).

588 **Pharmacogenes selection and phenotypes definition.**

589 We selected the pharmacogenes to test based on the guidelines provided by the Clinical
590 Pharmacogenetics Implementation Consortium [24]. The CPIC assigns levels to gene/drug pairs
591 based on PharmGKB [39] annotation levels, recommendations by external groups (e.g. FDA,
592 EMA) or by other professional societies. We selected gene/drug pairs to match all the ATC
593 codes included in our analyses, and with a CPIC level A or B, meaning that prescribing actions
594 (e.g. alternative therapies or dosing) are recommended when the pharmacogene is present.
595 The gene/drug pairs we selected are: *ABCG2* for rosuvastatin*, CYP2C9* for fluvastatin*,*
596 *SLCO1B1* for all statins, *CYP2C19* for clopidogrel, and *CYP2D6* for tamoxifen.
597 For each gene and for each individual included in the analyses, we defined their star alleles
598 diplotypes to test starting from the imputed phased genotypes available in FinnGen and using
599 Stargazer [40,41]. Having genotype data available, we could define only diplotypes captured by
600 SNP and not by structural variations (gene deletions or copy number variations). We then
601 grouped individuals based on the phenotypes also provided as output by Stargazer and
602 reflecting the metabolic activity score associated to each specific diplotype
603 (poor/decreased/increased function for *SLCO1B1* and *ABCG2*, poor/intermediate/ultrarapid
604 metabolizer for *CYP2D6, CYP2C9, CYP2C19*). In the Estonian Biobank star alleles were called
605 from imputed phased genotypes available using PharmCAT[42] and an in-house pipeline[43].
606
607 We used a logistic/linear regression model to estimate the association of each phenotype with
608 persistence/adherence, adjusting by sex, age at initiation (standardized) and first 10 principal
609 genetic components (standardized). We only tested phenotypes when defined for at least 5
610 individuals in the adherence analysis, or for at least 5 persistent/non-persistent individuals in the
611 persistence analysis.
612

613 **Genetic risk factors of persistence and adherence.**

614 We used REGENIE [44] to run a GWAS of persistence and adherence to each medication (see
615 **Supplementary Methods** for details about genotyping and imputation in FinnGen). Persistence
616 was treated as a binary phenotype, while adherence was kept as continuous trait, but scaled so
617 to have mean 0 and variance equal to 1. For both analyses we adjusted for sex, genotyping
618 batch, first 10 principal genetic components, age at first purchase and year of birth. The

resulting summary statistics were post-processed to keep only SNPs with INFO score ≥ 0.8 and MAF ≥ 1%.

Genetic correlations between persistence and adherence and 33 publicly available traits (listed in **Supplementary Table 13**) were calculated using linkage disequilibrium score regression [45].

PGSs for the same traits were derived using PRS-CS [46]. Their association with persistence and adherence was assessed by running a logistic/linear regression model for each PGS individually, adjusting for sex, age at initiation and the first 10 genetic principal components (standardized).

Weights for the same traits (**Supplementary Table 13**) were applied to the Estonian Biobank individuals using PLINK[47]. Similarly, the association of the PGSs with statin persistence and adherence was assessed by running a logistic or linear regression model for each PGS individually, adjusting for sex, age at initiation, and the first 10 principal components. The PGS score, age at treatment initiation, and principal components were standardized.

We also derived a PGS for adherence to statins from the summary statistics of the GWAS in FinnGen using PRS-CS and applied it to the Estonian Biobank individuals using PLINK. We then tested the association between statin adherence PGS and the adherence observed in the Estonian Biobank using a linear regression model, adjusting for sex, age at initiation and the first 10 genetic principal components.

Acknowledgements

We are grateful to Finnish individuals, whose data made this study possible. We would also like to thank the entire FinRegistry team for making the data available for the study.

We want to acknowledge the participants and investigators of FinnGen study. The FinnGen project is funded by two grants from Business Finland (HUS 4685/31/2016 and UH 4386/31/2016) and the following industry partners: AbbVie Inc., AstraZeneca UK Ltd, Biogen MA Inc., Bristol Myers Squibb (and Celgene Corporation & Celgene International II Sàrl), Genentech Inc., Merck Sharp & Dohme LCC, Pfizer Inc., GlaxoSmithKline Intellectual Property Development Ltd., Sanofi US Services Inc., Maze Therapeutics Inc., Janssen Biotech Inc,

Novartis AG, and Boehringer Ingelheim International GmbH. Following biobanks are acknowledged for delivering biobank samples to FinnGen: Auria Biobank (www.auria.fi/biopankki), THL Biobank (www.thl.fi/biobank), Helsinki Biobank (www.helsinginbiopankki.fi), Biobank Borealis of Northern Finland (https://www.ppshp.fi/Tutkimus-ja-opetus/Biopankki/Pages/Biobank-Borealis-briefly-in-English.aspx), Finnish Clinical Biobank Tampere (www.tays.fi/en-US/Research_and_development/Finnish_Clinical_Biobank_Tampere), Biobank of Eastern Finland (www.ita-suomenbiopankki.fi/en), Central Finland Biobank (www.ksshp.fi/fi-FI/Potilaalle/Biopankki), Finnish Red Cross Blood Service Biobank (www.veripalvelu.fi/verenluovutus/biopankkitoiminta), Terveystalo Biobank (www.terveystalo.com/fi/Yritystietoa/Terveystalo-Biopankki/Biopankki/) and Arctic Biobank (https://www.oulu.fi/en/university/faculties-and-units/faculty-medicine/northern-finland-birth-cohorts-and-arctic-biobank). All Finnish Biobanks are members of BBMRI.fi infrastructure (www.bbmri.fi). Finnish Biobank Cooperative -FINBB (https://finbb.fi/) is the coordinator of BBMRI-ERIC operations in Finland. The Finnish biobank data can be accessed through the Fingenious® services (https://site.fingenious.fi/en/) managed by FINBB.

The authors for the Estonian Biobank would like to express our gratitude to all EstBB participants who made it possible to conduct this study. Data analysis was carried out in part in the High-Performance Computing Centre of University of Tartu. This work was written at writing retreats and writing days organized by the University of Tartu Institute of Genomics. The Estonian Biobank Research Team responsible for data collection, genotyping, QC and imputation: Andres Metspalu, Tõnu Esko, Reedik Mägi, Mari Nelis and Georgi Hudjashov.

This study has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101016775; from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant number 945733); from Academy of Finland fellowship grant N. 323116; from the European Union through the European Regional Development Fund Project No. 2014-2020.4.01.15-0012 GENTRANSMED; the European Union's Horizon 2020 Research and Innovation Programme under Grant agreement 847776; the Estonian Research Council grant number PSG615.

682 **Data and code availability**

683 Data dictionaries for FinRegistry are publicly available on the FinRegistry website
684 (www.finregistry.fi/finnish-registry-data). Access to FinRegistry data can be obtained by
685 submitting a data permit application for individual-level data for the Finnish social and health
686 data permit authority Findata (https://asiointi.findata.fi/). The application includes information on
687 the purpose of data use; the requested data, including the variables, definitions for the target
688 and control groups, and external datasets to be combined with FinRegistry data; the dates of
689 the data needed; and a data utilization plan. The requests are evaluated on a case-by-case
690 basis. Once approved, the data are sent to a secure computing environment Kapseli and can be
691 accessed within the European Economic Area (EEA) and countries with an adequacy decision
692 from the European Commission.

693 Access to individual level data from the Estonian Biobank can be obtained through the Estonian
694 Biobank, following standard data access procedures
695 (https://genomics.ut.ee/en/content/estonian-biobank). All research using the Estonian Biobank
696 data is regulated by the Human Genes Research Act and must be approved by the Estonian
697 Committee of Bioethics and Human Research.

698 Analysis code used to produce the results is available on GitHub at:
699 https://github.com/dsgelab/drugs-persistence-adherence

700

701 **Ethics declarations**

702 FinRegistry is a collaboration project of the Finnish Institute for Health and Welfare (THL) and
703 the Data Science Genetic Epidemiology research group at the Institute for Molecular Medicine
704 Finland (FIMM), University of Helsinki. The FinRegistry project has received the following
705 approvals for data access from the National Institute of Health and Welfare
706 (THL/1776/6.02.00/2019 and subsequent amendments), DVV (VRK/5722/2019-2), Finnish
707 Center for Pension (ETK/SUTI 22003) and Statistics Finland (TK-53-1451-19). The FinRegistry
708 project has received IRB approval from the National Institute of Health and Welfare (Kokous
709 7/2019).

710 Patients and control subjects in FinnGen provided informed consent for biobank research,
711 based on the Finnish Biobank Act. Alternatively, separate research cohorts, collected prior the

712 Finnish Biobank Act came into effect (in September 2013) and start of FinnGen (August 2017),
713 were collected based on study-specific consents and later transferred to the Finnish biobanks
714 after approval by Fimea (Finnish Medicines Agency), the National Supervisory Authority for
715 Welfare and Health. Recruitment protocols followed the biobank protocols approved by Fimea.
716 The Coordinating Ethics Committee of the Hospital District of Helsinki and Uusimaa (HUS)
717 statement number for the FinnGen study is Nr HUS/990/2017.
718
719 The FinnGen study is approved by Finnish Institute for Health and Welfare (permit numbers:
720 THL/2031/6.02.00/2017, THL/1101/5.05.00/2017, THL/341/6.02.00/2018,
721 THL/2222/6.02.00/2018, THL/283/6.02.00/2019, THL/1721/5.05.00/2019 and
722 THL/1524/5.05.00/2020), Digital and population data service agency (permit numbers:
723 VRK43431/2017-3, VRK/6909/2018-3, VRK/4415/2019-3), the Social Insurance Institution
724 (permit numbers: KELA 58/522/2017, KELA 131/522/2018, KELA 70/522/2019, KELA
725 98/522/2019, KELA 134/522/2019, KELA 138/522/2019, KELA 2/522/2020, KELA 16/522/2020),
726 Findata permit numbers THL/2364/14.02/2020, THL/4055/14.06.00/2020,
727 THL/3433/14.06.00/2020, THL/4432/14.06/2020, THL/5189/14.06/2020,
728 THL/5894/14.06.00/2020, THL/6619/14.06.00/2020, THL/209/14.06.00/2021,
729 THL/688/14.06.00/2021, THL/1284/14.06.00/2021, THL/1965/14.06.00/2021,
730 THL/5546/14.02.00/2020, THL/2658/14.06.00/2021, THL/4235/14.06.00/2021, Statistics Finland
731 (permit numbers: TK-53-1041-17 and TK/143/07.03.00/2020 (earlier TK-53-90-20)
732 TK/1735/07.03.00/2021, TK/3112/07.03.00/2021) and Finnish Registry for Kidney Diseases
733 permission/extract from the meeting minutes on 4th July 2019.
734
735 The Biobank Access Decisions for FinnGen samples and data utilized in FinnGen Data Freeze
736 10 include: THL Biobank BB2017_55, BB2017_111, BB2018_19, BB_2018_34, BB_2018_67,
737 BB2018_71, BB2019_7, BB2019_8, BB2019_26, BB2020_1, BB2021_65, Finnish Red Cross
738 Blood Service Biobank 7.12.2017, Helsinki Biobank HUS/359/2017, HUS/248/2020,
739 HUS/150/2022 § 12, §13, §14, §15, §16, §17, §18, and §23, Auria Biobank AB17-5154 and
740 amendment #1 (August 17 2020) and amendments BB_2021-0140, BB_2021-0156 (August 26
741 2021, Feb 2 2022), BB_2021-0169, BB_2021-0179, BB_2021-0161, AB20-5926 and
742 amendment #1 (April 23 2020)and it´s modification (Sep 22 2021), Biobank Borealis of Northern
743 Finland_2017_1013, 2021_5010, 2021_5018, 2021_5015, 2021_5023, 2021_5017, 2022_6001,
744 Biobank of Eastern Finland 1186/2018 and amendment 22 § /2020, 53§/2021, 13§/2022,
745 14§/2022, 15§/2022, Finnish Clinical Biobank Tampere MH0004 and amendments (21.02.2020

& 06.10.2020), §8/2021, §9/2022, §10/2022, §12/2022, §20/2022, §21/2022, §22/2022, §23/2022, Central Finland Biobank 1-2017, and Terveystalo Biobank STB 2018001 and amendment 25th Aug 2020, Finnish Hematological Registry and Clinical Biobank decision 18th June 2021, Arctic biobank P0844: ARC_2021_1001.

The analysis of individual-level data from the Estonian Biobank was carried out under ethical approval nr 1.1-12/624 from the Estonian Committee on Bioethics and Human Research (Estonian Ministry of Social Affairs), using data according to data release application nr T12 6-7/GI/11086 from the Estonian Biobank.

Authors contribution

Study design: MC, AC, SJ, MN, KL, LM and AG. Data analysis: MC, AC and HMK. Results interpretation: MC, AC, HMK, SJ, PK, TK, MF, SR, MN, KL, LM, AG. Writing – original draft: MC, AC and AG. All authors were involved in further drafts of the manuscript and revised it critically for content. All authors gave final approval of the version to be published. MC, AC and HMK contributed equally to this work.

Conflicts of interest: AG is CEO and founder of Real World Genetics Oy.

764 **References**

765 1. Osterberg, L. & Blaschke, T. Adherence to medication. *N. Engl. J. Med.* **353**, 487–497 (2005).

766 2. Simpson, R. J., Jr & Mendys, P. The effects of adherence and persistence on clinical outcomes in
767 patients treated with statins: a systematic review. *J. Clin. Lipidol.* **4**, 462–471 (2010).

768 3. Miller, N. H., Hill, M., Kottke, T. & Ockene, I. S. The multilevel compliance challenge:
769 recommendations for a call to action. A statement for healthcare professionals. *Circulation* **95**, 1085–
770 1090 (1997).

771 4. Kardas, P., Lewek, P. & Matyjaszczyk, M. Determinants of patient adherence: a review of systematic
772 reviews. *Front. Pharmacol.* **4**, 91 (2013).

773 5. Sabaté, E. *Adherence to Long-Term Therapies: Evidence for Action*. (World Health Organization,
774 2003).

775 6. Simpson, S. H. *et al.* A meta-analysis of the association between adherence to drug therapy and
776 mortality. *BMJ* **333**, 15 (2006).

777 7. Chan, D. C. *et al.* Patient, physician, and payment predictors of statin adherence. *Med. Care* **48**,
778 196–202 (2010).

779 8. Citarella, A. *et al.* Family history of cardiovascular disease and influence on statin therapy
780 persistence. *Eur. J. Clin. Pharmacol.* **70**, 701–707 (2014).

781 9. Shin, D. W. *et al.* Antihypertensive medication adherence in cancer survivors and its affecting factors:
782 results of a Korean population-based study. *Support. Care Cancer* **19**, 211–220 (2010).

783 10. Yang, Y. *et al.* Predictors of medication nonadherence among patients with diabetes in Medicare Part
784 D programs: a retrospective cohort study. *Clin. Ther.* **31**, 2178–88; discussion 2150-1 (2009).

785 11. Kariis, H. M. *et al.* The role of depression and antidepressant treatment in antihypertensive
786 medication adherence and persistence: Utilising electronic health record data. *J. Psychiatr. Res.* **168**,
787 269–278 (2023).

788 12. Rolnick, S. J., Pawloski, P. A., Hedblom, B. D., Asche, S. E. & Bruzek, R. J. Patient characteristics
789 associated with medication adherence. *Clin. Med. Res.* **11**, 54–65 (2013).

790 13. Yeaw, J., Benner, J. S., Walt, J. G., Sian, S. & Smith, D. B. Comparing adherence and persistence
791 across 6 chronic medication classes. *J. Manag. Care Pharm.* **15**, 728–740 (2009).

792 14. Pirmohamed, M. Pharmacogenomics: current status and future perspectives. *Nat. Rev. Genet.* **24**,
793 350–362 (2023).
794 15. Swen, J. J. *et al.* A 12-gene pharmacogenetic panel to prevent adverse drug reactions: an open-
795 label, multicentre, controlled, cluster-randomised crossover implementation study. *Lancet* **401**, 347–
796 356 (2023).
797 16. Hess, L. M., Raebel, M. A., Conner, D. A. & Malone, D. C. Measurement of adherence in pharmacy
798 administrative databases: a proposal for standard definitions and preferred measures. *Ann.*
799 *Pharmacother.* **40**, 1280–1288 (2006).
800 17. Vrijens, B. *et al.* A new taxonomy for describing and defining adherence to medications. *Br. J. Clin.*
801 *Pharmacol.* **73**, 691–705 (2012).
802 18. Cramer, J. A. *et al.* Medication compliance and persistence: terminology and definitions. *Value*
803 *Health* **11**, 44–47 (2008).
804 19. Andrade, S. E., Kahler, K. H., Frech, F. & Chan, K. A. Methods for evaluation of medication
805 adherence and persistence using automated databases. *Pharmacoepidemiol. Drug Saf.* **15**, 565–74;
806 discussion 575-7 (2006).
807 20. Viippola, E. *et al.* Data Resource Profile: Nationwide registry data for high-throughput epidemiology
808 and machine learning (FinRegistry). *Int. J. Epidemiol.* (2023) doi:10.1093/ije/dyad091.
809 21. Kurki, M. I. *et al.* FinnGen provides genetic insights from a well-phenotyped isolated population.
810 *Nature* **613**, 508–518 (2023).
811 22. Leitsalu, L. *et al.* Cohort Profile: Estonian Biobank of the Estonian Genome Center, University of
812 Tartu. *Int. J. Epidemiol.* **44**, 1137–1147 (2015).
813 23. DiMatteo, M. R. Variations in patients' adherence to medical recommendations: a quantitative review
814 of 50 years of research. *Med. Care* **42**, 200–209 (2004).
815 24. Relling, M. V. & Klein, T. E. CPIC: Clinical Pharmacogenetics Implementation Consortium of the
816 Pharmacogenomics Research Network. *Clin. Pharmacol. Ther.* **89**, 464–467 (2011).
817 25. Cooper-DeHoff, R. M. *et al.* The Clinical Pharmacogenetics Implementation Consortium Guideline for
818 SLCO1B1, ABCG2, and CYP2C9 genotypes and Statin-Associated Musculoskeletal Symptoms. *Clin.*
819 *Pharmacol. Ther.* **111**, 1007–1021 (2022).

820 26. Lee, C. R. *et al.* Clinical Pharmacogenetics Implementation Consortium Guideline for CYP2C19
821 Genotype and Clopidogrel Therapy: 2022 Update. *Clin. Pharmacol. Ther.* **112**, 959–967 (2022).
822 27. Goetz, M. P. *et al.* Clinical Pharmacogenetics Implementation Consortium (CPIC) Guideline for
823 CYP2D6 and Tamoxifen Therapy. *Clin. Pharmacol. Ther.* **103**, 770–777 (2018).
824 28. Dönertaş, H. M., Fabian, D. K., Valenzuela, M. F., Partridge, L. & Thornton, J. M. Common genetic
825 associations between age-related diseases. *Nat Aging* **1**, 400–412 (2021).
826 29. Ghoussaini, M. *et al.* Open Targets Genetics: systematic identification of trait-associated genes using
827 large-scale genetics and functional genomics. *Nucleic Acids Res.* **49**, D1311–D1320 (2021).
828 30. Tyrrell, J. *et al.* Genetic predictors of participation in optional components of UK Biobank. *Nat.*
829 *Commun.* **12**, 886 (2021).
830 31. Helin-Salmivaara, A. *et al.* Long-term persistence with statin therapy: a nationwide register study in
831 Finland. *Clin. Ther.* **30 Pt 2**, 2228–2240 (2008).
832 32. Zucker, I. & Prendergast, B. J. Sex differences in pharmacokinetics predict adverse drug reactions in
833 women. *Biol. Sex Differ.* **11**, 32 (2020).
834 33. Brabete, A. C. *et al.* A Sex- and Gender-Based Analysis of Adverse Drug Reactions: A Scoping
835 Review of Pharmacovigilance Databases. *Pharmaceuticals* **15**, (2022).
836 34. Kiiskinen, T. *et al.* Genetic predictors of lifelong medication-use patterns in cardiometabolic diseases.
837 *Nat. Med.* **29**, 209–218 (2023).
838 35. Hartonen, T. *et al.* Nationwide health, socio-economic and genetic predictors of COVID-19
839 vaccination status in Finland. *Nat Hum Behav* **7**, 1069–1083 (2023).
840 36. Shiraly, R., Khani Jeihooni, A. & Bakhshizadeh Shirazi, R. Perception of risk of hypertension related
841 complications and adherence to antihypertensive drugs: a primary healthcare based cross-sectional
842 study. *BMC Prim Care* **23**, 303 (2022).
843 37. Harden, K. P. & Koellinger, P. D. Using genetics for social science. *Nat Hum Behav* **4**, 567–576
844 (2020).
845 38. Aarnio, E. J. *et al.* Register-based predictors of adherence among new statin users in Finland. *J.*
846 *Clin. Lipidol.* **8**, 117–125 (2014).
847 39. Whirl-Carrillo, M. *et al.* An Evidence-Based Framework for Evaluating Pharmacogenomics

848 Knowledge for Personalized Medicine. *Clin. Pharmacol. Ther.* **110**, 563–572 (2021).
849 40. Lee, S.-B. *et al.* Stargazer: a software tool for calling star alleles from next-generation sequencing
850 data using CYP2D6 as a model. *Genet. Med.* **21**, 361–372 (2019).
851 41. Lee, S.-B., Wheeler, M. M., Thummel, K. E. & Nickerson, D. A. Calling Star Alleles With Stargazer in
852 28 Pharmacogenes With Whole Genome Sequences. *Clin. Pharmacol. Ther.* **106**, 1328–1337
853 (2019).
854 42. Sangkuhl, K. *et al.* Pharmacogenomics Clinical Annotation Tool (PharmCAT). *Clin. Pharmacol. Ther.*
855 **107**, 203–210 (2020).
856 43. Reisberg, S. *et al.* Translating genotype data of 44,000 biobank participants into clinical
857 pharmacogenetic recommendations: challenges and solutions. *Genet. Med.* **21**, 1345–1354 (2019).
858 44. Mbatchou, J. *et al.* Computationally efficient whole-genome regression for quantitative and binary
859 traits. *Nat. Genet.* **53**, 1097–1103 (2021).
860 45. Bulik-Sullivan, B. *et al.* An atlas of genetic correlations across human diseases and traits. *Nat. Genet.*
861 **47**, 1236–1241 (2015).
862 46. Ge, T., Chen, C.-Y., Ni, Y., Feng, Y.-C. A. & Smoller, J. W. Polygenic prediction via Bayesian
863 regression and continuous shrinkage priors. *Nat. Commun.* **10**, 1776 (2019).
864 47. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage
865 analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
866