Drug and Alcohol Dependence 263 (2024) 112415

Contents lists available at ScienceDirect

Drug and Alcohol Dependence

journal homepage: www.elsevier.com/locate/drugalcdep

Short Communication

Polygenic risk scores for nicotine use and family history of smoking are associated with smoking behaviour

Jerome C. Foo [a,b,c,d,1], Maja P. Völker [a,1], Fabian Streit [a,e,f,g], Josef Frank [a], Norman Zacharias [h], Lea Zillich [a,g,i,j,k], Lea Sirignano [a], Peter Nürnberg [l], Thomas F. Wienker [m], Michael Wagner [n,o], Markus M. Nöthen [p], Michael Nothnagel [q], Henrik Walter [r], Bernd Lenz [s], Rainer Spanagel [b], Falk Kiefer [s], Georg Winterer [t,u,v], Marcella Rietschel [a], Stephanie H. Witt [a,g,*,2]

[a] Department of Genetic Epidemiology in Psychiatry, Central Institute of Mental Health, Medical Faculty Mannheim, Heidelberg University, Mannheim, Germany
[b] Institute of Psychopharmacology, Central Institute of Mental Health, Medical Faculty Mannheim, Heidelberg University, Mannheim, Germany
[c] College of Health Sciences, Department of Psychiatry, University of Alberta, Edmonton, Canada
[d] Neuroscience and Mental Health Institute, University of Alberta, Edmonton, Canada
[e] Department of Psychiatry and Psychotherapy, Central Institute of Mental Health, Medical Faculty Mannheim, Heidelberg University, Mannheim, Germany
[f] Hector Institute for Artificial Intelligence in Psychiatry, Central Institute of Mental Health, Medical Faculty Mannheim, Heidelberg University, Mannheim, Germany
[g] German Center for Mental Health (DZPG), partner site Mannheim/Heidelberg/Ulm, Germany
[h] Department of Otorhinolaryngology, Head and Neck Surgery, Charité-Universitätsmedizin Berlin, Freie Universität Berlin, Humboldt-Universität zu Berlin, and Berlin Institute of Health, Campus Benjamin Franklin, Berlin, Germany
[i] Department of Translational Brain Research, Central Institute of Mental Health, Medical Faculty Mannheim, Heidelberg University, Mannheim, Germany
[j] HITBR Hector Institute for Translational Brain Research gGmbH, Mannheim, Germany
[k] German Cancer Research Center (DKFZ), Heidelberg, Germany
[l] Cologne Center for Genomics (CCG), University of Cologne, Faculty of Medicine and University Hospital Cologne, Cologne 50931, Germany
[m] Department of Molecular Human Genetics, Max Planck Institute for Molecular Genetics, Berlin, Germany
[n] Department of Psychiatry and Psychotherapy, University Hospital Bonn, Bonn, Germany
[o] Department of Neurodegenerative Diseases and Geriatric Psychiatry, University Hospital Bonn, Bonn, Germany
[p] Institute for Human Genetics, University Hospital Bonn, Bonn, Germany
[q] Department of Statistical Genetics and Bioinformatics, Cologne Center for Genomics, University of Cologne, Cologne, Germany
[r] Department of Psychiatry and Psychotherapy, Division of Mind and Brain Research, Charité- Universitätsmedizin Berlin, Berlin, Germany
[s] Department of Addictive Behavior and Addiction Medicine, Central Institute of Mental Health, Medical Faculty Mannheim, Heidelberg University, Mannheim, Germany
[t] Department of Anesthesiology and Intensive Care Medicine, Campus Virchow-Klinikum and Campus Charité Mitte, Charité-Universitätsmedizin Berlin, Berlin, Germany
[u] Pharmaimage Biomarker Solutions GmbH, Berlin, Germany
[v] PI Health Solutions GmbH, Berlin, Germany

ABSTRACT

Introduction: Formal genetics studies show that smoking is influenced by genetic factors; exploring this on the molecular level can offer deeper insight into the etiology of smoking behaviours.

Methods: Summary statistics from the latest wave of the GWAS and Sequencing Consortium of Alcohol and Nicotine (GSCAN) were used to calculate polygenic risk scores (PRS) in a sample of ~2200 individuals who smoke/individuals who never smoked. The associations of smoking status with PRS for Smoking Initiation (i.e., Lifetime Smoking; SI-PRS), and Fagerström Test for Nicotine Dependence (FTND) score with PRS for Cigarettes per Day (CpD-PRS) were examined, as were distinct/additive effects of parental smoking on smoking status.

Results: SI-PRS explained 10.56% of variance (Nagelkerke-R^2) in smoking status ($p=6.45x10^{-30}$). In individuals who smoke, CpD-PRS was associated with FTND score ($R^2=5.03\%$, $p=1.88x10^{-12}$). Parental smoking alone explained $R^2=3.06\%$ ($p=2.43\times10^{-12}$) of smoking status, and 0.96% when added to the most informative SI-PRS model (total $R^2=11.52\%$).

Conclusion: These results show the potential utility of molecular genetic data for research investigating smoking prevention. The fact that PRS explains more variance than family history highlights progress from formal to molecular genetics; the partial overlap and increased predictive value when using both suggests the importance of combining these approaches.

* Correspondence to: Central Institute of Mental Health, Dept. of Genetic Epidemiology in Psychiatry, J5, 68159 Mannheim, Germany.
E-mail address: stephanie.witt@zi-mannheim.de (S.H. Witt).
[1] Shared First Authors
[2] ORCID ID: 0000-0002-1571-1468

https://doi.org/10.1016/j.drugalcdep.2024.112415
Received 28 December 2023; Received in revised form 22 July 2024; Accepted 12 August 2024
Available online 15 August 2024

1. Introduction

Smoking behaviours are related to a multitude of diseases including cancer, chronic obstructive pulmonary disease (COPD), and heart disease (U.S. Department of Health and Human Services, 2014). Despite the fact that the adverse effects of smoking on health are well-known and considerable efforts to lower its prevalence have been made by many countries, the global number of individuals who smoke is still rising (Reitsma et al., 2021), leading to increasing preventable mortality worldwide.

Twin and family studies have shown that smoking behaviour is influenced by genetic factors (Li et al., 2003; Sullivan and Kendler, 1999; Vink et al., 2005), and it is thought that a large proportion of variance in smoking-related behaviours is accounted for by genetic effects (Quach et al., 2020). The different stages of smoking (e.g. initiation, regular smoking, nicotine dependence, cessation) have strong genetic components which overlap (6,7,8). Better understanding of the genetic factors involved in the development of smoking is needed to take more effective steps towards prevention.

One method of exploring the influence of genetic load on phenotypes is polygenic risk score (PRS) profiling, which takes genetic variants identified in genome-wide association studies (GWAS) as being associated with a given phenotype and combines them into a PRS that captures part of an individual's susceptibility to developing that phenotype (Lewis and Vassos, 2020; Choi et al., 2020). In polygenic risk scoring, summary results of risk variants and effect sizes identified in GWAS of particular phenotypes are used as "training samples", and used to generate risk scores reflecting the genetic load for the phenotype in an independent "target sample". PRS are beginning to show utility not only in research-based case-control studies, but also at the level of population-based cohort studies (Lewis and Vassos, 2020). While it has been suggested that PRS are beginning to have clinical utility in terms of disease risk prediction (Lewis and Vassos, 2020), for many phenotypes, predictive ability remains modest, and more evidence is still needed on the value of adding PRS to prediction scenarios in clinical settings (Lambert et al., 2019).

Family history (FH) is another measure capturing genetic risk, but also comprises environmental familial influences (Boardman et al., 2010). Several recent studies have started to explore the interplay of genome-wide PRS and family history, finding that for a variety of common diseases, the two measures are complementary, partially independent, and not interchangeable (Mars et al., 2022).

In the present study, as a proof-of-principle, we aimed to explore the influence of genetic load on smoking phenotypes by investigating whether PRS of nicotine phenotypes could predict smoking status and intensity of dependence, and how this compared to prediction using FH. PRS were calculated based on summary statistics from the largest GWAS of nicotine phenotypes (smoking initiation, cigarettes per day) to date (Saunders et al., 2022), which did not include the present sample, and were used to predict smoking status and behaviour in a population-based case-control study of individuals who smoke and individuals who never smoked (n= 2396) (Lindenberg et al., 2011).

2. Methods

2.1. Data collection

Demographic and genetic data from the population-based German multi-center study "Genetics of Nicotine Dependence and Neurobiological Phenotypes" (German Research Foundation, SPP1226) were used. Full details on the data collection in the cohort can be found in Lindenberg et al., 2011 (Lindenberg et al., 2011). Briefly, in 2007–2009, active individuals who smoke (n=1116) and individuals who never smoked (n=1280) aged 18–65 years were recruited from the general population. Exclusion criteria included being a former individual who smoke, alcohol or substance use/dependence, non-German ethnicity, pregnancy, and medical conditions or medication that might interfere with the study.

Classification of individuals who smoke was conducted using DSM-IV criteria (American Psychiatric Association (APA) (APA), 1994), Structured Clinical Interview for DSM Disorders (SCID) (First et al., 1996) and the German version of the Fagerström Test for Nicotine Dependence (FTND) (Lindenberg et al., 2011). FTND scores were available for 2228 individuals (n=1027 individuals who smoke, n=1201 individuals who never smoked). Parental history of smoking was assessed via questionnaire.

2.2. Genotyping and quality control of the target sample

Genetic data was available from 2312 individuals. DNA was extracted from whole blood and prepared using the Qiagen FlexiGene DNA Kit (Qiagen, Hilden, Germany) according to manufacturer protocols. Genotyping was performed using the Infinium OmniExpress Exome Array (Illumina, *San Diego, USA*), which includes a GWAS backbone of ~ 680,000 markers selected to capture the greatest amount of common SNP variation.

Quality control was performed using PLINK 1.90 (Chang et al., 2015). Individuals were excluded for being ancestry outliers (>4.5 standard deviations on any of the first 20 ancestry principal components), relatedness (Pi-HAT>0.2) and person–missing rate ($\geq$ 0.02); Single nucleotide polymorphisms (SNPs) were excluded based on missing rate ($\geq$ 0.02), deviation from Hardy-Weinberg disequilibrium (HWE–P-value $\leq 1\times10^{-6}$), and minor allele frequency (< 0.01).

Imputation of non-typed markers was performed using a locally installed instance of the Michigan Imputation Server with the 1000 Genomes backbone(Das et al., 2016).

The final data set available for PRS analyses contained n=2169 individuals (997 active individuals who smoke, 1172 individuals who never smoked; former individuals who smoke not included) and 7004,791 SNPs.

2.3. Polygenic risk scoring and summary statistics

GWAS meta-analysis summary statistics (European ancestry summary statistics) from the second phase of the GWAS and Sequencing Consortium of Alcohol and Nicotine (GSCAN2) excluding the 23andme subset were used, comprising the phenotypes: 1) smoking initiation (i.e. ever being a individual who smoke vs. never; SI; n=805,431) and 2) cigarettes per day (CpD; n=326,497).

Details on the phenotype definitions in GSCAN2 can be found in Saunders et al., 2022 (Saunders et al., 2022) and at https://genome.psych.umn.edu/images/d/da/GSCAN_GWAS_Phenotype_Definitions-2–24-2016.pdf (Accessed April 10, 2024). Additionally, calculations were performed with GWAS meta-analysis summary statistics from the first phase of GSCAN (GSCAN1, not including 23andMe) (Liu et al., 2019) to compare with results from GSCAN2. The total number of markers for GSCAN2 was 13,595,219 for SI and 13,763,312 for CpD, and the total number of markers for GSCAN1 was 11,802,365 for SI and 12,003,613 for CpD.

PRS were based on posterior effect sizes estimated using the PRS-CS method (Ge et al., 2019). The method uses Bayesian regression models to apply continuous shrinkage priors to the effect sizes to account for LD structure among SNPs. PRS-CS maximizes the number of SNPs included in PRS, as it does not require SNPs to be independent from one another as in other clumping and threshold methods. The LD pattern of the 1000 Genomes European reference panel and the default priors for effect sizes were used. Based on the posterior effect sizes estimated with PRS-CS, scores were calculated in PRSice2 (Choi and O'Reilly, 2019), using the –no-clump option, and excluding the extended MHC region. The number of SNPs included in the PRS from GSCAN2 was 1019,639 for SI and 1019,641 for CpD, in GSCAN1, it was 1014,896 for SI and 1021,339 for CpD.

The individual who smoke/individual who never smoked dataset described above was used as the target sample. Regression analyses were used to test for association of PRS with phenotypes of interest, namely with smoking status (individuals who smoke vs. individuals who never smoked: SI), and FTND score (in individuals who smoke: CpD). Analyses were adjusted for population stratification by including the first 10 ancestry principle components (PCs). R^2 (Nagelkerke pseudo R^2 for categorical phenotypes) was used to estimate explained variability. Model fits were calculated as difference of R^2 of the full model (e.g. Pheno ~ PRS + PCs) and R^2 of the null model (Pheno ~ PCs). Regression analyses were performed in R 3.6.3.

In an additional step, specifying an incremental variable in the regression, we investigated the contribution of family history of smoking (for SI-PRS: parental history of smoking (FH), 0=neither, 1=one, 2=both parents).

In addition, we calculated receiver operating characteristic (ROC) curves and estimated the area under the curve (AUC) to examine the ability to predict smoking status using SI-PRS and FH. Calculations were performed using the caret package.

Generalized Linear Models (GLM) with smoking status as a binary response variable and logit link function were used as prediction models (null model: ~ PCs; model 1: ~ SI-PRS + PCs; model 2: ~ SI-PRS + FH + PCs).

3. Results

3.1. Smoking status

PRS for SI from GSCAN2 explained R^2=10.56% of variance in smoking status ($p=6.45\times10^{-30}$) (Fig. 1A). Quantile plots show the increasing effect of higher SI-PRS (Fig. 1B).

FH alone explained R^2=3.06 % ($p=2.43\times10^{-12}$) of smoking status. In the model combining FH and SI-PRS, FH explained an additional and independent R^2=0.96% ($p=8.50\text{x}10^{-7}$), resulting in a combined R^2 of 11.52%.

Overall, the SI-PRS from GSCAN2 explained more phenotypic variance than GSCAN1 (R^2=10.56% vs. R^2=8.79%, respectively).

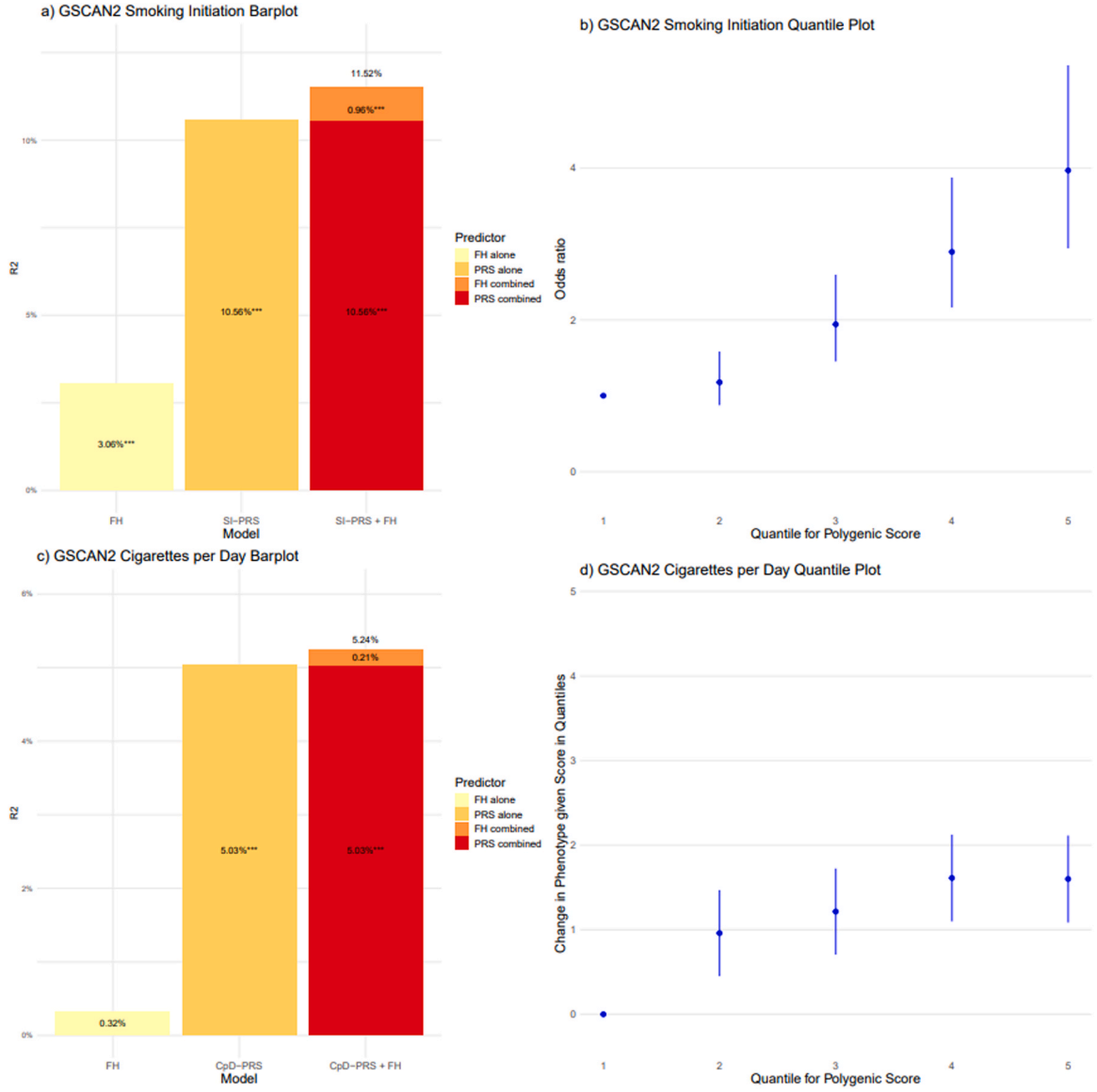

Fig. 1. a) individual who smoke/individual who never smoked Status predicted by PRS for Smoking Initiation and Parental Smoking; b) Quantile plot showing odds ratio for SI-PRS on ever being a individual who smoke/individual who never smoked status; c) FTND score predicted by PRS for Cigarettes per Day and Parental Smoking; d) Quantile plot showing change in FTND score given CpD-PRS. 1 is the reference quantile in all quantile plots.

3.1.1. FTND

PRS for CpD (R^2=5.03 % p=1.88×10^{-12}, Fig. 1c) were significantly associated with FTND scores in individuals who smoke. Quantile plots depict change in phenotype over PRS in quantiles (Fig. 1d). FH alone explained 0.32 % of variance in FTND scores (n.s., p= 0.08). In the combined model with CpD-PRS, FH explained an additional 0.21 % of variance (p = 0.143). More variance in FTND scores was explained with CpD-PRS from *GCSAN1* than *GSCAN2* (R^2= 5.81% vs R^2=5.03% respectively).

3.2. AUC and ROC

The full model including SI-PRS, FH, and PCs had better predictive ability (AUC = 0.68) than the models including PRS-SI and PCs only (AUC = 0.66), and the null model including PCs only (AUC = 0.52). Overall, the predictive value of each model ranged from modest (smoking status ~ SI-PRS + parental smoking + PCs, smoking status ~ SI-PRS + PCs) to low (smoking status ~ PCs) (Fig. 2).

4. Discussion

The present study demonstrates the informativeness of PRS and highlights their potential utility in the prediction of disease risk. We found that PRS for smoking behaviours were significantly associated with smoking status, and also FTND score in individuals who smoke; these associations suggest their importance for use in smoking research as well as the future potential to use PRS for purposes of prediction and preventative measures. History of parental smoking was also significantly associated with PRS for smoking initiation and overlap with the variance explained by PRS was observed. Research has found that PRS are able to predict a reliable but modest amount of complex genetic phenotypes and the amount of variance in smoking initiation behaviour phenotypes explained here is in the range of PRS to predict other disease-related phenotypes (Duncan et al., 2019).

Combining SI-PRS with history of parental smoking enabled a better prediction of smoking status, with results suggesting both independent and overlapping components of contribution of parental smoking and PRS. It is of interest to note that in terms of SI, while FH alone explained 3.06%, the combined model explained up to 11.52%. This highlights the strength of using such a molecular genetic approach, which is more direct than formal genetics. It is also important to note that with the inclusion of the PRS, the amount explained by parental smoking decreased. This overlap may suggest that PRS can cover a substantial share of the contributing genetic factors contained in phenotypic parameters such as parental smoking (i.e., family history), which are overall measures that capture not only genetic contributions but also environmental factors; when looking at 'purely' biological mechanisms, it may be advantageous to prioritize the use of techniques such as PRS. Used together, these approaches will allow more detailed analyses and enable exploration of environment-familial loading as well as dissection of biological pathways involved.

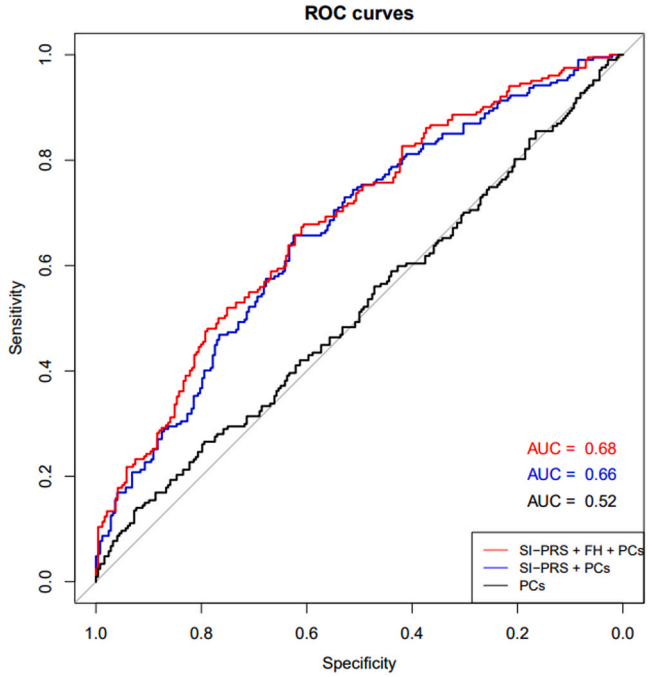

Fig. 2. ROC curves and AUCs of prediction models for smoking status.

We observed that CpD-PRS were associated with FTND score, which points to the ability to identify those who might be at the highest risk of heavy smoking. These results are in line with those of a previous study which used GSCAN summary statistics to generate PRS for amount of smoking in another European cohort (Bray et al., 2021). That study found that in a non-European cohort PRS did not predict smoking behaviours; the inclusion of only individuals with European ancestry also limits the generalizability of the results of the present study.

In the present sample, the results of predictive analyses found that addition of SI-PRS and parental smoking increased AUC (PCs only = 0.52, PCs + PRS = 0.66, PRs + PRS + FH = 0.68). The results are relatable to a previous study by Bray and colleagues (Bray et al., 2021) which examined PRS in smoking cessation, specifying predictive models with clinical and genetic predictors (PRS of smoking behaviors), finding that genetic predictors increased AUC 0.617–0.665 (Bray et al., 2021) (but not including FH in prediction models). Taken together, the results suggest that genetic factors are valuable for predicting smoking behaviors, although we note that even with a large discovery sample, the discriminatory power is limited, and sample size may not be the only factor. We also note that the variance explained by PRS based on GSCAN2 vs GSCAN1 increased for explaining smoking status, but not for FTND score, which may perhaps be related to the binary/continuous nature of the variables.

Towards the application of PRS in clinical and practical contexts, other research has found that PRS are informative not only about disease risk but also about time of occurrence; e.g., age/age of onset (Bray et al., 2021; Mars et al., 2020; Deutsch and Selya, 2020). Recent work has also demonstrated that PRS are significantly associated with trajectories of problematic substance use and may be informative about developmental progression of disease (Deak et al., 2022). It is hoped that PRS may be more powerful and potentially approach practical utility with refined phenotypes and PRS based on better characterized and larger GWAS. It will be important to extend the approach to different study populations (e.g., of other ancestries); the utility of PRS is evolving as novel methods and those for extension to diverse populations are constantly being developed (Kachuri et al., 2024). PRS can also be refined, e.g., by selecting only variants belonging to specific biological pathways to test whether these play a role in susceptibility for disease or treatment, also with the aim to identify specific subpopulations. Well-defined samples such as the present one offer the possibility to identify relevant pathways using the PRS approach.

Our findings, serving as a proof of principle, suggest that including PRS for smoking initiation in the exploration of an individual's risk of becoming a individual who smoke may have the potential to inform prevention and intervention strategies. Combining PRS with FH yielded additional predictive ability, suggesting that the two should be used together. At the same time, the present results suggest that the improvement in predictive power obtained by the use of smoking PRS remains limited, and that they may be currently more well-suited to help understand the genetic bases of smoking and related behavioural traits, than for use in prediction or clinical models.

Funding

This research was supported by the German Research Foundation (DFG: TRR SPP1226, DFG Wa731/8-1, Wi1316/9-1, DFG: TRR265 Heinz et al. (2020) (Project-ID 402170461)), and German Federal Ministry of Education and Research (BMBF: Target-OXY 031L0190A; e:Med SysMedSUDs 01ZX01909). Funders had no involvement in study design, collection, analysis, interpretation of data, writing of the report, or decision to submit the article for publication.

CRediT authorship contribution statement

Bernd Lenz: Writing – review & editing. **Fabian Streit:** Writing – review & editing, Methodology, Formal analysis, Conceptualization. **Henrik Walter:** Investigation, Funding acquisition. **Maja P. Völker:** Writing – review & editing, Visualization, Formal analysis. **Norman Zacharias:** Investigation, Data curation. **Rainer Spanagel:** Writing – review & editing, Resources, Investigation, Funding acquisition, Conceptualization. **Josef Frank:** Writing – review & editing, Methodology, Formal analysis. **Georg Winterer:** Resources, Investigation, Funding acquisition, Conceptualization. **Lea Zillich:** Writing – review & editing. **Falk Kiefer:** Resources, Investigation, Funding acquisition, Conceptualization. **Stephanie H. Witt:** Writing – review & editing, Supervision, Methodology, Investigation, Funding acquisition, Conceptualization. **Peter Nürnberg:** Investigation. **Marcella Rietschel:** Writing – review & editing, Supervision, Methodology, Investigation, Funding acquisition, Conceptualization. **Lea Sirignano:** Writing – review & editing. **Michael Wagner:** Writing – review & editing, Methodology, Investigation. **Thomas F. Wienker:** Conceptualization. **Michael Nothnagel:** Data curation. **Jerome Clifford Foo:** Writing – review & editing, Writing – original draft, Visualization, Methodology, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Markus M Nöthen:** Investigation.

Declaration of Competing Interest

None.

Acknowledgments

This research was supported by the German Research Foundation (DFG: TRR SPP1226, DFG Wa731/8–1, Wi1316/9–1, TRR265 Heinz et al. (2020) (Duncan et al., 2019) (Project-ID 402170461)), and German Federal Ministry of Education and Research (BMBF: Target-OXY 031L0190A, e:Med SysMedSUDs 01ZX01909).

Contributors

Jerome C. Foo, PhD, Maja P. Völker, MSc, Fabian Streit, PhD, Josef Frank, PhD, Norman Zacharias, PhD, Lea Zillich, PhD, Lea Sirignano, MSc, Peter Nürnberg, PhD, Thomas F. Wienker, MD, Michael Wagner, PhD, Markus Nöthen, MD, Michael Nothnagel, PhD, Henrik Walter, MD, PhD, Bernd Lenz, MD, Rainer Spanagel, PhD, Falk Kiefer, MD, Georg Winterer, MD, PhD, Marcella Rietschel, MD, Stephanie H. Witt, PhD

References

American Psychiatric Association (APA) (1994) Diagnostic and Statistical Manual of Mental Disorders, 4th edn, Washington, DC: American Psychiatric Association.

Boardman, J.D., Blalock, C.L., Pampel, F.C., 2010. Trends in the genetic influences on smoking. J. Health Soc. Behav. 51 (1), 108–123.

Bray, M.J., Chen, L.-S., Fox, L., Ma, Y., Grucza, R.A., Hartz, S.M., et al., 2021. Studying the utility of using genetics to predict smoking-related outcomes in a population-based study and a selected cohort. Nicotine Tob. Res 23 (12), 2110–2116.

Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M., Lee, J.J., 2015. Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience 4, 7.

Choi, S.W., Mak, T.S.-H., O'Reilly, P.F., 2020. Tutorial: a guide to performing polygenic risk score analyses. Nat. Protoc. 15 (9), 2759–2772.

Choi, S.W., O'Reilly, P.F., 2019. PRSice-2: polygenic risk score software for biobank-scale data. Gigascience 8, 7.

Das, S., Forer, L., Schönherr, S., Sidore, C., Locke, A.E., Kwong, A., Vrieze, S.I., Chew, E. Y., Levy, S., McGue, M., Schlessinger, D., Stambolian, D., Loh, P.R., Iacono, W.G., Swaroop, A., Scott, L.J., Cucca, F., Kronenberg, F., Boehnke, M., Abecasis, G.R., Fuchsberger, C., 2016. Next-generation genotype imputation service and methods. Nat. Genet 48 (10), 1284–1287 (Oct).

Deak, J.D., Clark, D.A., Liu, M., Schaefer, J.D., Jang, S.K., Durbin, C.E., et al., 2022. Alcohol and nicotine polygenic scores are associated with the development of alcohol and nicotine use problems from adolescence to young adulthood. Addiction 117 (4), 1117–1127.

Deutsch, A.R., Selya, A.S., 2020. Stability in effects of different smoking-related polygenic risk scores over age and smoking phenotypes. Drug Alcohol Depend. 214, 108154.

Duncan, L., Shen, H., Gelaye, B., Meijsen, J., Ressler, K., Feldman, M., et al., 2019. Analysis of polygenic risk score usage and performance in diverse human populations. Nat. Commun. 10 (1), 3328.

First M.B., Spitzer R.L., Gibbon M., Williams J.B.W. (1996) Structured Clinical Interview for DSM-IV Axis I Disorders Research Version (SCID-1). New York: Biometrics Research.

Ge, T., Chen, C.Y., Ni, Y., Feng, Y.A., Smoller, J.W., 2019. Polygenic prediction via Bayesian regression and continuous shrinkage priors. Nat. Commun. 10 (1), 1776. Apr 16.

Heinz, A., Kiefer, F., Smolka, M.N., Endrass, T., Beste, C., Beck, A., et al., 2020. Addiction Research Consortium: losing and regaining control over drug intake (ReCoDe)-From trajectories to mechanisms and interventions. Addict. Biol. 25 (2), e12866.

Kachuri, L., Chatterjee, N., Hirbo, J., Schaid, D.J., Martin, I., Kullo, I.J., Kenny, E.E., Pasaniuc, B., Polygenic Risk Methods in Diverse Populations (PRIMED) Consortium Methods Working Group, Witte, J.S., Ge, T., 2024. Principles and methods for transferring polygenic risk scores across global populations. Nat. Rev. Genet 25 (1), 8–25. Jan.

Lambert, S.A., Abraham, G., Inouye, M., 2019. Towards clinical utility of polygenic risk scores. Hum. Mol. Genet 28 (R2). R133-R42.

Lewis, C.M., Vassos, E., 2020. Polygenic risk scores: from research tools to clinical instruments. Genome Med 12 (1), 44.

Li, M.D., Cheng, R., Ma, J.Z., Swan, G.E., 2003. A meta-analysis of estimated genetic and environmental effects on smoking behavior in male and female adult twins. Addiction 98 (1), 23–31.

Lindenberg, A., Brinkmeyer, J., Dahmen, N., Gallinat, J., de Millas, W., Mobascher, A., et al., 2011. The German multi-centre study on smoking-related behavior-description of a population-based case-control study. Addict. Biol. 16 (4), 638–653.

Liu, M., Jiang, Y., Wedow, R., Li, Y., Brazel, D.M., Chen, F., et al., 2019. Association studies of up to 1.2 million individuals yield new insights into the genetic etiology of tobacco and alcohol use. Nat. Genet 51 (2), 237–244.

Mars, N., Koskela, J.T., Ripatti, P., Kiiskinen, T.T.J., Havulinna, A.S., Lindbohm, J.V., et al., 2020. Polygenic and clinical risk scores and their impact on age at onset and prediction of cardiometabolic diseases and common cancers. Nat. Med 26 (4), 549–557.

Mars, N., Lindbohm, J.V., Della Briotta Parolo, P., Widén, E., Kaprio, J., Palotie, A., FinnGen, Ripatti, S., 2022. Systematic comparison of family history and polygenic risk across 24 common diseases. Am. J. Hum. Genet 109 (12), 2152–2162. Dec 1.

Quach, B.C., Bray, M.J., Gaddis, N.C., Liu, M., Palviainen, T., Minica, C.C., et al., 2020. Expanding the genetic architecture of nicotine dependence and its shared genetics with multiple traits. Nat. Commun. 11 (1), 5562.

Reitsma, M.B., Flor, L.S., Mullany, E.C., Gupta, V., Hay, S.I., Gakidou, E., 2021. Spatial, temporal, and demographic patterns in prevalence of smoking tobacco use and initiation among young people in 204 countries and territories, 1990-2019. Lancet Public Health 6 (7), e472–e481 (Jul).

Saunders, G.R.B., Wang, X., Chen, F., et al., 2022. Genetic diversity fuels gene discovery for tobacco and alcohol use. Nature 612, 720–724.

Sullivan, P.F., Kendler, K.S., 1999. The genetic epidemiology of smoking. Nicotine Tob. Res 1. Suppl 2:S51-7; discussion S69-70.

U.S. Department of Health and Human Services. The Health Consequences of Smoking—50 Years of Progress: A Report of the Surgeon General. Atlanta: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health, 2014 (Accessed 2024 July 20).

Vink, J.M., Willemsen, G., Boomsma, D.I., 2005. Heritability of smoking initiation and nicotine dependence. Behav. Genet 35 (4), 397–406.