Am. J. Hum. Genet. 71:906–922, 2002

Recombination between Palindromes P5 and P1 on the Human Y Chromosome Causes Massive Deletions and Spermatogenic Failure

Sjoerd Repping,[1,2] Helen Skaletsky,[1] Julian Lange,[1] Sherman Silber,[3] Fulco van der Veen,[2] Robert D. Oates,[4] David C. Page,[1] and Steve Rozen[1]

[1]Howard Hughes Medical Institute, Whitehead Institute, and Department of Biology, Massachusetts Institute of Technology, Cambridge; [2]Center for Reproductive Medicine, Department of Obstetrics and Gynecology, Academic Medical Center, Amsterdam; [3]Infertility Center of St. Louis, St. Luke's Hospital, St. Louis; and [4]Department of Urology, Boston University Medical Center, Boston

It is widely believed that at least three nonoverlapping regions of the human Y chromosome—*AZFa*, *AZFb*, and *AZFc* ("azoospermia factors" a, b, and c)—are essential for normal spermatogenesis. These intervals are defined by interstitial Y-chromosome deletions that impair or extinguish spermatogenesis. Deletion breakpoints, mechanisms, and lengths, as well as inventories of affected genes, have been elucidated for deletions of *AZFa* and of *AZFc* but not for deletions of *AZFb* or of *AZFb* plus *AZFc*. We studied three deletions of *AZFb* and eight deletions of *AZFb* plus *AZFc*, as assayed by the STSs defining these intervals. Guided by Y-chromosome sequence, we localized breakpoints precisely and were able to sequence nine of the deletion junctions. Homologous recombination can explain seven of these deletions but not the remaining two. This fact and our discovery of breakpoint hotspots suggest that factors in addition to homology underlie these deletions. The deletions previously thought to define *AZFb* were found to extend from palindrome P5 to the proximal arm of palindrome P1, 1.5 Mb within *AZFc*. Thus, they do not define a genomic region separate from *AZFc*. We also found that the deletions of *AZFb* plus *AZFc*, as assayed by standard STSs heretofore available, in fact extend from P5 to the distal arm of P1 and spare distal *AZFc*. Both classes of deletions are massive: P5/proximal-P1 deletions encompass up to 6.2 Mb and remove 32 genes and transcripts; P5/distal-P1 deletions encompass up to 7.7 Mb and remove 42 genes and transcripts. To our knowledge, these are the largest of all human interstitial deletions for which deletion junctions and complete intervening sequence are available. The restriction of the associated phenotype to spermatogenic failure indicates the remarkable functional specialization of the affected regions of the Y chromosome.

Introduction

Approximately 15% of couples face difficulty conceiving a child, and inadequate or absent sperm production is a primary factor in a substantial fraction of these cases (MacLeod and Gold 1951; Zuckerman et al. 1977; David et al. 1979; MacLeod and Wang 1979; Hull et al. 1985). Various terminal and interstitial deletions of the Y-chromosome long arm constitute the most common molecularly identified causes of such spermatogenic failure (Tiepolo and Zuffardi 1976; Ma et al. 1992; Reijo et al. 1995; Vogt et al. 1996; Simoni et al. 1997; Yen 1998). Study of these deletions has led to the view that there are three nonoverlapping Y-chromosome intervals necessary for normal spermatogenesis: *AZFa*, *AZFb*, and *AZFc* ("azoospermia factors" a, b, and c) (Vogt et al. 1996) (MIM 400024) (fig. 7).

Received May 30, 2002; accepted for publication July 10, 2002; electronically published September 20, 2002.

Address for correspondence and reprints: Dr. Steve Rozen, Whitehead Institute, 9 Cambridge Center, Cambridge MA 02142. E-mail: rozen@wi.mit.edu

0002-9297/2002/7104-0018$15.00

Recently, the breakpoints and lengths of *AZFa* and *AZFc* deletions were defined in the context of finished Y-chromosome sequence. These results revealed ectopic (nonallelic) homologous recombination between flanking, direct repeats to be the likely mechanism of both deletions (Blanco et al. 2000; Kamp et al. 2000; Sun et al. 2000; Kuroda-Kawaguchi et al. 2001). This sequence-level characterization of *AZFa* and *AZFc* deletions has led to a detailed understanding of the gene content of the affected regions, as well as an understanding that both regions contain more than one gene (or gene family) likely to play an important role in spermatogenesis.

Deletions of *AZFb* and of *AZFb* plus *AZFc*, which have been associated with azoospermia (no sperm observed in semen) in all studies to date, have not been characterized beyond plus/minus results for a handful of widely spaced STSs, many of which amplify several sites on the Y chromosome (Vogt et al. 1996; Girardi et al. 1997; Brandell et al. 1998; Ferlin et al. 1999; Krausz et al. 1999; Lin et al. 2000; Martinez et al. 2000; Van Landuyt et al. 2000; Friel et al. 2001; Maurer et al. 2001). Here we aim to characterize deletions of *AZFb* and of *AZFb* plus *AZFc* at the level of genomic detail achieved for *AZFa* and *AZFc*—namely, to determine

deletion breakpoints and lengths, to elucidate deletion mechanisms, and to provide a detailed understanding of the genes and transcripts affected by these deletions.

Patients, Material, and Methods

Patients

We studied 11 unrelated, azoospermic men who, on the basis of STSs available heretofore, were determined to have interstitial deletions involving *AZFb*. Specifically, three men had plus/minus STS results matching the original definition of *AZFb:* They possessed the flanking STSs sY105 and sY149 but lacked the *AZFb* STSs sY117, sY127, and sY143 (Vogt et al. 1996). Seven other men had plus/minus STS results that indicated deletion of *AZFb* plus *AZFc:* They possessed sY105 (proximal to *AZFb*) but lacked sY117, sY127, sY143, and the *AZFc* STS sY149; however, they possessed the STS sY1201, which lies just distal to *AZFc* (Kuroda-Kawaguchi et al. 2001). The remaining man had STS results identical to those of the previous seven except at sY117, which he possessed.

For 3 of these 11 men, a DNA sample from a patrilineal relative was available. All had intact Y chromosomes. The 11 men with deletions had been identified at the Whitehead Institute and at the Academic Medical Center during Y-chromosome deletion screening of 602 men with idiopathic, nonobstructive azoospermia, 393 men with severe oligozoospermia (sperm density below 5×10^6/ml at the Whitehead Institute or total sperm count $<20 \times 10^6$ at the Academic Medical Center), and 106 men with sperm density $>20 \times 10^6$/ml and normal sperm motility and morphology (World Health Organization 1992).

This study was approved by the institutional review boards of the Massachusetts Institute of Technology and the Academic Medical Center. Informed consent was obtained from all participants.

Plus/Minus STS Analysis

All STSs used in low-resolution breakpoint mapping (fig. 1*B*) and the associated PCR conditions have been deposited in GenBank (see the "Electronic-Database Information" section). All STSs used in high-resolution breakpoint mapping are shown in tables A4–A6. PCR conditions were as described elsewhere for sY1201 (see GenBank).

New PCR primers were designed using Primer3 (Rozen and Skaletsky 2000). Because of the many Y-chromosome repeated sequences, it is often difficult to design primer pairs that specifically amplify a single Y-chromosome site. Therefore, we tested potentially nonspecific primer pairs against two kinds of negative control templates: first, we confirmed that the primer pairs resulted in no product when amplifying DNA from BAC clones lacking the STS target site itself but containing sequence similar to it; second, we confirmed that the primer pairs resulted in no product when amplifying genomic DNA samples from men with previously described deletions of the STS target site.

FISH

Interphase nuclei from AMC0110 and WHT4396 were hybridized with cosmid 18E8 by methods described elsewhere (Saxena et al. 2000; de Vries et al. 2002).

Amplification and Sequencing of Deletion Junctions

Tables A7 and A8 list primers used to amplify deletion breakpoints. The resulting PCR products were sequenced using the PCR primers (and additional, internal primers when necessary), a BigDye Terminator kit, and an ABI 3700 automated sequencer (Applied Biosystems).

Dot Plots

Dot plots were generated using the program self_dot_plot.pl (available at the Page Lab's Web site [The Human Y Chromosome: Annotated Sequence of the *AZFc* Palindromic Complex]).

Y-Chromosome Sequence

The sequence of the euchromatic, male-specific portion of the human Y-chromosome long arm is available in GenBank (contigs NT_011875 and NT_011903) (see the "Electronic-Database Information" section).

Results

In the course of screening for Y-chromosome deletions, we identified 11 azoospermic men who lacked STSs that define *AZFb* (Vogt et al. 1996) and whose deletions were restricted to *AZFb* and *AZFc*. Three of these men had deletions of *AZFb*, seven had deletions of *AZFb* plus *AZFc,* and one lacked two STSs in *AZFb* and was also deleted for *AZFc* (table 1 and fig. 1).

The localization of deletion breakpoints in these men was complicated by the massive and nearly identical ampliconic repeats that constitute 41% (5.9 Mb) of the euchromatic portion of the Y-chromosome long arm (Kuroda-Kawaguchi et al. 2001) (GenBank sequence contigs NT_011875 and NT_011903) (fig. 1*A*). To precisely localize these breakpoints, we used a three-stage strategy consisting of (1) low-resolution breakpoint mapping by using plus/minus STSs; (2) high-resolution breakpoint mapping by using additional, more closely spaced plus/minus STSs; and (3) PCR amplification of deletion junctions.

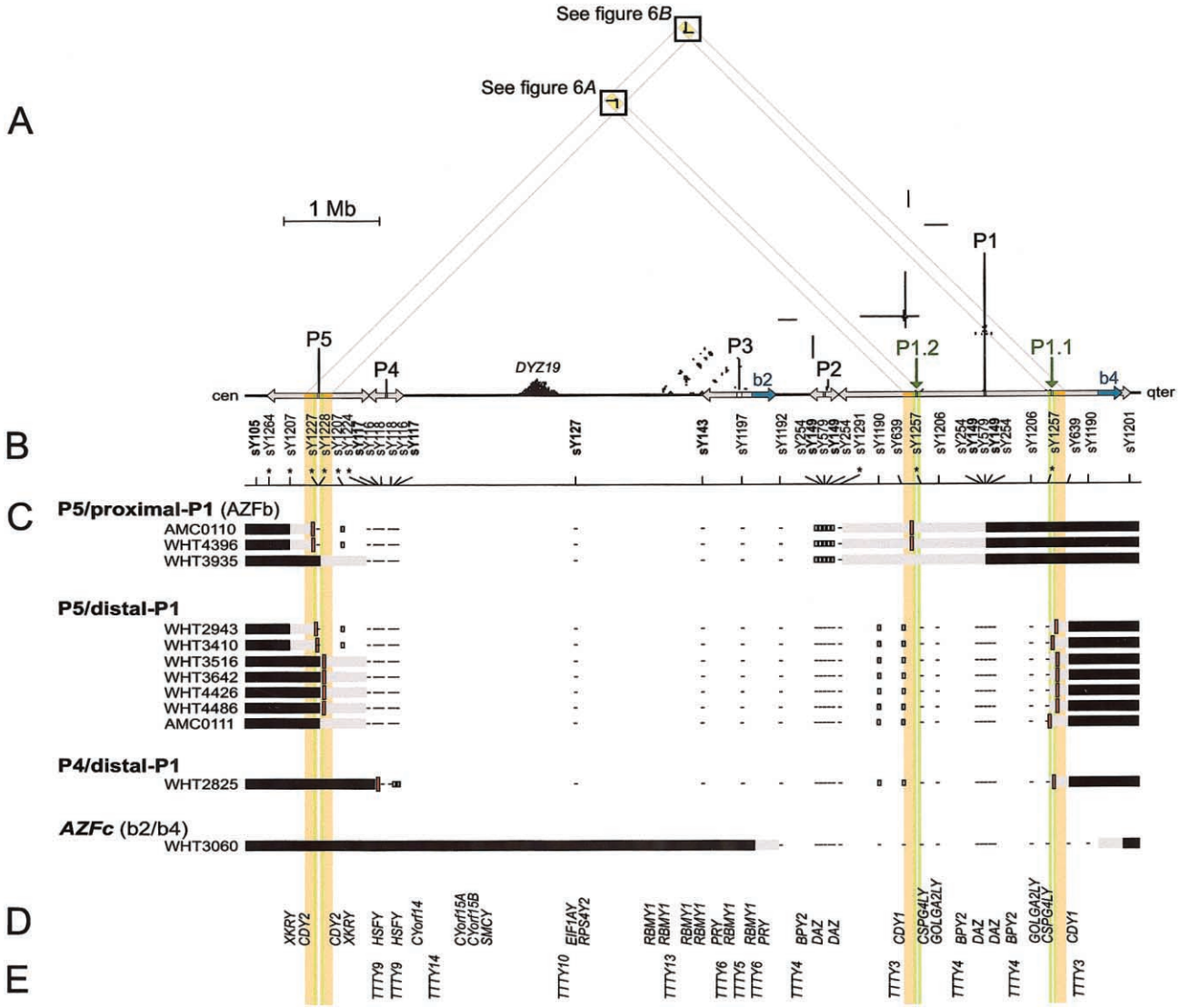


Figure 1 Deletions between P5 and P1 and between P4 and P1 in relation to the sequence of the Y chromosome. *A,* Triangular dot plot, encompassing P5, P4, and *AZFc* (which contains P1, P2, and part of P3). The baseline represents 8 Mb of Y-chromosome long-arm sequence. This triangular display avoids the redundant, artifactual symmetries that would appear in a square self-versus-self plot. Each dot within the plot represents 100 bp of identity between two parts of the Y chromosome in a window of 100 bp. Direct repeats appear as horizontal lines of dots, and inverted repeats appear as vertical lines. Palindromes appear as vertical lines that almost intersect the baseline. Five major palindromes are labeled "P1" through "P5" from right to left, and arrows along the baseline indicate the extents of their inverted-repeat arms. P1.1 and P1.2 are minipalindromes within P1 (see also fig. 6). The arms of P1.1 and P1.2 are too short (10 kb) to be visible at this scale. The "b2" and "b4" direct repeats that bound *AZFc* are shaded blue on the baseline. Sequences that are homologous between P5 and P1 are shaded orange and green on the baseline. Diagonal gray guide lines connect the P5-P1 homologous sequences on the baseline to the two areas within the plot that contain the corresponding dots. These areas are shaded orange and boxed. The prominent triangle of dots near the baseline and labeled "*DYZ19*" is a satellite repeat array. *B,* STSs used for low-resolution breakpoint mapping. Tick marks show STS positions; asterisks indicate new STSs. STSs used in the original definition of *AZFb* are in boldface. Results for sY1227 and sY1228 are difficult to represent at this scale; see table A1. *C,* Low-resolution plus/minus STS results and deletion breakpoint positions for 12 men with spermatogenic failure. At the left are the identifiers of the men studied, and to the right of each man's identifier is a representation of those parts of his Y chromosome that were determined to be present or absent. Horizontal black bars indicate confirmed presence of Y-chromosome DNA, and minuses indicate confirmed absence of Y-chromosome DNA, as assayed by low-resolution STSs. White boxes represent STS positives that were disregarded because of cross-amplifying loci elsewhere and because of negative results for flanking STSs. Horizontal gray bars represent the intervals to which breakpoints were localized by low-resolution breakpoint mapping. (Where STSs fall within gray bars, their results were positive). Short red vertical lines indicate the locations of amplified breakpoint junctions for nine of the patients and, for AMC0111, the 10-kb interval to which high-resolution STS mapping localized the distal breakpoint. *AZFc*-deleted patient WHT3060 is shown for comparison. *D,* Genes with significant confirmed or predicted ORFs (see the "Electronic-Database Information" section). *E,* Spliced but apparently noncoding transcripts (see the "Electronic-Database Information" section).

Table 1

Summary of Deletions between P5 and P1 and between P4 and P1

Deletion and Patient	Length of Deletion in Reference Sequence (Mb)	Length of Sequence Identity at Deletion Junction (bp)	Deletion-Junction Sequence Alignment	Deletion-Junction Sequence GenBank Accession Number
P5/proximal P1:[a]				
AMC0110	6.23	933	Fig. 3*A*	AF395664
WHT4396	6.23	73	Fig. 3*C*	AF437293
WHT3935	4.96–6.92	ND		
P5/distal P1:[b]				
WHT2943	7.68	25	Fig. 3*D*	AF395669
WHT3410	7.66	3	Fig. 4	AF395667
WHT3516	7.66	933	Fig. 3*E*	AF395665
WHT3642	7.66	933	Fig. 3*E*	AF395666
WHT4426	7.66	933	Fig. 3*E*	AF480412
WHT4486	7.66	933	Fig. 3*E*	AF480413
AMC0111	7.15–7.66	ND		
P4/distal P1:				
WHT2825	7.03	0[c]	Fig. 5	AF395668
AZFc (b2/b4):				
WHT3060[d]	3.25–3.75	ND		

NOTE.—ND = no data.

[a] Formerly *AZFb*.

[b] Formerly *AZFb* plus *AZFc*.

[c] Contains a 31-bp insertion relative to reference sequence at site of deletion junction.

[d] Shown for comparison purposes (Kuroda-Kawaguchi et al. 2001).

Low-Resolution Breakpoint Mapping

Guided by Y-chromosome sequence, we first assembled a panel of plus/minus STSs that would provide efficient low-resolution breakpoint localization in the regions of interest (fig. 1*B*). Although of low resolution, this panel still provides denser and more-informative coverage than the STSs heretofore used to assay deletions of *AZFb* and *AZFc*. Some of these new STSs were designed to amplify outer or inner edges of the very large, nearly perfect, palindromic inverted repeats that figure prominently in this region (figs. 1*A* and 1*B*). These palindromes consist of two inverted repeat "arms" surrounding relatively short "spacer" sequences that are not part of the inverted repeats. For the palindromes relevant to this study (P5, P4, P2, P1, and the minipalindromes P1.2 and P1.1), arm-to-arm divergence ranges from 0.02% to 0.03% (Kuroda-Kawaguchi et al. 2001) (GenBank sequence contigs NT_011875 and NT_011903) (figs. 1*A* and 1*B*). Consequently, the spacers and the boundaries between the palindromes and surrounding sequence offer the only possible locations for single-copy STSs within these palindromes.

Low-resolution mapping localized the proximal breakpoints to three areas: the proximal arm of palindrome P5 (in four men), the distal arm of P5 (in six men), and the proximal arm of P4 (in one man) (figs. 1*B* and 1*C* and table A1 [see appendix A]). Low-resolution mapping localized the distal breakpoints to two areas: the proximal arm of P1 (in the three men initially characterized as having deletions of *AZFb*) and the distal arm of P1 (in the remaining eight men) (figs. 1*B* and 1*C* and table A2). Thus, these deletions can be broadly categorized as (1) P5/proximal-P1 deletions (extending from P5 to the proximal arm of P1; previously thought to define *AZFb*), (2) P5/distal-P1 deletions (extending from P5 to the distal arm of P1; previously identified as deletions of *AZFb* plus *AZFc*), and (3) a P4/distal-P1 deletion (extending from P4 to the distal arm of P1).

The STS results clearly indicated that P5/proximal-P1 deletions remove part of *AZFc,* including all of P2 and the embedded proximal DAZ cluster—two closely spaced copies of the *DAZ* gene family in 3′←5′::5′→3′ orientation that straddle the center of P2 (Saxena et al. 2000; Kuroda-Kawaguchi et al. 2001) (fig. 1). (There is also a homologous, distal *DAZ* cluster that contains two additional *DAZ* genes that straddle the center of P1.) FISH analysis with a probe that produces one dot per *DAZ* cluster confirmed the absence of palindrome P2 in two men (AMC0110 and WHT4396) with P5/proximal-P1 (*AZFb*) deletions (fig. 2). No cells were available from the third man (WHT3935) with a P5/proximal-P1 deletion. Low-resolution mapping also clearly showed that all men who appeared to have deletions of both *AZFb* and *AZFc* in fact retained several hundred kilobases of distal *AZFc,* including one copy of *CDY1* (fig. 1).

High-Resolution Breakpoint Mapping

Plus/minus STSs cannot localize a breakpoint in one copy of an essentially perfect ampliconic duplication if another, intact copy is retained. The reason is that the

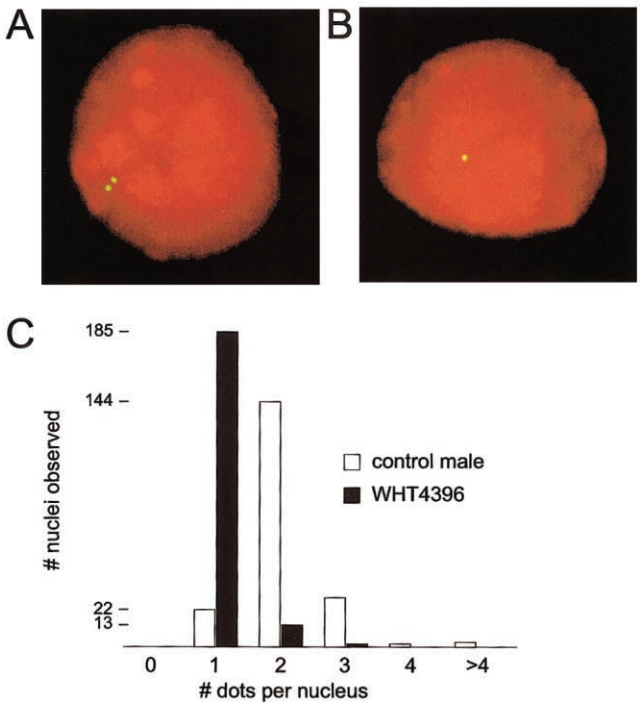


Figure 2 FISH, showing *DAZ* clusters in an unaffected man and in a man with a P5/proximal-P1 (*AZFb*) deletion. *A,* Representative interphase nucleus from a fertile, male control individual hybridized with cosmid 18E8 (*yellow*), showing two dots (corresponding to two *DAZ* clusters, each containing two *DAZ* genes). *B,* Representative interphase nucleus from P5/proximal-P1–deleted patient WHT4396, showing one dot (corresponding to one *DAZ* cluster). *C,* Histogram of number of nuclei observed grouped by number of dots per nucleus.

STSs amplify the intact copy of the amplicon and therefore cannot reveal which portion of the partial copy is absent. For example, in men with P5/proximal-P1 deletions, plus/minus STSs cannot localize distal breakpoints to an interval <1.5 Mb because of the masking effects of the retained, distal arm of P1 (figs. 1*B* and 1*C* and table A2). However, low-resolution STS breakpoint mapping showed that, for 10 of the 11 men, at least one of the two breakpoints was not masked by the arm of a major palindrome and was therefore amenable to more-precise localization by high-resolution breakpoint mapping. For all men except WHT3935 (in whom both breakpoints were masked), we used new, closely spaced STSs to localize at least one of the two breakpoints to an interval small enough to attempt PCR amplification of the deletion junction (table A3).

Deletion-Junction Amplification and Sequencing

For three men (WHT2943, WHT3410, and WHT2825), both the proximal and distal deletion breakpoints were precisely localized by high-resolution STS mapping (table A3). For these men, we designed PCR primers on each side of the deletion, to amplify deletion junctions. These primers resulted in products in the men with deletions but not in control individuals (tables A7 and A8).

For two men (AMC0110 and WHT4396), only the proximal breakpoint could be precisely localized by high-resolution mapping (table A3). For these men, we designed PCR primer pairs for deletion-junction amplification, on the basis of the hypothesis that the deletions were caused by homologous recombination: we designed one primer just proximal to the interval known to contain the proximal breakpoint and another primer just distal to the homologous interval (presumably containing the distal breakpoint) (table A7). Indeed, primer pairs designed in this way generated products in the men with deletions but not in control individuals, including AMC0110's father.

We designed deletion-junction primers in an analogous fashion for four other men (WHT3516, WHT3642, WHT4426, and WHT4486), in whom only the distal breakpoints could be precisely localized (table A3). Again, the deletion-junction primers resulted in products in these men but not in control individuals, including WHT3516's brother and WHT4486's father. This strategy was attempted for AMC0111 but was unsuccessful. This man's proximal breakpoint lies in the distal arm of P5 and is therefore masked by the proximal arm of P5. His distal breakpoint falls within the 10-kb proximal arm of minipalindrome P1.1 in the distal arm of P1 (fig. 1*A*), and finer localization was impossible because of masking by the distal arm of P1.1.

We sequenced all deletion-junction products to confirm their identity, to establish the locations of deletions within the products, to gather additional information on deletion mechanisms, and to estimate deletion lengths accurately (figs. 3–5 and table 1). For three men (AMC0110, WHT3516, and WHT4486), DNA from a patrilineal relative was available. We used DNA from these relatives to check whether, prior to the deletions, the sequences of the deletion sites had been the same as the Y-chromosome reference sequence. In all instances, the sequences of 1.5-kb PCR products surrounding the deletion breakpoints indeed agreed with the reference sequence.

Evidence for Nonhomologous Recombination in Addition to Homologous Recombination

The alignment of deletion-junction sequences with corresponding proximal and distal Y-chromosome reference sequences reveals homologous recombination as the likely mechanism of seven of the nine deletions for which we amplified deletion junctions (fig. 3). However, examination of the junction sequences from the two remaining deletions indicates that homologous recom-

Table 2

Genes Affected by P5/Proximal-P1, P5/Distal-P1, and P4/Distal-P1 Deletions

			No. of Copies Deleted		
Gene[a]	Expression Pattern[b]	No. of Copies	P5/Proximal P1	P5/Distal P1	P4/Distal P1
CDY2	Testis	2	1	1	0
XKRY	Testis	2	1	1	0
HSFY	Testis	2	2	2	1
CYorf14	W	1	1	1	1
CYorf15A	W	1	1	1	1
CYorf15B	W	1	1	1	1
SMCY	W	1	1	1	1
EIF1AY	W	1	1	1	1
RPS4Y2	Testis, prostate	1	1	1	1
RBMY1	Testis	6	6	6	6
PRY	Testis	2	2	2	2
BPY2	Testis	3	1	3	3
DAZ	Testis	4	2	4	4
CDY1	Testis	2	1	1	1
CSPG4LY	Testis	2	0	2	2
GOLGA2LY	Testis	2	0	2	2
TTTY9[c]	Testis	2	2	2	1[d]
TTTY14[c]	Testis, kidney, fetal brain	1	1	1	1
TTTY10[c]	Testis, prostate, fetal brain	1	1	1	1
TTTY13[c]	Testis	1	1	1	1
TTTY6[c]	Testis	2	2	2	2
TTTY5[c]	Testis	1	1	1	1
TTTY4[c]	Testis, prostate	3	1	3	3
TTTY3[c]	Testis	2	1	1	1
Total		46	32	42	38

[a] For GenBank accession numbers, see the "Electronic-Database Information" section.

[b] W = widely expressed (including testis).

[c] Spliced transcript without significant ORF.

[d] Proximal breakpoint lies only 0.9 kb 5′ of the 5′-most available sequence of the *TTTY9* transcript; possibly the deletion affects *TTTY9* transcription or 5′ transcript sequence that has not been identified.

bination could not have been the deletion mechanism (figs. 4 and 5).

Breakpoints Cluster near Central P5, P1.1, and P1.2

To further investigate the role of sequence homology in the deletions we were studying, we generated a dot plot that compares the Y-chromosome region between sY105 and sY1201 to itself (fig. 1*A*). As expected, this plot showed that *AZFc* and immediately flanking sequences contained the largest direct and inverted ampliconic duplications (Kuroda-Kawaguchi et al. 2001). Otherwise, the only significant similarities occur between (1) central P5 and (2) two 92-kb sequences in the proximal and distal arms of P1 that contain minipalindromes P1.2 and P1.1, respectively. All amplified P5/proximal-P1 and P5/distal-P1 deletion junctions are located within these sequences. The dots corresponding to these homologies are boxed in figure 1*A* and are shown as close-up views in figure 6.

Within these regions of homology are four copies of a 933-bp sequence. One copy, in the proximal arm of P5, is in the same orientation as a copy just proximal to P1.2, and one deletion appears to be due to homologous recombination between these copies (AMC0110) (figs. 1, 3*A*, 3*B*, and 6*A*). Another copy of this 933-bp sequence, in the distal arm of P5, is in the same orientation as a copy just distal to P1.1, and four of the deletions appear to be due to homologous recombination between these two copies (figs. 1, 3*E*, and 6*B*). Three of the remaining deletions also have proximal breakpoints near the center of P5 and distal breakpoints either near P1.2 (WHT4396) or in P1.1 (WHT2943 and WHT3410) (figs. 1 and 6). Two other deletions have distal breakpoints in P1.1 (AMC0111 and WHT2825) (fig. 1 and table A6).

Discussion

In the present study, we have localized deletion breakpoints in three men initially characterized as having deletions of *AZFb* and in eight men initially characterized as having deletions of *AZFb* plus *AZFc*. We then sequenced the deletion junctions in nine of these men, and the sequences indicated that seven of the deletions can be explained by homologous recombination, whereas two cannot. We found that all proximal breakpoints except one are clustered near the center of palindrome P5 and that all distal breakpoints are clustered near one of two

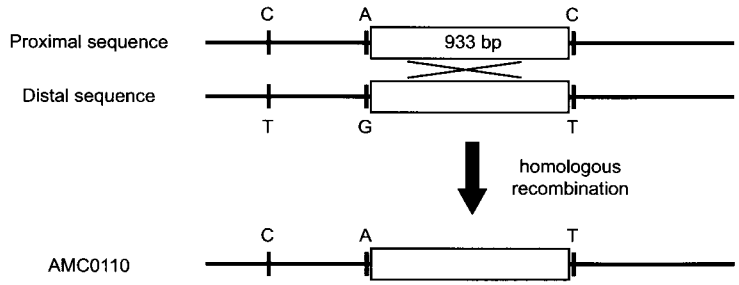
A

```
Proximal sequence  ..........
AMC0110            ATTAAGAGCCAGCACTTATCACCTGCCAAGGGCAAGCATGGTGGAGTGGTGAGTGTTGTA
Distal sequence    ............T.....................................

Proximal sequence  ..........
AMC0110            TGAGCAAGTGTGGGCTCCAACCACTAAACTCATCCAGGCATGCCAGCTGTGGCAAAACAA
Distal sequence    ..........

Proximal sequence  ..............[933 bp sequence identity]C.......
AMC0110            GAAATTTTAGGTGCCAA                        TTCTTGAT
Distal sequence    ..............G                         ........
```

B



C

```
Proximal sequence  ..........
WHT4396            AGAGGCAAGGTGGGTGTAGGTACCATAATGGACAGCTGTCGCAAAGCAGCAATCAGAAGA
Distal sequence    .A..........

Proximal sequence  ............[73 bp sequence identity]C..................
WHT4396            GTCTGACACCTGTG                        TTAAGTTATAGAAACAGAAA
Distal sequence    ............A                        ....................

Proximal sequence  ..........................................G.............
WHT4396            ACTTTGGTGTCAAATAGATGAAGAACTATTCTAAATTATAAAAACAGAACCACAGCTTCT
Distal sequence    ..........
```

D

```
Proximal sequence  ..........
WHT2943            ATAGACAAGTAACAGGCTCTGAAATTCAGGCAATAAGTAATATTGCAAAAAAAAAAGTCCA
Distal sequence    .....A.....................................................-.....

Proximal sequence  ..........
WHT2943            GGACCAGATGGATTCACATCCGAATTCTACCAGAGGTACAAGGAGGAGCTGGTACCATTC
Distal sequence    ................................C........T...C.............

Proximal sequence  ............[25 bp sequence identity]C.....................
WHT2943            CTTCTGAAACGAT                        TCCTCTCTAACTCATTTTATGAG
Distal sequence    .........A..G                        ......................

Proximal sequence  ..........
WHT2943            GCCAGCATCATCCTGATACCAAAGCCTGGCAGAGACACAGCAAAAATAGAGAATTTTAGA
Distal sequence    ..........

Proximal sequence  ..........................................T..............T.
WHT2943            CCAAAATCCCTGATGAACATCAATGCAAGAATCGTCAATAAAAAACTGGCAAACCGAACC
Distal sequence    ..........
```

E

```
Proximal sequence                                  ............................
Patients = WHT3516 / WHT3642 / WHT4426 / WHT4486    CAAAGATGCGGGTACTGCTCTATTTCACAAG
Distal sequence                                    .........A...........GC........

Proximal sequence  ............[933 bp sequence identity]T..........
Patients           GCATTGATCAAGAG                        CTGGCACCTAA
Distal sequence    ............A                        ...........

Proximal sequence  ..........
Patients           AATTTCTTGTTTTGCCACAGCTGGCATGCCTGGATGAGTTTAGTGGTTGGAGCCCACACT
Distal sequence    ..........

Proximal sequence  ..................................................G.........
Patients           TGCTCATACAACACTCACCACTCCACCATGCTTGCCCTTGGCAGGTGATAAATGCTGGCT
Distal sequence    ..........
```

minipalindromes, P1.1 and P1.2, in palindrome P1. These results support several significant conclusions: (1) factors in addition to homology (perhaps related to the nearby palindrome centers) underlie these deletions; (2) the deletions previously thought to define the *AZFb* region actually extend 1.5 Mb into the *AZFc* region; (3) P5/distal-P1 deletions, previously thought to be deletions of *AZFb* plus *AZFc,* spare the distal portion of *AZFc* and constitute a distinct class of recurrent deletions; and (4) all deletions studied are massive and remove numerous genes, although the only reported phenotype in men with these deletions is spermatogenic failure.

Factors in Addition to Homology Underlie P5/P1 Deletions

For *AZFa* and *AZFc,* high sequence similarity between flanking repeats, including long segments of identity, is the only known factor accounting for the frequency of these recurrent deletions (Blanco et al. 2000; Kamp et al. 2000; Sun et al. 2000; Kuroda-Kawaguchi et al. 2001). In contrast, although homology is important in many deletions between P5 and proximal or distal P1, our results provide evidence for involvement of an additional important factor:

First, every precisely localized proximal breakpoint (except WHT2825's) lies in a hotspot within 30 kb of the center of P5, and every precisely localized distal breakpoint lies in one of two hotspots within 25 kb of either P1.2 or P1.1 (figs. 1 and 6). Homology alone does not account for these hotspots, since no deletion breakpoints were observed in 65 kb of immediately adjacent homologous sequence (fig. 6); these homologies have approximately the same level of overall similarity as those involved in the deletions and contain longer segments of perfect identity.

Second—and perhaps most significantly—homologous recombination provides no explanation for two of the deletions. One of these has breakpoints in the P5 and P1.1 hotspots (WHT3410) (figs. 4 and 6*B*), and the other has its distal breakpoint in the P1.1 hotspot (WHT2825) (fig. 5). Indeed, these are the first identified examples of nonhomologous recombination causing interstitial Y-chromosome deletions.

Although the molecular factors responsible for the P5, P1.2, and P1.1 hotspots are unknown, their association with the center of P5 and the immediate vicinity of P1.2 and P1.1 is striking (fig. 6). Perfect palindromes (those without central, nonrepeated spacers) appear to be highly unstable in the mammalian germline, showing rearrangement in up to 56% of progeny (Collick et al. 1996; Akgun et al. 1997; Lobachev et al. 2000). Palindromes like P5, P1.2, and P1.1 that have central spacers are far more stable. Nevertheless, we conjecture that such palindromes could be subject to a slight propensity for breaks and that this propensity could account for the P5, P1.2, and P1.1 deletion hotspots. In support of this conjecture, we note that a palindrome-like inverted repeat containing the *NEMO* and *LAGE2* genes also appears to be unusually susceptible to rearrangements (Aradhya et al. 2001).

No Distinct AZFb *Interval*

Although the *AZFb* and *AZFc* intervals were thought to be nonoverlapping (fig. 7) (Vogt et al. 1996), the results of the present study establish that the deletions taken to define an *AZFb* interval are deletions between P5 and the proximal arm of P1 and overlap *AZFc* by 1.5 Mb (figs. 1 and 7). These results are in fact consistent with *AZFb*'s original definition—negative results for sY117, sY127, and sY143 coupled with positive results for flanking markers sY105 and sY149 (Vogt et al. 1996). Heretofore, however, lack of single-copy STSs in and near *AZFc* precluded detection of the overlap. The analysis that here establishes this overlap was made possible by recently available Y-chromosome genomic sequence; this sequence clarifies the organization of the large (0.2–1.5 Mb), nearly identical repeats that contain the breakpoints of P5/proximal-P1, P5/distal-P1, and P4/distal-P1 deletions (GenBank contigs NT_011875 and NT_011903) (Kuroda-Kawaguchi et al. 2001) (figs. 1*A* and 1*B*).

P5/Distal-P1 Deletions Constitute an Important Class of Recurrent Y-Chromosome Deletions

Our results also show that P5/distal-P1 deletions (formerly identified as deletions of *AZFb* plus *AZFc*) constitute a class of recurrent deletions distinct from deletions of *AZFc* and from P5/proximal-P1 (*AZFb*) deletions. These deletions spare the distal portion of *AZFc* and are generated by a mechanism unrelated to the b2/b4 recombination responsible for *AZFc* dele-

Figure 3 P5/P1 deletion junctions, indicating homologous recombination as the cause of the deletion. *A,* Junction sequence in AMC0110 aligned with proximal and distal Y-chromosome reference sequence. Dots indicate base pairs identical to the junction sequence. *B,* Model of production of the deletion-junction sequence in AMC0110 by homologous recombination. The bases shown (C/T, A/G, and C/T) differ between the proximal and distal copies of the sequence and therefore define the location of the deletion breakpoint within the junction sequence. *C,* Junction sequence in WHT4396 aligned with proximal and distal Y-chromosome reference sequence. *D,* Junction sequence in WHT2943 aligned with proximal and distal Y-chromosome reference sequence. *E,* Junction sequence in WHT3516, WHT3642, WHT4426, and WHT4486 aligned with proximal and distal Y-chromosome reference sequence.

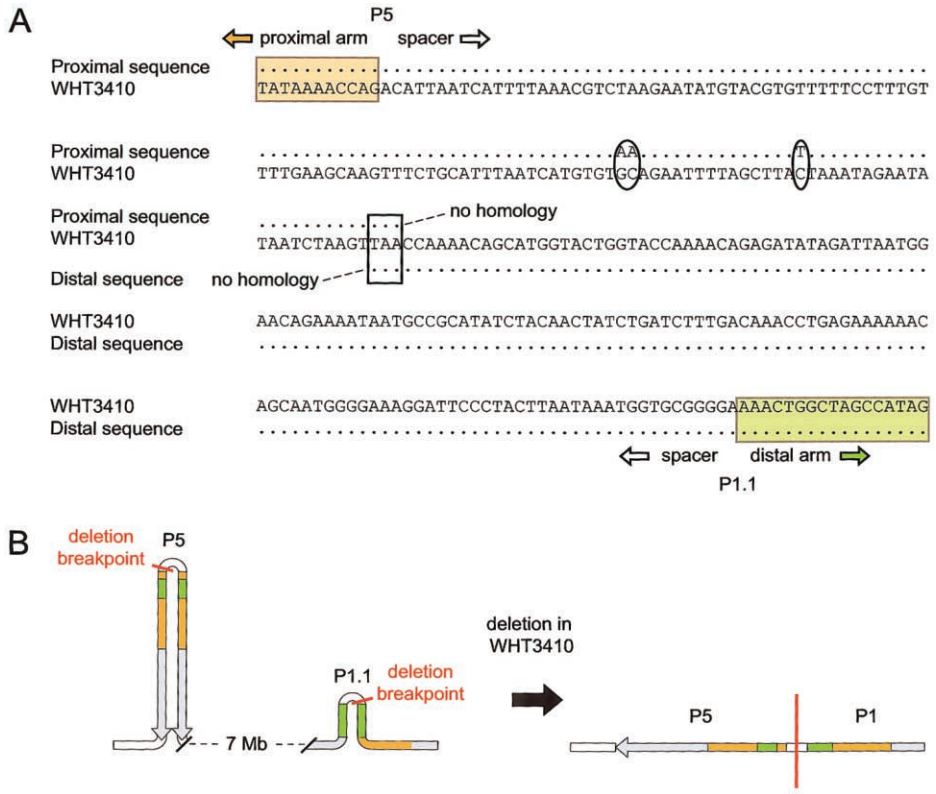


Figure 4 P5/distal-P1 deletion junction in WHT3410. *A*, Junction sequence aligned with proximal and distal Y-chromosome reference sequences. Ovals mark 3 bp that differ between WHT3410 and the Y-chromosome reference sequence. The white box marks the 3 bp that contain the deletion junction. *B*, Breakpoint locations within palindrome P5 and minipalindrome P1.1, and the resulting sequence organization in WHT3410. Orange and green shading is as in figures 1 and 6.

tions (fig. 1*A*) (Kuroda-Kawaguchi et al. 2001). In the present sample of azoospermic men, P5/distal-P1 deletions are more frequent than *AZFa* or P4/distal-P1 deletions are: 7/602 versus 1/602 ($P < .07$, by Fisher's exact test, two sided). The combined frequency of deletions between P5 and either proximal or distal P1 is significantly higher than the frequencies of *AZFa* or P4/distal-P1 deletions are in the present sample: 10/602 versus 1/602 ($P < .012$, by Fisher's exact test, two sided). However, the frequency of all P5/P1 deletions among men with nonobstructive azoospermia (1.7%, 95% CI 0.8%–3.0%) is still markedly lower than the

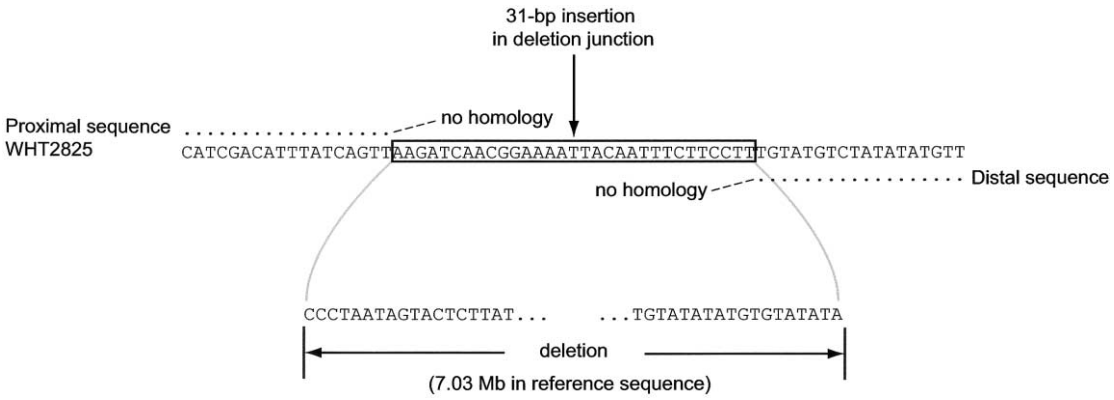


Figure 5 P4/distal-P1 deletion junction in WHT2825. Junction sequence is aligned with proximal and distal Y-chromosome reference sequence.

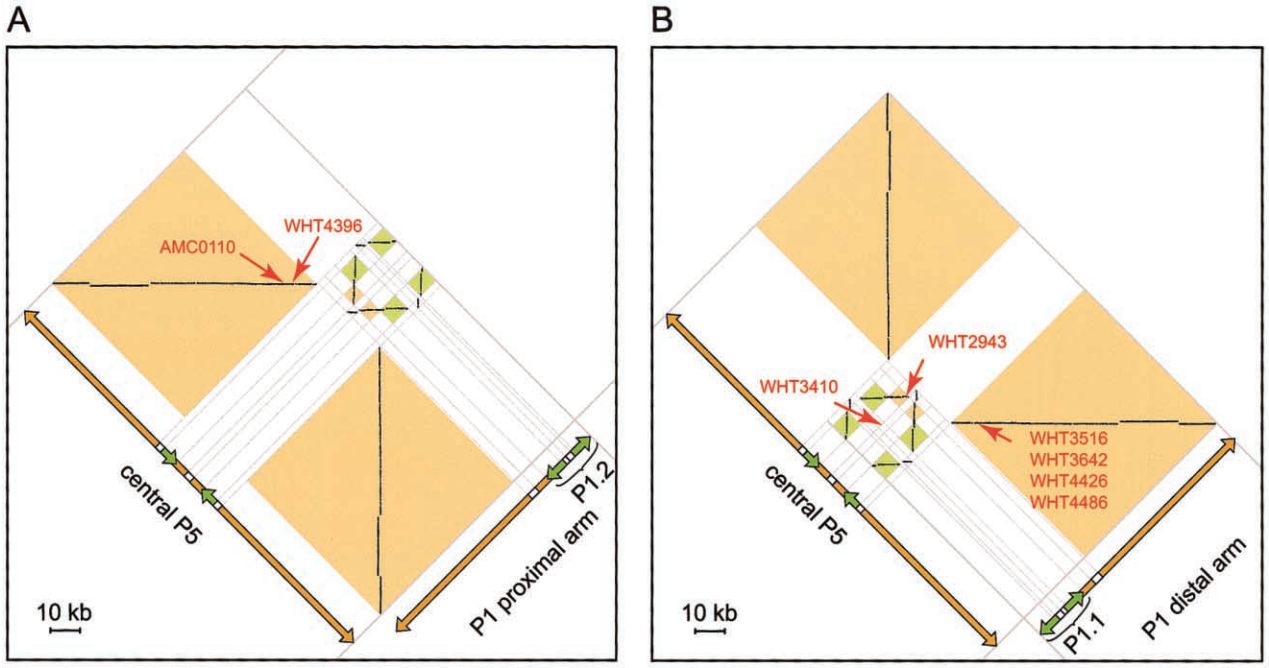


Figure 6 "High-magnification" dot plots of the two regions of high similarity between P5 and P1 (see fig. 1*A*). Each dot represents 50 bp of identity in a window of 50 bp. Orientation and orange and green shading is as in figure 1*A*. *A*, Central P5 versus P1.2 and the neighboring region of the proximal arm of P1. Locations of the P5/proximal-P1 deletions in AMC0110 and WHT4396 are indicated by red arrows. (The coordinate of a deletion in the plot is the location of the proximal breakpoint paired with the location of the distal breakpoint. Deletions that extend between direct repeats map to horizontal lines of dots.) *B*, Central P5 versus P1.1 and the neighboring region of the distal arm of P1. Locations of the P5/distal-P1 deletions in WHT2943, WHT3410, WHT3642, WHT3516, WHT4426, and WHT4486 are indicated. WHT3410's deletion does not extend between homologous sequences and therefore does not map to a line of dots.

frequency of *AZFc* deletions among these men (6.0%, 95% CI 4.0%–8.7%) (Oates et al., in press).

P5/Proximal-P1 and P5/Distal-P1 Deletions Are Much Larger than AZFc

All deletions studied are massive, removing one-fourth to one-third of the euchromatic, male-specific region of the Y chromosome: P5/proximal-P1 deletions excise up to 6.23 Mb, P5/distal-P1 deletions excise up to 7.66 Mb, and the P4/distal-P1 deletion excises 7.03 Mb (table 1 and fig. 7). They dwarf human Y-chromosome *AZFc* deletions (~3.5 Mb) (table 1 and figs. 1 and 7) (Kuroda-Kawaguchi et al. 2001) and *AZFa* deletions (0.8 Mb) (Blanco et al. 2000; Kamp et al. 2000; Sun et al. 2000). They are also substantially larger than the estimated sizes of the largest recurrent human interstitial deletions of which we are aware—for example, those responsible for Smith-Magenis syndrome (MIM 182290) (for review, along with other interstitial deletions, see Mazzarella and Schlessinger 1998; Ji et al. 2000; Shaffer and Lupski 2000; Stankiewicz and Lupski 2002). To our knowledge, the recurrent deletions described here are the largest, in the human genome, for which amplified breakpoints and sequence-based size estimates have been reported.

Finished Y-chromosome genomic sequence was essential for the precise localization and sequencing of these deletion junctions. Our results illustrate the importance of continued efforts to accurately finish the sequence of other large-scale ampliconic duplications in the human genome. These duplications present formidable technical obstacles to mapping and complete sequencing (Kuroda-Kawaguchi et al. 2001). Nevertheless, the regions containing such duplications, like the Y-chromosome region extending from P5 to P1, are biologically important because they are often exposed to rearrangement that causes genetic disease (Eichler 2001).

P5/P1 Deletions Remove as Many as One-Fourth of Y-Chromosome Transcripts

P5/proximal-P1 deletions remove 32 genes and transcripts of a total of ~150 on the male-specific region of the Y chromosome, including all members of seven testis-specific families and some members of six additional testis-specific families (figs. 1*C*–1*E* and table 2). P5/distal-P1 deletions remove a total of 42 genes or transcripts, including all members of 11 testis-specific families and some members of 4 additional testis-specific families (figs. 1*C*–1*E* and table 2). In addition, all the deletions

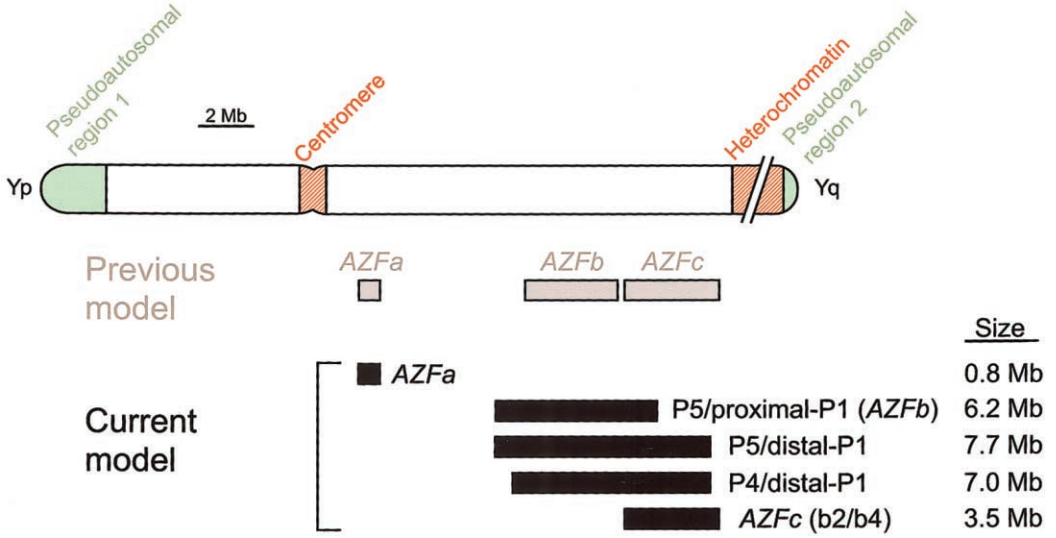


Figure 7 Previous model of recurrent, interstitial Y-chromosome deletions that cause infertility in men (Vogt et al. 1996) contrasted with current model.

studied here result in the loss of a number of genes expressed in multiple tissues. These include five widely expressed genes, one of them a translation initiation factor (*EIF1AY*).

Nevertheless, the only reported phenotype is impaired spermatogenesis. This singularity of phenotype reflects the remarkable functional specialization of the affected regions of the human Y chromosome and contrasts starkly with the multiplex phenotypes—"contiguous gene syndromes"—usually associated with large genomic deletions.

Dissection of Gene Function Will Require Smaller Deletions

Neither P5/proximal-P1 nor P5/distal-P1 deletions allow us to dissect the functions of the many affected genes. These deletions provide no evidence that a single gene or gene family—for example, *RBMY1* (Elliott et al. 1997, 2000)—encodes an azoospermia factor b. On the contrary, the P5/proximal-P1 (*AZFb*) deletion phenotype may be the aggregate effect of complete or partial loss of function of many genes and gene families.

It is to be hoped, however, that small, as-yet-undetected deletions will offer critical opportunities to dissect the roles that individual Y-chromosome genes and gene families play in spermatogenesis (Yen 2001). These deletions are likely to show lower penetrance or expressivity than *AZFa, AZFc,* P5/proximal-P1, and P5/distal-P1 deletions. Consequently, they may be less enriched among infertile men, and many thousands of Y chromosomes may have to be examined to find them.

Whenever such a deletion is found, substantial work will be needed to characterize the extent of the deletion, because of the technical difficulties presented by the nearly identical, ampliconic repeats that are so abundant in the male-specific region of the Y chromosome. In addition to this difficulty, it will be necessary to characterize any other genetic variations in the male-specific region of the Y chromosome bearing the deletion, since these too may affect the observed phenotype and will be recombinationally inseparable from the deletion. Thus, gathering these deletions will require the efforts of many clinics and laboratories over the coming years, and fully characterizing them will demand the utmost technical care. The sequence-guided analytical approaches developed here should provide a foundation for this research.

Note added in proof.—We have identified an additional testis-specific transcript family, *TTTY17* (GenBank accession number AF527829), which is affected by P5/P1 deletions. *TTTY17* occurs in three copies, each located near one of the three copies of *TTTY4*. With the inclusion of *TTTY17,* the total number of transcripts removed by P5/proximal-P1 deletions becomes 33, the number removed by P5/distal-P1 deletions becomes 44, and the number affected by the P4/distal-P1 deletion becomes 41.

Acknowledgments

We thank L. Brown, S. K. M. van Daalen, C. M. Korver, T. Pyntikova, B. Redeker, and J. de Vries, for technical contributions; C. Brunning and W. A. Hogge, for a patient sample; and N. Halim, T. P. Rasmussen, P. Tomasi, and P. J. Wang, for comments on the manuscript. This work was supported by the National Institutes of Health, the Howard Hughes Medical Institute, and the Academic Medical Center.

Appendix A

Table A1

Low-Resolution Mapping of Proximal Breakpoint

Deletion and Patient	Results for STSs Flanking Proximal Breakpoint												Inferred Proximal Breakpoint Location
	Palindrome P5						Palindrome P4						
	Proximal Arm		Spacer-to-Arm Boundaries		Distal Arm		Proximal Arm			Distal Arm			
	sY1264	sY1207	sY1227[a]	sY1228[a]	sY1207	sY1224	sY117	sY116	sY118	sY118	sY116	sY117	
P5/proximal P1:[b]													
AMC0110	+	+	−	−	(+)	−	−	−	−	−	−	−	Proximal P5
WHT4396	+	+	−	−	(+)	−	−	−	−	−	−	−	Proximal P5
WHT3935	+	+	+	+	(+)	−	−	−	−	−	−	−	Distal P5
P5/distal P1:[c]													
WHT2943	+	+	−	−	(+)	−	−	−	−	−	−	−	Proximal P5
WHT3410[d]	+	+	−	−	(+)	−	−	−	−	−	−	−	Proximal P5
WHT3516	+	+	+	+	(+)	−	−	−	−	−	−	−	Distal P5
WHT3642	+	+	+	+	(+)	−	−	−	−	−	−	−	Distal P5
WHT4426	+	+	+	+	(+)	−	−	−	−	−	−	−	Distal P5
WHT4486	+	+	+	+	(+)	−	−	−	−	−	−	−	Distal P5
AMC0111	+	+	+	+	(+)	−	−	−	−	−	−	−	Distal P5
P4/distal P1:													
WHT2825	+	+	+	+	+	+	+	+	−	−	(+)	(+)	Proximal P4

Note.—+ = Positive result; − = negative result; (+) = uninformative positive result because of a cross-amplifying copy in the other palindrome arm.

[a] No distinction is made between proximal and distal inner-arm-to-spacer boundaries, because orientation of the P5 spacer is expected to be polymorphic owing to putative recurrent inversions between the palindrome arms (for an example, see table A3, WHT3410).

[b] Formerly *AZFb*.

[c] Formerly *AZFb* plus *AZFc*.

[d] For sY1227 and sY1228, the primer site in the arm is present, but the primer site in the spacer is deleted.

Table A2

Low-Resolution Mapping of Distal Breakpoint

Deletion and Patient	Results for STSs Flanking Distal Breakpoint									Inferred Distal Breakpoint Location
	Palindrome P1									
	Proximal Arm Boundary	Proximal Arm			Spacer	Distal Arm			Distal Arm Boundary	
	sY1291	sY1190	sY639	sY1257	sY579	sY1257	sY639	sY1190	sY1201	
P5/proximal P1:[a]										
AMC0110	−	(+)	(+)	(+)	+	+	+	+	+	Proximal P1
WHT4396	−	(+)	(+)	(+)	+	+	+	+	+	Proximal P1
WHT3935	−	(+)	(+)	(+)	+	+	+	+	+	Proximal P1
P5/distal P1:[b]										
WHT2943	−	(+)	(+)	−	−	−	+	+	+	Distal P1
WHT3410	−	(+)	(+)	−	−	−	+	+	+	Distal P1
WHT3516	−	(+)	(+)	−	−	−	+	+	+	Distal P1
WHT3642	−	(+)	(+)	−	−	−	+	+	+	Distal P1
WHT4426	−	(+)	(+)	−	−	−	+	+	+	Distal P1
WHT4486	−	(+)	(+)	−	−	−	+	+	+	Distal P1
AMC0111	−	(+)	(+)	−	−	−	+	+	+	Distal P1
P4/distal P1:										
WHT2825	−	(+)	(+)	−	−	−	+	+	+	Distal P1

Note.—+ = Positive result; − = negative result; (+) = uninformative positive result because of a cross-amplifying copy in the other palindrome arm.

[a] Formerly *AZFb*.

[b] Formerly *AZFb* plus *AZFc*.

Table A3

High-Resolution Mapping of Proximal and Distal Breakpoints

Patient(s)	Proximal Primer Pairs[a]	Proximal Interval Length[b] (kb)	Distal Primer Pairs[a]	Distal Interval Length[b] (kb)	Comments
AMC0110	Table A4	1.5			
WHT4396	Table A4	1.8			
WHT3935					Both proximal and distal breakpoints are masked
WHT2943	Table A4	See comments	Table A6	2	We initially mistook the small, 2.7-kb deletion in P5 for the P5/distal-P1 deletion but then amplified the deletion junction by using long-range PCR with primers that bracket both the 2.7-kb deletion and the P5/distal-P1 deletion (GenBank accession number AF395669) (table A8)
WHT3410	Table A4	See comments	Table A6	See comments	High-resolution breakpoint mapping was complicated, because spacer orientation in both P5 and P1.1 is reversed compared to the reference sequence (expected to be a common polymorphism); we therefore successfully attempted long-range PCR by using primers in the proximal arm of P5 and the distal arm of P1.1 (table A8)
WHT3516, WHT3642, WHT4426, and WHT4486			Table A6	1.5	
AMC0111			Table A6	10	Masking by the distal arm of P1.1 precluded finer localization of distal breakpoint
WHT2825	Table A5	1.2	Table A6	1.0	

[a] Table that shows primer pairs and PCR results used to refine breakpoint locations.
[b] Lengths of intervals to which breakpoints were localized by high-resolution breakpoint mapping.

Table A4

High-Resolution Breakpoint Mapping in the Proximal Arm of Palindrome P5

STS	Length (bp)	Forward Primer	Reverse Primer	Results for AMC0110	WHT4396	WHT2943	WHT3410
sY1207	…[a]	…[a]	…[a]	+	+	+	+
10662/3	312	CTCTTGCTTGATCCTCAATTTC	ATGCCAGGAATTGGTCACA	+			
10660/1	346	CCATATGCGCTTTCTCTTGC	AGAAGTCACTCTCGGTCGGA	+			
10664/5	443	AGGAAGCCCAGAAATCAGAA	CAGCCTGAGTAACGTGGTGA	+	+	+	+
10666/7	413	TGAGATATCTACGAATTACCTTCCTG	GGAGATCAGAACCAGCCTGA	+	+		
10741/2	207	AAAACCAGCTAGAGGATCAGGAC	CATCTTAGCTCTTAAAGTTTACGGG	+	+	+	+
10739/40	420	CAGGACATGCTGGTTCATTACA	CAATTTTGCAGAGGCCGC	+	+	+	+
10821/3	452	GGCAAGTTTCACAATCTGCTG	GCCCTTGGCAGGTGATAAG	+	+		
10823/4	1,121	GCTTGCATTAAGAGCCAGCAC	CAAAGATGCGGGTACTGCTCTAT	−	+		
10827/8	743	TCTTGATCAATGCCTTGTGAAAT	ACCCACTGTAAGTCTGACGGAA	−	−		
10737/8	185	GGCAAATGCCTTTTTCTTCAC	CAAGTGTGATATCTTGCAGCCA	−	−	+	+
10735/6	190	CATGAGAGAGCATAACAAATCGC	CCATTATAATAACCAAGCCAAAGC	−	−	+	+
10733/4	381	ATAGAGGCTGCCTAGCTTTGC	CCCTGAGAGATTAGAAAAGCTGAT	−	−	+	+
10731/2	577	GTTGCAGTCAGCCAAGGTG	GATGGGCAGTTAGTGTAGGCATA	−	−	+	+
10743/4	492	GCTTCCCGGGTTCACGA	AGAGCGGCAGTGGTTTCAA	−	−	+	+
10831/2	484	CAAGTTTTCAGAGACGGCTGAC	GCCTGAGCCTTGGGTGGT			+	
10899/900	657	CAGAGGTGACCACCCAAGG	CGTATTAATTATACAGAATGGGCTCC			+	
10901/2	687	CATTGAGGCAAAGTTGAGTTAAAA	TGATGGCCATGCTACCCAT			+	
10745/6	305	GAAACCTACTGCACAGGATGTTATG	ATAGGGAAAAAGTAAAGTATGCCCC	−	−	−[b]	+
10747/8	387	GAAGGTTTGTAATCTGTCCCTATG	CCAAGTGACTCACCTCCTACA	−	−	−[b]	+
11185/6	159	GTGCAGTGCACCGAGGG	CCACTGGAGAAGCGCAGTATTAA			+	
10751/2	244	GCTACAAGGCTACAACCAATGG	GGTTTTAGGTCTAACATGTAAGTCCC	−	−	−	+
10841/2	480	AACGCCAAACACAATGACAA	ACGGACTTGCACAATGGC				+
10843/4	462	TCAGCCATTGTGCAAGTCC	ACGCTATCCCTCACCCCTT				+
10847/8	306	CTGTCTTTGTAGGTTGGCCTACA	GAACTTTGCACAGCATACCCA				+
10974/5	199	TGATCCTTGGGTATGCTGTG	AAGAGCAGCTGAATCCACG				+
sY1227	…[a]	…[a]	…[a]	−	−	−	−
sY1228	…[a]	…[a]	…[a]	−	−	−	−

NOTE.—Markers are ordered from centromere to telomere within the proximal arm of P5. For location of sY1207 in the proximal arm of P5 and for the locations of sY1227 and sY1228 in the center of P5, see figure 1*B*.
[a] For length and primers, see GenBank.
[b] Additional 2.7-kb deletion; for discussion, see table A3.

Table A5

High-Resolution Breakpoint Mapping in the Proximal Arm of Palindrome P4

STS	Length (bp)	Forward Primer	Reverse Primer	Results for WHT2825
sY116	…[a]	…[a]	…[a]	+
10767/8	291	TGGTTGAAACCAGATCTCCA	TGAGCCCAGATGAACAAGTG	+
10765/6	289	ATCCTTTCAGCAATTGTGGC	ACCATGGAAAGGAAATCTCG	+
10815/6	254	TGCACATTGGTGTGGAAATC	CTACACACAGGCCTTGATCG	+
10817/8	192	GCCCTCCCACACTGTCATAA	ATGGGGCTTCACTGGAAAAA	+
10819/20	244	GGGGAAATAACAGCAAAGCA	AGAAGAGGACAACAAGCCCA	+
10887/8	475	ACCTGACTCGGATGAACACC	TTATGCTCAAAACAACCCAGG	+
10951/2	809	TTTTTAGATCCCTGGGGGC	AAGACCACAAGGCAAAGGG	−
10953/4	707	CCTTTTGCCTTGTAATCTTGC	AATGCAGGGACATCAAAAGC	−
10955/6	695	CTTCCTGTTTTGCCTTCCTG	GAGACAGCAGCTAACCGTCC	−
10889/90	191	GGGCTTAAGGGGTTTACTGC	GCCTTGCTAACACAAGAGGC	−
10891/2	296	TTCATTGCACCTGCATGTCT	AAGCCATGGGAGTCACAGAG	−
sY118	…[a]	…[a]	…[a]	−

NOTE.—STSs are ordered from centromere to telomere within the proximal arm of P4. For the locations of sY116 and sY118 within the proximal arm of P4, see figure 1*B*.

[a] For length and primers, see GenBank.

Table A6

High-Resolution Breakpoint Mapping in the Distal Arm of Palindrome P1

STS	Length (bp)	Forward Primer	Reverse Primer	Results for WHT-2943	WHT-3410	WHT-3516	WHT-3642	WHT-4426	WHT-4486	AMC-0111	WHT-2825
sY1257	…[a]	…[a]	…[a]	−	−	−	−	−	−	−	−
sY1289	…[a]	…[a]	…[a]		−			−	−	+	
10982/3	381	GACATGGTCTCATCATATAGGCTG	CTAAAAACAACTGTAAGTTGCAGGT		−					+	
11116/7	286	CTGCATTAAGTGGCTAGTGAGTTT	GAGTAAGAAAAAAAGAAGAGTGCATTT		−					+	
11112/3	173	TTTCCCTGCAATTATAATGTGTGT	GTTTGCCACAGCCACTCTCT		−					+	
10769/70	284	TTAGGGACCACCAGGATCAG	GGAGAGGATAGGCAAGCTCC	−	−	−	−			+	−
sY1290	…[a]	…[a]	…[a]		−			−	−	+	
10988/9	170	GCGGTGTATTGGCAGATTTT	CAGTGATTCAGCATTGCAAGA		+						
10990/1	368	GGATGAAATTAATCCCATGAATT	ATGTAAAAAATTAGCCAGTACCGG		+						
10829/30	500	CAGCTTCCCAGGTTCATGC	CCAGAGTGACAGTGGCTCCAT	−	+						−
10893/4	155	TGAACATAAAAGAGAGACATCAAAAC	AAAGCATACTCAAGTAATACAGCCA	−	+						−
10905/6	826	GAGGTGGCCACCCAAGC	GACTTCCTTCATGAGTCACACAGC	−	+						−
10909/10	203	CCTACTGCACAGGATGTTGCA	AGACGGATAATCTGTGTGGGG	−	+						−
10895/6	129	ATAATCTTTGAGATCTGCTACGATG	AAGTTGCTACATGATTTGAGAGCT	−	+						−
10957/8	637	GGCAAATCGGTCTGCTTTTA	CTGGAAGGTTTTGGGTTGAA		+						−
10837/8	298	TGGCTTCCCTTACTGTGGTC	CTCCCATGTCCCGCTAATAA	−							+
10968/9	438	AGATCAACGAGACAGAAAGTTAAGAAC	TACACATTTCCCTCTACATGCTGA	−							
10970/1	449	TCTACCAGACGTACAAGGTGGAC	ATTTGTCATGGATATCCCATCC	−				−	−		
10839/40	793	AACGCCAAACACAATGACAA	TGACTTGTACAATGGCCGAA	+	+			−	−		+
10911/2	533	ATCAATCACAGGTTTGGTCTTCTC	CTGAAGCCCTCAGTGGAGAG	+	+						+
10755/6	415	ACTGTGGGTGAGACAGACCC	CAGAACCCACTTAAGGGCAA	+	+	−	−				+
10851/2	606	CCAGGCAGCCTCTGTTTTGT	CATGAGATGAGCCTCATGAACTAC			−	−				
10853/4	189	CCACTGTAAGTCTGACGGGG	GGATTTTGTCAGGGTAACTATGCC			−	−				
10913/4	480	AAAGATGCAGGTACTGCTCTGC	CGGGCTCTTTACGGTGATAT			−	−	−	−		
10915/6	242	TGAGACACTGATAATTGAGTGTCAATA	ATTGGCAAGTTTCATAATCTGCTA			+	+	+	+		
10917/8	148	CTGTTCTAATATCCTGTCACTTACGC	GCACACATTTCAGTTACTATAACCTCA			+	+				
10919/20	327	TTCCCAGCATCCAAAGTTCAT	AAGCTAGTTGATTAAAAACAAAGCAAC			+	+				
10921/2	279	GGATCTATGTAAAGAGGAGAGTTTCCT	TAGCTCTATGTATGAAGACGTGCAT			+	+				
10857/8	465	TGGTTTGATTGTCTATTATGCCA	CCATCTTGGCCTCATTAAGTGA			+	+				
10859/60	274	GGCATTGGGAAAATACAGGC	GCCCAGAAAAGCCCATAGAT			+	+				
10861/2	230	CTTTTCCACATTGTCTGTACTGATTT	GAGCAGCATTTTGGGCCT			+	+				
10761/2	340	CCACTTGACCCCTGTCACATC	GTCCATCAGCCTAGAACCTAGACT	+	+	+	+			+	+
sY639	…[a]	…[a]	…[a]	+	+	+	+	+	+	+	+

NOTE.—Within the distal arm of P1, sY1257 amplifies the proximal boundary of P1.1 (fig. 1*B*), and sY1289 and sY1290 amplify the two arm-to-spacer boundaries in P1.1. In these patients, STSs 10988/9 through 10837/8 are ordered from centromere to telomere in the distal arm of P1.1, and STSs 10968/9 through sY639 are ordered from centromere to telomere distal to P1.1.

[a] For length and primers, see GenBank.

Table A7

Primer Pairs Used for PCR Amplification of Deletion Junctions

Patient(s)	Primer Identification Numbers	Forward Primer	Reverse Primer	Deletion-Junction Product Size (bp)
AMC0110	10821/6	GGCAAGTTTCACAATCTGCTG	CAAAGATGCAGGTACTGCTCTGC	1,534
WHT4396	10827/12088	TCTTGATCAATGCCTTGTGAAAT	CCACTGTAAGTCTGACGGGG	740
WHT2943[a]	11189/90	ACTAAATGCCCACAGGAGAAAA	TTGTCATGGATATCCCATCCA	976
WHT3410[a]	11075/115	GGTCTTCTCACGTAATCCCG	TGTCCTGGATGGTAATGCC	754
WHT3516, WHT3642, WHT4426, and WHT4486	10824/916[b]	CAAAGATGCGGGTACTGCTCTAT	ATTGGCAAGTTTCATAATCTGCTA	1,537
WHT2825	11086/7	TGAGCATAAGTTTGTTTCGGG	CGAGGATGTCAATGTTTTGC	421

NOTE.—For PCR conditions, see sY1201 (GenBank).

[a] We first used long-range PCR (table A8) and then, based on the sequence of the long-range PCR product, designed the primers shown in this table.

[b] Annealing at 63°C.

Table A8

Primer Pairs Used for Long-Range PCR Amplification of Deletion Junctions

Patient	Primer Identification Numbers	Forward Primer	Reverse Primer	Deletion-Junction Product Size (kb)
WHT2943	11130/1	GTAATGCCATTGAGGCAAAGTTGAGTTAAAA	CCACACTGACTTGTACAATGGCCGAA	6.7[a]
WHT3410	11126/7	GGCCTGTCTTTGTAGGTTGGCCTACA	AATTTTTGTCAGTGATTCAGCATTGCAAGA	2.3

NOTE.—Long-range PCR conditions were as follows: enzyme and buffer, Advantage2 kit (Clontech) using manufacturer's buffer; primer concentration, 1 μM each; cycling, 95°C for 1 min followed by 30 cycles of 95°C for 0.5 min, 68°C for 12 min.

[a] Product includes a 2.7-kb deletion proximal to WHT2943's P4/distal-P1 deletion; see table A3.

Electronic-Database Information

Accession numbers and URLs for data presented herein are as follows:

GenBank, http://www.ncbi.nlm.nih.gov/Genbank/ (for previously published STSs sY105 [accession number G11994], sY116 [accession number G66528], sY117 [accession number G11996], sY118 [accession number G66529], sY127 [accession number G11998], sY143 [accession number G38347], sY149 [accession number G73322], sY254 [accession number G38349], sY579 [accession number G63909], sY639 [accession number G67162], sY1190 [accession number G67165], sY1192 [accession number G67166], sY1197 [accession number G67168], sY1201 [accession number G67170], and sY1206 [accession number G67171]; for new STSs, sY1207 [accession number G72341], sY1224 [accession number G72342], sY1227 [accession number G72343], sY1228 [accession number G72344], sY1257 [accession number G72345], sY1264 [accession number G72346], sY1289 [accession number G73323], sY1290 [accession number G73324], and sY1291 [accession number G72340]; for deletion junctions in AMC0110 [accession number AF395664], WHT3516 [accession number AF395665], WHT3642 [accession number AF395666], WHT3410 [accession number AF395667], WHT2825 [accession number AF395668], WHT2943 [accession number AF395669], WHT4396 [accession number AF437293], WHT4426 [accession number AF480412], and WHT4486 [accession number AF480413]; for reference sequence of the euchromatic, male-specific region of the human Y-chromosome long arm [accession numbers NT_011875 and NT_011903]; and for genes and transcripts affected by P5/P1 deletions, *CDY2* [accession number AF080598], *XKRY* [accession number AF000997], *HSFY* [accession number AF332226], *CYorf14* [accession number AF119903], *CYorf15A* [accession number AF332224], *CYorf15B* [accession number AF332225], *SMCY* [accession number U52191], *EIF1AY* [accession number AF000987], *RPS4Y2* [accession number AF497481], *RBMY1* [accession number X76060], *PRY* [accession number AF000988], *BPY2* [accession number AF000980], *DAZ* [accession number U21663], *CDY1* [accession number AF000981], *CSPG4LY* [accession number AF332228], *GOLGA2LY* [accession number AF332229], *TTTY9* [accession number AF332238], *TTTY14* [accession number AF332243], *TTTY10* [accession number AF332239], *TTTY13* [accession number AF332242], *TTTY6* [accession number AF332237], *TTTY5* [accession number AF332236], *TTTY4* [accession number AF332231], and *TTTY3* [accession number AF332230])

The Human Y Chromosome: Annotated Sequence of the *AZFc* Palindromic Complex, http://staffa.wi.mit.edu/page/Y/azfc/ (for dot-plot code)

Online Mendelian Inheritance in Man (OMIM), http://www.ncbi.nlm.nih.gov/Omim/ (for *AZFc* [MIM 400024] and Smith-Magenis syndrome [MIM 182290])

References

Akgun E, Zahn J, Baumes S, Brown G, Liang F, Romanienko PJ, Lewis S, Jasin M (1997) Palindrome resolution and recombination in the mammalian germ line. Mol Cell Biol 17: 5559–5570

Aradhya S, Bardaro T, Galgoczy P, Yamagata T, Esposito T, Patlan H, Ciccodicola A, Munnich A, Kenwrick S, Platzer M, D'Urso M, Nelson DL (2001) Multiple pathogenic and benign genomic rearrangements occur at a 35 kb duplication involving the *NEMO* and *LAGE2* genes. Hum Mol Genet 10:2557–2567

Blanco P, Shlumukova M, Sargent CA, Jobling MA, Affara N, Hurles ME (2000) Divergent outcomes of intrachromosomal recombination on the human Y chromosome: male infertility and recurrent polymorphism. J Med Genet 37:752–758

Brandell RA, Mielnik A, Liotta D, Ye Z, Veeck LL, Palermo GD, Schlegel PN (1998) *AZFb* deletions predict the absence of spermatozoa with testicular sperm extraction: preliminary report of a prognostic genetic test. Hum Reprod 13:2812–2815

Collick A, Drew J, Penberth J, Bois P, Luckett J, Scaerou F, Jeffreys A, Reik W (1996) Instability of long inverted repeats within mouse transgenes. EMBO J 15:1163–1171

David G, Jouannet P, Martin-Boyce A, Spira A, Schwartz D (1979) Sperm counts in fertile and infertile men. Fertil Steril 31:453–455

de Vries JWA, Hoffer MJV, Repping S, Hoovers JMN, Leschot NJ, van der Veen F (2002) Reduced copy number of *DAZ* genes in subfertile and infertile men. Fertil Steril 77:68–75

Eichler EE (2001) Segmental duplications: what's missing, misassigned, and misassembled—and should we care? Genome Res 11:653–656

Elliott DJ (2000) *RBMY* genes and *AZFb* deletions. J Endocrinol Invest 23:652–658

Elliott DJ, Millar MR, Oghene K, Ross A, Kiesewetter F, Pryor J, McIntyre M, Hargreave TB, Saunders PTK, Vogt PH, Chandley AC, Cooke H (1997) Expression of *RBM* in the nuclei of human germ cells is dependent on a critical region of the Y chromosome long arm. Proc Natl Acad Sci USA 94:3848–3853

Ferlin A, Moro E, Garolla A, Foresta C (1999) Human male infertility and Y chromosome deletions: role of the AZF-candidate genes *DAZ, RBM* and *DFFRY.* Hum Reprod 14: 1710–1716

Friel A, Houghton JA, Maher M, Smith T, Noel S, Nolan A, Egan D, Glennon M (2001) Molecular detection of Y chromosome microdeletions: an Irish study. Int J Androl 24:31–36

Girardi SK, Mielnik A, Schlegel PN (1997) Submicroscopic deletions in the Y chromosome of infertile men. Hum Reprod 12:1635–1641

Hull MG, Glazener CM, Kelly NJ, Conway DI, Foster PA, Hinton RA, Coulson C, Lambert PA, Watt EM, Desai KM (1985) Population study of causes, treatment, and outcome of infertility. Br Med J (Clin Res Ed) 291:1693–1697

Ji Y, Eichler EE, Schwartz S, Nicholls RD (2000) Structure of chromosomal duplicons and their role in mediating human genomic disorders. Genome Res 10:597–610

Kamp C, Hirschmann P, Voss H, Huellen K, Vogt PH (2000) Two long homologous retroviral sequence blocks in proximal Yq11 cause *AZFa* microdeletions as a result of intrachromosomal recombination events. Hum Mol Genet 9: 2563–2572

Krausz C, Quintana-Murci L, Barbaux S, Siffroi JP, Rouba H, Delafontaine D, Souleyreau-Therville N, Arvis G, Antoine JM, Erdei E, Taar JP, Tar A, Jeandidier E, Plessis G, Bourgeron T, Dadoune JP, Fellous M, McElreavey K (1999) A high frequency of Y chromosome deletions in males with nonidiopathic infertility. J Clin Endocrinol Metab 84:3606–3612

Kuroda-Kawaguchi T, Skaletsky H, Brown LG, Minx PJ, Cordum HS, Waterston RH, Wilson RK, Silber S, Oates R, Rozen S, Page DC (2001) The *AZFc* region of the Y chromosome features massive palindromes and uniform recurrent deletions in infertile men. Nat Genet 29:279–286

Lin YM, Chen CW, Sun HS, Hsu CC, Chen JM, Lin SJ, Lin JS, Kuo PL (2000) Y-chromosome microdeletion and its effect on reproductive decisions in Taiwanese patients presenting with nonobstructive azoospermia. Urology 56:1041–1046

Lobachev KS, Stenger JE, Kozyreva OG, Jurka J, Gordenin DA, Resnick MA (2000) Inverted *Alu* repeats unstable in yeast are excluded from the human genome. EMBO J 19: 3822–3830

Ma K, Sharkey A, Kirsch S, Vogt P, Keil R, Hargreave TB, McBeath S, Chandley AC (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

MacLeod J, Gold RZ (1951) The male factor in fertility and infertility. II. Spermatozoon counts in 1000 men of known fertility and in 1000 cases of infertile marriage. J Urol 66: 436–449

MacLeod J, Wang Y (1979) Male fertility potential in terms of semen quality: a review of the past, a study of the present. Fertil Steril 31:103–116

Martinez MC, Bernabe MJ, Gomez E, Ballesteros A, Landeras J, Glover G, Gil-Salom M, Remohi J, Pellicer A (2000) Screening for AZF deletion in a large series of severely impaired spermatogenesis patients. J Androl 21:651–655

Maurer B, Gromoll J, Simoni M, Nieschlag E (2001) Prevalence of Y chromosome microdeletions in infertile men who consulted a tertiary care medical centre: the Munster experience. Andrologia 33:27–33

Mazzarella R, Schlessinger D (1998) Pathological consequences of sequence duplications in the human genome. Genome Res 8:1007–1021

Oates RD, Silber S, Brown LG, Page DC. Clinical characterization of 42 oligospermic or azoospermic men with microdeletion of the *AZFc* region of the Y chromosome, and of 18 children conceived via intracytoplasmic sperm injection. Hum Reprod (in press)

Reijo R, Lee TY, Salo P, Alagappan R, Brown LG, Rosenberg M, Rozen S, Jaffe T, Strauss D, Hovatta O, de la Chapelle A, Silber S, Page DC (1995) Diverse spermatogenic defects in humans caused by Y chromosome deletions encompassing a novel RNA-binding protein gene. Nat Genet 10:383–393

Rozen S, Skaletsky H (2000) Primer3 on the WWW for general

users and for biologist programmers. In: Krawetz S, Misener S (eds) Bioinformatics methods and protocols: methods in molecular biology. Humana Press, Totowa, NJ, pp 365–386

Saxena R, de Vries JWA, Repping S, Alagappan RK, Skaletsky H, Brown LG, Ma P, Chen E, Hoovers JMN, Page DC (2000) Four *DAZ* genes in two clusters found in *AZFc* region of human Y chromosome. Genomics 67:256–267

Shaffer LG, Lupski JR (2000) Molecular mechanisms for constitutional chromosomal rearrangements in humans. Ann Rev Genet 34:297–329

Simoni M, Gromoll J, Dworniczak B, Rolf C, Abshagen K, Kamischke A, Carani C, Meschede D, Behre HM, Horst J, Nieschlag E (1997) Screening for deletions of the Y chromosome involving the *DAZ* (Deleted in AZoospermia) gene in azoospermia and severe oligozoospermia. Fertil Steril 67: 542–547

Stankiewicz P, Lupski JR (2002) Genome architecture, rearrangements and genomic disorders. Trends Genet 18:74–82

Sun C, Skaletsky H, Rozen S, Gromoll J, Nieschlag E, Oates R, Page DC (2000) Deletion of azoospermia factor a (*AZFa*) region of human Y chromosome caused by recombination between HERV15 proviruses. Hum Mol Genet 9:2291–2296

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

Van Landuyt L, Lissens W, Stouffs K, Tournaye H, Liebaers I, Van Steirteghem A (2000) Validation of a simple Yq deletion screening programme in an ICSI candidate population. Mol Hum Reprod 6:291–297

Vogt PH, Edelmann A, Kirsch S, Henegariu O, Hirschmann P, Kiesewetter F, Köhn FM, Schill WB, Farah S, Ramos C, Hartmann M, Hartschuh W, Meschede D, Behre HM, Castel A, Nieschlag E, Weidner W, Gröne HJ, Jung A, Engel W, Haidl G (1996) Human Y chromosome azoospermia factors (AZF) mapped to different subregions in Yq11. Hum Mol Genet 5:933–943

World Health Organization (1992) WHO laboratory manual for the examination of human semen and sperm-cervical mucus interaction. Cambridge University Press, Cambridge, United Kingdom

Yen PH (1998) A long-range restriction map of deletion interval 6 of the human Y chromosome: a region frequently deleted in azoospermic males. Genomics 54:5–12

——— (2001) The fragility of fertility. Nat Genet 29:243–244

Zuckerman Z, Rodriguez-Rigau LJ, Smith KO, Steinberger E (1977) Frequency distribution of sperm counts in fertile and infertile males. Fertil Steril 28:1310–1313