A de novo paradigm for male infertility

Joris Veltman (✉ joris.veltman@newcastle.ac.uk)  
Newcastle University  https://orcid.org/0000-0002-3218-8250

Manon Oud  
Radboudumc  https://orcid.org/0000-0001-9513-3030

Roos Smits  
Radboudumc

Hannah Smith  
Newcastle University

Francesco Mastrorosa  
Newcastle University  https://orcid.org/0000-0003-0579-1895

Giles Holt  
Newcastle University

Brendan Houston  
The University of Melbourne

Petra de Vries  
Radboud University Medical Center

Bilal Alobaidi  
Newcastle University  https://orcid.org/0000-0003-2718-4826

Lois Batty  
Newcastle University

Hadeel Ismail  
Newcastle University

Jackie Greenwood  
Newcastle University

Harsh Sheth  
Foundation for Research in Genetics and Endocrinology  https://orcid.org/0000-0001-9626-0971

Aneta Mikulasova  
Newcastle University

Galuw Astuti  
Radboudumc

Christian Gilissen  
Radboud University Medical Center  https://orcid.org/0000-0003-1693-9699

Kevin McEleny  
Newcastle Fertility Centre, The Newcastle upon Tyne Hospitals NHS Foundation Trust

Helen Turner  
Department of Cellular Pathology, The Newcastle upon Tyne Hospitals NHS Foundation Trust

Jonathan Coxhead  
Newcastle University

Simon Cockell  
Newcastle University

Didi Braat  
Radboudumc
Article

Keywords: De novo mutations, male infertility, RBM5

DOI: https://doi.org/10.21203/rs.3.rs-332732/v1

License: ©  This work is licensed under a Creative Commons Attribution 4.0 International License.  Read Full License
Abstract

De novo mutations (DNMs) are known to play a prominent role in many sporadic disorders with reduced fitness. We hypothesize that DNMs play an important role in male infertility and explain a significant fraction of the genetic causes of this understudied disorder. We performed a trio-based exome-sequencing study in a unique cohort of 185 infertile males and their unaffected parents. Following a systematic analysis, 29 of 145 rare protein altering DNMs were classified as possibly causative of the male infertility phenotype. We observed a significant enrichment of Loss-of-Function (LoF) DNMs in LoF-intolerant genes \( (p=1.00\times 10^{-5}) \) as well as predicted pathogenic missense DNMs in missense-intolerant genes \( (p=5.01\times 10^{-4}) \). One DNM gene identified, RBM5, is an essential regulator of male germ cell pre-mRNA splicing. In a follow-up study, 5 rare pathogenic missense mutations affecting this gene were observed in a cohort of 2,279 infertile patients, with no such mutations found in a cohort of 5,784 fertile men \( (p=0.009) \). Our results provide the first evidence for the role of DNMs in severe male infertility and point to many new candidate genes affecting fertility.

Main

Male infertility contributes to approximately half of all cases of infertility and affects 7% of the male population. For the majority of these men the cause remains unexplained\(^3\). Despite a clear role for genetic causes in male infertility, there is a distinct lack of diagnostically relevant genes and at least 40% of all cases are classified as idiopathic\(^3–6\). Previous studies in other conditions with reproductive lethality, such as neurodevelopmental disorders, have demonstrated an important role for de novo mutations (DNMs) in their etiology\(^1\). In line with this, recurrent de novo chromosomal abnormalities play an important role in male infertility. Both azoospermia Factor (AZF) deletions on the Y chromosome as well as an additional X chromosome, resulting in Klinefelter syndrome, occur de novo. Collectively, these de novo events explaining up to 25% of all cases of non-obstructive azoospermia (NOA)\(^3,6\). Interestingly, in 1999 a DNM in the Y-chromosomal gene USP9Y was reported in a man with azoospermia\(^7\). Until now, however, a systematic analysis of the role of DNMs in male infertility had not been attempted. This is partly explained by a lack of basic research in male reproductive health in general\(^6,8\), but also by the practical challenges of collecting parental samples for this disorder, which is typically diagnosed in adults.

In this study, we investigated the role of DNMs in 185 unexplained cases of oligozoospermia \(<5\) million sperm cells/ml \(n=74\) and azoospermia \(n=111\) by performing whole exome sequencing (WES) in all patients and their parents (see Supplementary Figure 1 and 2, Supplementary notes and tables for details on methods and clinical description). In total, we identified and validated 192 rare DNMs, including 145 protein altering DNMs. All de novo point mutations were autosomal, except for one on chromosome X, and all occurred in different genes (Supplementary Table 1). Two de novo copy number variations (CNVs) were also identified affecting a total of 7 genes (Supplementary Figure 3).

None of the 145-protein altering DNMs occurred in a gene already known for its involvement in autosomal dominant human male infertility. This is not unexpected as only 4 autosomal dominant genes have so far been linked to isolated male infertility in humans\(^5,9\). Broadly speaking, across genetic disorders, dominantly acting disease genes are usually intolerant to loss-of-function (LoF) mutations, as represented by a high pLI score\(^10\). The median pLI score of genes with a LoF DNM \(n=17\) in our cohort of male infertility cases was significantly higher than that of genes with 181 LoF DNMs identified in a cohort of 1,941 control cases from de novo-db v1.6.111 \(p=1.00\times 10^{-5}\) (Figure 1). This observation indicates that LoF DNMs likely play an important role in male infertility, similar to what is known for developmental disorders and severe intellectual disability\(^12,13\). As an example, a heterozygous likely pathogenic frameshift DNM was observed in the LoF intolerant gene GREB1L \(p=1\) of Proband_076. Homozygous Greb1L knock-out mice appear to be embryonic lethal, however, typical male infertility phenotypic features such as abnormal fetal testis morphology and decreased fetal testis volume are observed\(^14\). Interestingly, this patient has a reduced testis volume and severe oligospermia (Supplementary Notes Table 1). Nonsense and missense mutations in GREB1L in humans are known to cause renal agenesis\(^15\) (OMIM: 617805), not known to be present in our patient. Of note, all previously reported damaging mutations in GREB1L causing renal agenesis are either maternally inherited or occurred de novo. This led the authors of one of these renal agenesis studies to speculate that
disruption to \textit{GREB1L} could cause infertility in males\cite{14}. A recent WES study involving a cohort of 285 infertile men also noted several patients presenting with pathogenic mutations in genes with an associated systemic disease where male fertility is not always assessed\cite{16}. We also assessed the damaging effects of the two \textit{de novo} CNVs by looking at the pLI score of the genes involved. Proband\_066 presented with a large 656 kb \textit{de novo} deletion on chromosome 11, spanning 6 genes in total. This deletion partially overlapped with a deletion reported in 2014 in a patient with cryptorchidism and NOA\cite{17}. Two genes affected in both patients, \textit{QSER1} and \textit{CSTF3}, are extremely LOF-intolerant with pLI scores of 1 and 0.98, respectively. In particular, \textit{CSTF3} is highly expressed within the testis and is known to be involved in pre-mRNA 3’ end cleavage and polyadenylation\cite{18}.

To systematically evaluate and predict the likelihood of these DNMs causing male infertility and identify novel candidate disease genes, we assessed the predicted pathogenicity of all DNMs using three prediction methods based on SIFT\cite{19}, MutationTaster\cite{20} and PolyPhen2\cite{21}. Using this approach, 84/145 protein altering DNM were predicted to be pathogenic, while the remaining 61 were predicted to be benign. To further analyse the impact of the variants on the genes affected, we looked at the missense Z-score of all 122 genes affected by a missense variant, which indicates the tolerance of genes to missense mutations\cite{22}. Our data highlights a significantly higher missense Z-score in genes affected by a missense DNM predicted as pathogenic (n=63) when compared to genes affected by predicted benign (n=59) missense DNMs (p-value=5.01x10^{-4}, Figure 2, Supplementary Figure 4). Furthermore, using the STRING database\cite{23}, we found a significant enrichment of protein interactions amongst the 84 genes affected by a protein altering DNM predicted to be pathogenic (PPI enrichment p-value = 2.35 x 10^{-2}, Figure 3). No such enrichment was observed for the genes highlighted as likely benign (n=61, PPI enrichment p-value=0.206) or those affected by synonymous DNMs (n=35, PPI enrichment p-value=0.992, Supplementary Figure 5). These two findings suggest that (1) the predicted pathogenic missense DNMs detected in our study affect genes sensitive to missense mutations, and (2) the proteins affected by predicted pathogenic DNMs share common biological functions.

The STRING network analysis also highlighted a central module of interconnected proteins with a significant enrichment of genes required for mRNA splicing (Supplementary Figure 6). The genes U2AF2, HNRNPL, CDC5L, CWC27 and RBM5 all contain predicted pathogenic DNMs and likely interact at a protein level during the mRNA splicing process. Pre-mRNA splicing allows gene functions to be expanded by creating alternative splice variants of gene products and is highly elaborated within the testis\cite{24}. One of these genes, RBM5 has been previously highlighted as an essential regulator of haploid male germ cell pre-mRNA splicing and male fertility\cite{2}. Mice with a homozygous ENU-induced allele point mutation in \textit{RB5M} present with azoospermia and germ cell development arrest at round spermatids. Whilst in mice a homozygous mutation in \textit{RB5M} is required to cause azoospermia, this may not be the case in humans as is well-documented for other genes\cite{25}, including the recently reported male infertility gene \textit{SYCP2}\cite{26}. Of note, \textit{RB5M} is a tumour suppressor in the lung\cite{27}, with reduced expression affecting RNA splicing in patients with non-small cell lung cancer\cite{28}. \textit{HNRNPL} is another splicing factor affected by a possible pathogenic DNM in our study. One study implicated a role for \textit{HNRNPL} in patients with Sertoli cell only phenotype\cite{29}. The remaining three mRNA splicing genes have not yet been implicated in human male infertility. However, mRNA for all three is expressed at medium to high levels in human germ cells and all are widely expressed during spermatogenesis\cite{30}. Specifically, CDC5L is a component of the PRP19-CDC5L complex that forms an integral part of the spliceosome and is required for activating pre-mRNA splicing\cite{31}. U2AF2 plays a role in pre-mRNA splicing and 3’-end processing\cite{32}. Interestingly, \textit{CSTF3}, one of the genes affected by a \textit{de novo} CNV in Proband\_066, affects the same mRNA pathway\cite{17}.

Whilst DNMs most often cause dominant disease, they can contribute to recessive disease, usually in combination with an inherited variant on the trans allele. This was observed in Proband\_060, who carried a DNM on the paternal allele, in trans with a maternally inherited variant in Testis and Ovary Specific PAZ Domain Containing 1 (\textit{TOPAZ1}) (Supplementary Figure 7). \textit{TOPAZ1} is a germ-cell specific gene which is highly conserved in vertebrates\cite{33}. Studies in mice revealed that \textit{Topaz1} plays a crucial role in spermatocyte, but not oocyte progression through meiosis\cite{34}. In men, \textit{TOPAZ1} is expressed in germ cells in both sexes\cite{29,35,36}. Analysis of the testicular biopsy of this patient revealed a germ cell arrest in early spermiogenesis (Figure 4).

In addition to all systematic analyses described above, we evaluated the function of all DNM genes to give each a final pathogenicity classification (Table 1, details in Material & Methods). Of all 145 DNMs, 29 affected genes linked to male
reproduction and were classified as possibly causative. For replication purposes, unfortunately no other trio-based exome data are available for male infertility, although we note that a pilot study including 13 trios was recently published \(^{37}\). While this precluded a genuine replication study, we were able to study these candidate genes in exome datasets of infertile men \((n=2,279)\), in collaboration with members of the International Male Infertility Genomics Consortium and the Geisinger Regeneron DiscovEHR collaboration \(^{38}\). The 33 candidate genes selected for this analysis include the 29 genes mentioned above and 4 additional LoF intolerant genes carrying LoF DNMs with an ‘unclear’ final pathogenicity classification. For comparison, we included an exome dataset from a cohort of 11,587 fertile men and women from Radboudumc.

In the additional infertile cohorts, we identified only 2 LoF mutations in our DNM LoF intolerant genes (Supplementary table 2). Next, we looked for an enrichment of rare predicted pathogenic missense mutations in these cohorts (Table 2). A burden test revealed a significant enrichment in the number of such missense mutations present in infertile men compared to fertile men in the RBM5 gene (adjusted p-value=0.009). In this gene, 5 infertile men were found to carry a distinct rare pathogenic missense mutation, in addition to the proband with a \textit{de novo} missense mutation (Supplementary figure 8, Supplementary table 3). Importantly, no such predicted pathogenic mutations were identified in men in the fertile cohort. In line with these results, RBM5, already highlighted above as an essential regulator of male germ cell pre-mRNA splicing and male infertility \(^{2}\), is highly intolerant to missense mutations (missense Z-score 4.17).

Given the predicted impact of these DNMs on spermatogenesis, we were interested in studying the parental origin of DNMs in our trio-cohort. We were able to phase 29% of all our DNMs using a combination of short-read WES and targeted long-read sequencing (Supplementary Table 4). In agreement with literature \(^{39–42}\), 72% of all DNMs occurred on the paternal allele. Interestingly, phasing of 8 likely causative DNMs showed that 6 of these were of paternal origin (75%). This suggests that DNMs with a deleterious effect on the future germline can escape negative selection in the paternal germline. This may be possible because the DNM occurred after the developmental window in which the gene is active, or the DNM may have affected a gene in the gamete's genome that is critical for somatic cells supporting the (future) germline. Transmission of pathogenic DNMs may also be facilitated by the fact that from spermatogonia onwards, male germ cells form cysts and share mRNAs and proteins \(^{43}\). As such, the interconnectedness of male germ cells, which is essential for their survival \(^{44}\), could mask detrimental effects of DNMs occurring during spermatogenesis.

In 2010, we published a pilot study pointing to a \textit{de novo} paradigm for mental retardation \(^{45}\) (now more appropriately termed developmental delay or intellectual disability). This work contributed to the widespread implementation of patient-parent WES studies in research and diagnostics for neurodevelopmental disorders \(^{46}\), accelerating disease gene identification and increasing the diagnostic yield for these disorders. The data presented here suggest that a similar benefit could be achieved from trio-based sequencing in male infertility. This will not only help to increase the diagnostic yield for men with infertility but will also enhance our fundamental biological understanding of human reproduction and natural selection.

\textbf{Declarations}

\textbf{Data access}

Raw and processed exome sequencing data of our 185 patient-parent trios is available under controlled access and requires a Data Transfer Agreement from the European Genome-Phenome Archive (EGA) repository: EGAS00001004945.

\textbf{Acknowledgements}

We are grateful for the participation of all patients and their parents in this study. We thank Laurens van de Wiel (Radboudumc), Sebastian Judd-Mole (Monash University), Arron Scott and Bryan Hepworth (Newcastle University) for technical support, and Margot J Wyrwoll (University of Münster) for help with handling MERGE samples and data. This project was funded by The Netherlands Organisation for Scientific Research (918-15-667) to JAV as well as an Investigator Award in Science from the Wellcome Trust (209451) to JAV, a grant from the Catherine van Tussenbroek Foundation to MSO, a UUKi Rutherford Fund Fellowship awarded to BJH and the German Research Foundation Clinical Research Unit ‘Male Germ Cells’
(DFG, CRU326) to CF and FT. This project was also supported in part by funding from the Australian National Health and Medical Research Council (APP1120356) to MKOB, by grants from the National Institutes of Health of the United States of America (R01HD078641 to D.F.C and K.I.A, P50HD096723 to D.F.C.) and from the Biotechnology and Biological Sciences Research Council (BB/S008039/1) to DJE.

Author contributions

This study was designed by MSO, LELMV, LR and JAV. RMS, JG, HT and GWvdH provided all clinical data and performed the TESE histology and cytology analysis under supervision of LR, DDMB, ES, KDH and KM. JC performed the exome sequencing with support from BA, and bioinformatics support was provided by MJX, GA, CG and SC. Sanger sequencing was performed by PFdV, HI, HES, LEB and BKSA. MSO and HES performed the SNV analyses with support from MJX, FKM performed CNV analysis with support from AM and MSK, and GSH and LEB performed the phasing. DJE, HS, BJH and MKOB provided support on the functional interpretation of mutations. DFC, LN, CF, SK, FT, KIA, ARE, CK, and CG-J were involved in the replication study. The first draft of the manuscript was prepared by MSO, HES, RMS, MJX, GWvdH, and JAV. All authors contributed to the final manuscript.

References

1. Veltman, J. A. & Brunner, H. G. De novo mutations in human genetic disease. Nat. Rev. Genet.13, 565–575 (2012).
2. O’Bryan, M. K. et al. RBM5 Is a Male Germ Cell Splicing Factor and Is Required for Spermatid Differentiation and Male Fertility. PLoS Genet.9, e1003628 (2013).
3. Krausz, C. & Riera-Escamilla, A. Genetics of male infertility. Nat. Rev. Urol.15, 369–384 (2018).
4. Tüttelmann, F., Ruckert, C. & Röpke, A. Disorders of spermatogenesis. medizinische Genet.30, 12–20 (2018).
5. Oud, M. S. et al. A systematic review and standardized clinical validity assessment of male infertility genes. Hum. Reprod.34, 932–941 (2019).
6. Kasak, L. & Laan, M. Monogenic causes of non-obstructive azoospermia: challenges, established knowledge, limitations and perspectives. Hum. Genet.140, 135–154 (2021).
7. Sun, C. et al. An azoospermic man with a de novo point mutation in the Y-chromosomal gene USP9Y. Nat. Genet.23, 429–432 (1999).
8. De Jonge, C. & Barratt, C. L. R. The present crisis in male reproductive health: an urgent need for a political, social, and research roadmap. Andrology7, 762–768 (2019).
9. Schilit, S. L. P. et al. SYCP2 Translocation-Mediated Dysregulation and Frameshift Variants Cause Human Male Infertility. Am. J. Hum. Genet.106, 41–57 (2020).
10. Lek, M. et al. Analysis of protein-coding genetic variation in 60,706 humans. Nature536, 285–291 (2016).
11. denovo-db, Seattle, WA (denovo-db.gs.washington.edu) [Aug 2020].
12. Gu, Y. et al. Three intellectual disability-associated de novo mutations in MECP2 identified by trio-WES analysis. BMC Med. Genet.21, 99 (2020).
13. Fritzen, D. et al. De novo FBXO11 mutations are associated with intellectual disability and behavioural anomalies. Hum. Genet.137, 401–411 (2018).
14. De Tomasi, L. et al. Mutations in GREB1L Cause Bilateral Kidney Agenesis in Humans and Mice. Am. J. Hum. Genet.101, 803–814 (2017).
15. Brophy, P. D. et al. A Gene Implicated in Activation of Retinoic Acid Receptor Targets Is a Novel Renal Agenesis Gene in Humans. Genetics207, 215–228 (2017).
16. Alhathal, N. et al. A genomics approach to male infertility. Genet. Med.22, 1967–1975 (2020).
17. Seabra, C. M. et al. A novel Alu-mediated microdeletion at 11p13 removes WT1 in a patient with cryptorchidism and azoospermia. Reprod. Biomed. Online29, 388–391 (2014).
18. Grozdanov, P. N., Li, J., Yu, P., Yan, W. & MacDonald, C. C. Cstf2t Regulates expression of histones and histone-like proteins in male germ cells. *Andrology* **6**, 605–615 (2018).

19. Vaser, R., Adusumalli, S., Leng, S. N., Sikic, M. & Ng, P. C. SIFT missense predictions for genomes. *Nat. Protoc.* **11**, 1–9 (2016).

20. Schwarz, J. M., Rödelsperger, C., Schuelke, M. & Seelow, D. MutationTaster evaluates disease-causing potential of sequence alterations. *Nat. Methods* **7**, 575–576 (2010).

21. Adzhubei, I. A. *et al.* A method and server for predicting damaging missense mutations. *Nat. Methods* **7**, 248–249 (2010).

22. Samocha, K. E. *et al.* A framework for the interpretation of de novo mutation in human disease. *Nat. Genet.* **46**, 944–950 (2014).

23. Szklarczyk, D. *et al.* The STRING database in 2017: quality-controlled protein–protein association networks, made broadly accessible. *Nucleic Acids Res.* **45**, D362–D368 (2017).

24. Song, H., Wang, L., Chen, D. & Li, F. The Function of Pre-mRNA Alternative Splicing in Mammal Spermatogenesis. *Int. J. Biol. Sci.* **16**, 38–48 (2020).

25. Elsea, S. H. & Lucas, R. E. The Mousetrap: What We Can Learn When the Mouse Model Does Not Mimic the Human Disease. *ILAR J.* **43**, 66–79 (2002).

26. Jamsai, D. *et al.* In vivo evidence that RBM5 is a tumour suppressor in the lung. *Sci. Rep.* **7**, 16323 (2017).

27. Li, J. *et al.* HnRNPL as a key factor in spermatogenesis: Lesson from functional proteomic studies of azoospermia patients with sertoli cell only syndrome. *J. Proteomics* **75**, 2879–2891 (2012).

28. Dewey, F. E. *et al.* Distribution and clinical impact of functional variants in 50,726 whole-exome sequences from the DiscovEHR study. *Science (80-. ).* **354**, aaf6814 (2016).

29. Francioli, L. C. *et al.* Genome-wide patterns and properties of de novo mutations in humans. *Nat. Genet.* **47**, 822–826 (2015).

30. Goldmann, J. M. *et al.* Parent-of-origin-specific signatures of de novo mutations. *Nat. Genet.* **48**, 935–939 (2016).
42. Jónsson, H. et al. Parental influence on human germline de novo mutations in 1,548 trios from Iceland. *Nature* **549**, 519–522 (2017).

43. Braun, R. E., Behringer, R. R., Peschon, J. J., Brinster, R. L. & Palmiter, R. D. Genetically haploid spermatids are phenotypically diploid. *Nature* **337**, 373–376 (1989).

44. Greenbaum, M. P., Iwamori, T., Buchold, G. M. & Matzuk, M. M. Germ Cell Intercellular Bridges. *Cold Spring Harb. Perspect. Biol.* **3**, a005850–a005850 (2011).

45. Vissers, L. E. L. M. et al. A de novo paradigm for mental retardation. *Nat. Genet.* **42**, 1109–12 (2010).

46. Vissers, L. E. L. M., Gilissen, C. & Veltman, J. A. Genetic studies in intellectual disability and related disorders. *Nat. Rev. Genet.* **17**, 9–18 (2016).

**Tables**

**Table 1: De novo mutation classification summary.**

|                        | Possibly causative | Unclear | Unlikely causative | Not Causative | Total |
|------------------------|-------------------|---------|--------------------|---------------|-------|
| Missense               | 21                | 38      | 50                 | 13            | 122   |
| Frameshift             | 4                 | 8       | 1                  | 0             | 13    |
| Stop gained            | 1                 | 3       | 0                  | 0             | 4     |
| In-frame indels        | 3                 | 1       | 1                  | 1             | 6     |
| Splice site variant    | 0                 | 0       | 0                  | 11            | 11    |
| Synonymous             | 0                 | 0       | 0                  | 36            | 36    |
| **TOTAL**              | **29**            | **50**  | **52**             | **61**        | **192**|

A total of 192 rare DNMs were classified based on pathogenicity scores as well as functional data into 4 categories, ‘Possibly causative’, ‘Unclear’, ‘Unlikely Causative’ and ‘Not causative’. 

---

**Page 8/14**
| Gene   | Missense Z-score | NIJ/NCL Cohort | NIJ/NCL Cohort | MERGE GEMINI Cohort | Geisinger-Regeneron Italian Cohort | Total NOA Cohort of Infertile Men (n=2,279) | Fertile Dutch Men (n=5,784) | Fertile Dutch Women (n=5,803) | Burden test Infertile Men vs Fertile Men (Bonf) | Burden test Infertile Men vs Fertile Women (Bonf) |
|--------|------------------|----------------|----------------|---------------------|-------------------------------------|-----------------------------------------------|---------------------------------|---------------------------------|-----------------------------------------------|-----------------------------------------------|
| ABLIM1 | 1.62             | 1              | 1              | 1                   | 1                                   | 0                                             | 5                              | 1                               | 1.15                           | 1                                             |
| ATP1A1 | 6.22             | 0              | 0              | 0                   | 1                                   | 0                                             | 1                              | 0                               | 1                               | 1                                             |
| CDC5L  | 2.78             | 1              | 1              | 1                   | 3                                   | 0                                             | 6                              | 2                               | 4.85                           | 0.15                           |
| CDK5RAP2 | -0.37           | 1              | 0              | 1                   | 1                                   | 0                                             | 2                              | 3                               | 2                              | 3.15                           |
| HUWE1  | 8.87             | 1              | 0              | 2                   | 0                                   | 0                                             | 3                              | 0                               | 0                               | 0.41                           |
| IHO80  | 3.53             | 1              | 0              | 1                   | 0                                   | 0                                             | 2                              | 3                               | 3                               | 1                              |
| MAP3K3 | 2.04             | 1              | 0              | 2                   | 0                                   | 0                                             | 3                              | 1                               | 2                              | 1                              |
| MCM6   | 1.07             | 1              | 1              | 1                   | 3                                   | 0                                             | 6                              | 4                               | 8                               | 0.64                           |
| PPP1R7 | 1.86             | 0              | 0              | 0                   | 1                                   | 0                                             | 1                              | 1                               | 1                              | 1                              |
| QSER1  | 1.34             | 1              | 0              | 1                   | 0                                   | 0                                             | 2                              | 8                               | 1                               | 1                              |
| RASAL2 | 1.40             | 0              | 1              | 1                   | 2                                   | 1                                             | 0                              | 5                               | 25                              | 1                              |
| RRM1   | 4.17             | 1              | 2              | 2                   | 0                                   | 1                                             | 0                              | 6                               | 0                               | 2                              |
| RPA1   | 1.22             | 1              | 0              | 0                   | 1                                   | 0                                             | 0                              | 2                               | 3                               | 3.15                           |
| SDF4   | 0.53             | 0              | 0              | 0                   | 0                                   | 0                                             | 1                              | 2                               | 1                               | 1                              |
| SOGA1  | 2.27             | 1              | 0              | 1                   | 1                                   | 0                                             | 3                              | 15                              | 5                               | 1                              |
| STARD10| 1.34             | 1              | 0              | 2                   | 0                                   | 0                                             | 0                              | 3                               | 4                               | 5                              |
| TENM2  | 3.30             | 1              | 0              | 2                   | 2                                   | 0                                             | 2                              | 7                               | 16                              | 1                              |
| ZFHX4  | 1.01             | 0              | 0              | 3                   | 3                                   | 0                                             | 0                              | 6                               | 14                              | 8                              |

Table 2: Rare potentially pathogenic missense mutations in exome data from various cohorts of infertile men and fertile control cohorts.

The genes included in this analysis were among the strongest candidate genes affected by a DNM (either missense or LoF mutation). The missense Z-score is included here to indicate a relative (in)tolerance to missense mutation. For the original NIJ/NCL discovery cohort, only the missense DNMs are included in this Table (7 of these genes were affected by a LoF DNM). A burden test was done to compare the total number of predicted pathogenic missense mutations observed in the infertile vs. fertile men, as well as between fertile men and fertile women (Fisher’s Exact test, adjusted for multiple testing following Bonferroni correction).

Figures
Figure 1

Analysis of the intolerance to loss-of-function variation for DNM genes. Violin plots represent the distribution of the pLI scores of all genes in gnomAD, all genes affected by DNMs and all LoF DNM in this study and in a control population (http://denovo-db.gs.washington.edu/denovo-db/). The observed median pLI score is displayed for each category as a black circle. The closer the pLI score is to 1, the more intolerant to LoF variation a gene is\(^\text{10}\). Comparison between LoF DNMs in our study and control populations shows a significance difference (p-value=1.00\times10^{-5}).
Intolerance to missense variants for genes with a DNM. Violin plots show the distribution of Z-scores of genes containing a missense DNM in our cohort, where an enrichment can be observed for predicted pathogenic DNMs in genes more intolerant to missense mutations based on their mean z-score with a p-value of 5.01x10^{-4}. 

Figure 2
Figure 3

Protein-protein interactions predicted for proteins encoded by damaging DNM genes. A protein-protein interaction analysis was performed for all 84 genes containing a DNM scored as damaging using the STRING tool. A significantly larger number of interactions is observed between our damaging DNM genes than is expected for a similar sized dataset of randomly selected genes (PPI enrichment p-value $2.35 \times 10^{-2}$) with the number of expected edges being 25 and the observed being 36. The central module of the main interaction network within the figure contains 5 genes which are all involved in the process of mRNA splicing (Supplementary figure 6)
Figure 4

Description of control and TOPAZ1 proband testis histology and aberrant acrosome formation: (a,b): H&E stainings of (a) control and (b) Proband_060 with DNM in TOPAZ1 gene. The epithelium of the seminiferous tubules in the TOPAZ1 proband show reduced numbers of germ cells and an absence of elongating spermatids. (c,d): Immunofluorescent labelling of DNA (magenta) and the acrosome (green) in control sections (c) and TOPAZ1 proband sections (d). (c) The arrowhead indicates the acrosome in an early round spermatid and the arrows the acrosome in elongating spermatids. Spreading of the acrosome and nuclear elongation are hallmarks of spermatid maturation. (d) No acrosomal spreading (see arrowheads) or nuclear elongation is observed in the TOPAZ1 proband. The asterisk indicates an example of progressive acrosome accumulation without spreading. Size bar in a, b: 40 µm, c, d: 5 µm.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryNotes2GEMINIParticipatingAuthors.docx
- Supplementaryfigures18.docx
- Supplementaryfigures18.docx
