Journal of Assisted Reproduction and Genetics (2021) 38:3195–3212
https://doi.org/10.1007/s10815-021-02352-5

GAMETE BIOLOGY

Novel variations in spermatogenic transcription regulators RFX2 and TAF7 increase risk of azoospermia

Samudra Pal[1] · Pranab Paladhi[1] · Saurav Dutta[1] · Gunja Bose[2] · Papiya Ghosh[3] · Ratna Chattopadhyay[2] · Baidyanath Chakravarty[2] · Indranil Saha[4] · Sujay Ghosh[1]

Received: 23 August 2021 / Accepted: 29 October 2021 / Published online: 11 November 2021

Abstract

Purpose Genetic etiology of idiopathic male infertility is enigmatic owing to involvement of multiple gene regulatory networks in spermatogenesis process. Any change in optimal function of the transcription factors involved in this process owing to polymorphisms/mutations may increase the risk of infertility. We investigated polymorphisms/mutations of spermatogenic transcription regulators TAF7 and RFX2 and analysed their association with incidence of azoospermia among the men from West Bengal, India.

Methods Genotyping was carried by Sanger's dideoxy sequencing of 130 azoospermic men who were detected negative in Y chromosome microdeletion screening and 140 healthy controls. Association study was done by suitable statistical methods. In silico analysis was performed to infer the intuitive damaging effects of detected variants at transcripts and protein level.

Results We found significant association of TAF7 C16T (MW827584 G > A), RFX2 562delT (MZ560629delA), rs11547633 A > C, rs17606721 A > G, MW827583 C > T, and MZ379836 C > T variants with the incidence of azoospermia. In silico analysis predicted that the variants either alter the natural splice junctions of the transcript or cause probable damage in the structure of proteins of respective genes.

Conclusion Polymorphisms/mutations of TAF7 and RFX2 genes increase risk of male infertility in Bengali population. The novel variants may be used as markers for male infertility screening in ART practise.

Keywords Male infertility · Polymorphisms · Spermatogenesis · TAF7 · RFX2

Introduction

Infertility is the inability to conceive a child after a year or more of regular unprotected sexual intercourse. Infertility affects nearly 15% of couples globally, amounting to 48.5 million as estimated in 2015 [1]. Males are found to be solely responsible for 20–30% of infertility cases and contribute to overall 50% of all the reported cases. According to published literature, approximately 50% of reported infertility cases from India are related to the reproductive anomalies or disorders in the male [2]. The cases for which the reasons are mostly unknown are termed as 'idiopathic'. The etiology of idiopathic male infertility is enigmatic owing to its multifactorial nature. It includes Y chromosome microdeletion, mutations and/or polymorphisms of genes regulating spermatogenesis or sex determination signal which have not been studied properly yet. The microdeletion in AZF region of Y chromosome has been identified as common cause of male infertility

✉ Sujay Ghosh
sgzoo@caluniv.ac.in

1 Cytogenetics & Genomics Research Unit, Department of Zoology, University of Calcutta, Taraknath-Palit-Siksha-Prangan, Ballygunge Science College Campus, 35 Ballygunge Circular Road, Kolkata, West Bengal 700019, India

2 Institute of Reproductive Medicine (IRM), HB-36/A/3 1st Cross Rd Bidhannagar, Sector III, Bidhannagar, Kolkata, West Bengal 700106, India

3 Department of Zoology, Bijoy Krishna Girls' College (Affiliated to University of Calcutta), Howrah, West Bengal, India

4 Genome-The Fertility Centre, 61-E, Sarat Bose Road, Kolkata, West Bengal 700025, India

across all ethnicities [3]. Besides, sex chromosome copy number variation and autosomal gene polymorphisms are also been found associated with male infertility [4]. Surprisingly, reports on genetic etiology of male infertility in Bengali population from India are limited and it can be hypothesized that some novel genetic variations may increase the risk of male infertility in this population.

Human testes express several regulatory genes for spermatogenesis. The transcription profiles of the genes differ from one cell type to another during the course of spermatogenesis [5]. It is intuitive that alteration in optimal functioning of the transcription factors may affect the spermatogenesis process and become the underpinning cause of idiopathic male infertility. In this regard, the transcription factors for haploid male germinal cells "***Regulatory Factor X2"*** (**RFX2**) (Ensembl:ENSG00000087903 MIM:142765) is of particular interest. RFX2 is a member of regulatory factor X gene family residing on X chromosome. RFX proteins were identified initially in mammals as the regulatory factor that binds to the X-box motif in MHCII gene promoter [6]. The family consists of seven members (RFX1-7), all of which are characterized by a highly conserved 76-residue winged-helix DNA binding domain [7] with similar DNA-binding specificities [8]. The target sites of these transcription regulators are mostly located at the promoters of Testis-specific genes [9, 10] having a sequence 5'-GTNRCC(0-3 N) RGYAAC-3'. The most important conserved function of the RFX protein is regulation of ciliogenesis. The evidence in favour of implication of RFX2 in spermatogenesis failure has been obtained primarily from the work on mouse model. It has been reported that in RFX2 knockout mouse spermatogenesis process was found to be arrested at the round spermatid phase [11]. This observation was further supported by studies in mice [12]. The role of this gene in ciliogenesis has been revealed in knockdown Xenopus model that exhibited cilia-defective embryonic phenotypes.[13, 14]. However, the role of RFX2 transcription regulator in human spermatogenesis and its implication in male infertility have not been addressed yet through scientific intervention.

Another important transcription activator in mammalian systems is ***TATA-Box Binding Protein Associated Factor 7*** **(TAF7)**(Ensembl:ENSG00000178913 MIM:600573), a component of the TFIID protein complex that binds to the TATA box in class II promoters recruits RNA polymerase II and other factors. Previous study [15] has suggested the role of TAF7l, a paralogue of TAF7 and TBP-related factor 2 (*Trf2*) in regulation of spermiogenesis. But studies are limited in exploring implication of TAF7 in human spermatogenesis impairment and male infertility. To get more insight into the genetic etiology of idiopathic male infertility, we for very first time analysed the variations of RFX2 and TAF7 from the genome of males reported for azoopermia at ART clinic.

Materials and methods

The study design was reviewed and approved by the ethical committee constituted by the University of Calcutta. Ethical compliance was followed as outlined by declaration of Helsinki and ICMR (Indian Council of Medical Research).

Sample cohort

The sample includes 140 control males and 130 azoospermic cases reported to Institute of Reproductive Medicine, Kolkata (IRM). All the cases were proven clinically as 'azoospermia' from the semiograms. The males who were reported for some other health issues to the clinic have been recruited as 'control' in this study. The control males were defined by the criteria of fathering at least one child, have normal seminogram and without any issue related to fertility of the female partner. All the volunteers consented to participate in the study and donated their blood and semen samples. The epidemiological detail and family history of individual male was recorded through personal interview. All the participating males were reported randomly that eliminate chance of any potential bias in sampling. All participating males (cases and controls) belonged to the same ethnic group, Bengali population which is an Indo-Aryan ethnolinguistic group residing in the province of West Bengal in the eastern region of India. Subjects with very poor health condition, suffering from infectious disease, cryptorchidism, malignant or benign testicular tumours, testicular torsion, addicted to anabolic steroids or certain drugs, not able to provide sufficient epidemiological details. Cases with diagnosed female infertility are excluded from the study. All clinical and epidemiological were kept secretly with a code. Seminograms were made following the WHO 2010 guidelines (World Health Organization, 2010). We collected samples of few azoospermic men of Malayali and Gujrati individuals during course of our study, but did not present their data along with the result of Bengali population data to maintain the homogeneity of ethnicity of the participating cases. The detail of these Malayali and Gujrati cases is presented separately in the supplementary table S1.

Tissue collection

Nearly 2 ml of venous blood samples was collected by venipuncture method in EDTA coated vacutainer tubes and were stored at – 20 °C refrigerator for DNA isolation. Blood samples were collected only after obtaining full consent from the participating families. Semen collections were done within 2 days and maximum 7 days from the collection date (World Health Organization, 2010). Approximately, 1 ml of semen

sample was collected from each participating male patient by self-masturbation method in a private place. To categorise the semen samples qualitatively and quantitatively, samples are analysed following liquefaction at IRM laboratory, Kolkata. Semen test with no sperm (no pellets by any means) were perceived as 'azoospermic' (World Health Organization, 2010).

DNA isolation

Genomic DNA was isolated from all blood samples by using QIAamp Blood Mini Kit (QIAGEN, Hilden, Germany) and HiPurA Sperm Genomic DNA Purification Kit (HiMedia, Mumbai, India) according to the manufacturer's instructions. Aliquots were prepared by diluting the DNA with T.E. Buffer in variable amount to achieve a final concentration of 25 ng/µl for use, which is optimum for PCR based sequencing. DNA concentration was measured using the Nanodrop (Thermo Fisher, Waltham, MA).

Microdeletion study

We have cross checked the azoospermic samples to negate the chance of potential Y chromosome microdeletions. We used panel of STS markers (SY121, SY182, DFFRY, SY86, DBY, SY83 for Azf a, SY34, SY124, SY130, RBM-1, RBM-2, SY143, SY134, SY127 for Azf b, BPY2, SY254, SY255, SY239, SY242, SY1197 FOR Azfc to scan the length of Y chromosome.

Genotyping

The DNA concentration is measured by nanophotometer and is diluted to a final concentration of 25 ng/µl. Sanger's dideoxy sequencing was employed for genotyping of the conserved domain of RFX2 and TAF7 reading frame.The conserved domains of respective genes were determined through NCBI conserved domain search(https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi) web-based software. We focused on the conserved domain owing to greater possibility of finding mutations and polymorphisms that may imperil the optimum function of the said genes. The domain specific primers were designed by using integrated DNA technologies (IDT) and Primer3 web software. We analysed sequence of conserved regions from both the genes, TAF7 gene exon 1 and 5´ UTR (Transcript TAF7-201, ID-ENST00000313368.8., CCDS4259) and RFX2 exon 6 (Transcript RFX2-201, ID-ENST00000303657.10, CCDS 12157). The detail of primers and their products are presented in supplementary table S2. The polymerase chain reactions (PCR) were carried out with reaction mixtures (30 µl) containing 15 µl of PROMEGA master mix (comprising PCR buffer, $MgCl_2$, dNTPs & Taq DNA polymerase), 1 µl each forward and reverse primer, 12 µl nuclease free water and 1 µl DNA solution (concentration ~ 25 ng/µl). One cycle consisted of denaturation at 94 °C for 5 min, and then 30 cycles of—94 °C for 30 s, 60 °C (this temperature may vary according to different sets of primers) for 30 s, and 72 °C for 1 min and a final elongation at 72 °C for 5 min. For sequencing, PCR products were analysed using a Taq Dye Deoxy Terminator sequencing kit (Applied Biosystems, Foster City, USA) with an ABI Prism 377 DNA sequencer (Applied Biosystems, Foster City, USA). All six polymorphic variations were deduced by DNA sequencing. The chromatogram data was analysed by FinchTV software. All the sequencing and genotyping were carried out blindly, without knowing the status of semiogram of the subjects. Cross checking of the sequencing results was done with the reference sequences of these genes retrieved from the database and with the help of NCBI BLAST and EMBOSS NEEDLE programme that confirmed the occurrence of specific variants in the genome of the case samples.

In silico functional prediction of polymorphic variants

To predict the probable damaging effect of the genetic variants, several in silico methods were applied. These include **PROVEAN** (**Pro**tein **V**ariation **E**ffect **An**alyzer)[16], predicting impact of amino acid substitution or indel on the biological function of a protein, **PolyPhen-2** (**Poly**morphism **Phen**otyping v**2**)[17] for predicting possible impact of an amino acid substitution on the structure and function of protein, ***Mutation t@ster***,[18] for analysing the disease-causing potential of DNA variants, **Regulation Spotter**[19] to annotate of extratranscriptic variant, **SIFT**(**S**orting **I**ntolerant **F**rom **T**olerant)[20], for analysing effect of amino acid substitution on protein function as functionally neutral or deleterious. Moreover, **Human Splice Finder**[21] and **SpliceAid** [22] was used to predict any change in transcript structure that imperils potential splice sites owing to polymorphic variations or mutations.. A deleterious effects of mutation or variation on protein structure was developed by **Robetta**[23] server (http://robetta.bakerlab.org) and analysed by **Swiss PDB viewer**[24] along with **UCSF Chimera.** This helps to visualize the interactive 3D structure of the protein and to detect any possible structural motif alteration in protein.

Statistical analyses

Pearson's χ^2 test was performed to determine differences in allele and genotype distributions between case and control groups and to verify that allele frequencies were in Hardy–Weinberg equilibrium. Fisher's exact test was performed to calculate Odds ratios with respective

95% confidence interval (CI). Throughout, a two-tailed *P*-value < 0.02 was considered to be of statistically significant after Bonferroni's correction. All the statistical analyses were performed using GraphpadInstat software (GraphpadInstat software, SanDiego, CA) with a significance level of 95%. We have made a genotyping table showing odd ratio, risk ratio, and *P* value.

Results

Microdeletion analysis

We did not find any microdeletion among the case samples. This result suggests involvement some other genetic factors as the cause of azoospermia (Table 1).

Polymorphism analysis

The result of sequencing is presented in the Fig. 1. We genotyped the sequence of Exon 1 and 5´ UTR region of TAF7 gene and exon 6 of RFX2 gene among all 130 case males and found six variations. Out of all detected variants four are found 'novel' and two are found previously reported as far as NCBI genome database is concerned. The four novel variants include three single nucleotide variants and one deletion. We received accession numbers for each 'novel' variants following submission to National Centre for Biotechnology Information (NCBI) Gen-Bank for validation. The data supporting our findings are available at (https://www.ncbi.nlm.nih.gov/). The novel variants are MW827584, MW827583, MZ379836 and MZ560629. We found MW827584 G > A (Risk Allele), MW827583 C > T (Risk Allele), MZ379836 C > T (Risk Allele), rs11547633 A > C (Risk Allele), rs17606721 A > G (Risk Allele) in TAF7 and a deletion MZ560629 delA in RFX2 that causes frame shift in RFX2 reading frame. (Fig. 1) The anticipated changes in the amino acid sequence of TAF7 for MW827584 G > A is Aspartic acid to Tyrosine. The other variants MW827583 C > T, MZ379836 C > T, rs11547633 A > C and rs17606721 A > G all are located in the 5´ UTR regions of TAF7 and probably alters the regulatory feature of the promoter region, which in turn affect transcription of the gene. The deletion and frameshift mutation MZ560629 delA of RFX2 causes generation of premature termination signal at exon 6 and intuitively results in synthesis of truncated protein (Fig. 2). The wild-type RFX2 protein is composed of 729 amino acids, but this frameshift mutation causes production of shorter 212 amino acid long RFX2 protein that lacks several secondary motifs and domains (Fig. 3).

The genomic locations along with HGVS nomenclature corresponding altered cDNA sequence and altered amino acid data of the single nucleotide variations/mutations are given at Table 2. All the HGVS nomenclature of the variants are generated by Position Converter tool of LUMC Mutalyzer program (https://mutalyzer.nl/position-converter?) and Variant Validator tool (https://variantvalidator.org/service/validate/).

Genotypes and risk association

The genotypes and their association with azoospermia is presented in the Table 1. We found significant association of all the polymorphic variants (minor alleles) and mutations with the incidence of azoospermia and so we can consider them as 'risk alleles'. We calculated the odd ratio for each heterozygous and homozygous genotype considering the homozygous major allele genotype as the reference.

The novel variant MW827584 G > A (NM_005642.2:c.16C > T; Asp6Tyr) in exon 1 of the TAF7 gene for which we reported minor allele 'A' has shown elevated risk of azoospermic against the odd 5.065 (95% CI = 3.019–8.497, RR = 2.199, RR 95% CI = 1.676–2.885, $p < 0.0001$). We have not found any homozygous 'AA' genotype for this variant in the study population. This missense variant causes replacement of aspartic acid with tyrosine at the position 6th (Table 2). The novel intronic variant MW827583 C > T (NM_005642.2:c.-295G > A), with its minor allele 'T' has exhibited significant association with azoospermia in heterozygous (CT genotype) condition against the odd 4.003 (OR 95% CI = 2.402–6.672, RR = 1.903,RR 95% CI = 1.495–2.424, $p < 0.0001$).We did not found any homozygous TT genotype for this variant in the sample cohort. Another Novel intronic variant MZ379836 C > T (NM_005642.2:c.-170C > T) with its minor allele 'T' has also shown significant association with spermatogenic failure in heterozygous (genotype CT) condition against the odd 5.254 (OR 95% CI = 3.061–9.017, RR = 2.152, RR 95% CI = 1.638–2.825, *p* value < 0.0001). We did not find any homozygous TT for this variant in the sample cohort as well. We found deletion MZ560629 delA of RFX2 only among the case sample cohort. Out of 130 azoospermic cases, we recorded only 6 cases with this novel deletion in heterozygous condition. The deletion causes occurrence of a premature termination codon which results in production of a truncated transcript (Fig. 2) which may subject to nonsense-mediated mRNA decay (NMD).

The variants rs11547633 A > C (NM_005642.2:c.-229A > C) and rs17606721 A > G (NM_005642.2:c.-68 T > C), both exhibited significant risk association with the incidence of azoospermia. The. SNP rs11547633 in both heterozygous genotype AC and homozygous genotype CC elevates the risk of spermatogenic failure against the odds 4.834 (OR 95% CI = 2.794–8.364, RR = 2.067, RR 95%

Table 1 Distribution of TAF7 and RFX2 risk variants among the azoospermic ($N=130$) and control samples ($N=140$). Foot Note: † *C.I*, confidence interval. *Significance after Bonferroni's Correction for *P* value; Corrected *P* value is 0.02

VARIANTS	Gene	TYPE	LOCATION	AMINO ACID CHANGE	BASE POSITION	GENOTYPE	CONTROL	CASE	ODDS RATIO	95% C.I†	P VALUE	Relative risk (RR)	C.I of RR†
							N = 140	**N = 130**					
MW827584 G > A	TAF7	Missense variant	Exon 1	D6Y	5:141320029	GG	98	41	Reference				
			(ENSE00001388076)			GA	42	89	5.065	3.019–8.497	< 0.0001*	2.199	1.676–2.885
						AA	0	0					
rs11547633 A > C	TAF7	5' UTR VARIANT	5' UTR	N.A	5:141320273	AA	83	32	Reference				
			(ENSR00000187846)			AC	44	82	4.834	2.794–8.364	< 0.0001*	2.067	1.587–2.691
						CC	0	0					
rs17606721 A > G	TAF7	5' UTR VARIANT	5' UTR	N.A	5:141320112	AA	93	34	Reference				
			(ENSR00000187846)			AG	47	96	5.587	3.303–9.450	< 0.0001*	2.228	1.723–2.880
						GG	0	0					
MW827583 C > T	TAF7	5' UTR variant	Exon 1 + 5' UTR	N.A	5:141320339	CC	86	37	Reference				
			(ENSR00000187846)			CT	54	93	4.003	2.402–6.672	< 0.0001*	1.903	1.495–2.424
						TT	0	0					
MZ379836 C > T	TAF7	5' UTR variant	Exon 1 + 5' UTR	N.A	5:141320214	CC	97	36	Reference				
			(ENSR00000187846)			CT	40	78	5.254	3.061–9.017	< 0.0001*	2.152	1.638–2.825
						TT	0	0					
MZ560629 delA	RFX2	frameshift variant	Exon6	GITS..> GSHH…	19:6026199	AA	140	0	Reference				
			(ENSE00003790276)	[Frameshift]		A_	134	6	-	-	-	-	-
						_ _	0	0					

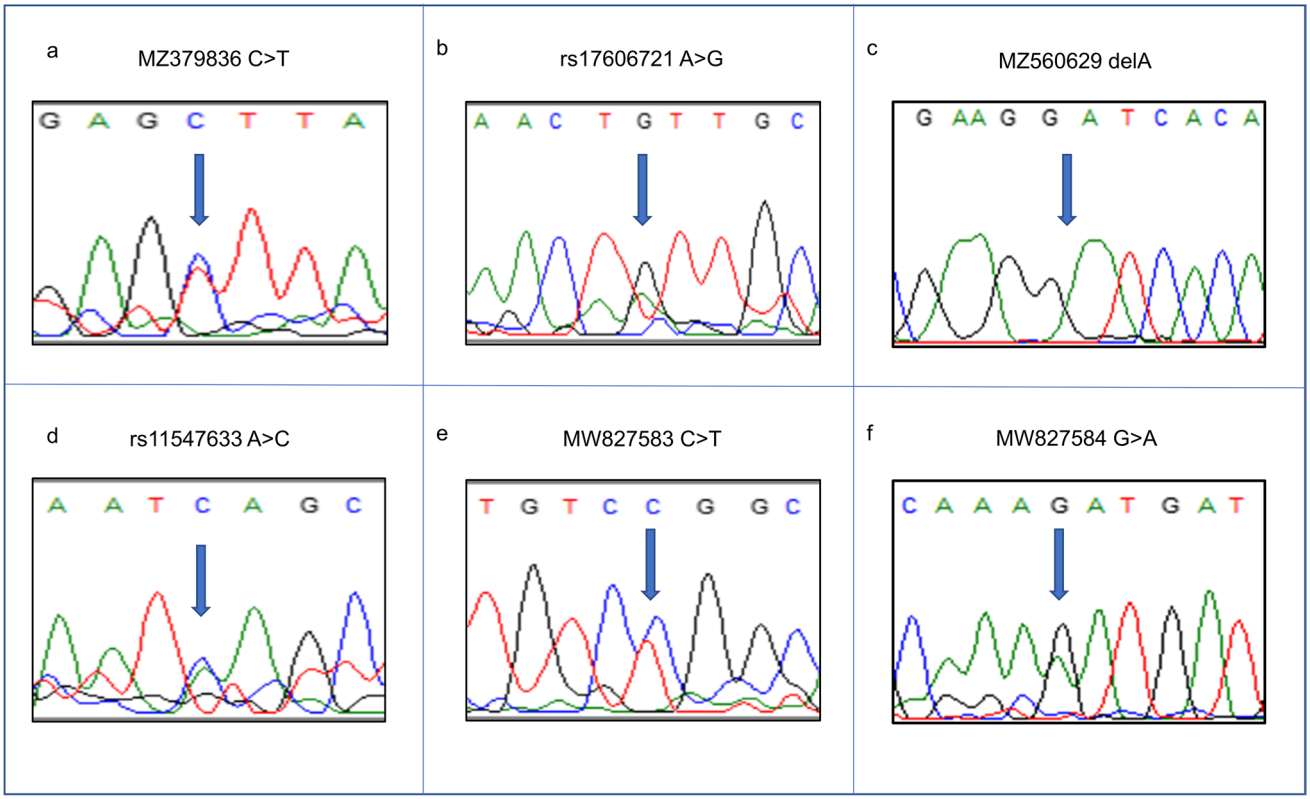


Fig. 1 Chromatograms show genotypes MZ379836 CT (a), rs17606721 AG (b), MZ560629 delA (c), rs11547633 AC (d), MW827583 CT (e), and MW827584 GA (f)

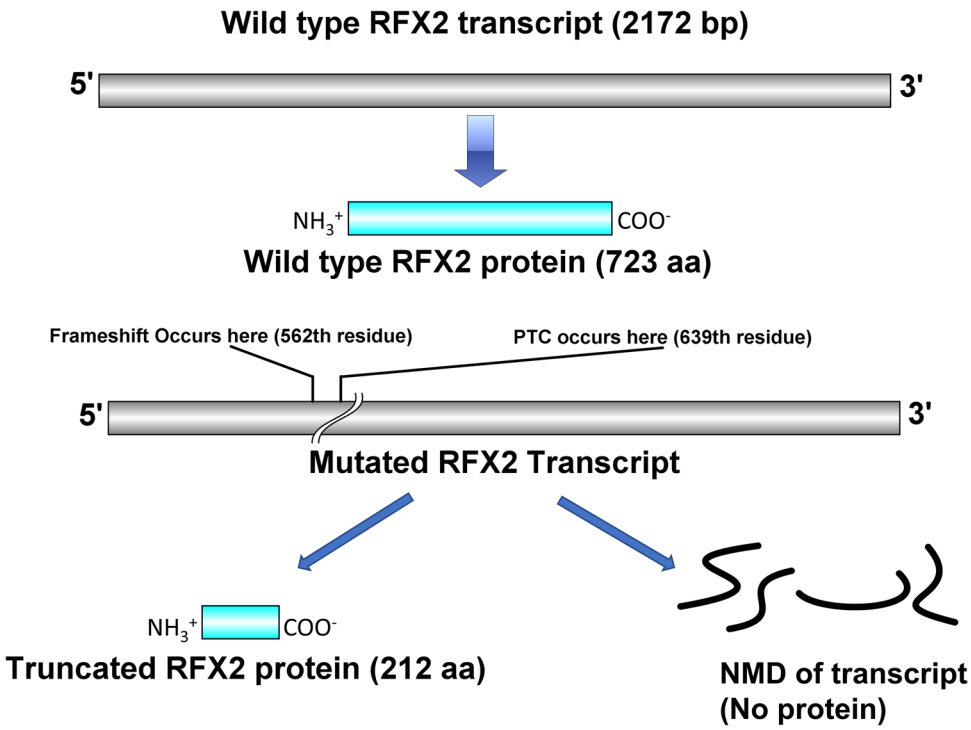


Fig. 2 Illustration of wild type (723 amino acid) and truncated RFX2 protein (212 amino acid). Due to MZ560629 delA mutation, frameshift occurs at 187th residue of RFX2 protein, resulted in appearance of PTC at 212th residue, causes production of non-functional truncated RFX2 protein

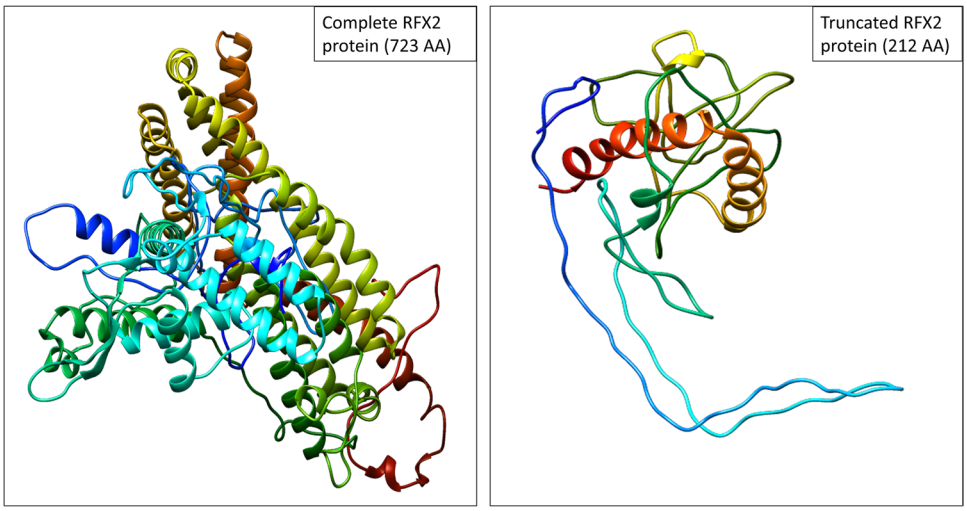


Fig. 3 3D structure of complete RFX2 protein and Truncated RFX2 protein, designed by UCSF Chimera software

Table 2 Genetic variations of TAF7 & and RFX2 gene with HGVS nomenclature, characteristic of the alteration, respective amino acid change and both wild and mutant cDNA sequence snippet

Variants	Gene	Type	Amino Acid change	cDNA sequence snippet
MW827584 G > A (NM_005642.2:c.16C > T)	Taf7	Missense variant	D6Y	Wild Type: TAAAGATGAGTAAAAGCAAAGATGATGCTC CTCACGAACTG Mutant Type: TAAAGATGAGTAAAAGCAAATATGATG CTCCTCACGAACTG
rs11547633 A > C (NM_005642.2:c.-229A > C)	Taf7	5' UTR VARIANT	N.A	Wild Type: ATGAACTTAAGTCTGTGAATAAGCCTTTGT GTTAACGACTG Mutant Type: ATGAACTTAAGTCTGTGAATCAGCCTT TGTGTTAACGACTG
rs17606721 A > G (NM_005642.2:c.-68 T > C)	Taf7	5' UTR VARIANT	N.A	Wild Type: GCAAACTGGTGTTTTAAACTATTGCAGTAG CTGGAACTTTT Mutant Type: GCAAACTGGTGTTTTAAACTGTTGCAG TAGCTGGAACTTTT
MW827583 C > T (NM_005642.2:c.-295G > A)	Taf7	5' UTR VARIANT	N.A	Wild Type: CTAGCTCTCTCCGCGTTGTCCGGCAGCGGC ACCTAGAGGTT Mutant Type: CTAGCTCTCTCCGCGTTGTCAGGCAGC GGCACCTAGAGGTT
MZ379836 C > T (NM_005642.2:c.-170C > T)	Taf7	5' UTR VARIANT	N.A	Wild Type: TTTAGAGAAAAGACTTGGAGCTTAAATAAA AACTAAGGCAA Mutant Type: TTTAGAGAAAAGACTTGGAGATTAAAT AAAAACTAAGGCAA
MZ560629 delA (NM_000635.3:c.562delT)	RFX2	frameshift variant	GITS.. > GSHH	Wild type: AACCTCCAAAAAAGCGAAGGAATCACATCA CACAAAAGCGG Altered type: AACCTCCAAAAAAGCGAAGGATCACAT CACACAAAAGCGG

CI = 1.587–2.691, p value = < 0.0001) and 5.188 (OR 95% CI = 2.376–11.327, RR = 2.165, RR 95% CI = 1.369–3.424, p value = < 0.0001), respectively.

In addition to Bengali population, few cases of Malayali and Gujrati infertile men were also analysed similarly and compared with fertile men of identical ethnicity (Table S2). We observed association of all the polymorphisms and mutation with these infertile groups too, though small sample size limits to draw any inference.

In silico analyses of polymorphic variants

We conducted in silico analyses of the variants and deletion mutation to predict probable damaging effects on transcript and proteins. All these programs use different algorithms to infer the deleterious effect of polymorphism or mutation; so, the result may be counterintuitive. We have considered any given variant as 'damaging' when any two of these detected it as damaging or having deleterious effect. The summary

of the results of analyses by PolyPhen-2, SIFT, PROVEAN, and SNP&GO is provided in the Table 3.

Prediction of polyphen-2 software

Polyphen-2 predicted TAF7 exonic variant MW827584 G > A (NM_005642.2:c.16C > T) (Asp6Tyr) as damaging (according to HumDiv model- 'Probably damaging', score -0.997, according to HumVar model- 'Possibly damaging', score-0.819).

Prediction of SIFT software

The SIFT predicted TAF7 exonic variant MW827584 G > A (NM_005642.2:c.16C > T) (Asp6Tyr) as 'damaging' with a score 0.

Prediction of PROVEAN software

The PROVEAN program predicted TAF7 exonic variant MW827584 G > A (Asp6Tyr) as 'deleterious' with a score of -6.18.

Prediction of mutation T@ster software

Mutation taster predicted MW827584 G > A (NM_005642.2:c.16C > T) as 'disease causing' with probability score 0.99999995926752 (simple_aae), rs11547633 A > C (NM_005642.2:c.-229A > C) as 'polymorphism' with probability score 0.999969688491278(without_aae), rs17606721 A > G (NM_005642.2:c.-68 T > C) as 'polymorphism' with probability score 0.999926065832026 (without_aae), MW827583 C > T (NM_005642.2:c.-295G > A) as 'polymorphism' with probability score 0.999694031420779 (without_aae), MZ379836 C > T (NM_005642.2:c.-170C > T) as 'polymorphism' with a score of 0.999970070131294 and MZ560629 delA (NM_000635.3:c.562delT) as 'Disease Causing' with a probability score of 1.00 (complex_aae) and suggests NMD decay. The outcome of the analyses is presented in the Table 4.

Prediction of regulation spotter software

Regulation Spotter suggested MW827584 G > A (NM_005642.2:c.16C > T) as 'Likely effect, functional region (much evidence)' with a score 53.58 (extra-transcriptic) and its occurrence in the TAF7 exon1 results in histone modification (in 3 + cell lines), might affect genomic interaction and resides within transcription factor binding site (TFBS).

The variant rs11547633 A > C (NM_005642.2:c.-229A > C) was predicted as 'known variant polymorphism'

Table 3 Predicted outcome of the polymorphisms found in TAF7 and RFX2 gene using bioinformatics tools PolyPhen-2, SIFT, PROVEAN and SNPs&GO

Variants	Gene	Type	Amino Acid change	Polyphen-2			SIFT		PROVEAN		SNPs&GO	
				Status	Hum Var value	Hum Div Value	Status	Score	Status	Score	Status	Score
MW827584 G > A	Taf7	Missense variant	D6Y	Damaging	0.819	0.997	Damaging	0	Deleterious	-6.18	Disease	0.521
rs11547633 A > C	Taf7	5' UTR VARIANT	N.A									
rs17606721 A > G	Taf7	5' UTR VARIANT	N.A									
MW827583 C > T	Taf7	5' UTR VARIANT	N.A									
MZ379836 C > T	Taf7	5' UTR VARIANT	N.A									
MZ560629 delA	RFX2	frameshift variant	GITS.. > GSHH				Neutral, Causes NMD	0.915				

Table 4 Predicted outcome of the polymorphisms found in TAF7 and RFX2 gene using bioinformatics tools Mutation Taster and RegulationSpotter

Variants	Gene	Amino acid change	Mutation Taster			Regulation Spotter		
			Prediction	Model	Probability	Status	Model	Probability
MW827584 G > A (NM_005642.2:c.16C > T)	TAF7	D6Y	Disease Causing	simple_aae	0.99	Likely effect: Functional region (much evidence)	Extratranscriptic	Score: 53.58
rs11547633 A > C (NM_005642.2:c.-229A > C)	TAF7	N.A	Polymorphism	without_aae	0.99	Known variant: Polymorphism	Extratranscriptic	Score: 82.72
rs17606721 A > G (NM_005642.2:c.-68 T > C)	TAF7	N.A	Polymorphism	without_aae	0.99	Known Variant: Polymorphism	Extratranscriptic	Score: 47.08
MW827583 C > T (NM_005642.2:c.-295G > A)	TAF7	N.A	Polymorphism	without_aae	0.99	Likely effect: Functional region (much evidence)	Extratranscriptic	Score: 103.50
MZ379836 C > T (NM_005642.2:c.-170C > T)	TAF7	N.A	Polymorphism	without_aae	0.99	Likely effect: Functional region (much evidence)	Extratranscriptic	Score: 73.97
MZ560629 delA (NM_000635.3:c.562delT)	RFX2	GITS.. > GSHH	Disease Causing	complex_aae	1	Likely effect: Non Functional Region	Extratranscriptic	Score: 17.28

with a score 82.72 (extratranscriptic) and its occurrence at 5´ UTR region(within active Ensembl promoter of ENSG00000255729 and TAF7 gene) probably results in histone modification (in 3 + cell lines), belongs to DNAse1 and H3K4me3 epigenetic marked region. It shows active coexistence with CTCF, TAF1, Yy1, Rad21 binding region.

The variant rs17606721 A > G (NM_005642.2:c.-68 T > C) was predicted as 'known variant polymorphism' with a score 47.08 (extra-transcriptic). Being present in the promoter region of TAF7 gene and ENSG00000255729 regulatory region, it plays significant role in histone modification (in 3 + cell lines) and might affect genomic interactions as it is present in DNAse1 open chromatin and within CTCF binding site.

The polymorphisms MW827583 C > T (NM_005642.2:c.-295G > A) and MZ379836 C > T (NM_005642.2:c.-170C > T) were treated as 'Likely effect functional region (much evidence)' with a score of 103.50 and 73.97, respectively (extratranscriptic) and their presence at 5´ UTR region as well as in the promoter region of ENSG00000255729 and TAF7 gene causes histone modification (in 3 + cell lines). Both showed positive interaction with CTCF, TAF1,Yy1, Rad21 binding region potentially interacts with DNAse1 and H3K4me3 epigenetic marks. The summary of the outcome is presented in the Table 4.

Prediction of human splicing finder (HSF)

Human splicing finder (HSF) [21] predicted whether the polymorphisms and mutations affect the splicing of respective transcripts. The summary of the outcome is presented in Table 5. The exonic variant MW827584 G > A (NM_005642.2:c.16C > T) breaks exonic splice enhance (ESE) sequence and creates mutant motif. Overall, it alters auxiliary sequence and the ratio between exonic splice enhancer and exonic splicing silencer (ESS) (-7). The variant rs17606721 A > G (NM_005642.2:c.-68 T > C) breaks ESE_ site and create respective mutant motifs. It alters auxiliary sequence and ESE/ESS motifs ratio (-9). Similarly the variant MZ379836 C > T (NM_005642.2:c.-170C > T) breaks EIE, and build mutant motifs. It alters auxiliary sequence and ESE/ESS motifs ratio (-3). The deletion MZ560629 delA (NM_000635.3:c.562delT) creates a new splice site (REF: CTCCAAAAAAGCGA, score: 15.22, -ALT: CCTCCAAAAAAGCG, score: 68.82). It activates a cryptic acceptor site, causing potential alteration

Table 5 Predicted outcome of the polymorphisms found in TAF7 and RFX2 gene using bioinformatics tool human splicing finder (HSF)

Variants	Type of New Splice Site				Type of New Broken Site		
	Reference Motif value	Linked protein (with Score)	Mutant Motif	Status	Reference motif	Mutant Motif	Status
Taf7 MW827584 G > A (NM_005642.2:c.16C > T)	-	MBNL1(-1) SLM-2(-2) Sam68(-5)	-	-	ESS_hnRNPA1 (New ESS Site) RESCUE ESE (ESE Site Broken) EIE (New ESE Site) RESCUE ESE (ESE Site Broken) RESCUE ESE (ESE Site Broken) ESE_9G8 (ESE Site Broken) RESCUE ESE (ESE Site Broken) EIE (ESE Site Broken) Sironi_motif1 (ESS Site Broken) RESCUE ESE (ESE Site Broken) EIE (ESE Site Broken)	TATGAT GATGAT TATGAT AGATGA AAGATG AAAGAT AAAGAT AAAGAT AAAGATGA CAAAGA CAAAGA	**Alteration of Auxiliary sequence** **Significant alteration of ESE / ESS motifs ratio (-7)**
Taf7 rs11547633 A > C (NM_005642.2:c.-229A > C)	-	-	-		-	-	-
Taf7 rs17606721 A > G (NM_005642.2:c.-68 T > C)	-	SC35(+7) ETR-3(+5) SLM-2(-5) Sam68(-5)	-	-	ESE_Tra2 (ESE Site Broken) EIE (ESE Site Broken) ESE_9G8 (ESE Site Broken) EIE (ESE Site Broken) ESE_SRp55 (New ESE Site) EIE (ESE Site Broken) RESCUE ESE (ESE Site Broken) EIE (ESE Site Broken) EIE (ESE Site Broken) ESE_ASF (ESE Site Broken) ESE_ASFB (ESE Site Broken)	AATAT AATATG GAATAT GAATAT AGCATA AGAATA CAGAAT CAGAAT CCAGAA CTCCAGA CTCCAGA	**Alteration of Auxiliary sequence** **Significant alteration of ESE / ESS motifs ratio (-9)**
Taf7 MW827583 C > T (NM_005642.2:c.-295G > A)	-		-	-	-	-	-

Table 5 (continued)

Variants	Type of New Splice Site				Type of New Broken Site		
	Reference Motif value	Linked protein (with Score)	Mutant Motif	Status	Reference motif	Mutant Motif	Status
Taf7 MZ379836 C > T (NM_005642.2:c.-170C > T)	-	SLM-2(+10) Sam68(-10) SRp30C(+1) SC35(+8)	-	**-**	EIE (ESE Site Broken) Fas ESS (New ESS Site) PESE (ESE Site Broken) EIE (ESE Site Broken) IIE (ESS Site Broken) ESE_SRp40 (New ESE Site) Sironi_motif2 (ESS Site Broken) PESE (ESE Site Broken) ESS_hnRNPA1 (ESS Site Broken) ESE_ASF (ESE Site Broken) ESE_ASFB (ESE Site Broken) Sironi_motif1 (ESS Site Broken) EIE (ESE Site Broken)	GAGTGG TAGTGG GAGTGGAC AGAGTG AGAGTG ATAGTGG AGAGTGG AGAGTGGA CAGAGT CAGAGTG CAGAGTG TCAGAGTG GTCAGA	**Alteration of Auxiliary sequence** **Significant alteration of ESE / ESS motifs ratio (-3)**
RFX2 MZ560629 delA (NM_000635.3:c.562delT)	- REF: CTCCAA AAAAGCGA (15.22)	Sam68(-5) SLM-2(-8) SRp30C(+2) SRp20(+3) PSF(+5) NOVA-2(+10) HTra2beta1 (+5) hnRNP M(+5)	-ALT: CCT CCAAAA AAGCG (68.82)	**Activation of a cryptic Acceptor site. Potential alteration of splicing (variation: 352.17%)**	**-**	**-**	-

of splicing (variation: 352.17%). Human splice finder suggests rs11547633 A > C (NM_005642.2:c.-229A > C) and MW827583 C > T (NM_005642.2:c.-295G > A) have no impact on splicing.

Prediction of SpliceAid software

The 'SpliceAid' program [22] provides a database of strictly experimentally assessed target RNA sequences in humans and signifies the effect of altered binding sites for splicing factors. Altered Splicing event can give rise to production of altered splice variant of a gene. The altered 'A' allele of exonic variant MW827584 (NM_005642.2:c.16C > T) creates binding site for SLM-2, Sam68 whereas there was only MBNL1 binding site at the wild type. Similarly, the 'C' allele of the SNP rs11547633 A > C (NM_005642.2:c.-229A > C) creates binding site for SLM-2 and Sam68 in contrast to wild type splice sites for ETR-3, TIAL-1, TIA-1, NOVA-1 binding. The 'G' allele of the rs17606721 A > G (NM_005642.2:c.-68 T > C) creates a new splice site for ETR-3 binding, whereas wild type site is target for SLM-2, Sam68 and SC35. The 'T' allele of MW827583 C > T (NM_005642.2:c.-295G > A) didn't alter the wild type splice sites (dedicated for YB-1, 9G8, SRp40, ETR-3) too much except creating a binding site for SRp30C.The 'T' allele of MZ379836 C > T (NM_005642.2:c.-170C > T) produces a new splice site for SC35 along with the wild type splice sites for SRp30C, Sam68 and SLM-2 binding. For MZ560629 delA (NM_000635.3:c.562delT), the wild type sequence have splice sites for Sam68, SLM-2, hnRNP M, PSF, HTra-2beta1, SRp30c, SRp20 and NOVA- 2; but the deletion causes destruction of SRp30C binding site and creation of new splice site for HTra2alpha, hnRNP H1 and hnRNP H2 (Figs. 4, 5, 6).

Structure analysis of wild-type and mutant proteins through molecular modelling

We used Phyre2 homology modelling tool and Robetta web server to produce the 3D structure of TAF7 and RFX2 protein. As there were no dedicated PDB ID for both of the proteins, we have developed 3D model by using the FASTA amino acid sequence derived from Ensembl database (http://www.ensembl.org/). The mutation of TAF7 MW827584G > A and the 562delT Frameshift mutation of RFX2 (MZ560629 del A) were individually replaced with the wild type sequence to generate the altered protein model.

We predicted a change in 3D configuration of TAF7 protein owing to MW827584 G > A variation that results in substitution of aspartic acid with tyrosine at 6th position. The altered amino acid is bigger in size and more hydrophobic than the normal wild-type residue and resulting in incorrect protein folding. therefore, the local conformation will be destabilized. Using Swiss PDB Viewer program we predicted change in structural stability of TAF7. Upon computing electrostatic potential, we found the orientation of the acidic area (red mesh in Fig. 6) has been changed along with the amino acid change. Electrostatic energy around the mutant residue is higher ($E_{Tyr} = +92.428$ kJ/mol) than the wild type residue ($E_{Asp} = +16.212$ kJ/mol) which results to slight overall energy change of the mutated TAF7 protein ($E_{Total=}$ -13687.530 kJ/mol).

The wild-type RFX2 protein consists 723 amino acids and the 562delT frameshift mutation causes the production of a truncated version of the RNA which may generate a shorter peptide of 212 amino acid long and have lost functional domains (Fig. 3). This is due to the appearance of a premature stop codon at 639th base at the coding region. Mutation Taster predicts that the mutated protein has lost DNA_BIND RFX-type winged-helix (location:199-274th amino acid residue) and a modified phosphoserine residue (location: 416th amino acid residue) (Fig. 2). Swiss PDB viewer predicted that the force field energy around the mutated RFX2 protein is -5644.217 kJ/mol whereas the wild type RFX2 protein has a force field energy value of -26885.752 kJ/mol.

Discussion and conclusion

The genetic etiology of idiopathic male infertility is intriguing. Beside Y chromosome microdeletion that constitutes a proportion of underpinning causes of spermatogenic failure across the population divides, point mutation and variations of genes that regulate spermatogenesis may constitute rest of the proportions. But identification of these variants is difficult owing to multifactorial nature of spermatogenesis regulation and their cryptic expression under variable genomic backdrop. Some of these variants are population specific and some others are ubiquitous. We initiated the study with the working hypothesis that variations or mutation in transcription regulators of spermatogenesis process may affects the optimum fertility among men. Following review of literatures, we preferred to focus of two transcription factors that are involved in regulation of transcription of many target genes linked with haploid spermatocyte maturation. Both of these genes were reported for spermatogenesis failure in mouse model in earlier studies [11, 12, 25], but attempt to explore their role in infertility in human is limited.

We report for the first time ever that some novel polymorphisms and mutations of TAF7 (TATA-Box Binding Protein Associated Factor 7) and RFX2(regulatory factor X2) gene increase risk of azoospermia among men of West Bengal,

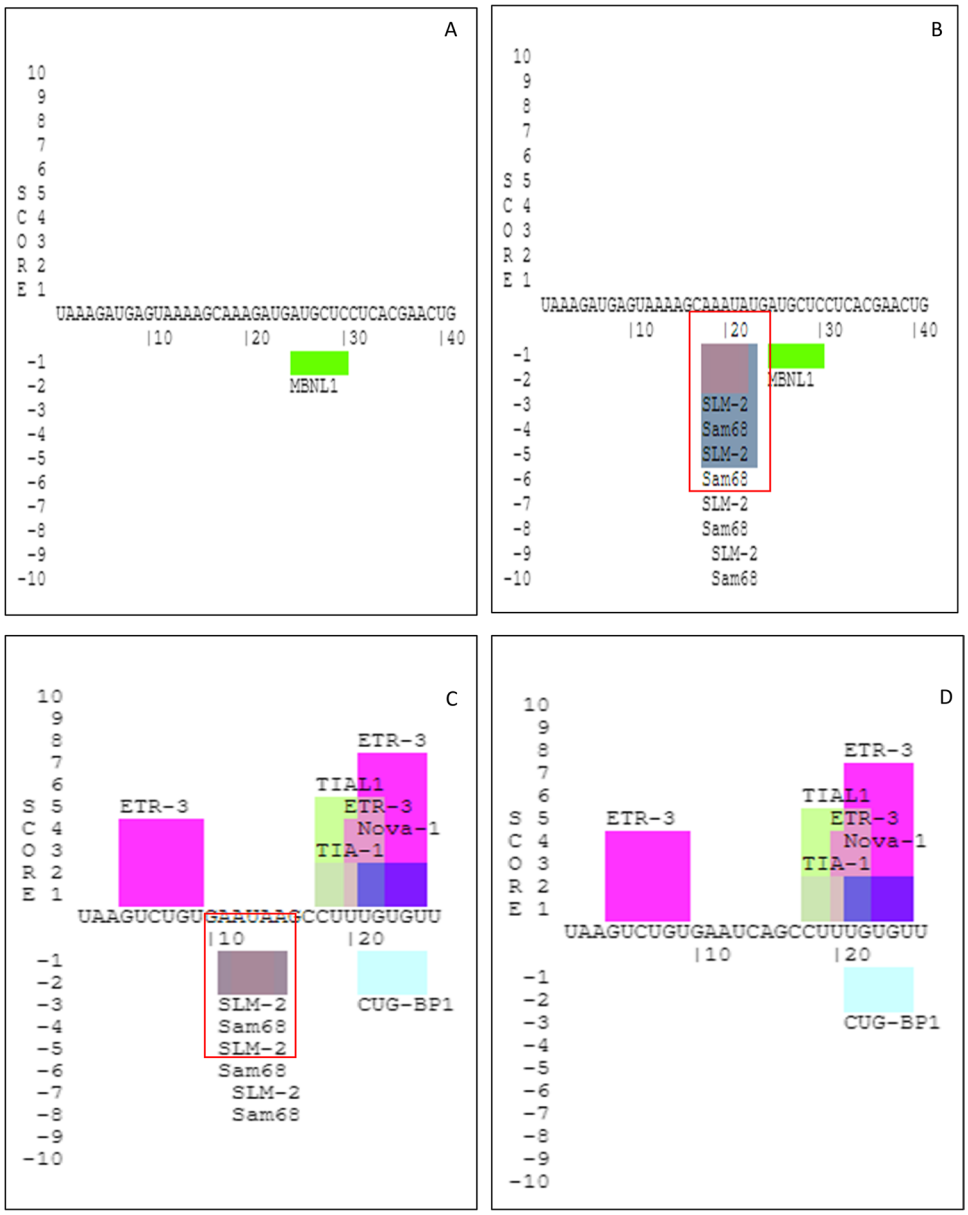
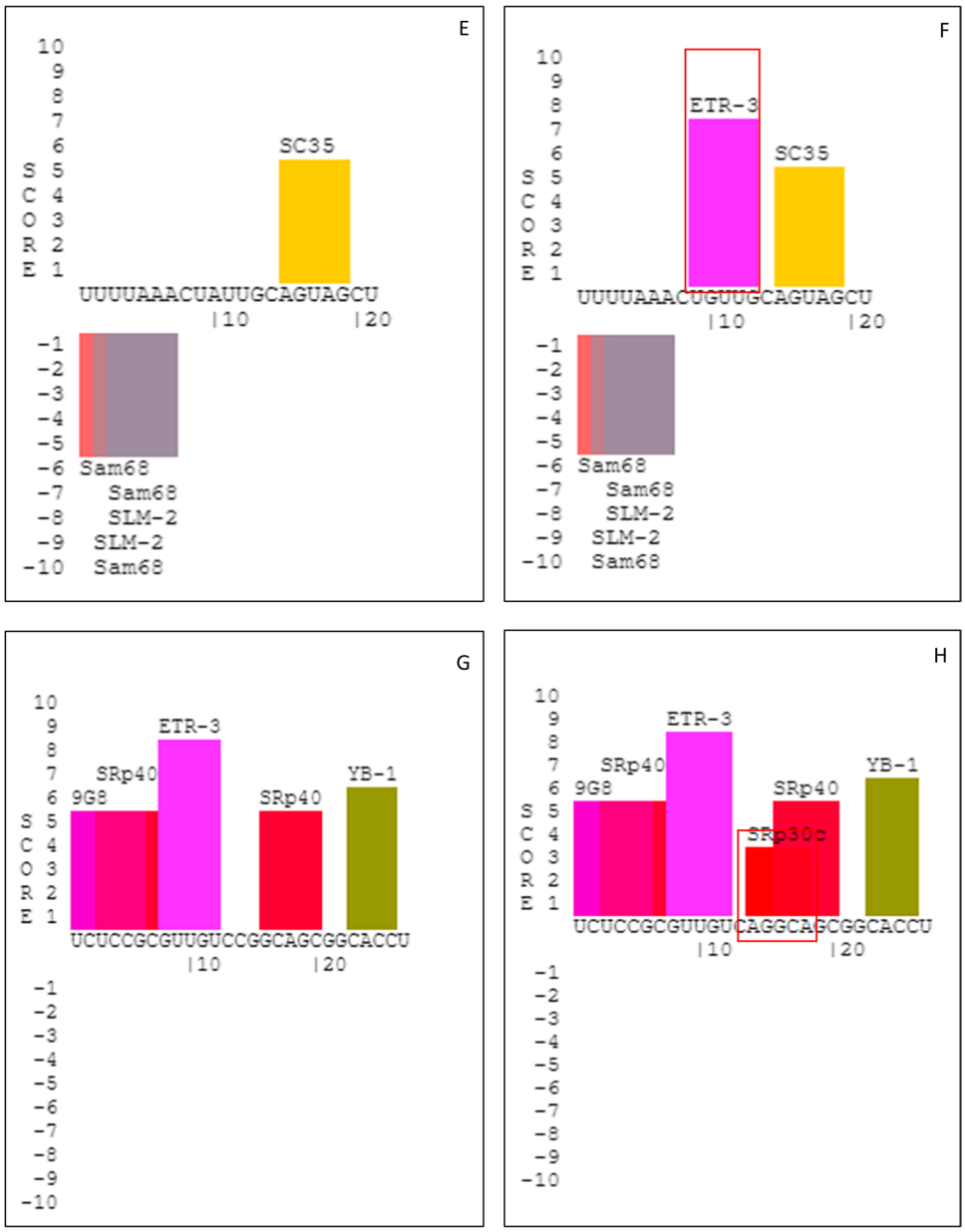


Fig. 4 Results show predicted splicing factor binding generated with Spliceaid software. SpliceAid results for MW827584 G > A reference sequence (**a**) and mutated sequence (**b**), rs11547633 A > C reference sequence (**c**), and mutated sequence (**d**), results for rs17606721 A > G reference sequence (**e**) and mutated sequence (**f**), MW827583 C > T reference sequence (**g**) and mutated sequence (**h**), results for rs17606721 A > G reference sequence (**e**) and mutated sequence (**f**), MW827583 C > T reference sequence (**g**) and mutated sequence (**h**), results for MZ379836 C > T reference sequence (**i**) and mutated sequence (**j**), MZ560629 delA reference sequence (**k**) and mutated sequence (**l**)

India. Beside novel variants and mutation that we report first time, we found significant association of two previously reported variants, rs1154763 A > C (NM_005642.2:c.-229A > C) and rs17606721 A > G, both of which exhibited many fold increase in risk of azoospermia in case cohort. Interestingly, occurrence of the variant rs1154763 A > C (NM_005642.2:c.-229A > C) is new to any Asian population as far as data from gnomAD[26] is concerned. The other variant rs17606721 A > G (NM_005642.2:c.-68 T > C), which has merged with the variant rs2075264 is known to occur among 16% of the south Asian population only.

The MW827584 G > A (NM_005642.2:c.16C > T) was the only missense variant found in our study cohort and this transition causes Aspartic acid to Tyrosine replacement and predicted as 'disease causing' risk factor for infertility among men. Occurrence of the MW827583 C > T (NM_005642.2:c.-295G > A) and the MZ379836 C > T (NM_005642.2:c.-170C > T) within core promoter sequence ENSR00000187846 of TAF7 (Chromosome 5:141,317,801–141,322,200) shows significant alteration in regulatory feature that may disrupt gene expression. The deletion MZ560629 delA (NM_000635.3:c.562delT) leads to a novel frameshift mutation of RFX2 transcript that may result into generation of truncated RNA that either translated into a shorter form of peptide as predicted from structural analysis or subjected to non-sense mediated decay as predicted from Mutation T@ster program. We found this mutation only among azoospermic male and not among the healthy controls. It is difficult at this point to ascertain the exact effect of this deletion mutation on the RNA as the

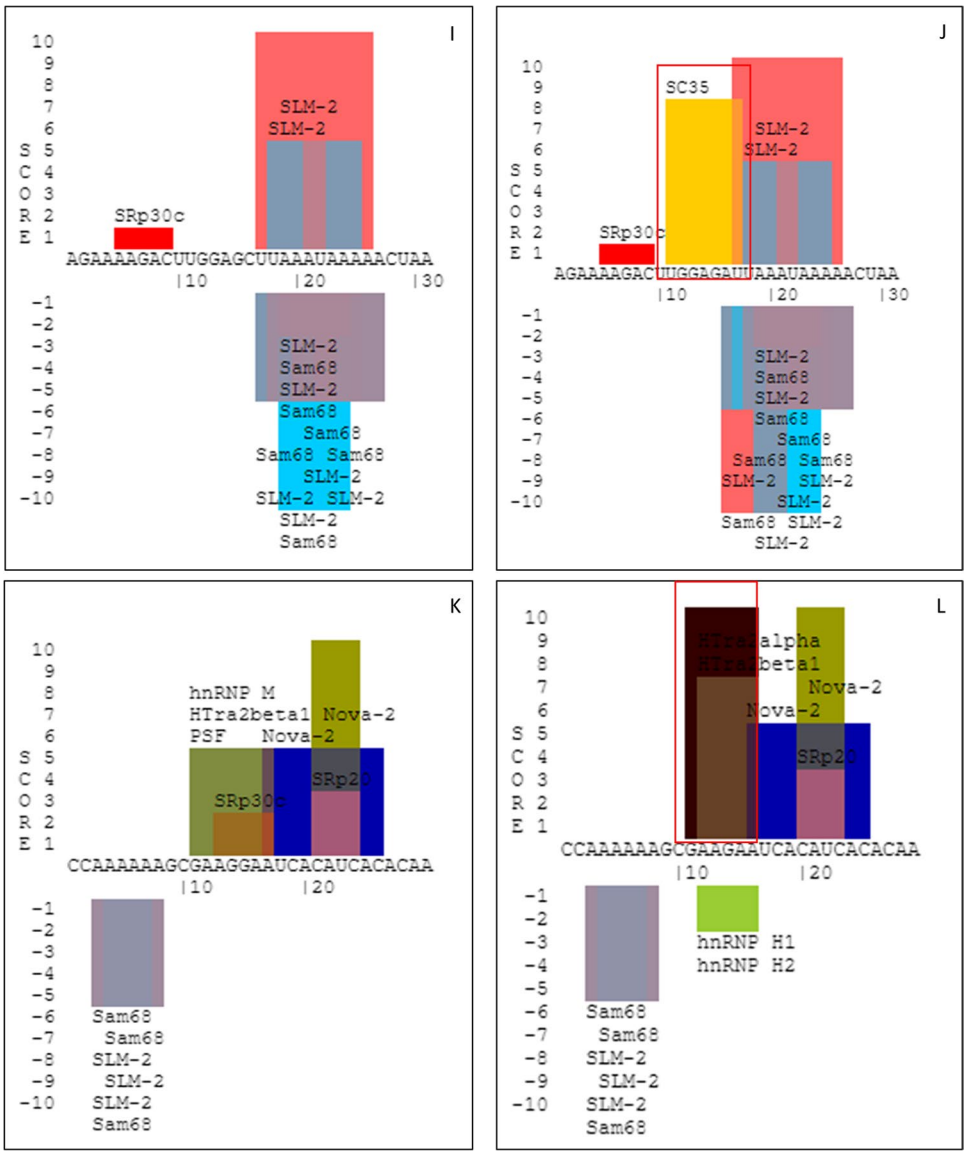
Fig. 4 (continued)



predictions by two programs are apparently contradictory. We can infer at least that the mutation imperils optimum function of wild-type RFX2 in either of these two ways.

We used in silico approach to infer the damaging effect of all the six variants and mutations. This approach helps us to validate the outcome of genetic association study. The missense variant TAF7 MW827584 G > A (NM_005642.2:c.16C > T) causes substitution of aspartic acid by tyrosine that may leads to protein denaturation[27] owing to change in biochemical properties of the peptide chain. The variants rs11547633 A > C, MW827583 C > T and MZ379836 C > T are located in 5´UTR and overlaps core promoter of TAF7 and enhancer motif EH38E2414630 (chr5:141,320,176–141,320,522). Change in this sequence may affect the local chromatin organization. 'Human Splicing Finder' and 'SpliceAid' programme predicted probable damaging effects on these variants on transcripts of both the genes that may leads to generation of alternate splice variants and suboptimal functioning of the genes. We have observed that MW827584 G > A, rs17606721 A > G, MZ379836 C > T and MZ560629 delA have probable damaging effects on splicing. Additionally, these two proteins with their mutant variants may affects the functioning of other regulators as they have several cross talks with other transcription modifiers as predicted from Cytoscape protein network . It is intuitive that the adverse effects of detected variants and mutations of TAF7 and RFX2 are not only

Fig. 4 (continued)



restricted in altered function of these two genes, but may modify the functional status of other genes as well.

Our study has some potential limitations. First, the sample size is comparatively small; study on larger sample size is warrant to reconfirm this association further. Second, at moment we cannot infer the exact molecular implications of TAF7 and RFX2 polymorphisms and mutations on the process of spermatogenesis. Third, some of the 'controls' from our study cohort carried the polymorphic variants of the genes. We do not have rationale to justify their fatherhood if the variants are deleterious as predicted through in silico analyses. It is possible that variations in individual genetic makeup may generate difference in manifestation of the phenotype 'infertility'. In other words, the polymorphic variants of TAF7 and RFX2 cause infertility in interaction with specific combination other genetic modifiers and their allelic combination. Confounding effects of allelic variation of those genetic modifiers may cause difference in manifestation of fertility state along with the allelic variants of TAF7 and RFX2 genes. These confounding allelic variations of other genes can only be identified by whole genome sequencing which is beyond the scope of present study. Fifth, unavailable transcriptomics and proteomics data of RFX2 gene expression limit us to confirm the fate of MZ560629 delA deletion mutation at molecular level. Sixth, we focused on specific regions RFX2 reading frame in present study and screening of entire reading is warrant.

Nevertheless, our findings provide groundwork for future research that will address fundamental concerns, such as how TAF7 and RFX2 regulate spermatogenesis, how it interacts with other spermatogenic regulators, and how certain mutations induce azoospermia. This study can be replicated in other populations and in mouse models

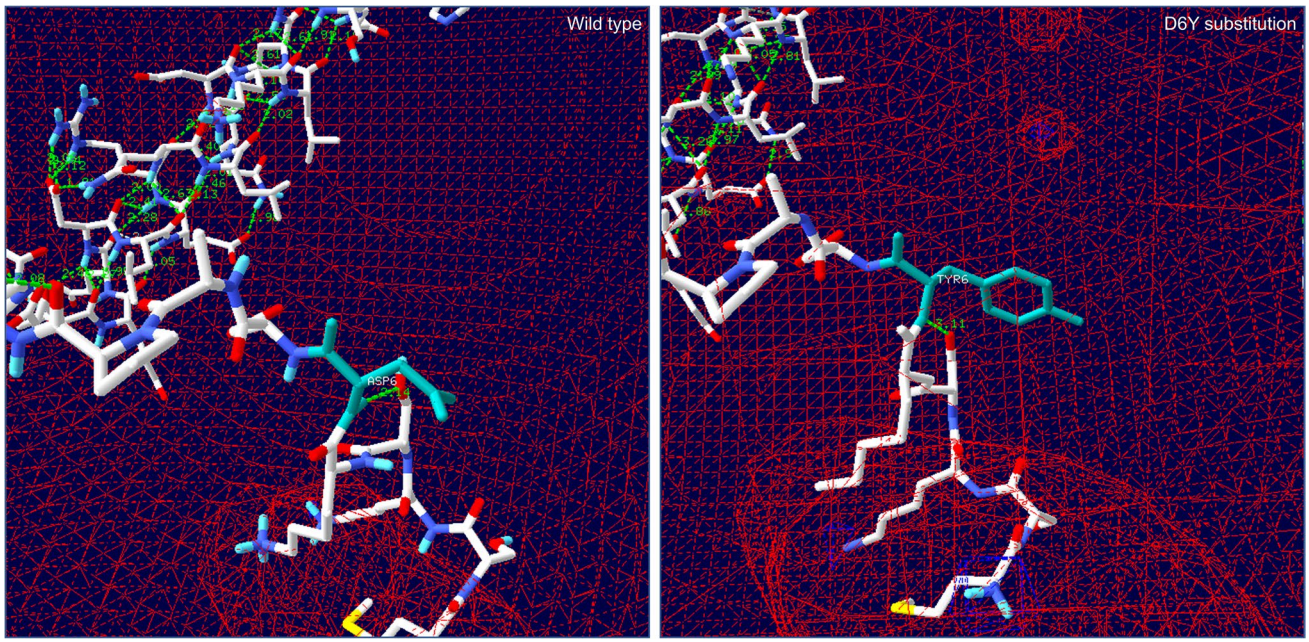


Fig. 5 3D structure of wild type and mutated TAF7. Upon computing mutation effect, the no of hydrogen bonds remains unchanged but the bong length increases from 2.14 Å to 3.11 Å. Upon computing electrostatic potential, the orientation of the acidic area (red mesh) has been changed along with the amino acid change. Aspartic acid substitution with tyrosine causes appearance of basic area (blue mesh) around the residue. Substitution of aspartic acid (negatively charged, acidic amino acid) with tyrosine (neutral amino acid) at 6th residue causes the change of orientation of the electrostatic field around the residue

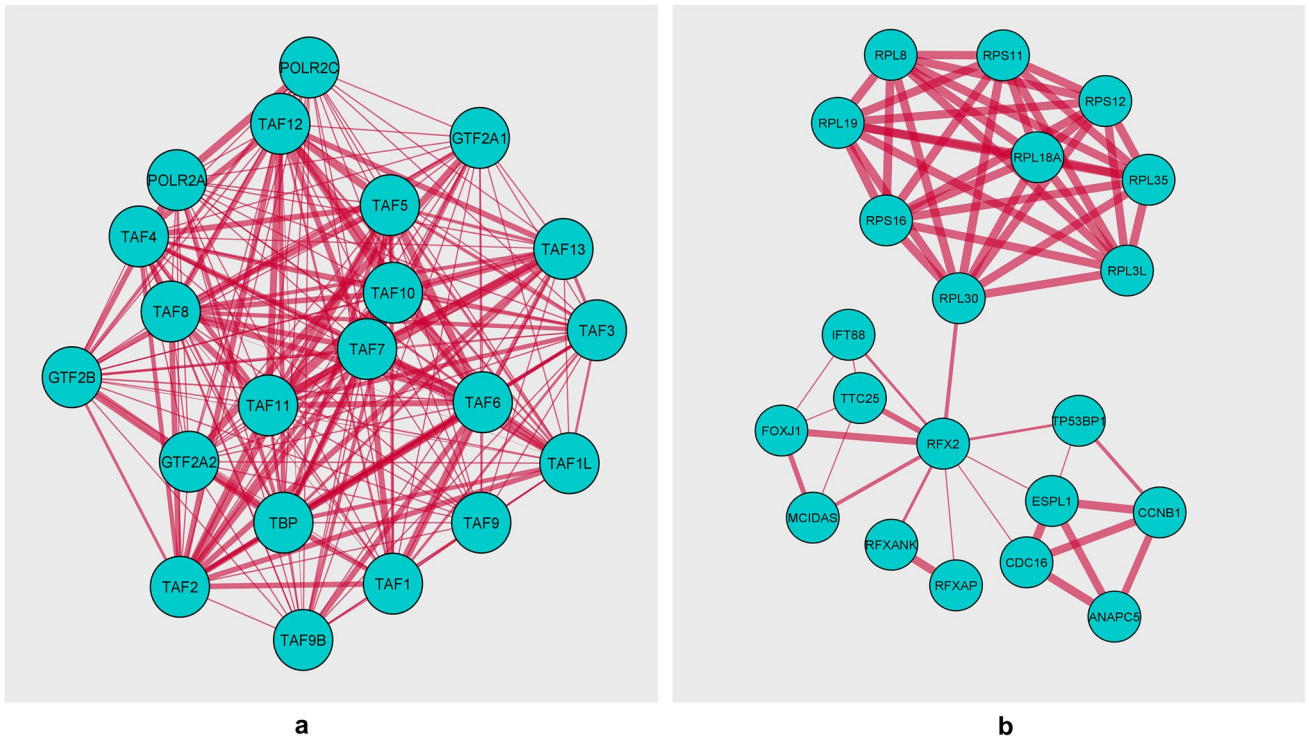


Fig. 6 Protein interaction network of (**a**) TAF7 and (**b**) RFX2 Protein interaction network of (**a**) TAF7 and (**b**) RFX2

with an extended approach to apply ChIP/RNASeq analysis to resolve the issue at incisive level.This study also provide insight into previously unknown genetic risk related to idiopathic male infertility and clinicians may also consider these genes for routine screening before they will offer ART support to an infertile patient.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10815-021-02352-5.

Acknowledgements We are thankful to all the participants who have provided us their tissue samples along with their consent. We are thankful to the entire medical faculties during their active cooperation during sample collection and data recording.

Author contribution Sujay Ghosh conceived and apprehended the project, designed the experiments and wrote the paper. Samudra Pal performed the major experiments, analysed data, reported the novel genetic variants and wrote the paper. Papiya Ghosh, Saurav Dutta and Pranab Paladhi helped in sample collection, experiments and data analysis. Baidyanath Chakravarty, Ratna Chattopadhyay, Gunja Bose and Indranil Saha helped in recruitment of infertile case samples as well as control samples, completed initial diagnosis, performed seminogram to characterize the samples and recorded the epidemiological data.

Funding The project is financed by ICMR (Indian Council for Medical Research), New Delhi, India (Grant No: 5/10/FR/10/2015-RCH).

Samudra Pal is thankful to CSIR (Council of scientific and industrial research) for providing him fellowship.

Data availability The data that support the findings of this study are available on request from the corresponding author.

Code availability Not applicable

Declarations

Ethics approval The study design was reviewed and approved by the institutional ethics committee of the University of Calcutta, West Bengal, Kolkata, India (Dated 04/12/2017; Approval No. CU/BIOETHICS/HUMAN/2306/3044). For working with human subjects, we followed the criteria outlined in the "Declaration of Helsinki" and the Indian Council of Medical Research (ICMR).

Consent to participate All the Samples have been collected after obtaining informed consent from the respective participants under the guidance of trained clinicians. At the laboratory, all records were preserved anonymously and with the utmost confidentiality.

Consent for publication All authors gave their consent to publish this manuscript.

Conflict of interest The authors declare no competing interests.

References

1. Agarwal A, Mulgund A, Hamada A, Chyatte MR. A unique view on male infertility around the globe. Reprod Biol Endocrinol. 2015;26(13):37.
2. Kumar TCA. In vitro fertilization in India. Curr Sci. 2004;86(2):254–6. http://www.jstor.org/stable/24107860. Accessed 25 Jan 2004.
3. Kim SY, Kim HJ, Lee BY, Park SY, Lee HS, Seo JT. Y Chromosome Microdeletions in Infertile Men with Non-obstructive Azoospermia and Severe Oligozoospermia. J Reprod Infertil. 2017;18(3):307–15.
4. Liu S-Y, Zhang C-J, Peng H-Y, Sun H, Lin K-Q, Huang X-Q, et al. Strong association of SLC1A1 and DPF3 gene variants with idiopathic male infertility in Han Chinese. Asian J Androl. 2017;19(4):486–92.
5. Soumillon M, Necsulea A, Weier M, Brawand D, Zhang X, Gu H, et al. Cellular source and mechanisms of high transcriptome complexity in the mammalian testis. Cell Rep. 2013;3(6):2179–90.
6. Reith W, Satola S, Sanchez CH, Amaldi I, Lisowska-Grospierre B, Griscelli C, et al. Congenital immunodeficiency with a regulatory defect in MHC class II gene expression lacks a specific HLA-DR promoter binding protein RF-X. Cell. 1988;53(6):897–906.
7. Emery P, Durand B, Mach B, Reith W. RFX proteins, a novel family of DNA binding proteins conserved in the eukaryotic kingdom. Nucleic Acids Res. 1996;24(5):803–7.
8. Reith W, Ucla C, Barras E, Gaud A, Durand B, Herrero-Sanchez C, et al. RFX1, a transactivator of hepatitis B virus enhancer I, belongs to a novel family of homodimeric and heterodimeric DNA-binding proteins. Mol Cell Biol. 1994;14(2):1230–44.
9. Nelander S, Larsson E, Kristiansson E, Månsson R, Nerman O, Sigvardsson M, et al. Predictive screening for regulators of conserved functional gene modules (gene batteries) in mammals. BMC Genomics. 2005;9(6):68.
10. Smith AD, Sumazin P, Zhang MQ. Tissue-specific regulatory elements in mammalian promoters. Mol Syst Biol. 2007;16(3):73.
11. Kistler WS, Baas D, Lemeille S, Paschaki M, Seguin-Estevez Q, Barras E, et al. RFX2 is a major transcriptional regulator of spermiogenesis. PLoS Genet. 2015;11(7):e1005368.
12. Wu Y, Hu X, Li Z, Wang M, Li S, Wang X, et al. Transcription factor RFX2 is a key regulator of mouse spermiogenesis. Sci Rep. 2016;8(6):20435.
13. Chung M-I, Peyrot SM, LeBoeuf S, Park TJ, McGary KL, Marcotte EM, et al. RFX2 is broadly required for ciliogenesis during vertebrate development. Dev Biol. 2012;363(1):155–65.
14. Bisgrove BW, Makova S, Yost HJ, Brueckner M. RFX2 is essential in the ciliated organ of asymmetry and an RFX2 transgene identifies a population of ciliated cells sufficient for fluid flow. Dev Biol. 2012;363(1):166–78.
15. Zhou H, Grubisic I, Zheng K, He Y, Wang PJ, Kaplan T, et al. Taf7l cooperates with Trf2 to regulate spermiogenesis. Proc Natl Acad Sci USA. 2013;110(42):16886–91.
16. Choi Y, Chan AP. PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. Bioinformatics. 2015;31(16):2745–7.
17. Adzhubei I, Jordan DM, Sunyaev SR. Predicting functional effect of human missense mutations using PolyPhen-2. Curr Protoc Hum Genet. 2013;Chapter 7:Unit7.20.
18. Schwarz JM, Rödelsperger C, Schuelke M, Seelow D. MutationTaster evaluates disease-causing potential of sequence alterations. Nat Methods. 2010;7(8):575–6.
19. Schwarz JM, Hombach D, Köhler S, Cooper DN, Schuelke M, Seelow D. RegulationSpotter: annotation and interpretation of extratranscriptic DNA variants. Nucleic Acids Res. 2019;47(W1):W106–13.
20. Ng PC, Henikoff S. SIFT: Predicting amino acid changes that affect protein function. Nucleic Acids Res. 2003;31(13):3812–4.
21. Desmet F-O, Hamroun D, Lalande M, Collod-Béroud G, Claustres M, Béroud C. Human Splicing Finder: an online bioinformatics tool to predict splicing signals. Nucleic Acids Res. 2009;37(9):e67.

22. Piva F, Giulietti M, Nocchi L, Principato G. SpliceAid: a database of experimental RNA target motifs bound by splicing proteins in humans. Bioinformatics. 2009;25(9):1211–3.
23. Kim DE, Chivian D, Baker D. Protein structure prediction and analysis using the Robetta server. Nucleic Acids Res. 2004;32(Web Server issue):W526-31.
24. Johansson MU, Zoete V, Michielin O, Guex N. Defining and searching for structural motifs using DeepView/Swiss-PdbViewer. BMC Bioinformatics. 2012;23(13):173.
25. Kistler WS, Horvath GC, Dasgupta A, Kistler MK. Differential expression of Rfx1-4 during mouse spermatogenesis. Gene Expr Patterns. 2009;9(7):515–9.
26. Collins RL, Brand H, Karczewski KJ, Zhao X, Alföldi J, Francioli LC, et al. A structural variation reference for medical and population genetics. Nature. 2020;581(7809):444–51.
27. Tanford C. Protein denaturation. C. Theoretical models for the mechanism of denaturation. Adv Protein Chem. 1970;24:1–95.