



Editor's choice
Scan to access more
free content

ORIGINAL ARTICLE

Truncating mutations in *TAF4B* and *ZMYND15* causing recessive azoospermia

Özgecan Ayhan,¹ Mahmut Balkan,² Ayse Guven,¹ Renin Hazan,¹ Murat Atar,³ Atalay Tok,¹ Aslihan Tolun¹

► Additional material is published online only. To view please visit the journal online (<http://dx.doi.org/10.1136/jmedgenet-2013-102102>).

¹Department of Molecular Biology and Genetics, Boğaziçi University, Istanbul, Turkey
²Medical Faculty, Department of Molecular Biology and Genetics, Dicle University, Diyarbakir, Turkey
³Medical Faculty, Department of Urology, Dicle University, Diyarbakir, Turkey

Correspondence to

Professor Aslihan Tolun, Department of Molecular Biology and Genetics, Boğaziçi University, Bebek, Istanbul 34342, Turkey; tolun@boun.edu.tr

Received 8 October 2013
Revised 29 November 2013
Accepted 9 December 2013
Published Online First
15 January 2014

ABSTRACT

Background Azoospermia is the absence of a measurable level of spermatozoa in the semen. It affects approximately 1% of all men, and the genetic basis of the majority of idiopathic cases is unknown. We investigated two unrelated consanguineous families with idiopathic azoospermia. In family 1, there were three azoospermic brothers and one oligozoospermic brother; and in family 2, there were three azoospermic brothers. Testis biopsy in the brothers in family 2 had led to the diagnosis of maturation arrest in the spermatid stage.

Methods Candidate disease loci were found via linkage mapping using data from single nucleotide polymorphism genome scans. Exome sequencing was applied to find the variants at the loci.

Results We identified two candidate loci in each family and homozygous truncating mutations p.R611X in *TAF4B* in family 1 and p.K507Sfs*3 in *ZMYND15* in family 2. We did not detect any mutations in these genes in a cohort of 45 azoospermic and 15 oligozoospermic men. Expression studies for *ZMYND15* showed that the highest expression was in the testis.

Conclusions Both genes are known to have roles in spermatogenesis in mice but neither has been studied in humans. To our knowledge, they are the first genes identified for recessive idiopathic spermatogenic failure in men. Assuming that recessive genes for isolated azoospermia are as numerous in men as in mice, each gene is possibly responsible for only a small fraction of all cases.

INTRODUCTION

Azoospermia is the absence of a measurable amount of sperm in the ejaculate, and oligozoospermia is the term used when the number of sperm in the ejaculate is <15 million/mL.^{1–2} Azoospermia affects approximately 1% of all men and can be caused by either a physical blockage in the genital track, known as obstructive azoospermia, or spermatogenic failure, known as non-obstructive azoospermia. Isolated azoospermia results mostly from spermatogenic failure, some in pretesticular form as a result of mutations in genes related to hormones. The testicular form of isolated spermatogenic failure accounts for approximately 20% of cases of male infertility in populations of European ancestry.³ Identifying genetic factors for idiopathic male infertility will provide valuable insights into the aetiology and aid the development of targeted treatments. Several genes have been identified as responsible for the condition. The best-known gene defects are heterogeneous microdeletions at Yq11,

which may involve several genes within the azoospermia factor (AZF) region;⁴ such microdeletions account for approximately 13% of all azoospermic or severely oligospermic men.^{5–6} Most, but not all, men carrying AZF mutations are affected. In addition, dominant mutations were identified in a few genes on the autosomes by candidate gene approach, that is, screening infertile men for mutations in genes assessed to have roles in spermatogenesis. Two of nineteen azoospermic men with sperm maturation arrest were found to carry the same truncating mutation in *SYCP3* (SPGF4, MIM 270960).⁷ Missense mutations in *NR5A1* were identified in 7 of 315 men with idiopathic spermatogenic failure (SPGF8, MIM 613957).⁸ Splicing and missense mutations in *KLHL10* were identified in 7 of 556 men tested (SPFG11; MIM 615081).⁹ *CFTR* has a special place among the male infertility genes because both gene mutations and polymorphisms can lead either to obstructive azoospermia due to congenital bilateral absence of the vas deferens (CBAVD, MIM 277180) or non-obstructive azoospermia.¹⁰ Variants in several other genes have been shown to be associated with an increased risk of idiopathic spermatogenic failure, including *MTHFR*,¹¹ *SHBG*,¹² *FSHR*¹³ and *YBX2*.^{14–15} as well as *TAF7L* on the X-chromosome.¹⁶ Despite the many genes known to contribute to isolated testicular spermatogenic failure, the total contribution of these genes to the condition is very small. Most of the remaining cases likely have genetic causes as well. We report here the results of a genetic study in two unrelated consanguineous families. By linkage mapping and exome sequencing analysis, we identified truncating mutations in two different genes with mouse orthologs known to have roles in spermatogenesis.

MATERIALS AND METHODS

Families and subjects

Two unrelated families were the subjects of the search for azoospermia genes (figure 1). Family 1 was from a village where inbreeding was common, and the parents were first cousins. All four infertile brothers were available for linkage analysis and one unaffected brother for mutation testing only. None of the infertile brothers agreed to diagnostic testicular biopsy, or testicular sperm extraction (TESE) or testicular sperm aspiration (TESA) for assisted reproduction. They have been married for 23–37 years with no children except that the oligospermic brother has one child (4-years old). The unaffected brother has nine children.

To cite: Ayhan Ö, Balkan M, Guven A, et al. *J Med Genet* 2014;**51**: 239–244.

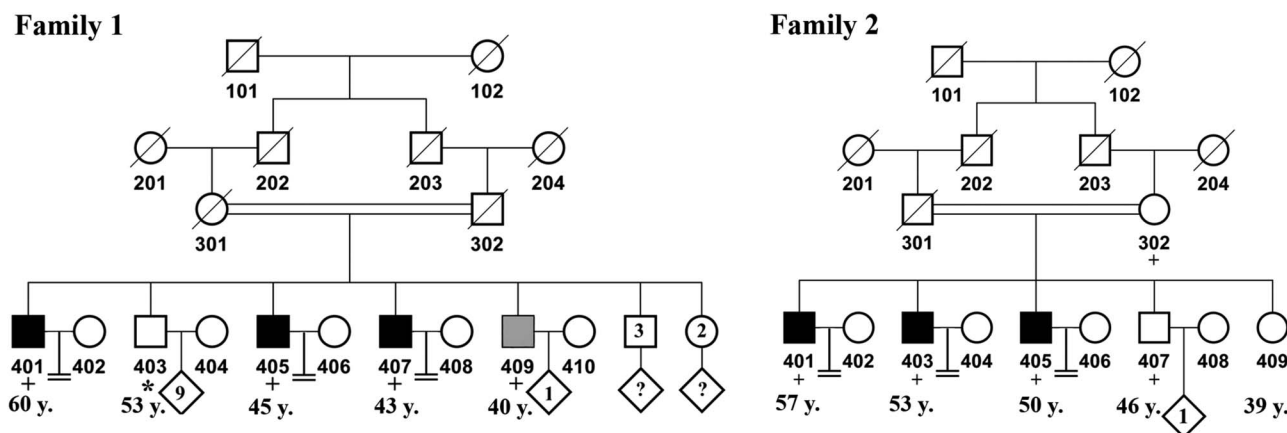


Figure 1 Partial pedigrees of the families. +, individuals included in genome scans; *, individual who underwent mutation testing. The oligospermic brother is shown in grey.

In family 2, the parents were highly consanguineous, both first and second cousins, and consanguinity was reported for all marriages in previous generations. All three affected brothers, the unaffected brother, sister and mother were available for the genetic study. The affected brothers have been married for 20–26 years with no children conceived naturally. In one brother, TESA had resulted in twins, but positive results were not achieved in the others.

A cohort of 45 unrelated azoospermic men and 15 oligozoospermic men were analysed for mutations in three candidate genes. Population control samples were screened for the three candidate mutations; these samples were randomly selected from the old, unrelated DNA collection in our laboratory at Boğaziçi University and used anonymously. Informed consent was obtained from all participants. The Boğaziçi University Institutional Review Board for Research with Human Participants approved the study protocol.

Clinical investigations

Participating members of family 1 and the cohort of infertile men were clinically investigated at Dicle University Medical Faculty Department of Molecular Biology and Genetics and Urology Department. The clinical findings for the cohort were reported previously.¹⁷ All subjects underwent an andrologic investigation that included medical history, physical examination and measurement of hormone levels. Serum concentrations of follicle stimulating hormone (FSH), luteinising hormone and testosterone were measured by electrochemiluminescence immunoassay using Elecsys 1010 (Roche Diagnostics, Germany). Semen analysis was performed according to the WHO guidelines¹ with a computer-assisted semen analysis (CASA) system (SM-CMA), and azoospermia was confirmed by semen analysis (sperm concentration, motility and morphology) under a light microscope according to the WHO guidelines.¹ A diagnosis of non-obstructive azoospermia was based on clinical findings of small testicles plus an increased FSH level. Clinical evaluations for family 2 were based on medical records.

Genotyping and linkage analysis

DNA samples from the four affected brothers in family 1 were subjected to a single nucleotide polymorphism (SNP) genome scan using the Illumina Human 610 Quad BeadChip and five samples (the mother, three affected sons and the unaffected son) from the members of family 2 using the Illumina Human 370

Quad. High parental consanguinity prompted us to assume recessive inheritance in both families. Because the original pedigrees exceeded the capacity of the programme, simplified pedigrees were used that assumed first cousin parents as the only consanguinity. The Allegro v1.2c program in the easyLINKAGE package V.5.08 was used for parametric multipoint logarithm of the odds (LOD) score calculations. We assumed full penetrance and a disease frequency of 0.0001, and selected markers at 0.07 cM intervals in sets of 50. Subsequently, a detailed linkage analysis that included all markers was performed at the loci that yielded relatively high LOD scores. Further, online tool HomozygosityMapper (<http://www.homozygositymapper.org>) was used to detect shared homozygosity regions >100 kb in the affected brothers. In family 2, regions of >100 SNPs, where the unaffected brother shared the homozygosity or where homozygosity, was due to deduced non-informative paternal genotypes were excluded.

Haplotypes were constructed using Allegro. SNP data were formatted on Excel sheets to confirm shared homozygosity in affected brothers. At the loci where azoospermic brothers were homozygous for haplotypes that were possibly identical by descent (IBD) in family 2 for which LOD scores did not reach significance, microsatellite markers were used to investigate descent from a recent common ancestor. The Hg19/GRCh37. p5 map was used throughout the study.

Exome sequencing analysis

DNA samples from one azoospermic man (405) in each family were subjected to exome sequencing. The captured exome library was created using an Illumina TruSeq Capture kit for family 1 and Agilent SureSelect Target Enrichment System for family 2 and sequenced on Illumina HiSeq2000 platforms for both families. Raw reads generated by sequencing were aligned to the reference genome using BWA-0.5.9, and variant calling was performed with SAMtools-0.1.14. The output variants (SNPs and indels) within regions where shared homozygosity was assessed as IBD in affected brothers (and in family 2, but not in unaffected brother) were considered. Those that were reported in the dbSNP and 1000 Genomes databases with frequencies ≥ 0.05 (so that a not-so-rare variant was not missed), alternative depth <60% of total depth, not validated by inspection of sequence reads by BamView, or found in other in-laboratory exome sequencing results were filtered out. The remaining novel/rare variants that were deduced to affect protein structure

(missense, truncating and splicing variants) were considered and prioritised with respect to severity of the mutation, gene function, expression in relevant tissues and evidence in animal models. Predicted severity of missense/splicing and truncating mutations were investigated *in silico* using online tools PolyPhen, MutPred, SIFT and Mutation Taster. Candidate variants that were assessed to possibly contribute to the condition were validated by Sanger sequencing, and inheritance in other family members was investigated.

Mutational analysis

The infertile cohort was screened for mutations in the coding regions of *TAF4B*, *ZMYND15* and *SPAG7* using either single-strand conformational polymorphism assay on polyacrylamide gels or high-resolution melting curve analysis on the LightCycler 480 system (Roche Applied Science, Germany). The sizes of the PCR products for the analyses were <280 bp. A population control panel of 120 individuals was similarly screened for the mutations in each of the genes identified in the study families, which corresponded to a power of 80% to detect a normal sequence variant with a frequency of 0.01.¹⁸

RESULTS

All affected subjects had normal male karyotype and no micro-deletions in the AZF region at Yq11. Serum levels were within normal ranges for luteinising hormone, prolactin and testosterone, but the FSH level was increased. Semen analysis revealed azoospermia.

Clinical findings

Family 1

Four of the eight brothers had reduced sperm levels, three with azoospermia and one with oligozoospermia (figure 1). The sperm count in the oligozoospermic brother was 6 million/mL, and sperm morphology was normal.

Family 2

The three azoospermic brothers had undergone thorough clinical investigation, and the available medical records showed a similar clinical phenotype in all. The results of a bilateral testis biopsy led to the diagnosis of spermatogenesis arrest or, more specifically, maturation arrest in the spermatid stage. There were spermatogonia and spermatocytes in phase I and phase II but very few spermatids and no spermium in the lumen of germ epithelia. Tubulus membranes were partly normal and partly thickened, to the greatest extent in 403. Thickening was so excessive in some regions that the tubules were blocked.

Genetic analyses

The results of the multipoint LOD score calculations are presented in online supplementary figure S1, the candidate loci in table 1, the haplotypes at candidate loci in online supplementary table S1 and the candidate gene variants in online supplementary table S2.

Family 1

Multipoint LOD scores reached 3.01 at 4p16 for the assumed pedigree (see online supplementary figure S1), and haplotype analysis revealed a shared homozygosity region of 14 Mb in the affected brothers (see online supplementary table S1). At 18p11.21–18q21.1, the locus with the next highest score (2.98), shared homozygosity region was 0.23 Mb. A third region at 7q22.1 yielded a score of 2.80, where shared homozygosity was 0.44 Mb. A more detailed linkage analysis using all

markers yielded a maximum LOD score of 3.01 at all three loci (table 1). Haplotype analysis showed that the haplotypes were compatible with IBD from a recent common ancestor at those loci but not at the other three loci that yielded scores >2 (see online supplementary table S1). It seemed that two crossovers in patient 407 at 18q11.1 resulted in a small region (0.23 Mb) of homozygosity. Noticing many such small regions of homozygosity throughout a region of 30.19 Mb at 18p11.21–q12.3 and hypothesising that it could be any of those regions where the putative crossovers occurred, we detected all those small regions using online tool HomozygosityMapper. A similar situation was noted in a 35.15 Mb region at 7q11.21–22.1, where the maximal LOD score was 2.8 only for a small region (0.44 Mb). Also, a 0.39 Mb region at 5p15.31 where linkage analysis did not yield a significant LOD score was found by HomozygosityMapper.

The exome sequencing output listed in total 88 variants at the candidate loci, and all but two of them were filtered out for various reasons: 66 because they had frequencies ≥ 0.05 , 10 were in intergenic or intronic regions or in ncRNA genes, and 10 in UTR regions. The two remaining variants (see online supplementary table S2) were rare missense NM_153717.2 (*EVC*):c.1369G>A (p.E457K) at the larger candidate locus 4p16.2 and novel nonsense variant NM_005640.1(*TAF4B*):c.1831C>T (p.R611X) at 18q11.2 (see online supplementary figure S2). A similar analysis was performed for the three locus with many small regions of shared homozygosity, and 27 variants remained after filtering potentially deleterious variants with allele frequencies <0.05 and alternative depth/total depth ratios >0.6. We investigated whether any of those variants was in a homozygosity region, and no new candidates were found. Online tools predicted the variant in *EVC* as rather benign, and the gene is responsible for recessive skeletal dysplasia (OMIM 225500) and dominant acrofacial bone development (OMIM 193530). In contrast, the other variant is a deleterious, severe mutation in a gene with a role in spermatogenesis in mice. All brothers with compromised fertility were homozygous for it, whereas the unaffected brother tested was heterozygous, and it was not found in the control samples. We concluded that the mutation underlies the disorder in the family.

Family 2

Five loci yielded initial multipoint LOD scores >2 for the assumed pedigree (see online supplementary figure S1), and haplotype analysis did not support IBD in two of them. Detailed analysis at the remaining three loci yielded LOD scores around 2.5 (table 1). At locus 17pter-p22.1, shared homozygosity in the infertile brothers was approximately 7.2 Mb. At 14q11.2–12, the maximal shared homozygosity was approximately 2.8 Mb (see online supplementary table S1). Genotyping with microsatellite markers confirmed possible IBD at both loci. The shared parental haplotype at the remaining locus was very small (<300 kb), indicating that it was likely a common haplotype in the inbred family.¹⁹ No other candidate locus possibly IBD in the affected brothers but not in the unaffected one was detected with the online tool used.

The exome sequencing output listed a total of 824 variants at the candidate loci, all but three of which were filtered out for various reasons: 592 because they were in intergenic or intronic regions or in ncRNAs, 110 had alternative depths <60% of the total depth, 15 were in UTR regions, 57 were synonymous, 43 had frequencies ≥ 0.05 , 3 evaluated as false calls after inspection with BamView and 1 was found in other in-laboratory exome sequence samples. Two of the three candidate variants remaining, all at 17p (see online supplementary table S2), were

Table 1 The candidate loci where only the affected brothers in the respective families are homozygous for a haplotype possibly IBD from a recent ancestor

Locus	LOD score*	Shared patient haplotype			Shared parental haplotype		
		Begins (bp)	Ends (bp)	Size (Mb)	Begins (bp)	Ends (bp)	Size (Mb)
Family 1							
4p16.2	3.01 (3.01)	rs4689888 (4 569 460)	rs17779675 (6 449 514)	1.88	rs4689888 (4 569 460)	rs1400931 (18 752 731)	14.18
7q22.1	2.8 (3.01)	rs4236541 (98 830 439)	rs28365067 (99 272 310)	0.44	rs4729320 (96 551 820)	rs6967487 (99 602 191)	3.05
5p15.31	0 (3.01)	rs186893 (8 473 747)	rs17225452 (8 873 310)	0.39	Same as in patients (three different haplotypes)		0.39
18q11.1	2.98 (3.01)	rs2339102 (23 797 853)	rs1609839 (24 027 419)	0.23	rs12959318 (11 643 985)	rs12456229 (75 118 412)	63.47
Family 2							
17pter-p22.1	2.53 (2.53)	pter (0)	rs3744405 (7 193 255)	7.19	Same as in patients		7.19
14q11.2–12	2.53 (2.53)	rs8010057 (23 506 110)	rs2332572 (26 341 293)	2.83	rs8010057 (23 506 110)	rs1952267 (29 889 699)	6.38
7q36.2	2.48 (2.51)	rs4621714 (153 356 432)	rs10229774 (153 652 188)	0.29	Same as in patients		0.29

*LOD scores obtained after detailed calculations are given in parentheses. IBD, identical by descent; LOD, logarithm of the odds.

missense, NM_014853.2(SGSM2):c.2078G>A (p. R693Q) and NM_004890.2(SPAG7):c.406C>T (p.R136W). The former was assessed to be benign and the latter benign/pathogenic using online tools. We identified the latter by the candidate gene approach before exome sequencing and before *ZMYND15* was reported, and at the time *SPAG7* (sperm-associated antigen 7), a homologue of fox sperm acrosomal protein-1²⁰ seemed a good candidate gene. The remaining variant is a novel exonic deletion of four nucleotides in *ZMYND15* (see online supplementary figure S2), and a much better candidate because the mutation is definitely deleterious and male null mice are infertile. The mutation was deduced to affect all three validated isoforms of the gene, and its designation in isoform 1 is NM_001136046.2 (*ZMYND15*):c.1520_1523delAACA (p.K507Sfs*3). All infertile brothers were homozygous for it, whereas the unaffected brother was heterozygous, and it was not found in the control samples. We concluded that the mutation underlies the disorder in the family.

The cohort of infertile men

The two truncating mutations assessed to underlie the disorders in the families were not found in the cohort of infertile men. The cohort was also screened for mutations in the coding regions of *SPAG7* before exome sequence analysis and later in *TAF4B* and *ZMYND15*. No novel or rare variants were found in any of the subjects.

DISCUSSION

To the best of our knowledge, we have identified two novel genes for isolated testicular spermatogenic failure as the first recessive genes for the condition in man. None of the genes have been studied in humans. Linkage mapping was not straightforward. Simplified pedigrees were used for both families because the capacity of the linkage programme was not sufficient to handle the original pedigrees with multiple parental consanguinities. Such a strategy would overestimate LOD scores but finds the loci that are best candidates. We later used an online tool to detect regions of shared homozygosity in affected brothers. Exome sequencing analysis revealed the causative mutations in both families. The genes are considered to be responsible for infertility in the respective families for several reasons: the mutations are very severe and not found in 180 unrelated subjects (the control panel plus the infertile cohort) tested, and mouse knockout models are azoospermic.^{21 22} We screened the cohort of infertile men for the two genes plus

SPAG7 for mutation but did not find any potentially pathogenic variant.

Nonsense mutation *TAF4B* p.R611X in exon 9 is deduced to truncate the protein product of the validated isoform by 252 residues (the native isoform has 862 amino acids). The truncated protein lacks the histone fold domain (residues 653–702) that increases the DNA binding activity of TAFs²³ plus the TAF12 interaction domain (residues 830–862) that facilitates interaction of the protein with TAF12, which is essential for DNA binding at the core promoters of several genes.²⁴ Mutation *ZMYND15* p.K507Sfs*3 in exon 9 is deduced to shift the translational reading frame and lead to premature termination after the synthesis of two non-native amino acids in all three validated protein isoforms; the mutant protein isoforms lack 236 or 244 native residues (reference isoforms have 742, 703 or 750 amino acids). The truncated proteins lack the Pro-rich domain essential for binding of some signal transduction and cytoskeleton proteins.²⁵

Studies on mouse knockout models have contributed substantially to the identification of genes involved in spermatogenesis, and human orthologs of several such genes were later shown to be associated with human infertility. *TAF4B* and *ZMYND15* identified here are now added to the list of such genes. *TAF4B* has the additional feature that knockout female mice are also infertile due to folliculogenesis failure.²⁶ Future studies could unravel whether defects in the gene also lead to infertility in human females. Male mice deficient in either *Taf4b* or *Zmynd15* exhibit misregulation of several genes required for normal spermatogenesis.^{20 21} *TAF4B*, also called *TAFII105* (RNA polymerase II, TATA box-binding protein-associated factor), has 15 exons and encodes an 862-amino acid protein. UniGene (EST Profiles) reports variable expression in several tissues including testis, whereas Freiman *et al* reported predominant expression in testis and very weak expression in the other organs tested, and found that the protein was enriched in mouse gonadal tissues.²⁶ Falender *et al* reported that *Taf4b*-null young mice were initially fertile but became infertile by 3 months, with impaired gonocyte proliferation and reduced expression of spermatogonial stem cell markers, indicating that the gene protein was required for normal spermatogenic maintenance in the adult.²¹ Despite its severity, homozygous *TAF4B* mutation in the four brothers reported here exhibited phenotypic variability in that one brother was oligospermic and the other three were azoospermic. Yan *et al* reported that *Zmynd15* (zinc finger mynd-containing protein 15) in mice was expressed

exclusively in haploid germ cells.²² We found that, in men, the highest expression of validated isoforms 1 and 2 was in testis and much lower in the other tissues tested (see online supplementary figure S3). Yan *et al* found also that *Zmynd15* expression was during late spermatogenesis, and the protein product was a transcriptional repressor essential for spermiogenesis.²² *Zmynd15*-deficient mice displayed late spermatogenic disruption. Thus, the phenotype of the brothers with homozygous *ZMYND15* mutation presented here is compatible with *ZMYND15* deficiency.

Male infertility is a common and socially important health issue, and isolated spermatogenic failure contributes substantially to the condition. The best-known genetic cause of isolated spermatogenic failure is Yq11 microdeletions, and heterozygous mutations or polymorphisms in a few other genes have been described. Considering that more than 2300 genes are estimated to play a role in human spermatogenesis²⁷ and that knockout mouse models indicate that close to 400 genes are involved in spermatogenesis,²⁸ the number of known genes with roles in monogenic inheritance for spermatogenic failure in men is very small. Perhaps because most research has been conducted on mainly outbred populations, no recessive genes for idiopathic azoospermia had been identified up to now. We have identified two such genes, but neither seems to be responsible for the condition in any of the 60 idiopathic azoospermic/oligozoospermic subjects who originated from eastern Turkey, a region with a high level of inbreeding. Testing a larger group of men with idiopathic spermatogenesis failure could perhaps identify other causative mutations.

Genetic studies on spermatogenesis failure in man are not simple because the trait is practically lethal since the affected individuals do not reproduce. A very large family is needed to localise a dominant trait, and consanguinity is very helpful in the case of a recessive trait. Perhaps because such families are scarce in western countries where most infertility studies are conducted, the traditional strategy used for gene identification (gene localisation and subsequent candidate gene approach) has not been applied, but rather particular genes have been screened for mutations in infertile men. The two genes we identified for spermatogenic failure are the first recessive genes identified for this disorder, and, in the near future, the powerful new generation genetic technologies could facilitate the identification of other recessive genes involved in spermatogenesis, as was the case for *CATSPER1*, which was identified by linkage mapping in two small families with a total of three members with reduced sperm motility, low sperm count and increased abnormally structured spermatozoa.²⁹ Nonetheless, our findings that *TAF4B* and *ZMYND15* are responsible for spermatogenesis failure in men can help to unravel the roles of these genes in male fertility and are expected to stimulate new research on the genes. Our findings that none of the presumably inbred men with idiopathic azoospermia we tested carried mutations in any of these genes support the hypothesis that the genes that are responsible for the disorder are numerous and each gene is responsible for only a small fraction of cases. It seems that the challenge for genetic studies on male infertility will persist in the near future.

Acknowledgements We thank all the individuals who participated in the study for their cooperation.

Contributors AG, ÖA, RH and A Tok generated and analysed genetic data and cowrote the manuscript. MB and MA performed the clinical investigations and cowrote the manuscript. A Tolun supervised the genetic studies, obtained funding and cowrote the manuscript.

Funding This work was supported by the Boğaziçi University Research Fund, grant number 5708, and the Scientific and Technological Research Council of Turkey, grant number 110T252.

Competing interests None.

Patient consent Obtained.

Ethics approval Boğaziçi University Institutional Review Board for Research with Human Participants approved the study protocol.

Provenance and peer review Not commissioned; externally peer reviewed.

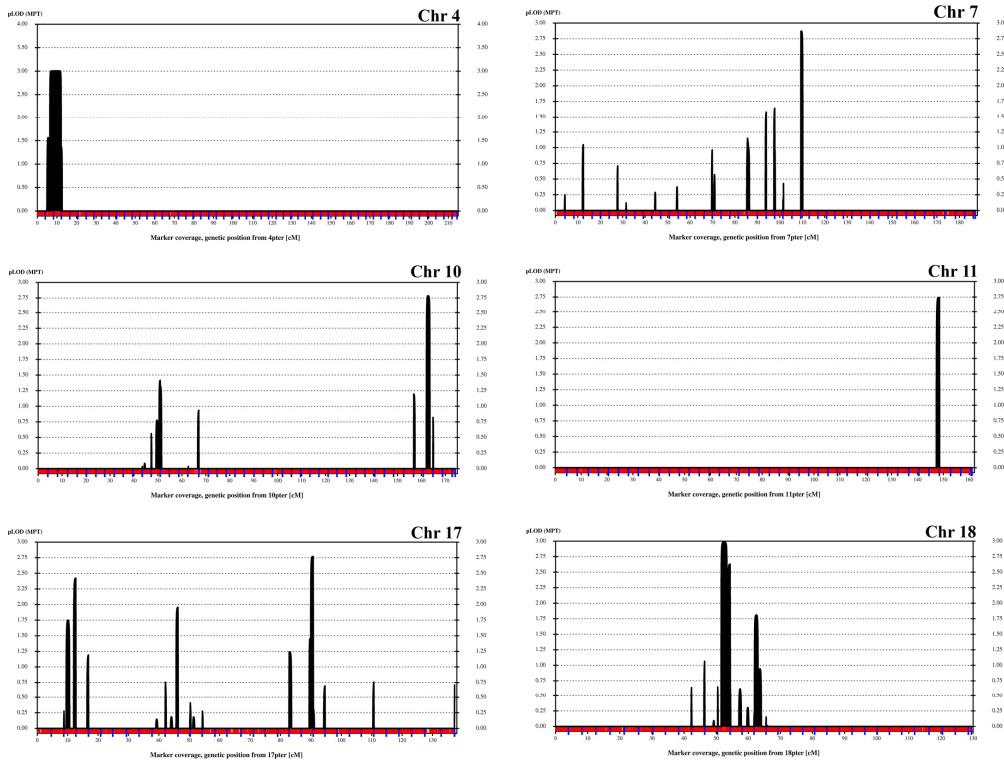
REFERENCES

- 1 World Health Organization. *WHO laboratory manual for the examination of human semen and sperm-cervical mucus interaction*. 4th edn. Cambridge: Cambridge University Press, 1999.
- 2 Jungwirth A, Giwercman A, Tournaye H, Diemer T, Kopa Z, Dohle G, Krausz C. European Association of Urology Working Group on Male Infertility. European Association of Urology guidelines on Male Infertility: the 2012 update. *Eur Urol* 2012;62:324–32.
- 3 Bhasin S, de Kretser DM, Baker HW. Clinical review 64: pathophysiology and natural history of male infertility. *J Clin Endocrinol Metab* 1994;79:1525–9.
- 4 Vogt PH, Edelmann A, Kirsch S, Henegariu O, Hirschmann P, Kiesewetter F, Köhn FM, Schill WB, Farah S, Ramos C, Hartmann M, Hartschuh W, Meschede D, Behre HM, Castel A, Nieschlag E, Weidner W, Gröne HJ, Jung A, Engel W, Haidl G. Human Y chromosome azoospermia factors (AZF) mapped to different subregions in Yq11. *Hum Mol Genet* 1996;5:933–43.
- 5 Reijo R, Lee TY, Salo P, Alagappan R, Brown LG, Rosenberg M, Rozen S, Jaffe T, Straus D, Hovatta O, de la Chapelle A, Silber S, Page DC. Diverse spermatogenic defects in humans caused by Y chromosome deletions encompassing a novel RNA-binding protein gene. *Nat Genet* 1995;10:383–93.
- 6 McLachlan RI, Mallidis C, Ma K, Bhasin S, de Kretser DM. Genetic disorders and spermatogenesis. *Reprod Fertil Dev* 1998;10:97–104.
- 7 Miyamoto T, Hasuike S, Yogev L, Maduro MR, Ishikawa M, Westphal H, Lamb DJ. Azoospermia in patients heterozygous for a mutation in SYCP3. *Lancet* 2003;362:1714–19.
- 8 Bashamboo A, Ferraz-de-Souza B, Lourenco D, Lin L, Sebire NJ, Montjean D, Bignon-Topalovic J, Mandelbaum J, Siffroi J-P, Christin-Maitre S, Radhakrishna U, Rouba H, Ravel C, Seeler J, Achermann JC, McElreavey K. Human male infertility associated with mutations in NR5A1 encoding steroidogenic factor 1. *Am J Hum Genet* 2010;87:505–12.
- 9 Yatsenko AN, Roy A, Chen R, Ma L, Murthy LJ, Yan W, Lamb DJ, Matzuk MM. Non-invasive genetic diagnosis of male infertility using spermatozoal RNA: KLHL10 mutations in oligozoospermic patients impair homodimerization. *Hum Molec Genet* 2006;15:3411–19.
- 10 van der Ven K, Messer L, van der Ven H, Jeyendran RS, Ober C. Cystic fibrosis mutation screening in healthy men with reduced sperm quality. *Hum Reprod* 1996;11:513–17.
- 11 Wu W, Shen O, Qin Y, Lu J, Niu X, Zhou Z, Lu C, Xia Y, Wang S, Wang X. Methylentetrahydrofolate reductase C677 T polymorphism and the risk of male infertility: a meta-analysis. *Int J Androl* 2012;35:18–24.
- 12 Lazaros L, Xita N, Kaponis A, Zikopoulos K, Sofikitis N, Georgiou I. Evidence for association of sex hormone-binding globulin and androgen receptor genes with semen quality. *Andrologia* 2008;40:186–91.
- 13 Tapanainen JS, Aittomäki K, Min J, Vaskivuo T, Huhtaniemi IT. Men homozygous for an inactivating mutation of the follide-stimulating hormone (FSH) receptor gene present variable suppression of spermatogenesis and fertility. *Nat Genet* 1997;15:205–6.
- 14 Deng Y, Zhang W, Su D, Yang Y, Ma Y, Zhang H, Zhang S. Some single nucleotide polymorphisms of MSY2 gene might contribute to susceptibility to spermatogenic impairment in idiopathic infertile men. *Urology* 2008;71:878–82.
- 15 Hammoud S, Emery BR, Dunn D, Weiss RB, Carrell DT. Sequence alterations in the YBX2 gene are associated with male factor infertility. *Fertil Steril* 2009;91:1090–5.
- 16 Akinloye O, Gromoll J, Callies C, Nieschlag E, Simoni M. Mutation analysis of the X-chromosome linked, testis-specific TAF7 L gene in spermatogenic failure. *Andrologia* 2007;39:190–5.
- 17 Balkan M, Tekes S, Gedik A. Cytogenetic and Y chromosome microdeletion screening studies in infertile males with oligozoospermia and azoospermia in Southeast Turkey. *J Assist Reprod Genet* 2008;25:559–65.
- 18 Collins JS, Schwartz CE. Detecting polymorphisms and mutations in candidate genes. *Am J Hum Genet* 2002;71:1251–2.
- 19 Yang W, Wang Z, Wang L, Sham PC, Huang P, Lau YL. Predicting the number and sizes of IBD regions among family members and evaluating the family size requirement for linkage studies. *Eur J Hum Genet* 2008;16:1535–43.
- 20 Mao M, Fu G, Wu JS, Zhang QH, Zhou J, Kan LX, Huang QH, He KL, Gu BW, Han ZG, Shen Y, Gu J, Yu YP, Xu SH, Wang YX, Chen SJ, Chen Z. Identification of genes expressed in human CD34(+) hematopoietic stem/progenitor cells by expressed sequence tags and efficient full-length cDNA cloning. *Proc Natl Acad Sci USA* 1998;95:8175–80.

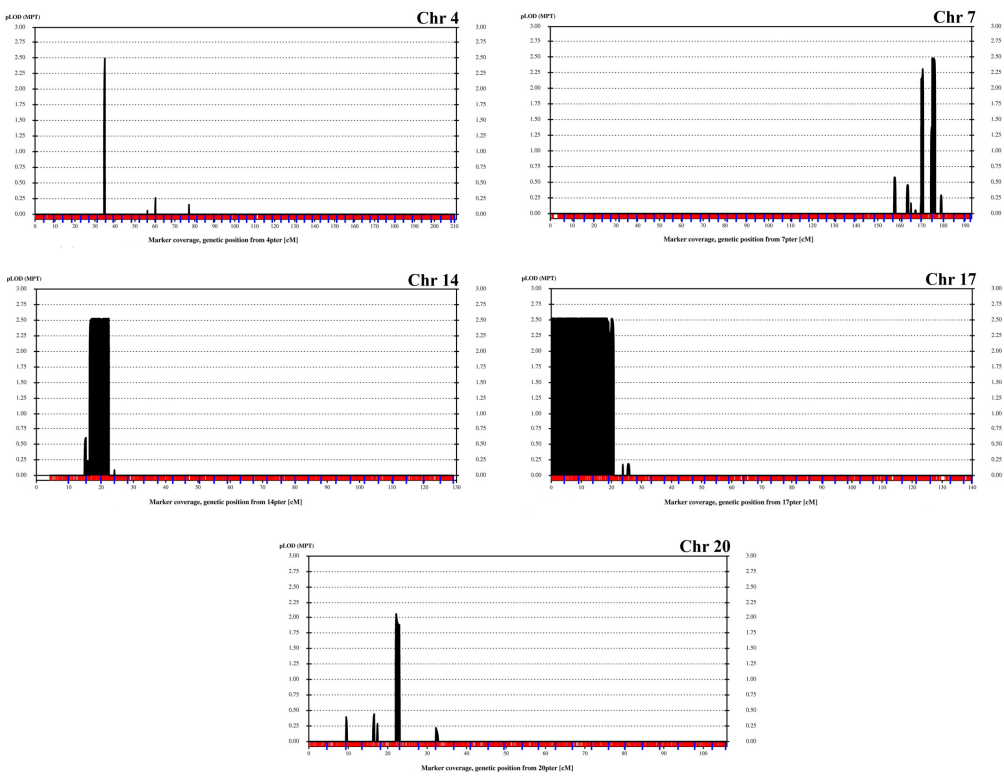
- 21 Falender AE, Freiman RN, Geles KG, Lo KC, Hwang K, Lamb DJ, Morris PL, Tjian R, Richards JS. Maintenance of spermatogenesis requires TAF4b, a gonad-specific subunit of TFIID. *Genes Dev* 2005;19:794–803.
- 22 Yan W, Si Y, Slaymaker S, Li J, Zheng H, Young DL, Aslanian A, Saunders L, Verdin E, Charo IF. Zmynd15 encodes a histone deacetylase-dependent transcriptional repressor essential for spermiogenesis and male fertility. *J Biol Chem* 2010;285:31418–26.
- 23 Shao H, Revach M, Moshonov S, Tzuman Y, Gazit K, Albeck S, Unger T, Dikstein R. Core promoter binding by histone-like TAF complexes. *Mol Cell Biol* 2005;25:206–19.
- 24 Gazit K, Moshonov S, Elfakess R, Sharon M, Mengus G, Davidson I, Dikstein R. TAF4/4b×TAF12 displays a unique mode of DNA binding and is required for core promoter function of a subset of genes. *J Biol Chem* 2009;284:26286–96.
- 25 Kay BK, Williamson MP, Sudol M. The importance of being proline: the interaction of proline-rich motifs in signaling proteins with their cognate domains. *FASEB J* 2000;14:231–41.
- 26 Freiman RN, Albright SR, Zheng S, Sha WC, Hammer RE, Tjian R. Requirement of tissue-selective TBP-associated factor TAFII105 in ovarian development. *Science* 2001;293:2084–7.
- 27 Tamowski S, Aston KI, Carrell DT. The use of transgenic mouse models in the study of male infertility. *Syst Biol Reprod Med* 2010;56:260–73.
- 28 Massart A, Lissens W, Tournaye H, Stouffs K. Genetic causes of spermatogenic failure. *Asian J Androl* 2012;14:40–8.
- 29 Avenarius MR, Hildebrand MS, Zhang Y, Meyer NC, Smith LL, Kahrizi K, Najmabadi H, Smith RJ. Human male infertility caused by mutations in the CATSPER1 channel protein. *Am J Hum Genet* 2009;84:505–10.

Supplementary Figure S1 Multipoint LOD scores based on a recessive model with full penetrance. Only chromosomes that yield scores >2 are presented.

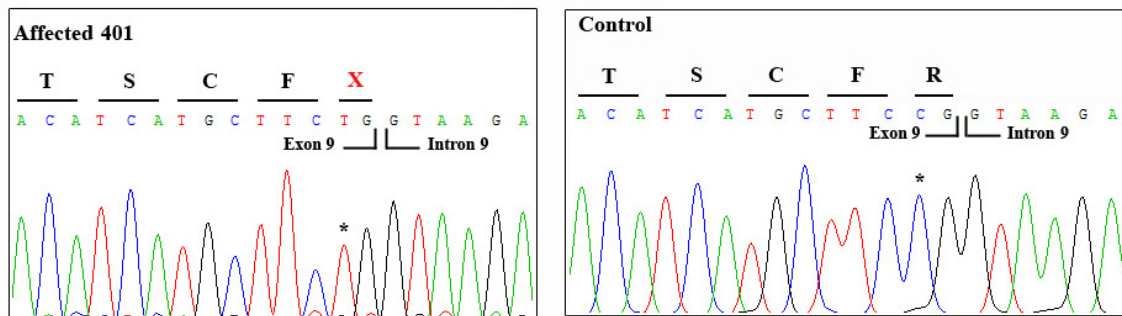
Family 1



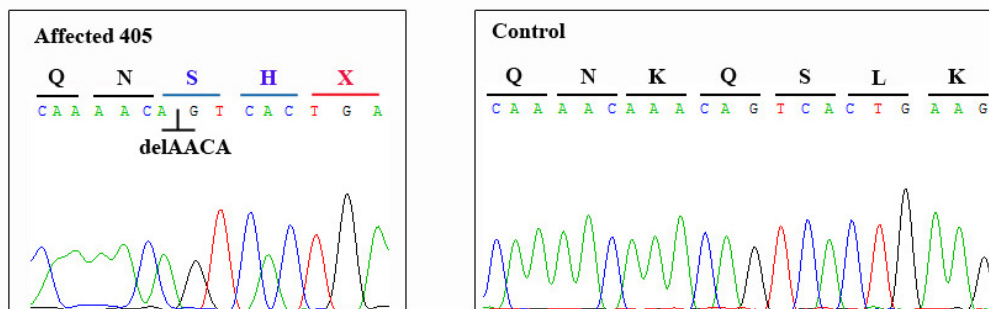
Family 2



Supplementary Figure S2 Chromatograms showing the identified mutations in a homozygous individual and the reference sequence.

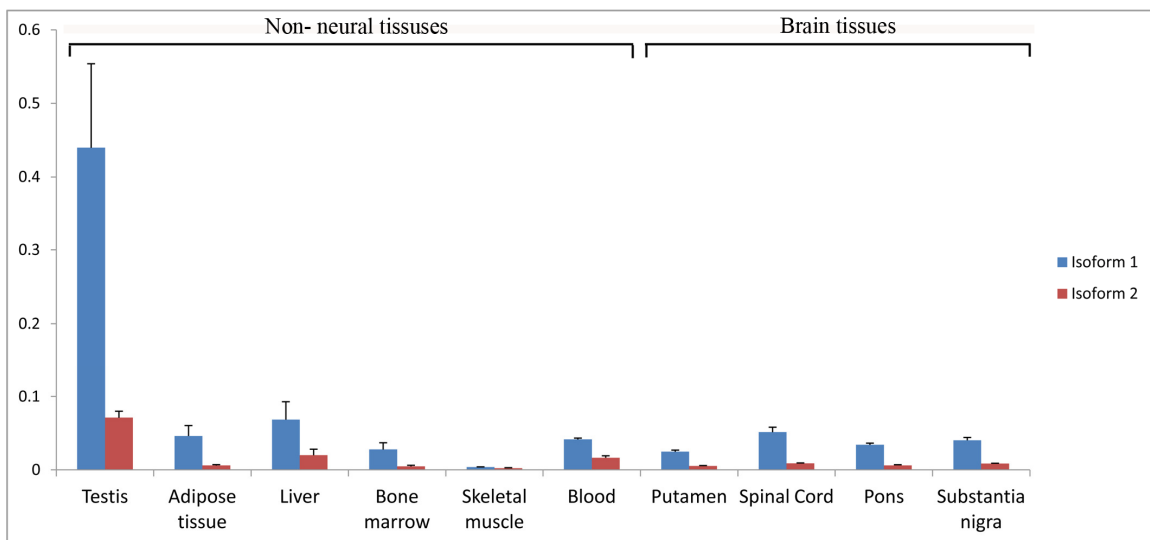


TAF4B c.1831 C>T (p.R611X) in family 1



ZMYND15 c.1520delAACA (p.K507Sfs*3) in family 2

Supplementary Figure S3 Quantification of transcript levels for ZMYND15 isoforms 1 and 2 relative to POLR2A in various tissues.



Supplementary Table S1. Haplotypes with respect to selected SNP and microsatellite markers at the candidate loci. Affected subjects are highlighted in yellow and disease haplotypes are highlighted in gray. In family 2, the father's haplotypes are deduced.

Family 1

Marker	Position (Mb)	Individual			
		401 AS	405 AS	407 AS	409 AS
Locus 4p16.2-4p15.32					
rs4689888	4.56	A B	A B	A B	A B
rs6446663	4.57	A A	A A	A A	A A
rs10005569	4.72	B B	B B	B B	B B
rs2968690	4.90	B B	B B	B B	B B
rs7692195	5.06	A A	A A	A A	A A
rs13128441	5.16	A A	A A	A A	A A
rs7669364	5.28	A A	A A	A A	A A
rs2654508	5.42	B B	B B	B B	B B
rs13138291	5.55	A A	A A	A A	A A
rs2276877	5.85	B B	B B	B B	B B
rs4398594	6.15	A A	A A	A A	A A
rs4689393	6.28	B B	B B	B B	B B
rs9291117	6.447	B B	B B	B B	B B
rs17779675	6.449	A B	B B	B B	B B
rs1574131	8.04	B A	A A	A A	A A
rs2014303	10.58	A B	B B	B B	B B
rs10029959	12.57	B A	A A	A A	A A
rs10939485	13.53	B B	A A	A A	A B
rs4698300	14.86	B A	B B	B B	B A
rs7359	16.16	A B	A A	A A	A B
rs2597749	17.47	A B	A A	A A	A B
rs1568228	18.75	B A	B B	B B	B A
rs1400931	18.75	A B	B A	B A	B B
Locus 7q22.1					
rs4729320	96.55	A B	A B	B B	A B
rs4727353	96.56	A A	A A	A B	A A
rs11973256	97.09	B B	B B	B A	B B
rs10085387	97.38	B B	B B	B A	B B
rs11772060	97.78	A A	A A	A B	A A
rs4729475	98.31	A A	A A	A B	A A
rs4236541	98.83	B B	B B	B A	B B
rs10274646	98.83	A A	A A	A A	A A
rs12705058	98.93	B B	B B	B B	B B
rs952319	99.17	B B	B B	B B	B B
rs10249369	99.24	A A	A A	A A	A A
rs776746	99.27	B B	B B	B B	B B
rs28365067	99.27	A A	A A	B A	A A
rs4646437	99.36	A A	A A	B A	A A
rs472660	99.46	B B	B B	A B	B B
rs6967487	99.60	A B	A B	A A	A B
Locus 5p15.31					
rs186893	8.47	A B	A B	B B	B B
rs180503	8.47	A A	A A	A A	A A
rs284422	8.52	B B	B B	B B	B B

D17S559	4.76	3	2	3	1	3	3	3	3	3	3	2	3
D17S1854	5.66	2	2	2	1	2	2	2	2	2	2	2	2
D17S1832	5.97	3	1	3	2	3	3	3	3	3	3	1	3
D17S1881	6.52	2	1	2	1	2	2	2	2	2	2	1	2
D17S906	6.69	1	2	1	2	1	1	1	1	1	1	2	1
rs2654189	7.17	B	A	B	A	B	B	B	B	B	B	A	B
rs3744405	7.19	A	A	B	A	A	B	A	B	A	B	A	B
D17S960	7.25	1	2	2	1	1	2	1	2	1	2	2	2
D17S720	7.69	3	4	2	1	3	2	3	2	3	2	4	2
Locus 14p11.1-14q11.2													
rs8010057	23.50	A	A	B	B	A	B	A	B	A	B	A	B
rs10143875	23.58	A	B	A	B	A	A	A	A	A	A	B	A
D14S581	24.29	1	2	1	2	1	1	1	1	1	1	2	1
D14S64	24.56	3	2	3	1	3	3	3	3	3	3	2	3
D14S264	25.28	2	1	2	1	2	2	2	2	2	2	1	2
rs17278202	26.33	B	A	B	B	B	B	B	B	B	B	A	B
rs2332572	26.34	A	B	A	A	B	A	A	A	A	A	B	A
rs12894029	27.61	B	A	B	A	A	B	B	B	B	B	A	B
rs10142324	28.64	A	B	A	B	B	A	A	A	A	A	B	A
rs4141682	29.20	B	A	B	A	A	B	B	B	B	B	A	B
rs8018346	29.80	B	A	B	A	A	B	B	B	B	B	A	B
rs1952267	29.88	B	A	A	B	A	A	B	A	B	A	A	A
Locus 7q36.2													
rs4621714	153.35	A	A	B	A	A	B	A	B	A	B	A	B
rs4629774	153.37	A	B	A	A	A	A	A	A	A	A	B	A
rs1558300	153.39	B	A	B	A	B	B	B	B	B	B	A	B
rs10487688	153.60	B	A	B	A	B	B	B	B	B	B	A	B
rs6973850	153.63	A	B	A	B	A	A	A	A	A	A	B	A
rs1860218	153.64	A	A	A	B	A	A	A	A	A	A	A	A
rs10229774	153.65	B	B	A	B	B	A	B	A	B	A	B	A

Abbreviations: F, father; M, mother, AS, affected son; US, unaffected son. Alleles A and B are for SNPs and alleles 1–4 are for microsatellites.

Supplementary Table S2. Candidate variants at the candidate loci. Novel/rare exonic (except synonymous) and splicing variants were selected, and those with alternative depth <60% of the total depth or not validated by inspection of sequence reads on BamView were filtered out. The strongest candidates for causative mutations are highlighted in pink.

A. In patient 402 in family 1.

chr_name	chr_start	chr_end	ref_base	alt_base	hom_het	np_qualit	tot_depth	alt_depth	region	gene	change
Chromosome 4 rs4689888 (4569460 bp) - rs17779675 (6449514 bp)											
chr04	5755565	5755565	G	A	hom	196	31	31	exonic	EVC	nonsynonymous
Chromosome 18 rs2339102 (23797853 bp) - rs1609839 (24027419 bp)											
chr18	23873494	23873494	C	T	hom	194	85	84	exonic;splicing	TAF4B	stopgain

Chromosome 7 rs4236541 (98830439 bp) - rs28365067 (99272310 bp)

No candidate variants

Chromosome 5 rs186893 (8473747 bp) - rs17225452 (8873310 bp)

No variants because there are no genes in the region.

B. In patient 402 in family 2.

chr_name	chr_start	chr_end	ref_base	alt_base	hom_het	np_qualit	tot_depth	alt_depth	region	gene	change
Chromosome 17 pter to rs3744405 (7193255 bp)											
17	2276785	2276785	G	A	hom	105	80	80	exonic	SGSM2	nonsynonymous
17	4647504	4647507	AAAC	-	hom	7893	112	101	exonic	ZMYND15	frameshift del
17	4863328	4863328	G	A	hom	195	58	58	exonic	SPAG7	nonsynonymous

Chromosome 14 rs8010057 (23506110 bp) - rs2332572 (26341293 bp)

No candidate variants

Chromosome 7 rs4621714 (153356432 bp) - rs10229774 (153652188 bp)

No candidate variants

PloyPhen score : Ranges from 0 to 1, and 0 is benign and 1 is damaging.

MutPred: Score is the probability that the mutation is deleterious.

SIFT score ranges from 0 to 1. Amino acid substitution is predicted damaging if the score is ≤ 0.05 and tolerated if the score is ≥ 0.05 .

Mutation Taster: The probability value refers to the prediction, i.e. a value close to 1 indicates a high 'security' of the prediction.

dbSNP132	freq	Sanger Seq	Phenotype	PolyPhen	MutPred	SIFT Score	Mutation Taster
rs141859946	0.001	not done	EVC	BENIGN, 0.123	BENIGN, p= 0.309	TOLERATED, 0.71	BENIGN, p=0.99

(novel) . validated -

dbSNP132	freq	Sanger Seq	Phenotype	PolyPhen	MutPred	SIFT Score	Mutation Taster
rs140107244	0.00	not done	-	BENIGN, 0.121	BENIGN, p = 0.352	TOLERATED, 0.59	DISEASE CAUSING, p=0.99
(novel)	-	validated	-				
rs201256045	NA	validated	-	BENIGN, 0.013	BENIGN, p = 0.252	DAMAGING, 0.01	DISEASE CAUSING, p=0.99