



Analysis of mRNA Expression Levels in FFPE Testicular Biopsies Using Real-time PCR Arrays

Predrag Noveski¹, Katerina Popovska-Jankovic¹, Katerina Kubelka², Vanja Filipovski², Slobodan Lazarevski², Toso Plaseski³ and Dijana Plaseska-Karanfilska^{1*}

¹Research Center for Genetic Engineering and Biotechnology “Georgi D. Efremov”, Macedonian Academy of Sciences and Arts, Skopje, Republic of Macedonia; ²Clinical Hospital “Acibadem Sistina”, Skopje, Republic of Macedonia; ³Clinic of Endocrinology, Diabetes and Metabolic Disorders, Faculty of Medicine, University “Ss Cyril and Methodius”, Skopje, Republic of Macedonia

Abstract

Background and objective: Tissues archived for long-term storage as formalin-fixed, paraffin-embedded (FFPE) samples represent rich source of material for genomic studies, but the nucleic acid isolation and downstream analysis is technically difficult. This is especially true when conducting mRNA expression studies, which strongly depend on the quality and quantity of the starting material for successful data normalization. Our objective was to investigate the mRNA expression levels in testicular FFPE samples using real-time PCR arrays and to present our experience with some technical challenges. **Methods:** Total RNA was extracted from FFPE samples from six patients with hypospermatogenesis and three controls using the Qiagen AllPrep DNA/RNA FFPE Kit. The integrity of the isolated RNA was assessed by an Agilent Bioanalyzer 2100. Qiagen Human Male Infertility RT² Profiler PCR Arrays were used to study the expression of 84 genes. **Results:** Our experience with the analysis of the FFPE testicular samples using the RT² Profiler PCR Array showed difficulties with the data normalization, despite using the same amount of starting material and similar values for the RNA integrity numbers (RINs) across all the samples, as recommended by the manufacturer. By using the percentage of the relatively intact RNA (area between 150 bp and 4,000 bp on the electropherograms of the Bioanalyzer 2100), we observed a negative correlation between amount of the intact RNA and average Ct values of the analyzed genes. Our initial results using PCR Array analysis on FFPE testicular tissues revealed 38 differentially expressed genes that

are enriched in interactions, as well as for Gene Ontology molecular and cellular function terms. **Conclusions:** Stratification of the FFPE samples according to their percentage of intact RNA could improve the quantitative real-time PCR Array analysis.

Introduction

Infertility is a worldwide problem affecting 5–7% of the general male population, and in about 50% of the reported cases the main cause for male infertility is idiopathic.¹ In the majority of cases of idiopathic male infertility the causes might be of a genetic nature because of the vast number of genes and gene interactions involved in spermatogenesis and male fertility. Testicular biopsies archived as formalin-fixed, paraffin-embedded (FFPE) samples represent a valuable resource for the analysis of the influence of DNA sequence changes and RNA expression on spermatogenesis. However, the isolation and downstream analysis of nucleic acids from these samples are technically difficult due to chemical modifications and degradation associated with the fixation process and storage.² In particular, the presence of formaldehyde induces cross-linking between nucleic acids and proteins, and the low pH (<1) induces fragmentation.^{3,4} Nevertheless, FFPE-archived tissues are more easily accessible than fresh or frozen tissues, thus much effort has been made to use FFPE tissues for different genetic analysis.

One approach in the search for genetic causes of male infertility is to analyze testicular gene expression profiles of patients with idiopathic male infertility and perform a comparative analysis to the profiles of subjects with normal spermatogenesis. Several different methodologies are available for this, including northern blotting, reverse transcription-quantitative polymerase chain reaction (RT-qPCR), microarray analysis and RNA sequencing. Although RT-qPCR is considered as one of the gold-standard methods for detection and measurement of mRNA transcripts, it has a disadvantage since only a limited number of genes can be analyzed simultaneously. Recently, PCR Arrays have been introduced that combine the real-time PCR performance with the ability of arrays to detect the expression of many genes simultaneously. Such PCR Array methodology that employs SYBR[®] Green chemistry was successfully validated against *TaqMan* PCR, microarrays and other gene expression measurement technologies.^{5,6} However, these validation studies were performed on two high quality RNA reference samples, the first one being from normal human brain tissue and the second one being a pool of 10 human cell lines. Studies using the PCR Arrays on FFPE tissues are scarce.

Keywords: FFPE; RT² Profiler PCR Array; Male infertility; Testis; Spermatogenesis; Hypospermatogenesis.

Abbreviations: FFPE, formalin-fixed, paraffin-embedded samples; mRNA, messenger ribonucleic acid; PCR, polymerase chain reaction; DNA, deoxyribonucleic acid; RNA, ribonucleic acid; RIN, RNA integrity number; Ct, cycle threshold; pH, molar concentration of hydrogen ions in the solution; RT-qPCR, reverse transcription-quantitative polymerase chain reaction; SYBR Green, cyanine dye, [2-[N-(3-dimethylaminopropyl)-N-propylamino]-4-[2,3-dihydro-3-methyl-(benzo-1,3-thiazol-2-yl)-methylidene]-1-phenyl-quinolinium]; *TaqMan*, fluorogenic 5' nuclease chemistry; miRNA, microRNA, small non-coding RNA molecule; cDNA, complementary DNA; qPCR, quantitative polymerase chain reaction; bp, nucleotide base pairs; SPSS, Statistical Package for Social Sciences; GO, Gene Ontology; STRING, Search Tool for the Retrieval of Interacting Genes/Proteins; WebGestalt, WEB-based Gene Set Analysis Toolkit; SOD2, superoxide dismutase 2, mitochondrial.

Received: 9 July 2016; Revised: 10 September 2016; Accepted: 14 September 2016

*DOI: 10.14218/ERHM.2016.00006

*Correspondence to: Dijana Plaseska-Karanfilska, Research Centre for Genetic Engineering and Biotechnology “Georgi D. Efremov”, Macedonian Academy of Sciences and Arts, Krste Misirkov 2, 1000 Skopje, Republic of Macedonia. Tel: +389-2-3235-410, E-mail: dijana@manu.edu.mk

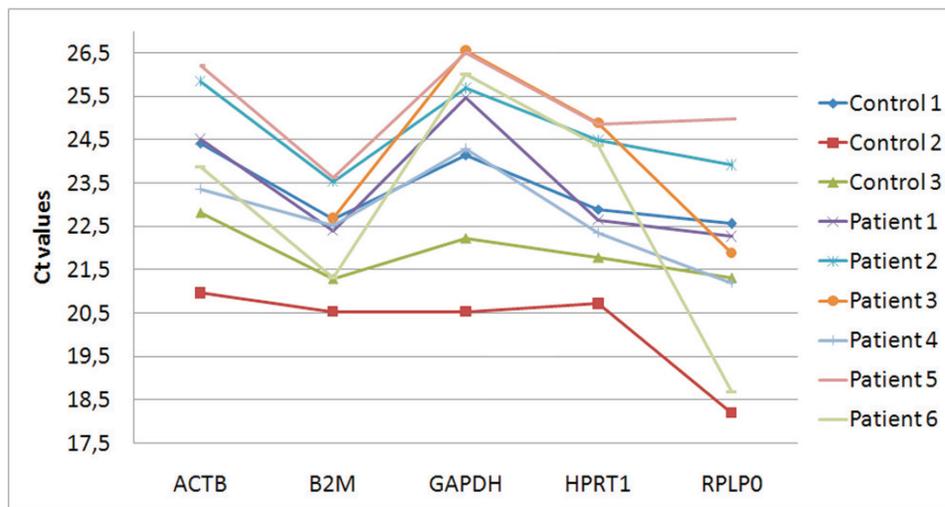


Fig. 1. Graphical presentation of the Ct values of five housekeeping genes included in the RT² Profiler PCR Array plates testing the nine analyzed samples [three patients with normal spermatogenesis (controls) and six patients with hypospermatogenesis].

Here, we are reporting our experience with the analysis of gene expression of 84 genes required for normal male fertility in testicular biopsies archived as FFPE tissues from male infertile patients. Our goal was to address some possible technical challenges when working with low-quality RNA during gene expression studies and to present initial results of differentially-expressed genes in testicular tissues of male infertile patients with hypospermatogenesis.

Materials and methods

Samples

FFPE-archived tissues from three men diagnosed with normal spermatogenesis (patients with obstructive azoospermia) and six men diagnosed with hypospermatogenesis (infertile patients) based on histopathological findings were used in the study. While an ideal study design would have been to use biopsies from healthy fertile patients with normal testicular function as controls, in practice this is very hard to achieve; thus, as shown in other studies, testicular biopsies from men with obstructive azoospermia and histopathological finding of normal spermatogenesis were used as an acceptable alternative.⁷ This study was approved by the Ethical Committee of the Macedonian Academy of Sciences and Arts. All subjects gave written informed consent for participation, in accordance with the Declaration of Helsinki.

RNA isolation, quality control and performance of real-time qPCR array analysis

The total RNA was extracted using the AllPrep DNA/RNA FFPE Kit from Qiagen (Hilden, Germany), with the simultaneous isolation of DNA and RNA being performed after deparaffinization of the FFPE samples using xylene. We have previously compared this kit with the RNA isolation method using the Trizol reagent and with two other commercial kits for RNA isolation only; we had found that the AllPrep DNA/RNA FFPE Kit was best suited for our needs because the obtained RNA samples were of consistent quality and that the RNA samples contained small RNAs of sufficient

quality and quantity that enabled us to perform miRNA microarray expression study.⁸ RNA quantity and purity were determined using the NanoDrop ND-2000 (NanoDrop Technologies, Wilmington, DE, USA), and RNA integrity (as RNA integrity number, RIN) was assessed using the RNA 6000 Nano Chip on a Bioanalyzer 2100 (Agilent Technologies, Santa Clara, CA, USA). The isolated RNA was analyzed using the Human Male Infertility RT² Profiler PCR Array (Cat# PAHS-165ZC; Qiagen), which profiles the expression of 84 key gene transcripts detected in the male germline that are known to be involved in different fertility- and sperm-related processes in the cell, including spermatogenesis, fertilization, male sex differentiation, cell motility, cell cycle and response to stress. The cDNA synthesis, pre-amplification and real-time PCR analyses were conducted according to the instructions given by the manufacturer. A starting amount of 640 ng of total RNA (maximum amount available from the sample with the lowest RNA concentration) was used for all samples for the reverse transcription and qPCR. The real-time qPCR was performed using a 7500 Fast Real-Time PCR System (Applied Biosystems, Foster City, CA, USA).

Data analysis

Following the real-time qPCR, the data from all samples were analyzed using the same Ct threshold. The Ct threshold was set to an arbitrary value of 0.067761 after averaging all plates and performing manual inspection for best-fit, while the baseline was set to 3 cycles below the lowest Ct. Afterwards, the Ct values for each sample were exported into separate Microsoft Excel files for web-based RT² Profiler PCR Array Data Analysis 3.5 (<http://pcrdata-analysis.sabiosciences.com/pcr/arrayanalysis.php>). Quantification of the relative changes of gene expression levels (fold-change) was performed using the $2^{-\Delta\Delta Ct}$ method. The *p*-values were calculated using a Student's *t*-test (two-tail distribution and equal variances between the two samples) on the $2^{-\Delta Ct}$ values. The percentage of intact RNA, which was arbitrarily set between 150–4,000 base pairs (bp), was determined by 2100 Expert Software (Agilent Technologies). Pearson's correlation analysis was performed using the Statistical Package for Social Sciences, Version 19 (SPSS, Chicago, IL, USA). The level of the two-tailed test of significance was set at $p < 0.05$. Analysis for possible interactions between genes

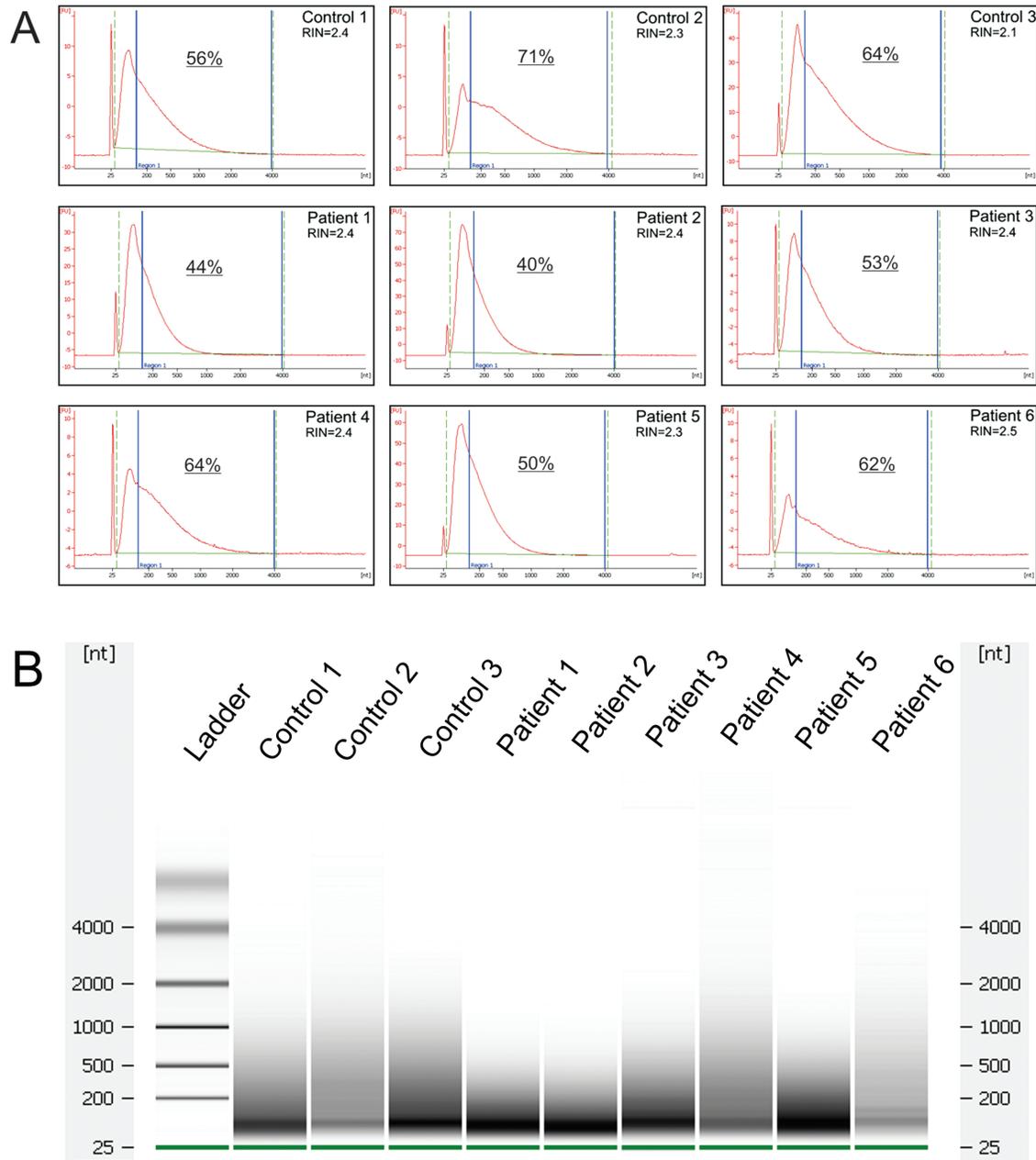


Fig. 2. Illustration of electropherograms and gel electrophoresis from 2100 Expert Software for all nine samples. A) The blue lines define the region in the electropherogram between 150 bp and 4,000 bp. The percentage of the region between the lines is also shown. B) Gel electrophoresis of the nine analyzed samples.

and the building of network interactions was carried out using STRING v10, while Gene Ontology (GO) analysis was performed using WebGestalt.⁹⁻¹¹

Results

Normalization of gene expression between plates

The analysis of relative expression in the RT² profiler PCR Arrays was based on defining the same Ct threshold values across all runs

and then performing normalization with stably expressed genes in all samples. Although the RIN numbers (which ranged between 2.1 and 2.5) and the total RNA input amount was similar in all samples, there were large differences in the expression levels of the five housekeeping genes in the nine studied samples (Fig. 1). There was no correlation between average Ct values of the five housekeeping genes and the RIN numbers in the nine analyzed samples.

In order to understand the discrepancies in the Ct values of the housekeeping genes between the samples, we determined the percentage of the relatively intact RNA (set between 150 bp and 4,000 bp) in each sample by analyzing the electropherograms (Fig. 2).

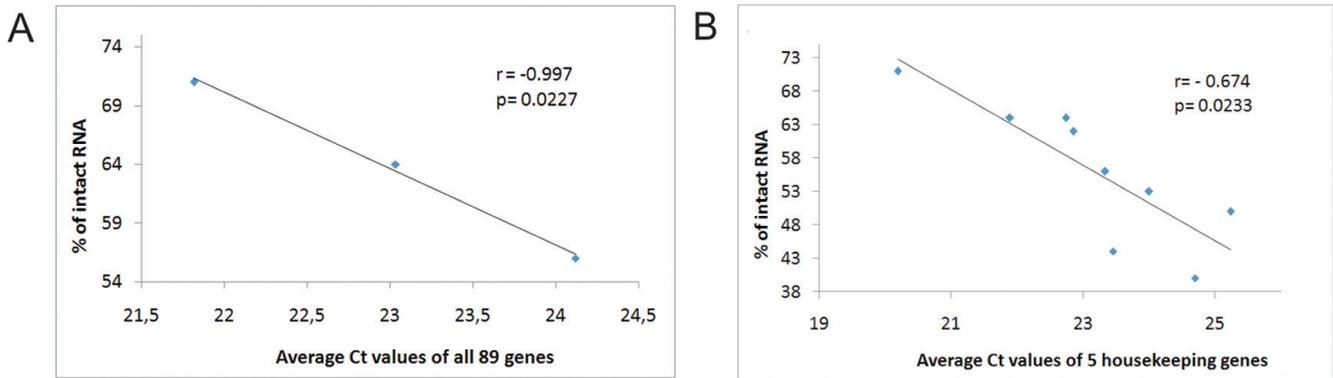


Fig. 3. Illustration of the correlation of average Ct values and percentage of intact RNA. A) Correlation between average Ct values of all 89 genes analyzed and percentage of intact RNA. Each dot in the scatterplot represents one control sample. B) Correlation between the average Ct values of five housekeeping genes analyzed in all nine samples and percentage of intact RNA. Each dot in the scatterplot represents one subject (patients with normal spermatogenesis (controls) or patients with hypospermatogenesis).

The percentages of the selected area varied between 40% and 71% in the different samples.

Control samples, from patients with normal spermatogenesis, are expected to have similar expression levels of all analyzed genes; hence, we averaged the Ct values of all 89 genes in the three controls and compared them with the area percentage of the intact RNA. There was a strong significant negative correlation between the averaged Ct values and the amount of intact RNA in the controls (Fig. 3A). The same comparison was made for the five housekeeping genes in all nine samples, since they are expected to be expressed at similar levels; the results showed a moderate significant negative correlation (Fig. 3B).

When analyzing the average Ct values of the five housekeeping genes in the group of patients with hypospermatogenesis, the correlation was found to be similarly moderate but without significance ($r=-0.686$, $p=0.1324$). When analyzing the average Ct

values of the 84 targeted genes in the group of patients with hypospermatogenesis, the correlation found to be weak and without significance ($r=-0.246$, $p=0.6383$). This weak correlation might be due to the different severity of hypospermatogenesis in the patients that may have influenced the expression levels of the 84 analyzed genes.

Analysis of differentially expressed genes

According to the manufacturer’s protocol, normalization of the expression data of the analyzed samples can be performed using one of the housekeeping genes or any other of the 84 genes, provided that the Ct values of the gene used for normalization does not differ more than 1.5 cycles in all samples (plates). In our study, none of the housekeeping genes satisfied this criterion. However,

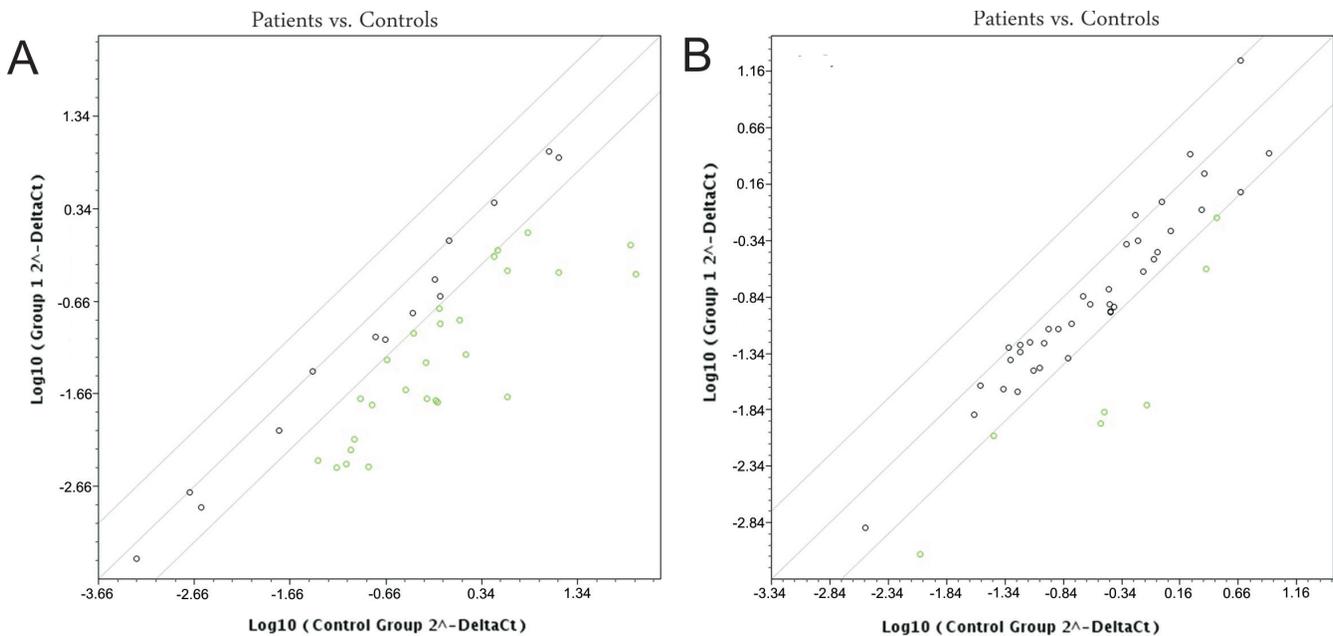


Fig. 4. Scatterplot comparing the normalized expression of genes between the patients with normal spermatogenesis (controls) and patients with hypospermatogenesis. A) Genes involved in spermatogenesis and fertilization are shown. B) Genes involved in male sex differentiation, cell motility, cell cycle and response to stress are shown. The central line indicates unchanged gene expression. The boundary (fold regulation cut-off) was set to 4. Green dots represents genes with fold regulation <4.

in six out of nine samples (plates) normalization was possible with the *SOD2* gene (superoxide dismutase 2, mitochondrial), which plays a key role in the protection against oxidative stress in the mitochondria.¹² The best strategy for choosing the right genes for data normalization, as previously shown, is to test the larger number of genes and to select those with lowest variability between the studied groups of samples.¹³ The *SOD2* gene is expressed in various tissues, including different testicular cells, so we chose this gene for normalization in our mRNA expression studies of testicular biopsies. This might be a limitation of our study, since we cannot exclude the possibility that the *SOD2* gene is differentially expressed between physiological and pathological conditions. The inclusion of experimentally-validated control genes, instead of the widely used housekeeping genes, might improve these commercially available PCR arrays.

When the fold-change cut-off was set at 4, a number of genes showed lower expression among the patients with hypospermatogenesis than in the patients (controls) with normal spermatogenesis. Furthermore, the genes involved in spermatogenesis and fertilization showed a higher number of down-regulated genes, as compared to genes involved in male sex differentiation, cell motility, cell cycle and response to stress (Fig. 4).

All 84 genes were down-regulated in the patients with hypospermatogenesis, but the difference was significant only for 38 of them. The fold-changes and *p*-values for all of the 84 analyzed genes for the six patients with hypospermatogenesis are given in Supplementary Table S1.

Discussion

The analysis of gene expression using real-time PCR is a standard and widely used technique because of the wide dynamic range of quantification, high sensitivity, accuracy and robustness.^{14,15} The quantification of changes in gene expression could be absolute, using standard curves, or relative, using a calibrator gene whose expression level between samples is constant.^{16,17} In order to investigate the expression of many genes simultaneously, we used a RT² Profiler PCR Array from Qiagen for simultaneous amplification and expression detection of 84 genes detected in the male germline.

One of the recommendations for analysis using PCR Arrays is to use the same input amount of RNA and the same quality of RNA samples, as ascertained through RIN number, as well as the same Ct threshold between plates. Even though we fulfilled these requirements, our analysis of nine FFPE samples still showed high variation of Ct values between plates for the five pre-selected housekeeping genes. We found no correlation between RIN numbers and variation in Ct values.

After a visual inspection of BioAnalyzer electropherograms, we noticed differences between samples in the curve shapes, and for that reason we decided to investigate a possible correlation between the amount of relatively intact RNA (150–4,000 bp) and variation in Ct values. We found a strong negative correlation between the amount of intact RNA and the average Ct values of all 89 genes in the three control samples. Furthermore, a moderate negative correlation was observed when we investigated five housekeeping genes in all samples. Fragments with sizes between 70–150 bp, even though abundant, could not contain one of the primer sequences required for initial PCR amplification. Consequently, the samples with same RNA concentration but different percentage of intact RNA will generate different Ct values in a PCR reaction.

In the majority of studies of the RNA expression levels in FFPE tissue samples, the RIN number has been used as a primary measurement for RNA integrity. This applies not only for real-time PCR-based studies, as in our case, but also for microarray genome-wide expression studies. We propose methodology that takes into account not only the integrity of the RNA as a whole, but also the percentage of the relatively intact RNA between samples. In heavily-degraded samples, differences in this percentage could have a significant impact on the final result.

Conclusion

Data normalization of the relative gene expression levels using a reference gene with stable expression across various conditions is an essential step in real-time quantitative PCR analysis. However, in samples with degraded RNA, especially for RNA from FFPE tissues which is often heavily degraded, erroneous results are possible due to different concentrations of intact RNA molecules in the different samples. Our experience with analysis of FFPE testicular samples using the RT² Profiler PCR Array from Qiagen showed difficulties with data normalization, despite using the same amount of starting material and similar values for the RIN numbers across all the samples, as recommended by the manufacturer. We detected high correlation between the amount of intact RNA and the average Ct values of analyzed genes in the control samples.

Stratification of the FFPE samples according to their percentage of intact RNA could improve the quantitative real-time PCR Array analysis.

Prospective

Simultaneous analysis of RNA expression levels of a large number of genes is a powerful tool for discovering the affected biological pathways and underlying genetic causes of complex conditions, such as male infertility. Targeted analysis of the genes previously found to be primarily expressed in the testicular tissue and related to the spermatogenesis process could be a cost-effective and robust method for analysis of a larger number of samples. However, design of such panels of genes should be accompanied with adequate solutions to overcome common technical difficulties, such as experimental validation of housekeeping genes, and especially addressing and improving the data analysis of samples from which RNA with sufficient quality is difficult to obtain.

Our initial results using PCR Array analysis on FFPE testicular tissues revealed 38 differentially-expressed genes. The network analysis for possible interactions between these 38 selected genes, using the STRING web tool showed that the resulting network is enriched in interactions ($p=1.73E-14$) (Fig. 5).

GO enrichment analysis for molecular function showed a significant enrichment in a subset of genes involved in nucleic acid binding (GO: 0003676), transcription cofactor activity (GO: 0003712), translation activator activity (GO: 0008494) and histone acetyl-lysine binding (GO: 0070577). For the cellular component, the analysis showed a significant enrichment in a subset of genes for the chromosomal portion (GO: 0044427), the microtubule-based flagellum (GO: 0009434), the CatSper complex (GO: 0036128) and a chromatoid body (GO: 0033391).

Further analyses on a larger number of samples are warranted to confirm our initial findings on the affected biological pathways in testicular tissues of patients with hypospermatogenesis.

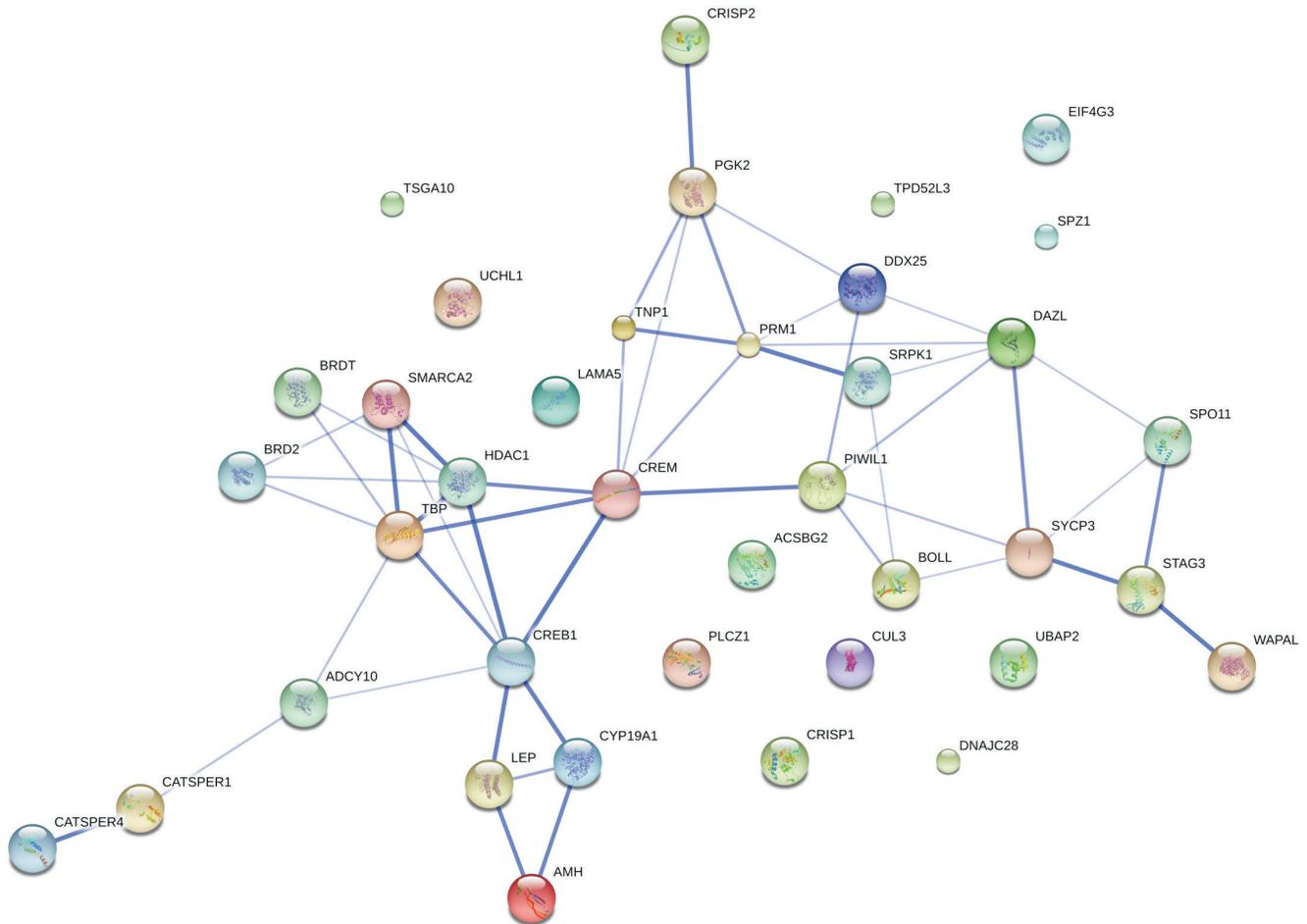


Fig. 5. Illustration of interactions between the 38 differentially-expressed genes in patients with hypospermatogenesis (STRING v.10, confidence view).

Acknowledgements

This study was supported by project CRP/MAC09-01 from ICGEB, Trieste (to D. Plaseska-Karanfilska).

Conflict of interest

The authors have no conflict of interests related to this publication.

Author contributions

Designed the study (DPK), Performed the experiments (PN, KPJ), Analyzed the data (PN, DPK), Contributed reagents/materials/analysis tools (KKS, VF, SL, TP, DPK), Wrote the manuscript (PN, DPK).

Supplementary information

(Supplementary information is linked to the online version of the paper on the *Exploratory Research and Hypothesis in Medicine* website.)

References

- [1] Plaseska-Karanfilska D, Noveski P, Plaseski T, Maleva I, Madjunkova S, Moneva Z. Genetic causes of male infertility. *Balkan J Med Genet* 2012;15(Suppl):31–34. doi:10.2478/v10034-012-0015-x.
- [2] Roberts L, Bowers J, Sensinger K, Lisowski A, Getts R, Anderson MG. Identification of methods for use of formalin-fixed, paraffin-embedded tissue samples in RNA expression profiling. *Genomics* 2009;94:341–348. doi:10.1016/j.ygeno.2009.07.007.
- [3] Srinivasan M, Sedmak D, Jewell S. Effect of fixatives and tissue processing on the content and integrity of nucleic acids. *Am J Pathol* 2002;161:1961–1971. doi:10.1016/s0002-9440(10)64472-0.
- [4] Lin J, Kennedy SH, Svarovsky T, Rogers J, Kemnitz JW, Xu A, *et al*. High-quality genomic DNA extraction from formalin-fixed and paraffin-embedded samples deparaffinized using mineral oil. *Anal Biochem* 2009;395(2):265–267. doi:10.1016/j.ab.2009.08.016.
- [5] Canales RD, Luo Y, Willey JC, Austermler B, Barbacioru CC, Boysen C, *et al*. Evaluation of DNA microarray results with quantitative gene expression platforms. *Nat Biotechnol* 2006;24(9):1115–1122. doi:10.1038/nbt1236.
- [6] Arikawa E, Sun Y, Wang J, Zhou Q, Ning B, Dial SL, *et al*. Cross-platform comparison of SYBR Green real-time PCR with TaqMan PCR, microarrays and other gene expression measurement technologies evaluated in the MicroArray Quality Control (MAQC) study. *BMC Genomics* 2008;9:328. doi:10.1186/1471-2164-9-328.
- [7] Zhuang X, Li Z, Lin H, Gu L, Lin Q, Lu Z, *et al*. Integrated miRNA and mRNA expression profiling to identify mRNA targets of dysregulated miRNAs in non-obstructive azoospermia. *Sci Rep* 2015;5:7922. doi:10.1038/srep07922.
- [8] Noveski P, Popovska-Jankovic K, Kubelka-Sabit K, Filipovski V, Lazarevski S, Plaseski T, *et al*. MicroRNA expression profiles in testicular biopsies of patients with impaired spermatogenesis. *Andrology* 2016;4(6):1020–1027. doi:10.1111/andr.12246.

- [9] Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, *et al.* STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res* 2015;43(D1):D447–D452. doi:10.1093/nar/gku1003.
- [10] Zhang B, Kirov S, Snoddy J. WebGestalt: an integrated system for exploring gene sets in various biological contexts. *Nucleic Acids Res* 2005;33(suppl 2):W741–W748. doi:10.1093/nar/gki475.
- [11] Wang J, Duncan D, Shi Z, Zhang B. WEB-based GEne SeT AnaLysis Toolkit (WebGestalt): update 2013. *Nucleic Acids Res* 2013;41(W1):W77–W83. doi:10.1093/nar/gkt439.
- [12] Luk E, Yang M, Jensen LT, Bourbonnais Y, Culotta VC. Manganese activation of superoxide dismutase 2 in the mitochondria of *Saccharomyces cerevisiae*. *J Biol Chem* 2005;280:22715–22720. doi:10.1074/jbc.M504257200.
- [13] Aerts JL, Gonzales MI, Topalian SL. Selection of appropriate control genes to assess expression of tumor antigens using real-time RT-PCR. *Biotechniques* 2004;36:84–91.
- [14] Vandenbroucke II, Vandesompele J, Paepe AD, Messiaen L. Quantification of splice variants using real-time PCR. *Nucleic Acids Res* 2001;29:e68. doi:10.1093/nar/29.13.e68.
- [15] Bustin SA. Quantification of mRNA using real-time reverse transcription PCR (RT-PCR): trends and problems. *J Mol Endocrinol* 2002;29:23–39. doi:10.1677/jme.0.0290023.
- [16] Bustin SA. Absolute quantification of mRNA using real-time reverse transcription polymerase chain reaction assays. *J Mol Endocrinol* 2000;25:169–193. doi:10.1677/jme.0.0250169.
- [17] Ginzinger DG. Gene quantification using real-time quantitative PCR: an emerging technology hits the mainstream. *Exp Hematol* 2002;30:503–512. doi:10.1016/S0301-472X(02)00806-8.