

## **Correlation between the aberrant human testicular germ-cell gene expression and disruption of spermatogenesis leading to male infertility**

Arka Baksi<sup>1</sup>, Ruchi Jain<sup>1</sup>, Ravi Manjithaya<sup>2</sup>, Vasana S S<sup>3</sup>, Paturu Kondaiah<sup>1</sup> and Rajan R. Dighe<sup>1\*</sup>

<sup>1</sup>Department of Molecular Reproduction, Development and Genetics, Indian Institute of Science, Bangalore, India

<sup>2</sup>Molecular Biology and Genetics Unit, Jawaharlal Nehru Centre for Advanced Scientific Research, Bangalore, India

<sup>3</sup>Manipal Fertility, Bangalore, India

\* Corresponding Author

### **Corresponding Author Contact:**

#### **Prof. Rajan R Dighe**

Department of Molecular Reproduction, Development and Genetics

Indian Institute of Science

Bangalore 560012, India

Phone 91 80 2293 3261/2660

Fax: 91 80 2360 0999

E-mail: [rdighe@alumni.iisc.ac.in](mailto:rdighe@alumni.iisc.ac.in); [digherr@gmail.com](mailto:digherr@gmail.com)

### **Abstract:**

Spermatogenesis is characterized by sequential gene-expression at precise stages in progression of differentiation of the germ cells. Any alteration in expression of the critical genes is responsible for arrest of spermatogenesis associated with infertility. In spite of advances in the differential gene expression accompanying spermatogenesis, the corresponding regulatory mechanisms and their correlation to human infertility have not been clearly established. This study aims to identify the gene expression pattern of the human testicular germ cells from the patients either with obstructive azoospermia with complete intra-testicular spermatogenesis or non-obstructive azoospermia with spermatogenesis arrested at different stages and correlate the same to infertility. The

testicular transcriptomes of 3 OA and 8 NOA patients and pooled testicular RNA (commercial source) were analyzed for their differential gene expression to identify potential regulators of spermatogenesis and the results were further validated in all of the 44 patients clinically diagnosed with azoospermia undergoing sperm retrieval surgery over the study period and 4 control samples included in this study. Analyses of the differential transcriptome led to identification of genes enriched in a specific testicular cell type and subsequently, several regulators of the diploid- double-diploid- haploid transitions in the human spermatogenesis were identified. Perturbations in the expression of these genes were identified as the potential causes of the spermatogenic arrest seen in azoospermia and thus the potential mediators of human male infertility. Another interesting observation was the increased autophagy in the testes of patients with non-obstructive azoospermia. The present study suggests that the regulation of the diploid-double-diploid-haploid transition is multigenic with the tandem alteration of several genes resulting in infertility. In conclusion, this study identified some of the genetic regulators controlling spermatogenesis using comparative transcriptome analyses of testicular tissues from azoospermic individuals and showed how alterations in several genes results in disruption of spermatogenesis and subsequent infertility. This study also provides interesting insights into the gene expression patterns of the Indian population that were not available earlier.

## Introduction

Spermatogenesis is an extraordinary process of cell differentiation in which the testicular germ cells undergo series of mitotic divisions, followed by a meiotic division and transformation of the haploid cells into spermatozoa which eventually attain motility. Regulation of spermatogenesis occurs at different levels, the first being the extrinsic level wherein the gonadotropins and testosterone regulate gene expression in the germ cells sustaining their survival and progression of differentiation<sup>1</sup> and blockades in their action results in apoptotic death of the germ cells<sup>2-4</sup>. The second level is the interactive regulation that involves communications between the somatic cells such as the Sertoli cells and the germ cells. The third level of regulation is the intrinsic gene expression associated with the germ cell differentiation<sup>1,5</sup>. These differentiation events are characterized by sequential gene-expression wherein the specific genes are turned ‘on’ or ‘off’ at precise stages facilitating the progression of differentiation of the germ cells. Any alteration in the expression of the genes appears to be responsible for arrest of spermatogenesis associated with infertility<sup>6-12</sup>. However, the critical genes required for progression of spermatogenesis and the genes whose altered expression leads to its arrest have not been clearly identified<sup>13</sup>. Identification of such genes is essential for understanding the molecular defects in arrested spermatogenesis and possible treatment of the consequent infertility. However, such a task has proved to be difficult due paucity of the human testicular tissue from the normal, fertile individuals, as well as, the infertile males with arrested spermatogenesis. The other contributing factor is the complex organization of spermatogenesis and the unique environment present in the seminiferous

tubules that has not been replicated *in vitro*. It is also not possible to culture the germ cells and make them undergo differentiation *in vitro*.

The infertile patients with the spermatogenesis arrested at any stage of differentiation provide an interesting paradigm for investigating alterations in the testicular transcriptome that could be correlated to spermatogenic arrest. In the present study, we investigated gene-expression changes in the testes of the infertile patients with non-obstructive azoospermia having spermatogenic arrest either at the diploid or double-diploid stages and compared it to the germ cell gene expression patterns of the obstructive azoospermic (OA) patients who exhibited complete spermatogenesis. This allowed identification of genes which could potentially be responsible for the arrest. The study demonstrates that alterations in the expression of several regulators of spermatogenic progression are the potential causes of germ-cell arrest leading to infertility.

## **Material and Methods**

### **Patient Samples**

The testicular biopsies of infertile patients were obtained from Manipal Ankur Fertility Clinic, Bangalore with appropriate clearances from the Institutional Review Board of Manipal Ankur Fertility Clinic and the Human Ethics Committee of the Indian Institute of Science, Bangalore. All experiments were performed in accordance with the relevant guidelines and regulations with prior approval.

## **RNA isolation, RT-PCR and real-time PCR**

The total RNA from the testicular tissue was isolated using TRI reagent (Sigma-Aldrich, USA) as per the manufacturer's protocol. cDNA was generated from 2µg of RNA using the Revert Aid First Strand cDNA Synthesis kit (Thermo Fisher Scientific, USA). qRT-PCR was performed using the ABI Prism 7900HT sequence detection system. cDNA equivalent to 20ng RNA was used for qRT-PCR analysis using Dynamo SYBER green 2X mix (Finnzymes, Finland) along with gene-specific primer pairs (Table S1). RPL35 expression was used for normalization. Differential expression of genes was determined using the following formula:

$$\delta Ct = Ct_{gene} - Ct_{RPL35}$$

$$\delta\delta Ct [Log_2 Fold Change (L_2FC)] = \delta Ct_{arrested\ population} - \delta Ct_{normal\ population}$$

$$Fold\ Change\ (FC) = 2^{-\delta\delta Ct}$$

## **Transcriptome Profiling using Microarray**

The quality and quantity of RNA was analyzed using Qubit RNA Assay Kit in Qubit fluorometer (Thermo Fisher Scientific, USA) followed by Agilent 2100 Bioanalyzer (Agilent Technologies, USA). Samples with RNA-integrity values >6 were chosen for microarray and biotin-labeled complementary-RNA (c-RNA) was synthesized from each using the TargetAmp-Nano Labeling Kit for Illumina Expression BeadChip (Illumina Inc. USA). The cRNA was hybridized onto an Illumina HumanHT-12v4 chip (Illumina Inc. USA) and signal detection was performed by incubating with Cy3-streptavidin conjugate in HiScanSQ System (Illumina Inc. USA). The total human testicular RNA pooled from 5 individuals of Asian origin (Clontech, USA) was used as the control. The

preprocessing and identification of differentially expressed genes are provided in the supplementary section.

### **Data Availability**

The complete microarray data are available at GEO database, accession No GSE10886

### **Gene-Enrichment analysis**

The enrichment of the population specific gene sets to GO terms was analyzed using the ‘Enrichr’ tool (<http://amp.pharm.mssm.edu/Enrichr/>)<sup>14</sup>. Enrichment of biological processes was tested for the haploid, diploid and the double-diploid-enriched genes. Pathway analysis was performed by uploading to KEGG (<http://www.genome.jp/kegg/>)<sup>15</sup>.

### **Transcription factor Binding site analysis and Network construction**

The testicular cell population-enriched genes were searched for the putative transcription factor binding sites using Mapper2<sup>16</sup>. Network maps were created through String database<sup>17</sup> for each population-enriched gene set (diploid, double-diploid and haploid) and their respective differentially-expressed transcription factors (only the putative transcription factors that were differentially expressed in the microarray analysis were selected).

### **Immunohistochemistry and TUNEL assay**

Immunohistochemistry was performed as described in Pant *et al*<sup>18</sup>. The primary antibodies used were Anti-EGR2 (abcam), Anti-RFX2 (Sigma-Aldrich), Anti-CDKN1A/p21 (Cell Signaling Technology) and Anti-LC3B (Sigma-Aldrich). In-situ

Apoptosis Detection Kit (Abcam) was used to check for the apoptotic cells in the tissue sections according to the Manufacturer's protocol. Additional details are provided in the supplementary section.

### **Image Capture:**

All images were captured using the Zeiss microscope and processed by Zeiss AxioCam 4.3 software.

### **Statistical Analysis:**

Unpaired t-test with Welch's correction was performed to assess statistical significance for the expression of the selected genes in qRT-PCR. GraphPad Prism V software was used to analyze data for statistical significance. p value  $\leq 0.01$ , 0.001, 0.0001 are represented as \* \*\* \* respectively.

## **Results**

### **Patients**

The classification the infertile patients analyzed in this study has been described in details in Baksi et al<sup>19</sup>. Forty four infertile patients were first classified into OA or NOA based on clinical inputs and further classified into three distinct groups based their germ cell flow cytometry coupled with marker gene analysis as shown in the previous publication from the laboratory. In addition, 4 patients, previously known to be fertile, but undergoing surgical procedures unrelated to infertility were considered as the controls. Six patients exhibited only diploid cells indicating diploid/pre-meiotic arrest (Group III or Group D), while 24 patients showed presence of the diploid and double-diploid cells indicating double-diploid/meiotic arrest (Group II or Group DT). Fourteen patients with OA showed presence of the diploid, double-diploid and haploid cells (Group I or Group

DTH) indicating complete spermatogenesis. The control, fertile individuals also exhibited full spermatogenesis with all three germ cell populations<sup>19</sup>.

### **Transcriptome profiling**

The total RNA samples from Groups D (3), DT (5) and DTH (3) and the total testicular RNA (N) (pooled from 8 individuals, Clontech, USA) (Figure S1) were subjected to transcriptome analysis by Microarray and the genes expressed differentially between the groups were identified (Table S2). Seventeen genes representing the highest differential-expression and the testicular cell-specific markers were validated by qRT-PCR (Figure S2). The gene-expression profiles were further analyzed to identify the testicular cell-specific enriched genes (Figure S3 and S4).

### **Gene set enrichment and pathway analysis**

The population-enriched gene sets were analyzed for enrichment to GO terms. The diploid-enriched genes showed predominance for early developmental processes active in the spermatogonia<sup>3,4,20,21</sup>. The double-diploid-enriched genes were enriched for the biological processes such as inflammation, stress response and movement of cellular components. The haploid-enriched genes showed enrichment for the male-gamete formation, spermatid development, ciliary movement and sperm motility (Table S3). Pathway analysis through KEGG also provided the pathways regulated by the population-enriched genes (Table S4). These results indicated that the annotation of the cell population-enriched gene sets was in accordance with their function.



## **Transcription Factors and Network Maps for population-enriched genes**

Network maps were constructed using the population-enriched genes and their TFs (Figures S5-7). The genes having  $\geq 4$  connections were identified from these networks of which 15, 13 and 58 genes were unique to the diploid, double-diploid and haploid subsets respectively, while 12 genes were common to all three groups (Figure S8).

## **Identification of the population-enriched genes possibly responsible for arrest of spermatogenesis and validation at the transcript levels:**

The diploid-enriched genes were mainly the regulators of cell-cycle and proliferation, while the double-diploid and haploid-enriched genes were potential regulators of the double-diploid to haploid transition<sup>22-33</sup> (Figure S9A-C). As these observed alterations could potentially cause the spermatogenic arrest, these genes have been referred to as ‘crucial’ population-enriched genes. Subsequently, these crucial genes were validated at the transcript level in all 4 control and 44 patients’ samples and their expression patterns corroborated the microarray analysis (Figure 1A-C).

## **Validation of selected genes using immunohistochemistry:**

The hubs of the diploid and double-diploid-enriched altered genes, CDKN1A (Figure S9A) and EGR2 (Figure S9B), and RFX2 (Figure S9C), whose aberration could have wide-ranging effects on the cell-cycle were validated at the protein level using specific antibodies. Staining of CDKN1A was seen only in Group D, whereas Groups DT and DHT exhibited poor staining (Figure 2A). The nuclear staining for EGR2 was seen in the Group DT with no staining in Group DTH and only a faint signal seen in Group D (Figure 2B). RFX2 staining was observed in the double-diploid and haploid cells of

Group DTH, while no expression was seen in Group DT spermatocytes. Group D sections also did not stain for RFX2 due to absence of the double-diploid cells (Figure 2C). Thus, the differences seen at the transcript level were further confirmed at the protein level.

### **Status of apoptosis and autophagy in the azoospermic patients:**

The status of apoptosis and autophagy was ascertained in the patient samples as possible clearance mechanisms for the arrested germ-cells. No appreciable difference was seen in the status of apoptosis in the patients belonging to all the groups (Table-S5 and Figure 3A). For autophagy, no difference was observed at the transcript level of the autophagy-related genes (Table S6). However, a significant increase in LC3B staining was observed in the Groups D and DT while the Group DTH showed a faint staining (Figure 3B).

### **Discussion:**

Arrest of spermatogenesis constitutes approximately 10% of infertile male patients and devising novel therapeutic strategies to treat such infertility has been an extremely challenging task. Lack of understanding of the molecular mechanisms that govern the spermatogenic differentiation and inability to manipulate germ cell differentiation *in vitro* have been the major impediments in developing such strategies. The testicular tissues from the infertile patients with spermatogenesis arrested at different stages of differentiation provide an interesting model for understanding for understanding the molecular mechanisms operative in the human testis. In this study, an extensive analysis of testicular gene expression patterns of the NOA patients led to identification of some of the critical genes that play roles in progression of spermatogenesis and aberrant

expression of these genes may be responsible for arrest of germ cell differentiation. Presented below is the discussion on some of the genes with aberrant expression which could be responsible for arrest of spermatogenesis based on the information available in the literature with both human and rodent models.

CDKN1A, an inhibitor of cyclins<sup>34</sup>, exhibited higher transcript levels in all Group D patients and formed the hub of the network of the diploid-enriched genes (Figure S9A). Over-expression of CDKN1A is associated with inhibition of cell-proliferation and the cell-cycle arrest (mediated by nuclear CDKN1A)<sup>22,35</sup>. In mice, CDKN1A maintains the spermatogonial stem cell proliferation as its overexpression inhibits germ-line stem cell proliferation<sup>36</sup>. The CDKN1A-null mice did not undergo G1-arrest in response to induced DNA damage<sup>37,38</sup> identifying it as a mediator of the G1 check-point. Interestingly, CDKN1A deficiency does not affect the stem cell number or self-renewal mechanism<sup>36</sup>, but the undifferentiated spermatogonia exhibit up-regulation of CDKN1A upon DNA damage<sup>39</sup>. The CDKN1A knockout mice showed 35% increase in their testicular weight due to increased proliferation of the germ cells. In the ATM-deficient mice spermatogenesis is halted at the diploid stage via CDKN1A mediated inhibition of germ-cell proliferation<sup>39</sup>. In human testis, CDKN1A levels were virtually undetectable in the germ cells<sup>40</sup> while the non-dividing Leydig cells showed a moderate expression<sup>41</sup>. In the present study, all 6 Group D patients showed an elevated expression of CDKN1A compared to DTH and N, both at the transcript (Figure 1A) and protein levels (Figure 2A, panel 3) indicating that there is a DNA damage induced proliferation check in the diploid cells preventing their entry into meiosis. Expression of Cyclin E involved in G<sub>1</sub>-S transition is depleted in the Group D patients, further supporting this conclusion. In

addition, in human testicular cancer CDKN1A is also present in the cytoplasm where it acts as an anti-apoptotic factor <sup>42</sup>. Thus, higher expression of CDKN1A can potentially cause cell-cycle arrest while preventing apoptosis of the arrested spermatogonia and consequently impeding their entry into meiosis.

KLF4 is generally associated with the post-mitotic differentiation of the epithelial cells in both humans and mice <sup>43-46</sup>. In this study, its expression was found to be much higher in the Group D patients. In humans, KLF4 is induced in response to cellular stress and up-regulates the transcription of CDKN1A and subsequently arrests cell cycle by repressing the cyclins <sup>47</sup>. The over-expression of KLF4 (predominantly expressed in the non-dividing cells <sup>48</sup>) in the actively dividing cells results in cell-cycle arrest with corresponding up-regulation of CDKN1A. A similar occurrence is seen in the azoospermic men with a pre-meiotic arrest (Figure 2A). KLF4 action mediated by CDKN1A induces proliferation arrest, while lack of CDKN1A results in tumorigenesis <sup>23,48</sup>. Overexpression of both these genes, as seen in Group D patients, probably results in arrest of spermatogonial proliferation and differentiation halting the entire differentiation at the spermatogonial stage. This induction of CDKN1A is independent of p53 which corroborates our observation that there was no change in p53 transcript levels across the different patient groups.

Another up-regulated gene in the Group D with a potential role in spermatogenic arrest was GADD45A. In mice, this gene is expressed in the primordial germ cells (PGC) during cell fate determination, causing cell-cycle arrest <sup>49</sup>. The PGCs arrested at the G2-M phase migrate to the gonadal ridge where subsequent down-regulation of GADD45A allows further proliferation confirming its negative correlation to cell proliferation <sup>50</sup>.

GADD45A is also involved in the murine adult stem cell differentiation<sup>51</sup> and is up-regulated in the cells exposed to genotoxic stress causing cell-cycle arrest<sup>24</sup>. Thus, the higher level of GADD45A in Group D (Figure 1A) is probably one of the factors responsible for spermatogonial cell-cycle arrest. Correspondingly, the levels of Cyclin-B are also depleted in these patients. Further, GADD45A-null cells show higher sensitivity to apoptosis under conditions of stress<sup>52</sup> indicating its role as an anti-apoptotic factor and supports lack of apoptosis seen in the diploid arrested patients (Figure 3A).

There was a very significant increase in FOS and JUN expression in the Group D patients. Such an increase in FOS expression causes loss of mouse germ cells<sup>25</sup> with decreased number of spermatocytes and spermatids<sup>53</sup> which correlates with the decreased testicular cell numbers as observed in this study.

The network of diploid-enriched genes provides a list of genes that act in tandem to regulate the diploid cell population and alteration in their expression halts diploid to double-diploid transition (Figure. S9A). From our observations, both cell-cycle check points (G1-S: CDKN1A, KLF4 and G2-M: GADD45A) appear to be operational in patients with a pre-meiotic arrest. Further, the altered expression of KLF4, CDKN1A, GADD45A and FOS (important regulators of mitotic cell-cycle) could potentially be responsible for the arrest. Besides the genes validated in this study, microarray analysis and subsequent network mapping revealed that the components of the cell-cycle such as CDC25B<sup>26</sup> and FOXM1<sup>54</sup> (crucial for expression of cell-cycle genes in G2 phase; depletion causes cell-cycle arrest) were also down regulated in the arrested patients. CDKN1A and FOXM1 are known to act in opposition to each other with CDKN1A overexpression leading to low levels of FOXM1 and disruption of the normal cell-

cycle<sup>27</sup>. Taken together, these genes represent the potential causes of the pre-meiotic arrest.

The double-diploid-enriched genes identified in the study showed enrichment for regulation of the inflammatory response and several related pathways such as toll like receptor, TNF and MAPK signaling (Tables S3 and S4). The testes are primarily immune-privileged organs and any inflammation altering this immune-suppressed microenvironment leads to disruption of spermatogenesis<sup>28</sup>. Secretion of inflammatory cytokines by circulating macrophages invading the testis in orchitis, affect the integrity of the Sertoli cell tight junctions and facilitate shedding of the immature germ cells into the lumen of the tubule causing infertility<sup>55-57</sup>. CCL3 is known to attract and induce extravasation of immune cells within the testis<sup>58</sup> and its elevated levels (Figure 1B) directly correlates with the accumulation of immature double-diploid cells in the tubular lumen of Group DT patients. Among the interleukins, IL1A is normally expressed in the testis regulating the function of the blood-testis barrier<sup>59</sup> while the inflammatory cytokines show minimal expression<sup>29</sup>. In rats, infertility induced by ionizing radiations (IR) is mediated by IL1B (not expressed in the normal rat testis). Further, IL1B injection alone was able to mimic the effects of IR mediated loss of fertility<sup>60</sup>. IL1B also showed increased expression in the Group DT patients indicating that there is an acute inflammation in these patients (Figure 1B). Thus, the observed overexpression of the pro-inflammatory genes such as CCL3 (chemokine ligand 3), IL1B and IL8 (interleukins), could potentially alter the micro-environment within the testis and be responsible for the meiotic arrest of spermatocytes in the Group DT patients.

EGR2, an early gene expressed in the Spermatogonium-A, is known to be responsible for maintenance of spermatogonia<sup>61</sup> by controlling mitosis<sup>62</sup> and its expression must be down regulated at the onset of meiosis<sup>30</sup>. As seen in the Figure 2B, expression of EGR2 persists in the nucleus of the spermatocytes in Group DT where its continued expression might provide conflicting differentiation cues to the cell resulting in disruption of meiosis and subsequent arrest. Additionally, the transcripts of FOXM1, JUND (genes involved in mitotic cell-cycle control), comprising the double-diploid network, show no change in the Group DT as compared to the Group DTH. This further validates the classification of genes as enriched in a particular testicular cell population.

Expression of majority of the haploid-enriched genes was not seen in the Groups D and DT (Figure. S9C) and can be attributed to the absence of haploid cells. To understand the role of the haploid-enriched genes in the context of spermatocyte differentiation, only those genes reported to be expressed in spermatocytes and persisting till spermatid stage were selected from the haploid-enriched genes for further analyses. These selected genes showed enrichment for pathways involved in sperm energy-metabolism and architecture (Table S4) confirming the strategy for population-enriched (D, T and H) classification employed.

One of the most interesting genes identified in this study is RFX2, a member of the RFX family of transcription factors. Expression of RFX2 is testis-specific, particularly in the meiotic and post-meiotic cells<sup>63</sup> and controls transcription of various pachytene spermatocyte-specific genes including histoneH1<sup>31</sup>. RFX2 binds to MYBL1 and broadens its action in the control of meiosis<sup>64</sup> and MYBL1 knockout mice exhibit arrested spermatogenesis at the pachytene stage<sup>53</sup>. Expression of RFX2 is down-

regulated in both Groups DT and D (Figure 1C). As shown in Figure 2C, expression of RFX2 was seen in the double-diploid cells of Group DTH, but not that of the Group DT suggesting its crucial role in the progression of meiosis. The complete absence of any double-diploid cells in Group D corresponds to the observed lack of staining. The role of RFX2 as the master regulator of spermiogenesis<sup>65</sup>, controlling the genes required for ciliogenesis<sup>66</sup> is well established. However, its role in the control of meiosis is less studied and requires investigation. The present study suggests a greater role for RFX2 in regulating the double-diploid-haploid transition which is in agreement with Kistler *et al* who suggested a similar role for RFX2 in rats<sup>64</sup>.

CST8 belonging to the cystatin superfamily of cysteine protease-inhibitors is mainly expressed in the testis and epididymis<sup>67</sup>. CST8-null mice show a defect in the Sertoli-germ-cell adhesion and premature release of germ cells into the lumen with the severity increasing with age<sup>32</sup>. In the histological sections of the DT patients, a large number of the double-diploid cells were present in the lumen of the tubules with a disruption of proper tubular architecture similar to that seen in CST8-null mice (unpublished data) indicating a similar role for CST8 in human spermatogenesis. Lack of expression of CST8 (Figure 1C) might be responsible for loss of germ-cell integrity, disruption of the Sertoli cell-germ cell interactions and shedding of spermatocytes in the tubular lumen of the Group DT patients.

The alterations in expression of genes such as DDIT3, ATFs, GADD45A, and CDKN1A in the Group DT patients strongly suggest that DNA damage is extant in the germ cells. In this context, the loss of GGN (gametogenetin) in the Group DT (Figure 1C) indicates its probable contribution towards the genotoxic stress. GGN, a mouse pachytene



spermatocyte specific gene, along with POG (proliferation of germ cells), controls meiosis by regulating the DNA double strand break repair<sup>68</sup>. GGN-null mice are embryonic lethal while GGN<sup>+/-</sup> mice are infertile with an arrest at the pachytene stage as the DNA double strand breaks during homologous recombination are not repaired in these mice<sup>33</sup>. The Group DT patients show an arrest in spermatogenesis at the meiotic stage and a down-regulation of GGN at the transcript level indicating that the hampered DNA damage repair and induction of DNA damage inducible cell-cycle arrest together cause the meiotic arrest.

Intriguingly, the androgen receptor gene, which is expressed in the Sertoli cells, was observed to be differentially expressed across all the networks indicating that disruption of germ-cell development has a direct bearing on the Sertoli cells.

As seen in Figure 3, autophagy was up-regulated in the Groups D and DT patients as compared to the Group DTH. This observation indicated that autophagy was a probable mechanism for clearance of the arrested cells in NOA and the regulation seems to be at the protein level. Interestingly, CDKN1A is reported to induce autophagy<sup>69</sup> and increased autophagy could be a consequence of its observed high levels. Our data suggests a role for autophagy in NOA and male infertility which requires further investigation.

This study suggests that the major disruption in azoospermia is in the transition from one phase to another. In Group D patients, the arrest seems to be a result of alterations of genes such as CDKN1A, KLF4, GADD45A and FOS regulating mitosis, consequently, the diploid to double-diploid transition. In Group DT patients, alterations of genes such as RFX2, GGN and CST8 that were down-regulated and persistence of EGR2 could be

the potential causes of meiotic arrest. The present study suggests that the regulation of the diploid - double-diploid - haploid transition are multigenic with the tandem alteration of several genes resulting in infertility (Figure 4). The exact point of arrest in humans within a phase still remains to be identified.

In conclusion, this study provides glimpses into understanding of the possible mechanisms involved in the regulation of human spermatogenesis and their relation to infertility. The genes that have been analyzed using both human and rodent data in this study are restricted to those with most striking changes in expression. There is a possibility that the genes with relatively smaller differences in expression levels may have profound effects on spermatogenesis and hence need to be analyzed further. The microarray analysis associated with this study provides a huge resource for analyzing such genes and further analysis may lead to identification of other potential targets for infertility treatment. More detailed analysis of genes responsible for arrest of spermatogenesis can also lead to identification of potential targets for male contraception.

#### **Additional Information:**

#### **Author Contributions:**

The research described here was designed and executed by AB and RRD. SSV identified the infertile patients and provided the biopsies. RM was involved in autophagy experiments. RJ and PK analyzed the microarray data and RRD supervised the entire project. AB, PK and RRD prepared the manuscript. All authors read and approved the manuscript.

#### **Acknowledgements:**

The authors thank Prof. M.R.S. Rao, JNCASR for scientific discussions and Prof. N.V. Joshi, CES, IISc for help with the statistics.

## Funding:

The work described here was funded by DBT-IISc partnership program and J.C. Bose fellowship to RRD

## Conflict of interest:

The authors declare no conflict of interest

## References:

- 1 Eddy, E. M. Male germ cell gene expression. *Recent progress in hormone research* **57**, 103-128 (2002).
- 2 Marathe, C. K., Shetty, J. & Dighe, R. R. Selective immunoneutralization of luteinizing hormone results in the apoptotic cell death of pachytene spermatocytes and spermatids in the rat testis. *Endocrine* **3**, 705-709, doi:10.1007/BF03000201 (1995).
- 3 Pareek, T. K., Joshi, A. R., Sanyal, A. & Dighe, R. R. Insights into male germ cell apoptosis due to depletion of gonadotropins caused by GnRH antagonists. *Apoptosis* **12**, 1085-1100 (2007).
- 4 Shetty, J., Marathe, G. & Dighe, R. Specific immunoneutralization of FSH leads to apoptotic cell death of the pachytene spermatocytes and spermatogonial cells in the rat. *Endocrinology* **137**, 2179-2182 (1996).
- 5 Matzuk, M. M. & Lamb, D. J. Genetic dissection of mammalian fertility pathways. *Translocations* **45**, 46XY (2002).
- 6 Krausz, C., Escamilla, A. R. & Chianese, C. Genetics of male infertility: from research to clinic. *Reproduction* **150**, R159-R174 (2015).
- 7 Kleene, K. C. Patterns of translational regulation in the mammalian testis. *Molecular reproduction and development* **43**, 268-281 (1996).
- 8 Kierszenbaum, A. *et al.* Patterns, mechanisms, and functions of translation regulation in mammalian spermatogenic cells. *Cytogenetic and genome research* **103**, 217-224 (2003).
- 9 Sassone-Corsi, P. Unique chromatin remodeling and transcriptional regulation in spermatogenesis. *Science* **296**, 2176-2178 (2002).
- 10 Kleene, K. C. A possible meiotic function of the peculiar patterns of gene expression in mammalian spermatogenic cells. *Mechanisms of development* **106**, 3-23 (2001).
- 11 Sassone-Corsi, P. Transcriptional Checkpoints Determining the Fate of Male Germ Cells. *Cell* **88**, 163-166 (1997).
- 12 Zorrilla, M. & Yatsenko, A. N. The genetics of infertility: current status of the field. *Current genetic medicine reports* **1** (2013).
- 13 Neto, F. T. L., Bach, P. V., Najari, B. B., Li, P. S. & Goldstein, M. Genetics of male infertility. *Current urology reports* **17**, 70 (2016).
- 14 Chen, E. Y. *et al.* Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC bioinformatics* **14**, 128 (2013).
- 15 Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y. & Morishima, K. KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic acids research* **45**, D353-D361 (2017).

- 16 Marinescu, V. D., Kohane, I. S. & Riva, A. MAPPER: a search engine for the computational identification of putative transcription factor binding sites in multiple genomes. *BMC bioinformatics* **6**, 1 (2005).
- 17 Szklarczyk, D. *et al.* STRING v10: protein–protein interaction networks, integrated over the tree of life. *Nucleic acids research*, gku1003 (2014).
- 18 Pant, I., Rao, S. G. & Kondaiah, P. Role of areca nut induced JNK/ATF2/Jun axis in the activation of TGF- $\beta$  pathway in precancerous Oral Submucous Fibrosis. *Scientific reports* **6**, 34314 (2016).
- 19 Baksi, A., Vasan, S. & Dighe, R. R. DNA Flow cytometric analysis of the human testicular tissues to investigate the status of spermatogenesis in azoospermic patients. *Scientific Reports (Nature Publisher Group)* **8**, 1-9 (2018).
- 20 Xiao, X., Mruk, D. D., Wong, C. K. & Cheng, C. Y. Germ cell transport across the seminiferous epithelium during spermatogenesis. *Physiology* **29**, 286-298 (2014).
- 21 Linde, R. *et al.* Reversible inhibition of testicular steroidogenesis and spermatogenesis by a potent gonadotropin-releasing hormone agonist in normal men: an approach toward the development of a male contraceptive. *New England Journal of Medicine* **305**, 663-667 (1981).
- 22 Abukhdeir, A. M. & Park, B. H. P21 and p27: roles in carcinogenesis and drug resistance. *Expert reviews in molecular medicine* **10**, e19 (2008).
- 23 Chen, X. *et al.* Krüppel-like factor 4 (gut-enriched Krüppel-like factor) inhibits cell proliferation by blocking G1/S progression of the cell cycle. *Journal of Biological Chemistry* **276**, 30423-30428 (2001).
- 24 Johnen, H. *et al.* Gadd45g is essential for primary sex determination, male fertility and testis development. *PLoS One* **8**, e58751 (2013).
- 25 Kaur, P. & Bansal, M. P. Upregulation of AP1 by tertiary butyl hydroperoxide induced oxidative stress and subsequent effect on spermatogenesis in mice testis. *Molecular and cellular biochemistry* **308**, 177-181 (2008).
- 26 Lammer, C. *et al.* The cdc25B phosphatase is essential for the G2/M phase transition in human cells. *Journal of cell science* **111**, 2445-2453 (1998).
- 27 Laoukili, J. *et al.* FoxM1 is required for execution of the mitotic programme and chromosome stability. *Nature cell biology* **7**, 126-136 (2005).
- 28 Zhao, S., Zhu, W., Xue, S. & Han, D. Testicular defense systems: immune privilege and innate immunity. *Cellular & molecular immunology* **11**, 428-437 (2014).
- 29 Guazzone, V. A., Jacobo, P., Theas, M. S. & Lustig, L. Cytokines and chemokines in testicular inflammation: a brief review. *Microscopy research and technique* **72**, 620-628 (2009).
- 30 Pellegrini, M. *et al.* ATRA and KL promote differentiation toward the meiotic program of male germ cells. *Cell Cycle* **7**, 3878-3888 (2008).
- 31 Horvath, G. C., Kistler, W. S. & Kistler, M. K. RFX2 is a potential transcriptional regulatory factor for histone H1t and other genes expressed during the meiotic phase of spermatogenesis. *Biology of reproduction* **71**, 1551-1559 (2004).
- 32 Parent, A. D., Cornwall, G. A., Liu, L. Y., Smith, C. E. & Hermo, L. Alterations in the Testis and Epididymis Associated With Loss of Function of the Cystatin-Related Epididymal Spermatogenic (CRES) Protein. *Journal of andrology* **32**, 444-463 (2011).
- 33 Jamsai, D. *et al.* Loss of GGN leads to pre-implantation embryonic lethality and compromised male meiotic DNA double strand break repair in the mouse. *PLoS one* **8**, e56955 (2013).
- 34 Xiong, Y. *et al.* p21 is a universal inhibitor of cyclin kinases. *Ried* **49**, 370 (1991).

- 35 Harper, J. W., Adami, G. R., Wei, N., Keyomarsi, K. & Elledge, S. J. The p21 Cdk-interacting protein Cip1 is a potent inhibitor of G1 cyclin-dependent kinases. *Cell* **75**, 805-816 (1993).
- 36 Kanatsu-Shinohara, M., Takashima, S. & Shinohara, T. Transmission distortion by loss of p21 or p27 cyclin-dependent kinase inhibitors following competitive spermatogonial transplantation. *Proceedings of the National Academy of Sciences* **107**, 6210-6215 (2010).
- 37 Brugarolas, J. *et al.* Radiation-induced cell cycle arrest compromised by p21 deficiency. *Nature* **377**, 552-557 (1995).
- 38 Zhou, B. P. *et al.* Cytoplasmic localization of p21Cip1/WAF1 by Akt-induced phosphorylation in HER-2/neu-overexpressing cells. *Nature cell biology* **3**, 245-252 (2001).
- 39 Takubo, K. *et al.* Stem cell defects in ATM-deficient undifferentiated spermatogonia through DNA damage-induced cell-cycle arrest. *Cell stem cell* **2**, 170-182 (2008).
- 40 Oldereid, N., Angelis, P. D., Wiger, R. & Clausen, O. Expression of Bcl-2 family proteins and spontaneous apoptosis in normal human testis. *Molecular human reproduction* **7**, 403-408 (2001).
- 41 Uhlén, M. *et al.* Tissue-based map of the human proteome. *Science* **347**, 1260419 (2015).
- 42 Koster, R. *et al.* Cytoplasmic p21 expression levels determine cisplatin resistance in human testicular cancer. *The Journal of clinical investigation* **120**, 3594-3605 (2010).
- 43 Shields, J. M., Christy, R. J. & Yang, V. W. Identification and characterization of a gene encoding a gut-enriched Krüppel-like factor expressed during growth arrest. *Journal of Biological Chemistry* **271**, 20009-20017 (1996).
- 44 Garrett-Sinha, L. A., Eberspaecher, H., Seldin, M. F. & de Crombrughe, B. A gene for a novel zinc-finger protein expressed in differentiated epithelial cells and transiently in certain mesenchymal cells. *Journal of Biological Chemistry* **271**, 31384-31390 (1996).
- 45 Segre, J. A., Bauer, C. & Fuchs, E. Klf4 is a transcription factor required for establishing the barrier function of the skin. *Nature genetics* **22**, 356-360 (1999).
- 46 Behr, R. & Kaestner, K. H. Developmental and cell type-specific expression of the zinc finger transcription factor Krüppel-like factor 4 (Klf4) in postnatal mouse testis. *Mechanisms of development* **115**, 167-169 (2002).
- 47 Wei, D., Kanai, M., Huang, S. & Xie, K. Emerging role of KLF4 in human gastrointestinal cancer. *Carcinogenesis* **27**, 23-31 (2005).
- 48 Glynne, R., Ghandour, G., Rayner, J., Mack, D. H. & Goodnow, C. C. B-lymphocyte quiescence, tolerance and activation as viewed by global gene expression profiling on microarrays. *Immunological reviews* **176**, 216-246 (2000).
- 49 Kurimoto, K. *et al.* Complex genome-wide transcription dynamics orchestrated by Blimp1 for the specification of the germ cell lineage in mice. *Genes & development* **22**, 1617-1635 (2008).
- 50 Mochizuki, K. & Matsui, Y. Epigenetic profiles in primordial germ cells: global modulation and fine tuning of the epigenome for acquisition of totipotency. *Development, growth & differentiation* **52**, 517-525 (2010).
- 51 Zhang, R.-p., Shao, J.-z. & Xiang, L.-x. GADD45A protein plays an essential role in active DNA demethylation during terminal osteogenic differentiation of adipose-derived mesenchymal stem cells. *Journal of Biological Chemistry* **286**, 41083-41094 (2011).

- 52 Yan, S.-J. *et al.* Deficiency in TR4 nuclear receptor abrogates Gadd45a expression and increases cytotoxicity induced by ionizing radiation. *Cellular & molecular biology letters* **17**, 309 (2012).
- 53 Toscani, A. *et al.* Arrest of spermatogenesis and defective breast development in mice lacking A-myb. *Nature* **386**, 713 (1997).
- 54 Wierstra, I. & Alves, J. FOXM1, a typical proliferation-associated transcription factor. *Biological chemistry* **388**, 1257-1274 (2007).
- 55 Frungieri, M. B. *et al.* Number, distribution pattern, and identification of macrophages in the testes of infertile men. *Fertility and sterility* **78**, 298-306 (2002).
- 56 Theas, M. *et al.* Tumour necrosis factor- $\alpha$  released by testicular macrophages induces apoptosis of germ cells in autoimmune orchitis. *Human reproduction* **23**, 1865-1872 (2008).
- 57 Rival, C. *et al.* Functional and phenotypic characteristics of testicular macrophages in experimental autoimmune orchitis. *The Journal of pathology* **215**, 108-117 (2008).
- 58 Guazzone, V. A., Jacobo, P., Denduchis, B. & Lustig, L. Expression of cell adhesion molecules, chemokines and chemokine receptors involved in leukocyte traffic in rats undergoing autoimmune orchitis. *Reproduction* **143**, 651-662 (2012).
- 59 Sarkar, O., Mathur, P. P., Cheng, C. Y. & Mruk, D. D. Interleukin 1 alpha (IL1A) is a novel regulator of the blood-testis barrier in the rat. *Biology of reproduction* **78**, 445-454 (2008).
- 60 Bergh, A. & Söder, O. Interleukin-1 $\beta$ , but not interleukin-1 $\alpha$ , induces acute inflammation-like changes in the testicular microcirculation of adult rats. *Journal of reproductive immunology* **17**, 155-165 (1990).
- 61 Rossi, P. *et al.* Transcriptome analysis of differentiating spermatogonia stimulated with kit ligand. *Gene Expression Patterns* **8**, 58-70 (2008).
- 62 Joseph, L. J. *et al.* Molecular cloning, sequencing, and mapping of EGR2, a human early growth response gene encoding a protein with "zinc-binding finger" structure. *Proceedings of the National Academy of Sciences* **85**, 7164-7168 (1988).
- 63 Wolfe, S. A., van Wert, J. & Grimes, S. R. Transcription factor RFX2 is abundant in rat testis and enriched in nuclei of primary spermatocytes where it appears to be required for transcription of the testis-specific histone H1t gene. *Journal of cellular biochemistry* **99**, 735-746 (2006).
- 64 Horvath, G. C., Kistler, M. K. & Kistler, W. S. RFX2 is a candidate downstream amplifier of A-MYB regulation in mouse spermatogenesis. *BMC developmental biology* **9**, 63 (2009).
- 65 Kistler, W. S. *et al.* RFX2 is a major transcriptional regulator of spermiogenesis. *PLoS genetics* **11**, e1005368 (2015).
- 66 Wu, Y. *et al.* Transcription factor RFX2 is a key regulator of mouse spermiogenesis. *Scientific reports* **6**, 20435 (2016).
- 67 Cornwall, G., Orgebin-Crist, M. & Hann, S. The CRES gene: a unique testis-regulated gene related to the cystatin family is highly restricted in its expression to the proximal region of the mouse epididymis. *Molecular Endocrinology* **6**, 1653-1664 (1992).
- 68 Lu, B. & Bishop, C. E. Mouse GGN1 and GGN3, two germ cell-specific proteins from the single gene Ggn, interact with mouse POG and play a role in spermatogenesis. *Journal of Biological Chemistry* **278**, 16289-16296 (2003).
- 69 Capparelli, C. *et al.* CDK inhibitors (p16/p19/p21) induce senescence and autophagy in cancer-associated fibroblasts, "fueling" tumor growth via paracrine interactions, without an increase in neo-angiogenesis. *Cell cycle* **11**, 3599-3610 (2012).

## Figure Legends:

**Figure 1. Transcript levels of testicular cell population-enriched genes crucial for diploid to double-diploid to haploid transition across all patient samples.** The transcript levels of the crucial testicular cell population-enriched genes were ascertained in all the 44 patient samples (samples included or not-included in the microarray) and the control using specific primers in qRT-PCR and the significance was calculated using the unpaired t-test with Welch's correction. **A. Transcript levels of diploid-enriched genes crucial for diploid to double-diploid transition. B. Transcript levels of double-diploid-enriched genes crucial for double-diploid to haploid transition. C. Transcript levels of haploid-enriched genes crucial for double-diploid to haploid transition.** N: Pooled RNA control, Control: patients with proven fertility, DTH: obstructive azoospermia patients with complete intra-testicular spermatogenesis, DT: meiotic arrest patients with diploid and double-diploid cells in their testes, D: pre-meiotic arrest patients with only diploid cells in their testes;  $P > 0.05$ - ns;  $P \leq 0.05$ -\*;  $P \leq 0.01$ -\*\*;  
 $P \leq 0.001$ -\*\*\*

**Figure 2. Representative photomicrographs of expression of A. CDKN1A, B. EGR2 and C. RFX2 proteins in testicular tissue sections of Groups DTH, DT and D patients.** Expression of CDKN1A, EGR2 and RFX2 was ascertained at the protein level by immuno-histochemistry. **B. Inset:** magnified image showing nuclear staining of EGR2 in the Group DT sample. **C. Inset:** magnified image [200X] showing differences in expression of RFX2 in the spermatocytes present in Group DTH and DT patients. The spermatocytes or meiotic cells are marked with black arrows. DTH: obstructive azoospermia patients with complete intra-testicular spermatogenesis, DT: meiotic arrest

patients with diploid and double-diploid cells in their testes, D: pre-meiotic arrest patients with only diploid cells in their testes; Scale: 50 $\mu$ m, Magnification: 100X.

**Figure 3. Representative photomicrographs showing the status of A. Apoptosis and B. Autophagy in testicular tissue sections of patients from Groups DTH, DT and D.**

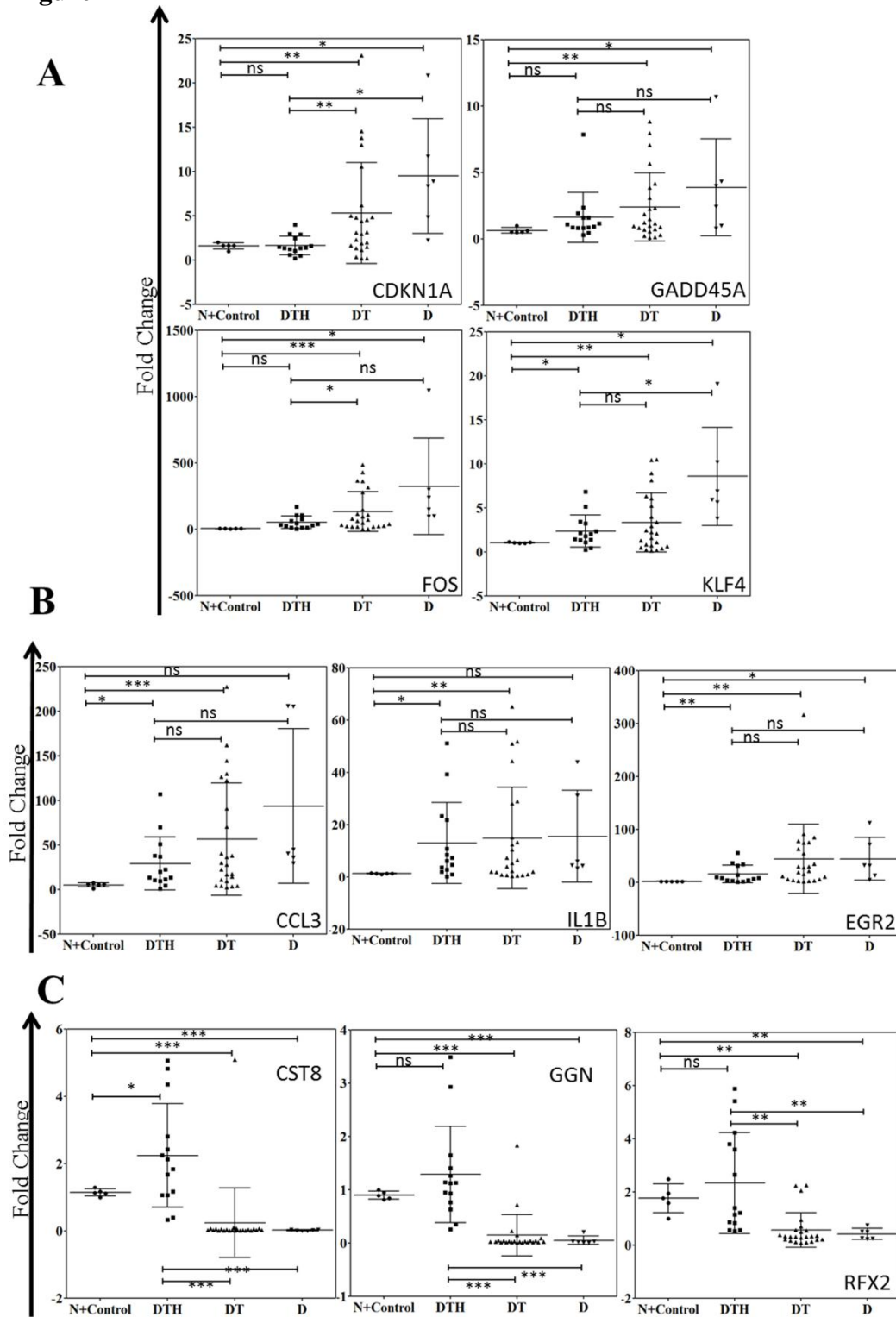
**A.** The status of apoptosis was ascertained by TUNEL assay. **B.** The status of autophagy was determined in the different groups of patients by immuno-histochemical staining for LC3B. DTH: obstructive azoospermia patients with complete intra-testicular spermatogenesis, DT: meiotic arrest patients with diploid and double-diploid cells in their testes, D: pre-meiotic arrest patients with only diploid cells in their testes; Scale: 50 $\mu$ m, Magnification: 100X.

**Figure 4. Schematic representation of the progression of human spermatogenesis and its relationship with the altered population-enriched genes identified from this study.**

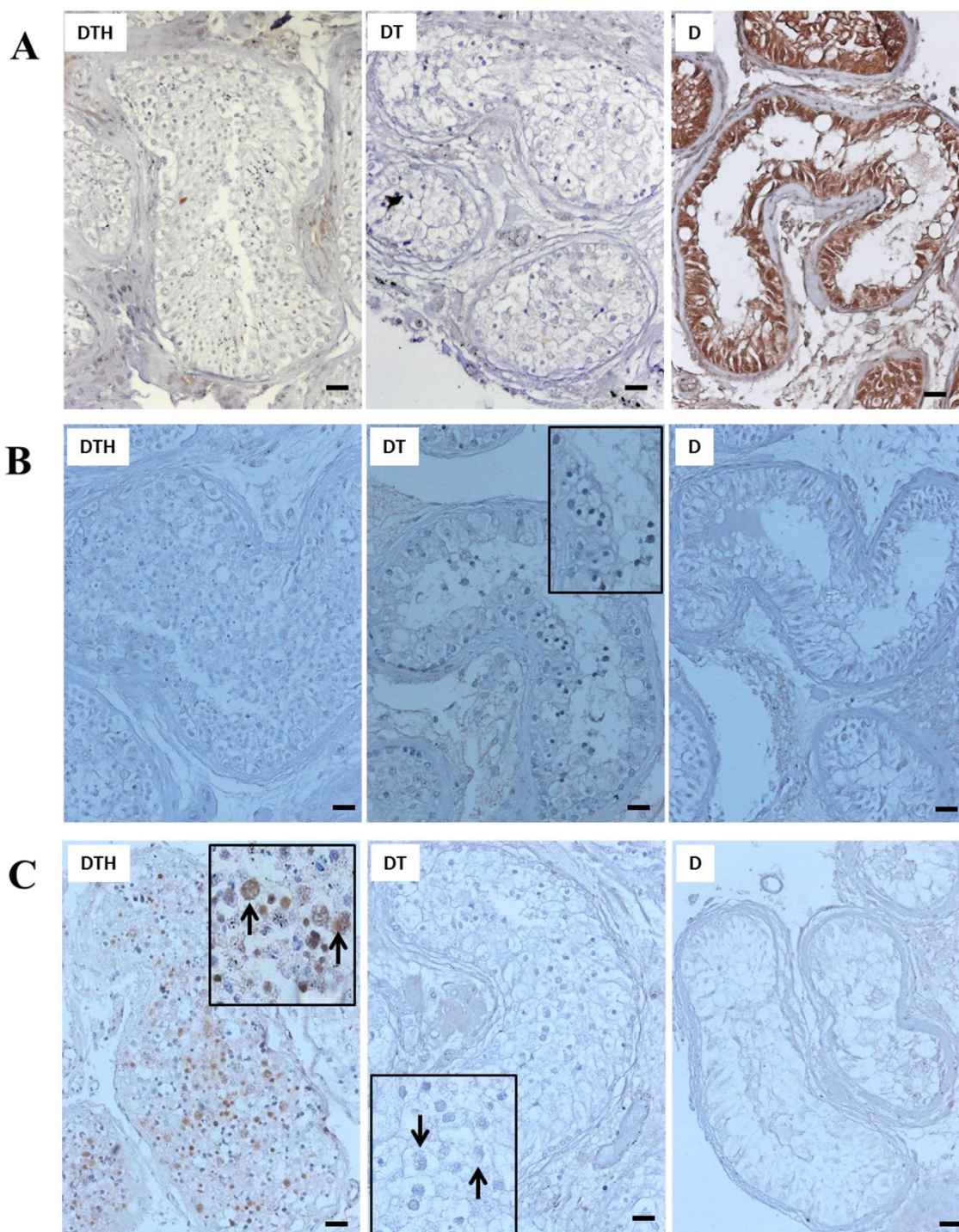
The phases of human spermatogenesis as well the points of transition between each phase are shown along with the altered genes identified from this study as potential causes of the spermatogenic arrest. Adark: dark A type Spermatogonia, Apale: pale A type Spermatogonia capable of self-renewal marked by arrow, B: Type B spermatogonia, P-L: pre-leptotene spermatocytes, Lep: leptotene spermatocytes, Z: zygotene spermatocytes, Pearly: early-pachytene spermatocytes, Pmid: mid-pachytene spermatocytes, Plate: late-pachytene spermatocytes, SS- secondary spermatocytes.



**Figure 1**

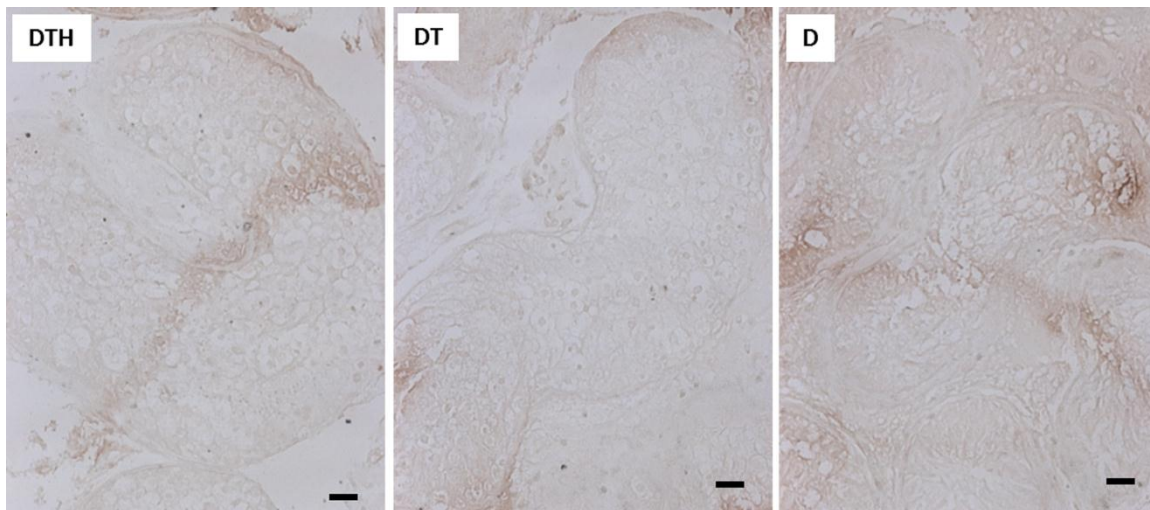


**Figure 2**

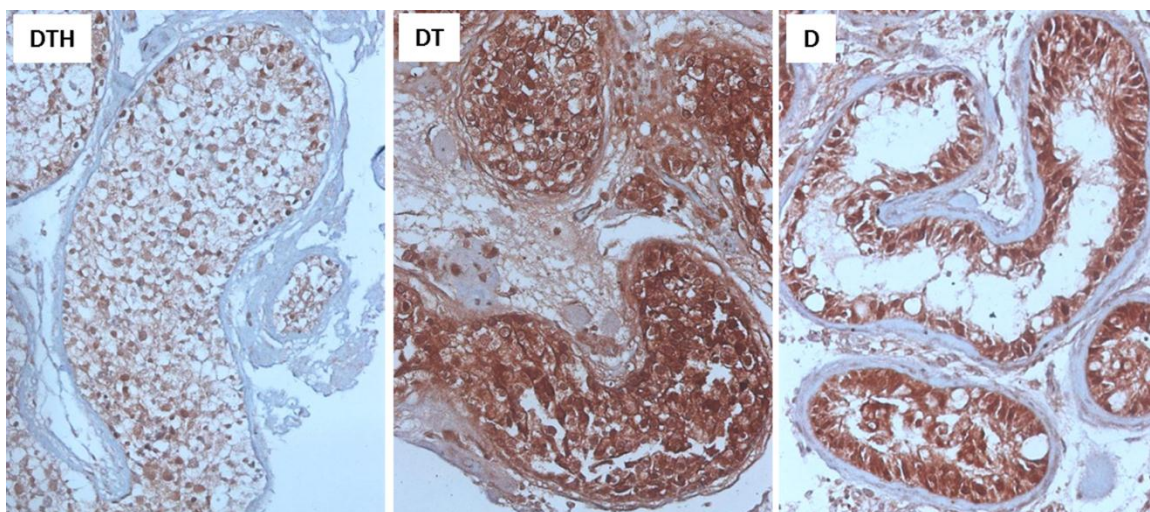


**Figure 3**

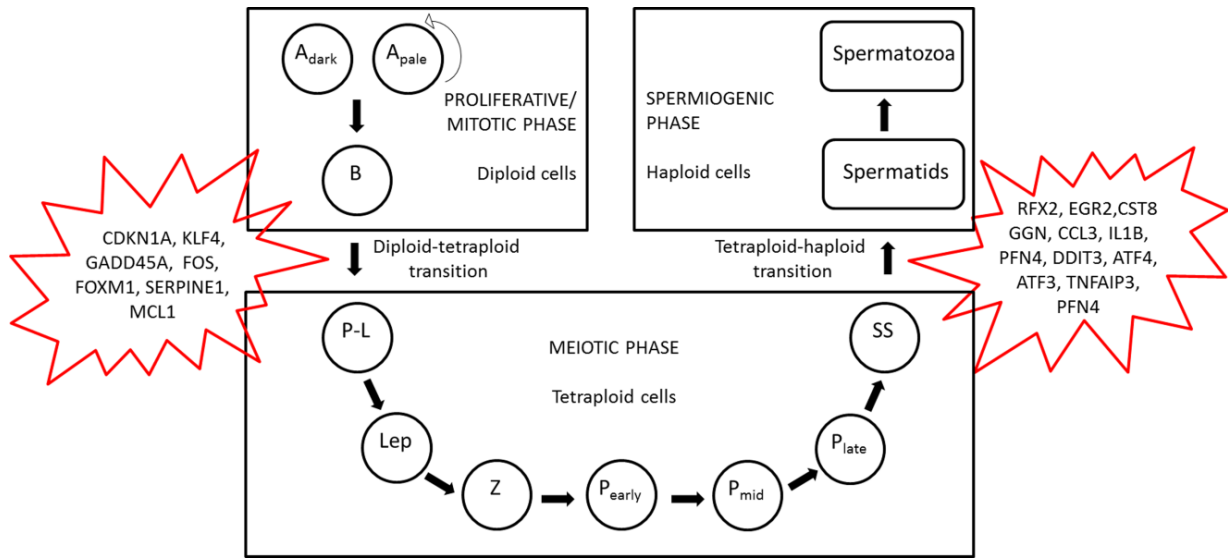
**A**



**B**



**Figure 4**



**Supplementary material for manuscript - ‘Correlation between the aberrant human testicular germ-cell gene expression and disruption of spermatogenesis leading to male infertility’**

**Table S 1 (Supplementary Table 1)**

**List of Primers used for real time PCR**

<b>Primer</b>	<b>5'&lt;-----Sequence-----&gt;3'</b>
PRM1 Forward	GCCAGGTACAGATGCTGTTCGCAG
PRM1 Reverse	TTAGTGTCTTCTCACTCTCGGTCTG
PRM2 Forward	TCTCGAAGGAGGCTGCAC
PRM2 Reverse	TCTGCATGTTCTCTTCCTGG
TNP1 Forward	TACCGTAAGGGCAACCTGAA
TNP1 Reverse	TCCCCTTCTGCTGTTCTTGT
ZBP2 Forward	CGAACGTGCGTCCTACTCTC
ZBP2 Reverse	CCATAAGTAGGTGGGGTCCA
DDX25 Forward	TGTCCTGGATGAAGCAGATG
DDX25 Reverse	TGCCACACAGAGTCCTCAA
USP2 Forward	CTACACGCTGTGGGAGACG
USP2 Reverse	GTTTCGAAGACCAGCCAGAC
ACRBP Forward	CGTGGAGAATACCTGCATCA
ACRBP Reverse	CAGAAGTCACAGAGGGCACA
GK2 Forward	TGGAGGAATGACCAACAACA
GK2 Reverse	CCAAACGCTTACTCCCTCTG
YBX1 Forward	GGGTGCAGGAGAACAAGGTA
YBX1 Reverse	TCTTCATTGCCGTCCTCTCT
JUN Forward	TAACAGTGGGTGCCAACTCA
JUN Reverse	TTTTCTCTCCGTCGCAACTT
JUNB Forward	TTAACAGGGAGGGGAAGAGG
JUNB Reverse	TTCCACAGTACGGTGCAGAG
IGFBP6 Forward	ATCCGCCCAAGGACGAC
IGFBP6 Reverse	GCCTGCTTGGGGTTTACTCT

GADD45A Forward	GAGAGCAGAAGACCGAAAGGA
GADD45A Reverse	CACAACACCACGTTATCGGG
MAP3K6 Forward	GGGAGCAGAGGAGCACATT
MAP3K6 Reverse	GGCTCTGGATGCCTTCTACA
TGFB3 Forward	AAATTGACATGATCCAGGG
TGFB3 Reverse	TTCTGCTCGGAATAGGTTGG
MYC Forward	CTCCTCCTCGTCGCAGTAGA
MYC Reverse	GCTGCTTAGACGCTGGATTT
VEGFB Forward	AAGTCCGGATGCAGATCCT
VEGFB Reverse	GGCTTCACAGCACTGTCCTT
RPL35 Forward	GAACCAAAGGGAGCACACAG
RPL35 Reverse	CAATGGCCTTAGCAGGAAGA
CDKN1A Forward	ACCGAGGCACTCAGAGGAG
CDKN1A Reverse	GCCATTAGCGCATCACAGT
KLF4 Forward	GCGGCAAACCTACACAAAG
KLF4 Reverse	CCCCGTGTGTTTACGGTAGT
FOS Forward	CTGTCTTCGGTGGACTCCTT
FOS Reverse	GTTGCACAAGCCACTGGAG
CCL3 Forward	TGCTCAGAATCATGCAGGTC
CCL3 Reverse	AAATTCTGTGGAATCTGCCG
IL1B Forward	GAGCTCGCCAGTGAAATGAT
IL1B Reverse	GGAGATTCGTAGCTGGATGC
EGR2 Forward	GTGACCATCTTTCCCAATGC
EGR2 Reverse	AGCAAAGCTGCTGGGATATG
CST8 Forward	CTGCGCCATTCAAGAAAACCT
CST8 Reverse	GGATGCCTCAAACACCATT
GGN Forward	ACTCCTGTCCCTGCAAGTGT
GGN Reverse	GAGAGCCGGACAGCTTGAT
RFX2 Forward	TACTTACCAACGGGGACCAA
RFX2 Reverse	GAGGGACGCAGAGCCAC

## **Extended Materials and Methods:**

### **Preprocessing and Identification of differentially-expressed genes**

Average signal intensities of all the genes were background-subtracted and Quantile normalization was applied between the arrays. The Linear model and empirical Bayes methods were performed to assess the differentially-expressed genes between the Normal and the patients' samples with arrested spermatogenesis using "Limma" (Linear Models for microarray data) package in R-3.3.2 (1). The Benjamini and Hochberg method was used to correct for multiple testing(2) within Limma. The genes with adjusted P-values <0.05 and fold change >1.5 were considered significantly regulated. All analyses were represented as 'Y'VS'X' where X is the arrested population (variable) and Y is the normal population (constant). (Table S1)

The differentially-expressed genes with fold changes >3 from the above analysis were further analyzed using 'Venny' ([bioinfogp.cnb.csic.es/tools/venny/index.html](http://bioinfogp.cnb.csic.es/tools/venny/index.html))(3) and their individual fold changes (from Microarray) were plotted manually for identifying the genes commonly and differentially-expressed across all sample groups.

### **Immunohistochemistry**

Paraffin sections (4  $\mu$ m) were embedded on silane-coated slides for Immunohistochemistry. Paraffin was removed and sections were rehydrated by passing through xylene (15 minutes), decreasing concentrations of ethanol (100%, 95%, and 70%; 5 minutes each) and double distilled water (5 minutes). Antigen-retrieval was carried out in citrate buffer (pH-6) by heating the sections in a microwave oven at 600 W for 20 minutes. The sections were cooled to room temperature, washed in 1X PBS and the intracellular peroxidase was quenched by methanol- $H_2O_2$  treatment for 20 minutes at room temperature. The sections were washed again with PBS

and non-specific binding was blocked by 5% skim milk and incubated with the primary antibodies (at recommended dilutions) overnight at 4 °C. The signal was developed using the PathSitu PolyExcel HRP/Dab Detection System kit. The sections were subsequently dehydrated and mounted in DPX. Sections incubated with only secondary antibody served as negative controls.

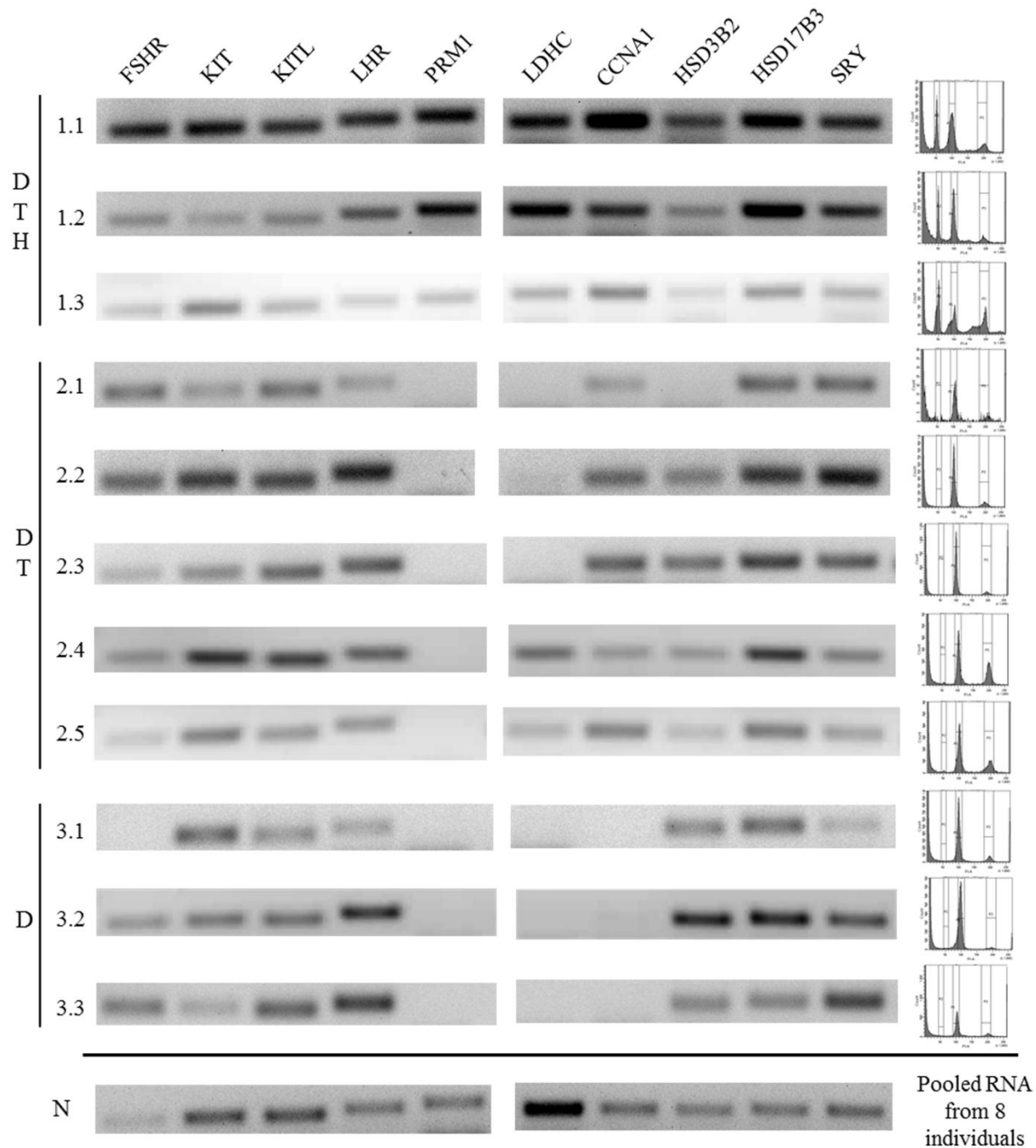
### **TUNEL assay:**

The slides were rehydrated in different gradients of alcohol and finally immersed in TBS. The sections were permeabilized using a Proteinase K solution followed by the quenching of the endogenous peroxidases by H<sub>2</sub>O<sub>2</sub>-methanol. The sections were labeled with a TdT labeling mix in a humidified chamber for 90 minutes. The reaction was stopped by adding the Stop buffer and the samples were washed in TBS. Streptavidin-HRP was added and the reaction developed using Dab solution. The slides were washed, air-dried and mounted in DPX post development. DNase treated section served as positive control whereas TdT untreated sections were served as negative control.

1. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic acids research* 2015:gkv007.
2. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the royal statistical society Series B (Methodological)* 1995:289-300.
3. Oliveros J. An interactive tool for comparing lists with venn diagrams. Available at: <http://bioinfogp.cnb.csic.es/tools/venny/index.html> Accessed on 15th January 2015.



**Figure S1 (Supplementary Figure 1)**



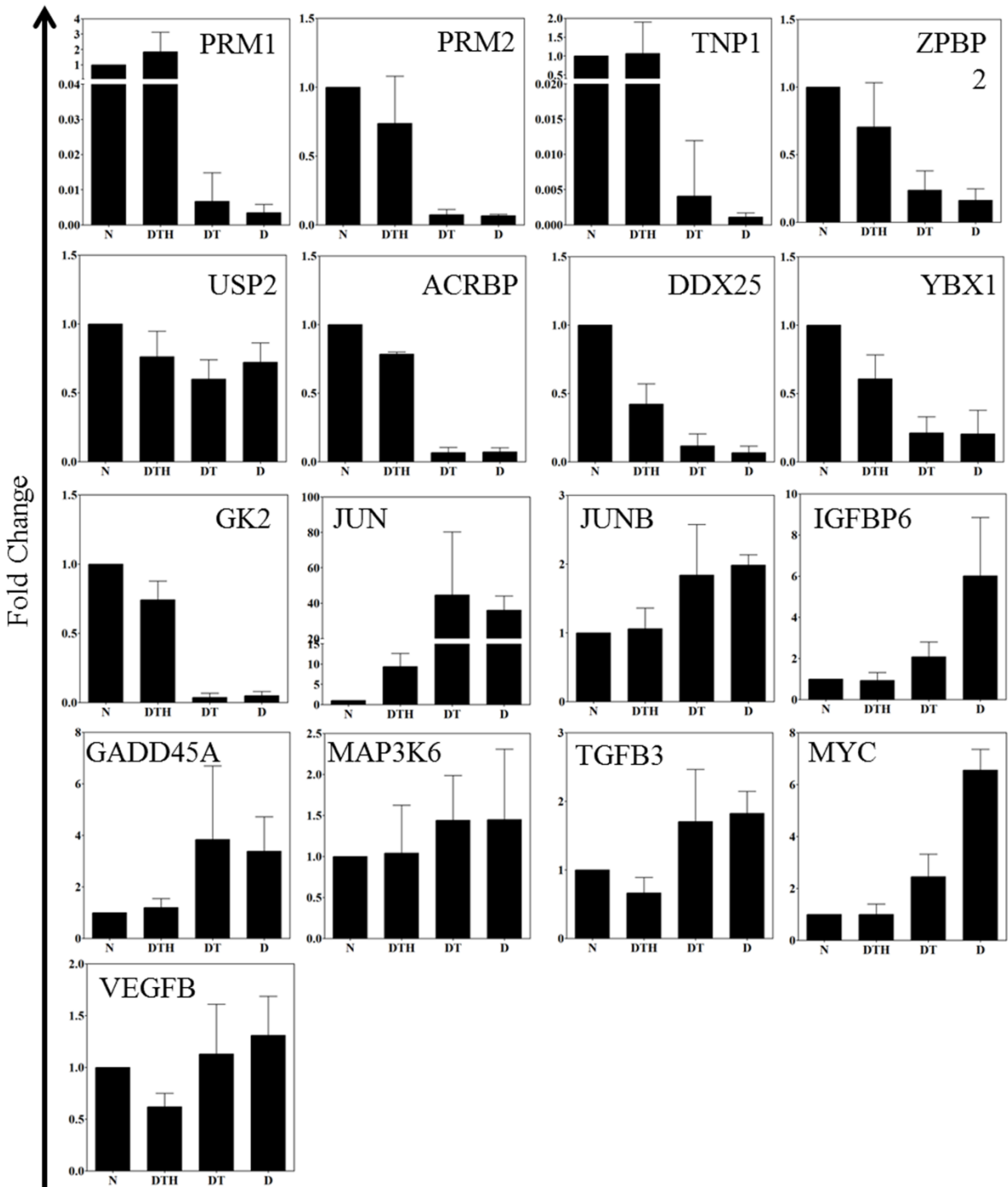
**Legend: Germ cell patterns of patient samples used for transcriptome analysis ascertained by flow cytometry and amplification of marker genes.** DTH: obstructive azoospermia patients with complete intra-testicular spermatogenesis, DT: meiotic arrest patients with diploid and double-diploid cells in their testes, D: pre-meiotic arrest patients with only diploid cells in their testes, N: Pooled RNA control.

**Table S 2 (Supplementary Table 2)**

<b>Analysis</b>	<b>Differentially expressed genes (p&lt;0.05)</b>	<b>Differentially expressed genes (Fold Change &gt; 1.5)</b>	<b>Differentially expressed genes (Fold Change &gt; 3)</b>
N VS DTH	1332	390	60
		High expression in N – 236	High expression in N - 29
		Low expression in N - 154	Low expression in N - 31
N VS DT	5764	3867	1216
		High expression in N – 2436	High expression in N – 1071
		Low expression in N – 1431	Low expression in N – 145
N VS D	8494	6473	1863
		High expression in N – 3236	High expression in N – 1323
		Low expression in N – 3237	Low expression in N – 540
DTH VS DT	10141	4910	1107
		High expression in DTH – 2692	High expression in DTH – 1070
		Low expression in DTH – 2218	Low expression in DTH – 37
DTH VS D	14049	7779	1803
		High expression in DTH – 3624	High expression in DTH – 1406
		Low expression in DTH – 4155	Low expression in DTH – 397
DT VS D	3682	1210	37
		High expression in DT – 480	High expression in DT – 12
		Low expression in DT – 730	Low expression in DT – 25

N: Pooled RNA control, DTH: obstructive azoospermia patients with complete intra-testicular spermatogenesis, DT: meiotic arrest patients with diploid and double-diploid cells in their testes, D: pre-meiotic arrest patients with only diploid cells in their testes.

Figure S2 (Supplementary Figure 2)

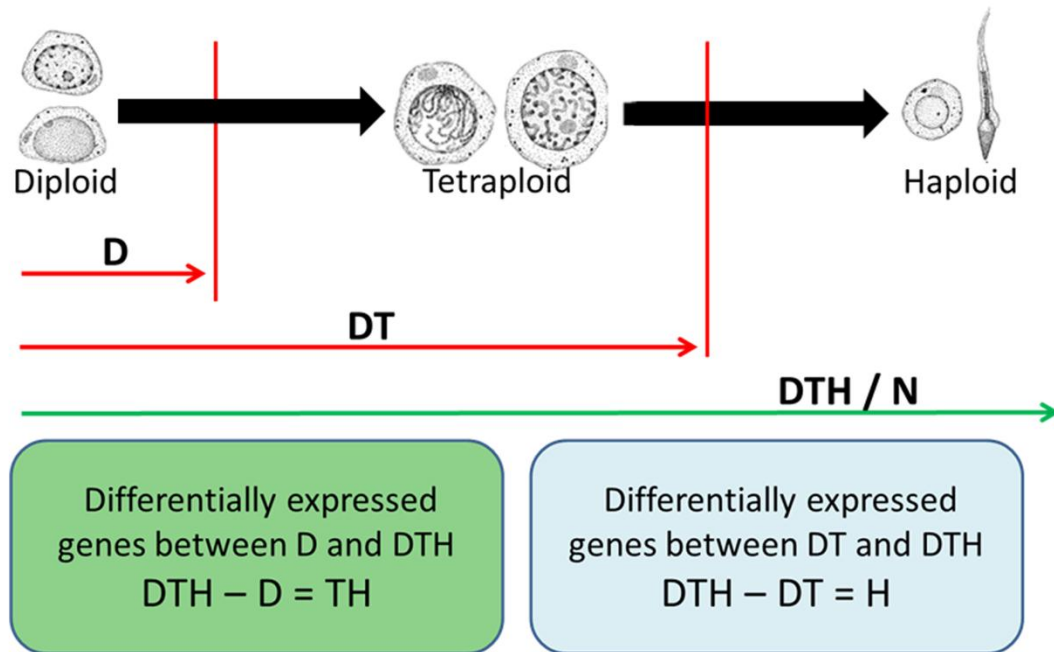


**Legend: Validation of microarray.** A total of top 17 differentially expressed genes, identified from the transcriptome analysis (D Vs N and DT Vs N) that also are the known markers of individual testicular cell populations, were quantitated in qRT-PCR. The Y-axis represents the

fold changes as compared to N (which is plotted as 1). PRM1 and PRM2 and TNP1 (4-6), the known markers for the haploid population were not expressed in samples from Group DT and D. ZPBP2, ACRBP, USP2 involved in sperm motility, DDX25 essential for spermatid development, GK2 energy pathway, and YBX1 (7-9), known to be expressed in the mature spermatids, were also expressed more in N and DTH as compared to the groups with a spermatogenic arrest. Important transcription factors such as FOS, JUN and growth factors such TGFB3 and IGFBP6 (involved in growth inhibitory signal from the Sertoli Cell) mostly were expressed more in DT and D as compared to DTH and N. MYC, a marker for spermatogonial cells (10) was expressed more in the testes of patients belonging to Group DT and D. Also there was no significant difference in the expression levels of the validated genes between N and patient samples belonging to Group DTH. However, for further analysis N and Group DTH were maintained as separate entities and all comparisons of the arrested samples (Group DT and D) were made with respect to N. [N: Pooled RNA control, DTH: obstructive azoospermia patients with complete intra-testicular spermatogenesis, DT: meiotic arrest patients with diploid and double-diploid cells in their testes, D: pre-meiotic arrest patients with only diploid cells in their testes.]

1. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic acids research* 2015:gkv007.
2. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the royal statistical society Series B (Methodological)* 1995;289-300.
3. Oliveros J. An interactive tool for comparing lists with venn diagrams. Available at: [http p. bioinfogp Cnb Csic Es/tools/venny/index.html](http://bioinfogp.cnb.csic.es/tools/venny/index.html) Accessed on 15th January 2015.
4. Wykes SM, Nelson JE, Visscher DW, Djakiew D, Krawetz SA. Coordinate expression of the PRM1, PRM2, and TNP2 multigene locus in human testis. *DNA and cell biology* 1995;14:155-61.
5. Steger K, Fink L, Failing K, Bohle RM, Kliesch S, Weidner W *et al.* Decreased protamine-1 transcript levels in testes from infertile men. *Molecular Human Reproduction* 2003;9:331-6.
6. Steger K, Klonisch T, Gavenis K, Behr R, Schaller V, Drabent B *et al.* Round Spermatids Show Normal Testis-Specific Hit but Reduced cAMP-Responsive Element Modulator and Transition Protein 1 Expression in Men With Round-Spermatid Maturation Arrest. *Journal of andrology* 1999;20:747-54.
7. Bedard N, Yang Y, Gregory M, Cyr DG, Suzuki J, Yu X *et al.* Mice lacking the USP2 deubiquitinating enzyme have severe male subfertility associated with defects in fertilization and sperm motility. *Biology of reproduction* 2011;85:594-604.
8. Sheng Y, Tsai-Morris C-H, Dufau ML. Cell-specific and hormone-regulated expression of gonadotropin-regulated testicular RNA helicase gene (GRTH/Ddx25) resulting from alternative utilization of translation initiation codons in the rat testis. *Journal of Biological Chemistry* 2003;278:27796-803.
9. Fukushima M, Villar J, Tsai-Morris C-H, Dufau ML. Gonadotropin-regulated Testicular RNA Helicase (GRTH/DDX25), a Negative Regulator of Luteinizing/Chorionic Gonadotropin Hormone-induced Steroidogenesis in Leydig Cells CENTRAL ROLE OF STEROIDOGENIC ACUTE REGULATORY PROTEIN (StAR). *Journal of Biological Chemistry* 2011;286:29932-40.
10. Uetani N, Kosemura I, Yamaura K, Sato K. Expression in situ of c-myc mRNA in adult mouse testis and ovary. In: *Proceedings of Annual Meeting of JSIR: Japan Society for Immunology of Reproduction*, 1993:129-32.

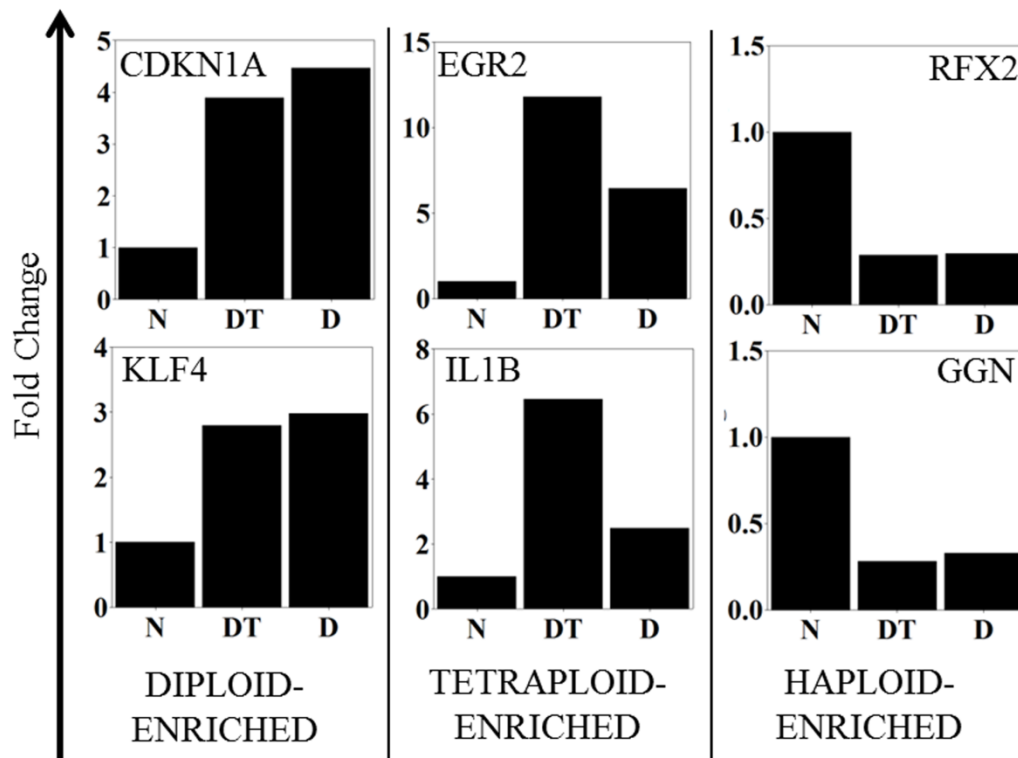
**Figure S3 (Supplementary Figure 3)**



**Legend: Schematic of microarray data analyses.** The Double-diploid and haploid-enriched genes were arrived at by subtracting the differentially expressed genes in the Groups D and DT from the differentially expressed genes in the control pooled RNA (N). The differential gene expression between N (cell populations present: DTH) and Group D (cell populations present: D) correspond to the double-diploid and haploid-enriched genes ( $DTH - D = TH$ ) [N VS D], while the genes differentially expressed between N and Group DT (cell populations present: DT) represent the haploid population-enriched genes ( $DTH - DT = H$ ) [N VS DT]. Further, the comparison of these two differentially expressed gene sets was carried out to identify the double-diploid-enriched genes (the common genes between N VS D and N VS DT represent the set of haploid-enriched genes and the remainder would be double-diploid-enriched genes).

D: diploid cells; T: double-diploid cells, H: haploid; Group D: pre-meiotic arrest patients with only diploid cells in their testes; Group DT: meiotic arrest patients with diploid and double-diploid cells in their testes; Group DTH: obstructive azoospermia patients with complete intratesticular spermatogenesis; N: Pooled RNA control

**Figure S4 (Supplementary Figure 4)**



**Legend:** Representative graphs showing differential gene-expression pattern of the testicular cell population (diploid, double-diploid and haploid) enriched genes. The median fold change of all the 1754 genes (from N Vs D and N Vs DT analyses; Supplementary Table 2) were computed and the genes showing enrichment in a particular cell population (diploid, double-diploid and haploid) were annotated as shown. This was done to confirm that the differences seen in the transcriptome analysis were due to alterations in the gene expression itself and not a result of the presence of different cell populations in the different groups. The genes that showed minimal expression in both Groups DT and D patients but showed elevated expression in N and Group DTH patients were annotated as haploid-enriched genes while the genes with elevated expression in either Groups DT or D and lower expression in Group DTH and N were annotated as double-diploid-enriched or diploid-enriched respectively. A total of 74 diploid-enriched, 32 double-diploid-enriched and 794 haploid-enriched genes were identified.

**Table S3 (Supplementary Table 3)**

**Gene enrichment analysis for the cell population specific genes:** The different cell population enriched gene sets were analyzed for enrichment to GO terms to verify whether the classification as diploid, double-diploid and haploid specific genes was accurate

<b>DIPLOID-ENRICHED</b>		<b>DOUBLE-DIPLOID-ENRICHED</b>		<b>HAPLOID-ENRICHED</b>	
<b>Biological process</b>	<b>p value</b>	<b>Biological Process</b>	<b>p value</b>	<b>Biological process</b>	<b>p value</b>
response to steroid hormone	1.02E-06	regulation of inflammatory response	4.23E-05	sperm motility	6.7E-09
response to corticosteroid	1.32E-05	negative regulation of apoptotic signaling pathway	0.000217	cilium movement	2.08E-05
regulation of cell growth	2.48E-05	MAPK cascade	0.00015	male gamete generation	1.39E-05
negative regulation of extrinsic apoptotic signaling pathway	0.000292	cellular response to gonadotropin stimulus	0.000373	spermatogenesis	1.31E-05
response to progesterone	0.000245	regulation of cellular response to stress	0.000421	spermatid development	1.49E-05
transcription from RNA polymerase II promoter	0.001754	cellular process involved in reproduction	0.00046	gamete generation	9.14E-05
germ cell migration	0.000727	positive regulation of cell cycle	0.00096	sperm-egg recognition	1.53E-05
cell cycle arrest	0.015158	positive regulation of cellular component movement	0.001219	axonemal dynein complex assembly	0.001794

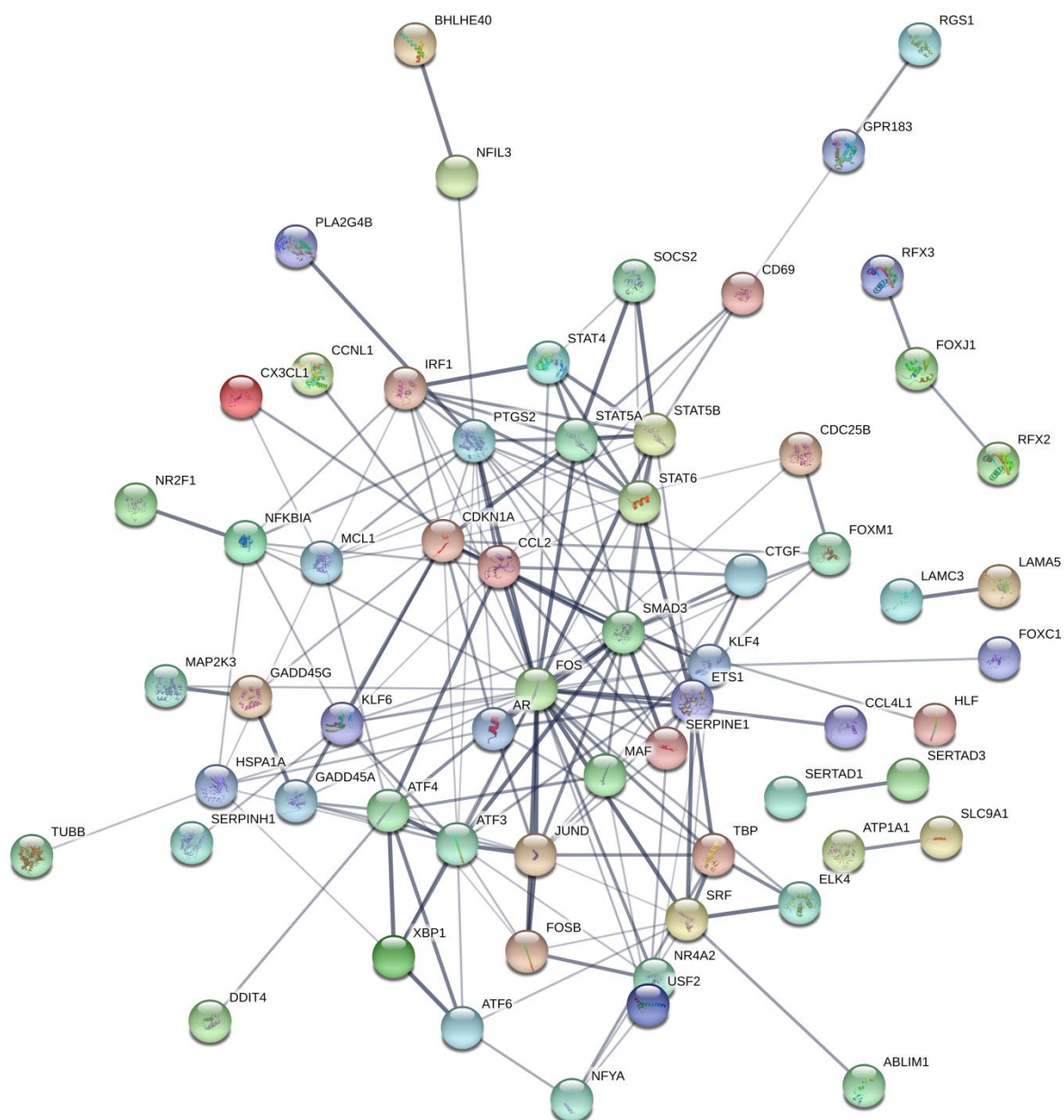


**Table S4 (Supplementary Table 4)**

**Enriched pathways for the cell population enriched genes:** Pathway analysis through KEGG provided the pathways regulated by the population enriched genes. In case of the diploid-enriched genes, the major associated pathways included growth development and the double-diploid-enriched genes showed enrichment for the meiotic, inflammatory and apoptotic pathways. The genes that showed enrichment for pathways such as spermatid maturation, sperm formation and architecture, ciliary and flagellar organization as well as energy production all belonged to the haploid-enriched gene set

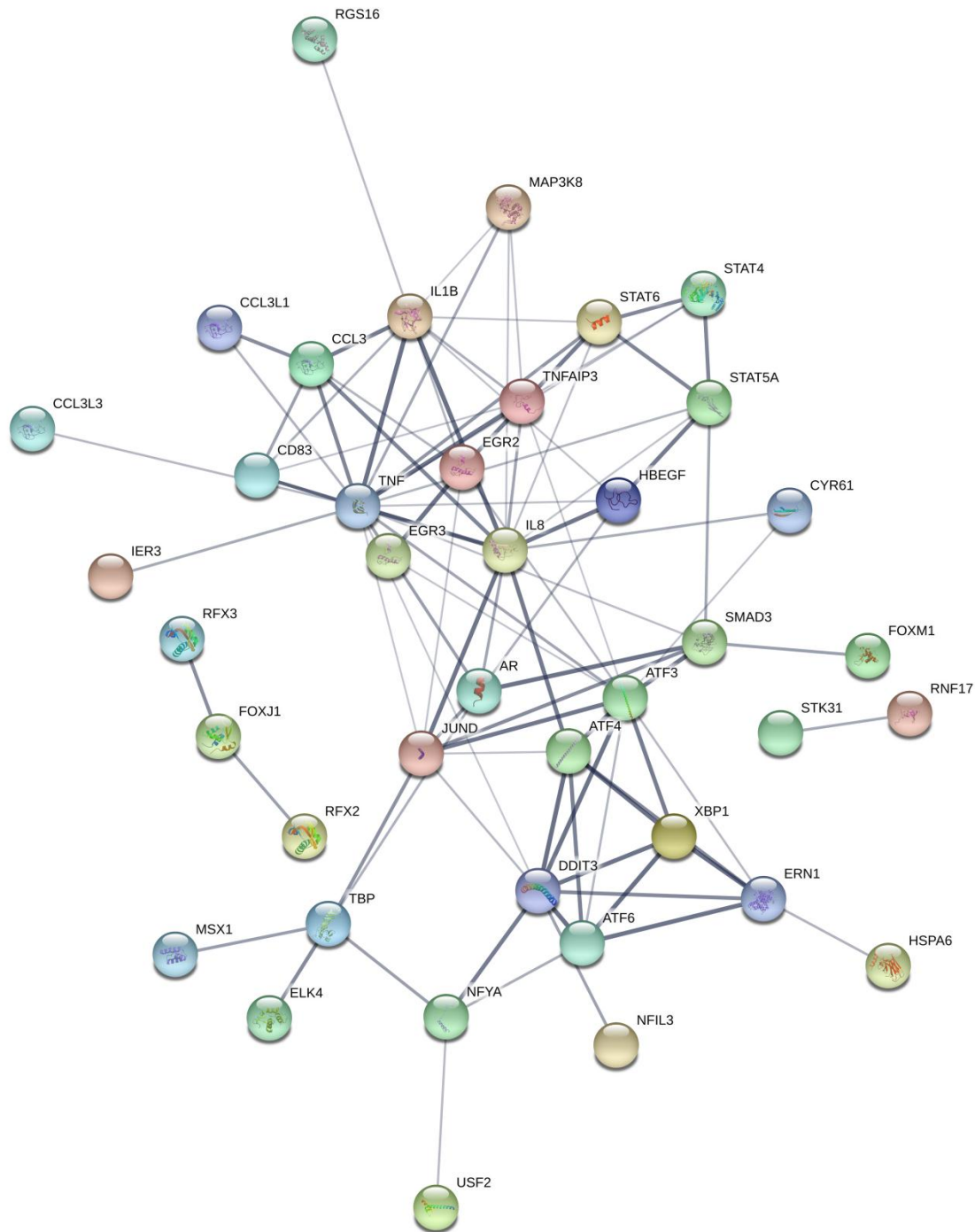
<b>DIPLOID-ENRICHED</b>	<b>DOUBLE-DIPLOID-ENRICHED</b>	<b>HAPLOID-ENRICHED</b>
<b>Biological Process</b>	<b>Biological Process</b>	<b>Biological process</b>
TNF signaling pathway	Toll-like receptor signaling pathway	Glycolysis / Gluconeogenesis
MAPK signaling pathway	Rheumatoid arthritis	Pentose phosphate pathway
p53 signaling pathway	TNF signaling pathway	Biosynthesis of amino acids
Cell cycle	MAPK signaling pathway	Carbon metabolism
PI3K-Akt signaling pathway	Graft-versus-host disease	Tight junction
Oxytocin signaling pathway	Cytokine-cytokine receptor interaction	Metabolic pathways
GnRH signaling pathway	Oocyte meiosis	Endocytosis
NF-kappa B signaling pathway	NOD-like receptor signaling pathway	Starch and sucrose metabolism
Estrogen signaling pathway	Protein processing in endoplasmic reticulum	Regulation of actin cytoskeleton
VEGF signaling pathway	Chemokine signaling pathway	Fanconi anemia pathway
Jak-STAT signaling pathway	Estrogen signaling pathway	Fatty acid biosynthesis

**Figure S5 (Supplementary Figure 5)**



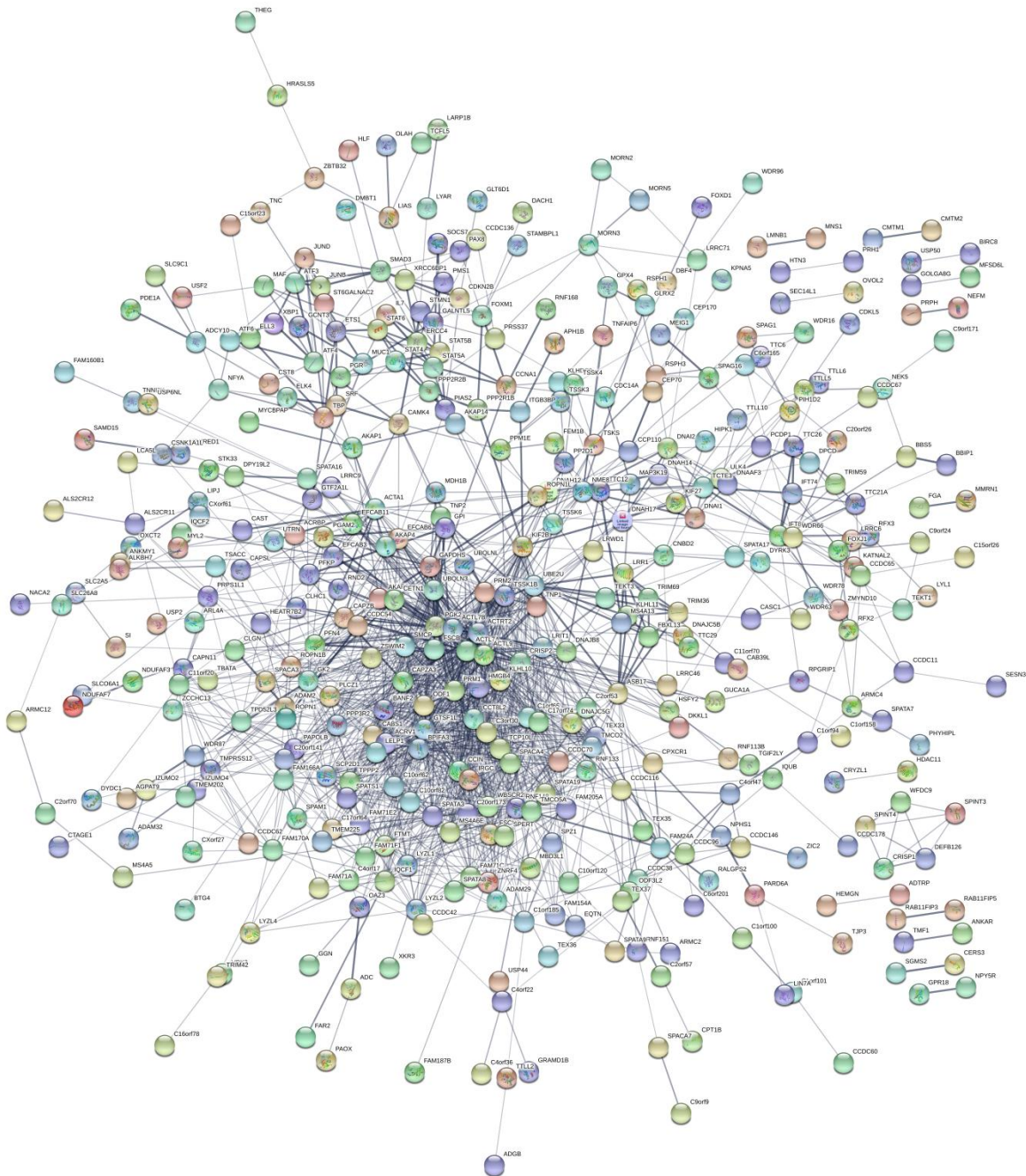
**Legend: Unbiased network map of the diploid-enriched genes before selection of genes and TFs with >4 connections**

**Figure S6 (Supplementary Figure 6)**



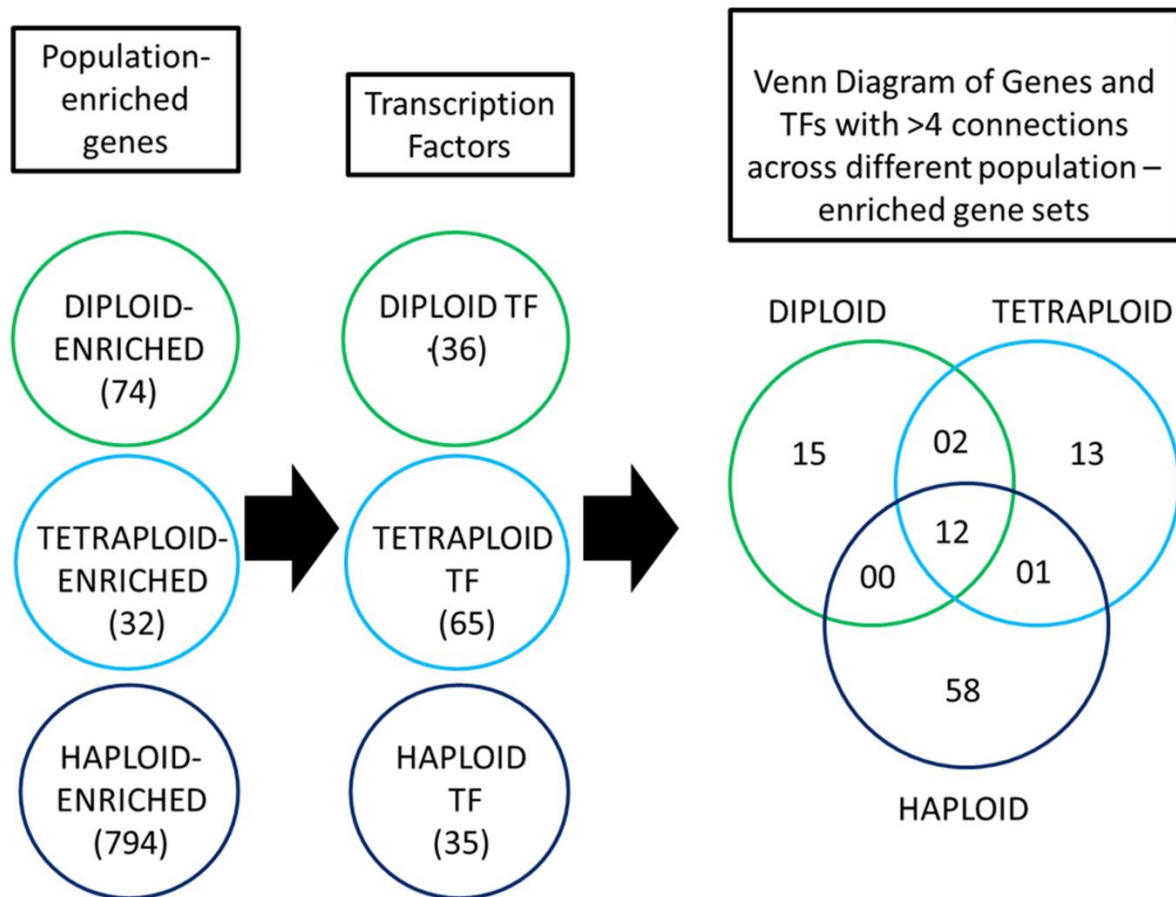
**Legend: Unbiased network map of the double-diploid-enriched genes before selection of genes and TFs with >4 connections**

**Figure S7 (Supplementary Figure 7)**



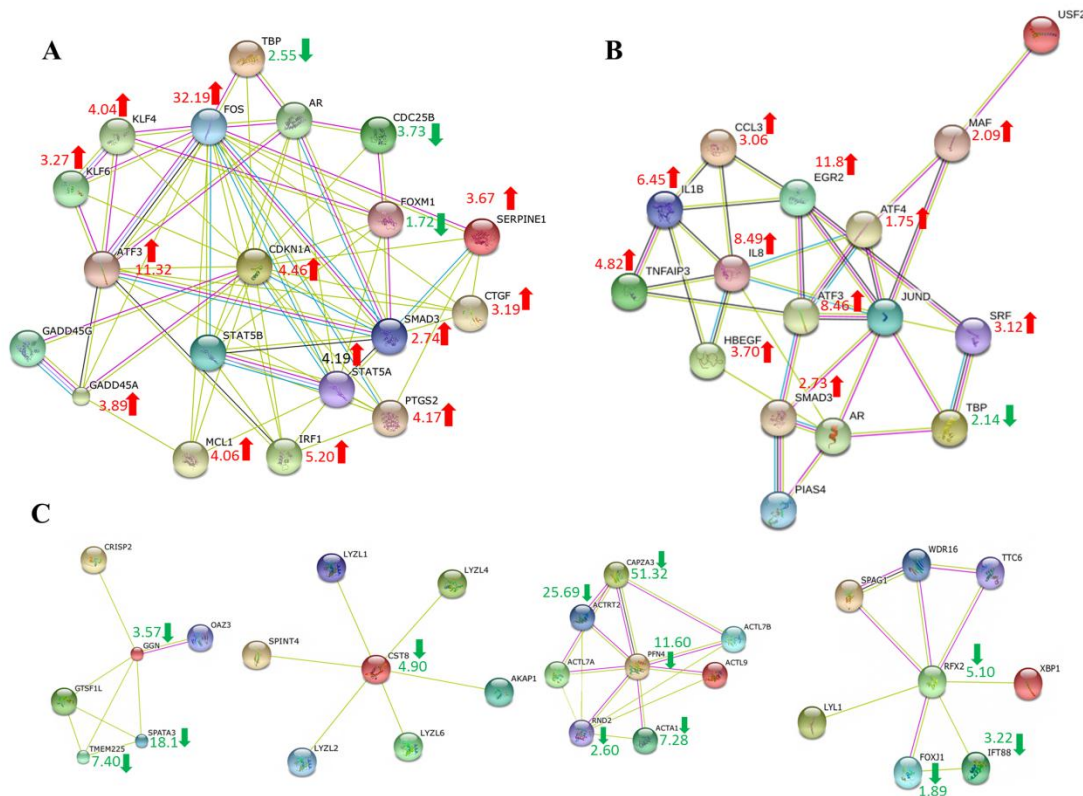
**Legend: Unbiased network map of the haploid-enriched genes before selection of genes and TFs with >4 connections**

**Figure S8 (Supplementary Figure 8)**



**Legend: Schematic representation of workflow to identify the population-enriched genes with putative role in the spermatogenic arrest.** The population-enriched gene sets were analyzed for the presence of possible transcription factor (TF) binding sites. From the putative TFs identified, the unique transcription factors differentially expressed among the different groups of patients were identified. A total of 36 diploid-enriched, 65 double-diploid-enriched and 35 haploid-enriched TF's were identified and used to construct network maps with their respective population-enriched gene sets. The transcription factors and genes that have a large number of connections with other genes in a network represent hubs that either exert control over several genes or are the genes under tight regulation and their disruption would have wider reaching effects on the constituent pathway. Subsequent to the construction of network maps, the genes and TF's with >4 connections were chosen. The Venn diagram represents the distribution of differentially expressed genes and TFs (both with >4 connections) across the different population-enriched gene sets. These genes represented in the Venn diagram were then further analyzed for their putative role in the spermatogenic arrest.

**Figure S9 (Supplementary Figure 9)**



**Legend: Network maps for testicular cell population-enriched genes. A. Diploid-enriched genes involved in diploid to double-diploid transition.** The network map for the subset of the diploid-enriched genes involved in diploid to double-diploid transition is shown. The red arrows indicate over-expression while the green arrows indicate down-regulation in the Group D (pre-meiotic arrest patients with only diploid cells in their testes). The numerals represent the fold change of the corresponding genes (from microarray; D Vs N analysis) in Group D with respect to N (Pooled RNA control). **B. Double-diploid-enriched genes involved in double-diploid to haploid transition.** The network map for the subset of the double-diploid-enriched genes involved in double-diploid to haploid transition is shown. The red arrows indicate over-expression while the green arrows indicate down-regulation in the Group DT (meiotic arrest patients with diploid and double-diploid cells in their testes). The numerals represent the fold change of the corresponding genes (from microarray; DT Vs N analysis) in Group DT with respect to N. **C. Haploid-enriched genes involved in double-diploid to haploid transition.** The network maps for the subset of haploid-enriched genes involved in double-diploid to haploid transition is shown. The green arrows indicate down-regulation in the Group DT. The numerals represent the fold change of the corresponding genes (from microarray; DT Vs N analysis) in Group DT with respect to N.

**Table S5 (Supplementary Table 5)**

**Status of Apoptotic pathway genes across transcriptome analyses**

<b>Gene</b>	<b>N VS DTH</b>	<b>N VS DT</b>	<b>N VS D</b>	<b>DTH VS DT</b>	<b>DTH VS D</b>	<b>DT VS D</b>
BCL2	NC	-0.89	-0.76	NC	-0.88	-0.73
BCL2L1	NC	NC	NC	NC	NC	NC
BAX	NC	NC	NC	NC	-0.76	NC
BAD	NC	NC	NC	NC	NC	NC
BID	NC	NC	NC	NC	NC	NC
APAF1	NC	NC	NC	NC	NC	NC
CASP8	NC	NC	NC	NC	NC	NC
CASP9	NC	NC	NC	NC	NC	NC
CASP3	NC	NC	NC	NC	NC	NC
CASP6	NC	NC	NC	NC	-0.59	NC
CASP7	NC	NC	NC	NC	NC	NC
FAS	NC	NC	-1.06	-0.63	-1.20	NC
FASLG	NC	NC	NC	NC	NC	NC

*Note:* NC- no change in expression; Fold change (FC) is represented as  $\log_2$  FC; Differential expression is from samples analyzed in microarray.

**Table S6 (Supplementary Table 6)**

**Status of Autophagy genes across transcriptome analyses**

<b>Gene</b>	<b>N VS DTH</b>	<b>N VS DT</b>	<b>N VS D</b>	<b>DTH VS DT</b>	<b>DTH VS D</b>	<b>DT VS D</b>
ATG3	NC	1.44	0.86	NC	1.08	NC
ATG4C	NC	1.03	NC	NC	NC	NC
ATG16L2	NC	-0.73	NC	NC	-0.67	NC
MAP1LC3A	NC	NC	NC	NC	NC	NC
ATG2A	NC	-0.97	-0.86	NC	NC	NC
ATG9A	NC	NC	NC	0.96	0.83	NC
ATG 7	NC	NC	NC	0.61	NC	NC
DRAM1	NC	-1.14	NC	-0.97	-1.43	NC
GABARAPL1	NC	-1.14	NC	-0.85	NC	NC
BECN1	NC	NC	NC	NC	-0.59	NC
PIK3R2	NC	-0.85	-0.84	NC	NC	NC
PIK3AP1	NC	NC	NC	NC	-0.63	NC
PIK3R1	NC	NC	NC	NC	-0.81	NC

*Note:* NC- no change in expression; Fold change (FC) is represented as  $\log_2$  FC; Differential expression is from samples analyzed in microarray.