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[5], Raaji Alagappan[1], Laura G.Brown[1], Michael Rosenberg[1,3], Steve Rozen[2], Tom Jaffe[1], Donald Straus[3], Outi Hovatta[6], Albert de la Chapelle[5], Sherman Silber[4] and David C.Page[1]

[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, MA 02142, [3]Department of Biology, Brandeis University, Waltham, MA 02254, [4]In Vitro Fertilization Program, St Luke's Hospital, St Louis, MO 63017, USA, [5]Department of Medical Genetics, University of Helsinki and [6]Family Federation of Finland, Kalevankatzu 16, Helsinki, Finland

We have detected deletions of portions of the Y chromosome long arm in 12 of 89 men with azoospermia (no spermatozoa in their 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 the complete absence of germ cells to spermatogenic arrest with the occasional production of condensed spermatids. We found no evidence of *YRRM* genes, recently proposed as *AZF* candidates in the *AZF* region. The region contains a single-copy gene, *DAZ* (deleted in azoospermia), 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.
Key words: AZF/DAZ/RNA-binding protein gene/spermatogenic defects/Y chromosome deletions

Introduction

Human spermatozoa are produced via a complex developmental process. Progression from spermatogonial stem cells to mature spermatozoa requires 65 days and involves an elaborate succession of distinct cell types (Clermont, 1966; Dym, 1994). The process is punctuated by at least three mitotic and two meiotic divisions. Meanwhile, the genome is repackaged — with protamines rather than

***Previously published in *Nature Genetics* (1995) 10, 383–393 Reijo *et al.* Reprinted by kind permission.**

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.

Some 2% of human males are infertile because of severe defects in sperm production (Hull *et al.*, 1985; Silber, 1989). 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, immunological factors, anatomic malformations 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. The 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 (1976). 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 (1976) 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 (Fitch *et al.*, 1985; Hartung *et al.*, 1988). 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 a 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 (Johnson *et al.*, 1989; Skare *et al.*, 1990; Ma *et al.*, 1992, 1993; Vogt *et al.*, 1992; Kobayashi *et al.*, 1994). These interstitial deletions do not involve regions of the Y chromosome known to pair and recombine with the X chromosome, and therefore seemed less likely to perturb sex chromosome behaviour in meiosis. Indeed, the finding of overlapping interstitial Yq deletions in three azoospermic males led Ma *et al.* (1993) 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.

Several aspects of the *AZF* hypothesis merited further scrutiny, motivating this study. First, it has proved 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.* (1993). 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. In reality, however, 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 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 sequence-tagged sites (STS), all detectable by polymerase chain reaction (PCR) and widely available to the research community. More than 100 such STS have been incorporated in a comprehensive physical map of the chromosome (Foote *et al.*, 1992; Vollrath *et al.*, 1992).

The phenotypes associated with such Y deletions also warrant further exploration. Azoospermia, the absence of spermatozoa 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 (Silber, 1995). Thus, azoospermia is a non-specific finding associated with an array of histologically distinct spermatogenic disorders. If *AZF* exists, the histological nature of testicular defects resulting from its absence would be of great interest. Therefore we have 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?

Materials and methods

Testing for Y-specific STS

Many Y chromosomal STSs for which we tested (Figure 1) have been described previously (Kobayashi *et al.*, 1992). The remaining STSs are listed in Table I and were generated by nucleotide sequencing of ends of YAC inserts, YAC subtraction products or exon-trappings products. YAC insert ends were captured by inverse PCR (Haldi *et al.*, 1995) following digestion with H*ae*III, A*lu*I and

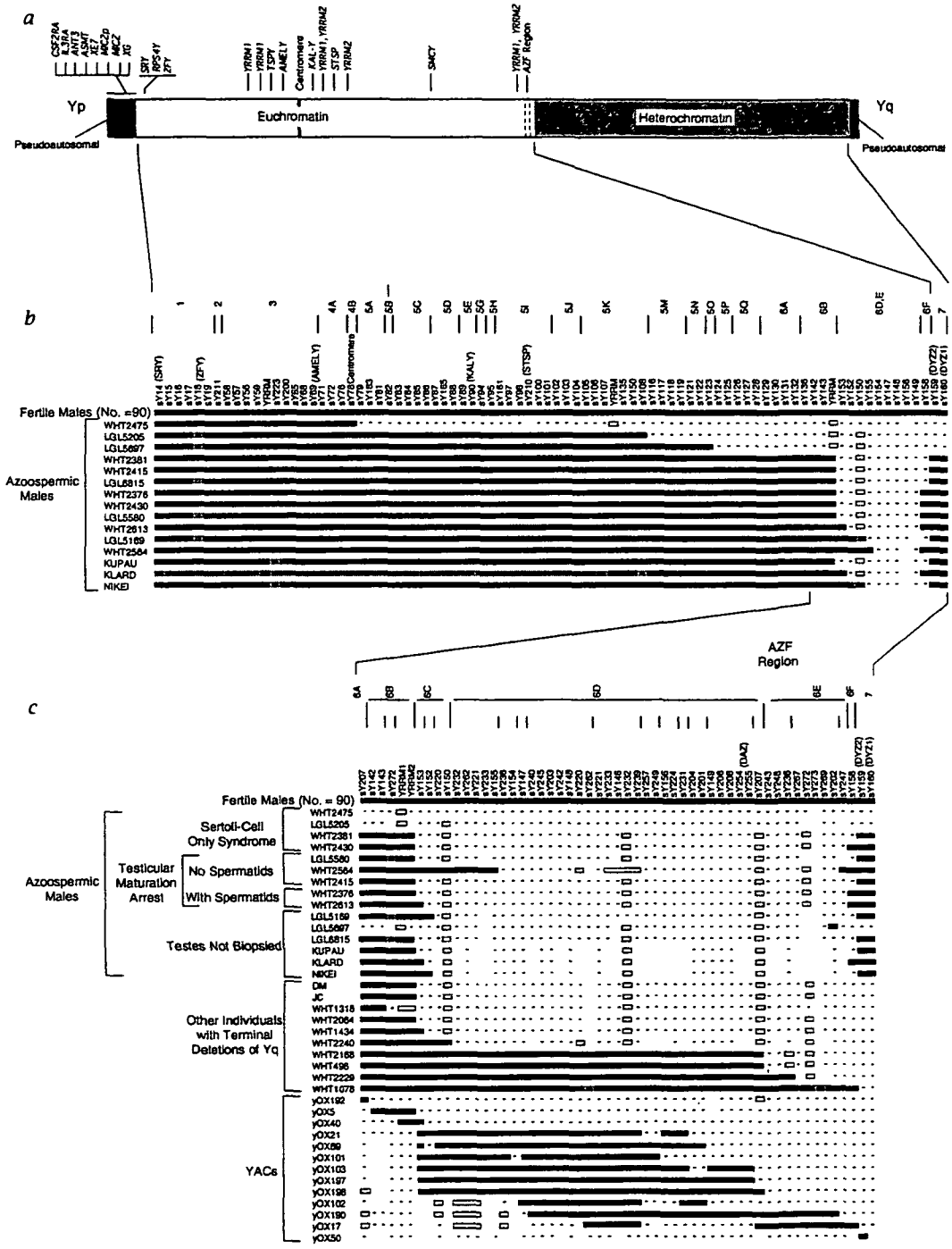


Figure 1.

*Taq*I. Oligonucleotide primers were selected so that nearly all PCR assays could be carried out under identical conditions (Vollrath *et al.*, 1992). *YRRM* primers were as described in Ma *et al.* (1993) and corrected in Kobayashi *et al.* (1994).

Human genomic DNAs were prepared from blood or lymphoblastoid cell lines (Page *et al.*, 1987). PCR was performed in v-bottomed, 96-well plates (MJ Research) in 20 μl volumes in 1.5 mM $MgCl_2$, 5 mM NH_4Cl, 10 mM Tris (pH 8.2), 50 mM KCl, 100 μM dNTPs, with 1 U *Taq* DNA polymerase, 100–200 ng 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 I, certain primer pairs were annealed at 62°C. Reaction products were stored at 4°C until they were loaded onto 2–4% agarose gels for analysis.

We also tested individual Y-derived YAC (Foote *et al.*, 1992) for STSs, in which case we employed 5–10 ng total yeast genomic DNA as template and an annealing temperature of 62°C.

YAC subtraction

The subtraction protocol of Rosenberg *et al.* (1994) was modified for use with YAC DNA. DNA from 66 overlapping YACs, spanning most of the Y chromosome's euchromatic region (Foote *et al.*, 1992), were separated from

Figure 1. Deletion mapping the *AZF* region on human Y chromosome. (**a**) Diagram of chromosomes euchromatic, heterochromatic and pseudoautosomal regions with the Yp telomere to the left and the Yq telomere to the right. Above are indicated the centromere, the *AZF* region and previously cloned genes and pseudogenes. (**b**) A low-resolution analysis of Y chromosomes of azoospermic men. Along the top border are listed deletion intervals 1–7 and, immediately below, 84 Y-chromosomal sequence-tagged sites (STS), all reported previously (Vollrath *et al.*, 1992; *YRRM*, Ma *et al.*, 1993, as corrected by Kobayashi *et al.*, 1994); gene, pseudogene and locus names are in parentheses. Shown below are the results of testing men for the presence (solid black box) or absence (–) of each STS. The first horizontal (solid black) line denotes the 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 the remaining 12 azoospermic men, as well as interstitial deletions observed in Chandley and Hargreave's patients KUPAU, KLARD and NIKEI. Blank spaces or grey boxes indicate, respectively, an inferred absence or presence (by interpolation) of markers for which the assay was not performed. White boxes represent positive results to be interpreted in the light of the Y-specific repeat nature of the sequences assayed; these positive results probably reflect the presence of closely related sequences elsewhere on the chromosome. (**c**) A higher resolution map of the *AZF* region. Along the 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 the results of testing for STSs in the following samples: 90 fertile men, 12 azoospermic men in whom deletions were detected (testicular histologies indicated to the 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 (Vollrath *et al.*, 1992) and 13 YAC (Foote *et al.*, 1992). A few of the infertile men were not tested for sY262, sY267, sY269, sY272 and sY273. STSs are ordered so as to minimize the number of apparent breakpoints in this set of patients and YACs. Some errors in the ordering are likely given the repeat-rich nature of the 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.* (1992) and Foote *et al.* (1992), though the WHT2168 breakpoint appears to be more proximal than reported previously. All available male relatives of azoospermic men with deletions (fathers of WHT2415, WHT2475, WHT2613, LGL5169 and LGL5697; brother of WHT2381; paternal uncle of WHT2376) were tested and found to carry all markers listed in (**b**) and (**c**). Male relatives of WHT2430, WHT2564, LGL5580, LGL5205 and LGL6815 were not available.

Table I. Y chromosomal sequence-tagged sites (STS)

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	TGGGAAAGCCTCAACTGCC	109
sY221	GTAAGCCCCAGATACCCTCC	AAATTGTTTGGAAAAGGACACC	113
sY224	ATAGTTAGTTTTGTGGTAACAT	CATAGCCTCTATGCAGATGGG	158
sY231	ATTGATGTGTTGCCCCAAAT	AGAGTGAACTTTAAATCCCAGCC	149
sY232	GACTCTACCACTTGGGCTCAATTT	AGATGTACCCAAGGCCACTG	91
sY233	AGTTAGTAAGCCCCAGTTATCCTCC	TTTGGAAAAGGACACCTTATTAGCCA	115
sY236	CCCCATCGGTAAACCAAATCA	CCATTGAAGTTTGAAGGTGTCA	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]	CTGGACAAAGCCTTGGAAA	CTGCATGTCAATTGTGGGAC	114
sY248[a]	CATTGGCATGAATGTGTATTC	CTCTGGGACAAGTGTTCCTT	94
sY249	GACAAAGGGCTGATGATTTA	CATCACCTTTACTTTTTAAATGG	114
sY254[b]	GGGTGTTACCAGAAGGCAAA	GAACCGTATCTACCAAAGCAGC	350
sY255[b]	GTTACAGGATTCGGCGTGAT	CTCGTCATGTGCAGCCAC	126
sY257	AGGTTGTTTGGCCTTGAGC	TCTATGATCTGTACCCGGTGC	123
sY262	AGCTCACTGCAAGCAACAGA	CCACCATCCCCCTTCTTC	100
sY267	GAATGTGTATTCAAGGACTTCTCG	TACTTCCTTCGGGGCCTCT	102
sY269	CTCTGGGACAAGTGTTCCTTG	CATGGCATGAATGTGTATTCA	94
sY272	GGTGAGTCAAATTAGTCAATGTCC	CCTTACCACAGGACAGAGGG	93
sY273	GGTCTTTAAAAGGTGAGTCAAATT	AGACAGAGGGAACTTCAAGACC	95

[a]Annealed at 62°C; otherwise polymerase chain reaction conditions as indicated in Materials and methods.
[b]Within *DAZ* gene.

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 and yOX198) blanketing the *AZF* region. 100 ng of this DNA were digested with *Sau*3A and ligated to a *Sau*3A-compatible PCR adapter (an equimolar mixture of GACACTCTCGAGACATCACCGTCC and phosphorylated GATCGGACGGTGATGTCTCGAGAGTG). '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 DNA were then amplified separately using oligonucleotides OL25 and OL31DB respectively as PCR primers (Rosenberg *et al.*, 1994). Subtractive hybridizations were carried out with the following in a total volume of 4 μl: 4 ng amplified tracer DNA; 7 μg

amplified, biotinylated YAC driver DNA; 3 μg amplified, biotinylated yeast genomic driver DNA; 20 μg yeast tRNA; 5 μg oligonucleotide OL30; and 2 μg 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 DNA from 130 subtraction clones that had been pooled, amplified and biotinylated as described above. The resulting subtraction product, in bulk, was radiolabelled and hybridized to high-density arrays of an 11 700 clone, Y-enriched cosmid library (LLOYNCO3; Human Genome Center, Lawrence Livermore National Laboratory) according to the procedure of Holland *et al.* (1993), resulting in the identification of 120 cosmid clones.

Exon trapping

Substrates for exon trapping (Duyk *et al.*, 1990) 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 of 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 be derived 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 previously (Fisher *et al.*, 1990). 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 previously (Fisher *et al.*, 1990). Nucleotide sequencing of *DAZ* cDNA clones was performed as described previously (Fisher *et al.*, 1990). Because the composite length of *DAZ* cDNA clones was considerably shorter than the 3.5 kb transcript observed on Northern blots (Figure 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 a starting template and the following two *DAZ* oligonucleotides as gene-specific primers: AACGAAACAAATCCATA-

GCCTTTG (for cDNA synthesis) and CTCGCTCGCCCAGAACCGTATCTACC-AAAGCA (for secondary amplification). The resulting PCR products (~500 bp) were cloned (TA cloning system; Invitrogen) and sequenced.

Results

Y chromosome deletions in azoospermic men

We studied 89 men in whom a 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 an absence of germ cells (Sertoli cell-only syndrome; $n = 42$) or a preponderance of premeiotic spermatogenic cells (testicular maturation arrest; $n = 36$). 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 origin.

The human Y chromosome is divided into euchromatic and heterochromatic halves (Figure 1a). The heterochromatin, comprising distal Yq, is dispensable with regard to male fertility (Borgaonkar and Hollander, 1971; Andersson *et al.*, 1988). We focused on the 30 Mb euchromatic region, which includes proximal Yq, the centromere and Yp, and for which a physical map of ordered STS and overlapping yeast artificial chromosome (YAC) clones has been constructed (Foote *et al.*, 1992; Vollrath *et al.*, 1992).

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 shown previously 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 >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 (Figure 1b). 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 (Figure 1c). 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 (Ma *et al.*, 1993). Using genomic DNA (kindly provided by A.Chandley and T.Hargreave), we determined that the STSs common to the 12 deletions we identified are also absent in KLARD, NIKEI and KUPAU (Figure 1b). These results provide a consistent definition of the *AZF* region, the deletion of which appears to account for ~13% (12/89) 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.)

Refined map of the AZF region

Unforeseen hazards can imperil efforts to map precisely phenotypes on the human Y chromosome. A 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 (e.g. the deletion of *ZFY* in female WHT1014; Page *et al.*, 1987) may also be deleted for a second, non-contiguous region (the deletion of *SRY* in WHT1014; Page *et al.*, 1990; Sinclair *et al.*, 1990). Such difficulties may be compounded in the vicinity of *AZF* because Y-specific repetitive DNA sequences comprise most of this portion of the chromosome (Foote *et al.*, 1992); 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 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 shown previously to have breakpoints in this region (Foote *et al.*, 1992; Vollrath *et al.*, 1992). The results allowed us to refine the physical map of the region deleted in the azoospermic men (Figure 1c). Overlapping YAC yOX198 and yOX17 (500 and 900 kb respectively) appear to span most of the mapped region, suggesting that it encompasses ~10^6 bp. The map incorporates 56 loci at an estimated average spacing of ~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 (Figure 1c). 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 ~5×10^5 bp.

Although the resolution of this physical map is limited, our findings in azoospermic men with Y chromosomal deletions suggest that their breakpoints may be clustered (Figure 1c). 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 (Figures 1c and 2). We were surprised to discover that the histological 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 (Figure 2d). There is no obvious correlation between the size of the Y deletion and the severity of the spermatogenic defect (Figure 1c).

Cosmid cloning and exon trapping

Although the *YRRM1* and *YRRM2* genes had been reported to be deleted in azoospermic males KLARD, NIKEI and KUPAU (Ma *et al.*, 1993), 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. Because 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 5- to 10-fold coverage of the area. Of these cosmids, 120 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') (Lamar and Palmer, 1984; Straus and Ausubel, 1990; Rosenberg *et al.*, 1994). 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 tracers; 58 YACs spanning the remainder of the euchromatic Y were used as drivers. This subtraction was intended to yield a pool of *AZF*

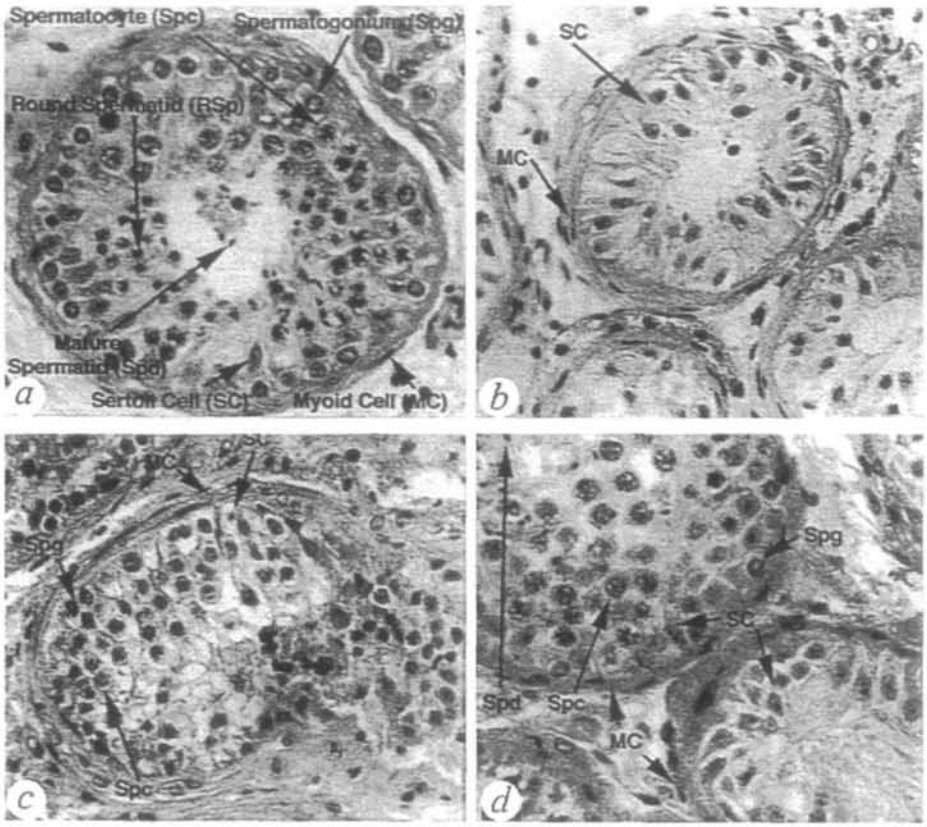


Figure 2. Testicular histologies associated with *AZF* deletions. (**a**) A photomicrograph of a normal seminiferous tubule (in cross-section) from a fertile human male. The tubule is ringed by myoid cells and contains somatic (Sertoli) cells and the following germ cells: spermatogonia. spermatocytes,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

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 (Figure 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 pre-annealed 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

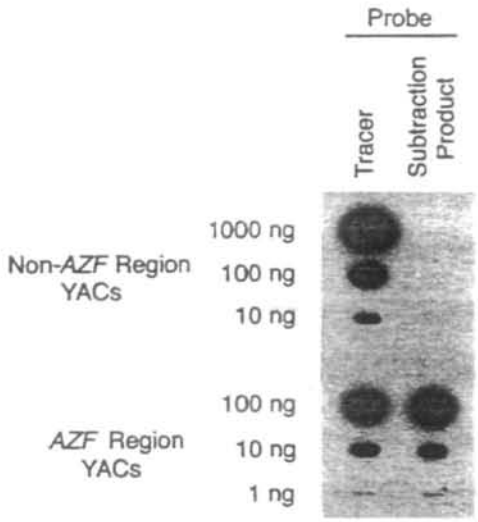


Figure 3. *AZF* region specificity of the YAC subtraction product. The autoradiogram was produced by hybridizing ^{32}P-labelled tracer DNA or subtraction product to membrane-bound DNA from pooled YACs — 58 YAC from outside the *AZF* region or eight YACs from the *AZF* region and its immediate environs. The indicated quantities of pooled YAC DNA were spotted onto nylon membrane; 10-fold greater quantities were used for the non-*AZF* region YAC to compensate for the higher complexity of this YAC pool. Hybridization: 20 h at 65°C in 5× SSC (1× = 0.15 M NaCl, 15 mM Na-citrate, pH 7.4), 5× Denhardt's solution (1× = 0.02% Ficoll 400, 0.02% polyvinylpyrrolidone, 0.02% bovine serum albumin), 1% sodium dodecyl sulphate (SDS) and 50 $\mu g ml^{-1}$ salmon sperm DNA. Wash: three times for 15 min each at 65°C in 0.1× SSC, 0.1% SDS.

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 the trapping products revealed 16 potential exons.

No evidence of YRRM sequences in the AZF region

As judged by a nucleotide sequence analysis, none of the potential exons recovered from the *AZF* region or its immediate environs was derived from the *YRRM* genes, which have been proposed as *AZF* candidates (Ma *et al.*, 1993). 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 reported previously (Ma *et al.*, 1993; Kobayashi *et al.*, 1994), *YRRM* sequences are absent from YACs spanning the *AZF* region (YAC 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,

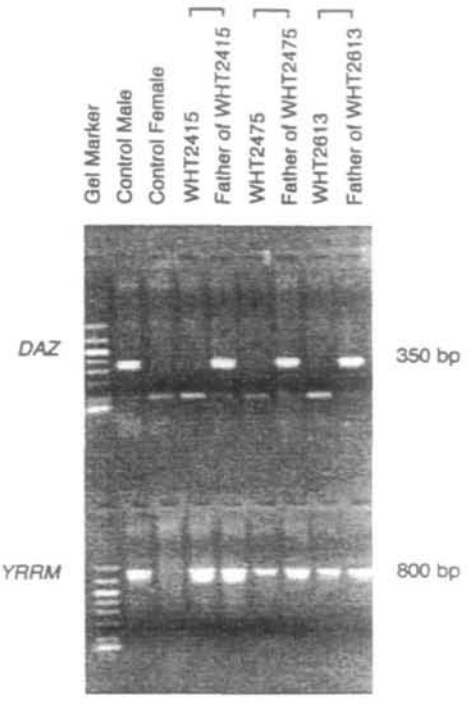


Figure 4. Polymerase chain reaction testing of genomic DNA from three azoospermic men and their fathers for *DAZ* and *YRRM*. Top: sY254, a sequence-tagged site within *DAZ*. Bottom: *YRRM1* (Ma *et al.*, 1993, as corrected by Kobayashi *et al.*, 1994). Products were separated by electrophoresis in 2% agarose gel and visualized by ethidium bromide staining.

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 (Figure 4). Indeed, we found both *YRRM1* and *YRRM2* to be present in patients KUPAU, NIKEI and KLARD. We conclude that *YRRM1* sequences are dispersed to several locations on the human Y chromosome, but we find no evidence of *YRRM* sequences in the *AZF* region.

An AZF-region gene transcribed in testes

An 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 the *AZF* region were hybridized to Southern blots of restriction-digested human male and female genomic DNA 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

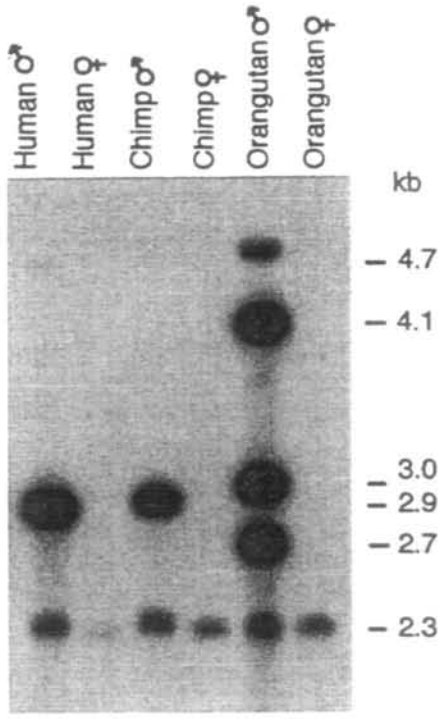


Figure 5. A single copy of the *DAZ* gene on human and chimpanzee Y chromosomes. The autoradiogram was produced by hybridizing *DAZ* exon 325.7 to a Southern blot of *Eco*RI-digested genomic DNA. The sizes (in kb) of hybridizing fragments are indicated to the right. Hybridization: 20 h at 42°C in 50%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

and hybridized to clones in the cDNA library. These four products recognized overlapping sets of cDNA clones, and a subsequent analysis confirmed that all four were derived from a single transcription unit. Approximately 1 in 5000 clones in the testis cDNA library derives from this gene, which we will refer to as *DAZ* (deleted in azoospermia).

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 (Figure 4). *DAZ* was also absent in KUPAU and NIKEI. As DNA was limited, 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 DNA from either species, exon 325.7 detected a single male-specific fragment, corresponding to *DAZ* (Figure 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.

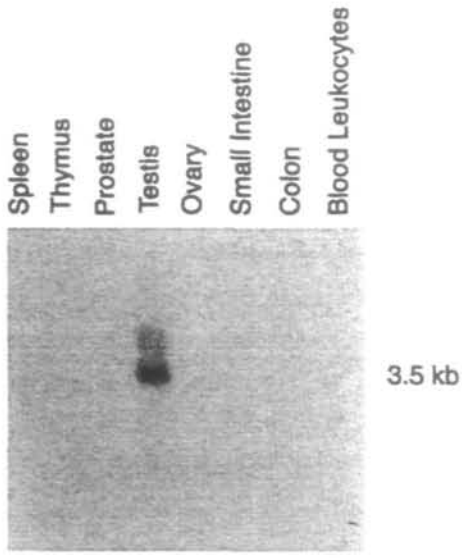


Figure 6. Transcription of *DAZ* gene in human adult testis. The autoradiogram was produced by hybridizing *DAZ* exon 325.7 to a Northern blot of poly(A)$^+$ RNA (2 μg/lane) from human tissues (Clontech). Additional negative results were obtained with RNA from heart, brain, placenta, lung, liver, skeletal muscle, kidney and pancreas (data not shown). Hybridization was at 47°C, otherwise as in the legend to Figure 5.

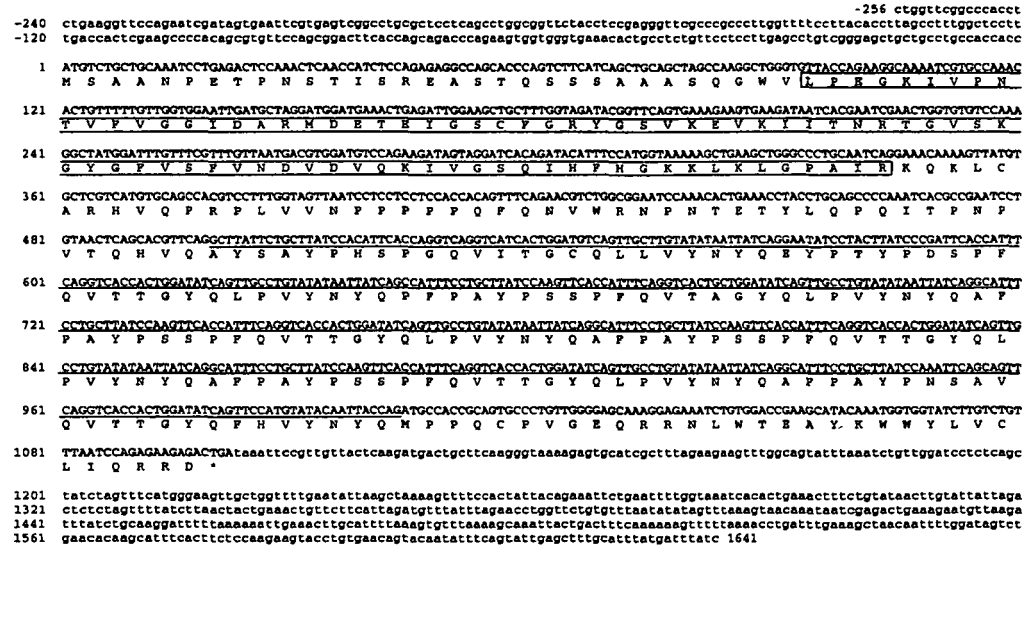
The results of screening cDNA libraries indicated that the human *DAZ* gene is transcribed in the adult testis. To confirm this result and to assess whether the gene is transcribed elsewhere, we hybridized exon 325.7 to Northern blots of poly(A)-selected RNA from 16 different adult human organs. We observed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

A putative RNA-binding protein

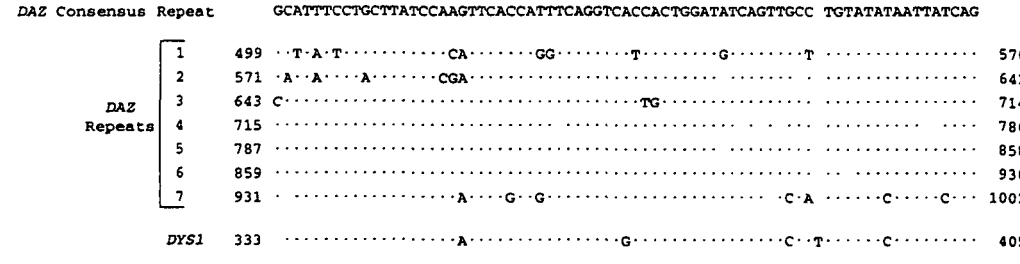
A nucleotide sequence analysis of *DAZ* cDNA clones isolated from the human adult testis library revealed a single long open reading frame (Figure 7a). The first ATG in this open reading frame (position 1 in Figure 7a) occurs in a sequence context that is favourable for the initiation of translation (Kozak, 1986). Beginning at this ATG, the transcript appears to encode a protein of 366 amino acids, with a predicted molecular weight of 41 257. Although 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 (Figure 7a and b). The repeats differ from each other by at most a few nucleotides, suggesting that they 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 (Figure 7b) (Lucotte and Ngo, 1985). To

a



b



c

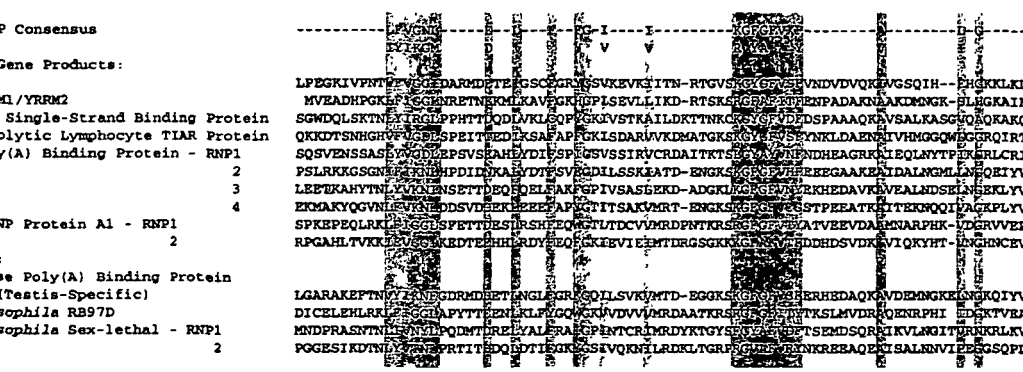


Figure 7.

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 C-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 (Figure 7c). Similar RNP/RRM domains are found, for example, in the mammalian polyadenylate binding proteins 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 (Kenan *et al.*, 1991; Burd and Dreyfuss, 1994). We conclude that the DAZ protein probably functions by binding RNA, or possibly single-stranded DNA.

Discussion

Frequent de-novo deletion of an AZF

We examined the Y chromosomes of men with spermatogenic defects, so severe as to result in the absence of spermatozoa 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 is likely to harbour one or more genes required for 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 (Figure 1). These results do not exclude the existence of genes essential for spermatogenesis elsewhere on the human Y

Figure 7. The *DAZ* cDNA sequence and the predicted amino acid sequence of encoded protein. (**a**) The nucleotide sequence is a composite of (i) a cDNA insert of plasmid pDP1577 and (ii) a 5′ RACE product obtained using adult human testis RNA as a template. (5′ RACE products overlapped the insert of pDP1577 by 470 nucleotides and extended 143 nucleotides further 5′.) The composite cDNA sequence is incomplete at the 3′ end, which may account for it being smaller than the 3.5 kb transcript observed by Northern blotting (Figure 6). The predicted 366 amino acid sequence is immediately beneath the nucleotide sequence; the RNP/RRM domain (Figure 1c) is boxed. Seven tandem repeats of 72 nucleotide units (Figure 1b) are underlined. The numbering of nucleotides and amino acids begins with the first in-frame AUG codon. GenBank accession number U21663. (**b**) Tandem repeats within the *DAZ* coding sequence. At the top is the consensus nucleotide sequence of 72 bp *DAZ* repeats. Below are the seven *DAZ* repeats (numbering of the nucleotides as in Figure 1a) and the portion of the nucleotide sequence of plasmid p49a (*DYS1*; Lucotte and Ngo, 1985). Dots represent identity to the *DAZ* repeat consensus. Apart from a single nucleotide insertion, the portion of *DYS1* shown is colinear with the *DAZ* repeats. (**c**) Amino acid sequences of the RNP/RRM domains in *DAZ* and other proteins. At the top is the consensus sequence (Burd and Dreyfuss, 1994); dashes indicate the positions where no consensus is apparent. The regions most highly conserved are shaded. The list of other proteins is representative but not exhaustive.

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; alternatively they may result in phenotypes less severe than azoospermia. Vogt *et al.* (1992) 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 1 in 1000 men is azoospermic because of severe spermatogenic defects (Hull *et al.*, 1985). As described here, *AZF* is absent in ~1 in 8 such men, although present in their fathers. Thus, it appears that at least 1 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 (Yen *et al.*, 1990; Lefebvre *et al.*, 1995; Roy *et al.*, 1995). A similar mechanism may be operating on the Y chromosome to produce deletions of *AZF*. This hypothesis is attractive because 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 (Figure 1c). It remains to be seen whether these apparent deletion hot spots 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 histological 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 the occasional production of mature condensed spermatids (Figure 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 (Figure 1c). 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, histological 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 (Figure 2d). Because 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 the 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 the 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, although 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 (Lindsley and Tokuyasu, 1980; Magram and Bishop, 1991). It seems unlikely that *AZF* functions in the migration of primordial germ cells to the gonad, because this process occurs even in the Y chromosome's absence (e.g. in XX or XO embryos) (Carr *et al.*, 1968). Given the testicular histologies observed in *AZF*-deleted men, it is conceivable that *AZF* facilitates the 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 (i.e. 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, because the presence of distal Yq (including the entirety of the *AZF* region) has been reported previously in several chromosomally aberrant, but nonetheless fertile, healthy women (Vollrath *et al.*, 1994).

DAZ and YRRM

Our mapping studies indicate that the *YRRM* genes are unlikely candidates for *AZF*. Ma *et al.* (1993) reported de-novo deletions of one or more *YRRM* genes in several azoospermic males, including KLARD, NIKEI and KUPAU. On this basis they proposed the *YRRM* genes as *AZF* candidates. However, we find no evidence of *YRRM* sequences in the *AZF* region despite the disposal of such sequences to other locations on Yp and Yq. Of course, we cannot exclude the formal possibilities that (i) *YRRM* genes retained on deleted Y chromosomes are transcriptionally silenced by position effects or (ii) 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* 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 ~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*.

Although *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 (Figure 7c) (Ma *et al.*, 1993). (Outside this domain, the proteins exhibit little sequence similarity.) By analogy to well characterized proteins containing such domains (Kenan *et al.*, 1991; Burd and Dreyfuss, 1994), 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 sizeable Y-specific gene family in humans and gorillas (Ma *et al.*, 1993), while *DAZ*, although single-copy in humans and chimpanzees, may have been amplified to form a Y-specific family in orangutan (Figure 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 summation, 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 the degeneration of early spermatogenic cells and azoospermia (Karsch-Mizrachi and Haynes, 1993). 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 (Kwon *et al.*, 1993), and other genes expressed during spermatogenesis may also be post-transcriptionally regulated (Hecht, 1993). 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 (Butler *et al.*, 1993; Siomi *et al.*, 1993). Perhaps RNA-binding proteins and post-transcriptional mechanisms figure prominently in the regulation of male germ cell development in mammals.

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 DNA from patients KLARD, NIKEI and KUPAU; A.McMurray and J.Segre for advice on exon trapping; the 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. This work was supported by National Institutes of Health, US Department of Agriculture, Academy of Finland, Sigrid Juselius Foundation and the Finnish Cultural Foundation. R.R. was the recipient of a Damon-Runyon/Walter Winchell fellowship.

References

Andersson, M. *et al.* (1988) Y:autosome translocations and mosaicism in the aetiology of 45,X maleness: assignment of fertility factor to distal Yq11. *Hum. Genet.*, **79**, 2–7.

Borgaonkar, D.S. and Hollander, D.H. (1971) Quinacrine fluorescence of the human Y chromosome. *Nature*, **230**, 52.

Burd, C.G. and Dreyfuss, G. (1994) Conserved structures and diversity of functions of RNA-binding proteins. *Science*, **265**, 615–621.

Butler, M.G. *et al.* (1991) Anthropometric comparison of mentally retarded males with and without the fragile X syndrome. *Am. J. Med. Genet.*, **38**, 260–268.

Carr, D.H., Haggar, R.A.S. and Hart, A.G. (1968) Germ cells in the ovaries of XO female infants. *Am. J. Clin. Pathol.*, **49**, 521–526.

Clermont, Y. (1996) Renewal of spermatogonia in man. *Am. J. Anat.*, **118**, 509–524.

Duyk, G.M., Kim, S., Meyers, R.M. and Cox, D.R. (1990) Exon trapping: a genetic screen to identify candidate transcribed sequences in cloned mammalian genomic DNA. *Proc. Natl. Acad. Sci. USA*, **87**, 8995–8999.

Dym, M. (1994) Spermatogonial stem cells of the testis. *Proc. Natl. Acad. Sci. USA*, **91,** 11287–11289.

Fisher, E.M.C. *et al.* (1990) Homologous ribosomal protein genes on the human X and Y chromosomes: escape from X inactivation and possible implications for Turner syndrome. *Cell*, **63,** 1205–1208.

Fitch, N., Richer, C.-L., Pinsky, L. and Kahn, A. (1985) Deletion of the long arm of the Y chromosome and review of Y chromosome abnormalities. *Am. J. Med. Genet.*, **20,** 31–42.

Foote, S., Vollrath, D., Hilton, A. and Page, D.C. (1992) The human Y chromosome: overlapping DNA clones spanning the euchromatic region. *Science*, **258,** 60–66.

Haldi, M. *et al.* (1995) Large human YACs constructed in a *rad52* strain show a reduced rate of chimerism. *Genomics*, **24,** 478–484.

Hartung, M., Devictor, M., Codaccioni, J.L. and Stahl, A. (1988) Yq deletion and failure of spermatogenesis. *Ann. Genet.*, **31,** 21–26.

Hecht, N.B. (1993) In Desjardins, C. and Eing, L.L. (eds), *Cell and Molecular Biology of the Testis.* Oxford University Press, New York, NY, USA, pp. 400–432.

Holland, J., Coffey, A.J., Giannelli, F. and Bentley, D.R. (1993) Vertical integration of cosmid and YAC resources for interval mapping on the X-chromosome. *Genomics*, **15,** 297–304.

Hull, M.G.R. *et al.* (1985) Population study of causes, treatment, and outcome of infertility. *Br. Med. J.*, **291,** 1693–1697.

Johnson, M.D., Tho, S.P.T., Behzadian, A. and McDonough, P.G. (1989) Molecular scanning of Yq11 (interval 6) in men with Sertoli cell-only syndrome. *Am. J. Obstet. Gynecol.*, **161,** 1732–1737.

Karsch-Mizrachi, I. and Haynes, S.R. (1993) The *Rb97D* gene encodes a potential RNA-binding protein required for spermatogenesis in *Drosophila. Nucleic Acids Res.*, **21,** 2229–2235.

Kenan, D.J., Query, C.C. and Keene, J.D. (1991) RNA recognition: towards identifying determinants of specificity. *Trends Biochem.*, **16,** 214–220.

Kobayashi, K. *et al.* (1994) PCR analysis of the Y chromosome long arm in azoospermic patients: evidence for a second locus required for spermatogenesis. *Hum. Mol. Genet.*, **3,** 1965–1967.

Kozak, M. (1986) Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. *Cell*, **44,** 283–292.

Kwon, Y.K., Murray, M.T. and Hecht, N.B. (1993) 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.

Lamar, E.E. and Palmer, E. (1984) Y-encoded species-specific DNA in mice: evidence that the Y chromosome exists in two polymorphic forms in inbred strains. *Cell*, **37,** 171–177.

Lefebvre, S. *et al.* (1995) Identification and characterization of a spinal muscular atrophy-determining gene. *Cell*, **80,** 155–165.

Lindsley, D. and Tokuyasu, K.T. (1980) In Ashburner, M. and Wright, T.R.F. (eds), *The Genetics and Biology of Drosophila.* Academic Press, London, UK, pp. 226–294.

Lucotte, G. and Ngo, Y.Y. (1985) p491, a highly polymorphic probe, that detects *Taq*1 RFLPs on the human Y chromosome. *Nucleic Acids Res.*, **13,** 8285.

Ma, K. *et al.* (1992) Towards the molecular localisation of the AZF locus: mapping of microdeletions in azoospermic men within 14 subintervals of interval 6 of the human Y chromosome. *Hum. Mol. Genet.*, **1,** 29–33.

Ma, K. *et al.* (1993) A Y chromosome gene family with RNA-binding protein homology: candidates for the azoospermia factor AZF controlling human spermatogenesis. *Cell*, **75,** 1287–1295.

Magram, J. and Bishop, J.M. (1991) Dominant male sterility in mice caused by insertion of a transgene. *Proc. Natl. Acad. Sci. USA*, **88,** 10327–10331.

Page, D.C. *et al.* (1987) The sex-determining region of the human Y chromosome encodes a finger protein. *Cell*, **51,** 1091–1104.

Page, D.C., Fisher, E.M.C., McGillivray, B. and Brown, L.G. (1990) Additional deletion in sex-determining region of human Y chromosome resolves paradox of X,t(Y;22) female. *Nature*, **346,** 279–281.

Rosenberg, M., Przybylska, M. and Straus, D. (1994) RFLP subtraction: a method for making libraries of polymorphic markers. *Proc. Natl. Acad. Sci. USA*, **91,** 6113–6117.

Roy, N. *et al.* (1995) The gene for neuronal apoptosis inhibitory protein is partially deleted in individuals with spinal muscular atrophy. *Cell*, **80,** 167–178.

Silber, S.J. (1989) The relationship of abnormal semen parameters to male fertility. *Hum. Reprod.*, **4,** 947–953.

Silber, S.J. (1995) Sertoli cell-only revisited. *Hum. Reprod.*, **10,** 1031–1032.

Sinclair, A.H. *et al.* (1990) A gene from the human sex-determining region encodes a protein with homology to a conserved DNA-binding motif. *Nature*, **346,** 240–244.

Siomi, H., Siomi, M.C., Nussbaum, R.L. and Dreyfuss, G. (1993) The protein product of the fragile X gene, FMR1, has characteristics of an RNA-binding protein. *Cell*, **74,** 291–298.

Skare, J. *et al.* (1990) Interstitial deletion involving most of Yq. *Am. J. Med. Genet.*, **36,** 394–397.

Straus, D. and Ausubel, F.M. (1990) Genomic subtraction for cloning DNA corresponding to deletion mutations. *Proc. Natl. Acad. Sci. USA*, **87,** 1889–1893.

Tiepolo, L. and Zuffardi, O. (1976) Localization of factors controlling spermatogenesis in the nonfluorescent portion of the human Y chromosome long arm. *Hum. Genet.*, **34,** 119–124.

Vogt, P. *et al.* (1992) 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.

Vollrath, D. *et al.* (1992) The human Y chromosome: a 43-interval map based on naturally occurring deletions. *Science*, **258,** 52–59.

Yen, P.H. *et al.* (1990) Frequent deletions of the human X chromosome distal short arm result from recombination between low copy repetitive elements. *Cell*, **61,** 603–610.

Discussion

Edwards: Could you explain to us why only in some cases is the deletion present in spermatozoa? In most patients, the deletion causes Sertoli-cell-only syndrome or maturation arrest but occasionally it is present in spermatozoa. Do you have any explanation why it can persist in the spermatozoa of some men?

Page: I guess the question is: why do deletions of this region of the Y chromosome, independent whether DAZ is the right factor, result in such a wide range of spermatogenetic defects? I think that this is not in fact an unusual observation in genetics, to see a wide range of manifestations associated with what appears to be the same mutation. It could be that there are facts of genetic background; it could be that there are other genes in the genome that modify the consequences of being deleted for this gene. That is one example. I suspect that there may be micro-environmental variables that we do not yet have a good feel for. There was one case where Silber did the testicular histology i.e., one of the cases where a deletion was present. You could see on the testis biopsy a Sertoli-cell-only tubule immediately adjacent to a tubule that had occasional condensed spermatids. So this individual had a Y-deletion, and right next door there were tubules with two very different pictures. I think we just have to think of these disorders to the extent that they are linked to Y-deletions as representing a continuum and we should not be wedded to previous classifications.

Edwards: The deletion could have arisen between fertilization and some time before spermatogenesis began, although this is not necessarily true in all cases. The deletion could also arise somewhere in the primordial germ cells, spermatogonia, spermatocytes, or spermatids. If it did arise in the spermatic cell line, it could occur at any stage of development in the germinal cell line. In the maturation arrest cases you describe, the deletion could thus occur in the primordial germ cells, in which case primordial germ cells might never form spermatogonia. In the cases where spermatogonia were present and maturation arrest occurred, the deletion could arise in the early spermatogonia or sometimes even occurs in spermatids. The latter would be too late to stop the spermatozoa forming.

I believe that in some cases of the amplified triple repeats in fragile X, that some amplifications occur in the primordial germ cells and others may arise in the testis. Am I right on that?

Page: I think you are right.

Edwards: Do you therefore believe that some sort of similar genetic activity recurs in primordial germ cells and in spermatogonia, i.e. that the same effect can arise at different stages. Would that explain the results you have?

Page: Let us think about where we are doing the DNA testing. The DNA testing is primarily being done on blood DNA from the infertile male. Then if we find a deletion in blood DNA, we check the father's blood DNA and we find that he has an intact Y. Somewhere in the link that connects the father's blood DNA, we have to find the way back to his own zygote and then connect up to the son's blood DNA.

I think your question suggests that the mutation must have occurred after the fertilization that gave rise to the infertile male. I do not think we can make that assumption; I do not think it is clear whether these deletions on the Y occurred in the father in a pre-zygotic mutation, prior to fertilization or whether it occurred actually after the zygote that would go on to make the infertile male. I do not think that the actual situation is all clear.

Edwards: You could obtain these data. You have some patients with spermatogonia and therefore you could amplify the spermatogonia present in a testis biopsy to find out if the deletion was present. You would thus gain the data that you need, and since so many testis biopsies are carried out now you could discover whether in maturation arrest the deletion is present in some spermatogonia. Otherwise, you have a difficult situation to explain the interaction between various testicular compartments, e.g. between supporting cells or germinal cells. At present, you must postulate a variable expression of the gene at different stages in the history of primordial germ cells or of the sperm cell. We need evidence of the existence of the deletion at different stages in testis formation and differentiation.

Page: Essentially, what you are postulating is that the deletions might arise post-zygotically. They would be somatic mutations, although in this stage they might actually be occurring in the germ cells, and postulating then that the infertile individual is actually a mosaic for the Y-deletion. That would be, I agree, a very tidy explanation. I think this is one among the explanations that we should entertain, but I by no means think that this is the only explanation. There are plenty of other examples of the same mutations giving variable manifestations in different individuals.

Edwards: But you can prove this situation very quickly by taking your maturation arrest cases. Take spermatogonia or the spermatids in these cases, or even spermatocytes, to find out if a deletion is present.

Page: In my heart of hearts I suspect that the answer will be both: yes and no. I suspect that there will be some pre-fertilization mutations and some post-fertilization mutations. The other thing to realize is that in fact a lot of these mutations arise after fertilization. There may be men who are infertile because of the deletion of this gene, yet we do not even detect them, they test out as having an intact Y in blood DNA. So we may be underestimating the fraction of men who have deletions in the Y chromosome which are responsible for infertility.

Edwards: Could you tell us then about the fragile X syndrome? The same sort of model might exist there. In one case there is an amplification of triplets and in the other case these are deletion sequences. I wonder if, in fragile X, the amplification takes place in the primordial germ cells and then they are passed on to the testis. That model could be the exact model for your situation and explain how genetic factors cause deletion or repeat amplification in those early stages.

Page: The situation is a little bit different in that here we are deleting material. It is completely gone. In the case of fragile X, something is being amplified. In fragile X, the focus is on what is happening in the father of the individuals who

will themselves be affected. In our work, we are trying to explain the clinical features of afflicted individuals. It may be that we may have to focus on the fathers; it may actually be—with regard to understanding the basic biology—that it is just as interesting to focus on the fathers of these infertile male individuals.

Edwards: Have you looked at the spermatozoa of the fathers?

Page: Not yet. We will do that.

Silber: How precise is your mapping methodology and what is the possibility of picking up these 12 out of 89 men with deletions? The proportion of men you have identified may represent a small fraction of the total. Until you have more precise sign posts, is it possible that we are underestimating the number of azoospermic men with maturation arrest or Sertoli-cell-only syndrome that have defects in this region?

Page: Basically we see that about 10 to 15% of men with non-obstructive azoospermia are missing about 1% of the euchromatic or functional portions of the Y-chromosome. Even though this is far below the level of microscopic detection, it is what in the USA we would say is hitting the side of a barn, using DNA technology. This is a big target, using DNA probes, although it is far below the level of microscopic detection. It is possible that there are even smaller deletions within this critical region that we are failing to observe using our present DNA markers.

There is another possibility that could lead us to underestimate the frequency of affected genes. If DAZ is the critical gene, then there could be infertile men who have very subtle mutations within the gene which we would yet have to detect, e.g. basepair substitutions. So there are lots of reasons to think that the frequency of Y-chromosome mutations accounting for infertility could be higher. I also would not want to claim that it has to be a lot higher; we just do not know.

Silber: How many basepairs constitute this region and in between each one of these markers?

Page: The Y-chromosome is a very tiny chromosome; its euchromatic region is only 30,000,000 nucleotides. This region is only about half a million nucleotides, so that gives us half-a-million nucleotides to check out one by one in each of these individuals. Our DNA markers are spaced about every thirty-thousand nucleotides, so deletions that are smaller than thirty-thousand nucleotides could easily be missed. That leaves a lot of latitude.

Simpson: In the fragile X situation the molecular pathogenesis is that if there are too many repeats, hypermethylation occurs; therefore the message is not transcribed and translated. So, in essence, deletions would be comparable in that in the fragile X situation, the protein is not made. It would not matter whether it was deleted as in the other mechanism.

Even aside from the likelihood that there must be a lot of point mutations, this would not be evident from the techniques that you have shown. Are you not surprised at the extraordinary high mutation rate that must occur in this gene? I mean, one in 5000 is a very high frequency even if there are several different genes. This is orders of magnitude higher than what we would expect for most

somatic mutations, yet we are dealing with something that is pretty important for the species.

Page: At the last minute, I tried not to make my lecture too long and omitted a few slides that addressed this point. This is an extraordinarily high mutation rate, i.e. this one in 5000. We are not talking about the allele frequency in the population, I think we often get fixated on questions like: what is the allele frequency for CF mutant alleles? We are not talking here about stable inherited alleles whose frequency in the population does not change. These mutations effectively have an allele frequency of zero in the population. They are not present to the next generation; every time you see one of these mutations—there may be one or two exceptions—it represents a new lightning strike. The lightning is striking at a frequency of one in 5000 at least.

There are few other instances of genes that have comparable mutation rates. The Duchenne muscular dystrophy gene suffers a comparable rate of mutations. It turns out that the spinal muscular atrophy gene undergoes mutations at a comparable rate, but we can basically count on less than one hand the number of examples of genes that have mutation rates approaching this. I believe that the Duchenne muscular dystrophy gene it is the largest gene in the genome, so it is a mapping target. I believe this gene has some aspects similar to the spinal muscular atrophy gene.

It also turns out that the Y chromosome contains many chromosome-specific repetitive sequences. This makes it a mappers' nightmare, trying to map portions of the Y chromosome. It is like walking into a house of mirrors, walking out and being given a piece of paper and being asked to sketch out the floor plan. Everywhere you turn, on many parts of the Y chromosome the situation looks the same and this is also true in the worst part of the Y chromosome. This part of the Y chromosome contains virtually nothing but Y-specific repetitive sequences, and I think these sequences are the seeds of the Y's own destruction in this case. This is no way to run a chromosome. It is just absurd to have something that is required for fertility, located in such a precarious position. Is there some reason for it being located in this bad neighbourhood? I do not know. But I know that the repetitive sequences in this vicinity make this a very unstable neighbourhood.

Edwards: The fact that there is such high mutation rate would make one very suspicious about mutations being involved. Is that a fair conclusion? If you get a high mutation rate, you immediately start to think of the mechanism that might be involved. Now in this particular case, since the father produces spermatozoa, the deletion must arise very late in his gametogenesis if it occurs there. So that could occur during the spermatocyte stage and if it occurs in the spermatocyte stage, it would be a meiotic factor. Could you explain if it could be some chromosomal inversion or cross-over or some other mechanism which would give such a high rate of anomaly? That would be more sensible, would it not?

Page: I think one of the simplest ways to think why some of these deletions might occur is that deletion might occur before or after the fertilization event that gives rise to the infertile male. That is the first big discussion. Of course, if

it occurs after the fertilization it must be a meiotic event. If before fertilization, it could be either a mitotic or meiotic event.

Let us say, we take two Y-chromosomes in either mitosis or meiosis. At one stage or another there are two copies of the Y-chromosome and I imagine that they can become a little bit misaligned. A cross-over event might occur between repetitive sequences which has the effect of simply creating a Y that has chopped out a little intervening segment. I must admit that it is not obvious to me whether we should be thinking about the deletions occurring in mitosis or occurring in meiosis.