HHS Public Access

Author manuscript

Environ Int. Author manuscript; available in PMC 2023 March 01.

Published in final edited form as:

Environ Int. 2022 March ; 161: 107086. doi:10.1016/j.envint.2022.107086.

Secondhand smoke affects reproductive functions by altering the mouse testis transcriptome, and leads to select intron retention in *Pde1a*

Stella Tommasi[1,*], **Tevfik H. Kitapci**[1], **Hannah Blumenfeld**[1], **Ahmad Besaratinia**[1]

[1]Department of Population and Public Health Sciences, USC Keck School of Medicine, University of Southern California, M/C 9603, Los Angeles, CA 90033, USA

Abstract

Background: Human exposure to secondhand smoke (SHS) is known to result in adverse effects in multiple organ systems. However, the impact of SHS on the male reproductive system, particularly on the regulation of genes and molecular pathways that govern sperm production, maturation, and functions remains largely understudied.

Objective: We investigated the effects of SHS on the testis transcriptome in a validated mouse model.

Methods: Adult male mice were exposed to SHS (5 h/day, 5 days/week for 4 months) as compared to controls (clean air-exposed). RNA-seq analysis was performed in the testis of SHS-exposed mice and controls. Variant discovery and plink association analyses were also conducted to detect exposure-related transcript variants in SHS-treated mice.

Results: Exposure of mice to SHS resulted in the aberrant expression of 131 testicular genes. Whilst approximately two thirds of the differentially expressed genes were protein-coding, the remaining (30.5%) comprised noncoding elements, mostly lncRNAs (19.1%). Variant discovery analysis identified a homozygous frameshift variant that is statistically significantly associated

[*]**Corresponding Author: Tel:** (323) 442-7753; **Fax:** (323) 865-0103. tommasi@med.usc.edu.

Author Contributions

ST: conceived and designed the study, funding acquisition, performed experiments and collected data, analyzed data, and interpreted the results; wrote the manuscript; THK: performed experiments and collected data, performed bioinformatics analysis, wrote methods; HB: performed experiments and collected data, analyzed data; AB: initial study design, funding acquisition, performed experiments and collected data, analyzed data, and interpreted the results, co-wrote the manuscript. All authors approved the final version of the manuscript.

Conflicts of Interest

All authors declare no competing interests. The sponsors of the study had no role in study design, data collection, data analysis, data interpretation, writing of the report, or in the decision to submit for publication.

Conflicts of Interest: All the authors declare no conflict of interest. A completed declaration of competing financial interest form has been submitted together with this manuscript. The authors declare they have no actual or potential competing financial interests.

Declaration of interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

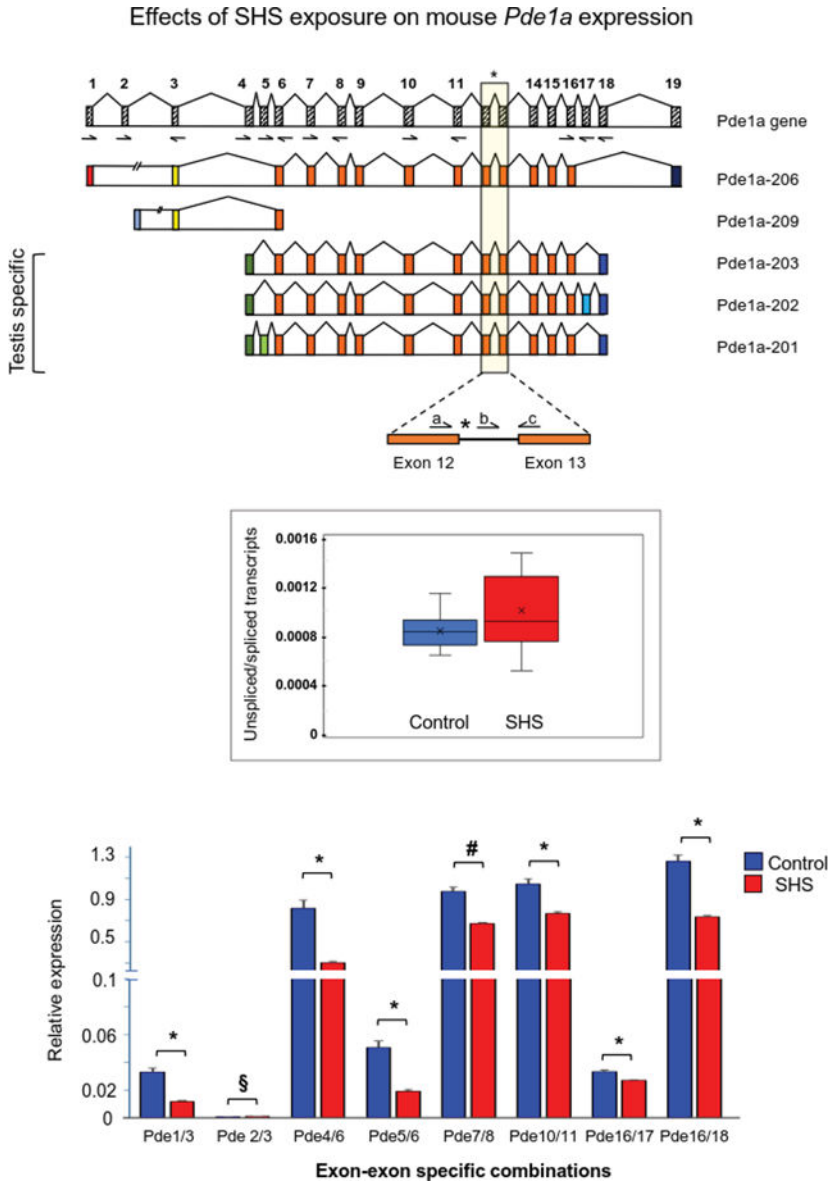
Supplemental Materials

Supplemental Materials include Supplementary Tables S1 and S2, and Supplementary Figure S1 with corresponding legend.

with SHS exposure (P= 7.744e-06) and is generated by retention of a short intron within *Pde1a*, a key regulator of spermatogenesis. Notably, this SHS-associated intron variant harbors an evolutionarily conserved, premature termination codon (PTC) that disrupts the open reading frame of *Pde1a*, presumably leading to its degradation *via* nonsense-mediated decay.

Discussion: SHS alters the expression of genes involved in molecular pathways that are crucial for normal testis development and function. Preferential targeting of lncRNAs in the testis of SHS-exposed mice is especially significant considering their crucial role in the spatial and temporal modulation of spermatogenesis. Equally important is our discovery of a novel homozygous frameshift variant that is exclusively and significantly associated with SHS-exposure and is likely to represent a safeguard mechanism to regulate transcription of *Pde1a* and preserve normal testis function during harmful exposure to environmental agents.

Graphical Abstract



Keywords

Gene expression; Male reproductive system; Variant; RNA-seq; Secondhand smoke

1. Introduction

Exposure to secondhand smoke (SHS) is a pressing global health concern as it continues to cause substantial morbidities and mortalities, worldwide (IARC 2004, 2012; Max et al. 2012; Oberg et al. 2011; UDHHS 2006, 2014). It is currently estimated that 93% of the world's population still lives in low- and middle-income countries with non-existent or poorly enforced smoke-free laws, and that 40% of the world's children (≤14 years old) are exposed to SHS (Oberg et al. 2011; Ritchie and Roser 2013). As a result, more than 880,000 individuals are expected to die globally of SHS-related diseases every year, according to the World Health Organization (Yousuf et al. 2020).

SHS is a mixture of "sidestream" smoke (SSS) and a small fraction of "mainstream" smoke (MSS), exhaled by smokers (IARC 2012; UDHHS 2006). Although the chemical composition of SHS and MSS differs from one another, because of ageing and dilution with ambient air, they both contain a wide variety of toxicants and carcinogens, albeit at varying concentrations (FDA 2012; IARC 2012; Lodovici and Bigagli 2009).

Several studies have investigated the reproductive effects of active smoking in male smokers as approximately 20% of the world's population aged ≥15 years smoke cigarettes daily, and the highest prevalence of smoking is in young men during their reproductive age (WHO 2018). Tobacco smoking has been shown to cause a wide range of genotoxic effects in smokers' sperm, including oxidative DNA damage, DNA adduction, DNA strand breaks, DNA fragmentation, gene mutations, and chromosomal aberrations (DeMarini 2004; Ji et al. 2013; La Maestra et al. 2014, 2015; Marchetti et al. 2011). Epigenetic changes, including aberrant DNA methylation (Gunes et al. 2018; Jenkins et al. 2017; Laqqan et al. 2017; Xu et al. 2013) and dysregulation of microRNAs have also been observed in smokers' sperm (Marczylo et al. 2012). Furthermore, active smoking has been shown to adversely affect sperm quality, fertilization ability, reproduction, and early development (ASRM 2012; Dai et al. 2015; Esakky et al. 2018; Gunes et al. 2018; Harlev et al. 2015). To date, however, the effects of SHS exposure on the reproductive system and functions remain largely understudied.

Epidemiological studies investigating the impact of SHS exposure on the male reproductive system have been hampered mainly by the difficulties in collecting biological specimens from this organ system. Added to these challenges are the uncertainties regarding the assessment of chronic exposure to SHS in human populations. The latter is conventionally achieved by administering questionnaires or interviewing study subjects, a practice prone to recall bias. Smoke inhalation experiments in mouse models offer a viable alternative to human studies, whereby well-defined and highly controlled exposure conditions allow to replicate human exposure to SHS and extend the investigation to any organ/tissue of interest (Coggins 2007; Hecht 2005). According to the Mouse Genomic Informatics (MGI) database, over 90% of murine genes have a correspondent human ortholog (Bult et al. 2016).

Moreover, gene expression patterns between human and mouse tissues are highly conserved, especially for genes expressed in a tissue-specific manner (Dowell 2011).

In the present study, we have used advanced sequencing technologies (RNA-seq) and bioinformatics analysis to investigate, for the first time, the transcriptomic effects of SHS in the testis of mice. Using a microprocessor-controlled smoking machine, we have sub-chronically exposed mice to SHS (5 hours/day, 5 days/week for 4 months), as previously described (Tommasi et al. 2020). As a primary area of interest, we have first analyzed the global gene expression to identify genes and associated molecular pathways and functional networks that were impacted in SHS-exposed mice as compared to controls (clean air-exposed). Subsequently, we have performed variant discovery and association analysis to detect sequence variations, at a transcriptional level, strictly linked to the SHS exposure status and relevant to reproductive functions. Finally, we have validated the results, at a single gene level, by quantitative reverse-transcription polymerase chain reaction (RT-qPCR). A schematic representation of the workflows used for gene expression profiling and variant discovery is shown in Figure 1 (Fig. 1).

2. Materials and Methods

2.1 Animal care and maintenance

The study was conducted according to the National Institutes of Health (NIH) Guide for the Care and Use of Laboratory Animals (Council 2011), and the Animal Research: Reporting of *In Vivo* Experiments (ARRIVE) guidelines. All efforts were made to minimize animal suffering. The study was reviewed and approved by the Institutional Animal Care and Use Committee (IACUC) of the City of Hope (Protocol # 09012). All mice were maintained in polypropylene cages in groups of 2–3 animals per cage and had access to food and water *ad libitum* at all times.

2.2 Study design

Adult male mice (6–8 weeks old), on a C57BL/6 genetic background, were randomly divided into two groups: (1) ‘experimental’ (SHS exposure) and (2) ‘control’ (sham-treatment in clean air). Each group consisted of five mice. The experimental group was exposed daily to SHS for four months as described in the section below and in refs. (Tommasi et al. 2015b; Tommasi et al. 2020). Similarly, the control group was exposed to filtered high-efficiency particulate-air (HEPA) (*see*, section below). At the end of all experiments, SHS-exposed and control mice were euthanized by CO_2 asphyxiation and various tissues and organs, including the testes, were harvested, and kept at − 80 °C until further analysis. Based on a study by Flurkey et al. (Flurkey et al. 2007), investigating mouse models of aging as compared to humans, 24-month-old mice (C57BL/6J) are equivalent to human beings of 69 years of age. Applying this mouse-human age conversion, we have calculated that a 4-month exposure of mice to SHS corresponds to approximately 12 years of human exposure to SHS, which is a biologically relevant duration of exposure in humans. In our previous experiments, we have confirmed that five mice per group ensure a study power of at least 80% to detect a gene expression fold change (FC) of 1.5 between SHS-exposed and control mice at a False Discovery Rate (FDR) < 0.05 (Tommasi et al. 2015b).

2.3 SHS exposure

The smoking machine and exposure protocol have been described in detail in refs. (Tommasi et al. 2015b; Tommasi et al. 2020). Briefly, SHS was generated using a microprocessor-controlled TE-10 smoking machine (Teague Enterprises, Woodland, CA, USA). The TE-10 smoking machine was programmed to burn 3R4F Reference Kentucky cigarettes (Tobacco Research Institute, University of Kentucky, Lexington, KY, USA), producing a mixture of sidestream smoke (89%) and mainstream smoke (11%). This formulation is conventionally used to simulate SHS exposure for *in vivo* experiments (Besaratinia and Pfeifer 2008; D'Agostini et al. 2006; IARC 2012; Tommasi et al. 2015b). Following an acclimatization period, the mice were exposed, whole body, to SHS for 5 h/day, 5 days/week, for four months. This exposure regimen was achieved by simultaneously burning 7–9 cigarettes at any given time during the experiments. The average concentration of total suspended particulate (TSP) in the exposure chambers was 233.0 ± 15.4 mg/m^3 during the exposure period (Kim et al. 2011).

We note that whole body smoke exposure in rodents may result in residual transdermal and gastrointestinal absorption of smoke particles due to grooming (Kim et al. 2012). However, 'nose-only' exposure can cause stress and discomfort for the animals, which would be pronounced in long-term studies, such as the present one. Therefore, we chose whole body exposure of mice to SHS based on tolerability as well as comparability to humans being exposed to SHS in real life (Kim et al. 2012). We also note that a daily SHS exposure of mice, for 7 days/week (La Maestra et al. 2014, 2015; Polyzos et al. 2009), would more closely replicate human exposure to SHS, and that a shorter exposure time could be a limitation. However, we would also like to point out that a regimen of 7 days/week for prolonged periods of time is logistically demanding and technically challenging. As such, the overwhelming majority of published studies on SHS in animal models have used the exposure protocol used in the present study, which consists of 5 days/week of chronic exposure to SHS. We would like to stress that this protocol has and continues to be used routinely by well-established research groups throughout the world (Bivalacqua et al. 2009; Hartney et al. 2012; IARC 2004; Teague et al. 1994; Witschi et al. 2002). Finally, we would like to emphasize that, in our previously published studies (Kim et al. 2011; Kim et al. 2012; Tommasi et al. 2012; Tommasi et al. 2014; Tommasi et al. 2015a, b; Tommasi et al. 2020) we have verified that SHS exposure (5 h/day 5 days/week) for four months is sufficient to elicit significant genotoxic, epigenetic, and transcriptomic responses in various organs and tissues of male C57BL/6 mice and that these effects can persist over extended periods of time (Tommasi et al. 2015b; Tommasi et al. 2020).

2.4 Library construction and sequencing

Total RNA was isolated from mouse testis using the RNeasy Mini Kit (Qiagen, Valencia, CA, USA). RNA samples were checked for quality control using the RNA 6000 Nano Chip kit in an Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA). Libraries for RNA-seq were prepared from total RNA (300 ng per sample) using the Kapa Hyper Prep Kits with RiboErase (Kapa Biosystems). The workflow consisted of ribosomal RNA (rRNA) depletion, cDNA generation, and end repair to generate blunt ends, A-tailing, adaptor ligation, and PCR amplification. Different adaptors were used for multiplexing

samples in one lane. Sequencing was performed on an Illumina HiSeq3000 platform, in paired-end format with 150 bp reads. To minimize technical variance, library construction and data acquisition for all samples were done in a single run, not in different batches, and in a 'blind' fashion. The RNA-seq data will be deposited in the Gene Expression Omnibus database at NCBI (htttp://www.ncbi.nlm.nih.gov/geo/), and accession number will be provided as soon as it becomes available.

2.5 Differential gene expression analysis

Sequence analysis of the libraries prepared from SHS-exposed and control mice yielded an average of 43 million raw reads per sample. The quality of raw sequencing data was checked with fastQC (Andrews et al. 2010). Following filtering and trimming, analysis-ready RNA-seq reads were mapped to the *Mus Musculus* reference genome (Gencode_mouse/release_M22/GRCm38) using the STAR aligner tool. The reference genome for *Mus musculus* (https://ftp.ebi.ac.uk/pub/databases/gencode/Gencode_mouse/release_M22/GRCm38.primary_assembly.genome.fa.gz) and the annotation file (ftp://ftp.ebi.ac.uk/pub/databases/gencode/Gencode_mouse/release_M22/gencode.vM22.chr_patch_hapl_scaff.annotation.gtf.gz) were downloaded from the EMBL-EBI genome browser. STAR aligner was run for each sample with default settings except for the following parameters: --twopassMode Basic --quantMode GeneCounts. GeneCounts parameter generates read count for every gene. A custom python script was used to merge the individual files to form a matrix so that i^{th} row represents the i^{th} gene and j^{th} columns represents the j^{th} sample. The resulting matrix was saved as a CSV file, with 18,851 transcripts, to be used as input for the R package DESeq2 analysis (Love et al. 2014). DESeq2 implements a generalized linear model with a negative-binomial link function to model the count data from high-throughput sequencing experiments.

2.6 Clustering analysis

To assess the clustering of samples within a group, we used Principal Component Analysis (PCA) dimensionality reduction and k-means clustering. DESeq2 results were filtered using an FDR of 0.05 and a FC of 1.5. The resulting subset of genes was used as input for the PCA using the R function prcomp(). K-means clustering algorithm was run based on the first two PCs.

2.7 Functional analysis

We used the Ingenuity® Pathway Analysis (IPA®) application (QIAGEN Bioinformatics, Redwood City, CA, USA; www.qiagenbioinformatics.com) to analyze the gene list generated by DESeq2. The IPA® software searches for overlaps between the list of DEGs and the information available in the Ingenuity knowledge database, thus allowing exploration of relationships, canonical pathways, upstream regulators and function and disease categories relevant to the dataset. The Fisher exact test was used to calculate the overlap *p*-value, and the *z*-score was used to assess potential functional outcomes, positive for '*activation*' and negative for '*inhibition*'. To sort the results, Benjamini-Hochberg (BH) corrected FDR was set to 5% (Reiner et al. 2003), and pathways were sorted by bias corrected *z*-score.

2.8 Reverse Transcription-quantitative Polymerase Chain Reaction (RT-qPCR)

We used RT-qPCR to validate randomly selected DEGs identified by RNA-seq in SHS-exposed mice relative to controls. The RT-qPCR protocol has been described in refs (Tommasi et al. 2015b; Tommasi et al. 2020). Briefly, gene-specific primers were designed using the NCBI primer blast software available at https://www.ncbi.nlm.nih.gov/tools/primer-blast/. Total RNA (250 ng) isolated from the testis of mice was treated with DNase I and reverse transcribed into cDNA using the iScript™ Reverse Transcription Supermix for RT-qPCR (iScript RT Supermix) (Bio-Rad laboratories, Inc., Hercules, CA). The synthesized cDNA was further diluted with low-EDTA TE buffer (10 mmol/L Tris-HCl, 0.1 mmol/L EDTA, pH 8.0), and two µl of the diluted cDNA were used per reaction in a master mix containing gene-specific primers and SsoAdvanced™ Universal SYBR® Green Supermix (Bio-Rad laboratories, Inc.). The mouse actin beta (*Actb*), glyceraldehyde-3-phosphate dehydrogenase (*Gadph*), and TATA-Box Binding Protein (*Tbp*) genes were used as references. All PCR assays were performed using the CFX96 Touch™ Real-Time PCR detection system (Bio-Rad Laboratories, Inc.). The cycling conditions included a pre-incubation at 95 °C for 2 minutes, followed by 40 cycles at 95 °C for 5 seconds, and 58 °C for 30 seconds. Three to five samples per experimental and control group were run in triplicate for every target gene and with the reference gene. Fold-changes in transcript level were calculated between the experimental and control group using the gene study feature of CFX Maestro™ software (BioRad Laboratories, Inc.). The relative expression of the Pde1a transcripts was calculated using exon-exon specific combinations and applying the formula $2^{-\Delta Cq}$. The primer sets used for RT-qPCR are listed in Supplementary Table 2 (Suppl. Table 2).

2.9 Genotyping and association analysis

We used the Genome Analysis Tool Kit (GATK) pipeline, developed by the Broad Institute, to identify transcript variants, *i.e.*, single nucleotide polymorphisms (SNPs) and indel, in SHS-exposed mice relative to controls (Fig. 1). The latest version of the GATK software (GATK4) is available at (https://gatk.broadinstitute.org/hc/en-us) with end-to-end workflows called "*GATK Best Practices*". Read alignment files were generated using STAR aligner, as described above. Read groups were added using "AddOrReplaceReadGroups", and duplicates were marked with "MarkDuplicates". "SplitNCigarReads" was then used to split reads that contained Ns in their CIGAR strings. HaplotypeCaller was run on every chromosome in parallel using the -ERC GVCF parameter to generate genomic variant calling files (gVCF). The results from each chromosome were merged with the "GatherVcfs" function. gVCF files from each sample were merged with the "CombineGVCFs", and genotypes were called for all samples using "GenotypeGVCFs". Variants were filtered using the hard filtering criteria suggested by the GATK best practices using the Variant Filtration tool. Once the variants were detected, a transcriptome-wide association analysis was performed using the plink software (http://pngu.mgh.harvard.edu/purcell/plink/) (Purcell et al. 2007). Association analysis was run using the genotypes as predictor variable and the treatment (SHS-exposure versus control) as binary response variable. A second association analysis was run using differential gene expression as the response variable. The "CombineGVCFs" file displaying sequence variations was

uploaded into the 'Integrated Genomic Viewer' (IGV_2.9.4) for visualization of the variants (Robinson et al. 2011).

3. Results

3.1 Genome-wide gene expression analysis

To investigate the effects of SHS exposure on global gene expression, we performed transcriptome analysis of strand specific RNA-seq libraries generated from the whole testis of mice from the experimental and control group. Of the 18,851 genes/transcripts identified by DESeq2 (Fig. 1), 131 were found to be differentially expressed in SHS-exposed mice as compared to controls (at FDR-adjusted *p*-value of < 0.05 and FC of 1.5). Of these 131 genes, ninety were up-regulated and forty-one were down-regulated, representing 68.7% and 31.3%, respectively, of the total pool of differentially expressed genes (DEGs) in SHS-exposed mice (Fig. 2A and 2B). Of the up-regulated genes, 73.3% were protein coding whilst 20% belonged to the class of long non-coding RNAs (lncRNAs) (based on the Ensembl gene/transcript biotype classification). The remaining 6.7% included elements of unknown function, *i.e.*, pseudogenes and 'to be experimentally confirmed' (TEC) elements (Fig. 2C). Similarly, down-regulated genes were largely protein coding (61%), though lncRNAs (17.1%) and pseudogenes (21.9%) were highly represented (Fig. 2C). The relative high percentage of lncRNAs is consistent with a recent microarray analysis of transcripts in multiple mouse tissues indicating that the testis harbors the highest number of tissue-specific lncRNAs (11% *versus* 5%, in brain and 1% in other tissues) (Hong et al. 2018). Almost all dysregulated lncRNAs were long intergenic non-coding RNAs (lincRNAs; based on the Ensembl categorization), which is the most abundant biotype detected in the mouse testis (Hong et al. 2018; Sahlu et al. 2020).

Principal component analysis (PCA) (Fig. 2D) and hierarchical clustering (Fig. 2E) of the DEGs showed a clear separation of the experimental group (S6 to S10) from the control group (S1 to S5). The full list of DEGs identified in SHS-exposed *versus* control mice is shown in Supplementary Table 1. The list includes related information on genomic loci (where annotation is available), transcript biotype, FC, and BH-corrected *P* value (Suppl. Table 1).

3.2 Functional network and pathway analysis

We used the Ingenuity Pathway Analysis® software (IPA® v.9.0) to gain insight into the molecular pathways and functional networks associated with the aberrantly expressed genes in SHS-exposed mice as compared to controls. Of the 131 DEGs detected by RNA-seq, 115 mapped to known IDs in the IPA® database (78 up-regulated and 37 down-regulated genes, respectively). As shown in Table 1, "*cancer*" and "*organismal injury and abnormalities*" were the top diseases associated with the DEGs (based on a BH-corrected *P* value < 0.05 and sorted by bias corrected *z*-score; *see*, Table 1). Also, the "*cell to cell signaling*", "*reproductive system development and function*" and "*emotional behavior*" categories were significantly associated with the DEGs (Table 1). Gene network analysis by IPA® revealed that the most affected functional network included DEGs that converge into the 'Transforming growth factor-beta' (TGF-β) pathway and the 'Ras-dependent extracellular

signal-regulated kinase (ERK)1/2 mitogen-activated protein (MAP) kinase' pathway (Fig. 3A). Both signaling pathways are known to play a central role in regulating growth and cell proliferation in mammalian cells and are often impacted in cancer, including genitourinary malignancies (Boguslawska et al. 2019; Dhillon et al. 2007; Meloche and Pouyssegur 2007). Notably, the ERK1/2 and the TGF-β pathways are among the most important signal transduction pathways involved in Sertoli cell proliferation, a process that is essential for normal spermatogenesis (Meroni et al. 2019).

Next, we used the 'Upstream Regulator Analysis' tool in IPA® to identify potential upstream regulators that are likely to cause aberrant expression of the target genes identified in SHS-exposed mice. Based on a *P*-value of overlap of <0.05 and an activation *z*-score of > 2 in the IPA® prediction tool, the following top master regulators were identified: (1) the transcription factor HNF1 homeobox A (HNF1A; *z*-score: 2.449), (2) the phosphatase and tensin homolog (PTEN; *z*-score: 2.236), (3) the transforming growth factor beta 1 (TGFB1; *z*-score: 2.201), and (4) the insulin-like growth factor 1 (IGF1; *z*-score: 2.160) (Fig. 3B). Our prediction analysis shows that HNF1A, PTEN, TGFB1, and IGF1 can work together to elicit, either by direct binding (as in the case of the HNF1A transcription factor) or indirectly, the SHS-induced transcriptional changes observed in the present study (Fig. 3B). Notably, HNF1A, PTEN, TGFB1, and IGF1 and several of their downstream targets are known to be involved in the morphology and development of the male reproductive system and often associated with genitourinary carcinoma (Fig. 3B and Table 2) (Mallik et al. 2020; Neirijnck et al. 2019; Ni et al. 2019).

3.3 Gene validation of RNA-seq data by RT-qPCR

To independently validate the RNA-seq results, we randomly selected several targets from the 131-gene list (Suppl. Table 1) and examined their transcription levels by standard RT-qPCR. In agreement with the RNA-seq data, we confirmed up-regulation of the nuclear paraspeckle assembly transcript 1, formerly known as nuclear-enriched abundant transcript 1 (*Neat1*), the RIKEN cDNA 4930560O18 gene (*4930560O18Rik*), the fibrillarin-like 1 gene (*Fbll1*), the six transmembrane epithelial antigen of the prostate 1 gene (*Steap1*) and the integrator complex subunit 6 like gene (*Ints6l*) (Fig. 3C). Down-regulation of the kallikrein 1-related peptidase b21 gene (*Klk1b21*), a serine protease expressed exclusively in the Leydig cells (the interstitial cells adjacent to the seminiferous tubules) of adult mouse testis (Matsui and Takahashi 2001), was also verified in SHS-exposed mice by RT-qPCR (Fig. 3C).

In all cases, the median normalized expression levels of the target genes determined by RT-qPCR were directly correlated to the normalized read counts from RNA-seq analysis. (Fig. 3D). The strong positive correlation (r^2 =0.97; P < 0.0001) between the expression results of the target genes measured independently by RT-qPCR and RNA-seq (Fig. 3D) confirmed the robustness and reproducibility of the whole transcriptome analysis performed in this study.

3.4 Detection of transcript variants

Besides investigating differential gene expression, RNA-seq has proven useful for detecting low-frequency and/or functionally important variants, such as single nucleotide polymorphisms (SNPs) and indels as well as longer structural variants (SV; ≥50 bp), within untranslated regions (UTRs), coding exons, and introns (Piskol et al. 2013). Using the GATK Best Practices workflow as outlined in Figure 1 (Fig. 1; *see*, Methods), we further analyzed the RNA-seq data and called, with high confidence, 54,550 genotypes (hereafter referred to a*s* '*variants*') across the transcriptome of SHS-exposed mice and controls. Association analysis by plink, followed by multiple testing correction in R, enabled us to identify a novel frameshift variant (indel) at position 79,875,122 on chromosome 2 that was statistically significantly associated with exposure to SHS (*P* value: 7.744e-06; Fig. 4A). As shown in Figure 4B, all samples derived from SHS-exposed mice contained an insertion of 106 nucleotides (nt) at position chr2: 79,875,122 (forward strand) on both alleles [(T)CCTGT…ACCTA-/(T)CCTGT…ACCTA-n)], while counterpart samples from control mice were all homozygous for the [(T)-/(T)-n] genotype (Fig. 4B). Sequence alignment with the reference genome indicates that this SHS-associated variant is likely the result of a splice defect leading to retention of a short intron (intron 12) between exons 12 and 13 of the murine *Pde1a* gene (ENSMUSG00000059173; encoded on the reverse strand) (Fig. 4C) (Howe et al. 2021). Importantly, retention of intron 12 disrupts the open reading frame of the *Pde1a* gene by introducing a premature termination codon (PTC) immediately after the 5'end splice site at position chr2: 79,875,227 (on the reverse strand). The PTC occurs within the consensus sequence 'GURAGU', where URA can be one of the stop codons (UAA, UGA, or UAG) (Fig. 4C). Of significance, the sequence surrounding the PTC is evolutionary conserved among several species, strongly suggesting that this stop codon plays a functionally and biologically important role (Fig. 4D).

The allele frequency of the SHS-associated variant is very low, as calculated by the allele depth/depth of coverage ratio (AD/DP = ~ 1%) in samples from SHS-exposed mice. This is consistent with previous reports showing that intron retention (IR) occurs at a low percentage in most mammalian genes, making its detection difficult (Edwards et al. 2016; Grabski et al. 2021). To independently quantify the levels of this unique variant in SHS-exposed mice relative to controls, we designed a RT-qPCR assay for the *Pde1a* intron 12 using primers specific for the amplification of either the exon12-exon13 junction (*a/c*, intron 12 spliced out) or the intron12-exon13 junction (*b/c*, intron 12 retained) (Fig. 5A; inset). The size of the PCR products was verified on agarose gel and the sequence of the respective fragments (with or without intron 12) was confirmed by Sanger sequencing (data not shown). As expected, *Pde1a* transcripts retaining the intron (IR transcripts/unspliced) are present at extremely low levels. This is reflected by the much higher number of PCR cycles needed to amplify the intron12-exon13 junction (unspliced transcripts, Cq ~32) than the exon12-exon13 junction (spliced transcripts, Cq ~12) (Fig. 5B).

Next, we calculated the ratio of the unspliced transcripts to the spliced transcripts in samples from SHS-exposed mice relative to controls, as described in ref. (Edwards et al. 2016). We observed a slight but not significant increase in IR-transcript levels in SHS-exposed mice relative to controls (Fig. 5C), which is consistent with the variant being strongly

associated with exposure, but not with differential gene expression (based on our plink association analysis). It is important to note that transcripts with a premature stop codon are often subjected to targeted degradation by nonsense-mediated decay (NMD), a surveillance mechanism that selectively degrades abnormal transcripts, which may give rise to harmful truncated proteins (Ge and Porse 2014). Together, the NMD-targeted degradation of the IR transcripts, and the extremely low frequency of the variant are likely to translate into a limited number of aberrant templates available for PCR amplification. As a result, the RT-qPCR analysis is expected to underestimate the actual number of the IR transcripts in samples from SHS-exposed mice relative to controls.

3.5 mRNA quantification of *Pde1a* isoforms

Several studies have demonstrated that an inverse correlation exists between IR and mRNA levels (Braunschweig et al. 2014; Dvinge and Bradley 2015; Edwards et al. 2016), and that IR can regulate gene/protein isoform production (Jung et al. 2015; Zheng et al. 2020). At least ten different mouse *Pde1a* transcripts, derived from differential promoter usage and alternative splicing, have been described to date (*i.e.*, Pde1a-201 to Pde1a-210 based on Ensembl genome browser) (*see*, Suppl. Fig. 1). Pde1a-204, −205, −206 represent the longest transcripts which encode the longest PDE1A isoform (545 aa). Pde1a-201, −202, −203 transcripts have shorter and distinct N- and C-termini compared to the long transcripts and encode proteins of 420 aa, 465 aa and 456 aa, respectively. Pde1a-201, −202, −203 are highly expressed in the testis (Vasta et al. 2005). The group includes the protein coding Pde1a-208 (476 aa), two processed transcripts (Pde1a-207 and −209), and one transcript that is predicted to be the primary target for NMD degradation (Pde1a-210) (*see*, Suppl. Fig. 1).

We performed RT-qPCR using several combinations of exon-binding primers that are specific for the various *Pde1a* transcripts, as described in ref. (Vasta et al. 2005). The location of the exon-specific primers is illustrated in Figure 5A (Fig. 5A), while primer sequence and amplicon size are provided in Supplementary Table 2 (Suppl. Table 2). In line with previous reports (Michibata et al. 2001; Vasta et al. 2005), our PCR analysis confirmed that Pde1a-203 is the most abundant transcript detected in the testis of untreated control mice, as indicated by the high level of expression observed for exons 4/6, which are unique for this gene isoform (Figs. 5D). Likewise, lower expression of two additional testis-specific *Pde1a* transcripts was confirmed in samples from control mice based on the detection level of exons 5/6 and 16/17, which are specific for Pde1a-201 and Pde1a-202, respectively. This indicates that the signal for exons 16/18, which are shared by all three testis specific isoforms, is mostly attributable to Pde1a-203 (Fig. 5D). Signal for exons 1/3 and 2/3, which target the unique 5’ end of the long transcripts (Pde1a-204, −205, −206) and the processed transcript containing a unique exon 2 (Pde1a-209), is almost undetectable in the testis of untreated control mice (Fig. 5D). Overall, we observed decreased expression of various *Pde1a* isoforms in samples from SHS-exposed mice as compared to controls. Notably, the most pronounced decrease was in the expression of Pde1a-203 transcripts as indicated by the 63.1% drop in signal detected by PCR for exons 4/6, and the 41.8% drop for exons 16/18. A more moderate reduction in signal was also observed for exons 7/8 (31.2% down) and exons 10/11 (26.6% down) in samples from SHS-exposed mice relative to controls. The latter is likely due to the contribution of other splicing isoforms, which also contain exons 7/8 and

10/11 (Fig. 5A and 5D). Altogether, our RT-qPCR results show that transcription of the main testis specific isoform (Pde1a-203) is significantly reduced consequent to exposure to SHS. Our findings support a major role for IR-mediated mRNA degradation in the regulation of the testis *Pde1a*.

4. Discussion

We provide evidence that *in vivo* exposure of mice to SHS induces changes in the testicular transcriptome that are likely to affect normal testis development and function. As shown in Figures 2B and Supplementary Table 1, sub-chronic exposure of mice to SHS resulted in the aberrant expression of 131 testicular genes (Figs. 2B and Suppl. Table 1). Whilst approximately two thirds of the DEGs were protein-coding, the remaining (30.5%) comprised noncoding elements, *i.e.*, lncRNAs (19.1%) and putatively nonfunctional sequences with unknown roles, such as TEC and pseudogenes (11.4%) (Fig. 2C). LncRNAs represent an abundant class of RNAs longer than 200 bp (as opposed to small noncoding RNAs) that are not translated into proteins and are tissue- and cell-type specific (Rinn and Chang 2012). Although their modes of action are not fully characterized, lncRNAs are known to regulate a multitude of biological processes, including transcriptional activation or repression, chromatin remodeling, mRNA splicing, editing and degradation, modulation of miRNAs, regulation of protein activity or abundance, and other functions (Kopp and Mendell 2018; Quinn and Chang 2016; Rinn and Chang 2012; St Laurent et al. 2015). The observed high percentage of dysregulated lncRNAs suggests the preferential targeting of these regulatory elements by SHS *in vivo*. Such a number is also consistent with the pervasive transcription of lncRNAs normally observed in the testis of untreated mice (and mammals in general) as compared to other tissues, which is reflective of the biological relevance of lncRNAs in the highly dynamic and complex program of spermatogenesis (Hong et al. 2018; Li et al. 2021; Sahlu et al. 2020; Soumillon et al. 2013; Trovero et al. 2020; Wichman et al. 2017). Therefore, any aberration in the expression of lncRNAs induced by SHS is likely to have an impact on testis development and spermatogenesis.

Neat1, for example, the top up-regulated noncoding transcript detected in SHS-exposed mice (log2 FC:1.4; *see*, Fig. 3C and Suppl. Table 1), is an abundant lncRNA that serves as core structural component of the paraspeckles (Adriaens et al. 2016; Clemson et al. 2009; Fox et al. 2018; Taiana et al. 2020). Paraspeckles are distinct nuclear domains, rich in RNA-binding proteins and RNA transcripts, that are thought to control gene expression through protein sequestration and retention of hyper-edited RNAs, thus promoting survival during cellular stress (Bond and Fox 2009; Yamazaki and Hirose 2015). A recent study by Major et al. has shown that spatial-temporal production of paraspeckles and paraspeckle-associated proteins (including NEAT1) contributes to diverse cellular functions during testis development and spermatogenesis (Major et al. 2019). NEAT1 has been shown to be up-regulated through binding of the p53 transcription factor, in response to several stress signals in a variety of mouse and human cell lines (Mello et al. 2017). Up-regulation of NEAT1 promotes formation of paraspeckles, which is crucial to replication stress response and chemosensitivity (Adriaens et al. 2016). Notably, NEAT1 and the adjacent noncoding RNA MALAT1, which is also over-expressed in SHS-exposed mice relative to controls (Suppl. Table 1), were shown to be up-regulated in normal human epithelial cells treated

with NNK (4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone), a potent carcinogen present in SHS as well as tobacco smoke (Silva et al. 2010). Once activated, NEAT1 may also act as a transcriptional regulator by promoting active chromatin states of specific promoters, including promoters of genes involved in cancer progression (Chakravarty et al. 2014; Major et al. 2019). NEAT1 has been shown to increase in several cancer types, including prostate cancer, and high levels of NEAT1 are often associated with poor prognosis (Chakravarty et al. 2014; Major et al. 2019).

Pathway analysis of the aberrant transcripts identified in SHS-exposed mice shows that "*cancer*", "*organismal injury and abnormalities*", "*cell-to-cell signaling*", and "*reproductive system development*" are the most affected 'function and disease' categories (Table 1). Upstream Regulator Analysis by IPA® identified HNF1A, PTEN, TGFB1, and IGF1 as the top master regulators, that are likely to account for the aberrant expression of several downstream genes in samples from SHS-exposed mice (Fig. 3B). The involvement of HNF1A, PTEN, TGFB1, and IGF1 in the morphology and development of the reproductive system is well-established (Griffeth et al. 2014; Mallik et al. 2020; Neirijnck et al. 2019; Ni et al. 2019; Pitetti et al. 2013; Selfe and Shipley 2019; Yu and Rohan 2000) (Table 2). For instance, a recent integrative analysis framework developed by Mallik et al. to investigate the molecular signatures of two testicular germ cell tumor (TGCT) subtypes: seminoma (SE) and non-seminoma (NSE), has identified the transcription factor HNF1A as one of the major hub molecules (among transcription factors, genes, and miRNAs) involved in the SE-specific regulatory network, where it may function as an oncogene (Mallik et al. 2020). The dominant functions of the IGF, PTEN, and TGF-β signaling pathways, often operating in synergy with one other, in the Sertoli cells during spermatogenesis have also been extensively described (Neirijnck et al. 2019; Ni et al. 2019; Pitetti et al. 2013). Sertoli cells are the only somatic cells in the seminiferous epithelium that sustain germ cells development by providing them with a unique microenvironment, physical support, and appropriate nutrients, growth factors and hormones, thus supporting spermatogenesis (Escott et al. 2014). Abnormal activity of these signaling pathways has been observed in testicular tumors or infertile patients, supporting a crucial role for these pathways in reproductive functions (Griffeth et al. 2014; Neirijnck et al. 2019; Ni et al. 2019; Selfe and Shipley 2019; Yu and Rohan 2000). In our study, the predicted activation of HNF1A, PTEN, TGFB1, and IGF1 is in accordance with the expression status of several downstream gene targets observed in SHS-exposed mice (Fig. 3B). Together, these findings underscore selective targeting of signaling pathways by SHS, which can have functional consequences on the reproductive system.

Another important category of disease and functions affected by SHS exposure is 'emotional behavior'. This is not surprising considering that several studies have shown the detrimental effects of SHS on the adult brain (Akhtar et al. 2013; Heffernan and O'Neill 2013a, b; Raber et al. 2021). Nonsmokers exposed to SHS are at increased risk of developing mild cognitive impairment and various forms of dementia (Chen 2012; Langa and Levine 2014). A recent study by Raber et al. has shown that chronic exposure of mice to SHS affects main brain and body functions. Cognitive injury and neuropathological changes were found to be associated with alterations of metabolic pathways including those involved in oxidative stress response, which are perturbed in human neurodegenerative disease (Raber et al.

2021). The identification of pathways associated with behavioral and emotional functions underscores another important yet unexplored feature: the incredible similarity in gene expression between brain and testis observed in both human and mouse (Guo et al. 2005; Matos et al. 2021). Notably, brain and testis are the organs that mostly exploit the potential of alternative splicing, thus expressing the largest repertoire of splice variants (Naro et al. 2021).

A novel and important finding of our study is the discovery of a unique 'frameshift' variant within the murine *Pde1a* that is exclusively and statistically significantly associated with SHS exposure (*P* value = 7.744e-06; Fig. 4A and 4B). *Pde1a* belongs to a large and diverse family of cyclic nucleotide phosphodiesterases (PDEs) that plays a crucial role in signal transduction by regulating the intracellular concentration of cyclic nucleotides through the hydrolysis of cAMP and/or cGMP to their respective nucleoside 5' monophosphates (Azevedo et al. 2014; Conti and Beavo 2007). cAMP and cGMP act as second messengers and are important in regulating the development and functional activity of spermatozoa, including capacitation, acrosome reaction, hyperactivation and motility (Hess et al. 2005; Vasta et al. 2005; Yan et al. 2001). Consistent with its role in normal testis development and spermatogenesis, Pde1a mRNA (specifically the testis-specific isoform Pde1a-203) is highly expressed in post-meiotic spermatids, with protein expression found in the tail of elongated and mature spermatids. No detectable signals were found in spermatocytes and spermatogonia (Vasta et al. 2005; Yan et al. 2001). Of relevance, *PDE1A* is one of the 336 genes (ranked # 32; standardized value 0.978) significantly associated with testicular cancer in GWAS and other genetic association datasets from the GWASdb SNP-Disease Associations dataset (Rouillard et al. 2016). Furthermore, PDE1A is among the 3,130 genes significantly associated with tobacco use disorders in GWAS and other genetic association datasets from the GAD Gene-Disease Associations dataset (no rank available) (Rouillard et al. 2016).

The SHS-associated *Pde1a* variant identified in our study is likely due to alternative splicing and consequent retention of intron 12 within the transcripts (*see*, Fig. 4C). Notably, this variant creates an in-frame PTC (UAG) immediately after the first nucleotide of the retained intron (chr2: 79,875,227; reverse strand), which disrupts the coding frame of the gene within the PDEase catalytic domain, presumably leading to NMD-activated degradation of IR-transcripts or a loss-of-function PDE1A (Fig. 4C). Comparative genomic analysis and sequence alignment have shown that the genomic sequence surrounding this PTC is highly conserved even among lower organisms (Fig. 4D), thereby suggesting a strong evolutionary pressure to preserve an early stop codon exactly at this position. Moreover, the percentage of the entire intron sequence overlaps ranges from 72.38% to 89.38% across various species. This is in accordance with the data by Galante et al., showing that sequence identity of orthologous retained introns is much higher than the average sequence identity found for orthologous non retained introns (60%) (Galante et al. 2004). Of relevance, there is a ~82% intronic sequence identity between the mouse *Pde1a* and the human *PDE1A*, which is also located on chromosome 2 at 2q32.1.

Intron retention is a form of alternative splicing that has been long overlooked as a marginal byproduct of unspliced or partially spliced pre-mRNAs. Only recently, IR has been

recognized as an important regulatory mechanism capable of fine-tuning gene expression and transcriptome diversity in a variety of cell types and tissues, including the testis (Monteuuis et al. 2019). In fact, IR has emerged as the most predominant form of alternative splicing in spermatogenesis, where it plays a crucial role in controlling the timely expression of certain genes during germ cell differentiation (Naro et al. 2017). IR is thought to regulate normal gene expression or cause disease by inserting an amino acid sequence, producing a truncated protein, or detaining a transcript in the nucleus to prevent translation (Edwards et al. 2016; Ge and Porse 2014; Monteuuis et al. 2019; Wong et al. 2017). Often, IR promotes destabilization of mRNA via NMD in the cytoplasm or other mechanisms in the nucleus. The NMD machinery degrades IR-transcripts due to the presence of PTCs introduced by the retained introns (Monteuuis et al. 2019; Zheng et al. 2020).

IR can play a crucial role in response to a variety of stress conditions (Dutertre et al. 2011). Under hypoxic conditions for example, IR can become the prevalent form of alternative splicing in cancer cells and lead to reduced expression of HDAC6 and TP53BP1, two proteins involved in cytotoxic response pathway and DNA repair (Memon et al. 2016; Monteuuis et al. 2019). A comprehensive analysis aimed at assessing the impact of a broad range of environmental perturbations on regulation of RNA processing, has recently shown that IR is a common mechanism in cells responding to environmental stimuli, likely operating through modulation of splicing factor expression (Richards et al. 2017). Of interest, Richards et al. observed a positive shift towards intron enrichment for 18 environmental stressors, including acrylamide, selenium, zinc, and copper, which are all key constituents of tobacco smoke (Richards et al. 2017). In a similar context, SHS, like its above-mentioned constituents, can trigger retention of a specific intron as a safeguard mechanism to fine tune the expression of a key gene such as *Pde1a,* and preserve normal testis function. The presence of a PTC within the splice consensus sequence GUAGGU of the intron also supports a role for SHS in the misregulation of the RNA splicing machinery. A number of studies has shown that IR can regulate gene/protein isoform production (Zheng et al. 2020) and that an inverse correlation may exist between IR and mRNA levels (Braunschweig et al. 2014; Dvinge and Bradley 2015; Edwards et al. 2016; Jung et al. 2015). In line with those reports, we observed a significant reduction in the expression of testis specific *Pde1a* transcripts (retaining intron 12 with a PTC) in SHS-exposed mice relative to controls (Fig. 5D), which is highly suggestive of NMD-activated degradation of mRNAs. Based on a widely accepted rule, NMD is predicted to be triggered when the in-frame termination codon is found, such as in this case, more than 50–55 nucleotides upstream of the final exon-exon junction (Popp and Maquat 2013).

While the novel findings of our transcriptome analysis have significant translational implications for male fertility and reproduction, we also acknowledge the limitations of the present study as no phenotypic data were collected in parallel with transcriptomic data. However, we would like to stress that SHS-induced morphological alterations in mouse testis and spermatozoa have been thoroughly investigated in a previous report (La Maestra et al. 2015) and were beyond the scope of the present study. Therefore, in line with the goal of this study and given the limited amount of testis tissues available from mice, we had to prioritize and perform genome-wide transcriptional analysis and variant detection and avoid other analyses that were beyond the scope of the study and required

additional source materials and tissues. That being said, we would like to confirm that we are planning follow up investigations in larger groups of mice to obtain sufficient quantities of source material in order to allow cellular, sub-cellular and molecular analyses which are required for simultaneous detection of genotypic and phenotypic changes consequent to SHS exposure.

5. Conclusions

In summary, we have demonstrated that exposure of mice to SHS alters the expression of genes involved in molecular pathways that are crucial for normal testis development and function. Preferential targeting of lncRNAs in the testis of SHS-exposed mice is especially significant considering their crucial role in the spatial and temporal modulation of spermatogenesis. Equally important is our discovery of a novel homozygous frameshift variant that is exclusively and significantly associated with SHS-exposure. This variant introduces a premature stop codon within the reading frame of the *Pde1a* transcripts. Since the location of the PTC is a major factor determining the fate of the transcripts, we hypothesized that SHS-induced PTC-containing transcripts are subjected to degradation by NMD. Alternatively, *Pde1a* transcripts can escape NMD and generate a loss-of-function truncated protein. Further studies are needed to elucidate the biological role of this SHS-associated intron variant and its role in the pathogenesis of reproductive diseases in humans.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We would like to thank Andrew Caliri for technical support and help with the mouse tissues. We are also grateful to Dr. Yibu Chen and Mrs. Meng Li (USC Libraries Bioinformatics Services) for assisting with data analysis.

Funding

This study was supported by grants from the National Cancer Institute of the National Institutes of Health (1R21CA268197 to AB) and the University of California Tobacco-Related Disease Research Program (26IP-0051 to ST, and 28IR-0058 and 31IR-1839 to AB). The bioinformatics software and computing resources used in the analysis are funded by the USC Office of Research and the Norris Medical Library.

Data availability

The raw RNA-seq data have been deposited in the Gene Expression Omnibus database at NCBI (https://www.ncbi.nlm.nih.gov/geo/), and accession number will be provided as soon as it becomes available.

Abbreviations

Actb	actin beta
BH	Benjamini-Hochberg
DEGs	differentially expressed genes

FC	fold change
FDR	False Discovery Rate
Gadph	glyceraldehyde-3-phosphate dehydrogenase
GATK	Genome Analysis Tool Kit
HEPA	high-efficiency particulate-air
HNF1A	transcription factor HNF1 homeobox A
IACUC	Institutional Animal Care and Use Committee
IGF1	insulin-like growth factor 1
Ints6l	integrator complex subunit 6 like gene
IPA	Ingenuity Pathway Analysis
IR	intron retention
Klk1b21	kallikrein 1-related peptidase b21 gene
MSS	mainstream smoke
Neat1	nuclear-enriched abundant transcript 1
NMD	nonsense-mediated decay
PCA	Principal Component Analysis
PDE1a	phosphodiesterase 1A
PTC	premature termination codon
PTEN	phosphatase and tensin homolog
RNA-seq	RNA-sequencing
RT-qPCR	reverse transcription-quantitative polymerase chain reaction
SHS	secondhand smoke
SNPs	single nucleotide polymorphisms
SSS	sidestream smoke
Steap1	six transmembrane epithelial antigen of the prostate 1 gene
SV	structural variant
Tbp	TATA-box binding protein
TEC	To be Experimentally Confirmed
TSP	total suspended particulate

TGFB1	transforming growth factor, beta 1
UTRs	untranslated regions
WHO	World Health Organization

References

Adriaens C, Standaert L, Barra J, Latil M, Verfaillie A, Kalev P, et al. 2016. P53 induces formation of neat1 lncrna-containing paraspeckles that modulate replication stress response and chemosensitivity. Nat Med 22:861–868. [PubMed: 27376578]

Akhtar WZ, Andresen EM, Cannell MB, Xu X. 2013. Association of blood cotinine level with cognitive and physical performance in non-smoking older adults. Environ Res 121:64–70. [PubMed: 23199696]

Andrews S, F. K, A. S-P, B L, K C, S. W 2010. Fastqc: A quality control tool for high throughput sequence data.Babraham Bioinformatics, Babraham Institute, Cambridge, United Kingdom.

ASRM PC, American Society for Reproductive Medicine. 2012. Smoking and infertility: A committee opinion. Fertility and Sterility 98.

Azevedo MF, Faucz FR, Bimpaki E, Horvath A, Levy I, de Alexandre RB, et al. 2014. Clinical and molecular genetics of the phosphodiesterases (pdes). Endocr Rev 35:195–233. [PubMed: 24311737]

Besaratinia A, Pfeifer GP. 2008. Second-hand smoke and human lung cancer. The lancet oncology 9:657–666. [PubMed: 18598930]

Bivalacqua TJ, Sussan TE, Gebska MA, Strong TD, Berkowitz DE, Biswal S, et al. 2009. Sildenafil inhibits superoxide formation and prevents endothelial dysfunction in a mouse model of secondhand smoke induced erectile dysfunction. J Urol 181:899–906. [PubMed: 19095260]

Boguslawska J, Kryst P, Poletajew S, Piekielko-Witkowska A. 2019. Tgf-beta and microrna interplay in genitourinary cancers. Cells 8.

Bond CS, Fox AH. 2009. Paraspeckles: Nuclear bodies built on long noncoding rna. J Cell Biol 186:637–644. [PubMed: 19720872]

Braunschweig U, Barbosa-Morais NL, Pan Q, Nachman EN, Alipanahi B, Gonatopoulos-Pournatzis T, et al. 2014. Widespread intron retention in mammals functionally tunes transcriptomes. Genome Res 24:1774–1786. [PubMed: 25258385]

Bult CJ, Eppig JT, Blake JA, Kadin JA, Richardson JE, Mouse Genome Database G. 2016. Mouse genome database 2016. Nucleic Acids Res 44:D840–847. [PubMed: 26578600]

Chakravarty D, Sboner A, Nair SS, Giannopoulou E, Li R, Hennig S, et al. 2014. The oestrogen receptor alpha-regulated lncrna neat1 is a critical modulator of prostate cancer. Nat Commun 5:5383. [PubMed: 25415230]

Chen R 2012. Association of environmental tobacco smoke with dementia and alzheimer's disease among never smokers. Alzheimers Dement 8:590–595. [PubMed: 22197095]

Clemson CM, Hutchinson JN, Sara SA, Ensminger AW, Fox AH, Chess A, et al. 2009. An architectural role for a nuclear noncoding rna: Neat1 rna is essential for the structure of paraspeckles. Mol Cell 33:717–726. [PubMed: 19217333]

Coggins CR. 2007. An updated review of inhalation studies with cigarette smoke in laboratory animals. Int J Toxicol 26:331–338. [PubMed: 17661224]

Conti M, Beavo J. 2007. Biochemistry and physiology of cyclic nucleotide phosphodiesterases: Essential components in cyclic nucleotide signaling. Annu Rev Biochem 76:481–511. [PubMed: 17376027]

Council NR. 2011. Guide for the care and use of laboratory animals: Eighth edition. Washington, DC:The National Academies Press.

D'Agostini F, Izzotti A, Balansky R, Zanesi N, Croce CM, De Flora S. 2006. Early loss of fhit in the respiratory tract of rodents exposed to environmental cigarette smoke. Cancer research 66:3936–3941. [PubMed: 16585223]

Dai JB, Wang ZX, Qiao ZD. 2015. The hazardous effects of tobacco smoking on male fertility. Asian J Androl 17:954–960. [PubMed: 25851659]

DeMarini DM. 2004. Genotoxicity of tobacco smoke and tobacco smoke condensate: A review. Mutation research 567:447–474. [PubMed: 15572290]

Dhillon AS, Hagan S, Rath O, Kolch W. 2007. Map kinase signalling pathways in cancer. Oncogene 26:3279–3290. [PubMed: 17496922]

Dowell RD. 2011. The similarity of gene expression between human and mouse tissues. Genome biology 12:101. [PubMed: 21241524]

Dutertre M, Sanchez G, Barbier J, Corcos L, Auboeuf D. 2011. The emerging role of pre-messenger rna splicing in stress responses: Sending alternative messages and silent messengers. RNA Biol 8:740–747. [PubMed: 21712650]

Dvinge H, Bradley RK. 2015. Widespread intron retention diversifies most cancer transcriptomes. Genome Med 7:45. [PubMed: 26113877]

Edwards CR, Ritchie W, Wong JJ, Schmitz U, Middleton R, An X, et al. 2016. A dynamic intron retention program in the mammalian megakaryocyte and erythrocyte lineages. Blood 127:e24–e34. [PubMed: 26962124]

Esakky P, Hansen DA, Drury AM, Felder P, Cusumano A, Moley KH. 2018. Testicular cells exhibit similar molecular responses to cigarette smoke condensate ex vivo and in vivo. FASEB journal : official publication of the Federation of American Societies for Experimental Biology 32:63–72. [PubMed: 28842431]

Escott GM, da Rosa LA, Loss Eda S. 2014. Mechanisms of hormonal regulation of sertoli cell development and proliferation: A key process for spermatogenesis. Curr Mol Pharmacol 7:96–108. [PubMed: 25620228]

FDA. 2012. Harmful and potentially harmful constituents in tobacco products and tobacco smoke; established list. 77 FR 20034.Food and Drug Administration.

Flurkey K, Currer JM, Harrison DE. 2007. Mouse models in aging research. In: The mouse in biomedical research (second edition), Vol. III, (Fox JG ea, editors. American College Laboratory Animal Medicine (Elsevier), ed). Burlington, MA, 637–672.

Fox AH, Nakagawa S, Hirose T, Bond CS. 2018. Paraspeckles: Where long noncoding rna meets phase separation. Trends Biochem Sci 43:124–135. [PubMed: 29289458]

Galante PA, Sakabe NJ, Kirschbaum-Slager N, de Souza SJ. 2004. Detection and evaluation of intron retention events in the human transcriptome. RNA 10:757–765. [PubMed: 15100430]

Ge Y, Porse B. 2014. The functional consequences of intron retention: Alternative splicing coupled to nmd as a regulator of gene expression. BioEssays : news and reviews in molecular, cellular and developmental biology 36.

Grabski DF, Broseus L, Kumari B, Rekosh D, Hammarskjold ML, Ritchie W. 2021. Intron retention and its impact on gene expression and protein diversity: A review and a practical guide. Wiley Interdiscip Rev RNA 12:e1631. [PubMed: 33073477]

Griffeth RJ, Bianda V, Nef S. 2014. The emerging role of insulin-like growth factors in testis development and function. Basic Clin Androl 24:12. [PubMed: 25780585]

Gunes S, Metin Mahmutoglu A, Arslan MA, Henkel R. 2018. Smoking-induced genetic and epigenetic alterations in infertile men. Andrologia 50:e13124. [PubMed: 30132931]

Guo JH, Huang Q, Studholme DJ, Wu CQ, Zhaoa SY. 2005. Transcriptomic analyses support the similarity of gene expression between brain and testis in human as well as mouse. Cytogenet Genome Res 111:107–109. [PubMed: 16103650]

Harlev A, Agarwal A, Gunes SO, Shetty A, du Plessis SS. 2015. Smoking and male infertility: An evidence-based review. World J Mens Health 33:143–160. [PubMed: 26770934]

Hartney JM, Chu H, Pelanda R, Torres RM. 2012. Sub-chronic exposure to second hand smoke induces airspace leukocyte infiltration and decreased lung elastance. Front Physiol 3:300. [PubMed: 22934051]

Hecht SS. 2005. Carcinogenicity studies of inhaled cigarette smoke in laboratory animals: Old and new. Carcinogenesis 26:1488–1492. [PubMed: 15930027]

Heffernan TM, O'Neill TS. 2013a. Exposure to second-hand smoke damages everyday prospective memory. Addiction 108:420–426. [PubMed: 22913297]

Heffernan TM, O'Neill TS. 2013b. Everyday prospective memory and executive function deficits associated with exposure to second-hand smoke. J Addict 2013:160486. [PubMed: 24804137]

Hess KC, Jones BH, Marquez B, Chen Y, Ord TS, Kamenetsky M, et al. 2005. The "soluble" adenylyl cyclase in sperm mediates multiple signaling events required for fertilization. Dev Cell 9:249–259. [PubMed: 16054031]

Hong SH, Kwon JT, Kim J, Jeong J, Kim J, Lee S, et al. 2018. Profiling of testis-specific long noncoding rnas in mice. BMC Genomics 19:539. [PubMed: 30012089]

Howe KL, Achuthan P, Allen J, Allen J, Alvarez-Jarreta J, Amode MR, et al. 2021. Ensembl 2021. Nucleic Acids Res 49:D884–D891. [PubMed: 33137190]

IARC. 2004. Tobacco smoke and involuntary smoking:Lyon, France, World Health Organization, International Agency for Research on Cancer.

IARC. 2012. Tobacco smoking, second-hand tobacco smoke, and smokeless tobacco.:Lyon, France, World Health Organization, International Agency for Research on Cancer.

Jenkins TG, James ER, Alonso DF, Hoidal JR, Murphy PJ, Hotaling JM, et al. 2017. Cigarette smoking significantly alters sperm DNA methylation patterns. Andrology 5:1089–1099. [PubMed: 28950428]

Ji G, Yan L, Liu W, Qu J, Gu A. 2013. Ogg1 ser326cys polymorphism interacts with cigarette smoking to increase oxidative DNA damage in human sperm and the risk of male infertility. Toxicol Lett 218:144–149. [PubMed: 23376476]

Jung H, Lee D, Lee J, Park D, Kim YJ, Park WY, et al. 2015. Intron retention is a widespread mechanism of tumor-suppressor inactivation. Nat Genet 47:1242–1248. [PubMed: 26437032]

Kim SI, Arlt VM, Yoon JI, Cole KJ, Pfeifer GP, Phillips DH, et al. 2011. Whole body exposure of mice to secondhand smoke induces dose-dependent and persistent promutagenic DNA adducts in the lung. Mutation research 716:92–98. [PubMed: 21925188]

Kim SI, Yoon JI, Tommasi S, Besaratinia A. 2012. New experimental data linking secondhand smoke exposure to lung cancer in nonsmokers. FASEB journal : official publication of the Federation of American Societies for Experimental Biology 26:1845–1854. [PubMed: 22318968]

Kopp F, Mendell JT. 2018. Functional classification and experimental dissection of long noncoding rnas. Cell 172:393–407. [PubMed: 29373828]

La Maestra S, De Flora S, Micale RT. 2014. Does second-hand smoke affect semen quality? Arch Toxicol 88:1187–1188. [PubMed: 24838294]

La Maestra S, De Flora S, Micale RT. 2015. Effect of cigarette smoke on DNA damage, oxidative stress, and morphological alterations in mouse testis and spermatozoa. Int J Hyg Environ Health 218:117–122. [PubMed: 25260855]

Langa KM, Levine DA. 2014. The diagnosis and management of mild cognitive impairment: A clinical review. JAMA 312:2551–2561. [PubMed: 25514304]

Laqqan M, Tierling S, Alkhaled Y, Porto CL, Solomayer EF, Hammadeh ME. 2017. Aberrant DNA methylation patterns of human spermatozoa in current smoker males. Reprod Toxicol 71:126–133. [PubMed: 28576685]

Li K, Xu J, Luo Y, Zou D, Han R, Zhong S, et al. 2021. Panoramic transcriptome analysis and functional screening of long noncoding rnas in mouse spermatogenesis. Genome Res 31:13–26. [PubMed: 33328167]

Lodovici M, Bigagli E. 2009. Biomarkers of induced active and passive smoking damage. Int J Environ Res Public Health 6:874–888. [PubMed: 19440419]

Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for rna-seq data with deseq2. Genome biology 15:550. [PubMed: 25516281]

Major AT, Hogarth CA, Young JC, Kurihara Y, Jans DA, Loveland KL. 2019. Dynamic paraspeckle component localisation during spermatogenesis. Reproduction 158:267–280. [PubMed: 31299635]

Mallik S, Qin G, Jia P, Zhao Z. 2020. Molecular signatures identified by integrating gene expression and methylation in non-seminoma and seminoma of testicular germ cell tumours. Epigenetics : official journal of the DNA Methylation Society:1–15.

Marchetti F, Rowan-Carroll A, Williams A, Polyzos A, Berndt-Weis ML, Yauk CL. 2011. Sidestream tobacco smoke is a male germ cell mutagen. Proceedings of the National Academy of Sciences of the United States of America 108:12811–12814. [PubMed: 21768363]

Marczylo EL, Amoako AA, Konje JC, Gant TW, Marczylo TH. 2012. Smoking induces differential mirna expression in human spermatozoa: A potential transgenerational epigenetic concern? Epigenetics : official journal of the DNA Methylation Society 7:432–439.

Matos B, Publicover SJ, Castro LFC, Esteves PJ, Fardilha M. 2021. Brain and testis: More alike than previously thought? Open Biol 11:200322. [PubMed: 34062096]

Matsui H, Takahashi T. 2001. Mouse testicular leydig cells express klk21, a tissue kallikrein that cleaves fibronectin and igf-binding protein-3. Endocrinology 142:4918–4929. [PubMed: 11606460]

Max W, Sung HY, Shi Y. 2012. Deaths from secondhand smoke exposure in the united states: Economic implications. Am J Public Health 102:2173–2180. [PubMed: 22994180]

Mello SS, Sinow C, Raj N, Mazur PK, Bieging-Rolett K, Broz DK, et al. 2017. Neat1 is a p53-inducible lincrna essential for transformation suppression. Genes & development 31:1095–1108. [PubMed: 28698299]

Meloche S, Pouyssegur J. 2007. The erk1/2 mitogen-activated protein kinase pathway as a master regulator of the g1- to s-phase transition. Oncogene 26:3227–3239. [PubMed: 17496918]

Memon D, Dawson K, Smowton CS, Xing W, Dive C, Miller CJ. 2016. Hypoxia-driven splicing into noncoding isoforms regulates the DNA damage response. NPJ Genom Med 1:16020. [PubMed: 28480052]

Meroni SB, Galardo MN, Rindone G, Gorga A, Riera MF, Cigorraga SB. 2019. Molecular mechanisms and signaling pathways involved in sertoli cell proliferation. Front Endocrinol (Lausanne) 10:224. [PubMed: 31040821]

Michibata H, Yanaka N, Kanoh Y, Okumura K, Omori K. 2001. Human ca2+/calmodulin-dependent phosphodiesterase pde1a: Novel splice variants, their specific expression, genomic organization, and chromosomal localization. Biochim Biophys Acta 1517:278–287. [PubMed: 11342109]

Monteuuis G, Wong JJL, Bailey CG, Schmitz U, Rasko JEJ. 2019. The changing paradigm of intron retention: Regulation, ramifications and recipes. Nucleic Acids Res 47:11497–11513. [PubMed: 31724706]

Naro C, Jolly A, Di Persio S, Bielli P, Setterblad N, Alberdi AJ, et al. 2017. An orchestrated intron retention program in meiosis controls timely usage of transcripts during germ cell differentiation. Dev Cell 41:82–93 e84. [PubMed: 28366282]

Naro C, Cesari E, Sette C. 2021. Splicing regulation in brain and testis: Common themes for highly specialized organs. Cell Cycle 20:480–489. [PubMed: 33632061]

Neirijnck Y, Kuhne F, Mayere C, Pavlova E, Sararols P, Foti M, et al. 2019. Tumor suppressor pten regulates negatively sertoli cell proliferation, testis size, and sperm production in vivo. Endocrinology 160:387–398. [PubMed: 30576429]

Ni FD, Hao SL, Yang WX. 2019. Multiple signaling pathways in sertoli cells: Recent findings in spermatogenesis. Cell Death Dis 10:541. [PubMed: 31316051]

Oberg M, Jaakkola MS, Woodward A, Peruga A, Pruss-Ustun A. 2011. Worldwide burden of disease from exposure to second-hand smoke: A retrospective analysis of data from 192 countries. Lancet 377:139–146. [PubMed: 21112082]

Piskol R, Ramaswami G, Li JB. 2013. Reliable identification of genomic variants from rna-seq data. Am J Hum Genet 93:641–651. [PubMed: 24075185]

Pitetti JL, Calvel P, Zimmermann C, Conne B, Papaioannou MD, Aubry F, et al. 2013. An essential role for insulin and igf1 receptors in regulating sertoli cell proliferation, testis size, and fsh action in mice. Mol Endocrinol 27:814–827. [PubMed: 23518924]

Polyzos A, Schmid TE, Pina-Guzman B, Quintanilla-Vega B, Marchetti F. 2009. Differential sensitivity of male germ cells to mainstream and sidestream tobacco smoke in the mouse. Toxicol Appl Pharmacol 237:298–305. [PubMed: 19345701]

Popp MW, Maquat LE. 2013. Organizing principles of mammalian nonsense-mediated mrna decay. Annu Rev Genet 47:139–165. [PubMed: 24274751]

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. 2007. Plink: A tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 81:559–575. [PubMed: 17701901]

Quinn JJ, Chang HY. 2016. Unique features of long non-coding rna biogenesis and function. Nat Rev Genet 17:47–62. [PubMed: 26666209]

Raber J, Perez R, Torres ERS, Krenik D, Boutros S, Patel E, et al. 2021. Effects of chronic secondhand smoke (shs) exposure on cognitive performance and metabolic pathways in the hippocampus of wild-type and human tau mice. Environ Health Perspect 129:57009. [PubMed: 34009016]

Reiner A, Yekutieli D, Benjamini Y. 2003. Identifying differentially expressed genes using false discovery rate controlling procedures. Bioinformatics 19:368–375. [PubMed: 12584122]

Richards AL, Watza D, Findley A, Alazizi A, Wen X, Pai AA, et al. 2017. Environmental perturbations lead to extensive directional shifts in rna processing. PLoS Genet 13:e1006995. [PubMed: 29023442]

Rinn JL, Chang HY. 2012. Genome regulation by long noncoding rnas. Annu Rev Biochem 81:145–166. [PubMed: 22663078]

Ritchie H, Roser M. 2013. “Smoking”. Published online at OurWorldInData.org. Retrieved from: ‘https://ourworldindata.org/smoking‘ [Online Resource]

Robinson JT, Thorvaldsdottir H, Winckler W, Guttman M, Lander ES, Getz G, et al. 2011. Integrative genomics viewer. Nat Biotechnol 29:24–26. [PubMed: 21221095]

Rouillard AD, Gundersen GW, Fernandez NF, Wang Z, Monteiro CD, McDermott MG, et al. 2016. The harmonizome: A collection of processed datasets gathered to serve and mine knowledge about genes and proteins. Database (Oxford) 2016.

Sahlu BW, Zhao S, Wang X, Umer S, Zou H, Huang J, et al. 2020. Long noncoding rnas: New insights in modulating mammalian spermatogenesis. J Anim Sci Biotechnol 11:16. [PubMed: 32128162]

Selfe J, Shipley JM. 2019. Igf signalling in germ cells and testicular germ cell tumours: Roles and therapeutic approaches. Andrology 7:536–544. [PubMed: 31179642]

Silva JM, Perez DS, Pritchett JR, Halling ML, Tang H, Smith DI. 2010. Identification of long stress-induced non-coding transcripts that have altered expression in cancer. Genomics 95:355–362. [PubMed: 20214974]

Soumillon M, Necsulea A, Weier M, Brawand D, Zhang X, Gu H, et al. 2013. Cellular source and mechanisms of high transcriptome complexity in the mammalian testis. Cell Rep 3:2179–2190. [PubMed: 23791531]

St Laurent G, Wahlestedt C, Kapranov P. 2015. The landscape of long noncoding rna classification. Trends Genet 31:239–251. [PubMed: 25869999]

Taiana E, Ronchetti D, Todoerti K, Nobili L, Tassone P, Amodio N, et al. 2020. Lncrna neat1 in paraspeckles: A structural scaffold for cellular DNA damage response systems? Noncoding RNA 6.

Teague SV, Pinkerton KE, Goldsmith M, Gerbremicheal A, Chang S, Jenkins RA, et al. 1994. Sidestream cigarette smoke generation and exposure system for environmental tobacco smoke studies. Inhalation Toxicology 6:79–93.

Tommasi S, Zheng A, Yoon JI, Li AX, Wu X, Besaratinia A. 2012. Whole DNA methylome profiling in mice exposed to secondhand smoke. Epigenetics : official journal of the DNA Methylation Society 7:1302–1314.

Tommasi S, Zheng A, Yoon JI, Besaratinia A. 2014. Epigenetic targeting of the nanog pathway and signaling networks during chemical carcinogenesis. Carcinogenesis.

Tommasi S, Zheng A, Besaratinia A. 2015a. Expression of epigenetic modifiers is not significantly altered by exposure to secondhand smoke. Lung Cancer 90:598–603. [PubMed: 26525280]

Tommasi S, Zheng A, Besaratinia A. 2015b. Exposure of mice to secondhand smoke elicits both transient and long-lasting transcriptional changes in cancer-related functional networks. Int J Cancer 136:2253–2263. [PubMed: 25346222]

Tommasi S, Yoon JI, Besaratinia A. 2020. Secondhand smoke induces liver steatosis through deregulation of genes involved in hepatic lipid metabolism. Int J Mol Sci 21.

Trovero MF, Rodriguez-Casuriaga R, Romeo C, Santinaque FF, Francois M, Folle GA, et al. 2020. Revealing stage-specific expression patterns of long noncoding rnas along mouse spermatogenesis. RNA Biol 17:350–365. [PubMed: 31869276]

UDHHS. 2006. The health consequences of involuntary exposure to tobacco smoke: A report of the surgeon general. Atlanta:Centers for Disease Control and Prevention, Office on Smoking and Health.

UDHHS. 2014. The health consequences of smoking: 50 years of progress: A report of the surgeon general. Atlanta:Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health.

Vasta V, Sonnenburg WK, Yan C, Soderling SH, Shimizu-Albergine M, Beavo JA. 2005. Identification of a new variant of pde1a calmodulin-stimulated cyclic nucleotide phosphodiesterase expressed in mouse sperm. Biol Reprod 73:598–609. [PubMed: 15901640]

WHO. 2018. Who global report on trends in prevalence of tobacco smoking 2000–2025, second edition. . Geneva:World Health Organization

Wichman L, Somasundaram S, Breindel C, Valerio DM, McCarrey JR, Hodges CA, et al. 2017. Dynamic expression of long noncoding rnas reveals their potential roles in spermatogenesis and fertility. Biol Reprod 97:313–323. [PubMed: 29044429]

Witschi H, Espiritu I, Dance ST, Miller MS. 2002. A mouse lung tumor model of tobacco smoke carcinogenesis. Toxicol Sci 68:322–330. [PubMed: 12151628]

Wong JJ, Gao D, Nguyen TV, Kwok CT, van Geldermalsen M, Middleton R, et al. 2017. Intron retention is regulated by altered mecp2-mediated splicing factor recruitment. Nat Commun 8:15134. [PubMed: 28480880]

Xu W, Fang P, Zhu Z, Dai J, Nie D, Chen Z, et al. 2013. Cigarette smoking exposure alters pebp1 DNA methylation and protein profile involved in mapk signaling pathway in mice testis. Biol Reprod 89:142. [PubMed: 24198121]

Yamazaki T, Hirose T. 2015. The building process of the functional paraspeckle with long non-coding rnas. Front Biosci (Elite Ed) 7:1–41. [PubMed: 25553361]

Yan C, Zhao AZ, Sonnenburg WK, Beavo JA. 2001. Stage and cell-specific expression of calmodulin-dependent phosphodiesterases in mouse testis. Biol Reprod 64:1746–1754. [PubMed: 11369604]

Yousuf H, Hofstra M, Tijssen J, Leenen B, Lindemans JW, van Rossum A, et al. 2020. Estimated worldwide mortality attributed to secondhand tobacco smoke exposure, 1990–2016. JAMA Netw Open 3:e201177. [PubMed: 32181828]

Yu H, Rohan T. 2000. Role of the insulin-like growth factor family in cancer development and progression. J Natl Cancer Inst 92:1472–1489. [PubMed: 10995803]

Zheng JT, Lin CX, Fang ZY, Li HD. 2020. Intron retention as a mode for rna-seq data analysis. Frontiers in genetics 11:586. [PubMed: 32733531]

Highlights

- There is a paucity of data on the effects of SHS on the male reproductive system.
- We investigated the effects of SHS on the testis transcriptome in a mouse model.
- Testis development & function-specific genes were dysregulated in SHS-exposed mice.
- SHS-exposed mice showed a novel intron variant in a key regulator of spermatogenesis.
- SHS exposure adversely affects reproductive system-related genes and pathways in mice.

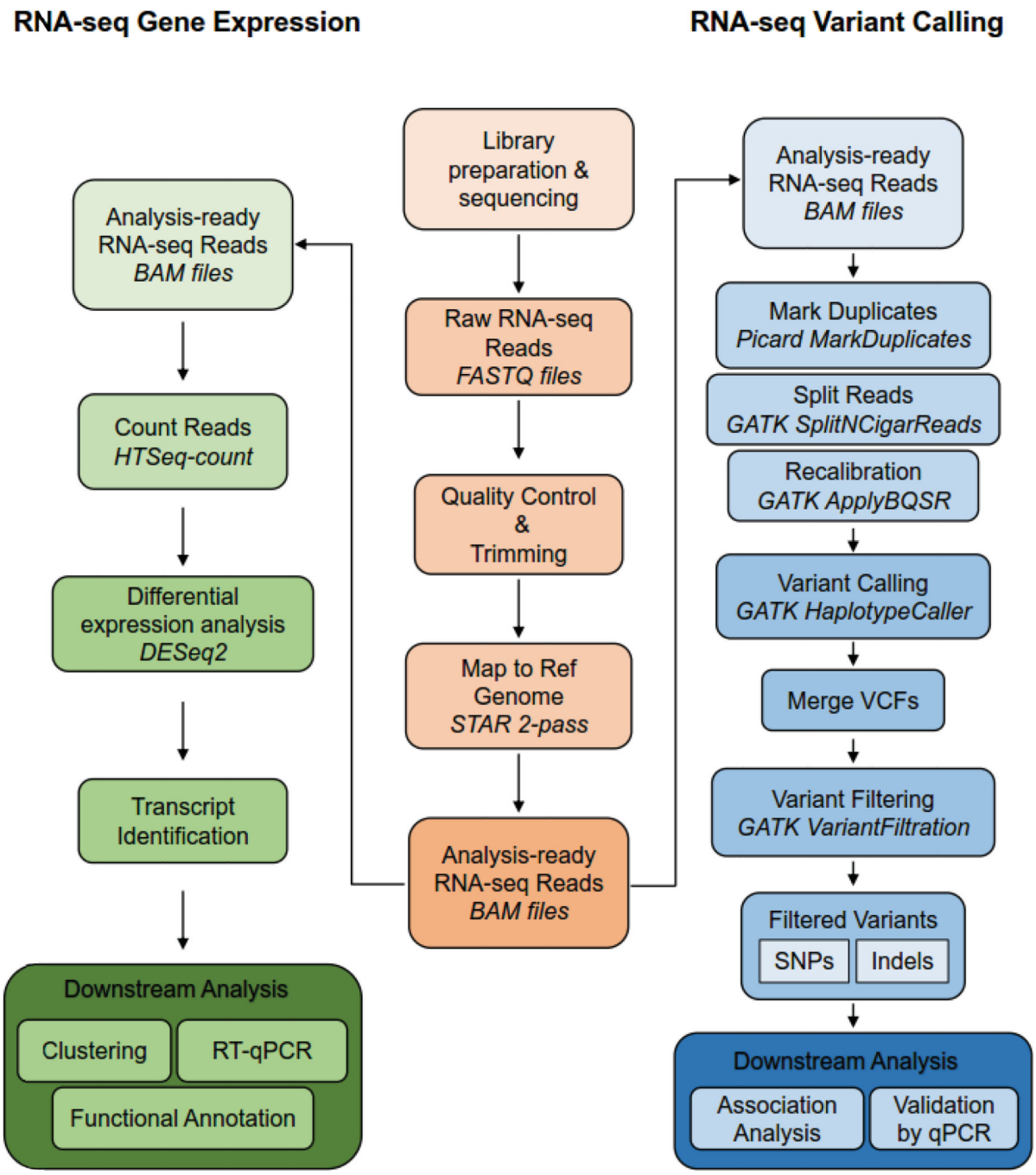


Figure 1. Flowchart of the RNA-seq based gene expression profiling and variant calling workflows.

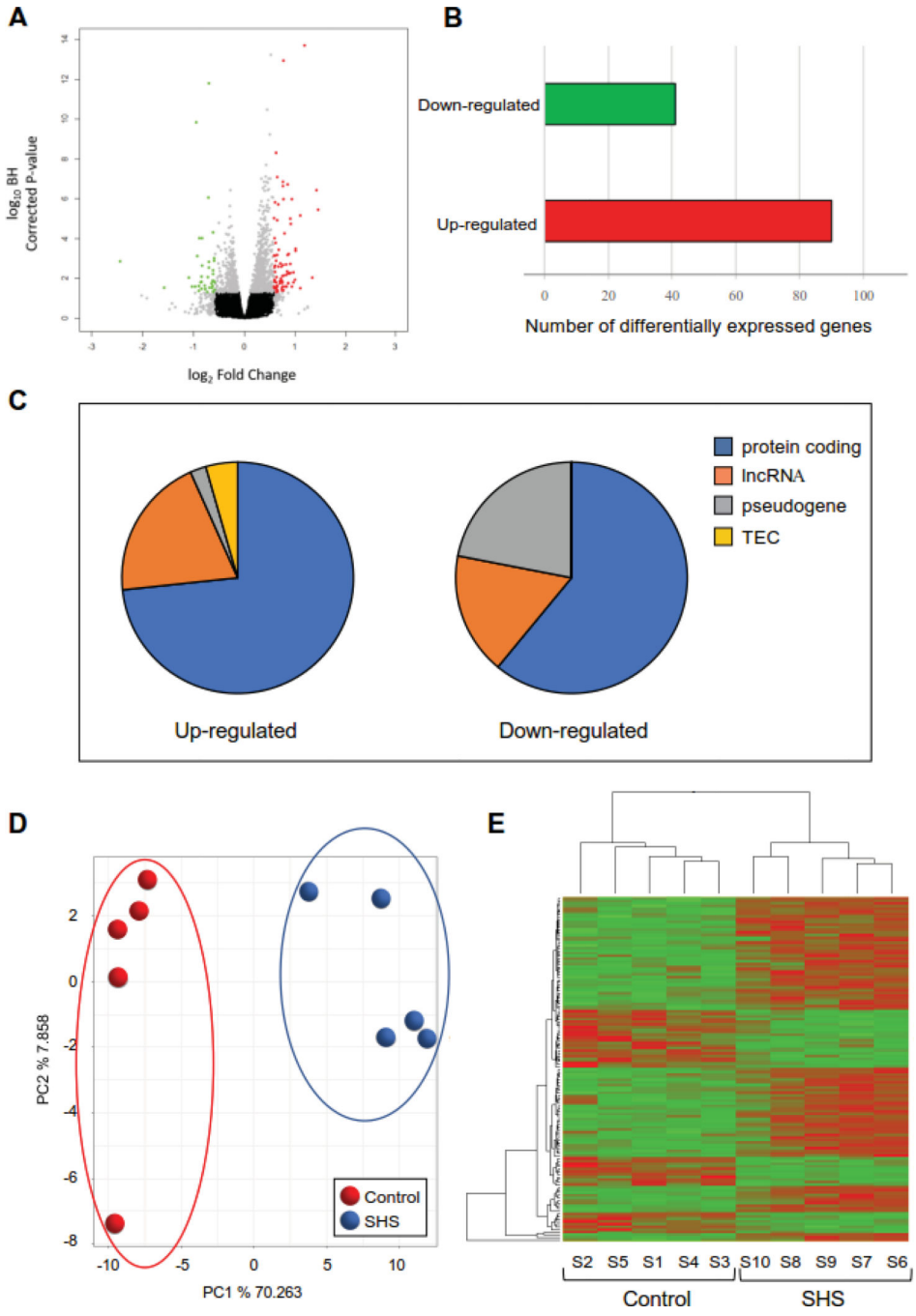


Figure 2. RNA-seq analysis in SHS-exposed mice *versus* controls.

Whole transcriptome analysis was performed on total RNA extracted from the testis of SHS-exposed mice and controls, as described in Materials and Methods. The volcano plot (A) and the bar chart (B) show the differentially expressed genes identified in SHS-exposed mice relative to controls. Fold change threshold was set to 1.5 and FDR threshold was set to 5%. Red and green dots/bars represent statistically significant up-regulated (n=90) and down-regulated (n=41) genes, respectively. C) Distribution of gene/transcript biotypes (based on the Ensembl classification) in up- and down-regulated targets. The majority of DEGs is

composed of protein-coding genes (69.5% of total DEGs), with lncRNAs being overall the second largest category (19.1% of total DEGs). D) Principal Component Analysis (PCA) of RNA-seq data obtained from SHS-exposed mice and controls. The PCA and k-means clustering (k=2) based on PC1 and PC2 coordinates, clearly separate the treatment (in blue) from the control group (in red). E) Hierarchical clustering of the 131 DEGs identified in the testis of SHS-exposed mice *versus* controls. Genes are depicted in rows and samples in columns. Up-regulated genes are shown in red, while down-regulated genes are highlighted in green. As for the PCA, hierarchical clustering showed high reproducibility between biological replicates and distinct clustering of the treatment *versus* the control group.

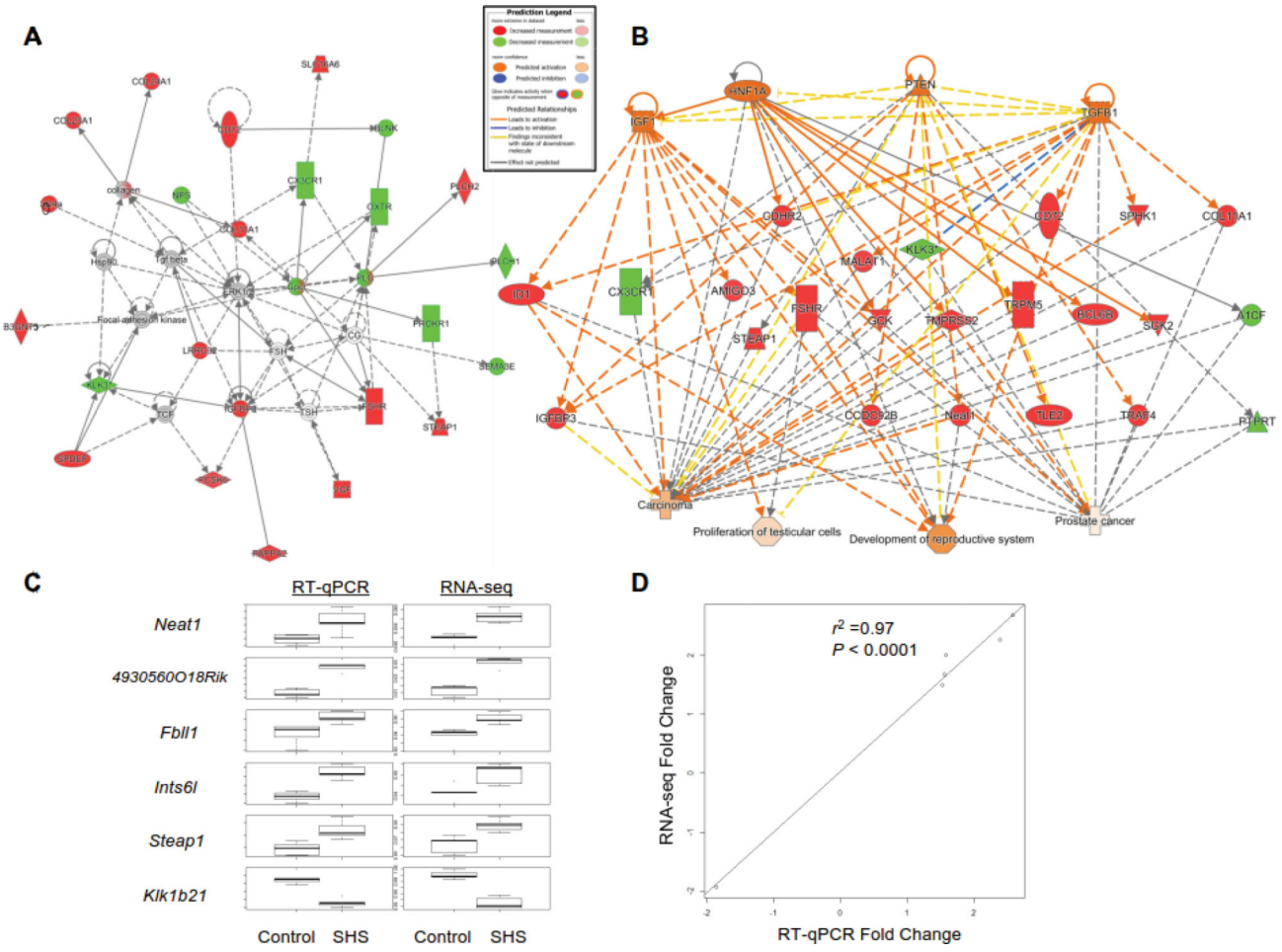


Figure 3. Functional network analysis and gene validation by RT-qPCR.

A) Gene network analysis by IPA® showed that the most affected functional network included DEGs that converge into the ERK1/2 and the TGF-β pathways. Red and green nodes represent up-regulated and down-regulated DEGs, respectively. White nodes show molecules that are not included in the dataset but interact with the other components of the network. Solid line, direct interaction; dashed line, indirect interaction. B) The Upstream Regulator Analysis function in IPA® was used to identify the upstream regulators that are likely to account for the transcriptional changes observed in the testis of SHS-exposed mice. Based on a prediction *z*-score > 2, we identified a quartet of molecules (HNF1A, IGF1, PTEN and TGFB1) as the top master regulators. The predicted networks regulated by HNF1A, IGF1, PTEN, and TGFB1, their targeted genes and downstream biological effects are shown. Orange nodes indicate prediction of activation. Red and green nodes represent up-regulated and down-regulated DEGs, respectively. For more indicators, please refer to the Prediction Legend. C) Expression levels of randomly selected up- and down-regulated genes. The box plots show the median normalized expression ± Standard Deviation (±SD) of *Neat1*, *4930560O18Rik*, *Fbll1*, *Ints6l*, *Steap1*, and *Klk1b21*. All gene targets examined by RT-qPCR (left panel) showed the same pattern of differential expression detected by RNA-seq (right panel). D) Normalized read counts for each target gene in RNA-seq data

were correlated to the median normalized expression levels of the corresponding gene determined by RT-qPCR using the Pearson's correlation analysis.

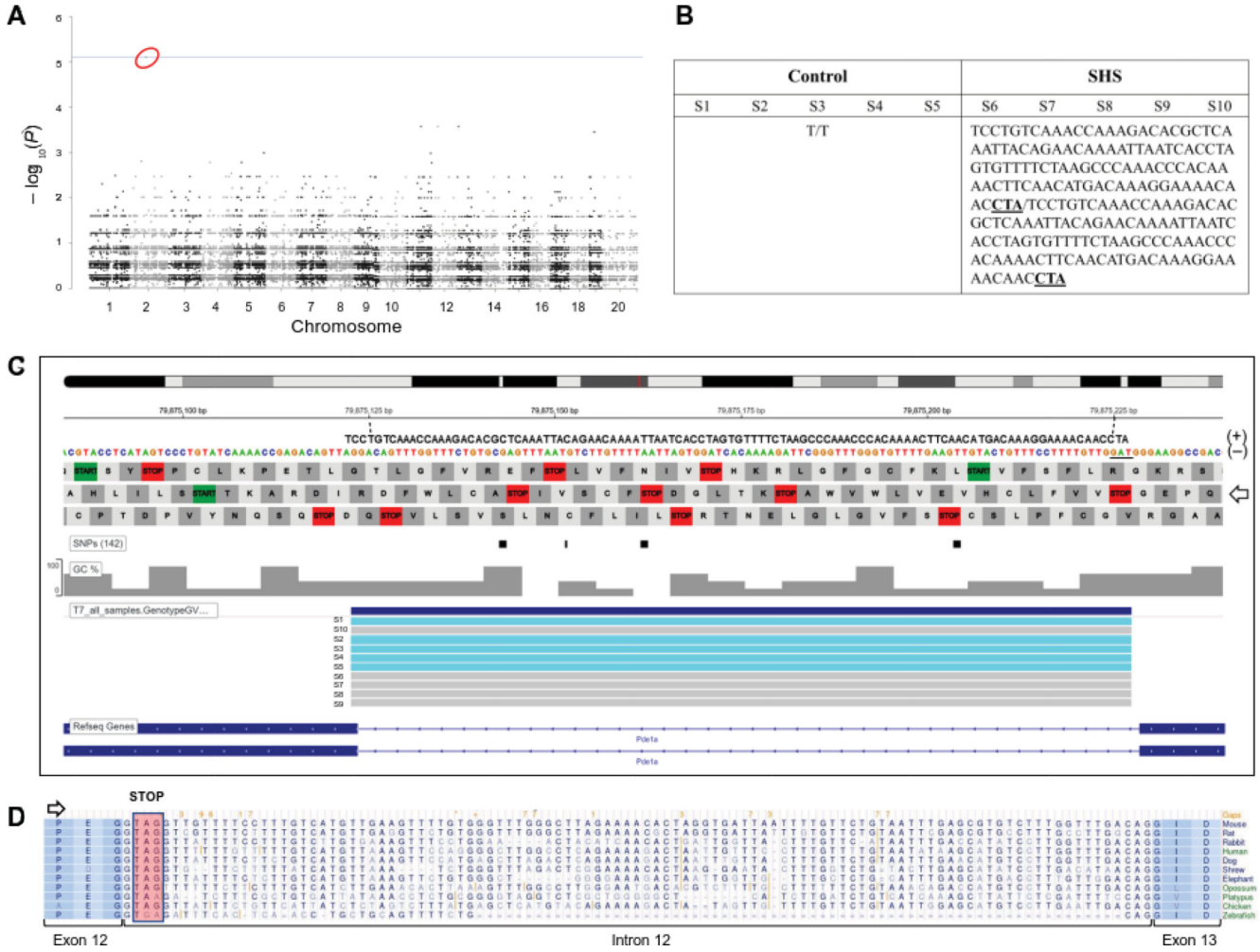


Figure 4. Detection of a *Pde1a* frameshift variant strongly associated with SHS exposure.

The GATK Best Practices workflow was used for variant discovery analysis. A) Manhattan plot showing the statistical significance of the association with SHS exposure, as –log10(p-value) on the y-axis against chromosomes on the x-axis. A red circle shows the *Pde1*a variant identified on chromosome 2 (*P* value = 7.744e-06). B) Genotypes detected in SHS-exposed mice and controls using a combination of GATK and plink association analysis. The sequences for the reference and alternate allele are conventionally shown on the forward strand (+). S1 to S5, samples from control mice; S6 to S10, samples from SHS-exposed mice. C) The “CombineGVCFs” file displaying sequence variations was uploaded into the ‘Integrated Genomic Viewer’ (IGV_2.9.4) for visualization of the variants. The location and sequence of the variant, which corresponds to a 106 bp-retained intron, at position chr2: 79875122–79875227 within the *Pde1a* gene, is shown. The retained intron creates an in-frame stop codon immediately after the first nucleotide of the intron (GTAG, underlined) on the reverse (−) strand where *Pde1a* is encoded. The direction of the *Pde1a* reading frame is indicated by a white arrow. Genotypes for each sample are shown by horizontal bars in IGV. *Magenta*, [(T)/(T)-n] genotype; *gray*, [(T)CCTGT…ACCTA/(T)CCTGT…ACCTA-n] genotype. D) Multiple alignments were generated using multiz in the UCSC comparative genomics alignment pipeline. The genomic sequence surrounding the PTC (boxed) is highly

conserved even among lower organisms. The double lines indicate aligning species with one or more unalignable bases in the gap region. The numbers and symbols on the Gaps line indicate the lengths of gaps in the mouse sequence at those alignment positions relative to the longest non-mouse sequence. If there is sufficient space in the display, the size of the gap is shown. If the space is insufficient and the gap size is a multiple of 3, a “*” is displayed; other gap sizes are indicated by “+”.

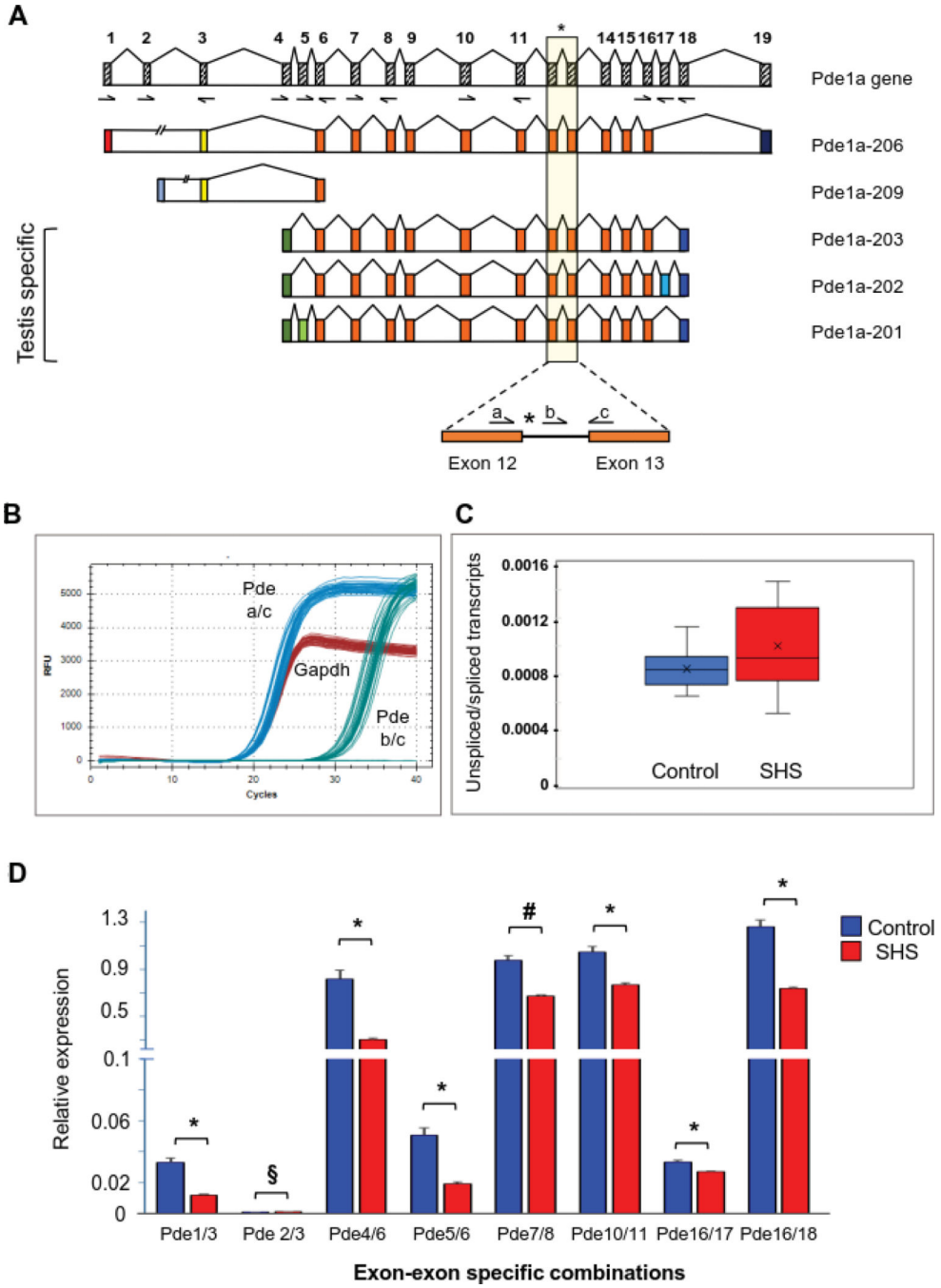


Figure 5. Analysis of IR-transcripts and relative expression of *Pde1a* gene isoforms

A) Ten distinct transcripts, resulting from differential promoter usage and alternative splicing, have been described for the murine *Pde1a* gene (Ensembl, Pde1a-201 to Pde1a-210; *see*, Suppl. Fig. 1). For clarity, only the long isoform (Pde1a-206), the processed transcript (Pde1a-208), and the testis-specific isoforms (Pde1a-201, −202, −203) are shown. Coding exons are numbered based on their genomic location on chromosome 2 (hatched boxes). Several primer sets (arrows) were designed to detect common and alternate regions of the different *Pde1a* transcripts, as described in ref. (Vasta et al. 2005). A detailed list of

primers is provided in Supplementary Table 2 (Suppl. Table 2). The area overlapping exon 12 to exon 13 is highlighted, and the retained intron 12 is indicated by an *. B) RT-qPCR was performed to analyze IR-transcripts in SHS-exposed mice and controls. The primers *a-c* or primers *b-c* were used to amplify the exon12-exon13 junction (intron 12 spliced out) or the intron12-exon13 junction (intron 12 retained), respectively. C) The level of IR-transcripts was calculated by taking the ratio of the unspliced transcripts to the spliced transcripts in SHS-exposed mice relative to controls. Distribution of data within each group is shown by box and whisker plots; the 'lower' and 'upper' edges of boxes represent the 1st and 3rd quartiles, respectively (25 and 75 percentiles, resp.). Horizontal lines within the boxes represent the medians (2nd quartile or 50 percentile). The 'lower' and 'upper' vertical lines extending from the boxes, also known as the "whiskers", represent the lowest and highest data points, respectively, excluding any outliers (minimum and maximum values, resp.). $P < 0.215$ by Mann-Whitney U Test. D) The relative expression of several *Pde1a* transcripts was calculated in SHS-exposed mice and controls by RT-qPCR. Exon-specific primer sets were used, as outlined in (A). The data represent the means of five samples per group, with each sample tested in triplicate, ± SE. *Gapdh* was used as the reference gene. * P value < 0.00001, § P value = 0.00022, # P value = 0.0001.

Table 1.

Categories and functions affected by exposure to SHS.

Categories	Diseases or Functions Annotation	B-H p-value	Biascorrected z-score
Cancer, Organismal Injury and Abnormalities	Cancer of cells	0.0468	1.922
Cancer, Organismal Injury and Abnormalities	Neoplasia of cells	0.0468	1.728
Cancer, Organismal Injury and Abnormalities	Epithelial neoplasm	0.041	1.327
Organismal Injury and Abnormalities	Benign lesion	0.0474	1.324
Cancer, Organismal Injury and Abnormalities	Non-melanoma solid tumor	0.0468	1.115
Cancer, Organismal Injury and Abnormalities	Carcinoma	0.0468	1.112
Behavior	Emotional behavior	0.0148	0.934
Cancer, Organismal Injury and Abnormalities	Adenocarcinoma	0.00354	0.926
Cell-To-Cell Signaling and Interaction	Binding of breast cancer cell lines	0.0404	0.497
Reproductive System Development and Function	Litter size	0.0191	0.147

To sort the results, Benjamini-Hochberg (BH) corrected FDR was set to 5%.

Table 2.

Diseases and/or functions impacted in SHS-exposed mice of relevance to the reproductive system.

Categories	Diseases or Functions Annotation	p-Value	Molecules
Cancer, Organismal Injury and Abnormalities	Genitourinary carcinoma	1.82E-05	A1CF, BCL6B, COL11A1, FSHR, GCK, HNF1A, IDI, IGF1, IGFBP3, KLK3, MALATI, PTEN, PTPRT, SPHK1, TGFB1, TMPRSS2, TRAF4, TRPM5
Cancer, Organismal Injury and Abnormalities, Reproductive System Disease	Prostate cancer	1.04E-05	Al CE, COLI 1A1, FSHR, GCK, HNF1A, IDI, IGF1, IGFBP3, KLK3, PTEN, PTPRT, TGFB1, TMPRSS2, TRAF4
Reproductive System Development and Function	Morphology of reproductive system	1.33E-05	FSHR, HNF1A, IGF1, KLK3, Neatl, PTEN, TGFB1
Organismal Development, Reproductive System Development and Function	Development of reproductive system	2.38E-05	FSHR, HNF1A, IDI, IGF1, Neatl, PTEN, TGFB1
Organismal Injury and Abnormalities, Reproductive System Development and Function, Reproductive System Disease	Abnormal morphology of reproductive system	1.81E-05	FSHR, HNF1A, IGF1, KLK3, Neatl, PTEN
Cell Signaling	Protein kinase cascade	1.27E-05	FSHR, IDI, IGF 1, IGFBP3, TGFB1, TRAF4
Cellular Growth and Proliferation, Organ Development, Reproductive System Development and Function	Proliferation of gonadal cells	2.7E-06	FSHR, IGF1, PTEN, TGFB1
Organ Development, Reproductive System Development and Function	Growth of testis	8.06E-07	FSHR, IGF1, TGFB1
Cellular Development, Cellular Growth and Proliferation, Organ Development, Reproductive System Development and Function	Proliferation of prostate cells	3.42E-05	PTEN,TGFB1
Cellular Development, Cellular Growth and Proliferation, Organ Development, Reproductive System Development and Function, Tissue Development	Proliferation of Leydig precursor cells	1.09E-05	IGF1,TGFB1

*The upstream master regulators predicted to be activated in IPA® are highlighted in red.