The critical region of overlap defining the *AZFa* male infertility interval of proximal Yq contains three transcribed sequences

Carole A Sargent, Catherine A Boucher, Stefan Kirsch, Graeme Brown, Birgit Weiss, Anita Trundley, Paul Burgoyne, Neomie Saut, Christine Durand, Nicolas Levy, Philippe Terriou, Timothy Hargreave, Howard Cooke, Michael Mitchell, Gudrun A Rappold, Nabeel A Affara

Human Molecular Genetics Group, University of Cambridge, Department of Pathology, Tennis Court Road, Cambridge CB2 1QP, UK
C A Sargent
C A Boucher
G Brown
A Trundley
N A Affara

Ruprecht-Karls-Universitat Heidelberg, Institut fur Humangenetik, Im Neuenheimer Feld 328, 69120 Heidelberg, Germany
S Kirsch
B Weiss
G A Rappold

National Institute for Medical Research, The Ridge Way, Mill Hill, London NW17 1AA, UK
P Burgoyne

INSERM Genetique Medicale et Developpment, Faculte de Medecine de la Timone, 27 Boulevard Jean Moulin, 13385 Marseille Cedex 05, France
N Saut
C Durand
N Levy
P Terriou
M Mitchell

Department of Urology and University of Edinburgh Department of Surgery (Urology), Western General Hospital, Crewe Road, Edinburgh EH4 2XU, UK
T Hargreave

MRC Human Genetics Unit, Department of Surgery (Urology), Western General Hospital, Crewe Road, Edinburgh EH4 2XU, UK
H Cook

Correspondence to:
Dr Affara.

Revised version received
11 May 1999
Accepted for publication
13 May 1999

Abstract

The position of deletion breakpoints in a series of four *AZFa* male infertility patients has been refined using new markers derived from BAC clone DNA sequence covering the *AZFa* male infertility interval. The proximal half of the *AZFa* interval is occupied by pseudogene sequences with homology to Xp22. The distal half contains an anonymous expressed sequence tag (named AZFaT1) found transcribed in brain, testis, and skeletal muscle and the *DFFRY* and *DBY* genes. All the patients have AZFaT1 and *DFFRY* deleted in their entirety and three patients additionally have *DBY* deleted. The three patients with AZFaT1, *DFFRY*, and *DBY* deleted show a severe Sertoli cell only syndrome type I phenotype, whereas the patient that has retained *DBY* shows a milder oligozoospermic phenotype. The expression of *DBY* in a cell line from this latter patient is unaltered; this shows that it is the loss of genes lying within the deletion that is responsible for the observed oligozoospermia. RT-PCR analysis of mouse testis RNA from normal and XX*Sxr*^{a} mice (devoid of germ cells) has shown that *Dby* is expressed primarily in somatic cells and that the level of expression is unaltered during germ cell differentiation. This contrasts with *Dffry* where no transcripts are detectable in XX*Sxr*^{a} mouse testis and expression occurs specifically in testis mRNA in a germ cell dependent fashion.
(*J Med Genet* 1999;**36**:670–677)

Keywords: *AZFa*; Y chromosome; infertility

It is estimated that 10% of infertile men have terminal or interstitial deletions including critical segments of the Y chromosome long arm.[1–15] At least three critical regions have been defined by deletion analysis, *AZFa* in proximal Yq and both *AZFb* and *AZFc* in the distal Yq euchromatin.[4] Recent analysis has suggested that a fourth interval (*AZFd*) is present between *AZFb* and *AZFc*.[16] In the majority of cases involving the *AZFa* interval, a Sertoli cell only (SCO) syndrome phenotype is observed[4]; either no germ cells are visible in any seminiferous tubules (SCO I) or germ cells are present in a minority of tubules. This latter variant arises from a failure to complete differentiation and maturation of spermatocytes and spermatids, leading to degeneration of germ cells within most tubules (SCO II). Inability to recover mature sperm in men with *AZFb* deletions[15] suggests that this region is also associated with early blocks in germ cell differentiation.

The *AZFa* interval has been estimated to span 400-600 kb of DNA[17] and includes at least one functional gene, *DFFRY*.[17 18] The *DFFRY* gene has an X chromosome homologue in Xp11.4 that escapes X inactivation and both genes are expressed in a wide range of tissues.[19] Comparative mapping studies have shown that the mouse orthologue of *DFFRY* (*Dffry*) is located in the *Sxr*^{b} deletion interval of the mouse Y chromosome short arm.[17]. This interval is associated with the *Spy* spermatogenic phenotype, which is characterised by a failure of spermatogonial proliferation.[20 21] Very few germ cells beyond the spermatogonial stages can be found and this resembles more closely the SCO II phenotype. More recent comparative mapping studies have shown that two further X-Y homologous genes, *Dby* (Dead box on the Y) and *Uty* (Ubiquitous transcribed tetratricopeptide repeat gene on the Y chromosome), are located in the *Sxr*^{b} interval.[22 23] For both of these genes the human orthologues are located in proximal Yq adjacent to *DFFRY* and are also expressed ubiquitously.[18] The centromere to telomere order of these genes (*DFFRY-DBY-UTY*) is the same on both the mouse and human Y chromosomes, indicating that the *Sxr*^{b} interval and proximal Yq11.2 represent a conserved syntenic segment.[23] This suggests that these genes represent an ancient organisation on the Y chromosome which predates the divergence of the human and mouse lineages and that a shared conserved gene(s) may underlie the spermatogenic phenotypes observed in both species.

The existing *AZFa* PAC contig has been extended using previously published STS markers and by end clone analysis. Sequence of bacterial artificial chromosome (BAC) clones, contributed to GENBANK by the Whitehead Institute/MIT Center for Genome Research, has allowed further markers to be developed for this region. These new STSs allowed confirmation of the PAC contig and refinement of four patients' breakpoints, three of which had been reported previously.[17]

The BAC clone sequences were analysed for predicted gene content. The critical region of

Table 1 New STS markers

STS	*P1*	*P2*	T_{anneal} (°C)	*Size (bp)*
83D22T7	tcagagcagcctgagtagca	tgggactaaaggaacatgcc	57	123
203-44K	gaagaaaggtctgcctgtgc	atcacctgctggactcgttc	59	154
203-93K	tctgaggacacactggttgg	atatgaatgggacgaggagg	57	144
203TEL	tgatgaattctctgggaggg	gctatgttgcaaggcagtca	57	205
264TEL	ctattttggcttttgctgcc	ttgcatcccttccttacacc	57	140
475TEL	gcccagtcctggtttcagac	gagaccctctcatattgttttatattg	57	142
494CEN	tccccaatccctccctatac	ttacaccaccaaggagggag	59	296
494-130K	gaggaaatacggatttgggg	cttcaccatcccttgtctcc	58	107
494-146K	cacacaacccatgctaaacg	atggcttcatcccaactgag	58	294
UTY exon1	atgaaatcctgcgcagtgtc (omj476)	ctcgactgtcaggctaacag (omj477)	57	123

overlap of all four patients contains only three transcribed sequences. In addition to *DFFRY* and *DBY*, an anonymous expressed sequence tag (AZFaT1) has been mapped. Correlation of patient phenotype to the presence or absence of these three sequences suggests that deficiency of *DFFRY* or AZFaT1 or both is associated with an oligozoospermia phenotype in one case. The more severe SCO I phenotype seen in the other three patients may reflect the additional loss of *DBY*.

The expression patterns of mouse *Dby* in mouse testis mRNA and human AZFaT1 in a range of human tissue mRNAs have also been investigated and compared with those of mouse *Dffry* and human *DFFRY* in the same tissues. Human *DBY* expression in deleted and non-deleted cell lines has also been examined.

Materials and methods

SCREENING THE PAC LIBRARY

The RPCI1 library was screened by PCR using primers for established STS markers[24 25] and new STSs derived from BAC sequence. For first round screening the pools were grown in 10 ml of culture medium as recommended by the HGMP Resource Centre and DNA was prepared using the standard protocol supplied. An additional phenol:chloroform extraction and ethanol precipitation was used to remove RNase. Second and subsequent PCR reactions were performed on cell suspensions with RNase added to the PCR reaction mix at a final concentration of 50 ng/μl. PCR reactions were carried out in 10 μl in the presence of 250 nmol/l primers, 200 μmol/l dNTPs, and 1 × buffer (10 mmol/l Tris-HCl, pH 9.0, 1.5 mmol/l $MgCl_2$, 50 mmol/l KCl, 1% Triton X-100, 0.1% gelatin) with "Super Taq" *Taq* polymerase (HT Biotechnology Ltd, Cambridge, UK). Cycling conditions for all PCRs were 95°C for 2.5 minutes followed by 35 cycles of 95°C for 20 seconds, T_{anneal} for 30 seconds, 72°C for 30 seconds, and a final extension at 72°C for 10 minutes. Annealing temperatures for new STS markers are given in table 1.

Additional clones from RPCI 3, 4, and 5 libraries were isolated by hybridisation with existing probes and end clones from the contig arrays. Overlaps were confirmed by hybridisation, PCR, and Alu fingerprint analysis.

ANNOTATION OF BAC DATA

BLASTN searching with known genes and STS markers identified BAC clone sequences from Yq. Segments of the BAC sequences were analysed through the NIX program at the HGMP resource centre. Results were reassessed through standard BLASTN searches to allow final alignment upon the genomic map.

PATIENTS

The three patients SAYER, JOLAR, and ELTOR are described in Brown *et al*[17] and references therein. Briefly, JOLAR and ELTOR presented with azoospermia and infertility. Testicular biopsy of ELTOR showed the absence of germ cells, medium sized seminiferous tubules with minimally thickened basal laminae, and normal numbers of interstitial cells in the stroma. This is typical of SCO I. Analysis of testicular biopsies for JOLAR and AZ539 indicated the typical features of SCO I. Patient SAYER presented with infertility and oligozoospermia with poor sperm motility. Testicular biopsy of patient SAYER showed seminiferous tubules of normal dimensions, with most containing small or moderate numbers of mature spermatozoa. Occasional tubules contained only spermatids, spermatocytes, or spermatogonia. Sertoli cell content was found to be increased. Interstitial cell numbers were normal. The sperm count was 3 $\times 10^6$/ml. FISH analysis of patient line SAYER and W2 normal male lymphoblastoid cell line with PAC 290l19 was performed as described in Zheng *et al.*[26]

PATIENT PCR

Patient DNA was analysed by PCR under the same conditions used for screening the PAC libraries. Each was carried out in a 20 μl reaction contained 100 ng of total human DNA.

RT-PCR

Total RNA was prepared from patient cell line SAYER, normal male lymphoblastoid cell line W2, and normal female lymphoblastoid cell line NF1 using TRI-Reagent (Sigma) following the manufacturer's protocols. Additional poly(A)$^+$ RNA from skeletal muscle, liver, brain, and testis was purchased from Clonetech. One μg of total cell line RNA or tissue specific poly A(+) RNA was reverse transcribed using the reverse transcription system from Promega. For AZFaT1 amplification, patient and control male and female RT products were used undiluted and tissue specific RT products were diluted tenfold, before PCR amplification. Actin amplification was used as a control.

Linear phase was achieved for PCR of control male first strand cDNA with actin, DBY, and UTY primers. PCR reactions were sampled at points between 21 and 33 cycles, with varying dilutions of template. Patient and control samples were standardised by PCR with the actin primers in the linear phase (25 cycles, 1/100 dilution of template) before PCR with *DBY* and *UTY*. All PCR reactions were carried out under standard conditions, as described above. For PCR of *DBY* the primers DBYF (5'-CTAGTATTCATGGAGACCGG-3') and DBYR (5'-TGTAAGGGTGACTTTGCTGC-3') were used with annealing temperature of 55°C. UTY 3'-UTR[21] and actin primers[17] were

also used (DBY, 31 cycles and 1/10th dilution of the template; UTY, 31 cycles and stock RT template).

RT-PCR expression analysis of mouse staged testis tissue used total RNA from mouse testis at 17 and 18 days post coitum, 0.5, 3.5, 7.5, 10.5, and 21 days post partum isolated using TRI-reagent (Sigma). Also, poly(A)$^+$ RNA was prepared from fresh mouse ovary tissue from the MF1 random bred strain and the testes of XX*Sxr*^{a} mice using the Dynabeads mRNA Direct kit (Dynal) following the manufacturer's protocol. These isolated RNA samples were reverse transcribed as described above. For PCR of *Dby*, the primers *DbyF* 5'-TTGGTGGCATTGTGTCCTGC-3' and *DbyR* 5'-AGAGGTGGCTTATGAGTATTT CTTC-3'[21] were used on 1 µl RT product in a standard PCR (see above) with an annealing temperature of 60°C. Primer sequences and reaction conditions for *Dffrx*, *Dffry*, and *Pgk2* are as previously described.[17]

SEQUENCING

All sequencing reactions were carried out using the Thermo Sequenase dye terminator cycle sequencing premix kit from Amersham Pharmacia. Samples were run using a stretch 373 Perkin-Elmer sequencer.

Results

CHARACTERISATION OF *AZFa* PAC CONTIGS AND BAC CLONE DNA SEQUENCES

Previous work from this laboratory had developed a PAC and YAC contig across the *AZFa* region and identified the *DFFRY* gene.[17] Recently, the complete DNA sequence (800 kb) of a series of BAC clones that cover the *AZFa* region has been contributed to GENBANK and this has provided sequence to design new STS markers. The PAC contig has been characterised and extended using new and established STS markers and is shown in fig 1. The internal consistency between marker order based on the depth of coverage in the PAC contig and the predominantly single fold coverage of the BAC sequence contig increases confidence in the results of gene prediction on this body of sequence data.

Little annotation had been performed on the DNA sequence derived from these BAC clones. Through the application of the NIX program (available from the UK Human Gene Mapping Project Resource Centre, HGMPRC), the potential genes and pseudogenes encoded by this 800 kb of DNA have been documented and these are also summarised in fig 1. The first 400 kb of the sequence appears to be almost entirely occupied by non-functional pseudogenes that have related sequences mapping to Xp22, and this region is not likely to be involved in contributing to the *AZFa* phenotype. The remaining 400 kb of sequence contains homology to an anonymous EST (named AZFaT1, see below), the *DFFRY*[17–19] and *DBY*[18] genes. *DFFRY* covers 170 kb of DNA and comparison to the known cDNA sequence[17 18] has shown that it is composed of at least 46 exons. It should be noted that the first 719 bp of the *DFFRY* transcript described by Lahn and Page[18] is absent from the BAC clone sequence, suggesting that this 5' UTR region may be a cloning artefact. *DBY* is a much smaller gene with the coding region covering 17 kb of DNA and is composed of 17 exons. The established marker GMGY6 spans exon 1 of the *DBY* gene. The white box above the *DFFRY* gene indicates the position of sequences related to the *CDY* gene contained within intron 27 of *DFFRY*. Two further Y linked members of this family have been mapped to Yq.[27]

PATIENT BREAKPOINT ANALYSIS

Genomic DNA from three patients was analysed using established and new STS markers. For patient JOLAR, fig 1 summarises historical data because of the unavailablity of an established cell line. The proximal breakpoint in the new patient (AZ539) considerably reduces the critical *AZFa* region to approximately 400 kb as determined from BAC DNA sequencing. AZFaT1 and *DFFRY* are deleted in all four patients, whereas patients AZ539, JOLAR, and ELTOR additionally have *DBY* deleted. None of the three patients tested have the *UTY* gene deleted. It is interesting to note that patient SAYER, lacking AZFaT1 and *DFFRY* but retaining *DBY*, has a milder oligozoospermia phenotype as compared to the Sertoli cell only syndrome found in the other three cases.

AZFaT1 PARTIAL SEQUENCE AND EXPRESSION

BLASTN interrogation of the sequence databases with BAC clone 69h8 sequence showed almost 100% identity to two ESTs (IMAGE clone 647985; 5' EST AA204659 and 3' EST AA207105)[28] present in DBEST. This expressed sequence has been named AZFaT1. Sequence analysis of the IMAGE Consortium (LLNL) cDNA clone (from a Stratagene human hNT neurone cDNA library) and comparison to BAC clone sequence has shown the intron/exon structure of the 3' segment of this potential gene (fig 2A). The sequence from the cDNA clone has a short open reading frame and a large 3' UTR (containing a consensus poly A addition signal) terminated with a poly A tail. The 3' end of AZFaT1 is 12 kb upstream from the 5' end of *DFFRY* and the cDNA sequence extends over 27 kb of genomic DNA. Despite several attempts at 5' RACE and screening for more cDNA clones, it has not been possible to determine further 5' sequences. No significant matches to known genes have been identified with the AZFaT1 sequence. A second IMAGE Consortium (LLNL) cDNA clone (clone ID 649085) starts 9 kb upstream of *DFFRY* showing similarities to several ESTs containing Alu-like sequence. This clone ends 158 bp from the 3' end of AZFaT1. The Alu content of this clone and the similarity in size between the genomic interval and the clone insert suggests that it represents a genomic clone contaminant of the cDNA library. This is not shown in fig 1.

Expression of AZFaT1 was assessed by RT-PCR of mRNA from four tissues using the

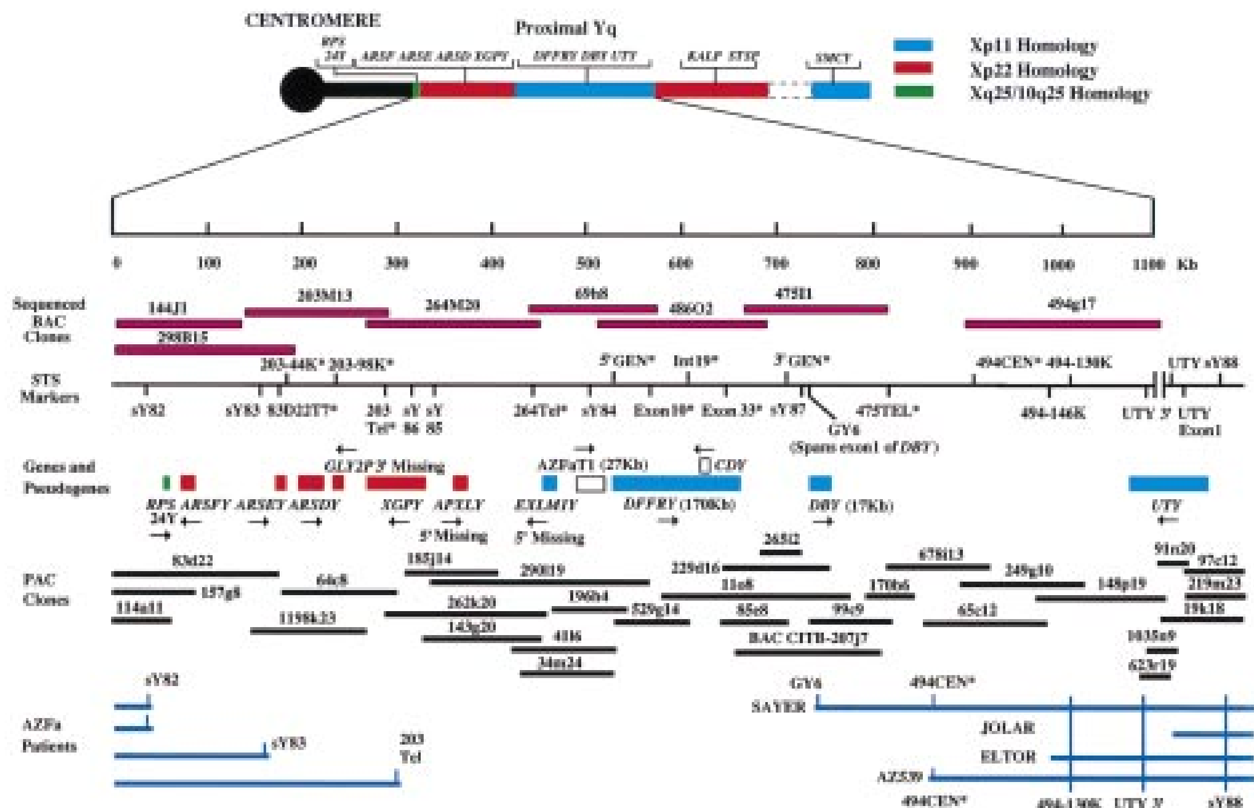

Figure 1 The figure summarises the proximal region of the human Y chromosome long arm in successive layers of complexity and shows the position of deletion breakpoints determined for four AZFa patients. The first layer shows the general organisation of Xp22 and Xp11 homologous sequences located in Yq11.2. The second layer depicts the organisation of the BAC clones sequenced by the Whitehead Institute/MIT Center of Genome Research to give 800 kb of sequence covering most of the AZFa critical region. The third layer illustrates the STS markers used in this study. New markers are marked with an asterisk. The fourth layer marks the position of genes that had either been mapped previously (for example, DFFRY, DBY, UTY) or pseudogenes identified by the use of the NIX program. The white box above the DFFRY gene marks the location of sequences related to the CDY gene family. The fifth layer shows the PAC contigs constructed with the aid of the new STS markers. The sixth layer shows the position of the deletion breakpoints determined for the four AZFa patients. The accession numbers of BAC clones are: 264M20, AC004617; 475I1, AC004474; 203M13, AC002992; 144J1, AC004772; 298B15, AC005942; 69H8, AC004810; 486O2, AC002531; 494G17, AC005820.

primers AZFaT1 forward and reverse designed to produce an inter exon PCR product (primers underlined in fig 2A). AZFaT1 is expressed in testis, brain, and skeletal muscle but not adult liver (fig 2B). No transcripts can be detected in mRNA derived from the SAYER lymphoblastoid cell line and a normal female lymphoblastoid cell, but are readily detectable in mRNA from a normal male lymphoblastoid cell line. This shows that the transcript is male specific and derived from the Y chromosome. Two products are generated from skeletal muscle mRNA, suggesting the existence of differential splicing with the larger transcript containing an additional exon. Purification and sequencing of the two PCR products generated from skeletal muscle confirms their provenance from the AZFaT1 sequence. The additional exon in the larger transcript is boxed in fig 2A and introduces an earlier stop codon into the open reading frame. Thus different C-terminal amino acid sequences could arise from this potential gene.

DBY EXPRESSION IN CELL LINE FROM PATIENT SAYER DELETED FOR DFFRY AND AZFaT1

The milder oligozoospermia phenotype observed in patient SAYER raises the question of whether the *AZFa* phenotype is caused by the combined deficiency of the AZFaT1 transcript and the *DFFRY* and *DBY* genes or by loss of *DBY* alone. If *DBY* is the causative gene, then the oligozoospermia in patient SAYER may arise as a result of a down regulation of *DBY* expression because of the close proximity to the adjacent deletion removing AZFaT1 and *DFFRY*. Alternatively, the full *AZFa* phenotype may require expression of all three genes to be ablated and the observed oligozoospermia reflects the modifying influence of normal *DBY* expression. The third possibility is that *DBY* is not involved in the *AZFa* phenotype and that patient SAYER is mosaic bearing a mixed population of germ cells, with a proportion carrying an intact Y chromosome.

These issues were investigated by exploiting a lymphoblastoid cell line derived from peripheral blood lymphocytes obtained from patient SAYER. In situ hybridisation to metaphase chromosomes of Y PAC clone 290l19, contained wholly within the SAYER deletion, showed no evidence of mosaicism in the cell line upon examination of 50 metaphases (data not shown). All metaphases showed deletion of the PAC clone from the Y chromosome. As a positive control, the same PAC clone was hybridised to normal male metaphase chromosomes on the same slide and 10 from 10 meta-

A

```
ttggagggtcgcaggtcgcggcgctggggatggtgggagcctgcgcagctggctgggcgg
tctctttctctctctcctttttttttttttttttttttttttttttgtcagccgggttcaa
tgcgccatattgaagaggacgggtctaatagatcgctggagacacaatttaactgaaccc
cgcccgttgtggactgactttgatgctctgagtccctccctccttcacgccgctagCAGG
                                                           G
CCCTGATGTAGATTGCCTTTGTCTTACTTGGGACGTTTACCTGAGCGCTTGGTGCTGGTG
  P  D  V  D  C  L  C  L  T  W  D  V  Y  L  S  A  W  C  W  C
TCGGGACCGGGAGATAGGAGTGTCTCAGGAGAGACCTGGCCGAAAACCGCGAGAAAGAAA
  R  D  R  E  I  G  V  S  Q  E  R  P  G  R  K  P  R  E  R  K
                                                AZFaT1forward
AGTGAAGCCTAGTGAAACTGCCTTTGCAGTGACTCAAGAAAAACTCATCACCTGGAGTCC
  V  K  P  S  E  T  A  F  A  V  T  Q  E  K  L  I  T  W  S  P
GTGTAAGCTCGGCGACAGCCCTAGCAGCGAGGCCAAAACAGgtaaagagggctactgaat
  C  K  L  G  D  S  P  S  S  E  A  K  T
g...1903bp....ttcaaactttattcttacagTTTGGGAAGAAAGAAAACCTAAAGTA
                                     V  W  E  E  R  K  P  K  V
TTTGCCGTTGGTGATTCAAGGgtgagttgggagtatttttt...16952bp...ttttttt
 F  A  V  G  D  S  R
ttcttgcttctag[GATGGTAGAGACCAGTCTACAGCCTGAGACCCATCACTA]gtaagaaa
                D  G  R  D  Q  S  T  A  *
ctggtttggaac...4763bp...atgatgtgcttccctttaagGAATCAAACTTGACGT
                                              E  S  N  L  T
ATGGAGCCAAGAAAGCCCTTGGAAAAACTGGCCTCATATTTGTGTACACAGTCCCTGTA
Y  G  A  K  K  A  L  G  K  T  G  L  I  F  C  V  H  S  P  C
CAGGGTTTCTGACCTGTGgtaagtaaagaatgttactt...817bp...ccccattttt
T  G  F  L  T  C
                                      AZFaT1reverse
tgaattcagAGTCACTGAAAACTAAGCTGCGCTTTCCTAAAGTCCTGCGAACTGAAGCCA
             E  S  L  K  T  K  L  R  F  P  K  V  L  R  T  E  A
GACAACTTAAACCTCAGAAGAAAATAACAGCAACCTATTTATATACATAAGCCACTTTCA
R  Q  L  K  P  Q  K  K  I  T  A  T  Y  L  Y  T  *
TACCTGCCTACTGTTGTATAGACTTCAGAGTAATGTGGCCTGTATCGATTTTCCAGGAGT
ATTCTTTTGTGTGTTGTTTTTTCTCAATTCCTCCTATTTTCTCTTTACAGGATGTGAGAC
TTCACAACCTGCTAAAAATGAGCTTTCAGGACCTACCCATATAGGAATAAACCATCCTAG
CCATGAGAGATCAGATGAAACCTGAGACCAGAGAGACTCATTTGTTTCAAAATAGTTTCT
CCAAAAGATTTTATAAAAGAAAAGGCTGGGGGGAGTGGGATATGAAAGGAAAATGAATCT
TGGGGCCCCCAAATCACTAAGCTCAAGGGATAAGTCAAGTTAGAAACTGTTCAGGGCCAA
CTTACCTTGCATTCTATTCAAATTCACCCCTCTGCTCACTTAGATGCATATCTGATTGTA
ATCAGAAACTCAAAAGAATGCAGCAGTTTGTCTCTCACCTATCTATGACCTGGAAGCCCC
CTTCCCCGTTTGAGTCTTCCTGCCTTTGCTTCACTTTATCCCTGCCTTTCTAGACTGAAC
CAACATACTTCTTAGATATATTGATTGATGTCTCATGTCTCCCTAAAATGTATAAAACTA

AGCTGGGCCCCAACCACCTTGGGCACATGTCGTTAGGACTTCCTGAGACTGTGTCACAAG
TTTGTGTCCACAACTTTGACAAAATAAACTTTTTAAATTAACTGAGACCTGTCCTAAATT
TTTAGGGTTCCATCAGATATGCATTTCTCTCACCTGAGCATCAGAGGGATGATTTCGAGT
TCTCTGTGTTTTTTGTCCACAGGAATTTCCTTGTGGGCAAATTCTGAGGGAGGGATGTAG
CTTTTTATGTTTGGAGCTATTTTATTTAGAAATAAAATGGGAGGCAGGTTTGCCTGAGT
CAGTTGCCTGCTTGACTTCCTTTGGCTTAGTCATTTTGGAGTCCTGAGATTTATTTTCTT
TTCACATACCTGTTTTATTTATGATGGTTTATGGTCATTGAACCTTGTAACATTGATTAC
ATTTCTTTTCACTGGTCTTTTTTCTTAAGTAGATAAACTTATTTTTAACTTGGGCACAAG
AGAGCATAAAAATTATATGAGGTACATAGCCTGCTTAAATAGAATATTTAAAATACGTTT
TGTCCACAGAGCTGGAATTCTTTGTAAATGAAAACCAAATTTTACTTAATTTTTAGAGAT
AGGATCTTGGCTTTGTCGCCTAGGCTAGAGTGCTGTGGGGGCAATCATAGCCCACTGCAA
CCTCGAATTCCTGGCTTAAGTGATCCTCCCATCTCAGCCTCCTGGGTGCTTGGACAACAG
ACATGACACCTTACCTGGCTAGTTAAAAAAAATTTTTTTTGTAGAGATGGGCTTTGCCTG
TGTGGCCCAGGCTGCTCTTGAACTCCTAGGTTCAAGTGATCCTCCAGCTTGGGCCTTCCG
AAGTGCTGGGATTACATTGCTTCCAGCTTTAGTTTTTTAAATGTTTAACTTGTTTAACTT
ATTTAACTTGCTGTTTATTATTTAATGGAGAGCTTTTAGTGTTTGTCTTTTGAGGTGGGA
TCTTGCCATGCCCAGACTAGCCTTGAAGTCCTACAGTCAAGTGATCTTTGCATCTCAGTC
TCCCAAGAAGTAGTGGTAGAGAATTTTTATGTTACATTCCTTTCTCTGTGTATATGTTTG
TGGGAGGACATATGTCCCTTTGTATGTTTGGGCATATATATTTTTTTGTGTGAAAGTTATG
CATGTTATTGTTGATCAATACAAGAGGTTTAGAGGCCAGAGAAGAGAAATAAAATGGGAA
AAACTACAAAC(A)n
```

B

L S M F Sm Lv Br Ts G cD Bl

Figure 2 (A) Sequence of AZT1 with exonic sequence in upper case and intronic sequence in lower case bold (the number of bases of intronic sequence not included are also shown). The alternatively spliced muscle specific exon is shown boxed. Within the 5' most exon the highlighted G is the first base of image clone 647985. The exonic and intronic sequence upstream of this base was predicted by comparison with consensus branch point and acceptor site sequences. The predicted polyadenylation signal is in bold, the poly A tail in bold, and the primers AZFaT1 forward and reverse are underlined, italicised, and bold. (B) The expression of AZFaT1 in human lymphoblastoid cell lines, adult liver, brain, skeletal muscle, and testis mRNA was examined by RT-PCR using the primers AZFaT1 forward and reverse. The 220 bp product was visualised on a 0.8% agarose gel. No product is generated from genomic DNA because of the large introns between the forward and reverse primers. L=1 kb ladder, S=SAYER cell line deleted for AZFaT1, M=normal male lymphoblastoid cell line, F=normal female cell line, Sm=skeletal muscle, Lv=adult liver, Br=brain, Ts= testis, G= male genomic DNA, cD=IMAGE cDNA clone, Bl=water only control.

phases were found to be positive. The absence of mosaicism was also confirmed by PCR analysis of cell line DNA. The expression of *DBY* and *UTY* (both expressed ubiquitously) was examined by quantitative RT-PCR in both the deleted cell line and a lymphoblastoid cell line derived from a normal male using PCR primers specific for the Y copies of these genes. For both *UTY* and *DBY*, no difference in the level of expression in comparison to the actin control is observed (fig 3) as a consequence of the adjacent deletion. The expression of *UTY* in a lymphoblastoid cell line from patient AZ539 is also not affected by the adjacent deletion (data not shown).

EXPRESSION OF *DBY* IN MOUSE TESTIS RNA AT DIFFERENT STAGES OF SPERMATOGENESIS

In previous studies it had been shown that *Dffry* transcription followed a germ cell dependent pattern of expression and was first detectable at 7.5 days after birth.[17] Fig 4 shows the pattern of *Dby* expression over the same time period. *Dby* is expressed at all stages from

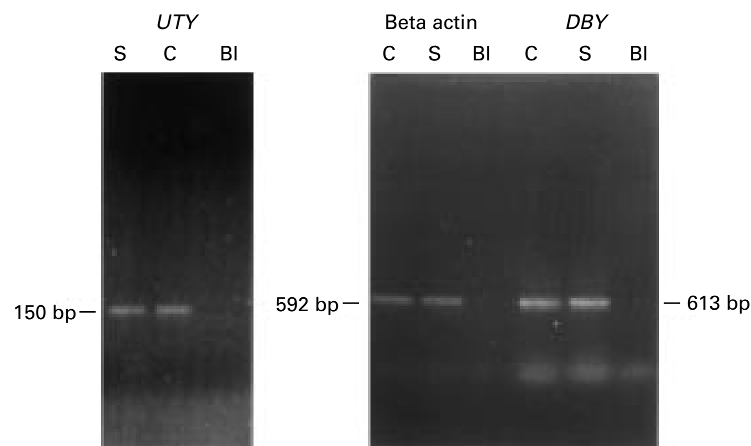

Figure 3 Analysis of the expression of DBY, UTY, and actin by RT-PCR in mRNA from normal male and patient SAYER lymphoblastoid cell lines. The DBY primers generate a 613 bp product, the UTY primers a 150 bp product, and the actin primers a 592 bp product. PCR products were analysed on a 2% agarose gel at different times in the amplification cycle to ensure accurate quantitation of gene expression. S=SAYER cell line, C=normal male cell line, Bl= water only control.

17 days post coitum to 21 days post partum and in the testis RNA from XX*Sxr*^{a} mice (lacking germ cells), but is not detectable in mouse ovary RNA. This is similar to the expression pattern of *Dffrx* (shown) but not *Dffry*[17] and indicates that *Dby* is expressed in somatic cells of the testis. The germ cell specific *Pgk2* gene is detectable at 21 days, consistent with its expression at the round spermatid stage of differentiation. Analysis of RNA from the testes of four independent XX*Sxr*^{a} mice with both *Dby* and *Dffry* confirms that expression levels of *Dby* are not altered by the presence of germ cells, in stark contrast to *Dffry* (fig 4B).

Discussion

The mapping of deletion breakpoints in this study has succeeded in reducing the critical region of proximal Yq involved with the *AZFa* phenotype. Much of the proximal half of the interval has been excluded by patient AZ539 and the refined position of the genes mapped to the region in relation to the breakpoints in patients has focused attention on *DFFRY* and *DBY*. From the careful analysis of DNA sequence centromeric to the 5' end of *DFFRY* using the NIX program, there is no evidence for the existence of potentially functional genes apart from the transcribed sequence AZFaT1. The sequence centromeric to this point appears to be a monolithic segment of degenerated genes with related sequences mainly in Xp22. As exon prediction programs cannot predict all exons it is possible that the 150 kb, spanning the proximal breakpoint of patient AZ539 to AZFaT1, may contain one or more functional genes that remain as yet undetected.

Patient SAYER with the milder oligozoospermia phenotype has proved to be pivotal in helping to understand the potential contribution of genes mapping to the *AZFa* interval. In patient SAYER, the promoter region of the *DBY* gene is functional and expression levels of both *DBY* and *UTY* are unaffected by the adjacent deletion including AZFaT1 and *DFFRY*. This is shown by the equivalence of expression level between the deleted and non-deleted subjects. Thus, down regulation of *DBY* can be excluded as an explanation of the milder phenotype in patient SAYER. This indicates that a gene or genes (*DFFRY* or AZFaT1) located in the interval between the proximal breakpoint in AZ539 and the distal breakpoint in SAYER could be contributing to the *AZFa* phenotype. This result could be explained by a number of possibilities. A less severe phenotype is caused by deletion of AZFaT1 or *DFFRY* or both and that the additional deficiency of an early acting gene (possibly *DBY*) or genes beyond the distal SAYER breakpoint is required for the full *AZFa* SCO I syndrome phenotype. The gap between *UTY* and *DBY* may contain further contributory genes. Alternatively, gonadal mosaicism for an intact Y chromosome (not found in the SAYER lymphoblastoid cell line, but cannot be excluded from testicular tissue) may have moderated the severity of spermatogenic impairment. One also has to consider the possibility that when dealing with a small sample size, phenotypic variation may reflect differences in genetic background, environment, and age. Under these circumstances, this would imply that the critical gene(s) lie in the interval defined by the proximal AZ539 breakpoint and the distal SAYER breakpoint. The *UTY* gene can be excluded from involvement in the phenotype as it is present in all four patients. The analysis of further patients with deletions of the AZFa region and the new markers available will help to refine the critical interval.

The analysis of AZFaT1 is incomplete as it has not been possible to determine whether this potential gene possesses a functional 5' region. The characterised portion of AZFaT1 is organised as an interrupted gene and contains a short open reading frame that can participate in differential splicing. This indicates that its location on the Y has not occurred by a retroposition event, and the failure to detect similarities to either other genes or EST sequences suggests that it may be unique to the Y. Studies are in progress to determine whether this transcript possesses a functional 5' end, homologous sequences on the human X chromosome, and if there are conserved sequences on the mouse Y chromosome. If this gene is non-functional, this places even greater emphasis on *DFFRY* as a candidate gene.

Expression analysis of *Dby* and *Dffry* in RNA from mouse testis singles out *Dffry* as a good candidate for a role in spermatogenesis. The mouse *Spy* phenotype located in the *Sxr*^{b} interval is characterised by an early block in spermatogonial proliferation and results in an almost complete absence of germ cells.[20 21] This resembles the human *AZFa* phenotype. In contrast to *Dffry*, *Dby* expression is not specific to the testis or to germ cells[22] (and as evidenced by the readily detectable expression in XX*Sxr*^{a} mouse testis RNA) and appears to be unaltered in a significant way by the changing population of germ cells as spermatogenesis proceeds.

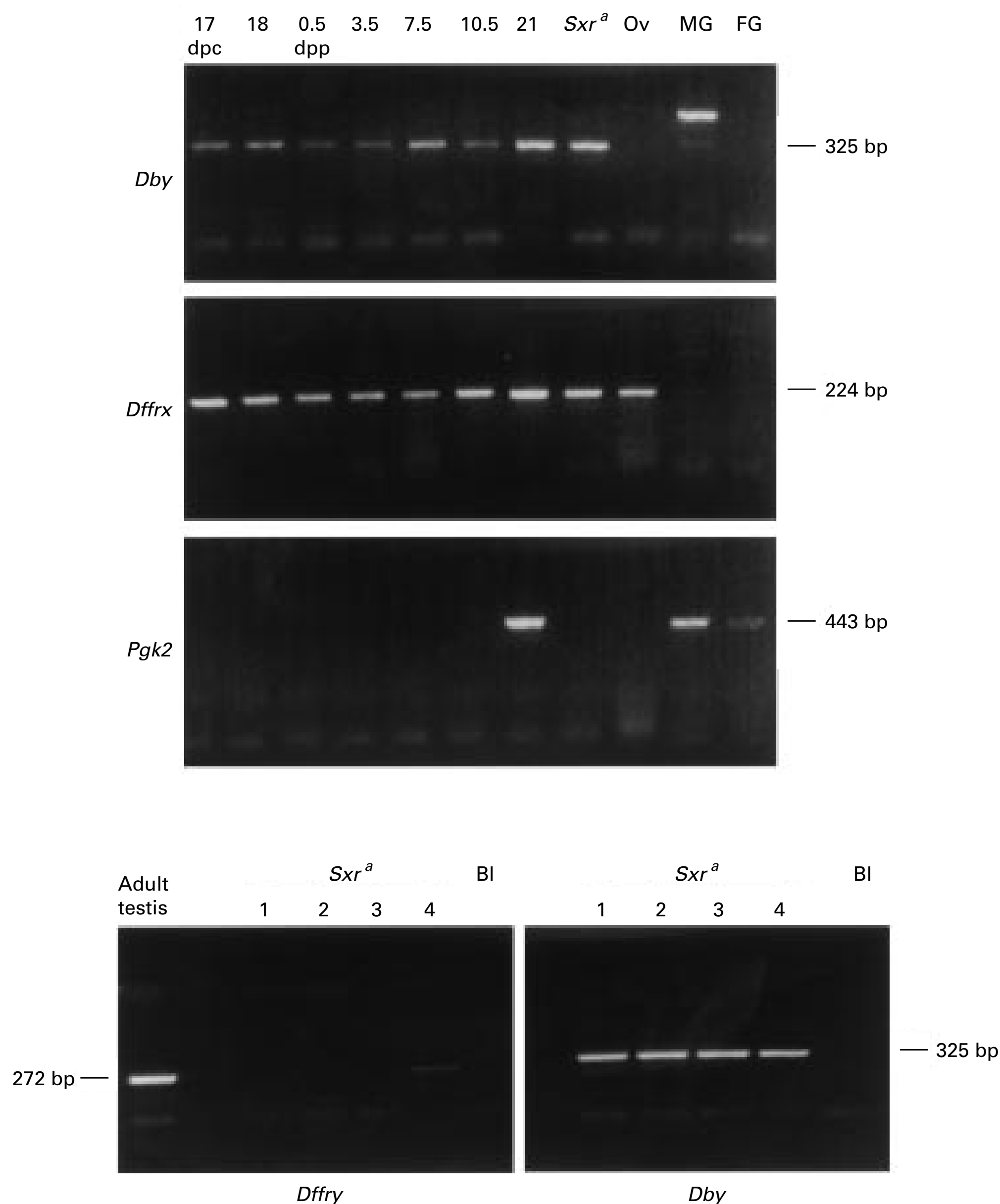

Figure 4 (Above) RT-PCR analysis of Dby expression (using the Y specific PCR primers that generate a 335 bp product) in mouse testis mRNA at 17 and 18 days post coitum and various days after birth. As controls, the same mRNAs were analysed with Dffrx and Pgk2 primers used in a previous study. The failure to obtain a PCR product from female genomic DNA and ovary mRNA shows that these primers are Y specific. The PCR products were analysed on 1% agarose gels. Sxr^a=mRNA from the testis of $XXSxr^a$ mice, Ov=ovary mRNA, MG=male mouse genomic DNA, FG=female mouse genomic DNA. (Below) RT-PCR analysis of the testis mRNA extracted from four different $XXSxr^a$ mice with primers from the Dby and Dffry genes. Except for mouse 4 where there may be a very low level of germ cells, no Dffry expression is detectable by RT-PCR.

However, *Dby* may be necessary at early stages in spermatogenesis and small changes in expression in germ cells or supporting lineages could be important. Thus the combined contribution of *Dby* and *Dffry* may be a component of the *Spy* phenotype. The expression of *Dby*/*DBY* in Sertoli cells may be necessary for germ cell differentiation to proceed correctly and this could explain the difference in phenotype seen in SAYER in relation to the other patients in this study. Five other genes have been mapped to the Sxr^b interval,[23 28–32] but none of these is contained within deletions associated with the *AZFa* phenotype; indeed the genes *Ubely*[29 30] and *Eif2γy* do not detect any related sequences on the human Y chromosome.[31 33] Although these genes do not contribute to the infertile phenotype caused by deletions of the *AZFa* region on the human Y chromosome, this does not exclude them from being involved in the *Spy* phenotype in mouse. On the basis of its germ cell specific pattern of expression, it has been suggested that *Ubely* is a good candidate for a role in the *Spy* phenotype.[30 34] However, experiments with a *Ube1y* transgene alone have failed to rescue the Spy phenotype in Sxr^b mice.[35] It is possible that several of the genes in the Sxr^b region may

cooperate to bring about the production of mature germ cells and that combinations of transgenes (including *Ube1y* and *Dffry*) will be necessary to rescue the Spy phenotype. It is rather intriguing to note that *Ube1y and Dffry* are concerned with the addition (*Ube1y* - ubiquitin activating enzyme) and removal (*Dffry* - ubiquitin hydrolase) of ubiqutin from proteins. Ubiquitination marks a protein for degradation which, in order to occur, requires removal of the ubiquitin just before entry into the proteasome. The protein encoded by *DFFRY*/*Dffry* belongs to the group of ubiquitin hydrolases that enable entry of polypeptides into the proteasome for degradation. Is it possible that the overlap in the expression of *Ube1y* and *Dffry* in germ cells permits fine control over the turnover of protein(s) that are necessary for successful spermatogenesis?

We thank HGMP Resource Centre, Hinxton, for access to the RPCI1 PAC library and computing facilities. This work was supported by grants from the Wellcome Trust and the Medical Research Council.

1 Tiepolo L, Zuffardi O. Localisation of factors controlling spermatogenesis in the non-fluorescent portion of the human Y chromosome long arm. *Hum Genet* 1976;**34**:119-24.
2 Ma K, Sharkey A, Kirsch S, *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* 1982;**1**:29-33.
3 Vogt PH, Chandley AC, Hargreave TB, Keil R, Ma K, Sharkey A. Microdeletions in interval 6 of the Y chromosome of males with idiopathic sterility point to disruption of AZF, a human spermatogenesis gene. *Hum Genet* 1992;**89**:491-6.
4 Vogt PH, Edelmann A, Kirsch S, *et al*. Human Y chromosome azoospermia factors mapped to different subregions of in Yq11. *Hum Mol Genet* 1996;**5**:933-43.
5 Reijo R, Lee TY, Salo P, *et al*. Diverse spermatogenic defects in humans caused by Y chromosome deletions encompassing a novel RNA-binding protein gene. *Nat Genet* 1995;**10**: 383-93.
6 Reijo R, Alagappan RK, Patrizio P, Page DC. Severe oligospermia resulting from deletions of azoospermia factor gene on Y chromosome. *Lancet* 1996;**347**:1290-3.
7 Qureshi SJ, Ross AR, Ma K, *et al*. Polymerase chain reaction screening for Y chromosome microdeletions: the first step towards the diagnosis of genetically-determined spermatogenic failure in men. *Mol Hum Reprod* 1996;**2**:775-9.
8 Stuppia L, Mastroprimiano G, Calabrese G, Peila R, Tenaglia R, Palka G. Microdeletions in interval 6 of the Y chromosome detected by STS-PCR in 6 of 33 patients with idiopathic oligo- or azoospermia. *Cytogenet Cell Genet* 1996;**72**:155-8.
9 Najmabadi H, Huang V, Yen P, *et al*. Substantial prevalence of microdeletions of the Y-chromosome in infertile men with idiopathic azoospermia and oligozoospermia detected using a sequence-tagged site-based mapping strategy. *J Clin Endocrinol Metab* 1996;**81**:1347-52.
10 Pryor JL, Kent-First M, Muallen A, *et al*. Microdeletions in the Y chromosome of infertile men. *N Engl J Med* 1997;**336**:534-9.
11 Foresta C, Ferlin A, Garolla A, Rossato M, Barbaux S, De Bortoli A. Y-chromosome deletions in idiopathic severe testiculopathies. *J Clin Endocrinol Metab* 1997;**82**:1075-80.
12 Simoni M, Gromoll J, Dworniczak B, *et al*. Screening for deletions of the Y chromosome involving the DAZ (Deleted in AZoospermia) gene in azoospermia and severe oligozoospermia. *Fertil Steril* 1997;**67**:542-7.
13 Girardi SK, Mielnik A, Schlegel PN. Submicroscopic deletions in the Y chromosome of infertile men. *Hum Reprod* 1997;**12**:1635-41.
14 Stuppia L, Gatta V, Calabrese G, *et al*. A quarter of men with idiopathic oligo-azoospermia display chromosomal abnormalities and microdeletions of different types in interval 6 of Yq11. *Hum Genet* 1998;**102**:566-70.
15 Brandell RA, Mielnik A, Liotta D, *et al*. AZFb deletions predict the absence of spermatozoa with testicular sperm extraction: preliminary report of a genetic test. *Hum Reprod* 1998;**13**:2812-15.
16 Kent-First M, Muallem A, Shultz J, *et al*. Defining regions of the Y chromosome responsible for male infertility and identification of a fourth AZF region(AZFd) by Y chromosome microdeletion detection. *Mol Reprod Dev* 1999;**53**:27-41.
17 Brown GM, Furlong RA, Sargent CA, *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* 1998;**7**:97-108.
18 Lahn BT, Page DC. Functional coherence of the human Y chromosome. *Science* 1997;**278**:675-80.
19 Jones MH, Furlong RA, Burkin HI, *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* 1996;**5**:1693-702.
20 Burgoyne PS, Levy ER, McLaren A. Spermatogenic failure in male mice lacking the H-Y antigen. *Nature* 1986;**320**: 170-2.
21 Sutcliffe MJ, Burgoyne PS. Analysis of the testes of H-Y negative XO*Sxr*^{b} mice suggests that the spermatogenesis gene (*Spy*) acts during the differentiation of the A spermatogonia. *Development* 1989;**107**:373-80.
22 Greenfield A, Scott D, Pennisi D, *et al*. An H-Y epitope is encoded by a novel mouse Y chromosome gene. *Nat Genet* 1996;**14**:474-8.
23 Mazeyrat S, Saut N, Sargent CA, *et al*. The mouse Y chromosome interval necessary for spermatogonial proliferation is gene dense with syntenic homology to the human AZFa region. *Hum Mol Genet* 1998;**7**:1713-24.
24 Vollrath D, Foote S, Hilton A, *et al*. The human Y chromosome: a 43-interval map based on naturally occurring deletions. *Science* 1992;**258**:52-9.
25 Jones MH, Khwaja OSA, Briggs H, *et al*. A set of ninety-seven overlapping yeast artificial chromosome clones spanning the human Y chromosome euchromatin. *Genomics* 1994;**24**:266-75.
26 Zheng YL, Ferguson-Smith MA, Warner JP, Ferguson-Smith ME, Sargent CA, Carter NP. Analysis of chromosome 21 copy number in uncultured amniocytes by fluorescence in situ hybridisation using a cosmid contig. *Prenat Diagn* 1992;**12**:931-43.
27 Lahn BT, Page DC. Retroposition of autosomal mRNA yielded testis-specific gene family on human Y chromosome. *Nat Genet* 1999;**21**:429-33.
28 Lennon G, Auffray C, Polymeropoulos M, Soares NB. The I.M.A.G.E. Consortium: an integrated molecular analysis of genomes and their expression. *Genomics* 1996;**33**:151-2.
29 Mitchell MJ, Woods DR, Tucker PK, Opp JS, Bishop CE. Homology of a candidate spermatogenic gene from the mouse Y chromosome to the ubiquitin-activating enzyme E1. *Nature* 1991;**354**:483-6.
30 Kay GF, Ashworth A, Penny GD, *et al*. A candidate spermatogenesis gene on the mouse Y chromosome is homologous to ubiquitin-activating enzyme E1. *Nature* 1991;**354**:486-9.
31 Ehrmann IE, Ellis PS, Mazeyrat S, *et al*. Characterisation of genes encoding translation initiation factor eIF-2gamma in mouse and human: sex chromosome localisation, escape from X-inactivation and evolution. *Hum Mol Genet* 1998;**7**: 1725-37.
32 Agulnik AI, Mitchell MJ, Lerner JL, Woods DR, Bishop CE. A mouse Y chromosome gene encoded by a region essential for spermatogenesis and expression of male-specific minor histocompatibility antigens. *Hum Mol Genet* 1994;**3**: 873-8.
33 Mitchell MJ, Wilcox SA, Watson JM, *et al*. The origin and loss of the ubiquitin activating enzyme gene on the mammalian Y chromosome. *Hum Mol Genet* 1998;**7**:29-34.
34 Odorisio T, Mahadevaiah SK, McCarrey JR, Burgoyne, PS. Transcriptional analysis of the candidate spermatogenesis gene Ube1y and of the closely related Ube1x shows that they are coexpressed in spermatogonia and spermatids but are repressed in pachytene spermatocytes. *Dev Biol* 1996;**180**:336-43.
35 Burgoyne PS. The role of Y encoded genes in mammalian spermatogenesis. *Semin Cell Dev Biol* 1998;**9**:423-32.