An azoospermic man with a *de novo* point mutation in the Y-chromosomal gene *USP9Y*

Chao Sun[1], Helen Skaletsky[1], Bruce Birren[2], Keri Devon[2], Zhaolan Tang[1], Sherman Silber[3], Robert Oates[4] & David C. Page[1]

In humans, deletion of any one of three Y-chromosomal regions—*AZFa*, *AZFb* or *AZFc*—disrupts spermatogenesis, causing infertility in otherwise healthy men[1–5]. Although candidate genes have been identified in all three regions[3,6–8], no case of spermatogenic failure has been traced to a point mutation in a Y-linked gene, or to a deletion of a single Y-linked gene. We sequenced the *AZFa* region of the Y chromosome and identified two functional genes previously described: *USP9Y* (also known as *DFFRY*) and *DBY* (refs 7,8). Screening of the two genes in 576 infertile and 96 fertile men revealed several sequence variants, most of which appear to be heritable and of little functional consequence. We found one *de novo* mutation in *USP9Y*: a 4-bp deletion in a splice-donor site, causing an exon to be skipped and protein truncation. This mutation was present in a man with nonobstructive azoospermia (that is, no sperm was detected in semen), but absent in his fertile brother, suggesting that the *USP9Y* mutation caused spermatogenic failure. We also identified a single-gene deletion associated with spermatogenic failure, again involving *USP9Y*, by re-analysing a published study.

While screening infertile men for Y-chromosome deletions, we identified one azoospermic individual, WHT2996, with an interstitial deletion on proximal Yq that encompassed sequence-tagged sites (STSs) sY86 and sY87. The deletion was not present in the father of WHT2996, but had arisen *de novo*, suggesting that it was the cause of spermatogenic failure in WHT2996 (Fig. 1). The deletion in WHT2996 was similar in location and extent to previously reported deletions that defined the *AZFa* region[4].

Using a genomic library generated from a normal male, we constructed a BAC contig spanning the deletion and hence the *AZFa* region (Fig. 1). Three genes had been mapped previously to the *AZFa* region or its immediate vicinity: *USP9Y*, *DBY* and *UTY* (refs 7,8). To place these genes more precisely, and to investigate whether other genes are found nearby, we determined the complete nucleotide sequence of the *AZFa* region, which we found to span approximately 0.8 Mb. *USP9Y* and *DBY* are located entirely within the *AZFa* region (Fig. 1). *USP9Y* is composed of 46 exons distributed across 159 kb of genomic DNA. *DBY* is composed of 17 exons spanning 16 kb of genomic DNA. Electronic analysis of the sequence revealed no addi-

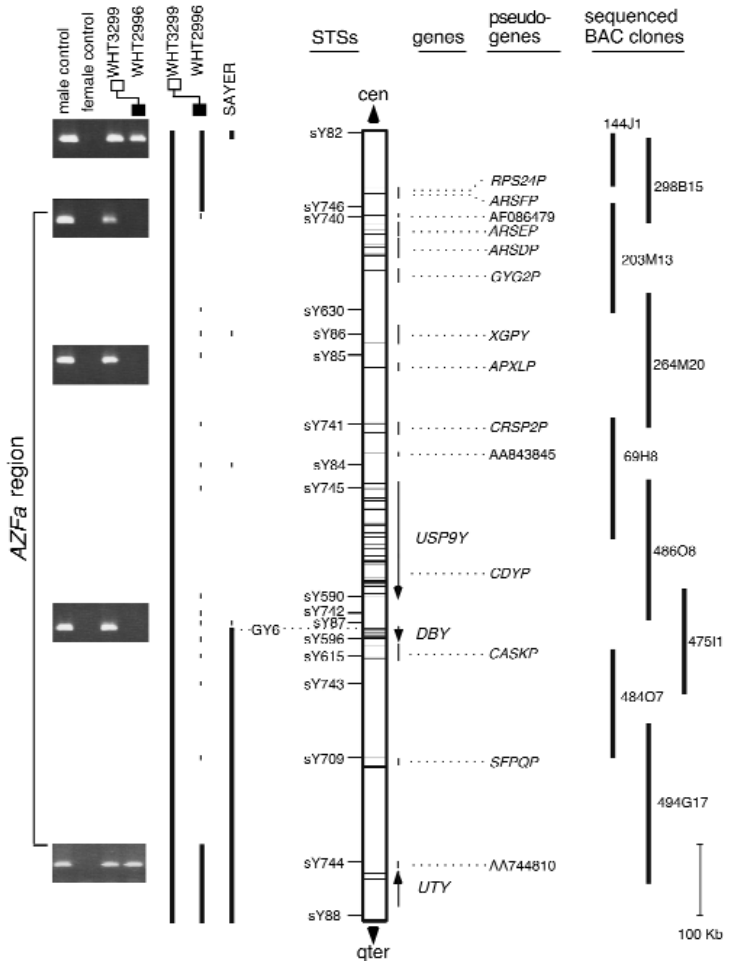

Fig. 1 *AZFa* region of the human Y chromosome. The vertical open bar at the centre of the figure represents this portion of chromosome, oriented with respect to the centromere (top) and the long-arm telomere (bottom). Within the vertical open bar, exons of genes and pseudogenes are shown as horizontal black lines; gene and pseudogene names are indicated (right). (Only the 3′ portion of *UTY* has been sequenced; two exons are shown.) At far right is a tiling path of BAC clones whose nucleotide sequences have been determined. Left, 18 STSs used to isolate the sequenced BAC clones and to characterize the deletion in azoospermic individual WHT2996. Results of testing genomic DNAs from WHT2996 and his father, WHT3299, for presence or absence of these 18 STSs are summarized further to the left; filled black bars encompass STSs found to be present, whereas minus signs represent STSs found to be absent. Gel images of five of these tests are shown at the extreme left (PCR products aligned with chromosomal positions of corresponding STSs), where approximate boundaries of the *AZFa* region are indicated. Published STS results for patient SAYER, who is deleted for part of the *AZFa* region[8], are included for comparison; note the presence of GY6. Many of the genes and pseudogenes shown have homologues on the X chromosome[22]. Three apparent pseudogenes with EST matches but no characterized functional homologues are listed as EST accession numbers.

[1]*Howard Hughes Medical Institute, Whitehead Institute, and Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA.* [2]*Whitehead Institute/MIT Center for Genome Research, Cambridge, Massachusetts, USA.* [3]*Infertility Center of St. Louis, St. Luke's Hospital, St. Louis, Missouri, USA.* [4]*Department of Urology, Boston University School of Medicine, Boston, Massachusetts, USA. Correspondence should be addressed to D.C.P. (e-mail: dcpage@wi.mit.edu).*

Table 1 • Nucleotide sequence variants in *USP9Y*

Location of variant	Nucleotide change	Amino acid change	No. infertile men (of 576 tested) in whom variant was detected	*De novo* or transmitted[a]	No. unrelated fertile control men (of 96 tested) in whom variant was detected
intron 7	GTAA (splice donor) deleted	frameshift at aa 220; truncation at aa 225	1 (WHT2780)	*de novo*	0
exon 21	C→T (nt 3,106)	Pro→Ser (aa 1,036)	1	transmitted	0
exon 22	G→A (nt 3,178)	Ala→Thr (aa 1,060)	3	not tested	1
exon 24	G→A (nt 3,636)	no change	12	transmitted	1
exon 24	C→T (nt 3,670)	deletion of aa 1,188–1,229[b]	1	transmitted	0

Nucleotides are numbered with respect to AUG start codon in published cDNA sequence. [a]*De novo*, detected in an infertile man but not in his father or fertile brother (if father was unavailable for study). Transmitted, detected in both an infertile man and his father or fertile brother. [b]Exon 24 is spliced out, causing in-frame deletion of 42 aa, as revealed by sequencing of RT–PCR products (data not shown).

tional genes in the *AZFa* region, but at least 11 pseudogenes. The 3′ end of *UTY* is located no more than 150 kb distal to the deletion in WHT2996 (Fig. 1), raising the possibility that *UTY* may be subject to position-effect silencing in WHT2996. We tested and rejected this possibility through RT–PCR analysis, which revealed indistinguishable *UTY* transcripts in cell lines prepared from WHT2996 and his father (data not shown). These results suggested that spermatogenic failure in WHT2996 and other *AZFa*-deleted men is probably due to loss of *USP9Y* or *DBY* function, or both.

We then sought more definitive evidence that either gene is required for normal spermatogenesis. We searched for *USP9Y* or *DBY* point mutations in 576 infertile men who met two criteria: (i) they had either nonobstructive azoospermia or severe oligospermia (<5 million sperm/ml semen); and (ii) their Y chromosomes, including the *AZFa* region, were grossly intact, as indicated by the presence of numerous Y-DNA landmarks[9]. As controls, we examined 96 unrelated, fertile men. All *USP9Y* and *DBY* exons and splice sites were screened by single-strand conformation polymorphism (SSCP) analysis[10] or denaturing high-performance liquid chromatography[11] (DHPLC).

We discovered five different variants in the nucleotide sequence of *USP9Y*, but none in *DBY* (Table 1). We next examined whether these variants were present in fertile male relatives; any variant that affected fertility severely should not be found among such relatives but should appear *de novo* in the infertile man. Of the *USP9Y* variants, four appeared unlikely to cause spermatogenic failure for several reasons: they were transmitted from the father or were present in a fertile brother; they were found in a fertile control; or no amino acid was changed. One such variant—found in infertile male WHT3394 and his fertile brother—was a nonsense mutation in exon 24, which appeared to be skipped during splicing of most *USP9Y* transcripts in these brothers (data not shown). Skipping of exon 24, containing 126 of the 7,668 coding nucleotides of *USP9Y*, would not disturb the reading frame downstream.

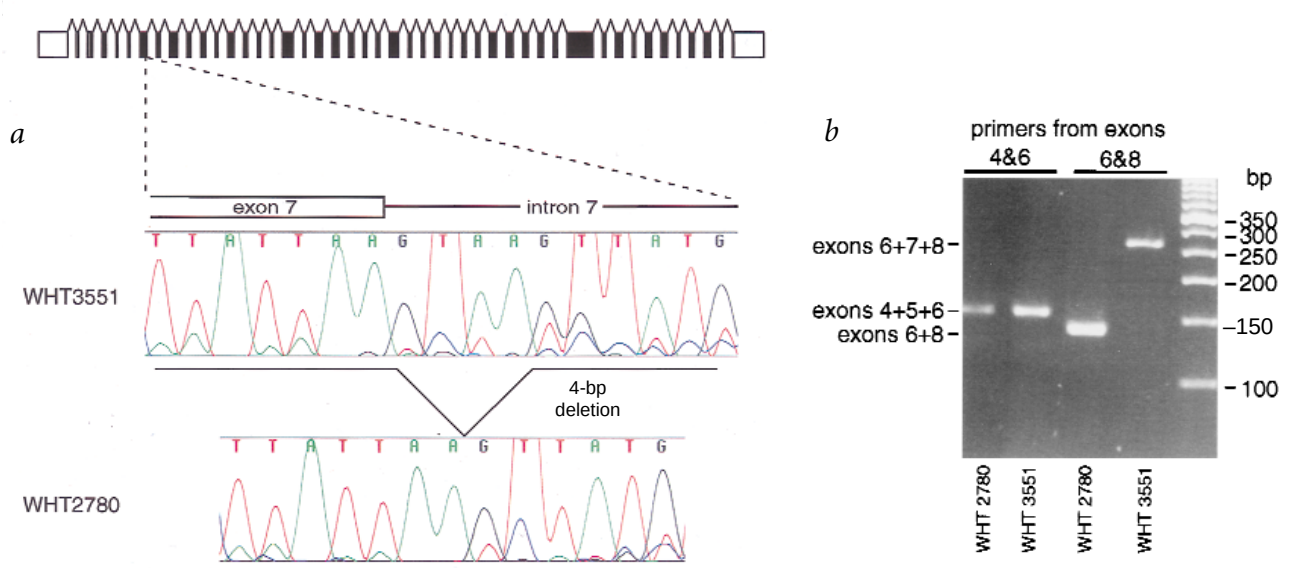

Fig. 2 *De novo* deletion at *USP9Y* splice-donor site in azoospermic man WHT2780. ***a***, Top, intron/exon structure of *USP9Y*; coding exons in black; introns not drawn to scale. Bottom, sequence at exon 7/intron 7 boundary from fertile male WHT3551 and his azoospermic brother WHT2780, demonstrating deletion of first 4 bp of intron 7 (GTAA) in WHT2780. This 4-bp deletion was found in both peripheral leukocytes and a lymphoblastoid cell line from WHT2780. ***b***, Characterization of *USP9Y* transcripts in WHT2780 and WHT3551 by RT–PCR amplification using primers specific to exons 4 and 6 (left) or exons 6 and 8 (right). Sequencing of the exon 6/8 product from WHT2780 confirmed that exon 7 had been spliced out.

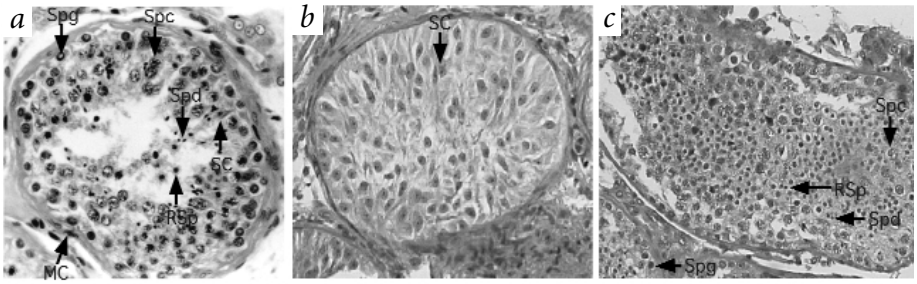

Fig 3 Testicular histologies associated with *AZFa* deletion and *USP9Y* point mutation. ***a***, Photomicrograph of normal seminiferous tubule (in cross section) from a fertile man[3]. Tubule is ringed by myoid cells (MC) and contains Sertoli cells (SC) and germ cells: spermatogonia (Spg), spermatocytes (Spc), round spermatids (RSp), and mature spermatids with condensed nuclei (Spd). ***b***, Tubule from *AZFa*-deleted man WHT2996. No germ cells are visible. ***c***, Tubule from *USP9Y*-mutant man WHT2780. Spermatogenic cells at various developmental stages are present. Staining was done with haematoxylin and eosin.

The fifth *USP9Y* variant—the only *de novo* mutation—was expected to severely truncate the USP9Y protein. In WHT2780, an azoospermic but otherwise healthy male, we discovered a 4-bp deletion in the splice-donor site of *USP9Y* intron 7 (Fig. 2*a*). This deletion was not present in the man's brother, who had fathered two children (the father of WHT2780 was not available for study). We typed both WHT2780 and his brother for 19 Y-DNA polymorphisms[11] and found that the two men share a rare Y haplotype (including a T→A substitution at nt 112 in the sY65 PCR product) not reported previously. These findings suggest that the two men inherited the same Y chromosome apart from what is evidently a *de novo USP9Y* mutation in the azoospermic man. The splice-site deletion in WHT2780 predicts skipping of exon 7, shifting the reading frame and causing USP9Y to be truncated by approximately 90% (Fig. 2*a*). To test this, we prepared RNAs from lymphoblastoid cell lines from WHT2780 and his brother and carried out RT–PCR using primers corresponding to exons 6 and 8. Sizing and sequencing of the RT–PCR product demonstrated that exon 7 is skipped in the azoospermic man, WHT2780, but not in his fertile brother (Fig. 2*b*). We would expect this splice-site/frameshift mutation, falling near the 5′ end of the *USP9Y* coding sequence, to result in loss of *USP9Y* function.

What is the effect of this *USP9Y* mutation on spermatogenesis? A biopsy of the testis of WHT2780 revealed premeiotic and meiotic germ cells in most seminiferous tubules, with small numbers of post-meiotic cells (spermatids) in a few tubules (Fig. 3*c*). These findings suggested a histologic diagnosis of hypospermatogenesis and spermatogenic arrest.

To what degree does loss of *USP9Y* function account for the spermatogenic failure observed in *AZFa*-deleted men? In WHT2996, in whom the entire *AZFa* region is deleted, we observed no testicular germ cells (Sertoli-cell-only syndrome; Fig. 3*b*). Identical findings have been reported for other patients with deletions of the entire *AZFa* region[4,5,12,13]. Thus, deletion of the entire *AZFa* region appears to result consistently in a more severe spermatogenic defect than in our *USP9Y*-mutant patient WHT2780. The simplest explanation is that loss of *DBY*, the only other gene in the *AZFa* region, exacerbates the spermatogenic consequences of loss of *USP9Y*. This model is corroborated by observations of infertile patient SAYER, who differed from other reported *AZFa*-deleted men in that his testis biopsy revealed hypospermatogenesis (like our *USP9Y* mutant WHT2780) rather than Sertoli-cell-only syndrome[5]. SAYER is deleted for *USP9Y*, but retains more distal DNA landmarks, including STS marker GY6 (ref. 8). Our sequencing of the *AZFa* region revealed that GY6, originally identified as an anonymous Y-DNA landmark[14], actually contains the first (and most proximal) exon of *DBY* (Fig. 1). Thus, the distal breakpoint of the deletion in SAYER falls between *USP9Y*, which is deleted, and *DBY*, which is retained. Assuming that the partial *AZFa* deletion in SAYER does not silence *DBY* by eliminating upstream regulatory elements, then SAYER and WHT2780 probably both lack *USP9Y* but not *DBY* function—a likely explanation for their similar testicular histologies.

How does loss of *USP9Y* function impair spermatogenesis—apparently without disturbing other developmental or physiologic processes? These are challenging queries because *USP9Y* has a highly similar homologue, *USP9X*, on the X chromosome, and both genes are expressed throughout the body[7,8,15]. Further work will determine whether *USP9Y* and *USP9X* perform the same function and their aggregate dosage is critical in germ-cell development, or whether *USP9Y* has a role in spermatogenesis not provided by *USP9X*.

Methods

Physical map of *AZFa* region. Six of the STSs (sY82, sY84–sY88) were previously reported[9]. Two more STSs, sY590 and sY596, were previously developed from the sequences of the *USP9Y* and *DBY* genes, respectively[7]. Five additional STSs (sY740–sY744) were generated during ongoing efforts to construct a high-resolution map of the Y chromosome (C. Tilford, D.C.P. *et al.*, unpublished data). We identified BAC clones containing these 13 STSs by PCR screening of DNA pools (Research Genetics) prepared from CITB BAC libraries of human male genomic DNA (ref. 16). To close gaps among the resulting BAC contigs, five additional STSs (sY615, sY630, sY709, sY745, sY746) were generated by PCR-amplification and sequencing of BAC ends[17]. We assembled a contig incorporating the tiling path of BACs (and many other redundant BAC clones; data not shown) based on the STS content of individual BACs.

Nucleotide sequence of *AZFa* region. A path of BAC clones for sequencing was selected based on *Hin*dIII and *Eco*RI fingerprints[18]. We performed large-scale shotgun sequencing and assembly of BACs as described (http://www-seq.wi.mit.edu).

We searched for putative exons and genes by analysing the genomic sequence of the *AZFa* region using GRAIL (ref. 19) and Genscan software[20]. We also searched for homologous sequences in the non-redundant (nr) and dbEST segments of GenBank using BLASTN software[21]. Most pseudogenes could be recognized as such because they contain disrupted versions of long ORFs that are present in functional homologues located on the X chromosome[22] or elsewhere in the genome.

Detection of DNA sequence variants. We searched for *USP9Y* and *DBY* sequence variants in genomic DNAs prepared from lymphoblastoid cell lines or peripheral blood leukocytes. Exons and flanking splice sites were PCR amplified from the genomic DNAs using oligonucleotide primer pairs derived from introns. In some cases, PCR products were screened for sequence variants using a DHPLC (ref. 11) instrument (Transgenomic), in which case genomic DNAs from two unrelated infertile men were mixed before PCR amplification. In other cases, PCR products were screened for sequence variants on SSCP (ref. 10) gels (FMC Bioproduct) electrophoresed for 16 h at RT. In these SSCP experiments, multiplex PCR was employed, with each PCR reaction containing equal amounts of genomic DNA from two or three patients (total of 40 ng genomic DNA/20 μl reaction) and primer pairs for three or four target fragments. All variants detected by DHPLC or SSCP were verified by sequencing of PCR products. In all cases where samples were available, we also tested the fathers or brothers of infertile men in whom we had detected sequence variants. Paternity or fraternity was confirmed by genotyping of polymorphic mark-

ers located on autosomes (*D1S80*, *D10S595*, *D17S5*, *APOB*) and the Y chromosome (M2–M17, M20–M22; ref. 11).

RT–PCR analysis. Trizol reagent (Gibco BRL) was used to prepare RNAs from lymphoblastoid cell lines of WHT2996 and his father, WHT3394 and his brother, and WHT2780 and his brother. For WHT3394 and WHT2780, lymphoblastoid cultures were treated with puromycin (100 μg/ml) for 15 h before harvesting to inhibit nonsense-coupled RNA degradation[23]. In all cases, first-strand cDNA was synthesized using random hexamers as primers. In the case of WHT2996 and his father, *UTY* transcripts were PCR amplified using the primers 5′–GTGCACGAAAAACAAGCAAA–3′ and 5′–TCTTGGAAGGTTGCATAGACA–3′ (derived from the two 3′-most exons). In the case of WHT3394 and his brother, *USP9Y* transcripts were amplified using primers 5′–GGGGTGCTTATTTAAATGCTC–3′ (exon 23) and 5′–CCACCTCCAGCTTATTGCTTCC–3′ (exon 26). In the case of WHT2780 and his brother, *USP9Y* transcripts were amplified using the following two primer pairs: 5′–GATCTTAGTGTAAAAGGCCTTG–3′ (exon 4) and 5′–GGCGATGAGTATTGTTAATAATAC–3′ (exon 6); 5′–GTGTGAATTAATTTCCTCAAATGC–3′ (exon 6) and 5′–CTGGAATGAAGTACTTTTTCAG–3′ (exon 8). Thermocycling conditions were 30 cycles of 1 min at 94 °C, 45 s at 60 °C, 45 s at 72 °C.

GenBank accession numbers. *USP9Y* cDNA, AF000986; BAC 144J1, AC004772; BAC 298B15, AC005942; BAC 203M13, AC002992; BAC 264M20, AC004617; BAC 69H8, AC004810; BAC 486O8, AC002531; BAC 475I1, AC004474; BAC 484O7, AC006565; BAC 494G17, AC005820; sY590, G34983; sY596, G34990; sY740–sY744; G49202, G49203, G49206, G49208, G49209, respectively; sY615, G49204; sY630, G49201; sY709, G49212; sY745, G49211; sY746, G49213.

Note added in proof: Direct evidence that DBY is expressed in patient SAYER has recently been published24.

Acknowledgements

We thank C. Tilford for developing several STSs used here; L. Brown for maintaining patient archives; E. Choi for screening BAC libraries; J. Jaruzelska for participation in variant screening; R. Alagappan, L. Pooler and M. Velez-Stringer for preparing genomic DNAs; M. Moore, D. Altshuler, B. Lahn, E. Lander and S. Rozen for support and advice; and R. Saxena and C. Tilford for comments on the manuscript. C.S. is the recipient of an NIH postdoctoral fellowship. This work was supported in part by NIH.

Received 3 August; accepted 25 October 1999

1. Tiepolo, L. & Zuffardi, O. Localization of factors controlling spermatogenesis in the nonfluorescent portion of the Y chromosome long arm. *Hum. Genet.* **34**, 119–124 (1976).
2. Ma, K. *et al.* Towards the molecular localisation of the AZF locus: mapping of microdeletions in azoospermic men within 14 subintervals of interval 6 of the human Y chromosome. *Hum. Mol. Genet.* **1**, 29–33 (1992).
3. Reijo, R. *et al.* Diverse spermatogenic defects in humans caused by Y chromosome deletions encompassing a novel RNA-binding protein gene. *Nature Genet.* **10**, 383–393 (1995).
4. Vogt, P.H. *et al.* Human Y chromosome azoospermia factors (AZF) mapped to different subregions in Yq11. *Hum. Mol. Genet.* **5**, 933–943 (1996).
5. Qureshi, S.J. *et al.* Polymerase chain reaction screening for Y chromosome microdeletions: a first step towards the diagnosis of genetically-determined spermatogenic failure in men. *Mol. Hum. Reprod.* **2**, 775–779 (1996).
6. Elliott, D.J. *et al.* Expression of RBM in the nuclei of human germ cells is dependent on a critical region of the Y chromosome long arm. *Proc. Natl Acad. Sci. USA* **94**, 3848–3853 (1997).
7. Lahn, B.T. & Page, D.C. Functional coherence of the human Y chromosome. *Science* **278**, 675–680 (1997).
8. Brown, G.M. *et al.* Characterisation of the coding sequence and fine mapping of the human *DFFRY* gene and comparative expression analysis and mapping to the Sxr^b interval of the mouse Y chromosome of the *Dffry* gene. *Hum. Mol. Genet.* **7**, 97–107 (1998).
9. Vollrath, D. *et al.* The human Y chromosome: a 43-interval map based on naturally occurring deletions. *Science* **258**, 52–59 (1992).
10. Orita, M., Suzuki, Y., Sekiya, T. & Hayashi, K. Rapid and sensitive detection of point mutations and DNA polymorphisms using the polymerase chain reaction. *Genomics* **5**, 874–879 (1989).
11. Underhill, P.A. *et al.* Detection of numerous Y chromosome biallelic polymorphisms by denaturing high-performance liquid chromatography. *Genome Res.* **7**, 996–1005 (1997).
12. Pryor, J.L. *et al.* Microdeletions in the Y chromosome of infertile men. *N. Engl. J. Med.* **336**, 534–539 (1997).
13. Grimaldi, P. *et al.* Analysis of Yq microdeletions in infertile males by PCR and DNA hybridization techniques. *Mol. Hum. Reprod.* **4**, 1116–1121 (1998).
14. Jones, M.H. *et al.* A set of ninety-seven overlapping yeast artificial chromosome clones spanning the human Y chromosome euchromatin. *Genomics* **24**, 266–275 (1994).
15. Jones, M.H. *et al.* The *Drosophila* developmental gene fat facets has a human homologue in Xp11.4 which escapes X-inactivation and has related sequences on Yq11.2. *Hum. Mol. Genet.* **5**, 1695–1701 (1996).
16. Shizuya, H.B. *et al.* Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in *Escherichia coli* using an F-factor-based vector. *Proc. Natl Acad. Sci. USA* **89**, 8794–8797 (1992).
17. Nelson, D.L. *et al. Alu*-primed polymerase chain reaction for regional assignment of 110 yeast artificial chromosome clones from the human X chromosome: Identification of clones associated with a disease locus. *Proc. Natl Acad. Sci. USA* **88**, 6157–6161 (1991).
18. Marra, M. *et al.* High throughput fingerprint analysis of large-insert clones. *Genome Res.* **7**, 1072–1084 (1997).
19. Uberbacher, E.C. & Mural, R.J. Locating protein-coding regions in human DNA sequences by a multiple sensor-neural network approach. *Proc. Natl Acad. Sci. USA* **88**, 11261–11265 (1991).
20. Burge, C. & Karlin, S. Prediction of complete gene structures in human genomic DNA. *J. Mol. Biol.* **268**, 78–94 (1997).
21. Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
22. Lahn, B.T. & Page, D.C. Four evolutionary strata on the human X chromosome. *Science* **286**, 964–967 (1999).
23. Andreutti-Zaugg, C., Scott, R.J. & Iggo, R. Inhibition of nonsense-mediated messenger RNA decay in clinical samples facilitates detection of human MSH2 mutations with an *in vivo* fusion protein assay and conventional techniques. *Cancer Res.* **57**, 3288–3293 (1997).
24. Sargent, C.A. *et al.* The critical region of overlap defining the AZFa male infertility interval of proximal Yq contains three transcribed sequences. *J. Med. Genet.* **36**, 670–677 (1999).