Diverse spermatogenic defects in humans caused by Y chromosome deletions encompassing a novel RNA-binding protein gene

Renee Reijo[1], Tien-Yi Lee[1], Pia Salo[3], Raaji Alagappan[1], Laura G. Brown[1], Michael Rosenberg[1,4], Steve Rozen[2], Tom Jaffe[1], Donald Straus[4], Outi Hovatta[5], Albert de la Chapelle[3], Sherman Silber[6] & David C. Page[1]

We have detected deletions of portions of the Y chromosome long arm in 12 of 89 men with azoospermia (no sperm in semen). No Y deletions were detected in their male relatives or in 90 other fertile males. The 12 deletions overlap, defining a region likely to contain one or more genes required for spermatogenesis (the *Azoospermia Factor*, *AZF*). Deletion of the *AZF* region is associated with highly variable testicular defects, ranging from complete absence of germ cells to spermatogenic arrest with occasional production of condensed spermatids. We find no evidence of *YRRM* genes, recently proposed as *AZF* candidates, in the *AZF* region. The region contains a single-copy gene, *DAZ* (*D*eleted in *AZ*oospermia), which is transcribed in the adult testis and appears to encode an RNA binding protein. The possibility that *DAZ* is *AZF* should now be explored.

[1]*Howard Hughes Medical Institute and* [2]*Center for Genome Research, Whitehead Institute and Department of Biology, Massachusetts Institute of Technology, 9 Cambridge Center, Cambridge, Massachusetts 02142, USA*
[3]*Department of Medical Genetics, University of Helsinki, Helsinki, Finland*
[4]*Department of Biology, Brandeis University, Waltham, Massachusetts 02254, USA*
[5]*Family Federation of Finland, Kalevankatu 16, Helsinki, Finland*
[6]*In Vitro Fertilization Program, St. Luke's Hospital, St. Louis, Missouri 63017, USA*

Human sperm are produced via a complex developmental process. The progression from spermatogonial stem cells to mature spermatozoa requires 65 days and involves an elaborate succession of distinct cell types[1,2]. The process is punctuated by at least three mitotic and the two meiotic divisions. Meanwhile, the genome is repackaged — with protamines rather than histones — and re-imprinted. Spermatogenesis begins at puberty and continues throughout adult life; a human male may produce 10^{12} to 10^{13} gametes during his lifetime.

Two percent of human males are infertile because of severe defects in sperm production[3,4]. Most of these men are otherwise healthy, and the cause of spermatogenic anomalies is usually not identified with certainty. Such isolated defects in fertility have often been ascribed to infection, immunologic factors, anatomic malformation, or chemical insult. Relatively little research has focused on possible genetic aetiologies. While spermatogenesis must require many gene products, no human mutations specifically disrupting spermatogenesis have been defined at the molecular level. Identification of genes specifically involved in sperm production — and analysis of the mutant phenotypes — could provide both insight into this developmental process and a more rational basis for therapy of male infertility.

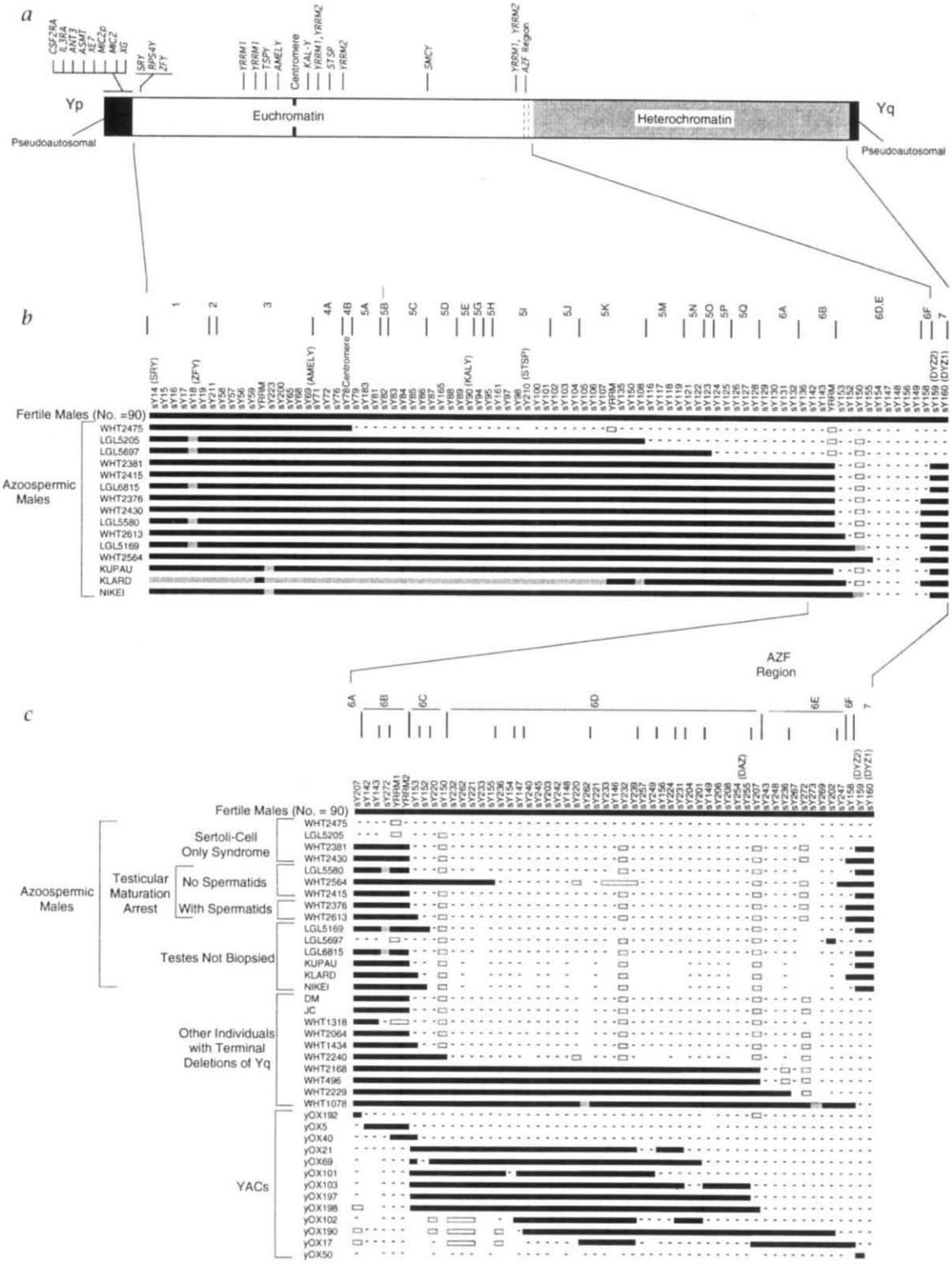
A role for the human Y chromosome in spermatogenesis — quite apart from determining the sex of the gonad's somatic components — was first suggested by the studies of Tiepolo and Zuffardi[5]. Having karyotyped 1170 subfertile men, they reported six azoospermic individuals with microscopically detectable deletions of distal Yq. In four of these cases, the fathers were tested, and all were found to carry intact Y chromosomes. On the basis of these *de novo* deletions in azoospermic men, Tiepolo and Zuffardi proposed the existence of a spermatogenesis gene, or *Azoospermia Factor* (*AZF*), on Yq.

The hypothesis of one or more Y-borne structural genes required for spermatogenesis was generally favoured in subsequent reports of terminal deletions and other microscopically-detectable Y anomalies in azoospermic men[6,7]. It remained possible, however, that azoospermia in the case of such gross Y abnormalities resulted not from structural gene loss but from perturbations of sex chromosome pairing or segregation during meiosis. The two theories need not be mutually exclusive.

Somewhat stronger evidence for the existence of spermatogenesis gene(s) on the human Y chromosome was provided by the detection, using DNA probes, of interstitial, usually submicroscopic Yq deletions in azoospermic men[8–13]. These interstitial deletions do not involve regions of the Y known to pair and recombine with the X chromosome and, therefore, seemed less likely to perturb sex chromosome behavior in meiosis. Indeed, the finding of overlapping interstitial Yq deletions in three azoospermic males led Ma and colleagues to initiate a gene search, culminating in their identification of *YRRM1* and *YRRM2*, closely related genes whose absence, they proposed, might be the cause of azoospermia[11].

Several aspects of the *AZF* hypothesis merited further scrutiny, motivating our present study. First, it has proven difficult to define the regions of the Y chromosome whose deletion results in azoospermia. As we report here, the *YRRM1* and *YRRM2* genes are present in most if not all azoospermic men, including those with interstitial Yq deletions — findings that are completely at odds with the mapping studies of Ma *et al.*[11]. In the absence of a consistent

and reproducible map localization for *AZF*, the very existence of such a gene remains in doubt. The difficulties stem in part from inadequate attention to the possibility that putative 'mutations' are actually polymorphisms of no functional consequence. In principle, *de novo* mutations and polymorphisms can be distinguished by examining the Y chromosomes of immediate male relatives, but in reality this critical control has been performed for few azoospermic men in whom putative deletions have been reported. This problem is compounded by the fact that many of the DNA probes employed in recent studies detect families of Y-specific repetitive sequences whose organization and number vary dramatically among normal human males. Interpretation of these Y-specific-repetitive markers, usually scored by Southern blotting, is inherently troublesome, requiring careful controls for polymorphism. Different laboratories employed different, sometimes quite limited sets of Y-DNA markers, further complicating



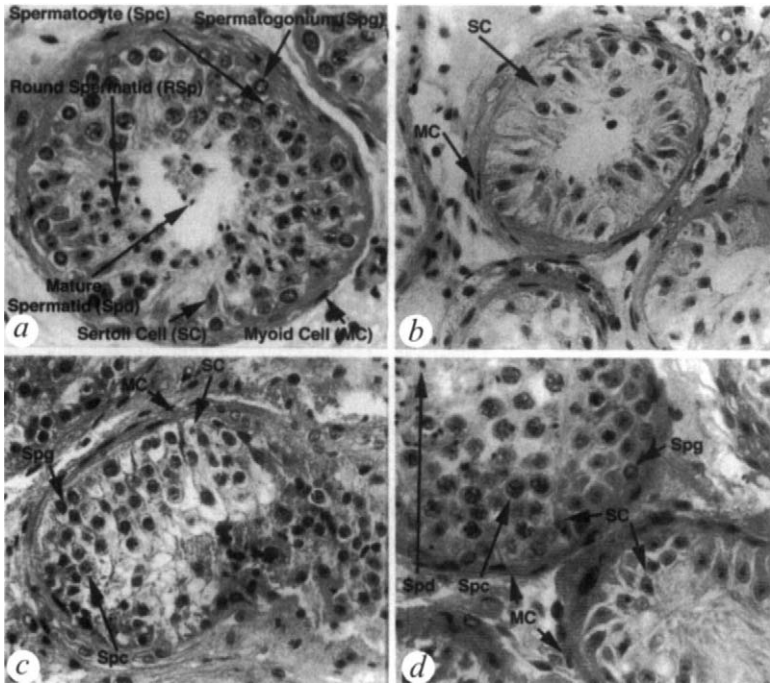


Fig. 2 Testicular histologies associated with *AZF* deletions. *a*, Photomicrograph of normal seminiferous tubule (in cross section) from fertile human male. Tubule is ringed by myoid cells and contains somatic (Sertoli) cells and following germ cells: spermatogonia, spermatocytes, round spermatids, and mature spermatids with condensed nuclei. *b*, Sertoli-cell-only syndrome; tubule from *AZF*-deleted male WHT2475. *c*, Testicular maturation arrest with no mature spermatids; tubule from *AZF*-deleted male WHT2415. *d*, Testicular maturation arrest with condensed spermatids in tubule at upper left; only Sertoli cells seen in tubule at lower right; from *AZF*-deleted male WHT2376. Staining; hematoxylin and eosin.

efforts to compare and integrate results from various centres. We sought to address these difficulties by systematically testing azoospermic men (and, as appropriate, male relatives) for the presence of a large collection of Y-specific STSs (sequence-tagged-sites), all detectable by PCR and widely available to the research community. More than 100 such STSs have been incorporated into a comprehensive physical map of the chromosome[14,15].

The phenotypes associated with such Y deletions also warrant further exploration. Azoospermia, the absence of sperm in semen, can be associated with a variety of abnormal testis histologies, ranging from the complete lack of germ cells to meiotic arrest with few or no mature spermatids[16]. Thus, azoospermia is a nonspecific finding associated with an array of histologically distinct spermatogenic disorders. If *AZF* exists, the histologic nature of testicular defects resulting from its absence would be of great interest. We have therefore focused our study on azoospermic men whose testes had been biopsied.

In this study, we set out to address the following questions: Is any part of the Y chromosome frequently and consistently deleted in association with severe spermatogenic defects? Do those deletions represent new mutations or heritable variations? What testicular histologies are observed in azoospermic men with Y deletions? And what genes are present in the deleted regions?

Y chromosome deletions in azoospermic men

We studied 89 men in whom semen analysis revealed no spermatozoa and in whom physical obstruction of the seminiferous pathways had been ruled out. These men were otherwise generally healthy. Of the 89 men, 78 had undergone testis biopsy, in all cases revealing absence of germ cells (Sertoli-cell-only syndrome; 42 cases) or a preponderance of premeiotic spermatogenic cells (testicular maturation arrest; 36 cases). These men were ascertained solely on the basis of semen analysis and testis biopsy. (Of the 89 men, 84 had undergone no previous chromosomal studies; the remaining five had been found to have normal 46,XY karyotypes.) As controls, we studied 90 men who had fathered children. The azoospermic and fertile men were of diverse ethnic origins.

The human Y chromosome is divided into euchromatic and heterochromatic halves (Fig. 1*a*). The heterochromatin, comprising distal Yq, is dispensable with regard to male fertility[17,18]. We focused on the 30-

◀ Fig. 1 Deletion mapping the *AZF* region on human Y-chromosome. *a*, Diagram of chromosome euchromatic, heterochromatic and pseudoautosomal regions with Yp telomere to left and Yq telomere to right. Above are indicated the centromere, *AZF* region, and previously cloned genes and pseudogenes. *b*, Low-resolution analysis of Y chromosomes of azoospermic men. Along top border are listed deletion intervals one through seven and, immediately below, 84 Y-chromosomal STSs, all previously reported (ref. 15; YRRM: ref. 11 as corrected in ref. 13); gene, pseudogene and locus names in parentheses. Shown below are results of testing men for presence (solid black box) or absence (minus sign) of each STS. First horizontal (solid black) line denotes presence of all 84 loci tested, as found in all 90 fertile men and in 77 of 89 azoospermic men tested. Also depicted are terminal and interstitial deletions (three and nine cases, respectively) in remaining 12 azoospermic men and, finally, interstitial deletions observed in Chandley and Hargreave's patients KUPAU, KLARD, NIKEI. Blank spaces or grey boxes indicate, respectively, inferred absence or presence (by interpolation) of markers for which assay was not performed. White boxes represent positive results to be interpreted in light of Y-specific-repeat nature of sequences assayed; these positive results likely reflect presence of closely related sequences elsewhere on chromosome. *c*, Higher resolution map of *AZF* region. Along top border are listed 22 intervals (defined by patient or YAC endpoints) and 47 markers, 30 of which were derived in this study. Shown below are results of testing for STSs in following samples: 90 fertile men, 12 azoospermic men in whom deletions were detected (testicular histologies indicated at right; results not shown for 77 azoospermic males with no deletions detected), three azoospermic patients of Chandley and Hargreave, 10 other individuals with Y breakpoints in the region[15], and 13 YACs[14]. A few of the fertile men were not tested for sY262, sY267, sY269, sY272 and sY273. STSs ordered so as to minimize number of apparent breakpoints in this set of patients and YACs. Some errors in ordering are likely given repeat-rich nature of region, and there is no information as to STS order within an interval. We refrain from naming newly defined subintervals until their order and STS content have been further verified. STS, patient, and YAC endpoint orderings are generally in agreement with those of Vollrath *et al.*[15] and Foote *et al.*[14], though WHT2168 breakpoint appears more proximal than previously reported. All available male relatives of azoospermic men with deletions (fathers of WHT2415, WHT2475, WHT2613, LGL5169, and LGL5697; brother of WHT2381; and paternal uncle of WHT2376) were tested and found to carry all markers listed in Fig. 1*b,c*. Male relatives of WHT2430, WHT2564, LGL5580, LGL5205 and LGL6815 not available.

megabase (Mb) euchromatic region, which includes proximal Yq, the centromere, and Yp, and for which a physical map of ordered STSs and overlapping YAC (yeast artificial chromosome) clones has been constructed[14,15].

We made no assumptions as to the number or location of spermatogenesis genes on the Y chromosome, but instead tested each azoospermic or fertile male for the presence of 83 Y-specific STSs previously shown to blanket most of the euchromatic region. Given that the absence of even a single STS might be biologically significant and that we would perform more than 14,000 tests, we took two precautions to minimize the number of false negative results. First, we employed only those STSs (a total of 84) whose PCR assays reliably yielded positive results on normal males; we avoided previously mapped STSs whose PCR assays were prone to inconsistency. Second, we did not record an STS as absent from a male until at least three successive attempts to PCR amplify the locus yielded negative results.

Using this set of Y-DNA markers, deletions of portions of Yq were found in 12 of the 89 azoospermic men. No deletions were detected in the 90 fertile men (Fig. 1*b*). Three deletions were of terminal portions of Yq, while the other nine were interstitial. If the deletions are the cause of azoospermia, then one would expect them to represent new mutations not present in the azoospermic males' fathers or other paternal relatives. For seven of the 12 deletions, samples were available from fathers, brothers, or paternal uncles, and in all seven cases the male relatives were found to carry intact Y chromosomes (Fig. 1*c*). We conclude that the deletions are probably the cause of azoospermia in these men.

All 12 deletions overlap a region likely to harbour one or more spermatogenesis genes. We wished to know whether this region was also absent in the three azoospermic patients (KLARD, NIKEI, and KUPAU) whose Y deletions provided the foundation for the identification of the *YRRM* genes as *AZF* candidates[11]. Using genomic DNAs (kindly provided by A. Chandley and T. Hargreaves), we determined that the STSs common to the 12 deletions we identified are also absent in KLARD, NIKEI and KUPAU (Fig. 1*b*). These results provide a consistent definition of the *AZF* region, deletion of which appears to account for roughly 13% (12 of 89 cases) of spermatogenic defects so severe as to result in azoospermia. (As discussed below, these results do not exclude the existence of genes essential for spermatogenesis elsewhere on the human Y chromosome.)

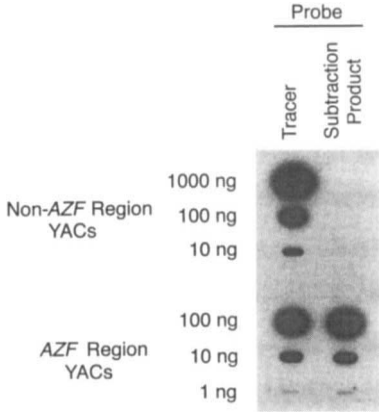


Fig. 3 *AZF*-region specificity of YAC subtraction product. Autoradiogram produced by hybridizing ^{32}P-labelled tracer DNA or subtraction product to membrane-bound DNA from pooled YACs — 58 YACs from outside *AZF* region or 8 YACs from *AZF* region and immediate environs. Indicated quantities of pooled YAC DNAs were spotted onto nylon membrane; tenfold greater quantities were used for non-*AZF*-region YACs to compensate for higher complexity of this YAC pool. Hybridization: 20 hr at 65 °C in 5X SSC (1× = 0.15M NaCl, 15 mM Na citrate pH7.4), 5× Denhardt's solution (1X = 0.02% Ficoll 400, 0.02% polyvinyl-pyrrolidone, 0.02% bovine serum albumin), 1% SDS, 50 μg ml^{-1} salmon sperm DNA. Wash: three times for 15 min each at 65°C in 0.1× SSC, 0.1% SDS.

Refined map of *AZF* region

Unforeseen hazards can imperil efforts to map precisely phenotypes on the human Y chromosome. Linkage analysis, so useful with regard to other nuclear chromosomes, cannot be employed to validate or refute conclusions drawn from studies of Y deletions. An individual in whom all available markers indicate the absence of a single, interstitial portion of the Y chromosome (for example, the deletion of *ZFY* in female WHT1014; ref. 19) may also be deleted for a second, noncontiguous region (the deletion of *SRY* in WHT1014; refs 20,21). Such difficulties may be compounded in the vicinity of *AZF*, since Y-specific repetitive DNA sequences comprise most of this portion of the chromosome[14]; the organization of DNA sequences in this region is difficult to deduce and may vary among normal males.

Recognizing such hazards, we set out to scrutinize and potentially redefine the *AZF* region using an expanded collection of Y-DNA landmarks from a larger region encompassing all nine interstitial deletions we had identified. We generated 30 additional markers specific to this portion of the chromosome by various methods, including sequencing the ends of YAC inserts, exon trapping, and 'YAC subtraction' (described below). We tested for the presence of these STSs in all 90 fertile and all 89 azoospermic men — those with and those without deletions already detected. To improve the resolution and accuracy with which the STSs were ordered, we also tested for their presence in 10 other individuals with partial Y chromosomes and nine YACs previously shown to have breakpoints in this region[14,15]. The results allowed us to refine the physical map of the region deleted in the azoospermic men (Fig. 1*c*). Overlapping YACs yOX198 and yOX17 (500 and 900 kb, respectively) appear to span most of the mapped region, suggesting that it encompasses roughly 10^6 bp. The map incorporates 56 loci at an estimated average spacing of roughly 20 kb.

No additional deletions were detected using our enhanced map. The 12 deletions we had detected using the initial set of Y-specific STSs — and the deletions in Chandley and Hargreave's patients KLARD, NIKEI, and KUPAU — were also detected by many of the supplemental STSs (Fig. 1*c*). The smallest deletion found was in azoospermic male WHT2564. His deletion, which encompasses 35 Y-DNA loci, appears to be contained in its entirety within each of the other deletions associated with azoospermia. We will use the term '*AZF* region' to denote the portion of the Y chromosome deleted in WHT2564. We estimate that the *AZF* region encompasses roughly 5×10^5 bp.

Though the resolution of this physical map is limited, our findings in azoospermic men with Y chromosomal deletions suggest that their breakpoints may be clustered (Fig. 1*c*). Seven of the 15 azoospermia-associated deletions have proximal breakpoints between *YRRM1*/*YRRM2* and

sY153. Seven of the 12 interstitial deletions have distal breakpoints between sY158 and sY159.

Histology of spermatogenic defects

Nine of the 12 azoospermic men in whom we detected deletions of the *AZF* region had undergone testis biopsy (Figs 1*c*, 2). We were surprised to discover that the histologic appearance of the testis differed dramatically among these nine men. Five of the men appeared to have no germ cells (Sertoli-cell-only syndrome), while the other four had spermatogonia and premeiotic spermatogenic cells in at least some seminiferous tubules (testicular maturation arrest). Most surprisingly, in two of the men with spermatogenic arrest (WHT2376 and WHT2613), testis biopsy revealed occasional mature, condensed spermatids (Fig. 2*d*). There is no obvious correlation between the size of the Y deletion and the severity of the spermatogenic defect (Fig. 1*c*).

Cosmid cloning and exon trapping

Although the *YRRM1* and *YRRM2* genes had been reported as deleted in azoospermic males KLARD, NIKEI, and KUPAU[11], the *AZF* region was not searched systematically for transcription units. Indeed, given the difficulty of mapping this portion of the Y chromosome, we were unsure that the *AZF* region as defined here had been included in any previous gene hunt. To identify transcripts that might encode *AZF*, we used cosmids from the *AZF* region as substrates for exon trapping. Since the region as defined by WHT2564 need not contain the entirety of the *AZF* transcription unit, we included the adjoining regions of the chromosome in this search for genes.

We began by identifying 180 cosmid clones providing five to tenfold coverage of the area. 120 of these cosmids were isolated from a Y-enriched library using a complex hybridization probe prepared by 'YAC subtraction', a novel application of DNA subtraction technology. Subtraction methods allow one to purify DNA fragments that are present in one population ('tracer') but absent in another ('driver')[22–24]. In YAC subtraction, tracer and driver consist of YAC (or multiple YAC) DNAs. In this case, eight overlapping YACs spanning the *AZF* region were pooled and used as tracer; 58 YACs spanning the remainder of the euchromatic Y were used as driver. This subtraction was intended to yield a pool of *AZF* region sequences from which had been removed (i) Y-specific repeats represented outside the *AZF* region and (ii) interspersed repeats scattered throughout the genome. These goals were met. The subtraction product hybridized exclusively to *AZF*-region YACs (Fig 3), while the tracer from which it derived hybridized strongly to both *AZF* and non-*AZF* region YACs. Hybridization of the subtraction product to the Y-enriched cosmid library identified 120 clones, 107 of which were found to contain DNA landmarks mapped to the *AZF* region or its immediate environs. The map location of the remaining 13 cosmids was not determined. (About 900 cosmid clones were detected when tracer (*AZF* region) YACs were preannealed with human placental DNA and hybridized to the same library, probably because the YACs contain Y-specific repeats not blocked efficiently by human placental DNA.)

The Y-enriched cosmid library did not contain clones corresponding to certain loci in the most distal portion of the *AZF* region. To ensure representation of this distal portion in our exon-trapping experiments, we identified three P1 phage clones from the region and subcloned YAC yOX17 to obtain 60 cosmids.

Each of the 180 cosmid and three P1 clones from the *AZF* region and its environs was individually subcloned and subjected to exon trapping. Nucleotide sequencing of trapping products revealed 16 potential exons.

No evidence of *YRRM* sequences in *AZF* region

As judged by nucleotide sequence analysis, none of the potential exons recovered from the *AZF* region or its immediate environs derived from the *YRRM* genes, which have been proposed as *AZF* candidates[11]. Although surprising, the absence of *YRRM* from among the *AZF*-region exons was consistent with several other observations. As judged by PCR using primer sequences previously reported[11,13], *YRRM* sequences are absent from YACs spanning the *AZF* region (YACs yOX58, yOX69, yOX101, yOX102, yOX103, yOX105, yOX134, yOX197 and yOX198). Similarly, *YRRM* sequences are not present in any of the *AZF*-region cosmid or P1 clones we identified. Instead, *YRRM*-related sequences appear to be present in diverse locations across the Y chromosome, including proximal Yp (YACs yOX75, yOX76, yOX98, and yOX99), proximal Yq (YACs yOX119, yOX120, yOX124, yOX127 and yOX162) and more distal Yq (yOX5, yOX19, yOX40, yOX97 and yOX140). None of these *YRRM*-positive YACs appears to overlap the *AZF* region, although the overlapping YACs yOX5 and yOX40 lie just proximal to it. As judged by PCR, both *YRRM1* and *YRRM2* are present in all 89 azoospermic men we studied, including the 12 in whom we had detected deletions (Fig. 4). Indeed, we found both *YRRM1* and *YRRM2* to be present in patients KUPAU, NIKEI and KLARD. We conclude that *YRRM* sequences are dispersed to several locations on the human Y chromosome, but we find no evidence of *YRRM* sequences in the *AZF* region.

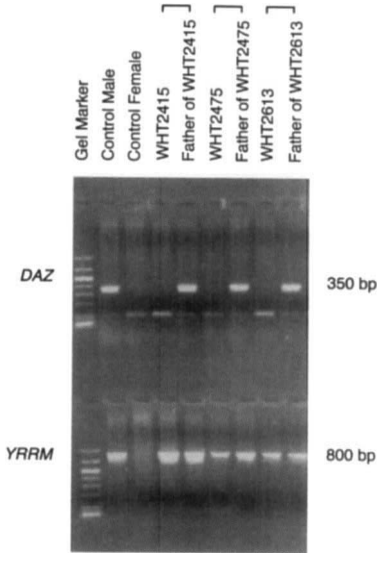


Fig. 4 PCR testing of genomic DNAs from three azoospermic men and their fathers for *DAZ* and *YRRM*. At top: sY254, an STS within *DAZ*. At bottom: *YRRM1* (ref. 11 as corrected in ref. 13). Products separated by electrophoresis in 2% agarose gel and visualized by ethidium bromide staining.

An *AZF*-region gene transcribed in testes

Analysis of the exon trapping products revealed a novel transcription unit in the *AZF* region. Trapping products that mapped to the region were identified by their presence in normal males and absence in both normal females and *AZF*-deleted azoospermic males (as judged by PCR amplification). Products mapping to

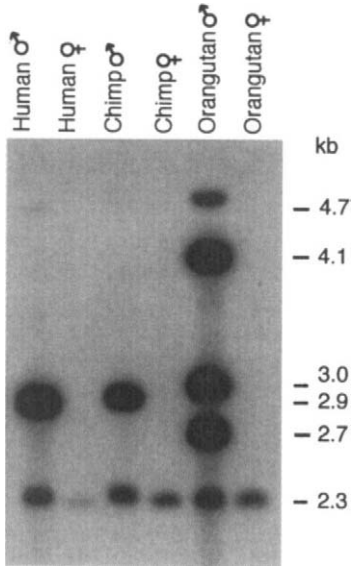


Fig. 5 A single copy of *DAZ* gene on human and chimpanzee Y chromosomes. Autoradiogram produced by hybridizing *DAZ* exon 325.7 to Southern blot of *Eco*RI-digested genomic DNAs. Sizes (in kb) of hybridizing fragments indicated at right. Hybridization: 20 h at 42 °C in 50% formamide, 5× SSC, 1× Denhardt's, 20 mM Na phosphate pH 6.6, 0.005% denatured salmon sperm DNA, 1% SDS, 10% dextran sulphate. Wash as in Fig. 3.

the *AZF* region were hybridized to Southern blots of restriction-digested human male and female genomic DNAs and to plaque lifts of a cDNA library prepared from human adult testis, where it seemed likely that *AZF* would be expressed. Four exon-trapping products fulfilled these criteria: they mapped to the *AZF* region; detected one or more male-specific bands by Southern blotting; and hybridized to clones in the cDNA library. These four products recognized overlapping sets of cDNA clones, and subsequent analysis confirmed that all four derived from a single transcription unit. Approximately one in 5000 clones in the testis cDNA library derives from this gene, which we will refer to as *DAZ* (*D*eleted in *AZ*oospermia).

PCR assays with primers corresponding to one of the trapped exons (325.7) confirmed the absence of the *DAZ* gene in all 12 azoospermic men in whom we had detected Yq deletions; *DAZ* is present in their fathers and other male relatives (Fig. 4). *DAZ* is absent also in KUPAU and NIKEI. As DNA was limiting, KLARD was not tested.

As judged by Southern blotting, there appears to be a single copy of *DAZ* on the human and chimpanzee Y chromosomes. When hybridized to *Eco*RI-digested genomic DNAs from either species, exon 325.7 detected a single male-specific fragment, corresponding to *DAZ* (Fig. 5). (On this overexposed autoradiogram, one also observes a much less intensely hybridizing fragment common to males and females. The nature and origin of this fragment are not known.) *DAZ* sequences may be amplified on the orangutan Y chromosome, as suggested by the presence of three intensely hybridizing, male-specific fragments in this primate.

The results of screening cDNA libraries indicated that the human *DAZ* gene is transcribed in the adult testis. To confirm this result and to assess whether the gene is transcribed elsewhere, we hybridized exon 325.7 to northern blots of poly(A)-selected RNAs from 16 different adult human organs. We observed a 3.5-kb transcript in the testis and detected no evidence of transcription elsewhere (Fig. 6).

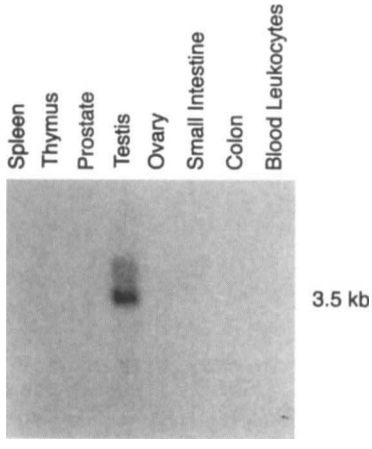


Fig. 6 Transcription of *DAZ* gene in human adult testis. Autoradiogram produced by hybridizing *DAZ* exon 325.7 to Northern blot of poly(A)⁺ RNAs (2 μg/lane) from human tissues (Clontech). Additional negative results obtained with RNAs from heart, brain, placenta, lung, liver, skeletal muscle, kidney and pancreas (not shown). Hybridization at 47 °C, otherwise as in Fig. 5.

A putative RNA binding protein

Nucleotide sequence analysis of *DAZ* cDNA clones isolated from the human adult testis library revealed a single long open reading frame (Fig. 7*a*). The first ATG in this open reading frame (position 1 in Fig. 7*a*) occurs in a sequence context that is favourable for initiation of translation[25]. Beginning at this ATG, the transcript appears to encode a protein of 366 amino acids, with a predicted molecular weight of 41,257. Though the library from which *DAZ* clones were isolated was constructed using poly(A)⁺ RNA and oligo(dT) priming, we have yet to identify a 3' poly(A) tail in any *DAZ* cDNA clone.

The features of the *DAZ* coding region include seven tandem repeats of a 72-nucleotide unit (Fig. 7*a,b*). The repeats differ from each other by at most a few nucleotides, suggesting that the repeats have been generated or homogenized by unequal crossing over during primate evolution. Curiously, these *DAZ* repeats exhibit remarkable nucleotide identity to *DYS1*, an extremely polymorphic family of repetitive sequences specific to Yq (Fig. 7*b*)[26]. To our knowledge, there is no evidence that *DYS1* sequences are transcribed. The tandem repeats in the *DAZ* nucleotide sequence appear to be translated into seven repeats of a 24-amino acid unit, which comprise most of the carboxy-terminal half of the predicted protein.

Within the N-terminal half of the DAZ protein is an 85-residue domain whose amino acid sequence matches the RNP/RRM consensus observed in many proteins that bind RNA or single-stranded DNA (Fig. 7*c*). Similar RNP/RRM domains are found, for example, in the mammalian polyadenylate binding proteins (PABP) and the *Drosophila* sex-lethal protein. The remainder of the predicted amino acid sequence (including the tandem repeats described above) is characterized by a high concentration of proline, glutamine and tyrosine residues, as is typical of many RNP/RRM proteins[27,28]. We conclude that the DAZ protein likely functions by binding RNA, or possibly single-stranded DNA.

Discussion

Frequent *de novo* deletion of an *Azoospermia Factor*. We examined the Y chromosomes of men with spermatogenic defects so severe as to result in absence of sperm in semen, who were otherwise healthy, and who had undergone no previous chromosomal testing. About 13% of such azoospermic men have *de novo* deletions of interstitial or terminal portions of Yq. All 12 such deletions we detected overlap, defining an '*AZF* region' which appears to measure several hundred kb and which is likely to harbour one or more genes required for

a

```
                                                                                                  -256 ctggttcggcccacct
-240  ctgaaggttccagaatcgatagtgaattcgtgagtcggcctgcgctcctcagcctggcggttctacctccgagggttcgcccgcccttggttttcctcttacaccttagccttttggctcctt
-120  tgaccactcgaagccccacagcgtgttccagcggacttcaccagcagacccagaagtggtgggtgaaacactgccttctgttcctccttgagcctgtcgggagctgctgcctgccaccacc

   1  ATGTCTGCTGCAAATCCTGAGACTCCAAACTCAACCATCTCCAGAGAGGCCAGCACCCAGTCTTCATCAGCTGCAGCTAGCCAAGGCTGGGTGTTACCAGAAGGCAAAATCGTGCCAAAC
      M  S  A  A  N  P  E  T  P  N  S  T  I  S  R  E  A  S  T  Q  S  S  S  A  A  A  S  Q  G  W  V  L  P  E  G  K  I  V  P  N

 121  ACTGTTTTTGTTGGTGGAATTGATGCTAGGATGGATGAAACTGAGATTGGAAGCTGCTTTGGTAGATACGGTTCAGTGAAAGAAGTGAAGATAATCACGAATCGAACTGGTGTGTCCAAA
      T  V  F  V  G  G  I  D  A  R  M  D  E  T  E  I  G  S  C  F  G  R  Y  G  S  V  K  E  V  K  I  I  T  N  R  T  G  V  S  K

 241  GGCTATGGATTTGTTTCGTTTGTTAATGACGTGGATGTCCAGAAGATAGTAGGATCACAGATACATTTCCATGGTAAAAAGCTGAAGCTGGGCCCTGCAATCAGGAAACAAAAGTTATGT
      G  Y  G  F  V  S  F  V  N  D  V  D  V  Q  K  I  V  G  S  Q  I  H  F  H  G  K  K  L  K  L  G  P  A  I  R  K  Q  K  L  C

 361  GCTCGTCATGTGCAGCCACGTCCTTTGGTAGTTAATCCTCCTCCTCCACCACAGTTTCAGAACGTCTGGCGGAATCCAAACACTGAAACCTACCTGCAGCCCCAAATCACGCCGAATCCT
      A  R  H  V  Q  P  R  P  L  V  V  N  P  P  P  P  P  Q  F  Q  N  V  W  R  N  P  N  T  E  T  Y  L  Q  P  Q  I  T  P  N  P

 481  GTAACTCAGCACGTTCAGGCTTATTCTGCTTATCCACATTCACCAGGTCAGGTCATCACTGGATGTCAGTTGCTTGTATATAATTATCAGGAATATCCTACTTATCCCGATTCACCATTT
      V  T  Q  H  V  Q  A  Y  S  A  Y  P  H  S  P  G  Q  V  I  T  G  C  Q  L  L  V  Y  N  Y  Q  E  Y  P  T  Y  P  D  S  P  F

 601  CAGGTCACCACTGGATATCAGTTGCCTGTATATAATTATCAGCCATTTCCTGCTTATCCAAGTTCACCATTTCAGGTCACTGCTGGATATCAGTTGCCTGTATATAATTATCAGGCATTT
      Q  V  T  T  G  Y  Q  L  P  V  Y  N  Y  Q  P  F  P  A  Y  P  S  S  P  F  Q  V  T  A  G  Y  Q  L  P  V  Y  N  Y  Q  A  F

 721  CCTGCTTATCCAAGTTCACCATTTCAGGTCACCACTGGATATCAGTTGCCTGTATATAATTATCAGGCATTTCCTGCTTATCCAAGTTCACCATTTCAGGTCACCACTGGATATCAGTTG
      P  A  Y  P  S  S  P  F  Q  V  T  T  G  Y  Q  L  P  V  Y  N  Y  Q  A  F  P  A  Y  P  S  S  P  F  Q  V  T  T  G  Y  Q  L

 841  CCTGTATATAATTATCAGGCATTTCCTGCTTATCCAAGTTCACCATTTCAGGTCACCACTGGATATCAGTTGCCTGTATATAATTATCAGGCATTTCCTGCTTATCCAAATTCAGCAGTT
      P  V  Y  N  Y  Q  A  F  P  A  Y  P  S  S  P  F  Q  V  T  T  G  Y  Q  L  P  V  Y  N  Y  Q  A  F  P  A  Y  P  N  S  A  V

 961  CAGGTCACCACTGGATATCAGTTCCATGTATACAATTACCAGATGCCACCGCAGTGCCCTGTTGGGGAGCAAAGGAGAAATCTGTGGACCGAAGCATACAAATGGTGGTATCTTGTCTGT
      Q  V  T  T  G  Y  Q  F  H  V  Y  N  Y  Q  M  P  P  Q  C  P  V  G  E  Q  R  R  N  L  W  T  E  A  Y  K  W  W  Y  L  V  C

1081  TTAATCCAGAGAAGAGACTGAtaaattccgttgttactcaagatgactgcttcaagggtaaaagagtgcatcgctttagaagaagtttggcagtatttaaatctgttggatcctctcagc
      L  I  Q  R  R  D  *

1201  tatctagtttcatgggaagttgctggttttgaatattaagctaaaagttttccactattacagaaattctgaatttttggtaaatcacactgaaacttttctgtataacttgtattattaga
1321  ctctctagtttttatcttaactactgaaactgttcttcattagatgtttatttagaacctggtttctgtgtttaatatatagtttaaagtaacaaataatcgagactgaaagaatgttaaga
1441  tttatctgcaaggatttttaaaaaattgaaacttgcattttaaagtgtttaaaagcaaattactgactttcaaaaaagtttttaaaacctgatttgaaagctaacaattttggatagtct
1561  gaacacaagcatttcacttctccaagaagtacctgtgaacagtacaatatttcagtattgagctttgcatttatgatttatc 1641
```

b

```
DAZ Consensus Repeat      GCATTTCCTGCTTATCCAAGTTCACCATTTCAGGTCACCACTGGATATCAGTTGCC TGTATATAATTATCAG

                1   499 ··T·A·T··········CA·······GG·····T·····G·····T ·····  570
                2   571 ·A··A···A·····CGA·····························  642
DAZ             3   643 C·····························TG·····  714
Repeats         4   715 ·····  786
                5   787 ·····  858
                6   859 ·····  930
                7   931 ··················A····G··G·····C·A ·····C·····C··· 1002

DYS1            333 ··················A··············G··············C··T·····C·····  405
```

c

```
RRM/RNP Consensus

Human Gene Products:
   DAZ
   YRRM1/YRRM2
   Myc Single-Strand Binding Protein
   Cytolytic Lymphocyte TIAR Protein
   Poly(A) Binding Protein - RNP1
                              2
                              3
                              4
   hnRNP Protein A1 - RNP1
                         2
Others:
   Mouse Poly(A) Binding Protein
      (Testis-Specific)
   Drosophila RB97D
   Drosophila Sex-lethal - RNP1
                           2
```

```
----------LFVGNL-----E--L---F--FG-I----I---------KGFGFVXF--------A--------L-G----------
          IYIKGM     D       Y  V    V         R YA   Y        V        I

LPEGKIVPNTVFVGGIDARMDETEIGSCFGRYGSVKEVKIITN-RTGVSKGYGFVSFVNDVDVQKIVGSQIH--FHGKKLKLGPAIR
 MVEADHPGKLFIGGLNRETNEKMLKAVFGKHGPISEVLLIKD-RTSKSRGFAFITFENPADAKNAAKDMNGK-SLHGKAIKVEQAK
SGWDQLSKTNLYIRGLPPHTTDQDLVKLCQPYGKIVSTKAILDKTTNKCKGYGFVDFDSPAAAQKAVSALKASGVQAQKAKQQEQDP
QKKDTSNHGHVFVGDLSPEITTEDIKSAFAPFGKISDARVVKDMATGKSKGYGFVSFYNKLDAENAIVHMGGQWLGGRQIRTNWATR
SQSVENSSASLYVGDLEPSVSEAHLYDIFSPIGSVSSIRVCRDAITKTSLGYAYVNFNDHEAGRKAIEQLNYTPIKGRLCRIMWSQR
PSLRKKGSGNIFIKNLHPDIDNKALYDTFSVFGDILSSKIATD-ENGKSKGFGFVHFEEEGAAKEAIDALNGMLLNGQEIYVAPHLS
LEETKAHYTNLYVKNINSETTDEQFQELFAKFGPIVSASLEKD-ADGKLKGFGFVNYEKHEDAVKAVEALNDSELNGEKLYVGRAQK
EKMAKYQGVNLFVKNLDDSVDDEKLEEEFAPYGTITSAKVMRT-ENGKSKGFGFVCGSTPEEATKAITEKNQQIVAGKPLYVAIAQR
SPKEPEQLRKLFIGGLSFETTDESLRSHFEQWGTLTDCVVMRDPNTKRSRGFGFVTYATVEEVDAAMNARPHK-VDGRVVEPKRAVS
RPGAHLTVKKIEVGGIKEDTEEHHLRDYFEQFGKIEVIEIMTDRGSGKKKGFAFVTFDDHDSVDKIVIQKYHT-VNGHNCEVRKALS


LGARAKEFTNVYIKNFGDRMDDETLNGLFGRFGQILSVKVMTD-EGGKSKGFGFVSFERHEDAQKAVDEMNGKELNGKQIYVGRAQK
DICELEHLRKLFIGGLAPYTTEENLKLFYGQWGKVVDVVVMRDAATKRSRGFGFITYTKSLMVDRAQENRPHI IDGKTVEAKRALP
MNDPRASNTNLIVNYLPQDMTDRELYALFRAIGPINTCRIMRDYKTGYSFGYAFVDFTSEMDSQRAIKVLNGITVRNKRLKVSYARP
PGGESIKDTNLYVTNLPRTITDDQLDTIFGKYGSIVQKNILRDKLTGRPRGVAFVRYNKREEAQEAISALNNVIPEGGSQPLSVRLA
```

Fig. 7 *DAZ* cDNA sequence and predicted amino acid sequence of encoded protein. *a*, Nucleotide sequence is composite of 1) cDNA insert of plasmid pDP1577 and 2) 5' RACE product obtained using adult human testis RNA as template. (5' RACE products overlapped insert of pDP1577 by 470 nucleotides and extended 143 nucleotides further 5'.) Composite cDNA sequence is incomplete at 3' end, which may account for its being smaller than 3.5-kb transcript observed by northern blotting (Fig. 6). Predicted 366-amino-acid sequence is immediately beneath nucleotide sequence; RNP/RRM domain (Fig. 1*c*) is boxed. Seven tandem repeats of a 72-nucleotide unit (Fig. 1*b*) are underlined. Numbering of nucleotides and amino acids begins with first in-frame AUG codon. GenBank accession number U21663. *b*, Tandem repeats within *DAZ* coding sequence. At top is consensus nucleotide sequence of 72-bp *DAZ* repeats. Below are the seven *DAZ* repeats (numbering of nucleotides as in Fig. 1*a*) and portion of nucleotide sequence of plasmid p49f (*DYS1*; ref 26). Dots represent identity to *DAZ* repeat consensus. Apart from a single nucleotide insertion, portion of *DYS1* shown is collinear with *DAZ* repeats. *c*, Amino acid sequences of RNP/RRM domains in DAZ and other proteins. At top is consensus sequence[28]; dashes indicate positions where no consensus is apparent. Regions most highly conserved are shaded. List of other proteins is representative but not exhaustive.

Table 1 Y-chromosomal STSs

STS	Left primer	Right primer	Product size (bp)
sY201	TGTTGTACGTAGAAAAAGGATATTTTACC	ATATGGTAAACCACTTTTTAAAATTGCCA	99
sY202	ACAGTTTGAAATGAAATTTTAAATGTGTT	TGACAAAGTGAGACCCTACTACTA	121
sY203	AAGGATATTTTACCTTTGGTAAT	GTGGAGCAGTGACCTGAAAT	157
sY204	CCTTTGGTAATATTTTGGTTATAT	ACTTGGATAAGCAGGAAATGGCTG	119
sY206	ACAGAATTTCAGTTGTATTTTTATTT	ACCCTCCAAGATATTAATTCTTTG	143
sY207	AATTAAAGGACCCTTAAATTCATT	CCTCTGAAAGATTAATATATGGTTCT	153
sY208	GGACATAGTCCTGCTTAAGAAAAAGTGG	ACGTGGTTCAGGAGGTCTACTATTCTA	140
sY220	ATGGGTGAGAAGCCTGATTGT	TGGGAAAGCCTGAACTGCC	109
sY221	GTAAGCCCCAGATACCCTCC	AAATTGTTTGGAAAAGGACACC	113
sY224	ATAGTTAGTTTTGTGGTAACAT	CATAGCCTCTATGCAGATGGG	158
sY231	ATTGATGTGTTGCCCCAAAT	AGAGTGAACTTTAAATCCCAGCC	149
sY232	GACTCTACCACTTGGGCTCAATTT	AGATGTACCCAAGGCCACTG	91
sY233	AGTTAGTAAGCCCCAGTTATCCTCC	TTTGGAAAAGGACACCTTATTAGCCA	115
sY236	CCCCATCGGTAAACCAAATCA	CCCATTGAAGTTTCAAGGTGTCA	94
sY239[a]	CATTCATCTTCCCTTTTGAAGG	ATGCAAGTCGCAGGAAATCT	200
sY240[a]	TCAAATAGCAGCAATTTAATAT	GCACCTGAAGAGCTGCTTG	247
sY242	ACACAGTAGCAGCGGGAGTT	TCTGCCACTAAACTGTAAGCTCC	233
sY243	GTTTCTTCATAAGCAACCAAATTG	CAGATTATGCCACTGCCCTT	118
sY245	TTACTTCCTTAAGTCAAAGCGG	CTGAGACAGCAAGACCAATCC	101
sY247[a]	CTGGACAAAGCCTTGGAAAA	CTGCATGTCAATTGTGGGAC	114
sY248[a]	CATTGGCATGAATGTGTATTC	CTCTGGGACAAGTGTTCCTT	94
sY249	GACAAAGGGCTGATGATTTA	CATCACCTTTACTTTTTAAATGG	114
sY254[b]	GGGTGTTACCAGAAGGCAAA	GAACCGTATCTACCAAAGCAGC	107
sY255[b]	GTTACAGGATTCGGCGTGAT	CTCGTCATGTGCAGCCAC	126
sY257	AGGTTGTTTGGCCTTGAGC	TCTATGATCTGTACCCGGTGC	123
sY262	AGCTCACTGCAAGCAACAGA	CCACCATCCCCCTTCTTC	100
sY267	GAATGTGTATTCAAGGACTTCTCG	TACTTCCTTCGGGGCCTCT	102
sY269	CTCTGGGACAAGTGTTCCTTG	CATTGGCATGAATGTGTATTCA	94
sY272	GGTGAGTCAAATTAGTCAATGTCC	CCTTACCACAGGACAGAGGG	93
sY273	GGTCTTTAAAAGGTGAGTCAAATT	AGACAGAGGGAACTTCAAGACC	95

[a]Anneal at 62 °C; otherwise, PCR conditions as indicated in Methods. [b]Within *DAZ* gene.

spermatogenesis. We conclude that an *AZF* gene (or genes) does in fact exist on the human Y chromosome, and that its *de novo* deletion is among the most common causes of severe spermatogenic defects.

In the present series of 89 azoospermic men, we did not detect *de novo* Y deletions outside the *AZF* region (Fig. 1). These results do not exclude the existence of genes essential for spermatogenesis elsewhere on the human Y chromosome. If spermatogenesis genes exist elsewhere on the human Y chromosome, then *de novo* deletions involving those genes are probably less extensive or less common than those described here or, alternatively, result in phenotypes less severe than azoospermia. Vogt *et al.*[12] have reported an azoospermic male with a *de novo* interstitial deletion located more proximally on Yq.

Deletions of *AZF* arise in human populations at a remarkable frequency. Roughly one in 1,000 men is azoospermic due to severe spermatogenic defects[3]. As described here, *AZF* is absent in approximately one in eight such men, though present in their fathers. Thus, it appears that at least one in 10^4 newborn human males carries a *de novo* deletion of *AZF*.

By what mechanism do these deletions arise? Similar frequencies of *de novo* deletion are observed in steroid sulphatase deficiency, and spinal muscular atrophy. In both cases, deletions are thought to arise via recombination between duplicated or otherwise repeated sequences flanking the critical gene(s) and specific to the particular chromosomal region[29–31]. A similar mechanism may be operating on the Y chromosome to produce deletions of *AZF*. This hypothesis is attractive since the region surrounding *AZF* is rich in Y-specific repetitive sequences. Consistent with but not proof of this hypothesis is the apparent clustering of breakpoints observed among the interstitial Yq deletions (Fig. 1*c*). It remains to be seen whether these apparent deletion hotspots coincide precisely with Y-specific repetitive sequences.

A spectrum of spermatogenic defects. Spermatogenesis is marked by an orderly progression of distinct cell types. One might have anticipated that the absence of *AZF* would interrupt this pathway at some discrete point. Our histologic studies of testis biopsies from azoospermic men with *AZF* deletions overturn such expectations. We find that azoospermic men with deletions of *AZF* exhibit a wide spectrum of spermatogenic defects, ranging from the complete absence of germ cells (Sertoli-cell-only syndrome) to meiotic arrest with occasional production of mature, condensed spermatids (Fig. 2).

Two different models could account for this diversity of phenotypes. First, multiple genes in close proximity on Yq could contribute to the phenotype, with the severity of the spermatogenic defect determined by the combination of Yq genes deleted. Alternatively, phenotypic diversity might reflect variable expressivity among individuals bearing functionally equivalent *AZF* null mutations; such variable expressivity could be in response to genetic background, environmental, or stochastic effects.

We favour Model 2 for two reasons. First, Model 1 would predict some correlation between the size of the Y deletion and the severity of the spermatogenic defect. No such correlation can be seen (Fig. 1*c*). Men who completely

lacked germ cells did not necessarily have the most extensive deletions, and men who produced occasional mature (postmeiotic) spermatids did not have the smallest deletions. Second, histologic variability can be observed not only between different *AZF*-deleted men but also between adjacent seminiferous tubules in a single individual. For example, in *AZF*-deleted individuals diagnosed as having testicular maturation arrest (with spermatogonia and immature, premeiotic cells in some tubules), it was not unusual to observe other tubules with no germ cells. In one such case, a tubule containing condensed spermatids was seen immediately adjacent to a tubule containing only Sertoli cells (Fig. 2*d*). Since the tubules within an individual are presumably genetically identical (mosaicism being a formal but unlikely possibility), this tubule-to-tubule variation in histology is not readily explained on genetic grounds and appears to imply the existence of important stochastic or microenvironmental influences. We suspect that the observed range of testis histologies reflects variable expressivity of functionally equivalent deletions of *AZF*, which is either a single gene or multiple genes in close proximity.

Our experimental observations suggest several additional conclusions. First, germ stem cells can persist, at least in some males, in the absence of *AZF*. Second, *AZF* is not absolutely required for the production of mature, condensed spermatids; it is not essential for progression of male germ cells through meiosis. Third, Sertoli-cell-only syndrome and testicular maturation arrest are not distinct disorders — at least when associated with Yq deletions — but represent different manifestations of the same underlying defect.

A 'pure male sterile' locus? *AZF*-deleted males, though azoospermic, are otherwise healthy, suggesting that *AZF* function may be restricted to or at least essential only for male germ cell development. To our knowledge, no other 'pure male sterile' locus has been identified in humans, although such genes have been identified in *Drosophila*, mice and other organisms[32,33]. It seems unlikely that *AZF* functions in the migration of primordial germ cells to the gonad, since this process occurs even in the Y chromosome's absence (for example, in XX or XO embryos)[34]. Given the testicular histologies observed in *AZF*-deleted men, it is conceivable that *AZF* facilitates differentiation of primordial germ cells into the spermatogonial stem cells present in adults. Alternatively, *AZF* might influence the destiny of these stem cells, which in normal males confront three alternative fates: proliferation, degeneration or differentiation (that is, entry into the spermatogenesis pathway). Future experiments may reveal which if any of these stem cell processes is altered in men lacking *AZF*. It seems likely that *AZF* would be expressed in the fetal and/or adult testis, but we have little basis on which to predict whether *AZF* should be expressed in germ cells or in somatic cells that support male germ cell proliferation and differentiation.

One might have supposed that the location of *AZF* on the Y chromosome would serve to prevent its expression in females, which might otherwise have had deleterious effects. This appears not to be the case, as the presence of distal Yq (including the entirety of the *AZF* region) has previously been reported in several chromosomally aberrant but nonetheless fertile, healthy women[15].

***DAZ* and *YRRM*.** Our mapping studies indicate that the *YRRM* genes are unlikely candidates for *AZF*. Ma *et al.* reported *de novo* deletions of one or more *YRRM* genes in several azoospermic males, including KLARD, NIKEI, and KUPAU, and on this basis they proposed the *YRRM* genes as *AZF* candidates[11]. However, we find no evidence of *YRRM* sequences in the *AZF* region despite the dispersal of such sequences to other locations on Yp and Yq. Of course, we cannot exclude the formal possibility (i) that *YRRM* genes retained on deleted Y chromosomes are transcriptionally silenced by position effects or (ii) that the *AZF* region contains a diverged homologue of *YRRM* not detected by presently available assays. These formal possibilities aside, there remains little basis for entertaining the *YRRM* genes as *AZF* candidates.

The *DAZ* (*D*eleted in *AZ*oospermia) gene is an attractive candidate for *AZF*. A single-copy gene located in the *AZF* region, *DAZ* is transcribed in the testis. *DAZ* is the only transcription unit that we have found to be deleted consistently in azoospermic males with *de novo* Yq deletions. However, we cannot exclude the existence of other transcription units in the *AZF* region. Nor do we have definitive evidence that loss of *DAZ* function was the primary or even a contributing cause of azoospermia in cases with Yq deletions. Our data suggest that approximately 87% of azoospermic men with Sertoli-cell-only syndrome or testicular maturation arrest retain *AZF* (and *DAZ*). Perhaps some of these men will be found to harbour *de novo* point mutations in *DAZ*.

Though *DAZ* is not a member of the *YRRM* gene family, the *DAZ* and *YRRM* genes are similar in certain respects. First, both *DAZ* and *YRRM* encode proteins with a single RNP/RRM domain (Fig. 7*c*)[11]. (Outside this domain, the proteins exhibit little sequence similarity.) By analogy to well characterized proteins containing such domains[27,28], both the DAZ and YRRM proteins are likely to function by binding RNA or possibly single-stranded DNA. Second, both the *DAZ* and *YRRM* coding sequences contain a series of near-perfect tandem repeats. *DAZ* contains seven tandem repeats of a 72-nucleotide unit, while *YRRM* contains four tandem repeats of a 111-nucleotide unit. (The sequences of the *DAZ* and *YRRM* repeats are dissimilar.) Third, both the *DAZ* and *YRRM* genes reside in regions of the Y chromosome rich in Y-specific repetitive sequences. *YRRM* itself comprises a sizable Y-specific gene family in humans and gorillas[11], while *DAZ*, though single-copy in humans and chimpanzees, may have been amplified to form a Y-specific family in orangutan (Fig. 5). The repeats within the *DAZ* coding sequence display remarkable nucleotide similarity to *DYS1*, a highly polymorphic family of Y-specific repetitive sequences. Fourth, both *DAZ* and *YRRM* appear to be expressed specifically in the testis. In sum, there are many molecular parallels between *DAZ* and *YRRM*.

It is tempting to speculate that testis-specific RNA-binding proteins encoded by *DAZ* and *YRRM* might function in male germ cell development. (*YRRM* may play a role in spermatogenesis even though it is not *AZF*, a locus to which attention is drawn because of its frequent deletion in human populations.) A precedent may be provided by the *Drosophila Rb97D* gene, which like human *DAZ* and *YRRM* encodes a protein with a single RNP/

RRM domain. Loss of *Rb97D* function results in degeneration of early spermatogenic cells and azoospermia[35]. Indeed, there is evidence that RNA-binding proteins function in mammalian spermatogenesis. In mice, protamine expression is translationally regulated by a protein that binds the protamine mRNA's 3' untranslated region[36] and other genes expressed during spermatogenesis may also be posttranscriptionally regulated[37]. It is interesting that the testes are grossly abnormal in males with fragile X syndrome, the only heritable human disease traced to a defective RNA-binding protein[38,39]. Perhaps RNA-binding proteins and posttranscriptional mechanisms figure prominently in the regulation of male germ cell development in mammals.

Methods

Testing for Y-specific STSs. Many Y chromosomal STSs for which we tested (Fig. 1) were described[15]. The remaining STSs are listed in Table 1 and were generated by nucleotide sequencing of (1) ends of YAC inserts, (2) YAC subtraction products, or (3) exon trapping products. YAC-insert ends were captured by inverse PCR[40] following digestion with *Hae*III, *Alu*I and *Taq*I. Oligonucleotide primers were selected so that nearly all PCR assays could be carried out under identical conditions[15]. *YRRM* primers were as described in ref 11. and corrected in ref. 13.

Human genomic DNAs were prepared from blood or lymphoblastoid cell lines[19]. PCR was performed in v-bottom, 96-well plates (MJ Research) in 20-µl volumes in 1.5 mM $MgCl_2$, 5 mM NH_4Cl, 10 mM tris (pH8.2), 50 mM KCl, 100 µM dNTP's, with 1 U of *Taq* DNA polymerase, 100 to 200 ng of human genomic DNA per reaction, and each primer at 1 µM. Thermocycling usually consisted of an initial denaturation of 5 min at 94 °C; 35 cycles of 1 min at 94 °C, 1.5 min at 58 °C,1 min at 72 °C; and, finally, 5 min at 72 °C. As indicated in Table 1, certain primer pairs were annealed at 62 °C. Reactions were stored at 4 °C until they were loaded onto 2 to 4% agarose gels for analysis.

We also tested individual Y-derived YACs[14] for STSs, in which case we employed 5 to 10 ng of total yeast genomic DNA as template and an annealing temperature of 62 °C.

YAC subtraction. The subtraction protocol of Rosenberg *et al.*[24] was modified for use with YAC DNAs. DNAs from 66 overlapping YACs spanning most of the Y chromosome's euchromatic region[14] were separated from yeast chromosomes by pulsed-field electrophoresis on 1.2% low-melt agarose gels, excised, and purified using Geneclean (Bio 101). 'Tracer' was prepared using DNA pooled from eight overlapping YACs (yOX69, yOX101, yOX102, yOX103, yOX104, yOX190, yOX192, yOX198) blanketing the *AZF* region. 100 ng of this DNA was digested with *Sau*3A and ligated to *Sau*3A-compatible PCR adapter (an equimolar mixture of GACACTCTCG-AGACATCACCGTCC and phosphorylated GATCGGACGGTGA-TGTCTCGAGAGTG). 'Drivers' were prepared from total yeast genomic DNA (strain AB1380) and from DNA pooled from 58 YACs spanning the remainder of the euchromatic portion of the Y chromosome. Yeast genomic DNA (1 µg) or pooled YAC DNA (100 ng) was sonicated to an average length of 1 kb, treated with Klenow fragment of DNA polymerase to produce blunt ends, and ligated to blunt-end PCR adapter (an equimolar mixture of AATTCTTGCGCCTTAAACCAAC and phosphorylated GTTGGTTTAAGGCGCAAG). Tracer and driver DNAs were then amplified separately using oligonucleotides OL25 and OL31DB, respectively, as PCR primers[24]. Subtractive hybridizations were carried out with the following in a total volume of 4 µl: 4 ng of amplified tracer DNA; 7 µg of amplified, biotinylated YAC driver DNA; 3 µg of amplified, biotinylated yeast genomic driver DNA; 20 µg of yeast tRNA; 5 µg of oligonucleotide OL30; and 2 µg of oligonucleotide OL25. Individual products of subtraction were sequenced after digesting bulk product with *Sau*3A and cloning into the *Bam*HI site of plasmid pBluescript KS(+) (Stratagene). To increase the sequence complexity of the subtraction product, an additional round of subtractive hybridization was performed using, as a third driver, 2 µg of DNA from 130 subtraction clones that had been pooled, amplified, and biotinylated as described above. The resulting subtraction product, in bulk, was radiolabeled and hybridized to high-density arrays of an 11,700-clone, Y-enriched cosmid library (LL0YNC03; Human Genome Center, Lawrence Livermore National Laboratory) according to the procedure of Holland *et al.*[41], resulting in identification of 120 cosmid clones.

Exon trapping. Substrates for exon trapping[42] included 120 cosmids identified by hybridization to YAC subtraction product, 60 cosmids constructed by subcloning YAC yOX17 in SuperCos1 (Stratagene), and three P1 clones identified by commercial screening (Genome Systems). These genomic clones were digested with *Bam*HI and *Bgl*II, individually subcloned into pSPL3 (Gibco-BRL) and transfected into COS7 cells. After 48 h growth, RNA was harvested using Trizol (Gibco-BRL). cDNA was synthesized, and clones that contained potential intron-exon boundaries were identified by PCR using primers flanking the cloning sites. These exon trapping products were sequenced, and from these sequences STSs were developed.

Characterization of potential exons. We further characterized exon trapping products whose corresponding STSs were male-specific and mapped to the *AZF* region, including exon 325.7 (subcloned as plasmid pDP1593), which proved to derive from the *DAZ* gene. To confirm male specificity and to look for evidence of transcription, potential exons were labelled with ^{32}P-dCTP by random priming and hybridized to Southern and northern blots as described [43]. Putative exons were then used as hybridization probes in screening a cDNA library (HL1161X, Clontech) constructed by oligo(dT) priming of mRNA from the testes of four human adults; hybridization (at 47 °C) and washing conditions were as published[43]. Nucleotide sequencing of *DAZ* cDNA clones was performed as described[43]. Since the composite length of *DAZ* cDNA clones was considerably shorter than the 3.5-kb transcript observed on northern blots (Fig. 6), we used a RACE protocol (5'-Amplifinder; Clontech) to capture the 5' portion of the *DAZ* transcript. We employed human adult testis RNA as starting template and the following two *DAZ* oligonucleotides as gene-specific primers: AACGAAACAAATCCATAGCCTTTG (for cDNA synthesis) and CTCGCTCGCCCAGAACCGTATCTA-CCAAAGCA (for secondary amplification). The resulting PCR products (approximately 500 bp) were cloned (TA cloning system; Invitrogen) and sequenced.

Acknowledgements

We thank B. Raphael, A. Hashem, R. Dredge, M. Velez-Stringer and C. Rosenberg for experimental and analytic contributions; A. Chandley and T. Hargreave for DNAs from patients KLARD, NIKEI and KUPAU; A. McMurray and J. Segre for advice on exon trapping, Lawrence Livermore National Laboratory for the flow-sorted cosmid library; and P. Bain, G. Fink, K. Jegalian, N. Kenmochi, B. Lahn, R. Polakiewicz and J. Seligman for comments on the manuscript. Supported by National Institutes of Health, U.S. Dept. of Agriculture, Academy of Finland, Sigrid Juselius Foundation and the Finnish Cultural Foundation. R.R. was recipient of a Damon-Runyon/Walter Winchell fellowship.

Received 3 May; accepted 19 June 1995.

1. Clermont, Y. Renewal of spermatogonia in man. *Am. J. Anat.* **118**, 509–524 (1966).
2. Dym, M. Spermatogonial stem cells of the testis. *Proc. natn. Acad. Sci. U.S.A.* **91**, 11287–11289 (1994).
3. Hull, M.G.R. *et al.* Population study of causes, treatment, and outcome of infertility. *Brit. med. J.* **291**, 1693–1697 (1985).
4. Silber, S.J. The relationship of abnormal semen parameters to male fertility. *Hum. Reprod.* **4**, 947–953 (1989).
5. Tiepolo, L. & Zuffardi, O. Localization of factors controlling spermatogenesis in the nonfluorescent portion of the human Y chromosome long arm. *Hum. Genet.* **34**, 119–124 (1976).
6. Fitch, N., Richer, C.-L., Pinsky, L. & Kahn, A. Deletion of the long arm of the Y chromosome and review of Y chromosome abnormalities. *Am. J. med. Genet.* **20**, 31–42 (1985).
7. Hartung, M., Devictor, M., Codaccioni, J.L. & Stahl, A. Yq deletion and failure of spermatogenesis. *Ann. Genet.* **31**, 21–26 (1988).
8. Johnson, M.D., Tho, S.P.T., Behzadian, A. & McDonough, P.G. Molecular scanning of Yq11 (interval 6) in men with Sertoli-cell-only syndrome. *Am. J. Obstet. Gynecol.* **161**, 1732–1737 (1989).
9. Skare, J. *et al.* Interstitial deletion involving most of Yq. *Am. J. med. Genet.* **36**, 394–397 (1990).
10. 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. molec. Genet.* **1**, 29–33 (1992).
11. Ma, K. *et al.* A Y chromosome gene family with RNA-binding protein homology: Candidates for the azoospermia factor AZF controlling human spermatogenesis. *Cell* **75**, 1287–1295 (1993).
12. Vogt, P. *et al.* Microdeletions in interval 6 of the Y chromosome of males with idiopathic sterility point to disruption of AZF, a human spermatogenesis gene. *Hum. Genet.* **89**, 491–496 (1992).
13. Kobayashi, K. *et al.* PCR analysis of the Y chromosome long arm in azoospermic patients: evidence for a second locus required for spermatogenesis. *Hum. molec. Genet.* **3**, 1965–1967 (1994).
14. Foote, S., Vollrath, D., Hilton, A. & Page, D.C. The human Y chromosome: Overlapping DNA clones spanning the euchromatic region. *Science* **258**, 60–66 (1992).
15. Vollrath, D. *et al.* The human Y chromosome: A 43-interval map based on naturally occurring deletions. *Science* **258**, 52–59 (1992).
16. Silber, S.J. Sertoli cell only revisited. *Hum. Reprod.* **10**, 1031–1032 (1995).
17. Borgaonkar, D.S. & Hollander, D.H. Quinacrine fluorescence of the human Y chromosome. *Nature* **230**, 52 (1971).
18. Andersson, M. *et al.* Y;autosome translocations and mosaicism in the aetiology of 45,X maleness: assignment of fertility factor to distal Yq11. *Hum. Genet.* **79**, 2–7 (1988).
19. Page, D.C. *et al.* The sex-determining region of the human Y chromosome encodes a finger protein. *Cell* **51**, 1091–1104 (1987).
20. Page, D.C., Fisher, E.M.C., McGillivray, B. & Brown, L.G. Additional deletion in sex-determining region of human Y chromosome resolves paradox of X,t(Y;22) female. *Nature* **346**, 279–281 (1990).
21. Sinclair, A.H. *et al.* A gene from the human sex-determining region encodes a protein with homology to a conserved DNA-binding motif. *Nature* **346**, 240–244 (1990).
22. Lamar, E.E. & Palmer, E. Y-encoded, species-specific DNA in mice: Evidence that the Y chromosome exists in two polymorphic forms in inbred strains. *Cell* **37**, 171–177 (1984).
23. Straus, D. & Ausubel, F.M. Genomic subtraction for cloning DNA corresponding to deletion mutations. *Proc. natn. Acad. Sci. .U.S.A.* **87**, 1889–1893 (1990).
24. Rosenberg, M., Przybylska, M. & Straus, D. RFLP subtraction: A method for making libraries of polymorphic markers. *Proc. natn. Acad. Sci. U.S.A.* **91**, 6113–6117 (1994).
25. Kozak, M. Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. *Cell* **44**, 283–292 (1986).
26. Lucotte, G. & Ngo, N.Y. p49f, A highly polymorphic probe, that detects Taq1 RFLPs on the human Y chromosome. *Nucl. Acids Res.* **13**, 8285 (1985).
27. Kenan, D.J., Query, C.C. & Keene, J.D. RNA recognition: towards identifying determinants of specificity. *Trends Biochem.* **16**, 214–220 (1991).
28. Burd, C.G. & Dreyfuss, G. Conserved structures and diversity of functions of RNA-binding proteins. *Science* **265**, 615–621 (1994).
29. Yen, P.H. *et al.* Frequent deletions of the human X chromosome distal short arm result from recombination between low copy repetitive elements. *Cell* **61**, 603–610 (1990).
30. Lefebvre, S. *et al.* Identification and characterization of a spinal muscular atrophy-determining gene. *Cell* **80**, 155–165 (1995).
31. Roy, N. *et al.* The gene for neuronal apoptosis inhibitory protein is partially deleted in individuals with spinal muscular atrophy. *Cell* **80**, 167–178 (1995).
32. Lindsley, D. & Tokuyasu, K.T. in *The Genetics and Biology of Drosophila* (eds Ashburner, M. & Wright, T.R.F.) 226–294 (Academic Press, London, 1980).
33. Magram, J. & Bishop, J.M. Dominant male sterility in mice caused by insertion of a transgene. *Proc. natn. Acad. Sci. U.S.A.* **88**, 10327–10331 (1991).
34. Carr, D.H., Haggar, R.A. & Hart, A.G. Germ cells in the ovaries of XO female infants. *Am. J. clin. Pathol.* **49**, 521–526 (1968).
35. Karsch-Mizrachi, I. & Haynes, S.R. The *Rb97D* gene encodes a potential RNA-binding protein required for spermatogenesis in Drosophila. *Nucl. Acids Res.* **21**, 2229–2235 (1993).
36. Kwon, Y.K., Murray, M.T. & Hecht, N.B. Proteins homologous to the Xenopus germ cell-specific RNA-binding proteins p54/p56 are temporally expressed in mouse male germ cells. *Dev. Biol.* **158**, 90–100 (1993).
37. Hecht, N.B. in *Cell and Molecular Biology of the Testis* (eds Desjardins, C. & Ewing, L.L.) 400–432 (Oxford University Press, New York, 1993).
38. Butler, M.G. *et al.* Anthropometric comparison of mentally retarded males with and without the fragile X syndrome. *Am. J. med. Genet.* **38**, 260–268 (1991).
39. Siomi, H., Siomi, M.C., Nussbaum, R.L. & Dreyfuss, G. The protein product of the fragile X gene, FMR1, has characteristics of an RNA-binding protein. *Cell* **74**, 291–298 (1993).
40. Haldi, M. *et al.* Large human YACs constructed in a *rad52* strain show a reduced rate of chimerism. *Genomics* **24**, 478–484 (1995).
41. Holland, J., Coffey, A.J., Giannelli, F. & Bentley, D.R. Vertical integration of cosmid and YAC resources for interval mapping on the X-chromosome. *Genomics* **15**, 297–304 (1993).
42. Duyk, G.M., Kim, S., Meyers, R.M. & Cox, D.R. Exon trapping: A genetic screen to identify candidate transcribed sequences in cloned mammalian genomic DNA. *Proc. natn. Acad. Sci. U.S.A.* **87**, 8995–8999 (1990).
43. Fisher, E.M.C. *et al.* Homologous ribosomal protein genes on the human X and Y chromosomes: Escape from X inactivation and possible implications for Turner syndrome. *Cell* **63**, 1205–1218 (1990).