RESEARCH

Allelic variation in TUSC1 gene: rs1462218557 is associated with male infertility and azoospermia

Zahra Vahidi Emami[1], Masoud Sheidai[2*] and Naser Kalhor[3]

Abstract

Background Male infertility is rapidly growing, and single nucleotide polymorphism (SNP) association studies are of critical importance. Tumor suppressor candidate 1 (TUSC1) gene is associated with azoospermia. We investigated association between rs1462218557 in TUSC1 with azoospermia.

Methods We performed tetra-ARMS PCR analysis and sequencing on healthy and infertile individuals.

Results Tetra-ARMS PCR results revealed that the allele frequency of the SNP for T alleles was 0.66, while the same value for C allele was 0.36. It was different from the previous reports of the other countries because they reported the allele C as the frequent allele and it was considered an ancestral allele. The genotype frequencies obtained showed that 67% of the samples were heterozygotes (T/C), while 33% were homozygotes (TT). SNP-STAT analysis showed a significant association ($p = 0.04$) with the studied SNP and azoospermia in Iranian samples.

Conclusions The present study reports a new allele frequency of rs1462218557 in TUSC1 gene, which may be associated with azoospermia in Iranian people. Moreover, no association was observed between neighboring sequences/SNPs with male infertility.

Keywords Allele frequency, SNP-STAT, LDA, LFMM, RDA

Introduction

Population and landscape genetic studies provide detailed information on genetic structure and genetic diversity within target populations and produce data on the effects of ecological and environmental variables on genetic admixture and gene flow versus reproductive isolation among the studied populations. It also may identify specific genotypes or genes in local geographical or ethnic populations, which are associated with both target disease and environmental conditions [1]. Spermatogenic failure is considered the main cause of male infertility for all three general types of azoospermia, oligozoospermia, and asthenozoospermia. However, other factors like gene mutations and chromosomal aberrations, and lifestyle (smoking and alcohol abuse), as well as environmental factors, such as exposure to certain chemicals, are potential causes of spermatogenic failure [2].

Although tumor suppressor candidate 1 (TUSC1) gene was previously introduced as a possible tumor suppressor in lung cancer [3], new studies revealed the association of TUSC1 and DPF3 gene polymorphisms with male infertility [4]. TUSC1, which is an intron-less gene, contains two major transcripts TUSC1-L (2.0 kb long) and TUSC1-S (1.5 kb long), which have been detected in a wide range of human adult tissues and are most strongly expressed in human and mouse testes [3].

*Correspondence:
Masoud Sheidai
msheidai@yahoo.com
[1] Department of Genetic, Qom Branch, Islamic Azad University, Qom, Iran
[2] Department of Plant Sciences and Biotechnology, Faculty of Life Sciences and Biotechnology, Shahid Beheshti University, Tehran, Iran
[3] Department of Mesenchymal Stem Cell, Academic Center for Education, Culture and Research, (ACECR), Qom Branch, Qom, Iran

Vahidi Emami *et al. Egyptian Journal of Medical Human Genetics*   (2023) 24:58

Page 2 of 8

We performed the present study to investigate association between rs1462218557 of TUSC1 gene and azoospermia patients in Iran and provide data on its allele frequency, any association between geographical variables and the clinical aspects of the studied samples, and any association between neighboring SNPs with azoospermia by sequencing a few random samples by PCR. For rs1462218557 SNP analysis, we used tetra-ARMS method.

Material and methods

Samples studied

For tetra-ARMS study, the sample size was determined according to Kirby et al. [5], as performed in package PWR in R.ver.4.1. We used a power of 0.8 and a significance level of 0.05. The blood samples were obtained from ACECR of Qom University (informed consent was obtained from all participants). Fifty out of 100 samples were diagnosed with non-obstructive azoospermia. Sample size can be determined regarding population size, an acceptable margin of error and confidence level needed for the research. For this sample size, margin of error was calculated at 9.68% using Raosoft webtool (http://www.raosoft.com/samplesize.html). All subjects were between 28 and 32 years old and had signed a consent form for genetics studies. Semen analysis was performed due to diagnostic value, and then, azoospermic samples were found normal in karyotyping and did not have any microdeletion on the Y chromosome either. People participating in this study are from 5 different Iranian ethnicities, including Fars, Turk, Arab, Kord, and Lor. The majority of them were Fars and Turk. Blood samples were kept in tubes containing EDTA to prevent clot formation and at − 20 °C. For sequencing, we used 19 people. Different bioinformatics methods were used for association studies like logistic regression, latent factor mixed model (LFMM), redundancy analysis (RDA), and canonical analysis (CCA). For genotype versus phenotype grouping of the studied individuals, we used bot clustering method and heat-map analysis.

DNA extraction and PCR details

The standard salting-out method was used for extraction of genomic DNA from blood samples. The quality of DNA samples was examined by 1% agarose gel electrophorese. In this study, the association between rs1462218557 (TUSC1 gene) and azoospermia in the Iranian men population was examined by genotyping this variant using tetra-ARMS PCR. Two outer primers which produce a 540 bp fragment together and two inner primers that each one of which is specific for one of the alleles (Table 1) were designed for this method by using Oligo7 software [6]. Tetra-ARMS PCR was performed using

Table 1 Sequence of primers used in this study

Primers	Sequence	Product size
Forward	5′ GAGTCGGGTTCCTGTAGAG 3′	19
Reverse	5′ GGCAAATTCCCGGCCTCCTTTG 3′	22
Inner forward	5′ GCGCCACATCGGTCCCAACC 3′	20
Inner reverse	5′ GCTGGCGGCGCGGTAGA 3′	17

Prime thermal cycler and Parstous 2 × Taq PreMix, which contains PCR's necessary materials such as $MgCl_2$ 2 mM, Taq DNA polymerase, and dNTPs. For each sample, 3 μl of DNA was mixed with 15 μl of 2 × Taq premix, 0.5 μl of each primer (all four primers are used in this step), and 9 μl sterile water in microtubes. Reaction's steps were an initial denaturation at 94 °C for 5 min, followed by 33 cycles for replication which includes denaturation at 94 °C for 40 s, primer annealing at 62 °C for 50 s, and extension at 72 °C for 40 s, and after cycles there was a final extension at 72 °C for 5 min. Eventually, the products were analyzed through agarose gel electrophorese with 2% agarose concentration.

Some PCR products (20 samples), using only the outer primers, which produced one 740bp fragment, were sent to CODON genetic laboratory for sequencing test to confirm the accuracy of the Tetra-ARMS PCR for rs1462218557 to investigate the association between adjacent SNPs with non-obstructive azoospermia, and to perform some other phylogenic analysis. The sequencing results of nine samples (ten of control and nine of case groups) were chosen to be analyzed. The ingredients used for PCR and its steps were the same as tetra-ARMS PCR, except for omitting the inner primers.

Data analyses

Allele and genotype frequencies determined by tetra-ARMS PCR were determined by SNP-stat package accordingly. Similarly, the association test between allele frequency and azoospermia was determined by logistic regression method as performed in the same program. Association between phenotypic or clinical features of the studied samples with geographical variables was determined by RDA. DNA sequences of representative samples obtained were aligned and trimmed by MUSCLE program implemented in MEGA 7 software [7]. A maximum likelihood (ML) phylogenetic tree was constructed for the studied samples based on Kimura 2-parameter sequence genetic differences. Haplotype groups and haplotype diversity were determined by TCS networking as performed in Population Analysis with Reticulate Trees (POPART) program (http://popart.otago.ac.nz), ver. 3 [8]. The SNPs and

Table 2 The SNPs and neighboring sequences adjacent to rs1462218557

Population	Group	Sample size	Reference allele	Alternative allele
European	Sub	9690	C = 0.9999	T = 0.0001
African	Sub	2898	C = 1.0000	T = 0.0000
African others	Sub	114	C = 1.000	T = 0.000
African American	Sub	2784	C = 1.0000	T = 0.0000
Asian	Sub	112	C = 1.000	T = 0.000
East Asian	Sub	86	C = 1.00	T = 0.00
Other Asian	Sub	26	C = 1.00	T = 0.00
Latin American 1	Sub	146	C = 1.000	T = 0.000
Latin American 2	Sub	610	C = 1.000	T = 0.000
South Asian	Sub	98	C = 1.00	T = 0.00
Other	Sub	496	C = 1.000	T = 0.000

neighboring sequences adjacent to rs1462218557 were determined by NCBI SNP data (Table 2) base and were used to investigate their association with azoospermia by using both LDA (linear discriminant analysis) and LFMM (latent factor mixed model) approach, which are appropriate for SNP data [9]. These analyses were performed by PAST software and LFMM package in R program ver. 4.2.

Results

Tetra-ARMS results

Based on the tetra-ARMS PCR design in this experiment, the length of DNA fragment amplified by external primers is 747 bp, the fragment amplified for C nucleotide by internal primers is 173 bp, and for T nucleotide is 611 bp (Fig. 1). If the sample studied is heterozygote for this SNP, it will show all these bands.

Based on 100 individuals (200 alleles) studied, the allele frequency of the SNP for T alleles was 0.66, while the same value for C allele was 0.36. These values were 0.62 and 0.38 for azoospermia samples, respectively. Moreover, they were 0.71 and 0.29 in control samples, respectively. These results reveal that allele frequencies obtained in Iranian sample studied differ from what has been reported from the other countries because they reported the allele C as the frequent allele and it was considered an ancestral allele (about 0.999 in Ensemble site data).

The genotype frequencies obtained showed that 67% of the samples were heterozygote (T/C), while 33% were homozygotes (TT). We obtained no CC homozygote in the studied individuals. These values for azoospermia samples were 76% T/C and 24% TT, while they were 58% and 42% in control samples, respectively. SNP-STAT analysis of data showed a significant association ($p = 0.04$) with the studied SNP and azoospermia in Iranian samples.
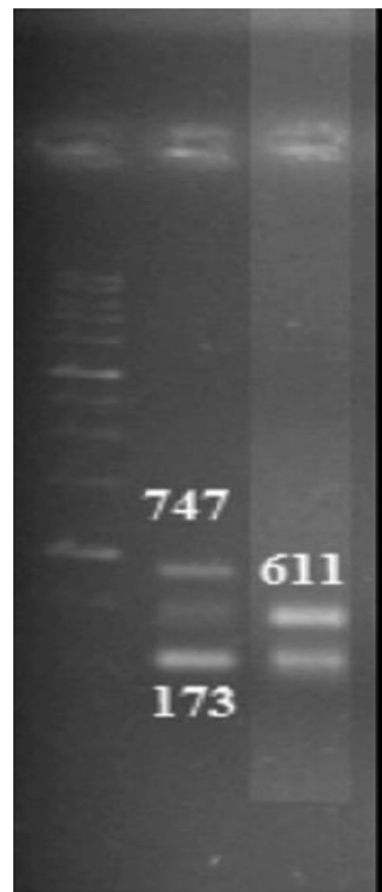
Fig. 1 DNA amplification products in tetra-ARMS PCR

Sequence variability and haplotype groups

In total, we sequenced twenty samples. Accordingly, we obtained 698 bp length DNA after alignment and curation, which showed 11 polymorphic sites among individuals (Fig. 2). The Kimura 2-parameter genetic distance

Fig. 2 Representative sequences of the studied samples showing variation in their nucleotides

varied from 0.001 to 0.004 among the studied samples. The mean nucleotide diversity obtained was 0.004, and Tajima's D statistic was D=− 0.3486. All these results indicated a low degree of nucleotide substitution and that these substitutions are not under selective pressure. The studied samples differed in sequences as they formed different clusters/clades in TCS network (Fig. 3).

Some of the sequences were placed close to each other and formed three main haplotype groups (Group 1–3, in Fig. 3) due to sequence similarity.

Maximum likelihood phylogenetic tree of the studied individuals based on DNA sequences also produced three main clusters supporting the TCS network result (Fig. 4). Moreover, this phylogenetic tree showed that both case
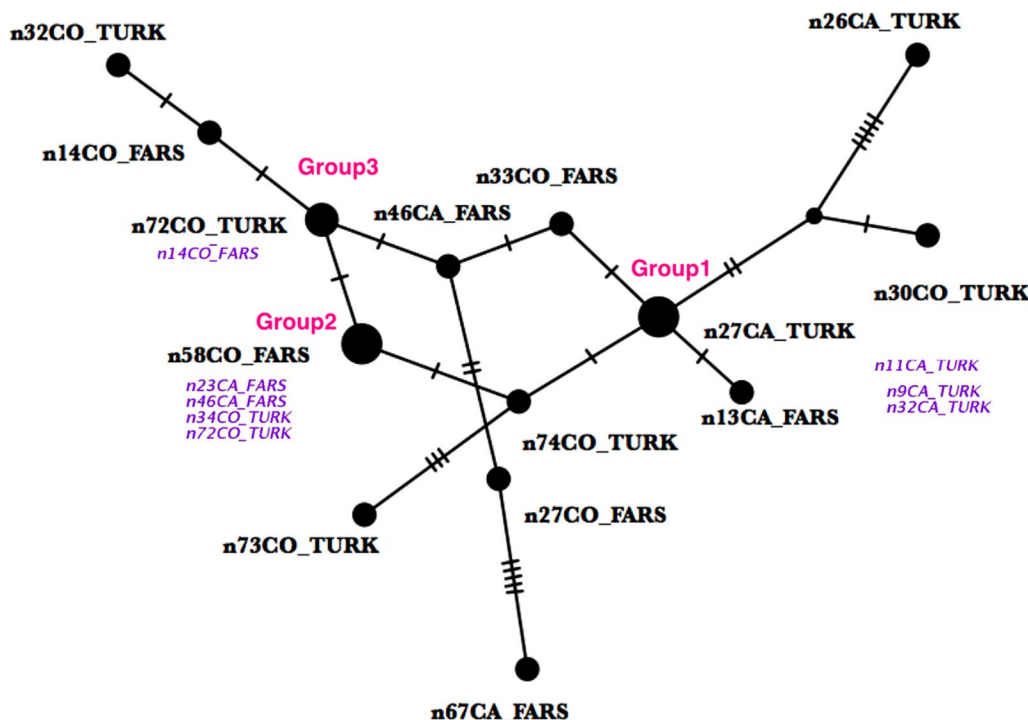

Fig. 3 TCS network showing three main haplotype groups (dash lines show the number of nucleotide substitutions)
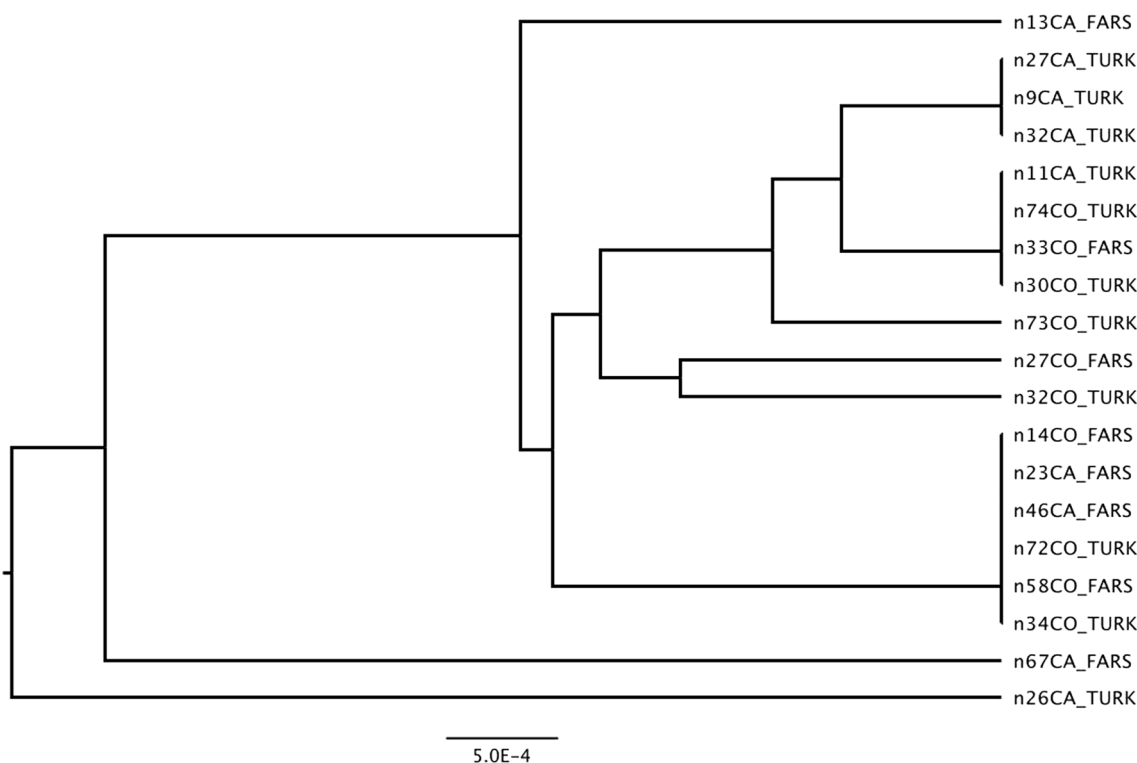
Fig. 4 Maximum likelihood tree of the studied individuals based on DNA sequences showing that the samples from both case (CA samples) and control (CO samples) are intermixed

and control samples were mixed and were placed close to each other. Similarly, samples with different ethnic group backgrounds were also scattered throughout the phylogenetic tree. These results indicated that the sequences obtained from the queried SNP, as well as the neighboring sequences, were not associated with azoospermia. NCBI search for the sequence of present SNP (rs1462218557) showed that the following six known SNPs were in close vicinity of the queried SNP. They were rs555576178 INDEL (in-frame deletion), rs533994707 SNP (missense variant), rs577073776 SNP (missense variant), rs562944441 SNP (5′ UTR variant), rs574590943 SNP (5′ UTR variant), and rs541953116 SNP (5′ UTR variant). Therefore, sequence-based analyses contain rs1462218557 as well as these six SNPs.

LDA and LFMM results
We performed two types of association analyses to study association between different sequences of the studied samples and azoospermia. The LDA plot (Fig. 5) showed that there is no distinction between case and control samples and therefore DNA sequences do not differentiate the azoospermia patients from the controls. The Manhattan plot of LFMM analysis which is a Bayesian approach method also produced low p values for the

studied DNA sequences showing no association between sequences and azoospermia (Fig. 6). Association between clinical features of the studied samples with geographical variables (Longitude and latitude) of their ethnic group also produced no significant association ($p = 0.57$, Fig. 7). This shows that though the clinical feature of the studied samples (particularly sperm characteristics which are absent in azoospermia individuals) differentiates these samples from each other, these features are not related to ethnic groups geographical features.

Discussion
The present study reports nucleotide variability for the SNP rs1462218557 of TUSC1 gene and shows association of these alleles with azoospermia in Iranian samples studied. Population genetic investigations are important for both association studies and personalized medicine in case of human diseases. Our report is different from the other countries which report C allele of rs1462218557 as the ancestral allele with about 99% occurrence. All these results indicated a low degree of nucleotide substitution and that these substitutions are not under selective pressure. TCS network and maximum likelihood phylogenetic tree based on DNA sequences produced three main clusters/groups. Moreover, phylogenetic tree shows that
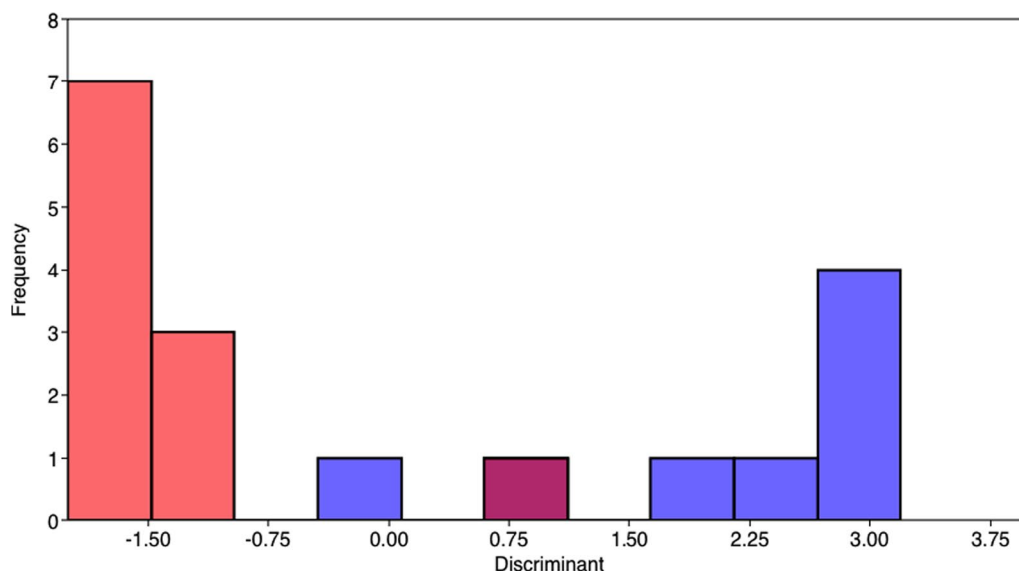
Fig. 5 LDA plot showing no association between DNA sequences and azoospermia individuals. The control individuals (red colored) are not differentiated from azoospermia samples (blue colored samples)
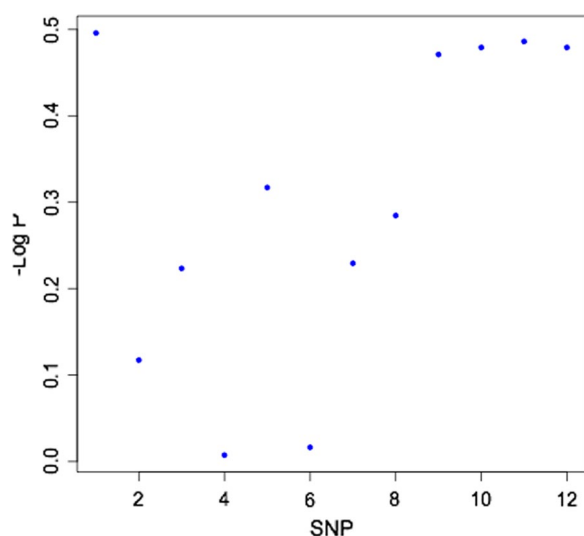

Fig. 6 Manhattan plot of LFMM analysis, showing non-significant *p* value for the studied sequences

both case and control samples are mixed and are placed close to each other. Similarly, samples with different ethnic group backgrounds were also scattered throughout the phylogenetic tree. The findings indicate that the sequences obtained from the queried SNP, as well as the neighboring sequences, are not associated with azoospermia. The following LDA and Bayesian approach method of LFMM showed no significant association between neighboring sequences and the six SNPs with azoospermia. RDA showed no significant association

between clinical feature of the studied samples and longitude and latitude features of ethnic samples.

We show that this SNP has different allele frequencies with a higher value of occurrence for C allele.

We found no report on association of this rs with azoospermia. Therefore, the present study presents association of this SNP with azoospermia for the first time. TUSC1 has been suggested as a candidate tumor suppressor gene in several cancers, and it is reported that variation in TUSC1 and the surrounding region is associated with risk of cancers. However, it was reported that rs12348 in TUSC1 is significantly associated with azoospermia and oligozoospermia [4]. Therefore, the present study adds up to the importance of TUSC1 gene SNPs in male infertility.

The sequence variability obtained here shows almost a very low magnitude of nucleotide substitution ranging from 0.001 to 0.004. This probably shows that this part of genome may be a conserved region. Based on this low level of sequence variability, we obtained three haplotype groups that do not correspond to case and control or ethnic background of the studied samples. Therefore, it seems that the neighboring sequences involved may have no role in azoospermia within Iranian samples studied. This was supported by both LDA and LFMM analyses. Tajima's D statistics showed a negative value and low degree of nucleotide. Substitution is not under positive selection and occurs due to nucleotide polymorphism associated with a selective sweep. The clinical features analysis by RDA also showed no significant association with longitude and latitude variables of the ethnic groups.
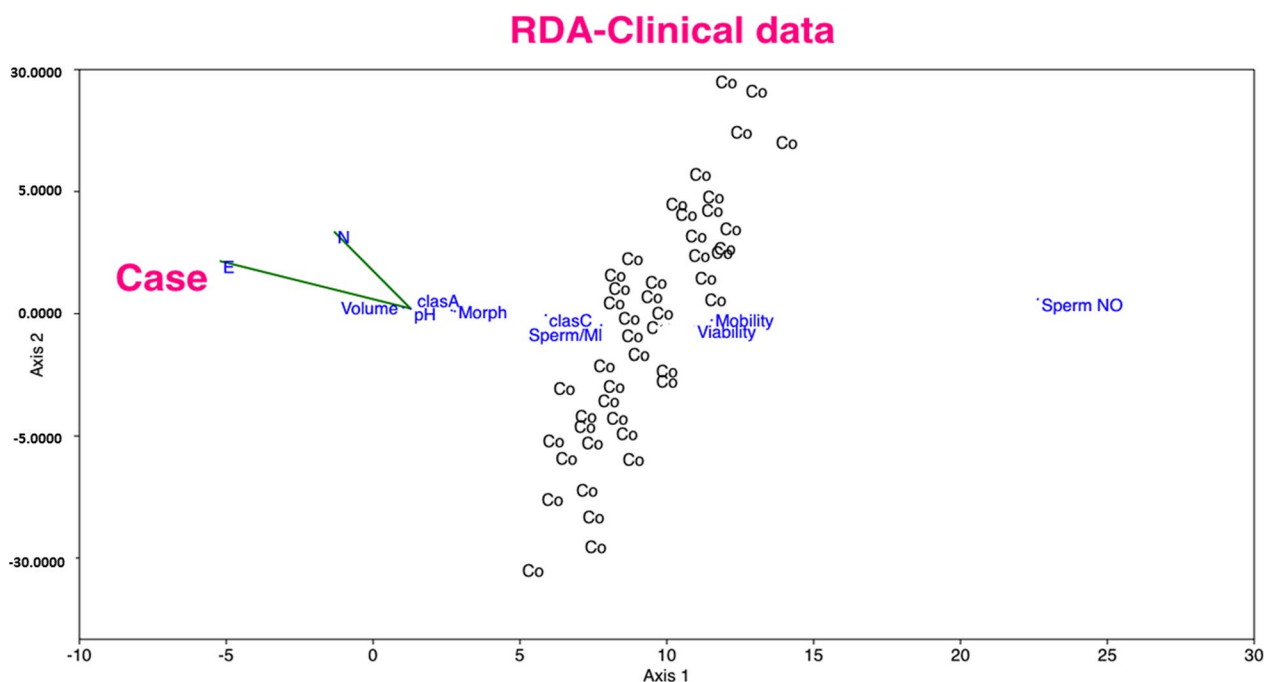
Fig. 7 RDA plot of clinical features and geographical variables of the studied individuals. (*N* = Longitude, *E* = Latitude)

Therefore, no selection exists over the clinical features studied. It has been suggested that the joint effects of selection and linkage are important in shaping patterns of nucleotide variation in humans [10]. This study has some limitations, including high costs of this research and small sample population due to the lack of case samples.

Conclusion

The present study reports a new allele frequency of rs1462218557 in TUSC1 gene, which may be associated with azoospermia in Iranian people. Moreover, no association was observed between neighboring sequences/SNPs with male infertility. Further studies are recommended to confirm the findings of this study.

Declarations

References
1. Freeland JR (2020) Molecular ecology. Wiley, USA
2. Walczak-Jedrzejowska R, Wolski JK, Slowikowska-Hilczer J (2013) The role of oxidative stress and antioxidants in male fertility. Cent Eur J Urol 66(1):60
3. Shan Z, Parker T, Wiest JS (2004) Identifying novel homozygous deletions by microsatellite analysis and characterization of tumor suppressor candidate 1 gene, TUSC1, on chromosome 9p in human lung cancer. Oncogene 23(39):6612–6620
4. Sato Y, Hasegawa C, Tajima A, Nozawa S, Yoshiike M, Koh E et al (2018) Association of TUSC1 and DPF3 gene polymorphisms with male infertility. J Assist Reprod Genet 35(2):257–263
5. Kirby A, Gebski V, Keech AC (2002) Determining the sample size in a clinical trial. Med J Aust 177(5):256–257
6. Poursalehi F, Aghasizadeh M, Ghorbanzadeh S, Kazemi T, Sharifi F, Moodi M et al (2022) Association of the ANGPTL3 gene polymorphisms and haplotypes with cardiovascular diseases in Birjand longitudinal aging study (BLAS). Egypt J Med Hum Genet 23(1):1–9
7. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: molecular evolutionary genetics analysis using maximum

likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 28(10):2731–2739

8. Leigh JW, Bryant D (2015) POPART: full-feature software for haplotype network construction. Methods Ecol Evol 6(9):1110–1116
9. Legendre P, Legendre L (2012) Numerical ecology. Development in environmental modelling. Elsevier, Amsterdam
10. Nachman MW, Bauer VL, Crowell SL, Aquadro CF (1998) DNA variability and recombination rates at X-linked loci in humans. Genetics 150(3):1133–1141

Publisher's Note