Characterization of *HSFY*, a novel AZFb gene on the Y chromosome with a possible role in human spermatogenesis

A.Tessari¹*, E.Salata¹*, A.Ferlin¹, L.Bartoloni¹, M.L.Slongo² and C.Foresta¹,³

¹Department of Medical and Surgical Sciences, Centre for Male Gamete Cryopreservation and ²Department of Experimental Biomedical Sciences, University of Padova, Padova, Italy

³To whom correspondence should be addressed: University of Padova, Department of Medical and Surgical Sciences, Centre for Male Gamete Cryopreservation, Via Ospedale 105, 35128 Padova, Italy. E-mail: carlo.foresta@unipd.it

*These authors contributed equally to this work.

**HSFY** (heat shock transcription factor, Y chromosome) has been mapped in the AZFb region of the Y chromosome, whose deletion results in severe male infertility. **HSFY** belongs to the heat shock factor family that has been shown to be implicated in spermatogenesis both in animals and humans. We report the characterization of the genomic structure, the number of copies on the Y chromosome and the expression of the gene. By comparison of data obtained from expression analysis and from GenBank cDNA sequences, seven exons were identified. Alternative splicing generates three different transcripts and proteins, each containing an HSF domain typical of HSF proteins. Two identical and functional full-length copies of **HSFY** map in palindrome P4 of AZFb, whereas four similar sequences mapping in two clusters in palindrome P1 of AZFc and P3 seem to represent pseudogenes. Sequences similar to few **HSFY** exons are also located in the short arm of chromosomes Y, X and 22. Expression analysis revealed that the three **HSFY** transcripts are differentially expressed, transcript 1 being present in many tissues including testis and ejaculated sperm, and transcripts 2 and 3 being testis-specific. These data suggest that **HSFY** could have an important role in human spermatogenesis.

**Key words:** AZFb/HSFY/male infertility/spermatogenesis/Y chromosome

---

**Introduction**

Microdeletions of the Y chromosome long arm (Yq) are the most common mutations in infertile males, occurring in 10–15% of idiopathic primary testicularFailures (azoospermia and severe oligozoospermia) (Foresta et al., 2001). Microdeletions localize to three non-overlapping spermatogenesis loci, defined as ‘azoospermia factors’ (AZFa, b and c from proximal to distal Yq) (Vogt et al., 1996). Candidate genes for these loci have been identified in recent years (Ma et al., 1993; Reijo et al., 1995; Sargent et al., 1999; Sun et al., 1999; Foresta et al., 2000). However, many genes mapping to these regions may be deleted in infertile men and have testis-specific expression (Lahn and Page, 1997; Kuroda-Kawaguchi et al., 2001; Repping et al., 2002; Ferlin et al., 2003; Skaletsky et al., 2003), and therefore all could be considered hypothetical AZF candidate genes. In particular, deletions within AZFb not removing RBMY1 have been sparsely reported (Najmabadi et al., 1996; Girardi et al., 1997; Krausz et al., 1999; Foresta et al., 1999) and we recently identified four severely infertile subjects with apparently similar breakpoint deletion in AZFb not involving RBMY1 (Ferlin et al., 2003). These data suggest that, at least in these cases, other known or unknown genes could be responsible for the testicular phenotype. Analysis of the complete 2.8 Mb deletion interval in these patients demonstrated that at least five single copy genes and two duplicated genes are removed (Ferlin et al., 2003). Among these novel genes deleted in infertile men, **HSFY** (heat shock transcription factor, Y chromosome) could be a good AZFb candidate for several reasons. **HSFY** transcripts are predicted to be predominantly expressed in the testis (Ferlin et al., 2003; Skaletsky et al., 2003), and the gene belongs to the heat shock factor (HSF) family that has been shown to have a role in spermatogenesis (Neuer et al., 2000).

The HSF family is a group of highly conserved regulators, which play a role as transcriptional activators of heat shock proteins (HSP) genes. HSP are stress-induced proteins whose expression usually results in repair of damaged proteins and survival of the cell, through a chaperone function (Hartl, 1996). Inducible HSP expression is regulated by the heat shock transcription factors, which in response to various inducers acquire DNA-binding activity to the heat shock element (HSE), thereby mediating transcription of the heat shock genes (Wu, 1995; Morano et al., 1999). The HSF family consists of multiple genes in mammals and it is thought to be involved in physiological pathways related to development and differentiation, other than in stress response (Sistonen et al., 1992; Sarge et al., 1994; Rallu et al., 1997; Wang et al., 2003).

**HSFY** is characterized by an HSF-type DNA-binding domain related to the HSF2 gene on chromosome 6 (Ferlin et al., 2003). **HSF2** is expressed at high levels and is only active in two developmental pathways: embryogenesis and spermatogenesis (Pirkkala et al., 2001). Hsf2 mRNA and protein are expressed in a stage-specific manner during spermatogenesis in the rodent, being present in the nuclei of early spermatocytes and post-meiotic round spermatids (Sarge et al., 1994; Alastalo et al., 1998). **HSF2** seems to regulate the expression of many genes and in particular it controls the hsp70 gene family.
promoter (Sistonen et al., 1992), members of which (such as hsp70.2) are germ cell specific (Dix et al., 1996). Mouse mutant for hsf2 and hsp70.2 show spermatogenic impairment, resulting in reduced spermat production and increased germ cell apoptosis (Dix et al., 1996; Kallio et al., 2002; Wang et al., 2003). The importance of this system in humans is underlined by the fact that the human HSPA2 gene (the homologue of the murine hsp70.2) is significantly expressed in testis with normal spermatogenesis, whereas weak expression is detected in testis with abnormal spermatogenesis and no expression is seen in Sertoli cell-only syndrome (SCOS) (Son et al., 1999, 2000; Feng et al., 2001). These observations suggest that this system could play a role during spermatogenesis, by regulating the meiotic division with mechanisms still unknown.

In this study we have better characterized the HSFY gene, in terms of genomic organisation and expression pattern in different tissues with particular focus at the testis.

Materials and methods
Informed consent was obtained by each subject. The Hospital Ethical Committee approved the study.

Sequences and electronic analysis
Using data available at GenBank for the human Y chromosome and sequences obtained by our expression analysis, we performed electronic analysis with the BLASTN program (Achtschel et al., 1990). To determine HSFY gene structure, data obtained from expression analysis and the two mRNA sequences annotated in GenBank for HSFY (transcript 1, 1426 nt, accession no. AF332227; transcript 2, 1058 nt, accession no. AF332226) were aligned with contig NT_011875 (which covers AZFa and AZFb sequences). The existence of homologous sequences on other chromosomes and other regions of the Y chromosome was evaluated by alignment of HSFY exonic sequences on the human genome.

RNA extraction, reverse transcription and PCR amplification
Specific intron-spanning primers were designed to evaluate HSFY expression: HSFY1f (5'-TTTTTGTTCAGACGCTCAACCT-3') from 6796429 to 6796449 nt on NT_011875 contig; HSFY1r (5'-GGTTTCAGGAAGGCTCCCTAT-3') from 6797206 to 6797187 nt on NT_011875 contig; HSFY2r (5'-TTGAAGATCTGGAAAGTTTCATTTT-3') from 6837815 to 6837785 nt on NT_011875 contig. HSFY1f/HSFY2r amplify a 300 bp fragment from transcript 1 and HSFY1f/HSFY2r a 152 bp fragment from transcript 2. A fourth primer, HSFY3r (5'-GGTTTCAGGAAGGCTCCCTAT-3') from 6831950 to 6831930 nt on NT_011875 contig, was designed to amplify, together with HSFY1f, a specific 501 bp fragment from a novel 1496 nt transcript (transcript 3, see Results).

Total RNA was extracted from Sertoli cells (testicular biopsy from SCOS patients) and ejaculated sperm (isolated by Swin up technique from fertile men) by RNA-Be (ISO-TEX Diagnostics, Inc., USA) according to the manufacturer's instructions. The RNA was reverse-transcribed in a reaction performed on 50 ng total RNA/sample using Senscript RT (1 U/reaction; Qiagen, Germany) with 2 mg random hexamer primers (Amersham Pharmacia Biotech Inc., USA) and RNAsin (10 IU/reaction), in a final volume of 20 µl, according to the manufacturer's instructions. The absence of contaminating genomic DNA in cDNA samples was confirmed using reactions in which the RT enzyme was omitted.

cDNA from Human Multiple Tissue cDNA (NTC) Panel (Clontech, USA) and from reverse-transcribed Sertoli cells and sperm RNA was subjected to PCR. Amplification conditions (using PTC 100 MJ Research Inc thermalcycler, Germany) by electrophoresis in 1×TAE buffer (Tris-acetate/EDTA). The products were visualized by ethidium bromide staining followed by exposure to UV light. All PCR products were also directly sequenced on both strands.

The quality of cDNA was evaluated by amplification of the housekeeping gene GAPD (glyceraldehyde-3-phosphate dehydrogenase). Additional controls include use of H2O instead of cDNA and amplification of protamine mRNA (specific for sperm).

Northern blot analysis
Northern hybridization was performed on multiple human tissue mRNA blot (MTN II; Clontech). A 501 bp cDNA fragment specific for HSFY transcript 3 (from base 496 to base 996), obtained by amplification with primers HSFY1f/HSFY3r, was radiolabelled with [α-32P]dCTP (3000 Ci/mmol) using a random priming Kit (Pharmacia, USA) and was purified with Sephadex G50 Nick-Column (Pharmacia). Hybridization and washing conditions were as recommended by the manufacturer. Signal was visualized by autoradiography with Kodak X-Omatic film at −80°C using an intensifying screen for 28 days. In order to evaluate the quality of Northern blot, hybridization was compared with a 432 bp GAPD probe.

Results
Bioinformatic analysis
Two almost identical copies of full-length HSFY spanning 42 kb in the AZFa region on the Yq exist (from 6795934 to 6838206 bp and from 7022958 to 6980683 bp in contig NT_011875). The two HSFY copies lie on plus and minus strands on palindrome P4 as defined by Repping et al. (2002), corresponding to C1/C2 repeats as previously defined by our group (Ferlin et al., 2003) (Figure 2).

Electronic analysis with data obtained from expression analysis (see below) and from GenBank data reveals for HSFY a seven exon structure with alternative splicing generating three different transcripts and proteins (Figure 1). Transcript 1 (1426 bp) results from exon 1 and 2, transcript 2 (1058 bp) results from exon 1 and 7; transcript 3 (1496 bp) results from exon 1 and exon 3–7. The corresponding proteins are 401, 203 and 214 amino acids long respectively. Exon 1 encodes a sequence with an amino acid identity of 80.2% with HSF domain typical of HSF proteins (from 79 to 154 amino acids), and this sequence is shared by the three hypothetical proteins. Only the protein coded by transcript 1 has an HSF-type DNA-binding domain motif, with an amino acid identity of 61.5% (from 79 to 213 amino acids).

BLAST analysis conducted on the whole Human Genome revealed positive match regions on contig NT_011903 (AZFc region on Yq), NT_011896 (short arm of the Y chromosome), NT_011519 (chromosome 22p11-q11), and NT_011757 (chromosome Xp22) (Figures 2 and 3). An HSFY similar region (sequences homologous to exon 1 and 2) of ~3.2 kb with an average percentage nucleotide identity of 82% was also found in NT_011875 distally to the two original copies (between P4 and P3) (Figures 2 and 3). No EST was found for this region, and the predicted translation product of the first exon, based on HSFY frame, results in a 57 amino acid protein with low identity with HSFY and no HSF domain. In AZFc four copies of putative HSFY sequences are detectable, each with an extension of ~53 kb (Figures 2 and 3). These regions conserve the HSFY seven exon structure, with an average percentage identity at the nucleotidelic level of 84% with respect to those in AZFb, and they are organized in two clusters on P1 and P3 palindromes. No EST is annotated in GenBank about these HSFY copies on AZFc, and predicted translation products of the first exon, based on HSFY frame, reveal the presence of a stop codon at the 16th codon. These data let us hypothesize that these four copies are pseudogenes.

On the short arm of the Y chromosome there is a single HSFY homologous region of ~6.3 kb (exon 1–5 similar sequences), with an average percentage identity at the nucleotide level of 80% with HSFY.
copies on AZFb. No EST was found around this region and also in this case the predicted translation product, based on HSFY frame, seems to have a stop on the 133rd codon. Therefore, this region also appears to be a pseudogene. On chromosome 22 p11-q11 there is a 4 kb region in which similar sequences are detectable in exon 1 to 4. A 1381 bp mRNA (supported by four EST) corresponding to this region is annotated in GenBank (accession no AY026053) and it shows high homology (86%) with transcript 1 of the HSFY gene. Finally a homologous sequence of HSFY exon 7 is also present on chromosome Xp, with an average percentage identity of 81%.

The HSFY gene in AZFb partially overlaps with TTTY9 (testis transcript Y9), a 9.3 kb duplicated gene oriented in antisense versus HSFY (Figure 2). Closely located to HSFY there are also sequences homologous to exons of PRY (PTP-BL related, Y gene) (Lahn and Page, 1997; Stouffs et al., 2001), while full-length copies of this gene map to palindrome P3 near the RBMY1A1 cluster. BLAST analysis conducted on the region around HSFY gene showed that the entire region including TTTY9, HSFY and PRY sequences is duplicated in two clusters in AZFc (P1 and P3 palindromes), maintaining the same disposition observed in AZFb (Figure 2). No match was found for
TTTY9 gene on chromosomes Yp, X or 22. The two loci LOC140024 and LOC254261 mapping in the P4 palindrome in the AZFb region are also present in the P1 cluster of the AZFc, where there are also additional copies of these loci but, contrary to HSFY, TTTY9 and PRY sequences, they are not present in the P3 cluster. These data confirmed the homology between sequences in P4 and P1 (Repping et al., 2002), including TTTY9, HSFY, PRY, LOC140024 and LOC254261, but highlighted that a smaller region encompassing TTTY9, HSFY and PRY is also present in P3. Furthermore, the block PRY, LOC254261 and LOC140024 is also found in both arms of P5.

Expression analysis (Figures 4 and 5)
Transcript 1 was detected in testis, brain, pancreas and sperm, but not in other tissues nor in Sertoli cells. Transcript 2 was detected only in testis cDNA library. With HSFY1f/HSFY2r primers used to detect transcript 2, an additional product of ~600 bp was amplified, whose sequence data confirmed that it belongs to HSFY gene. In this manner, we identify a third new HSFY transcript (transcript 3) composed by the two exons also present in transcript 2 and four new exons located in the middle of the transcript, which all respected the AG–GT rule about splicing sites (Figure 1). A specific PCR amplification reaction was performed to confirm the presence of this new transcript by using primers designed on the sequenced transcript and we obtained the expected 501 bp size product. This 501 bp-specific product was detected only in testis cDNA, in accordance with previous data. A Northern blot analysis performed using the 501 bp-specific product as a probe confirmed the presence of an ~1.5-1.6 kb mRNA only on the lane relative to human testis, while no amplification was detectable in other tissues (Figure 5). A testis-specific EST (BU569236), corresponding to nt 30–810 of transcript 3 confirmed our results.

Discussion
Microdeletions of AZF regions remove a large number of genes and transcripts, including testis-specific families and widely expressed genes. Nevertheless, the only phenotype resulting from such deletions is spermatogenic impairment. This finding reflects the remarkable functional specialization of these Yq regions, but makes it very difficult to dissect the roles that individual Y-chromosome gene and gene families play in spermatogenesis. There is increasing evidence that the infertile phenotype is the aggregate effect of loss of function of many genes and gene families and that no single gene could be considered as unique AZF candidate, especially if we consider AZFb (Repping et al., 2002). Recent reports clarified that AZFb boundaries actually extend from palindrome P5 to the proximal arm of palindrome P1 (6.2 Mb) and overlap AZFc by 1.5 Mb (Repping et al., 2002), but smaller deletions with a proximal breakpoint in palindrome P4, as well as deletions not removing the candidate gene RBMY1, may be present (Repping et al., 2002; Ferlin et al., 2003). Many additional genes could therefore be responsible for the testicular phenotype.

In order to better understand the function of individual AZFb genes, we decided to characterize the genomic organization and expression profile of HSFY, a novel gene with a potential role in spermatogenesis previously mapped to repeats C1/C2 (Ferlin et al., 2003) of palindrome P4 (Repping et al., 2002; Skaletsky et al., 2003).
Palindrome P4 contains copies of only two full-length gene families: HSFY and TTTY9. Therefore, deletions with a proximal breakpoint in P5 remove both copies of HSFY and TTTY9, whereas deletions with a proximal breakpoint in P4 remove only one copy of HSFY and TTTY9 genes. Since TTTY9 was found to represent spliced transcript without significant open reading frame (ORF) (Repping et al., 2002), HSFY seems to be the most important gene family of this region, and possibly, as yet undetected, differences in testicular phenotype between patients with deletions involving P5 or not could be related to the loss of function of a different number of copies of this gene.

While preparing this manuscript, Skaletsky et al. (2003) reported the complete sequence and genomic organization of the Y chromosome male-specific region (MSY). In agreement with these authors, we found two functional copies of HSFY in P4. However, we obtained different results on the exon–intron structure of the gene and on the presence of similar copies or sequence homologues to some exons outside the Y chromosome. These differences are probably due to the different approach we used, as it was based on expression analysis and not only on computational data. Expression analysis of HSFY in the testis demonstrated a novel transcript in addition to the two described by Skaletsky et al., and this finding allowed us to determine the exon–intron boundaries on contig NT_011875 and to find the seven exon structure. Data in GenBank and in the Skaletsky et al. paper show only three exons, corresponding to exons 1, 2 and 7 of the present paper. In fact, the novel transcript actually contains exons 3–6. Different splicing seems to produce three transcripts with significant ORF coding for three proteins with different properties, each of which contains an HSF domain. After that, we blasted each single exon with human genomic DNA and found sequence homology on the chromosomes X and Y, and autosomes, and this analysis allowed us to find similar sequences on both the sex chromosomes and on chromosome 22. On the X chromosome, only sequences similar to exon 7 were found. The two homologue sequences in P1 and P3 are probably pseudogenes, like the homologous sequences in chromosomes Yp and those distal to P4. Further analyses are required to assess whether the sequences similar to exons 1–4 on chromosome 22 represent the autosomal homologous gene of HSFY. These data perfectly agree with the evolutionary routes described in the Skaletsky et al. paper, as they proposed that HSFY belongs to the class of ampliconic sequences originated from autosomes or X-degenerated genes, although they could not find homologous sequences outside the Y chromosome.

A preliminary analysis of HSFY genomic blocks and their relatives suggests that HSFY originates from chromosome 22, where four out of seven exons are present. This block was duplicated on chromosome Yp, where exon 5 was added, and then it was duplicated in AZFb (corresponding to the P4 region), where it probably gained exons from
TTTY9 gene and acquired its specific function. Afterwards, this chromosomal region (containing also PRY and TTTY9 genes) was rearranged several times, so now it is present in six copies in Yq. The first duplication and ‘jump’ of HSFY was in the RBMY1A1 region (corresponding to the P3 region), the second in the DAZ region (corresponding to the P1 region). We believe that these duplications were independent events because the genes migrating together were different. Finally, very recently in the evolutionary scale, each of these three HSFY regions on Yq was inverse-duplicated, and now this gene is present in three separate copies. These evolutionary events further suggest that HSFY has acquired functional specialization on the Y chromosome from an autosomal ancestor.

Expression analysis revealed that the three HSFY transcripts are differentially expressed in the different tissues and in the testsis, with transcripts 2 and 3 being testis-specific. Further studies are necessary to clarify the specific cell types expressing HSFY, and it has to be stressed that evaluation of protein expression would better clarify these aspects that are only based on analysis of transcripts. The function of HSFY proteins is still unknown, and it is possible that the three proteins could have different roles. This is also suggested by the absence of DNA-binding domains in HSFY transcripts 2 and 3, suggesting that these proteins may not act as transcription factors.

In conclusion, several features suggest that HSFY could have an important role in human spermatogenesis. Functional copies of this gene map in a region frequently deleted in severely infertile men, it has testis-specific transcripts with expression in germ cells, and it belongs to the heat shock factor family that has been shown to be implicated in spermatogenesis, together with different heat shock proteins, both in animals and humans.

Acknowledgement

The financial support of the University of Padova to A.Ferlin is gratefully acknowledged.

References


Submitted on 1 October 2003; accepted on 3 December 2003