



Linschooten, Joost O, Van Schooten, Frederik J, Baumgartner, Adi
ORCID logoORCID: <https://orcid.org/0000-0001-7042-0308>,
Cemeli, Eduardo, Van Delft, Joost, Anderson, Diana and
Godschalk, Roger WL (2009) Use of spermatozoal mRNA profiles
to study gene-environment interactions in human germ cells.
Mutation Research/Fundamental and Molecular Mechanisms of
Mutagenesis, 667 (1-2). pp. 70-76.

Downloaded from: <https://ray.yorks.ac.uk/id/eprint/6053/>

The version presented here may differ from the published version or version of record. If
you intend to cite from the work you are advised to consult the publisher's version:

Research at York St John (RaY) is an institutional repository. It supports the principles of
open access by making the research outputs of the University available in digital form.
Copyright of the items stored in RaY reside with the authors and/or other copyright
owners. Users may access full text items free of charge, and may download a copy for
private study or non-commercial research. For further reuse terms, see licence terms
governing individual outputs. [Institutional Repositories Policy Statement](#)

RaY

Research at the University of York St John

For more information please contact RaY at
ray@yorks.ac.uk

Use of spermatozoal mRNA profiles to study gene-environment interactions in human germ cells

Joost O. Linschooten¹, Frederik J. Van Schooten¹, Adolf Baumgartner², Eduardo Cemeli², Joost van Delft¹, Diana Anderson² and Roger W.L. Godschalk^{1*}

¹Department of Health Risk Analysis and Toxicology, Maastricht University, PO Box 616, 6200 MD Maastricht, the Netherlands and ²Division of Biomedical Sciences, University of Bradford, Bradford, BD7 1DP, UK

Keywords: paternal exposure, spermatozoal mRNA, gene expression

*To whom correspondence should be addressed:

Roger W.L. Godschalk

Department of Health Risk Analysis and Toxicology

Maastricht University

PO Box 616

6200 MD Maastricht

The Netherlands

Tel: +31 43 3881104

Fax: +31 43 3884146

Email: R.Godschalk@grat.unimaas.nl

Abstract

Paternal exposure to genotoxic compounds is thought to contribute to diseases in their offspring. Therefore, it is of importance to develop biomarkers of male germ cell exposure to genotoxins. Unfortunately, the testis cannot be reached for routine biomonitoring, but mRNA-profiles in spermatozoa may reflect the processes that have occurred in the testis after exposures to genotoxins, since spermatozoa are largely transcriptionally inactive. Therefore, mRNA profiles from sperm in ejaculates of cigarette smokers (N=4) were compared with unexposed controls (N=4). Smoking behaviour was verified by assessing cotinine levels in seminal plasma. High expression of the germ cell specific gene *protamine 2* (*PRM2*) was observed in spermatozoal mRNA isolates by Q-PCR, which was absent in reference mRNA isolates obtained from a pool of other organs. Gene-expression analysis was subsequently performed using microarray technology and a total of 781 genes were found to be differentially expressed in spermatozoa of smokers compared to non-smokers (fold change >40%; $p < 0.05$). To further limit the number of false positive results, genes were additionally selected on basis of their correlation with cotinine levels in seminal plasma ($r > 0.80$ as arbitrary cut-off value, $p < 0.05$), and a total of 197 transcripts remained, of which the germ cell specific transcription factor *SALF* was the highest up-regulated gene (5.4-fold) and the zinc finger encoding gene *TRIM26* most down regulated (7.4 fold). Although no altered pathways could be identified for the differentially expressed genes, an enrichment was observed for NF- κ B regulated genes (43% vs. 26%, $P = 0.003$) playing a central role in stress response. Indeed, subsequent analysis of transcription factor networks suggests that apoptosis was inhibited in smokers. These data show the feasibility of using gene-expression profiles in mature sperm to elucidate gene-environment interactions in male testis.

Introduction

Childhood cancer incidences have increased over the last few decades [1], suggesting a prominent role of environmental exposures to the parents in disease aetiology. We all are exposed to a variety of environmental genotoxins, such as constituents of cigarette smoke, exhaust fumes, and food contaminants. Most research has focussed on the effects of such exposures on somatic cells [2], but far less research has been performed on the effect of exposures to genotoxins in relation to parental germline mutations and possible health effects in the next generation. Genotoxic effects in the germline have been described for ionising radiation [3], and there is growing evidence that environmental / dietary exposures to chemicals can also be involved in the induction of heritable mutations [4].

Epidemiological studies have shown that exposure to tobacco smoke and environmental tobacco smoke during pregnancy increases the risk of severe health impairments in the newborn children [5, 6]. However, few studies have investigated the effect of cigarette smoking of the father [7] before conception. These studies have found no overall effect of paternal smoking, though some report a positive association between paternal smoking and the development of childhood cancer. Although the effect of parental exposures to chemical carcinogens seems to be small in general [8], it may still be particularly relevant for a susceptible subgroup of individuals [9].

Most germline mutations seem to arise after paternal exposure and indeed, several environmental and food contaminants are known to reach the testis in significant concentrations, such as polycyclic aromatic hydrocarbons (PAH) and the food-derived carcinogen acrylamide [10,11]. Further studies of the relationship between exposure to environmental genotoxins and the formation of germ line mutations require a reliable assessment of the exposure in the testis and subsequent biological effects in the testis. Assessment of external exposures is probably insufficient, because many individual factors may affect the dose that actually reaches the male germ cells. Exposure to genotoxic agents may cause differences in gene expression in several tissues as compared to unexposed subjects. For example, expression of the cytochrome P450 proteins (CYP's) 1A1 and 1B1 is higher in subjects exposed to PAH [12]. The fact that exposure to genotoxins can affect the tissue-specific regulation of gene expression [13] suggests a possible use of gene expression profiles as biomarkers of exposure of that particular organ / tissue. Since many internal organs cannot be reached in a non-invasive manner, these studies rely on surrogate

tissues, such as peripheral blood. In theory, the testis can also only be reached in an invasive manner. However, since the discovery of mRNA in mature spermatozoa, there is a growing interest in results obtained from these cells [14]. Ostermeier *et al.* described the absence of ribosomal subunits in the RNA obtained from spermatozoa, which suggests that spermatozoa are largely transcriptionally inactive and mRNA isolated from these cells will thus reflect processes that have taken place earlier in the testis [15]. Some genes are specifically transcribed in this tissue during spermatogenesis, for example genes involved in histone replacement, like *protamine 1*, *2* and *3*, during spermatogenesis [16]. Together with a different expression of several sperm specific proteins, it was also reported that some cell specific transcription factors are expressed during spermatogenesis [17].

mRNA profiles in mature spermatozoa may thus reflect the testicular response to exposures to genotoxins, representing an interesting potential retrospective biomarker of exposure as spermiogenesis and sperm maturation in the epididymis lasts for approximately 45-50 days in men. The aim of this study is therefore to test whether gene expression profiles in spermatozoa can be considered as new biomarkers for exposure of the testis to genotoxic compounds.

Material and methods

Study population

Four healthy smoking and 4 healthy non-smoking subjects gave their informed consent to provide an ejaculate by masturbation. Smoking behaviour was assessed by a questionnaire and by cotinine levels in seminal plasma (see below). Each semen sample was processed within 2 hours and analyzed for volume, number of spermatozoa per ml, motility, pH and morphology according to WHO criteria (1999). Semen samples were snap-frozen in liquid nitrogen and subsequently stored in aliquots at -80°C until analysis.

Cotinine assessment

Sperm samples were thawed at room temperature and centrifuged at 10,000 rpm for 5 min. Supernatants were transferred to clean tubes, and cell pellets were used for RNA isolation. Seminal plasma was stored at -20°C until cotinine levels were assessed by a radio-immuno assay according to the method described by van Vunakis *et al.* [18]. Cotinine levels in the seminal plasma were expressed in ng/ml.

RNA isolation from spermatozoa and cDNA synthesis

Before RNA isolation from spermatozoa, somatic cells in the ejaculate were lysed using SDS and Triton-X, as described in Goodrich *et al.* [19] with minor modifications. After first lysis of the somatic cells, a small amount of the cell suspension was transferred onto a slide and stained with Giemsa to check whether lysis of somatic cells was successful. Lysis was followed with a washing step in lysis buffer, and the cell pellets were subsequently resuspended in Trizol solution (Invitrogen, UK). RNA was isolated from these solutions using a RNeasy Minelute kit (Qiagen Westburg bv., Leusden, the Netherlands), according to the producers' manuals. mRNA quantity was measured spectrophotometrically using a Nanodrop spectrophotometer (Nanodrop technologies, Wilmington, USA) and stored at -80°C until use for microarray and real time polymerase chain reaction (RT-PCR). The quality of the RNA was checked using an Agilent BioAnalyzer (Agilent Technologies, Breda, The Netherlands). In contrast to RNA from other cell types, the RNA isolated from sperm cells is fragmented and does not contain the 18S and 28S ribosomal subunits. Nonetheless, this RNA could be used for cDNA synthesis and microarray analysis. From each

sample, 500ng of RNA was used for cDNA synthesis in combination with the iScript cDNA Synthesis Kit (Bio-Rad Laboratories, Hercules, CA)

Microarray

The arrays were hybridised using two groups, smokers and non-smokers. All samples were compared to a common reference sample obtained from one extra non-smoking donor. The subjects were matched for age, caffeine intake and alcohol use. Arrays were performed using Agilent human 4x44K arrays (Agilent technologies), according to the manufacturers' instructions. The microarray slides were scanned on a GenePix 4000B (Molecular Devices, Sunnyvale, CA). Processing of array data was done as described by Staal *et al.* [20]. The images were processed with ImaGene 8.0.1 software (Biodiscovery, Los Angeles, CA) to quantify spot signals. Flagged spots, consisting of poor quality spots and negative and positive control spots, were excluded. Data from ImaGene were transported to GeneSight software version 4.1.6 (Biodiscovery) for transformations and normalisations. For each spot, median local background intensity was subtracted from the median spot intensity and spots from low expression genes (with a net intensity of <10 in both channels), were excluded from further analysis. The corrected median intensities were log base 2 transformed and normalised using the LOWESS algorithm. The gene expression ratios were imported into Excel (Microsoft Corporation, Redmond, WA) for further analysis, where genes with availability of 100% of the values per gene were used.

Real time PCR

To validate a selection of genes from the microarray experiments, quantitative real time polymerase chain reaction (RT-PCR) was performed. For cDNA synthesis, 200ng of each sample was used in combination with the iScript cDNA Synthesis Kit (Bio-Rad Laboratories, Hercules, CA). RT-PCR reactions were carried out using iQ SYBR Green Supermix, which contains iTaq DNA Polymerase, deoxynucleoside triphosphates, MgCl₂ and SYBR Green I (Bio-Rad laboratories). The cDNA was 10 times diluted before analysis and an amount of 5µl sample was added to each reaction well, and all samples were measured in duplicate. Samples were analysed on a MyiQ Single-Colour Real-Time PCR detection System (Bio-Rad laboratories), using the following parameters: 3min at 95°C, 40 cycles at 95°C for 15s and at 60°C for 45s.

Expression of only a few genes could be verified, due to the limited amount of mRNA available from the spermatozoa. Several toxicologically relevant genes (including *CYP1A1* and *CYP1B1*) were selected for validation by Q-PCR as potential biomarkers of exposure. To normalize the amount of mRNA's, the housekeeping gene *β -actin* was used, and each sample was tested in duplicate. *Protamine 2*, a sperm specific protein, was selected as potential additional housekeeping gene. This protein is involved in replacement of histones during spermatogenesis, and there is no current evidence that expression of this gene is altered after exposure to environmental factors. The normalized values from *β -actin* and *PRM2* were used to define differences in gene expression levels in smokers and non-smokers.

Statistical analysis

Differences in gene expression were calculated and assessed as significant when the fold change is higher than 40% (i.e., 0.5 on a log base 2 transformed scale) and $p < 0.05$, using Student's t-Test. Thereafter, data of the highest up- or down-regulated genes were Pearson-correlated with the log-transformed cotinine levels in the seminal plasma, to limit the number of false-positive results. Fatigo+ was used for analysis of transcription factors, with search criteria of 1 kb upstream from the gene. T-Profiler changes were scored using t-test's, in the average activity of predefined groups of genes, as described in Boorsma *et al.* [21]. MetaCore from GeneGo Inc. (<http://www.genego.com>) was used to study transcription factor networks with the available expression data. Each network analysis is based on one specific transcription factor that has target genes in the list of differentially regulated genes, together with genes from that same list that directly influence activation or inhibition of this particular transcription factor. The generated networks centred on one transcription factor, are subsequently interpreted in terms of GO-processes.

Results

Characteristics of study population

Non-smoking (N=4) and smoking (N=4) subjects were matched for age, ethnicity and the weekly usage of alcohol or caffeine containing beverages. Matching for alcohol and caffeine consumption was necessary, since both caffeine and alcohol can affect sperm quality [22-26]. The characteristics of the subjects are presented in Table 1. According to WHO-criteria, the sperm count of smokers was lower than that of non-smokers, but this difference did not reach statistical significance. There were no significant differences in sperm motility and morphology. Smoking was assessed by questionnaire, but no detailed information was available on the amount of cigarettes smoked per day. Nonetheless, the cotinine levels in the seminal plasma showed a clear difference in exposure to nicotine, since these levels were over 30 times higher in the smoking subjects than in non-smoking volunteers.

RNA quantity and quality control

The amount of mRNA isolated from the semen samples was low, which was expected since these cells contain less mRNA compared to somatic cells (approximately 2000 times lower) [27]. For this reason it is of great importance that somatic cells were removed before the spermatozoa were lysed to release the mRNA, since this could negatively influence gene expression data. To check whether lysis of these cells was successful, a small amount of each sample was microscopically checked. It appeared that hardly any somatic cells were still present and therefore it is expected that the contribution of somatic cell RNA is << 1% to the total amount of RNA isolated. The BioAnalyzer data showed that the spermatozoal mRNA is highly fragmented and the 18S and 28S ribosomal subunits are absent, which is typical for this cell type (as illustrated in Figure 1).

Microarray analysis

Of the 44,000 genes on the slides, 13,994 were detected on all 8 arrays. These data showed that a total of 781 genes were differentially expressed in smokers compared to non-smokers (fold change >40%; $p < 0.05$). This list of genes was correlated to the log-transformed cotinine values in seminal plasma, which resulted in a large variation in correlation coefficients, ranging from 0.48 up to 0.99. Of the 781

differentially regulated genes, 197 showed a significant correlation of $r > 0.80$ with the cotinine levels and only 21 showed a correlation coefficient higher than $r > 0.90$ (these 21 genes are presented in Table 2). *SALF*, a germ cell specific transcription factor, showed the highest fold change, in combination with the highest correlation with log-transformed cotinine levels in the seminal plasma. The most strongly down regulated gene was found to be the tripartite motif-containing gene *TRIM26*, which encodes three zinc-binding domains.

Analysis of pathways and transcription factors

The list of 13,994 genes without missing values on all 8 arrays was uploaded into T-Profiler, a pathway analysis tool that focuses on shifts in total gene expression profiles, not only in that of differentially expressed genes [21]. The analysis showed that modulated genes are typical of testis/germ cell related processes, which provides evidence for the specificity of the spermatozoal RNA. Unfortunately, no differentially regulated pathways were detected by this tool. Accordingly, no exposure-related pathways were identified by several other 'pathway finding' programmes that focus only on the differentially regulated genes, like Fatigo+ [28] and Metacore (<http://www.genego.com>).

However, the analysis of hypothetical transcription factor binding sites up to 1kb upstream of the group of differentially regulated genes showed that a higher percentage of these genes were regulated by NF- κ B as compared to the remaining genes on the array (43.1% vs. 26.3%, $p=0.003$). A similar effect was found for forkhead transcription factors like FOX (15.7% vs. 6.5%, $p=0.006$) and transcription factor YY1 (5.8% vs. 0.4%, $p=0.008$). These transcription factors remained significantly enriched after the cut-off value for the correlation with cotinine was increased from $r > 0.80$ to $r > 0.90$ (see Figure 2).

Since some specific transcription factors seemed differentially regulated, a transcription factor network analysis was performed (using MetaCore). Anti-apoptosis proved to be the main significant process in the gene list, which was confirmed by the down regulation of genes involved in the induction of proliferation and apoptosis in smokers, e.g. *cyclin D1* (fold change -2.24; $p=0.042$), *cyclin D3* (fold change -3.06; $p=0.008$) and *FasL* (fold change -2.99; $p=0.024$), whereas anti-apoptotic genes were up regulated, e.g. *Annexin A5* (fold change 1.73; $p=0.043$) and *NAP1* (fold change 2.12; $p=0.002$).

Real time PCR

High expression of the sperm specific gene *protamine 2* (*PRM2*) was observed in spermatozoal mRNA by Q-PCR, which was absent in reference mRNA obtained from a pool of mRNA from 10 different human cell lines (Stratagene, La Jolla, CA, cat.nr. 750500), which can be regarded as further evidence that the RNA was isolated from spermatozoa without a major contribution of somatic cells. Since *SALF* was highly differentially regulated, according to microarray data, this gene was also selected to be validated on Q-PCR. The expression of this germ cell specific transcription factor indeed shows the same direction of regulation on both microarray and real time PCR (see Table 3).

The cytochrome P450 enzymes 1A1 and 1B1 known to be up regulated by exposure to aryl hydrocarbon (Ah) receptor agonists like PAH (PAH are constituents of cigarette smoke) and were therefore included in the validation by real time PCR. From the microarray data, it could be expected that these genes were not significantly up or down regulated. Indeed, using real time PCR, these genes show only a marginal increase in gene expression.

Discussion

Unfortunately, internal organs such as the testis cannot be reached for biomonitoring studies, and researchers have to rely on surrogate tissues. Recent insights into the presence of mRNA in mature spermatozoa obtained from an ejaculate [14] open new possibilities for the assessment of gene expression profiles that can be relevant for germ cell exposure and subsequent health effects in the offspring. In the present work, we showed that it is possible to isolate mRNA from spermatozoa which represents germ cell and testis related processes. Gene-environment interactions largely depend on the activation of specific transcription factors that mediate stress responses after exposure to toxic and genotoxic compounds (e.g. [29]). Thus, studying the activation of transcription factors in certain organs may reflect the actual exposure of that particular organ to genotoxic compounds. Therefore, gene expression profiles seem to represent better biomarkers than the mere assessment of exogenous levels of exposure (for instance, number of cigarettes per day). Accordingly, we showed that several transcription factors can indeed be identified as regulators in the stress response after exposure to cigarette smoke and the toxins therein.

In our study, cigarette smokers were used as a model of exposure to environmental genotoxins, since cigarette smoke is the source of a variety of toxic and genotoxic compounds [30]. Although the direct effects of this type of exposure on sperm quality have been intensively studied [26, 31], the evidence is still not very strong for transgenerational effects. There is some evidence for a higher risk of childhood leukaemia after paternal smoking in the preconceptional period [32], but these data are not consistent in the scientific literature [33]. Boffetta *et al.* [8] performed a meta-analysis in which epidemiological studies were reviewed of paternal smoking and the possible development of specific types of childhood cancer; a positive correlation was performed between paternal smoking in the preconceptional period and elevated risks of several childhood cancers [7]. *In vitro* studies do indeed indicate that cigarette smoke condensate can induce mutations that are typically analysed in sperm (tandem repeats) [34]. Moreover, it has been shown that cigarette smoke derived compounds can reach the testis and semen, and are able to bind to DNA in these cells [22]. As it is illustrated in Figure 3, spermatids and mature spermatozoa in the testis are thought to be exposed to exogenous genotoxins in three different ways during spermatogenesis [35]. First, the exogenous substances can pass the blood-testis barrier, which is a physical barrier between the

blood vessels and the seminiferous tubules in the testis formed by tight junctions between the Sertoli cells. Secondly, after transport via the Sertoli cells, contaminants may be able to induce damage in spermatids and mature spermatogonia. Finally, genotoxic compounds could reach the lumen via the Sertoli cells and reach the mature spermatozoa, but in this case it is not yet known to what level these agents are able to induce damage to the DNA, because it is tightly packed in the sperm nucleus. For instance, Gallagher et al. [36] isolated DNA from sperm cells of smoking and non-smoking subjects, but were not able to detect smoking-related DNA adducts. As the sperm still have to pass through the epididymis after being released into the testicular lumen of the seminiferous tubules, genotoxins could have an effect on the sperm during the epididymal maturation. However, as mature sperm is transcriptionally silenced, this route of exposure is expected to have no effect on mRNA profiles.

During spermiogenesis, protamines replace histones in order to pack the DNA more compactly [16], making the male genome small enough to fit into the spermatid nucleus (the nucleus of a mature spermatozoon is 6 times smaller than that of somatic cells) [37]. This high level of 'packaging' of DNA in spermatozoa also affects the transcription in developing germ cells. Spermatozoa are assumed to be transcriptionally silenced for a large part of the genome during maturation, with some exceptions (such as transcription in mitochondria and transcription in parts of the genome that have not been packaged by protamines but by histones). It was therefore postulated that the RNA present in the ejaculated spermatozoa is most likely residual and non-functional [38], and may be a reflection of gene expression during earlier stages of spermatogenesis [39]. Another observation supporting this idea is the absence of the ribosomal subunits 18S and 28S in the spermatozoal RNA (also in this study, see Figure 1) [19]. Thus, mRNA profiles in sperm provide a retrospective view into processes that have taken place during earlier stages of spermatogenesis, and could also reflect the processes that occur after exposure to toxins and genotoxins in the testis. Moreover, there is also evidence that spermatogenesis-related gene expression profiles can also be used in clinical assessments of sperm quality [40].

Several genes were differentially expressed in smokers' spermatozoa than in non-smokers, including *SALF*, a germ cell specific transcription factor, which also showed the highest correlation with cotinine levels in seminal plasma (which are considered to be good indicators of exposure to cigarette smoke). *SALF* is especially expressed in pachytene spermatocytes and haploid spermatids [41]. Pathway analysis

of all the genes that were differentially regulated in smokers as compared to non-smokers did not reveal any enriched pathway. Part of this inability to detect altered pathways may be related to the degradation of mRNA in spermatozoa (see Figure 1), and as a result signals of genes with relatively low levels of expression may be lost, making pathway analysis incomplete. However, the remaining detectable genes can still provide information regarding the pathways that were activated / inhibited by analyzing the transcription factors that have recognition sequences in the upstream sequences of genes (up to 1kb). The transcription factors that were identified to play a role in the transcription of differentially regulated genes by exposure to cigarette smoke were NF- κ B, forkhead transcription factors (FOX) and YY1. These transcription factors are known regulators in stress responses. For instance, NF- κ B is known to be a key regulator in the inflammatory response, but has also been linked to the control of apoptosis, cell cycle, differentiation and cell migration [29]. NF- κ B can be activated by various stressors, including bacteria, viruses, oxidative stress and certain chemicals. Forkhead proteins are a family of transcription factors with more than 100 members of functionally diverse transcription factors that have commonly been associated with the regulation of foetal development. FOX genes were found to be involved in the regulation of NF- κ B activity [42] (via transcription of its inhibitor IKK) and also DNA damage inducible genes (e.g. *GADD45*). Finally, YY1 has fundamental roles in differentiation, replication, and cellular proliferation [43]. Oei and Shi [44] noted a physical interaction between YY1 and poly(ADP-ribose) polymerase (PARP is a nuclear enzyme involved in DNA repair and transcription). Moreover, YY1 seems to be a cofactor for *MDM2* in the regulation of p53 homeostasis. The transcription factor network analysis showed that apoptosis appeared to be the most significantly influenced process (downregulation in smokers), which means that damaged cells in smoking subjects may not be stopped during the cell cycle. This would predict that ejaculates of smokers would contain more spermatozoa with damaged DNA. Indeed, sperm of smokers showed more genetic abnormalities, like aneuploidy [25]. These findings suggest that cigarette smoke components actually reach the testis and the developing germ cells. However, it is not yet clear to what extent environmental genotoxins can reach the testis to induce damage in spermatozoa.

mRNA in spermatozoa is delivered to the oocyte together with the DNA during fertilisation [45], and some studies show that this RNA may have a significant impact on the developing foetus [46]. The role of this mRNA is still not clear, but it has been suggested that it plays an important role during early

embryogenesis [47]. Nonetheless, it is not yet possible to determine the effect on the health of newborns up to adult ages.

Two well-known genes that are up-regulated after exposure to Ah-receptor agonists, like polycyclic aromatic hydrocarbons and dioxins, are *CYP1A1* and *CYP1B1*, and therefore the expression of these genes can be used as biomarker for exposure to these compounds [48]. PAH are important genotoxins in cigarette smoke, but we did not observe altered expression of *CYP1A1* or *1B1* in spermatozoal mRNA. It is of course possible that parts of the compounds that reach the testis are metabolised by the Sertoli cells and that the germ cells do not have to metabolise the compounds themselves. Nonetheless, the mRNA of both genes was detectable in all subjects by microarray analysis as well as Q-PCR, indicating that germ cells may have the ability to produce these cytochrome P450's to subsequently metabolise environmental genotoxins. The spermatozoal specific enzyme *protamine 2* (*PRM2*) was included as reference gene in these Q-PCR experiments on spermatozoal mRNA, since it was not differentially regulated after exposure to cigarette smoke in the microarray results. As it is known that *PRM2* expression can be abnormal in infertile men [16], *PRM2* should be used as reference gene with care. Still, our data (microarray and real time quantitative PCR) show that there is no differential gene expression of all three protamines (*PRM1*, *PRM2* and *PRM3*) as a result of smoking.

The main goal of this study was to examine the feasibility of using gene expression profiles in mRNA from spermatozoa as biomarkers of exposure. The germ cell specific transcription factor *SALF* appeared to be such a biomarker in subjects exposed to cigarette smoke, where the exposure resulted in different gene-expression profiles. It might be possible that this shift in gene expression can result in decreased sperm quality. Further studies are needed to further validate this biomarker, also for other important environmental and dietary contaminants. It remains to be elucidated whether these changes in spermatozoal mRNA can induce or reflect (negative) health effects in the next generation.

Acknowledgements

The authors would like to thank Ralph Gottschalk and Lou Maas for technical assistance. This work was supported by the Cefic Long-Range Research Initiative Innovative Science award 2004 to R.W.L.G.; as well as by the EU Integrated Project NewGeneris, 6th Framework Programme, Priority 5: Food Quality and Safety (Contract no. FOOD-CT-2005-016320). NewGeneris is the acronym of the project 'Newborns and Genotoxic exposure risks' <http://www.newgeneris.org>.

Conflict of interest statement: None declared.

References

1. Steliarova-Foucher, E., et al., *Geographical patterns and time trends of cancer incidence and survival among children and adolescents in Europe since the 1970s (the ACCISproject): an epidemiological study*. Lancet, 2004. **364**(9451): p. 2097-105.
2. Shelby, M.D., et al., *Fertility, reproduction, and genetic disease: studies on the mutagenic effects of environmental agents on mammalian germ cells*. Environ Health Perspect, 1993. **100**: p. 283-91.
3. Dubrova, Y.E., et al., *Human minisatellite mutation rate after the Chernobyl accident*. Nature, 1996. **380**(6576): p. 683-6.
4. Yauk, C.L., et al., *Induced minisatellite germline mutations in herring gulls (*Larus argentatus*) living near steel mills*. Mutat Res, 2000. **452**(2): p. 211-8.
5. Perera, F.P., et al., *Molecular epidemiologic research on the effects of environmental pollutants on the fetus*. Environ Health Perspect, 1999. **107 Suppl 3**: p. 451-60.
6. Perera, F.P., et al., *A summary of recent findings on birth outcomes and developmental effects of prenatal ETS, PAH, and pesticide exposures*. Neurotoxicology, 2005. **26**(4): p. 573-87.
7. Ji, B.T., et al., *Paternal cigarette smoking and the risk of childhood cancer among offspring of nonsmoking mothers*. J Natl Cancer Inst, 1997. **89**(3): p. 238-44.
8. Boffetta, P., J. Tredaniel, and A. Greco, *Risk of childhood cancer and adult lung cancer after childhood exposure to passive smoke: A meta-analysis*. Environ Health Perspect, 2000. **108**(1): p. 73-82.
9. Godschalk, R.W. and J.C. Kleinjans, *Characterization of the exposure-disease continuum in neonates of mothers exposed to carcinogens during pregnancy*. Basic Clin Pharmacol Toxicol, 2008. **102**(2): p. 109-17.
10. Gaspari, L., et al., *Polycyclic aromatic hydrocarbon-DNA adducts in human sperm as a marker of DNA damage and infertility*. Mutat Res, 2003. **535**(2): p. 155-60.
11. Bjorge, C., et al., *A comparative study of chemically induced DNA damage in isolated human and rat testicular cells*. Reprod Toxicol, 1996. **10**(6): p. 509-19.
12. Mollerup, S., et al., *Sex differences in risk of lung cancer: Expression of genes in the PAH bioactivation pathway in relation to smoking and bulky DNA adducts*. Int J Cancer, 2006. **119**(4): p. 741-4.
13. van Leeuwen, D.M., et al., *Cigarette smoke-induced differential gene expression in blood cells from monozygotic twin pairs*. Carcinogenesis, 2007. **28**(3): p. 691-7.
14. Pessot, C.A., et al., *Presence of RNA in the sperm nucleus*. Biochem Biophys Res Commun, 1989. **158**(1): p. 272-8.
15. Ostermeier, G.C., et al., *Spermatozoal RNA profiles of normal fertile men*. Lancet, 2002. **360**(9335): p. 772-7.
16. Carrell, D.T., B.R. Emery, and S. Hammoud, *Altered protamine expression and diminished spermatogenesis: what is the link?* Hum Reprod Update, 2007. **13**(3): p. 313-27.
17. Sassone-Corsi, P., *Unique chromatin remodeling and transcriptional regulation in spermatogenesis*. Science, 2002. **296**(5576): p. 2176-8.
18. Van Vunakis, H., H.B. Gijka, and J.J. Langone, *Radioimmunoassay for nicotine and cotinine*. IARC Sci Publ, 1993(109): p. 293-9.
19. Goodrich, R., G. Johnson, and S.A. Krawetz, *The preparation of human spermatozoal RNA for clinical analysis*. Arch Androl, 2007. **53**(3): p. 161-7.
20. Staal, Y.C., et al., *Binary PAH mixtures cause additive or antagonistic effects on gene expression but synergistic effects on DNA adduct formation*. Carcinogenesis, 2007. **28**(12): p. 2632-40.

21. Boorsma, A., et al., *T-profiler: scoring the activity of predefined groups of genes using gene expression data*. Nucleic Acids Res, 2005. **33**(Web Server issue): p. W592-5.
22. Horak, S., J. Polanska, and P. Widlak, *Bulky DNA adducts in human sperm: relationship with fertility, semen quality, smoking, and environmental factors*. Mutat Res, 2003. **537**(1): p. 53-65.
23. Schmid, T.E., et al., *The effects of male age on sperm DNA damage in healthy non-smokers*. Hum Reprod, 2007. **22**(1): p. 180-7.
24. Aitken, R.J., et al., *Influence of caffeine on movement characteristics, fertilizing capacity and ability to penetrate cervical mucus of human spermatozoa*. J Reprod Fertil, 1983. **67**(1): p. 19-27.
25. Robbins, W.A., et al., *Effect of lifestyle exposures on sperm aneuploidy*. Cytogenet Genome Res, 2005. **111**(3-4): p. 371-7.
26. Marinelli, D., et al., *Mini-review of studies on the effect of smoking and drinking habits on semen parameters*. Int J Hyg Environ Health, 2004. **207**(3): p. 185-92.
27. Miller, D., G.C. Ostermeier, and S.A. Krawetz, *The controversy, potential and roles of spermatozoal RNA*. Trends Mol Med, 2005. **11**(4): p. 156-63.
28. Al-Shahrour, F., et al., *BABELOMICS: a systems biology perspective in the functional annotation of genome-scale experiments*. Nucleic Acids Res, 2006. **34**(Web Server issue): p. W472-6.
29. Li, X. and G.R. Stark, *NFkappaB-dependent signaling pathways*. Exp Hematol, 2002. **30**(4): p. 285-96.
30. Godschalk, R.W., et al., *Differences in aromatic-DNA adduct levels between alveolar macrophages and subpopulations of white blood cells from smokers*. Carcinogenesis, 1998. **19**(5): p. 819-25.
31. Gandini, L., et al., *The in-vitro effects of nicotine and cotinine on sperm motility*. Hum Reprod, 1997. **12**(4): p. 727-33.
32. Chang, J.S., et al., *Parental smoking and the risk of childhood leukemia*. Am J Epidemiol, 2006. **163**(12): p. 1091-100.
33. Pang, D., R. McNally, and J.M. Birch, *Parental smoking and childhood cancer: results from the United Kingdom Childhood Cancer Study*. Br J Cancer, 2003. **88**(3): p. 373-81.
34. Yauk, C.L., et al., *Mainstream tobacco smoke causes paternal germ-line DNA mutation*. Cancer Res, 2007. **67**(11): p. 5103-6.
35. Holstein, A.F., W. Schulze, and M. Davidoff, *Understanding spermatogenesis is a prerequisite for treatment*. Reprod Biol Endocrinol, 2003. **1**: p. 107.
36. Gallagher, J.E., et al., *³²P-postlabeling analysis of DNA adducts in human sperm cells from smokers and nonsmokers*. Cancer Epidemiol Biomarkers Prev, 1993. **2**(6): p. 581-5.
37. Kramer, J.A. and S.A. Krawetz, *RNA in spermatozoa: implications for the alternative haploid genome*. Mol Hum Reprod, 1997. **3**(6): p. 473-8.
38. Miller, D., et al., *A complex population of RNAs exists in human ejaculate spermatozoa: implications for understanding molecular aspects of spermiogenesis*. Gene, 1999. **237**(2): p. 385-92.
39. Ostermeier, G.C., et al., *Toward using stable spermatozoal RNAs for prognostic assessment of male factor fertility*. Fertil Steril, 2005. **83**(6): p. 1687-94.
40. Wang, H., et al., *A spermatogenesis-related gene expression profile in human spermatozoa and its potential clinical applications*. J Mol Med, 2004. **82**(5): p. 317-24.
41. Han, S.Y., et al., *TFIIAalpha/beta-like factor is encoded by a germ cell-specific gene whose expression is up-regulated with other general transcription factors during spermatogenesis in the mouse*. Biol Reprod, 2001. **64**(2): p. 507-17.
42. Grant, C., et al., *Foxp3 represses retroviral transcription by targeting both NF-kappaB and CREB pathways*. PLoS Pathog, 2006. **2**(4): p. e33.

43. Gordon, S., et al., *Transcription factor YY1: structure, function, and therapeutic implications in cancer biology*. Oncogene, 2006. **25**(8): p. 1125-42.
44. Oei, S.L. and Y. Shi, *Transcription factor Yin Yang 1 stimulates poly(ADP-ribosyl)ation and DNA repair*. Biochem Biophys Res Commun, 2001. **284**(2): p. 450-4.
45. Ostermeier, G.C., et al., *Reproductive biology: delivering spermatozoan RNA to the oocyte*. Nature, 2004. **429**(6988): p. 154.
46. Rassoulzadegan, M., et al., *RNA-mediated non-mendelian inheritance of an epigenetic change in the mouse*. Nature, 2006. **441**(7092): p. 469-74.
47. Krawetz, S.A., *Paternal contribution: new insights and future challenges*. Nat Rev Genet, 2005. **6**(8): p. 633-42.
48. Lemm, F., M. Wilhelm, and P.H. Roos, *Occupational exposure to polycyclic aromatic hydrocarbons suppresses constitutive expression of CYP1B1 on the transcript level in human leukocytes*. Int J Hyg Environ Health, 2004. **207**(4): p. 325-35.

Table 1: Characteristics of study population		
	Smokers	Non-smokers
n	4	4
Age (yrs)	38.8 ± 4.9*	35.8 ± 3.9*
Cotinine in seminal plasma (ng/ml)	276.8 ± 107.2*	8.3 ± 0.5**
Sperm concentration (×10 ⁶ cells/ml)	19.4 ± 1.3*	42.6 ± 14.9*
pH	8	8
Motility [#]	40/32/5/23	59/13/6/22
Morphology (% normal)	72.5 ± 7.5	81.3 ± 2.4
Alcohol intake (consumptions)	1 – 10 drinks per week	1 – 10 drinks per week
Caffeine intake (consumptions)	1 – 5 cups per day	1 – 5 cups per day

*Average values ± SE of the mean are shown.

**Difference in cotinine levels in seminal plasma proved to be significant ($p < 0.05$) using Student's t-Test.

[#]Motility Format: 3/2/1/0 in average percentages, with 3 = fast-progressive; 2 = slow-progressive; 1 = non-progressive, but tail movement; 0 = immobile, no tail movement.

Table 2: Overview of significantly regulated genes that also showed a high correlation ($R > 0.90$) with log-transformed cotinine levels in seminal plasma of smokers and non-smokers.

Gene name	UniGeneID	Proposed biological function	Fold change	R with cotinine
<i>SALF</i>	Hs.44385	Germ line specific transcription factor	5.474	0.986*
<i>UBE4B</i>	Hs.386404	Ubiquitination factor E4B	-1.993	-0.976
<i>TRIM26</i>	Hs.485041	Zinc finger protein	-7.356	-0.950
<i>FADS2</i>	Hs.502745	Fatty acid desaturase	-1.959	-0.944
<i>ARMC1</i>	Hs.269542	Signal transduction, development, cell adhesion and mobility, tumor initiation and metastasis	2.118	0.941
<i>ACTL6A</i>	Hs.435326	Actin-like 6A	3.209	0.940
<i>NBPF15</i>	Hs.512037	Neuroblastoma breakpoint family	1.848	0.937
	Hs.143408	Unknown	2.220	0.935
<i>NBPF14</i>	Hs.515947	Neuroblastoma breakpoint family	1.771	0.934
<i>SMYD4</i>	Hs.514602	SET and MYND domain; Transcriptional regulation	-4.265	-0.929
<i>LRRC51</i>	Hs.317243	Leucine rich repeat containing 51	-2.101	-0.925
<i>FLJ40852</i>	Hs.17589	Hypothetical protein	1.734	0.915
<i>C3orf48</i>	Hs.585048	Chromosome 3 open reading frame 48	1.853	-0.915
<i>SH2D5</i>	Hs.166270	Intracellular signaling	-2.468	-0.908
	Hs.530461	Unknown	-4.581	0.908
<i>LOC644246</i>	Hs.463231	Hypothetical protein	1.923	-0.906
<i>SRPK1</i>	Hs.443861	Phosphorylation of SR proteins	1.737	0.906
<i>RNF125</i>	Hs.458449	Negative regulator of IFN production	-4.898	-0.905
	Hs.125434	Unknown	1.881	0.900
<i>GALK2</i>	Hs.122006	Galactokinase	-2.423	-0.900

* Remains statistically significant when fold changes are correlated with cotinine levels in smokers only ($p < 0.05$; $R > 0.90$)

Table 3: Real time PCR validation data; fold change as compared to non-smokers (NS as reference =1)		
Gene name	Fold change Q-PCR	Fold change array
<i>PRM-2</i>	1.01 ± 0.37	0.88
<i>Cyp1A1</i>	1.26 ± 0.31	1.08
<i>Cyp1B1</i>	1.77 ± 1.04	1.39
<i>SALF</i>	2.29 ± 2.61	5.48

Values are corrected using the geometrical average of *β-actin* and *protamine 2* as reference genes.

Figure 1: BioAnalyzer results showing the presence (lane 1) and absence (lane 2) of 18S and 28S ribosomal subunits (A), and the chromatograms of a sperm sample (B) with fragmented mRNA and a lung RNA sample containing the 18S and 28S ribosomal subunits (C)

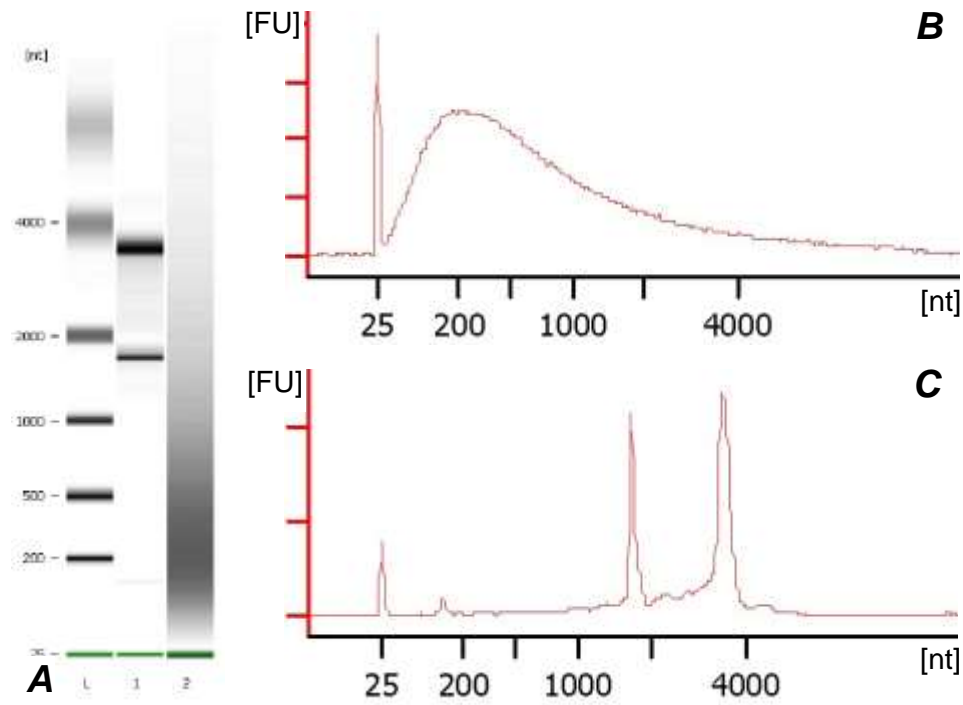


Figure 2: A significantly higher percentage of genes is regulated by NF- κ B, FOX or YY1 in the set of differentially regulated genes by cigarette smoking as compared to the remaining genes on the array. The percentage of genes, potentially regulated by one of these transcription factors increases when the correlation cut-off value with the cotinine levels is set stricter.

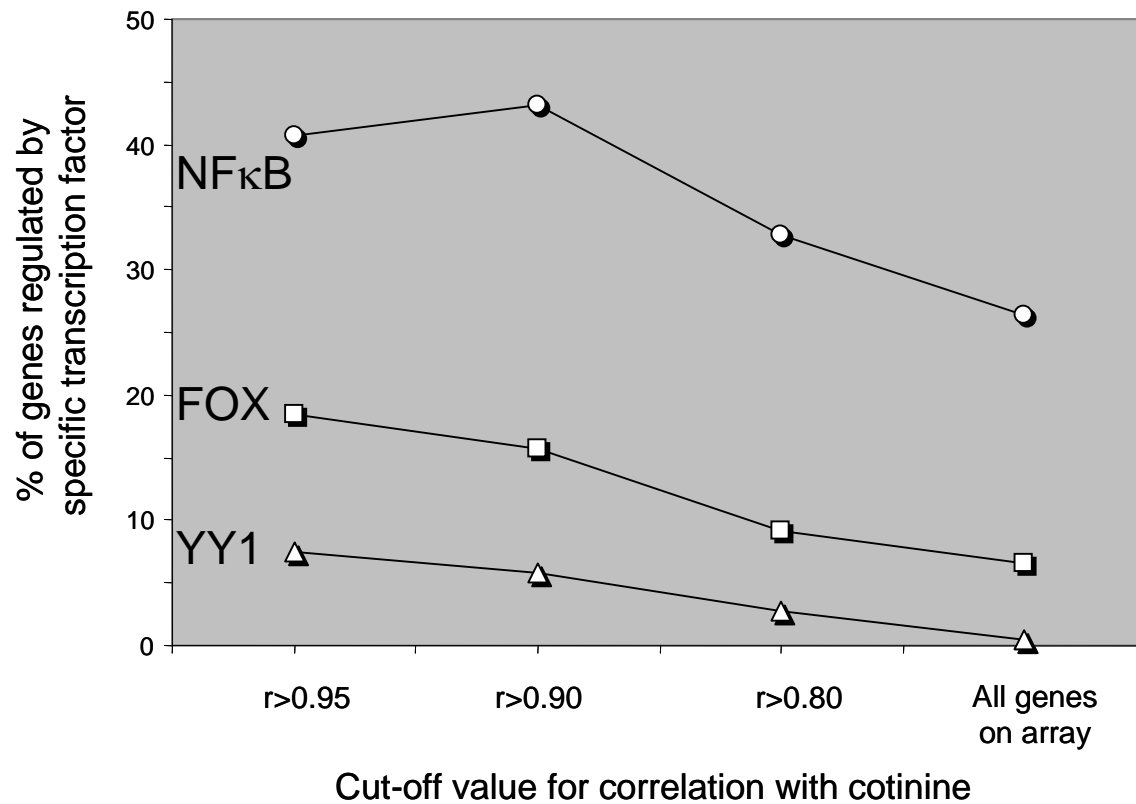


Figure 3: Three different routes of exposure of germ cells to exogenous substances that could ultimately influence mRNA profiles in sperm. Arrows indicate the exposure route to exogenous compounds, 1: directly from the blood into the basal compartment, 2: via the Sertoli cell into the adluminal compartment, or 3: through the Sertoli cell directly into the lumen of the seminiferous tubule.

