Genomic prediction of alcohol-related morbidity and mortality

Tuomo Kiiskinen, Nina J. Mars, Teemu Palviainen, Jukka Koskela, Joel T. Rämö, Pietari Ripatti, Sanni Ruotsalainen, FinnGen, GSCAN Consortium, Aarno Palotie, Pamela A. F. Madden, Richard J. Rose, Jaakko Kaprio, Veikko Salomaa, Pia Mäkelä, Aki S. Havulinna and Samuli Ripatti

Abstract
While polygenic risk scores (PRS) have been shown to predict many diseases and risk factors, the potential of genomic prediction in harm caused by alcohol use has not yet been extensively studied. Here, we built a novel polygenic risk score of 1.1 million variants for alcohol consumption and studied its predictive capacity in 96,499 participants from the FinnGen study and 39,695 participants from prospective cohorts with detailed baseline data and up to 25 years of follow-up time. A 1 SD increase in the PRS was associated with 11.2 g (=0.93 drinks) higher weekly alcohol consumption (CI = 9.85–12.58 g, p = 2.3 × 10^-8). The PRS was associated with alcohol-related morbidity (4785 incident events) and the risk estimate between the highest and lowest quintiles of the PRS was 1.83 (95% CI = 1.66–2.01, p = 1.6 × 10^-30). When adjusted for self-reported alcohol consumption, education, marital status, and gamma-glutamyl transferase blood levels in 28,639 participants with comprehensive baseline data from prospective cohorts, the risk estimate between the highest and lowest quintiles of the PRS was 1.58 (CI = 1.26–1.99, p = 8.2 × 10^-5). The PRS was also associated with all-cause mortality with a risk estimate of 1.33 between the highest and lowest quintiles (CI = 1.20–1.47, p = 4.5 × 10^-6) in the adjusted model. In conclusion, the PRS for alcohol consumption independently associates for both alcohol-related morbidity and all-cause mortality. Together, these findings underline the importance of heritable factors in alcohol-related health burden while highlighting how measured genetic risk for an important behavioral risk factor can be used to predict related health outcomes.

Introduction
Alcohol drinking is a major dose-dependent contributor to morbidity and mortality. Globally, 3 million annual deaths (5% of all deaths) result from alcohol consumption, and is also linked to more than 200 disease and injury outcomes. As ethanol is a psychoactive substance with addictive properties, alcohol consumption can lead to the development of alcohol use disorders (AUDs), globally prevalent mental disorders of pathological addictive or abusive drinking patterns, which are linked to worse health outcomes, negative socioeconomic effects, and increased mortality. There is a strong connection between the health burden and the level of alcohol consumed, and in total, alcohol has been estimated to be the most damaging of all substances of abuse, in terms of harm caused to self and others.

Alcohol-related behaviors are also affected by genetic factors and the estimated heritability of alcohol consumption in twin studies has ranged between 35% and 65% (weighted average 37%) and its single nucleotide polymorphism-based heritability has been estimated to be 10%. Recent large-scale genome-wide association studies (GWAS) have identified multiple loci associated with alcohol consumption, underlining the importance of large study populations for unraveling the genetic architecture underlying alcohol-related traits. Similarly, GWAS of...
alcohol dependence, AUD, and the Alcohol Use Disorders Identification Test (AUDIT) scores have shown the traits to be genetically distinct but positively correlated9–11.

Polygenic risk scores (PRSs) derived from GWAS summary statistics have showcased improved performance in disease prediction12. PRSs for known risk factors have also been shown to associate with the related disease13, and recently associations between multiple risk factor PRSs and related traits were confirmed and reported14,15. However, the link between PRSs for behavioral traits and associated health outcomes remains poorly understood.

The assessment of potential health risks related to alcohol has so far relied on traditional risk factors, including family history, without explicit measurement of genetic risk. Here we developed a highly polygenic risk score for alcohol consumption and studied whether alcohol-related polygenic burden predicts alcohol-use disorders and other alcohol-related morbidity and mortality in Finnish biobank cohorts (n = 96,499) linked to electronic health records. Furthermore, we studied whether the PRS for alcohol consumption predicts alcohol-related outcomes beyond self-reported alcohol consumption and other related risk factors, thus providing more objective information independent of individual reporting bias or temporal fluctuations.

Methods

Study sample and definition of alcohol-related morbidity

The data are comprised of 96,499 Finnish individuals from FinnGen Data Freeze 2 (https://www.finngen.fi/), which includes prospective epidemiological and disease-based cohorts as well as hospital biobank samples (Contributors S1, Table S1). The data were linked by the unique national personal identification numbers assigned to all Finnish citizens and residents. The final alcohol-related morbidity endpoint was defined as any event coded as any of the conditions in the composite endpoint.

Genotyping and imputation

FinnGen, FINRISK, Health 2000, and Finnish Twin Cohort samples were genotyped with Illumina and Affymetrix genomewide SNP arrays. Individuals with non-European ancestry or uncertain sex were excluded. Within each cohort, every genotyping batch was first imputed separately and then merged together for association analyses. The details about the genotype calling, quality controls, and imputation are provided in the Supplementary material (Methods S1).

Polygenic risk scores

Summary statistics from the largest existing GWAS meta-analysis on alcohol consumption (8) were used for constructing the PRS. To avoid overfitting, a separate ad hoc meta-analysis was performed by GSCAN (Contributors S2), excluding all Finnish and 23andMe samples (n = 527,282 after exclusions). LDpred-method20 was used to account for linkage disequilibrium (LD) among loci with whole-genome sequencing data on 2690 Finns serving as the external LD reference panel. We compared the PRSs generated with LDpred-parameters and their predictive ability in FINRISK (Fig S1). Any threshold above 0.003 worked practically similarly, and for simplicity we chose to use the LDpred-inf PRS in all the analyses. The final scores were generated with PLINK2 (ref. 21) by calculating the weighted sum of risk allele dosages for each variant. The number of variants in the final scores was 1.1 million (1,134,960 in FinnGen, 1,143,220 in FINRISK and Health 2000, and 1,143,138 in the Twin Cohort).

Statistical analysis

The Cox proportional hazard model was used to estimate survival curves, hazard ratios (HRs), and 95% confidence interval (95% CI) in the survival analyses where
age was used as the time scale. R’s cox.zph function was used to test whether the proportional assumption criteria applied in our models. Linear regression in FINRISK and Health 2000 and linear mixed model in the Twin Cohort was used for estimating the relationship between the PRS and alcohol consumption. Logistic regression in the FINRISK and Health 2000 cohorts and linear mixed model in the Twin Cohort was used to estimate the relationship between alcohol abstinence and the PRS.

All the cohorts (FinnGen, FINRISK, Health 2000, and the Twin Cohort) were analyzed independently as single datasets where age, sex, genotyping array, and the first ten principal components of ancestry were used as core covariates. Additionally, body mass was used as a covariate in the model estimating the PRS–alcohol consumption relationship. Self-reported weekly average alcohol consumption from the past year (when available, the past week’s consumption) was used as the estimate for alcohol consumption. In the fully adjusted survival model analyses, \( \log(x + 1) \)-transformed alcohol consumption-estimate, current smoking status, binary higher education status, binary marital/cohabitation status, and gamma-glutamyl transferase (GGT) blood levels at baseline served as covariates. The GGT levels were measured following uniform recommendations of the European Committee for Clinical Laboratory Standards (ECCLS)\(^2\) enabling comparability between the cohorts.

In the survival analyses, all prevalent cases (in FINRISK and Health 2000) and individuals with covariate missingness were excluded. The PRS was normalized and included as a continuous variable in the models. In the survival analysis, the highest and lowest genetic risk for alcohol consumption were compared using PRS quintiles.

In analyses using baseline consumption data, the analyses were performed separately in the Health 2000, FINRISK Study, and Twin Cohorts and then meta-analyzed using fixed effects model.

In risk prediction, FINRISK cohorts with at least 10 years of follow-up (from 1992 to 2002) were used to train the model, and the predictive performance was tested in the Health 2000 cohort. The maximal follow-up window was restricted to 10 years. The change in the predictive performance was assessed by comparing models with and without the PRS using the correlated C-index approach\(^3\) along with calculating the continuous reclassification improvement (NRI)\(^4\) and integrated discrimination improvement (IDI)\(^5\). The Hosmer–Lemeshow goodness-of-fit test was used to test model calibration.

**Ethical approval**

The study was conducted in accordance with the principles of the Helsinki declaration. Written informed consent was obtained from all the study participants. For the Finnish Institute of Health and Welfare (THL)-driven FinnGen preparatory project and FinnGen project, all patients and control subjects had provided informed consent for biobank research, based on the Finnish Biobank Act. Alternatively, FINRISK and Health 2000 cohorts were based on study specific consents and later transferred to the THL Biobank after approval by Valvira, the National Supervisory Authority for Welfare and Health. Recruitment protocols followed the biobank protocols approved by Valvira. The Biobank Access Decisions for FinnGen samples and data utilized in FinnGen Data Freeze 2 include: Auria Biobank AB17-5154, THL Biobank BB2017_55, BB2017_111, BB2018_19, BB_2018_34, Finnish Red Cross Blood Service Biobank 7.12.2017, Helsinki Biobank HUS/359/2017 and Northern Finland Biobank Borealis BB_2017_1013. The Ethical Review Board of the Hospital District of Helsinki and Uusimaa approved the FinnGen study protocol Nr HUS/990/2017. The FinnGen preparatory project as well as the FinnGen project was approved by THL, approval numbers THL/2031/6.02.00/2017, amendments THL/341/6.02.00/2018, THL/2222/6.02.00/2018, THL/1101/5.05.00/2017, VRK43431/2017-3, KELA 131/522/2018, and Statistics Finland TK-53-1041-17. The Twin Cohort studies were approved by the Coordinating Ethical Committee of the Helsinki and Uusimaa Hospital District, reference numbers 246/13/03/00/15, 113/E3/2001, and HUS/1169/2016. The transfer of the FINRISK and Health 2000 sample collections to the THL biobank has been approved by the Coordinating Ethics Committee of Helsinki University Hospital on 10 October 2014 and by the Ministry of Social Affairs and Health on 9 March 2015. This study was conducted under the THL biobank permission BB2017_64 (FINRISK and Health 2000). No additional ethical approval was needed for meta-analyzing the results. All DNA samples and data in this study were pseudonymized.

**Results**

**Cohorts**

Our primary dataset (FinnGen) is comprised of 96,499 unrelated individuals (54,262 women) with a total of 55,484,114 person-years of registry-based follow-up and 4785 first-observed alcohol-related major health events. Alcohol consumption estimates were available for a total of 39,695 individuals from the prospective cohorts (FINRISK, Health 2000, and Twin Cohort, Fig. S2). Two cohorts, FINRISK and Health 2000, have full registry data and information on self-reported alcohol consumption and related baseline data, and consist of 28,639 individuals (94.5% of the participants after excluding 964 prevalent alcohol-related morbidity cases), with 424,053 person-years of registry-based follow-up and 988 first ever alcohol-related events (Table 1). The interview-based DSM-IV AUD-status was available in a subset of the Twin cohort for 713 cases and 1460 controls.
Alcohol consumption

In a meta-analysis of the three cohorts with alcohol consumption estimates available \((n = 39,695)\), the PRS for alcohol consumption was strongly associated with self-reported alcohol consumption. A 1 SD increase in the PRS was associated with an 11.2 g \((=0.93 \text{ drinks} \times 12 \text{ g})\) increase in weekly pure alcohol intake \((\beta = 11.2 [9.85–12.6 g], \rho = 2.3 \times 10^{-58})\) (Fig. 1, cohort-specific figures: Fig. S3). Adding the PRS to the model improved \(r^2\) by ~0.6 percentage points (from 9.17% to 9.80%). In addition, the PRS was negatively associated with alcohol abstinence (reported alcohol consumption 0). In FINRISK and Health2000, a 1 SD increase in the PRS for alcohol consumption was associated with a 13.7% reduced likelihood of being a nondrinker \((\text{OR} = 0.863 [0.833–0.895], \rho = 6.1 \times 10^{-16})\) while this was not the case in the Twin Cohort where there were only 30 nondrinkers \((\text{OR} = 0.999 [0.998–1.00], \rho = 0.31)\).

Alcohol-related morbidity

The PRS for alcohol consumption was strongly associated with increased risk for lifelong major alcohol-related events derived from electronic health-records in the FinnGen dataset \((n = 96,499, \text{cases} = 4785)\) (Fig. 2). The FinnGen dataset was divided into three groups consisting of the lowest quintile, three middle quintiles, and the highest quintile of the alcohol consumption PRS. The cumulative disease rate of alcohol-related morbidity is displayed as a function of age \((n = 96,499)\).

### Table 1 Population characteristics of FinnGen, FINRISK, Health 2000, and Twin Cohort datasets.

|                        | FinnGen                      | FINRISK                      | Health 2000                  | Twin Cohort                  |
|------------------------|------------------------------|------------------------------|------------------------------|------------------------------|
| \(N\)                  | 96,499 (54,262 women)        | 23,824 (12,513)              | 5945 (3260)                  | 9926 (5036)                  |
| \(N\) (incident events)| 4785                         | 817                          | 171                          | NA                           |
| Age (years)            | 57.5 (end of follow-up)      | 48.6 (baseline)              | 54.0 (baseline)              | 49.3 (baseline)              |
| Alcohol drinking (g/week)| NA                          | 76.5                         | 73.9                         | 84.8                         |
| Non-drinkers           | NA                           | 2874 (12%)                   | 1298 (22%)                   | 30 (0.3%)                    |
| Current smokers        | NA                           | 5929 (25%)                   | 1538 (26%)                   | 3478 (35%)                   |
| Higher education       | NA                           | 8612 (36%)                   | 1721 (28%)                   | 1192 (12%)                   |
| Marriage or co-habitation| NA                         | 17,468 (73%)                 | 4151 (68%)                   | 6661 (67%)                   |
| GGT (U/I)              | NA                           | 33.8                         | 36.6                         | NA                           |

Fig. 1 Alcohol drinking (g/week) for the deciles of the alcohol consumption polygenic risk score. The association is shown for males \((n = 18,887)\) and females \((n = 20,808)\) with 95% confidence interval error bars \((n = 39,695)\).

Fig. 2 Cumulative hazard risk comparison of alcohol consumption polygenic risk score quintiles. The FinnGen dataset was divided into three groups consisting of the lowest quintile, three middle quintiles, and the highest quintile of the alcohol consumption PRS. The cumulative disease rate of alcohol-related morbidity is displayed as a function of age \((n = 96,499)\).
a 1 SD increase in the PRS was associated with a 26% increase in risk (HR = 1.26 [1.23–1.30], p = 5.7 × 10^{-56}). The association was similar in both males (HR = 1.26 [1.22–1.31], p = 1.3 × 10^{-38}) and females (HR = 1.27 [1.20–1.33], p = 4.8 × 10^{-19}).

In the cohorts where alcohol consumption estimates and other related baseline data were available at the cohort entry time, the PRS was associated with an increased risk of incident major alcohol-related events and the association was maintained also in the fully adjusted model (n = 28,639, cases = 911). In a meta-analysis of the two cohorts, 1 PRS SD was associated with a 26% increased risk of incident alcohol-related events when the consumption-estimate was not in the model (HR = 1.26 [1.18–1.34], p = 1.1 × 10^{-12}) and with a 15% increase when alcohol consumption was in the model (HR = 1.15 [1.08–1.22], p = 2.1 × 10^{-5}). In a fully adjusted model, including marital status, education, smoking status, and GGT, the estimate was unchanged (HR = 1.15 [1.08–1.22], p = 2.0 × 10^{-5}) (Table 2). The risk estimate between the highest and lowest quintiles of the PRS in the fully adjusted model was 1.58 (HR = 1.58 [1.26–1.99], p = 8.2 × 10^{-5}).

Mortality

We observed a similar increase in the risk of alcohol-related and all-cause mortality. In FinnGen with 7249 deaths, 1 SD increase in the PRS for alcohol consumption was associated with 8% increase in the risk of death (HR = 1.08 [1.06–1.11], p = 2.0 × 10^{-11}). The risk estimate between the highest and lowest 20% in the PRS was 1.27 (HR = 1.27 [1.18–1.37], p = 2.1 × 10^{-16}). In our prospective cohorts, with cause-of-death information available, 4125 deaths were recorded. For all-cause mortality, there was 11% increase in the risk of death per 1 PRS SD in the basic model (HR = 1.11 [1.07–1.14], p = 3.2 × 10^{-10}) and 9% in the fully adjusted model (HR = 1.09 [1.06–1.12], p = 1.1 × 10^{-5}). The risk difference between the highest and lowest quintiles of the PRS was 33% (HR = 1.33 [1.2–1.47], p = 4.5e-08) in the fully adjusted model.

Of the 4125 deaths, 335 were known to be alcohol-related. Without alcohol consumption in the model, the increase in alcohol-related mortality was 26% per 1 PRS SD (HR = 1.26 [1.13–1.4], p = 3.7 × 10^{-5}). When alcohol consumption was included in the model, the increase was 13% (HR = 1.13 [1.01–1.26], p = 0.027) and in a model with all co-variates, 11% (HR = 1.11 [0.996–1.24], p = 0.058) (Table 2). Similarly, the PRS was associated with a higher risk of death from other than alcohol-related causes (n = 3790) when fully adjusted for all covariates (HR = 1.08 [1.05–1.12], p = 1.4 × 10^{-5}).

**DSM-IV alcohol-use disorder**

The PRS was also associated with an interview-based DSM-IV AUD diagnosis in the Nicotine Addiction Genetics Family cohort (440 cases, 1140 controls) and a subset of FinnTwin16 cohort (273 cases, 320 controls). A meta-analysis of the two cohorts (713 cases) resulted in a combined 20% increase in the prevalence of AUD per 1 PRS SD (OR = 1.20 [1.11–1.31], p = 2.29 × 10^{-5}) in the unadjusted model. Adjusting for marital status, education, and smoking explained part of the effect (OR = 1.14 [1.02–1.28], p = 0.023) and further adjusting with maximal amount of drinks taken explained most of the effect (OR = 1.06 [0.94–1.19], p = 0.35).

**Prediction**

The predictive performance of the PRS was evaluated in the Health 2000 cohort (5732 complete cases, 110 events) with a follow-up-window of 10 years based on the Cox model trained in the FINRISK cohort (18,427 complete cases with ≥ 10 years of follow-up, 628 events). In a model not including the alcohol consumption estimate, adding
the PRS to the model increased the C-index by 0.020, from 0.69 to 0.71 \((p = 0.017)\). Both IDI \((0.00242 [0.00102–0.00383], p = 7.3 × 10^{-5})\) and NRI \((0.335 [0.146–0.523], p = 5.1 \times 10^{-5})\) shifts were positive and statistically significant. When the log-transformed alcohol consumption estimate was included, a minimal improvement of prediction was observed (C-index = 0.0022 from 0.812 to 0.814, \(p\)-value = 0.30; NRI = 0.308 [0.119–0.497], \(p = 0.014\) and IDI = 0.00173 [0.000726–0.00305], \(p = 0.017\)). Similarly, a minimal gain was observed when adding PRS to a model with all available covariates including also marital status, education status, smoking status, and GGT (C-index = 0.00183 from 0.847 to 0.849, \(p = 0.44\); NRI = 0.235 [0.0461–0.423], \(p = 0.015\); IDI = 0.00331 [0.0000254–0.00659], \(p = 0.048\)).

**Discussion**

We developed a highly polygenic risk score for alcohol consumption by obtaining weights from a recently published large-scale discovery sample and showed that the PRS was strongly associated with alcohol consumption in independent biobank cohort samples. An increased polygenic burden for alcohol consumption was associated with higher incidence of major alcohol-induced health events. The associations remained significant when we accounted for self-reported alcohol consumption and other relevant covariates; in a fully adjusted model, the relative risk-estimate between the highest and lowest quintiles of the PRS was 1.6. Furthermore, the PRS was also associated with both alcohol-related, non-alcohol related, and all-cause mortality.

Our PRS shows the utility of genetic information for prediction of alcohol-related harm. The PRS, developed from a genetic analysis of cross-sectional self-reported alcohol consumption, was associated with future risk of major alcohol-related health events. While a large number of PRSs have already been established for various traits and diseases\(^1\), the development of PRSs for behavioral traits, such as substance use, has until now been limited\(^26–29\) and the studies have not assessed their impact on future major health events.

Our results show that using a large sample size with long follow-up, we were able to build a PRS of alcohol consumption that is associated not only with alcohol consumption in independent samples, but also with future incident alcohol-related health events. In line with the knowledge that alcohol consumption is a major contributor to the worldwide burden of death, especially among working-age adults\(^9\), we found the PRS to be associated also with all-cause mortality, further highlighting the importance of alcohol drinking as a cause of premature death.

Our score provides a genetic basis for potentially identifying a subset of high-risk individuals even early on in life, with potential for more targeted prevention of AUDs and other alcohol-related morbidity. Prevention is a cost-effective and efficient strategy to reduce alcohol-related harms\(^30\) and it is labeled one of the United Nations main health-related worldwide strategies of sustainable development \(\text{(https://sustainabledevelopment.un.org/sdg3)}\). A higher genetic predisposition for alcohol-related harms was detected both in the presence and absence of alcohol consumption data, as our PRS predicted alcohol-related harms beyond self-reported alcohol consumption. Health services are encouraged to support initiatives for screening and brief interventions for harmful drinking\(^31\) as an effective strategy for tackling alcohol-related harm\(^32\). Thus, genetic information could potentially be used to improve the arsenal of possible strategies to detect high-risk individuals for targets of brief interventions. The fact that individuals in the highest PRS quintile showed an elevated risk for alcohol-related health events even in fully adjusted models could justify the use of genetic information even in clinical settings where a detailed history of alcohol consumption estimates, AUDIT-scores, or similar information are attainable. Communicating the information of higher risk for alcohol-related harm to patients could serve as a motivator for reducing drinking or committing to abstinence. However, the true effects of informing patients about their alcohol-related genomic risk warrants further research.

Self-reported alcohol consumption is known to be biased and problematic in terms of reliability and validity for predicting alcohol-related risks\(^33,34\). Also, GGT is known to be less-than-ideal biochemical measure of drinking\(^35\). Some inaccuracy derives from true measurement error, but another source is the lifelong temporal fluctuation of alcohol-drinking patterns not captured by a measure at one single timepoint. Our PRS was associated with alcohol-related harms even when adjusting for self-reported alcohol consumption estimate. One potential reason for this is that the PRS contains information from the latent genetic predisposition for alcohol consumption, thus overriding both the true measurement error and temporal fluctuations in alcohol drinking volume.

Furthermore, it has been hypothesized that alcohol consumption-based genetic discovery might inform more about low-level drinking than about problematic drinking and AUDs\(^36\). However, we built a PRS for alcohol consumption and successfully used it to predict alcohol-related harms. Due to the robustness of a self-reported single timepoint alcohol consumption estimate and the fact that different alcohol-related traits are to some degree genetically distinct\(^37–41\), it is expected that a PRS developed directly for alcohol-related morbidity will outperform our PRS in predicting alcohol-related health burden. Supporting this assumption, the general pattern is that PRSs are more strongly associated with their respective
diseases than with related phenotypes. Unfortunately, no high-quality summary statistics for alcohol-related harms including both somatic and psychiatric outcomes yet exist; the performed GWAS have only covered AUD and alcohol dependence and been smaller in size than our discovery sample of choice, thus making future efforts for large-scale GWAS discovery based on alcohol-related harms more than necessary.

Our PRS was derived using European ancestry discovery samples and tested in the Finnish population. Its applicability in other populations therefore needs further evaluation as the alcohol-related genetic mechanisms may vary between populations. However, it has to be noted that the PRS derived from a non-Finnish sample performed well in the Finnish dataset, even though Finns are somewhat genetically different from the rest of the Europeans.

Our design allowed us to study outcomes prospectively. Our registry-based follow-up captures alcohol-related outpatient and inpatient visits, withdrawal treatment prescription for alcoholism, and deaths, thus covering major alcohol-related health events over several decades. Nonetheless, some of the milder cases of alcohol-related health problems could have gone undetected.

In conclusion, a PRS for alcohol consumption was associated with elevated risk for incident alcohol-related health events and all-cause mortality. These findings underline the importance of heritable factors driving alcohol-related behavior. A successful attempt to predict alcohol-related health outcomes with a PRS shows promise in possible future utilization of genetic information in risk estimation and prediction of alcohol-related harms.

Acknowledgements
We would like to thank Lea Urpa for proofreading, and Sari Kivikko, Hui-Yi Shen, and Ulla Tuomainen for management assistance. We would like to thank all participants of the study cohorts for their generous participation. SRI was supported by the Academy of Finland Center of Excellence in Complex Disease Genetics (Grant No. 312062), Academy of Finland (Grant No. 285380), the Finnish Foundation for Cardiovascular Research, the Sigrid Juselius Foundation, and University of Helsinki HiLIFE Fellow grant. The funding agencies had no role in the design and conduct of the study; collection, analysis, and interpretation of data; or the writing of the manuscript or the decision to submit it for publication. The FinnGen project is funded by two grants from Business Finland (HUS 4685/31/2016 and UH 4386/31/2016) and nine industry partners (AbbVie, AstraZeneca, Biogen, Celgene, Genentech, GSK, MSD, Pfizer, and Sanofi). Following biobanks are acknowledged for collecting the FinnGen project samples: Aura Biobank (https://www.aurabiobank.fi/), THL Biobank (https://www.thl.fi/web/thl-biobank), Helsinki Biobank (https://www.terveyskyla.fi/helsinginbiopankki/en), Northern Finland Biobank Borealis (https://www.ppsph.fi/Tukimuus-ja-opetus/Biopankki), Finnish Clinical Biobank Tampere (https://www.tays.fi/en-US/Research_and_development/Finnish_Clinical_Biobank_Tampere), Biobank of Eastern Finland (https://ita-suomenbiokliinit.fi/), Central Finland Biobank (https://www.kshp.fi/fi/FI-Potilaille/Biopankki), Finnish Red Cross Blood Service Biobank (https://www.bloodservice.fi/Research%20Projects/Biobanking), The Finnish Twin Cohort Nicotine Addictions Genetics family study has been supported by NIH DA18254 to P.A.F.M., genotyping in the twin cohort by Global Research Awards for Nicotine Dependence (GRAND) funded by Pfizer Inc. to JKa and the Welcome Trust Sanger Institute, and the FinnTwin16 study by NIH AA-12502, AA-00145, and AA-09203 to R.J.R. JKa has been supported by the Academy of Finland (grants 265240, 263278, 308248, 312073). A.S.H. was supported by the Academy of Finland (Grant no. 321356). T.K., JTR, SRu, and P.R were supported by the Doctoral Programme in Population Health, University of Helsinki. JTR was supported by the MD/PhD Program of the Faculty of Medicine, University of Helsinki.

Author details
1 Institute for Molecular Medicine Finland (FIMM), HILIFE, University of Helsinki, Helsinki, Finland. 2 Finnish Institute for Health and Welfare (THL), Helsinki, Finland. 3 Analytic and Translational Genetics Unit, Department of Medicine, Department of Neurology and Department of Psychiatry Massachusetts General Hospital, Boston, MA, USA. 4 The Stanley Center for Psychiatric Research and Program in Medical and Population Genetics, The Broad Institute of MIT and Harvard, Cambridge, Boston, MA, USA. 5 Department of Psychiatry, Washington University School of Medicine in St.Louis, St.Louis, MO, USA. 6 Department of Psychological and Brain Sciences, Indiana University Bloomington, Bloomington, IN, USA. 7 Department of Public Health, Clinicum, Faculty of Medicine, University of Helsinki, Helsinki, Finland. 8 The Broad Institute of MIT and Harvard, Cambridge, MA, USA.

Data availability
The FinnGen data may be accessed through Finnish Biobanks (www.fnbbf.fi) and THL Biobank data through THL Biobank (https://thl.fi/en/web/thl-biobank). Summary statistics from Liu et al. (2019) “Association Studies of up to 1.2 million individuals yield new insights into the genetic etiology of tobacco and alcohol use” published in Nature Genetics are archived at the University of Minnesota library (https://conservancy.umn.edu/handle/11299/20156).

Code availability
The full genotyping and imputation protocol for FinnGen is described at https://doi.org/10.17504/protocols.io.nmndc5e.

Conflict of interest
A.P. reports that part of his salary is received from the FinnGen project that is partially funded by nine industry partners AbbVie, AstraZeneca, Biogen, Celgene, Genentech, GSK, Merck/MSD, Pfizer, and Sanofi. JKa reports grants from Academy of Finland, US-PHS NIH/NIDA, US-PHS NIH/NIAAA, Pfizer Inc./GRAND program grant, and EU FP7 during the conduct of the study. V.S. reports a conference trip and an honorarium for participating in an advisory board meeting from Novo Nordisk (unrelated to the present study) and a grant to his institute for research collaboration from Bayer (unrelated to the present study). No other relationships or activities that could appear to have influenced the submitted work.

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

Supplementary Information accompanies this paper at (https://doi.org/10.1038/s41398-019-0676-2).

Received: 20 July 2019 Revised: 23 October 2019 Accepted: 14 November 2019
Published online: 21 January 2020

References
1. World Health Organization. Global Status Report on Alcohol And Health 2018. (WHO, 2018).
2. Vengelene, V., Bilbao, A., Molander, A. & Spanagel, R. Neuropharmacology of alcohol addiction. Br. J. Pharmacol. 154, 299–315 (2008).
3. Grant, B. F. et al. Epidemiology of DSM-5 alcohol use disorder: results from the National Epidemiologic Survey on Alcohol and Related Conditions III. JAMA Psychiatry 72, 757–766 (2015).
4. Wood, A. M. et al. Risk thresholds for alcohol consumption: combined analysis of individual-participant data for 599,912 current drinkers in 83 prospective studies. Lancet 391, 1513–1523 (2018).
