# scientific reports

OPEN

Genetic regulation of spermine oxidase activity and cancer risk: a Mendelian randomization study

João Fadista[1,2,3 ✉], Victor Yakimov[4,7], Urmo Võsa[5], Christine S. Hansen[4,7], Silva Kasela[5], Line Skotte[1], Frank Geller[1], Julie Courraud[6,7], Tõnu Esko[5,11], Viktorija Kukuškina[5], Alfonso Buil[7,8], Mads Melbye[9,10], Thomas M. Werge[7,8], David M. Hougaard[4,7], Lili Milani[5], Jonas Bybjerg-Grauholm[4,7], Arieh S. Cohen[6,7] & Bjarke Feenstra[1 ✉]

Spermine oxidase (SMOX) catalyzes the oxidation of spermine to spermidine. Observational studies have reported SMOX as a source of reactive oxygen species associated with cancer, implying that inhibition of SMOX could be a target for chemoprevention. Here we test causality of SMOX levels with cancer risk using a Mendelian randomization analysis. We performed a GWAS of spermidine/spermine ratio to identify genetic variants associated with regulation of SMOX activity. Replication analysis was performed in two datasets of *SMOX* gene expression. We then did a Mendelian randomization analysis by testing the association between the SMOX genetic instrument and neuroblastoma, gastric, lung, breast, prostate, and colorectal cancers using GWAS summary statistics. GWAS of spermidine/spermine ratio identified *SMOX* locus ($P = 1.34 \times 10^{-49}$) explaining 32% of the variance. The lead SNP rs1741315 was also associated with *SMOX* gene expression in newborns ($P = 8.48 \times 10^{-28}$) and adults ($P = 2.748 \times 10^{-8}$) explaining 37% and 6% of the variance, respectively. Genetically determined SMOX activity was not associated with neuroblastoma, gastric, lung, breast, prostate nor colorectal cancer ($P > 0.05$). A PheWAS of rs1741315 did not reveal any relevant associations. Common genetic variation in the *SMOX* gene was strongly associated with SMOX activity in newborns, and less strongly in adults. Genetic down-regulation of SMOX was not significantly associated with lower odds of neuroblastoma, gastric, lung, breast, prostate and colorectal cancer. These results may inform studies of SMOX inhibition as a target for chemoprevention.

Abbreviations

SMOX	Spermine oxidase
GWAS	Genome-wide association study
eQTL	Expression quantitative trait locus
mQTL	Methylation quantitative trait locus
PheWAS	Phenome-wide association study
ODC	Ornithine decarboxylase
H_2O_2	Hydrogen peroxide
3-AP	3-Aminopropanal
ROS	Reactive oxygen species
IHPS	Infantile hypertrophic pyloric stenosis
EGCUT	Estonian Genome Center University of Tartu

[1]Department of Epidemiology Research, Statens Serum Institut, Artillerivej 5, 2300 Copenhagen, Denmark. [2]Department of Clinical Sciences, Lund University Diabetes Centre, Malmö, Sweden. [3]Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki, Finland. [4]Section of Neonatal Genetics, Danish Centre for Neonatal Screening, Department of Congenital Diseases, Statens Serum Institut, Copenhagen, Denmark. [5]Estonian Genome Centre, Institute of Genomics, University of Tartu, Tartu, Estonia. [6]Section of Clinical Mass Spectrometry, Danish Center for Neonatal Screening, Statens Serum Institut, Copenhagen, Denmark. [7]The Lundbeck Foundation Initiative for Integrative Psychiatric Research, Aarhus, Denmark. [8]Institute of Biological Psychiatry, Mental Health Center Sct. Hans, Roskilde, Denmark. [9]Department of Clinical Medicine, University of Copenhagen, Copenhagen, Denmark. [10]Department of Medicine, Stanford University School of Medicine, Stanford, CA, USA. [11]Program in Medical and Population Genetics, Broad Institute, Cambridge, MA, USA. ✉email: jpsf@ssi.dk; fee@ssi.dk

YOB	Year of birth
GA	Gestational age
RNA-seq	RNA sequencing
TMM	Trimmed mean of M
MDS	Multidimensional scaling
PC	Principal component
BW	Birth weight
bp	Base pairs
LD	Linkage disequilibrium

Although polyamines are essential for normal cell growth and development[1,2], dysregulation of polyamine metabolism is involved in tumorigenesis[3], and hence recognized as a potential target for chemotherapy and chemoprevention[4]. Already in the 1960s, ornithine decarboxylase (ODC), the first rate-limiting enzyme in polyamine biosynthesis, was demonstrated to be at high levels in human cancer specimens[5]. Moreover, a genetic variant in ODC, regulating its enzymatic activity, confirmed the role of this enzyme in human colon cancer risk[6,7], and ODC levels have been shown to be elevated in human skin, breast and prostate cancer specimens[8–10]. Other polyamine metabolic enzymes have not yet been reported to be associated with tumorigenesis in humans.

Spermine oxidase (*SMOX*), a member of the mammalian polyamine catabolic pathway, encodes an enzyme with cytoplasmic and nuclear expression in most tissues[11], that catalyzes the oxidation of spermine to spermidine with the production of hydrogen peroxide (H_2O_2) and 3-aminopropanal (3-AP)[12,13]. SMOX has been reported as a source of induced reactive oxygen species (ROS) associated with neuroblastoma, gastric, lung, breast, prostate and colon cancers[14–20], implying that inhibition of SMOX could be a target for chemoprevention[3]. However, so far, studies of SMOX inhibition have been observational in their nature with no direct inference on causation[14–20]. Genome-wide association studies (GWAS) of SMOX activity are also lacking. A GWAS could potentially detect genetic determinants of SMOX activity that might serve as instruments to assess the causality of SMOX on cancer using a Mendelian randomization framework[21].

To our knowledge, there has not been any Mendelian randomization study assessing the potential causal relationship between SMOX activity and risk of any type of cancer. In this study, we present the first GWAS of spermidine/spermine ratio as a proxy for SMOX activity. We then use the genetic determinants of SMOX activity to perform Mendelian randomization analysis to assess causal relationships between SMOX activity and risk of neuroblastoma, gastric, lung, breast, prostate and colon cancer. In addition, we perform a phenome-wide association study (PheWAS) to query possible deleterious effect of altered SMOX activity through all disease categories and to test possible pleiotropic effects.

Methods

Study population.   The GWAS of spermidine/spermine ratio reflecting SMOX activity, was based on dried blood spot samples taken during routine newborn screening from 534 individuals of Danish ancestry, recruited in a matched case–control study of infantile hypertrophic pyloric stenosis (IHPS), as previously described[22,23] (SSI-IHPS cohort). In addition, dried blood spot samples of 262 newborns of Danish ancestry were included for expression quantitative trait locus (eQTL) analysis (PSYCH-twin cohort). These 262 newborns were taken at random from each of 262 monozygotic twin pairs discordant for later diagnosis of psychiatric disorders as previously described[24]. Of these 262 newborns, 235 were included for methylation quantitative trait loci (mQTL) analysis. Furthermore, 671 whole blood samples taken from adults of the Estonian Genome Center University of Tartu (EGCUT) cohort[25] were also used for eQTL (N = 508)[26,27] and mQTL (N = 305; overlap with gene expression dataset 141 samples) lookups. The Danish Scientific Ethics Committee, the Danish Data Protection Agency and the Danish Neonatal Screening Biobank Steering Committee approved the PSYCH-twins study as well as the GWAS of metabolites in newborns. Usage of EGCUT RNA-seq dataset was approved by Estonian Committee on Bioethics and Human Research, protocol nr. 1.1–12/624, data extraction nr. N26.

The study also included GWAS summary statistics from four cancer-specific consortia cohorts: The ELLIPSE consortium (prostate cancer)[28] with 79,194 cases and 61,112 controls; the BCAC consortium (breast cancer)[29] with 62,533 cases and 60,976 controls; the TRICL consortium (lung cancer)[30] with 29,266 cases and 56,450 controls, and the North American-based Children's Oncology Group (neuroblastoma)[31] with 2,101 cases and 4,202 controls. GWAS summary statistics for gastric and colorectal cancers relied on population-based cohorts: UK Biobank with 5,693 colorectal cancer cases and 386,740 controls[32], and the BioBank Japan Project with 6,563 gastric cancer cases and 195,745 controls[33]. Analyses were restricted to individuals of European ancestry, except for gastric cancer due to the lack of publicly available European gastric cancer GWAS data.

Patients or the public were not involved in the design, or conduct, or reporting, or dissemination plans of our research.

Metabolite measurements and GWAS.   The metabolites spermidine and spermine were quantified in whole blood of dried blood spots using the AbsoluteIDQ® p400 Kit (Biocrates Life Sciences AG, Innsbruck, Austria) (Supplementary Table S3, Figures S10-S12). The AbsoluteIDQ® p400 Kit allows quantification of 408 metabolites which were described in our previous study[23]. In our previous study[23], we quantified 148 of the 408 metabolites, from which spermidine and spermine measurements were also used for this study. Briefly, since all samples were run sequentially, the only batch effect we noticed was plate effect, for which we corrected this by dividing the raw concentration values of all samples by a plate-specific correction factor[23]. We also did metabolite filtering by selecting the 148 metabolites that were above the limit of detection (LOD) in at least 80% of the samples and had a coefficient of variation of the replicates below 25%[23]. In this study we performed genome-

wide association scans of all these 148 metabolites and their biologically relevant ratios (e.g. direct substrate and product of an enzyme), including spermidine to spermine ratio, which had the most significant genome-wide hit of all the metabolites and ratios tested. Our GWAS p-value threshold, correcting for the 148 metabolites tested, is $P < 5 \times 10^{-8}/148 = 3.4 \times 10^{-10}$. For computing time issues, the GWAS analysis was first implemented in PLINK[34] and then repeated in R[35] only for the genome-wide significant variants. The PLINK[34] and R[35] analyses were done using the following linear model: spermidine/spermine inverse normally transformed concentration ratios ~ SNP + sex + YOB + GA + Parity + IHPS. YOB pertains to year of birth, GA as gestational age in weeks, Parity as number of completed pregnancies, and IHPS as case/control status. Of note, the GWAS analysis was also repeated in R[35] in order to obtain the variance explained (R^2) of the genome-wide significant variants and to allow YOB to be coded as a factor. As described before[22], the SSI-IHPS cohort was comprised of Danish ancestry and unrelated individuals of which DNA samples were array-genotyped with the Illumina Multi-Ethnic Global_v2_A2 array, and after genotyping QC, unobserved genotypes were imputed from the Haplotype Reference Consortium panel[36]. Altogether, 6,846,507 SNPs with minor allele frequency > = 1% and imputation info score > = 0.8 were used in the analysis. Being unrelated and of Danish ancestry was determined from the genome-wide SNP array data as previously described[22], using the kinship and principal component analysis, respectively. The GWAS summary statistics of spermidine/spermine ratio of the SSI-IHPS newborn cohort is available here: danishnationalbiobank.com/gwas.

SMOX eQTL and mQTL identification. To replicate the genetic association with SMOX activity at the gene expression level, eQTL analysis was performed in whole blood from two additional data sets of 262 newborns (a subset of the PSYCH-twin cohort) and 508 adults (a subset of the EGCUT cohort) that had previously been RNA-sequenced and GWAS-genotyped[24,26,27]. Preprocessing of the EGCUT expression data is described elsewhere[26]. Briefly, in EGCUT, gene expression matrix was normalized by trimmed mean of M values (TMM)[37], log2 transformed, gene expression values were centered and scaled, and genetic PCs were regressed out[26]. In addition, in order to remove non-genetic variance from the data, up to 20 gene expression-based PCs which were not associated with genetics were regressed out from the residuals of the expression data. Finally, inverse normal transformation was applied on the residuals from the previous step. In order to detect a possible age-dependent eQTL effect, the adult cohort (EGCUT) was stratified into different age groups (10-years intervals). Formal interaction analysis was also performed, in R[35], between donor age at blood draw and rs1741315 genotype on SMOX expression.

Moreover, 235 out of 262 individuals from the newborn PSYCH-twin cohort[24] also had genome-wide methylation array data from the Infinium HumanMethylation450 BeadChip (450 k) array, as previously described[24]. For this study, all samples were filtered by Call Rate > 0.99 (for detection P < 0.01 and bead count > 2) and median methylation and unmethylation intensity signal > 2000 to ensure proper sample measurement quality. All probes (CpGs) were filtered by Call Frequency > 0.99 (for detection P < 0.01 and bead count > 2) to ensure sufficient probe representation across most samples. The data was normalized with the noob background dye correction method after sample filtering[38], implemented in the R[35] package minfi. For all the 480 CpG sites within 1 Mb of the *SMOX* gene, we tested their methylation fraction against the lead (most significant) *SMOX* SNP. The analysis was implemented in R[35] as the following linear model: CpG methylation fraction ~ lead *SMOX* SNP + sex + YOB + GA + BW (birth weight). An association was deemed significant if P-value $< 0.05/480 = 1.04 \times 10^{-4}$ to correct for the number of CpGs tested. Furthermore, 305 individuals from the adult EGCUT cohort also had genome-wide methylation array data from the Infinium HumanMethylation450 BeadChip (450 k) array. Briefly, methylation data was normalized according to the GoDMC (http://www.godmc.org.uk/) pipeline[39] using functional normalization method as implemented in the R package meffil[40]. In step 1, we adjusted normalized methylation proportion for age, sex, predicted cell counts, predicted smoking and genetic PCs. Predicted smoking was performed as the methylation fraction of known CpG sites that are well correlated with smoking status[41]. In step 2, PCA was applied on the residuals from the previous step, and top non-genetic PCs were regressed out. To be able to estimate the SNPxAge effect on methylation, inverse normally transformed methylation proportions were regressed against covariates from step 1 and methylation PCs from step 2.

Variance explained, Mendelian randomization, and PheWAS analyses. The lead SNP located on chromosome 20q12 (in cis with *SMOX*) being genome-wide significantly ($P < 5 \times 10^{-8}$) associated with spermidine/spermine ratio was used as the genetic instrument in the Mendelian randomization analyses. Variance explained by the genetic determinant of SMOX activity was calculated as the adjusted R^2 of lm function in R[35] implemented as spermidine/spermine inverse normally transformed concentrations ratio ~ SNP, or normalized and inverse normally transformed gene expression value of *SMOX* ~ SNP. The Wald ratio method[42], implemented in the R[35] package MendelianRandomization (v0.3.0) was used to test whether the association between SMOX genotype and incident cancers could be caused by differences in SMOX activity (in the SSI-IHPS newborn cohort) or *SMOX* gene expression (in the EGCUT adult cohort). To confirm the results of the MR Wald ratio test, we also used another implementation of this method from the TwoSampleMR (v0.5.6) R[35] package. The primary analysis measured the association between genetically determined SMOX activity and the risk of neuroblastoma, gastric, lung, breast, prostate and colorectal cancers. The PhenoScanner bioinformatic tool[43], a curated database of publicly available results from large-scale genetic association studies, was used for the PheWAS analysis and detect possible concurrent risk factors of the investigated cancer types (pleiotropy). The PhenoScanner database was queried from the MendelianRandomization R[23] package. In our main analyses we tested six key outcomes, namely neuroblastoma, gastric, lung, breast, prostate and colorectal cancers. A P-value of less than 0.05/6 (number of cancers tested) = 0.008 was considered statistically significant. Of note, Table S1 contains rs1741315 allele frequencies for reference populations relevant to our study.

Pos	rsID	Cohort	Analysis	N	EAF	Beta	SE	P Value	R^2
20:4,155,948	rs1741315	Newborns (SSI-IHPS)	SMOX activity	534	0.458	−0.757	0.046	1.34×10^{-49}	0.318
		Newborns (PSYCH-twins)	*SMOX* expr	262	0.471	−0.830	0.062	2.47×10^{-30}	0.366
			cg07472708	235		−0.012	0.003	4.63×10^{-05}	0.005
		Adults (EGCUT)	*SMOX* expr	508	0.452	−0.366	0.065	2.75×10^{-08}	0.057
			cg07472708	305		−0.062	0.069	0.371	0.002

Table 1. SNP rs1741315 associations with SMOX activity, *SMOX* gene expression and cg07472708 methylation fraction. SMOX activity and *SMOX* gene expression estimates were based on inverse normally transformed values (Methods). SNP rs1741315 'A' allele is the effect allele. Pos, genomic position of rs1741315 (hg19); rsID, variant identifier; EAF, effect allele frequency; N, sample size; Beta, effect estimate; SE, standard error; R^2, variance explained of rs1741315 on SMOX activity, *SMOX* gene expression and cg07472708 methylation fraction. An increased spermidine/spermine ratio represents increased SMOX activity.

				SSI-IHPS instrument		EGCUT instrument	
Outcome	Cases, No	Controls, No	EAF	Odds ratio (95% CI)	P	Odds ratio (95% CI)	P
Prostate cancer	79,194	61,112	0.430	0.98 (0.96–1.00)	0.046	0.96 (0.92–1.00)	0.046
Breast cancer	62,533	60,976	0.434	0.99 (0.96–1.02)	0.465	0.98 (0.93–1.03)	0.465
Lung cancer	29,266	56,450	0.442	0.97 (0.94–1.00)	0.084	0.95 (0.89–1.01)	0.084
Colorectal cancer	5,693	386,740	0.439	0.96 (0.91–1.01)	0.097	0.92 (0.83–1.02)	0.097
Gastric cancer	6,563	195,745	0.361	1.02 (0.97–1.07)	0.543	1.03 (0.93–1.14)	0.543
Neuroblastoma	2,101	4,202	0.412	1.07 (0.97–1.19)	0.182	1.16 (0.93–1.44)	0.182

Table 2. Associations between genetically predicted SMOX activity from SSI-IHPS newborn cohort (SSI-IHPS instrument) and *SMOX* expression from the EGCUT adult cohort (EGCUT instrument), based on SNP rs1741315, and six site-specific cancers (prostate, breast, lung, colorectal and neuroblastoma). All estimations were based on the Wald ratio method, with A being the effect allele. EAF, effect allele frequency in each cohort (cases and controls combined); P, p-value. An increased spermidine/spermine ratio represents increased SMOX activity.

To study the possible causal effects of reticulocyte count, lymphocyte count, and hemoglobin concentration on the cancers tested, MR-Base[44] platform was used with default settings (accessed 29 April 2021). Briefly, genome-wide significant hits of GWAS of reticulocyte count[45], lymphocyte count[45] and hemoglobin concentration[45] were used as genetic instruments (clumping was used to prune SNPS for LD) of those exposures to test their causal role on the six cancers tested.

All methods were carried out in accordance with relevant guidelines and regulations.

Ethical approval and consent to participate. The Danish Scientific Ethics Committee, the Danish Data Protection Agency and the Danish Neonatal Screening Biobank Steering Committee approved the PSYCH-twins study as well as the GWAS of metabolites in newborns. The Danish Scientific Ethics Committee granted exemption from obtaining informed consent from participants as this research project was based on genotyping samples from biobank material (H-4-2013-055). Usage of EGCUT RNA-seq dataset was approved by Estonian Committee on Bioethics and Human Research, protocol nr. 1.1-12/624, data extraction nr. N26. Patients or the public were not involved in the design, or conduct, or reporting, or dissemination plans of our research.

**Consent for publication.** All authors consent this study for publication.

Results
Participant characteristics. A total of 534 newborn participants (SSI-IHPS cohort) were included in the GWAS of spermidine/spermine ratio to find the genetic instrument of SMOX activity (Methods). An additional 262 newborns (PSYCH-twin cohort) and 508 adults (EGCUT cohort) were included to confirm the genetic instrument of SMOX activity at the gene expression level (Table 1 and Methods). A total of 950,575 participants, including 79,194 participants who had prostate cancer, 62,533 with breast cancer, 29,266 with lung cancer, 5,693 with colorectal cancer, 6,563 with gastric cancer and 2,101 with neuroblastoma were included as the cancer outcome cohorts (Table 2 and Methods). Demographic characteristics of each cohort are described in the original studies[22–24,26,28–30,32,33].

SMOX activity GWAS. We performed a GWAS of spermidine/spermine ratio, with one locus reaching genome-wide significance, after correction for the 148 metabolites tested ($P < 5 \times 10^{-8}/148 = 3.4 \times 10^{-10}$) (Fig. 1, Methods). This locus was on chromosome 20q12, with the most significant SNP rs1741315 ($P = 1.34 \times 10^{-49}$)
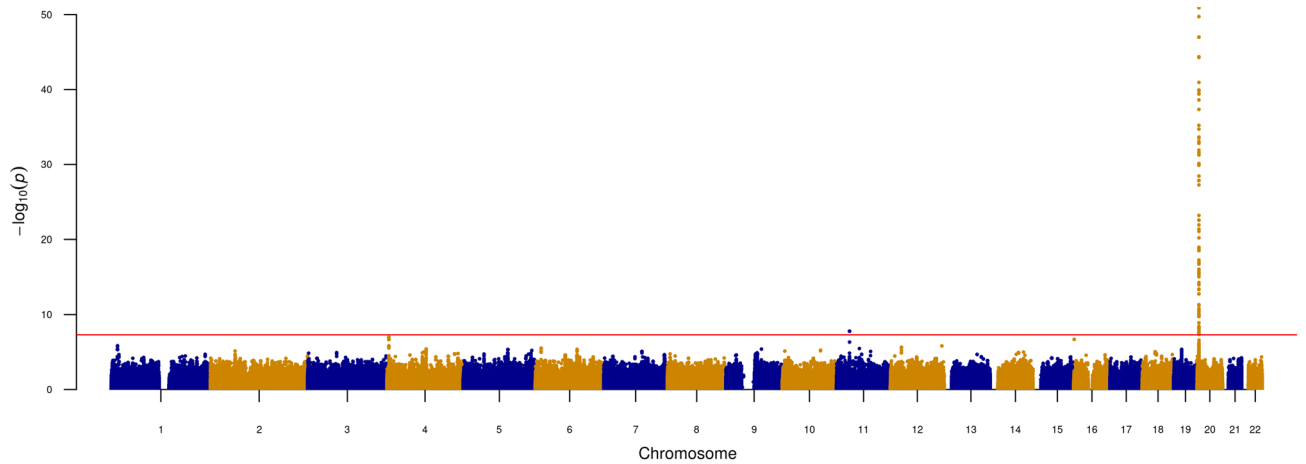
Figure 1. Genetic determinants of SMOX activity. Manhattan plot of the genome-wide association study (GWAS) of spermidine/spermine ratio (534 samples; SSI-IHPS cohort).
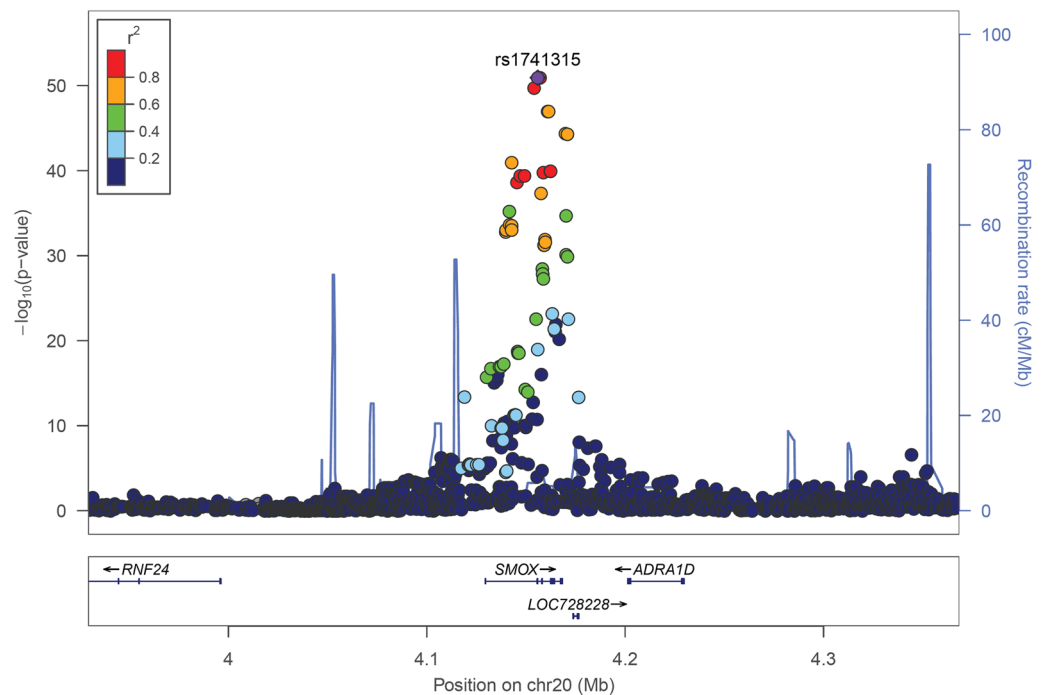
Figure 2. Regional association plot of the new genome-wide significant loci for spermidine/spermine ratio at the SMOX locus. Color-coded linkage disequilibrium (LD) is shown for the top SNP rs1741315 (LD determined with 1000Genomes EUR population). The x-axis represents the genomic region (hg19 assembly) surrounding 200 kb of SMOX gene, while the y-axis represents the strength of the association in $-\log10$(P-value).

being an intronic variant in the *SMOX* gene (Fig. 2). This lead cis-acting SNP explained 32% of the variance in the spermidine/spermine ratio (Figure S9, Table 1), which is regarded as a powerful genetic instrument of SMOX activity for the Mendelian randomization analysis. Of note, there was no sex-difference in SMOX activity (Wilcoxon rank sum test P = 0.10, Figure S8).

SMOX eQTL and mQTL analyses.   To assess effects of rs1741315 on SMOX activity at the gene expression level, we performed eQTL analyses in two additional cohorts. In a newborn cohort of 262 individuals (PSYCH-twin)[24] rs1741315 was associated with *SMOX* gene expression (P = 2.47×10^{-30}) explaining 37% of the variance (Table 1). In a cohort of 508 adults (EGCUT)[26], rs1741315 was also associated with *SMOX* gene expression (eQTL P = 2.75×10^{-8}) explaining 6% of the variance (Table 1, Figure S1). In the eQTLGen Consortium (https://www.eqtlgen.org/), which the EGCUT cohort is a part of, rs1741315 is also associated with *SMOX* gene expression (eQTL P = 3.27×10^{-310}). Of note, rs1741315 was array-genotyped in all cohorts, in contrast to being

just imputed. We stratified the adult cohort (EGCUT) into different age groups (10-years intervals) in order to detect a possible age-dependent eQTL effect in early vs. late adulthood (Figure S2), but found no significant interaction between donor age at blood draw and rs1741315 genotype (P = 0.50) on *SMOX* expression (Figure S3). Due to some missing values in the data, this age-stratified analysis was performed on 494 out of the total 508 adults. Moreover, 235 out of 262 individuals from the newborn PSYCH-twin cohort[24] had genome-wide methylation array data (Methods). Zooming in on CpGs at the SMOX locus, we found rs1741315 to be associated with the methylation proportion of the CpG cg07472708 (mQTL P = 4.63×10^{-5}) explaining 0.5% of the variance in methylation proportion for this CpG (Table 1). Contrary to the newborn data, we did not find rs1741315 to be associated with the methylation proportion of the CpG cg07472708 (mQTL P = 0.371) in 305 individuals from the adult EGCUT cohort who also had genome-wide methylation array data (Figure S4, Table 1 and Methods). Similar to the eQTL data, we also stratified the adult cohort (EGCUT) into different age groups (10-years intervals) in order to detect a possible age-dependent mQTL effect (Figure S5), but found no significant interaction between donor age at blood draw and rs1741315 genotype (P = 0.48) on cg07472708 methylation levels (Figure S6). The CpG cg07472708 is located in the same *SMOX* intron as rs1741315 and is under a region of transcription factor binding site for the transcription factor GATA-1 (Methods). It should be noted that the 50-mer probe containing cg07472708 overlaps with rs1741317 (37 bp apart from the C nucleotide of cg07472708), a high LD ($r^2 = 0.996$) SNP with our lead rs1741315. This could lead to a technical artifact from poorer measurement conditions in those individuals carrying the minor allele, so the results from mQTL analysis should be interpreted with caution.

PheWAS association of SMOX genetic instrument.

One of the assumptions of Mendelian randomization is that genetic variants should only affect the outcome through their effect on the risk factors, i.e. no pleiotropic effects[46,47]. To test if this assumption holds, we checked the association of the SMOX genetic instrument rs1741315 against the PhenoScanner database of genotype–phenotype associations[43] at P < 1×10^{-5} (to correct for the 2998 traits/diseases tested in the database). With rs1741315 'A' allele being the effect allele, we detected reticulocyte count (β = -0.150; 95% CI: -0.157, -0.143; P = 4.12×10^{-285}), lymphocyte count (β = 0.016; 95% CI: 0.009, 0.023; P = 6.44×10^{-6}) and hemoglobin concentration (β = 0.016; 95% CI: 0.009, 0.023; P = 8.05×10^{-6}) as traits associated with this variant[45]. One potential way to test and adjust for pleiotropic effects through associations with these blood traits would be by multivariable Mendelian randomization analysis[48]. However, such analysis is only possible for multi-SNP genetic instruments, which we did not have. Moreover, running MR using genetics instruments for reticulocyte count[45], lymphocyte count[45], and hemoglobin concentration[45] as exposures for cancer, provided no evidence of these potential confounders being causal for the cancers tested (Table S4, Methods).

Association of SMOX genetic instrument with cancer risk.

By using the effect estimate form the SSI-IHPS newborn cohort, we found little evidence that the genetic instrument of SMOX activity, rs1741315, was associated with any of the five non-pediatric cancers evaluated (Table 2). Although genetically lower levels of SMOX were associated with slightly lower risk of prostate cancer at P = 0.047, this finding should be interpreted with caution as it becomes non-significant when correcting for the six different cancers tested.

Since the SMOX genetic instrument, rs1741315, explained 5 to 6 times more variance of *SMOX* expression in newborns than in adults, we used MR to test the causal association between *SMOX* expression and neuroblastoma, a pediatric cancer in which SMOX has been reported to play a role[19,20]. Genetically lower levels of SMOX did not associate with lower risk of developing neuroblastoma[31] (OR = 0.95; 95% CI:0.88, 1.03; P = 0.182) (Table 2). We repeated the MR analysis for all 6 cancers tested with the effect estimates from the EGCUT adult cohort, and similar non-significant associations were obtained as for the SSI-IHPS newborn cohort (Table 2).

Discussion

We conducted the first GWAS of spermidine/spermine ratio to identify genetic variants regulating spermine oxidase (SMOX) activity. Variants in the *SMOX* gene explained a large proportion of the variance in this ratio in newborns and were eQTLs for *SMOX* expression in newborns as well as adults. We did not find genetically determined SMOX activity to be associated with risk of either pediatric (neuroblastoma) or adult cancers (gastric, lung, breast, prostate, and colorectal cancer).

Observational studies have reported elevated SMOX levels in gastric, lung, breast, prostate, and colorectal cancers[14–18], and it has been hypothesized that SMOX inhibition could be an effective target for chemoprevention[3]. Our results do not indicate that genetically driven variation in SMOX activity plays a major role in cancer risk. Several factors and limitations of our study might explain this discrepancy between the observational epidemiological evidence[14–18] and our Mendelian randomization results.

First, our results indicate that the genetic instrumental variable more accurately reflects SMOX activity in infancy, compared with later life, where most cancers occur. This variability in the degree to which *SMOX* expression is genetically regulated over the life course might violate an important MR assumption of monotonicity or homogeneity[49]. However, we also used the estimate of SMOX activity from the EGCUT adult cohort for the MR analysis and found similar results as when using the estimate of SMOX activity from the newborn cohort (Table 2).

Second, the positive association between observed SMOX levels and cancer in observational studies might be driven by reverse causation, in which cancer in an individual could cause SMOX to be elevated due to inflammation.

Third, elevated SMOX levels in cancer could also be due to environmental factors not captured by genetics, e.g. short-term pharmacological changes, induced by SMOX inhibitors. Our genetic instrument was developed

based on normal range SMOX activity data, thus additional genetic variants might play a role in aberrant expression of this enzyme.

Fourth, despite we have corrected for disease status, half the samples from the discovery SSI-IHPS cohort were selected to be IHPS cases and are therefore not representative of the general population. However, the association of rs1741315 with spermidine/spermine ratio did not differ between IHPS cases and controls (Figure S7).

Fifth, the instrumental variable SNP is also associated with hematological traits[45] that could potentially affect cancer risk. To assess whether pleiotropic effects drove our results, we performed a PheWAS analysis, with no diseases or non-hematological traits detected. We detected reticulocyte count, lymphocyte count and hemoglobin concentration associated with rs1741315. However, it is unclear if and how the variability in those blood cell phenotypes could affect the cancer predisposition. Moreover, MR analysis of those hematological traits as exposure, revealed no evidence that these potential blood trait confounders are causal for the cancers tested. Therefore, there was no reason to conclude that pleiotropy would affect our Mendelian Randomization analyses.

In addition, the GWAS of gastric cancer was performed on a different ancestry (Japanese, East Asian) than the *SMOX* GWAS (Danish, Western Northern European), which could have biased our MR estimates. However, the top correlated variants of rs1741315 in Europeans are also the same as in the Japanese population (Table S2), which means the MR estimate bias should be negligible.

Finally, SMOX activity and expression levels were measured in blood in this study, compared to cancer tissue cells in the observational studies.

Our study had important strengths. Several previous studies which examined SMOX activity and cancer risk were susceptible to recall bias, confounding and reverse causation[13–18,21] none of which are concerns of Mendelian randomization studies[21]. Furthermore, using spermidine to spermine ratio to measure SMOX activity was here validated by the orthogonal measurement of *SMOX* gene expression and its genetic regulation. In addition, the fact that our genetic instrument explained a sizeable proportion of the variance of SMOX activity together with the fact that we used summary statistics from the largest meta-analyses of primary GWAS of these cancer types to date[28–33] are factors enhancing the statistical power to detect causal effects[50].

Conclusions

In conclusion, lifelong genetic exposure to low levels of SMOX was not associated with lower cancer risk. However, this finding cannot be assumed to represent a short-term strong drug inhibition of SMOX, thus establishing causality would require a randomized clinical trial. To better interpret the complexity of the relationship between SMOX activity and cancer, future studies should effectively distinguish newborn versus adulthood SMOX activity and short versus long exposures. This study is the first to use Mendelian randomization to assess the possible benefits of SMOX inhibitors on cancer risk, paving the way for other polyamine catabolic pathway enzymes to be tested with the same methodology.

References

1. Pegg, A. E. Polyamine metabolism and its importance in neoplastic growth and a target for chemotherapy. *Can. Res.* **48**(4), 759–774 (1988).
2. Gerner, E. W. & Meyskens, F. L. Jr. Polyamines and cancer: Old molecules, new understanding. *Nat. Rev. Cancer* **4**(10), 781–792 (2004).
3. Casero, R. A. Jr., Murray Stewart, T. & Pegg, A. E. Polyamine metabolism and cancer: treatments, challenges and opportunities. *Nat. Rev. Cancer.* **18**(11), 681–695 (2018).
4. Casero, R. A. Jr. & Marton, L. J. Targeting polyamine metabolism and function in cancer and other hyperproliferative diseases. *Nat. Rev. Drug Discov.* **6**(5), 373–390 (2007).
5. Russell, D. & Snyder, S. H. Amine synthesis in rapidly growing tissues: ornithine decarboxylase activity in regenerating rat liver, chick embryo, and various tumors. *Proc. Natl. Acad. Sci. U.S.A.* **60**(4), 1420–1427 (1968).
6. Martinez, M. E. *et al.* Pronounced reduction in adenoma recurrence associated with aspirin use and a polymorphism in the ornithine decarboxylase gene. *Proc. Natl. Acad. Sci. U.S.A.* **100**(13), 7859–7864 (2003).
7. Hubner, R. A. *et al.* Ornithine decarboxylase G316A genotype is prognostic for colorectal adenoma recurrence and predicts efficacy of aspirin chemoprevention. *Clin. Cancer Res.* **14**(8), 2303–2309 (2008).
8. Gupta, S. *et al.* Chemoprevention of prostate carcinogenesis by alpha- difluoromethylornithine in TRAMP mice. *Can. Res.* **60**(18), 5125–5133 (2000).
9. Elmets, C. A. & Athar, M. Targeting ornithine decarboxylase for the prevention of nonmelanoma skin cancer in humans. *Cancer Prev. Res. (Phila Pa).* **3**(1), 8–11 (2010).
10. Manni, A. *et al.* Prognostic influence on survival of increased ornithine decarboxylase activity in human breast cancer. *Clin. Cancer Res.* **2**(11), 1901–1906 (1996).
11. Uhlén, M. *et al.* Tissue-based map of the human proteome. *Science* **347**(6220), 1260419 (2015).
12. Wang, Y. *et al.* Cloning and characterization of a human polyamine oxidase that is inducible by polyamine analogue exposure. *Cancer Res.* **61**(14), 5370–5373 (2001).
13. Vujcic, S., Diegelman, P., Bacchi, C. J., Kramer, D. L. & Porter, C. W. Identification and characterization of a novel flavin-containing spermine oxidase of mammalian cell origin. *Biochem. J.* **367**(Pt 3), 665–675 (2002).
14. Babbar, N. & Casero, R. A. Jr. Tumor necrosis factor-alpha increases reactive oxygen species by inducing spermine oxidase in human lung epithelial cells: A potential mechanism for inflammation-induced carcinogenesis. *Cancer Res.* **66**(23), 11125–11130 (2006).
15. Goodwin, A. C. *et al.* Increased spermine oxidase expression in human prostate cancer and prostatic intraepithelial neoplasia tissues. *Prostate* **68**, 766–772 (2008).
16. Chaturvedi, R. *et al.* Spermine oxidase mediates the gastric cancer risk associated with Helicobacter pylori CagA. *Gastroenterology* **141**, 1696–1708 (2011).

17. Goodwin, A. C. *et al.* Polyamine catabolism contributes to enterotoxigenic Bacteroides fragilis-induced colon tumorigenesis. *Proc. Natl Acad. Sci. USA* **108**, 15354–15359 (2011).
18. Pledgie, A. *et al.* Spermine oxidase SMO(PAOh1), not N1-acetylpolyamine oxidase PAO, is the primary source of cytotoxic H2O2 in polyamine analogue-treated human breast cancer cell lines. *J. Biol. Chem* **280**, 39843–39851 (2005).
19. Amendola, R. *et al.* Direct oxidative DNA damage, apoptosis and radio sensitivity by spermine oxidase activities in mouse neuroblastoma cells. *Biochem. Biophys. Acta.* **1755**(1), 15–24 (2005).
20. Fratini, E. *et al.* Link between spermine oxidase and apoptosis antagonizing transcription factor: A new pathway in neuroblastoma. *Int. J. Oncol.* **55**(5), 1149–1156 (2019).
21. Ebrahim, S. & Davey Smith, G. Mendelian randomization: can genetic epidemiology help redress the failures of observational epidemiology?. *Hum. Genet.* **123**, 15–33 (2008).
22. Fadista, J. *et al.* Genome-wide meta-analysis identifies BARX1 and EML4-MTA3 as new loci associated with infantile hypertrophic pyloric stenosis. *Hum. Mol. Genet.* **28**(2), 332–340. https://doi.org/10.1093/hmg/ddy347 (2019).
23. Fadista, J. *et al.* Integrating genetics with newborn metabolomics in infantile hypertrophic pyloric stenosis. *Metabolomics* **17**(1), 7 (2021).
24. Hansen, C.S., *et al.* Elevated DNA methylation gestational age is associated with the risk of later bipolar disorder and anorexia nervosa in twins. *medRxiv* 2020.07.16.20155010
25. Leitsalu, L. *et al.* Cohort profile: Estonian biobank of the estonian genome center University of Tartu. *Int. J. Epidemiol.* **44**(4), 1137–1147 (2015).
26. Võsa U, *et al.* Unraveling the polygenic architecture of complex traits using blood eQTL metaanalysis. 2018. *bioRxiv* 447367.
27. Lepik, K. *et al.* C-reactive protein upregulates the whole blood expression of CD59 - an integrative analysis. *PLoS Comput. Biol.* **13**(9), e1005766 (2017).
28. Schumacher, F. R. *et al.* Association analyses of more than 140,000 men identify 63 new prostate cancer susceptibility loci. *Nat. Genet.* **50**(7), 928–936. https://doi.org/10.1038/s41588-018-0142-8 (2018).
29. Michailidou, K. *et al.* Genome-wide association analysis of more than 120,000 individuals identifies 15 new susceptibility loci for breast cancer. *Nat. Genet.* **47**(4), 373–380. https://doi.org/10.1038/ng.3242 (2015).
30. McKay, J. D. *et al.* Large-scale association analysis identifies new lung cancer susceptibility loci and heterogeneity in genetic susceptibility across histological subtypes. *Nat. Genet.* **49**(7), 1126–1132. https://doi.org/10.1038/ng.3892 (2017).
31. McDaniel, L. D. *et al.* Common variants upstream of MLF1 at 3q25 and within CPZ at 4p16 associated with neuroblastoma. *PLoS Genet.* **13**(5), e1006787 (2017).
32. Pan-UKB team. https://pan.ukbb.broadinstitute.org. 2020.
33. Ishigaki, K. *et al.* Large-scale genome-wide association study in a Japanese population identifies novel susceptibility loci across different diseases. *Nat. Genet.* **52**(7), 669–679 (2020).
34. Chang, C. C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience.* **25**(4), 7 (2015).
35. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/ (2020).
36. McCarthy, S. *et al.* A reference panel of 64,976 haplotypes for genotype imputation. *Nat. Genet.* **48**, 1279–1283 (2016).
37. Robinson, M. D. & Oshlack, A. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol.* **11**(3), R25 (2010).
38. Triche, T. J. Jr. *et al.* Low-level processing of illumina infinium DNA methylation BeadArrays. *Nucleic Acids Res.* **41**(7), e90 (2013).
39. Min JL, *et al.* Genomic and phenomic insights from an atlas of genetic effects on DNA methylation. *medRxiv* 2020.09.01.20180406.
40. Min, J. L. *et al.* Meffil: Efficient normalization and analysis of very large DNA methylation datasets. *Bioinformatics* **34**(23), 3983–3989 (2018).
41. Wahl, A. *et al.* IgG glycosylation and DNA methylation are interconnected with smoking. *Biochim. Biophys. Acta Gen. Subj.* **1862**(3), 637–648 (2018).
42. Stephen, B., Butterworth, A. S. & Thompson, S. G. Mendelian randomization analysis with multiple genetic variants using summarized data. *Genet. Epidemiol.* **37**, 658–665. https://doi.org/10.1002/gepi.21758 (2013).
43. Kamat, M. A. *et al.* PhenoScanner V2: an expanded tool for searching human genotype-phenotype associations. *Bioinformatics* **35**(22), 4851–4853. https://doi.org/10.1093/bioinformatics/btz469 (2019).
44. Hemani, G. *et al.* The MR-Base platform supports systematic causal inference across the human phenome. *Elife* **30**(7), e34408 (2018).
45. Astle, W. J. *et al.* The allelic landscape of human blood cell trait variation and links to common complex disease. *Cell* **167**(5), 1415-1429.e19. https://doi.org/10.1016/j.cell.2016.10.042 (2016).
46. VanderWeele, T. J., Tchetgen Tchetgen, E. J., Cornelis, M. & Kraft, P. Methodological challenges in mendelian randomization. *Epidemiology* **25**, 427–435 (2014).
47. Thomas, D. C. & Conti, D. V. Commentary: The concept of 'Mendelian randomization'. *Int. J. Epidemiol.* **33**, 21–25 (2004).
48. Stephen, B. & Thompson, S. G. Multivariable mendelian randomization: The use of pleiotropic genetic variants to estimate causal effects. *Am. J. Epidemiol.* **181**(4), 251–260. https://doi.org/10.1093/aje/kwu283 (2015).
49. Burgess, S. & Thompson, S. G. Interpreting findings from Mendelian randomization using the MR-Egger method. *Eur. J. Epidemiol.* **32**(5), 377–389 (2017).
50. Burgess, S. Sample size and power calculations in Mendelian randomization with a single instrumental variable and a binary outcome. *Int. J. Epidemiol.* **43**, 922–929 (2014).

Acknowledgements

Author contributions

J.F. conceptualized, designed the study, coordinated data collection and carried out the initial analyses and drafted the initial manuscript. V.Y., U.V., C.S.H., S.K., J.C. and V.K. collected data and also carried initial analyses. L.S., F.G., T.E., A.B., M.M., T.M.W., D.M.H., L.M., J.B.G., A.S.C. conceptualized and designed the study, and critically reviewed the manuscript for important intellectual content. B.F. conceptualized, designed the study, coordinated and supervised data collection and analysis, and critically reviewed the manuscript for important

intellectual content. All authors approved the final manuscript as submitted and agree to be accountable for all aspects of the work.

Funding

Competing interests

The JF works for Novo Nordisk A/S since January 2021. No other author has a competing interest regarding this study.

Data availability

Supporting data is available in Supplementary Material and on request.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-021-97069-x.

Correspondence and requests for materials should be addressed to J.F. or B.F.

Reprints and permissions information is available at www.nature.com/reprints.

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