HAL
open science

Unraveling the sperm transcriptome by next generation sequencing and the global epigenetic landscape in infertile men

Fadi Choucair

To cite this version:

Fadi Choucair. Unraveling the sperm transcriptome by next generation sequencing and the global epigenetic landscape in infertile men. Molecular biology. COMUE Université Côte d'Azur (2015 - 2019); Université Libanaise, 2018. English. NNT : 2018AZUR4058 . tel-01958881

HAL Id: tel-01958881

https://theses.hal.science/tel-01958881v1

Submitted on 18 Dec 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

UNIVERSITÉ CÔTE D'AZUR

ÉCOLE DOCTORALE
SCIENCES DE LA VIE ET DE LA SANTÉ

Université Libanaise
École Doctorale Sciences et Technologies
Doyen

THÈSE DE DOCTORAT

Exploration du transcriptome spermatique par le séquençage nouvelle génération et le portrait épigénétique de l'infertilité masculine

Unraveling the sperm transcriptome by next generation sequencing and the global epigenetic landscape in infertile men

Fadi CHOUCAIR

INSERM U1065, C3M

Présentée en vue de l'obtention du grade de docteur en interactions moléculaires et cellulaires
de l'Université Côte d'Azur
et de l'Université Libanaise
Dirigée par : Valérie Grandjean / Mira Hazzouri
Soutenue le : 6 Septembre 2018

Devant le jury, composé de :

Mme RACHEL LEVY, PR, UMRS 938, UPMC
M. FABIEN MONGELARD, MC, CRCL, ENS Lyon
Mme NINA SAADALLAH-ZEIDAN, PR, Université Libanaise
Mme Anne-Amandine CHASSOT, CR, iBV, Université de Nice
Mme VALÉRIE GRANDJEAN, CR, C3M, Université de Nice
Mme MIRA HAZZOURI, PR, Université Libanaise

Exploration du transcriptome spermatique par le séquençage nouvelle génération et le portrait épigénétique de l'infertilité masculine

Jury

Président du Jury
Mme Anne-Amandine CHASSOT, CR, iBV, Université de Nice

Rapporteurs
Mme RACHEL LEVY, PR, UMRS 938, UPMC
M. FABIEN MONGELARD, MC, CRCL, ENS Lyon

Examinateurs
Mme NINA SAADALLAH-ZEIDAN, PR, Université Libanaise
Mme VALÉRIE GRANDJEAN, CR, C3M, Université de Nice
Mme MIRA HAZZOURI, PR,Université Libanaise

Invités
M. IMAD ABOU JAOUDE, M.D., Hôpital Abou Jaoude, Liban
M. JOHNNY AWWAD, M.D., PR, American University of Beirut

Exploration du transcriptome spermatique par le séquençage nouvelle génération et le portrait épigénétique de l'infertilité masculine

Résumé

L'infertilité masculine est actuellement considérée comme un problème majeur qui pose une situation alarmante sur la santé publique. L'oligozoospermie, l'asthénozoospermie et la tératozoospermie sont les trois anomalies les plus connues des spermatozoïdes. Elles affectent, respectivement, la densité, la motilité et la morphologie des spermatozoïdes. Un spermatozoïde anormal est très souvent corrélé à des altérations génétiques et épigénétiques qui peuvent affecter considérablement le transcriptome. Dans ce sens, le sequençage aléatoire du transcriptome entier des spermatozoïdes ou RNA-seq constitue un outil puissant pour caractériser ces maladies. Jusqu'à présent, il n'existe aucune étude exploitant des données RNA-seq chez des hommes présentant de telles anomalies spermatiques. L'objectif principal de notre étude fût d'identifier des profils distincts des modifications du transcriptome de chaque phénotype d'infertilité pour ainsi révéler des gènes-signatures qui tamponnent une spermatogenèse pathologique

Pour ce faire, les transcriptomes des spermatozoïdes de 60 sujets infertiles atteints soit d'oligozoospermie, d'asthénozoospermie ou de tératozoospermie ont été comparés à ceux de 20 patients fertiles. Ces analyses supervisées nous ont conduit à identifier: (i) les gènes clés spécifiques aux différentes anomalies des spermatozoïdes (ii) les voies de signalisation associées, (ii) les différents longs ARNs non codants dérégulés dans ces anomalies.

Au niveau de l'oligozoospermie, les transcrits de spermatozoïdes dérégulés étaient associées à divers stades de la spermatogenèse, y compris le cycle cellulaire méiotique, l'assemblage du complexe synaptonémal, la cohésion des chromatides sœurs, les processus métaboliques de piRNA, le processus catabolique protéique dépendant de la voie de l'ubiquitine, à la réponse aux dommages de l'ADN et particulièrement le processus de fécondation. Quant à l'asthenozoospermia, la spermatogenèse, l'assemblage du cil, des voies métaboliques reliées à la spermatogenèse, la chimiotaxie et la physiologie des cellules immunitaires ont été significativement dérégulés.

De plus, ce qui nous a intéressé au plus était l'analyse des transcrits sous-exprimés qui a permis l'identification de nombreux transcrits associées aux modifications des histones.

Nous avons aussi mis en évidence une sous expression des gènes différentiellement exprimés qui définit la tératozoospermie. Cette sous expression est associée au système ubiquitine-protéasome, à l'organisation du cytosquelette, au cycle cellulaire, à la SUMOylation en réponse aux dommages de l'ADN et aux protéines de réparation ainsi qu'à de nombreux modulateurs épigénétiques.

Les gènes signature de l'oligozoospermie ont été liés au processus de fécondation et les composants de la matrice extracellulaire, tandis que ceux de la tératozoospermie sont liés à la spermatogenèse et la morphogenèse cellulaire, alors que les gènes signature de l'asthénozoospermie sont impliqués dans l'assemblage du ribosome et du flagelle.

En complément de cette étude, nous avons réalisé une étude très globale du paysage épigénétique du sperme des hommes infertiles. Nous avons, ainsi comparé les niveaux des espèces réactives de l'oxygène (ERO), de méthylation de l'ADN, ainsi que l'intégrité de la chromatine dans les spermatozoïdes de 30 individus infertiles avec ceux de 33 individus fertiles. Nos analyses montrent des niveaux élevés d'espèces réactives de l'oxygène chez les individus infertiles. Ces niveaux sont d'une part négativement corrélés avec les niveaux de méthylation globale de l'ADN et d'autre part négativement corrélés avec ceux de la 5-hydroxyméthylcytosine et de la 5-formylcytosine (intermédiaire dans le processus de déméthylation active). Ces derniers suggèrent qu'une infertilité associée au stress oxydatif conditionne l'épigénome du sperme.

En conclusion, l'ensemble de notre travail apporte des ressources précieuses et originales dans la compréhension des pathologies de sperme.

Mots clés : spermatozoïdes ; infertilité masculine ; transcriptomique ; RNA-seq ; épigénome.

Unraveling the sperm transcriptome by next generation sequencing and the global epigenetic landscape in infertile men

Abstract

Male infertility is actually considered as a public alarming health problem. The sperm pathologies spectrum ranges between different phenotypes including oligozoospermia, asthenozoospermia and teratozoospermia depending on the sperm conventional parameters abnormalities. Abnormal sperm is characterized by genetic alterations and epigenetic alterations which can affect the transcriptome extensively. These alterations in RNA profiles are retrospectively indicative of aberrant spermatogenic events. RNA-seq is a powerful tool for comprehensive characterization of whole transcriptome. To date, RNA-seq analysis of sperm from infertile men has not been reported. Our objectives are: (i) recognize key clusters, key pathways and specific gene transcripts for different sperm abnormalities; (ii) catalog the spermatozoal lncRNAs in different sperm pathologies; (iii) identify signature genes which are mechanistically important in the cascade of events driving a pathological spermatogenesis; (iii) portray the global epigenetic landscape in sperm from infertile men. Expression data from 60 sperm samples from 3 groups of infertile men (oligozoospermia, asthenozoospermia, and teratozoospermia) were generated on Illumina HiSeq platform, compared to 20 fertiles, and the resulting gene expression patterns were analyzed for functional enrichment.

Our supervised analyses identified numerous differentially expressed genes between fertile and infertile men. In oligozoospermia, the deregulated spermatozoal transcripts were associated with various stages of spermatogenesis including meiotic cell cycle, synaptonemal complex assembly, sister chromatid cohesion, piRNA metabolic process, ubiquitin-dependent protein catabolic process, cellular response to DNA damage stimulus and interestingly fertilization. As for asthenozoospermia, spermatogenesis, cilium assembly, metabolic-related pathways, chemotaxis and immune cell physiology were most significantly differentially expressed. Interestingly, numerous transcripts associated with histone modifications were highly down-regulated.

With regards to teratozoospermia, we evidenced sperm-specific differentially expressed genes which are involved in the ubiquitin-proteasome, cytoskeleton organization, the cell cycle pathway, SUMOylation of DNA damage response and repair proteins, as well as many putative epigenetic modulators of gene expression.. We also attempted to identify distinct patterns of gene expression changes that were definite to the different abnormal sperm phenotypes in infertile men relative to controls. Signature genes of oligozoospermia were over-enriched by genes involved in fertilization and extracellular matrix components, while signature genes of teratozoospermia were enriched by genes involved in spermatogenesis and cellular components involved in morphogenesis, whilst signature genes of asthenozoospermia were enriched by genes implicated in ribosome and cilium assembly.

We complemented this work by a parallel epigenetic analysis of the global epigenetic landscape in infertile men. We compared the levels of reactive oxygen species (ROS), DNA integrity and global epigenetic parameters in sperm from 33 infertile subjects with abnormal semen parameters compared to fertile individuals. We pointed out that infertile men are characterized by strikingly high levels of reactive oxygen species (ROS) which were in part negatively correlated with the global DNA methylation, and positively correlated with the levels of 5-hydroxymethylcytosine and 5-formylcytosine (active demethylation intermediates). These findings suggest that male infertility associated with oxidative stress shapes the sperm epigenetic landscape.

In summary, this original work yielded a transcriptional portrait of sperm abnormalities and provided valuable resources that would further elucidate sperm pathologies.

Keywords: spermatozoa; male infertility; transcriptomics; RNA-seq; epigenome.

ACKNOWLEDGMENTS

As I come to close this chapter in my life, I would like to thank many unique people, with whom I was lucky to interact and who made my time at graduate school a wonderful experience.

First, I would like to express a deep gratitude to my supervisors, 'my academic moms', Mira and Valérie. Thank you both for enabling me to grow and believe in myself as an independent scientist, for patiently guiding me through different steps of this endeavor and for always being there when I needed.

Thank you Mira, for creating new opportunities for me, and providing endless intellectual and emotional support. Thank you Valérie, for providing invaluable support, insightful comments and ideas. I am very fortunate to have you both as my mentors.

I would also like to thank Khaled Hazzouri for generously sharing your time and knowledge: this has been an enriching learning experience.

I thank my dear collaborators and colleagues: Imad Abou Jaoude, Johnny Awwad, Youssef Gemayel, Elias Saliba, Nahida Azar, Rita Assaf, Christina Karam, Layal Hamdar, Aurore Assaker, Dalia Khalifeh and all members of the molecular biology lab at the Faculty of Public Health in the Lebanese University. Thank you for taking the time to help me with different aspects of my science.

I thank the members of my dissertation and examining committees: Rachel Levy, Fabien Mongelard, Nina Saadallah-Zeidan and Anne-Amandine Chassot. Thank you, for your time, support and insights.

TABLE OF CONTENTS

LIST OF TABLES ..... **XI**

LIST OF FIGURES ..... **XII**

LIST OF ABBREVIATIONS ..... **XV**

INTRODUCTION **1**

CHAPTER I **3**

MOLECULAR CONTROL OF SPERMATOGENESIS AND POTENTIAL CAUSES OF MALE INFERTILITY **3**

1.1. "Drawing" the molecular portrait of human spermatogenesis - a review of the past familiar faces and a glimpse into the new features 3

1.1.1. Molecular control of spermatogenesis 3

1.1.1.1 Introduction 3

1.1.1.2. Male germ cells development during embryogenesis 4

1.1.1.3. Male germ cells during the mitotic phase 5

1.1.1.4. Male germ cells during the meoitic phase 7

1.1.1.5. Post-meoitic phase/ spermiogenesis 8

1.1.1.5.1. Acrosome biogenesis 8

1.1.1.5.2. Manchette biogenesis 9

1.1.1.5.3. Axoneme development 9

1.1.1.5.4. Cytoplasmic removal 10

1.1.1.5.5. Chromatin remodeling 10

1.1.1.5.6. The expression of late spermatid genes 10

1.1.1.5.7. Expression of cellular elements for energy production and motility 11

1.1.1.6. Spermiation 11

1.1.1.7. Gene expression pattern during spermatogenesis 12

1.2 Molecular and epigenetic causes of infertility 12

1.2.1. Molecular insights into the causes of male infertility 12

1.2.2. Epigenetic layers and players in the context of male infertility: an extensive review of the general theory and the current state-of-the-art 15

1.2.2.1. Introduction 15

1.2.2.2. The sperm methylome 16
1.2.2.2.1. Basic concepts 16
1.2.2.2.2. MTHFR enzyme activity 17
1.2.2.2.3. Role of the DNMTs in spermatogenesis 18
1.2.2.2.4. Tet-dependent DNA demethylation 18
1.2.2.2.5. Global DNA methylation 19
1.2.2.2.6. Methylation levels at promoter of CpG islands 19
1.2.2.2.7. Methylation status of non-imprinted genes 19
1.2.2.2.8. Methylation at repetitive elements 21
1.2.2.2.9. Methylation at imprinted genes 21
1.2.2.3. The spermatozoal chromatin proteome 23
1.2.2.3.1. Basic concepts 23
1.2.2.3.2. Histones: post-translational modifications and variants 23
1.2.2.3.3. Chromatin remodeling 26
1.2.2.4. Non-coding RNAs 27
1.2.2.4.1. Basic concepts 27
1.2.2.4.2. MicroRNAs (miRNAs) and spermatogenesis 28
1.2.2.4.3. Piwi-interacting RNAs (piRNAs) and spermatogenesis/infertility 28
1.2.2.4.4. Long non-codong RNAs (lncRNAs) 29
1.2.2.5. High order nuclear organization in sperm 29
Conclusion 30

CHAPTER II 33
A REVIEW OF THE TRANSCRIPTOMICS STUDIES 33

2.1. Review in literature data 33
2.1.1. Data collection 33
2.1.2. Data mining results 34
2.1.2.1. Teratozoospermia 34
2.1.2.2. Normozoospermia 34
2.1.2.3. Asthenozoospermia 36
2.1.2.4. Oligoasthenozospermia 39
2.1.2.5. Oligozoospermia 40
2.2. New emerging technology approaches: Sperm transcriptome profiling using Next generation sequencing 42
Conclusion 46

CHAPTER III 49
MATERIALS AND METHODS 49

3.1. Optimization of sperm RNA isolation 49
3.1.1. Materials and methods 51
3.1.1.1 Isolation of sperm cells 51
3.1.1.2. Cell lysis 51
3.1.1.3. Testing the effectiveness of lysis 51
3.1.1.4. RNA isolation 52
3.1.1.5. DNA isolation 52
3.1.1.6. Nucleic acids QC metrics 52
3.1.1.7. RNA library preparation and sequencing 52
3.1.1.8. Global 5-methylcytosine (5-mC) measurement 53
3.1.1.9. Statistical analysis 53
3.1.2. Results and discussion 54
3.2. The ABCs of RNA-seq: An explanatory review on the basics of experimental design, laboratory performance and computational analysis in RNA-seq experiments 61
3.2.1. Design considerations 63
3.2.1.1. Library type 63
3.2.1.2. Replicates 63
3.2.1.3. Sample pairing 64
3.2.1.4. Pooling 64
3.2.1.5. Normalization of RNA-seq data 64
3.2.1.6. Sequencing decisions 65
3.2.2. Sequencing, at a glance (Illumina) 65
3.2.3. RNA-seq data analysis workflow 67
3.2.3.1. Quality control of raw data 67
3.2.3.2. Preprocessing 68
3.2.3.3. Alignment (mapping) 69
3.2.3.4. Quality control for aligned reads 72
3.2.3.5. Reads assembly and quantification 74
3.2.3.6. Differential expression analysis: normalizing and transforming read counts 77
3.2.3.7. Variability within the experiment: data exploration 78
3.2.3.8. Differential gene expression analysis 80
3.2.3.9. Exploratory plots following DGE analysis 80
3.3. Experimental work for sperm transcriptomic analysis 81
3.3.1 Patient population 81
3.3.2. Semen analysis 81

3.3.3. Sperm purification 82
3.3.4. RNA preparation 82
3.3.5. RNA-Seq library preparation 83
3.3.6. Sequencing Libraries 83
3.3.7. Bioinformatics 83
3.4. Experimental work for sperm epigenetic analysis 84
3.4.1. Study participants 84
3.4.2. Sperm damage assessment 84
3.4.2.1. Measurement of reactive oxygen species production 84
3.4.2.2. Sperm Chromatin dispersion (SCD) test 85
3.4.3. Sperm chromatin abnormalities assessment 85
3.4.3.1. Sperm chromatin structure assessment 85
3.4.3.2. Histone retention assessment 85
3.4.4.1. DNA extraction 86
3.4.4.2. Global 5-mC, 5-hmC and 5-fC measurement 86
3.4.5. Statistical analysis 86

CHAPTER IV 90
RESULTS 90

4.1. Comprehensive transcriptomic analysis of human spermatozoa in male infertility 90
4.1.1. Sperm transcriptome in oligozoospermia 90
4.1.1.1. Total number of transcripts and their expression level 90
4.1.1.2. Nature of spermatozoa transcripts 90
4.1.1.3. Functions of spermatozoal transcripts 91
4.1.1.3.1. Down-regulated genes 91
4.1.1.3.2. Up-regulated genes 96
4.1.1.3.3. Long non-coding RNAs 97
4.1.2. Sperm transcriptome in asthenozoospermia 104
4.1.2.1. Total number of transcripts and their expression level 104
4.1.2.2. Nature of spermatozoa transcripts 104
4.1.2.3. Enrichment analysis of protein-coding genes for identification of biological and functional differences 104
4.1.2.3.1. Down-regulated genes 104
4.1.2.3.2. Up-regulated genes 109
4.1.2.3.3. Long non-coding RNAs 109
4.1.3. Sperm transcriptome in teratozoospermia 116

4.1.3.1. Total number of transcripts and their expression level 116
4.1.3.2. Nature of spermatozoa transcripts 116
4.1.3.3. Functions of spermatozoal transcripts 116
4.1.3.3.1. Complete genes differentially expressed 116
4.1.3.3.2. Down-regulated genes 119
4.1.3.3.3. Up-regulated genes 123
4.1.3.3.4. Long non-coding RNAs 123
4.1.4. Dissection of genes expression signatures associated with abnormal sperm phenotypes in infertile men by RNA seq analysis 128
4.1.4.1. Repertoire of DEG's in sperm from infertile men as determined by Next Generation Sequencing of cDNA libraries 128
4.1.4.2. Data analysis strategy 128
4.1.4.3. Molecular characterization of sperm pathologies types of infertile men 129
4.1.4.4. Identification of the best set of genes discriminating the different phenotypes of sperm pathologies of infertile men 130
4.1.4.5. Transcriptional landscape of long non-coding RNAs in sperm from infertile men 132
4.2. Characterization of the sperm global epigenetic landscape associated with oxidative stress in infertile men 136
Conclusion 143

CHAPTER V 145
GENERAL DISCUSSION 145

5.1. Differential analysis of spermatozoal transcripts from infertile men with different sperm pathologies using RNA-seq 145
5.2. Assessment of global epigenetic changes and oxidative damage in sperm from infertile men 151

CONCLUSION AND FUTURE DIRECTIONS 155

REFERENCES 161

Supplementary materials 187

LIST OF TABLES

Table 1. *Male infertility-associated epigenetic modifications: DNA methylome alterations. 23*
Table 2. *Male infertility-associated epigenetic modifications: Aberrations in the spermatozoal chromatin proteome. 26*
Table 3. *Male infertility-associated epigenetic modifications: Altered sperm non-coding RNAs. 29*
Table 4. *Schematic representation of the microarray-based transcriptome studies evaluated in the present review.41*
Table 5. *The percentage of spermatozoa with intact cell membranes, after three physical treatments (means±SEM). 55*
Table 6. *Factors and Levels Assayed in Fractional Factorial Design for the optimisation of sperm RNA extraction: Factors Screening. 58*
Table 7. *Summary of the experiments to find optimal conditions of sperm RNA isolation. 58*
Table 8. *Functional enrichments of the protein-protein interaction (PPI) network for the mouse orthologs associated with male infertility phenotype. 92*
Table 9. *The spermatozoal enriched transcripts from oligozoospermic men associated with the oligozoospermia mouse phenotype obtained from functional annotation analysis (Toppgene) based on their abundancy. 93*
Table 10. *The spermatozoal enriched transcripts from oligozoospermic men associated with the abnormal sperm number mouse phenotype obtained from functional annotation analysis (Toppgene) based on their abundancy. 94*
Table 11. *Functional enrichments of the protein-protein interaction (PPI) network for the mouse orthologs associated with abnormal sperm number. 96*
Table 12. *List of long non-coding RNAs associated with oligozoospermia. 100*
Table 13. *The molecular functions, biological processes, cellular localization, and pathways implicated of spermatozoal transcripts in asthenozoospermia obtained from functional annotation analysis (Toppgene) of the abundantly down-regulated set of genes significantly differentially expressed.. ..... 105*
Table 14. *The spermatozoal enriched transcripts from asthenozoospermic men associated with the abnormal sperm motility mouse phenotype obtained from functional annotation analysis (Toppgene) based on their abundancy. .. 107*
Table 15. *Functional enrichments of the protein-protein interaction (PPI) network for the mouse orthologs associated with abnormal sperm motility phenotype. 107*
Table 16. *The molecular functions, biological processes involved, cellular localization, and pathways implicated of spermatozoal transcripts in teratozoospermia obtained from functional annotation analysis (Toppgene) of the complete set of genes significantly differentially expressed.. ..... 118*
Table 17. *Functional enrichments of the protein-protein interaction (PPI) network for the mouse orthologs associated with teratozoospermia phenotype. 119*
Table 18. *The spermatozoal enriched down-regulated transcripts from teratozoospermic men associated with the teratozoospermia mouse phenotype obtained from functional annotation analysis (Toppgene) based on their abundancy. 121*
Table 19. *Biological pathway analysis of the set of common signature genes differentially expressed orthologs related to "male infertility" phenotype identified by RNA seq. 131*
Table 20. *Comparison of sperm quality between the infertile and fertile study groups. 138*

LIST OF FIGURES

Figure 1. *Key molecular events for PGCs specification and migration.* *5*
Figure 2. *Stages and molecular events associated with human spermatogenesis. Sequence of the stages spanning spermatogonial stem cells differentiation during embryogenesis and the release of mature spermatozoa.* *11*
Figure 3. *Timeline of studies involving transcriptomic profiling using Next generation sequencing from human spermatozoa.* *46*
Figure 4. *Sperm head morphometry analysis (mean ± SEM) after three physical treatments (osmotic choc, snap freeze, osmotic choc followed by snap freezing) compared with spermatozoa in fresh semen.* *56*
Figure 5. *Histogram of relative fluorescence intensities (DAPI), reported as arbitrary fluorescence units (a.f.u.), after adhering to sperm nucleus under different physical treatment conditions.* *56*
Figure 6. *Electropherograms (in a 1% agarose gel run for 40 minutes at 150 Volts) of DNA from sperm cells.* *61*
Figure 7. *The levels of methylated DNA (5-mC) in total sperm DNA.* *61*
Figure 8. *Illumina sequencing workflow using the sequencing by synthesis method adapted from Illumina.* *66*
Figure 9. *FASTQC format and reports.* *67*
Figure 10. *Quality control checks a: before quality trimming and b: after quality trimming.* *68*
Figure 11. *Three RNA mapping strategies: a: De novo assembly; b: Align to reference genome; c: Align to transcriptome.* *69*
Figure 12. *An overview of cufflinks workflow* *71*
Figure 13. *a: Tuxedo software components; b: Tuxedo protocol for differential gene expression* *72*
Figure 14. *Post alignment quality control representations.* *73*
Figure 15. *a: HTseq different modes for reads quantification; b: Gene transfer format (GTF) features* *75*
Figure 16. *Overview of Cufflinks; b: Scheme of a simple deBruijn graph-based transcript assembly* *76*
Figure 17. *Options for differential gene expression (DGE) analysis.* *77*
Figure 18. *Illustration of variability within the experiment.* *79*
Figure 19. *Exploratory plots following DGE analysis* *81*
Figure 20. *Protein interaction map. Map was prepared using the STRING web tool (https://string-db.org) for the proteins encoded by the set of genes associated with the male infertility phenotype..* *92*
Figure 21. *Protein interaction map. Map was prepared using the STRING web tool (https://string-db.org) for the proteins encoded by the set of genes associated with the abnormal sperm number phenotype..* *96*
Figure 22. *(a) RNA–protein interaction network map (v1.0) prepared using the RAIN web tool (https://rth.dk/resources/rain/) for the long non-coding RNAs (lncRNAs) associated with oligozoospermia. (b) lncRNAs- top ranked candidate genes encoding proteins interaction map associated with oligozoospermia..* *98*
Figure 23. *Protein interaction map. Map was prepared using the STRING web tool (https://string-db.org) for the proteins encoded by the set of orthologous genes associated with abnormal sperm motility phenotype..* *106*
Figure 24. *(a) RNA–protein interaction network map (v1.0) prepared using the RAIN web tool (https://rth.dk/resources/rain/) for the long non-coding RNAs (lncRNAs) associated with asthenozoospermia. (b) lncRNAs- top ranked candidate genes encoding proteins interaction map associated with asthenozoospermia.* *110*
Figure 25. *Pathway analysis of the ribosome (Kegg pathway) using the Pathview server (https://pathview.uncc.edu/). Genes significantly differentially expressed (DE) in asthenozoospermia* *112*
Figure 26. *Pathway analysis of oxidative phosphorylation (a) and glycolysis (b) (Kegg pathway) using the Pathview server (https://pathview.uncc.edu/). Genes significantly differentially expressed (DE) in asthenozoospermia.* *113*
Figure 27. *Pathway analysis of the calcium signaling pathway (Kegg pathway) using the Pathview server (https://pathview.uncc.edu/). Genes significantly differentially expressed (DE) in asthenozoospermia.* *114*
Figure 28. *Protein interaction map. Map was prepared using the STRING web tool (https://string-db.org) for the proteins encoded by the set of mouse orthologous genes associated with teratozoospermia phenotype. Colored lines denote interactions and network nodes represent proteins.* *117*
Figure 29. *Pathway analysis of Proteasome and Ubiquitin-mediated proteolysis (Kegg pathway) using the Pathview server (https://pathview.uncc.edu/). Genes significantly differentially expressed (DE) in teratozoospermia.* *121*
Figure 30. *(a) RNA–protein interaction network map (v1.0) prepared using the RAIN web tool (https://rth.dk/resources/rain/) for the long non-coding RNAs (lncRNAs) associated with teratozoospermia. These candidate proteins were further prioritized based on topological features in protein-protein interaction network. (b) lncRNAs- top ranked candidate genes encoding proteins interaction map associated with asthenozoospermia..* ... *124*
Figure 31. *Pie chart representing the alignment statistics of reads which align to reference genome.* *128*

Figure 32. *Transcriptional landscape of human sperm RNAs in ejaculates of infertile men with different sperm abnormalities types.* *130*
Figure 33. *Identification of the best set of genes discriminating the different phenotypes of sperm pathologies of infertile men.* *131*
Figure 34. *Relationship between sperm global DNA methylation with sperm histone retention (abnormal chromatin packaging) a) by Aniline blue test and sperm chromatin alteration b) by Toluidine blue assay.* *139*
Figure 35. *Relationship between sperm global DNA methylation with sperm ROS a) by NBT test and sperm DNA fragmentation (SDF) b) by SCD test.* *139*
Figure 36. *Relationship between sperm global DNA hydroxymethylation with sperm ROS by NBT test (a) and sperm DNA fragmentation by SCD test (b).* *139*

LIST OF ABBREVIATIONS

5-fC: 5-formylcytosine
5-hmC: 5-hydroxymethylcytosine
5meC: 5th carbon of cytosine
8-OHdG: oxidative DNA adduct
Ad: adult dark
aES: apical ectoplasmic specialization
BAM: binary Alignment/Map
BH: benjamini-hochberg
BTB: blood testis barrier
CaM: calcium/calmodulin
CAR: chromatin-associated RNA
CIGAR: Concise Idiosyncratic Gapped Alignment Report
CSF1: colony stimulating factor 1
DAPI: 4′,6-diamidino-2-phenylindole
DGE: differential gene expression
DM-CpG: differentially methylated CpG
DNA: deoxyribonucleic acid
DNMT: de novo methyltransferases
DTT: dithiothreitol
ECM: extracellular matrix
EG: exclusive gene
ELISA: enzyme-linked immunosorbent assay
Enc: Encore NGS Multiplex System I
ERO: espèces réactives de l'oxygène
FC: fold change
FDR (false discovery rate)
FFPES: Formalin Fixed Paraffin Embedded System
FPKM: fragments per kilobase of exons per million mapped
GDE: genes differentially expressed
gDNA: genomic DNA
GnRH: gonadotropin releasing hormone
GO: Gene ontology
GTF: Gene transfer format
HAT: histone aceltyltransferase
HDAC: histone deacetylase
HMR: hypomethylated DNA region
ICSI: intracytoplasmic sperm injection
IFT: intraflagellar transport
IMT: intra-manchette transport
IQR: Inter-Qartile Range
IVF: in vitro fertilization
KEGG: Kyoto Encyclopedia of Genes and Genomes
lincRNA: long intergenic noncoding RNA
lncRNA: long non-coding RNA
MBP: methyl-CpG binding protein
MDS: multidimentional scaling
miRNA: microRNA
MS: mass spectrometry
MSI: meiotic sex chromosome inactivation
MTHFR: methylenetetrahydrofolate reductase
NBT: Nitro Blue Tetrazolium

ncRNA: non-coding RNA
NGS: next generation sequencing
NMD: Nonsense-Mediated Decay
Nt: nucleotides
ODF : outer dense fiber
OR: Ovation® RNA-Seq System V2
PCA: Principle Component Analysis
PCR: polymerase chain reaction
PE: Paired-end
PGC : primordial germ cells
piRNA: Piwi-interacting RNA
PPI: protein–protein interaction
PRM : protamine
QC: quality control
qPCR: quantitative polymerase chain reaction
RA: retinoic acid
RAIN: RNA-protein Association and Interaction Network
RBP: RNA-binding protein
RIN: RNA integrity number
RNA: ribonucleic acid
RNA-seq : RNA sequencing
RNP: ribonucleoprotein particle
ROS: reactive oxygen species
RPKM: reads per kilobase of exons per million mapped
rRNA: ribosomal RNA
RT-PCR: real time polymerase chain reaction
SAC: spindle assembly checkpoint
SAM: Sequence Alignment/Map
SCD: Sperm Chromatin dispersion
SCF: stem cell factor
SEM:
SEP: sperm expression profiles
Sf1: Steroidogenic factor 1
SP: SeqPlex RNA Amplification
SR: Single read
SREs: sperm RNA elements
sSMC : small supernumerary marker chromosome
STRING : Search Tool for the Retrieval of Interacting Genes/Proteins
SVM : support vector machine
TB: toluidine blue
TBC : tubulobulbar complex
TBDC: toluidine blue dark violet cell
TBLC: toluidine blue light blue cell
TET : ten-eleven-translocation
TMM: Trimmed Mean of M-values
TP : transition protein
UL: Ovation Ultralow Library Systems
UPP: ubiquitin-proteasomal pathway
VDAC : voltage-dependent anion channels
WHO: world health organization
βME: beta mercaptoethanol

INTRODUCTION

Spermatozoa transfers transcripts as well as epigenetic information to the oocyte during fertilization. In this thesis I present two studies. In the first one, I applied an emerging technology next generation sequencing technology (NGS)) to systematically study human sperm Ribonucleic Acids (RNAs) in infertile men with different sperm pathologies and to gain insights into genes associated with male infertility. In this manner, paternal RNAs can serve as a biomarker to diagnose patients with subfertility or predict pregnancy outcomes [1]. In the second part, I characterized the sperm global epigenetic landscape in infertile men. as anyn changes in the epigenome may carry a high risk to be transmitted to the offspring, thus this examination of the epigenetic facet of sperm is valuable in part for characterizing the mechanisms of sperm pathogenesis, and in determining the risks of using a pathological sperm in assisted reproductive technologies (ART).

Outline of the thesis

The work presented in this thesis is divided into four major parts: (1) (Chapters 1 & 2) Introduction including extensive reviews of the literature and updated evidence regarding: the molecular basis of spermatogenesis, the sperm epigenetic layers and players in infertile men and the transcriptomic studies of male infertility; (2) (Chapter 3) Materials and methods including descriptions of the population, and the experimental methodology; (3) Results of the transcriptomic analysis of the sperm from infertile patients presenting different semen analysis abnormalities (count, motility, morphology) (Chapter 4) and assessment of the global epigenetic landscape in sperm from infertile men (Chapter 4) ; (4) A general conclusion of the main findings followed by the future directions (Chapter 5).

Chapters 1 & 2 present the molecular basis of spermatogenesis, the sperm epigenetic layers and players in infertile men and the transcriptomic studies of male infertility.

In **chapter 3,** we describe the laboratory and computational methods used in this work. The aim is to provide an overview of the protocol of RNA extraction refinement and optimization for expression profiling, and how genomic data of sperm compiled in this work has been generated and analyzed. Therefore, we first describe methods used for sperm RNA

purification which are based on the experimental work performed in the framework of this thesis, followed by a description of the basics of RNA sequencing (RNA-seq) analysis. Finally, we will present the experimental work methodologies used for sperm transcriptomic and epigenetic analysis.

In **chapter 4,** we present results from a genome-wide expression analysis of sperm from infertile men at different pathologic conditions (oligozoospermia, asthenozoospermia, teratozoospermia) using RNA-Seq. Additionally, in the second section of this chapter, we characterized the global epigenetic landscape in sperm from infertile men associated with oxidative stress.

In **Chapter 5,** we discussed the rationale for appreciating the use of NGS for transcriptome analysis. Next, we presented our empirical evidence about the aberrant transcriptome in sperm for each sperm pathology.

In **Chapter 6** we discussed the association between oxidative stress and an aberrant sperm epigenome in infertile men.

Finally, a **conclusion** highlights the significant strengths of this research that have been covered in this thesis as well as future directions for expanding this field of research, including some suggested future experiments.

In the same context we highlight the publication of two scientific publications listed below that tackled the sperm Deoxyribonucleic Acid (DNA) fragmentation in infertile men and antioxidants régime for sperm DNA damage. They are posted in supplememntary materials at the end of this manuscript.

Choucair, Fadi B., Eliane G. Rachkidi, Georges C. Raad, Elias M. Saliba, Nina S. Zeidan, Rania A. Jounblat, Imad F. Abou Jaoude, and Mira M. Hazzouri. "High level of DNA fragmentation in sperm of Lebanese infertile men using Sperm Chromatin Dispersion test." *Middle East Fertility Society Journal* 21, no. 4 (2016): 269-276.

Choucair, Fadi, Elias Saliba, Imad Abou Jaoude, and Mira Hazzouri. "Antioxidants modulation of sperm genome and epigenome damage: Fact or fad? Converging evidence from animal and human studies." *Middle East Fertility Society Journal* 23, no. 2 (2018): 85-90.

CHAPTER I

MOLECULAR CONTROL OF SPERMATOGENESIS AND POTENTIAL CAUSES OF MALE INFERTILITY

In this section, we will review in-depth the molecular control mechanisms of spermatogenesis. Spermatogenesis is governed by timely genetic and epigenetic events that we will try to piece together into a coherent complete biological picture.

An ever-expanding number of molecules have been implicated in the regulation of epigenetic processes during spermatogenesis. The second section will provide an extensive review of aberrant epigenetic regulation/modification and their association with fertility.

1.1. "Drawing" the molecular portrait of human spermatogenesis - a review of the past familiar faces and a glimpse into the new features

1.1.1. Molecular control of spermatogenesis

1.1.1.1 Introduction

Spermatogenesis is a highly conserved process among many organisms consisting of three key phases: a mitotic amplification phase, a meiotic phase, and a post-meiotic phase also known as spermiogenesis. Almost all the events during this process are orchestrated by complex molecular pathways, which are endowed by multiple genetic and/or epigenetic factors. A systematic understanding of the critical molecular events involved in spermatogenesis may help to pave the way for the diagnosis of sperm pathologies. In this paper, we will retrospectively review the dominant events operating during the initiation and progression of spermatogenesis in humans. We will provide an update on the orchestrated genetic and epigenetic networks in male germ cells during the mitotic phase, meiotic phase and spermiogenesis. We have considered genetic players based on human gene expression, and point mutations, data and mouse genetic and molecular aspects. Defining these broad mechanisms is a necessary prelude to determine the timing of aberrant events in the context of infertility.

1.1.1.2. Male germ cells development during embryogenesis

Lessons gleaned from mouse models revealed that primordial germ cells (PGCs) originate from epiblast cells in the yolk sac in response to the adjacent endoderm and extra-embryonic ectoderm signals. Several transcriptional regulator factors are required for the acquisition of the germ cell fate including intrinsic factors, such as Prdm1 (Positive Regulatory Domain I-Binding Factor 1) also known as Blimp1, a known transcriptional repressor which has a critical role in the foundation of the germ cell lineage [2], and extrinsic factors, such as BMPs (bone morphogenetic proteins) embryonic-derived signals [3].

In humans, it was suggested that BMP4- and WNT3A (expressed by embryonic epiblast cells) -mediated upregulation of PRDM [4]. Worth noting, WNT3A was suggested to be required for the responsiveness of epiblast cells to BMP4. PRDM1 in turns modulate the divergence of neural or germline fates through repression of *SOX*2 (SRY box-2). SOX2 is a transcription factor that is essential for maintaining self-renewal, or pluripotency, of undifferentiated embryonic stem cells [5]. In addition, SOX17 may also act with other factors, such as OCT4, in the generation of human germ cells. Following BMP4 signaling, epiblast cells begin to express Stella, which permits the specification of PGCs [4]. Thereafter, PGCs migrate through the hindgut to the genital ridges. Both migration and proliferation/apoptosis of PGCs during migration are directed by interactions between receptors, expressed by PGCs such as c-kit, and ligands, secreted by the surrounding somatic cells namely SCF (stem cell factor).

The migration process occurs within a rigidly controlled time frame aimed to a specific site. If PGCs end up in the wrong place at the wrong time they underwent apoptosis. The PGCs express the tyrosine kinase receptor c-KIT and the surrounding cells express its ligand the stem cell factor (SCF) [6]. In addition, another receptor the G-protein coupled receptor CXCR4 interacted with the chemokine CXCL12 (SDF1) is responsible for the maintenance of adult stem cell niches [7]. Moreover, several germ cell progenitor and pluripotency-associated genes are expressed for the maintenance of PGCs during migration such as Oct-4 or Nanos3 [8-9]. At this level, PGCs begin to express the migratory germ cell marker VASA which is the hallmark of postmigrating germ cells [10]. The above-mentioned events are summarized in **Figure 1**.

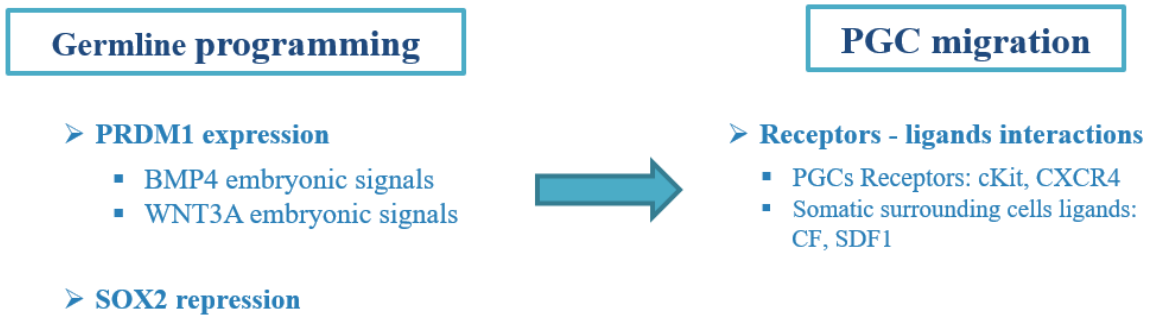


Figure 1. Key molecular events for PGCs specification and migration.

1.1.1.3. Male germ cells during the mitotic phase

PGCs now colonize the genital ridge, surrounded by somatic cells. They are now referred to as "gonocytes". A genetically controlled masculinizing signal imposes the testicular differentiation, and promotes male germ cells fate through the secretome of the somatic surrounding compartment. Several genes are implicated in the male path sexual development including *DHH*, *FGF9*, *M33*, *DMRT1*, *AMH*, *SRY* and *SOX9*[11] which are regulatory genes known to be expressed in the mammalian genital ridge prior to sexual differentiation [12]. SRY and SOX9 are the 2 key genes involved in testis development. In fact, *Sry* expression regulates its downstream target SOX9, which in turn promotes male development and prevents ovarian pathways. *Sry* is expressed by pre-Sertoli cells and promotes their differentiation. *GATA4*, *WT1* and *SF1* are transcriptional factors involved in sex determination, have been shown to play a major role in controlling the gonadal specific *Sry* promoter [13-14]. Expression of Sry in the gonadal primordium causes the somatic supporting cell precursors of the gonad to differentiate as Sertoli cells. The expression of Steroidogenic factor 1 (Sf1) labels the cells of the mesonephros as Leydig precursors [15].

Gonocytes represent the fetal reservoir of stem cells. They are maintained in a quiescent undifferentiated state owing to the NOTCH1 signaling in Sertoli cells [16-17]. NOTCH is a negative regulator of GDNF and CYP26B1, which balances the effects of FGF2 (fibroblast growth factor 2) - a Sertoli cells growth factor activating GDNF- and SOX9/SF1 activator of CYP26B1 thus keeping GDNF and CYP26B1 at basal levels for gonocytes maintenance. Normally, CYP26B1 in Sertoli and Leydig cells regulates the concentration of retinoic acid which subsequently controls germ cell progression toward meiosis. However, the presence of CYP26B1 in the perinatal testis prevents meiotic entry. More so, GDNF induces the

activation of RET signaling which is required for gonocyte proliferation and maintenance. Thus, down-regulation of GDNF may contribute to the entry of gonocytes into mitotic arrest. The transformation of gonocytes into adult spermatogonia is usually completed postnatally by the age of 6 months. This conversion of gonocytes into adult dark (Ad) spermatogonia is driven by a significant increase in gonadotropin releasing hormone (GnRH) secretion which induces gonadotropin and testosterone production. Ad spermatogonia appear at three months of age, and have a characteristic dark nucleus. This period is termed mini puberty. Several factors are found to be involved in the gonocyte-to-Ad spermatogonia transition: transcription factors such as *DMRTC2*, *EGR2*; growth factor like proteins such as *NRG1*, *NRG3*; and RNA binding and Y chromosome encoded genes such as *RBMY1B*, *RBMY1E*, *RBMY1J* and *TSPY4* [18]. It has been suggested that the NOTCH signaling in Sertoli cells is also involved in the transition of quiescent prospermatogonia into differentiated spermatogonia. In this manner, overactivation of NOTCH1 signaling would induce in part the up-regulation of the KIT receptor, and suppresses the expression of GDNF and CYP26B1.

This induces direct differentiation into KIT-expressing proliferating spermatogonia. It is worth noting that CYP26B1 is down-regulated and represses meiosis until puberty [16]. Add to that, the chromatin-modifying protein Swi-independent 3a (SIN3A) [19] and several members of the transforming growth factor beta (TGF-β) superfamily, such as the activins, inhibins, and bone morphogenetic proteins (BMPs) [20] play a role in the regulation of mitotic re-entry. During the first 2–3 years in boys, gonocytes transition into reserve active adult dark and adult pale spermatogonia [21]. It is important to distinguish between these two subpopulations, one responsive to Gdnf which adopts a self-renewing, stem cell phenotype, while the other advances to a differentiated, Kit-positive cell type, incapable of self-renewal [22]. A_{dark} are considered the reserve stem cells whereas the A_{pale} were the renewing stem cells [23]. The niche factor GDNF is secreted by Sertoli cells and promotes SSC (spermatogonial stem cells) renewal and maintenance.

GDNF binds to the GDNF-receptor α1 (GFRα1) and catalyses the activation of the c-RET receptor, which activates several signalling cascades, such as PI3K/AKT, MEK and SCR kinases [24]. This influence transcription factors such as BCL6B, ETV5, ID4 [23], and NANOS2 [25-26].Other genes not regulated by GDNF (e.g., *Zbtb16*, *Taf4b*, and *Lin28*), are likely controlled by different signals, promote SSC self-renewal and block differentiation [23]. Sertoli cell also secrete CXCL12 signals and via the CXCR4 receptor present on SSCs are able to regulate self-renewal through an undefined intercellular signaling pathway [27]. In

addition, Sertoli cells secreted-FGF2 activates MAP2K1and subsequently upregulates of the *Etv5*, *Bcl6b*, and *Lhx1* genes expressed in spermatogonial stem cells [24, 28]. Furthermore, a new set of proteins have been found involved in spermatogonial self-renewal are the PIWI proteins (PIWIL2 and TDRD1) [3, 29].

In parallel, other extrinsic factors , such as colony stimulating factor 1 (CSF1) which is produced by Leydig cells and some peritubular myoid cells also drive SSC maintenance and self-renewal [30]. However, formation of the male germinal epithelium and SSC homing are achieved mainly through adhesion molecules such as integrins which mediate the binding of cells to the extracellular matrix or surrounding cells namely integrin beta 1 (*ITGB1*) [31]. It was proposed that two waves of spermatogonial proliferation occur: one between 4 and 8 years that leads to a modest increase of total spermatogonia numbers and another one around puberty which sharply expands the spermatogonia pool [13]. By 5 years of age, the adult dark and adult pale spermatogonia begin to differentiate into type B spermatogonia which by the age of 10 represent only 10% of all spermatogonia [16]. Type B spermatogonia undergo mitotic proliferation until puberty. In GDNF-responsive Ap spermatogonia, GDNF works synergistically with Neuregulin-1 (*NRG1)* induce spermatogonial differentiation to regulate the balance between SSC self renewal and differentiation [32]. Likewise, the inhibition of FOXO transcription factors through the activation of the PI3K/AKT pathway in the presence of extrinsic factors, results in down-regulation of the GDNF receptor *c-Ret* and commitment to spermatogenic differentiation [1, 28]. However, in Kit-positive Ap spermatogonia binding of Kit ligand also known as SCF (stem cell factor) secreted by Sertoli cells to the Kit receptor is critical for spermatogonial differentiation [3]. Additionally, *Sohlh1* (spermatogenesis and oogenesis helix-loop-helix 1) and *Sohlh2* are transcription factors that play a pivotal role in the differentiation of spermatogonia [33] by controlling Kit expression.

1.1.1.4. Male germ cells during the meoitic phase

Meiotic differentiation of spermatogonia is mediated by retinoic acid (RA) secreted by Sertoli cells [34]. In fact, RA signaling activates the PI3K/AKT/mTOR signaling pathway to induce the translation of *Kit* mRNAs [35]. Moreover, RA induces *Stra8* expression which was suggested to play roles in premeiotic DNA synthesis and in meiotic progression [35-36] . Expression of *DAZ1*, *DAZL* and *BOULE [37]* is required for meiotic entry, where type B spermatogonia are divided into pre-leptotene spermatocytes. DAZ proteins can regulate the expression, transport and localization of target mRNAs of proteins. Such targets in mouse are genes related to: cell cycle progression such as *Cdc25A;* germ cell adhesion to Sertoli cells

such as *Tpx-1*, cytoskeleton assembly such as *Cappβ1*, protein degradation such as proteosome α7/C8 subunit, *Pα7/c8*, and microtubule-associated proteins such as *Tpx2*.
The meiotic prophase I further progresses into leptotene, zygotene, pachytene and diplotene spermatocytes. In leptotene spermatocytes, chromatin condensation is initiated and elements necessary for synapsis are recruited such as axial elements of the synaptonemal complex SYP3 [38]. Initiation of synapsis and meiotic recombination occur in zygotene spermatocytes.

During pachynema, we have full chromosomes synapsis and meiotic recombination. Interestingly, X and Y chromosomes lack homology and remain unsynapsed, thus forming the XY-body. Several proteins namely BRCA1, γ-H2AX and ATR silence the sex chromosomes by a mechanism called meiotic sex chromosome inactivation (MSI) [3]. Briefly, it was proposed that BRCA1 is involved in recruiting ATR and phosphorylate the histone H2AX, and that it is this phosphorylation that triggers chromatin condensation and transcriptional repression [39]. Finally, in diplotene spermatocytes the synaptonemal complex is dissociated. Recently, it was proposed that several checkpoints and surveillance mechanism exist, to eliminate chromosomally aberrant spermatocytes or silence unsynapsed chromosomes, in cases of improper synapsis or lower expression of cell cycle regulating genes [40]. During subsequent meiotic divisions a spindle assembly checkpoint (SAC) monitors the chromosomal alignment. After the second meiotic division, haploid round spermatids are formed.

1.1.1.5. Post-meoitic phase/ spermiogenesis

When meiosis is completed, round spermatids undergo complex morphogenesis processes through which they differentiate into elongated spermatids and eventually spermatozoa. Several cytological and morphological changes take place namely: acrosome and manchette biogenesis, expression of late spermatid genes, flagella formation, acquisition of signaling pathways related to motility and energy production, chromatin remodeling, cytoplasmic removal, and formation of the spermatid-Sertoli cell junction.

1.1.1.5.1. Acrosome biogenesis

It involves consecutive phases: the Golgi phase, also called the granule phase, where many pro-acrosomic granules are formed in trans-Golgi stacks followed by a fusion into a single, large acrosomic granule that associates on the top of the nuclear envelope by an F-actin structure the acroplaxome [41]. It starts to develop in round spermatids. Several genes have been involved in vesicle-to-vesicle fusion, transport and sorting of the acrosomal granules, namely *Gopc, Agfg1* (previously called *Hrb*), *CSNK2A2* [41-42] and *SPATA16*

[43]. Add to that, acrosome condensation following acrosomal vesicle formation, particularly *ZPBP1* is required for sperm acrosome formation, compaction and sperm:oocyte binding [44].

1.1.1.5.2. Manchette biogenesis

Manchette is a transient microtubular structure that plays a role in shaping the nucleus and development of the sperm head. HOOK1 protein encoded by this gene implicated in the interaction between the Golgi apparatus to microtubules, connects the manchette to the nucleus and stabilizes this structure [45-46]. RIM-BP3, a manchette associated protein, interacts with HOOK1 and may be involved in linking manchette-bound HOOK1 to the nucleus [47]. The intra-manchette transport (IMT) process is involved in sperm head development, which encompasses several molecular motors (e.g. Kinesin II, dynein, myosin Va) associating with cargoes that are walking along the cystoskeletal elements, just like cytoskeletal rail-tracks for intracellular transport. IMT consists of both microtubule based and F-actin based transport [48], particularly LRGUK-1 is implicated in manchette formation and movement by binding to HOOK2 [49].

1.1.1.5.3. Axoneme development

Centrioles are microtubule-based cylindrical organelles that nucleate flagellum in round spermatids. It is composed of a microtubular structure, the flagellar axoneme that drives sperm motility. Tektins are cytoskeletal proteins that are associated with the axonemal outer doublet microtubules which are thought to play a fundamental role in ciliary movement [50]. In particular, tektin B1or tektin 2 [51], tektin 3[52], and tektin 1 [53] are localized in flagella of mature sperm in humans. In addition, SPAG6 is a protein associated with the central axonemal apparatus [54].

Ccdc181 is a linker protein between microtubules of the axoneme [55]. In order to maintain sperm motility, several proteins are localized on the sperm tail and are found involved in metabolite transport and signal transduction such as VDAC-1 and VDAC-2 (Voltage-dependent anion channels (VDAC), also known as mitochondrial porins) [56]. Recently VDAC3 has been detected in the sperm outer dense fiber, a structure surrounding the microtubular structure of the sperm tail [57]. Hence, it was speculated that VDACs might have an alternative structural organization and different functions in outer dense fiber (ODF) than in mitochondrial bioenergetics. AKAPs are scaffold proteins for kinases and phosphatases which can are involved in sperm motility signal transduction. AKAP3 and AKAP4 are abundantly expressed in the fibrous sheath of developing spermatids which encases the axoneme in the principle piece of the sperm. Intraflagellar transport (IFT) is

involved in sperm tail development using molecular motors associating with cargoes [48]. It was proposed that LRGUK-1 plays a role in the initiation of the axoneme growth to form the core of the sperm tail [49].

1.1.1.5.4. Cytoplasmic removal

Excess of cytoplasm is removed to ensure the development of a compact structure as well as to remove ribosomal RNAs to prevent spurious protein synthesis. It is noteworthy to mention that expulsion through the cytoplasmic droplet is not complete, and thus remnants of rRNA are cleaved to ensure complete translational cessation in sperm at fertilization. Accordingly, it was proposed that this mechanism may confer protection of some paternal RNAs required for early development [58]. Most of the excess of cytoplasm from spermatids is removed by the tubulobulbar complexes (TBCs) which are actin-filament-related structures that form at adhesion junctions between Sertoli cells and the maturing spermatids in the seminiferous epithelium [59]. In fact, CAPZA3 protein which interacts with F-actin in the TBC, plays a role in cytoplasmic removal and acrosomal biogenesis [60-61]. In addition, SPEM1 protein is implicated in residual cytoplasm removal from spermatids [62]. It was proposed that SPEM1may serve as a scaffold protein that can introduce ubiquitinated unwanted/defective proteins in the excess of cytoplasm to UBQLN1, which then delivers these proteins to the proteasome for degradation present in the manchette [63].

1.1.1.5.5. Chromatin remodeling

Chromatin remodeling is a major event that occurs during spermiogenesis leading to the formation of a mature spermatozoa with a compacted nucleus. Briefly, somatic histones are replaced by transition proteins (TP1 and TP2) which are subsequently replaced by protamines (PRM1 and PRM2) [64].. Several factors are implicated in the correct incorporation of protamines into sperm chromatin such asthe post-translational modifications of protamines.For example, PRM2 requires phosphorylation of its mature protein by a Ca^{2+}/calmodulin-dependent protein kinase IV (Camk4) before it can be incorporated [65].

1.1.1.5.6. The expression of late spermatid genes

During late spermiogenesis several genes need to be expressed. This switch for expression of these genes is achieved through transcriptional regulators. Spermiogenes is governed by multiple independent mechanisms. The main modes for gene expression regulation target the two transcription factors, CREM (cAMP-responsive element modulator) and TBPL1 (TATA-binding protein like-1)[41].$CREM_T$ modulates the transcription of many important postmeiotic genes, such as *Prm1*, *Prm2*, *Tnp1* and *Tnp2* [37, 64-65].

TATA box binding protein-like 1 (TBPL1) protein is involved in transcription initiation, and many late spermiogenic genes are governed by TBPL1 [37, 66]. Additionally, PAPOLB (also called TPAP) is a testis-specific, cytoplasmic polyadenylate (polyA) polymerase which modulates specific transcription factors at posttranscriptional and posttranslational levels by extending the poly(A) tail of specific transcription factors mRNAs in round spermatids and, hence functioning in switching late spermiogenic genes [41, 66]. A recent study reported that MIWI proteins interact with piRNA and hence post-transcriptionally regulating gene expression of late spermatids [67].

1.1.1.5.7. Expression of cellular elements for energy production and motility

During spermiogenesis numerous genes essential for energy fuelling of sperm motility from glycolysis such as GAPDH2 glyceraldehyde 3-phosphate dehydrogenase 2 [68] are expressed. In addition many genes coding for signaling enzymes pivotal for sperm motility are expressed, namely *Catsper* calcium channels [69]. The aforementioned findings are summarized *in* ***Figure 2***.

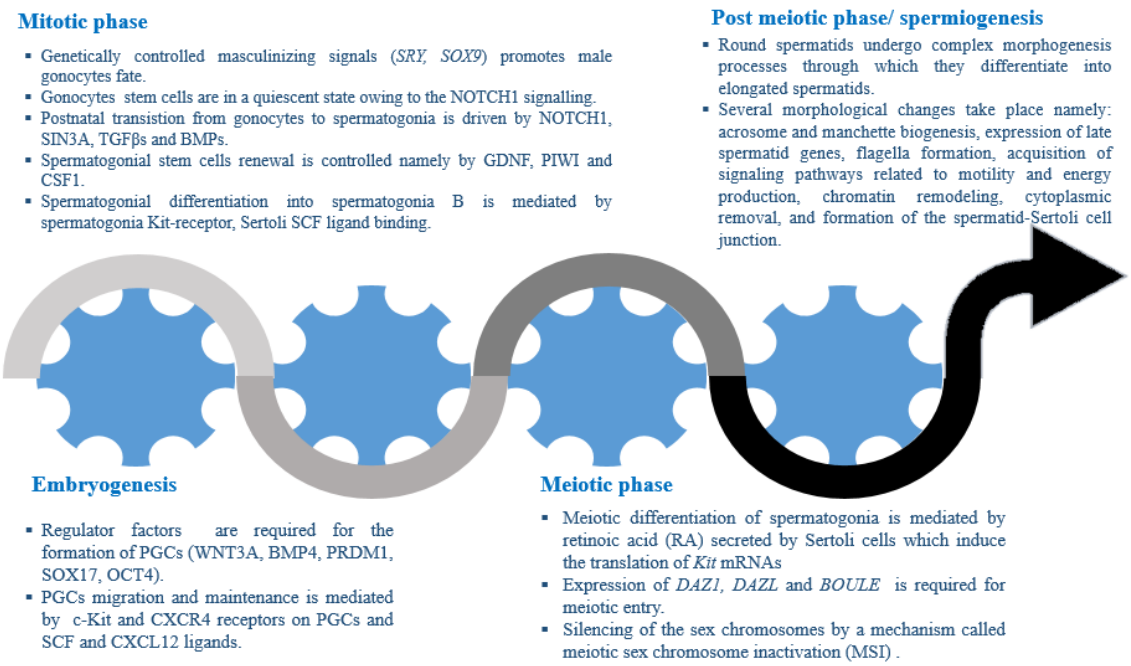


Figure 2. Stages and molecular events associated with human spermatogenesis. Sequence of the stages spanning spermatogonial stem cells differentiation during embryogenesis and the release of mature spermatozoa.

1.1.1.6. Spermiation

The apical ectoplasmic specialization (aES) is implicated in spermiation, and is composed of various cell adhesion molecules and confers cell adhesion between Sertoli cells and spermatids. It is dissembled to allow the release of fully developed elongated spermatids

(i.e., spermatozoa) into the tubule lumen [41]. It was proposed that *c-Yes* regulates the phosphorylation status of the adhesion proteins which mediates their endocytosis, thereby destabilizing this junction at spermiation. This key regulator (c-Yes) has a specific spatial and temporal expression pattern during the epithelial cycle of spermatogenesis to mediate apical ES and blood testis barrier (BTB) dynamics [70]. It is noteworthy that c-*Yes* is present in human spermatozoa [71].

1.1.1.7. Gene expression pattern during spermatogenesis

Transcription is inactive and blocked during 2 stages of spermatogenesis: during meiotic recombination in lepto/zygotene spermatocytes and during the late steps of spermiogenesis in elongated spermatids owing to their tightly compacted DNA structure [72]. During this transcriptional repression the transcriptionally competent pachytene spermatocytes and round spermatids mRNAs are stored before being translated, through their association with RNA-binding proteins (RBPs) to form ribonucleoprotein particles (RNP) complexes [73]. Indeed, spermiogenesis-relevant mRNAs required for post-meiotic spermatid differentiation are synthesized in round spermatids, and they are subsequently translationally repressed. During spermiogenesis, translationally repressed mRNAs relevant to spermatid differentiation are gradually released and synthesized [74].

Many RNA-binding proteins are required during spermatogenesis to mediate post-transcriptional regulation at different cellular levels such as splicing, polyAdenylation, nuclear export, translational regulation (activation or repression) by targeting the 3'UTR of target mRNAs or the ribosome [72]. In human germ cells, PUM2/NANOS2 [75], DAZL, DAZ, BOULE [76], SAM68 [77-78], HIWI [79], RBMY [80] are different RBPs involved in post-transcriptional regulation.

1.2 Molecular and epigenetic causes of infertility

1.2.1. Molecular insights into the causes of male infertility

Infertility is a multifactorial and genetically complex disorder [81]. Several factors influence the risk for the development of male infertility and the course of this alteration. These factors may be genetic, lifestyle (e.g. aspects of nutrition, BMI), environment (e.g. occupational exposure), clinical (e.g. varicocele), and medications [82]. Clinically, the sperm pathologies spectrum ranges between different phenotypes including oligozoospermia, asthenozoospermia and teratozoospermia depending on the sperm conventional parameters abnormalities [83]

Oligozoospermia is one of the most common types of male infertility. Generally, it is associated with a markedly reduced sperm count. A value of 15 million/ml was determined to be the cut-off between fertile and oligozoospermia [83]. It is very heterogeneous, thus can be stratified and varies between mild/moderate to severe oligozoospermia. Several factors have been postulated as underlying cause of this pathology. Oligozoospermia represents a typical multicausal multifactorial “disease” with an intense crosstalk of the genetic background with the environment, including lifestyle habits, environment and diet. Certain genetic disorders can cause oligozoospermia such as Klinefelter's Syndrome, Kallmann's syndrome, and Testicular Malposition (cryptorchidism) [84]. Notably, oligozoospermia has also been associated with a number of genetic risk factors [85], namely specific microdeletions in the chromosome Y and polymorphisms [86]. Along this perspective, an elevated risk for chromosome abnormalities in their sperm was noticed, particularly sex chromosome abnormalities [87]. Likewise, many lines of evidence have suggested that oligozoospermia is related to various risk factors such as varicocele [88], obesity [89], oxidative stress [90-91], alcohol [92] and tobacco [93].

However, the vast majority of studies have been concerned with the cytogenetic analysis and Y chromosome microdeletions screening of sperm chiefly from severe oligozoospermic men in different ethnic groups. These studies highlight the genetic homogeneity of severe oligozoospermia susceptibility among different ethnic populations [94-97]. To further characterize oligozoospermia, some spermatozoa transcripts were evaluated in infertile subjects using Real-Time Polymerase Chain Reaction (RT-PCR) approach. The ubiquitin-conjugating enzyme 2B (UBE2B) RNA was shown altered in severe oligozoospermia [98]. Similarly, the expression of some heat shock proteins and heat shock factors [88], VASA gene [99] and apoptosis-related genes [100] in oligozoospermic infertile sperm was also deregulated.

Over the past decade, RNA microarrays have been invaluable tools for the deconvolution of complex biological mechanisms, resulting in a new understanding of male infertility pathogenesis and providing a foundation for the generation of new biomarkers for diagnosis, prognosis and the prediction of therapeutic responses. Few microarray-based studies have contributed significantly to our understanding of the development of oligozoospermia. In this manner, several dysregulated genes involved in oligozoospermia at the mRNA expression level were identified [101-103]. In regards to asthenozoospermia, this condition arises from damage to sperm motility [83]. A value of <40% of sperm progressive motility was

determined by the World Health Organisation (WHO) in 2010 to be the cut-off between fertile and asthenozoospermic individuals [83].

In the vast majority of cases, the cause of disease is not completely known, namely considered idiopathic. Despite a strong genetic component owing to various flagellar pathologies [104], mitochondrial and nuclear genomic alterations [105], the identification of genes important in "disease" susceptibility has progressed slowly, resulting in a limited understanding of the molecular basis of asthenozoospermia. Numerous environmental risk factors including, oxidative stress [105], lifestyle habits [106], food [107], and heavy metals [108] may contribute to asthenozoospermia. This slow progress can be attributed, in part, to the involvement of multiple of causes. In spite of this scanty information several attempts have been made to treat asthenozoospermia [109-111], treatment strategies has barely been improved.

Currently, several global approaches including array-based expression profiling based on RNA microarrays [112-114], proteomics based on 2D electrophoresis (2DG) and/or mass spectrometry (MS) [115-118], DNA methylation arrays [116, 119] have been used to detect the changes at different molecular levels (DNA, RNA, or protein) in asthenozoospermic men. In particular, while proteomics-based approaches the most commonly used method for studies on asthenozoospermia, microarray-based gene profiling has been rarely used. Nevertheless, it was demonstrated that asthenozoospermia bears a specific gene-expression signature [112, 114]. However, these studies have focused on comparative aspects of gene expression of specific spermatozoa transcripts by RT-PCR namely related to sperm function and structure [120-127]. As for teratozoospermia, it refers to abnormal sperm morphology. The WHO established a cut-off value of sperm normal forms of 4%, below this value individuals can be labeled teratozoospermic infertile [83]. Alterations may vary widely across the sperm structure affecting the head, mid-piece, and/or the tail [83]. Teratozoospermia, once regarded as multifactorial [128], can have both genetic and acquired causes, with interaction of these factors in many cases. Familial aggregations were early indicators of a genetic component in teratozoospermia [129]. Several mutations were pointed out to be associated with distinct sperm pathological phenotypes, such as globozoospermia, large headed spermatozoa, etc [129]. Accordingly, it may also involve a strong genetic susceptibility associated with unfavorable environmental conditions. The genetic component itself is complex and heterogeneous and thus requires a sensitive wide approach to recapitulate the genetic milieu of teratozoospermia.

In this manner, genetic expression array analysis allows the characterization of sperm by assessing the expression patterns of thousands of genes. Such investigations of altered sperm have been used to elucidate which molecular pathways are altered. Very few array-based studies about the sperm transcriptional profile in teratozoospermic men have been devised [130-131], and they have pointed out that the ubiquitin-proteasome pathway is disrupted in teratozoospermia. In summary, these physiological and molecular studies give insight into the clinical conditions that can serve as resources for further investigations. Moreover, a large fraction of the sperm is still functionally uncharacterized. As described in the outlines of this thesis, we sequenced the transcriptome of sperm from infertile men with abnormal sperm parameters using next generation sequencing to get further insights into these conditions.

1.2.2. Epigenetic layers and players in the context of male infertility: an extensive review of the general theory and the current state-of-the-art

1.2.2.1. Introduction

It is thought that several environmental and intrinsic factors could lead to epigenetic variations over time, the so-called “epigenetic drift”[132].

Accordingly, increasing evidence suggests that complex interactions between genes and environment may play a major role in infertility. Hence, the rules of how epigenetic modifications impact gene expression in the context of male infertility have yet to be fully understood. Here, we review the current understanding about the individual layers of epigenetic information as well as the epigenetic regulators that will control gene expression in spermatozoa.

We define “layers” as structural components of DNA and Chromatin; namely DNA methylation, histone modifications, chromatin loops, chromosomal organization and location within the nucleus, and noncoding RNAs (bound to DNA). Conversely, “players” or epigenetic regulators are active complexes and enzymes required for Epigenetic layers. These encompass a large group of proteins including chromatin-binding factors, chromatin remodeling complex proteins, non-coding RNAs, DNA methyltransferases, and histone modifying enzymes[133].

Therefore, we will examine the current state-of-the-art in epigenetics-andrology to decipher the underlying epigenetic events leading to aberrant spermatogenesis and infertility as well as long-lasting effects on development. More interestingly, epigenetic modifications and alterations have been yet used as therapy targets for many diseases[134]. Thus, exploring a role for epigenetics in male infertility, could allow new insights to clarify the interplay

between environment and gene regulation, and identify new emerging therapeutic targets. To better understand the molecular etiology of male infertility, we will rely primarily on data from human studies, and occasionally from animal studies. This will provide substantial evidence that might piece together the complete epigenome puzzle in spermatozoa.

1.2.2.2. The sperm methylome

1.2.2.2.1. Basic concepts

DNA methylation consists of the addition of a methyl group to the 5th carbon of cytosine (5meC). DNA methylation plays a crucial role in gene expression regulation. Indeed,_methylation changes the interactions between proteins and DNA, which leads to alterations in chromatin structure and either a decrease or an increase in the rate of transcription [135].

The study of DNA methylation has been stimulated by the identification of the key enzymes that methylate DNA and their interactions with DNA and DNA binding proteins. Thus,_DNA methylation is very complex process involving several groups of enzymes, namely methylenetetrahydrofolate reductase (MTHFR), *de novo* methyltransferases (DNMT), and ten-eleven-translocation (TET) proteins

Briefly, the methyl pool is brought by the activity of the Methylenetetrahydrofolate reductase (MTHFR). MTHFR is a key regulatory enzyme involved remethylation reactions. MTHFR catalyses the reduction of 5,10-methylenetetrahydrofolate to 5-methyltetrahydrofolate, which is the methyl donor for remethylation of homocysteine to methionine. Methionine is in turn converted to S-adenosylmethionine, a universal methyl donor for any acceptor including DNA[136]. From a mechanistic point of view, DNA methylation is laid down by *de novo* methyltransferases (DNMT 3a, 3b). Throughout DNA replication, hemi-methylated DNA is formed. To maintain DNA methylation, hemi-methylated DNA is substrate for DNMT1[137]. Therefore, DNA methylation is considered a stable, and mitotically heritable epigenetic mark.

Methylation exclusively occurs at CpG dinucleotides usually unmethylated, although non-CpG methylation is also observed in mammalian, to note in gametes[138]. Mostly at gene promoters, CpGs sometimes are clustered into CpG islands. Particularly, their methylation lead to silencing of gene expression[139]. Methylated Cytosine at CpG island are associated with the formation of a repressive chromatin structure; by recruiting methylated CpG binding proteins MeCP 1&2 [140] . The latter may alter transcription or condense chromatin. Occasionally, methylated CpGs can also prohibit transcription factor binding[139].

Furthermore, DNA methylation may occur at intergenic regions and repetitive elements -usually methylated - in order to maintain genomic integrity[141]. However, DNA demethylation can occur in a passive manner throughout cell replication by dilution; if DNMT1 is not expressed; or an active demethylation by enzymatic removal of methyl group via intermediates (5-hydroxymethylcytosine (5hmC)) using Ten-eleven-translocation (TET) [142] and AID proteins [143]. Interestingly, some genes known as imprinted genes are differentially methylated allowing for mono-allelic parent-specific gene expression. Correctly imprinted maternal and paternal genes are needed for embryo viability. Specifically, during germ cells development a “wave” of reprogramming involves widespread erasure of somatic-like patterns of DNA methylation followed by establishment of sex-specific patterns by de novo DNA methylation. This is marked by a demethylation followed by a remethylation of the genome and at imprinted genes[144].

DNA methylation has been implicated in a number of distinct cellular processes during development and differentiation, in particular spermatogenesis.The canonical view of sperm DNA methylation as an epigenetic mark was interrogated at CpG island promoters of genes, repetitive elements and imprinting control regions. Interestingly, the sperm methylome patterns, at these definite fingerprints of the genome, were shown perturbed in cases of aberrant spermatogenesis. It was suggested that a broad elevation of DNA methylation in abnormal spermatozoa may be due to improper erasure of DNA methylation in primordial germ cells [145].

1.2.2.2.2. MTHFR enzyme activity

Recently, MTHFR activity has attracted the interest and research attention in epigenetic regulation of gene expression. Hence, any alteration in the *MTHFR* gene may lead to an imbalance in the pool of methyl groups and therefore affects the DNA methylation status.

Numerous studies have addressed this hypothesis, and interrogated either the polymorphic variants of *MTHFR* or the status of its promoter in sperm from infertile men. First off, several *MTHFR* polymorphic variants were found associated with male infertility susceptibility. Along these lines, many studies and systematic review evoked a role of genetic mutations in folate-related enzymes genes, namely *MTHFR* with the risk of male infertility [146-147]. This link was documented in cases of azoospermia [148-149], oligozoospermia [149] and idiopathic normozoospermia[150]. Taking into consideration any potential ethnic variation, these studies explored the genetic polymorphisms of *MTHFR* in different populations [418-421].

Beyond that, it has also been suggested that male infertility might be linked with aberrant promoter methylation of *MTHFR*, namely hypermethylation which is considered as a form of epigenetic silencing. A low methyl donor pool in spermatogonial cells, owed to *MTHFR* gene dysfunction, drives methylation errors at the paternal imprinted genes, such as *H19* imprinted gene [151] which in turns has drawbacks on the development. In view of that, *MTHFR* gene hypermethylation correlated with recurrent spontaneous abortion[152]. The tendency towards *MTHFR* promoter hypermethylation in infertile patients was accentuated in spermatozoa with low motility [153-154]and poor morphology[154], and to a lesser extent in cases of oligozoospermia [155]and idiopathic normozoospermia[156-157].

1.2.2.2.3. Role of the DNMTs in spermatogenesis

At the level of the DNA methylation machinery, a study using a mouse model reported the role of DNMT3L in male germ cell development. *Dnmt3L* mutant mice presented defects in spermatogenesis and hypomethylation of several regions in the genome[158]. In the same year, Dhillon et al. interrogated DNMT3b polymorphism in Indian men with oligoasthenoteratozoospermia. No association was found between the frequency of DNMT3b polymorphism and the risk of infertility [159]. More so, there is an altered expression of DNMT3A in low motility sperm. As for the methylation status, 1CpG loci associated with *DNMT3A* and *DNMT3B*was shown altered in low motility sperm samples [153]. Furthermore, genetic analysis in oligozoospermic patients permitted to identify an amino acid change in DNMT3A in one case and in DNMT3L in eight men with altered methylation profiles out of 24[160].

1.2.2.2.4. Tet-dependent DNA demethylation

The 3 TET enzymes (TET1, TET2, TET3) were successfully expressed in human spermatozoa [161]. However, the role of TETs in fertility issues was questioned by two studies. The first study reported a strong association between TET2- and TET3-mRNA in sperm cells and the age of men, motility of sperm, and positive Intracytoplasmic sperm Injection/Invitro Fertilization (ICSI/IVF) outcomes[162]. Similarly, Ni et al. 2016 [161] stated that, TET1- and TET3-mRNA levels in sperm were significantly correlated with age and progressive sperm motility, while TET2- mRNA levels were associated with positive-ICSI outcomes. Interestingly, TET1-3 showed a significant association with sperm concentration. A significant reduction of TET1-3 in sperm was reported in cases of oligozoospermia and/or asthenozoospermia when compared to fertile individuals.

1.2.2.2.5. Global DNA methylation

Firstly, some studies took a global approach to determine DNA methylation alterations in infertile patients. To assess alterations in DNA methylation density in global DNA a simple immunoassay method, targeting 5-methylcytosine residues, can be used. Along this perspective, it was shown that oxidative damage to sperm DNA, a contributing factor to male infertility, is in part responsible for modifying sperm global DNA hypomethylation [432-434]. Contrastingly, others did not found any significant large-scale alterations across the whole genome in sperm between infertile men with reduced semen parameters[163], normozoospermic infertile men[164], asthenozoospermic and fertile men[119].

It could be speculated that oxidative DNA adducts (8-OHdG) incorporation in the methyl-CpG binding protein (MBP) recognition sequence, result in the inhibition of MBP binding, impeding the process of DNA methylation. Besides that, it appears that sperm morphological variations assessed in the same sample; as recognized at high magnification and scored based on head shape, the presence of a vacuole in the nucleus, and the head base; are associated with differences in DNA methylation levels[165]. Thus, a poor sperm head presents a high level of chromatin decondensation[166], and is associated with hypermethylated sperm nuclear DNA[165]. In the same manner, defective apoptotic spermatozoa may be linked to global hypermethylation of its nuclear DNA[167]. Anyway, whether defective spermatozoa of infertile men exhibit relatively hyper- or hypo-DNA methylation remains an ongoing debate.

1.2.2.2.6. Methylation levels at promoter of CpG islands

Secondly, methylome was examined in spermatozoa of infertile men interrogating sequences in promoters of CpG islands, repetitive elements and imprinted genes. From a global perspective, several approaches identified the total number differentially methylated CpG (DM-CpGs) in sperm DNA between fertile and infertile subjects. Using an array-based method the Illumina Infinium 450K array, 9189 DM-CpG were recognized in low motility samples[153], 2752 DM-CPGs in normozoospermic infertile men using the aforementioned approach[164], more so, 696 DM-CpGs in individuals with abnormal sperm parameters[168]. Whereas, using the liquid hybridization capture-based bisulfite sequencing method, 134 associated with asthenozoospermia were identified[119].

1.2.2.2.7. Methylation status of non-imprinted genes

At first, we will present evidence for the methylation status in non-imprinted genes associated with impaired spermatogenesis. The promoter of *GSTM1*, a gene which plays a role in protecting cells from oxidative stress[169], was aberrantly hypermethylated at CpG

islands in infertile men diagnosed with oligoasthenozoospermia[159]. Higher rate of methylation of *CREM* (cyclic AMP-responsive element modulator) promoter- a key transcription factor that targets several important genes needed for normal spermatogenesis, including the protamine genes [170]- in patients with abnormal protamination and oligozoospermia[171].

The methylation pattern of two germline regulator genes, DAZ (deleted in azoospermia) and DAZL (deleted in azoospermia-like),were characterized in oligoasthenoteratozoospermic (OAT) patients and controls. The promoter CpG islands of DAZL displayed a significantly increased methylation defects in the OAT group compared with controls, but DAZ remains unmethylated [172]. Likewise, in the same year Wu et al. suggested that promoter methylation of DAZ is unlikely involved in the etiology of spermatogenic failure[155].

More so, sperm abnormalities may be associated with a broad epigenetic defects of elevated DNA methylation at numerous sequences of non-imprinted genes (e.g. *HRAS, NTF3, PAX8, SFN*)[145]. In contrast, DNA hypomethylation was shown largely maintained at developmental genes promoters (*CDX4, GZF1, HOTAIR, EVX2,* HOXC10) of a small group of infertile men[173].

Furthermore, several genes associated with spermatogenesis and epigenetic regulation (e.g. *HDAC1, SIRT3, DNMT3A*) presented aberrant DNA methylation, majorly hypomethylation, in low motility sperm, parallel transcriptomic analysis interrelated analteredexpressionof some genes with an altered methylome[153]. Additionally, aberrant DNA methylation at the spermatogenesis-specific gene *VASA* was found associated with oligozoospermia[174]. In a recent study, analysis of methylation at single CpGs and genomic regions revealed that alterations namely hypermethylation in 12 CpGs were associated with samples with poor sperm motility and alterations at 898 CpGs associated with sperm viability[163].

In the same year, Du et al. 2015 [119]employing promoter targeted bisulfite sequencing, analyzed the DNA methylation profiles associated with low sperm motility in asthenozoospermia. This work identified 134 differentially methylated CpGs, 41 differentially methylated regions and 134 differentially variable CpGs, harboring genes involved in spermatogenesis or sperm motility or expressed specifically in testis (e.g.*CDK16, SPAG1, TCTE3, SPATS2…*). It is pertinent to mention that more hypomethylated CpGs and regions were localized in the promoters, while more hypermethylated were in the gene body in sperm cells from asthenozoospermic patients[119].

In a very recent study by Camprubí et al. in 2016 [168], a set of differentially methylated genes related to spermatogenesis was determined. It is noteworthy to mention that, a strong association was identified between *RPS6KA2* hypermethylation and advanced age, *APCS* hypermethylation and oligozoospermia, *JAM3/NCAPD3* hypermethylation and numerical chromosome sperm anomalies, and *ANK2* hypermethylation and lower pregnancy.

1.2.2.2.8. Methylation at repetitive elements

Methylation status of repetitive elements (i.e. Alu and LINE-1) is a major contributor of global DNA methylation patterns and has been addressed in relation to male infertility. At the LINE-1 sequences, no differences in the methylation status was found between normozoospermic infertile men[164], oligozoospermic patients and fertile controls [175-176] and fertile controls. However, it has been found to be hypomethylated in infertile patients with severe spermatogenic disorders[29]. For the Alu sequences, they showed significantly lower methylation levels in normozoospermic infertile subjects, namely for Alu Yb8 [164], and in men with abnormal semen parameters[177]. In contrast, Kobayshi et al. 2007 [175] did not found any difference in the methylation condition of Alu between oligozoospermic subjects and controls.More so, at the NBL2 and D4Z4 sequences, no differences in the methylation status were noticed between normozoospermic infertile males and fertile[164].

Promoter targeted bisulfite sequencing in asthenozoospemic individuals revealed that the percentage of differentially methylated CpGs either hypomethylated or hypermethylated in the repetitive elements are~40%. Additionally, there are more differentially variable CpGs associated with asthenozoospermia, which could suggest a state of genomic instability[119].

1.2.2.2.9. Methylation at imprinted genes

Imprinting defects were found associated with disturbed spermatogenesis. Interestingly, a very recent study by Kobayashi et al. 2017 [178], attempted to identify factors associated with aberrant imprint methylation and oligozoospermia. It was found that defective methylation at imprinted loci and a particular lifestyle (current smoking, excess consumption of carbonated drinks) were associated with severe oligozoospermia, while sperm with abnormal imprint methylation causes a decrease in the live birth rate and an increase in the miscarriage rate.

As for the methylation status of the maternally methylated imprinted genes, *MEST* gene is the well-studied example. *MEST* methylation levels remained indifferent between normal and oligozoospermic patients[179]. Aberrations of the maternally imprinted gene *MEST/PEG1* DNA methylation patterns- commonly hypermethylation- are associated with oligozoospermia[290, 414, 446-447, 450-452], reduced progressive sperm motility [290, 322,

450] and abnormal sperm morphology [163, 180]. Furthermore, *MEST* DNA methylation was significantly associated with clinical parameters such as oligozoospermia, decreased bi-testicular volume and increased FSH levels, thus making it a potential marker in andrology [174] However, El Hajj et al. 2011[177] did not find this strong association between imprinting error and abnormnal sperm parameters. On the other hand, epimutations in PEG1/MEST-Differentially Methylated Regions were found in 3% of oligozoospermic men[160].

Other maternally imprinted genes showed aberrant methylation pattern: in oligozoospermic patients at*LIT1, ZAC, PEG3* and *SNRPN*loci [175],at *PEG3*[178], or at *PLAG1*[145], or at *PEG1, LIT1, ZAC, PEG3,* and *SNRPN* regions[181], at *APAB1, SNRPN* associated with low motility[153], at *SNRPN* in infertile males with abnormal semen parameters [177]or in patients with reduced sperm motility and morphology[154],at *SNRPN, KCNQ1, LIT1* in abnormal protamine patients[182], and at *GNAS* associated with low sperm count, motility and vitality[163].

Besides that, defective imprinting in spermatozoa of infertile men may be owed to aberrant DNA methylation of paternally imprinted genes. *H19* imprinted gene methylation pattern is relatively well-documented in spermatozoa of infertile individuals.Methylation errors at *H19* loci – regularly hypomethylation- were shown associated with oligozoospermia[260, 446-447, 450-452], low sperm motility[153], and with abnormal semen parameters [177].

Other paternally imprinted genes presented DNA methylation alterations were also associated with hypospermatogenesis, to note: *GTL2*[175, 178], additionally *ZDBF2* and *GTL2* [181] were associated with oligozoospermia, *GTL2* linked to abnormal semen parameters [177], *PHPT1, ALDH1L1, LDB1,* and *PEX10* associated with low motility sperm[153], and *PEG10* in low sperm viability, motility and count [163]. It is noteworthy to mention that, it is clearly unveiled the existence of an association between epigenetic deregulation especially of imprinting and diseases[183]. In summary, the sperm DNA methylome has been studied extensively from a mechanistic point of view in the context of male infertility. All the above-mentioned body of evidence established the association between aberrant sperm DNA methylome and pathological spermatogenesis. The aforementioned findings are summarized in ***Table 1***.

Table 1. Male infertility-associated epigenetic modifications: DNA methylome alterations.

The table lists the molecular landscape of DNA methylation in sperm from infertile men

Epigenetic component	Findings
MTHFR enzyme activity	▪ Polymorphic variants of *MTHFR* ▪ Aberrant *MTHFR* promoter methylation (hypermethylation)
DNMT	▪ Polymorphism in *DNMT3b* ▪ Amino acid change in DNMT3A, DNMT3L ▪ Altered CpG methylation at *DNMT3A*, *DNMT3B*
Global DNA methylation	▪ Levels changes (+/-)
Methylation levels at promoters of CpG islands	▪ Differentially methylated CpGs
Methylation status of non-imprinted genes	▪ Aberrant methylation patterns of promoters in genes involved in spermatogenesis ▪ Aberrant DNA methylation in genes associated with epigenetic regulation ▪ Hypomethylation at developmental genes promoters
Methylation at imprinted genes	▪ Aberrations in the methylation pattern of some maternally imprinted genes (*MEST, PEG3, SNRPN…)* ▪ Epimutations in PEG1/MEST-Differentially methylated regions ▪ Methylation errors at some paternally imprinted genes (*H19, GTL2…*)
Methylation at repetitive elements	▪ Low methylation level at LINE-1 (+/-), Alu (+/-) sequences ▪ Differentially methylated CpGs in repetitive elements
Tet- dependent DNA demethylation	▪ Reduced Tet (1,2,3) transcripts levels

1.2.2.3. The spermatozoal chromatin proteome

1.2.2.3.1. Basic concepts

Sperm chromatin is highly specialized and is the end product of a highly complex differentiation program during which an impressive number of different testis-specific histone variants and protamine genes are expressed. The role of many of these sperm histone variants and histone-like proteins on gene expression has been recently investigated [184].

Basically, in order to control gene expression, either by silencing or activating, there is a considerable interplay among various factors. Numerous proteins that bind to DNA, recruited to specific sites, are involved in the epigenetic modification of the genome. The most abundant and best characterized factors are the histone core proteins and the chromatin remodeling proteins. The spermatozoal chromatin proteome is another important epigenetic layer because it influences the degree of compaction and accessibility of the DNA to transcription factors and the transcriptional machinery.

1.2.2.3.2. Histones: post-translational modifications and variants

Several post-translational modifications of nuclear histones occur on the N-terminal tails which protrude from the nucleosome. These chemical changes took place mainly on the

tails of H3 and H4. Therefore, acetylation -at the lysine (K) residues- and methylation -at the lysine and arginine (R) residues- are the well characterized. Interestingly, various combinations of histone modifications at a specific site are called “histone code”, may exert several effects on the genomic function.

Histones acetylation is laid down and up respectively by histone aceltyltransferases (HAT) and histone deacetylases (HDAC). Additionally, this specific modification is generally linked to gene activity , as in part it reduces the positive charge of histone and it acts as a docking site for other chromatin proteins (e.g.bromodomain proteins), to subsequently open the chromatin structure. It is noteworthy that, this modification could be considered a chromatin modification rather than an epigenetic modification due to the lack of a mechanism of mitotic heritability. Histone methylation is set down and up by specific lysine methyltransferases KMT and lysine demethylases KDM. This modification is associated with either transcriptional silencing or activity depending on the context (e.g. H3K9me, H3K27me correlated with gene silencing, H3K4me linked to transcriptional activity). Modified histone tails act as docking sites for other chromatin proteins that maybe chromatin remodellers or others[185].

Furthermore, different histones called “histone variants”-with different stabilities or domains –were reported, where some of which could affect gene expression [186-187].

The driving hypothesis is that aberrations of histone modifications may contribute to disease pathogenesis[188]. In this manner, we will attempt to check for possible implication of histones in male infertility. During spermatogenesis, while approximately 85% of histones are replaced by protamines, the remnants have been proposed to carry essential epigenetic marks transmitted to the offspring [189].

Genome-wide total histone retention was noticed in the sperm of infertile men with abnormal semen analysis or altered protamination[173]. Post-translational modifications of histones, namely acetylation and methylation, were found relevant in spermatogenesis [460-461]. Nuclear histone acetylation protein level in sperm, particularly H3K9ac was found unchanged between oligozoospermic men and fertile individuals [190] , a reduced protein expression for H4hac was detected in spermatids of impaired spermatogenesis[191]. Looking more deeply, a change in the H3K9ac binding pattern in ejaculated spermatozoa of oligozoospermic men was noticed, mainly in some genes associated with embryo development (e.g.*G6PD, SOX2, PLXNA3*)[190]).

Moreover, a lack of H4K12ac binding within some developmentally important promoters (*AFF4, EP300, LRP5, RUVBL1, USP9X, NCOA6, NSD1,* and *POU2F1*) was

detected in sperm from normozoospermic infertile men, while DNA methylation levels remain unchanged. This may explain the impaired chromatin condensation in sperm from subfertile individuals[192].It was suggested that histone hyperacetylation may be indispensable for correct histone-to-protamine exchange. In the mouse model, post-meiotic packaging of the male genome from round spermatids to mature spermatozoa involves not only removal of histones and assembly of protamines but also pericentric heterochromatin re-programing. Pericentric heterochromatin is a distinct region that harbors somatic features by retaining histone marks namely histone acetylation [96, 464]. Thus, complete male genome reprogramming during spermiogenesis requires several waves of acetylation of different histones variants. From an andrological point of view, the levels of only six acetylation marks (H3K9ac, H3K14ac, H4K5ac, H4K8ac, H4K12ac, and H4K16ac) were found not related with conventional sperm parameters including sperm DNA fragmentation index (DFI) in normozoospermic men. As for the enzymatic machinery: while sperm DFI was positively correlated with HAT activity, strict morphology was negatively correlated with acetylation enzyme index (=HAT activity/HDAC activity)[193].

More so, in low motility sperm samples, several epigenetic regulatory genespresent an altered expression (*HDAC1*) and an aberrant DNA methylation (*HDAC2, BRDT*)[153].

That being said, acetylated histones are “read” by other chromatin proteins, notably BRDT (bromodomain testis-specific). Particularly, in cases of azoospermia: the *BRDT* gene was not expressed in testicular tissue from patients with Sertoli Cells Only, additionally BRDT protein expression was absent in testicular tissue specimens with spermatocyte maturation arrest. Accordingly, it could be speculated that BRDT is required in chromatin compaction during spermiogenesis and embryo development owing to its plausible epigenetic roles[194].

As for histone methylation, the localization patterns of H3K4me or H3K27me was generally normal in the gametes of infertile men presenting an abnormal semen analysis or an altered protamination[173].Conversely, there was a lack or reduction in the levels of H3K4me or H3K27me retained at developmental transcription factors and certain imprinted genes in sperm from infertile patients. Furthermore, it was shown that modified histones may regulate the transcription of several genes. We will emphasize this feature at the *BRDT* gene.

In this context, a reduction in histone methylation at the promoter of *BRDT* caused an increase in the transcript level in sperm of infertile men, with teratozoospermia and asthenozoospermia associated with a defective chromatin condensation[195]. It has been suggested that sperm from asthenozoospermic men possesses a significant high proportion of

histone H2B and their variants TSH2B andH2BFWT to protamine (PRM1+PRM2) compared to fertile controls [196]. A recent finding in the mouse model, revealed the contribution of a histone variant H3.3 to viability and fertility [197].

1.2.2.3.3. Chromatin remodeling

Numerous ATP-dependent multi-protein complexes are considered as epigenetic factors, since they can reposition the nucleosomes and contribute to chromatin remodeling. Chromatin remodelers recognize modified histones and particularly some can modify them[198]. Differential expression study in testicular tissues between normal spermatogenesis and round spermatid maturation arrest revealed a significant different expression of chromatin remodelling factors between the 2 groups. A total of 22 genes out of 84 (*BRDW3, CTCF, BAZ2A, BAZ1A, CHD3, BRPF1, PBRM1, EED, RING1, PCGF5, PHF7, EZH2, BMI1, NSD1, CHD5, PCGF6, PHF21A, PCGF1, MBD4, ING1, MTA1* and *PHF2*) were noticed, where the majority (21 genes) were down-regulated in the developmental arrest [199].

Moreover, the *CTCF* gene was shown hypomethylated in low motility sperm samples compared to controls[153]. It is noteworthy that, CTCF has various roles in transcriptional repression, insulator function, and imprinting genetic information[200]. Furthermore, alterations in the *Sly* gene, part of a chromatin remodeling complex, in sperm from mice resulted in infertility [472-473]. The aforementioned findings are summarized in ***Table 2***.

Table 2. Male infertility-associated epigenetic modifications: Aberrations in the spermatozoal chromatin proteome.

Epigenetic component	Findings
Histones: Post-translational modifications & Variants	▪ Global histone retention ▪ Reduced protein level of some acetylated histones ▪ Changes in acetylated histones binding pattern ▪ Decreased acetylation enzyme index (=HAT/HDAC) ▪ Reduced levels of some methylated histones at specific promoters of, *BRDT* (coding for a chromatin protein) and certain imprinted genes ▪ High proportion of H2B and their variants
Chromatin remodelers	▪ Differential expression (majorly down-regulation) ▪ Hypomethylated *CTCF*(codes for a chromatin protein, linked to epigenetic remodeling)

1.2.2.4. Non-coding RNAs

1.2.2.4.1. Basic concepts

Non-coding RNAs are functional RNA molecules not translated into proteins. This type of RNAs makes up the majority of the transcriptome, where the coding exons account for only 2% of the genome[201].They encompass many classes of small, mid-sized and long non-coding RNAs. Recent lines of evidence have shown that non-coding RNAs appear to be as epigenetic regulators of gene expression[202]. Many different forms of RNA that regulate transcription have been described, including microRNAs (miRNAs), Piwi-interacting RNAs (piRNAs), and long non-coding RNAs (lncRNAs).

Basically, miRNAs play a critical role in post-transcriptional gene silencing, which is mediated through Drosha and Dicer protein complexes. Of note, they are not epigenetic layersper se, but can regulate other components of the epigenetic machinery as well as be epigenetically regulated themselves[203]. Piwi-interacting RNAs (piRNAs) are involved in silencing transposable elements in order to maintain genomic stability. They can also silence non-repetitive regions such as some imprinted genes[204]. Briefly, transposons and repeats remnants are found in piRNA clusters. The later are transcribed in precursors of piRNA. While the primary processing involves shortening of these precursors, the secondary processing entails loading onto PIWI proteins and degradation of transposon RNA, accordingly this will lead to post-transcriptional silencing. Beyond that, piRNA can be bound to MIWI2, which imports the piRNA-PIWI protein complex back to the nucleus to execute DNA methylation of transposon/repetitive element DNA. This path permits transcriptional silencing via RNA directed DNA methylation[205].

As for long-non coding RNAs (lncRNAs), they are characterized by several features: length >200nt, spliced, capped polyadenylated and constrained in the nucleus. Interestingly, lncRNAs appear to direct epigenetic complexes. Various presumptive mechanisms of functioning could be suggested: as a guide by tethering epigenetic complexes to a particular region, or as a scaffold to bring together several complexes, as molecular decoys, or as signals to regulate gene expression[206].

The discovery of significant amounts of non-coding RNAs in sperm led to speculation regarding its possible role in various processes namely sperm function [207] and embryonic development [208]. Recent microarray-based expression studies undoubtedly have contributed to our understanding of the sperm non-coding transcriptome. Furthermore,the unprecedented advances in genomics, particularly next generation sequencing offered novel insights into the exploration of some small non-coding RNAs in the spermatozoa[209]. It was

shown that many non-coding RNAs might contribute to the development of sperm pathologies. However, we are only beginning to understand the nature and extent of the involvement of non-coding RNAs in spermatogenesis.

1.2.2.4.2. MicroRNAs (miRNAs) and spermatogenesis

Firstly, a specific set of miRNAs, 52 miRNAs,was shown differentially expressed between the semen of infertile males, majorly asthenozoospermic, and fertile[210]. In the same manner, several miRNAs, some of which are novel, were also found differentially expressed, between infertile men (normozoospermic, asthenozoospermic, oligoasthenozoospermic) and fertile individuals. In particular, some miRNAs were elected specific to a sperm pathology[211].

The aforementioned work of Abu-Halima was continued in the same framework of sperm microRNAs between infertile men, mostly oligozoospermic and oligoasthenozoospermic and fertile controls. A panel of five microRNAs was determined as potential biomarkers for the diagnosis and assessment of male infertility[212]. Additionally, a dysregulation of miRNAs was associated with hypospermatogenesis in testicular tissues from patients with non-obstructive azoospemia[213-214]. In the mouse model, while deletion of both miR34b/c and miR-449 loci resulted in oligoasthenoteratozoospermia[215], deletion in *Dicer1* caused meiotic and spermiogenic defects[216], demonstrating that the dysregulation of micro-RNAs has relevance to fertility.

1.2.2.4.3. Piwi-interacting RNAs (piRNAs) and spermatogenesis/infertility

Secondly, single-nucleotide polymorphisms of 2 Piwi genes (*PIWIL3/HIWI3* and *PIWIL4/HIWI2*) were correlated with an increased risk of oligozoospermia in a Chinese population[217]. Furthermore, altered DNA methylation patterns were observed in piRNA-processing genes (*PIWIL2, TDRD1*) from testicular tissues of patients presenting spermatogenic disorders. This entails promoter hypermethylation-associated silencing of PIWIL2 and TDRD1[29].Very recently, mutations in *Piwi* were detected in patients with azoospermia, and then modeled in transgenic mice which led to abnormalities in spermatozoa. It is worth pointing out that, this mutation prevents the elimination of PIWI by ubiquitination. Mechanistically, it was found that accumulated PIWI binds the histone ubiquitin ligase RNF8 in the cytoplasm of late spermatids, therefore this complex prevents histone ubiquitination. Accordingly, the resultant sperm will present retained histones in its DNA[218]. The aberrant spermatozoal piRNAs imply that the PIWI pathway has profound function in male germ cells.

1.2.2.4.4. Long non-codong RNAs (lncRNAs)

Thirdly, a recent microarray-based expression study unveiled that many lncRNAs, in the testes of infertile men with maturation arrest or hypospermatogenesis, showed altered expression[219]. Notably, *NLC1-C* (narcolepsy candidate-region 1 genes) was down-regulated in the cytoplasm. Surprisingly, it was accumulated in the nucleus of spermatogonia and primary spermatocytes from patients with maturation arrest. This nuclear retention of NLC1-C leads to a repression of miR-320a andmiR-383 transcripts, thus causing a hyperactive proliferationof spermatogonia and primary spermatocytes. Ultimately, this dysregulation will result in male infertility. This new model proposed suggests that lncRNAs can modulate miRNA expression at the transcriptional level to regulate human spermatogenesis. In the same perspective, HOTAIR – a lncRNA – was found downregulated in sperm of patients with asthenozoospermia or oligoasthenozoospermia. Consequently, this will reduce histone H4 acetylation in *Nrf2* promoter and thus *Nrf2* expression[220]. To mention that, the transcription factor of Nrf2 (nuclear factor erythroid 2-related factor 2) can promote antioxidant enzymes expression[221]. More so, new evidence revealed a novel spermatogonia-specific lncRNA critical for the survival of murine spermatogonial stem cells [222]. The aforementioned findings are summarized in ***Table 3***. The table lists the different classes of non-coding RNAs in sperm from infertile men.

Table 3. Male infertility-associated epigenetic modifications: Altered sperm non-coding RNAs.

Class of non-coding RNA	Findings
microRNAs (miRNAs)	▪ Differential expression
Piwi-interacting RNAs (piRNAs)	▪ Single nucleotide polymorphisms of 2 *PIWI* genes ▪ Altered promoter DNA methylation of some genes ▪ Mutations
Long non-coding RNAs (lncRNAs)	▪ Altered expression

1.2.2.5. High order nuclear organization in sperm

The last epigenetic layer addresses the high degree of nuclear architecture. Chromosomes are non-randomly arranged within the nucleus, they reside in specific defined area called chromosome territories [223]. However, the nucleus is made up of several nuclear subcompartments. The nuclear periphery formed by the nuclear lamina, to note that heterochromatin preferentially localizes to the nuclear periphery [224].The transcription

factories localize at the center of the nucleus, and consist of concentrated regions of RNA Pol II [225]. The nuclear pores are essential for exporting to the cytoplasm, and they tend to localize adjacent to euchromatic regions[226]. The nucleolus is composed of aggregation of ribosomal DNA, thus a site of rRNA transcription [227]. The peri-nucleolar space contains several RNA processing proteins and several RNA pol III transcripts [228].

In the context of spermatozoa, it appears that sperm nuclei, from a small sample of infertile males with abnormal sperm parameters, presented an altered sex chromosome centromere position[229]. Likewise, sperm nuclei with abnormal sperm associated with an increased level of aneuploidy lead an abnormal positioning of chromosome 15, 18, X and Y centromeres of sperm cells[230].Otherwise, changes in the preferential topology of the chromosome 7, 9, 15, 18, X and Y centromeres were observed in sperm nucleus from infertile patients with high DNA fragmentation[231]. While, another group did not find any important alteration of nuclear reorganization of centromeric loci in sperm heads in severely compromised sperm[232].

In a case study, a patient with a moderate oligoasthenoteratozoospermia and carrier of sSMC (small supernumerary marker chromosome which isstructurally an abnormal chromosome) presented alterations in the positioning of the human chromosomes X and Y, particularly in close proximity to the sSMC, which ultimately affected chromosome segregation[233].

Conclusion

Within this fairly extensive review, we have endeavored to describe as comprehensively as possible the major “key events” regarding spermatogenesis in humans. Growing evidence shows that epigenetic abnormalities participate to cause dysregulation in spermatogenesis and ultimately male infertility. The exciting links between epigenetics and male infertility are just becoming manifest. Since the causes of spermatogenic defects in infertile men are known multifactorial, the interplay between epigenetics and genetics may be involved in the etiology of male infertility. In the next chapter we will review the common approach to analyze transcriptome-wide studies data and then appraise these studies to learn more the plausible causes of sperm pathologies.

CHAPTER II

A REVIEW OF THE TRANSCRIPTOMICS STUDIES

In this section, we will review gene expression studies in male infertility and then meticulously appraise each paper. We will mostly focus on microarray-based studies; next we survey the current state-of-art of next generation sequencing for sperm transcriptomics research.

2.1. Review in literature data

2.1.1. Data collection

We focused on studies of sperm transcriptome using different approaches to provide a global survey of all spermatozoa transcripts present within an ejaculate. To systematically collect transcriptomic studies published until February 2017, Pubmed (URL:http://www.ncbi.nlm.nih.gov/pubmed/) was searched with keywords related to the study background. This search was limited to English-languages publications. Only studies comparing fertile and infertile men were included. Specifically, microarray and Next generation sequencing studies were included. Studies carried out using human ejaculate sperm were the only included, thus excluding human testicular tissue analysis. Moreover, our focused outcome was set as the transcriptomic differences between the 2 groups. Several studies compiled the mentioned criteria. Each selected study was thoroughly analyzed paying particular attention to: (i) the transcriptomic approach employed, (ii) the population studied,(iii) the sample size ,(iv) the number of mapped genes , (v) the fold-change , (vi) the deregulated genes and their statistical significance , (vii) and the deregulated pathways.

2.1.2. Data mining results

All this body of information will be represented according to sperm pathologies in a timeline sequence.

2.1.2.1. Teratozoospermia

In order to trace the origins of teratozoospermia, Platts et al. 2007 [130] evaluated the molecular basis of spermatogenesis in 14 strictly teratozoospermic men (≤3% percent ideal forms) and compared them to 17 fertile men. Microarray quantifications were performed using either Affymetrix U133 version 2 arrays (detects ~54000 target genes) or Illumina Sentrix WG/6-8 bead chip array (detects ~46000 target genes). Two technical replicates were tested. Thus, it was found that 516 abundant transcripts were differentially present between the 2 groups with a significant fold change of ≥10. In fact, the majority of the genes of the ubiquitin-proteasomal pathway (UPP) were repressed in the teratozoospermic group, namely the E2 conjugating and the E3 ligases proteins. Moreover, protein phosphatases transcripts such as CDKN3, DUSP13, PPM1B, PPP1, myosin phosphatase, and elements associated with centrosomal reduction in late spermatid development, ODF 1-4 components of sperm tails, ACRV1and SPAM1 acrosomal components were also reduced.

However, it was noted an abundance of several transcripts involved in the apoptotic pathway including INSR, CASP1, CASP8, BIAP1, WWP2, RERE. Beyond that, analysis of human orthologs showed a reduction in late –stage spermatogenic transcripts (spermatocytes and spermatids) such as HSPA2- involved in chromosome desynapsis and reconfiguration of sperm plasma membrane- which indicated a failure of late stage spermatogenesis. Thus, it can be hypothesized that in cases of teratozoospermia, there is an alteration of the UPP pathway and an initiation of the apoptotic pathway. To note that this pathway is essential for a successful spermiogenesis by eliminating morphologically incompetent spermatozoa. Accordingly, this failure hinders, the positioning of the acrosomal cap, sperm nuclear condensation and sperm tail and mid-piece maturation causing teratozoospermia

2.1.2.2. Normozoospermia

Garrido et al, 2009 [234] performed a comparison of complete sperm expression profiles (SEP) between 5 normozoospermic infertile men and 5 fertile donors. The experiments were performed in duplicate. Sperm RNAs were analyzed using the Human Genome Bioarray, which tests 55.000 genes. Interestingly, hundreds of genes were significantly differentially expressed at least two folds. But, it was only documented a list of the mostly differentially expressed of ten folds and more. Three genes were overexpressed

whereas 136 genes were underexpressed. Among the mostly down-regulated genes, several ribosomal proteins (ranging between 17-34) and more drastically the ornithine decarboxylase antizyme3, the suppressor of potassium defect 3 and the ribosomal protein S3A (fold change between 34.2 and 68.77).

Additionally, a quantitative PCR validation of the mostly differently expressed genes confirmed the microarray results. Moreover, bioinformatic functional analysis showed that these differentially underexpressed genes are involved in spermatid development, gametogenesis, spermatid differentiation and male gamete generation processes. To conclude, the sperm formation and maturation processes were shown deregulated in infertile men without basic sperm analysis abnormalities.

From the study of Garcia-Herrero et al in 2010 [235], 5 semen samples from normozoospermic infertile men (more than 10 million/ml motile sperms and a total number of inseminated sperm higher than 3 million) undergoing assisted reproduction techniques compared to 5 fertile sperm donors were analyzed by microarray technology to subsequently perform ontological evaluation. The microarray platform used is the Human Whole Genome Bioarray which targets 55,000 gene. Microarrays were performed in technical duplicates. The samples were splitted between fresh and processed by swim-up samples, analyzed then compared for transcriptomic differences. The abundance of transcripts from microarray data was determined as, Genes differentially expressed (GDE) presenting a fold change of five or more, and exclusive genes (EG) solely expressed in one group. These data were subsequently submitted to ontological analysis, to determine the biological functions of a gene by comparing the results to the database GO term reproduction GO:0000003.

First off, they identified 1644 GDE in controls and 1568 GDE the infertile group. Among the 10 mostly differentially expressed with a possible role in fertility, to mention: SEMG1, CALM3, CAB39L and PPM2C with a highly significant fold change (FC>5) between 12 and 360. Secondly, they detected 985 exclusive genes in the controls and 692 exclusive genes in the infertile group. For instance, it is noteworthy to mention the exclusive genes with a reproductive role, for the controls: ODF3, SPAG16, CYLC2, TEX14, and as for the infertile group: KLK10 and HIST1H4C.When comparing the list of genes affected in the fertile/infertile from fresh samples, more GO terms related to reproduction were statistically affected in fertile samples. However, no GO terms were statistically affected in the list of genes in post-swim-up samples. Furthermore, while no pathways were affected in the post-swim-up samples comparison, 21 metabolic pathways were affected in fresh fertile men samples and 10 in the infertile group.

Thus it could be noted that, in fertile men high biological processes related to reproduction were shown more operational likewise, for cellular components related to mitochondria and the ubiquitin-proteasome pathway compared to infertile patients. In contrast, for the infertile men the exclusive genes affected were related to embryonic development.

Moreover, swim-up was shown to normalize the differences between fertile and infertile men. A recent study from Bansal et al. 2015 [112] explored the sperm transcriptome of 20 controls fertile, 20 normozoospermic infertile tagged as idiopathic infertile and 20 asthenozoospermic. Expression profiles were generated with the use of the Affymetrix-GeneChip Human Gene 1.0 ST Array. This microarray platform analyzes >30.000 coding transcripts and >11.000 long intergenic non-coding transcripts. After taking 2 technical replicates, they found 2081 significantly differentially expressed transcripts between the 3 groups. Comparisons showed that some transcripts were up-regulated in the normozoospermic group such as ribosomal proteins (RPS25, RPS11, RPS13, RPL30, RPL34, RPL27, RPS5) involved in the assembly of the ribosomes in the mitochondria, HINT1, HSP90AB1, SRSF9, EIF4G2, ILF2. In addition, some transcripts [RPL9, OAZ1, RPL18, RPL35, and FAU] were up-regulated in both the normozoospermic and asthenozoospermic groups, while DAD1 (defense against cell death) gene was the only down-regulated in both studied groups. Moreover, some transcripts were specific to the normozoospermic infertile group (up-regulated: CAPNS1, FAM153C, ARF1, CFL1, RPL19, USP22; down-regulated: ZNF90, SMNDC1, c14orf126, HNRNPK). Overall, some of these deregulated genes were found by gene ontological analysis related to reproduction (n = 58) and development (n= 210). Moreover, other transcripts were related to heat shock proteins (DNAJB4, DNAJB14), testis specific genes (TCP11, TESK1, TSPYL1, ADAD1), and Y-chromosome genes (DAZ1, TSPYL1).

This study investigated the differential expression of some coding and non-coding transcripts in sperm of infertile men, and used gene ontology to set genes according to their common function. This analysis was conducted on Indian men, to keep in mind any potential ethnic variations of these transcripts.

2.1.2.3. Asthenozoospermia

Jodar et al. 2012 [113] has performed the sperm transcriptome analysis on the pathogenesis of asthenozoospermia. Twenty-nine infertile patients (17 asthenozoospermic and 12 normozoospermic) versus 7 fertile controls were recruited, 4 asthenozoospermic and 4 fertile patients were then selected for microarray studies. The microarray-based strategy for

differential transcripts analysis was performed using the Affymetrix Human genome 133 plus 2 arrays which targets 38500 genes. It was found that 17 transcripts were down-regulated and 2 transcripts up-regulated in asthenozoospermic patients. The majority of these transcripts (10 out of 19) correspond to uncharacterized or predicted proteins and the others were found related to spermatogenesis. Subsequently, 5 genes related to spermatogenesis and sperm motility were selected for validation with a quantitative PCR. The selected genes are ANXA2, BRD2, mtND2, mtND3and OAZ3. To mention, ANXA2 encodes a calcium-binding protein involved in sperm motility regulation and other calcium related events such as fertilization and acrosomal reaction. BRD2 encodes a transcriptional regulator that may play a role in spermatogenesis by epigenetically modifying the chromatin structure.

The small set of ANXA2, BRD2 and OAZ3 transcripts showed a significant differential abundance between asthenozoospermic and fertile patient. Moreover, it was highlighted a significant positive correlation between sperm motility and mRNA levels of ANXA2 and BRD2 transcripts. Interestingly, the dysregulation of these genes was also detected in normozoospermic patients. Concerning the protamines transcripts PRM1and PRM2, a significant lower content of these transcripts was detected in asthenozoospermic patients using Quantitative Polymerase Chain Reaction (qPCR), and this reduction did not reach significance in the microarray approach. This finding was congruent with that determined in normozoospermic infertile patients. Therefore, it could be suggested that an abnormal chromatin condensation may initiate the apoptotic pathway thereby affecting the mitochondria. Interestingly, pathway mapping detected that spermatid development and the ubiquinone biosynthesis pathways are affected. To sum up, microarray studies of sperm RNAs elucidated the genetic basis of asthenozoospermia.

Recently, a new class of RNAs, the microRNAs (miRNAs) -which are short (20–24 nucleotides), single stranded non-coding RNA molecules, functioning in transcriptional and post-transcriptional regulation of gene expression- was characterized. The expression profiles of this class were studied in human sperm by Liu et al. 2012 [210]. It included 86 infertile males with abnormal semen analysis majorly asthenozoospermic versus 86 fertiles. Microarray analysis was conducted using CapitalBio mammalian miRNA array V3.0 which contains 2844 mature miRNA gene oligonucleotide probes in triplicate, corresponding to 1823 human, 648 mouse and 373 rat miRNAs, and 16 controls. miRNA microarray results were validated by qRT-PCR and Northern blot. Each experiment was performed in triplicate. It was found that 52 miRNAs were differentially expressed between the semen of infertile males and normal males.Analysis of the microarray expression levels confirmed that

21 miRNAs were significantly overexpressed in the abnormal semen with a fold change fixed to >1.5 . Conversely, 31 miRNAs were significantly underexpressed with a fold change fixed to <0.667 .In addition, qPCR validation confirmed7 out of 11 selected overexpressed transcripts and 6 out of 10 under-represented. The above miRNA expression patterns were also validated by Northern blot.

Thus, a specific signature or set of miRNAs showed differential expression in infertile men. Meanwhile, Abu-Halima et al. in 2013 [211] have performed the sperm microRNA expression profiles from 27 patients attending infertility clinic for treatment. In detail, 9 were normozoospermic, 9 asthenozoospermic and 9 oligoasthenozoospermic patients. Microarray assay analysis was conducted using Sureprint G3 Human v16 miRNA, 8x60K (release 16.0) microarray platform to analyzethe level of 1205 human miRNA.The level of significance was fixed to greater than 2 fold change. The miRNAs that exhibited the highest fold changes and common between asthenozoospermic and oligoasthenozoospermic men are 11: miR-141, miR-193b, miR-26a (targets the estrogen receptor 1 ER1 essential for spermatogenesis), miR-29a, miR-429, miR-200a, miR-99a, miR-363, miR-34b, miR-197, and miR-122.

In parallel, 4 miRNAs were elected specific to asthenozoospermic men, namely,miR-30a, miR-24, miR-1274a, and miR-4286.In addition, Let-7 family was shown deregulated in the 2 case groups, specifically it was shown that Let-7d and Let-7e were predicted to target high-mobility group AT-hook 2proteins involved in the spermatogenesis process and in sperm motility regulation [236]. Subsequently, 6 selected miRNAs were further tested for quantitative expression by the results were largely concordant with the microarray data. In detail, the expression of miR-122 and miR-34b were down-regulated and miR-141 and miR-200aexpression were up-regulated in asthenozoospermic men. It was earlier demonstrated that miR-122 plays a role in the post-transcriptional down-regulation of transition protein 2(TNP2) through targeting the 3' untranslated region of TNP2mRNA, which is synthesized only in round spermatids [237]. As for miR-34b, it was shown highly expressed in adult testis and was found to regulate NOTCH1 gene required for differentiation and survival of germ cells in the rat testis [238].

As a result, novel miRNAs were differentially expressed in infertile men compared to normozoospermic males: miR-429 and miR-1973 differentially expressed in oligoasthenozoospermic and asthenozoospermic men and miR-1274a and miR-4286 in asthenozoospermic men. From the study of Bansal et al. 2015 above-described, they noticed some transcripts down-regulated in the asthenozoospermic group such as ribosomal proteins (RPS25, RPS11, RPS13, RPL30, RPL34, RPL27, RPS5), HINT1, HSP90AB1, SRSF9,

EIF4G2, ILF2. To note, the ribosomal proteins are involved in the assembly of the ribosomes in the mitochondria and a disrupted function of the mitochondria can cause asthenozoospermia. Additionally, some transcripts [RPL9, OAZ1, RPL18, RPL35, and FAU] were up-regulated in asthenozoospermic groups, while DAD1 (defense against cell death) gene was the only down-regulated.Moreover some transcripts were specific(up-regulated:RPL24, HNRNPM, RPL4, PRPF8, HTN3, RPL11, RPL28, RPS16, SLC25A3, C2orf24,RHOA, GDI2, NONO, PARK7; down-regulated: HNRNPC, SMARCAD1, RPS24, RPS24,RPS27A, KIFAP3).When conducting a two-group comparison between normozoospermic infertile patients and asthenozoospermic, they found a down-regulation of several transcripts in the asthenozoospermic group ,of them several documented in the literature linked to fertility such as, PRM1 which is linked to fertilization and embryo development [239], SPATA18 which is linked to development and homeostasis [240], TNP1 which is associated with chromosomal packaging during spermiogenesis [241] , and RANBP9 associated with mRNA splicing [242] and EIF4GIt associated with spermatic differentiation[243]. Moreover, PIWIL1 transcript was shown down-regulated in asthenozoospermic patients, which may play a role in RNA silencing during the chromosomal repackaging [244-245].

2.1.2.4. Oligoasthenozospermia

The above-mentioned study of Abu-Halima et al. 2013 [211] also revealed from microarray data 7 specific microRNAs to oligoasthenozoospermic men, miR-200c, miR-34b*, miR- 15b, miR-34c-5p, miR-449a, miR-16, and miR-19a.A qRT-PCR validation showed that specific miRNAs: miR-122, miR-34b, miR-34c-5p, and miR-16 were down-regulatedandmiR-141 and miR-200a were up-regulated in oligoasthenozoospermic men. Interestingly, a novel miRNA miR-34b* showed the greatest expression change (34 fold-change) in oligoasthenozoospermic samples compared to normozoospermic. The aforementioned work of Abu-Halima was continued in the same framework of sperm microRNAs and in the same group in 2014 [212].

Sperm samples were collected from 80 infertile men mostly oligozoospermic and oligoasthenozoospermic, and 90 normal males. Five miroRNAs were selected from the previous findings with the use of microarray [211].The expression levels of the five miRNAs:hsa-miR-34b*, hsa-miR-34b, hsa-miR-34c-5p, hsa-miR-122 and hsa-miR-429,were analyzed by qRT-PCR and showed that 4 were significantly down-regulated and 1 up-regulated respectively. It is worth to notice that hsa-miR-429 miRNA could be used to monitor the progression of spermatogenesis as demonstrated in mouse testicular tissues [184-

186]. Furthermore, these results were analyzed by support vector machines (SVM) to compute the classification parameters for the combined five miRNAs. Thus, interstingly this microRNAs panel discriminated individuals with subfertility from control subjects with an accuracy of 98.65%, a specificity of 98.44% and a sensitivity of 98.83%.Accordingly, this set of microRNAs can be considered as potential biomarkers of male infertility.

2.1.2.5. Oligozoospermia

A study was conducted by Montjean et al. in 2012 [102] on 8 oligozoospermic patients versus 3 normozoospermic fertile. Using a microarray platform (Affymetrix, HG-U133 Plus 2.0 ChiP), 47000 transcripts were analyzed and 5382 transcripts were detected.Among them, 157 transcripts were found to be significantly deregulated between the two groups (up-, or down-). The majority of them are at least 2 fold down-regulated. They found a drastic down-regulation (up to 43 fold) in genes involved in spermatogenesis, sperm motility and anti-apoptotic process (PRM2, SPZ-1, SPATA-4, MEA-1, CREM). Additionally, a lesser striking drop (15-29 fold) was shown in genes involved in DNA repair (NIPBL), oxidative stress regulation (PARK7) and histone modification (DDX3X, JMJD1A). In contrast, a small group of regulated genes was up-regulated (1-4 folds) which are involved in oxido-reductase activity, oxidative stress response and abortive spermatogenesis (LMNA). Furthermore, the results of the microarray were confirmed by quantitative RT-PCR for the 5 mostly regulated genes and using 2 replicates. Accordingly, this study established the association between oligozoospermia and gene expression deregulation.

The afore-mentioned panel of five miRNAs: hsa-miR-34b*, hsa-miR-34b, hsa-miR-34c-5p, hsa-miR-122 and hsa-miR-429 considered in the study of Abu-Halima et al. 2014 [212] was studied in oligozoospermic patients. It can be re-mentioned that this set of microRNAs could be used as biomarkers of human infertility.

Table 4. Schematic representation of the microarray-based transcriptome studies evaluated in the present review.

Sperm pathology	References	Total transcripts deregulated	Major findings
Teratozoospermia	Platts et al. 2007 [130]	516	Alteration in ubiquitin-proteasome and apoptotic pathways which consequently affects spermiogenesis.
Normozoospermia	Garrido et al. 2009 [234]	139	Sperm formation and maturation processes are deregulated
	Garcia-Herrero et al. 2010 [235]	692	Differential expression of genes related to energy production processes and ubiquitin-proteasome pathway. Ontological analysis showed that embryonic development is affected.
	Bansal et al. 2015 [112]	2081	Some specific coding and non-coding transcripts related to fertility were shown deregulated.
Oligozoospermia	Montjean et al. 2012 [102]	157	Spermatogenesis, sperm motility, anti-apoptotic process and oxidative stress regulation processes were found altered.
	Abu-Halima et al. 2014 [212]	5	Panel of five microRNAs were identified as potential biomarkers of fertility.
Asthenozoospermia	Jodar et al. 2012 [113]	19	Genes related to spermatogenesis, sperm motility and chromatin condensation were affected.
	Liu et al. 2012 [210]	52	A specific set of miRNAs showed differential expression
	Abu Halima et al. 2013 [211]	15	Novel miRNAs were shown differentially expressed.
	Bansal et al. 2015 [112]	2081	Several coding and non-coding transcripts linked to fertility were found down-regulated.
Oligoasthenozoospermia	Abu Halima et al. 2013 [211]	18	Altered microRNAs expression profiles were found.
	Abu-Halima et al. 2014 [212]	5	Panel of five microRNAs were identified as potential biomarkers of fertility.

2.2. New emerging technology approaches: Sperm transcriptome profiling using Next generation sequencing

While most of these studies focused on the microarray-based approach for transcriptomic profiling, data using other global approaches is limited. In this contest, the advent of the so-called ‘‘next generation sequencing’’ (NGS) will permits an in-depth analysis that will provide even more information about the sperm molecular characterization from a global genomic perspective. In the last few years very few studies attempted NGS on sperm samples, therefore we will dissect them in order to better ascertain this unprecedented approach. Johnson et al 2011 [58] tried to recognize the fate of ribosomal RNAs (rRNA) in human sperm. This study is the only one who addressed this poorly understood condition. Briefly, sperms were purified by centrifugation through a 50% Percoll gradient. Purified sperm RNAs from 3 fertile men were used to construct RNA-seq libraries using the Illumina mRNA-Seq kit (Illumina, San Diego, CA, USA). Subsequently sperm libraries were analyzed by NGS using the Illumina Genome Analyzer GAIIx (Illumina, San Diego CA, USA) for 36 cycles.

It was shown that the absence of the full length 18S and 28S rRNA is ascribed to their cleavage. The majority (>75%) of total sequence reads aligned to 18S and 28S rRNAs, thus the highly abundance of these transcripts do not resolve their absence electrophoretically. The sequencing data demonstrated that the rRNAs in sperm are guanine-cytosine rich regions which form secondary structures and consequently they are considered cleaved. These potential cleavage sites were tested by RT-PCR. It is worth noting that cleavage of rRNA ensures translational cessation in sperm at fertilization.

In the same year, Krawetz et al. 2011 [209] conducted a survey of the small RNAs (<200 bases) in sperm. Sperm RNAs from 3 fertile donors were used to prepare sequencing libraries with the Illumina’s Small RNA differential gene expression (DGE) v1.0 kit. Each sample was sequenced on the Illumina’s GAII sequencer. Only the shared sequences between the samples were mapped to the genome.

Classification of the human sperm small non coding RNAs (sncRNAs) permitted to distinguish different classes of small RNA included microRNA (miRNAs) (≈7%), piwi-interacting piRNAs (≈17%), repeat-associated small RNAs (≈65%). A minor-subset of short RNAs within the transcription start site TSS/promoter fraction (≈11%) frames the histone promoter-associated regions enriched in genes of early embryonic development. These have been termed quiescent RNAs.

The 2 major types of sncRNA: miRNA and piRNA, map majorly to repeats or TSS/promoters. Furthermore, targets of miRNA were identified by computational bioinformatic analysis. Of the 2575 transcripts measured only 46 were identified as targeted genes. This highlighted their potential role in translational suppression and/or degradation and perhaps they could provide a rheostat for the maternally stored RNAs. Additionally, among the abundant miRNA the hsa-mir-34c having a role in spermatogenesis, and 4 epi-miRNAs that represses the expression of transcripts encoding proteins involved in the epigenetic machinery: hsa-miR-140, hsa-miR-21, hsa-miR- 152 and hsa-miR-148a.

The first study that used RNAseq by NGS to characterize the population of coding and non-coding transcripts in human sperm was accomplished by Sendler et al. in 2013 [246]. Sperm samples from 6 fertile men were used to characterize various classes and features of sperm RNA. In brief, libraries from purified RNAs were constructed using Single Primer Isothermal Amplification (SPIA) Nugen Ovation kit (Nugen Inc., San Carlos, CA, USA). All samples were subjected to sequencing 2x36 cycles of paired-end sequencing on the Illumina GAIIx platform and reads are compared to the Human reference genome. Multiple- mapped reads at different genomic locations were considered. This in-depth analysis discovered several transcripts. Some novel different additional (i) exons isoforms formed by alternative polyadenylation were detected. For example, the PKM2 pyruvate kinase related to energy pathways presents 2 exons isoforms [247]. Moreover, around 102 (ii) sperm/testis-associated transcripts were detected by RNAseq and absent in the genome reference. These were considered as critical for sperm gametogenesis, we note the PCYT2 in this category [248].

Furthermore, 726 transcripts were found (iii) abundant out of 22302 surveyed. Ontological analysis showed that they were linked to the testes, spermatids, spermatogenesis and infertility. Some are residuels like the protamines genes are required in early spermatogenesis and nuclear organization, while others were considered as sperm-delivered transcripts like the RANBP2 are required for nucleocytoplasmic transport and mitosis at post-fertilization stages [249]. In addition, some (iv) sperm-enhanced transcripts were detected markedly higher in sperm versus testes RNAs, which are required for final stages of maturation. For example, the structural components keratins KRT33A, KRT5 and JUP, the SIX6 transcription factors for spermiogenesis, the HSFY 1/2 affecting early embryonic growth, and the NARP affecting biological rhythm.

Novel classes of (v) non coding RNAs were detected, including the (v.1) RNU family involved in spliceosomes formation and polyadenylation. Despite the absence of spliceosomes activity in sperm, these RNAs (RNU11 in sperm) may be delivered to the

oocyte for further roles. Several (v.2) snaR (small NF90-associated RNAs) transcripts including snaR-C3, -C4, -E, -F, -G1 and -I are also abundant in sperm. At least 250 transcripts were considered as (v.3) intron-encoded RNAs, and they may function as precursors of miRNAs or as regulatory factors. Additionally, some RNAs present in sperm (v.4) overlap regions were detected and may be either lncRNAs or as antisense elements. Final class are the (v.5)chromatin-associated RNAs (CARs) transcribed from intergenic or intronic regions of the genome, bind to DNA and may play a role in regulation of DNA architecture or transcriptional regulation. It contains also 200–300 nt (vi) micro-RNA precursors (pri-miRNAs) namely pri-mir-1181, -miR-3648, -miR-3687,-mir-663 and -mir-181c. It is worth noting that pri-mir-181c putative targets include : FIGN, ONECUT2, ZNF197, POU2F1, ROD1, ACRV2A, ACRV2B, TCERG1 and FOXP1 associated with differentiation [250] , while TNRC6B is linked to the regulation of miRNA pathway [251].

Finally, numerous (vii) epigenetic marks were detected including transcripts encoding histone variants H1FNT, H2AFJ, H3F3C and HILS. In detail, H1FNT (H1T2) plays a role in spermatid elongation and DNA condensation [195-196] and HILS1 is involved in chromatin remodelling during spermiogenesis [252]. Sperm transcripts abundance was associated with other epigenetic elements such as HMRs (hypomethylated DNA regions), H3K4me3 marks and histone-retained regions. Thus, these histone modifications affect several genes expression implicated in embryonic development [253].

To conclude, the pool of RNAs detected by RNAseq may serve a role in the final stages of spermiogenesis or at fertilization. An application note was published in 2014 by Mao et al. [254] where a comparison of different sperm RNAseq methods was carried out as a function of RNA profiling. The challenge was to use as input material low quantity or degraded RNAs from pooled sperm RNAs (8 mixed samples) to successfully enable sequencing. Four commercially available RNA-seq amplification and library protocols for the preparation of samples were assessed:(i)SMARTer™ Ultra Low RNA (SU) for Illumina® Sequencing, (ii) SeqPlex RNA Amplification (SP), (iii) Ovation® RNA-Seq System V2 (OR), and (iv)Ovation® RNA-Seq Formalin Fixed Paraffin Embedded System (FFPES).

Moreover, 2 different libraries preparations methods were investigated: Encore NGS Multiplex System I (Enc) and Ovation Ultralow Library Systems (UL). In brief, the pooled RNA samples were either purified or not, then reverse transcribed and amplified using the above-mentioned kit-based amplification protocols. In detail, the priming methods are: oligodT priming method (SU), random priming (SP) or a combination (OR, FFPES), whereas the amplification methods was either PCR (SU, SP) or isothermal amplification SPIA (OR,

FFPES). Afterwards, 2 kit-based libraries were used for each approach to form a total of 16 libraries. All the kits differ by their input material. Later, all samples were loaded onto the llumina HiSeq 2500 sequencer, then all the reads were mapped against the human reference genome. It was shown that the differences in the initial amount of input RNA and choice of RNA purification step for each library preparation protocol will not affect the RNA profiling. While, substantial discrepancy was introduced by individual amplification methods prior to library construction in terms of fragments length, alignments, number of reads mapping to ribosomal and mitochondrial sequences,these variations introduced by the technical choice of protocol render questionable the interpretation of results of any RNA-seq experiment.

The last recent paper of Jodar et al.2015 [255] addressed the condition of idiopathic infertility by NGS. Spermatozoal RNAs were assessed in 96 patients with normal semen to retrospectively correlate the transcriptomic status by NGS to reproductive outcomes. Specifically, the amplification method used was SeqPlex RNA Amplification (Sigma-Aldrich Co.), libraries were prepared using the NEBNext Ultra DNA Library Prep Kit for Illumina (New England Biolabs) and the sequencer used was the Illumina HiSeq-2500.

At baseline, sperm RNA elements (SREs) set or specific signature reflective of fecundity status were defined. Namely, 648 SRE of which the majority (585) were in exonic regions of fertility-related ,cell-function related or unknown genes, and the others were intergenic regions, sperm-specific intronic elements, and non-coding RNAs. It was found that the absence of these specific SREs drastically the probability of achieving live birth by timed intercourse or intrauterine insemination decreased from 73 to 27% and not in assisted reproductive techniques. Overall, 30% of idiopathic infertile couples presented an aberrant SRE set.

A technical comment by Cappallo-Obermann was published in August 2016 [256] about somatic transcripts contamination bias due to inefficient removal of round cells. Accordingly, a response to the comment by the authors was published to highlight the efficiency of the 50% pure sperm gradient method in removing round cells, and the stringent criteria for selection of SREs to avoid dilutive effect of few somatic transcripts. To mention, that if these somatic transcripts were present they result from interaction of extracellular vesicles originated from accessory glands.

After reviewing the abstract of a recent study of Pereira et al. 2016 [257] to investigate the epigenetic basis of idiopathic infertility using next-generation sequencing in 5 infertile men versus 3 controls. In detail, Illumina stranded RNA-Seq library preparation with Ribozero purification was used to construct the paired-end libraries, and subsequent pilot

sequencing expanded to 50-60M reads at 2x76bp utilizing an Illumina platform (NextSeq 500). Interestingly, 86 genes with statistically different expression were identified, among them 7 differentially expressed genes (APLF, CYUB5R4, ERCC4, MORC1, PIWIL1, TNFRSF21, and ZFAND6) known to regulate DNA repair mechanisms or protect spermatozoa from reactive oxygen species throughout spermatogenesis were significantly under-represented in the study group compared to the control study.

The timeline of studies utilizing Next generation sequencing technology for transcriptomic analysis of human spermatozoa are illustrated in ***Figure 3.***

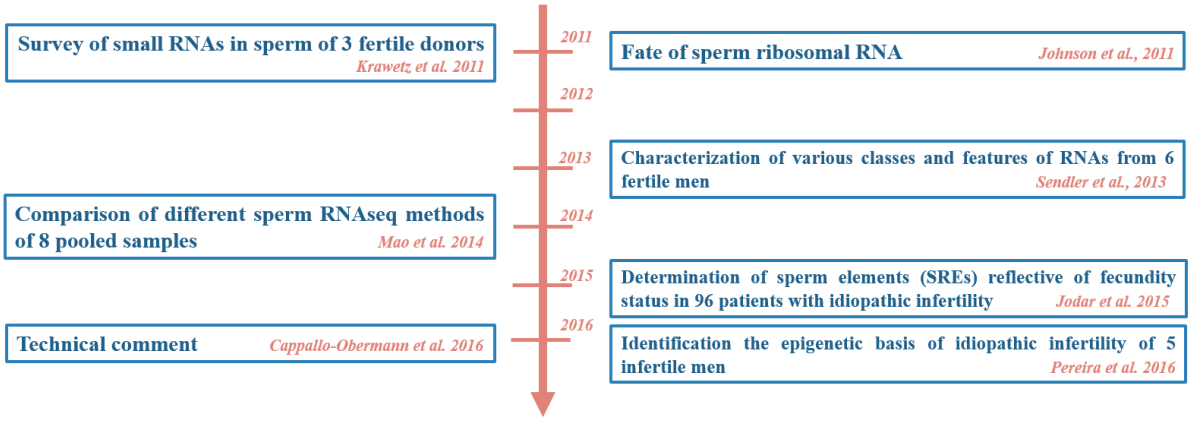


Figure 3. Timeline of studies involving transcriptomic profiling using Next generation sequencing from human spermatozoa.

Conclusion

In conclusion, this extensive literature review, allowed detecting several potentially interesting transcripts that may explain the pathogenesis of the aberrant spermatogenesis and some may serve as biomarkers for male infertility. In some cases, we identified a specific RNA and miRNAs signature differentially expressed in infertile men. Most of the studies used the microarray approach to identify potentially informative molecular set of transcripts in different subset of infertile males categorized according to their sperm pathology.

In the next chapter, we will describe the methods and methodologies used in this research work.

CHAPTER III

MATERIALS AND METHODS

In this section, we will describe the laboratory and computational methods used in this work. An overview of the optimized protocol of RNA will be presented in the first section, followed by a description of the basics of RNA-seq analysis. Finally, a description of the material and methods, which are based on the experimental work performed in the framework of this thesis will be detailed including the population characterization, the technical procedures and the bioinformatic analysis.

3.1. Optimization of sperm RNA isolation

Spermatogenesis is a complex process which permits the generation of mature spermatozoa. A fully mature spermatozoa contains DNA and a heterogeneous population of RNA molecules. Today, genomics is a thriving field of inquiry. In fact, several breakthroughs in molecular biology are happening at an unprecedented rate. Accordingly, comprehensive characterization of sperm genome and transcriptome outstrips our understanding of the molecular mechanisms of a physiological and pathological spermatogenesis and strengths our knowledge to find new molecular biomarkers or signatures for male fertility potential. Incredibly, it is about turning raw data from DNA and RNA sequences into deeper insights of male fertility.

Extraction of high quality DNA and RNA from spermatozoa is a challenging step, due to the high degree of DNA compaction. Therefore, this requires alternative methods and refinements of protocols, which remains very tedious. Currently, a number of approaches have been published to deal with these problems that hinder the use of conventional isolation

methods for extracting either DNA or RNA from human sperm cells. Improved and modified methods have been detailed in several works in order to efficiently lyse sperm cells and gain access to the nucleus. As for sperm RNAs, the last well established protocol is that of Goodrich and his coworkers in 2013 [258]. This protocol relies on the use of a modified column-based method by combining the conventional kit method RNeasy with minor modifications using an additional Trizol step for sperm cell extra-lysis. Very recently, a detailed methodology for human and rat spermatozoa RNA isolation by Bianchi et al. in 2018 was developed [259].

Regarding sperm DNA, several studies attempted to overcome the inefficient DNA isolation. These methods varied in the type of the lysis buffer used to digest the cell and the nuclear proteins; which may comprise detergents and/or chaotropic salts and proteinase K; and reducing agent to break disulfide bonds between protamines. Subsequently, DNA may be isolated by ethanol precipitation or silica-based spin columns [260-261].

However, we were unable to find a published method that allows isolation of DNA and RNA so that the same pool of sperm cells can be assayed for genome and transcriptome profiling, respectively. Instead, DNA and RNA are usually isolated from the same patient but from different pellets, thus in research using human clinical samples the amount of material available is most often scarce, allowing limited downstream applications. Therefore, the collection of a proper amount of nucleic acids for subsequent analysis becomes crucial. In addition, the simultaneous isolation of DNA and RNA from a single sample facilitates the meaningful interpretation and correlation of genome and transcriptome data, while reducing time, cost and sampling error.

Here we describe the extraction of both DNA and RNA from human spermatozoa using a simple column-method and present an option for extracting both DNA and RNA from the same cells for genome and transcriptome profiling. Firstly, we will test a simple disruption method of the sperm nucleus that will allow efficient subsequent isolation of nucleic acids. The proposed method includes a brief hypotonic shock followed by a flash-freezing in liquid nitrogen. Then, we will try to identify important experimental variables affecting the nucleic acids yield using an experimental design technique: the incomplete factorial design matrix.

This methodology was introduced into biological experiments such as PCR optimization by Cobb et al. in 1994 [262]. Finally, we will present the RNA and DNA metrics results after using the aforementioned co-extraction method, as well as its suitability for downstream applications.

3.1.1. Materials and methods

3.1.1.1 Isolation of sperm cells

Ten healthy male participants donated a sperm sample for this study. All participants have at least 48 h of abstinence. The ejaculate samples were washed with sperm washing medium (Origio, Denmark) and the purity of the human sperm was >0.01% (less than 1 round cell per 10 000 sperm) based on phase-contrast microscopic observation. Sperm pellets were counted on a Makler cell chamber using the average of ten square areas, and visually inspected for somatic cell contamination. The number of sperm cells was adjusted to 5 million cells.

3.1.1.2. Cell lysis

A simple disruption technique was evaluated: sperm pellets were subject to an osmotic shock in 500µl distilled water for 1minute, and then pelleted. The pellets were flash-frozen in liquid nitrogen for cellular lysis then stored at -80°C. Each sample was thawed on ice and subsequently sperm nucleic acids were extracted.

3.1.1.3. Testing the effectiveness of lysis

To test the effectiveness of this disruption method, aliquots pre and post-treatment were assessed for membrane integrity and morphology.

- Sperm membrane integrity was assessed using the supra vital eiosin-Y staining. Briefly, stained smears were evaluated under microscopy and the results of membrane integrity are expressed as the percentage of unstained or negative sperm (having an intact head membrane) after examination of 200 sperms.
- Sperm smears were stained with 4′,6-diamidino-2-phenylindole (DAPI) fluorescent staining. Sperm head morphometry dimensions were obtained by measuring captured sperm images using ImageJ software v.1.41 (National Institutes of Health, Bethesda, MD, USA). Measures included sperm head area (A) and sperm head perimeter (P).

More so, we have tried to address sperm membrane injury by using DAPI fluorescent staining which stains nuclear DNA. Initially, this approach was adopted from Harrison et al. 1990 [263]. That being said, we will monitor sperm membrane damage at which an efficient permeabilization is inflicted using several different approaches.

In brief, air-dried smears were stained with DAPI. Images were analyzed using analysis routines written for NIH ImageJ (http://rsb.info.nih.gov/ij/). Fluorescence intensities in individual cells were quantified by integrating fluorescence intensities under a digital mask corresponding to the area of each cell. On average 200 sperm cells were counted per slide.

3.1.1.4. RNA isolation

Sperm RNA was extracted using a detergent-based lysis followed by an in-column purification using the RNeasy plus minikit (#74134; Qiagen, The Netherlands). The below protocol was performed as outlined by Goodrich et al.[258].

Briefly, thawed pellets were lysed in 500 µl Buffer RLT Plus with βME (7.5µL). Lysates were combined with equal volume of warmed 37°C Qiazol (Qiagen), incubated for 5 minutes then vortexed. A volume of 200µL of chloroform was added to the mixture, followed by a cold centrifugation. The aqueous phase was transferred to a genomic DNA (gDNA) eliminator spin column, which was centrifuged for 30 s at ≥8000 x g. The flow-through was combined with 100% ethanol (1:0.8), loaded onto the spin column, and the latter was centrifuged at 12000 × g for 1 min to bind RNA. Wash and elution steps followed the manufacturer's protocol, including 2 separate 50 µL elutions to maximize yield.

3.1.1.5. DNA isolation

During RNA isolation, the gDNA eliminator columns were placed on ice. The gDNA eliminator columns method is a novel fast *non-enzymatic* way for removal of genomic DNA.

In order to recover DNA, the column was washed twice with 500µl of RPE buffer (Qiagen) and incubated for 4min, then centrifuged for 4min at 5600g. Before elution, the column was dried in an empty tube and centrifuged for 4min. DNA was eluted twice in 50µl RNase/DNAse free water incubated for 10min then centrifuged for 4minutes.

3.1.1.6. Nucleic acids QC metrics

RNA yields and quality were determined using the Nanodrop 2000 Spectrophotometer (#E112352; Thermo Scientific, Somerset, NJ) and Agilent 2100 Bioanalyzer (Agilent Technologies, Palo Alto, Calif.). DNA yield was determined using the Nanodrop 2000 spectrophotometer. As for DNA quality, it was analyzed on a 1% agarose gel electrophoresis for 40 minutes at 150 Volts.

3.1.1.7. RNA library preparation and sequencing

The total RNA was treated with Ribo-Zero™Magnetic Kit (Epicentre) to deplete rRNA. The retrieved RNA is fragmented by adding First Strand Master Mix (Invitrogen）. First-strand cDNAis generated by First Strand Master Mix and Super Script II reverse

transcription (Invitrogen) (Reaction condition：25℃ for 10 min；42℃ for 50 min；70℃ for 15 min). Purify the product (Agencourt RNAClean XP Beads, AGENCOURT), then add Second Strand Master Mix and dATP,dGTP, dCTP, dUTP mix to synthesize the second-strand cDNA (16℃for 1h).Then the purified cDNA combine with End Repair Mix, incubate at 30℃for 30min. After purification with beads, add A-Tailing Mix , incubate at 37℃for 30 min. Combine the Adenylate 3'Ends DNA, Index Adapter and Ligation Mix, incubate at 30℃for 10 min. After that, add the Uracil-N-Glycosylase enzyme into the purified ligation product, incubate at 37℃for 10min. Several rounds of PCR amplification with PCR Primer Cocktail and PCR Master Mix are performed to enrich the cDNA fragments. Then the PCR products are purified with Ampure XP Beads. The libraries will amplify on cBot to generate the cluster on the flowcell (TruSeq PE Cluster Kit V3–cBot–HS，Illumina). The amplified flowcell will be sequenced on the HiSeq 2000 System (TruSeq SBS KIT-HS V3，Illumina).

Primary sequencing data, called as raw reads, is subjected to quality control (QC) to determine if a resequencing step is needed. After QC, raw reads are filtered into clean reads which will be aligned to the reference sequences. QC of alignment is performed to determine if resequencing is needed. The alignment data is utilized to calculate distribution of reads on reference genes and mapping ratio. If alignment result passes QC, we can proceed with downstream analysis.

3.1.1.8. Global 5-methylcytosine (5-mC) measurement

We used Methylated DNA Quantification Kit (Colorimetric) (ab117128; abcam, USA) for the quantification of 5-mC. These enzyme-linked immunosorbent assay (ELISA) kits are highly sensitive and accurate. The assays were performed according to the manufacturer's recommendations. Briefly, 100ng of genomic DNA was used for 5-mC quantification, respectively. It is important to note that each sample was run in duplicate. After incubation with the input DNA, the wells were washed, and a capture antibody was applied to each well, after which the wells were again washed and detection antibody applied. Use of enhancer solution and development solution created a color change proportional to the quantity of 5-mC, and the samples were read on an automated plate reader at 450 nm absorbance. The amounts of 5-mC were calculated using a simple formula provided by the manufacturer.

3.1.1.9. Statistical analysis

Statistical analyses of results were used for treatment comparisons and carried out by one-way of variance (ANOVA) using the using the SPSS statistical package (SPSS Inc., Chicago, IL). If the P-value was less than 0.05 by ANOVA, the Tukey-Kramer Honestly

Significant Difference test was utilized using the same program. Values of $P < 0.05$ and $P<0.01$ were considered significant.

3.1.2. Results and discussion

It is often quite challenging to extract RNA from sperm cells. In order to efficiently extract RNA from sperm nucleus, conventional protocols needs improvement, thus several considerations must be kept in mind. Firstly, a heterogeneous population of cells is present in the ejaculate which demands a purification step to isolate only a population of sperm cells devoid of somatic cells contaminants [264]. As for sperm RNAs, its amount is 200 times less than the quantity in other cells [265], thus the removal of somatic cell is essential.

Here we used a simple wash centrifugation method. Pelleting sperm cells as separation method was recently used by Bianchi et al. [259]. Although, the mainstream transcriptome studies used gradient, either single or double layer, as sperm purification technique [266]. A technical article evaluated the effectiveness of sperm purification method, revealed that gradient centrifugation method yield higher RNA than somatic cells lysis. Conversely, the latter recovered more sperm cells [267]. On the contrary, others showed that high quality sperm RNA was overly isolated from whole semen and swim-up samples [268]. It is worthy to note, that isolation of sperm cells using swim-up technique showed absence of some transcripts which will mislead any interpretation of the subsequent data [269].

Moreover, sperm has a unique nucleus packaging resistant to isolation techniques [270], thus disruption and lysis steps are of paramount importance to efficiently get access to the nucleus. So, sperm can be considered as a 'tough nut to crack'. Although several technical modifications can be exploited, given the compacted structure of human sperm, we reasoned that the disruption step can be improved.

Sperm cell lysis was performed, in various RNA purification protocols, by using either a guanidine-based RLT lysis buffer supplemented with β-mercaptoethanol (BME) [271] or a bead-based homogenization [258, 267] to release cellular contents of sperm. Herein, the primary attempt was to monitor sperm membrane alteration inducing its destabilization, in order to evaluate at which the injury is overtly inflict. For this purpose, we have tried to address sperm membrane damage using different physical treatments including, osmotic shock, snap-freezing and the combination of both treatments. We examined the effectiveness of osmotic-shock and snap-freezing on lysis of sperm cells by checking the cytological characteristics of the sperm cell: the sperm membrane integrity and sperm head morphology.

The percentage of spermatozoa with intact cell membranes showed significant intergroup differences in all treatments ($p<0.01$) ***(Table 5)***. Fascinatingly, the combination of both treatments exhibited a very extreme level, and therefore rendered the membrane "unintact" of the whole cell population.

Table 5. The percentage of spermatozoa with intact cell membranes, after three physical treatments (means±SEM).

Physical treatment	Mean value %
Fresh	$90.2\pm5.4^{a,b,c}$
Osmotic choc	$27.6\pm0.9^{a,d,e}$
Snap freeze	$0.2\pm0.1^{b,d,e}$
Osmotic choc+snap freeze	0^{c}

Mean values of percentage of dead sperm in human fresh semen samples stained with eosin-nigrosin after three different physical treatments (osmotic choc, snap freeze, osmotic choc followed by snap freezing)
Within a column, different superscripts denote values that differ significantly ($P < 0.01$).

The morphometric dimensions of the human sperm head were meticulously measured following the three physical treatments. We found that using the above-mentioned different treatments in sperm cell suspensions caused no substantial changes in the sperm morphometry notably head area and perimeter ***(Figure 4)***. More so, we have measured the fluorescence intensities of DAPI to further evaluate spermatozoa intactness. Penetration of the stain into the sperm heads suggested that treated sperm cells had compromised membranes. Our results revealed that, treating samples with osmotic shock in concert with snap freezing intensified significantly the membrane disruption ***(Figure 5)***.

Furthermore, gross microscopic examination revealed the absence of somatic cells contaminants. Accordingly, it can be speculated that osmotic shock treatment caused burst of the somatic cells. To sum up, sperm cell membranes become mechanically weakened by osmotic shock (osmotic injury), by abrupt cooling (cold shock injury) or by both.

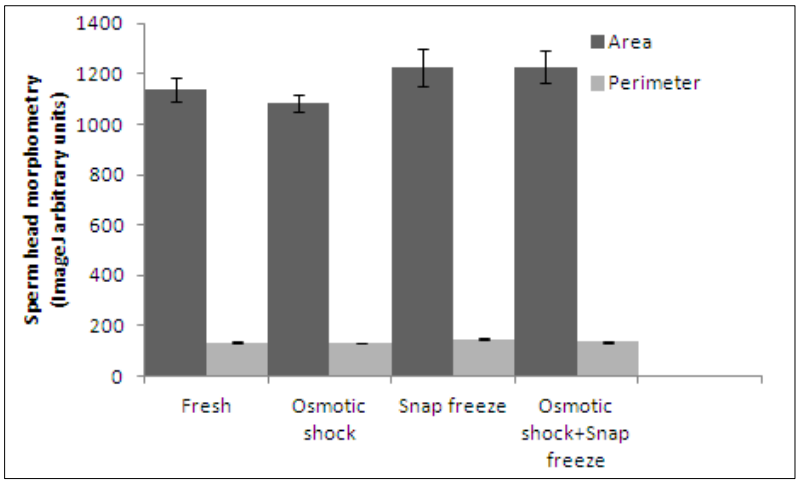
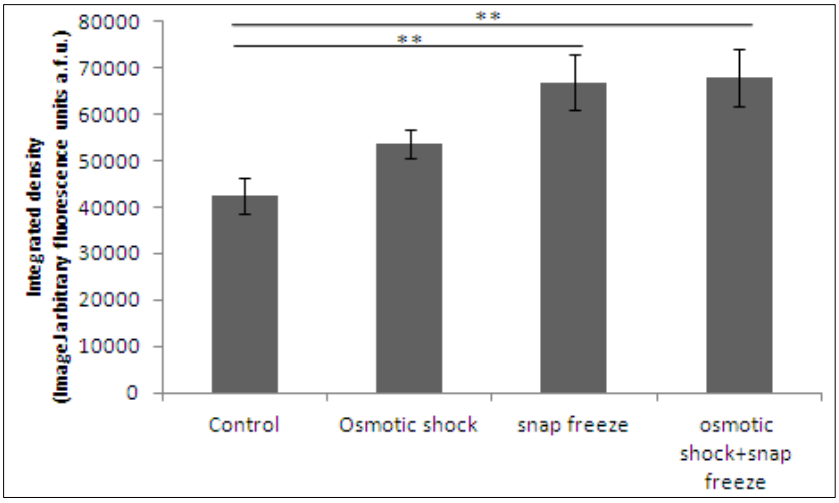


Figure 4. Sperm head morphometry analysis (mean ± SEM) after three physical treatments (osmotic choc, snap freeze, osmotic choc followed by snap freezing) compared with spermatozoa in fresh semen.



$^{**}p<0.01$

Figure 5. Histogram of relative fluorescence intensities (DAPI), reported as arbitrary fluorescence units (a.f.u.), after adhering to sperm nucleus under different physical treatment conditions.

Next, we incorporated the above-mentioned treatments in miniaturized experiments to determine its efficiency in terms of RNA yield. We used, for the first time, the fractional factorial design approach to investigate the effectiveness of these modifications in conventional RNA extraction protocols. A screening experiment was implemented to identify the important disruption strategies that will influence RNA yield during RNA extraction. The 2^k fractional factorial design (ref), where k represents the number of factors, was used to evaluate multiple components of the protocol of RNA extraction in different independent experiments.

Various strategies such as snap freezing [272], osmotic choc and Trizol warming [268], were identified from the literature as useful treatments for cell disruption in nucleic acid

extraction protocols. We also reasoned that the addition of DTT (dithiothreitol) to the lysis buffer; which have been very early identified as possible disulfide bond reducer [273], will further disrupt sperm cells. After many experimental trials we empirically tested and determined the optimal DTT concentration at 100mM.

Three variables were assayed and included the following: 1) The sperm cells conditions either physically treated or not; 2) the addition of DTT (100mM); 3) warming of Trizol. All factors were studied at two levels ***(Table 6)***. If a full factorial experiment were to be undertaken (3 factors at 2 levels), 8 trials would be required. These 8 trials were performed in duplicates.

Table 6. Factors and Levels Assayed in Fractional Factorial Design for the optimisation of sperm RNA extraction: Factors Screening.

Factor	Level +	Level -
1. Samples condition	Treated pellet: *Osmotic choc + Snap freezing*	Fresh pellet
2. Addition of DTT	Yes	No
3. Trizol warming	Yes	No

In order to determine the main effects of the 3 factors, a factorial design matrix is sufficient to determine the most important factors that would optimize the RNA yield, which requires only 8 experiments. A summary of the experiments is given in ***Table 7a***.

Table 7. Summary of the experiments to find optimal conditions of sperm RNA isolation.

a: Matrix for Fractional Factorial Screening Design

Trial	Samples condition	Addition of DTT	Trizol warming	RNA yield (ng per million ejaculated sperm)
1	-	-	-	17
2	+	-	-	30
3	-	+	-	36
4	+	+	-	62
5	-	-	+	20
6	+	-	+	33
7	-	+	+	45
8	+	+	+	81

b: Effects of the Factors on the RNA Yield

Factor	Effect
1	+22
2	+31
3	+8.5

This statistical design of experiment approach was adopted to identify which factor positively or negatively influencing RNA yield in RNA extraction protocols. We evaluated the RNA yield in ng/µl obtained from the 8 experiments. The effect of each factor is estimated by taking the sum of the experimental result for all experiments, with its appropriate sign, and dividing the total by 4. The same result can be obtained by subtracting the average experimental value of the 4 (-) trials from that of the 4 (+) trials. The values of the effects are given in ***Table 7b***. For example, the effect of samples condition before RNA extraction on the subsequent RNA yield is calculated in the following way:

Factor 1 effect = Average of factor 1 level (+) – Average of factor 1 level (-)

It can be concluded from **Table 3b** that the addition of DTT and the pre-extraction physical treatment of sperm cells have a major influence on RNA yield, whereas warming Trizol have a lesser effect. Taken together, these initial experiments suggested that using physically disrupted sperm cells prior to nucleic acid extraction, DTT treatment and warming Trizol to 37°C, enhances the yield of RNA from sperm cells.

Next, we characterized the RNA quality after applying the 3 factors to the original protocol on 10 samples. RNA metrics associated with the latter RNA isolation method, by analysis on the bioanalyzer (Agilent 2100) and the Nanodrop 2000 Spectrophotometer (#E112352; Thermo Scientific, Somerset, NJ), showed the following:

- Concentration (mean ±SD) = 75.6±13.11 ng per million ejaculated sperm.
- A260/A280 ratio ranging between 1.86 and 1.91.
- RNA integrity number (RIN) between 2.3 and 3.2 with a mean of 2.67 ± 0.35 (mean ±SD).
- 28S/18S was null in all samples.

For example, we yielded from one sperm sample 81ng per million sperms with an A260/A280 ratio of 1.89, a RIN of 2.3 and a null rRNA [28S/18S] ratio.

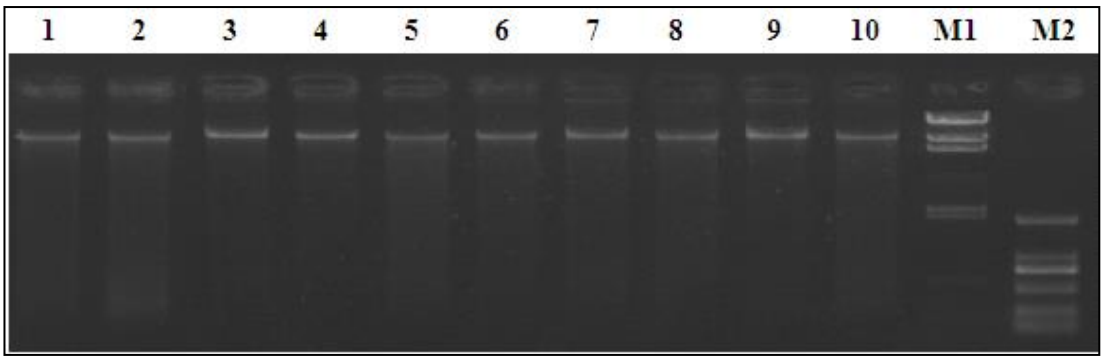
The mean of RNA recovery was reported to be around 60ng per million sperm cells [271]. This value was congruent with that reported by Schuster et al. (2016) [274] with values varied between 45–60 ng per million ejaculated sperm. The ratio for pure RNA A260/280 is ~2.0 [275].The RIN value for RNA derived from spermatozoa was recently suggested in a technical paper reporting a protocol for sperm extraction from numerous species and consequently establishing a database dedicated to sperm-borne RNA profiling of multiple species, the “SpermBase” [274]. A RIN between 2 and 4 is indicative of a good sperm RNA

quality, whereas RIN<2 is suggestive of excessive RNA degradation and RIN>4 is indicative of somatic cell contamination. Contrastingly, Georgiadis et al. (2015) considered a "high quality" RNA from sperm presented a RIN>3 and 28S/18S rRNA ratio greater than 0 as alternative indicator of 18S and 28S rRNA presence [268]. It is known that spermatozoal RNA is inherently fragmented owing to the cleavage of sperm ribosomal RNAs [58]. This reflects the unusual 28S/18S rRNA ratio [276] and the absence of the distinct rRNA bands on the electropherograms. Our results indicate that the isolated sperm RNA is of good quantity and quality.

1. Clean read Q20 (%) [≥90]: Sequencing quality metrics is used to assess the accuracy of a sequencing platform, it is represented by the Phred quality score (Q score). The latter indicates the probability that a given base is called incorrectly by the sequencer. The Q20 means that the base call accuracy (i.e., the probability of a correct base call) is ≥90%. A lower base call accuracy of 90% (Q20) will have an incorrect base call probability of 1 in 100, meaning that every 100 bp sequencing read will likely contain an error.
2. Gene unique mapping ratio (%) [≥80] which is the proportion that can only map to one gene position in total mapped reads.
3. Genome mapping ratio (%) [≥50] which is the proportion of reads that can map to genome in total reads.

All the samples tested passed this quality control. We relied primarily on the results of the QC, since this work solely discusses the effectiveness of the protocol. We then attempted to isolate sperm DNA from the gDNA eliminator columns. Parallel DNA extraction yielded DNA that could be potentially used in subsequent applications. The DNA metrics associated with the subsequent DNA isolation method, by analysis on the Nanodrop 2000 Spectrophotometer (#E112352; Thermo Scientific, Somerset, NJ), and agarose gel electrophoresis showed the following:

- DNA yield ranged from 0.4 to 1.4µg (median 1µg; mean ±SD= 0.9±0.35 µg per million ejaculated sperm.
- A260/A280 ratio ranging between 1.85 and 1.9.
- Electrophoretograms showed that the samples were slightly degraded ***(Figure 6)***.



Molecular DNA markers: Lane M1L: λ-Hind IIIdigest (Takara); Lane M2: D2000 (Tiangen)

Figure 6. Electropherograms (in a 1% agarose gel run for 40 minutes at 150 Volts) of DNA from sperm cells.

Conventional DNA extraction protocols yielded between 2.5-3 pg/ cell, clearly 2.5-5 µg per million sperm [261]. A 260/280 ratio of ~1.8 is generally accepted as "pure" for DNA [275].

Accordingly, we were able to use sufficient DNA for ELISA colorimetric assays. We evaluated the global DNA methylation by quantifying the 5-methylcytosine levels. The DNA for all samples has been used successfully to determine the relative methylation status using a colorimetric assay ***(Figure 7)***. Additionally, the measured values presented a coherent pattern and were concordant with the values reported by Jenkins et al. (2013) using the same approach [277].

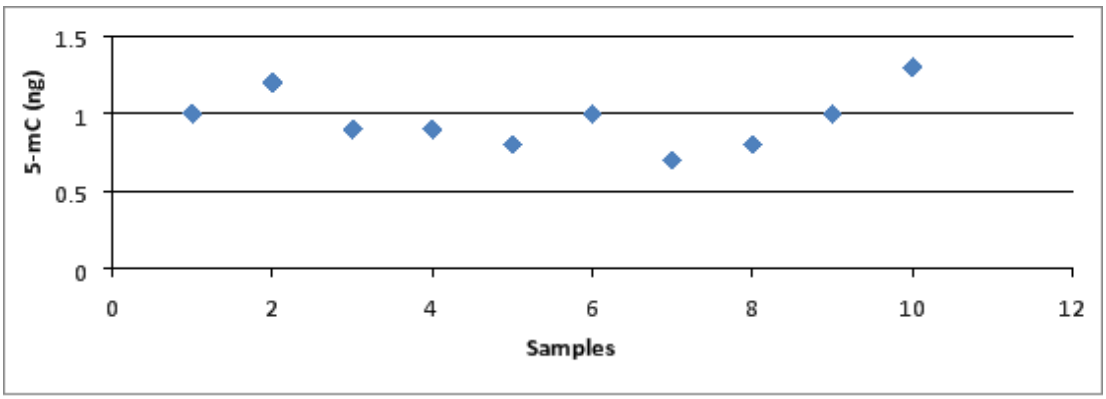


Figure 7. The levels of methylated DNA (5-mC) in total sperm DNA.

3.2. The ABCs of RNA-seq: An explanatory review on the basics of experimental design, laboratory performance and computational analysis in RNA-seq experiments

Nowadays, Next Generation Sequencing (NGS) permits the investigation of an entire genome and transcriptome with unprecedented resolution and throughput. Along these lines, second-generation sequencing of the transcriptome (RNA-seq) allows the identification of the

expressed genomic loci in a cell at a given time over the entire expression range [278]. However, this approach shows superiority over hybridization-based approaches and represents a new attempt to tackle multiple uncertainties and bias of other techniques. Interestingly, RNA-seq has a high sensitivity to detect genes expressed either at low or very high level [279]. This approach is not limited only to gene expression changes between different populations, but also gives unprecedented detail about additional transcriptional features such as non-coding transcripts, splice isoforms, novel transcripts and protein-RNA interaction sites [280]. While it can be performed in species for which genomes are available; it has also become a plausible strategy for non-model organisms [281] such as parasitized insects [282], or treated cultured animal cells [283] for example . However, its most common application is the detection of gene expression changes between different cell populations and/or experimental conditions.

In summary, RNA-seq analysis can help to investigate various transcriptional features, notably gene expression and differential expression. Furthermore, allele-specific expression, variants, mutation discovery, transcript discovery, fusion detection and RNA editing are other applications to profile the transcriptome [280]. The general workflow of a differential gene expression analysis; adapted from Conesa et al. 2016 [284] is:

1. Experimental design: i) Library type; ii) Numberof replicates; ii) Samples pairing; iii) Pooling; iv) Reference samples; v) Normalization of data; vi) Sequencing decisions.
2. Sequencing (biochemistry) 3. Bioinformatics: i) Quality control of raw data; ii) Preprocessing; iii) Alignment; iv) Quality control of aligned reads; v) Reads assembly and quantification; vi) Differential expression analysis: normalizing and transforming read counts; vii) Variability within the experiment: data exploration; viii) Differential gene expression analysis.

NGS facilities have proliferated and many biologists now have access to a service to which they can submit a sample and are handed back raw bioinformatic data. Whilst, biologists generally do not have the necessary background to analyze and critically interpret the results of these challenging experiments. Thus, the scientists wishing to use RNA-seq should have a solid understanding of the principal issues that are involved in computational analysis by RNA-seq. As it is difficult and time consuming to learn about this subject from the technical literature, and most reviews have other goals, in this review, we explain the principles of RNA sequencing that are important for the appreciation and interpretation of the outcome of transcriptomics experiments. In the main part of this review, we describe the

steps of a typical pipeline for all the different types of RNA-seq experiment. Additionally, the reader is referred to other reviews for more technical detail and specialized topics.

3.2.1. Design considerations

Several aspects should be taken into consideration when designing an RNA-seq experiment that will affect analysis strategy:

3.2.1.1. Library type

By definition *a library* is a collection of DNA or cDNA fragments that will be sequenced after hybridization on the flow cell. Due to the various types of RNAs, there are many RNA-seq library construction strategies. The total RNA will allow detecting mRNAs as well as non-coding RNAs. However, mRNA exploration solely is by far the most commonly used in RNA-seq experiments. Indeed, mRNAs needs to be enriched since they are present in low abundance compared to rRNAs and tRNAs that can compose an overwhelming fraction of the total RNA population. This can be achieved either via enrichment of poly(A)-tail containing nucleic acids using oligo(dT) beads or by removal of ribosomal RNA [279].

Moreover, RNAs may be selected according to their size before and/or after cDNA synthesis [284].Another consideration is whether to generate a strand-specific library, such approach leads to a complete view of the transcriptome such as identification of antisense transcripts with potential regulatory roles, determination of the transcribed strand of other non-coding RNAs demarcate the exact boundaries of adjacent genes transcribed on opposite strands, and resolution of the correct expression levels of coding or non-coding overlapping transcripts [285]. Here, libraries could be prepared using different modified protocols than the common approaches and subsequently each strand will be barcoded [286]. Furthermore, normalization of cDNA libraries can sometimes be performed. This approach refers to the equalization of cDNA concentration to ensure that the reads are evenly distributed, by pooling samples to obtain equal amount of each read prior to sequencing. Thus, it is helpful for decreasing the prevalence of abundant transcripts, thereby facilitating the assessment of rare transcripts [287].

3.2.1.2. Replicates

We can distinguish between technical and biological replicates. First off, biological replicates are separate samples or individuals. More so, the ENCODE consortium defines biological replicates as “RNA from an independent growth of cells/tissue” [288]. It is worth considering that publication quality data needs at least 3 biological replicates per sample

group [289]. In a recent paper reporting the most exhaustive RNA-seq experiment (48 replicates per condition), Schurch et al. (2016) [290] recommended the use of at least six replicates per condition if the focus is on a reliable description of one condition's transcriptome or strongly changing genes between two conditions. However, if the goal of the experiment is to identify the majority of all differentially expressed genes; including slightly changing ones and those that are lowly expressed; as many as twelve replicates are recommended. Secondly, technical replicates are different library preparations or repeated sequencing runs using the same RNA isolate or sample. It should be stated that technical/experimental variability introduced by the sequencing protocol is quite low and well controlled [284]. As for batch effect reduction, it is recommended to sequence everything for an experiment at the same time. At a glance, biological replicates depict variations in biology, whereas technical replicates portray variations in chemistry (RNA extraction, library preparation, hybridization…) and physics (images analysis and base calls). In particular, power analysis can be used to estimate sample sizes.

3.2.1.3. Sample pairing

The use of paired or matched samples reduces variance between groups. Several mathematical models have been proposed to deal with paired samples during RNA-seq experiments analysis [291].

3.2.1.4. Pooling

Pooling biological replicate RNA samples can maintain the biological information, while reducing the cost of sequencing. Although, validity and utility of sample pooling for gene expression analyses using NGS have been rarely evaluated. It is highly desirable to consider that pools should be similar as possible (concentration, characteristics…), nevertheless the presence of outliers or contamination in samples can lead to bias in the data. Sometimes reference samples can be used across batches in every run.

3.2.1.5. Normalization of RNA-seq data

To ensure accurate inference of expression levels or estimation of absolute transcripts concentrations, a factor analysis on suitable sets of control genes (artificial RNA spike-in) or samples (replicate libraries) can be performed. To note that, artificial RNA spike-in,RNAs of known quantities, can also be used for the calibration of the RNA concentrations in each sample [292].

3.2.1.6. Sequencing decisions

- **Number of reads per sample**: while for differential expression 20-25M reads are needed, for other purposes such as allele specific expression, alternative splicing and de novo assembly >100M reads are needed [288]. It is important to note that heterogeneous samples need more reads.
- **Read length**: an appropriate read length should be used based on the final goal of the study and the genome size. In fact, whilst using 50 bp single-end reads is sufficient for differential expression analysis, longer reads are recommended for other analysis [293].
- **Paired-end or single read**: Single read (SR) sequencing determines the DNA sequence of just one endof each DNA fragment. Paired-end (PE) sequencing yields both ends of each cDNA fragment then aligning the forward and reverse reads as read pairs. It is necessary to state that PE reads are recommended for *de novo* transcript discovery or isoforms expression [284].
- It is worth considering that sequencing capacity depends on the device features; number of cycles, number of lanes per flow cell, number of reads per flow cell…

3.2.2. Sequencing, at a glance (Illumina)

During library preparation, cDNA fragments are tagged by adapters. DNA fragments are then hybridized to the flow cell. Each fragment is massively and clonally amplified by "bridge amplification", forming clusters of double-stranded DNA. Sequencing primers are annealed. The sequencing of the fragment ends is based on fluorophore-labelled dNTP with reversible terminator elements that will become incorporated and excited by a laser. Images are taken after each round of nucleotide incorporation and bases are identified based on the recorded excitation spectra (base calls) ***(Figure 8)***.

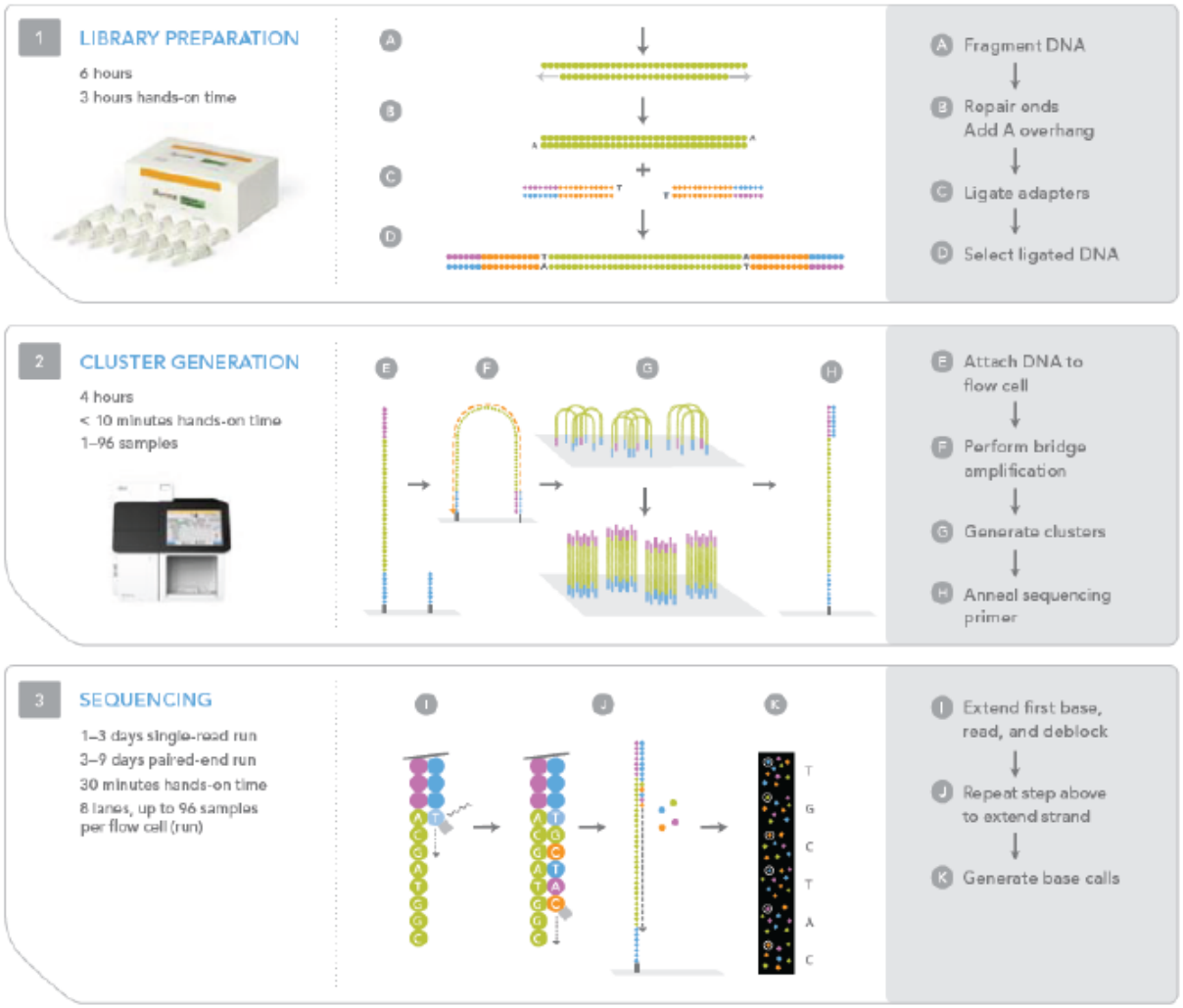


Figure 8. Illumina sequencing workflow using the sequencing by synthesis method adapted from Illumina.

3.2.3. RNA-seq data analysis workflow

Briefly, primary sequencing data, called raw reads are subjected to quality control (QC) to determine if a resequencing step is needed. After QC, raw reads are preprocessed and filtered into clean reads which will be aligned to the reference sequences. Various downstream analyses can be preceded including gene expression and deep analysis based on gene expression. We now describe in more detail the different themes of RNA-seq bioinformatics workflow:

3.2.3.1. Quality control of raw data

Raw reads are most commonly stored as FASTQC format. The latter displays 4 lines per read; the first line begins with a "@" character followed by a sequence identifier, the second is the sequence, the third line begins with a "+" character and can be followed by a sequence identifier, and the fourth line encodes the quality values of the sequence. Quality scores for each base of the sequence or base call quality scores are scores attributed to each base deducted from the images acquired during sequencing and represents the error probability of base calling. The score is called Phred score, Q, which is proportional to the probability p that a base call is incorrect, where Q = -10*log10 (p) [294]. For example, a Phred score of 10 corresponds to one error in every ten base calls (Q = -10*log10 (0.1)), or 90% accuracy; a Phred score of 20 corresponds to one error in every 100 base calls (p=0.01), or 99% accuracy.

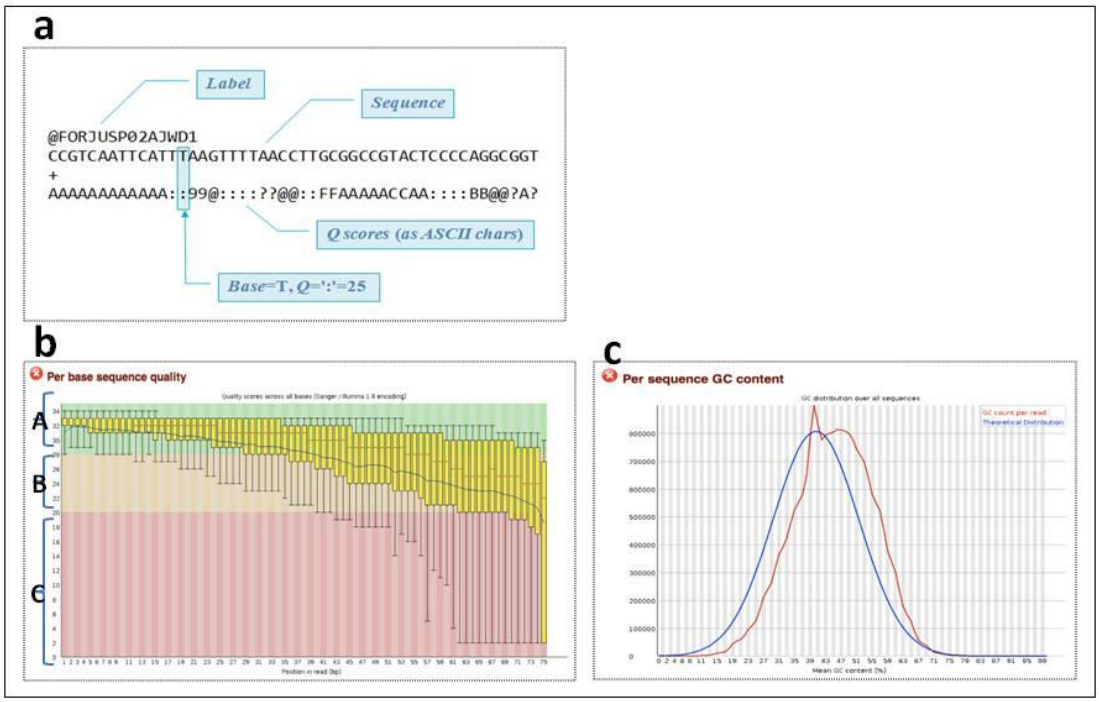


Figure 9. FASTQC format and reports; **a:** FASTQ read format [295]; b,c: Examples of FastQCreports [296].

Different sequencing platforms encoded Phred scores as ASCII characters to assign each base a unique score identifier. Particularly, several software packages perform the quality control such as FastQC, FastX and PRINSEQ. FastQC reads are reported as histograms per position base quality with a traffic light warning system ***(Figure 9b)***, good sequence quality (green**)** ***(Figure 9b, A)***, reasonable sequence quality (orange) ***(Figure 9b,B)*** or poor sequence quality (red) ***(Figure 9b, C)***, and as graphs per sequence content making it relatively easy to interpret results [297].

3.2.3.2. Preprocessing

Most of the times raw reads are considered as "dirty" reads owing to presence of the sequence of adaptors, and sometimes high content of unknown bases sequence specific bias (GC bias), sequence contamination and low quality reads. Therefore, they need to be removed before downstream analysis to decrease data noise. Accordingly, a preprocessing step is required to proceed. While filtering removes the entire read, trimming removes only the bad quality bases ***(Figure 10)***.

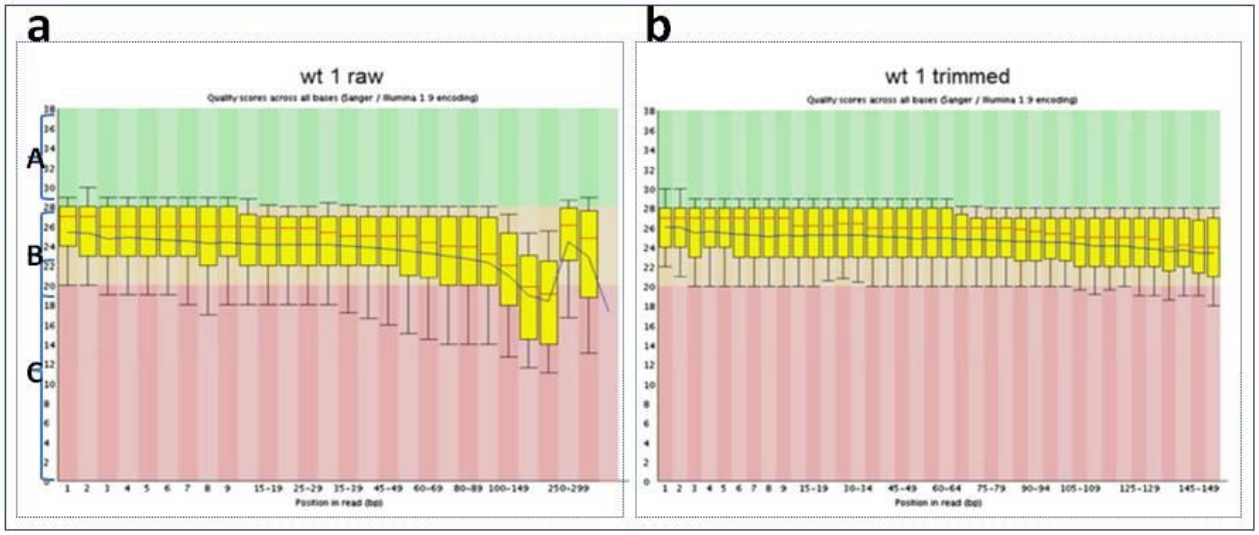


Figure 10. Quality control checks a: before quality trimming and b: after quality trimming [298].

Filtering steps are as follows: i) Remove reads with adaptors; ii) Remove reads in which unknown bases are more than 10%; iii) Remove low quality reads (the percentage of low quality bases is over 50% in a read). It should be considered that in case of paired end data, if a read is removed, its pair has to be removed as well.

After filtering, the remaining reads are called "clean reads" and are stored as FASTQ format [299]. It is important to note that to date there is no consensus about any base quality threshold to be used for reads trimming. Furthermore, it is worth considering that, aggressive trimming has large impact on RNA-seq gene expression estimates [300]. Once the trimmers

have been used, it is best to rerun the data through FastQC to check the resulting data. Numerous software packages for trimming/filtering are available such as, for example, FastX, PRINSEQ and Trimmomatic.

3.2.3.3. Alignment (mapping)

The purpose for alignment is to find out the origin of the read. However, several challenges will need to be addressed and overcome such as variants, sequencing errors and repetitive sequences. Therefore, three RNA-seq mapping strategies exist [301] ***(Figure 11)***:

- **De novo assembly:** this approach is used in case of absence of a reference genome. It involves the following: i) Finding overlaps between reads, ii) Assembling overlaps into contigs, iii) Assembling contigs into scaffolds.
- **Align to transcriptome:** use known and/or predicted gene or transcripts models to map the reads. It is used in case of short reads (<50bp)
- **Align to reference genome:** align directly sequenced reads to a reference genome.

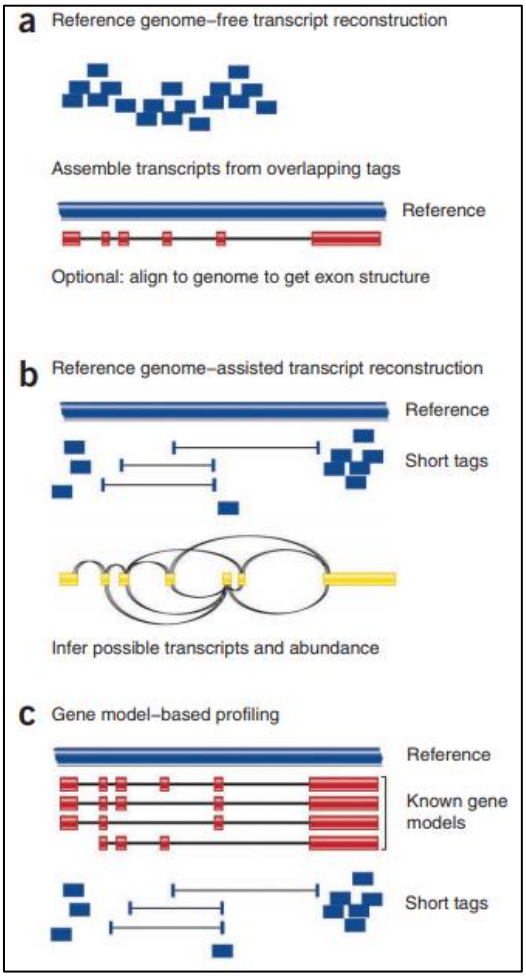


Figure 11. Three RNA mapping strategies: a: De novo assembly; b: Align to reference genome; c: Align to transcriptome [301].

Nevertheless, the reads will map to genome non-contiguously owing to the presence of introns in the sequence. Accordingly, it is important to perform subsequently spliced alignments. This approach is based on a computational algorithm to find multi-exon genes, detect unannotated exons or novel transcripts, and identify insertions and deletions. Briefly, initial read alignments are analyzed to discover exon junctions, and then the latter will serve as a guide for the final alignment.

It is worth noting that each strategy involves different alignment and assembly tools. Mainly 10 aligner tools are available [302] for mapping RNA-seq data, exist, Bowtie/Tophat and STAR being the major aligners used in publications-data. The afore-mentioned programs have splice-aware aligners to perform spliced alignments [303]. We will show below the mode of operation of two computational tools for alignment; Bowtie and Tophat versions 2.

First off, Bowtie is a fast tool, owed to the indexing of a reference genome that can perform short reads alignments.

Secondly, Tophat is a relatively fast spliced aligner that will scrutinize the unmapped, unannotated sequences.

In brief, we will detail the tophat2 pipeline workflow from Trapnell et al. (2010) [304]. Reads are primarily aligned to transcriptome, the unmapped reads will then be mapped to the genome. Using the resultant unmapped reads, Tophat will attempt to find novel splice sitesbased on known junction signals (GT-AG, GC-AG, and AT-AC). Reads are split into segments (25bp) then aligned against the genome, these small segments mapping will aid to identify potential splice sites. When the distance between the mapped positions of the left and right segments of a read are longer than the length of the read, it means that an unmapped intron or intergenic region exists and these segments are potential splice sites. Subsequently, the identified genomics sequences flanking a splice site are concatenated and the unmapped segments are aligned to them. The mapped segments against the genome (genome index) and the flanking sequences (junction flanking index) are stitched together to generate whole read alignments. Finally, the reads are re-aligned to exons alone ***(Figure 12)***.

To note that, Bowtie and Tophat are tools in the Tuxedo suite package ***(Figure 13)***.

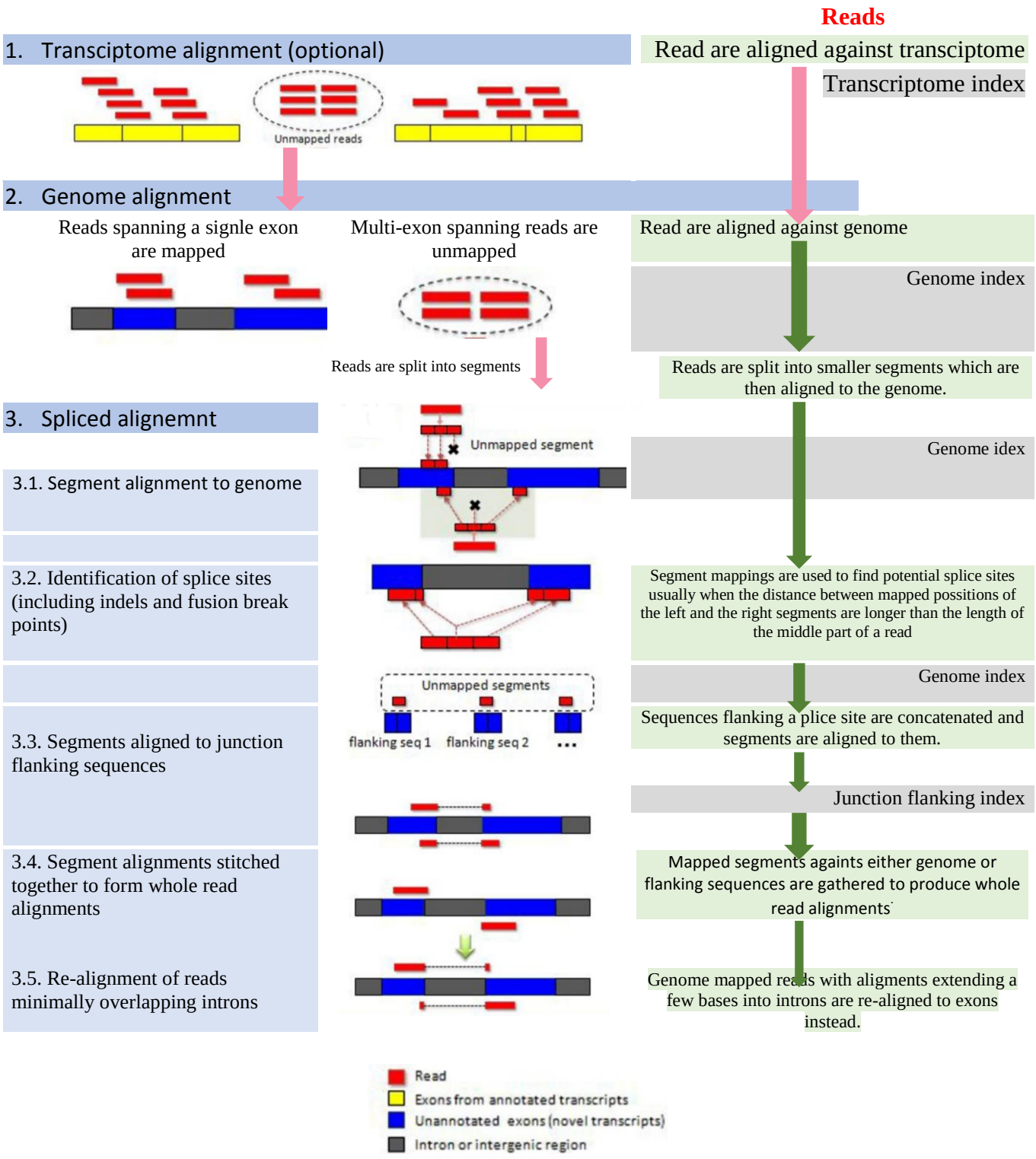


Figure 12. An overview of cufflinks workflow [305].

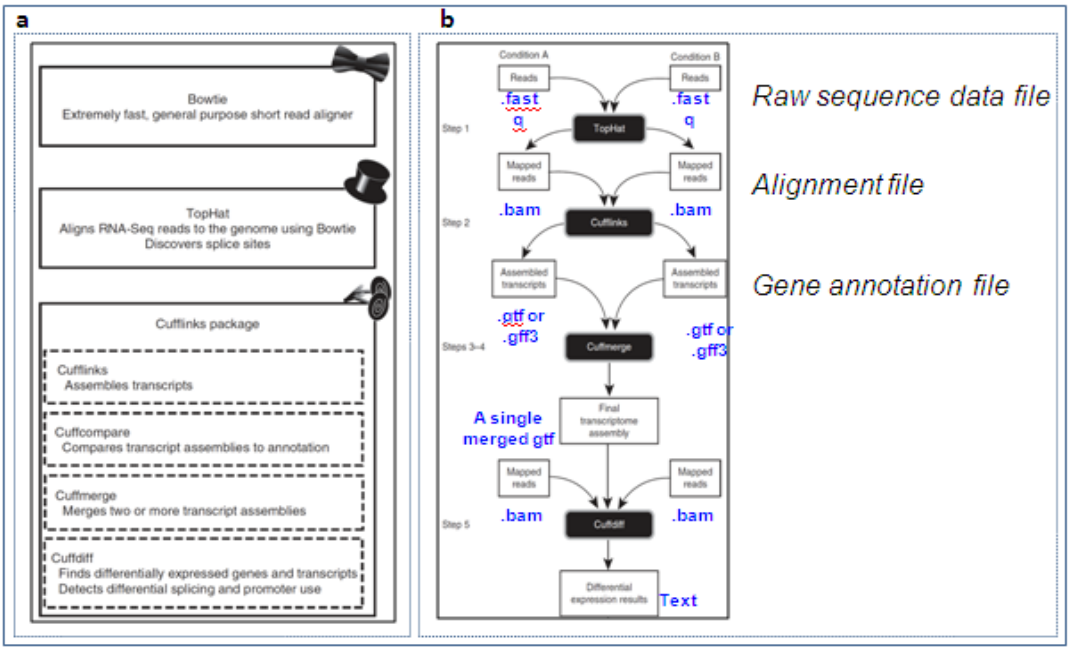


Figure 13. a: Tuxedo software components; b: Tuxedo protocol for differential gene expression [289].

It is highly desirable to consider that a Phred score is determined to report mapping quality. The latter describes the confidence in reads mapping position or the probability that the alignment location is wrong. Hence, this depends on the uniqueness of the mapped read, the length of the alignment and the number of mismatches. The read alignment data are generated in a SAM (Sequence Alignment/Map) file format or BAM format which is a binary form of SAM. The BAM file details the read name, position, mapping quality, and the gaps or mismatches such as insertions, deletions and introns, known as CIGAR string (Concise Idiosyncratic Gapped Alignment Report CIGAR indicates the necessary operations to map the read to the referencesequence at that particular locus) [306].

3.2.3.4. Quality control for aligned reads

Before downstream analysis several properties should be assessed:

- The saturation of sequencing depth measures the depth while interrogating the transcriptome. Particularly, we can determine the number of reads needed to detect a saturated number of transcripts. More so, it can assess the complexity of the sample by showing the number of transcripts expressed in a specific sample. It is illustrated by a saturation curve. Thus, an ideal saturation curve shows uniform curves where all samplesare sequenced to the same depth. Conversely, any other pattern may be attributed to biological differences, errors during library construction or sequencing [307] ***(Figures 14a,14b)***.

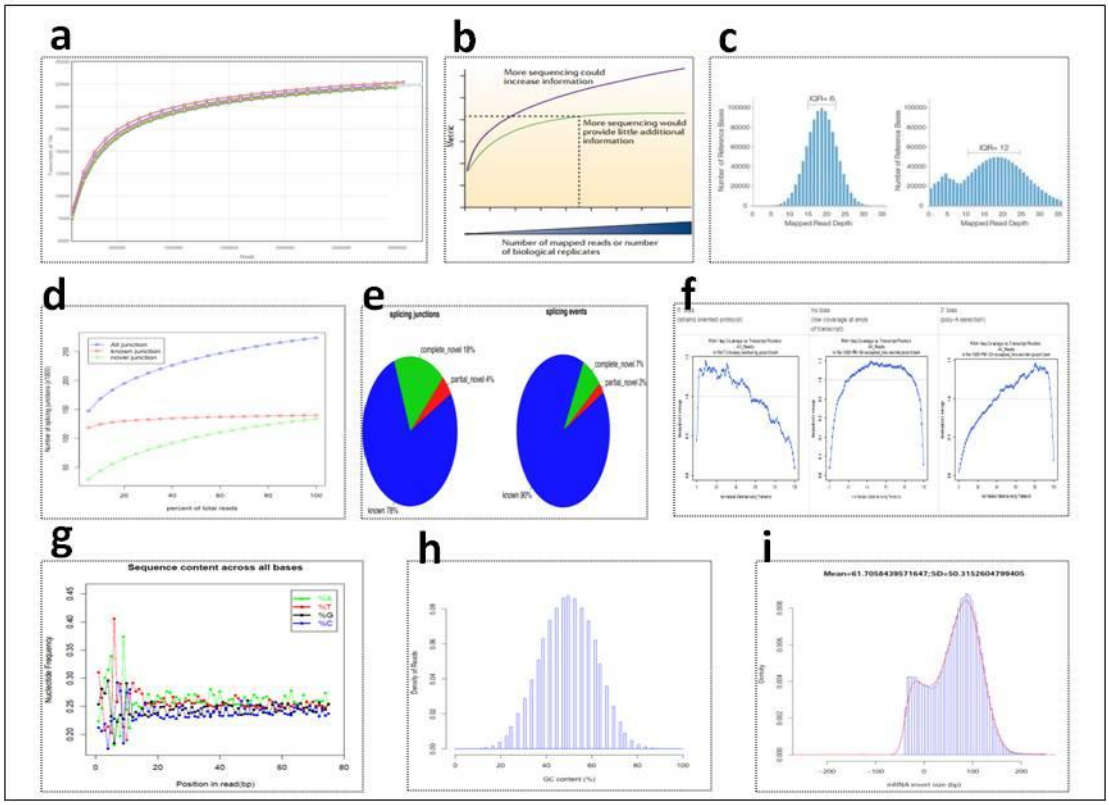


Figure 14. Post alignment quality control representations.

a;b: saturation of sequencing depth (a:[308], b:[307]); c:coverage uniformity histograms [295]); d,e: junctions coverage; f: read body coverage [309]; g:sequence content; h: distribution of read GC content; i: inner distance between read pairs (insert size) (d, e, g, g, i:[310]).

- Reads distribution between different genomic features such as exonic regions, intronic regions, coding 3'UTR and 5'UTR. This can determine the presence of any bias. These components are presented in a QC table.
- Coverage uniformity shows the read coverage uniformity along transcripts.
- It is illustrated using a plot that will take the form of a Poisson-like distribution with a small standard deviation. This distribution is valid under the assumption that reads are randomly distributed across the genome and that the ability to detect true overlaps between reads is constant within a sequencing run [311].

Histograms of ***Figure 14c*** show a good (left) and poor (right) sequencing coverage. Sometimes an IQR (Inter-Qartile Range) value can be represented on the scheme which is a measure of statistical variability. However, a low IQR reveals uniform sequence coverage. Moreover, junctions expression saturation can also be determined ***(Figures 14d, 14e)***. A series of curves can illustrate the junctions coverage or in other words the number of

junctions, either known or novel ones, that can be detected until reaching a plateau where deeper sequencing will not likely to detect additional junctions.

- Presence of 3' and 5'biais could be revealed by a lack of coverage at a specific end. For example, polyA capture and polyT priming can cause 3' bias [312] ***(Figure 14f)***.
- Sequence-specific bias due to a deviation in certain nucleotide content. For example, random primers can introduce such bias [312] ***(Figure 14g)***.
- Distribution of read GC content (%)[312] ***(Figure 14h)***.
- Distribution of observed inner distances in fragments, and find if it is consistent with the library size selection [313] ***(Figure 14i)***.

Several quality control programs are available; we mention RseQC, RNA-seqQC and Picard's CollectRnaSeqMetrics.

3.2.3.5. Reads assembly and quantification

In order to explore the whole transcriptome, we need to reconstruct the full-length transcripts by transcriptome assembly excluding small RNAs. However, transcriptome assembly is challenging due to discrepancy in the sequencing depth of transcripts, the presence of data from one strand only and the incidence of transcripts variants of isoforms. However, various transcriptome assemblers exist to reconstruct the transcriptome, in particular:

- **A reference based strategy or *ab initio* assembly:** after reads alignment using a splice-are aligner, overlapping reads are clustered to build a graph representing all possible isoforms. It is noteworthy that this strategy is highly sensitive for transcripts assembly.
- **The *de novo* transcriptome assembly:** this strategy does not use a reference genome but finds overlaps between short reads and assemble them. Therefore, it can detect novel and trans-spliced transcripts.
- **Combined strategy:** The above-mentioned strategies can merge together the 2 complementary strategies in order to get a more comprehensive transcriptome. Thus the data generated from the first strategy could also serve as input for the subsequent one.

Accordingly, several QC metrics for the sequencing assembly data should be assessed, such as: percentages for accuracy, completeness, contiguity, chimerism and variant resolution. To expand this topic, further details are presented in a review by Martin et al. (2011)[314].

The choice of the strategy depends on several factors, including the presence or completeness of a reference genome, the accessibility of sequencing resources, the type of

data generated and the main goal of the study [314]. For example, a study of the rice genome used Cufflinks (Bowtie and Tophat) by mapping reads against the whole reference genome, then the reads that did not align to the genome but that were mapped to these potential junctions. This led to the discovery of 649 genes that were missing from the rice annotation but that were found to be differently expressed in response to salinity stress [315].

In another study, Twine et al. (2011) [316] hypothesized that alternative splicing was involved in Alzeihmer pathogenesis. Thus, the transcriptome was assembled using a reference based assembler and discovered two genes with alternative start sites and splicing patterns that may help to explain the progression of Alzheimer's disease.

The obtained reference transcriptome is generated in GTF format (gene transfer format). The latter file delineates the locations of the genes [280]. It should be noted that a GTF file (gene transfer format) is a GFF (gene feature format) that stores genomic features, e.g. genomic intervals of genes and gene structure ***(Figure 15b)***.

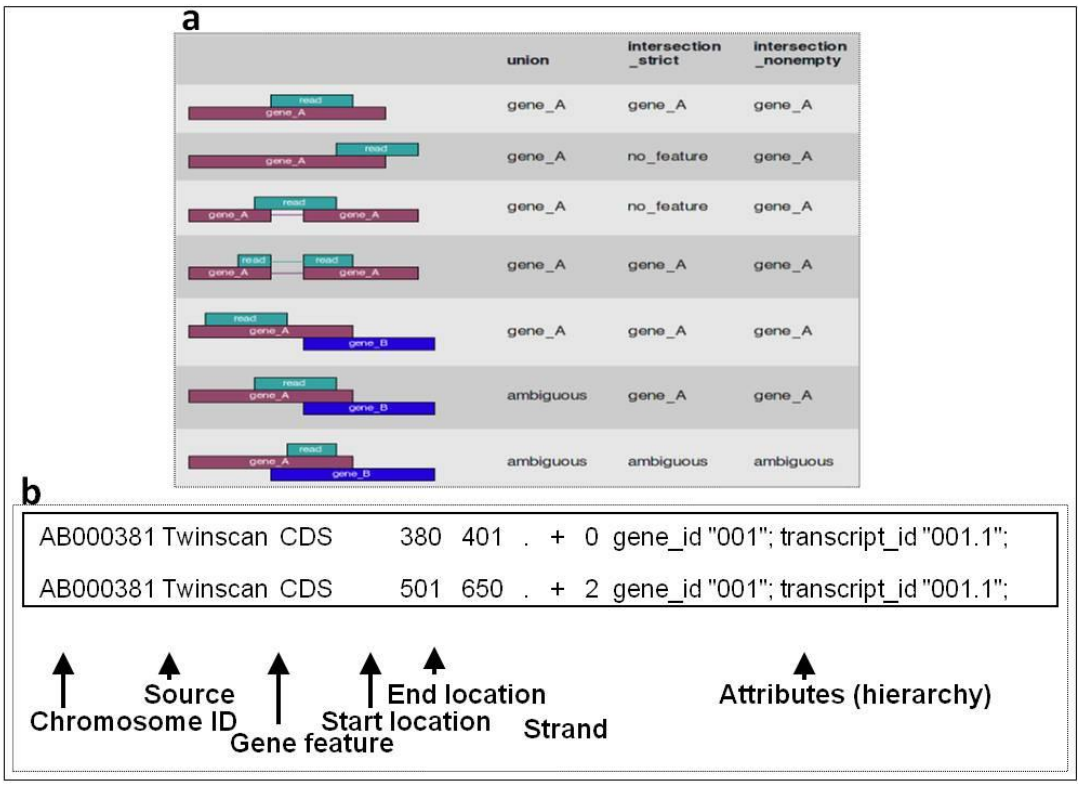


Figure 15. a: HTseq different modes for reads quantification [317]; b: Gene transfer format (GTF) features [318].

Hence, to perform differential gene expression between different conditions or samples, primarily gene-based reads should be counted [289]. Several software for counting reads are available namely HTSeq, Cufflinks (component of the Tuxedo suite package) ***(Figure 13b)***.

Several options for differential gene expression could be used depending on the availability of a reference genome or transcriptome ***(Figure 17)***.

The quantification can be performed on reads that map genes, exons, or transcripts. In this manner, Cufflinks counts transcripts, and HTSeq counts reads per genes.

In general, these software take as input BAM files derived from the alignment step as well as a GTF file. HTSeq can handle the gene reads features using different modes **(Figure 15a)**. By default, HTSeq uses the union mode matrix to execute the counting of reads. In particular, ambiguous positions of reads and multi-mapping reads are not counted [319].

Cufflinks is used to count transcripts and isoforms. As shown in ***Figure 16b***, this software uses a deBruijn graph approach to assign reads to a given isoform. Along these lines, we will show one simple deBruijn graph-based transcript assembly to illustrate clearly the algorithm for counting isoforms. Briefly, read sequences are split into subsequence k-mers of smaller length from the reads. A graph is constructed using unique k-mers as the nodes and overlapping k-mers connected by edges. The transcripts are assembled by traversing the paths in the graph [320] ***(Figure 16)***.

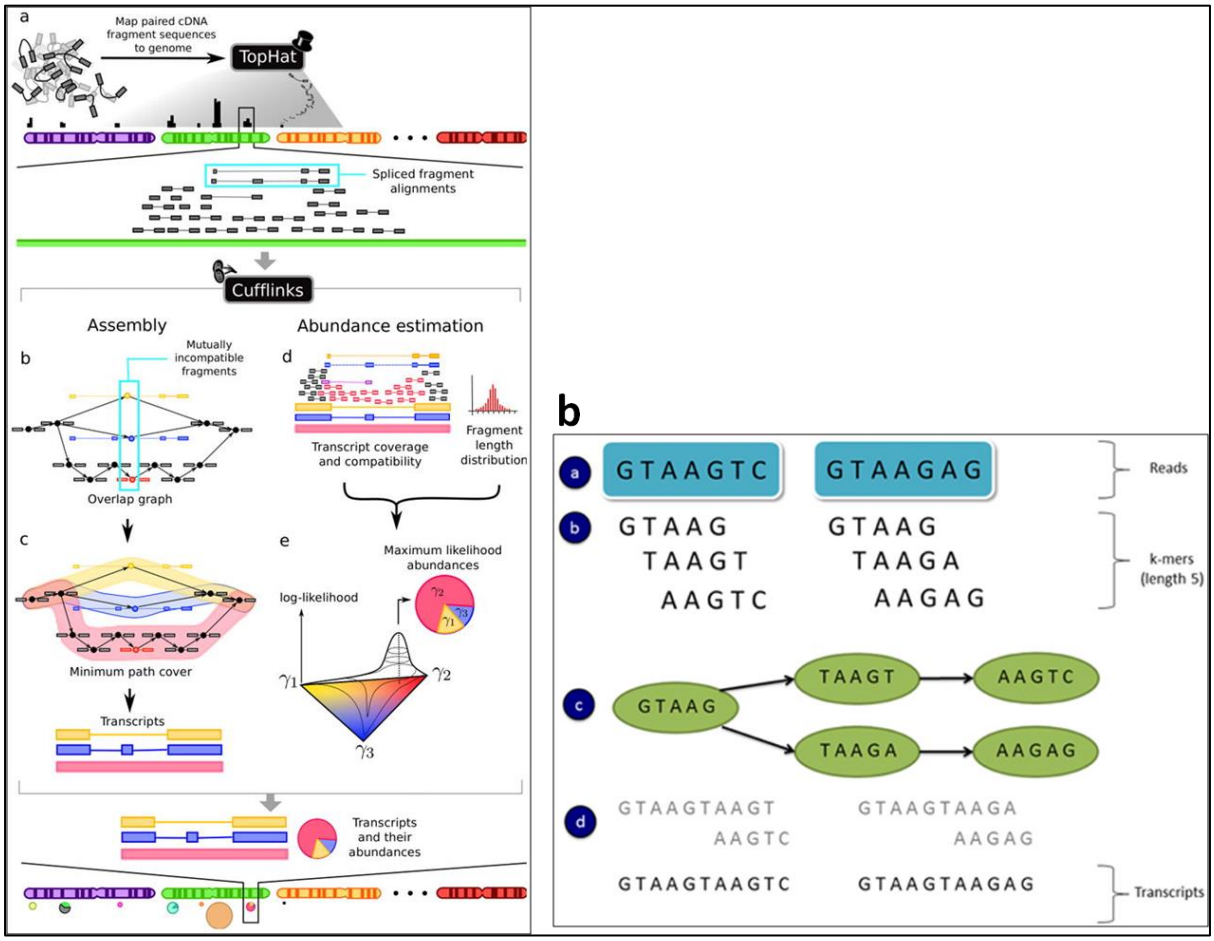


Figure 16. Overview of Cufflinks [304]; b: Scheme of a simple deBruijn graph-based transcript assembly [320].

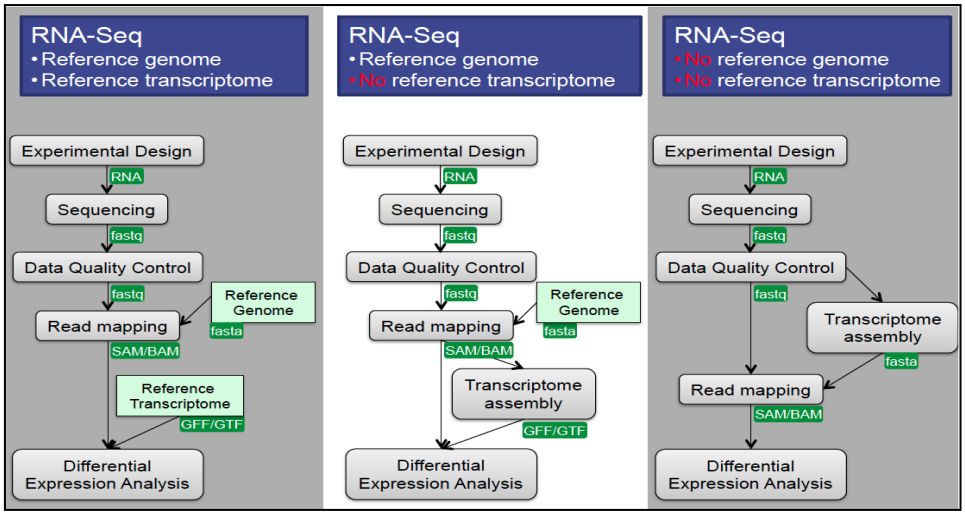


Figure 17. Options for differential gene expression (DGE) analysis [318].

3.2.3.6. Differential expression analysis: normalizing and transforming read counts

Finding genes that are differentially expressed between two conditions can be achieved by calculating the fraction of reads assigned to each gene compared to the total number of reads. Nevertheless, the number of sequenced reads of a gene depends on its expression level, length, gene GC content, the sequencing depth and the expression of all genes in the sample. Thus, sometimes variations in the number of sequenced reads may be due to non-biological factors such as library size or RNA composition bias caused by sampling. Accordingly, normalization is needed to eliminate effects not associated with the biological differences of interest. Several methods of normalization exist to adjust the number of reads using different formulas and find a normalization factor [321]; we will briefly expose some of them:

- **Total count:** All read counts are divided by the total number of reads (library size) and multiplied by the mean total count across all samples.
- **RPKM (reads per kilobase of exons per million mapped reads):** is a formula that takes into consideration the length of the gene and the total number of mapped reads.
- **FPKM (fragments per kilobase of exons per million mapped reads):** same as RPKM, but for paired-end reads. It attempts to normalize for gene size and library depth.
- **Trimmed Mean of M-values (TMM):** Samples that have the closest average expressions to mean of all samples is considered as reference samples, and all others are test samples. Thus, the scaling factor of test samples is calculated based on weighted mean (weighted by estimated asymptotic variance to account for lower variance of genes with larger counts) of log ratios between the test and reference, from

a gene set trimming most/lowest expressed genes and genes with highest/lowest log ratios (removal of upper and lower 30%).

- **DESeq's size factor:** The ratio of read count over geometric mean of read counts across all samples. The scaling factor is the median of these ratios. By definition, the geometric mean is the nth root of the product of n numbers. However, unlike an arithmetic mean, the geometric mean tends to dampen the effect of very high or low values, which might bias the mean.

It is worth noting that, RPKM (FPKM) and total count normalization should be avoided in the context of differential expression analysis, but they are only needed to report expression values within the same sample, where we must normalize to gene length or gene GC content. Conversely, DESeq and TMM methods are of choice in DE analysis owing to their low false positive rate and high power.

In order to assess measurement precision in DE analysis we need raw counts values. Downstream analysis requires transformation of read counts to the log scale. For example, $\log_2$ transformation permits an easier visualization and interpretation of data when plotted [321].

3.2.3.7. Variability within the experiment: data exploration

The main variability within the experiment is expected to come from biological differences between the samples. Accordingly, overviewing the similarities and dissimilarities between samples allow to detect any confounding factor that should be taken into consideration in the analysis, and to remove any sample outliers. In order to visualize sample-to-sample distances we can perform: (i) a hierarchical clustering of the whole sample set; (ii) a principle component analysis (PCA); (iii) a multidimentional scaling (MDS), depending on the software used. Although, we will not enter in details of the manipulation of the data and analysis but we will present the straightforward interpretation of these representations.

PCA is a mathematical algorithm that reduces the dimensionality of the data while retaining most of the variation in the data set. The data points (the samples) are projected onto the 2D plane such that they spread out in the two directions that explain most of the differences. The x-axis labeled *PC1* is the direction that separates the whole samples the most. The y-axis labeled *PC2* is a direction that separates the whole samples the second most. The percentage of the total variance that is contained in each direction is printed in the axis label. We expect to separate samples from the different biological conditions, meaning that

the biological variability is the main source of variance in the data [322]. After plotting data we can see how the groups are clearly separated ***(Figures 18a, 18b)***.

Likewise MDS shows samples distance, however a sample separated away from the other data points on the plot can be considered as outlier ***(Figure 18c)***. We can see in ***Figure 18c*** that the C and T groups are well separated, except T2 that is an outlier. More so, clustering dendogram also illustrates samples similarities and differences. Samples that are most similar occupy closer positions in the tree, while samples that are less similar are separated by larger numbers of branch points. This representation can show outliers and the batch effect. It will be wise to continue the subsequent analysis steps leaving out the confounding factors ***(Figure 18d)*** [323]. The below clustering dendogram shows samples that are most similar occupy closer positions in the tree, while samples that are less similar are separated by larger numbers of branch points. Particularly, samples C1 and T1 are more similar to each other than to the other samples, and there may be batch effects affecting samples C1 and T1.

Additionally, we can visualize the distances between samples in a heatmap ***(Figure 18e)***. Samples similar to each other share the same blocks color.

It is critical to consider that PCA and clustering should be done on normalized and preferably transformed read counts, so that the high variability of low read counts does not occlude potentially informative data trends.

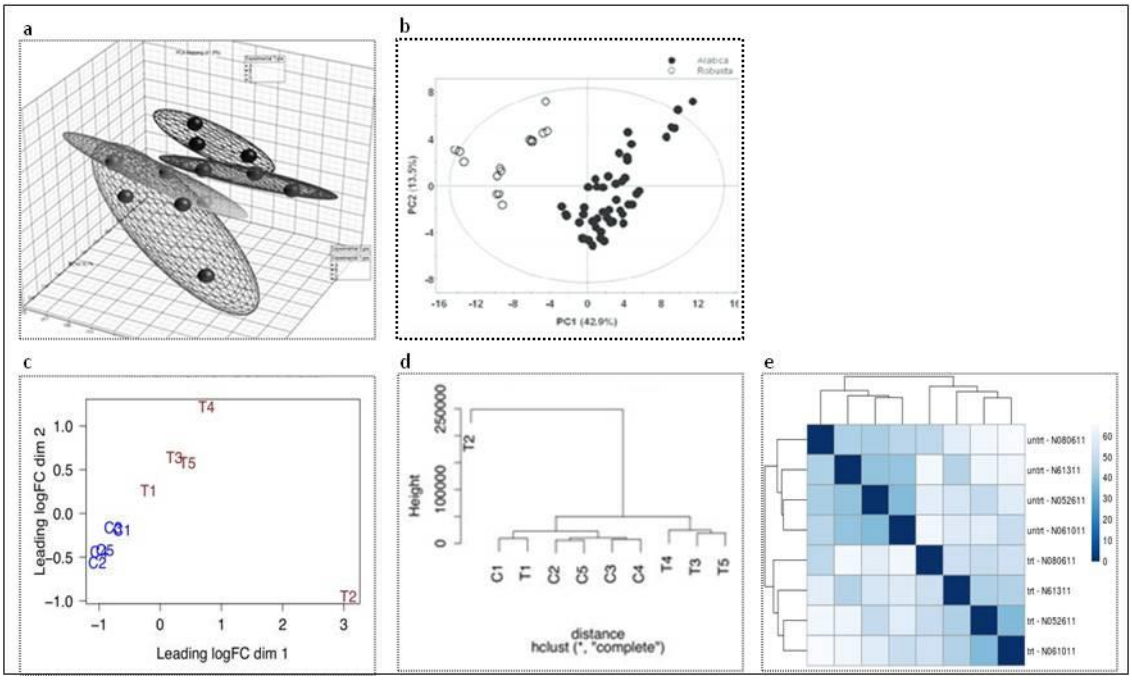


Figure 18. Illustration of variability within the experiment; a: principle component analysis (PCA) (a: [324], b: [325]); c: a multidimentional scaling (MDS) [325]; d: a cluster dendogram [323]; e: heatmap [326].

3.2.3.8. Differential gene expression analysis

DGE analysis workflow requires the normalization of raw count data, followed by dispersion estimation, then estimate the magnitude of DE or the log fold change and finally the estimation of the significance of the difference and correct for multiple testing. The read values are first normalized using the different methods discussed above. RNA-seq experiments lack replicates, thus we need to estimate within-group variance called dispersion estimation [327]. Therefore, we need to compensate this lack by pooling information across genes with similar expression values presuming that they have similar dispersion then shrink a given gene's variance towards the fitted values. Several tools can perform this step namely edger, DESeq2, Cuffdiff… RNA-seq experiments generate a number of counts that map a gene, thus it cannot be modeled as a normal distribution but as a Poisson distribution since we have data values. The latter assumes that the means and variances are equal, but in RNA-seq data this isn't the case.

For example, lowly expressed genes have much higher variance than highly expressed genes. To account for the replicates variability, we use the negative-binomial model that deals with discrete values and considers means and variances unequal [328]. All the above-mentioned programs can perform statistical testing for expression levels comparisons between groups. Further details concerning the different software packages for detecting differential expression was reviewed by Seyednasrollahet al. (2013) [329]. The significance of the difference will be then estimated. However, we would expect that there is a probability of having significant genes being due to random chance. To reduce this high risk false-positive results, p-values need to be corrected for multiple testing. Several corrections exist, the most common one is the Benjamini-Hochberg correction (BH) [330]. To decrease the severity of correction, a filtering threshold should be used to remove genes that have low chance for significant differential expression. The adjusted p-value is named FDR (false discovery rate).

3.2.3.9. Exploratory plots following DGE analysis

- **Histograms of p-values**: frequencies of p-values of all genes can be plotted on a histogram.
- **MA plot**: It shows the relationship between the expression change between conditions (log ratios, M), and the average expression strength of the genes (average mean, A). Moreover, it illustrates differential gene expression: genes that pass the significance threshold (adjusted p-value <0.05) are colored in red (in the grayscale figure shown below, the darker dots).

- **Heatmaps:** Illustrate the expression values across the individual samples ***(Figure 19).***

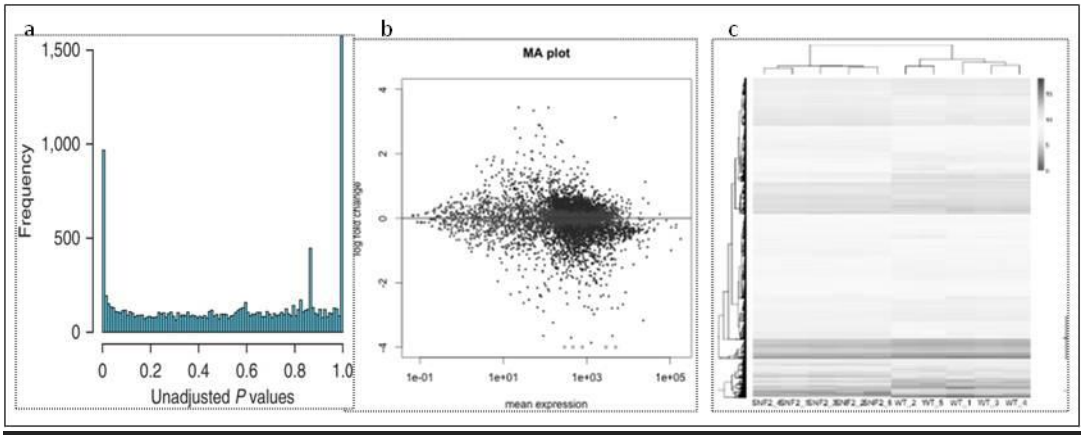


Figure 19. Exploratory plots following DGE analysis; **a:** Histogram of p-values [331] ; b: MA plot ; c: Heatmap (b,c: [326]).

3.3. Experimental work for sperm transcriptomic analysis

3.3.1 Patient population

We enrolled a total of 80 subjects in the study: 20 normozoospermic fertile men, 20 patients diagnosed with oligozoospermia (sperm count < 15million spermatozoa/ml), 20 patients diagnosed with asthenozoospermia (sperm motility < 40%), and 20 patients diagnosed with teratozoospermia (sperm normal forms < 4%). The inclusion criteria for the infertile patients were as follows: all subjects attended the male infertility clinic for fertility issues. All of these men were evaluated for proven male-factor infertility as assessed the male infertility specialist. None of them had female factor infertility in their partners. Our exclusion criteria were: azoospermia, incomplete semen analysis results.

The fertile group were males, 20 – 44 years old, of proven fertility (having established a successful pregnancy in the past) and whose semen samples fulfilled the criteria established by the WHO 2010 guidelines for semen analysis i.e. normal semen parameters [83] ***(Supplementary Tables S1, S2, S3)***.

3.3.2. Semen analysis

Semen samples were collected by masturbation after 2–3 days of sexual abstinence. After liquefaction, a complete semen analysis was performed to evaluate the sperm parameters according to the World Health Organization (WHO) guidelines. Sperm concentration and percentage motility analysis were done using a Makler counting chamber

(Sefi Medical Instruments, USA). The samples were also visually inspected for somatic cell contamination.

For assessment of sperm morphology, thin smears of the well-mixed ejaculated semen were prepared in duplicate by placing 2–5 μL (depending on the sperm concentration) on clean slides. After air drying, the slides were stained using Wright-Giemsa stain and graded on the basis of the cutoff value established by WHO 2010 guidelines. A total of 100 spermatozoa were scored per slide using bright field illumination and an oil immersion objective with a total magnification of × 2000.

3.3.3. Sperm purification

The ejaculate samples were washed with sperm washing medium (Origio, Denmark) and the purity of the human sperm was >0.01% (less than 1 round cell per 10 000 sperm) based on phase-contrast microscopic observation. Sperm pellets were subject to an osmotic shock in distilled water, and then pelleted. The pellets were flash-frozen in liquid nitrogen then stored at -80°C. Each sample was thawed on ice the sperm pellets were subjected to RNA isolation using the method described below.

3.3.4. RNA preparation

Total RNA was isolated using previously described protocol by Goodrich [271]. RNeasy columns (#74134; Qiagen, The Netherlands) were used to purify total RNA from the cells, and this RNA was further concentrated using RNA MinElute columns (#74204; Qiagen, The Netherlands) according to the manufacturer's protocol. The RNA concentration and purity were determined using a NanoDrop ND-1000 spectrophotometer (Nano-drop Technologies, Wilmington, DE) and an Agilent 2100 Bioanalyzer (Agilent RNA 6000 Nano Kit). RNA samples were stored at -80 °C until further use. We obtained an RNA purity A260/A280 ~1.9 and a RIN ranging between 2.1 and 3.2. In addition, sperm RNA electrophoretic profiles are characterized by a large population of RNA <200 nt in length. These findings are comparable to those of other studies [274]. In fact, a lack of intact 18S and 28S rRNAs in sperm renders the traditional RIN conventions inappropriate for assessing sperm RNA quality. Despite the absence of the ribosomal RNA in spermatozoa, spectral analysis of RNA abundance by size is still useful to assess sperm RNA quality and purity as the presence of the ribosomal subunits 18S and 28S may indicate somatic cell RNA contamination ***(Supplementary Figure S1)***.

3.3.5. RNA-Seq library preparation

The total RNA was treated with Ribo-Zero™Magnetic Kit (Epicentre) to deplete rRNA. The retrieved RNA is fragmented by adding First Strand Master Mix (Invitrogen). First-strand cDNA is generated by First Strand Master Mix and Super Script II reverse transcription (Invitrogen) (Reaction condition：25℃ for 10 min；42℃ for 50 min；70℃ for 15 min). The product was purified (Agencourt RNAClean XP Beads, AGENCOURT), then Second Strand Master Mix and dATP, dGTP, dCTP, dUTP mix were added to synthesize the second-strand cDNA (16℃for 1h).Then the purified cDNA was combined with End Repair Mix, and incubated at 30℃ for 30min. After purification with beads, the A-Tailing Mix was added, then incubated at 37℃ for 30 min. The Adenylate 3'Ends DNA, Index Adapter and Ligation Mix were combined, then incubated at 30℃ for 10 min. After that, the Uracil-N-Glycosylase enzyme was added into the purified ligation product, and incubated at 37℃ for 10min. Several rounds of PCR amplification with PCR Primer Cocktail and PCR Master Mix were performed to enrich the cDNA fragments. Then the PCR products were purified with Ampure XP Beads.

3.3.6. Sequencing Libraries

We accurately quantified the cDNA library templates and then normalize all the samples at specific matched molarities and absolute total molarity. We applied a biologically averaged design [225-227] where we pooled libraries in groups of 10 per phenotype, and used statistical methods (e.g. DEGseq [332]) that only adequately model technical variability. The libraries were amplified on a cBot to generate the cluster on the flowcell (TruSeq SE Cluster Kit V3–cBot–HS，Illumina). The amplified flowcell was sequenced on the HiSeq 2000 System (TruSeq SBS KIT-HS V3，Illumina). The sequencing reads were aligned to the human genome build 19 (hg19) using Tophat v1.0.14 [333]. The aligned bams were analyzed using HTSeq [319] to estimate the counts per gene. The Differentially expressed genes (DEGs) were calculated using DESeq2 [327, 334] included in the SARTools package developed at PF2 - Institut Pasteur [335].

3.3.7. Bioinformatics

Clusters of low, and high-expressed transcriptome genes from the RNA sequencing were assessed for GO enrichment using ToppGene (http://toppgene.cchmc.org/) [336]. Enriched terms were considered significant with $P < 0.05$ after false discovery rate correction. Functional protein association networks were investigated in silico by using

STRING 9.05 online database of protein interactions (https://string-db.org/) and GeneMANIA (https://genemania.org/). To identify lncRNA-protein interaction networks, we used RNA-protein Association and Interaction Network's (RAIN) web interface (http://rth.dk/resources/rain/). For Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway mapping visualization within, the Pathview server (https://pathview.uncc.edu/)[337] was used.

Venn diagrams were plotted using the InteractiVenn tool (http://www.interactivenn.net) [338]. Heatmapper (http://www.heatmapper.ca/) was used to generate the heatmap [339].

3.4. Experimental work for sperm epigenetic analysis

3.4.1. Study participants

Participants in the study were recruited from men with known male factor infertility ("infertile subjects"; n=33) defined as the presence of abnormal WHO semen quality criteria (WHO 2010)[83] and the inability to conceive despite more than 12 months of unprotected intercourse, attending an infertility clinic. "Fertile controls" (n=23) were men who had proven fertility within the last12 months and normal semen parameters according to WHO criteria. The age of the study population ranged between 19 and 40.

Semen samples were collected by masturbation in a sterile wide-mouthed calibrated container after an abstinence period of 2-5 days. Semen analysis was performed according to World Health Organization guidelines to evaluate conventional sperm parameters: sperm count, motility, viability, and percentage normal morphology.

3.4.2. Sperm damage assessment

3.4.2.1. Measurement of reactive oxygen species production

A modified colorimetric Nitro Blue Tetrazolium (NBT) test was used to evaluate reactive oxygen species (ROS) production of both leukocytes and sperm cells within semen [340]. Briefly, sperm was incubated with NBT reagent at 37°C for 45 min. Following incubation the samples were washed and centrifuged at 500 g for 10 min in PBS to remove all residual NBT solution, leaving only a cell pellet. Formazan, a blue water insoluble crystal produced from the yellow water-soluble tetrazolium salt by the action of cellular superoxide anions, is then deposited inside the sperm and leukocytes. The amount of formazan crystal present within a cell is closely related to its production of free radicals [341]. The air-dried smears were stained with Wright stain and a total of 100 spermatozoa were scored under 100 × magnifications. Sperms were scored as follows: formazan occupying 50% or less of the cytoplasm (+) and more than 50% of the cytoplasm (++).

3.4.2.2. Sperm Chromatin dispersion (SCD) test

The SCD test is based on the principle that when sperm immersed in an agarose matrix on a slide, treated with an acid solution to denaturate DNA, and then lysed with a buffer to remove membranes and proteins, the result is the formation of nucleoids with a central core and a peripheral halo of dispersed DNA loops. The dispersion halos that are observed in sperm nuclei with non-fragmented DNA after the removal of nuclear proteins are either minimally present or not produced at all in sperm nuclei with fragmented DNA. The agarose matrix allows working with unfixed sperm on a slide in a suspension-like environment. Nucleoids are visualized with bright-field microscopy with Wright's staining solution. Here, we performed the SCD test as proposed and improved by Fernandez et al. (2005) [342] , using the Halosperm® kit (INDAS laboratories, Madrid, Spain).

Four dispersion patterns were defined: (i) sperm with large halo: the halo has a 2 times larger width than that of the sperm core, with a darker spot (sperm head) in the middle. (ii) Sperm with moderate halo: having a halo size between large and small halos. (iii) Sperm with small halo: a very small, clear film, that has a halo appearance, surrounds the sperm head. (iv) Sperm with no halo. Sperm DNA fragmentation percentage (SCD percentage) was calculated as the proportion of sperm with small and no halos, to the total sperm count per slide. We assessed two slides for every patient, and a total of 100 sperms were counted per slide.

3.4.3. Sperm chromatin abnormalities assessment

3.4.3.1. Sperm chromatin structure assessment

Chromatin structure was assessed using the toluidine blue (TB) method [343]. Briefly, dried smears were fixed with ethanol: acetone mixture. Smears are then hydrolyzed using HCl, then stained with toluidine blue. Two hundred randomly selected sperms per sample were examined under high magnification. Different optical densities are distinguished by the TB test. These correspond to the following visual TB colors: dark violet cells (TBDCs; abnormal chromatin structure) and two intermediate forms: light violet and dark blue (probably with less damaged chromatin structure) versus light blue cells (TBLCs; normal chromatin structure).

3.4.3.2. Histone retention assessment

Nuclear maturity is evaluated using the aniline blue staining which discriminates between lysine-rich histones and arginine/cysteine-rich protamines [344].

Concisely, smears are fixed in glutaraldehyde than stained in aniline blue solution. Sperm heads containing immature nuclear chromatin stain blue and those with mature nuclei

do not take up the stain. The percentage of spermatozoa stained with aniline blue is determined by counting 200 spermatozoa per slide under bright field microscopy.

3.4.4. Global epigenetic DNA analysis

3.4.4.1. DNA extraction

The sperm pellets prior to freezing were subject to an osmotic shock and and flash freeze shock in liquid nitrogen for cellular lysis. Each sample was thawed on ice and sperm DNA was subsequently extracted. Sperm DNA was extracted using a detergent-based lysis followed by an in-column purification using the QIAamp® DNA Mini Kit (#51304; Qiagen, The Netherlands) according to the manufacturer’s protocol. DNA yields and quality were determined using the Nanodrop 2000 Spectrophotometer (#E112352; Thermo Scientific, Somerset, NJ).

3.4.4.2. Global 5-mC, 5-hmC and 5-fC measurement

We used colorimetric methylated DNA, hydroxymethylated DNA, and 5-formylcytosine DNA quantification kits (abcam) for the quantification of 5-mC, 5-hmC and 5-fC respectively. This enzyme-linked immunosorbent assay (ELISA) kit has been established as a quick, effective, and inexpensive method for the relative quantification of global 5-mC, 5-hmC, and 5-fC levels in many tissues types from various species [277]. The assays were performed according to the manufacturer's recommendations. Briefly, after incubation with the input DNA, the wells were washed, and a capture antibody was applied to each well, after which the wells were again washed and detection antibody applied. Use of enhancer solution and development solution created a color change proportional to the quantity of 5-mC, 5-hmC and 5-fC, and the samples were read on an automated plate reader at 450 nm absorbance. The use of a formula provided to calculate the relative 5-mC, 5-hmC and 5-fC statuses.

3.4.5. Statistical analysis

With the SPSS version 23.0 software (SPSS Inc., Chicago, IL), differences between groups for the different parameters were assessed by the non-parametric Mann Whitney U test. $P \leq 0.05$ was considered statistically significant. Whereas the potential associations were evaluated using the Spearman Rank Order correlation test.

Conclusion

In this chapter we detailed the methods and experimental methodologies used to conduct the transcriptomic and the epigenetic analysis. In this manner, we provided a

comprehensive overview about RNA-seq data analysis then detailed the optimized protocols and methods used including RNA extraction, sequencing, bioinformatic analysis and epigenetic analysis. In the next section, we will present the results of our analysis of the RNA-seq data, and the assessment of the global epigenome in sperm from infertile men.

CHAPTER IV

RESULTS

In this chapter we will detail the repertoire of RNA transcripts of sperm from infertile men at different pathologic conditions (oligozoospermia, asthenozoospermia, teratozoospermia) using RNA-Seq. Furthermore, we performed a complete functional enrichment analysis to distill pathways implicated in infertility pathogenesis. Next, we attempted to identify gene-expression signatures that discriminate the different sperm pathologies. Finally, we studied the sperm global epigenetic landscape in infertile men.

4.1. Comprehensive transcriptomic analysis of human spermatozoa in male infertility

4.1.1. Sperm transcriptome in oligozoospermia

4.1.1.1. Total number of transcripts and their expression level

Sequencing averagely generated 57,548,548 raw sequencing reads and then 53,143,056 after filtering low quality reads. The average mapping ratio with reference gene is 60.64%, and the average genome mapping ratio is 92.18%.

We identified 10087 differentially expressed sequences, and of these 2823 genes were significantly down-regulated while 7264 sequences were significantly up-regulated in oligozoospermia compared to fertile men. It is noteworthy to mention that for this analysis, a BH p-value adjustment was performed [330, 345] and the criteria set were: the level of controlled false positive rate set to 0.05, abs(log2FC)>1, and DeSeq2 read counts≥1.

4.1.1.2. Nature of spermatozoa transcripts

Characterization of the pool of purified spermatozoa RNAs longer than 200 nucleotides (nt) showed that spermatozoa carry protein coding transcripts (90%), long intergenic noncoding RNAs – lincRNA (3%), antisense RNA including some long non coding RNAs (2%), and pseudogenes (4%).

4.1.1.3. Functions of spermatozoal transcripts

To capture a range of meaningful findings in the biological aspects of oligozoospermia, we performed functional enrichment analyses using ToppGene. ToppGene detects functional enrichment of a given gene list based on ontologies, pathways, phenotypes, etc. All statistical results were corrected for multiple testing using the Benjamini-Hochberg method.

To identify shared associated concepts between the genes in general, , we clustered the up- and down-regulated genes separately. During genes sorting, narrowed filters were employed; including the highly abundant transcripts to study gene functions for relevant biological concepts.

4.1.1.3.1. Down-regulated genes

First, we consider the analysis of the highly abundant genes down-regulated. The latter spermatozoal transcripts, identified based on the DeSeq2 normalized read counts >100, were related to 914 genes. These transcripts were functionally related to spermatogenesis and male gamete generation (FDR= 9.783E-9), gamete generation (FDR= 2.988E-6), fertilization (FDR= 1.506E-5), fusion of sperm to egg plasma membrane involved in single fertilization (FDR= 1.135E-4), spermatid differentiation (FDR= 7.391E-3), and flagellated sperm motility (FDR= 1.603E-2). Hence, the cellular components of the highly abundant transcripts are cilium (FDR= 3.745E-6), sperm part (FDR= 1.047E-4), centrosome (FDR= 2.070E-3), microtubule organizing center (FDR= 3.373E-3), kinetochore (FDR=1.071E-2), cohesion complex (FDR=2.001E-2), and chromatoid body (FDR= 2.776E-2).

The major pathways affected are collagen formation (FDR= 2.282E-3), fertilization (FDR= 2.891E-3), extracellular matrix organization (FDR=2.790E-2), and sperm: oocyte membrane binding (FDR= 4.799E-2). Interestingly, direct comparison of human and mouse phenotypes allowed rapid recognition of disease causal genes. When employing mouse phenotype data, we detected several phenotypes related to spermatogenesis, namely abnormal spermatogenesis (p=3.344E-3), male infertility (p= 3.344E-3), abnormal fertilization (p= 4.043E-3), arrest of spermatogenesis (p= 6.471E-3), infertility (FDR= 9.953E-3), and increased male germ cell apoptosis (FDR= 4.886E-2).

After scrutinizing the complete set of genes associated with the male infertility phenotype in mouse ***(Supplementary Table S4)***, we performed a network analysis and found that 49.24% shared the same co-expression network. The protein–protein interaction (PPI) pairs of proteins that encoded by genes were retrieved using the Search Tool for the Retrieval of Interacting Genes/Proteins (STRING, http://string-db.org/). This interactive network to predict potential interactions between these genes showed a PPI enrichment p-value= 2.64e-

11. This indicates that these proteins are at least partially biologically connected ***(Figure 20).*** That been said, the majority of these genes are involved in the same or related biological pathway. In fact, this set of genes was found associated with biological processes, namely spermatogenesis (FDR= 4.33e-21; 23 hits out of 50), meiotic cell cycle (FDR=1.45e-12, 12 genes), synaptonemal complex assembly (FDR=0.00216, 3 genes), piRNA metabolic process (FDR=0.0095, 3 genes), sister chromatid cohesion (FDR=0.0103, 3 genes), and male gonad development (FDR=0.0349, 4 genes) ***(Table 8)***.

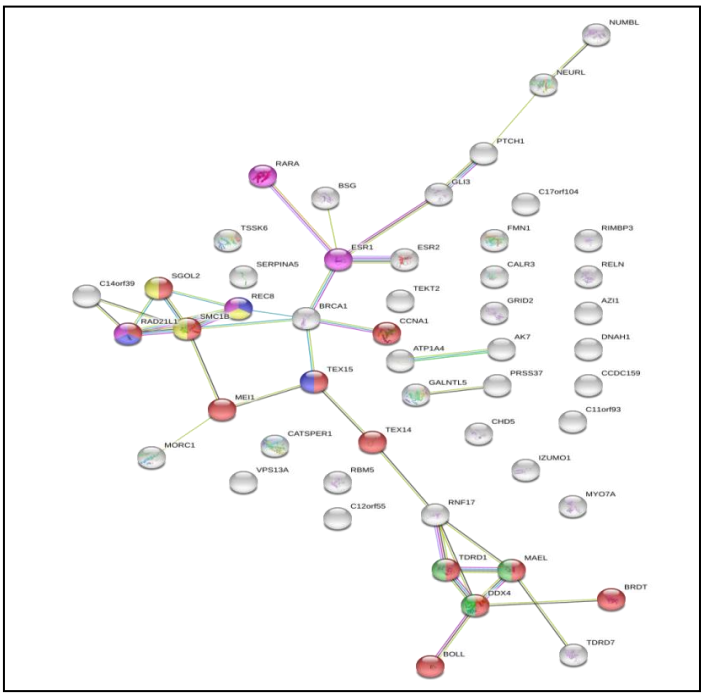


Figure 20. Protein interaction map. Map was prepared using the STRING web tool (https://string-db.org) for the proteins encoded by the set of genes associated with the male infertility phenotype. Colored lines denote interactions and network nodes represent proteins.

Table 8. Functional enrichments of the protein-protein interaction (PPI) network for the mouse orthologs associated with male infertility phenotype.

Biological process (GO)			
Pathway description	*Count in gene set*	*False discovery rate*	*Nodes legend*
Meotic cell cycle	12	1.45e-12	
Synaptonemal complex assembly	3	0.00216	
piRNA metabolic process	3	0.0095	
Sister chromatid cohesion	3	0.0103	
Male gonad development	4	0.0349	

These candidate genes were prioritized further, based on functional similarity to the abundant down-regulated transcripts to identify a subset of genes that were down-regulated. Notably, we identified 11 mouse orthologs associated with oligozoospermia phenotype ***(Table 9)***. In fact, this set of genes was related to spermatogenesis (FDR=1.245E-8), and specifically targeting the chromosomal parts (FDR=3.192E-2). Surprisingly, interaction enrichment revealed that the corresponding protein coding genes does not have significantly interactions in fact; this does not necessarily mean that it is not a biologically meaningful selection of proteins — it could simply be that these proteins have not been studied very much. However, 69.43% of these genes have shared protein domains. Additionally, we identified 26 enriched mouse orthologs related to the abnormal sperm number phenotype ***(Table 10)***.

Functional enrichments revealed that these genes were related to spermatogenesis (FDR=1.79e-13, 14 genes), male meiosis (FDR=7.04e-05, 3 genes), sister chromatid cohesion (FDR=0.00269, 3 genes), male gonads development (FDR=0.00526, 4 genes), and synaptonemal complex assembly (FDR=0.0438, 2 genes) ***(Figure 21) (Table 11)***. In addition, candidate genes were prioritized based on functional similarity to the abundant down-regulated transcripts. The top-ranked genes that are more likely interacting with these genes are the following: PLK1, DNMT1, NOS3, MAPK3, JAK2, PTCH1, POLD1, GATA4, GLI3, and CARM1. Therefore, functional analysis of the interactome showed that these genes are also related to cellular response to endogenous stimulus.

Table 9. The spermatozoal enriched transcripts from oligozoospermic men associated with the oligozoospermia mouse phenotype obtained from functional annotation analysis (Toppgene) based on their abundancy.

	Gene Symbol	**Gene Name**
1	RIMBP3	RIMS binding protein 3
2	TSSK6	testis specific serine kinase 6
3	CFAP54	cilia and flagella associated protein 54
4	RARA	retinoic acid receptor alpha
5	ESR1	estrogen receptor 1
6	ESR2	estrogen receptor 2
7	CHD5	chromodomain helicase DNA binding protein 5
8	CCNA1	cyclin A1

9	GALNTL5	polypeptide N-acetylgalactosaminyltransferase like 5
10	SGO2	shugoshin 2
11	AK7	adenylate kinase 7

Table 10. The spermatozoal enriched transcripts from oligozoospermic men associated with the abnormal sperm number mouse phenotype obtained from functional annotation analysis (Toppgene) based on their abundancy.

	Gene Symbol	**Gene Name**
1	MORC1	MORC family CW-type zinc finger 1
2	RIMBP3	RIMS binding protein 3
3	REC8	REC8 meiotic recombination protein
4	TSSK6	testis specific serine kinase 6
5	CFAP54	cilia and flagella associated protein 54
6	RARA	retinoic acid receptor alpha
7	BRCA1	BRCA1, DNA repair associated
8	MEIOC	meiosis specific with coiled-coil domain
9	BRDT	bromodomain testis associated
10	BSG	basigin (Ok blood group)
11	ESR1	estrogen receptor 1
12	ESR2	estrogen receptor 2
13	CHD5	chromodomain helicase DNA binding protein 5
14	CCNA1	cyclin A1
15	C14orf39	chromosome 14 open reading frame 39
16	GALNTL5	polypeptide N-acetylgalactosaminyltransferase like 5
17	RAD21L1	RAD21 cohesin complex component like 1
18	RBM5	RNA binding motif protein 5
19	SGO2	shugoshin 2
20	CEP131	centrosomal protein 131
21	MEI1	meiotic double-stranded break formation protein 1
22	RNF17	ring finger protein 17
23	TDRD1	tudor domain containing 1
24	AK7	adenylate kinase 7
25	BOLL	boule homolog, RNA binding protein

26	SMC1B	structural maintenance of chromosomes 1B

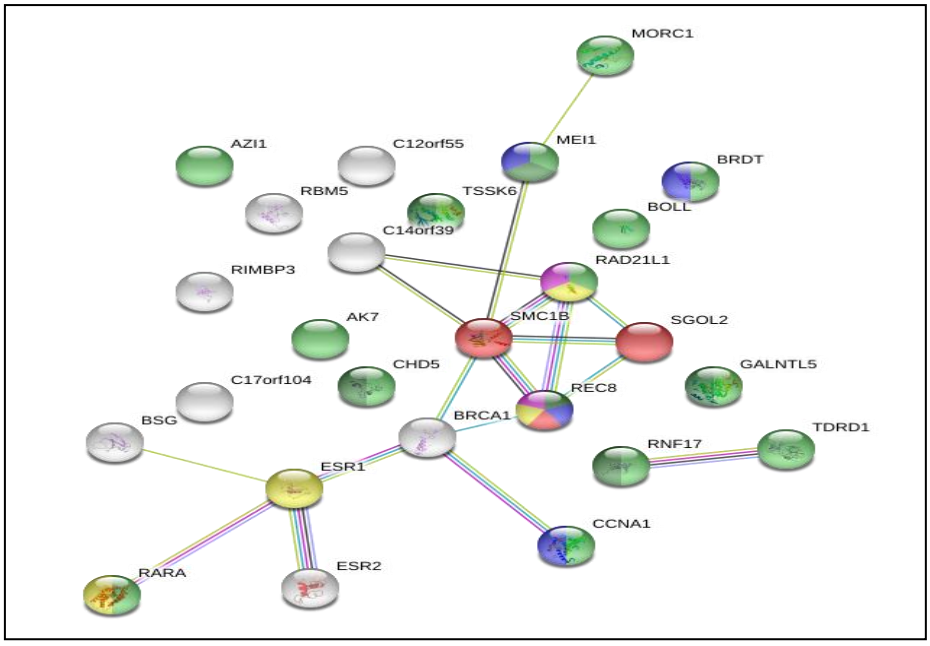


Figure 21. Protein interaction map. Map was prepared using the STRING web tool (https://string-db.org) for the proteins encoded by the set of genes associated with the abnormal sperm number phenotype. Colored lines denote interactions and network nodes represent proteins.

Table 11. Functional enrichments of the protein-protein interaction (PPI) network for the mouse orthologs associated with abnormal sperm number.

Biological process (GO)			
Pathway description	*Count in gene set*	*False discovery rate*	*Nodes legend*
Spermatogenesis	14	1.79e-13	
Spermatid development	6	4e-06	
Male meiosis	4	7.04e-05	
Sister chromatid cohesion	3	0.00269	
Male gonad development	4	0.00526	
Synaptonemal complex assembly	2	0.0438	

4.1.1.3.2. Up-regulated genes

Mining of the abundant up-regulated spermatozoa transcripts revealed 6907 enriched genes. These transcripts were functionally related to ubiquitin-dependent protein catabolic process (FDR=2.249E-36), cellular response to DNA damage stimulus (FDR= 3.579E-26), cellular respiration (FDR=4.917E-20), histone modification (FDR=8.727E-20), and histone

acetylation (FDR=8.210E-11). Notably, many orthologous phenotypes in mouse are detected, markedly increased apoptosis (FDR=2.351E-6), and oxidative stress (FDR=2.254E-4).

Enrichment analysis identified pathways that are overrepresented such as the cellular responses to stress (FDR=6.460E-25, 269 genes), chromatin modifying enzymes (FDR=1.104E-15, 166 genes), apoptosis (FDR=6.969E-12, 108 genes), and ubiquitin proteasome pathway (FDR=1.001E-7, 35 genes). It is noteworthy to mention, that we detected several genes enriched for reproductive system abnormality disease (FDR=1.164E-2, 13 genes). In particular, the latter genes were found related to the ribosome (FDR=3.845E-17, 10 genes).

Careful examination of the chromatin modifying enzyme genes revealed that 41 genes were involved in epigenetic regulation of gene expression (FDR= 3.941E-41) ***(Supplementary Table S5)*.** GeneMANIA was used to obtain the co expression and co localization coefficients of the corresponding proteins. Only 30.21% perform physical interaction. Importantly is the fact that 42.37% are co-expressed and 0.85% co-localized. We further analyzed these genes and then prioritized the top-ranked candidate genes based on functional similarity. Therefore, functional analysis of the interactome showed that these genes are related to the male infertility mouse phenotype (FDR=2.443E-5, 9 genes). These genes are the following: RB1, H3F3A, ESR1, TP53, STAT3, ATM, CDK2, CTNNB1, and RAD50. In fact, this set of genes is functionally related to the regulation of cell cycle (FDR=3.981E-7).

4.1.1.3.3. Long non-coding RNAs

The 290 detected lincRNA (long intergenic noncoding RNAs) included 26 genes belonging to the long non coding RNAs family (FDR=6.108E-48) ***(Table 12)***. Particularly, TP53TG1 – a lncRNA- was found associated with cellular response to DNA damage stimulus (FDR=4.296E-2). Additionally, we identified new 16 lncRNAs from the antisense RNAs group ***(Supplementary Table S6)***. Analysis of the Gene Ontology consortium terms of these genes did not reveal any enrichment for functional classes. The RNA–protein interaction pairs were retrieved using the Search Tool for RNA–protein Association and Interaction Networks (v1.0) (RAIN - https://rth.dk/resources/rain/). A network with 36 nodes and 12 PPI (protein-protein interaction) pairs was collected (PPI enrichment p-value=4.44e-16).

More so, candidate genes were identified and prioritized based on functional similarities to the differentially expressed genes. Accordingly, the top-ranked genes are: SFPQ, TRIAP1, NKX3-1, NONO, TP53, SETMAR, MYC, HNRNPK, XRCC6, and XRCC5

(Figure 22). These genes were enriched for the cellular response to DNA damage stimulus function and regulation of gene expression.

Furthermore, we collected a PPI network with 46 nodes and 44 PPI pairs, when we analyzed the total set of lncRNAs (42 genes) with the top-ranked candidate genes (PPI enrichment p-value=1.37e-13).Transcripts associated with biological processes such as Ku70:Ku80 complex (XRCC6 and XRCC5 genes, FDR=0.00809) were enriched in the spermatozoa.

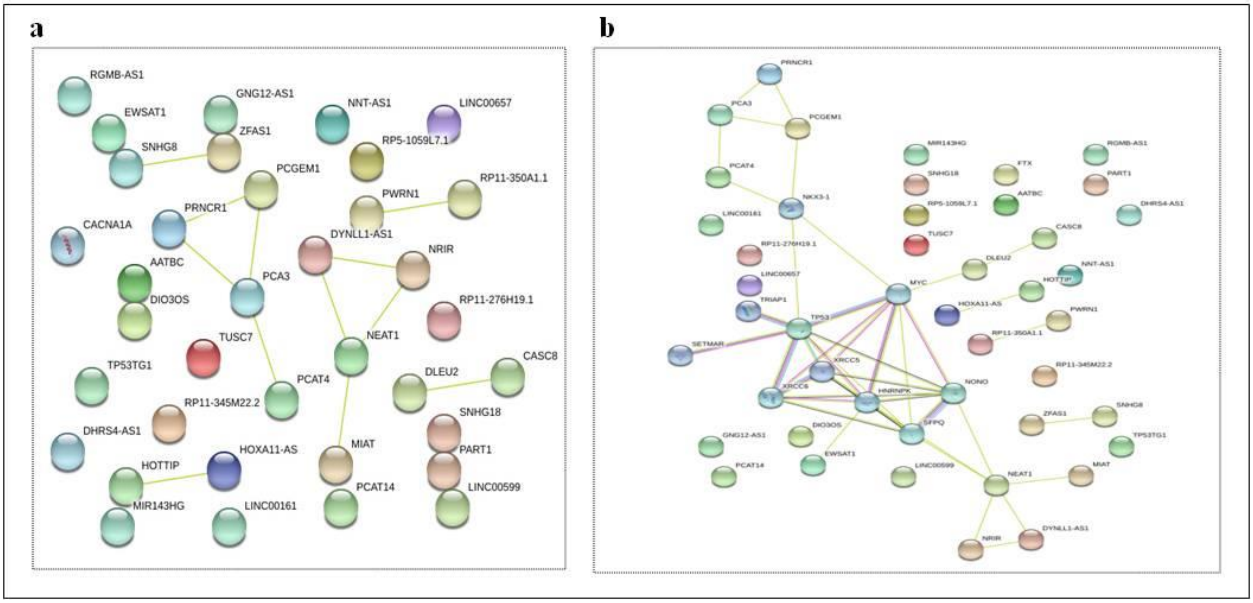


Figure 22. (a) RNA–protein interaction network map (v1.0) prepared using the RAIN web tool (https://rth.dk/resources/rain/) for the long non-coding RNAs (lncRNAs) associated with oligozoospermia. (b) lncRNAs- top ranked candidate genes encoding proteins interaction map associated with oligozoospermia. Colored lines denote interactions and network nodes represent proteins and lncRNAs.

Oligozoospermia is caused by a series of genetic [346] and epigenetic alterations [347]. Therefore, a more detailed genetic characterization will provide a better understanding of this condition and facilitate the development of new therapeutic strategies. Research is currently being focused on identifying novel molecular markers for male infertility. In this regard, spermatozoal transcriptomes have been investigated to assess male fertility [246, 348]. So far, the spermatozoal transcript profiling has been carried out in humans [102] and other species [349] to assess the role of spermatozoal RNA in fertility.

However, the role of spermatozoa RNAs in fertility is still under vigorous investigation. The presence of sperm mRNAs could be explained by the fact that these transcripts are stored not only as remnants of past spermatogenesis events, but for further zygotic functions [244]. Even though the genetic profiles of oligozoospermic patients

have been moderately studied [173, 187, 237], no studies have been performed using next generation sequencing (NGS) technology.

Table 12. List of long non-coding RNAs associated with oligozoospermia.

	Gene Symbol	Gene Name
1	PCGEM1	PCGEM1, prostate-specific transcript (non-protein coding)
2	PRNCR1	prostate cancer associated non-coding RNA 1
3	EWSAT1	Ewing sarcoma associated transcript 1
4	PWRN1	Prader-Willi region non-protein coding RNA 1
5	PWRN2	Prader-Willi region non-protein coding RNA 2
6	MAFTRR	MAF transcriptional regulator RNA
7	SOX21-AS1	SOX21 antisense divergent transcript 1
8	LINC00161	long intergenic non-protein coding RNA 161
9	DIO3OS	DIO3 opposite strand/antisense RNA (head to head)
10	PCAT4	prostate cancer associated transcript 4 (non-protein coding)
11	FTX	FTX transcript, XIST regulator (non-protein coding)
12	OLMALINC	oligodendrocyte maturation-associated long intergenic non-coding RNA
13	PCAT14	prostate cancer associated transcript 14 (non-protein coding)
14	CARMN	cardiac mesoderm enhancer-associated non-coding RNA
15	CASC22	cancer susceptibility 22 (non-protein coding)
16	MIR99AHG	mir-99a-let-7c cluster host gene
17	PART1	prostate androgen-regulated transcript 1 (non-protein coding)
18	NORAD	non-coding RNA activated by DNA damage
19	GAS1RR	GAS1 adjacent regulatory RNA
20	SNHG18	small nucleolar RNA host gene 18
21	SNHG8	small nucleolar RNA host gene 8
22	KCNIP4-IT1	KCNIP4 intronic transcript 1
23	LINC00599	long intergenic non-protein coding RNA 599
24	MIAT	myocardial infarction associated transcript (non-protein coding)
25	TP53TG1	TP53 target 1 (non-protein coding)
26	NEAT1	nuclear paraspeckle assembly transcript 1 (non-protein coding)

Transcripts quantitative PCR-based approaches of some candidate genes associated with oligozoospermia have been so far intensively addressed [218-220, 237-238]. In this study, we seek to address the genomic disparities by analyzing paired normal and oligozoospermic sperm cells using RNA-seq.

This study also reveals another transcriptomic aspect of oligozoospermia, such as the long non-coding RNA expression. Spermatozoa contain transcripts for 10087 differentially expressed genes, a number that is strikingly higher than the findings of earlier reports using a microarray-based approach. In particular, 2823 genes were significantly down-regulated while 7264 genes were significantly up-regulated. Gene ontology analysis of these transcripts revealed heterogeneous RNA population in spermatozoa. These transcripts included a large number of coding and non-coding RNAs.

The majority of the abundantly down-regulated spermatozoal transcripts might play a role in sperm production and function. In fact, transcripts associated with spermatogenesis, spermatid differentiation, flagellated sperm motility, fusion of sperm to egg plasma membrane involved in single fertilization, and fertilization were significantly enriched in the spermatozoa. However, the cellular components affected are sperm parts: centrosome, microtubule organizing center, kinetochore, cohesion complex, and chromatoid body. The presence of such spermatogenesis-associated differentially expressed (DE) transcripts in spermatozoa from oligozoospermic subjects has been previously suggested.

Several authors have suggested that sperm from oligozoospermic patients share several hallmarks, namely spermatogenic impairment [187, 237, 239], functional alterations such as decreased motility [350], and defective zona binding [351]. It has been known that oligozoospermic subjects showed a picture of testicular phenotype of severe hypospermatogenesis [352]. The results of our analysis were consistent with those reported above. This down-regulation of transcripts involved in spermatogenesis and sperm motility was congruent with previous findings from microarray studies [102-103].

Add to that, the altered cellular components are directly linked to male germ cells development [243-245]. In the same perspective, our results revealed that the collagen formation and the extracellular matrix (ECM) organization are 2 altered pathways in oligozoospermia. Recent findings illustrated the crucial role of ECM in supporting Sertoli and germ cell function in the seminiferous epithelium, including the blood-testis barrier (BTB) dynamics [353].

Consistent with our hypothesis, we found several mouse orthologs giving rise to the male infertility phenotype. The filtered ortholog set is enriched with genes that are implicated in spermatogenesis, meiotic cell cycle, synaptonemal complex assembly, piRNA metabolic process, sister chromatid cohesion, and male gonad development.

Disruptions in meiosis were found to be associated with oligozoospermia. From these DE transcripts, *CCNA1, TEX14 and BRDT*, observed in the present study have been reported to regulate spermatogenesis. The cyclin A1 (CCNA1) is expressed during meiosis and is required for spermatogenesis [354]. Targeted disruption of the CCNA1 gene in mice, showed reduced sperm production and spermatogenic arrest prior to the first meiotic division [355]. Furthermore, TEX14 is essential for intercellular bridge including germ cell communication during spermatogenesis [356]. The down-regulation of TEX14 causes spermatogenic arrest in pigs [357]. In addition, BRDT is required for male germ cells differentiation, spermiogenesis, and early embryogenesis [358]. Differential *BRDT* mRNA levels were observed between fertile donors and subfertile patients [195].

Failure of formation of a component of the synaptonemal complex lead to infertility [359]. In particular, a mutation in REC8, a key component of the meiotic cohesion complex, is considered a polymorphism and does not cause infertility [360]. The piRNA metabolic process was also found altered, due to their potential role in spermatogenic control. Thus, TDRD1 silencing contributes to an unsuccessful germ cell development that might explain male infertility [29].

Candidate genes identification and prioritization based on functional similarities to the abundant down-regulated set of genes will help increase the yield of downstream findings. Computational analysis using ToppGene, identified mouse orthologs for abnormal sperm number phenotype. Likewise, these genes are related to spermatogenesis, male meiosis, sister chromatid cohesion, male gonads development, and synaptonemal complex assembly. In addition, functional analysis of the down-regulated candidate genes showed that these genes are related to cellular response to endogenous stimulus. We can speculate that the attenuated responsiveness in male germ cells may be associated with hypospermatogenesis and oligozoospermia.

The spermatozoal transcripts that are significantly over represented are related to ubiquitin-dependent protein catabolic process, and cellular response to DNA damage stimulus. It has previously been suggested that proteasome subunit alpha type-3 (PSA3) protein levels were increased in azoospermia [361]. Similarly, the apoptotic pathway was

found deregulated in oligozoospermia [102-103]. It appears that the suppression of spermatogenesis was due to acceleration of germ cell apoptosis [362]. In fact the pro-apoptotic gene BAX [100, 363] and BARD1 [364] were found up-regulated in sperm from infertile subjects, which is consistent with our data . Cellular respiration transcripts, which are associated with oxidative stress, were also found over-represented in spermatozoa. It has been recently suggested that ROS is a common underlying mechanism of oligozoospermia [91, 102]. Transcriptome data derived from microarray revealed that LMNA, an important gene for spermatogenesis is up-regulated in germ cells of infertile men [102]. We also examined our data and found that this gene was over-represented. This may be indicative of an abortive spermatogenesis mechanism.

Interestingly, we identified several chromatin modifying enzyme genes namely related to histone modification and histone acetylation. These genes are associated with epigenetic regulation of gene expression [365]. There is evidence in previous studies demonstrating that sperm samples from patients with oligozoospermia often contain altered epigenetic modifications [179].

It is plausible that defects in genes that are key epigenetic regulators could lead to abnormal spermatogenesis and oligozoospermia. The interactome network between the epigenetic regulators and prioritized genes appears to be enriched for cell cycle regulation causing a male infertility mouse phenotype.

Apart from these findings, recent studies have suggested that sperm cells provide non-coding RNA (ncRNA) that may influence semen fertility [366]. One such class of ncRNA, long non-coding RNAs (lncRNAs), are abundant in spermatozoa. The lncRNA composition and their function in spermatozoa have not been completely studied. Our data showed that TP53TG1 which was enriched over 42 genes, was also associated with cellular response to DNA damage stimulus.

The interactome network analysis between the complete set of genes and the prioritized genes revealed that the Ku70:Ku80 complex was enriched. This complex is known to be involved in the repair of DNA double strand breaks [367]. It appears that lncRNAs may play a role in modulating the cellular response to DNA damage.

4.1.2. Sperm transcriptome in asthenozoospermia

4.1.2.1. Total number of transcripts and their expression level

Sequencing averagely generated 44,377,108 raw sequencing reads and then 40,169,798 after filtering low quality reads. The average mapping ratio with reference gene is 40.08%, and the average genome mapping ratio is 91.04%. We identified 14606 differentially expressed sequences, and of these 10559 genes were significantly down-regulated while 4047 sequences were significantly up-regulated in asthenozoospermia compared to fertile men. It is noteworthy to mention that for this analysis, a BH p-value adjustment was performed [330, 345] and the criteria set were: the level of controlled false positive rate set to 0.05, abs(log2FC)>1, and DeSeq2 read counts≥1.

4.1.2.2. Nature of spermatozoa transcripts

For this purpose, we used the Ensembl database as of January 2018 (http://www. ensembl. org) [368] to predict the transcripts types. Characterization of the pool of purified spermatozoa RNAs longer than 200 nucleotides (nt) showed that spermatozoa carry protein coding transcripts (97%), long intergenic noncoding RNAs – lincRNA (1%), antisense RNA including some long non coding RNAs (1%), and pseudogenes (1%).

4.1.2.3. Enrichment analysis of protein-coding genes for identification of biological and functional differences

To identify the differentially expressed genes, the transcriptome data of sperm from asthenozoospermic men were analyzed by using the DESeq package in R software.

4.1.2.3.1. Down-regulated genes

Gene ontology (GO)-enrichment analysis of abundant down-regulated genes, identified based on the DeSeq2 normalized read counts >500, in the asthenozoospermia group revealed a total of 1184 significantly enriched GO terms ***(Table 13)***. Most of these terms were linked to sperm function, as e.g. ”spermatogenesis”, or sperm structure, as e.g “cilium organization”, “cilium assembly”, “microtubule-based movement”, “cytoskeleton organization”. In addition, other GO terms corresponding to mRNA translation, as e.g “mRNA catabolic process”, “nuclear-transcribed mRNA catabolic process, nonsense-mediated decay”, and “ribosome biogenesis” were enriched. Add to that, these transcripts were significantly associated with “epigenetic regulation of gene expression”. To identify the biological pathways that are active in asthenozoospermia, the differentially expressed genes were mapped to canonical signaling pathways found in BioSystems: REACTOME, MSigDB C2 BIOCARTA (v6.0), and BioSystems: KEGG.

Table 13. The molecular functions, biological processes, cellular localization, and pathways implicated of spermatozoal transcripts in asthenozoospermia obtained from functional annotation analysis (Toppgene) of the abundantly down-regulated set of genes significantly differentially expressed. The analysis was carried out using *Homo sapiens* as background.

GO-Term	FDR B&H	Genes observed	Genes in annotation
Molecular function			
RNA binding	9.680E-23	208	1632
Cytoskeletal protein binding	6.439E-15	122	886
ATP binding	6.122E-9	153	1475
Structural constituent of ribosome	2.247E-4	31	216
Ubiquitin-protein transferase activity	3.518E-4	48	414
Microtubule motor activity	1.260E-3	15	77
ATPase activity	2.130E-3	48	446
Biological process			
mRNA catabolic process	1.565E-15	53	218
Cytoskeleton organization	1.142E-14	146	1164
Nuclear-transcribed mRNA catabolic process, nonsense-mediated decay	4.777E-14	37	121
Microtubule-based process	1.051E-13	97	658
Spermatogenesis	4.433E-5	63	151
Cilium assembly	1.097E-4	37	269
Cilium organization	1.610E-4	36	263
Regulation of gene expression, epigenetic	1.873E-4	36	265
Microtubule-based movement	2.186E-4	33	235
Ribosome biogenesis	4.585E-4	40	322
Cellular component			
Microtubule cytoskeleton	1.347E-18	151	1125
Microtubule organizing center	2.374E-9	83	646
Cytosolic ribosome	3.747E-9	29	119
Sperm part	1.411E-3	25	184
Sperm flagellum	2.030E-3	14	77
Dynein complex	3.200E-2	8	45
Pathway			
Nonsense-Mediated Decay (NMD)	7.780E-12	33	121
Translation	2.919E-11	38	65
Ribosome	1.460E-6	29	154
Cilium Assembly	1.427E-5	31	192
Glycolysis (Embden-Meyerhof pathway), glucose => pyruvate	4.087E-2	6	25
Ubiquitin mediated proteolysis	4.119E-2	17	137

There were a statistically significant amount of mapped genes for mRNA translation-related pathways, such as those mediated by “Nonsense-Mediated Decay (NMD)”, “translation”, and “ribosome”. More so, enriched down-regulated spermatozoal transcripts associated with the “cilium assembly” pathway as well as “glycolysis (Embden-Meyerhof pathway)” were also identified. Interestingly, direct comparison of human and mouse phenotypes allowed rapid recognition of disease causal genes. When employing mouse phenotype data we detected several phenotypes related to spermatozoa, namely abnormal spermatogenesis (FDR=9.938E-4), abnormal sperm physiology (FDR=4.518E-3), abnormal sperm motility (FDR=9.962E-3; 29 genes out of 193) ***(Table 14)***, male infertility (FDR=3.221E-2), and asthenozoospermia (FDR=3.887E-2).

The protein–protein interaction (PPI) pairs of proteins that are encoded by the 29 mouse orthologs related to the abnormal sperm motility phenotype were retrieved using the Search Tool for the Retrieval of Interacting Genes/Proteins (STRING, http://string-db.org/). A network with 29 nodes and 7 PPI (protein-protein interaction) pairs was collected ***(Figure 23)***. The aim of this interactive network which showed a PPI enrichment p-value= 0.011, is to predict potential interactions between these genes. This indicates that these proteins are at least partially biologically connected. Functional enrichments in this network revealed that the majority of the candidates are related to spermatogenesis, and cilium assembly ***(Table 15)***. While the major cellular components affected are namely “outer dense fiber” (FDR=5.24e-05), “axoneme” (FDR=5.24e-05), “sperm part” (FDR=0.00504), and “cell projection” (FDR=0.0278).

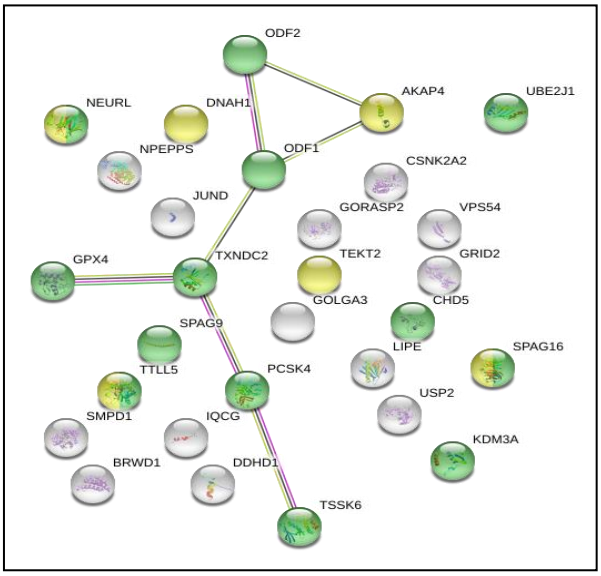


Figure 23. Protein interaction map. Map was prepared using the STRING web tool (https://string-db.org) for the proteins encoded by the set of orthologous genes associated with abnormal sperm motility phenotype. Colored lines denote interactions and network nodes represent proteins.

Table 14. The spermatozoal enriched transcripts from asthenozoospermic men associated with the abnormal sperm motility mouse phenotype obtained from functional annotation analysis (Toppgene) based on their abundancy.

	Gene Symbol	Gene Name
1	TSSK6	testis specific serine kinase 6
2	TXNDC2	thioredoxin domain containing 2
3	IQCG	IQ motif containing G
4	UBE2J1	ubiquitin conjugating enzyme E2 J1
5	VPS54	VPS54, GARP complex subunit
6	SMPD1	sphingomyelin phosphodiesterase 1
7	KDM3A	lysine demethylase 3A
8	TTLL5	tubulin tyrosine ligase like 5
9	AKAP4	A-kinase anchoring protein 4
10	TEKT2	tektin 2
11	GOLGA3	golgin A3
12	BRWD1	bromodomain and WD repeat domain containing 1
13	GPX4	glutathione peroxidase 4
14	GRID2	glutamate ionotropic receptor delta type subunit 2
15	SPAG9	sperm associated antigen 9
16	ODF1	outer dense fiber of sperm tails 1
17	ODF2	outer dense fiber of sperm tails 2
18	USP2	ubiquitin specific peptidase 2
19	DDHD1	DDHD domain containing 1
20	NEURL1	neuralized E3 ubiquitin protein ligase 1
21	NPEPPS	aminopeptidase puromycin sensitive
22	DNAH1	dynein axonemal heavy chain 1
23	GORASP2	golgi reassembly stacking protein 2
24	CSNK2A2	casein kinase 2 alpha 2
25	CHD5	chromodomain helicase DNA binding protein 5
26	PCSK4	proprotein convertase subtilisin/kexin type 4
27	JUND	JunD proto-oncogene, AP-1 transcription factor subunit
28	SPAG16	sperm associated antigen 16
29	LIPE	lipase E, hormone sensitive type

Table 15. Functional enrichments of the protein-protein interaction (PPI) network for the mouse orthologs associated with abnormal sperm motility phenotype.

Biological process (GO)			
Pathway description	*Count in gene set*	*False discovery rate*	*Nodes legend*
Spermatogenesis	13	1.3e-10	
Cilium assembly	6	5.74e-05	

Additionally, GeneMANIA was used to obtain the co expression and co localization coefficients of the corresponding proteins. Only 21.53% were predicted to perform physical interaction. Furthermore, the top 20 prioritized candidate genes identified based on functional similarities to the down-regulated genes (log2FC≥-2, FDR≥0.05) are the following: RSPH1, LMNA, SAXO1, PRKACA, CEP131, MARCKS, HSP90AA1, HIF1A, AKT1, SLIRP, SIRT1, NOS3, DYNLL1, CTNNB1, TTLL1, BBS4, SRC, GAPDHS, BCL2L11, TBPL1. Interestingly, functional analysis of the interactome showed that 5 genes are also related to estrogen signaling pathway (FDR=0.000591; PRKACA, SRC, NOS3, HSP90AA1, AKT1).

However, the analysis of genes with high log2FC (fold change) ≥-2 targeted 4472 genes. These transcripts were functionally related to cellular energy production, namely "oxidative phosphorylation" (FDR= 1.042E-10) "mitochondrial ATP synthesis coupled electron transport" (FDR= 1.440E-8), and "mitochondrial translation" (FDR= 3.586E-6). Add to that, we identified several transcripts implicated in histones modification (FDR= 1.433E-7) ***(Supplementary Table S7)***.

While, the analysis of genes with very high log2FC (fold change) ≥-3 targeted 369 genes. The highly differentially expressed genes in the cluster of the molecular function were found to be mainly related to RNA binding (FDR=2.006E-2). Whereas, the major cellular components deregulated were mitochondria (FDR=8.307E-3; 52 enriched genes out of 1769 annotated genes) and the cytoskeletal calyx (FDR= 4.675E-2; CYLC1 cylicin 1 and CCIN calicin).

We further narrowed the analysis; the criteria of a ≥log-2 fold change in expression, transcripts per million (TPM) ≥500 and *p*-value<0.05 (cut off at 5% FDR) were chosen to determine significantly down-regulated genes. Using these criteria, a total of 380 genes were significantly differentially down-regulated greater than log two-fold. To perform the annotation of the functions of the differentially expressed genes identified by DESeq analysis, GO analysis of differentially expressed genes was carried out using the ToppFun bioinformatic webtool. GO annotated differentially expressed genes mainly belonged to the three functional clusters (biological process, cellular component, and molecular function). The differentially expressed genes in the cluster of biological processes were found to be mainly related to nuclear-transcribed mRNA catabolic process, nonsense-mediated decay (FDR=5.081E-23), mRNA catabolic process (FDR= 5.257E-20), rRNA processing (FDR= 2.807E-13), ribosome biogenesis (FDR=1.069E-11), cytoskeleton organization (FDR= 3.255E-5), microtubule cytoskeleton organization (FDR= 1.285E-4), spermatogenesis (FDR=

1.755E-3), ribosomal large subunit assembly (FDR= 2.099E-3), ribosome assembly (FDR= 3.006E-3), actin filament depolymerization (FDR= 3.340E-3), and flagellated sperm motility (FDR= 1.025E-2). However, the DEG in the cluster of molecular function were found to be associated with structural constituent of ribosome (FDR=3.166E-13), cytoskeletal protein binding (FDR= 1.502E-12), ATP binding (FDR=5.655E-3), microtubule binding (FDR=1.636E-2), and calcium-dependent protein binding (FDR=2.926E-2). To understand the impact of the above-mentioned differentially expressed transcripts in asthenozoospermia, we mapped the genes that were involved in differential expression in our study to major deregulated pathways. The most frequently affected pathways are Nonsense-Mediated Decay (NMD) (FDR=3.083E-21), the ribosome (FDR= 3.451E-18), and the proteolytic involving calcium-dependent proteases (FDR=2.077E-2; CAPN2 calpain 2 and CAST calpastatin). It is noteworthy to mention that we recovered few members of the L and S ribosomal proteins gene families (25out of 85 genes) as well as 2 genes belonging to the heat shock 90kDa proteins family (HSP90AA1, HSP90AB1).

4.1.2.3.2. Up-regulated genes

Mining of the abundant up-regulated spermatozoa transcripts identified based on the DeSeq2 normalized read counts >100, in the asthenozoospermia group revealed a total of 131 genes. We found that enriched transcripts were barely detectable, particularly, 3 genes were biologically associated with cGMP-mediated signaling (FDR= 3.037E-2; PDE3A, THBS1, PRKG1). Interestingly, we only found 2 genes related to “fertility disorders”: CRISP2 (cysteine rich secretory protein 2), and CXCL8 (C-X-C motif chemokine ligand 8). Add to that, we further used the criteria of a ≥2 fold change in expression, transcripts per million (TPM)≥100 and *p*-value<0.05 (cut off at 5% FDR), a total of 63 genes were significantly differentially expressed. These transcripts were functionally associated with chemotaxis (FDR=2.119E-3). Moreover, we identified 11 mouse orthologs related to abnormal immune cell physiology (FDR=1.028E-2).

4.1.2.3.3. Long non-coding RNAs

The 78 detected lincRNA (long intergenic noncoding RNAs) from the complete set of DEGs included 11 genes belonging to the long non-coding RNAs family (FDR=9.840E-20). In addition, the 73 antisense RNA also included 8 long non-coding RNAs (FDR=1.174E-15) ***(Supplementary Tables S8, S9)***.

A total of 19 long ncRNAs were found to be notably down-regulated in asthenozoospermic patients. Analysis of the Gene Ontology consortium terms of these genes did not reveal any enrichment for functional classes. The RNA–protein interaction pairs were

retrieved using the Search Tool for RNA–protein Association and Interaction Networks (v1.0) (RAIN - https://rth.dk/resources/rain/). A network with 17 nodes and 4 PPI (protein-protein interaction) pairs were collected (PPI enrichment p-value= 2.24e-07). This indicates that these proteins are at least partially biologically connected ***(Figure 24a)***.

Subsequently, this gene network is used to prioritize candidate gene based on topological features in protein-protein interaction network. It was hypothesized that proteins associated with similar diseases would exhibit similar topological characteristics in PPI networks. Utilizing the location of a protein in the network with respect to other proteins (i.e., the "topological profile" of the proteins) we ranked the top 20 genes. This set of genes was functionally related to poly (A) RNA binding (FDR=0.00143) and RNA binding (FDR=0.00185) ***(Figure 24b)***. Interestingly, 1 gene HSP90AA1 (heat shock protein 90 alpha family class A member 1) was found enriched in the sperm mitochondrial sheath. Therefore, functional analysis of the interactome showed that these genes (9 out of 12) are related to gene expression.

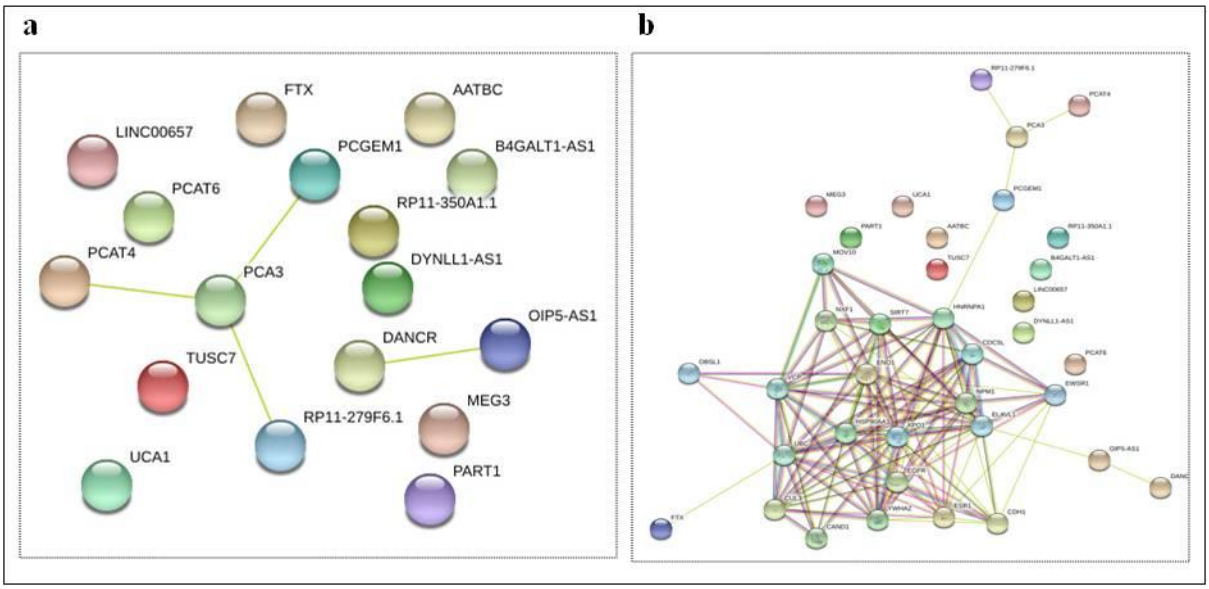


Figure 24. (a) RNA–protein interaction network map (v1.0) prepared using the RAIN web tool (https://rth.dk/resources/rain/) for the long non-coding RNAs (lncRNAs) associated with asthenozoospermia. (b) lncRNAs- top ranked candidate genes encoding proteins interaction map associated with asthenozoospermia. Colored lines denote interactions and network nodes represent proteins and lncRNAs.

Asthenozoospermia is caused by a series of genetic [113, 369] and epigenetic alterations [163]. Therefore, a more detailed characterization will provide a better understanding of this condition and facilitate the development of new personalized therapeutic strategies. Actually, research is being focused on identifying novel molecular

markers for male infertility. In this regard, spermatozoal transcriptome has been investigated to assess male fertility in humans [348] and other species [349]. Yet, the role of spermatozoa RNAs in fertility is still under vigorous investigation [370-371]. However, even though the genetic profiles of asthenozoospermic patients have been moderately studied, no studies have been performed using next generation sequencing (NGS) technology. Conversely, the characterization of the complete human sperm proteome from asthenozoospermic patients have been so far intensively addressed [275, 293-294]. In this study, we sought to address the genomic disparities by analyzing 20 normal and asthenozoospermic sperm cells using RNA-seq. This study also reveals another transcriptomic aspect of asthenozoospermia, such as the long ncRNA expression.

Spermatozoa contained transcripts for 14606 out of 30982 differentially expressed genes, a number that is intensively higher than the findings of earlier reports using a microarray-based approach [112-113]. In particular, 10559 genes were significantly down-regulated while 4047 genes were significantly up-regulated. The majority of the differentially expressed spermatozoal transcripts from asthenozoospermic men are directly related to spermatogenesis. It is evident that alterations in spermatogenesis are capable of altering sperm phenotype [112, 163]. A number of genes involved in this process has been reported to be dyregulated in asthenozoospermic patients [112-113]. It is evident that defects in spermatogenesis will profoundly alter structure, function, metabolism, and physiology of the spermatozoa.

Several authors have suggested that sperm from asthenozoospermic patients share several hallmarks, namely disruption of the mitochondrial function [112-113] as well as structural flagellar pathologies [171, 273, 295]. The molecular architecture of the sperm flagella was found altered in asthenozoospermia. In this manner, multiple defects affecting the sperm motility apparatus were identified such as; cilium assembly, cilium organization, and microtubule-based movement processes. Low motility spermatozoa carry high numbers of under-represented transcripts associated with cytoskeleton organization. For example, *TEKT2* is important for spermatozoal cytoskeleton development and it is a flagellar protein. In support of this notion, spermatozoa from mice null for the flagellar protein tektin 2 (Tekt2) demonstrated absence of inner dynein, and caused asthenozoospermia and subfertility [372]. Similarly, *ODF1* and *ODF2* proteins are down-regulated in asthenozoospermia [373-374], and *Odf2* cKO mice displayed asthenozoospermic characteristics as well [375]. Because tektins and ODFs form the main components of the cytoskeleton in sperm flagella, the results of the current study further indicate that the function of sperm flagella may be greatly

weakened by the abnormal expression of structural constituents of the cytoskeleton. More so, 2 spermatozoal transcripts are highly down-regulated in asthenozoospermia, CYLC1 cylicin 1 and CCIN calicin which are components of the cytoskeletal calyx in sperm head. They are thought to participate in acrosomal formation and nuclear morphogenesis during spermiogenesis [376-377]. The assembly of G-actin to form F-actin is controlled by several actin-binding proteins such as calicin, which is required during sperm capacitation and acrosomal reaction [378]. Similarly, the 2 proteins CYLC1 and CCIN were also identified to be down-regulated in obesity-associated asthenozoospermia [379].

Spermatozoal motility is correlated with mitochondrial function, which regulates energy generation. In this study, the presence of spermatozoal transcripts related to the ribosome and mitochondria which are significantly over represented and sturdily down-regulated, indicate that many of these transcripts are involved in cellular energy supply, wherein mitochondrial bioenergetics and organization take place and are related to sperm function. It is known that the so-called mitochondrial capsule generated around the axoneme in the sperm mid-piece drive sperm motility [380]. Accordingly, this functionally affected translation, oxidative phosphorylation, and ultimately sperm motility. In other words, the altered expression of these genes might cause mitochondrial dysfunction and impaired fertility. A paradigm was suggested according to which several nuclear genes notably ribosomal proteins promote the expression of key components of the mitochondrial transcription and translation machinery that are necessary for the mitochondrial biogenesis and function [381-382]. The ribosomal proteins transcripts are the major transcripts down-regulated in asthenozoospermia ***(Figure 25)***.

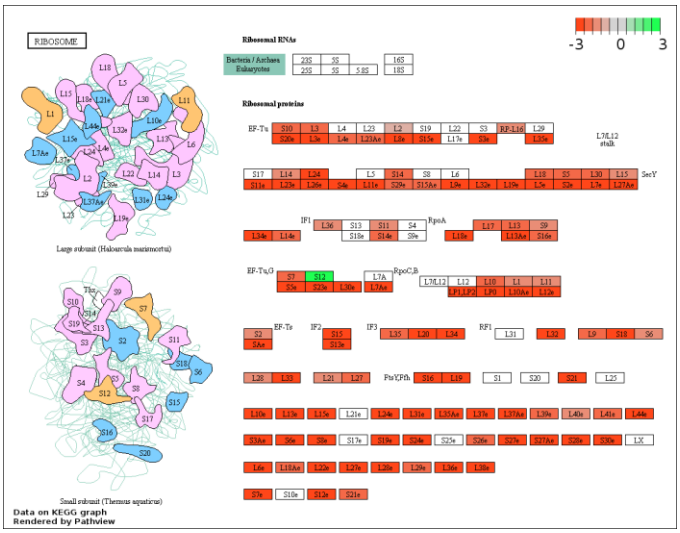


Figure 25. Pathway analysis of the ribosome (Kegg pathway) using the Pathview server (https://pathview.uncc.edu/)[337]. Highlighted genes are components of these pathways significantly

differentially expressed (DE) in asthenozoospermia. DE color code indicates genes that are up (green) or down (red) regulated.

Accordingly, this affected RNA and ATP binding since RNA binding proteins are involved in ribosome assembly [383] and ATP binding is the ultimate result of appropriate ribosomal function in sperm cells. This finding is concordant with the initial results of a microarray-based study [112]. Altogether, this will affect ATP production mainly through oxidative phosphorylation ***(Figure 26a)***, the sperm movement energizer.

Another metabolic pathway involved in energy production for sperm motility is glycolysis. In fact the principal piece of the sperm tail is devoid of mitochondria and enriched in glycolytic enzymes. We found that several transcripts associated with glycolysis (Embden-Meyerhof pathway) were also down-regulated, namely ALDOA, GAPDH, TPI1, HK1, PKM and ENO1 ***(Figure 26b)***. For example, *GAPDH* gene-knockout male mice have ATP levels that were only 10.4%, and were infertile [384]. It is noteworthy to mention that, ATP does not solely fuel sperm movement but also other critical cellular processes such as capacitation, hyperactivation and acrosome reaction [385].

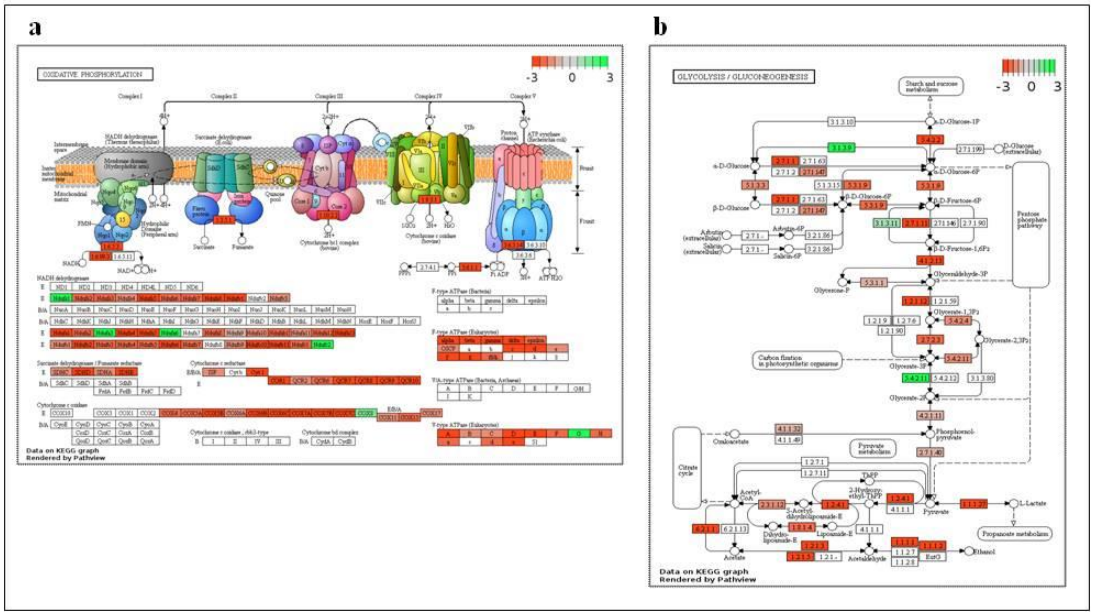


Figure 26. Pathway analysis of oxidative phosphorylation (a) and glycolysis (b) (Kegg pathway) using the Pathview server (https://pathview.uncc.edu/)[337]. Highlighted genes are components of these pathways significantly differentially expressed (DE) in asthenozoospermia. DE color code indicates genes that are up (green) or down (red) regulated.

In addition, the occurrence of alterations in the ubiquitin mediated proteolysis has also been previously suggested [386], but the functional relevance of this result is not clearly understood. Interestingly, we found that several abundant under-represented transcripts part of the Nonsense-Mediated Decay (NMD) pathway are significantly enriched in the

spermatozoa of asthenozoospermic patients. Recently, it was revealed that the NMD pathway is essential for spermatogenesis by contributing to shaping the male-germ-cell–specific transcriptome, which is typified by mRNAs with unusually short 3′UTRs [387-388]. A knockout for components of the NMD pathway in spermatogonia led to infertile mice [389].

Spermatozoa carry a complete calcium-dependent proteases system which include CAPN2 calpain 2 and CAST calpastatin, that are suggested to play a role in capacitation, acrosome reaction and egg-fusion [390-391]. In the present study we observed that these two transcripts were down-regulated in asthenozoospermia. It is known that, calcium ion is a key regulator of human sperm function [392]. We found several players in the calcium signaling pathway (CALM, CAMK, PDE1…) deregulated in spermatozoa of asthenozoospermic patients ***(Figure 27)***. Our results are in agreement with previous reports and suggest a strong relationship between calcium homeostasis and sperm motility. In fact, both reduced calcium/calmodulin (CaM) complex and intracellular calcium levels have been demonstrated in asthenozoospermic patients [316-317].

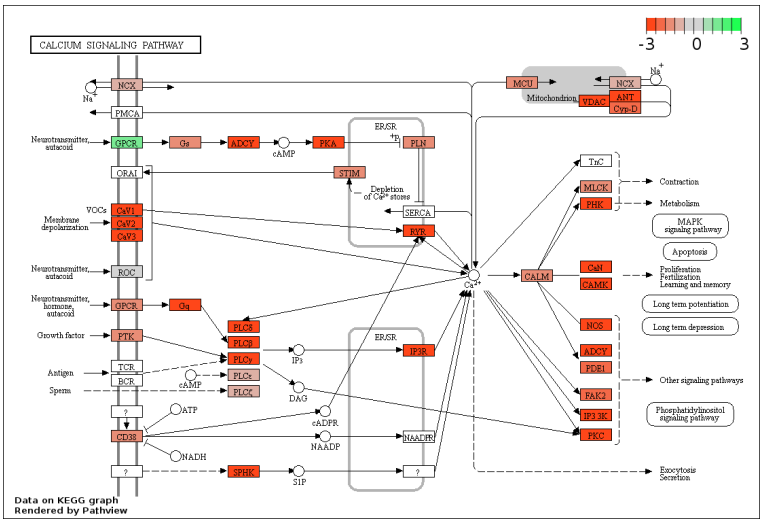


Figure 27. Pathway analysis of the calcium signaling pathway (Kegg pathway) using the Pathview server (https://pathview.uncc.edu/)[337]. Highlighted genes are components of these pathways significantly differentially expressed (DE) in asthenozoospermia. DE color code indicates genes that are up (green) or down (red) regulated.

In addition, we detected 5 genes related to the estrogen signaling pathway (PRKACA, SRC, NOS3, HSP90AA1, and AKT1) that were under-represented in the asthenozoospermia group. These transcripts are signaling modulator of several physiological processes, and are involved in different human sperm functions, namely sperm motility. Any alterations in the expressions of these transcripts might lead to male infertility. For example, NOS3 stimulates human sperm motility [393], while SRC is involved in capacitation and acrosome reaction [394].

Among various differentially expressed transcripts, we identified the down-regulation of 2 genes belonging to the heat shock 90kDa proteins family (HSP90AA1, HSP90AB1) in the asthenozoospermic group when compared with the fertile group. Bansal et al. (2015) reported that HSP90AB1 is also under-represented in asthenozoospermia [112].

Interestingly, we also identified numerous transcripts related to histone modification, which were detected in the highly down-regulated cluster in asthenozoospermia. Some of these genes are involved in shaping the epigenetic histone code. For example, HDAC1, TET1, TET2, TET3 were down-regulated in asthenozoospermic men. HDACs were reported in sperm of rats to modulate sperm motility by interacting with α-tubulin [395-396]. Notably, HDAC1 presented an altered expression in low motility sperm samples [153]. In addition, Ni et al. (2016) showed that sperm from infertile men present lower levels of TET enzymes mRNA (TET 1-3) when compared to fertile men [257]. However, the relation between epigenetic alterations and male subfertility has sparkled an avalanche of works on several layers and players characterizing the sperm epigenome [397].

In parallel, few abundant spermatozoal transcripts are significantly over-represented in spermatozoa from asthenozoospermic men. Only 3 genes related to cGMP-mediated signaling were found enriched. Sperm motility and other events are regulated by signal transduction systems involving cAMP as a second messenger. PDE3A is involved in cAMP degradation [398]. Experimentally, PDE inhibitors are able to enhance of human sperm motility [399].

In addition, several transcripts involved in chemotaxis and immune cell physiology were observed. These data suggest that the overexpression of the above-mentioned genes may represent an unspecific mechanism of reaction to spermatogenic damage, rather than the cause of defective spermatogenesis. Apart from these findings, studies have suggested that sperm cells provide non-coding RNA (ncRNA) that may influence semen fertility. One such class of ncRNA, long non-coding RNAs, were abundant in spermatozoa from asthenozoospermic men. Their function in spermatozoa have not been completely studied. We found that some lncRNAs can regulate transcription through not solely the interaction with RNA binding proteins. Although a role of long non-coding RNA (lncRNA) on sperm from asthenozoospermic patients has scarcely been reported, only one study addressed the role of HOTAIR lncRNA in asthenozoospermia [220]. Further studies are essential to better understand the roles of ncRNAs on spermatozoa fertility. Although the central dogma remains a core tenet of cellular and molecular biology, the appreciation of lncRNAs as functional genomic elements that defy the central dogma may be essential for fully

understanding biology and pathology. Our results indicate that the complexity of lncRNA transcription has been grossly underappreciated, and that myriad lncRNAs may be associated with male infertility. Studies have confirmed the presence of many RNAs in mature spermatozoa from asthenozoospermic men [170-171, 286], but the role of very few of these transcripts are correlated with asthenozoospermia etiology.

4.1.3. Sperm transcriptome in teratozoospermia

4.1.3.1. Total number of transcripts and their expression level

Sequencing averagely generated 53,288,668 raw sequencing reads and then 50,324,899 after filtering low quality reads. The average mapping ratio with reference gene is 36.49%, and the average genome mapping ratio is 91.55%. We identified 7345 differentially expressed sequences, and of these 4546 genes were significantly down-regulated while 2799 sequences were significantly up-regulated in teratozoospermia compared to fertile men. It is noteworthy to mention that for this analysis, a BH p-value adjustment was performed [330, 345] and the criteria set were: the level of controlled false positive rate set to 0.05, abs(log2FC)>1, and DeSeq2 read counts≥1.

4.1.3.2. Nature of spermatozoa transcripts

For this purpose, we used the Ensembl database as of January 2018 (http://www. ensembl. org) [368] to predict the transcripts types. Characterization of the pool of purified spermatozoa RNAs longer than 200 nucleotides (nt) showed that spermatozoa carry protein coding transcripts (90%), long intergenic non-coding RNAs – lincRNA (3%), antisense RNA including some long non coding RNAs (2%), and pseudogenes (5%).

4.1.3.3. Functions of spermatozoal transcripts

To capture a range of meaningful findings in the biological aspects of teratozoospermia, we performed functional enrichment analyses using ToppGene. ToppGene detects functional enrichment of a given gene list based on ontologies, pathways, phenotypes, etc. All statistical results were corrected for multiple testing using the Benjamini-Hochberg method.

4.1.3.3.1. Complete genes differentially expressed

From our extensive list of the complete differentially expressed genes, the most significant molecular function from the Gene Ontology analysis are ubiquitin-like protein transferase activity, ubiquitin-protein transferase activity, ubiquitin protein ligase activity, and ubiquitin binding. In addition, the most significant biological processes are RNA processing, cellular catabolic process, cell cycle, ubiquitin-dependent protein catabolic process, proteasome-mediated ubiquitin-dependent protein catabolic process, chromatin

organization, cellular response to DNA damage stimulus, and histone modification. The cellular component GO terms most significantly enriched spermatozoa transcripts from teratozoospermic men were catalytic complex, mitochondrial part, centrosome, ubiquitin ligase complex, proteasome complex, cytoskeletal part, sperm flagellum and sperm principle piece. Along these lines, the pathways affected are chromatin organization, ubiquitin mediated proteolysis, apoptosis, deubiquitination, cell cycle, protein ubiquitination, and HATs acetylate histones (**Table 16**).

Interestingly, direct comparison of human and mouse phenotypes allowed rapid recognition of disease causal genes. When employing mouse phenotype data we detected several phenotypes related to spermatogenesis, namely abnormal spermatogenesis (FDR= 6.160E-5), male infertility (FDR= 1.405E-4), and abnormal male germ cell morphology (FDR=4.178E-4; 262 genes out of 542). When mining the 262 genes associated with abnormal male germ cell morphology, 94 genes were found related to teratozoospermia. The protein–protein interaction (PPI) pairs of proteins that are encoded by these genes were retrieved using the Search Tool for the Retrieval of Interacting Genes/Proteins (STRING, http://string-db.org/). A network with 94 nodes and 65 PPI (protein-protein interaction) pairs was collected. This interactive network to predict potential interactions between these genes showed a PPI enrichment p-value= 5.15e-07. This indicates that these proteins are at least partially biologically connected ***(Figure 28)***. Functional enrichments in this network revealed that the majority of the candidates are related to spermatogenesis, organelle assembly and cell cycle ***(Table 17)***.

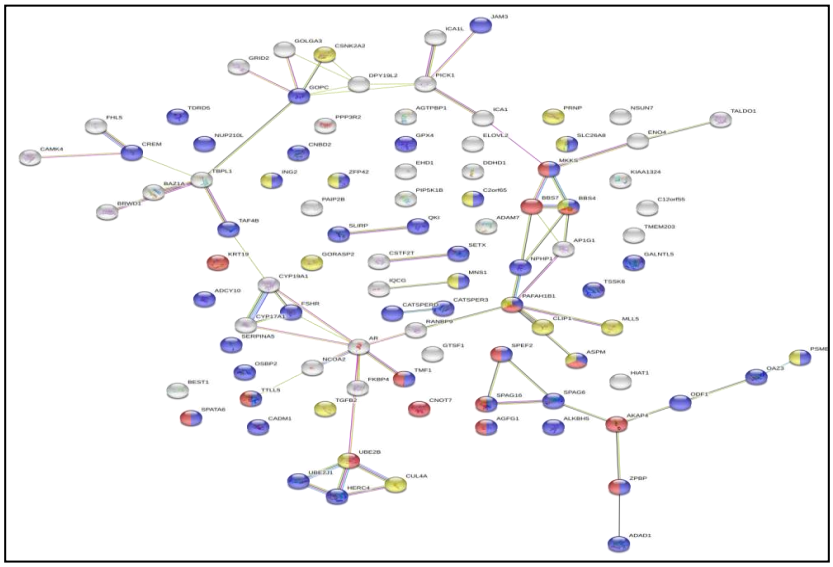


Figure 28. Protein interaction map. Map was prepared using the STRING web tool (https://string-db.org) for the proteins encoded by the set of mouse orthologous genes associated with teratozoospermia phenotype. Colored lines denote interactions and network nodes represent proteins.

Table 16. The molecular functions, biological processes involved, cellular localization, and pathways implicated of spermatozoal transcripts in teratozoospermia obtained from functional annotation analysis (Toppgene) of the complete set of genes significantly differentially expressed. The analysis was carried out using *Homo sapiens* as background.

GO-Term	FDR B&H	Genes observed	Genes in annotation
Molecular function			
Ubiquitin-like protein transferase activity	7.422E-6	222	441
Ubiquitin-protein transferase activity	2.101E-5	208	414
Ubiquitin protein ligase activity	1.534E-4	114	212
Ubiquitin binding	4.327E-4	64	108
Biological process			
RNA processing	1.518E-27	514	913
Cellular catabolic process	4.291E-23	903	1828
Cell cycle	1.332E-12	820	1766
Ubiquitin-dependent protein catabolic process	3.879E-12	313	590
Proteasome-mediated ubiquitin-dependent protein catabolic process	2.166E-10	217	393
Chromatin organization	3.231E-9	380	770
Cellular response to DNA damage stimulus	4.432E-10	397	800
Histone modification	8.498E-8	232	447
Cellular component			
Catalytic complex	5.393E-29	586	1077
Mitochondrial part	1.696E-25	534	987
Centrosome	2.813E-8	254	501
Ubiquitin ligase complex	2.954E-7	149	274
Proteasome complex	1.242E-4	49	78
Cytoskeletal part	2.672E-4	646	1520
Sperm flagellum	8.530E-3	43	77
Sperm principle piece	1.741E-2	16	23
Pathway			
Chromatin organization	1.710E-6	147	279
Ubiquitin mediated proteolysis	4.173E-6	81	137
Apoptosis	2.303E-5	96	174
Deubiquitination	3.769E-5	152	303

Cell cycle	9.102E-4	277	624
HATs acetylate histones	5.944E-3	73	143

Table 17. Functional enrichments of the protein-protein interaction (PPI) network for the mouse orthologs associated with teratozoospermia phenotype.

Biological process (GO)			
Pathway description	*Count in gene set*	*False discovery rate*	*Nodes legend*
Spermatogenesis	45	6.01e-45	
Organelle assembly	16	1.91e-08	
Cell cycle	17	0.00746	

We clustered the up- and down-regulated genes separately. During genes sorting, narrowed filters were employed; including the highly abundant transcripts and the higher fold change at different expression levels, to study gene functions for relevant biological concepts.

4.1.3.3.2. Down-regulated genes

First, we considered the analysis of the highly abundant down-regulated genes. The latter spermatozoal transcripts identified based on the DeSeq2 normalized read counts >100, were related to 1751 genes. Transcripts associated with biological processes such as, cellular catabolic process (FDR= 3.000E-41), cell cycle (FDR= 2.463E-21), posttranscriptional regulation of gene expression (FDR= 2.829E-21), chromatin organization (FDR= 8.874E-14), histone modification (FDR= 3.497E-11), proteasome-mediated ubiquitin-dependent protein catabolic process (FDR=9.158E-10), protein ubiquitination (FDR= 1.087E-9), cytoskeleton organization (FDR= 4.773E-8), and chromatin remodeling (FDR= 2.768E-7) were significantly under-represented in the spermatozoa. In addition, the major cellular component affected is the catalytic complex (FDR= 3.585E-25). Accordingly, the molecular functions of the highly abundant down-regulated spermatozoal transcripts are associated with cytoskeletal protein binding (FDR= 1.827E-11), ubiquitin-protein transferase activity (FDR=4.807E-5), ubiquitin protein ligase binding (FDR= 9.069E-4), and ubiquitin protein ligase activity (FDR= 2.047E-3). The abundantly down-regulated spermatozoal transcripts are associated with signaling pathways mainly the Nonsense Mediated Decay (NMD) (FDR= 3.450E-32), and Ubiquitin mediated proteolysis (FDR=2.533E-3). Of particular interest, the proteasome and ubiquitin-mediated proteolysis pathways were highlighted by our analysis to contain

differentially expressed genes (DEGs). **Figure 29** illustrates the KEGG pathway map of the proteasome and ubiquitin-mediated proteolysis pathway.

Interestingly, we identified from the list of the down-regulated genes 91 mouse orthologs related to the abnormal male germ cell morphology phenotype (FDR=1.567E-112), and 24 related to teratozoospermia (FDR= 8.092E-19) ***(Table 18)***. These candidate genes are functionally related to ubiquitin conjugating enzyme activity, and ubiquitin protein ligase activity. In addition, these transcripts are associated with biological processes such as spermatogenesis, spermatid development, spermatid differentiation, spermatid nucleus differentiation, and cytoskeleton organization.

The protein–protein interaction (PPI) pairs of proteins that are encoded by these genes were retrieved using the Search Tool for the Retrieval of Interacting Genes/Proteins (STRING, http://string-db.org/). A network with 23 nodes and 9 PPI (protein-protein interaction) pairs was collected. This interactive network to predict potential interactions between these genes showed a PPI enrichment p-value= 0.027. This indicates that these proteins are at least partially biologically connected.

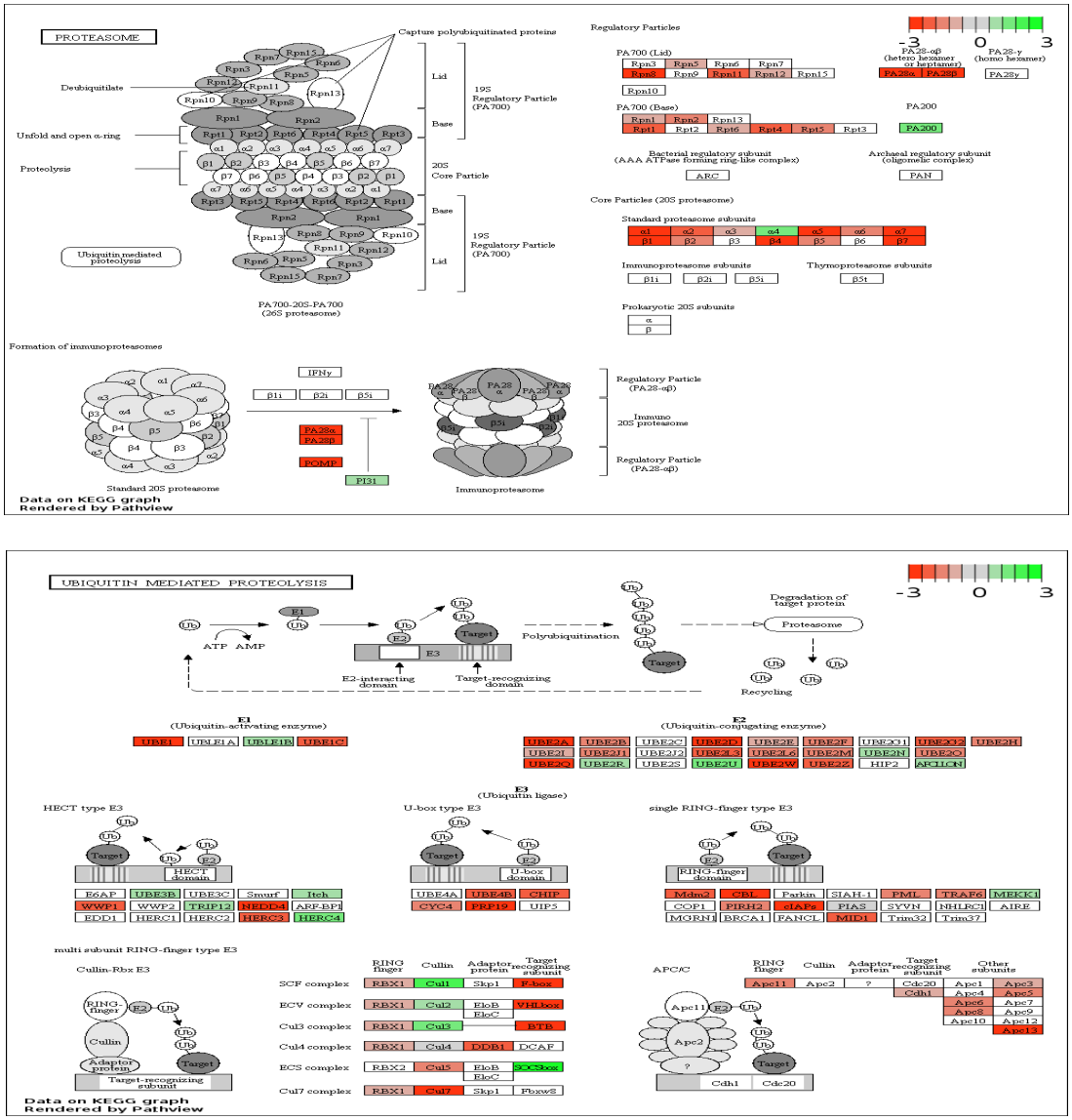


Figure 29. Pathway analysis of Proteasome and Ubiquitin-mediated proteolysis (Kegg pathway) using the Pathview server (https://pathview.uncc.edu/)[337]. Highlighted genes are components of these pathways significantly differentially expressed (DE) in teratozoospermia. DE color code indicates genes that are up (green) or down (red) regulated.

Table 18. The spermatozoal enriched down-regulated transcripts from teratozoospermic men associated with the teratozoospermia mouse phenotype obtained from functional annotation analysis (Toppgene) based on their abundancy.

`	Gene Symbol	Gene Name	References
1	MKKS	McKusick-Kaufman syndrome	[400]
2	CUL4A	cullin 4A	[401]
3	NCOA2	nuclear receptor coactivator 2	
4	UBE2J1	ubiquitin conjugating enzyme E2 J1	[402]
5	SELENOP	selenoprotein P	

6	UBE2B	ubiquitin conjugating enzyme E2 B	[403]
7	TMEM203	transmembrane protein 203	[404]
8	GOPC	golgi associated PDZ and coiled-coil motif containing	[42, 405]
9	AP1G1	adaptor related protein complex 1 gamma 1 subunit	[406]
10	ING2	inhibitor of growth family member 2	[407]
11	KRT19	keratin 19	
12	TBPL1	TATA-box binding protein like 1	[408]
13	CNOT7	CCR4-NOT transcription complex subunit 7	[344-345]
14	PAIP2B	poly(A) binding protein interacting protein 2B	[409]
15	BBS4	Bardet-Biedl syndrome 4	[410]
16	SLIRP	SRA stem-loop interacting RNA binding protein	[411]
17	TALDO1	transaldolase 1	
18	ALKBH5	alkB homolog 5, RNA demethylase	[412-413]
19	AR	androgen receptor	[414]
20	FKBP4	FK506 binding protein 4	
21	GOLGA3	golgin A3	[415]
22	ELOVL2	ELOVL fatty acid elongase 2	[416-417]
23	CSTF2T	cleavage stimulation factor subunit 2 tau variant	[418]
24	JAM3	junctional adhesion molecule 3	

When the criteria for sorting were set to log2FC (fold change) >-3 we targeted 829 genes. Surprisingly, we identified 16 genes belonging to the long non-coding RNAs family and 9 genes of the histones family which are recently recognized to potentially regulate the epigenome. The histones transcripts are biologically related to chromatin assembly or disassembly (FDR= 3.054E-17), and regulation of gene expression, epigenetic (FDR= 1.249E-5). Three mouse orthologs of the histones transcripts (H1F0, HIST1H1C, HIST1H1D) are found associated with reduced fertility phenotype (FDR=1.469E-3).

4.1.3.3.3. Up-regulated genes

Mining of the abundant up-regulated spermatozoa transcripts revealed 3301 enriched genes. These transcripts were significantly associated with spermatogenesis (FDR= 7.985E-16), cellular component morphogenesis (FDR=2.112E-4), piRNA metabolic process (FDR= 7.773E-3), sister chromatid segregation (FDR= 1.146E-2), DNA methylation involved in gamete generation (FDR= 3.598E-2), and male meiosis I (FDR=3.598E-2). Interestingly, there have been several phenotypes in mouse orthologs such as, abnormal male germ cell morphology (FDR= 3.686E-22), teratozoospermia (FDR= 4.967E-12), abnormal male germ cell apoptosis (FDR= 9.736E-10), abnormal sperm flagellum morphology (FDR= 4.863E-8), abnormal spermatocyte morphology (FDR= 3.820E-7), abnormal sperm midpiece morphology (FDR= 3.568E-6), abnormal sperm head morphology (FDR= 2.595E-5), abnormal spermatid morphology (FDR= 1.160E-4), abnormal acrosome morphology (FDR= 1.415E-4), and abnormal sperm mitochondrial sheath morphology (FDR= 9.670E-4). However, the major pathway affected is SUMOylation of DNA damage response and repair proteins (FDR= 4.466E-2). However, the analysis of genes with high log2FC (fold change) >2 targeted 476 genes. Notably, these transcripts were significantly associated with “fertility disorders” disease (FDR=6.831E-3; genes: CRISP2, CXCL8, TLR2).

Interestingly, we identified from the list of the up-regulated genes 171 mouse orthologs associated to the abnormal male germ cell morphology phenotype (FDR=3.538E-220), and 71 related to teratozoospermia (FDR= 1.194E-76) ***(Supplementary Table S10)***. These candidate genes are related to spermatogenesis (FDR= 9.042E-68), and piRNA metabolic process (FDR= 2.381E-7; genes: FKBP6, TDRD9, PIWIL2, TDRD1, MOV10L1, MAEL).

4.1.3.3.4. Long non-coding RNAs

The 282 detected lincRNA (long intergenic noncoding RNAs) and 181antisense RNAs included 52 genes belonging to the long non coding RNAs family ***(Supplementary Tables S11, S12)***. Analysis of the Gene Ontology consortium terms of these genes did not reveal any enrichment for functional classes. The RNA–protein interaction pairs were retrieved using the Search Tool for RNA–protein Association and Interaction Networks (v1.0) (RAIN - https://rth.dk/resources/rain/). A network with 45 nodes and 17 PPI (protein-protein interaction) pairs was collected (PPI enrichment p-value=0) ***(Figure 30a)***. This means that the derived proteins have more interactions among themselves than what would be expected for a random set of proteins of similar size, drawn from the genome. Such enrichment indicates that the proteins are at least partially biologically connected, as a group.

Subsequently, this gene network is used to prioritize candidate gene based on topological features in protein-protein interaction network. It was hypothesized that proteins associated with similar diseases would exhibit similar topological characteristics in PPI networks. Utilizing the location of a protein in the network with respect to other proteins (i.e., the "topological profile" of the proteins) we ranked the top 10 genes ***(Figure 30b)***. This set of genes was functionally related to cell cycle (FDR=5.140E-4), protein ubiquitination (FDR=6.886E-4), and apoptotic process (FDR=1.049E-2).

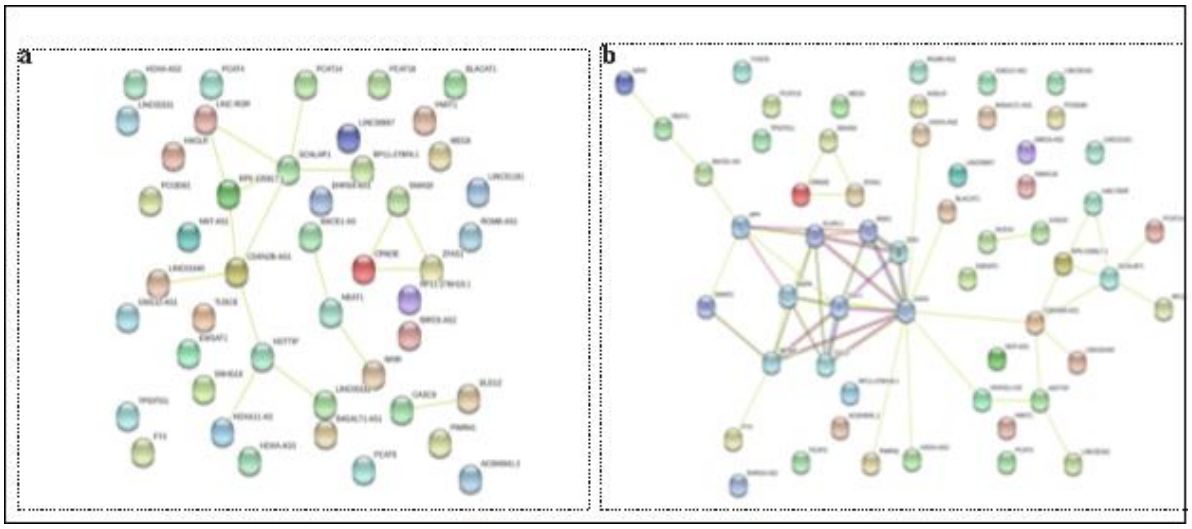


Figure 30. (a) RNA–protein interaction network map (v1.0) prepared using the RAIN web tool (https://rth.dk/resources/rain/) for the long non-coding RNAs (lncRNAs) associated with teratozoospermia. These candidate proteins were further prioritized based on topological features in protein-protein interaction network. (b) lncRNAs- top ranked candidate genes encoding proteins interaction map associated with asthenozoospermia. Colored lines denote interactions and network nodes represent proteins and lncRNAs.

Recent advances in high-throughput sequencing provide the opportunity to characterize sperm transcriptome at unprecedented depth, and identify putative key altered cellular function and pathways in infertile teratozoospermic subjects. We report whole transcriptome analysis of sperm RNA samples from infertile patients with teratozoospermia as compared to gene expression in sperm from matched-fertile men. Given the paucity of studies, we conducted this comprehensive transcriptome sequencing on sperm from fertile and teratozoospermic infertile men using next generation sequencing. Platts et al. 2007 [130] comprehensively analyzed the expression of mRNAs in teratozoospermia using microarray technology and pointed several perturbations in various pathways. In the same perspective, Fu et al. 2016 [419] identified around 2524 differentially expressed genes with a microarray approach. Therefore, our analysis detected around 7345 differentially expressed genes (DEG), which allows a meaningful interpretation of these results. Furthermore, a comparison

between our RNA-seq data and the microarray data illustrated that deep sequencing could detect a much wider dynamic range of gene expression than microarrays. Expression profiling analysis by arrays is limited by its low sensitivity due to background hybridization and occasionally reduced specificity due to cross-hybridization of probes and targets [420]. RNA-seq is a more sensitive technology and it is thus not surprising to see that we identified additional DEGs. To the best of our knowledge, this represents the most comprehensive characterization of the transcriptome in teratozoospermia, which is critical to understand the etiology of this condition at a wider level.

We first analyzed the expression difference at gene levels between infertile teratozoospermic and matched normozoospermic fertile sperm samples, and identified 7345 differentially expressed genes (DEGs). Notably, a small fraction of non-coding RNAs was also detected. As such, this suggested the power and sensitivity of the RNA-seq based approach for expression profiling.

Functional analyses indicated that these DEGs were mostly enriched in many bio-processes including: RNA processing, cellular catabolic process, ubiquitin-dependent protein catabolic process, proteasome-mediated ubiquitin-dependent protein catabolic process, chromatin organization, cell cycle, cellular response to DNA damage stimulus, and histone modification.

Along these lines, several pathways were found deregulated in teratozoospermia namely; chromatin organization, ubiquitin mediated proteolysis, deubiquitination, protein ubiquitination, cell cycle, apoptosis, and HATs acetylate histones. These alterations provided important clues for understanding the molecular mechanisms of teratozoospermia pathogenesis.

Among these pathways, are some that have been previously characterized as causal pathways in teratozoospermia pathogenesis such as the ubiquitin-proteasome pathway [130]. Several candidate genes regulating the proteasome and the cell cycle pathways were previously pointed on the basis of very few mapped genes [131, 419]. Actually, comparing to previous reports, we have identified more mapped genes with significant expression changes in several pathways. For example, more members of the ubiquitin mediated proteolysis pathway were found to be down-regulated in teratozoospermia compared to normozoospermia. In fact, KEGG enrichment in our study asserts that the majority of the ubiquitin mediated proteolysis and proteasome components are under-represented in teratozoospermia **(Figure 29)**. In this manner, the most abundant down-regulated transcripts were associated with biological processes such as: cellular catabolic process, proteasome-

mediated ubiquitin-dependent protein catabolic process, and protein ubiquitination. This clear disruption of the proteasome pathway was previously shown to be involved in the progress of teratozoospermia [130-131]. Hence, the ubiquitin-proteasome pathway was found to be involved in a subset of events related to spermatogenesis. These events include the ubiquitination of mitochondria which promotes mitochondrial shaping, and its relocation around the midpiece. Add to that, histones ubiquitination causes destabilization of nucleosomes which is a key step during meiosis and histone replacement by the transition proteins. More so, ubiquitination was found to be related to acrosome formation [421].

A further analysis of the highly abundant genes down-regulated revealed that transcripts related to the cell cycle, cytoskeleton organization, and chromatin remodeling were also significantly enriched in the spermatozoa of teratozoospermic men. It is known that sperm cell development is a cyclic process. In this perspective, aberrant expression of cell-cycle related proteins was found linked to spermatogenesis [422]. Recently, Huang et al. [131] showed that the cell cycle pathway was disrupted in teratozoospermia. In addition, it was also pointed out that “the making of abnormal spermatozoa” may be due to a pathological spermiogenesis marked by abnormalities of cytoskeletal components that cause a disorganization of organelles [359-361], and aberrant chromatin remodeling [423]. Another noteworthy finding was the down-regulation of many putative epigenetic modulators of gene expression, namely, transcripts related to post-transcriptional regulation of gene expression, chromatin organization, and histone modification.

Our data have also identified very few genes belonging to the long non-coding RNAs and histones (histone variants: H1F0, HIST1H1C, HIST1H1D) families which were very highly down-regulated in teratozoospermia. This is not surprising because several epigenetic players are important regulators of spermatogenesis [424-425]. Recently, Jenkins et al. [163] suggested that some specific epigenetic alterations in teratozoospermia may be indicative of perturbed cellular pathways during spermatogenesis. In the same perspective, the Nonsense Mediated Decay (NMD) pathway was also shown to be down-regulated. By definition, Nonsense-mediated mRNA decay (NMD) is a surveillance pathway that reduces errors in gene expression by eliminating mRNA transcripts that contain premature stop codons [426]. Until lately, it was evoked that NMD plays an essential role in spermatogenesis, spermiogenesis, and fertility [389]. It was shown that knocking-out NMD components in mice leads to infertility [426].

Accumulated sperm nuclear alteration is a hallmark of teratozoospermia genome, where a variety of chromosomal and DNA aberrations have been identified. In fact, several

published data showed that the rate of aneuploidy and DNA fragmentation in gametes of patients with severe teratozoospermia is elevated [213, 366-367]. Recently, a paradigm for the role of oxidative stress in promoting teratozoospermia has been addressed [427]. It could be postulated that teratozoospermia-associated nuclear damage is mediated by oxidative stress, apoptosis, and faulty meiosis. Spermatogenic flaws may include germ cell apoptosis [428], protamine deficiency [429] and meiotic segregation errors, resulting in DNA integrity and chromosomal segregation defects.

Particularly interesting is the observation of many mouse orthologs related to male infertility, abnormal male germ cell morphology, and teratozoospermia phenotypes. Using the dataset of the teratozoospermia phenotype, enrichment analysis showed that the majority of the transcripts are related to spermatogenesis, cell cycle and, organelle assembly. Markedly, genes associated with spermatogenesis, spermatid development, spermatid differentiation, and spermatid nucleus differentiation were all down-regulated. These findings are consistent with reported data on the origin of teratozoospermia that was traced as far back as the pachytene stage of spermatogenesis [130].

Our data have also identified several mouse orthologs related to abnormal male germ cell apoptosis. These transcripts were found to be over-represented in teratozoospermia. However, this defect in spermatogenesis is in part reflected by the significant up-regulation of the SUMOylation pathway. In fact, SUMO proteins are involved in the stress response of DNA damage and repair proteins during spermatogenesis [430].

Moreover, we reported that many genes implicated in piRNA metabolic process are significantly up-regulated in teratozoospermia. The PIWI/piRNA machinery has been implicated in suppressing transposable elements and protecting the integrity of the genome in germ cells [431]. Thus, abnormal accumulation of piRNAs would interfere with sperm maturation and the proper progression of male germ cell development [432]. Several lines of evidence strongly suggest that piRNAs also play an important role in epigenetic control [433]. These players can regulate the methylation levels of histones or DNA to influence the transcription of selected protein-encoding genes, imprinting loci, or transposons [434].

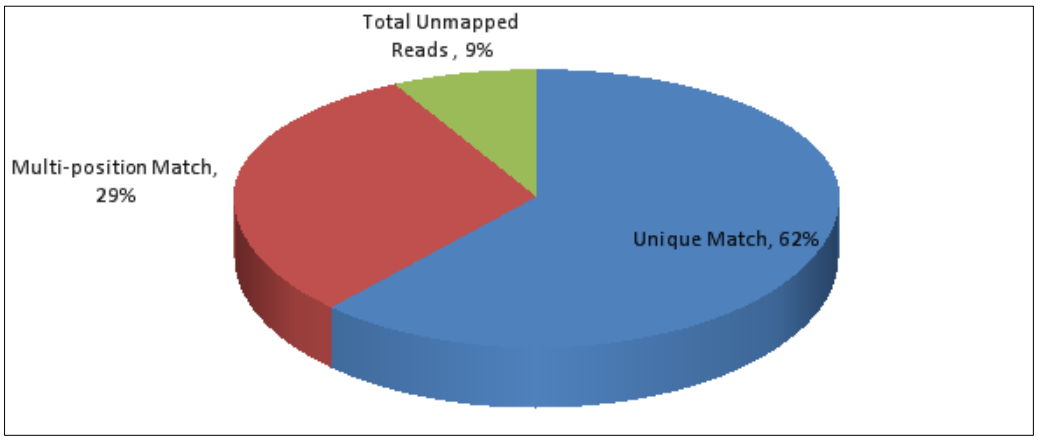
Taking advantage of the RNA-seq technology, we were able to analyze differential expression of some non-coding RNAs, including long non-coding RNAs, providing a way to understand fine scale regulation of gene expression. It is intriguing that our dataset also allows us to identify long non-coding RNAs (lncRNA) differentially expressed in teratozoospermia that are not previously reported. To our knowledge, our study is the first to address the deregulation of long non-coding RNAs in teratozoospermia using RNA-seq. We

have identified 52 long non-coding RNAs that were either overexpressed or under-expressed. The topological analyses using the interactome of lncRNAs suggest that lncRNAs can control various processes, notably cell cycle progression, ubiquitination and apoptosis. This is not unforeseen since the latter processes were deregulated in teratozoospermia as well.

4.1.4. Dissection of genes expression signatures associated with abnormal sperm phenotypes in infertile men by RNA seq analysis

4.1.4.1. Repertoire of DEG's in sperm from infertile men as determined by Next Generation Sequencing of cDNA libraries.

As previously mentioned, we identified from oligozoospermic samples a total of 10087 differentially expressed genes, 14606 DEG from asthenozoospermic samples, and 7345 differentially expressed genes from teratozoospermic patients. A summary of the RNA-seq annotation statistics for the transcriptome work is presented in ***Figure 31***.



A majority of RNA-seq reads map uniquely to the human reference genome sequence. A pie chart summarizing results of alignment of Illumina RNA-seq reads from sperm samples to the human reference genome sequence. Around 9% of all RNA-seq reads failed quality checks and were filtered out of our data set. The remaining 91.05% of RNA-seq reads passed quality checks. The blue portion of the chart represents reads that mapped uniquely to the potato genome sequence (61.9% of all RNA-seq reads). The red portion of the chart represents reads that mapped to multiple locations of the potato genome sequence (29.15%). The green portion of the chart represents reads that failed to map to the human genome sequence (8.95%).

Figure 31. Pie chart representing the alignment statistics of reads which align to reference genome.

4.1.4.2. Data analysis strategy

To determine whether spermatozoa isolated from patients with different sperm pathologies carry gene expression signatures that allow discrimination between each abnormal phenotype, a data analysis step was conducted. The patients included in the training sets used for the identification of the classifier genes, that are capable of discriminating each group, were selected very carefully. We filtered the complete set of genes to retrieve the reads (DeSeq2) which allowed the detection of the enriched transcripts largely representative

of the studied phenotype. In addition, we also used the set of mouse orthologs associated with “male infertility” phenotype since it is a common finding between the different phenotypes.

The classifier genes also described as transcriptional markers were considered the most representative of the realistic clinical phenotype profile. We intended to identify biomarker candidates which can discriminate different abnormal sperm phenotypes. Therefore, we applied a Venn diagram comparison of the three gene lists to select gene expression patterns unique for each phenotype.

4.1.4.3. Molecular characterization of sperm pathologies types of infertile men

To molecularly characterize these three sperm pathologies, we first identified the most abundantly differentially expressed genes as the "signature" genes. We carefully chosen the sorting criteria in order to recover a narrowed list of specific genes associated with each pathology. We recognized 865 genes in the asthenozoospermia group which were identified as the "signature" genes meeting the following criteria: (1) DeSeq2 normalized read counts >500 and (2) False discovery rate < 5%. In parallel, we also identified 2542 and 661 genes respectively in teratozoospermia and oligozoospermia. This set of genes was identified as the "signature" genes meeting the following criteria: (1) DeSeq2 normalized read counts >50 and (2) False discovery rate < 5% (3) padjvalue< 0.05 ***(Figure 32a)***.

To depict more detailed molecular portraits of these three pathologies, we analyzed which Gene Ontology terms associated with these signature genes were significantly enriched.

(i) The most significantly under-represented pathways in asthenozoospermia are the ribosome (FDR=2.664E-9; 30 genes) and cilium assembly (FDR= 2.580E-3; 22 genes).

(ii) Not surprisingly, very different processes are underlined by the signature genes of teratozoospermia: the most over-represented biological processes include spermatogenesis (FDR=2.295E-7; 99 genes), cellular component assembly involved in morphogenesis (FDR= 5.281E-7; 68 genes). Whereas the most over-enriched cellular components include sperm part, sperm principle piece, and sperm flagellum. Interestingly, among this list of deregulated genes we found 47 over-expressed genes in teratozoospermia group which encode mouse orthologs related to “teratozoospermia” phenotype (FDR= 5.237E-7).

(iii) Again, quite distinct pathways were found to underlie oligozoospermia. The most under-represented pathways are the Sperm: Oocyte Membrane Binding (FDR=5.006E-3; 3 genes: IZUMO1, IZUMO2, IZUMO4), fertilization (FDR=6.088E-3; 5 genes: IZUMO1,

IZUMO2, IZUMO4, CATSPER1, ADAM20), and ensemble of genes encoding core extracellular matrix including ECM glycoproteins, collagens and proteoglycans (FDR= 1.930E-7; 23 genes; e.g. COL1A1). Along these lines, the most significantly under-represented biological process is the fusion of sperm to egg plasma membrane involved in single fertilization (FDR= 2.554E-2; 5 genes: EQTN, CATSPER1, NOX5, IZUMO1, SPESP1).

Furthermore, we next plotted the common genes differentially expressed (DE) between the three data sets as heatmaps, and identified 13 genes (CABS1, IFT172, RBM5, REC8, SGO2, SMC1B, SOX30, SPATA31D1, SUN5, TDRD7, TLR3, TNP1, TSSK6) ***(Figure 32b)***. Functional analysis revealed that these genes are biologically related to spermatogenesis, sister chromatid cohesion, male meiotic nuclear division, and fertilization: exchange of chromosomal proteins.

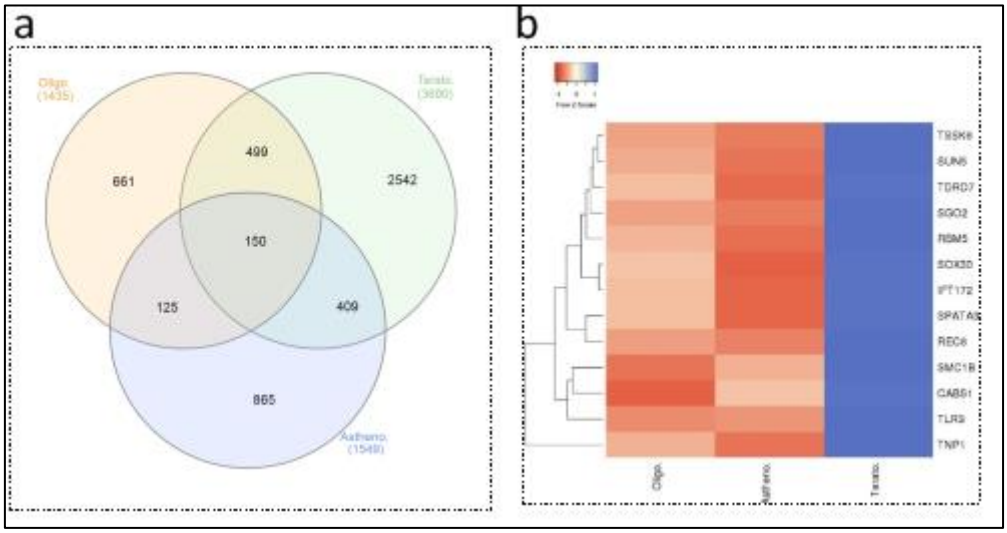


Figure 32. Transcriptional landscape of human sperm RNAs in ejaculates of infertile men with different sperm abnormalities types.

(a) *Venn Diagram of the abundantly differentially expressed genes identified by RNA seq; 150 genes were identified as the common signature genes. (b) Supervised hierarchical clustering analysis using 13 RNAs that were consistently up-regulated or down-regulated between the three data sets. Shades of red and blue are used to illustrate whether the expression value is above (blue) or below (red) the mean expression value across all samples (each row in the data was normalized from −1 to +1).*

4.1.4.4. Identification of the best set of genes discriminating the different phenotypes of sperm pathologies of infertile men

In an effort to identify a minimal set of genes that best characterize the three phenotypes and can form the basis for a distinct gene profile, in a similar fashion, we analyzed mouse orthologs related to “male infertility” phenotype to define each distinct phenotype.nFrom these, 10 genes were identified as the common signature genes: BSG, PRSS37, MEIOC, RBM5, TSSK6, MORC1, GRID2, and TEX15, DNAH1, and VPS13A

(Figure 33). As expected, the first eight genes are found associated with abnormal spermatogenesis (FDR=1.006E-6), one to the axonemal dynein complex assembly and one to the Golgi to vacuole transport ***(Table 19)***.

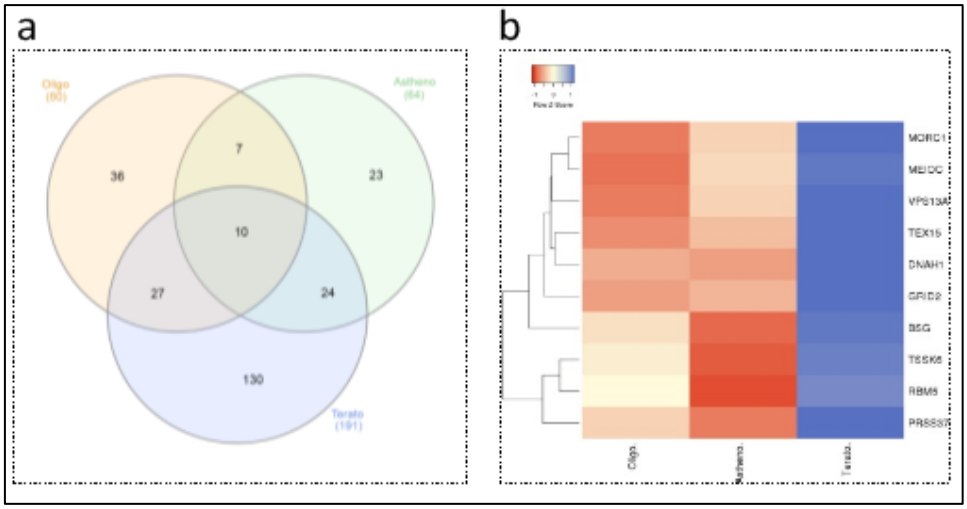


Figure 33. Identification of the best set of genes discriminating the different phenotypes of sperm pathologies of infertile men.

(a) Venn Diagram of the differentially expressed genes orthologs related to "male infertility" phenotype identified by RNA seq; 10 genes were identified as the common signature genes. (b) Supervised hierarchical clustering analysis using 10 orthologous genes associated with male infertility that were consistently up-regulated or down-regulated between the three data sets. Shades of red and green are used to illustrate whether the expression value is above (blue) or below (red) the mean expression value across all samples (each row in the data was normalized from −1 to +1).

Table 19. Biological pathway analysis of the set of common signature genes differentially expressed orthologs related to "male infertility" phenotype identified by RNA seq.

GO-Term	FDR B&H	Genes observed
Biological process		
Reproductive process	1.787E-8	10
Spermatogenesis	9.198E-8	8
Sister chromatid cohesion	3.945E-3	3
Spermatid differentiation	5.244E-3	3
Male meiotic nuclear division	1.316E-2	2
Fertilization, exchange of chromosomal proteins	1.339E-2	1

The most significant biological processes (ToppFun from the ToppGene Suite; https://toppgene.cchmc.org/) over-represented by the 10 signature genes characterizing the three sperm abnormalities.

(i) Among the 36 classifier genes deregulated in oligozoospermia, 11 genes are associated with "abnormal sperm number" mouse phenotype (RIMBP3, TYRO3, LHCGR,

NHLH2, CEP131, NR5A1, BRCA1, KIT, AXL, AK7, ESR1) and 3 genes (LHCGR, NR5A1, CATSPER1) linked to oligospermia human phenotype; that can be considered as specific signature genes. These genes are biologically related to spermatogenesis (FDR=3.163E-8, 8 genes).

(ii) Likewise, we also identified 130 genes specific to teratozoospermia. The spermatozoa carried many transcripts that might regulate different stages of spermatogenesis (FDR=4.810E-33; 33 genes), such as spermatid differentiation, acrosome assembly and organelle assembly; actin cytoskeleton organization (FDR=4.144E-2; NPHP1, MKKS, KRT19, PAFAH1B1, BBS4, JAM3, PICK1), and fertilization (FDR=4.654E-2, CATSPERD, ZPBP, SPEF2, SLIRP). Remarkably, we identified 4 enriched transcripts related to human male infertility (FDR= 3.748E-3; NPHP1, MKKS, BBS7, BBS4), and 1 transcript (DPY19L2) associated with sperm head anomaly (FDR=4.337E-2).

(iii) We also analyzed the set of genes deregulated in asthenozoospermia. The most enriched processes include spermatid development (FDR= 1.141E-4; KDM3A, STK11, H3F3B, AFF4, ODF2), spermatogenesis (FDR= 2.875E-4; KDM3A, STK11, H3F3B, AFF4, HTT ODF2, ARID4B), microtubule cytoskeleton organization (FDR= 9.488E-3; ATXN7, CDK5RAP2, DICER1, DYNC1H1, HTT), mitochondrial membrane organization (FDR= 1.055E-2; STAT3, HTT, HSP90AA1), and regulation of cilium assembly (FDR= 3.425E-2, HTT, ODF2).

Notably, 5 genes deregulated are related to “asthenozoospermia” mouse phenotype (FDR=1.127E-3; KDM3A, JUND, NPEPPS, LIPE, ODF2).

4.1.4.5. Transcriptional landscape of long non-coding RNAs in sperm from infertile men

A substantial fraction of transcripts in mammalian genomes do not code for proteins. These include lncRNAs (long non-coding RNAs), which is arbitrarily defined as greater than 200 bases in length, to distinguish them from functional small non-coding transcripts, such as microRNAs (miRNAs), and piwi-interacting RNAs (piRNAs).

LncRNAs (also known as processed transcripts) by definition are found within protein coding genes, overlapping with promoters, exons or introns in either sense or antisense orientations [435]. We identified several lncRNAs, then analyzed the global spermatozoal long ncRNA transcription profiles and found that few lncRNAs were differentially expressed in each group.

Five lncRNAs were found common between the 3 phenotypes from the set of lncRNAs: PCGEM1, PART1, PCAT4, FTX, and NORAD.

Next, we compared the long ncRNA expression levels between the different phenotypes and identified 9 DE-lncRNA specific to oligozoospermia, 9 specific to asthenozoospermia and 32 specific to teratozoospermia. Although the difference in expression between infertile and fertile men was statistically significant, analysis of the Gene Ontology consortium terms of these genes did not reveal any enrichment for functional classes. The importance of this finding will require further study.

Male infertility results from abnormalities in multiple processes that contribute to spermatozoa dysfunction. In this study, we tested the hypothesis that a set of genes with distinct expression patterns between fertile and infertile men can classify male infertility status for different sperm pathologies. Using RNA-Seq data, we identified genes that were differentially expressed between oligozoospermic, asthenozoospermic and teratozoospermic individuals.

We were the first to compare and identify that the three phenotypes of sperm pathologies, based on gene expression profiling using massively parallel genomic sequencing, were distinct biological entities and are associated with significant differences in dysregulated cellular and molecular features.

Subsequently, this has been validated both by us and other groups in different types of sperm abnormalities. There have been few microarray investigations performed to study globally the spermatozoal transcripts in infertile men with abnormal semen analysis [167, 170-171, 187, 272, 356].

Due to the filtering of genes from the complete extensive data set of DEGs, only few sets of genes could be robustly identified.

These three sperm abnormalities are easily distinguishable and their expression profiles seem to be different but were not integratively analyzed. However, the cellular pathways biological processes affected are not known in detail.

We show here that the differences in gene expression patterns between the three phenotypes reflect levels of alteration of distinct signaling pathways. These changes might have been pre-programmed already at a relatively early stage during spermatogenesis and hence, imply that the fate of the sperm is already set. Spermatogenesis involves mitosis and self-renewal of spermatogonia, meiosis in spermatocytes and the differentiation of haploid

spermatids (termed spermiogenesis) [436]. Any alterations in these processes may trigger complex deregulations in many signal transduction pathways.

Specifically in this study, a more in-depth molecular characterization of these phenotypes of spermatozoa was carried out and provided new insights into the biology of these abnormalities at the molecular level. The distinct and characteristic molecular mechanisms revealed by the protein classification and biological pathway analysis that the three phenotypes may represent biologically distinct entities. For example, our results indicated that asthenozoospermia showed down-regulation of genes involved in ribosome and cilium assembly. Altered sperm motility has been found to be largely related to altered sperm motility apparatus structure and mitochondrial bioenergetic deficit. A paradigm was suggested according to which several nuclear genes notably ribosomal proteins promote the expression of key components of the mitochondrial transcription and translation machinery that are necessary for the mitochondrial biogenesis and function [381].

This is in accordance with a previous microarray study [112]. Add to that, the molecular architecture of the sperm flagella was found altered in asthenozoospermia. In this manner, multiple defects affecting the sperm motility apparatus were identified such as; cilium assembly and cilium organization [171, 273, 382].

We also found that teratozoospermia is associated with deregulation in spermatogenesis and cellular components involved in morphogenesis. Several reports pointed out that “the making of abnormal spermatozoa” may be due to a pathological spermatogenesis and spermiogenesis marked by abnormalities of cytoskeletal components that cause a disorganization of organelles [167, 360-361].

Analysis of the dataset of oligozoospermia revealed that sperm oocyte membrane binding and fertilization are altered. In fact, the frequency of defective sperm–ZP interaction was significantly higher in subfertile men with abnormal semen than in those with normal semen, being highest with oligozoospermia [437]. Analysis of the clinical data showed an association between oligospermia and a decreased fertilization rate [438]. In addition, a case report revealed the occurrence of *CATSPER* gene deletion in human male infertility with impaired spermatogenesis [439]. Furthermore, in our study, ensemble of genes encoding core extracellular matrix including ECM glycoproteins, collagens and proteoglycans are under-represented in oligozoospermia. Collagen has been found to play a potential role in mice spermatogenesis and in human testes as well [440], in particular collagen mediates the detachment and migration of germ cells in the testes [441]. Furthermore, the down-regulation

of Col1a1 induced differentiation in mouse spermatogonia and suppressed their self-renewal [442].

Moreover, combined analysis showed that the three sperm abnormalities shared altered expression levels of a set of 13 sperm-specific genes, which are the following: *CABS1, IFT172, RBM5, REC8, SGO2, SMC1B, SOX30, SPATA31D1, SUN5, TDRD7, TLR3, TNP1, TSSK6.* The majority of these spermatozoal transcripts were related to spermatogenesis, and some might play a role in sister chromatid cohesion and fertilization. One such transcript, SOX30, observed in the present study has been recently reported to critically regulate spermatogenesis. Furthermore, SOX30 gene-knockout mice have been found to be infertile [443].

Another transcript, TNP1 is involved in spermatid chromatin condensation, and plays a role in male fertility. TSSK6 is expressed predominantly in the elongating spermatids and is involved in postmeiotic chromatin remodeling and fertilization [41]. Mice lacking TSSK6 had been found to have a compromised acrosome reaction, egg fusion and infertility[444]. Add to that, a polymorphism of the *TSSK6* gene was found associated with spermatogenic impairment in humans [445].

The minimal set of genes that best characterized the three different phenotypes was identified based on differential expressed orthologs associated with “male infertility”, the commonly enriched genes set related to a phenotype. Clustering of expression data from all three data sets grouped the three phenotypes together into a narrowed category to facilitate the biological interpretation and to visualize functionally grouped terms. Hence, these genes provide a robust set of potential molecular markers. The majority of these 10 genes are directly related to spermatogenesis. It is known that alterations in spermatogenesis are capable of altering sperm phenotype [167, 170, 290]. A number of genes involved in this process have been reported to be dyregulated in infertile men. It is evident that defects in spermatogenesis will profoundly alter structure, function, metabolism, and physiology of the spermatozoa.

For example, low levels of PRSS37 protein in human sperm was found associated with many cases of unexplained male infertility [446]. This protein was found to be involved in the fertilization process. In addition, male mice with deficient TSSK6 are infertile due to considerable sperm reduction and abnormal motility and morphology of spermatozoa [447]. A recent paper in 2016 by Pereira et al. revealed that MORC1 was differentially expressed infertile men using a next generation sequencing approach [257]. Morc1 is involved in

chromosomal pairing during the zygotene stage of meiosis and is a regulator of the epigenetic landscape of male germ cells during the period of global *de novo* methylation [448].

The parallel identification of specific RNA fingerprints for each sperm abnormality, on the other hand, revealed very subtle distinctions between the different phenotypes. Hence, a restrained number of genes were pointed out and were enriched for bioprocesses and mouse orthologs related directly to the phenotype under investigation.

It is intriguing to find long ncRNA expression patterns between the different groups. Interestingly, we identified 5 common lncRNAs shared by the three groups: PCGEM1, PART1, PCAT4, FTX, and NORAD. Recently, lncRNAs were considered to function as epigenetic modulators and hence impacting gene expression [449].

For example, FTX was found implicated in imprinting X-chromosome inactivation in mouse embryos [450]. NORAD lncRNAs are induced after DNA damage, which we termed "noncoding RNA activated by DNA damage". They preserve chromosomal stability by serving as a molecular decoy for highly conserved PUMILIO proteins, which repress the stability and translation of mRNAs to which they bind leading to a normal mitosis [451].

To our knowledge, our study is the first to address the deregulation of long ncRNAs in human sperm from infertile men presenting various sperm abnormalities using RNA-seq. Further studies are warranted to define the reason for these differences.

4.2. Characterization of the sperm global epigenetic landscape associated with oxidative stress in infertile men

In this study, we examined; in a population-based case-control study of men; the levels of sperm histone retention, sperm chromatin structure, DNA damage, global levels of DNA methylation, global DNA hydroxymethylation and 5-formylcytosine levels. With this epigenetic analysis, a clearer picture is emerging of the various factors at play in the etiology of male infertility. The clinical and semen parameters of the men included in this study are summarized in ***Table 20.*** As expected, mean sperm concentration, motility, morphology and viability are significantly lower in the infertile group compared with fertile men. Regarding sperm DNA damage markers, the mean level of semen reactive oxygen species production (NBT) and DNA fragmentation (SCD) were significantly increased in infertile men compared to fertile men.

Significant differences were observed among fertile and infertile subjects with respect to their epigenetic profiles. As compared with controls, patients presented significantly higher mean % histones retention (using Aniline blue test) and % chromatin condensation abnormality (using Toluidine blue assay). In the same optics, the level of global sperm DNA methylation (mean %) was significantly lower in sperm from infertile patients. Interestingly, a significant increment of the mean percentages of hydroxymethylation and 5-formylcytosine were observed in infertile subjects as compared with controls ***(Table 20)***.

Our results indicate that the level of global sperm DNA methylation is associated with sperm concentration (r=+0.714; *P*=0.000), motility (r=+0.602; *P=0.000*) and morphology (r=+0.523; *P=0.000*). Hence, infertile men have significantly lower levels of global sperm DNA methylation than fertile men.

Additionally, the level of global sperm hydroxymethylation was also associated with sperm count (r=-0.683; *P*=0.000), motility (r=-503; *P*=0.000) and morphology (r=-0.678; *P*=0.000). We have found a positive association between sperm ROS levels (NBT %) and sperm DNA fragmentation (SCD %) (r=+0.403; p=0.005). In addition, sperm levels of histone retention demonstrated significant negative correlation with global sperm DNA methylation (r=-0.675; *P*=0.000). Likewise, sperm packaging chromatin abnormalities (DNA integrity) presented a significant negative correlation (r=-0.598; *P*=0.000) ***(Figure 34)***. In the same perspective, we also identified a negative correlation between global sperm DNA methylation and both sperm ROS (r=-0.607; *P*=0.000) and sperm DNA fragmentation (r=-0.498; *P*=0.000) ***(Figure 35)***.

Contrastively, our data demonstrated also that there is a positive association between sperm global DNA hydroxymethylation and, both sperm ROS (r=+0.44; *P*=0.002) and sperm DNA fragmentation (r=+0.513; *P*=0.000) ***(Figure 36)***.

Table 20. Comparison of sperm quality between the infertile and fertile study groups. All values are expressed as mean ± standard deviation and analyzed using the Mann–Whitney U test for two non-paired data.

		Fertile Mean±SD	**Infertile Mean±SD**	**p-value**
Demographics	*N*	23	33	
	Age	34.13±5.38	35.12±5.03	N.S
Conventional sperm parameters	Sperm concentration ($x10^6$/ml)	67.61±17.54	11±8.83	0.000
	Motility (%)	54.82±12.7	28.61±15.4	0.000
	Normal morphology (%)	23.56±7.55	9.44±5.21	0.000
	Vitality (%)	69.47±13.5	54.11±19	0.004
Sperm damage	ROS production (NBT positive %)	11.19±4.6	25.42±13	0.000
	DNA fragmentation (SCD %)	19.7±3.05	36.07±8.13	0.000
Chromatin abnormalities	Histone retention (Aniline blue %)	21.91±5.3	44.67±13	0.000
	Chromatin integrity abnormality (Toluidine blue %)	23.21±6.2	44.34±11.7	0.000
Global epigenetic analysis	Global sperm DNA methylation (%)	31.4±2	26.11±3.1	0.000
	Global sperm DNA hydroxymethylation (%)	0.07±0.12	1.69±0.6	0.000
	Sperm 5-formylcytosine levels (%)	0.038±0.15	0.0468±0.4	0.000

N.S, not significant ($p \geq 0.05$)

The characterization of the epigenetic changes in the sperm of infertile men will provide new insights about the pathophysiological processes that are involved in this disorder. In the current work, we used several *in-situ* microscopy approaches to characterize the sperm chromatin as well as colorimetric ELISA assays to depict global DNA modifications. Understanding the complete sperm characteristics from conventional and nonconventional parameters in sperm from infertile men will foster a clearer picture of the epigenetic aberrations in “abnormal” sperm. Accordingly, we portrayed the global epigenetic landscape associated with male infertility.

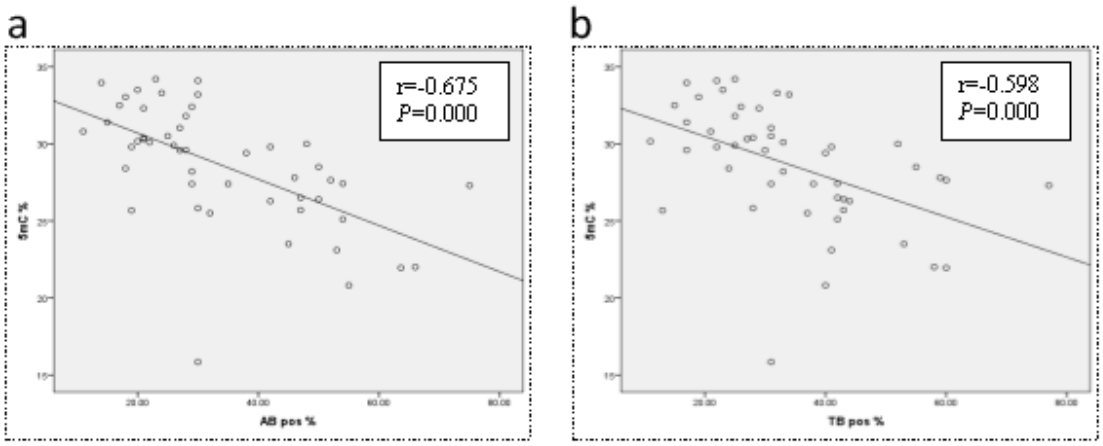


Figure 34. Relationship between sperm global DNA methylation with sperm histone retention (abnormal chromatin packaging) a) by Aniline blue test and sperm chromatin alteration b) by Toluidine blue assay.

Statistical analysis was performed using the Spearman Rank Order correlation test.

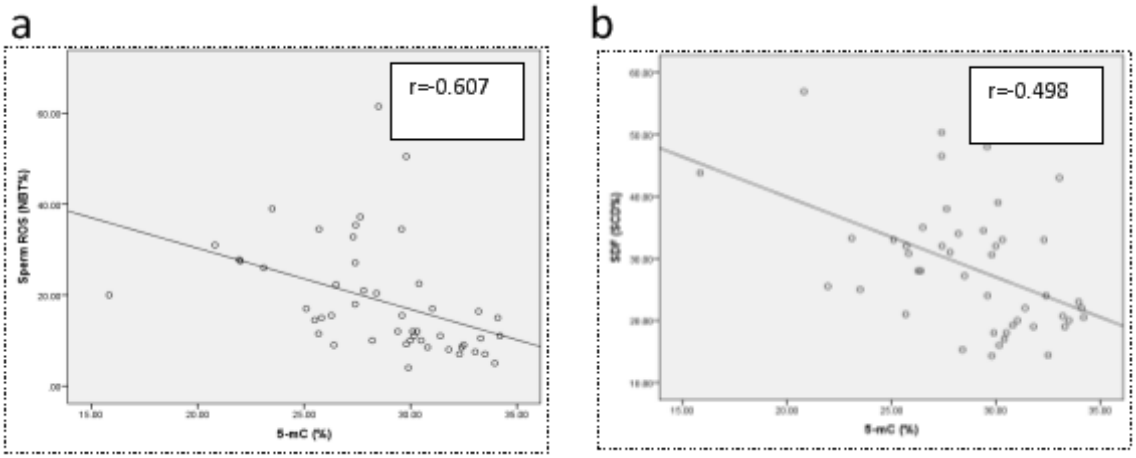


Figure 35. Relationship between sperm global DNA methylation with sperm ROS a) by NBT test and sperm DNA fragmentation (SDF) b) by SCD test.

Statistical analysis was performed using the Spearman Rank Order correlation test.

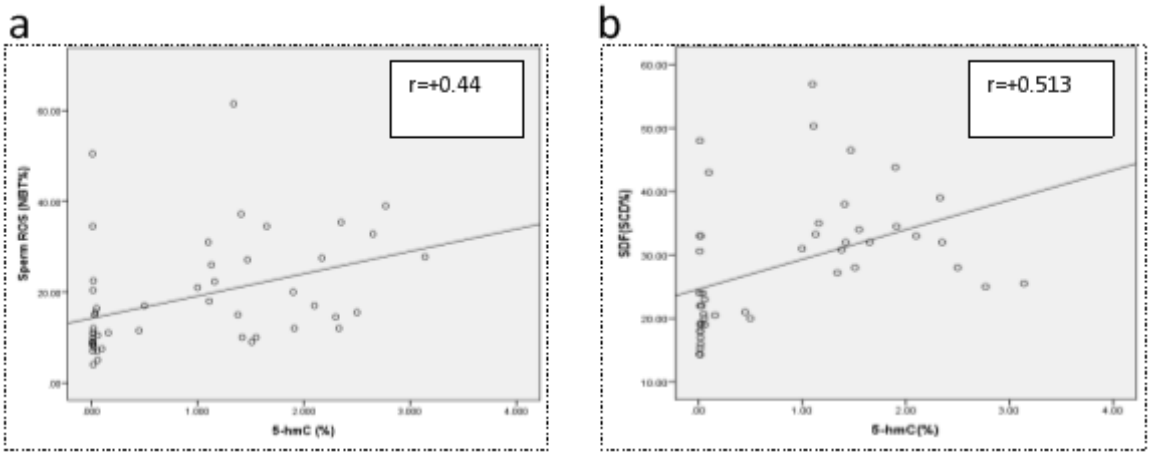


Figure 36. Relationship between sperm global DNA hydroxymethylation with sperm ROS by NBT test (a) and sperm DNA fragmentation by SCD test (b).

Statistical analysis was performed using the Spearman Rank Order correlation test.

In general, properly selected population of infertile men is characterized by a reduction of the sperm concentration, motility, morphology and vitality with values below the reference values set by the WHO for human semen characteristics [452]. Therefore, our results showed a strongly significant decrease in the aforesaid sperm conventional parameters.

To assess the whole characteristics of the infertile population, we elected to evaluate the parameters marking peculiarities happening during and post-spermiogenesis. First off, we have found alterations in the sperm chromatin, thus sperm from infertile patients presented significant high levels of histone retention (% Aniline blue positive=44.67% vs 21.91%; $p<0.01$) and abnormal chromatin (% Toluidine blue positive=44.34% vs 23.21%; $p<0.01$). These abnormalities reflect the defective remodeling state of sperm chromatin during spermiogenesis. Our results are in keeping with previous reports showing abnormal sperm chromatin structure in spermatozoa of infertile men as compared with controls, with persistence of histones [344, 453], and altered DNA integrity [454]. From an epigenetic perspective, histones play a pivotal role in epigenetic marking, thus the presence of residual histones will alter the epigenetic regulation. This finding was concordant with several previous reports of altered epigenome related to histones retention in the sperm of infertile men [173, 455].

More so, it is know that the dynamic states of chromatin structures tightly govern activation or repression, and function of the genome. Therefore, an “open” or “closed” structure of the chromatin is considered a keystone epigenetic event directed by various epigenetic layers and players [185]. In this study, we have found that sperm chromatin in infertile men presented an “open” structure, owing to the high percentage TB positive sperms, thereby chromatin might be less compact and more accessible to the dye. This state can be indirectly considered an epigenetic event.

Secondly, we also found that sperm from infertile men is characterized by ROS accumulation (NBT positive %=25.42; $p<0.01$). Therefore, several studies showed strong evidence for an association between sperm oxidative stress and male infertility [456-457]. From a classical view, looser chromatin packaging increased sperm oxidative damage susceptibility to genotoxic ROS assaults. Along these lines, Aitken et al. (2009) [458], proposed a new hypothesis “the two-step hypothesis” that assigns the sperm DNA damage to a defective remodelling of sperm chromatin which renders the nucleus more vulnerable to ROS attacks that will create DNA strand breaks. As previously pointed out, ROS attacks will generate DNA stand breaks as well as oxidized base adducts, 8-OHdG [459-460]. Likewise, our data indicate relative high levels of sperm DNA fragmentation in the infertile group.

Various studies assessed sperm DNA damage in sperm from infertile men using different assays reconcile our data [93, 461]. Additionally, we have found a positive association between sperm ROS levels (NBT %) and sperm DNA fragmentation (SCD %) ($r=+0.403$; $p=0.005$).

Currently, a massive effort is directed at providing better insight into sperm epigenetic signatures and their roles in infertility etiology. In this regard, we conducted a global analysis of DNA epigenetic changes using colorimetric assays. There was a significant decrease in global sperm DNA methylation levels (measured by quantification of 5-methylcytosine levels) in the infertile group (26.1% vs 31.4%; $p<0.01$). This finding is congruent with previous reports showing differential DNA methylation profiles in spermatozoa of infertile men as compared to normozoospermic controls [432, 434, 508].

As it was mentioned earlier, ROS can cause a wide range of DNA lesions including base modifications and strand breakage. Such DNA lesions have been shown to interfere with the ability of DNA to function as a substrate for the DNMTs [462-463], and more clearly by reducing the methyl-accepting ability of DNA. Thereby, this resulted in a global DNA hypomethylation.

Add to that, at the level of the CpG islands the incorporation of 8-OHdG and 5-hydroxymethylcytosine (the oxidation by-product of 5- methylcytosine) in the MBP (methyl-CpG binding protein) recognition sequence of the MeCP proteins (methylated CpG binding proteins) resulted in the significant inhibition of the binding affinity of MBP hence, impeding the process of DNA methylation [464]. Basically, methylated CpGs at CpG island are associated with remodeling of the chromatin structure; by recruiting methylated CpG binding proteins MeCP 1&2 [140].

Interestingly, global sperm DNA methylation demonstrated significant negative correlations with sperm chromatin integrity and histone retention ***(Figure 34)***. Such findings could be explained as the presence of methyl groups was found necessary to direct the assembly of DNA into a “closed” tightly packed chromatin structure [465].

In the light of these data, we have also identified significant negative relationships between global sperm DNA methylation levels and both sperm ROS ($r=-0.607$; $p<0.01$) and sperm DNA fragmentation ($r=-0.498$; $p<0.01$) ***(Figure 35)***. Hence, we could speculate that hypomethylated DNA is more prone to oxidative DNA damage, a finding consonant with the conclusions reached in several other studies [456, 466]. More so, we reported a positive association between global sperm DNA levels and conventional sperm parameters which is in

concert with a previous study [467]. It is proposed that DNA methylation is a key regulator of transcription [468], this paradigm explains why infertile men presented altered gene expression [170-171, 187]

Very recently, the emergence of DNA demethylation pathways drew attention to their potential relevance in epigenetic variation. Active DNA demethylation involves enzymatic removal of methyl group via intermediates (5-hydroxymethylcytosine and 5-formlycytosine) using TET enzymes [143]. Noteworthy, a significant reduction of TET1-3 mRNAs in sperm was reported in cases of oligozoospermia and/or asthenozoospermia compared to fertile individuals [161].

In the current research, the levels of the oxidized derivatives of 5methylcytosine: 5-hydroxymethylcytosine (5-hmC) and 5-formylcytosine (5-fC) were rised significantly in sperm derived from infertile patients (5-hmC%: 1.69 vs 0.07; 5fC%: 0.0468 vs 0.038; $p<0.01$). Apparently, only one study addressed this issue, and found a strongly negative association between the percentage of highly hydroxymethylated spermatozoa (measured by immunocytochemistry) and sperm normal morphology, and sturdy positive correlation to sperm DNA fragmentation [469]. In the same perspective, Wang et al. (2015) [470] revealed differences in genome-wide 5hmC patterns, notably at specific imprinted loci, in globozoospermia. In addition to being an intermediate in the active demethylation of 5mC, 5hmC appears to persist as a stable modification in many cell types [471]. Strikingly, our data demonstrated that there are positive relationships between sperm hydroxymethycytosine levels (5-hmC %) and, both sperm DNA fragmentation ($r=+0.513$; $p<0.01$) and sperm ROS ($r=+0.46$; $p<0.01$) ***(Figure 36)***. This finding is probably attributable to the modifications of DNA bases by the direct attack of ROS. In addition, our data demonstrated a negative association between hydroxymethylation levels of DNA and conventional sperm parameters.

It was pointed out that oxidation by hydroxyl radicals of cultured somatic cells can lead to the formation of 5hmC and 5fC from 5mC, initiated by abstraction of an H-atom from the methylgroup [472-473]. A possible integrated mechanistic scheme on the role of 5hmC in stress response was suggested by Chia et al. (2011) [474]. This calls for a new paradigm between 5-hydroxymethylcytosine (5-hmC) patterns and environmental factors.

Briefly, in this model oxidative stress alters the mitochondrial redox, thereby triggering series of biochemical signaling events, which affect TET proteins and change DNA methylation and hydroxymethylation patterns [475].

Regarding the global epigenetic landscape in sperm from infertile men, we will try to piece together events from our findings, by presenting the discussed sperm dimensions in turn. Sperm from infertile men is characterized by several epigenetic alterations that we will dissect and summarize them very briefly here. It is attractive to speculate that male infertility may ascribe to several intertwined factors. Accordingly, sperm from infertile men presents high histones retention, owing to a defective spermiogenesis, which will lead to poor sperm chromatin packaging. This renders the chromatin more susceptible to ROS attacks. In fact, male infertility is associated with high oxidative stress**.** Thus, the main alterations may prevail in mature spermatozoa due to oxidative direct damage. Reactive oxygen species will attack the loosely packed chromatin and induce the following effects: DNA fragmentation, abnormal DNA integrity, and formation of oxidative adducts. The latter will alter the epigenetic machinery engendering a massive DNA hypomethylation. More so, ROS increase the oxidized products of 5-mC (5-hmC, 5-fC) either by direct attack or through a complex pathway to activate TET proteins.

The strength and novelty of the current work are represented by the analysis of multiple sperm parameters from infertile and fertile patients. These analyses allowed us to identify several epigenetic signatures in sperm from infertile men. Interestingly, sperm global DNA hydroxymethylation and 5-formylcytosine levels between fertile and infertile patients were measured for the first time. Notably, all subjects studied are of the same age group. In fact, the association between age and both DNA methylation and hydroxymethylation profiles was previously reported [277]. Thus, here the methylation analysis is not affected by the paternal age factor.

Conclusion

In summary, this extensive analysis has expanded our knowledge of the spermatozoal transcriptome in infertile men, and emphasized the relevance of RNA profiles as a modality for clinical classification. Furthermore, the proposed aberrant epigenetic landscape could help to elucidate the molecular mechanisms involved in sperm pathogenesis associated with male infertility. Combined, these findings are an important step forward in our understanding of sperm abnormalities as discrete subsets for each abnormality and the mechanisms driving sperm pathogenesis. In the next chapter, we will discuss our findings and elicit the complete picture of sperm pathogenesis.

CHAPTER V

GENERAL DISCUSSION

In this chapter we will summarize briefly the results obtained from the transcriptomic and epigenetic analysis of spermatozoa. First, we will outline the rationale for appreciating the use of NGS for transcriptome analysis. Next, we will present our empirical evidence about the aberrant transcriptome in sperm for each sperm pathology. Furthermore, the association between oxidative stress-related infertility and aberrant sperm epigenome will be discussed.

CONTRIBUTIONS

5.1. Differential analysis of spermatozoal transcripts from infertile men with different sperm pathologies using RNA-seq

When we initiated our studies. Our understanding of the sperm transcriptome was limited. Yet, several groups had described the sperm transcriptome from infertile men solely using array-based approaches [102, 112-113, 130]. Until recently, new advances in transcriptomics namely the discovery of high throughput sequencing technologies permited an in-depth analysis of the transcriptome. This was ground breaking as it suggested a complete characterization of human spermatogenesis was about to be discovered.

In this manner, few attempts characterized the coding and the small non-coding healthy pool of transcripts from fertile males [209, 246]. To date, there is a paucity of studies from RNA-seq data of normozoospermic “idiopathic” infertile men [255, 257]. As this recent evidence made use of a sophisticated computational methodology, a comprehensive catalog detailing the differentially expressed genes between fertile and infertile men presenting different sperm pathologies including oligozoospermia, asthenozoospermia and teratozoospermia and what their specific molecular characteristics are was needed, as it could

serve as a guiding map for the scientific community that had a growing interest in sperm biology. The emergence of RNA-Seq and the development of ab initio transcriptome assemblers presented a unique opportunity to generate such a catalog at single nucleotide resolution in an effective manner and unprecedented scale. Our goal was to generate, for the first time, a detailed mRNA catalog describing a large set of properties such as expression patterns, cellular components, molecular functions, biological processes and pathways related to each distinct phenotype.

The whole transcriptome sequence of human spermatozoa provided novel insights into how sperm transcriptomic alterations can shape the sperm phenotype. Strikingly, we identified a very large number of genes differentially expressed between the infertile arm and fertile. For example, we found from oligozoospermic samples a total of 10087 differentially expressed genes, and of these, 2823 genes were significantly down-regulated while 7264 genes were significantly up-regulated. However, the spermatozoa from teratozoospermic samples contained in total 7345 differentially expressed genes, and of these, 4546 genes were significantly down-regulated while 2799 sequences were significantly up-regulated. In addition, we identified 14606 differentially expressed genes, and of these, 10559 genes were significantly down-regulated while 4047 sequences were significantly up-regulated in asthenozoospermia compared to fertile men. These numbers are strikingly higher than the findings of earlier reports using a microarray-based approach.

Protein-coding transcripts constituted a high proportion of the complete set of DE (differentially expressed) transcripts. Interestingly, few DE long non-coding RNAs members were also reported in each pathology cluster. One of the main advantages of our sperm compendium is that it allowed us to conclusively characterize the distinct genes and pathways altered in each pathology. We used a functional enrichment analysis of the datasets of DE transcripts to predict the molecular players labeling each specific phenotypic cluster.

In oligozoospermia, transcripts associated with spermatogenesis, spermatid differentiation, flagellated sperm motility, fusion of sperm to egg plasma membrane involved in single fertilization, and fertilization were found significantly down-regulated. This down-regulation of the transcripts that are involved in spermatogenesis and sperm motility was congruent with previous findings from microarray studies [102-103]. Orthologs sorting revealed that particularly; the meiotic cell cycle, synaptonemal complex assembly, piRNA metabolic process, sister chromatid cohesion are down-regulated biological processes that may alter spermatogenesis. In humans, failure of formation of the synaptonemal complex lead to azoospermia [359]. The piRNA metabolic process was also found altered, due to its

potential role in spermatogenic control. In this perspective, silencing of PIWI – proteins involved in piRNA processing - in humans contribute to an unsuccessful germ cell development and thus male infertility [29]. Our results also revealed that the collagen formation and the extracellular matrix (ECM) organization pathways are altered in oligozoospermia. Recently, it was highlighted the crucial role of ECM in supporting Sertoli and germ cell function in the seminiferous epithelium, including the blood-testis barrier (BTB) dynamics [353].

Further topological analysis between distinct oligozoospermia orthologs identified down-regulated candidate genes related to cellular response to endogenous stimulus. We can speculate that the attenuated responsiveness in male germ cells may be associated with hypospermatogenesis and oligozoospermia.

In parallel, we also identified spermatozoal transcripts significantly over represented that are related to ubiquitin-dependent protein catabolic process, and cellular response to DNA damage stimulus. Similarly, the apoptotic pathway was also reported to be deregulated in oligozoospermia [102-103]. It appears that the suppression of spermatogenesis was due to acceleration of germ cell apoptosis [362]. Surprisingly, we detected 290 detected lincRNA (long intergenic noncoding RNAs) which included 26 genes belonging to the long non coding RNAs family that were deregulated. However, due to the lack of a database for annotation of lncRNAs, we only pointed out only one lncRNA, TP53TG1 which was found associated with cellular response to DNA damage stimulus. We may speculate that lncRNAs might play a role in modulating the cellular response to DNA damage and thus this response is misregulated in oligozoospermia.

As for asthenozoospermia, we noticed that sperm from asthenozoospermic patients share several hallmarks, namely disruption of the mitochondrial function as well as structural flagellar alterations. In fact, low motility spermatozoa carry high numbers of under-represented transcripts associated with cytoskeleton organization of the sperm flagella. These findings are concordant with previous genomics and transcriptomics data [112-113, 115, 476]. Add to that, the presence of spermatozoal transcripts related to the ribosome and mitochondria which are sturdily down-regulated, indicate that many of these transcripts are involved in cellular energy supply, wherein mitochondrial bioenergetics and organization take place and are related to sperm motility. It is known that the so-called mitochondrial capsule generated around the axoneme in the sperm mid-piece drive sperm motility [380]. Along these lines, the main bioprocesses affected are the translation which promotes the expression of mitochondria components,, and the oxidative phosphorylation. The ribosomal

proteins transcripts which are the keystone components of the ribosomal translational machinery, are the major transcripts down-regulated in asthenozoospermia, similarly to the finding of a microarray-based study by Bansal et al. [112].

Additionally, we detected several players in the calcium signaling pathway (CALM, CAMK, PDE1…) down-regulated in spermatozoa of asthenozoospermic patients. Our results are in agreement with previous reports and suggest a strong relationship between calcium homeostasis and sperm motility [477-478]. Interestingly, we also identified numerous downregulated transcripts related to histones modifications. Some of these genes are involved in shaping the epigenetic histone code such as HDAC1, TET1, TET2, and TET3. Likewise, HDAC1 and TET1-3 presented an altered expression in low motility sperm samples [153, 257].

We also highlighted a small subset of 19 long ncRNAs that were found to be notably down-regulated in asthenozoospermic patients. Topological analysis revealed that the interactome candidate genes are related to gene expression. Although the role of long non-coding RNA (lncRNA) in sperm from asthenozoospermic patients has scarcely been reported, we can speculate from our data that some lncRNAs can regulate transcription through not solely the interaction with RNA binding proteins. Therefore, lncRNAs disturbance is a different layer impacting genes transcription deregulation in asthenozoospermia.

With regards to teratozoospermia, several pathways were found deregulated namely; ubiquitin mediated proteolysis, deubiquitination, and protein ubiquitination. The ubiquitin-proteasome pathway have been previously characterized as a causal pathway in teratozoospermia pathogenesis [130-131]. Several candidate genes regulating the proteasome were previously pointed on the basis of very few mapped genes [131, 419]. Compared to previous reports, we distinguished more members of the ubiquitin mediated proteolysis pathway to be down-regulated in teratozoospermia. This pathway was found to be involved in a subset of events related to spermatogenesis such as mitochondrial shaping, and acrosomal formation [421].

We also found that the cell cycle was down-regulated. Likewise, Huang and co-workers yielded the same conclusion [131]. This linkage between cell-cycle and spermatogenesis can be explained by the fact that spermatogenesis is a cyclic process, and therefore the expression of cell-cycle related proteins was found linked to spermatogenesis timelined events [422]. A noteworthy finding was the down-regulation of many putative epigenetic modulators of gene expression, namely transcripts related to post-transcriptional regulation of gene expression, chromatin organization, and histone modification. This is not surprising since several

epigenetic players are important regulators of spermatogenesis [424-425]. In the same context, the Nonsense Mediated Decay (NMD) pathway was also shown to be down-regulated. By definition, Nonsense-mediated mRNA decay (NMD) is a surveillance pathway that reduces errors in gene expression by eliminating mRNA transcripts that contain premature stop codons [426]. Until lately, it was evoked that NMD plays an essential role in spermatogenesis, spermiogenesis, and fertility [389]. Moreover, we reported many up-regulated genes implicated in piRNA metabolic process. The PIWI/piRNA machinery has been implicated in suppressing transposable elements and protecting the integrity of the genome in germ cells [431]. It was proposed that the abnormal accumulation of piRNAs would interfere with sperm maturation and the proper progression of male germ cell development [432].

We have identified 52 DE long non-coding RNAs. The topological analyses using the interactome of lncRNAs suggest that lncRNAs can control various processes, notably cell cycle progression, ubiquitination and apoptosis. We can speculate that the deregulations of many proteins may arise from a perturbed interactome. In this manner, these deregulations in lncRNAs is not unforeseen since the latter processes were deregulated in teratozoospermia as well. This parallel identification of specific RNA fingerprints for each sperm abnormality, revealed very subtle distinctions between the different phenotypes. Accordingly, the synoptic events for each sperm pathology can be drawn as follow:

In oligozoospermia, the hypospermatogenesis may be related to different levels: (i) the disrupted architecture of the germinal epithelium; (ii) the altered main bioprocesses during spermatogenesis namely meiosis; (iii) acceleration of germ cells apoptosis; (iv) deregulations of the superlayers of spermatogenesis control i.e., piRNAs, lncRNA.

In asthenozoospermia, low sperm motility may ascribe from: (i) a disorganized sperm flagella structure; (ii) functional deficiencies in sperm flagella; (iii) mitochondrial bioenergetic dysfunction; (iv) deregulation of the epigenetic controller on transcription i.e., HDAC, TET, lncRNAs.

In teratozoospermia, poor sperm morphology may be attributed to: (i) the altered temporal events during spermatogenesis; (ii) a disrupted spermiogenesis through the deregulation of protein ubiquitination; (iii) inefficient mRNA surveillance pathway (NMD) which results in the synthesis of abnormal proteins; (iv) an altered epigenetic regulatory control over the main bioprocesses.

In general, the above outlined pictures of each sperm abnormality are in concert with the mainstream findings of previous studies using microarray-based analysis. Surprisingly, we noticed in oligozoospermia DE transcripts predicted to be related to fertilization. Therefore, we can highlight that spermatozoa carry not only remnant mRNAs along with the corresponding proteins, which are retrospectively indicative of spermatogenic events, but also transcripts that might play vital roles in fertilization and embryogenesis. Taken together, we noticed that the epigenetic regulatory feature embraces the transcriptome. Therefore, we reported that the 3 sperm pathologies; oligozoospermia, asthenozoospermia and teratozoospermia; in addition to their altered transcriptome, exhibit an altered expression of many epigenetic players i.e. lncRNAs, piRNAs, HDACs…

Recently, lncRNAs were considered to function as epigenetic modulators and hence impacting gene expression [449]. In the same perspective, several lines of evidence strongly suggest that piRNAs also play an important role in the epigenetic control of gene expression [433]. Furthermore, we advanced our analysis by combining the three datasets and found that that the three sperm abnormalities shared altered expression levels of a set of 13 sperm-specific genes, which are the following: *CABS1, IFT172, RBM5, REC8, SGO2, SMC1B, SOX30, SPATA31D1, SUN5, TDRD7, TLR3, TNP1, TSSK6.*

The majority of these spermatozoal transcripts were related to spermatogenesis, and some might play a role in sister chromatid cohesion and fertilization namely based on predictions from mouse knockout studies [41, 443-444]. Alternatively, we also have proposed to identify a minimal set of genes that best characterize the three phenotypes. Therefore, we further scrutinized the data, then attempted to analyze only mouse orthologs related to “male infertility”, a common GO (gene ontology) term-phenotype between the three phenotypes. We have spotted 10 genes as the common signature genes: BSG, PRSS37, MEIOC, RBM5, TSSK6, MORC1, GRID2, and TEX15, DNAH1, and VPS13A. Similar to above, this clustering of expression data from all three datasets grouped together based on a narrowed category, also highlighted a clear disruption of spermatogenesis in infertile men.

Very few previous studies have documented a relationship between some of these genes and infertility. Indeed, the knockout of TSSK6 in mice caused a sperm reduction and abnormal motility and morphology [447]. In addition, MORC1 was identified to be differentially expressed in infertile “idiopathic” men using a next generation sequencing approach [257]. Morc1 is involved in chromosomal pairing during the zygotene stage of meiosis and is a regulator of the epigenetic landscape of male germ cells during the period of global *de novo* methylation as well [448].

Another key point introduced in our study was the identification of many long non-coding RNAs differentially expressed between infertile men and fertile. This was important since weak evidence of the engagement of lncRNAs in male infertility pathogenesis is reported in the literature. Since at the time of our study a publically available resource for functional investigation of human lncRNAs was not available, we could only provide a minimal functional characterization of lncRNA based on publically available catalogs of non-coding RNA–RNA and ncRNA–protein interactions (the RAIN database: RNA–protein Association and Interaction Networks database), and on topological protein–protein interaction networks. These provided few examples by which one can explore this new class of RNAs which can serve as a mean to understand lncRNAs-disease association. These would greatly benefit the scientific community and serve as an early guide for future mechanistic studies targeting specific genes.

5.2. Assessment of global epigenetic changes and oxidative damage in sperm from infertile men

Our first study provided a map that enabled us to understand the global properties of the sperm transcriptome in 3 categories of sperm pathologies (oligozoospermia, asthenozoospermia, teratozoospermia) which were highlighted to be deregulated using expression analyses. Still, a key step toward understanding these altered gene expressions is to track the epigenetic modifications underlying these alterations. More recently, intriguing evidence has emerged that genetic and epigenetic mechanisms are not separate events; thereby they can be considered the two sides of the same coin; thus they intertwine and intersect to dysregulate spermatogenic processes [479-481].

With this in mind, we have systematically applied several *in-situ* microscopy approaches to characterize the sperm chromatin; colorimetric ELISA assays to depict global DNA modifications (levels of 5-methylcytosine, 5-hydroxymethylcytosine and 5-formylcytosine) as well as assessed the sperm ROS levels in 33 infertile patients with abnormal semen analysis compared to 23 fertile men. We show that sperm from infertile men presents high histones retention (high % of aniline blue positive sperms), probably owing to a defective spermiogenesis, which lead to a poor sperm chromatin packaging. Our results are in keeping with previous reports showing abnormal sperm chromatin structure in spermatozoa of infertile men as compared with controls, with persistence of histones [344, 453]. This

renders the chromatin more susceptible to ROS assaults. Thus, the main alterations may prevail in mature spermatozoa due to oxidative direct damage. In fact, we also found that sperm from infertile men is characterized by ROS accumulation (high NBT% levels). Therefore, several studies confirmed the association between sperm oxidative stress and male infertility [456-457].

Indeed, reactive oxygen species will attack the loosely packed chromatin and induce the following effects: DNA fragmentation, abnormal DNA integrity, and formation of oxidative adducts. It is presumed that oxidative adducts can alter the epigenetic machinery engendering a massive DNA hypomethylation. This finding is congruent with previous reports showing differential DNA methylation profiles in spermatozoa of infertile men as compared to normozoospermic controls [456, 467, 482].

Add to that, it was proposed that ROS may increase the oxidized products of 5-mC (5-hmC, 5-fC) either by direct attack or through a complex pathway to activate TET proteins [475]. Here, we reported a significant rise in the levels of the oxidized derivatives of 5methylcytosine: 5-hydroxymethylcytosine (5-hmC) and 5-formylcytosine (5-fC) in sperm derived from infertile patients. Along these lines, while a negative association was established between the percentage of highly hydroxymethylated spermatozoa (measured by immunocytochemistry) and sperm normal morphology, a positive correlation to sperm DNA fragmentation was also pointed out [469].

While we were only able to survey a limited set of epigenetic parameters, our study demonstrates the great benefit of applying colorimetric ELISA approache at an early step when studying the global sperm epigenome. To date, there is a paucity of studies regarding the DNA demethylation intermediates patterns in infertile sperm. Interestingly, this is for the first time that the levels of DNA demethylation intermediates were examined, and these findings were interesting since they provided a coherent biological picture about the sperm biology. Altogether, our findings suggest that the global epigenetic landscape in sperm from infertile men is aberrant.

Conclusion

Overall, using systematic RNA-Seq and global epigenetic analysis we uncovered key aspects of different sperm abnormalities including their transcripts levels as well as their diverse global epigenetic patterns. Finally, we will end with concluding remarks including pointers to the significant strengths of this research that have been covered in this thesis, as well as some suggestion of future directions for expanding this field of research.

CONCLUSION AND FUTURE DIRECTIONS

The transcriptome of the infertile sperm using RNA-seq technology showed at unprecedented detail how fluctuations in gene expression might alter sperm density, architecture and function. In this regard, our high-throughput sequencing of sperm transcriptomes from infertile men with abnormal semen analysis revealed the complexity of transcriptomes as indicated by detection of plentiful protein coding genes and the identification of novel spermatogenesis-related active genes. Much remains unknown regarding the contributions of many cellular and biochemical events to any of the phenotypes.

This large scale study using NGS technology in conjunction to advanced computational analysis have proven extremely useful in capturing individual impacted signal transduction pathways and putative mechanisms involved in sperm pathologies. This dynamic and rapidly evolving area of investigation has and will undoubtedly continue to contribute greatly to the uncovering of the molecular events leading to mature and fertile spermatozoa. Furthermore, the identification of long non-coding RNAs combined with their expression patterns in infertile sperm provided novel insights into infertility pathogenesis. Long non-coding RNAs constitute a novel regulatory target in spermatogenesis, therefore they entail another facet that is receiving more attention which is the mediation of epigenetic control of transcription through their association with various epigenetic complexes [483]. In the same perspective, we also evidenced the contribution of sperm oxidative stress in shaping the sperm global epigenome in infertile men. Our study has several folds of significance that are the following:

i. Uncovering transcriptome complex in infertile sperm

Our study, for the first time, revealed the real complexity of sperm transcriptome in infertile men utilizing deep RNA-sequencing. During this sequencing effort, we have identified protein-coding transcripts in addition to some uncharacterized lncRNAs expressed in sperm.

ii. A fine-tuned functional transcriptomics in infertile sperm

Although the concepts of lncRNA have been established a decade ago, this transcriptome component has been poorly studied in sperm. One reason for this is that the tools to study are not well established. With the development of high throughput technology, especially deep RNA-sequencing, investigators were able to zoom into thisunderappreciated

transcriptome component. Based on this deep sequencing effort, it is suggested that the expression of these uncharacterized transcripts is deregulated in different sperm pathological phenotypes. Our deep RNA-sequencing also allowed functional prediction of lncRNAs in sperm from infertile men and the identification of controlled regulatory networks in sperm by subjecting lncRNA/mRNA transcripts to topological predictions.

iii. Aberrant global epigenetic landscape in infertile sperm

By focusing analysis on the global epigenetic layers we observed global DNA hypomethylation and for the first time increased levels of DNA 5-hydroxymethylcytosine and 5-formylcytosine. We further provided the first evidence that sperm oxidative stress mediated these global epigenetic alterations which have significant functional impacts on spermatogenesis. We characterized sperm chromatin abnormalities and epigenetic parameters in infertile sperm which are of translational value.

iv. Refinement and optimization of sperm RNA extraction protocol

We tested and modified various conditions in a standard protocol in an effort to isolate high-quality RNA from sperm cells. An overall optimum for the extraction conditions was found based on an experimental design technique.

v. Innovation

At the conceptual level, our study provides unpreceded delineation of the sperm transcriptome and epigenome in infertile men presenting different sperm pathologies by uncovering the multifaceted layers of transcriptomic and epigenomic regulation that may alter during a pathological spermatogenesis, providing a novel platform for discovery. Particularly, the findings from the present study provide two different insights. The first is that spermatozoal carry relics which are indicative of a pathological spermatogenesis. Second, spermatozoa contain transcripts that might play key roles in fertilization and embryogenesis. At the technical level, instead of narrowly focusing on a single molecule/pathway as in conventional studies, we leveraged the power of advanced genomics, bioinformatics, and systems biology to uncover novel or putative pathways associated with male infertility in an unbiased manner.

vi. Limitations

Our studies of spermatozoa transcriptome were designed to perform comprehensive transcriptome characterization of protein-coding mRNAs. A detailed expression study at of

small non-coding RNAs and long non-coding RNAs populations will likely be required to utterly unfold the heterogeneity of sperm cells and to understand transcriptome regulation in sperm. Our identifications of mRNA and lncRNAs in infertile sperm reflect potentially functional regulatory mechanisms that may link to disease. Further mechanistic studies are required to understand the underlying mechanisms of lncRNAs and to test the potential applications.

Future directions

Our RNAseq repertoire revealed interesting expression patterns for several transcripts namely lncRNAs and epigenetic players. These are promising candidates for follow-up experiments. One possible approach to learn more about these lncRNAs is to perform a chromatin localization assay to determine the genomic binding site of a noncoding RNA.

Another interesting candidate for future experiments is the characterization of of another facet of the sperm epigenome such as a colorimetric quantification of histones modifications using a commercially available kit, i.e. histone H3 modifications, which is underway.

Another intriguing avenue is to approach specific pathways. Thus, pathways analysis can draw attention to groups of genes involved in more specific cellular processes. One can also scrutinize a specific KEGG pathway generated from these datasets, for example the methionine pathway which donates methyl groups. In the light of this conception, we can study the effect of folate supplementation or other dietary sources on DNA methylation status. In this manner, we can incubate sperm in a medium supplemented with high folate then characterize the DNA methylation status. Another example might be to zoom into the nicotine metabolism pathway and glutathione metabolism, and link sperm phenotypes to smoking habits.

A possible approach is to choose a predicted unstudied mouse ortholog and perform knockout experiments in mice which will provide an excellent model for linking specific genes to a pathology such as the disruption of transcripts related to the piRNA metabolic process (e.g. *TDRD1*, *MAEL*, and *DDX4*) and oligozoospermia which to my knowledge were not genetically studied in sperm.

One future avenue that I find particularly interesting is understanding the function of a gene at the protein level. Thus, an important component of functional

annotation is characterizing protein interactions, along these lines analysis of gene expression data can be the starting point for further functional characterization, or high-throughput proteomic analysis. In this manner, sperm proteome has been extensively studied; however the interactome is poorly studied. Relying on the predicted protein-protein interaction maps generated from our RNAseq data some potentially interacting proteins can be specifically studied using a protein-protein interaction method or system (i.e. yeast two-hybrid, pull down), then we can integrate protein-protein interaction data with mRNA expression levels, for example the study of 3 transcripts related to the piRNA metabolic process (e.g. TDRD1, MAEL, and DDX4) in oligozoospermia are predicted to interact together.

Finally, with the rapidly decreasing cost of NGS, deep sequencing of sperm RNA has the potential to afford either fundamental or clinical research areas. Thus, different populations of interest can be candidate for large scale RNAseq analysis such as obese patients. We can also target very specific infertile individuals such as patients with familial globozoospermia, or patients with Kartagener's syndrome.

In the near future, combining transcriptomics and epigenetics with new emerging tools will help to decipher the complete function of a new and challenging class of genes. I am looking forward to see all the new discoveries on ncRNAs in the coming years. The stage is now set, ncRNAs are marching toward mechanism.

Finally, the scale of data being generated by such high-throughput experiments allow the use of hypothesis-driven approaches in future research, where the triage and selection of candidate genes or pathways will be based on some prior knowledge or assumption.

REFERENCES

1. Simon, L. and D.T. Carrell, *Current and Future Perspectives on Sperm RNAs*, in *Emerging Topics in Reproduction*. 2018, Springer. p. 29-46.
2. Ohinata, Y., et al., *Blimp1 is a critical determinant of the germ cell lineage in mice.* Nature, 2005. **436**(7048): p. 207.
3. Jan, S.Z., et al., *Molecular control of rodent spermatogenesis.* Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease, 2012. **1822**(12): p. 1838-1850.
4. Lin, I.-Y., et al., *Suppression of the SOX2 neural effector gene by PRDM1 promotes human germ cell fate in embryonic stem cells.* Stem cell reports, 2014. **2**(2): p. 189-204.
5. De Jong, J., et al., *Differential expression of SOX17 and SOX2 in germ cells and stem cells has biological and clinical implications.* The Journal of pathology, 2008. **215**(1): p. 21-30.
6. Mamsen, L.S., et al., *The migration and loss of human primordial germ stem cells from the hind gut epithelium towards the gonadal ridge.* International Journal of Developmental Biology, 2013. **56**(10-11-12): p. 771-778.
7. Gilbert, D., et al., *Clinical and biological significance of CXCL12 and CXCR4 expression in adult testes and germ cell tumours of adults and adolescents.* The Journal of pathology, 2009. **217**(1): p. 94-102.
8. Perrett, R.M., et al., *The early human germ cell lineage does not express SOX2 during in vivo development or upon in vitro culture.* Biology of reproduction, 2008. **78**(5): p. 852-858.
9. Julaton, V.T.A. and R.A. Reijo Pera, *NANOS3 function in human germ cell development.* Human molecular genetics, 2011. **20**(11): p. 2238-2250.
10. Castrillon, D.H., et al., *The human VASA gene is specifically expressed in the germ cell lineage.* Proceedings of the National Academy of Sciences, 2000. **97**(17): p. 9585-9590.
11. Stukenborg, J.-B., et al., *Male germ cell development in humans.* Hormone research in paediatrics, 2014. **81**(1): p. 2-12.
12. Raymond, C.S., et al., *Expression of Dmrt1 in the genital ridge of mouse and chicken embryos suggests a role in vertebrate sexual development.* Developmental biology, 1999. **215**(2): p. 208-220.
13. Dolci, S., F. Campolo, and M. De Felici. *Gonadal development and germ cell tumors in mouse and humans*. in *Seminars in cell & developmental biology*. 2015: Elsevier.
14. Biason-Lauber, A., *The Battle of the sexes: human sex development and its disorders*, in *Molecular Mechanisms of Cell Differentiation in Gonad Development*. 2016, Springer. p. 337-382.
15. Tilmann, C. and B. Capel, *Cellular and molecular pathways regulating mammalian sex determination.* Recent progress in hormone research, 2002. **57**(1): p. 1-18.
16. Garcia, T.X. and M.-C. Hofmann, *NOTCH signaling in Sertoli cells regulates gonocyte fate.* Cell Cycle, 2013. **12**(16): p. 2538-2545.
17. Garcia, T.X., et al., *Constitutive activation of NOTCH1 signaling in Sertoli cells causes gonocyte exit from quiescence.* Developmental biology, 2013. **377**(1): p. 188-201.

18. Gegenschatz-Schmid, K., et al., *DMRTC2, PAX7, BRACHYURY/T and TERT Are Implicated in Male Germ Cell Development Following Curative Hormone Treatment for Cryptorchidism-Induced Infertility.* Genes, 2017. **8**(10): p. 267.
19. Gallagher, S.J., et al., *Distinct requirements for Sin3a in perinatal male gonocytes and differentiating spermatogonia.* Developmental biology, 2013. **373**(1): p. 83-94.
20. Young, J.C., S. Wakitani, and K.L. Loveland. *TGF-β superfamily signaling in testis formation and early male germline development*. in *Seminars in cell & developmental biology*. 2015: Elsevier.
21. Culty, M., *Gonocytes, the forgotten cells of the germ cell lineage.* Birth Defects Research Part C: Embryo Today: Reviews, 2009. **87**(1): p. 1-26.
22. Nakagawa, T., et al., *Functional hierarchy and reversibility within the murine spermatogenic stem cell compartment.* Science, 2010. **328**(5974): p. 62-67.
23. Wu, X., et al., *Prepubertal human spermatogonia and mouse gonocytes share conserved gene expression of germline stem cell regulatory molecules.* Proceedings of the National Academy of Sciences, 2009. **106**(51): p. 21672-21677.
24. Chen, S.-R. and Y.-X. Liu, *Regulation of spermatogonial stem cell self-renewal and spermatocyte meiosis by Sertoli cell signaling.* Reproduction, 2015. **149**(4): p. R159-R167.
25. Kusz, K., et al., *The highly conserved NANOS2 protein: testis-specific expression and significance for the human male reproduction.* Molecular human reproduction, 2009. **15**(3): p. 165-171.
26. Sada, A., et al., *NANOS2 Acts Downstream of Glial Cell Line-Derived Neurotrophic Factor Signaling to Suppress Differentiation of Spermatogonial Stem Cells.* Stem Cells, 2012. **30**(2): p. 280-291.
27. McIver, S., et al., *The chemokine CXCL12 and its receptor CXCR4 are implicated in human seminoma metastasis.* Andrology, 2013. **1**(3): p. 517-529.
28. Wang, H., et al., *BMP6 Regulates Proliferation and Apoptosis of Human Sertoli Cells Via Smad2/3 and Cyclin D1 Pathway and DACH1 and TFAP2A Activation.* Scientific reports, 2017. **7**: p. 45298.
29. Heyn, H., et al., *Epigenetic disruption of the PIWI pathway in human spermatogenic disorders.* PloS one, 2012. **7**(10): p. e47892.
30. Waheeb, R. and M.C. Hofmann, *Human spermatogonial stem cells: a possible origin for spermatocytic seminoma.* International journal of andrology, 2011. **34**(4pt2).
31. Sá, R., et al., *Expression of stem cell markers: OCT4, KIT, ITGA6, and ITGB1 in the male germinal epithelium.* Systems biology in reproductive medicine, 2013. **59**(5): p. 233-243.
32. Dong, Y., et al., *Copy number variations in spermatogenic failure patients with chromosomal abnormalities and unexplained azoospermia.* Genet Mol Res, 2015. **14**(4): p. 16041-9.
33. Zhang, X., et al., *Immunohistochemical study of expression of Sohlh1 and Sohlh2 in normal adult human tissues.* PloS one, 2015. **10**(9): p. e0137431.
34. Raverdeau, M., et al., *Retinoic acid induces Sertoli cell paracrine signals for spermatogonia differentiation but cell autonomously drives spermatocyte meiosis.* Proceedings of the National Academy of Sciences, 2012. **109**(41): p. 16582-16587.
35. Busada, J.T., et al., *Retinoic acid regulates Kit translation during spermatogonial differentiation in the mouse.* Developmental biology, 2015. **397**(1): p. 140-149.
36. Bonache, S., et al., *Altered gene expression signature of early stages of the germ line supports the pre-meiotic origin of human spermatogenic failure.* Andrology, 2014. **2**(4): p. 596-606.
37. Fu, X.-F., et al., *DAZ family proteins, key players for germ cell development.* International journal of biological sciences, 2015. **11**(10): p. 1226.

38. Costa, Y., et al., *Two novel proteins recruited by synaptonemal complex protein 1 (SYCP1) are at the centre of meiosis.* Journal of cell science, 2005. **118**(12): p. 2755-2762.
39. Turner, J.M., et al., *BRCA1, histone H2AX phosphorylation, and male meiotic sex chromosome inactivation*. Current Biology, 2004. **14**(23): p. 2135-2142.
40. Jan, S.Z., et al., *Distinct prophase arrest mechanisms in human male meiosis.* Development, 2018: p. dev. 160614.
41. Yan, W., *Male infertility caused by spermiogenic defects: lessons from gene knockouts.* Molecular and cellular endocrinology, 2009. **306**(1-2): p. 24-32.
42. Christensen, G.L., et al., *Identification of polymorphisms in the Hrb, GOPC, and Csnk2a2 genes in two men with globozoospermia.* Journal of andrology, 2006. **27**(1): p. 11-15.
43. Dam, A.H., et al., *Homozygous mutation in SPATA16 is associated with male infertility in human globozoospermia.* The American Journal of Human Genetics, 2007. **81**(4): p. 813-820.
44. Yatsenko, A.N., et al., *Association of mutations in the zona pellucida binding protein 1 (ZPBP1) gene with abnormal sperm head morphology in infertile men.* Molecular human reproduction, 2011. **18**(1): p. 14-21.
45. Maldonado-Báez, L., et al., *Microtubule-dependent endosomal sorting of clathrin-independent cargo by Hook1*. J Cell Biol, 2013. **201**(2): p. 233-247.
46. Moreno, R.D., J. Palomino, and G. Schatten, *Assembly of spermatid acrosome depends on microtubule organization during mammalian spermiogenesis.* Developmental biology, 2006. **293**(1): p. 218-227.
47. Zhou, J., et al., *RIM-BP3 is a manchette-associated protein essential for spermiogenesis.* Development, 2009. **136**(3): p. 373-382.
48. Sun, X. and W.-X. Yang, *Mitochondria: transportation, distribution and function during spermiogenesis.* Advances in Bioscience and Biotechnology, 2010. **1**(02): p. 97.
49. Liu, Y., et al., *LRGUK-1 is required for basal body and manchette function during spermatogenesis and male fertility.* PLoS genetics, 2015. **11**(3): p. e1005090.
50. Amos, L.A., *The tektin family of microtubule-stabilizing proteins.* Genome biology, 2008. **9**(7): p. 229.
51. Wolkowicz, M.J., et al., *Tektin B1 demonstrates flagellar localization in human sperm.* Biology of reproduction, 2002. **66**(1): p. 241-250.
52. Roy, A., et al., *Tektin3 encodes an evolutionarily conserved putative testicular microtubules‐related protein expressed preferentially in male germ cells.* Molecular reproduction and development, 2004. **67**(3): p. 295-302.
53. Xu, M., et al., *Cloning and characterization of a novel human TEKTIN1 gene.* The international journal of biochemistry & cell biology, 2001. **33**(12): p. 1172-1182.
54. Zhang, Z., et al., *A heterozygous mutation disrupting the SPAG16 gene results in biochemical instability of central apparatus components of the human sperm axoneme.* Biology of reproduction, 2007. **77**(5): p. 864-871.
55. Schwarz, T., et al., *Ccdc181 is a microtubule-binding protein that interacts with Hook1 in haploid male germ cells and localizes to the sperm tail and motile cilia.* European journal of cell biology, 2017. **96**(3): p. 276-288.
56. Liu, B., et al., *Expression and localization of voltage-dependent anion channels (VDAC) in human spermatozoa.* Biochemical and biophysical research communications, 2009. **378**(3): p. 366-370.
57. De Pinto, V., et al., *Characterization of human VDAC isoforms: a peculiar function for VDAC3?* Biochimica et Biophysica Acta (BBA)-Bioenergetics, 2010. **1797**(6-7): p. 1268-1275.

58. Johnson, G., et al., *Cleavage of rRNA ensures translational cessation in sperm at fertilization.* Molecular human reproduction, 2011. **17**(12): p. 721-726.
59. Vogl, A.W., S.Y. J'Nelle, and M. Du, *New insights into roles of tubulobulbar complexes in sperm release and turnover of blood-testis barrier*, in *International review of cell and molecular biology*. 2013, Elsevier. p. 319-355.
60. TANAKA, H., Y. MIYAGAWA, and A. TSUJIMURA, *Genetic variation in the testis-specific GSG3/CAPZA3 gene encoding for the actin regulatory protein in infertile males.* Nagasaki International University Review, 2014. **14**: p. 269-274.
61. Jodar, M., E. Sendler, and S.A. Krawetz, *The protein and transcript profiles of human semen.* Cell and tissue research, 2016. **363**(1): p. 85-96.
62. Zheng, H., et al., *Lack of Spem1 causes aberrant cytoplasm removal, sperm deformation, and male infertility.* Proceedings of the National Academy of Sciences, 2007. **104**(16): p. 6852-6857.
63. Bao, J., et al., *UBQLN1 interacts with SPEM1 and participates in spermiogenesis.* Molecular and cellular endocrinology, 2010. **327**(1): p. 89-97.
64. Steger, K., et al., *Expression of mRNA and protein of nucleoproteins during human spermiogenesis.* Molecular human reproduction, 1998. **4**(10): p. 939-945.
65. Aoki, V.W. and D.T. Carrell, *Human protamines and the developing spermatid: their structure, function, expression and relationship with male infertility.* Asian journal of andrology, 2003. **5**(4): p. 315-324.
66. Tanaka, H., et al., *Genetic Variation in the Testis-Specific Poly (A) Polymerase Beta (PAPOLB) Gene Among Japanese Males.* The Open Reproductive Science Journal, 2015. **7**(1).
67. Gou, L.-T., et al., *Pachytene piRNAs instruct massive mRNA elimination during late spermiogenesis.* Cell research, 2014. **24**(6): p. 680.
68. Welch, J.E., et al., *Human Glyceraldehyde 3-Phosphate Dehydrogenase-2 Gene Is Expressed Specifically in Spermatogenic Cells.* Journal of andrology, 2000. **21**(2): p. 328-338.
69. Singh, A.P. and S. Rajender, *CatSper channel, sperm function and male fertility.* Reproductive biomedicine online, 2015. **30**(1): p. 28-38.
70. Xiao, X., et al., *c-Yes regulates cell adhesion at the blood–testis barrier and the apical ectoplasmic specialization in the seminiferous epithelium of rat testes.* The international journal of biochemistry & cell biology, 2011. **43**(4): p. 651-665.
71. Leclerc, P. and S. Goupil, *Regulation of the human sperm tyrosine kinase c-yes. Activation by cyclic adenosine 3', 5'-monophosphate and inhibition by Ca2+*. Biology of reproduction, 2002. **67**(1): p. 301-307.
72. Paronetto, M.P. and C. Sette, *Role of RNA-binding proteins in mammalian spermatogenesis.* International journal of andrology, 2010. **33**(1): p. 2-12.
73. Glisovic, T., et al., *RNA-binding proteins and post-transcriptional gene regulation.* FEBS letters, 2008. **582**(14): p. 1977-1986.
74. Rathke, C., et al., *Chromatin dynamics during spermiogenesis.* Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms, 2014. **1839**(3): p. 155-168.
75. Lai, F. and M.L. King, *Repressive translational control in germ cells.* Molecular reproduction and development, 2013. **80**(8): p. 665-676.
76. Kee, K., et al., *Human DAZL, DAZ and BOULE genes modulate primordial germ-cell and haploid gamete formation.* Nature, 2009. **462**(7270): p. 222.
77. Sette, C., V. Messina, and M.P. Paronetto, *Sam68: a new STAR in the male fertility firmament.* Journal of andrology, 2010. **31**(1): p. 66-74.
78. Li, L.-J., et al., *Decreased expression of SAM68 in human testes with spermatogenic defects.* Fertility and sterility, 2014. **102**(1): p. 61-67. e3.

79. Ehrmann, I. and D.J. Elliott, *Post-transcriptional control in the male germ line.* Reproductive biomedicine online, 2005. **10**(1): p. 55-63.
80. Guo, T.B., et al., *Spermatogenetic Expression of RNA-Binding Motif Protein 7, a Protein That Interacts With Splicing Factors.* Journal of andrology, 2003. **24**(2): p. 204-214.
81. Krausz, C., A.R. Escamilla, and C. Chianese, *Genetics of male infertility: from research to clinic.* Reproduction, 2015. **150**(5): p. R159-R174.
82. Olooto, W., *Infertility in male; risk factors, causes and management-A review.* Journal of Microbiology and Biotechnology Research, 2017. **2**(4): p. 641-645.
83. Organization, W.H., *WHO laboratory manual for the examination and processing of human semen.* 2010.
84. Dada, R., N. Gupta, and K. Kucheria, *Molecular screening for Yq microdeletion in men with idiopathic oligozoospermia and azoospermia.* Journal of biosciences, 2003. **28**(2): p. 163-168.
85. Dohle, G., et al., *Genetic risk factors in infertile men with severe oligozoospermia and azoospermia.* Human Reproduction, 2002. **17**(1): p. 13-16.
86. Aston, K.I., et al., *Evaluation of 172 candidate polymorphisms for association with oligozoospermia or azoospermia in a large cohort of men of European descent.* Human Reproduction, 2010. **25**(6): p. 1383-1397.
87. Martin, R.H., et al., *A comparison of the frequency of sperm chromosome abnormalities in men with mild, moderate, and severe oligozoospermia.* Biology of Reproduction, 2003. **69**(2): p. 535-539.
88. Ferlin, A., et al., *Heat shock protein and heat shock factor expression in sperm: relation to oligozoospermia and varicocele.* The Journal of urology, 2010. **183**(3): p. 1248-1252.
89. Sermondade, N., et al., *Obesity and increased risk for oligozoospermia and azoospermia.* Archives of internal medicine, 2012. **172**(5): p. 440-442.
90. Bhardwaj, A., et al., *Status of vitamin E and reduced glutathione in semen of oligozoospermic and azoospermic patients.* Asian journal of andrology, 2000. **2**(3): p. 225-228.
91. Agarwal, A., et al., *Mechanisms of oligozoospermia: an oxidative stress perspective.* Systems biology in reproductive medicine, 2014. **60**(4): p. 206-216.
92. Aswal, A., et al., *To analyse the semen for various parameters with special reference to lifestyle factors.* International Journal of Reproduction, Contraception, Obstetrics and Gynecology, 2017. **6**(6): p. 2589-2592.
93. Choucair, F.B., et al., *High level of DNA fragmentation in sperm of Lebanese infertile men using Sperm Chromatin Dispersion test.* Middle East Fertility Society Journal, 2016. **21**(4): p. 269-276.
94. Stahl, P.J., et al., *A decade of experience emphasizes that testing for Y microdeletions is essential in American men with azoospermia and severe oligozoospermia.* Fertility and sterility, 2010. **94**(5): p. 1753-1756.
95. Fu, L., et al., *Genetic screening for chromosomal abnormalities and Y chromosome microdeletions in Chinese infertile men.* Journal of assisted reproduction and genetics, 2012. **29**(6): p. 521-527.
96. Kim, S.Y., et al., *Y Chromosome Microdeletions in Infertile Men with Non-obstructive Azoospermia and Severe Oligozoospermia.* Journal of reproduction & infertility, 2017. **18**(3): p. 307.
97. Pylyp, L.Y., et al., *Chromosomal abnormalities in patients with oligozoospermia and non-obstructive azoospermia.* Journal of assisted reproduction and genetics, 2013. **30**(5): p. 729-732.

98. Yatsenko, A.N., et al., *UBE2B mRNA alterations are associated with severe oligozoospermia in infertile men.* MHR: Basic science of reproductive medicine, 2013. **19**(6): p. 388-394.
99. Guo, X., et al., *Differential expression of VASA gene in ejaculated spermatozoa from normozoospermic men and patients with oligozoospermia.* Asian journal of andrology, 2007. **9**(3): p. 339-344.
100. Amir, A. and F. Oenzil, *Alteration Expression of Bax, Bcl-2 and VDAC1 Genes in Oligozoospermic and Fertile Subjects.* Pakistan journal of biological sciences: PJBS, 2016. **19**(2): p. 71-76.
101. Li, Z., et al., *Integrated Analysis of DNA Methylation and mRNA Expression Profiles to Identify Key Genes in Severe Oligozoospermia.* Frontiers in physiology, 2017. **8**.
102. Montjean, D., et al., *Sperm transcriptome profiling in oligozoospermia.* Journal of assisted reproduction and genetics, 2012. **29**(1): p. 3-10.
103. Li, Z., et al., *Integrated analysis miRNA and mRNA profiling in patients with severe oligozoospermia reveals miR-34c-3p downregulates PLCXD3 expression.* Oncotarget, 2016. **7**(33): p. 52781.
104. CHEMES, H.E., *Phenotypes of sperm pathology: genetic and acquired forms in infertile men.* Journal of andrology, 2000. **21**(6): p. 799-808.
105. Bonanno, O., et al., *Sperm of patients with severe asthenozoospermia show biochemical, molecular and genomic alterations.* Reproduction, 2016. **152**(6): p. 695-704.
106. Kumar, S., et al., *Environmental & lifestyle factors in deterioration of male reproductive health.* The Indian journal of medical research, 2014. **140**(Suppl 1): p. S29.
107. Eslamian, G., et al., *Intake of food groups and idiopathic asthenozoospermia: a case–control study.* Human reproduction, 2012. **27**(11): p. 3328-3336.
108. Taha, E.A., et al., *Correlation between seminal lead and cadmium and seminal parameters in idiopathic oligoasthenozoospermic males.* Central European journal of urology, 2013. **66**(1): p. 84.
109. Amoako, A., et al., *Anandamide modulates human sperm motility: implications for men with asthenozoospermia and oligoasthenoteratozoospermia.* Human reproduction, 2013. **28**(8): p. 2058-2066.
110. Lenzi, A., et al., *A placebo-controlled double-blind randomized trial of the use of combined l-carnitine and l-acetyl-carnitine treatment in men with asthenozoospermia.* Fertility and Sterility, 2004. **81**(6): p. 1578-1584.
111. Balercia, G., et al., *Coenzyme Q10 treatment in infertile men with idiopathic asthenozoospermia: a placebo-controlled, double-blind randomized trial.* Fertility and Sterility, 2009. **91**(5): p. 1785-1792.
112. Bansal, S.K., et al., *Differential genes expression between fertile and infertile spermatozoa revealed by transcriptome analysis.* PloS one, 2015. **10**(5): p. e0127007.
113. Jodar, M., et al., *Differential RNAs in the sperm cells of asthenozoospermic patients.* Human reproduction, 2012. **27**(5): p. 1431-1438.
114. Yu, Q., et al., *SEMG1 may be the candidate gene for idiopathic asthenozoospermia.* Andrologia, 2014. **46**(2): p. 158-166.
115. Hashemitabar, M., et al., *A proteomic analysis on human sperm tail: comparison between normozoospermia and asthenozoospermia.* Journal of assisted reproduction and genetics, 2015. **32**(6): p. 853-863.
116. Amaral, A., et al., *Identification of proteins involved in human sperm motility using high-throughput differential proteomics.* Journal of proteome research, 2014. **13**(12): p. 5670-5684.

117. Martínez-Heredia, J., et al., *Identification of proteomic differences in asthenozoospermic sperm samples.* Human reproduction, 2008. **23**(4): p. 783-791.
118. Saraswat, M., et al., *Human spermatozoa quantitative proteomic signature classifies normo-and asthenozoospermia.* Molecular & Cellular Proteomics, 2017. **16**(1): p. 57-72.
119. Du, Y., et al., *Promoter targeted bisulfite sequencing reveals DNA methylation profiles associated with low sperm motility in asthenozoospermia.* Human Reproduction, 2015. **31**(1): p. 24-33.
120. Liu, B., et al., *Analysis and difference of voltage-dependent anion channel mRNA in ejaculated spermatozoa from normozoospermic fertile donors and infertile patients with idiopathic asthenozoospermia.* Journal of assisted reproduction and genetics, 2010. **27**(12): p. 719-724.
121. Chen, J., et al., *Differential expression of ODF1 in human ejaculated spermatozoa and its clinical significance.* Zhonghua nan ke xue= National journal of andrology, 2009. **15**(10): p. 891-894.
122. Cai, Z.M., et al., *Low expression of glycoprotein subunit 130 in ejaculated spermatozoa from asthenozoospermic men.* Journal of andrology, 2006. **27**(5): p. 645-652.
123. Zhou, J.-H., et al., *The expression of cysteine-rich secretory protein 2 (CRISP2) and its specific regulator miR-27b in the spermatozoa of patients with asthenozoospermia.* Biology of Reproduction, 2015. **92**(1).
124. Wu, W.-B., et al., *Expression of TEKT4 protein decreases in the ejaculated spermatozoa of idiopathic asthenozoospermic men.* Zhonghua nan ke xue= National journal of andrology, 2012. **18**(6): p. 514-517.
125. Said, L., A. Saad, and S. Carreau, *Differential expression of mRNA aromatase in ejaculated spermatozoa from infertile men in relation to either asthenozoospermia or teratozoospermia.* Andrologia, 2014. **46**(2): p. 136-146.
126. Paoli, D., et al., *Sperm glyceraldehyde 3-phosphate dehydrogenase gene expression in asthenozoospermic spermatozoa.* Asian journal of andrology, 2017. **19**(4): p. 409.
127. Kempisty, B., et al., *Evaluation of protamines 1 and 2 transcript contents in spermatozoa from asthenozoospermic men.* Folia Histochemica et cytobiologica, 2007. **45**(I): p. 109-113.
128. Brugo-Olmedo, S., C. Chillik, and S. Kopelman, *Definition and causes of infertility.* Reproductive biomedicine online, 2001. **2**(1): p. 173-185.
129. Coutton, C., et al., *Teratozoospermia: spotlight on the main genetic actors in the human.* Human reproduction update, 2015. **21**(4): p. 455-485.
130. Platts, A.E., et al., *Success and failure in human spermatogenesis as revealed by teratozoospermic RNAs.* Human molecular genetics, 2007. **16**(7): p. 763-773.
131. Huang, Z., et al., *Systematic tracking of altered modules identifies disrupted pathways in teratozoospermia.* Genetics and molecular research: GMR, 2016. **15**(2).
132. Feil, R. and M.F. Fraga, *Epigenetics and the environment: emerging patterns and implications.* Nature reviews genetics, 2012. **13**(2): p. 97.
133. Esteller, M., *Epigenetics in biology and medicine*. 2008: CRC Press.
134. Kelly, T.K., D.D. De Carvalho, and P.A. Jones, *Epigenetic modifications as therapeutic targets.* Nature biotechnology, 2010. **28**(10): p. 1069.
135. Jones, P.A. and D. Takai, *The role of DNA methylation in mammalian epigenetics.* Science, 2001. **293**(5532): p. 1068-1070.
136. Dattilo, M., C. Ettore, and Y. Ménézo, *Improvement of gamete quality by stimulating and feeding the endogenous antioxidant system: mechanisms, clinical results, insights on gene-environment interactions and the role of diet.* Journal of assisted reproduction and genetics, 2016. **33**(12): p. 1633-1648.

137. Li, E., *Chromatin modification and epigenetic reprogramming in mammalian development.* Nature reviews genetics, 2002. **3**(9): p. 662.
138. Pinney, S.E., *Mammalian non-CpG methylation: stem cells and beyond.* Biology, 2014. **3**(4): p. 739-751.
139. Razin, A., *CpG methylation, chromatin structure and gene silencing—a three-way connection.* The EMBO journal, 1998. **17**(17): p. 4905-4908.
140. Wade, P.A., *Methyl CpG-binding proteins and transcriptional repression.* Bioessays, 2001. **23**(12): p. 1131-1137.
141. Weber, M. and D. Schübeler, *Genomic patterns of DNA methylation: targets and function of an epigenetic mark.* Current opinion in cell biology, 2007. **19**(3): p. 273-280.
142. Ito, S., et al., *Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification.* nature, 2010. **466**(7310): p. 1129.
143. Bhutani, N., D.M. Burns, and H.M. Blau, *DNA demethylation dynamics.* Cell, 2011. **146**(6): p. 866-872.
144. Reik, W., W. Dean, and J. Walter, *Epigenetic reprogramming in mammalian development.* Science, 2001. **293**(5532): p. 1089-1093.
145. Houshdaran, S., et al., *Widespread epigenetic abnormalities suggest a broad DNA methylation erasure defect in abnormal human sperm.* PloS one, 2007. **2**(12): p. e1289.
146. Karimian, M. and A.H. Colagar, *Association of C677T transition of the human methylenetetrahydrofolate reductase (MTHFR) gene with male infertility.* Reproduction, Fertility and Development, 2016. **28**(6): p. 785-794.
147. Liu, K., et al., *Role of genetic mutations in folate-related enzyme genes on Male Infertility.* Scientific reports, 2015. **5**: p. 15548.
148. Shen, O., et al., *Association of the Methylenetetrahydrofolate Reductase Gene A1298C Polymorphism with Male Infertility: A Meta-Analysis.* Annals of human genetics, 2012. **76**(1): p. 25-32.
149. Chellat, D., et al., *Influence of methylenetetrahydrofolate reductase C677T gene polymorphisms in Algerian infertile men with azoospermia or severe oligozoospermia.* Genetic testing and molecular biomarkers, 2012. **16**(8): p. 874-878.
150. Weiner, A.S., et al., *Polymorphisms in folate-metabolizing genes and risk of idiopathic male infertility: a study on a Russian population and a meta-analysis.* Fertility and sterility, 2014. **101**(1): p. 87-94. e3.
151. Rotondo, J.C., et al., *Methylation loss at H19 imprinted gene correlates with methylenetetrahydrofolate reductase gene promoter hypermethylation in semen samples from infertile males.* Epigenetics, 2013. **8**(9): p. 990-997.
152. Rotondo, J., et al., *Methylenetetrahydrofolate reductase gene promoter hypermethylation in semen samples of infertile couples correlates with recurrent spontaneous abortion.* Human Reproduction, 2012. **27**(12): p. 3632-3638.
153. Pacheco, S.E., et al., *Integrative DNA methylation and gene expression analyses identify DNA packaging and epigenetic regulatory genes associated with low motility sperm.* PloS one, 2011. **6**(6): p. e20280.
154. Botezatu, A., et al., *Methylation pattern of methylene tetrahydrofolate reductase and small nuclear ribonucleoprotein polypeptide N promoters in oligoasthenospermia: a case-control study.* Reproductive biomedicine online, 2014. **28**(2): p. 225-231.
155. Wu, W., et al., *Lack of association between DAZ gene methylation patterns and spermatogenic failure.* Clinical chemistry and laboratory medicine, 2010. **48**(3): p. 355-360.

156. Wu, W., et al., *Idiopathic male infertility is strongly associated with aberrant promoter methylation of methylenetetrahydrofolate reductase (MTHFR).* PloS one, 2010. **5**(11): p. e13884.
157. Karaca, M., et al., *Association between methylenetetrahydrofolate reductase (MTHFR) gene promoter hypermethylation and the risk of idiopathic male infertility.* Andrologia, 2017. **49**(7).
158. La Salle, S., et al., *Loss of spermatogonia and wide-spread DNA methylation defects in newborn male mice deficient in DNMT3L.* BMC developmental biology, 2007. **7**(1): p. 104.
159. Dhillon, V.S., M. Shahid, and S.A. Husain, *Associations of MTHFR DNMT3b 4977 bp deletion in mtDNA and GSTM1 deletion, and aberrant CpG island hypermethylation of GSTM1 in non-obstructive infertility in Indian men.* MHR: Basic science of reproductive medicine, 2007. **13**(4): p. 213-222.
160. Montjean, D., et al., *Methylation changes in mature sperm deoxyribonucleic acid from oligozoospermic men: assessment of genetic variants and assisted reproductive technology outcome.* Fertility and sterility, 2013. **100**(5): p. 1241-1247. e2.
161. Ni, K., et al., *TET enzymes are successively expressed during human spermatogenesis and their expression level is pivotal for male fertility.* Human Reproduction, 2016. **31**(7): p. 1411-1424.
162. Dansranjavin, T., et al., *2032 ROLE OF TET2 AND TET3 IN FERTILITY: FIRST EVIDENCE FOR LINK BETWEEN 5-CYTOSINE-HYDROXYMETHYLATION AND PROPER SPERMATOGENESIS.* The Journal of Urology, 2013. **189**(4): p. e834.
163. Jenkins, T., et al., *Teratozoospermia and asthenozoospermia are associated with specific epigenetic signatures.* Andrology, 2016. **4**(5): p. 843-849.
164. Urdinguio, R.G., et al., *Aberrant DNA methylation patterns of spermatozoa in men with unexplained infertility.* Human Reproduction, 2015. **30**(5): p. 1014-1028.
165. Cassuto, N.G., et al., *Different levels of DNA methylation detected in human sperms after morphological selection using high magnification microscopy.* BioMed research international, 2016. **2016**.
166. Cassuto, N.G., et al., *Correlation between DNA defect and sperm-head morphology.* Reproductive biomedicine online, 2012. **24**(2): p. 211-218.
167. Barzideh, J., R. Scott, and R. Aitken, *Analysis of the global methylation status of human spermatozoa and its association with the tendency of these cells to enter apoptosis.* Andrologia, 2013. **45**(6): p. 424-429.
168. Camprubí, C., et al., *Spermatozoa from infertile patients exhibit differences of DNA methylation associated with spermatogenesis-related processes: an array-based analysis.* Reproductive biomedicine online, 2016. **33**(6): p. 709-719.
169. Mann, C., et al., *Glutathione S-transferase polymorphisms in MS Their relationship to disability.* Neurology, 2000. **54**(3): p. 552-552.
170. Nantel, F., et al., *Spermiogenesis deficiency and germ-cell apoptosis in CREM-mutant mice.* nature, 1996. **380**(6570): p. 159.
171. Nanassy, L. and D.T. Carrell, *Abnormal methylation of the promoter of CREM is broadly associated with male factor infertility and poor sperm quality but is improved in sperm selected by density gradient centrifugation.* Fertility and sterility, 2011. **95**(7): p. 2310-2314.
172. Navarro-Costa, P., et al., *Incorrect DNA methylation of the DAZL promoter CpG island associates with defective human sperm.* Human Reproduction, 2010. **25**(10): p. 2647-2654.
173. Hammoud, S.S., et al., *Genome-wide analysis identifies changes in histone retention and epigenetic modifications at developmental and imprinted gene loci in the sperm of infertile men.* Human Reproduction, 2011. **26**(9): p. 2558-2569.

174. Kläver, R., et al., *DNA methylation in spermatozoa as a prospective marker in andrology.* Andrology, 2013. **1**(5): p. 731-740.
175. Kobayashi, H., et al., *Aberrant DNA methylation of imprinted loci in sperm from oligospermic patients.* Human molecular genetics, 2007. **16**(21): p. 2542-2551.
176. Marques, C., et al., *Abnormal methylation of imprinted genes in human sperm is associated with oligozoospermia.* MHR-Basic Science of Reproductive Medicine, 2008. **14**(2): p. 67-74.
177. El Hajj, N., et al., *Methylation status of imprinted genes and repetitive elements in sperm DNA from infertile males.* Sexual Development, 2011. **5**(2): p. 60-69.
178. Kobayashi, N., et al., *Factors associated with aberrant imprint methylation and oligozoospermia.* Scientific reports, 2017. **7**: p. 42336.
179. Marques, C.J., et al., *Genomic imprinting in disruptive spermatogenesis.* The lancet, 2004. **363**(9422): p. 1700-1702.
180. Poplinski, A., et al., *Idiopathic male infertility is strongly associated with aberrant methylation of MEST and IGF2/H19 ICR1.* International journal of andrology, 2010. **33**(4): p. 642-649.
181. Sato, A., et al., *Assessing loss of imprint methylation in sperm from subfertile men using novel methylation polymerase chain reaction Luminex analysis.* Fertility and sterility, 2011. **95**(1): p. 129-134. e4.
182. Hammoud, S.S., et al., *Alterations in sperm DNA methylation patterns at imprinted loci in two classes of infertility.* Fertility and sterility, 2010. **94**(5): p. 1728-1733.
183. Butler, M.G., *Genomic imprinting disorders in humans: a mini-review.* Journal of assisted reproduction and genetics, 2009. **26**(9-10): p. 477-486.
184. Casas, E. and T. Vavouri, *Sperm epigenomics: challenges and opportunities.* Frontiers in genetics, 2014. **5**: p. 330.
185. Bannister, A.J. and T. Kouzarides, *Regulation of chromatin by histone modifications.* Cell research, 2011. **21**(3): p. 381.
186. Bönisch, C. and S.B. Hake, *Histone H2A variants in nucleosomes and chromatin: more or less stable?* Nucleic acids research, 2012. **40**(21): p. 10719-10741.
187. Maze, I., et al., *Every amino acid matters: essential contributions of histone variants to mammalian development and disease.* Nature reviews genetics, 2014. **15**(4): p. 259.
188. Bhaumik, S.R., E. Smith, and A. Shilatifard, *Covalent modifications of histones during development and disease pathogenesis.* Nature Structural and Molecular Biology, 2007. **14**(11): p. 1008.
189. Steger, K., *Transcriptional and translational regulation of gene expression in haploid spermatids.* Anatomy and embryology, 1999. **199**(6): p. 471-487.
190. Steilmann, C., et al., *Presence of histone H3 acetylated at lysine 9 in male germ cells and its distribution pattern in the genome of human spermatozoa.* Reproduction, Fertility and Development, 2011. **23**(8): p. 997-1011.
191. Sonnack, V., et al., *Expression of hyperacetylated histone H4 during normal and impaired human spermatogenesis.* Andrologia, 2002. **34**(6): p. 384-390.
192. Vieweg, M., et al., *Methylation analysis of histone H4K12ac-associated promoters in sperm of healthy donors and subfertile patients.* Clinical epigenetics, 2015. **7**(1): p. 31.
193. Kim, J.H., et al., *Histone acetylation level and histone acetyltransferase/deacetylase activity in ejaculated sperm from normozoospermic men.* Yonsei medical journal, 2014. **55**(5): p. 1333-1340.
194. Barda, S., et al., *Expression of BET genes in testis of men with different spermatogenic impairments.* Fertility and sterility, 2012. **97**(1): p. 46-52. e5.

195. Steilmann, C., et al., *The interaction of modified histones with the bromodomain testis-specific (BRDT) gene and its mRNA level in sperm of fertile donors and subfertile men.* Reproduction, 2010. **140**(3): p. 435-443.
196. Zhang, X., M.S. Gabriel, and A. Zini, *Sperm nuclear histone to protamine ratio in fertile and infertile men: evidence of heterogeneous subpopulations of spermatozoa in the ejaculate.* Journal of andrology, 2006. **27**(3): p. 414-420.
197. Tang, M.C., et al., *Contribution of the two genes encoding histone variant h3. 3 to viability and fertility in mice.* PLoS genetics, 2015. **11**(2): p. e1004964.
198. Kouzarides, T., *Chromatin modifications and their function.* Cell, 2007. **128**(4): p. 693-705.
199. Steilmann, C., et al., *Aberrant mRNA expression of chromatin remodelling factors in round spermatid maturation arrest compared with normal human spermatogenesis.* Molecular human reproduction, 2010. **16**(10): p. 726-733.
200. Dunn, K.L. and J.R. Davie, *The many roles of the transcriptional regulator CTCF.* Biochemistry and cell biology, 2003. **81**(3): p. 161-167.
201. Esteller, M., *Non-coding RNAs in human disease.* Nature reviews genetics, 2011. **12**(12): p. 861.
202. Kaikkonen, M.U., M.T. Lam, and C.K. Glass, *Non-coding RNAs as regulators of gene expression and epigenetics.* Cardiovascular research, 2011. **90**(3): p. 430-440.
203. Iorio, M.V., C. Piovan, and C.M. Croce, *Interplay between microRNAs and the epigenetic machinery: an intricate network.* Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms, 2010. **1799**(10): p. 694-701.
204. Watanabe, T., et al., *Role for piRNAs and noncoding RNA in de novo DNA methylation of the imprinted mouse Rasgrf1 locus.* Science, 2011. **332**(6031): p. 848-852.
205. Kim, V.N., J. Han, and M.C. Siomi, *Biogenesis of small RNAs in animals.* Nature reviews Molecular cell biology, 2009. **10**(2): p. 126.
206. Wang, K.C. and H.Y. Chang, *Molecular mechanisms of long noncoding RNAs.* Molecular cell, 2011. **43**(6): p. 904-914.
207. Ni, M.-J., et al., *Identification and characterization of a novel non-coding RNA involved in sperm maturation.* PloS one, 2011. **6**(10): p. e26053.
208. Kawano, M., et al., *Novel small noncoding RNAs in mouse spermatozoa, zygotes and early embryos.* PloS one, 2012. **7**(9): p. e44542.
209. Krawetz, S.A., et al., *A survey of small RNAs in human sperm.* Human Reproduction, 2011. **26**(12): p. 3401-3412.
210. Liu, T., et al., *Microarray analysis of microRNA expression patterns in the semen of infertile men with semen abnormalities.* Molecular medicine reports, 2012. **6**(3): p. 535-542.
211. Abu-Halima, M., et al., *Altered microRNA expression profiles of human spermatozoa in patients with different spermatogenic impairments.* Fertility and sterility, 2013. **99**(5): p. 1249-1255. e16.
212. Abu-Halima, M., et al., *Panel of five microRNAs as potential biomarkers for the diagnosis and assessment of male infertility.* Fertility and sterility, 2014. **102**(4): p. 989-997. e1.
213. Abu-Halima, M., et al., *MicroRNA expression profiles in human testicular tissues of infertile men with different histopathologic patterns.* Fertility and sterility, 2014. **101**(1): p. 78-86. e2.
214. Zhuang, X., et al., *Integrated miRNA and mRNA expression profiling to identify mRNA targets of dysregulated miRNAs in non-obstructive azoospermia.* Scientific reports, 2015. **5**: p. 7922.

215. Comazzetto, S., et al., *Oligoasthenoteratozoospermia and infertility in mice deficient for miR-34b/c and miR-449 loci.* PLoS genetics, 2014. **10**(10): p. e1004597.
216. Romero, Y., et al., *Dicer1 depletion in male germ cells leads to infertility due to cumulative meiotic and spermiogenic defects.* PloS one, 2011. **6**(10): p. e25241.
217. Gu, A., et al., *Genetic variants in Piwi-interacting RNA pathway genes confer susceptibility to spermatogenic failure in a Chinese population.* Human Reproduction, 2010. **25**(12): p. 2955-2961.
218. Gou, L.-T., et al., *Ubiquitination-deficient mutations in human piwi cause male infertility by impairing histone-to-protamine exchange during spermiogenesis.* Cell, 2017. **169**(6): p. 1090-1104. e13.
219. Lü, M., et al., *Downregulation of miR-320a/383-sponge-like long non-coding RNA NLC1-C (narcolepsy candidate-region 1 genes) is associated with male infertility and promotes testicular embryonal carcinoma cell proliferation.* Cell death & disease, 2015. **6**(11): p. e1960.
220. Zhang, L., et al., *Low long non-coding RNA HOTAIR expression is associated with down-regulation of Nrf2 in the spermatozoa of patients with asthenozoospermia or oligoasthenozoospermia.* International journal of clinical and experimental pathology, 2015. **8**(11): p. 14198.
221. Cheng, Q., et al., *A conserved antioxidant response element (ARE) in the promoter of human carbonyl reductase 3 (CBR3) mediates induction by the master redox switch Nrf2.* Biochemical pharmacology, 2012. **83**(1): p. 139-148.
222. Li, L., et al., *A long non-coding RNA interacts with Gfra1 and maintains survival of mouse spermatogonial stem cells.* Cell death & disease, 2016. **7**(3): p. e2140.
223. Cremer, T. and C. Cremer, *Chromosome territories, nuclear architecture and gene regulation in mammalian cells.* Nature reviews genetics, 2001. **2**(4): p. 292.
224. Reddy, K., et al., *Transcriptional repression mediated by repositioning of genes to the nuclear lamina.* nature, 2008. **452**(7184): p. 243.
225. Iborra, F.J., et al., *Active RNA polymerases are localized within discrete transcription "factories' in human nuclei.* Journal of cell science, 1996. **109**(6): p. 1427-1436.
226. Ibarra, A. and M.W. Hetzer, *Nuclear pore proteins and the control of genome functions.* Genes & development, 2015. **29**(4): p. 337-349.
227. Kobayashi, T., *A new role of the rDNA and nucleolus in the nucleus—rDNA instability maintains genome integrity.* Bioessays, 2008. **30**(3): p. 267-272.
228. Norton, J.T. and S. Huang, *The perinucleolar compartment: RNA metabolism and cancer*, in *RNA and Cancer*. 2013, Springer. p. 139-152.
229. Finch, K.A., et al., *Nuclear organization in human sperm: preliminary evidence for altered sex chromosome centromere position in infertile males.* Human Reproduction, 2008. **23**(6): p. 1263-1270.
230. Olszewska, M., E. Wiland, and M. Kurpisz, *Positioning of chromosome 15, 18, X and Y centromeres in sperm cells of fertile individuals and infertile patients with increased level of aneuploidy.* Chromosome research, 2008. **16**(6): p. 875-890.
231. Wiland, E., et al., *Topology of chromosome centromeres in human sperm nuclei with high levels of DNA damage.* Scientific reports, 2016. **6**: p. 31614.
232. Ioannou, D., et al., *Nuclear organisation of sperm remains remarkably unaffected in the presence of defective spermatogenesis.* Chromosome research, 2011. **19**(6): p. 741.
233. Olszewska, M., et al., *Genetic dosage and position effect of small supernumerary marker chromosome (sSMC) in human sperm nuclei in infertile male patient.* Scientific reports, 2015. **5**: p. 17408.

234. Garrido, N., et al., *Microarray analysis in sperm from fertile and infertile men without basic sperm analysis abnormalities reveals a significantly different transcriptome.* Fertility and sterility, 2009. **91**(4): p. 1307-1310.
235. García-Herrero, S., et al., *Ontological evaluation of transcriptional differences between sperm of infertile males and fertile donors using microarray analysis.* Journal of assisted reproduction and genetics, 2010. **27**(2-3): p. 111-120.
236. Zhao, C., et al., *Identification of several proteins involved in regulation of sperm motility by proteomic analysis.* Fertility and sterility, 2007. **87**(2): p. 436-438.
237. Yu, Z., T. Raabe, and N.B. Hecht, *MicroRNA Mirn122a reduces expression of the posttranscriptionally regulated germ cell transition protein 2 (Tnp2) messenger RNA (mRNA) by mRNA cleavage.* Biology of reproduction, 2005. **73**(3): p. 427-433.
238. HAYASHI, T., et al., *Requirement of Notch 1 and its ligand jagged 2 expressions for spermatogenesis in rat and human testes.* Journal of andrology, 2001. **22**(6): p. 999-1011.
239. Depa-Martynów, M., et al., *Association between fertilin beta, protamines 1 and 2 and spermatid-specific linker histone H1-like protein mRNA levels, fertilization ability of human spermatozoa, and quality of preimplantation embryos.* Folia histochemica et cytobiologica, 2007. **45**(I): p. 79-85.
240. Bornstein, C., et al., *SPATA18, a spermatogenesis-associated gene, is a novel transcriptional target of p53 and p63.* Molecular and cellular biology, 2011. **31**(8): p. 1679-1689.
241. Miyagawa, Y., et al., *Single-Nucleotide Polymorphisms and Mutation Analyses of the TNP1 and TNP2 Genes of Fertile and Infertile Human Male Populations.* Journal of andrology, 2005. **26**(6): p. 779-786.
242. Bao, J., et al., *RAN-binding protein 9 is involved in alternative splicing and is critical for male germ cell development and male fertility.* PLoS genetics, 2014. **10**(12): p. e1004825.
243. Baker, C.C. and M.T. Fuller, *Translational control of meiotic cell cycle progression and spermatid differentiation in male germ cells by a novel eIF4G homolog.* Development, 2007. **134**(15): p. 2863-2869.
244. Miller, D., G.C. Ostermeier, and S.A. Krawetz, *The controversy, potential and roles of spermatozoal RNA.* Trends in molecular medicine, 2005. **11**(4): p. 156-163.
245. Lalancette, C., et al., *Paternal contributions: new functional insights for spermatozoal RNA.* Journal of cellular biochemistry, 2008. **104**(5): p. 1570-1579.
246. Sendler, E., et al., *Stability, delivery and functions of human sperm RNAs at fertilization.* Nucleic acids research, 2013. **41**(7): p. 4104-4117.
247. Gupta, V. and R.N. Bamezai, *Human pyruvate kinase M2: a multifunctional protein.* Protein science, 2010. **19**(11): p. 2031-2044.
248. Fullerton, M.D., F. Hakimuddin, and M. Bakovic, *Developmental and metabolic effects of disruption of the mouse CTP: phosphoethanolamine cytidylyltransferase gene (Pcyt2).* Molecular and cellular biology, 2007. **27**(9): p. 3327-3336.
249. Dawlaty, M.M., et al., *Resolution of sister centromeres requires RanBP2-mediated SUMOylation of topoisomerase IIα.* Cell, 2008. **133**(1): p. 103-115.
250. Cox, G.A., et al., *The mouse fidgetin gene defines a new role for AAA family proteins in mammalian development.* Nature genetics, 2000. **26**(2): p. 198.
251. Lazzaretti, D., I. Tournier, and E. Izaurralde, *The C-terminal domains of human TNRC6A, TNRC6B, and TNRC6C silence bound transcripts independently of Argonaute proteins.* Rna, 2009. **15**(6): p. 1059-1066.
252. Yan, W., et al., *HILS1 is a spermatid-specific linker histone H1-like protein implicated in chromatin remodeling during mammalian spermiogenesis.* Proceedings of the National Academy of Sciences, 2003. **100**(18): p. 10546-10551.

253. Hammoud, S.S., et al., *Distinctive chromatin in human sperm packages genes for embryo development.* Nature, 2009. **460**(7254): p. 473.
254. Mao, S., et al., *A comparison of sperm RNA-seq methods.* Systems biology in reproductive medicine, 2014. **60**(5): p. 308-315.
255. Jodar, M., et al., *Absence of sperm RNA elements correlates with idiopathic male infertility.* Science translational medicine, 2015. **7**(295): p. 295re6-295re6.
256. Cappallo-Obermann, H. and A.-N. Spiess, *Comment on "Absence of sperm RNA elements correlates with idiopathic male infertility".* Science translational medicine, 2016. **8**(353): p. 353tc1-353tc1.
257. Pereira, N., et al., *Identifying the epigenetic basis of idiopathic infertility using next-generation sequencing of spermatozoal RNAs.* Fertility and sterility, 2016. **105**(2): p. e33-e34.
258. Goodrich, R.J., E. Anton, and S.A. Krawetz, *Isolating mRNA and small noncoding RNAs from human sperm*, in *Spermatogenesis*. 2013, Springer. p. 385-396.
259. Bianchi, E., et al., *High-quality human and rat spermatozoal RNA isolation for functional genomic studies.* Andrology, 2018. **6**(2): p. 374-383.
260. De Gannes, M.K., *Rapid method of processing sperm for nucleic acid extraction in clinical research*. 2014, University of Massachusetts Amherst.
261. Wu, H., et al., *Rapid method for the isolation of mammalian sperm DNA.* Biotechniques, 2015. **58**(6): p. 293-300.
262. Cobb, B.D. and J.M. CIarkson, *A simple procedure for optimising the polymerase chain reaction (PCR) using modified Taguchi methods.* Nucleic Acids Research, 1994. **22**(18): p. 3801-3805.
263. Harrison, R. and S.E. Vickers, *Use of fluorescent probes to assess membrane integrity in mammalian spermatozoa.* Journal of Reproduction and Fertility, 1990. **88**(1): p. 343-352.
264. Lewis, S.E., *Is sperm evaluation useful in predicting human fertility?* Reproduction, 2007. **134**(1): p. 31-40.
265. Krawetz, S.A., *Paternal contribution: new insights and future challenges.* Nature Reviews Genetics, 2005. **6**(8): p. 633.
266. El Fekih, S., et al., *Sperm RNA preparation for transcriptomic analysis: Review of the techniques and personal experience.* Andrologia, 2017. **49**(10).
267. Mao, S., et al., *Evaluation of the effectiveness of semen storage and sperm purification methods for spermatozoa transcript profiling.* Systems biology in reproductive medicine, 2013. **59**(5): p. 287-295.
268. Georgiadis, A.P., et al., *High quality RNA in semen and sperm: isolation, analysis and potential application in clinical testing.* The Journal of urology, 2015. **193**(1): p. 352-359.
269. Remohí, J. and A. Pellicer, *Profound Transcriptomic Differences Found between Sperm Samples from Sperm Donors vs. Patients Undergoing Assisted Reproduction Techniques Tends to Disappear after Swim-up Sperm Preparation Technique.* International Journal of Fertility & Sterility, 2010. **4**(3): p. 114.
270. Johnson, G.D., et al., *The sperm nucleus: chromatin, RNA, and the nuclear matrix.* Reproduction, 2011. **141**(1): p. 21-36.
271. Goodrich, R., G. Johnson, and S.A. Krawetz, *The preparation of human spermatozoal RNA for clinical analysis.* Archives of andrology, 2007. **53**(3): p. 161-167.
272. Binder, N.K., et al., *Male obesity is associated with changed spermatozoa Cox4i1 mRNA level and altered seminal vesicle fluid composition in a mouse model.* MHR: Basic science of reproductive medicine, 2015. **21**(5): p. 424-434.

273. Perreault, S.D., R.A. Wolff, and B.R. Zirkin, *The role of disulfide bond reduction during mammalian sperm nuclear decondensation in vivo.* Developmental biology, 1984. **101**(1): p. 160-167.
274. Schuster, A., et al., *SpermBase: a database for sperm-borne RNA contents.* Biology of reproduction, 2016. **95**(5).
275. Manchester, K.L., *Value of A260/A280 ratios for measurement of purity of nucleic acids.* Biotechniques, 1995. **19**(2): p. 208-210.
276. Cappallo-Obermann, H., et al., *Highly purified spermatozoal RNA obtained by a novel method indicates an unusual 28S/18S rRNA ratio and suggests impaired ribosome assembly.* Molecular human reproduction, 2011. **17**(11): p. 669-678.
277. Jenkins, T.G., et al., *Paternal aging and associated intraindividual alterations of global sperm 5-methylcytosine and 5-hydroxymethylcytosine levels.* Fertility and sterility, 2013. **100**(4): p. 945-951. e2.
278. Schuster, S.C., *Next-generation sequencing transforms today's biology.* Nature methods, 2007. **5**(1): p. 16.
279. Wang, Z., M. Gerstein, and M. Snyder, *RNA-Seq: a revolutionary tool for transcriptomics.* Nature Reviews Genetics, 2009. **10**(1): p. 57.
280. Han, Y., et al., *Advanced applications of RNA sequencing and challenges.* Bioinformatics and biology insights, 2015. **9**: p. BBI. S28991.
281. Hornett, E.A. and C.W. Wheat, *Quantitative RNA-Seq analysis in non-model species: assessing transcriptome assemblies as a scaffold and the utility of evolutionary divergent genomic reference species.* BMC genomics, 2012. **13**(1): p. 361.
282. Etebari, K., et al., *Deep sequencing-based transcriptome analysis of Plutella xylostella larvae parasitized by Diadegma semiclausum.* BMC genomics, 2011. **12**(1): p. 446.
283. Birzele, F., et al., *Into the unknown: expression profiling without genome sequence information in CHO by next generation sequencing.* Nucleic acids research, 2010. **38**(12): p. 3999-4010.
284. Conesa, A., et al., *A survey of best practices for RNA-seq data analysis.* Genome biology, 2016. **17**(1): p. 13.
285. Levin, J.Z., et al., *Comprehensive comparative analysis of strand-specific RNA sequencing methods.* Nature methods, 2010. **7**(9): p. 709.
286. Zhang, Z., et al., *Strand-specific libraries for high throughput RNA sequencing (RNA-Seq) prepared without poly (A) selection.* Silence, 2012. **3**(1): p. 9.
287. Bogdanova, E.A., et al., *Normalizing cDNA libraries.* Current protocols in molecular biology, 2010: p. 5.12. 1-5.12. 27.
288. Consortium, E., *Standards, guidelines and best practices for RNA-seq.* V1. 0, 2011: p. 1-7.
289. Trapnell, C., et al., *Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks.* Nature protocols, 2012. **7**(3): p. 562.
290. Schurch, N.J., et al., *How many biological replicates are needed in an RNA-seq experiment and which differential expression tool should you use?* Rna, 2016. **22**(6): p. 839-851.
291. Chung, L.M., et al., *Differential expression analysis for paired RNA-seq data.* BMC bioinformatics, 2013. **14**(1): p. 110.
292. Risso, D., et al., *Normalization of RNA-seq data using factor analysis of control genes or samples.* Nature biotechnology, 2014. **32**(9): p. 896.
293. Chhangawala, S., et al., *The impact of read length on quantification of differentially expressed genes and splice junction detection.* Genome biology, 2015. **16**(1): p. 131.

294. Cock, P.J., et al., *The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants.* Nucleic acids research, 2009. **38**(6): p. 1767-1771.
295. Anon. 2017; Available from: https://www.illumina.com/science/education/sequencing-coverage.html.
296. Ged.msu.edu. *Checking read quality with FastQC — ANGUS 2.0 documentation*. 2017; Available from: http://ged.msu.edu/angus/tutorials-2011/fastq_tutorial.html
297. Andrews, S., *FastQC: a quality control tool for high throughput sequence data.* 2010.
298. Kuenne, C., et al., *RNA-Seq analysis of isolated satellite cells in Prmt5 deficient mice.* Genomics data, 2015. **5**: p. 122-125.
299. Del Fabbro, C., et al., *An extensive evaluation of read trimming effects on Illumina NGS data analysis.* PloS one, 2013. **8**(12): p. e85024.
300. Williams, C.R., et al., *Trimming of sequence reads alters RNA-Seq gene expression estimates.* BMC bioinformatics, 2016. **17**(1): p. 103.
301. Cloonan, N. and S.M. Grimmond, *Simplifying complexity.* Nature methods, 2010. **7**(10): p. 793.
302. Fonseca, N.A., et al., *Tools for mapping high-throughput sequencing data.* Bioinformatics, 2012. **28**(24): p. 3169-3177.
303. Engström, P.G., et al., *Systematic evaluation of spliced alignment programs for RNA-seq data.* Nature methods, 2013. **10**(12): p. 1185.
304. Trapnell, C., et al., *Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation.* Nature biotechnology, 2010. **28**(5): p. 511.
305. Kim, D., et al., *TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions.* Genome biology, 2013. **14**(4): p. R36.
306. Li, H., et al., *The sequence alignment/map format and SAMtools.* Bioinformatics, 2009. **25**(16): p. 2078-2079.
307. Sims, D., et al., *Sequencing depth and coverage: key considerations in genomic analyses.* Nature Reviews Genetics, 2014. **15**(2): p. 121.
308. Messina, D. *1 graph that will give you a new perspective on your sequencing experiment - Cofactor Genomics*. 2017; Available from: https://cofactorgenomics.com/1-graph-will-give-you-new-perspective-sequencing-experiment.
309. Bio., B. *NGS Quality Control in RNA Sequencing- Some Free Tools - Bitesize Bio.* 2017; Available from: http://bitesizebio.com/13559/ngs-quality-control-in-rna-sequencing-some-free-tools/.
310. Rseqc.sourceforge.net. *RSeQC: An RNA-seq Quality Control Package — RSeQC documentation*. 2017; Available from: http://rseqc.sourceforge.net/
311. Wang, L., S. Wang, and W. Li, *RSeQC: quality control of RNA-seq experiments.* Bioinformatics, 2012. **28**(16): p. 2184-2185.
312. Roberts, A., et al., *Improving RNA-Seq expression estimates by correcting for fragment bias.* Genome biology, 2011. **12**(3): p. R22.
313. AC't Hoen, P., et al., *Reproducibility of high-throughput mRNA and small RNA sequencing across laboratories.* Nature biotechnology, 2013. **31**(11): p. 1015.
314. Martin, J.A. and Z. Wang, *Next-generation transcriptome assembly.* Nature Reviews Genetics, 2011. **12**(10): p. 671.
315. Mizuno, H., et al., *Massive parallel sequencing of mRNA in identification of unannotated salinity stress-inducible transcripts in rice (Oryza sativa L.).* BMC genomics, 2010. **11**(1): p. 683.

316. Twine, N.A., et al., *Whole transcriptome sequencing reveals gene expression and splicing differences in brain regions affected by Alzheimer's disease.* PloS one, 2011. **6**(1): p. e16266.
317. www-huber.embl.de. *HTSeq*. 2017; Available from: http://www-huber.embl.de/users/anders/HTSeq/doc/count.html.
318. Slideplayer.com. *RNA-Seq and transcriptome analysis - ppt download*. 2017; Available from: http://slideplayer.com/slide/6414081/.
319. Anders, S., P.T. Pyl, and W. Huber, *HTSeq—a Python framework to work with high-throughput sequencing data.* Bioinformatics, 2015. **31**(2): p. 166-169.
320. Moreton, J., A. Izquierdo, and R.D. Emes, *Assembly, assessment, and availability of de novo generated eukaryotic transcriptomes.* Frontiers in genetics, 2016. **6**: p. 361.
321. Dillies, M.-A., et al., *A comprehensive evaluation of normalization methods for Illumina high-throughput RNA sequencing data analysis.* Briefings in bioinformatics, 2013. **14**(6): p. 671-683.
322. Yeung, K.Y., D.R. Haynor, and W.L. Ruzzo, *Validating clustering for gene expression data.* Bioinformatics, 2001. **17**(4): p. 309-318.
323. Loraine, A.E., et al., *Analysis and visualization of RNA-Seq expression data using RStudio, Bioconductor, and Integrated Genome Browser*, in *Plant Functional Genomics*. 2015, Springer. p. 481-501.
324. Precisionbiomarker.com. *Explore data | Precision Biomarker Resources*. 2017; Available from: http://www.precisionbiomarker.com/explore-data.
325. Jumhawan, U., et al., *Selection of discriminant markers for authentication of asian palm civet coffee (kopi luwak): a metabolomics approach.* Journal of agricultural and food chemistry, 2013. **61**(33): p. 7994-8001.
326. Love, M.I., et al., *RNA-Seq workflow: gene-level exploratory analysis and differential expression.* F1000Research, 2015. **4**.
327. Love, M.I., W. Huber, and S. Anders, *Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2.* Genome biology, 2014. **15**(12): p. 550.
328. Schurch, N.J., et al., *Evaluation of tools for differential gene expression analysis by RNA-seq on a 48 biological replicate experiment.* arXiv preprint arXiv:1505.02017, 2015.
329. Seyednasrollah, F., A. Laiho, and L.L. Elo, *Comparison of software packages for detecting differential expression in RNA-seq studies.* Briefings in bioinformatics, 2013. **16**(1): p. 59-70.
330. Benjamini, Y. and Y. Hochberg, *Controlling the false discovery rate: a practical and powerful approach to multiple testing.* Journal of the royal statistical society. Series B (Methodological), 1995: p. 289-300.
331. Anders, S., et al., *Count-based differential expression analysis of RNA sequencing data using R and Bioconductor.* Nature protocols, 2013. **8**(9): p. 1765.
332. Wang, L., et al., *DEGseq: an R package for identifying differentially expressed genes from RNA-seq data.* Bioinformatics, 2009. **26**(1): p. 136-138.
333. Trapnell, C., et al., *Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks.* Nature protocols, 2012. **7**(3): p. 562-578.
334. Anders, S. and W. Huber, *Differential expression analysis for sequence count data.* Genome biology, 2010. **11**(10): p. R106.
335. Varet, H., et al., *SARTools: a DESeq2-and edgeR-based R pipeline for comprehensive differential analysis of RNA-Seq data.* PloS one, 2016. **11**(6): p. e0157022.
336. Chen, J., et al., *ToppGene Suite for gene list enrichment analysis and candidate gene prioritization.* Nucleic acids research, 2009. **37**(suppl_2): p. W305-W311.
337. Luo, W., et al., *Pathview Web: user friendly pathway visualization and data integration.* Nucleic acids research, 2017. **45**(W1): p. W501-W508.

338. Heberle, H., et al., *InteractiVenn: a web-based tool for the analysis of sets through Venn diagrams.* BMC bioinformatics, 2015. **16**(1): p. 169.
339. Babicki, S., et al., *Heatmapper: web-enabled heat mapping for all.* Nucleic acids research, 2016. **44**(W1): p. W147-W153.
340. Tunc, O., J. Thompson, and K. Tremellen, *Development of the NBT assay as a marker of sperm oxidative stress.* International journal of andrology, 2010. **33**(1): p. 13-21.
341. Esfandiari, N., et al., *Utility of the nitroblue tetrazolium reduction test for assessment of reactive oxygen species production by seminal leukocytes and spermatozoa.* Journal of Andrology, 2003. **24**(6): p. 862-870.
342. Fernández, J.L., et al., *Simple determination of human sperm DNA fragmentation with an improved sperm chromatin dispersion test.* Fertility and sterility, 2005. **84**(4): p. 833-842.
343. Erenpreisa, J., et al., *Toluidine blue test for sperm DNA integrity and elaboration of image cytometry algorithm.* Cytometry Part A, 2003. **52**(1): p. 19-27.
344. Foresta, C., et al., *Sperm nuclear instability and staining with aniline blue: abnormal persistance of histones in spermatozoa in infertile men.* International journal of andrology, 1992. **15**(4): p. 330-337.
345. Benjamini, Y. and D. Yekutieli, *The control of the false discovery rate in multiple testing under dependency.* Annals of statistics, 2001: p. 1165-1188.
346. O'Brien, K.L.F., A.C. Varghese, and A. Agarwal, *The genetic causes of male factor infertility: a review.* Fertility and sterility, 2010. **93**(1): p. 1-12.
347. Kitamura, A., et al., *Epigenetic alterations in sperm associated with male infertility.* Congenital anomalies, 2015. **55**(3): p. 133-144.
348. Lalancette, C., et al., *Identification of human sperm transcripts as candidate markers of male fertility.* Journal of molecular medicine, 2009. **87**(7): p. 735-748.
349. Margolin, G., et al., *Integrated transcriptome analysis of mouse spermatogenesis.* BMC genomics, 2014. **15**(1): p. 39.
350. Zheng, J., et al., *Decreased Sperm Motility Retarded ICSI Fertilization Rate in Severe Oligozoospermia but Good-Quality Embryo Transfer Had Achieved the Prospective Clinical Outcomes.* PloS one, 2016. **11**(9): p. e0163524.
351. Liu, D.Y. and H.G. Baker, *High frequency of defective sperm–zona pellucida interaction in oligozoospermic infertile men.* Human Reproduction, 2004. **19**(2): p. 228-233.
352. Foresta, C., et al., *Y-chromosome deletions in idiopathic severe testiculopathies.* The Journal of Clinical Endocrinology & Metabolism, 1997. **82**(4): p. 1075-1080.
353. Siu, M.K. and C.Y. Cheng, *Extracellular matrix and its role in spermatogenesis*, in *Molecular mechanisms in spermatogenesis*. 2009, Springer. p. 74-91.
354. Schrader, M., et al., *Cyclin A1 and gametogenesis in fertile and infertile patients: a potential new molecular diagnostic marker.* Human Reproduction, 2002. **17**(9): p. 2338-2343.
355. van der Meer, T., et al., *Cyclin A1 protein shows haplo-insufficiency for normal fertility in male mice.* Reproduction, 2004. **127**(4): p. 503-511.
356. Greenbaum, M.P., et al., *TEX14 is essential for intercellular bridges and fertility in male mice.* Proceedings of the National Academy of Sciences of the United States of America, 2006. **103**(13): p. 4982-4987.
357. Sironen, A., et al., *An exonic insertion within Tex14 gene causes spermatogenic arrest in pigs.* BMC genomics, 2011. **12**(1): p. 591.
358. Shang, E., et al., *The first bromodomain of Brdt, a testis-specific member of the BET sub-family of double-bromodomain-containing proteins, is essential for male germ cell differentiation.* Development, 2007. **134**(19): p. 3507-3515.

359. Judis, L., et al., *Meiosis I arrest and azoospermia in an infertile male explained by failure of formation of a component of the synaptonemal complex.* Fertility and sterility, 2004. **81**(1): p. 205-209.
360. Griffin, J., et al., *Analysis of the meiotic recombination gene REC8 for sequence variations in a population with severe male factor infertility.* Systems biology in reproductive medicine, 2008. **54**(3): p. 163-165.
361. Davalieva, K., et al., *Proteomic analysis of seminal plasma in men with different spermatogenic impairment.* Andrologia, 2012. **44**(4): p. 256-264.
362. LUE, Y.H., et al., *Mild testicular hyperthermia induces profound transitional spermatogenic suppression through increased germ cell apoptosis in adult cynomolgus monkeys (Macaca fascicularis).* Journal of Andrology, 2002. **23**(6): p. 799-805.
363. Mostafa, T., et al., *Seminal BAX and BCL2 gene and protein expressions in infertile men with varicocele.* Urology, 2014. **84**(3): p. 590-595.
364. Feki, A., et al., *BARD1 expression during spermatogenesis is associated with apoptosis and hormonally regulated.* Biology of Reproduction, 2004. **71**(5): p. 1614-1624.
365. Rice, J.C. and C.D. Allis, *Histone methylation versus histone acetylation: new insights into epigenetic regulation.* Current opinion in cell biology, 2001. **13**(3): p. 263-273.
366. Zhang, X., et al., *Systematic identification and characterization of long non-coding RNAs in mouse mature sperm.* PloS one, 2017. **12**(3): p. e0173402.
367. Dianatpour, A. and S. Ghafouri-Fard, *The Role of Long Non Coding RNAs in the Repair of DNA Double Strand Breaks.* International Journal of Molecular and Cellular Medicine (IJMCM), 2017. **6**(1): p. 1-12.
368. Zerbino, D.R., et al., *Ensembl 2018.* Nucleic acids research, 2017. **46**(D1): p. D754-D761.
369. Neesen, J., et al., *Disruption of an inner arm dynein heavy chain gene results in asthenozoospermia and reduced ciliary beat frequency.* Human molecular genetics, 2001. **10**(11): p. 1117-1128.
370. Hosken, D.J. and D.J. Hodgson, *Why do sperm carry RNA? Relatedness, conflict, and control.* Trends in ecology & evolution, 2014. **29**(8): p. 451-455.
371. Pereira, N., et al., *Investigating the role of sperm-specific RNA to screen men with unexplained infertility.* Fertility and Sterility, 2017. **108**(3): p. e46.
372. Tanaka, H., et al., *Mice deficient in the axonemal protein Tektin-t exhibit male infertility and immotile-cilium syndrome due to impaired inner arm dynein function.* Molecular and cellular biology, 2004. **24**(18): p. 7958-7964.
373. Lestari, S.W., et al., *Evaluation of outer dense fiber-1 and-2 protein expression in asthenozoospermic infertile men.* Medical Journal of Indonesia, 2015. **24**(2): p. 79.
374. Shen, S., et al., *Comparative proteomic study between human normal motility sperm and idiopathic asthenozoospermia.* World journal of urology, 2013. **31**(6): p. 1395-1401.
375. Zhao, W., et al., *Outer dense fibers stabilize the axoneme to maintain sperm motility.* Journal of cellular and molecular medicine, 2017.
376. Hess, H., H. Heid, and W.W. Franke, *Molecular characterization of mammalian cylicin, a basic protein of the sperm head cytoskeleton.* The Journal of cell biology, 1993. **122**(5): p. 1043-1052.
377. Kierszenbaum, A.L., E. Rivkin, and L.L. Tres, *Cytoskeletal track selection during cargo transport in spermatids is relevant to male fertility.* Spermatogenesis, 2011. **1**(3): p. 221-230.

378. Breitbart, H., G. Cohen, and S. Rubinstein, *Role of actin cytoskeleton in mammalian sperm capacitation and the acrosome reaction.* Reproduction, 2005. **129**(3): p. 263-268.
379. Liu, Y., et al., *Proteomic pattern changes associated with obesity-induced asthenozoospermia.* Andrology, 2015. **3**(2): p. 247-259.
380. Piomboni, P., et al., *The role of mitochondria in energy production for human sperm motility.* International journal of andrology, 2012. **35**(2): p. 109-124.
381. Scarpulla, R.C., *Transcriptional paradigms in mammalian mitochondrial biogenesis and function.* Physiological reviews, 2008. **88**(2): p. 611-638.
382. Gibbons, J.G., et al., *Ribosomal DNA copy number is coupled with gene expression variation and mitochondrial abundance in humans.* Nature communications, 2014. **5**.
383. Rozanska, A., et al., *The human RNA-binding protein RBFA promotes the maturation of the mitochondrial ribosome.* Biochemical Journal, 2017. **474**(13): p. 2145-2158.
384. Miki, K., et al., *Glyceraldehyde 3-phosphate dehydrogenase-S, a sperm-specific glycolytic enzyme, is required for sperm motility and male fertility.* Proceedings of the National Academy of Sciences of the United States of America, 2004. **101**(47): p. 16501-16506.
385. du Plessis, S.S., et al., *Oxidative phosphorylation versus glycolysis: what fuel do spermatozoa use?* Asian journal of andrology, 2015. **17**(2): p. 230.
386. Siva, A.B., et al., *Proteomics-based study on asthenozoospermia: differential expression of proteasome alpha complex.* Molecular human reproduction, 2010. **16**(7): p. 452-462.
387. Mühlemann, O., *Spermatogenesis studies reveal a distinct nonsense-mediated mRNA decay (NMD) mechanism for mRNAs with long 3′ UTRs.* PLoS genetics, 2016. **12**(5): p. e1005979.
388. MacDonald, C.C. and P.N. Grozdanov, *Nonsense in the testis: multiple roles for nonsense-mediated decay revealed in male reproduction.* Biology of Reproduction, 2017. **96**(5): p. 939-947.
389. Bao, J., et al., *UPF2-dependent nonsense-mediated mRNA decay pathway is essential for spermatogenesis by selectively eliminating longer 3'UTR transcripts.* PLoS genetics, 2016. **12**(5): p. e1005863.
390. Bastián, Y., et al., *Calpain modulates capacitation and acrosome reaction through cleavage of the spectrin cytoskeleton.* Reproduction, 2010. **140**(5): p. 673-684.
391. Rojas, F.J., M. Brush, and I. Moretti-Rojas, *Calpain–calpastatin: a novel, complete calcium-dependent protease system in human spermatozoa.* Molecular human reproduction, 1999. **5**(6): p. 520-526.
392. Ren, D., et al., *A sperm ion channel required for sperm motility and male fertility.* Nature, 2001. **413**(6856): p. 603-609.
393. Miraglia, E., et al., *Nitric oxide stimulates human sperm motility via activation of the cyclic GMP/protein kinase G signaling pathway.* Reproduction, 2011. **141**(1): p. 47-54.
394. Varano, G., et al., *Src activation triggers capacitation and acrosome reaction but not motility in human spermatozoa.* Human reproduction, 2008. **23**(12): p. 2652-2662.
395. Parab, S., et al., *HDAC6 deacetylates alpha tubulin in sperm and modulates sperm motility in Holtzman rat.* Cell and tissue research, 2015. **359**(2): p. 665-678.
396. Verma, P., et al., *Developmental Testicular Expression, Cloning, and Characterization of Rat HDAC6 In Silico.* BioMed Research International, 2017. **2017**.
397. Rajender, S., K. Avery, and A. Agarwal, *Epigenetics, spermatogenesis and male infertility.* Mutation Research/Reviews in Mutation Research, 2011. **727**(3): p. 62-71.

398. Lefievre, L., E. de Lamirande, and C. Gagnon, *Presence of cyclic nucleotide phosphodiesterases PDE1A, existing as a stable complex with calmodulin, and PDE3A in human spermatozoa.* Biology of Reproduction, 2002. **67**(2): p. 423-430.
399. Tardif, S., et al., *Clinically relevant enhancement of human sperm motility using compounds with reported phosphodiesterase inhibitor activity.* Human reproduction, 2014. **29**(10): p. 2123-2135.
400. Fath, M.A., et al., *Mkks-null mice have a phenotype resembling Bardet–Biedl syndrome.* Human molecular genetics, 2005. **14**(9): p. 1109-1118.
401. Yin, Y., et al., *Cell autonomous and nonautonomous function of CUL4B in mouse spermatogenesis.* Journal of Biological Chemistry, 2016. **291**(13): p. 6923-6935.
402. Koenig, P.-A., et al., *The E2 ubiquitin-conjugating enzyme UBE2J1 is required for spermiogenesis in mice.* Journal of Biological Chemistry, 2014. **289**(50): p. 34490-34502.
403. Lehti, M.S. and A. Sironen, *Formation and function of the manchette and flagellum during spermatogenesis.* Reproduction, 2016. **151**(4): p. R43-R54.
404. Shambharkar, P.B., et al., *TMEM203 is a novel regulator of intracellular calcium homeostasis and is required for spermatogenesis.* PLoS One, 2015. **10**(5): p. e0127480.
405. Ito, C., et al., *Failure to assemble the peri-nuclear structures in GOPC deficient spermatids as found in round-headed spermatozoa.* Archives of histology and cytology, 2004. **67**(4): p. 349-360.
406. Johnson, K.R., L.H. Gagnon, and B. Chang, *A hypomorphic mutation of the gamma-1 adaptin gene (Ap1g1) causes inner ear, retina, thyroid, and testes abnormalities in mice.* Mammalian genome, 2016. **27**(5-6): p. 200-212.
407. Saito, M., et al., *Targeted disruption of Ing2 results in defective spermatogenesis and development of soft-tissue sarcomas.* PLoS One, 2010. **5**(11): p. e15541.
408. Berkovits, B.D. and D.J. Wolgemuth, *The first bromodomain of the testis-specific double bromodomain protein Brdt is required for chromocenter organization that is modulated by genetic background.* Developmental biology, 2011. **360**(2): p. 358-368.
409. Yanagiya, A., et al., *The poly (A)-binding protein partner Paip2a controls translation during late spermiogenesis in mice.* The Journal of clinical investigation, 2010. **120**(9): p. 3389-3400.
410. Mykytyn, K., et al., *Bardet–Biedl syndrome type 4 (BBS4)-null mice implicate Bbs4 in flagella formation but not global cilia assembly.* Proceedings of the National Academy of Sciences of the United States of America, 2004. **101**(23): p. 8664-8669.
411. Colley, S.M., et al., *Loss of the nuclear receptor corepressor SLIRP compromises male fertility.* PLoS One, 2013. **8**(8): p. e70700.
412. Landfors, M., et al., *Sequencing of FTO and ALKBH5 in men undergoing infertility work-up identifies an infertility-associated variant and two missense mutations.* Fertility and sterility, 2016. **105**(5): p. 1170-1179. e5.
413. Kubo-Irie, M., et al., *Morphological abnormalities in the spermatozoa of fertile and infertile men.* Molecular reproduction and development, 2005. **70**(1): p. 70-81.
414. Milatiner, D., et al., *Associations between androgen receptor CAG repeat length and sperm morphology.* Human Reproduction, 2004. **19**(6): p. 1426-1430.
415. Bentson, L.F., et al., *New point mutation in Golga3 causes multiple defects in spermatogenesis.* Andrology, 2013. **1**(3): p. 440-450.
416. Zadravec, D., et al., *ELOVL2 controls the level of n-6 28: 5 and 30: 5 fatty acids in testis, a prerequisite for male fertility and sperm maturation in mice.* Journal of lipid research, 2011. **52**(2): p. 245-255.
417. Zadravec, D., et al., *Dominant negative effect on male fertility and sperm maturation by haploinsufficiency of ELOVL2 in mouse.* 2010.

418. Hockert, K.J., et al., *Spermatogenetic but not immunological defects in mice lacking the τCstF-64 polyadenylation protein.* Journal of reproductive immunology, 2011. **89**(1): p. 26-37.
419. Fu, G., et al., *Identification of candidate causal genes and their associated pathogenic mechanisms underlying teratozoospermia based on the spermatozoa transcript profiles.* Andrologia, 2016. **48**(5): p. 576-583.
420. Pawitan, Y., et al., *False discovery rate, sensitivity and sample size for microarray studies.* Bioinformatics, 2005. **21**(13): p. 3017-3024.
421. Hou, C.-C. and W.-X. Yang, *New insights to the ubiquitin–proteasome pathway (UPP) mechanism during spermatogenesis.* Molecular biology reports, 2013. **40**(4): p. 3213-3230.
422. Kierszenbaum, A., et al., *Temporal expression of cell cycle-related proteins during spermatogenesis: establishing a timeline for onset of the meiotic divisions.* Cytogenetic and genome research, 2003. **103**(3-4): p. 277-284.
423. Sakkas, D., et al., *Abnormal spermatozoa in the ejaculate: abortive apoptosis and faulty nuclear remodelling during spermatogenesis.* Reproductive biomedicine online, 2003. **7**(4): p. 428-432.
424. Teng, Y.-N., et al., *Histone gene expression profile during spermatogenesis.* Fertility and sterility, 2010. **93**(7): p. 2447-2449.
425. Sun, R. and H. Qi, *Dynamic expression of combinatorial replication-dependent histone variant genes during mouse spermatogenesis.* Gene Expression Patterns, 2014. **14**(1): p. 30-41.
426. Baker, K.E. and R. Parker, *Nonsense-mediated mRNA decay: terminating erroneous gene expression.* Current opinion in cell biology, 2004. **16**(3): p. 293-299.
427. Agarwal, A., E. Tvrda, and R. Sharma, *Relationship amongst teratozoospermia, seminal oxidative stress and male infertility.* Reproductive Biology and Endocrinology, 2014. **12**(1): p. 45.
428. Dehghanpour, F., et al., *Evaluation of sperm protamine deficiency and apoptosis in infertile men with idiopathic teratozoospermia.* Clinical and experimental reproductive medicine, 2017. **44**(2): p. 73-78.
429. Savadi-Shiraz, E., et al., *Quantification of sperm specific mRNA transcripts (PRM1, PRM2, and TNP2) in teratozoospermia and normozoospermia: New correlations between mRNA content and morphology of sperm.* Molecular reproduction and development, 2015. **82**(1): p. 26-35.
430. Shrivastava, V., et al., *SUMO proteins are involved in the stress response during spermatogenesis and are localized to DNA double-strand breaks in germ cells.* Reproduction, 2010. **139**(6): p. 999-1010.
431. Siomi, M.C., et al., *PIWI-interacting small RNAs: the vanguard of genome defence.* Nature reviews Molecular cell biology, 2011. **12**(4): p. 246-258.
432. Zhao, S., et al., *piRNA-triggered MIWI ubiquitination and removal by APC/C in late spermatogenesis.* Developmental cell, 2013. **24**(1): p. 13-25.
433. Huang, X.A., et al., *A major epigenetic programming mechanism guided by piRNAs.* Developmental cell, 2013. **24**(5): p. 502-516.
434. Peng, J.C. and H. Lin, *Beyond transposons: the epigenetic and somatic functions of the Piwi-piRNA mechanism.* Current opinion in cell biology, 2013. **25**(2): p. 190-194.
435. Kemenyova, P., et al., *Non-coding RNA in relation to autism spectrum disorder.* Non-coding RNA Investigation, 2017. **1**(4).
436. Eddy, E. *Regulation of gene expression during spermatogenesis*. in *Seminars in cell & developmental biology*. 1998: Elsevier.

437. Liu, D.Y., et al., *Comparison of the frequency of defective sperm–zona pellucida (ZP) binding and the ZP-induced acrosome reaction between subfertile men with normal and abnormal semen.* Human reproduction, 2007. **22**(7): p. 1878-1884.
438. Gonçalves, C., et al., *Y-chromosome microdeletions in nonobstructive azoospermia and severe oligozoospermia.* Asian journal of andrology, 2017. **19**(3): p. 338.
439. Jaiswal, D., et al., *Chromosome microarray analysis: a case report of infertile brothers with CATSPER gene deletion.* Gene, 2014. **542**(2): p. 263-265.
440. Dobashi, M., et al., *Distribution of type IV collagen subtypes in human testes and their association with spermatogenesis.* Fertility and sterility, 2003. **80**: p. 755-760.
441. He, Z., et al., *Expression of Col1a1, Col1a2 and procollagen I in germ cells of immature and adult mouse testis.* Reproduction, 2005. **130**(3): p. 333-341.
442. Chen, S.-H., D. Li, and C. Xu, *Downregulation of Col1a1 induces differentiation in mouse spermatogonia.* Asian journal of andrology, 2012. **14**(6): p. 842.
443. Feng, C.-W.A., et al., *SOX30 is required for male fertility in mice.* Scientific reports, 2017. **7**(1): p. 17619.
444. Sosnik, J., et al., *Tssk6 is required for Izumo relocalization and gamete fusion in the mouse.* Journal of cell science, 2009. **122**(15): p. 2741-2749.
445. Su, D., et al., *c. 822+ 126T> G/C: a novel triallelic polymorphism of the TSSK6 gene associated with spermatogenic impairment in a Chinese population.* Asian journal of andrology, 2010. **12**(2): p. 234.
446. Liu, J., et al., *Low levels of PRSS37 protein in sperm are associated with many cases of unexplained male infertility.* Acta biochimica et biophysica Sinica, 2016. **48**(11): p. 1058-1065.
447. Spiridonov, N.A., et al., *Identification and characterization of SSTK, a serine/threonine protein kinase essential for male fertility.* Molecular and cellular biology, 2005. **25**(10): p. 4250-4261.
448. Pastor, W.A., et al., *MORC1 represses transposable elements in the mouse male germline.* Nature communications, 2014. **5**: p. 5795.
449. Mercer, T.R. and J.S. Mattick, *Structure and function of long noncoding RNAs in epigenetic regulation.* Nature Structural and Molecular Biology, 2013. **20**(3): p. 300.
450. Ouimette, J.-F. and C. Rougeulle, *How Many Non-coding RNAs Does It Take to Compensate Male/Female Genetic Imbalance?*, in *Non-coding RNA and the Reproductive System*. 2016, Springer. p. 33-49.
451. Lee, S., et al., *Noncoding RNA NORAD regulates genomic stability by sequestering PUMILIO proteins.* Cell, 2016. **164**(1): p. 69-80.
452. Cooper, T.G., et al., *World Health Organization reference values for human semen characteristics.* Human reproduction update, 2010. **16**(3): p. 231-245.
453. Hammadeh, M., et al., *Predictive value of sperm chromatin condensation (aniline blue staining) in the assessment of male fertility.* Archives of andrology, 2001. **46**(2): p. 99-104.
454. Tsarev, I., et al., *Evaluation of male fertility potential by Toluidine Blue test for sperm chromatin structure assessment.* Human Reproduction, 2009. **24**(7): p. 1569-1574.
455. Oliva, R. and J.L. Ballescà, *Altered histone retention and epigenetic modifications in the sperm of infertile men.* Asian journal of andrology, 2012. **14**(2): p. 239.
456. Tunc, O. and K. Tremellen, *Oxidative DNA damage impairs global sperm DNA methylation in infertile men.* Journal of assisted reproduction and genetics, 2009. **26**(9-10): p. 537-544.
457. Aitken, R., *Oxidative stress and the etiology of male infertility.* Journal of assisted reproduction and genetics, 2016. **33**(12): p. 1691-1692.
458. Aitken, R. and G. De Iuliis, *On the possible origins of DNA damage in human spermatozoa.* MHR: Basic science of reproductive medicine, 2009. **16**(1): p. 3-13.

459. Enciso, M., et al., *A two-tailed Comet assay for assessing DNA damage in spermatozoa.* Reproductive biomedicine online, 2009. **18**(5): p. 609-616.
460. De Iuliis, G.N., et al., *DNA damage in human spermatozoa is highly correlated with the efficiency of chromatin remodeling and the formation of 8-hydroxy-2'-deoxyguanosine, a marker of oxidative stress.* Biology of reproduction, 2009. **81**(3): p. 517-524.
461. Evenson, D.P., *The Sperm Chromatin Structure Assay (SCSA®) and other sperm DNA fragmentation tests for evaluation of sperm nuclear DNA integrity as related to fertility.* Animal reproduction science, 2016. **169**: p. 56-75.
462. Turk, P.W., et al., *DNA adduct 8-hydroxyl-2'-deoxyguanosine (8-hydroxyguanine) affects function of human DNA methyltransferase.* Carcinogenesis, 1995. **16**(5): p. 1253-1255.
463. Weitzman, S.A., et al., *Free radical adducts induce alterations in DNA cytosine methylation.* Proceedings of the National Academy of Sciences, 1994. **91**(4): p. 1261-1264.
464. Valinluck, V., et al., *Oxidative damage to methyl-CpG sequences inhibits the binding of the methyl-CpG binding domain (MBD) of methyl-CpG binding protein 2 (MeCP2).* Nucleic acids research, 2004. **32**(14): p. 4100-4108.
465. Hashimshony, T., et al., *The role of DNA methylation in setting up chromatin structure during development.* Nature genetics, 2003. **34**(2): p. 187.
466. Tavalaee, M., S. Razavi, and M.H. Nasr-Esfahani, *Influence of sperm chromatin anomalies on assisted reproductive technology outcome.* Fertility and sterility, 2009. **91**(4): p. 1119-1126.
467. Montjean, D., et al., *Sperm global DNA methylation level: association with semen parameters and genome integrity.* Andrology, 2015. **3**(2): p. 235-240.
468. Attwood, J., R. Yung, and B. Richardson, *DNA methylation and the regulation of gene transcription.* Cellular and Molecular Life Sciences CMLS, 2002. **59**(2): p. 241-257.
469. Efimova, O.A., et al., *Genome-wide 5-hydroxymethylcytosine patterns in human spermatogenesis are associated with semen quality.* Oncotarget, 2017. **8**(51): p. 88294.
470. Wang, X.-X., et al., *Genome-wide 5-hydroxymethylcytosine modification pattern is a novel epigenetic feature of globozoospermia.* Oncotarget, 2015. **6**(9): p. 6535.
471. Klungland, A. and A.B. Robertson, *Oxidized C5-methyl cytosine bases in DNA: 5-hydroxymethylcytosine; 5-formylcytosine; and 5-carboxycytosine.* Free Radical Biology and Medicine, 2017. **107**: p. 62-68.
472. Madugundu, G.S., J. Cadet, and J.R. Wagner, *Hydroxyl-radical-induced oxidation of 5-methylcytosine in isolated and cellular DNA.* Nucleic acids research, 2014. **42**(11): p. 7450-7460.
473. Cadet, J. and J.R. Wagner, *Radiation-induced damage to cellular DNA: Chemical nature and mechanisms of lesion formation.* Radiation Physics and Chemistry, 2016. **128**: p. 54-59.
474. Chia, N., et al., *Hypothesis: environmental regulation of 5-hydroxymethylcytosine by oxidative stress.* Epigenetics, 2011. **6**(7): p. 853-856.
475. Efimova, O., et al., *Oxidized form of 5-methylcytosine—5-hydroxymethylcytosine: a new insight into the biological significance in the mammalian genome.* Russian Journal of Genetics: Applied Research, 2015. **5**(2): p. 75-81.
476. Jumeau, F., et al., *Defining the human sperm microtubulome: an integrated genomics approach.* Biology of Reproduction, 2016. **96**(1): p. 93-106.
477. Meseguer, M., et al., *Relationship between standard semen parameters, calcium, cholesterol contents, and mitochondrial activity in ejaculated spermatozoa from*

fertile and infertile males. Journal of assisted reproduction and genetics, 2004. **21**(12): p. 445-451.
478. Espino, J., et al., *Reduced levels of intracellular calcium releasing in spermatozoa from asthenozoospermic patients.* Reproductive Biology and Endocrinology, 2009. **7**(1): p. 11.
479. Cheng, P., et al., *Polymorphism in DNMT1 may modify the susceptibility to oligospermia.* Reproductive biomedicine online, 2014. **28**(5): p. 644-649.
480. Singh, K., et al., *Mutation C677T in the methylenetetrahydrofolate reductase gene is associated with male infertility in an Indian population.* International journal of andrology, 2005. **28**(2): p. 115-119.
481. Dada, R., et al., *Epigenetics and its role in male infertility.* Journal of assisted reproduction and genetics, 2012. **29**(3): p. 213-223.
482. Benchaib, M., et al., *Influence of global sperm DNA methylation on IVF results.* Human Reproduction, 2005. **20**(3): p. 768-773.
483. Lee, J.T., *Epigenetic regulation by long noncoding RNAs.* Science, 2012. **338**(6113): p. 1435-1439.

SUPPLEMENTARY MATERIALS

Table S1.
WHO 2010 spermatogenic semen parameters from Fertiles (F_s) and Oligozoospermic (O_z) sample groups

Test samples	**Concentration ($\times 10^6$/ml)**	**Motility (%)**	**Morphology (PIF)**
Normals			
F1	70	71	25
F2	97	72	28
F3	80	75	23
F4	62	72	19
F5	105	86	25
F6	100	90	30
F7	90	80	28
F8	67	54	17
F9	75	71	19
F10	110	85	21
F11	68	73	20
F12	70	62	15
F13	86	62	18
F14	54	55	14
F15	60	80	24
F16	80	60	16
F17	70	81	25
F18	95	80	23
F19	102	75	17
F20	85	68	22
Oligozoospermic			
O1	7	70	20
O2	9	44	12
O3	7	54	22
O4	5	60	15
O5	3	76	17
O6	6	66	14
O7	4	75	21
O8	5	60	15
O9	9	55	19
O10	8	50	16
O11	5	40	14
O12	6	67	18
O13	4	75	22
O14	7	43	12
O15	10	60	20
O16	4	63	17
O17	6	50	21
O18	10	42	16
O19	5	40	19
O20	7	43	12

WHO 2010 spermatogenic semen parameters from F_s and O_z sample groups. Sample from the F_s group were from subjects that ranged in age from 20 to 44 with a median age of 33 years, O_z age range was 19–41 with a median age of 29 years. All measures of appearance were within the normal range for both groups. Cell contamination was negligible. Motility was assessed manually and is the sum of rapid and slow movement. The percentage of ideal normal forms of spermatozoa in each sample was assessed using Wright-Giemsa stain. The threshold for normal forms is 4%. N_s individuals were from normally reproductive subjects having successfully fathered at least a single child.

Table S2.
WHO 2010 spermatogenic semen parameters from Fertiles (F_s) and Asthenozoospermic (A_z) sample groups

Test samples	**Concentration ($\times 10^6$/ml)**	**Motility (%)**	**Morphology (PIF)**
Fertiles			
F1	70	71	25
F2	97	72	28
F3	80	75	23
F4	62	72	19
F5	105	86	25
F6	100	90	30
F7	90	80	28
F8	67	54	17
F9	75	71	19
F10	110	85	21
F11	68	73	20
F12	70	62	15
F13	86	62	18
F14	54	55	14
F15	60	80	24
F16	80	60	16
F17	70	81	25
F18	95	80	23
F19	102	75	17
F20	85	68	22
Asthenozoospermic			
A1	17	12	18
A2	60	8	22
A3	22	13	25
A4	32	31	19
A5	51	29	35
A6	26	31	28
A7	20	25	16
A8	44	27	20
A9	47	26	22
A10	38	21	19
A11	31	16	27
A12	54	22	34
A13	40	13	21
A14	18	33	15
A15	26	19	29
A16	44	9	33
A17	21	24	18
A18	31	16	16
A19	25	28	22
A20	35	11	30

WHO 2010 spermatogenic semen parameters from F_s and A_z sample groups. Samples from the F_s group were from subjects that ranged in age from 20 to 44 with a median age of 33 years, A_z age range was 19–39 with a median age of 30 years. All measures of appearance were within the normal range for both groups.
Cell contamination was negligible. Motility was assessed manually and is the sum of rapid and slow movement. The percentage of ideal normal forms of spermatozoa in each sample was assessed using Wright-Giemsa stain.

The threshold for normal forms is 4%, and for normal motility is 40%. F_s individuals were from normally reproductive subjects having successfully fathered at least a single child.

Table S3.

WHO 2010 spermatogenic semen parameters from Fertiles (F_s) and Teratozoospermic (T_z) sample groups

`	**Concentration ($\times 10^6$/ml)**	**Motility (%)**	**Morphology (PIF)**
Fertiles			
F1	70	71	25
F2	97	72	28
F3	80	75	23
F4	62	72	19
F5	105	86	25
F6	100	90	30
F7	90	80	28
F8	67	54	17
F9	75	71	19
F10	110	85	21
F11	68	73	20
F12	70	62	15
F13	86	62	18
F14	54	55	14
F15	60	80	24
F16	80	60	16
F17	70	81	25
F18	95	80	23
F19	102	75	17
F20	85	68	22
Teratozoospermic			
T1	22	41	3
T2	34	57	1
T3	23	44	0
T4	17	44	2
T5	19	50	4
T6	21	45	3
T7	66	67	1
T8	81	62	0
T9	27	46	2
T10	18	53	1
T11	38	67	2
T12	47	58	1
T13	72	43	0
T14	26	82	4
T15	26	55	2
T16	43	48	1
T17	86	62	0
T18	53	67	1
T19	17	41	3
T20	28	83	1

WHO 2010 spermatogenic semen parameters from F_s and T_z sample groups. Sample from the F_s group were from subjects that ranged in age from 20 to 44 with a median age of 33 years, T_z age range was 24–39 with a median age of 31 years. All measures of appearance were within the normal range for both groups. Cell contamination was negligible. Motility was assessed manually and is the sum of rapid and slow movement. The percentage of ideal normal forms of spermatozoa in each sample was assessed using Wright-Giemsa stain. The threshold for normal forms is 4%. F_s individuals were from normally reproductive subjects having successfully fathered at least a single child.

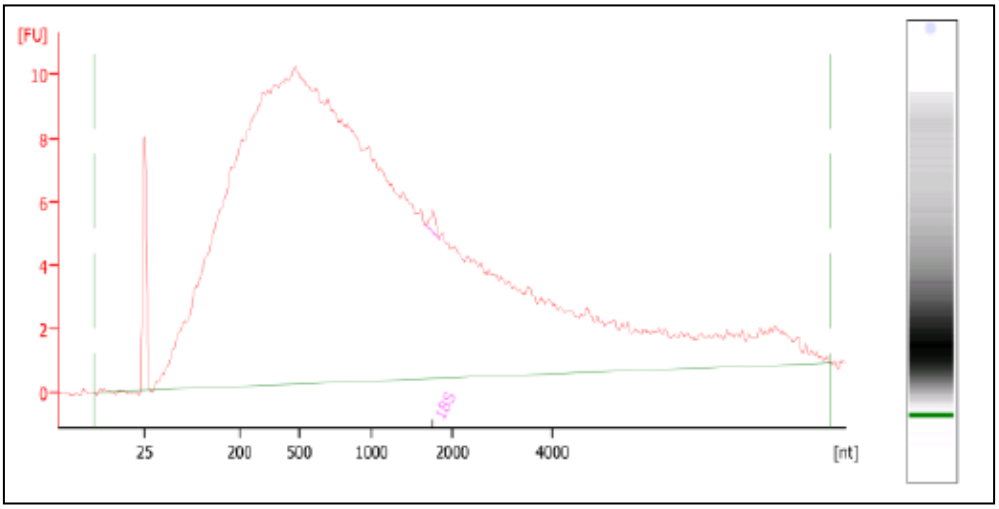


Figure S1. Example of electropherogram profile of total RNA isolated from human ejaculated sperm, as determined using an Agilent Bioanalyzer 2100 showing absence of rRNA.

Table S4. The spermatozoal enriched transcripts from oligozoospermic men associated with the male infertility mouse phenotype obtained from functional annotation analysis (Toppgene) based on their abundancy.

	Gene Symbol	Gene Name
1	TSSK6	testis specific serine kinase 6
2	ESR1	estrogen receptor 1
3	ESR2	estrogen receptor 2
4	FMN1	formin 1
5	C14orf39	chromosome 14 open reading frame 39
6	CATSPER1	cation channel sperm associated 1
7	GALNTL5	polypeptide N-acetylgalactosaminyltransferase like 5
8	CEP131	centrosomal protein 131
9	ATP1A4	ATPase Na+/K+ transporting subunit alpha 4
10	BOLL	boule homolog, RNA binding protein
11	SMC1B	structural maintenance of chromosomes 1B
12	MORC1	MORC family CW-type zinc finger 1
13	MYO7A	myosin VIIA
14	TEKT2	tektin 2
15	BRCA1	BRCA1, DNA repair associated
16	BRDT	bromodomain testis associated
17	BSG	basigin (Ok blood group)
18	GLI3	GLI family zinc finger 3
19	VPS13A	vacuolar protein sorting 13 homolog A
20	CCNA1	cyclin A1
21	GRID2	glutamate ionotropic receptor delta type subunit 2
22	TEX15	testis expressed 15, meiosis and synapsis associated
23	TEX14	testis expressed 14, intercellular bridge forming factor
24	MEI1	meiotic double-stranded break formation protein 1
25	RNF17	ring finger protein 17
26	TDRD1	tudor domain containing 1
27	TDRD7	tudor domain containing 7
28	NEURL1	neuralized E3 ubiquitin protein ligase 1
29	MAEL	maelstrom spermatogenic transposon silencer
30	SERPINA5	serpin family A member 5
31	CALR3	calreticulin 3
32	NUMBL	NUMB like, endocytic adaptor protein
33	PRSS37	protease, serine 37
34	CCDC159	coiled-coil domain containing 159
35	CFAP54	cilia and flagella associated protein 54
36	DDX4	DEAD-box helicase 4
37	DNAH1	dynein axonemal heavy chain 1
38	RIMBP3	RIMS binding protein 3
39	MEIOC	meiosis specific with coiled-coil domain
40	CHD5	chromodomain helicase DNA binding protein 5
41	RELN	Reelin
42	COLCA2	colorectal cancer associated 2
43	RAD21L1	RAD21 cohesin complex component like 1
44	PTCH1	patched 1
45	AK7	adenylate kinase 7
46	IZUMO1	izumo sperm-egg fusion 1
47	SGO2	shugoshin 2
48	REC8	REC8 meiotic recombination protein
49	RARA	retinoic acid receptor alpha
50	RBM5	RNA binding motif protein 5

Table S5. The spermatozoal enriched transcripts from oligozoospermic men associated with epigenetic regulation of gene expression obtained from functional annotation analysis (Toppgene) of the up-regulated chromatin modifying enzymes.

	Gene Symbol	Gene Name
1	MTA2	metastasis associated 1 family member 2
2	HIST1H2BJ	histone cluster 1 H2B family member j
3	H2AFV	H2A histone family member V
4	GATAD2A	GATA zinc finger domain containing 2A
5	RBBP4	RB binding protein 4, chromatin remodeling factor
6	RBBP7	RB binding protein 7, chromatin remodeling factor
7	SAP18	Sin3A associated protein 18
8	ACTB	actin beta
9	SAP30L	SAP30 like
10	CHD3	chromodomain helicase DNA binding protein 3
11	KAT2A	lysine acetyltransferase 2A
12	HIST1H4I	histone cluster 1 H4 family member i
13	SAP30	Sin3A associated protein 30
14	HIST1H2AC	histone cluster 1 H2A family member c
15	HIST2H2AC	histone cluster 2 H2A family member c
16	KAT2B	lysine acetyltransferase 2B
17	HIST1H2BG	histone cluster 1 H2B family member g
18	HIST1H2BF	histone cluster 1 H2B family member f
19	MTA1	metastasis associated 1
20	HIST1H2BC	histone cluster 1 H2B family member c
21	HIST2H2BE	histone cluster 2 H2B family member e
22	MTA3	metastasis associated 1 family member 3
23	HIST4H4	histone cluster 4 H4
24	HIST1H3E	histone cluster 1 H3 family member e
25	HIST1H4D	histone cluster 1 H4 family member d
26	SUDS3	SDS3 homolog, SIN3A corepressor complex component
27	HIST1H4C	histone cluster 1 H4 family member c
28	HIST1H4H	histone cluster 1 H4 family member h
29	HIST1H4B	histone cluster 1 H4 family member b
30	HIST1H4E	histone cluster 1 H4 family member e
31	AEBP2	AE binding protein 2
32	HIST1H2AE	histone cluster 1 H2A family member e
33	H2AFX	H2A histone family member X
34	H2AFZ	H2A histone family member Z
35	HIST1H2BD	histone cluster 1 H2B family member d
36	H2AFJ	H2A histone family member J
37	SUZ12	SUZ12 polycomb repressive complex 2 subunit
38	SAP130	Sin3A associated protein 130
39	HIST1H2BK	histone cluster 1 H2B family member k
40	HDAC1	histone deacetylase 1
41	HDAC2	histone deacetylase 2

Table S6. The spermatozoal enriched down-regulated transcripts from asthenozoospermic men associated with histone modification obtained from functional annotation analysis (Toppgene) based on their abundancy.

	Gene Symbol	Gene Name
1	KDM5C	lysine demethylase 5C
2	ACTL6A	actin like 6A
3	KDM5D	lysine demethylase 5D
4	TRRAP	transformation/transcription domain associated protein
5	BAP1	BRCA1 associated protein 1
6	OGT	O-linked N-acetylglucosamine (GlcNAc) transferase
7	C6orf89	chromosome 6 open reading frame 89
8	ARRB1	arrestin beta 1
9	KDM1B	lysine demethylase 1B
10	TBL1Y	transducin beta like 1 Y-linked
11	BCL6	B cell CLL/lymphoma 6
12	HDAC3	histone deacetylase 3
13	KAT2B	lysine acetyltransferase 2B
14	PYGO2	pygopus family PHD finger 2
15	LDB1	LIM domain binding 1
16	CAMK2D	calcium/calmodulin dependent protein kinase II delta
17	RNF8	ring finger protein 8
18	MTA2	metastasis associated 1 family member 2
19	CHD4	chromodomain helicase DNA binding protein 4
20	GTF3C4	general transcription factor IIIC subunit 4
21	BRMS1	breast cancer metastasis suppressor 1
22	ATG5	autophagy related 5
23	CLOCK	clock circadian regulator
24	NCOR1	nuclear receptor corepressor 1
25	MORF4L2	mortality factor 4 like 2
26	CTR9	CTR9 homolog, Paf1/RNA polymerase II complex component
27	CHD5	chromodomain helicase DNA binding protein 5
28	CTBP1	C-terminal binding protein 1
29	CTNNB1	catenin beta 1
30	SLK	STE20 like kinase
31	HDAC4	histone deacetylase 4
32	PHF19	PHD finger protein 19
33	NOC2L	NOC2 like nucleolar associated transcriptional repressor
34	RNF40	ring finger protein 40
35	DDB1	damage specific DNA binding protein 1
36	SUPT7L	SPT7 like, STAGA complex gamma subunit
37	USP15	ubiquitin specific peptidase 15
38	DR1	down-regulator of transcription 1
39	HDAC6	histone deacetylase 6
40	EP300	E1A binding protein p300
41	SAP18	Sin3A associated protein 18
42	EYA3	EYA transcriptional coactivator and phosphatase 3
43	HMG20B	high mobility group 20B
44	PRMT5	protein arginine methyltransferase 5
45	KAT8	lysine acetyltransferase 8
46	ARID5B	AT-rich interaction domain 5B
47	TADA3	transcriptional adaptor 3
48	CARM1	coactivator associated arginine methyltransferase 1
49	NCOA2	nuclear receptor coactivator 2
50	FMR1	fragile X mental retardation 1
51	USP16	ubiquitin specific peptidase 16
52	USP21	ubiquitin specific peptidase 21

Cont. Table S6. The spermatozoal enriched down-regulated transcripts from asthenozoospermic men associated with histone modification obtained from functional annotation analysis (Toppgene) based on their abundancy.

	Gene Symbol	Gene Name
53	TADA1	transcriptional adaptor 1
54	TAF5L	TATA-box binding protein associated factor 5 like
55	BRPF3	bromodomain and PHD finger containing 3
56	GATA2	GATA binding protein 2
57	KDM3B	lysine demethylase 3B
58	RUVBL2	RuvB like AAA ATPase 2
59	MSL3	MSL complex subunit 3
60	PCGF1	polycomb group ring finger 1
61	KMT5A	lysine methyltransferase 5A
62	WDR5	WD repeat domain 5
63	HIST1H1C	histone cluster 1 H1 family member c
64	HIST1H1E	histone cluster 1 H1 family member e
65	HCFC1	host cell factor C1
66	HDAC1	histone deacetylase 1
67	HLCS	holocarboxylase synthetase
68	PRMT2	protein arginine methyltransferase 2
69	PRMT1	protein arginine methyltransferase 1
70	TET3	tet methylcytosine dioxygenase 3
71	SMAD4	SMAD family member 4
72	DTX3L	deltex E3 ubiquitin ligase 3L
73	MECP2	methyl-CpG binding protein 2
74	MBD3	methyl-CpG binding domain protein 3
75	BRD7	bromodomain containing 7
76	FLCN	Folliculin
77	NOS1	nitric oxide synthase 1
78	POLE3	DNA polymerase epsilon 3, accessory subunit
79	PHB	Prohibitin
80	PHF2	PHD finger protein 2
81	BRCC3	BRCA1/BRCA2-containing complex subunit 3
82	PAF1	PAF1 homolog, Paf1/RNA polymerase II complex component
83	PRKAA2	protein kinase AMP-activated catalytic subunit alpha 2
84	PKN1	protein kinase N1
85	MAPK3	mitogen-activated protein kinase 3
86	TET2	tet methylcytosine dioxygenase 2
87	BCOR	BCL6 corepressor
88	CDC73	cell division cycle 73
89	PIH1D1	PIH1 domain containing 1
90	RAG1	recombination activating 1
91	KDM5A	lysine demethylase 5A
92	RBBP4	RB binding protein 4, chromatin remodeling factor
93	RBBP5	RB binding protein 5, histone lysine methyltransferase complex subunit
94	RBBP7	RB binding protein 7, chromatin remodeling factor
95	SAP30L	SAP30 like
96	RIOX1	ribosomal oxygenase 1
97	REST	RE1 silencing transcription factor
98	TBL1XR1	transducin beta like 1 X-linked receptor 1
99	MSL2	MSL complex subunit 2
100	RLF	rearranged L-myc fusion
101	PRMT6	protein arginine methyltransferase 6
102	RNF2	ring finger protein 2
103	COPRS	coordinator of PRMT5 and differentiation stimulator
104	JADE1	jade family PHD finger 1
105	CXXC1	CXXC finger protein 1

Cont. Table S6. The spermatozoal enriched down-regulated transcripts from asthenozoospermic men associated with histone modification obtained from functional annotation analysis (Toppgene) based on their abundancy.

	Gene Symbol	Gene Name
106	SATB1	SATB homeobox 1
107	SGF29	SAGA complex associated factor 29
108	SET	SET nuclear proto-oncogene
109	SETMAR	SET domain and mariner transposase fusion gene
110	SFPQ	splicing factor proline and glutamine rich
111	OTUB1	OTU deubiquitinase, ubiquitin aldehyde binding 1
112	NAA50	N(alpha)-acetyltransferase 50, NatE catalytic subunit
113	SKI	SKI proto-oncogene
114	SNW1	SNW domain containing 1
115	ZNF335	zinc finger protein 335
116	TET1	tet methylcytosine dioxygenase 1
117	EPC1	enhancer of polycomb homolog 1
118	SNAI2	snail family transcriptional repressor 2
119	WDR61	WD repeat domain 61
120	KDM1A	lysine demethylase 1A
121	PHF8	PHD finger protein 8
122	DMAP1	DNA methyltransferase 1 associated protein 1
123	RTF1	RTF1 homolog, Paf1/RNA polymerase II complex component
124	SUPT6H	SPT6 homolog, histone chaperone
125	TAF7	TATA-box binding protein associated factor 7
126	TAF10	TATA-box binding protein associated factor 10
127	TAF12	TATA-box binding protein associated factor 12
128	MAP3K7	mitogen-activated protein kinase kinase kinase 7
129	TBL1X	transducin beta like 1 X-linked
130	TCF3	transcription factor 3
131	USP22	ubiquitin specific peptidase 22
132	JADE2	jade family PHD finger 2
133	SIRT1	sirtuin 1
134	NACC2	NACC family member 2
135	SUDS3	SDS3 homolog, SIN3A corepressor complex component
136	SETD7	SET domain containing lysine methyltransferase 7
137	TP53	tumor protein p53
138	WBP2	WW domain binding protein 2
139	UBE2A	ubiquitin conjugating enzyme E2 A
140	UBE2E1	ubiquitin conjugating enzyme E2 E1
141	EID1	EP300 interacting inhibitor of differentiation 1
142	BRD1	bromodomain containing 1
143	UTY	ubiquitously transcribed tetratricopeptide repeat containing, Y-linked
144	MEAF6	MYST/Esa1 associated factor 6
145	NELFA	negative elongation factor complex member A
146	SMYD2	SET and MYND domain containing 2
147	ATXN7L3	ataxin 7 like 3
148	PRDM8	PR/SET domain 8
149	KAT6A	lysine acetyltransferase 6A
150	ZNF304	zinc finger protein 304

Table S7. List of long non-coding RNAs associated with asthenozoospermia derived from the antisense RNAs group.

	Gene Symbol	Gene Name
1	TUSC7	tumor suppressor candidate 7 (non-protein coding)
2	AATBC	apoptosis associated transcript in bladder cancer
3	B4GALT1-AS1	B4GALT1 antisense RNA 1
4	PCAT6	prostate cancer associated transcript 6 (non-protein coding)
5	RPS6KA2-AS1	RPS6KA2 antisense RNA 1
6	NRAV	negative regulator of antiviral response (non-protein coding)
7	PCA3	prostate cancer associated 3 (non-protein coding)
8	OIP5-AS1	OIP5 antisense RNA 1

Table S8. List of long non-coding RNAs associated with asthenozoospermia derived from the long intergenic non-coding RNAs (lincRNAs) group.

	Gene Symbol	Gene Name
1	PCGEM1	PCGEM1, prostate-specific transcript (non-protein coding)
2	PART1	prostate androgen-regulated transcript 1 (non-protein coding)
3	PCAT4	prostate cancer associated transcript 4 (non-protein coding)
4	FTX	FTX transcript, XIST regulator (non-protein coding)
5	NORAD	non-coding RNA activated by DNA damage
6	DRAIC	downregulated RNA in cancer, inhibitor of cell invasion and migration
7	PWRN2	Prader-Willi region non-protein coding RNA 2
8	UCA1	urothelial cancer associated 1 (non-protein coding)
9	DANCR	differentiation antagonizing non-protein coding RNA
10	MIR100HG	mir-100-let-7a-2-mir-125b-1 cluster host gene
11	MEG3	maternally expressed 3 (non-protein coding)

Table S9. List of long non-coding RNAs associated with asthenozoospermia derived from the long intergenic non-coding RNAs (lincRNAs) group.

	Gene Symbol	Gene Name
1	PCGEM1	PCGEM1, prostate-specific transcript (non-protein coding)
2	PART1	prostate androgen-regulated transcript 1 (non-protein coding)
3	PCAT4	prostate cancer associated transcript 4 (non-protein coding)
4	FTX	FTX transcript, XIST regulator (non-protein coding)
5	NORAD	non-coding RNA activated by DNA damage
6	DRAIC	downregulated RNA in cancer, inhibitor of cell invasion and migration
7	PWRN2	Prader-Willi region non-protein coding RNA 2
8	UCA1	urothelial cancer associated 1 (non-protein coding)
9	DANCR	differentiation antagonizing non-protein coding RNA
10	MIR100HG	mir-100-let-7a-2-mir-125b-1 cluster host gene
11	MEG3	maternally expressed 3 (non-protein coding)

Table S10. The spermatozoal enriched up-regulated transcripts from teratozoospermic men associated with teratozoospermia mouse phenotype obtained from functional annotation analysis (Toppgene) based on their abundancy.

	Gene Name	Original Symbol
1	NPHP1	nephrocystin 1
2	ADCY10	adenylate cyclase 10
3	ADAD1	adenosine deaminase domain containing 1
4	TDRD5	tudor domain containing 5
5	GTSF1	gametocyte specific factor 1
6	TSSK6	testis specific serine kinase 6
7	BEST1	bestrophin 1
8	ZFP42	ZFP42 zinc finger protein
9	SETX	Senataxin
10	DPY19L2	dpy-19 like 2
11	SPATA6	spermatogenesis associated 6
12	MNS1	meiosis specific nuclear structural 1
13	CATSPERD	cation channel sperm associated auxiliary subunit delta
14	NUP210L	nucleoporin 210 like
15	CAMK4	calcium/calmodulin dependent protein kinase IV
16	ZPBP	zona pellucida binding protein
17	CYP17A1	cytochrome P450 family 17 subfamily A member 1
18	CYP19A1	cytochrome P450 family 19 subfamily A member 1
19	ADAM7	ADAM metallopeptidase domain 7
20	SPEF2	sperm flagellar 2
21	TTLL5	tubulin tyrosine ligase like 5
22	ICA1	islet cell autoantigen 1
23	GPX4	glutathione peroxidase 4
24	RANBP9	RAN binding protein 9
25	GRID2	glutamate ionotropic receptor delta type subunit 2
26	CATSPER3	cation channel sperm associated 3
27	ODF1	outer dense fiber of sperm tails 1
28	CNBD2	cyclic nucleotide binding domain containing 2
29	KMT2E	lysine methyltransferase 2E
30	SPAG6	sperm associated antigen 6
31	CLIP1	CAP-Gly domain containing linker protein 1
32	CREM	cAMP responsive element modulator
33	NSUN7	NOP2/Sun RNA methyltransferase family member 7
34	TGFB2	transforming growth factor beta 2
35	ENO4	enolase family member 4
36	MFSD14A	major facilitator superfamily domain containing 14A
37	M1AP	meiosis 1 associated protein
38	SLC26A8	solute carrier family 26 member 8
39	GORASP2	golgi reassembly stacking protein 2
40	AKAP4	A-kinase anchoring protein 4
41	CFAP54	cilia and flagella associated protein 54
42	CADM1	cell adhesion molecule 1
43	PSME4	proteasome activator subunit 4
44	PPP3R2	protein phosphatase 3 regulatory subunit B, beta
45	BAZ1A	bromodomain adjacent to zinc finger domain 1A
46	BBS7	Bardet-Biedl syndrome 7
47	CSNK2A2	casein kinase 2 alpha 2

Table S10. The spermatozoal enriched up-regulated transcripts from teratozoospermic men associated with teratozoospermia mouse phenotype obtained from functional annotation analysis (Toppgene) based on their abundancy.

	Gene Name	Original Symbol
48	DDHD1	DDHD domain containing 1
49	PAFAH1B1	platelet activating factor acetylhydrolase 1b regulatory subunit 1
50	EHD1	EH domain containing 1
51	FSHR	follicle stimulating hormone receptor
52	KIAA1324	KIAA1324
53	ASPM	abnormal spindle microtubule assembly
54	AGFG1	ArfGAP with FG repeats 1
55	GALNTL5	polypeptide N-acetylgalactosaminyltransferase like 5
56	TMF1	TATA element modulatory factor 1
57	PIP5K1B	phosphatidylinositol-4-phosphate 5-kinase type 1 beta
58	OSBP2	oxysterol binding protein 2
59	TAF4B	TATA-box binding protein associated factor 4b
60	SPAG16	sperm associated antigen 16
61	QKI	QKI, KH domain containing RNA binding
62	OAZ3	ornithine decarboxylase antizyme 3
63	HERC4	HECT and RLD domain containing E3 ubiquitin protein ligase 4
64	ICA1L	islet cell autoantigen 1 like
65	SERPINA5	serpin family A member 5
66	FHL5	four and a half LIM domains 5
67	PRNP	prion protein
68	AGTPBP1	ATP/GTP binding protein 1
69	PICK1	protein interacting with PRKCA 1
70	IQCG	IQ motif containing G
71	BRWD1	bromodomain and WD repeat domain containing 1

Table S11. List of long non-coding RNAs associated with teratozoospermia derived from the long intergenic non-coding RNAs (lincRNAs) group.

	Gene Symbol	**Gene Name**
1	PCGEM1	PCGEM1, prostate-specific transcript (non-protein coding)
2	GAPLINC	gastric adenocarcinoma associated, positive CD44 regulator, long intergenic non-coding RNA
3	PCAT18	prostate cancer associated transcript 18 (non-protein coding)
4	EWSAT1	Ewing sarcoma associated transcript 1
5	LINC01191	long intergenic non-protein coding RNA 1191
6	PWRN1	Prader-Willi region non-protein coding RNA 1
7	CISTR	chondrogenesis-associated transcript
8	SOX21-AS1	SOX21 antisense divergent transcript 1
9	LINC00161	long intergenic non-protein coding RNA 161
10	PCAT4	prostate cancer associated transcript 4 (non-protein coding)
11	FTX	FTX transcript, XIST regulator (non-protein coding)
12	OLMALINC	oligodendrocyte maturation-associated long intergenic non-coding RNA
13	PCAT14	prostate cancer associated transcript 14 (non-protein coding)
14	MEG8	maternally expressed 8, small nucleolar RNA host gene
15	PART1	prostate androgen-regulated transcript 1 (non-protein coding)
16	TUSC8	tumor suppressor candidate 8 (non-protein coding)
17	LINC-ROR	long intergenic non-protein coding RNA, regulator of reprogramming
18	NORAD	non-coding RNA activated by DNA damage
19	GAS1RR	GAS1 adjacent regulatory RNA
20	SNHG18	small nucleolar RNA host gene 18
21	SNHG8	small nucleolar RNA host gene 8
22	CRNDE	colorectal neoplasia differentially expressed
23	SCHLAP1	SWI/SNF complex antagonist associated with prostate cancer 1 (non-protein coding)
24	BLACAT1	bladder cancer associated transcript 1 (non-protein coding)
25	DRAIC	downregulated RNA in cancer, inhibitor of cell invasion and migration
26	MIR155HG	MIR155 host gene
27	PICSAR	P38 inhibited cutaneous squamous cell carcinoma associated lincRNA
28	TP53TG1	TP53 target 1 (non-protein coding)
29	NEAT1	nuclear paraspeckle assembly transcript 1 (non-protein coding)

Table S12. List of long non-coding RNAs associated with teratozoospermia derived from the antisense RNAs group.

	Gene Symbol	Gene Name
1	RGMB-AS1	RGMB antisense RNA 1
2	NNT-AS1	NNT antisense RNA 1
3	HOTTIP	HOXA distal transcript antisense RNA
4	TH2LCRR	T helper type 2 locus control region associated RNA
5	BACE1-AS	BACE1 antisense RNA
6	HOXA-AS3	HOXA cluster antisense RNA 3
7	RPS6KA2-AS1	RPS6KA2 antisense RNA 1
8	ZFAS1	ZNFX1 antisense RNA 1
9	GNG12-AS1	GNG12 antisense RNA 1
10	CASC8	cancer susceptibility 8 (non-protein coding)
11	HAGLR	HOXD antisense growth-associated long non-coding RNA
12	DLEU2	deleted in lymphocytic leukemia 2 (non-protein coding)
13	NRIR	negative regulator of interferon response (non-protein coding)
14	NKILA	NF-kappaB interacting lncRNA
15	DHRS4-AS1	DHRS4 antisense RNA 1
16	BIRC6-AS2	BIRC6 antisense RNA 2
17	FOXP4-AS1	FOXP4 antisense RNA 1
18	B4GALT1-AS1	B4GALT1 antisense RNA 1
19	PCAT6	prostate cancer associated transcript 6 (non-protein coding)
20	HOXA11-AS	HOXA11 antisense RNA
21	EMX2OS	EMX2 opposite strand/antisense RNA
22	CDKN2B-AS1	CDKN2B antisense RNA 1
23	HOXA-AS2	HOXA cluster antisense RNA 2

Middle East Fertility Society Journal (2016) 21, 269–276

Middle East Fertility Society

Middle East Fertility Society Journal

www.mefsjournal.org
www.sciencedirect.com

ORIGINAL ARTICLE

High level of DNA fragmentation in sperm of Lebanese infertile men using Sperm Chromatin Dispersion test

Fadi B. Choucair [a], Eliane G. Rachkidi [a], Georges C. Raad [a], Elias M. Saliba [a], Nina S. Zeidan [b], Rania A. Jounblat [a], Imad F. Abou Jaoude [c], Mira M. Hazzouri [a,*]

[a] Lebanese University, Faculty of Sciences 2, Fanar, Lebanon
[b] Lebanese University, Faculty of Public Health, Fanar, Lebanon
[c] Abou Jaoude Hospital, Department of Obstetrics and Gynecology, Lebanon

Received 4 June 2016; accepted 15 June 2016
Available online 2 July 2016

KEYWORDS

Sperm DNA fragmentation;
Sperm chromatin dispersion test;
Male infertility;
Nuclear quality

Abstract Assessment of male fertility has been based on routine semen analysis established by the World Health Organization (WHO), evaluating sperm concentration, motility and morphology. Actually, these parameters become less reliable markers to evaluate male fertility potential. Therefore, a search for better markers has led to an increased focus on sperm chromatin integrity testing in fertility work-up and assisted reproductive technologies (ART). In this study, we evaluate sperm DNA fragmentation in 185 Lebanese infertile patients attending fertility clinics all over the country, in comparison with 30 control men of proven fertility, using the sperm chromatin disper-sion test (SCD).

Our results showed a significantly higher sperm DNA fragmentation in infertile group compared to control group (20.62% vs 12.96%; $p < 0.001$). In addition, we found that sperm DNA fragmen-tation is correlated with alterations in sperm parameters: count, motility and morphology. Moreover, we showed that 28% of normozoospermic patients have high sperm DNA fragmenta-tion. Also, a positive correlation was found between sperm DNA fragmentation and ART failure in infertile group patients. Finally, sperm DNA fragmentation is suggested to be associated with tobacco and environmental conditions.

2016 Middle East Fertility Society. Production and hosting by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

Infertility is recognized today as a public health issue. Approximately 9% of couples worldwide suffer from fertility problems, and 45% of them are due to the male partner (1). Male fertility evaluation is based primarily on routine semen analysis, assessing sperm count, motility and morphology

* Corresponding author.
E-mail address: hazzourim@yahoo.com (M.M. Hazzouri).

Peer review under responsibility of Middle East Fertility Society.

Production and hosting by Elsevier

http://dx.doi.org/10.1016/j.mefs.2016.06.005
1110-5690 2016 Middle East Fertility Society. Production and hosting by Elsevier B.V.
This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

(2). However, 15% of patients with male factor infertility showed a normal semen analysis (3), thus making it insufficient to predict male fertility potential. Consequently, science is shifting toward analyzing the molecular aspects of spermatozoa.

The status of sperm chromatin has a great importance, specifically in intracytoplasmic sperm injection (ICSI), where natural sperm selection barriers are overridden and a single sperm cell is selected based on its morphology, regardless of its nuclear quality (4). Several studies have shown that male fertility could be affected by sperm DNA damage; therefore, an assessment of sperm DNA content may be a useful adjunct to the routine semen analysis for a better male fertility predic-tion (5–7). Given the evidence supporting the presence of highly sperm DNA damage in infertile men compared to fertile ones, it is consequently postulated that fertility potential of infertile men is negatively affected (8–16).

While high levels of sperm DNA damage often correlate with poor semen parameters such as reduced count and motil-ity or abnormal morphology (14,17–20), it is also found in infertile men with normal semen parameters (21–23).

In addition to the association between sperm DNA damage and male infertility, sperm DNA fragmentation was also cor-related with high sexual chromosomes aneuploidy risk in sper-matozoa of infertile men (24–27). High sperm DNA fragmentation was also linked to a low fertilization rate and an altered embryo quality (13,20,28–36). Moreover, previous studies have shown that a paternal effect can be responsible for repeated failure of ART attempts (37–39) and some evi-dences suggest that abnormal integrity of sperm DNA possibly increases miscarriages (40–44).

Lebanon has been recognized for his contribution to infer-tility treatments within the Middle East region but the nuclear sperm quality is still ignored by most of the Lebanese clini-cians. The present study aims to explore, for the first time in Lebanon, the nuclear quality of spermatozoa in Lebanese infertile men, using the sperm chromatin dispersion test (SCD). A correlative relation between sperm DNA fragmenta-tion, male infertility, semen parameters, lifestyle factors and ART failures will also be investigated. Moreover, the study will match the introduction of a new technique into routine semen analysis for male infertility diagnosis and ART outcome prediction.

2. Materials and methods

2.1. Patients

This study included 30 fertile men (control group) and 185 men presenting to different infertility clinics (infertile group) from different regions in Lebanon between 2010 and 2013. A detailed medical history was obtained from the participants, including reproductive history and fertility evaluation, mainly semen analysis, ART history and lifestyle factors (smoking, professional risk of exposure to heat stress or toxicants, antioxidant treatment). We included in the study only males whose partners are clinically considered fertile, which have normal reproductive history after gynecological exam and endocrinological testings. It is interesting to mention that the two groups were matched at the same age range; infertile group: 30–45 years old, control group: 30–50 years old.

The ethics committee of the Lebanese University approved the study protocol.

2.2. Semen analysis

The semen samples of the two groups were collected by mas-turbation into a sterile container after 2–5 days of sexual absti-nence. After complete liquefaction of the sample, semen analysis was performed according to World Health Organization guidelines (2).

2.3. Sperm Chromatin dispersion (SCD) test

The SCD test is based on the principle that when sperm immersed in an agarose matrix on a slide, treated with an acid solution to denaturate DNA, and then lysed with a buffer to remove membranes and proteins, the result is the formation of nucleoids with a central core and a peripheral halo of dis-persed DNA loops. The dispersion halos that are observed in sperm nuclei with non-fragmented DNA after the removal of nuclear proteins are either minimally present or not pro-duced at all in sperm nuclei with fragmented DNA. The agar-ose matrix allows working with unfixed sperm on a slide in a suspension-like environment. Nucleoids are visualized either with fluorescent microscopy after staining with a DNA specific fluorochrome (DAPI), or with bright-field microscopy with Wright's staining solution.

SCD test protocol:

In this study, we performed the SCD test as proposed and improved by Fernandez et al., in 2005 (45), using the Halosperm kit (INDAS laboratories, Madrid, Spain).

In brief, an aliquot of a semen sample was diluted to 10 mil-lion/mL in phosphate-buffered saline (PBS). Gelled aliquots of low-melting point agarose in eppendorf tubes were provided in the kit, each one to process a semen sample. Eppendorf tubes were placed in a water bath at 90–100 LC for 5 min to fuse the agarose, and then in a water bath at 37 LC. After 5 min of incu-bation for temperature equilibration at 37 LC, 60 mL of the diluted semen sample was added to the eppendorf tube and mixed with the fused agarose. Of the semen–agarose mix,

20 LL was pipetted onto slides pre-coated with agarose, pro-vided in the kit, and covered with a 22 22 mm coverslip. The slides were placed on a cold plate in the refrigerator (4 LC) for 5 min to allow the agarose to produce a microgel with the sperm cells embedded within. The coverslips were gently removed and the slides immediately immersed horizon-tally in an acid solution, previously prepared by mixing 80 LL of HCl from an eppendorf tube in the kit with 10 mL of dis-tilled water, and incubated for 7 min. The slides were horizon-tally immersed in 10 mL of the lysing solution for 25 min. After washing 5 min in a tray with abundant distilled water, the slides were dehydrated in increasing concentrations of ethanol (70%, 90%, and 100%) for 2 min each and then air-dried. For bright field microscopy slides were horizontally covered with a mix of Wright's staining solution (Merck, Darmstadt, Germany) and PBS (Merck) (1:1) for 5–10 min with continuous airflow. Slides were briefly washed in tap water and allowed to dry.

Four dispersion patterns were defined:

1. Sperm with large halo: the halo has a 2 times larger width than that of the sperm core, with a darker spot (sperm head) in the middle.
2. Sperm with moderate halo: having a halo size between large and small halos.
3. Sperm with small halo: a very small, clear film, that has a halo appearance, surrounds the sperm head.
4. Sperm with no halo.

Sperm DNA fragmentation percentage (SCD percentage) was calculated as the proportion of sperm with small and no halos, to the total sperm count per slide. We assessed two slides for every patient, and a total of 1000 sperms were counted per slide.

2.4. Statistical analysis

We used a Wilcoxon Rank Sum test to determine the significance of differences in sperm DNA fragmentation rates obtained from the infertile group and sub-groups compared to the control. The extent and significance of correlations between semen analysis parameters, ART outcomes, lifestyle factors and sperm DNA fragmentation rates were determined using Spearman Rank Correlation test. The statistical tests were performed using Sigma Stat Version 3.5 software (Systat Software, Erkrath, Germany). Comparisons were considered to be statistically significant at $p < 0.05$.

3. Results

DNA fragmentation was evaluated by SCD test in sperm of 185 Lebanese infertile men in comparison with a control fertile group.

3.1. Sperm DNA fragmentation in Lebanese infertile men

Our results have shown a DNA fragmentation rate in sperm of infertile group significantly higher compared to the control group. The rates were distributed between 3% and 78.80% with a mean value of 20.62% ($p < 0.001$) (Fig. 1).

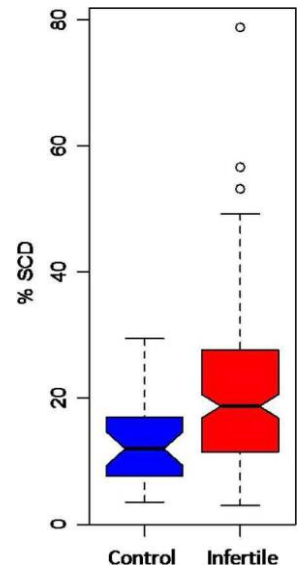


Figure 1 Comparison of sperm DNA fragmentation between the infertile group and the control group, measured by SCD test.

3.2. Sperm DNA fragmentation and semen analysis abnormalities

After finding a significant difference between fertile and infer-tile men, we suggested to divide the infertile group into sub-groups, based on semen analysis characteristics, notably sperm count, morphology and motility.

Seven sub-groups were established, and compared with the control group.

The results showed a significantly higher SCD percentage, and therefore a high sperm DNA fragmentation in the sub-groups shows altered semen analysis.

A very strong significant difference was established between the control group and the OAT (oligoasthenoteratozoosper mic), OA (oligoasthenozoospermic) and O (oligozoospermic) infertile sub-groups (p 6 0.001). A lower significant difference was also found in A (asthenozoospermic) and AT (asthenoteratozoospermic) infertile subgroups ($p = 0.025$, $p = 0.002$). No significant difference was noted with T (teratozoospermic) and N (normozoospermic) sub-groups ($p = 0.103$, $p = 0.269$) (Table 1).

3.3. DNA fragmentation and sperm count, motility and morphology

Sperm DNA fragmentation rate was also plotted with sperm count, motility and morphology.

A clear negative correlation was found between SCD per-centage and sperm count, with a very significant p value (<0.001) and a negative correlation coefficient (rho = 0.349) (Fig. 2A). A similar negative correlation was also shown between SCD percentage and sperm motility ($p < 0.001$; rho = 0.331) (Fig. 2B). The sperm motility and count decrease following a sharp slope when the DNA frag-mentation rate increases. Concerning abnormal sperm mor-phology, a positive correlation was found with sperm DNA fragmentation and teratozoospermia ($p = 0.003$, rho = +0.222) (Fig. 2C).

3.4. Sperm DNA fragmentation and ART failure

In our infertile group, a positive correlation was noted between the number of ART failures and sperm DNA fragmentation rate (rho = +0.352, $p < 0.001$) (Fig. 3).

3.5. Sperm DNA fragmentation and lifestyle

An association between sperm DNA fragmentation and life-style was then tested.

Within the infertile group, patients were identified regard-ing their exposure to heat stress or toxicants in their work as well as if they smoked or took antioxidant treatment. In every category two groups were obtained, and compared to each other according to their SCD percentages.

Table 1 Sperm DNA fragmentation comparison between control group and infertile sub-groups.

Groups comparison with control	A n = 43	AT n = 19	N n = 43	O n = 22	OA n = 27	OAT n = 26	T n = 5	Control group (C) n = 30
Mean ± SD	18.76% ±12.95	25.23% ±12.93	15.97% ±10.01	23.31% ±10.67	24.54% ±11.22	24.64% ±11.91	20.62% ±10.37	12.96% ±6.82
Median	14.65	23.60	12.86	20.44	21.58	24.58	23.80	12.00
Low/high values	6.88 78.80	6.94 48.34	3.16 53.17	3.00 49.27	9.24 48.11	3.67 56.64	5.78 29.69	3.54 29.45
p values	0.025	0.0022	0.269	<0.001	<0.001	<0.001	0.103	

C = control group; O = oligozoospermic; A = asthenozoospermic; T = teratozoospermic; N = normozoospermic.

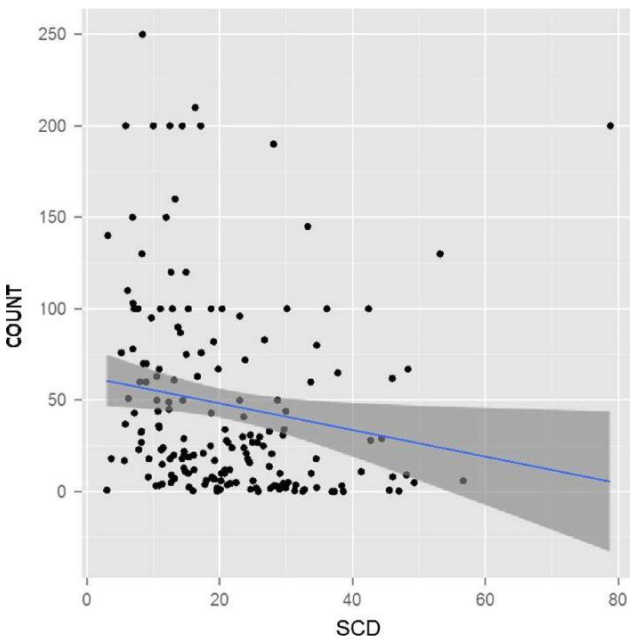


Figure 2A Correlation between sperm DNA fragmentation (SCD percentage) and Sperm Count.

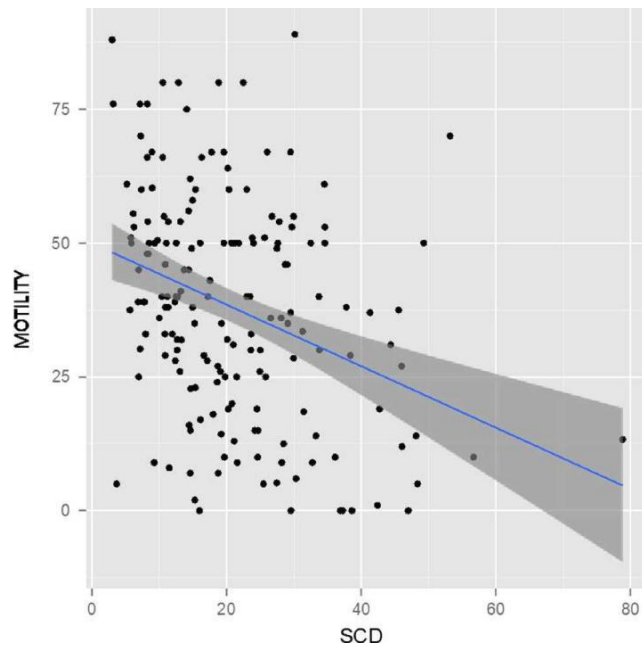


Figure 2B Correlation between sperm DNA fragmentation (SCD percentage) and Sperm Motility.

The analysis has shown a trend of positive correlation between smoking and sperm DNA fragmentation (rho = +0.050, p = 0.606). The same trend was also found between risky profession and sperm DNA fragmentation (rho = +0.059, p = 0.704). In contrast, a trend of negative correla-tion was noted between antioxidant treatment and sperm DNA fragmentation (rho = 0.079, p = 0.531).

4. Discussion

In the present study, we investigated the incidence of sperm DNA damage among 185 infertile men, attending Lebanese infertility clinics for infertility evaluation and treatment. We aimed to understand the association between sperm DNA damage, sperm characteristics, medical history and lifestyle factors. Although several studies have reported that sperm DNA damage could be associated with male infertility. The present study attempts to clarify the association between sperm DNA damage and specific categories of defects in the Lebanese infertile population.

Sperm DNA fragmentation levels measured by Sperm Chromatin Dispersion test (SCD), were compared between two groups of Lebanese fertile and infertile men. The infertile men group showed a significantly higher SCD percentage; thus, higher DNA fragmentation is compared to control. Indeed, these findings reveal that the average incidence of sperm DNA damage in fertile patients is 12.96% whereas patients from the infertile group show significantly higher level of sperm DNA damage (20.78%). Given that, this could pro-vide strong evidence for a negative effect of sperm DNA frag-mentation on male fertility potential.

Our result perfectly complies with previous studies that have reported the association between defects in sperm chro-matin and male infertility (23,25,45,46).

The group of infertile men is divided into seven sub-groups based on semen analysis phenotypes (O, A, T, OA, AT, OAT, and N). After comparing the sperm DNA fragmentation per-centages between infertile sub-groups and control, we found high rates of sperm DNA damage associated with astheno-zoospermia, and higher rates associated with oligozoospermia,

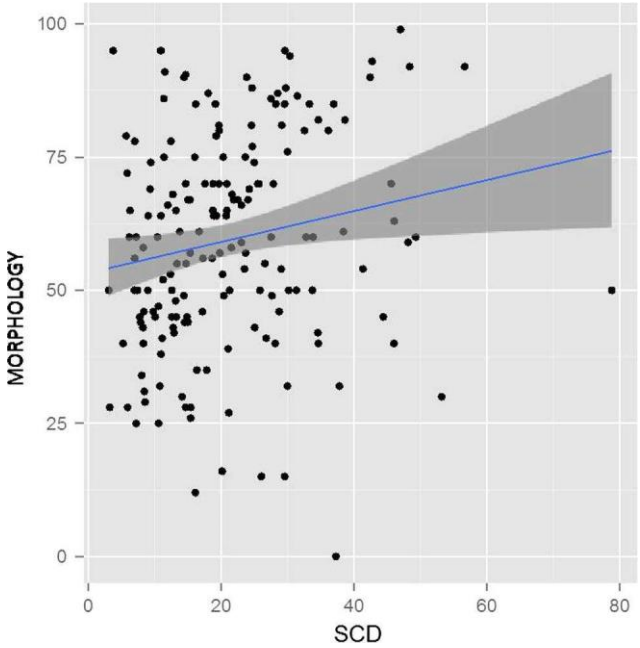


Figure 2C Correlation between sperm DNA fragmentation (SCD percentage) and Sperm Morphology.

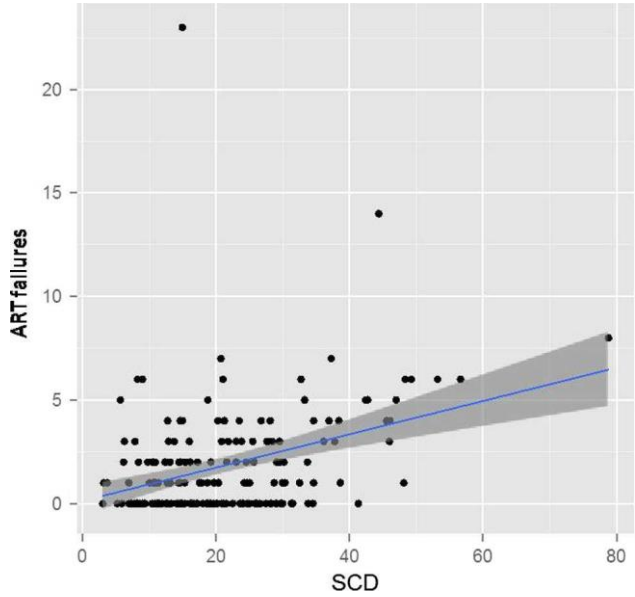


Figure 3 Relation between SCD percentage and ART failure.

oligoasthenozoospermia, asthenoteratozoospermia and oligoasthenoteratozoospermia. This finding was also reported by previous studies (21,47) and would suggest that sperm DNA fragmentation could be a promising predictive index of sperm alteration gravity, and therefore it is an important diagnostic tool for male fertility assessment.

When comparing the control group with the group of tera-tozoospermic infertile patients, there was no significant differ-ence (C/T: p = 0.103). It should be mentioned that this specific group of infertile men (T) only comprises five patients with moderated teratozoospermia, so a larger sample size will be needed in order to rule out significance.

In our study, the normozoospermic group does not appear to be significantly different from the control group (C/N: p = 0.269). Within this group, the mean value of SCD percent-age (15.97%) was higher than that of the control group (12.96%). Moreover, 28% of the normozoospermic infertile patients present a higher SCD percentage. Previous studies mentioned that around 10% of infertile patients with normal semen analysis showed a high degree of sperm DNA fragmen-tation (22,48), which could perfectly explain cases of male infertility, classified as idiopathic.

Besides the inter-group DNA fragmentation comparison, it was necessary to evaluate the association between DNA frag-mentation and semen analysis parameters. The SCD percent-age was negatively associated with sperm count and sperm motility but positively associated with abnormal sperm mor-phology (6,13,14,16,20,21,46,47,49–53). This finding suggests that an abnormal semen analysis reflects not only an altered spermatogenesis, but a deep negative effect on sperm nuclear quality.

As for teratozoospermia, its association with sperm DNA damage is much debated and has been the focus of many studies (54–57). Our study showed high SCD percentage in infertile men with teratozoospermia. This finding is extremely important, because abnormal sperm shape anomalies are easily noticed during clinical sperm processing and help to identify spermatozoa possibly carrying fragmented DNA. This result emphasizes the importance of sperm selection dur-ing ICSI as the damage carried by morphologically abnormal spermatozoa, could entail nuclear chromatin defect, and thus may affect ART outcomes such as fertilization rate, embryo cleavage, development, implantation and ongoing pregnancy (33).

Furthermore, the SCD percentage shows a positive correla-tion with failed ART attempts. ART cycles appear to fail more often with high sperm DNA fragmentation. Our finding is con-sistent with results from previous studies (9,35,43,58–60). Moreover, normal females whose partner had high sperm DNA fragmentation, suffered from recurrent pregnancy loss (RPL) (61,62).

The present study expanded into exploring the possible causes of sperm DNA fragmentation. While a number of stud-ies have been conducted regarding the association between life-style factors and sperm quality (56,63–65), studies on the Lebanese population are lacking.

Different lifestyle factors were studied. A trend of a positive correlation was observed between smoking and sperm DNA fragmentation in the Lebanese population. Smoking seems to be associated with increased sperm DNA fragmentation. Consequently, patients aiming to conceive should be encour-aged to reduce or even stop smoking which has already shown a deleterious effect on the motility of spermatozoa and its DNA integrity (66,67).

The impact of heat stress and toxicants on male fertility can be elucidated by the trend of a positive correlation between sperm DNA fragmentation and risky professions. The occupa-tions were classified as risky when the patient is exposed in his work, to toxicants, or heat. Many studies confirmed the nega-tive effect of heat stress (68–70), and exposure to toxicants (64,71,72) on sperm DNA quality.

It was denoted that high sperm DNA fragmentation after tobacco, heat and toxicants exposures, will cause testicular oxidative stress which is detrimental to sperm function

(73,74). This will be considered as a significant contributing factor to male infertility because it leads to germ cell apoptosis and dysfunction (75–77).

On the other hand, the antioxidant treatment was found to be negatively correlated with SCD percentage. Thus, it is con-sidered useful in preventing damage to spermatozoa and reducing DNA fragmentation levels (78–82). This could be a promising strategy in treating male infertility linked to nuclear sperm damage.

5. Conclusion

Our study is conducted in the Lebanese population to explore sperm DNA fragmentation. The SCD percentage is a potential discriminator between fertile and infertile patients. Moreover, sperm DNA fragmentation appears to be a powerful indicator of semen quality and ART outcomes because normal semen analysis is not sufficient to assess male fertility. We also suggested that tobacco, heat stress and toxicants exposure can affect sperm DNA integrity and we proposed the use of antioxidants as a treatment. The present study may serve as a platform for future meta-analysis studies, with a larger sample size and more variables to consider, in order to present a more definitive clarification about the role of sperm DNA fragmentation in male infertility.

Declaration of interests

The authors report no declarations of interest.

Funding

This work was funded by the National Council for Scientific Research – Lebanon (CNRS).

References

(1) Boivin J, Bunting L, Collins J, Nygren K. International estimates of infertility prevalence and treatment-seeking: potential need and demand for infertility medical care. Hum Reprod 2007;22:1506–12.

(2) Organization W. WHO laboratory manual for the examination and processing of human semen. Geneva: World Health Organi-zation; 2010.

(3) Agarwal A, Allamaneni S. Sperm DNA damage assessment: a test whose time has come. Fertil Steril 2005;84:850–3.

(4) Seli E. Spermatozoal nuclear determinants of reproductive outcome: implications for ART. Human Reprod Update 2005;11:337–49.

(5) Tarozzi N, Bizzaro D, Flamigni C, Borini A. Clinical relevance of sperm DNA damage in assisted reproduction. Reprod BioMed Online 2007;14:746–57.

(6) Evgeni E, Charalabopoulos K, Asimakopoulos B. Human sperm DNA fragmentation and its correlation with conventional semen parameters. J Reprod Infert 2014;15:2.

(7) Evgeni E, Lymberopoulos G, Gazouli M, Asimakopoulos B. Conventional semen parameters and DNA fragmentation in relation to fertility status in a Greek population. Euro J Obst Gynecol Reprod Biol 2015;188:17–23.

(8) Kodama H, Yamaguchi R, Fukuda J, Kasai H, Tanaka T. Increased oxidative deoxyribonucleic acid damage in the sperma-tozoa of infertile male patients. Fertil Steril 1997;68:519–24.

(9) Evenson D. Utility of the sperm chromatin structure assay as a diagnostic and prognostic tool in the human fertility clinic. Hum Reprod 1999;14:1039–49.

(10) Spano`M, Bonde JP, Hjøllund HI, Kolstad HA, Cordelli E, Leter G. Danish first pregnancy planner study team. Sperm chromatin damage impairs human fertility. Fertil Steril 2000;73:43–50.

(11) Zini A. Correlations between two markers of sperm DNA integrity, DNA denaturation and DNA fragmentation, in fertile and infertile men. Fertil Steril 2001;75:674–7.

(12) Saleh R, Agarwal A, Nada E, El-Tonsy M, Sharma R, Meyer A, et al. Negative effects of increased sperm DNA damage in relation to seminal oxidative stress in men with idiopathic and male factor infertility. Fertil Steril 2003;79:1597–605.

(13) Velez de la Calle J, Muller A, Walschaerts M, Clavere J, Jimenez C, Wittemer C, et al. Sperm deoxyribonucleic acid fragmentation as assessed by the sperm chromatin dispersion test in assisted reproductive technology programs: results of a large prospective multicenter study. Fertil Steril 2008;90:1792–9.

(14) Zhang L, Qiu Y, Wang K, Wang Q, Tao G, Wang L. Measurement of sperm DNA fragmentation using bright-field microscopy: comparison between sperm chromatin dispersion test and terminal uridine nick-end labeling assay. Fertil Steril 2010;94:1027–32.

(15) Chi H, Chung D, Choi S, Kim J, Kim G, Lee J, et al. Integrity of human sperm DNA assessed by the neutral comet assay and its relationship to semen parameters and clinical outcomes for the IVF-ET program. Clin Exp Reprod Med 2011;38:10.

(16) Venkatesh S, Kumar R, Deka D, Deecaraman M, Dada R. Analysis of sperm nuclear protein gene polymorphisms and DNA integrity in infertile men. Syst Biol Reprod Med 2011:1–9.

(17) Irvine S, Twigg P, Gordon L, Fulton N, Milne A, Aitken R. DNA integrity in human spermatozoa: relationships with semen quality. J Androl 2000;21:33–44.

(18) Muratori M, Piomboni P, Baldi E, Filimberti E, Pecchioli P, Moretti E, Gambera L, Baccetti B, Biagiotti R, Forti G, Maggi M. Functional and ultrastructural features of DNA-fragmented human sperm. J Androl 2000;21:903–12.

(19) Caglar GS, Koester F, Schoepper B, Asimakopoulos B, Nehls B, Nikolettos N, Diedrich K, Al-Hasani S. Semen DNA fragmen-tation index, evaluated with both TUNEL and Comet assay, and the ICSI outcome. In Vivo 2007;21:1075–80.

(20) Fang L, Lou LJ, Ye YH, Jin F, Zhou J. A study on correlation between sperm DNA fragmentation index and age of male, various parameters of sperm and in vitro fertilization outcome. Zhonghua yi xue yi chuan xue za zhi= Zhonghua yixue yichuanxue zazhi= Chinese. J Med Genet 2011;28:432–5.

(21) Varshini J, Srinag B, Kalthur G, Krishnamurthy H, Kumar P, Rao S, et al. Poor sperm quality and advancing age are associated with increased sperm DNA damage in infertile men. Andrologia 2011;44:642–9.

(22) Mayorga-Torres B, Cardona-Maya W, Cadavid A, Camargo M. Evaluation of sperm functional parameters in normozoospermic infertile individuals. Actas Urolo´gicas Espan˜olas (English Edi-tion) 2013;37:221–7.

(23) Oleszczuk K, Augustinsson L, Bayat N, Giwercman A, Bungum M. Prevalence of high DNA fragmentation index in male partners of unexplained infertile couples. Andrology 2012;1:357–60.

(24) Muriel L, Goyanes V, Segrelles E, Gosalvez J, Alvarez J, Fernandez J. Increased aneuploidy rate in sperm with fragmented dna as determined by the sperm chromatin dispersion (SCD) test and FISH analysis. J Androl 2006;28:38–49.

(25) Qiu Y, Wang L, Zhang L, Yang D, Zhang A, Yu J. Analysis of sperm chromosomal abnormalities and sperm DNA fragmenta-tion in infertile males. Zhonghua yi xue yi chuan xue za zhi= Zhonghua yixue yichuanxue zazhi= Chinese. J Med Genet 2008;25:681–5.

(26) Perrin A, Basinko A, Douet-Guilbert N, Gueganic N, Le Bris M, Amice V, et al. Aneuploidy and DNA fragmentation in sperm of

carriers of a constitutional chromosomal abnormality. Cytogenet Geno Res 2011;133:100–6.

(27) Younes B, Hazzouri K, Chaaban M, Karam W, Jaoude I, Attieh J, et al. High frequency of sex chromosomal disomy in sperma-tozoa of Lebanese infertile men. J Androl 2010;32:518–23.

(28) Tesarik J. Late, but not early, paternal effect on human embryo development is related to sperm DNA fragmentation. Hum Reprod 2004;19(3):611–5.

(29) Meseguer M, Santiso R, Garrido N, Gil-Salom M, Remohı´J, Fernandez J. Sperm DNA fragmentation levels in testicular sperm samples from azoospermic males as assessed by the sperm chromatin dispersion (SCD) test. Fertil Steril 2009;92:1638–45.

(30) Simon L, Castillo J, Oliva R, Lewis S. Relationships between human sperm protamines, DNA damage and assisted reproduc-tion outcomes. Reprod BioMed Online 2011;23:724–34.

(31) Avendano C, Oehninger S. DNA fragmentation in morpholog-ically normal spermatozoa: how much should we be concerned in the ICSI Era? J Androl 2010;32:356–63.

(32) Zini A, Jamal W, Cowan L, Al-Hathal N. Is sperm dna damage associated with IVF embryo quality? A systematic review. J Assist Reprod Genet 2011;28:391–7.

(33) Simon L, Murphy K, Shamsi M, Liu L, Emery B, Aston K, et al. Paternal influence of sperm DNA integrity on early embryonic development. Hum Reprod 2014;29:2402–12.

(34) Tandara M, Bajic A, Tandara L, Bilic-Zulle L, Sunj M, Kozina V, et al. Sperm DNA integrity testing: big halo is a good predictor of embryo quality and pregnancy after conventional IVF. Androl-ogy 2014;2:678–86.

(35) Sivanarayana T, Ravi Krishna C, Jaya Prakash G, Krishna K, Madan K, Sudhakar G, et al. Sperm DNA fragmentation assay by sperm chromatin dispersion (SCD): correlation between DNA fragmentation and outcome of intracytoplasmic sperm injection. Reprod Med Biol 2013;13:87–94.

(36) Lazaros L, Vartholomatos G, Pamporaki C, Kosmas I, Takenaka A, Makrydimas G, Sofikitis N, Stefos T, Zikopoulos K, Hatzi E, Georgiou I. Sperm flow cytometric parameters are associated with ICSI outcome. Reprod Biomed online 2013;26:611–8.

(37) Parinaud J, Mieusset R, Vieitez G, Labal B, Richoilley G. Influence of sperm parameters on embryo quality. Fertil Steril 1993;60:888–92.

(38) Janny L, Menezo Y. Evidence for a strong paternal effect on human preimplantation embryo development and blastocyst formation. Mol Reprod Dev 1994;38:36–42.

(39) Shoukir Y, Chardonnens D, Campana A, Sakkas D. Blastocyst development from supernumerary embryos after intracytoplasmic sperm injection: a paternal influence? Hum Reprod 1998;13:1632–7.

(40) Carrell DT, Liu L, Peterson CM, Jones KP, Hatasaka HH, Erickson L, Campbell B. Sperm DNA fragmentation is increased in couples with unexplained recurrent pregnancy loss. Arch Androl 2003;49:49–55.

(41) Borini A, Tarozzi N, Bizzaro D, Bonu M, Fava L, Flamigni C, et al. Sperm DNA fragmentation: paternal effect on early post-implantation embryo development in ART. Hum Reprod 2006;21:2876–81.

(42) Gil-Villa A, Cardona-Maya W, Agarwal A, Sharma R, Cadavid A. Assessment of sperm factors possibly involved in early recurrent pregnancy loss. Fertil Steril 2010;94:1465–72.

(43) Absalan F, Ghannadi A, Kazerooni M, Parifar R, Jamalzadeh F, Amiri S. Value of sperm chromatin dispersion test in couples with unexplained recurrent abortion. J Assist Reprod Genet 2011;29:11–4.

(44) AlKusayer G, Amily N, Abou-Setta A, Bedaiwy M. Live birth rate following IVF/ICSI in patients with sperm DNA fragmen-tation: a systematic review and meta-analysis. Fertil Steril 2014;102:e305.

(45) Ferna´ndez JL, Muriel L, Goyanes V, Segrelles E, Gosa´lvez J, Enciso M, LaFromboise M, De Jonge C. Simple determination of human sperm DNA fragmentation with an improved sperm chromatin dispersion test. Fertil Steril 2005;84:833–42.

(46) Shamsi M, Imam S, Dada R. Sperm DNA integrity assays: diagnostic and prognostic challenges and implications in man-agement of infertility. J Assist Reprod Genet 2011;28:1073–85.

(47) Belloc S, Benkhalifa M, Cohen-Bacrie M, Dalleac A, Chahine H, Amar E, et al. Which isolated sperm abnormality is most related to sperm DNA damage in men presenting for infertility evalua-tion. J Assist Reprod Genet 2014;31:527–32.

(48) Khodair H, Omran T. Evaluation of reactive oxygen species (ROS) and DNA integrity assessment in cases of idiopathic male infertility. Egypt J Dermatol Venerol 2013;33:51.

(49) Trisini A, Singh N, Duty S, Hauser R. Relationship between human semen parameters and deoxyribonucleic acid damage assessed by the neutral comet assay. Fertil Steril 2004;82:1623–32.

(50) Cohen-Bacrie P, Belloc S, Me´ne´zo Y, Clement P, Hamidi J, Benkhalifa M. Correlation between DNA damage and sperm parameters: a prospective study of 1633 patients. Fertil Steril 2009;91:1801–5.

(51) Varghese A, Bragais F, Mukhopadhyay D, Kundu S, Pal M, Bhattacharyya A, et al. Human sperm DNA integrity in normal and abnormal semen samples and its correlation with sperm characteristics. Andrologia 2009;41:207–15.

(52) Matsuura R, Takeuchi T, Yoshida A. Preparation and incubation conditions affect the DNA integrity of ejaculated human spermatozoa. Asian J Androl 2010;12:753–9.

(53) Curti G, Ca´nepa M, Cantu´ L, Montes J. Diagnosis of male infertility: a need of functional and chromatin evaluation. Actas Urolo´gicas Espan˜olas (English Edition) 2013;37:100–5.

(54) Avendan˜o C, Franchi A, Taylor S, Morshedi M, Bocca S, Oehninger S. Fragmentation of DNA in morphologically normal human spermatozoa. Fertil Steril 2009;91:1077–84.

(55) Avendan˜o C, Franchi A, Duran H, Oehninger S. DNA fragmen-tation of normal spermatozoa negatively impacts embryo quality and intracytoplasmic sperm injection outcome. Fertil Steril 2010;94:549–57.

(56) Elshal M, El-Sayed I, Elsaied M, El-Masry S, Kumosani T. Sperm head defects and disturbances in spermatozoal chromatin and DNA integrities in idiopathic infertile subjects: association with cigarette smoking. Clin Biochem 2009;42:589–94.

(57) Mehdi M, Khantouche L, Ajina M, Saad A. Detection of DNA fragmentation in human spermatozoa: correlation with semen parameters. Andrologia 2009;41:383–6.

(58) Duran E. Sperm DNA quality predicts intrauterine insemination outcome: a prospective cohort study. Hum Reprod 2002;17:3122–8.

(59) Speyer B, Pizzey A, Ranieri M, Joshi R, Delhanty J, Serhal P. Fall in implantation rates following ICSI with sperm with high DNA fragmentation. Hum Reprod 2010;25:1609–18.

(60) Jiang H, He R, Wang C, Zhu J. The relationship of sperm DNA fragmentation index with the outcomes of in-vitro fertilisation-embryo transfer and intracytoplasmic sperm injection. J Obstet Gynaecol 2011;31:636–9.

(61) Ribas-Maynou J, Garcı´a-Peiro´A, Fernandez-Encinas A, Amen-gual M, Prada E, Corte´s P, et al. Double stranded sperm DNA breaks, measured by Comet assay, are associated with unex-plained recurrent miscarriage in couples without a female factor. PLoS ONE 2012;7:e44679.

(62) Robinson L, Gallos I, Conner S, Rajkhowa M, Miller D, Lewis S, et al. The effect of sperm DNA fragmentation on miscarriage rates: a systematic review and meta-analysis. Hum Reprod 2012;27:2908–17.

(63) Soares S, Melo M. Cigarette smoking and reproductive function. Curr Opin Obstet Gynecol 2008;20:281–91.

(64) Calogero A, La Vignera S, Condorelli R, Perdichizzi A, Valenti D, Asero P, et al. Environmental car exhaust pollution damages human sperm chromatin and DNA. J Endocrinol Invest 2010;34: e139–43.

(65) Viloria T, Meseguer M, Martı´nez-Conejero J, O'Connor J, Remohı´J, Pellicer A, et al. Cigarette smoking affects specific sperm oxidative defenses but does not cause oxidative DNA damage in infertile men. Fertil Steril 2010;94:631–7.

(66) Sepaniak S, Forges T, Gerard H, Foliguet B, Bene M, Monnier-Barbarino P. The influence of cigarette smoking on human sperm quality and DNA fragmentation. Toxicology 2006;223:54–60.

(67) Anifandis G, Bounartzi T, Messini C, Dafopoulos K, Sotiriou S, Messinis I. The impact of cigarette smoking and alcohol consumption on sperm parameters and sperm DNA fragmenta-tion (SDF) measured by Halosperm . Arch Gynecol Obstet 2014;290:777–82.

(68) Paul C, Murray A, Spears N, Saunders P. A single, mild, transient scrotal heat stress causes DNA damage, subfertility and impairs formation of blastocysts in mice. Reproduction 2008;136 (1):73–84.

(69) Kanter M, Aktas C, Erboga M. Heat stress decreases testicular germ cell proliferation and increases apoptosis in short term: an immunohistochemical and ultrastructural study. Toxicol Ind Health 2011;29(2):99–113.

(70) Santiso R, Tamayo M, Gosa´lvez J, Johnston S, Marin˜o A, Ferna´ndez C, et al. DNA fragmentation dynamics allows the assessment of cryptic sperm damage in human: evaluation of exposure to ionizing radiation, hyperthermia, acidic pH and nitric oxide. Mutat Res/Funda Mol Mech Mutag 2012;734:41–9.

(71) Gomes M, Gonc¸ alves A, Rocha E, Sa´R, Alves A, Silva J, et al. Effect of in vitro exposure to lead chloride on semen quality and sperm DNA fragmentation. Zygote 2014;23:384–93.

(72) Hosseinpour E, Shahverdi A, Parivar K, Sedighi Gilani M, Nasr-Esfahani M, Salman Yazdi R, et al. Sperm ubiquitination and DNA fragmentation in men with occupational exposure and varicocele. Andrologia 2013;46:423–9.

(73) Cocuzza M, Sikka S, Athayde K, Agarwal A. Clinical relevance of oxidative stress and sperm chromatin damage in male infertility: an evidence based analysis. Int Braz J Urol 2007;33:603–21.

(74) Tremellen K. Oxidative stress and male infertility–a clinical perspective. Human Reprod Update 2008;14:243–58.

(75) Sakkas D, Mariethoz E, St. John J. Abnormal sperm parameters in humans are indicative of an abortive apoptotic mechanism linked to the fas-mediated pathway. Exp Cell Res 1999;251:350–5.

(76) Sakkas D. Nature of DNA damage in ejaculated human spermatozoa and the possible involvement of apoptosis. Biol Reprod 2002;66:1061–7.

(77) Novotny J, Aziz N, Rybar R, Brezinova J, Kopecka V, Filipcikova R, Reruchova M, Oborna I. Relationship between reactive oxygen species production in human semen and sperm DNA damage assessed by Sperm Chromatin Structure Assay. Biomed Papers 2013;157:383–6.

(78) Greco E. Reduction of the incidence of sperm DNA fragmenta-tion by oral antioxidant treatment. J Androl 2005;26:349–53.

(79) Greco E. ICSI in cases of sperm DNA damage: beneficial effect of oral antioxidant treatment. Hum Reprod 2005;20:2590–4.

(80) Me´ne´zo Y, Hazout A, Panteix G, Robert F, Rollet J, Cohen-Bacrie P, et al. Antioxidants to reduce sperm DNA fragmenta-tion: an unexpected adverse effect. Reprod BioMed Online 2007;14:418–21.

(81) Abad C, Amengual M, Gosa´lvez J, Coward K, Hannaoui N, Benet J, et al. Effects of oral antioxidant treatment upon the dynamics of human sperm DNA fragmentation and subpopula-tions of sperm with highly degraded DNA. Andrologia 2012;45:211–6.

(82) Talevi R, Barbato V, Fiorentino I, Braun S, Longobardi S, Gualtieri R. Protective effects of in vitro treatment with zinc, d-aspartate and coenzyme q10 on human sperm motility, lipid peroxidation and DNA fragmentation. Reprod Biol Endocrinol 2013;11:81.

Middle East Fertility Society Journal 23 (2018) 85–90

Contents lists available at ScienceDirect

Middle East Fertility Society Journal

journalhomepage:www.sciencedirect.com

Review

Antioxidants modulation of sperm genome and epigenome damage: Fact or fad? Converging evidence from animal and human studies

Fadi Choucair [a,b], Elias Saliba [b], Imad Abou Jaoude [c], Mira Hazzouri [b,⇑]

[a]Université de Nice Sophia Antipolis, France
[b]Lebanese University, Faculty of Sciences 2, Fanar, Lebanon
[c]Abou Jaoude Hospital, Department of Obstetrics and Gynecology, Lebanon

ARTICLE INFO

Article history:
Received 9 November 2017
Accepted 13 January 2018
Available online 2 February 2018

Keywords:
Sperm
Antioxidants
Micronutrients
Reactive oxygen species
Genome
Epigenome

ABSTRACT

Increasing evidence suggests that oxidative stress plays a major role in the pathogenesis of sperm DNA damage. Oxidative stress was also recently found to modulate the epigenetic make up of sperm. Along these lines, a growing body of evidence in both experimental and clinical studies has implicated several regimens of antioxidants, by oral administration or in vitro supplementation to sperm-preparation media, in improving various sperm parameters namely DNA damage. While these studies exhibited heterogeneity in treatment regimens, and variability in methodology, there remains a lack of quality evi-dence on the association between micronutrients and sperm DNA integrity. Another ancillary effect of antioxidants administration on sperm is the shaping of the epigenome. It is beginning to surface that micronutrients function as potent modulators of the sperm epigenome-regulated gene expression through regulation of mainly DNA methylation in humans and experimental models. However, the few promising experimental studies on mice supported the notion that epigenetic marks in spermatogenesis are dynamic and can be modulated by nutritional exposure. More so, the sperm epigenome was proposed to transfer a so-called epigenomic map to the offspring which can influence their development. Here, we review and summarize the current evidence in human and animal models research regarding the link between genome and epigenome micronutrients environment interactions on the sperm nuclear dam-age. Unfortunately, our conclusion is not very conclusive, rather, it opens an avenue to investigate the fortifying effect of antioxidants on sperm cells. Hopefully, further genome and epigenome-wide studies focusing on the prenatal environment, will serve as a promising route for embodying the possibility of "normalization" and restoration of some offspring health cues.

2018 Middle East Fertility Society. Production and hosting by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Contents

1. Introduction . . . 86
2. Influence of antioxidants administration on sperm genome . . . 86
   2.1. In-vivo supplementation. . . . 86
   2.2. In-vitro supplementation . . . 87
   2.3. New trends: natural antioxidants . . . 87
   2.4. Conclusions . . . 87
3. Influence of antioxidants administration on sperm epigenome . . . 88
   3.1. Human studies . . . 88
   3.2. Animal studies . . . 88
   3.3. Conclusions . . . 89

Peer review under responsibility of Middle East Fertility Society.

⇑ Corresponding author at: Lebanese University, Faculty of Sciences 2, Department of Biology, Fanar, Lebanon.
E-mail address: hazzourim@yahoo.com (M. Hazzouri).

https://doi.org/10.1016/j.mefs.2018.01.006
1110-5690/ 2018 Middle East Fertility Society. Production and hosting by Elsevier B.V.
This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

4. Overall conclusions and prospects 89
Conflict of interest. 89
References 89

1. Introduction

Once it is produced by the testicles, sperm starts a long journey of maturational processes and transits before it reaches the oocyte. Since, mammalian spermatozoa is rich in polyunsaturated fatty acids [1], it is very prone to reactive oxygen species (ROS) attacks.

As we know, the central tenet of oxidative stress is that an imbalance between oxidants and antioxidants in favor of the oxi-dants potentially leads to cellular damage.

Along these lines, oxidative-based damage in spermatozoa is characterized by several hallmarks of ‘‘incompetent” sperm pri-marily including DNA damage [2,3], reduced sperm numbers, diminished motility, abnormal morphology [4], and altered zona penetration [5].

However, in order for ROS to exhibit strong genotoxic damaging effects, several factors are implicated in sperm oxidative damage susceptibility. First off, a disruption in the sperm maturation pro-cess will make spermatozoa more susceptible to genotoxic ROS attacks. In this perspective, Aitken et al. 2009 [6], proposed a new hypothesis ‘‘the two-step hypothesis” that assigns the sperm DNA damage to a defective remodelling of sperm chromatin during spermiogenesis. This renders the nucleus more vulnerable to ROS attacks which will create DNA strand breaks. Furthermore, oxida-tive stress may further activate sperm endonucleases which will create DNA strand nicks [7,8]. Thus, the common management for sperm oxidative stress is micronutrients administration; namely antioxidants intake or supplementation. A very early attempt to study the impact of ROS on sperm was presented in a landmark paper by MacLeod et al. 1943 [9]. Interestingly, this study showed that hydrogen peroxide causes a rapid sperm motil-ity loss, an effect which was reversed by catalase supplementation of culture media. Another study performed by Fraga et al. in 1991

[10] showed that an ascorbic acid-based diet for 28 days decreased the levels of sperm oxodGDNA – a biomarker for oxidation. Fast-forwarding to this era, it has been postulated that oxidative stress can modulate the epigenetic make up of sperm [11]. By definition, epigenetics refers to the study of heritable changes in gene expres-sion that occur without a change in DNA sequence [12]. Hence, sperm epigenome was shown to regulate spermatogenesis, fertilization and embryogenesis [13]. Given the plausible role of antioxidants on oxidative-based sperm damage, we elected to examine the evidence in experimental and human studies involv-ing the effect of antioxidants supplementation on the sperm genome and epigenome. Studies included in this overview encom-pass investigations of sperm nuclear status, not discriminating between the types of antioxidants (single antioxidants, multiple antioxidants, . . .).

2. Influence of antioxidants administration on sperm genome

DNA damage is the most common effect of oxidative stress on human spermatozoa. Since an imbalance between scavenging antioxidants and pro-oxidants have been implicated in the patho-physiology of oxidative-based DNA damage, the administration of antioxidants has been practiced for years in an attempt to improve sperm quality necessary for reproductive success. Recently, a growing body of evidence has implicated antioxidants, by oral administration or in vitro supplementation in animals and humans, in improving various sperm parameters namely DNA damage. Here, our primary analysis was limited to studies report-ing DNA quality as a treatment endpoint. Thus, due to the relative plethora of data obtained from human studies, this section includes only discussions of human data.

2.1. In-vivo supplementation

The use of human models to evaluate the effects of antioxidants on sperm DNA damage has been emphasized over the years.

Many studies have used rats [14,15], dogs [16], and rabbits [17,18] as models. But human trial results that paralleled even exceeded those obtained from animal studies have further attested the impact of antioxidants on sperm.

A small placebo-randomized study by Greco et al. 2005 [19] examined the effects of antioxidants, on sperm with high DNA fragmentation (15% TUNEL – positive spermatozoa). The 64 patients enrolled were randomized between an antioxidant treat-ment (1 g vitamin C and 1 g vitamin E daily for 2 months) group and a placebo group. At the end of the treatment, the percentage of DNA-fragmented spermatozoa detected by TUNEL assay was found to be significantly lower in the treatment group. Another study specifically looked at sperm DNA integrity, protamination and apoptosis [20]. This prospective study included 45 men with oxidative stress confirmed by nitroblue tetrazolium (NBT) assay, who received one capsule of Menevit (Lycopene 6 mg, Vitamin E 400 IU, Vitamin C 100 mg, Zinc 25 mg, Selenium 26 μg, Folate

500 μg, Garlic oil 333 μg) for 3 months. At first, the levels of ROS significantly decreased post-treatment. In the light of this reduc-tion several improvements in sperm quality were marked. Sperm DNA integrity, assessed by TUNEL, improved significantly. Further-more, the observed improvements in sperm DNA quality were reflected by reductions in early apoptosis (annexin V positive, pro-pidium iodide negative). Moreover, the levels of sperm DNA pro-tamination, determined by Chromomycin A3 (CMA3), also statistically improved by the end of the treatment.

Surprisingly, Menezo et al. 2007 [21] noticed an ‘‘unexpected adverse effect”. A total of 58 patients with high DNA fragmentation or high degree of decondensation (>15%) – detected by Sperm Chromatin Structure Assay (SCSA) – were enrolled, then treated with oral antioxidants consisting of Vitamin C (400 mg), Vitamin E (400 mg), β-carotene (18 mg), zinc (500 μmol) and selenium (1 μmol) for 3 months. While a highly significant decrease in DFI was observed post-treatment, at the same extent sperm deconden-sation significantly increased. This iatrogenic effect may be related to antioxidants namely vitamin C that reduces cystine to two cys-teine moieties, and thus opening the interchain disulphide bridges. The most recent paper of Amar et al. 2015, treated 151 patients using a sequential regimen of treatment. A significant decrease in DFI and reduction in nuclear decondensation were noted, when those patients were treated with potent antioxidants for 5 weeks followed by a long duration of another combination of antioxidants components of the one carbon cycle [22]. In the one carbon cycle 1 carbon cycle, for example zinc and vitamins B2, 6 and 12 will nour-ish the homocysteine pathway this fortifies the endogenous antioxidant capacity as well as cysteine activity leading to an effect on methylation. In a study evaluating the effect of multivitamins on sperm DNA status specifically in the context of varicocele, Gual-Frau et al. 2015 [23] administrated multivitamins (1500 mg L-Carnitine, 60 mg vitamin C, 20 mg coenzyme Q10, 10 mg vitamin

E, 200 Lg vitamin B9, 1 Lg vitamin B12, 10 mg zinc, 50 Lg sele-nium) daily for three months to 20 infertile patients with grade I varicocele. An average reduction of 22.1% in sperm DNA fragmen-tation (assessed by Sperm Chromatin Dispersion SCD test), and a significant decrease by 31.3% of highly degraded sperm cells were reported compared to controls. Young et al. 2008 [24] explored the effect of antioxidants intake on chromosomal status in healthy men. Sperm samples from 89 healthy, non smoking men from a non-clinical setting were analyzed for aneuploidy using fluorescent in situ hybridization with probes for chromosomes X, Y and 21. Daily total intake (diet and supplements) for zinc, folate, vitamin C, vitamin E and b-carotene was derived from a food frequency questionnaire. Men with high folate intake had lower frequencies of sperm with disomies X, 21, sex nullisomy, and a lower aggregate measure of sperm aneuploidy compared with men with lower intake.

While the protective effect of antioxidants has been well-documented, antioxidants have also gained increasing attention in subgroups of infertile men, due to their antigenotoxic role in improving sperm DNA status. Studies have presented evidence of the antigenotoxic effect of antioxidants in infertile men presenting different pathologies of the sperm.

A small study specifically looked at the role of zinc supplementation in asthenozoospermic patients [25]. This study included 45 patients randomized and allocated into 4 groups: zinc only (200 mg zinc sulphate 2); zinc (200 mg 2) + vitamin E (10 mg 2) and zinc (200 mg 2) + vitamins E (10 mg 2) + C (5 mg 2) for 3 months, and non-therapy control group. Zinc therapy alone or with various antioxidants combinations significantly reduced the oxidative stress markers associated with asthenozoospermia (malonedialdehyde, tumour necrosis factor A), sperm apoptosis examined by light and electron microscopy and Bcl-2 Bax expres-sion. Parallel sperm in vitro experiment, by inducing oxidative stress (5 mM NADH) and adding Zinc (20 mM ZnCl). A significant decrease in sperm DNA fragmentation assayed by acridine orange was observed. It is worthy to note that there appears a lack of clear reporting of statistical results in the aforementioned study. In another study evaluating the effect of CoQ10 on sperm DNA dam-age (assessed by Comet), Tirabassi et al. 2015 [26] administrated (200 mg) and D-Asp (2660 mg) to 20 patients with idiopathic asthenozoospermia for 3 months. The results showed that CoQ10 plays a protective role against oxidative stress and DNA damage.

By contrast, another study examined sperm DNA in asthenoteratozoospermic men (n = 36) with leukocytospermia treated, with a formula of 20 mg beta-glucan, 50 mg fermented papaya, 97 mg lactoferrin, 30 mg vitamin C, and 5 mg vitamin E for 3 months, ver-sus 15 untreated controls [27]. DNA damage detected by acridine orange staining did not show any significant change. In the same context, a small study looked at the effect of oral antioxidant treat-ment upon the dynamics of sperm DNA fragmentation in asthenoteratozoospermic men [28].

Oral antioxidant treatment (1500 mg of L-Carnitine; 60 mg of vitamin C; 20 mg of coenzyme Q10; 10 mg of vitamin E; 10 mg of zinc; 200 Mg of vitamin B9; 50 Mg of selenium; 1 Mg of vitamin B12) during a time period of 3 months was administrated to 20 infertile men with asthenoteratozoospermia. This resulted in a sig-nificant improvement of DNA integrity (assessed by sperm chro-matin dispersion test (SCD)).

2.2. In-vitro supplementation

Several attempts have been also made to evaluate the effect of antioxidants supplementation on sperm DNA based on in vitro experiments.

Consequently, we will review the evidence about the impact of antioxidants supplementation on sperm during its processing and preparation, and will present additional evidence related to sperm cryopreservation which we have accumulated over the past years.

Hughes et al. 1998 [29] prepared spermatozoa from 150 patients in the presence of ascorbic acid, alpha tocopherol, urate, or acetyl cysteine. While, DNA damage was induced by ionizing irradiation, DNA strand breakage was measured using the comet assay. Sperm preparation using a supplemented medium with combination of antioxidants, significantly reduced sperm DNA damage. Similarly, another study evaluated the antigenotoxic effect of a dual combination of antioxidants (ascorbate and A-tocopherol) in vitro with graded doses. DNA damage was induced using hydrogen peroxide and DNA integrity was determined using the comet assay. Thus, the authors observed a significant reduction of DNA damage in a dose-dependent manner with both vitamins compared to controls [30]. The same group further evaluated the combination of glutathione and hypotaurine addition to sperm preparation medium on sperm DNA integrity. It appears that this combination did not affect baseline DNA integrity [31].

Moreover, other studies have also been conducted using various formulations of antioxidants: ethylenediaminetetraacetic acid (EDTA) and catalase [32], isoflavone genistein and equol [33], cata-lase [34], all showed a significant protective effect against sperm DNA damage.

While the protective effect of antioxidants has been widely studied during sperm processing, this effect has also been attribu-ted during sperm freezing. The enrichment of cryoprotective agents with antioxidants, protected sperm from cryodamage.

Several studies have been conducted to evaluate the addition of antioxidants to cyoprotectants during sperm cryopreservation. Experimental evidence specifically targeting sperm DNA cryodam-age post-thaw are numerous, this included supplementation with zinc [35,36], L-carnitine [37], isoflavonegenistein [38], ascorbic acid and resveratol [39], quercetin [40], ascorbate and catalase [41]and vitamin E [42]. Nevertheless, DNA damage in the aforementioned studies was appraising using different approaches mainly TUNEL assay, sperm chromatin dispersion test (SCD) and comet assay.

2.3. New trends: natural antioxidants

Antioxidants therapy has been shifted to being more natural as a recent trend. Accordingly several herbal plants extracts has been evaluated for their antioxidant and antigenotoxic effects on sper-matozoa. Several natural agents were tested on spermatozoa by in vitro incubation. Supplementation of hydroxytyrosol – a pheno-lic compound found in virgin oil and in olive mill waste (OMW) – in culture media significantly decreased sperm DNA oxidation assessed by flow cytometry and ROS levels in centrifugation-induced sperm damage group [43].

In vitro incubation of sperm from leukocytospermic patients with curcumin resulted in a significant decrease in sperm mito-chondrial DNA copy numbers thus mitigating mitochondrial DNA injury [44]. Moreover, administration of food saffron to profes-sional male cyclists alleviated sperm oxidative DNA damage assayed by TUNEL and increased seminal antioxidants biomarkers [45].

2.4. Conclusions

The evidence is insufficient to support the conclusion that antioxidants administration in vivo from clinical trials protects sperm genome from DNA damage, while there may be some ben-efit from culture medium enrichment with antioxidants in confer-ring an antigenotoxic effect on sperm cells.

In light of the above-mentioned findings, we noticed from the key publications relating to in vivo administration of antioxidants the following:

The duration of treatment with antioxidants was fixed to 3 months in the majority of trials, which would cover one full sper-matogenesis cycle. Hence, the limitations of the available evidence base have been referred to throughout this review. Notably, there is a lack of valid double-blinded, placebo-controlled trials, and there is scanty data on critical end-points to confirm oxidative stress in many published studies. Moreover, another constraint is the low power of this evidence owing to low-sample size and bias in patient's selection.

It is noteworthy that, these small studies did not evaluate the mechanism of action of antioxidants neither the long term effects of this short treatment course.

In parallel, in vitro experiments examined different doses as a crucial determinant of cellular effect, as herein, a basic principle in toxicology.

3. Influence of antioxidants administration on sperm epigenome

Yet, another ancillary effect of antioxidants administration on sperm is the shaping of the epigenome. To note that so far, epige-netics is a new promising strikingly evolving field of research. However, changes in the sperm epigenetic make up in response to antioxidants exposure appear to be a new perspective of epige-netic modulation linking dietary factors with the genome.

3.1. Human studies

To date, there is paucity of data that have examined this role. That been said, sperm DNA methylation is the only epigenetic mark studied post antioxidants treatments. While most of the studies nevertheless focused on studying the effect of antioxidants administration on sperm DNA quality, data on epigenetic marks is very limited. Tunc et al. 2009 [11] sought out to study global DNA methylation among infertile men – defined by the presence of abnormal WHO semen quality criteria (WHO 1999) and the inabil-ity to conceive despite more than 12 months of unprotected inter-course – in comparison to fertile controls. Infertile subjects were treated with antioxidants formula tablets for 3 months. This study included 45 infertile men who received one capsule of antioxi-dants, Menevit formula (Lycopene 6 mg, Vitamin E 400 IU, Vitamin C 100 mg, Zinc 25 mg, Selenium 26 ㎍, Folate 500 ㎍, Garlic oil 333 ㎍) for 3 months. Considerably, higher levels of seminal ROS and fragmented DNA TUNEL-positive sperms in infertile patients were noted. Interestingly, a significant negative correlation was observed between sperm DNA methylation and sperm DNA frag-mentation, as well as for seminal reactive oxygen species (ROS) production. After supplementation by antioxidants, seminal ROS and levels of sperm DNA fragmentation were significantly reduced, while sperm DNA methylation notably increased. Moreover, no correlation was found between serum homocysteine and sperm DNA methylation which rejects the potential role of abnormalities in the folate/homocysteine pathway in causing sperm DNA hypomethylation. Noteworthy that, this study lacks a placebo group and a clear genotyping of patients for MTHFR gene which can certainly affect DNA methylation results.

Since DNA methylation is affected by methyl groups production and pool through the one carbon folate metabolism, two studies looked at the effect of folate supplementation and sperm DNA methylation levels. But before examining the above-mentioned studies, we will describe the connection between folate metabolism and methionine production. Briefly, Tetrahydrofolate is reduced to 5-methyl-tetrahydrofolate via methyl-tetra-hydrofolate reductase (MTHFR) and the subsequent methyl group is transferred to homo-cysteine via methionine synthase to generate methionine. The methyl group of methionine is activated by adenylation to form S-Adenosyl-Methionine (SAMe) that acts as the universal methyl donor for any acceptor including DNA [46]. Aarabi et al. 2015 [47] measured methylation levels in 30 normozoospermic men with idiopathic infertility before and after high-dose folate supplemen-tation for 6 months. Methylation levels of the differentially methy-lated regions of several imprinted loci (H19, DLK1/GTL2, MEST, SNRPN, PLAGL1, KCNQ1OT1) were unchanged normal pre- and post-treatment. Interestingly, as for global DNA methylation, a sig-nificant global loss of methylation across different regions of the sperm genome was observed. Incredibly, a drastic loss of global methylation was detected by reduced representation bisulfite sequencing (RRBS) methylation in specific patients genotype, homozygous for MTHFR C6777T polymorphism. In particular, this polymorphism reduced by 50–60% the MTHFR activity [48] subse-quently modifying the response to high-dose folic acid supplemen-tation. In the light of these data, it could be postulated that high folic acid supplementation may down-regulate the MTHFR enzyme. Moreover, Ingenuity Pathway Analysis (IPA) to depict disease-related genes, showed that genes related to cancer and neurobehav-ioral disorders were affected by the supra-folic acid supplementation.

In the same context, a recent small study by Chan et al. 2017 [49] examined the effect of low-dose folic acid supplementation on sperm DNA methylation. This prospective, randomized, con-trolled trial included 10 patients receiving antioxidants supple-ments (272 mg ascorbic acid, 31 mg all-rac-A-tocopherol acetate and 400 Lg folic acid) daily for 90 days versus 9 subjects receiving a placebo. These subjects constituted the intervention study groups for short-term exposure to folic acid. Moreover, a third group of 8 patients (post-fortification cohort) reported being exposed to the recommended daily folate intakes for several years (long-term exposure to folic acid). Similar sperm DNA methylation levels in repeat regions were found between placebo and supple-mented individuals studied by mass spectrometry, RLGS, MCI and 450K arrays. To note that, the latter mentioned approaches are genome wide analyses of sperm DNA methylation to target a larger number of CpG sites across the genome. However, when compar-ing pre-fortification to post-fortification samples, no major differ-ences in sperm DNA methylation patterns were detected. Overall, short-term and long-term exposure to low-dose folic acid did not affect the sperm methylome.

3.2. Animal studies

Experimental animal studies were more expounding, and com-pelling evidence for the alteration in the sperm epigenome is pro-vided by studies in the mouse model.

Lambrot et al. 2013 [50], provided evidence that male mice maintained on a folic acid deficient diet in utero and throughout life (0.3 mg folic acid per kg versus 2 mg per kg in controls) pre-sented an altered sperm epigenome at specific gene regions and imprinted genes, located in promoters of genes and at specific CpG locations; implicated in development and in the function of a wide range of systems such as the central nervous system and digestive system. Interestingly, sperm from folate-deficient males showed differential methylation at sites associated to cancer and chronic human disease. Furthermore, histone modifications such as histone H3 lysine 4 (H3K4) and lysine 9 (H3K9) methylation had been severely affected which was highlighted by a significant reduction in methylation levels in folate deficient compared with folate sufficient sperm. Beyond that, follow-up of the offspring showed an altered gene expression in 380 placenta genes, and only 2 out of 10 studied genes were differentially methylated in folate-deficient males. Thus, this study highlights the effect of in utero

exposures to folate on the health of two generations: the fathers' reproductive health (F1) and the offspring (F2).

More recently, McPherson et al. 2016 [51] proposed that pater-nal antioxidants and vitamins food ''fortification" in mice can nor-malize sperm oxidative stress and methylome, and more advancely metabolic health alterations in offspring, all caused by an under-nutrition state. The experimental 40 mice were divided into 4 groups of 10 according to their nutritional state: control-diet, con-trol diet with micronutrients supplementation, restricted diet with vitamin and antioxidant supplementation for 17 weeks, and a restricted-diet with a 70% reduced diet from that of controls. The supplemented diet contained vitamin C, vitamin E, folate, lyco-pene, zinc, selenium and green tea extract. Diet restricted founder males showed significant increase in sperm oxidative damage, reduction in sperm global methylation. As for offspring outcomes, a delay in preimplantation embryo development, reduction in female offspring, reduction in postnatal weights and alteration in adult adiposity were noticed. All these aforementioned alterations were subsequently normalized in the micronutrients supple-mented group. Beyond that, diet restriction also profoundly influ-enced gene expression in pancreas of offspring in a sex specific manner with partial normalization by vitamin and antioxidant supplementation.

3.3. Conclusions

Recent evidence from animal and human studies explored the epigenome after antioxidants supplementation, invokes the so-called ''fortification" of the epigenome. However, very few studies in humans analyzed the sperm epigenome; specifically global methylation profile; following antioxidants supplementation. Some studies have showed favorable effects, while others have failed to show an interventional effect.

However, the promising experimental studies on mice were congruent and supported the notion that epigenetic marks in sper-matogenesis are dynamic and can be modulated by nutritional exposure. Moreover, it was proposed that the sperm epigenome transfers a so-called epigenomic map to the offspring which influ-ences their development.

4. Overall conclusions and prospects

Antioxidants have been widely assessed for their protective effects on sperm genome, and DNA integrity. More importantly, a careful review of the literature showed that the overwhelming majority of studies support this concept. Although the majority of studies have demonstrated convincing antigenotoxic effect of antioxidants in both animals and humans, controversial results have surfaced. However, we could not yield conclusive results from all studies owing to some limitations of these studies, namely clinical trials.

Nevertheless, most of the hypotheses raised about the environ-mental dietary shaping of the sperm epigenome to date, are based on animal experiments. Few attempts have been made to evaluate only the sperm methylome based on in vivo trials.

Noteworthy, there was a lack in all evidence exposed to propose mechanisms that could clearly explain the antigenotoxic effect of antioxidants through in vitro and in vivo studies.

Thus, more properly-designed in vivo trials may disclose addi-tional understanding and knowledge, to better understand the underlying mechanisms and to reach a definitive conclusion prior to consumption. However, this lack of strong clinical evidence on the beneficial effects of antioxidant and micronutrients supplemen-tation on sperm genome and epigenome, should not deter us from more basic and clinical research on this issue. Hopefully, further epigenome-wide studies focusing on the prenatal environment, to evaluate developmental influences and the role of antioxidant ther-apy, will serve as a promising route for embodying the possibility of ''normalization" and restoration of some offspring health cues.

Conflict of interest

The authors declare that there is no conflict of interest.

References

[1] A. Lenzi, M. Picardo, L. Gandini, F. Dondero, Lipids of the sperm plasma membrane: from polyunsaturated fatty acids considered as markers of sperm function to possible scavenger therapy, Hum. Reprod. Update 2 (1996) 246– 256.

[2] G. Barroso, M. Morshedi, S. Oehninger, Analysis of DNA fragmentation, plasma membrane translocation of phosphatidylserine and oxidative stress in human spermatozoa, Hum. Reprod. 15 (2000) 1338–1344.

[3] F.B. Choucair, E.G. Rachkidi, G.C. Raad, E.M. Saliba, N.S. Zeidan, R.A. Jounblat, I.F. A. Jaoude, M.M. Hazzouri, High level of DNA fragmentation in sperm of Lebanese infertile men using Sperm Chromatin Dispersion test, Middle East Fert. Soc. J. 21 (2016) 269–276.

[4] F.F. Pasqualotto, R.K. Sharma, D.R. Nelson, A.J. Thomas, A. Agarwal, Relationship between oxidative stress, semen characteristics, and clinical diagnosis in men undergoing infertility investigation, Fert. Steril. 73 (2000) 459–464.

[5] P.-C. Hsu, C.-C. Hsu, Y.L. Guo, Hydrogen peroxide induces premature acrosome reaction in rat sperm and reduces their penetration of the zona pellucida, Toxicology 139 (1999) 93–101.

[6] R.J. Aitken, G.N. De Iuliis, R.I. McLachlan, Biological and clinical significance of DNA damage in the male germ line, Int. J. Androl. 32 (2009) 46–56.

[7] S.M. Boaz, K. Dominguez, J.A. Shaman, W.S. Ward, Mouse spermatozoa contain a nuclease that is activated by pretreatment with EGTA and subsequent calcium incubation, J. Cell. Biochem. 103 (2008) 1636–1645.

[8] I. Har-Vardi, R. Mali, M. Breietman, Y. Sonin, S. Albotiano, E. Levitas, G. Potashnik, E. Priel, DNA topoisomerases I and II in human mature sperm cells: characterization and unique properties, Hum. Reprod. 22 (2007) 2183–2189.

[9] J. MacLeod, The role of oxygen in the metabolism and motility of human spermatozoa, Am. J. Physiol.-Legacy Content 138 (1943) 512–518.

[10] C.G. Fraga, P.A. Motchnik, M.K. Shigenaga, H.J. Helbock, R.A. Jacob, B.N. Ames, Ascorbic acid protects against endogenous oxidative DNA damage in human sperm, Proc. Natl. Acad. Sci. 88 (1991) 11003–11006.

[11] O. Tunc, K. Tremellen, Oxidative DNA damage impairs global sperm DNA methylation in infertile men, J. Assist. Reprod. Genet. 26 (2009) 537–544.

[12] D. Rodenhiser, M. Mann, Epigenetics and human disease: translating basic biology into clinical applications, Can. Med. Assoc. J. 174 (2006) 341–348.

[13] S.S. Hammoud, D.A. Nix, H. Zhang, J. Purwar, D.T. Carrell, B.R. Cairns, Distinctive chromatin in human sperm packages genes for embryo development, Nature 460 (2009) 473.

[14] M. Sönmez, G. Türk, A. Yüce, The effect of ascorbic acid supplementation on sperm quality, lipid peroxidation and testosterone levels of male Wistar rats, Theriogenology 63 (2005) 2063–2072.

[15] U.R. Acharya, M. Mishra, J. Patro, M.K. Panda, Effect of vitamins C and E on spermatogenesis in mice exposed to cadmium, Reprod. Toxicol. 25 (2008) 84–88.

[16] A. Michael, C. Alexopoulos, E. Pontiki, D. Hadjipavlou-Litina, P. Saratsis, C. Boscos, Effect of antioxidant supplementation on semen quality and reactive oxygen species of frozen-thawed canine spermatozoa, Theriogenology 68 (2007) 204–212.

[17] M. Yousef, G. Abdallah, K. Kamel, Effect of ascorbic acid and vitamin E supplementation on semen quality and biochemical parameters of male rabbits, Anim. Reprod. Sci. 76 (2003) 99–111.

[18] T. Gliozzi, L. Zaniboni, A. Maldjian, F. Luzi, L. Maertens, S. Cerolini, Quality and lipid composition of spermatozoa in rabbits fed DHA and vitamin E rich diets, Theriogenology 71 (2009) 910–919.

[19] E. Greco, M. Iacobelli, L. Rienzi, F. Ubaldi, S. Ferrero, J. Tesarik, Reduction of the incidence of sperm DNA fragmentation by oral antioxidant treatment, J. Androl. 26 (2005) 349–353.

[20] O. Tunc, J. Thompson, K. Tremellen, Improvement in sperm DNA quality using an oral antioxidant therapy, Reprod. Biomed. Online 18 (2009) 761–768.

[21] Y.J. Ménézo, A. Hazout, G. Panteix, F. Robert, J. Rollet, P. Cohen-Bacrie, F. Chapuis, P. Clément, M. Benkhalifa, Antioxidants to reduce sperm DNA fragmentation: an unexpected adverse effect, Reprod. Biomed. Online 14 (2007) 418–421.

[22] E. Amar, D. Cornet, M. Cohen, Y. Ménézo, Treatment for high levels of sperm DNA fragmentation and nuclear de condensation: sequential treatment with a potent antioxidant followed by stimulation of the one-carbon cycle vs one-carbon cycle back-up alone, Aust. J. Reprod. Med. Infert. 2 (2015) 1006.

[23] J. Gual-Frau, C. Abad, M.J. Amengual, N. Hannaoui, M.A. Checa, J. Ribas-Maynou, I. Lozano, A. Nikolaou, J. Benet, A. García-Peiró, Oral antioxidant treatment partly improves integrity of human sperm DNA in infertile grade I varicocele patients, Hum. Fert. 18 (2015) 225–229.

[24] S. Young, B. Eskenazi, F. Marchetti, G. Block, A. Wyrobek, The association of folate, zinc and antioxidant intake with sperm aneuploidy in healthy non-smoking men, Hum. Reprod. 23 (2008) 1014–1022.

[25] A. Omu, M. Al-Azemi, E. Kehinde, J. Anim, M. Oriowo, T. Mathew, Indications of the mechanisms involved in improved sperm parameters by zinc therapy, Med. Principles Pract. 17 (2008) 108–116.

[26] G. Tirabassi, A. Vignini, L. Tiano, E. Buldreghini, F. Brugè, S. Silvestri, P. Orlando, A. D'Aniello, L. Mazzanti, A. Lenzi, Protective effects of coenzyme Q10 and aspartic acid on oxidative stress and DNA damage in subjects affected by idiopathic asthenozoospermia, Endocrine 49 (2015) 549–552.

[27] P. Piomboni, L. Gambera, F. Serafini, G. Campanella, G. Morgante, V. De Leo, Sperm quality improvement after natural anti-oxidant treatment of asthenoteratospermic men with leukocytospermia, Asian J. Androl. 10 (2008) 201–206.

[28] C. Abad, M. Amengual, J. Gosálvez, K. Coward, N. Hannaoui, J. Benet, A. García-Peiró, J. Prats, Effects of oral antioxidant treatment upon the dynamics of human sperm DNA fragmentation and subpopulations of sperm with highly degraded DNA, Andrologia 45 (2013) 211–216.

[29] C.M. Hughes, S. Lewis, V.J. McKelvey-Martin, W. Thompson, The effects of antioxidant supplementation during Percoll preparation on human sperm DNA integrity, Hum. Reprod. (Oxford, England) 13 (1998) 1240–1247.

[30] E.T. Donnelly, N. McClure, S.E. Lewis, The effect of ascorbate and A-tocopherol supplementation in vitro on DNA integrity and hydrogen peroxide-induced DNA damage in human spermatozoa, Mutagenesis 14 (1999) 505–512.

[31] E.T. Donnelly, N. Mcclure, S.E. Lewis, Glutathione and hypotaurine in vitro: effects on human sperm motility, DNA integrity and production of reactive oxygen species, Mutagenesis 15 (2000) 61–68.

[32] H. Chi, J. Kim, C. Ryu, J. Lee, J. Park, D. Chung, S. Choi, M. Kim, E. Chun, S. Roh, Protective effect of antioxidant supplementation in sperm-preparation medium against oxidative stress in human spermatozoa, Hum. Reprod. 23 (2008) 1023–1028.

[33] J. Sierens, J. Hartley, M. Campbell, A. Leathem, J. Woodside, In vitro isoflavone supplementation reduces hydrogen peroxide-induced DNA damage in sperm, Teratogen. Carcinogen. Mutagen. 22 (2002) 227–234.

[34] A. Moubasher, A. El Din, M. Ali, W. El-sherif, H. Gaber, Catalase improves motility, vitality and DNA integrity of cryopreserved human spermatozoa, Andrologia 45 (2013) 135–139.

[35] A.P. Kotdawala, S. Kumar, S.R. Salian, P. Thankachan, K. Govindraj, P. Kumar, G. Kalthur, S.K. Adiga, Addition of zinc to human ejaculate prior to cryopreservation prevents freeze-thaw-induced DNA damage and preserves sperm function, J. Assist. Reprod. Genet. 29 (2012) 1447–1453.

[36] J. Wu, S. Wu, Y. Xie, Z. Wang, R. Wu, J. Cai, X. Luo, S. Huang, L. You, Zinc protects sperm from being damaged by reactive oxygen species in assisted reproduction techniques, Reprod. Biomed. Online 30 (2015) 334–339.

[37] W. Zhang, F. Li, H. Cao, C. Li, C. Du, L. Yao, H. Mao, W. Lin, Protective effects of l-carnitine on astheno-and normozoospermic human semen samples during cryopreservation, Zygote 24 (2016) 293–300.

[38] J.C. Martinez-Soto, J. De Dioshourcade, A. Gutiérrez-Adán, J.L. Landeras, J. Gadea, Effect of genistein supplementation of thawing medium on characteristics of frozen human spermatozoa, Asian J. Androl. 12 (2010) 431.

[39] C.S. Branco, M.E. Garcez, F.F. Pasqualotto, B. Erdtman, M. Salvador, Resveratrol and ascorbic acid prevent DNA damage induced by cryopreservation in human semen, Cryobiology 60 (2010) 235–237.

[40] N. Zribi, N.F. Chakroun, F.B. Abdallah, H. Elleuch, A. Sellami, J. Gargouri, T. Rebai, F. Fakhfakh, L.A. Keskes, Effect of freezing–thawing process and quercetin on human sperm survival and DNA integrity, Cryobiology 65 (2012) 326–331.

[41] Z. Li, Q. Lin, R. Liu, W. Xiao, W. Liu, Protective effects of ascorbate and catalase on human spermatozoa during cryopreservation, J. Androl. 31 (2010) 437–444.

[42] G. Kalthur, S. Raj, A. Thiyagarajan, S. Kumar, P. Kumar, S.K. Adiga, Vitamin E supplementation in semen-freezing medium improves the motility and protects sperm from freeze-thaw-induced DNA damage, Fert. Steril. 95 (2011) 1149–1151.

[43] S. Kedechi, N. Zribi, N. Louati, H. Menif, A. Sellami, S. Lassoued, R. Ben Mansour, L. Keskes, T. Rebai, N. Chakroun, Antioxidant effect of hydroxytyrosol on human sperm quality during in vitro incubation, Andrologia 49 (2017).

[44] L. Zhang, R. Diao, Y. Duan, T. Yi, Z. Cai, In vitro antioxidant effect of curcumin on human sperm quality in leucocytospermia, Andrologia (2017).

[45] B.H. Maleki, B. Tartibian, F.C. Mooren, F.Y. Nezhad, M. Yaseri, Saffron supplementation ameliorates oxidative damage to sperm DNA following a 16-week low-to-intensive cycling training in male road cyclists, J. Funct. Foods 21 (2016) 153–166.

[46] M. Dattilo, C. Ettore, Y. Ménézo, Improvement of gamete quality by stimulating and feeding the endogenous antioxidant system: mechanisms, clinical results, insights on gene-environment interactions and the role of diet, J. Assist. Reprod. Genet. 33 (2016) 1633–1648.

[47] M. Aarabi, M.C. San Gabriel, D. Chan, N.A. Behan, M. Caron, T. Pastinen, G. Bourque, A.J. MacFarlane, A. Zini, J. Trasler, High-dose folic acid supplementation alters the human sperm methylome and is influenced by the MTHFR C677T polymorphism, Hum. Mol. Genet. 24 (2015) 6301–6313.

[48] C.A. Sheppard, A Candidate Genetic Risk Factor for Vascular Disease: A Common Mutation in Methylenetetrahydrofolate Reductase, 1995.

[49] D. Chan, S. McGraw, K. Klein, L. Wallock, C. Konermann, C. Plass, P. Chan, B. Robaire, R. Jacob, C. Greenwood, Stability of the human sperm DNA methylome to folic acid fortification and short-term supplementation, Hum. Reprod. 32 (2017) 272–283.

[50] R. Lambrot, C. Xu, S. Saint-Phar, G. Chountalos, T. Cohen, M. Paquet, M. Suderman, M. Hallett, S. Kimmins, Low paternal dietary folate alters the mouse sperm epigenome and is associated with negative pregnancy outcomes, Nat. Commun. 4 (2013) 2889.

[51] N.O. McPherson, T. Fullston, W.X. Kang, L.Y. Sandeman, M.A. Corbett, J.A. Owens, M. Lane, Paternal under-nutrition programs metabolic syndrome in offspring which can be reversed by antioxidant/vitamin food fortification in fathers, Sci. Rep. 6 (2016).