Diverse Outcomes of Homologous Recombination in the Human Y Chromosome

by

Julian H. Lange

Bachelor of Science, Biochemistry
McGill University, 1997

Diplôme d'Etudes Approfondies, Bases fondamentales de l'oncogenèse
Université Paris 7, 1999

SUBMITTED TO THE DEPARTMENT OF BIOLOGY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY IN BIOLOGY

AT THE

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

FEBRUARY 2008

Signature of author _____
Department of Biology
January 14, 2008

Certified by _____
David C. Page
Professor of Biology
Howard Hughes Medical Institute
Thesis Supervisor

Accepted by _____
Stephen P. Bell
Professor of Biology
Chair, Biology Graduate Student Committee

Diverse Outcomes of Homologous Recombination in the Human Y Chromosome

by

Julian H. Lange

Submitted to the Department of Biology on January 14, 2008
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Biology

Abstract

Mammalian sex chromosomes began diverging from an ordinary pair of autosomes roughly 300 million years ago. Inversions in the evolving Y chromosome sequentially suppressed recombination with the X chromosome. While pseudoautosomal regions in the human Y chromosome still participate regularly in allelic homologous recombination, the male-specific region of the Y (MSY) – the only haploid portion of the nuclear genome – does not. It does, however, engage in non-allelic homologous recombination.

In this thesis, I examine modes and outcomes of non-allelic homologous recombination in the MSY. The predictions presented here are based on the double-strand break repair model of recombination between homologous chromosomes, in which a double-strand break (DSB) is the common precursor to crossing over and gene conversion. First, I show that massive MSY-specific palindromes, which maintain arm-to-arm sequence identity via gene conversion, are also the targets of crossing over. Crossover events in palindromes can lead to isochromosome formation and diverse reproductive disorders including sex reversal, male infertility, and Turner syndrome. Second, I demonstrate that a region of the MSY – thought to be recombinationally suppressed with the X chromosome – does undergo extensive X-Y gene conversion. This region encompasses hotspots of ectopic crossover events that lead to X-Y translocations associated with sex reversal syndromes.

Although sequences in the MSY engage in productive recombination via gene conversion, alternative resolution of DSBs by crossing over can produce evolutionary "dead ends".

Thesis Supervisor: David C. Page
Title: Professor of Biology, Howard Hughes Medical Institute

To my sisters

Acknowledgements

This thesis represents a seven-year journey, from my arrival at MIT through to my thesis defense. I have been accompanied on this journey by colleagues, friends, and family, whose support in my endeavor has emerged in many forms and reached me in different ways.

My thesis advisor, David Page, has mentored me in the numerous aspects of being a scientist. Thank you, David, for inviting close observation of your distinguished mode of science, for joining me in quadruple-checking that all i's are dotted and all t's are crossed, and, most importantly, for consistently encouraging me to produce work of which I have become very proud.

Helen Skaletsky and I have worked as a team since my first day. Helen, you have been a source of knowledge, support, understanding, and guidance. Your contributions to my learning and to this thesis have been immense.

The Page lab, constantly evolving in its cast of characters, has been a special place. The relationships built with my labmates have been truly wonderful, and I am grateful to each of you for the connections we have made. I would like to especially thank Jake Mueller, for your gentle friendship, innumerable moments of laughter, and valuable assistance with editing this thesis; Laura Brown, for the generosity of your kindness and for ready technical advice; Yanfeng Lim, for sharing chocolate, providing dance lessons, and inspiring me with your personality; Jessica Alföldi, for your sensibility and dry humor; Jana Koubová, for your kind heart and cheerful spirit; and my baymate Winston Bellott, for your calm presence at my back.

I would also like to thank the members of my thesis committee – Angelika Amon, David Housman, Iain Cheeseman, and Hunt Willard – for the interest you have taken in and enthusiasm you have shown for my work; my friends at the Academic Medical Center in Amsterdam, for the hearty welcome you gave me into your lab; and my classmates Andreas Hochwagen and Ezequiel Alvarez Saavedra, for your friendship and our squash matches.

In addition to learning within the lab, the past seven years have been marked by personal growth outside of MIT. My friends Javier Barrientos, Dean Ventola, Erin Maneri, Jim Shanahan, Brad Carver, Paul Murphy, Miguel de Baca, Natalie Boelman, Edoardo Saccenti, and Thierry Mousseigne have been sources of great impact in this sphere. You have created space for me to become who I am. You have made me laugh, watched me flail, stood by me, and accepted my love for you.

I thank R.G., for your patience with one step back amid two steps forward.

I thank Pierre Basiaux, for unlocking.

Finally, I thank my family – Mami, Papi, Nina, Maya – for your constant love.

Contents

Chapter 1

Introduction

Julian Lange

Among the human chromosomes, the Y chromosome is extraordinary in that its presence determines the most commonly recognized phenotype: sex. While males carry the chromosome, females do not[1]. The male-specific region of the Y chromosome (MSY) is also unique compared to all other nuclear DNA in that it is haploid and transmitted clonally. Despite these remarkable characteristics, the prevailing view of the Y chromosome in recent decades has been one of a genetic wasteland.

Human sex chromosomes co-evolved from an ordinary pair of sexually recombining autosomes beginning roughly 300 million years ago (mya)[2,3]. Large inversions in the proto-Y chromosome led to the sequential suppression of crossing over with the proto-X chromosome[3]. Increasingly bereft of recombination, the MSY suffered a dramatic reduction in size and has retained only a handful of its ancestral genes[4]. These findings have led to speculation that the Y chromosome will disappear in 10-15 million years[5,6].

Two axes of research have restored legitimacy to the Y chromosome. First, studies of aberrant Y chromosomes have identified MSY genes and their importance in gonadal sex reversal, male infertility, and germ cell tumor formation[7-27]. Second, sequence characterization has shown that the MSY engages in productive homologous recombination in the form of intrachromosomal gene conversion[4,28]. A feature shared by the two bodies of work upon which these findings are based is the prominent role of naturally occurring partial Y-chromosome deletions. For example, two decades ago, the hunt for the male-determining gene on the Y relied on the analysis of partial Y-deletions in sex-reversed individuals[15]. Similarly, partial deletions played an important role in the construction of physical maps of the Y chromosome[29-31]. These maps were subsequently used to build a scaffold of genomic clones for sequence characterization[4,32].

The work presented in this thesis explores the relationship between homologous recombination and partial deletions in the Y chromosome. The following introduction will establish a context in which to examine this relationship by: 1) briefly presenting Y-chromosome evolution, sequence, and gene content; 2) describing Y-chromosome deletions, their roles in early MSY sequence characterization and gene identification, and homology-mediated ectopic recombination as a prominent mechanism of deletion; and

3) reviewing the canonical double-strand break repair (DSBR) model of homologous recombination and literature on the distribution of sites of recombination in eukaryotic genomes. These topics will form the basis of my hypotheses on the role of non-allelic recombination in: 1) generating Y isochromosomes associated with reproductive disorders, and 2) modulating the divergence of formerly allelic sequences in the X and Y chromosomes.

The human Y chromosome: evolution, sequence, and gene content

Mammalian sex chromosomes have evolved from an ordinary pair of sexually recombining autosomes that began to diverge roughly 300 mya[2,3]. It has been proposed that acquisition of the male-determining gene *SRY* initiated the Y chromosome's differentiation from the X chromosome (reviewed in ref.[33]). Male-benefit alleles may have subsequently arisen on the proto-Y chromosome and been permanently linked to *SRY* via an inversion in the Y chromosome. This was evolutionarily advantageous because it ensured immediate cessation of crossing over with the X chromosome, enabling all Y-linked genes to segregate only in males. Following this initial inversion, the proto-Y chromosome underwent additional inversions that sequentially suppressed crossing over with the proto-X chromosome[3]. A lack of recombination led to the decay of the evolving Y chromosome via evolutionary processes such as genetic hitchhiking, Mullers's ratchet, and background selection (reviewed in ref.[34]). Today, meiotic recombination between the human X and Y chromosomes is restricted to pseudoautosomal regions (PARs) located at the termini of each chromosome. While the X chromosome participates fully in recombination during female meiosis, the 95% of the Y chromosome that is restricted to males – the MSY – is always haploid and is thought to not undergo any form of productive recombination with the X chromosome.

The evolutionary history of the Y chromosome is revealed by comparing its sequence to large, contiguous regions – or strata – in the X chromosome[3,4,35]. By comparing formerly allelic X- and Y-linked gene pairs, Lahn and Page[3] found that their divergence was correlated with position along the X

chromosome. Each gene pair began diverging once the region of the Y chromosome harboring the Y-linked member underwent inversion. Lahn and Page were therefore able to reconstruct the major evolutionary events of the Y chromosome's 300-million-year history and identified four strata in the X chromosome. Ross and colleagues[35] subsequently examined the divergence of genomic X and Y sequences in the youngest stratum and demonstrated that it could be parsed into two strata. Each stratum corresponds to one of five inversions in the Y chromosome that took place 240-320 mya, 130-170 mya, 80-130 mya, 38-44 mya, and 29-32 mya.

The recent sequencing of the human MSY revealed that it is not exclusively comprised of sequences that survive from the ancient autosomes. Rather, it is a patchwork of euchromatic sequences with different evolutionary origins and of heterochromatic sequences (Figures 1A and 1B)[4]. The euchromatic MSY is about 23 megabases (Mb), less than half of the Y chromosome's estimated total length of 59 Mb[36]. It contains 18 single-copy genes and nine MSY-specific gene families, which collectively encode roughly 78 protein-coding genes (Figure 1C). Whereas all 18 single-copy genes have a known homolog in the X chromosome, five of the nine multi-copy genes do not. Five large blocks of heterochromatin, each containing lengthy tandem arrays of short (5- to 171-basepair) sequence units, are located in the centromere and in the long arm (Yq). They range in size from about 200 kilobases (kb) (*DYZ17* arrays in the centromere) to more than 30 Mb (*DYZ18/DYZ1/DYZ2* arrays in distal Yq), and some of these arrays are polymorphic[4,37,38]. None of the 78 identified protein-coding genes are located in heterochromatic sequences.

Most of the euchromatic sequences fall into one of three sequence classes: X-transposed, X-degenerate, and ampliconic (Figure 1B)[4]. The 3.4 Mb of X-transposed sequences are 99% identical to sequences in Xq because they were brought to the Y chromosome roughly 3-4 mya via a single X-to-Y transposition event[39]. An inversion subsequently disrupted the transposed sequences such that they are now found in two non-contiguous blocks on the short arm (Yp)[40]. Two single-copy genes are located in X-transposed sequences. The X-degenerate sequences are the 300-million-year-old remains of the autosomes from which the X and Y chromosomes evolved. The 8.6 Mb of X-degenerate sequences,

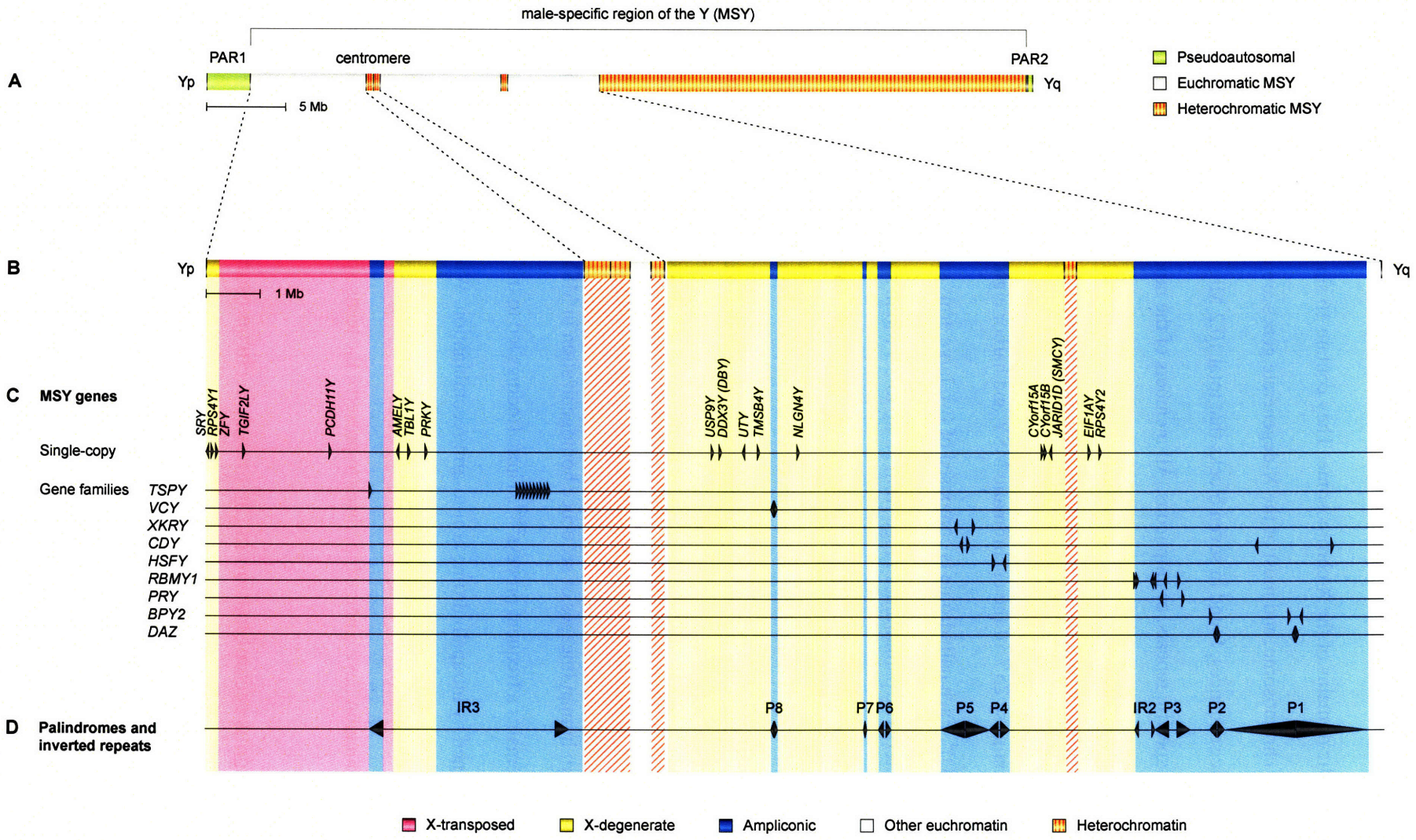
Figure 1. The MSY, the male-specific region of the human Y chromosome.

(**A**) Schematic of the entire Y chromosome, including pseudoautosomal regions (PAR1 and PAR2) in green and heterochromatic regions in orange.

(**B**) Expanded view of MSY euchromatin, highlighting three euchromatic sequence classes: X-transposed (pink), X-degenerate (yellow), and ampliconic (blue).

(**C**) MSY protein-coding genes and gene families; arrowheads indicate 5′-to-3′ orientation.

(**D**) Locations of eight palindromes (P1 through P8) and two inverted repeats (IR2 and IR3).


male-specific region of the Y (MSY)
A
PAR1
centromere
PAR2
Yp
Yq
5 Mb
Pseudoautosomal
Euchromatic MSY
Heterochromatic MSY
B
Yp
Yq
1 Mb
C
MSY genes
SRY
RPS4Y1
ZFY
TGIF2LY
PCDH11Y
AMELY
TBL1Y
PRKY
USP9Y
DDX3Y (DBY)
UTY
TMSB4Y
NLGN4Y
CYorf15A
CYorf15B
JARID1D (SMCY)
EIF1AY
RPS4Y2
Single-copy
Gene families
TSPY
VCY
XKRY
CDY
HSFY
RBMY1
PRY
BPY2
DAZ
D
Palindromes and inverted repeats
IR3
P8
P7
P6
P5
P4
IR2
P3
P2
P1
X-transposed
X-degenerate
Ampliconic
Other euchromatin
Heterochromatin

distributed among eight non-contiguous blocks on Yp and Yq, share varying levels of sequence identity with formerly allelic domains of the X chromosome. They contain 16 single-copy genes, each with a homolog on the X chromosome, and most of the X-degenerate genes are expressed in a wide range of tissues. Ampliconic sequences, found in seven blocks that total 10.2 Mb, are composed mainly of large, nearly identical Y-specific repeats, or amplicons. All members of the nine MSY-specific gene families are located in ampliconic sequences and are expressed predominantly or exclusively in the testis.

A striking feature of the ampliconic sequence class is the presence of eight massive palindromes that collectively comprise 25% of the euchromatic MSY and harbor eight of the nine testis-specific gene families (Figure 1D)[4]. Each palindrome is composed of two large arms separated by a small, unique "spacer" at the center. The palindromes range from 30 kb to 2.9 Mb in size and from 99.94% to 99.997% in arm-to-arm sequence identity. In addition, there are two large inverted repeats that display 99.95% and 99.75% sequence identity (Figure 1D). Molecular clock assumptions based on arm-to-arm sequence identity would indicate that the palindromic repeats arose through duplication events less than 100,000 years ago. However, Y-linked orthologues of these palindromes are found with equally high arm-to-arm sequence identity in chimpanzees[28], which last shared a common ancestor with humans over 6 mya[41]. Evidence in modern human lineages of sequence homogenization in MSY palindromes indicates that gene conversion remains a frequent process[28,42]. These findings led to the hypothesis that the palindromes have been maintained by intrapalindrome arm-to-arm recombination. Whereas much of the Y chromosome's evolutionary past has been spent eliminating recombination, palindromes that undergo productive recombination have emerged, perhaps to preserve the testis-specific genes located there.

Ectopic homologous recombination in the human Y chromosome

There are two broad classes of naturally occurring partial Y-chromosome deletions (Figure 2). The first is comprised of Y isochromosomes and X-Y translocations, which present as apparent terminal deletions (see legend of Figure 2). This class of deletions served as a tool to construct Y-chromosome physical

maps[29-31]. Known Y-specific loci were ordered and oriented by systematically testing for their presence or absence in genomic DNAs of individuals with such Y-chromosome anomalies. These loci were subsequently employed to characterize the second class: interstitial deletions that have played an important role in defining regions in Yq that are associated with male infertility[21-26].

Microscopically visible structural abnormalities of the Y chromosome

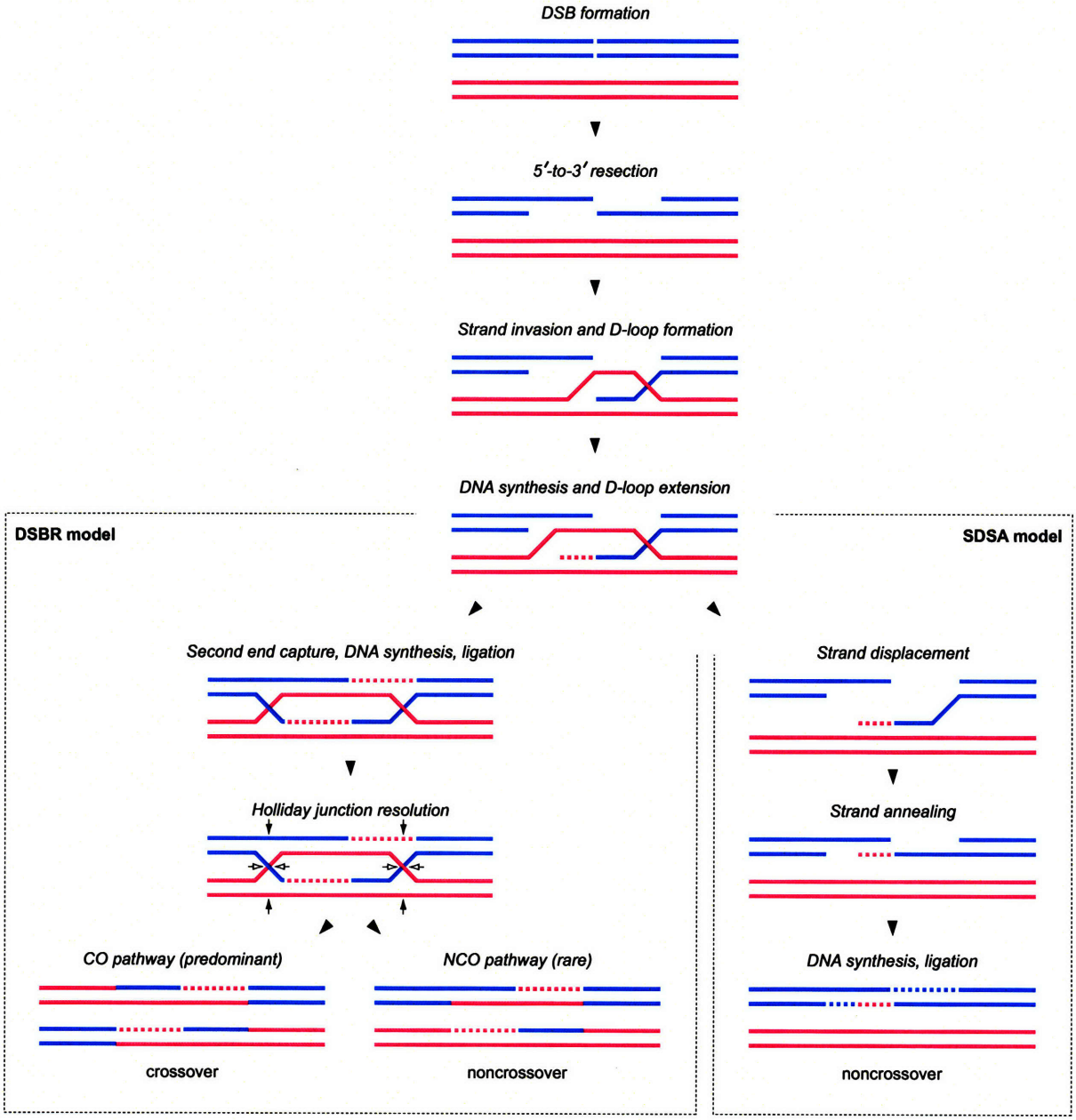
In 1959, studies by Jacobs[43] and by Ford[44] provided the first evidence for the role of the human Y chromosome in male sex determination. This finding spurred interest in the identification of Y-chromosome structural abnormalities. Initial descriptions of aberrant Y chromosomes were based on microscopic investigations (reviewed in ref.[45]). Cytogenetically, the Y chromosome could be identified by its small size and the presence of a large block of heterochromatin on Yq. Despite substantial variation in the length of this heterochromatin[37], it provided an important landmark.

Y isochromosomes were among the first structural abnormalities identified, found in males displaying dysgenetic gonads or features of Turner syndrome[46-48]. These chromosomes are composed of mirror-image duplications due to: 1) duplication of one arm and loss of the other arm, or 2) complete duplication of one arm and the centromere and partial duplication of the other arm (Figure 2). Jacobs and Ross[49] reported a particularly insightful finding of two females carrying Yq isochromosomes – the two copies of Yq were identified by the presence of a mirror-image duplication of the large heterochromatic block. These two females lacked Yp, enabling the authors to conclude the location of the male-determining gene. Research leading to the eventual discovery of *SRY* on distal Yp over two decades later relied on a different class of sex chromosome anomalies: X-Y translocations found in genetic disorders that interfere with gonad development.

Figure 2. Structural abnormalities of the human Y chromosome.

(A) Structures of anomalous Y chromosomes. Above, Y isochromosomes and X-Y translocations. Below, interstitial deletions.

(B) By PCR analysis of plus/minus STS assays, Y isochromosomes and X-Y translocations appear to be terminally deleted as compared with interstitial deletions. Solid black bars encompass MSY DNA that would be detected by PCR in each deletion. Notably, of the most distal MSY-specific sequences in Yp and Yq, only one is present in Y isochromosomes and X-Y translocations, whereas both are present in interstitial deletions.


DSB formation
5′-to-3′ resection
Strand invasion and D-loop formation
DNA synthesis and D-loop extension
DSBR model
SDSA model
Second end capture, DNA synthesis, ligation
Holliday junction resolution
CO pathway (predominant)
NCO pathway (rare)
crossover
noncrossover
Strand displacement
Strand annealing
DNA synthesis, ligation
noncrossover

Sex reversal syndromes due to homologous crossing over between X and Y chromosomes

Individuals with XX male syndrome, first described in 1964[50], are phenotypic males that are infertile and have small testes. Conversely, individuals with XY female syndrome do not develop female secondary sexual characteristics at puberty, do not menstruate, and have underdeveloped gonads[51]. XX males arise at a frequency of 1/20,000, while XY females occur more rarely[52]. By karyotyping, XX males appear to carry two normal X chromosomes and no Y chromosome, and XY females appear to carry one X chromosome and one intact Y chromosome.

If the Y chromosome determines male development, how could such XX males and XY female arise? When evidence for homology between the human X and Y chromosomes first surfaced in 1965[53], Ferguson-Smith[54] hypothesized that XX males could arise out of X-Y interchange during the first meiotic division. He proposed that homology-mediated crossing over might transfer Y-linked male-determining loci to an X chromosome, resulting in a gamete carrying an X-Y translocation. Subsequent experimentation validated this hypothesis. First, a series of observations demonstrated that in most XX males, a portion of the Y-specific DNA is translocated to one of the X chromosomes (Figure 2)[7-10,12,13]. Conversely, in XY females, a segment of the Y chromosome is replaced with a small segment of the X chromosome[11,14]. Second, the hunt for the male-determining gene spawned efforts to precisely determine the Y-chromosome DNA content of XX males and XY females. By deletion-mapping the Y chromosomes of XX males and XY females, researchers narrowed down the region present in XX males but absent in XY females, leading to the identification of *SRY*, the sex-determining gene[15]. Finally, amplification and sequencing of X-Y junction fragments isolated from XX males and XY females demonstrated that the X-Y translocations were due to crossing over between highly homologous loci in the X and Y chromosomes[16,20].

Interstitial deletions in Yq generated by ectopic homologous recombination

The distal Yq heterochromatic block proved to serve a second important purpose. In 1976, its absence in six men with azoospermia (i.e., sterile, no detectable sperm in semen) provided the first evidence for a

locus necessary for spermatogenesis in Yq[55]. Since this heterochromatin is not required for male fertility[56]. Tiepolo and Zuffardi inferred that these deletions in Yq included euchromatic sequences. They proposed the existence of an "*Azoospermia Factor*" (*AZF*) region adjacent to the Yq heterochromatin.

The identification of candidate *AZF* genes would only occur 15 years later, after physical mapping of the Y chromosome. Such maps served two purposes in characterizing the *AZF* region[57-61]. First, they enabled the detection, in men with spermatogenic failure, of submicroscopic interstitial deletions in Yq. Second, they were used to define the extent of deletion in each partially deleted Y chromosome. Analysis of the *AZF* region eventually led to the discovery of two spermatogenesis genes, *RBMY1* and *DAZ*[17,62].

Whereas initial studies assumed the presence of a single *AZF* locus, subsequent work demonstrated the presence of multiple *AZF* regions. In screening 370 infertile or subfertile men, Vogt and colleagues[19] observed 12 microdeletions that mapped to three regions in Yq. One of these subregions coincided with the previously identified *AZF* locus. The other two subregions were mapped proximally within Yq. The subregions, named *AZFa*, *AZFb*, and *AZFc*, were associated with different spermatogenic defects: deletion of *AZFa* results in azoospermia due to the complete absence of germ cells in the testis; deletion of *AZFb* results in azoospermia due to arrest of germ cells in meiosis; and deletion of *AZFc* results in either azoospermia with no germ cells in the testis or in low sperm production associated with subfertility.

Precise DNA-sequence characterization of *AZFa*, *AZFb*, and *AZFc* required the nucleotide sequence of the MSY. Mapping had previously hinted at the presence of many Y-specific repeats[31]; refinement of these deletions would likely be difficult in the absence of more sequence information. The *AZFa* region, spanning approximately 800 kb and containing two genes, was the first large, contiguous Y-chromosome segment to be sequenced[63]. With the availability of the complete MSY sequence[4], *AZFa*, *AZFb*, and *AZFc* deletions were defined at the nucleotide level (Figure 2), making it possible to determine the repertoires of genes lost and to gain insight into mechanisms of deletion[21-26,64,65].

Studies of *AZFa*, *AZFb*, and *AZFc* demonstrated that homologous recombination between direct repeats is a prominent cause of interstitial deletion in Yq. *AZFa* deletions arise by recombination between two 10-kb HERV15 retroviral sequence blocks that exhibit 94% sequence identity[21-23]. Similarly, most *AZFb* deletions are due to homologous recombination between palindromes P5 and P1[25]. Since highly homologous sequences in direct orientation lie in proximal and distal arms of palindomes P5 and P1, there are in fact two distinct types of *AZFb* deletion that are due to recombination: 1) "P5-proximal P1" deletions between the proximal arms of P5 and of P1 remove 6.2 Mb of sequence and 21 protein-coding genes; and 2) "P5-distal P1" deletions between the distal arms of P5 and of P1 remove 7.7 Mb and 25 protein-coding genes. *AZFc* deletions, the most common known cause of spermatogenic failure, arise via homologous recombination between two 229-kb direct repeats b2 and b4 that share 99.9% sequence identity. The deletions eliminate 3.5 Mb of ampliconic sequence and nine protein-coding genes[24].

The finding that the *AZFc* region is ampliconic and comprises numerous large Y-specific repeats in direct and inverted orientation prompted researchers to predict further homology-mediated deletions[66]. Indeed, three such deletions were identified and named after the amplicons that were targeted in each case: "gr/gr", "b1/b3", and "b2/b3"[26,64,65]. A systematic survey of 47 Y chromosomes encompassing worldwide diversity uncovered additional duplications and inversions that all evidently formed by homologous recombination[38]. Similarly, recently reported recurrent deletions in Yp are flanked by direct repeats[67].

The discovery of palindromes in the human MSY overturned the notion that productive recombination in the Y chromosome is limited to its two PARs[28]. Prior to this finding, the MSY was considered recombinationally inert. Whereas the palindromes provide evidence of productive recombination in the MSY, research into partial Y-chromosomal deletions has demonstrated the MSY's participation in ectopic homologous recombination. First, homology-mediated exchange between formerly allelic regions of the X and Y chromosomes – leading to X-Y translocations – was identified as a cause of sex reversal syndromes (i.e., XX males and XY females)[7-14,16,20]. Second, intrachromosomal homologous recombination between Y-specific amplicons – leading to interstitial deletions, duplications,

and inversions – emerged as a cause of structural variation in the MSY[21-26,38,64,65,67]. This raises the question: At what level are the events of gene conversion in palindromes and ectopic crossing over between amplicons related?

In order to probe the connection between productive and ectopic recombination in the Y chromosome, one must understand the mechanism of recombination. The most well understood such process is meiotic recombination between homologous chromosomes.

Meiotic homologous recombination

Gametes, the fruits of meiosis, differ from their progenitor cells in two fundamental ways. First, gametes are haploid, the products of a meiotic program in which diploid germ cells undergo one round of DNA replication followed by two successive rounds of cell division. Second, gametes acquire chromosomes that are comprised of novel assortments of parental alleles. Fusion of two gametes during sexual reproduction results in the reconstitution of a diploid chromosome complement and in offspring that are genetically distinct from their parents.

In sexually reproducing organisms, meiotic recombination lies at the heart of these events by serving to properly segregate homologous chromosomes and to promote genetic exchange between them (reviewed in ref.[68]). During the first – or reductional – division of meiosis, transient breaks introduced into each chromosome are repaired by genetic exchange with a homologous chromosome, creating new alleles and novel combinations of alleles on which selection can act. The physical connections established between each pair of homologs ensure their correct migration to opposite spindle poles. Models of how genetic exchange takes place have developed over the past 25 years.

Models of homologous recombination

Szostak and colleagues[69] first laid out the double-strand break repair (DSBR) model of homologous recombination in 1983, countering the prevailing model at the time. Meselson and Radding[70] had

previously proposed that recombination between homologs is initiated by a single-stranded nick in one of the two interacting chromosomes. However, additional observations favored an alternative model of DSBR in which genetic exchange between homologs is initiated by the deliberate introduction of double-strand breaks (DSBs).

In this model, DSBs are processed by either: 1) the crossover (CO) pathway, which results in crossing over, the reciprocal exchange of chromosome arms that flank the site; or 2) the noncrossover (NCO) pathway, which can result in gene conversion (Figure 3)[69,71,72]. After DSB formation, each of the two 5′ ends of the DSB is resected to expose 3′ single-stranded DNA (ssDNA) tails. One of the two 3′ ssDNA tails invades the homologous sequence of an intact, double-stranded homologous chromosome (i.e., non-sister chromatid), resulting in dissociation of the DNA duplex to form a displacement (D)-loop. Repair synthesis of resected DNA is primed by each of the two 3′ ssDNA tails and templated by the dissociated strands of the invaded homolog. Synthesis promotes D-loop extension in the invaded homolog, branch migration, and ligation to form two Holliday junctions. Non-random cleavage of each Holliday junction preferentially yields crossovers, although resolution occasionally results in noncrossovers. Gene conversion occurs via mismatch repair of heteroduplex DNA.

The DSBR model predicts that crossovers and noncrossovers should appear with similar timing. However, control of the two pathways is differentially timed[71,73]. To account for this paradox, the model of synthesis-dependent strand annealing (SDSA) has been proposed as the source of the majority of noncrossovers (reviewed in ref.[74]). As in the DSBR model, SDSA begins with DSB formation, resection of the two 5′ DNA ends to form two 3′ ssDNA tails, invasion of the intact, double-stranded homologous chromosome by one of the 3′ ssDNA tails, and D-loop formation (Figure 3). In contrast, the ensuing repair synthesis of resected DNA is predominantly conservative. Nascent DNA synthesis of one 3′ ssDNA tail uses a dissociated strand of the invaded DNA duplex as template. However, after synthesis has proceeded past the point of the DSB, the elongated 3′ ssDNA tail reanneals with the complementary strand. The two strands then use each other as templates for completion of synthesis, limiting genetic

exchange to gene conversion. Since the SDSA model forms noncrossovers only, this could account for the observed difference in levels of crossovers and noncrossovers.

Sites of recombination events: distribution and sequence features

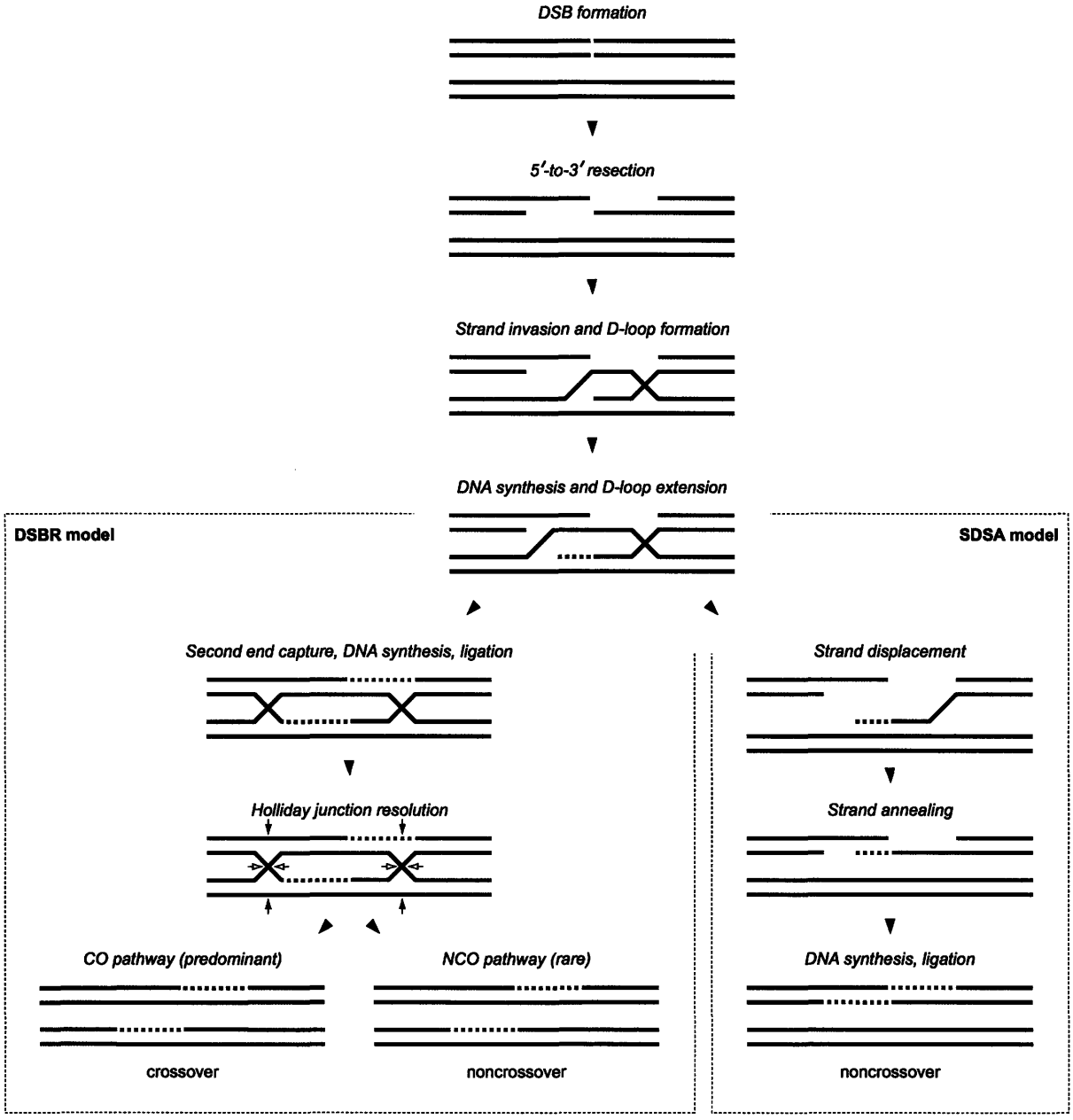
Meiotic recombination events are distributed nonrandomly in eukaryotic genomes. Locus-specific and genome-wide analyses in organisms from yeast to human have revealed that most recombination events take place in highly localized genomic domains termed "hotspots" and that the bulk of DNA is recombinationally "cold" (reviewed in ref.[75]). Such punctate distributions of recombination effectively dissociate physical and genetic linkage. This observation leads to several questions. What DNA sequence contexts govern the distribution of hotspots and coldspots? What are the characteristics of hotspots? What effect does recombination exert on the evolution of hotspots?

No single primary DNA-sequence feature or structural element has been identified as an absolute determinant of recombination hotspots in yeast. Nevertheless, several general conclusions have been drawn from analyses of DSB sites (reviewed in ref.[76]). Meiotic DSBs occur preferentially in intergenic promoter-containing regions rather than within genes[77-79]. Transcriptional activity may play a role in unfolding DNA. An open chromatin environment, as assayed by increased sensitivity to nucleases, correlates with increased frequency of recombination[77]. Furthermore, at some sites, hotspot activity is dependent on the binding of specific transcription factors[80-82]. Finally, DSBs are associated with G+C-rich regions, as revealed by a global analysis of hotspots in open reading frames[79]. Conversely, centromeres are coldspots of recombination[78].

In humans, two methodologies have recently been employed to analyze the footprints of recombination and to discover hotspots. Population genetic studies have examined historical crossover events to gain insight into the distribution of recombination events. These studies have both recapitulated the finding from yeast that recombination takes place at preferred sites[83-86] and identified features specific to hotspots in the human genome[87,88]. Myers and colleagues[88] analyzed polymorphism data to examine

Figure 3. Models of homologous recombination (adapted from refs.[73,89]).

A double-strand break (DSB) is resected to expose 3′ ssDNA tails. In the double-strand break repair (DSBR) model, these 3′ ssDNA tails invade the homolog and initiate DNA synthesis. Capture of the second end, DNA synthesis, and ligation generate Holliday junctions flanking the DSB site. This intermediate is resolved by cutting the outside strands (filled arrows) or crossed strands (open arrows) of each junction, leading predominantly to crossovers and occasionally to noncrossovers. Two of the four possible resolutions are shown here. In the synthesis-dependent strand annealing (SDSA) model, strand invasion occurs on only one side of the DSB. The newly synthesized DNA strand is displaced and anneals with the other DSB end. Break repair is completed by DNA synthesis and ligation. In this mechanism, only noncrossovers are formed. In both DSBR and SDSA models, gene conversion occurs by mismatch repair of heteroduplex DNA.


DSB formation
5′-to-3′ resection
Strand invasion and D-loop formation
DNA synthesis and D-loop extension
DSBR model
SDSA model
Second end capture, DNA synthesis, ligation
Holliday junction resolution
CO pathway (predominant)
NCO pathway (rare)
crossover
noncrossover
Strand displacement
Strand annealing
DNA synthesis, ligation
noncrossover

fine-scale recombination rates across the human genome and estimated that it contains 25,000 to 50,000 recombination hotspots, an average density of approximately one per 50 kb. In examining the relationship between genes and recombination, they found that, in contrast to the yeast genome where transcriptionally active regions are more highly recombinogenic, recombination rates in the human genome are on average lower within the transcribed portions of genes. A genome-wide comparison of the sequence composition of hotspots and coldspots demonstrated that hotspots are enriched for THE1A and THE1B elements, the long terminal repeats of two retrovirus-like retroposons, as well as for the seven-nucleotide motif CCTCCCT. Certain long interspersed nuclear elements (LINEs) are underrepresented in hotspots relative to coldspots. How these sequence features affect recombination rates is not understood.

A limited number of present-day crossover hotspots have been determined by high-resolution sperm typing, a PCR-based approach for recovering crossover molecules directly from sperm DNA[90-96]. Mapping of hotspots in this manner has shown that crossover events at each hotspot cluster within narrow regions of 1-2 kb[93]. For the most part, such modern hotspots correlate well with historical crossover hotspots[93]. Sperm typing assays have shown that even between men, crossover frequencies at a specific hotspot can vary substantially[97].

Two lines of evidence suggest that individual crossover hotspots are evolving rapidly. Sperm-typing assays have uncovered *bona fide* crossover hotspots concealed within regions of low historical recombination, evidence that emerging hotspots may not be captured in polymorphism data[98]. Second, orthologous regions in humans and chimpanzees that exhibit sequence similarities of ~99% do not share the same patterns of recombination[99,100]. Taken together, these results imply the continual evolution of the recombinogenic potential of hotspots.

Since all CO and NCO pathways derive from a common precursor, the formation of a DSB, hotspots of allelic crossing over and gene conversion should coincide. This expectation has been confirmed in yeast, mouse, and human[101-103]. Does this prediction hold true for recombination in the MSY?

Framework for a hypothesis

Two features of the DNA-sequence of the human Y chromosome convey opposing themes of its evolutionary history[4,28]. First, sequences with varying degrees of identity to the X chromosome mark the evolutionary decay of the Y chromosome. Second, a collection of massive Y-specific palindromes that harbor a suite of critical spermatogenesis genes and are maintained by gene conversion conversely augur a mode of preservation. Thus, the *suppression* of sexual recombination with the X chromosome, driven by functional specialization of the Y chromosome, has been counterbalanced by the *emergence* of productive recombination within palindromes. The significance of intrapalindrome recombination was recently highlighted when the region of the human Y chromosome that is restricted to males was rechristened from "non-recombining region of the Y", or NRY, to "male-specific region of the Y", or MSY[4].

The availability of the MSY's nucleotide sequence has led to insights into its gene content, function, and structure[4,28,38]. In addition, it has resulted in an understanding of the phenotypic effects and modes of origin of partial deletions in the Y chromosome. To date, several interstitial deletions have been defined at the DNA-sequence level, and ectopic homologous recombination between direct repeats has been identified as a prominent mechanism of MSY deletion[21-26,38,64,65,67]. However, the apparent terminal deletions, which served as the basis for the physical maps, have remained largely unexplored.

This thesis examines the role of homology-driven recombination in generating apparent terminal deletions in the human Y chromosome. More generally, it investigates the relationship between recombination in the MSY and the model of recombination between homologous chromosomes. The essence of this model is the processing of a double-strand break (DSB) by either the crossover (CO) pathway, leading to crossing over, or the noncrossover (NCO) pathway, which can lead to gene conversion[69,71-73]. The work presented here rests on two related suppositions: 1) that apparent terminal deletions are the products of a mechanism analogous to the CO pathway; and 2) that the maintenance of arm-to-arm sequence identity in MSY palindromes parallels the NCO pathway. We conjectured that since

a DSB is a common precursor to the CO and NCO pathways, there might be a relationship between sites of crossing over and gene conversion in the MSY.

To test this hypothesis, we investigated two territories within the MSY that are recurrent substrates for recombination. Arm-to-arm sequence identities of MSY palindromes are maintained by gene conversion[28]. We asked whether DSBs in MSY palindromes could sometimes be resolved by a CO-like pathway, leading to Y isochromosome formation. Conversely, we studied sex-reversal-associated X-Y translocations that are due to homology-mediated crossing over between formerly allelic domains of the X and Y chromosomes[7-14,16,20]. Could the initiating DSBs of these events occasionally be channeled into an NCO-like pathway, leading to gene conversion between non-allelic segments of the X and Y chromosomes?

We sought to address these questions by first identifying Y isochromosomes and X-Y translocations arising spontaneously in human populations. We began by reanalyzing genomic DNA samples from 2,380 individuals previously studied in our laboratory[17,24,25,29,31,104]. These individuals fall into two broad classes of ascertainment. The first class consists of 830 individuals obtained by either karyotypic identification of an aberrant Y chromosome, or discordance between sex chromosome constitution and sex phenotype. The second class consists of 1,550 men with spermatogenic failure, a phenotype often associated with anomalies of the Y chromosome. Based on our hypothesis, we formed specific predictions for how Y isochromosomes and X-Y translocations arise, and screened the 2,380 DNA samples to identify informative cases.

In order to systematically and efficiently determine the DNA-sequence content of each case of interest, we employed a catalog of Y-specific markers, or sequence-tagged sites (STSs), that are amplified by PCR. With each STS assay, we determined the presence or absence, in each human genomic DNA sample tested, of a specific site along the Y chromosome. During the mapping and sequencing of the human Y chromosome, and the refinement of deletion breakpoints – including those presented in this thesis – we and our colleagues developed 1287 MSY-specific STSs at an average spacing of less than one STS per 14 kb[4,17,21,25,31,32,38]. In Chapter 2 of this thesis, I describe MSY Breakpoint Mapper, an online,

publicly available database of these 1287 STSs[105]. This database, designed to facilitate mapping of partial deletions in the MSY, served to assist the characterization of deletions described in Chapter 3 and Chapter 4.

Acknowledgements

I thank J. Mueller, J. Alföldi, and D.C. Page for their insightful comments.

References

1. Painter TS (1921) The Y-chromosome in mammals. Science 53:503-504

2. Ohno S (1967) Sex chromosomes and sex-linked genes. Springer-Verlag, Berlin

3. Lahn BT, Page DC (1999) Four evolutionary strata on the human X chromosome. Science 286:964-967

4. Skaletsky H, Kuroda-Kawaguchi T, Minx PJ, Cordum HS, Hillier L, Brown LG, Repping S, et al. (2003) The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. Nature 423:825-837

5. Aitken RJ, Marshall Graves JA (2002) The future of sex. Nature 415:963

6. Graves JA, Koina E, Sankovic N (2006) How the gene content of human sex chromosomes evolved. Curr Opin Genet Dev 16:219-224

7. Guellaen G, Casanova M, Bishop C, Geldwerth D, Andre G, Fellous M, Weissenbach J (1984) Human XX males with Y single-copy DNA fragments. Nature 307:172-173

8. Page DC, de la Chapelle A, Weissenbach J (1985) Chromosome Y-specific DNA in related human XX males. Nature 315:224-226

9. Affara NA, Ferguson-Smith MA, Tolmie J, Kwok K, Mitchell M, Jamieson D, Cooke A, Florentin L (1986) Variable transfer of Y-specific sequences in XX males. Nucleic Acids Res 14:5375-5387

10. Andersson M, Page DC, de la Chapelle A (1986) Chromosome Y-specific DNA is transferred to the short arm of X chromosome in human XX males. Science 233:786-788

11. Disteche CM, Casanova M, Saal H, Friedman C, Sybert V, Graham J, Thuline H, Page DC, Fellous M (1986) Small deletions of the short arm of the Y chromosome in 46,XY females. Proc Natl Acad Sci USA 83:7841-7844

12. Petit C, de la Chapelle A, Levilliers J, Castillo S, Noel B, Weissenbach J (1987) An abnormal terminal X-Y interchange accounts for most but not all cases of human XX maleness. Cell 49:595-602

13. Page DC, Brown LG, de la Chapelle A (1987) Exchange of terminal portions of X- and Y-chromosomal short arms in human XX males. Nature 328:437-440

14. Levilliers J, Quack B, Weissenbach J, Petit C (1989) Exchange of terminal portions of X- and Y-chromosomal short arms in human XY females. Proc Natl Acad Sci USA 86:2296-2300

15. Sinclair AH, Berta P, Palmer MS, Hawkins JR, Griffiths BL, Smith MJ, Foster JW, Frischauf AM, Lovell-Badge R, Goodfellow PN (1990) A gene from the human sex-determining region encodes a protein with homology to a conserved DNA-binding motif. Nature 346:240-244

16. Weil D, Wang I, Dietrich A, Poustka A, Weissenbach J, Petit C (1994) Highly homologous loci on the X and Y chromosomes are hot-spots for ectopic recombinations leading to XX maleness. Nat Genet 7:414-419

17. Reijo R, Lee TY, Salo P, Alagappan R, Brown LG, Rosenberg M, Rozen S, Jaffe T, Straus D, Hovatta O, et al. (1995) Diverse spermatogenic defects in humans caused by Y chromosome deletions encompassing a novel RNA-binding protein gene. Nat Genet 10:383-393

18. Tsuchiya K, Reijo R, Page DC, Disteche CM (1995) Gonadoblastoma: molecular definition of the susceptibility region on the Y chromosome. Am J Hum Genet 57:1400-1407

19. Vogt PH, Edelmann A, Kirsch S, Henegariu O, Hirschmann P, Kiesewetter F, Kohn FM, Schill WB, Farah S, Ramos C, Hartmann M, Hartschuh W, Meschede D, Behre HM, Castel A, Nieschlag E, Weidner W, Grone 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

20. Schiebel K, Winkelmann M, Mertz A, Xu X, Page DC, Weil D, Petit C, Rappold GA (1997) Abnormal XY interchange between a novel isolated protein kinase gene, PRKY, and its homologue, PRKX, accounts for one third of all (Y+)XX males and (Y-)XY females. Hum Mol Genet 6:1985-1989

21. 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

22. 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

23. 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

24. 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

25. Repping S, Skaletsky H, Lange J, Silber S, Van Der Veen F, Oates RD, Page DC, Rozen S (2002) Recombination between palindromes P5 and P1 on the human Y chromosome causes massive deletions and spermatogenic failure. Am J Hum Genet 71:906-922

26. Repping S, Skaletsky H, Brown L, van Daalen SK, Korver CM, Pyntikova T, Kuroda-Kawaguchi T, de Vries JW, Oates RD, Silber S, van der Veen F, Page DC, Rozen S (2003) Polymorphism for a 1.6-Mb deletion of the human Y chromosome persists through balance between recurrent mutation and haploid selection. Nat Genet 35:247-251

27. Nathanson KL, Kanetsky PA, Hawes R, Vaughn DJ, Letrero R, Tucker K, Friedlander M, et al. (2005) The Y deletion gr/gr and susceptibility to testicular germ cell tumor. Am J Hum Genet 77:1034-1043

28. Rozen S, Skaletsky H, Marszalek JD, Minx PJ, Cordum HS, Waterston RH, Wilson RK, Page DC (2003) Abundant gene conversion between arms of palindromes in human and ape Y chromosomes. Nature 423:873-876

29. Vergnaud G, Page DC, Simmler MC, Brown L, Rouyer F, Noel B, Botstein D, de la Chapelle A, Weissenbach J (1986) A deletion map of the human Y chromosome based on DNA hybridization. Am J Hum Genet 38:109-124

30. Affara NA, Florentin L, Morrison N, Kwok K, Mitchell M, Cook A, Jamieson D, Glasgow L, Meredith L, Boyd E, et al. (1986) Regional assignment of Y-linked DNA probes by deletion mapping and their homology with X-chromosome and autosomal sequences. Nucleic Acids Res 14:5353-5373

31. Vollrath D, Foote S, Hilton A, Brown LG, Beer-Romero P, Bogan JS, Page DC (1992) The human Y chromosome: a 43-interval map based on naturally occurring deletions. Science 258:52-59

32. Tilford CA, Kuroda-Kawaguchi T, Skaletsky H, Rozen S, Brown LG, Rosenberg M, McPherson JD, Wylie K, Sekhon M, Kucaba TA, Waterston RH, Page DC (2001) A physical map of the human Y chromosome. Nature 409:943-945

33. Charlesworth D, Charlesworth B, Marais G (2005) Steps in the evolution of heteromorphic sex chromosomes. Heredity 95:118-128

34. Charlesworth B, Charlesworth D (2000) The degeneration of Y chromosomes. Philos Trans R Soc Lond B Biol Sci 355:1563-1572

35. Ross MT, Grafham DV, Coffey AJ, Scherer S, McLay K, Muzny D, Platzer M, et al. (2005) The DNA sequence of the human X chromosome. Nature 434:325-337

36. Morton NE (1991) Parameters of the human genome. Proc Natl Acad Sci USA 88:7474-7476

37. Bobrow M, Pearson PL, Pike MC, el-Alfi OS (1971) Length variation in the quinacrine-binding segment of human Y chromosomes of different sizes. Cytogenetics 10:190-198

38. Repping S, van Daalen SK, Brown LG, Korver CM, Lange J, Marszalek JD, Pyntikova T, van der Veen F, Skaletsky H, Page DC, Rozen S (2006) High mutation rates have driven extensive structural polymorphism among human Y chromosomes. Nat Genet 38:463-467

39. Page DC, Harper ME, Love J, Botstein D (1984) Occurrence of a transposition from the X-chromosome long arm to the Y-chromosome short arm during human evolution. Nature 311:119-123

40. Schwartz A, Chan DC, Brown LG, Alagappan R, Pettay D, Disteche C, McGillivray B, de la Chapelle A, Page DC (1998) Reconstructing hominid Y evolution: X-homologous block, created by X-Y transposition, was disrupted by Yp inversion through LINE-LINE recombination. Hum Mol Genet 7:1-11

41. Brunet M, Guy F, Pilbeam D, Mackaye HT, Likius A, Ahounta D, Beauvilain A, et al. (2002) A new hominid from the Upper Miocene of Chad, Central Africa. Nature 418:145-151

42. Adams SM, King TE, Bosch E, Jobling MA (2006) The case of the unreliable SNP: recurrent back-mutation of Y-chromosomal marker P25 through gene conversion. Forensic Sci Int 159:14-20

43. Jacobs PA, Strong JA (1959) A case of human intersexuality having a possible XXY sex-determining mechanism. Nature 183:302-303

44. Ford CE, Jones KW, Polani PE, De Almeida JC, Briggs JH (1959) A sex-chromosome anomaly in a case of gonadal dysgenesis (Turner's syndrome). Lancet 1:711-713

45. Jacobs PA (1969) Structural abnormalities of the sex chromosomes. Br Med Bull 25:94-98

46. Oikawa K, Blizzard RM (1961) Chromosomal studies of patients with congenital anomalies simulating those of gonadal aplasia, including a case of true gonadal and sex reversal. N Engl J Med 264:1009-1016

47. Klevit HD, Mellman WJ, Eberlein WR (1963) Triple mosaicism with an isochromosome derived from a partially deleted Y in a male pseudohermaphrodite. Pediatrics 32:56-62

48. Solomon IL, Hamm CW, Green OC (1964) Chromosome studies on testicular tissue cultures and blood leukocytes of a male previously reported to have no Y chromosome. N Engl J Med 271:586-592

49. Jacobs PA, Ross A (1966) Structural abnormalities of the Y chromosome in man. Nature 210:352-354

50. de la Chapelle A, Hortling H, Niemi M, Wennstroem J (1964) XX sex chromosomes in a human male. First case. Acta Med Scand Suppl 412:25-38

51. Swyer GIM (1955) Male pseudohermaphroditism: a hitherto undescribed form. Br Med J 2:709-712

52. de la Chapelle A (1981) The etiology of maleness in XX men. Hum Genet 58:105-116

53. Ferguson-Smith MA (1965) Karyotype-phenotype correlations in gonadal dysgenesis and their bearing on the pathogenesis of malformations. J Med Genet 39:142-155

54. Ferguson-Smith MA (1966) X-Y chromosomal interchange in the aetiology of true hermaphroditism and of XX Klinefelter's syndrome. Lancet 2:475-476

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

56. Borgaonkar DS, Hollander DH (1971) Quinacrine fluorescence of the human Y chromosome. Nature 230:52

57. Johnson MD, Tho SP, Behzadian A, McDonough PG (1989) Molecular scanning of Yq11 (interval 6) in men with Sertoli-cell-only syndrome. Am J Obstet Gynecol 161:1732-1737

58. Skare J, Drwinga H, Wyandt H, vanderSpek J, Troxler R, Milunsky A (1990) Interstitial deletion involving most of Yq. Am J Med Genet 36:394-397

59. Bardoni B, Zuffardi O, Guioli S, Ballabio A, Simi P, Cavalli P, Grimoldi MG, Fraccaro M, Camerino G (1991) A deletion map of the human Yq11 region: implications for the evolution of the Y chromosome and tentative mapping of a locus involved in spermatogenesis. Genomics 11:443-451

60. 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

61. Vogt P, Chandley AC, Hargreave TB, Keil R, Ma K, Sharkey A (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

62. Ma K, Inglis JD, Sharkey A, Bickmore WA, Hill RE, Prosser EJ, Speed RM, Thomson EJ, Jobling M, Taylor 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

63. Sun C, Skaletsky H, Birren B, Devon K, Tang Z, Silber S, Oates R, Page DC (1999) An azoospermic man with a de novo point mutation in the Y-chromosomal gene USP9Y. Nat Genet 23:429-432

64. Fernandes S, Paracchini S, Meyer LH, Floridia G, Tyler-Smith C, Vogt PH (2004) A large AZFc deletion removes DAZ3/DAZ4 and nearby genes from men in Y haplogroup N. Am J Hum Genet 74:180-187

65. Repping S, van Daalen SK, Korver CM, Brown LG, Marszalek JD, Gianotten J, Oates RD, Silber S, van der Veen F, Page DC, Rozen S (2004) A family of human Y chromosomes has dispersed throughout northern Eurasia despite a 1.8-Mb deletion in the azoospermia factor c region. Genomics 83:1046-1052

66. Yen P (2001) The fragility of fertility. Nat Genet 29:243-244

67. Jobling MA, Lo IC, Turner DJ, Bowden GR, Lee AC, Xue Y, Carvalho-Silva D, Hurles ME, Adams SM, Chang YM, Kraaijenbrink T, Henke J, Guanti G, McKeown B, van Oorschot RA, Mitchell RJ, de Knijff P, Tyler-Smith C, Parkin EJ (2007) Structural variation on the short arm of the human Y chromosome: recurrent multigene deletions encompassing Amelogenin Y. Hum Mol Genet 16:307-316

68. Roeder GS (1997) Meiotic chromosomes: it takes two to tango. Genes Dev 11:2600-2621

69. Szostak JW, Orr-Weaver TL, Rothstein RJ, Stahl FW (1983) The double-strand-break repair model for recombination. Cell 33:25-35

70. Meselson MS, Radding CM (1975) A general model for genetic recombination. Proc Natl Acad Sci USA 72:358-361

71. Börner GV, Kleckner N, Hunter N (2004) Crossover/noncrossover differentiation, synaptonemal complex formation, and regulatory surveillance at the leptotene/zygotene transition of meiosis. Cell 117:29-45

72. Guillon H, Baudat F, Grey C, Liskay RM, de Massy B (2005) Crossover and noncrossover pathways in mouse meiosis. Mol Cell 20:563-573

73. Allers T, Lichten M (2001) Differential timing and control of noncrossover and crossover recombination during meiosis. Cell 106:47-57

74. Pâques F, Haber JE (1999) Multiple pathways of recombination induced by double-strand breaks in Saccharomyces cerevisiae. Microbiol Mol Biol Rev 63:349-404

75. Kauppi L, Jeffreys AJ, Keeney S (2004) Where the crossovers are: recombination distributions in mammals. Nat Rev Genet 5:413-424

76. Petes TD (2001) Meiotic recombination hot spots and cold spots. Nat Rev Genet 2:360-369

77. Wu TC, Lichten M (1994) Meiosis-induced double-strand break sites determined by yeast chromatin structure. Science 263:515-518

78. Baudat F, Nicolas A (1997) Clustering of meiotic double-strand breaks on yeast chromosome III. Proc Natl Acad Sci USA 94:5213-5218

79. Gerton JL, DeRisi J, Shroff R, Lichten M, Brown PO, Petes TD (2000) Inaugural article: global mapping of meiotic recombination hotspots and coldspots in the yeast Saccharomyces cerevisiae. Proc Natl Acad Sci USA 97:11383-11390

80. White MA, Wierdl M, Detloff P, Petes TD (1991) DNA-binding protein RAP1 stimulates meiotic recombination at the HIS4 locus in yeast. Proc Natl Acad Sci USA 88:9755-9759

81. White MA, Dominska M, Petes TD (1993) Transcription factors are required for the meiotic recombination hotspot at the HIS4 locus in Saccharomyces cerevisiae. Proc Natl Acad Sci USA 90:6621-6625

82. Kon N, Krawchuk MD, Warren BG, Smith GR, Wahls WP (1997) Transcription factor Mts1/Mts2 (Atf1/Pcr1, Gad7/Pcr1) activates the M26 meiotic recombination hotspot in Schizosaccharomyces pombe. Proc Natl Acad Sci USA 94:13765-13770

83. Reich DE, Cargill M, Bolk S, Ireland J, Sabeti PC, Richter DJ, Lavery T, Kouyoumjian R, Farhadian SF, Ward R, Lander ES (2001) Linkage disequilibrium in the human genome. Nature 411:199-204

84. Daly MJ, Rioux JD, Schaffner SF, Hudson TJ, Lander ES (2001) High-resolution haplotype structure in the human genome. Nat Genet 29:229-232

85. Rioux JD, Daly MJ, Silverberg MS, Lindblad K, Steinhart H, Cohen Z, Delmonte T, et al. (2001) Genetic variation in the 5q31 cytokine gene cluster confers susceptibility to Crohn disease. Nat Genet 29:223-228

86. Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, Higgins J, DeFelice M, Lochner A, Faggart M, Liu-Cordero SN, Rotimi C, Adeyemo A, Cooper R, Ward R, Lander ES, Daly MJ, Altshuler D (2002) The structure of haplotype blocks in the human genome. Science 296:2225-2229

87. McVean GA, Myers SR, Hunt S, Deloukas P, Bentley DR, Donnelly P (2004) The fine-scale structure of recombination rate variation in the human genome. Science 304:581-584

88. Myers S, Bottolo L, Freeman C, McVean G, Donnelly P (2005) A fine-scale map of recombination rates and hotspots across the human genome. Science 310:321-324

89. Chen JM, Cooper DN, Chuzhanova N, Ferec C, Patrinos GP (2007) Gene conversion: mechanisms, evolution and human disease. Nat Rev Genet 8:762-775

90. Jeffreys AJ, Murray J, Neumann R (1998) High-resolution mapping of crossovers in human sperm defines a minisatellite-associated recombination hotspot. Mol Cell 2:267-273

91. Lien S, Szyda J, Schechinger B, Rappold G, Arnheim N (2000) Evidence for heterogeneity in recombination in the human pseudoautosomal region: high resolution analysis by sperm typing and radiation-hybrid mapping. Am J Hum Genet 66:557-566

92. Jeffreys AJ, Ritchie A, Neumann R (2000) High resolution analysis of haplotype diversity and meiotic crossover in the human TAP2 recombination hotspot. Hum Mol Genet 9:725-733

93. Jeffreys AJ, Kauppi L, Neumann R (2001) Intensely punctate meiotic recombination in the class II region of the major histocompatibility complex. Nat Genet 29:217-222

94. May CA, Shone AC, Kalaydjieva L, Sajantila A, Jeffreys AJ (2002) Crossover clustering and rapid decay of linkage disequilibrium in the Xp/Yp pseudoautosomal gene SHOX. Nat Genet 31:272-275

95. Cullen M, Perfetto SP, Klitz W, Nelson G, Carrington M (2002) High-resolution patterns of meiotic recombination across the human major histocompatibility complex. Am J Hum Genet 71:759-776

96. Jeffreys AJ, Neumann R (2005) Factors influencing recombination frequency and distribution in a human meiotic crossover hotspot. Hum Mol Genet 14:2277-2287

97. Jeffreys AJ, Neumann R (2002) Reciprocal crossover asymmetry and meiotic drive in a human recombination hot spot. Nat Genet 31:267-271

98. Jeffreys AJ, Neumann R, Panayi M, Myers S, Donnelly P (2005) Human recombination hot spots hidden in regions of strong marker association. Nat Genet 37:601-606

99. Winckler W, Myers SR, Richter DJ, Onofrio RC, McDonald GJ, Bontrop RE, McVean GA, Gabriel SB, Reich D, Donnelly P, Altshuler D (2005) Comparison of fine-scale recombination rates in humans and chimpanzees. Science 308:107-111

100. Ptak SE, Hinds DA, Koehler K, Nickel B, Patil N, Ballinger DG, Przeworski M, Frazer KA, Pääbo S (2005) Fine-scale recombination patterns differ between chimpanzees and humans. Nat Genet 37:429-434

101. Symington LS, Brown A, Oliver SG, Greenwell P, Petes TD (1991) Genetic analysis of a meiotic recombination hotspot on chromosome III of Saccharomyces cerevisiae. Genetics 128:717-727

102. Guillon H, de Massy B (2002) An initiation site for meiotic crossing-over and gene conversion in the mouse. Nat Genet 32:296-299

103. Jeffreys AJ, May CA (2004) Intense and highly localized gene conversion activity in human meiotic crossover hot spots. Nat Genet 36:151-156

104. Lahn BT, Ma N, Breg WR, Stratton R, Surti U, Page DC (1994) Xq-Yq interchange resulting in supernormal X-linked gene expression in severely retarded males with 46,XYq- karyotype. Nat Genet 8:243-250

105. Lange J, Skaletsky H, Bell GW, Page DC (2007) MSY Breakpoint Mapper, a database of sequence-tagged sites useful in defining naturally occurring deletions in the human Y chromosome. Nucleic Acids Res published online October 26

Chapter 2

MSY Breakpoint Mapper, a database of sequence-tagged sites useful in defining naturally occurring deletions in the human Y chromosome

Julian Lange, Helen Skaletsky, George W. Bell, David C. Page

Author contributions

J. Lange tested STS assays and designed the website layout. H. Skaletsky designed STSs and prepared data for the database. G.W. Bell performed all programming for the user interface. D.C. Page provided guidance throughout the project.

J. Lange and D.C. Page wrote the manuscript.

Summary

Y chromosome deletions arise frequently in human populations, where they cause sex reversal and Turner syndrome and predispose individuals to infertility and germ cell cancer. Knowledge of the nucleotide sequence of the male-specific region of the Y chromosome (MSY) makes it possible to precisely demarcate such deletions and the repertoires of genes lost, offering insights into mechanisms of deletion and the molecular etiologies of associated phenotypes. Such deletion mapping is usually conducted using polymerase chain reaction (PCR) assays for the presence or absence of a series of Y-chromosomal DNA markers, or sequence-tagged sites (STSs). In the course of mapping intact and aberrant Y chromosomes during the past two decades, we and our colleagues have developed robust PCR assays for 1287 Y-specific STSs. These PCR assays amplify 1698 loci at an average spacing of less than 14 kilobases across the MSY euchromatin. To facilitate mapping of deletions, we have compiled a database of these STSs, MSY Breakpoint Mapper (http://breakpointmapper.wi.mit.edu/). When queried, this online database provides regionally targeted catalogs of STSs and nearby genes. MSY Breakpoint Mapper is useful for efficiently and systematically defining the breakpoint(s) of virtually any naturally occurring Y chromosome deletion.

Introduction

Most boys and men carry the entirety of the Y chromosome, including the MSY, or male-specific region (Figure 1A)[1], while most girls and women carry none of it. However, some males and some females possess a portion but not all of the MSY. This is commonly the result of a translocation involving the Y chromosome and a second chromosome, or of an interstitial or other deletion within the Y chromosome[2-30]. Taken together, these naturally occurring deletions of the MSY are abundant, and they are diverse in structure and origin.

In recent decades, studies of these MSY deletions have propelled major advances in our understanding of Y chromosome function and structure. These advances have included the elucidation of

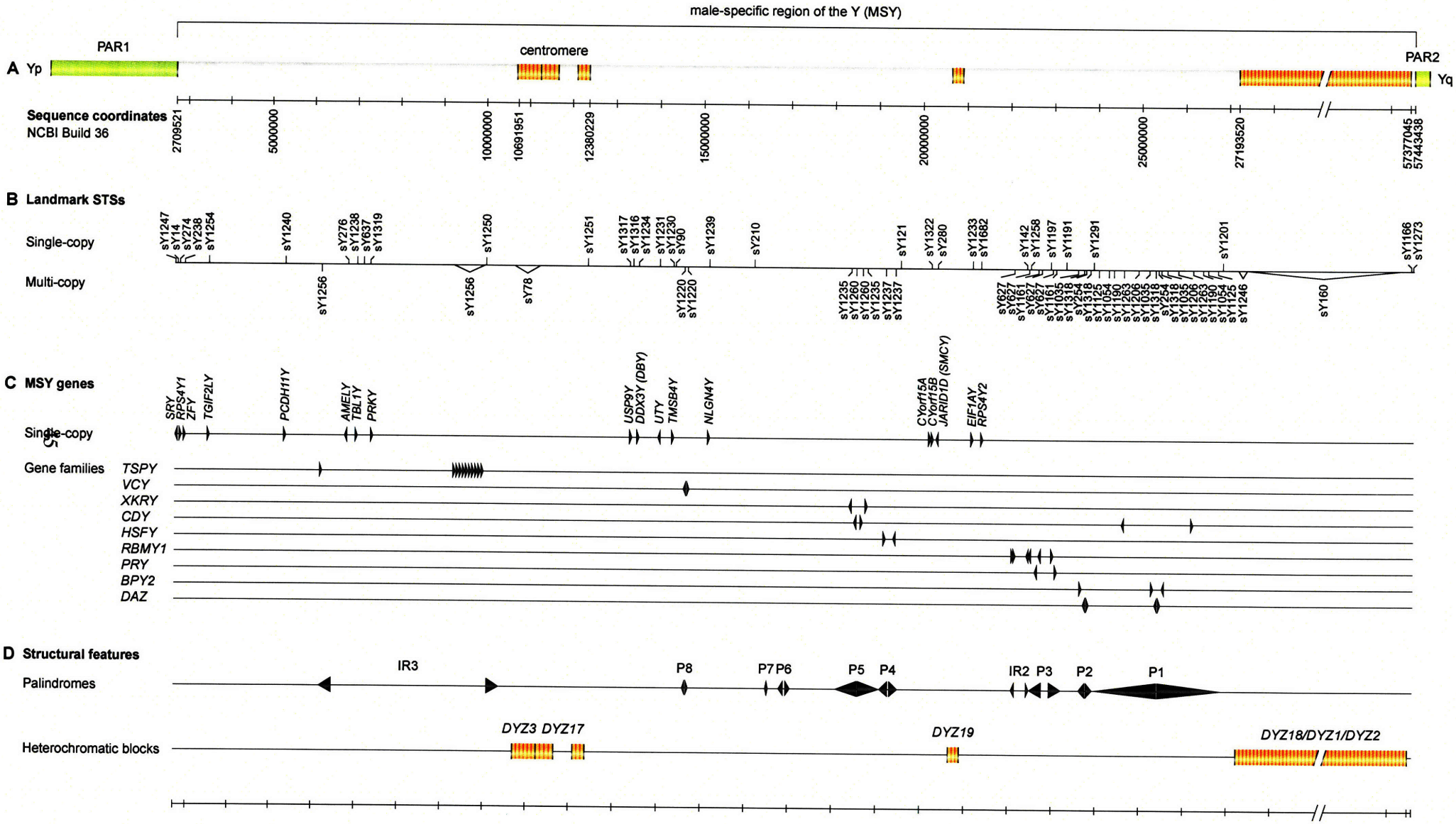
Figure 1. The MSY, the male-specific region of the human Y chromosome.

(A) Schematic of the entire Y chromosome, with pseudoautosomal regions (PAR1 and PAR2) in green, euchromatic regions of MSY in gray, and heterochromatic regions of MSY in orange. Immediately below are Y chromosome sequence coordinates based on NCBI Build 36.

(B) Locations of 51 landmark STSs.

(C) MSY protein-coding genes and gene families; arrowheads indicate 5′-to-3′ orientation.

(D) Major structural features.


male-specific region of the Y (MSY)
A
Yp
PAR1
centromere
PAR2
Yq
Sequence coordinates
NCBI Build 36
2709521
5000000
10000000
10691951
12380229
15000000
20000000
25000000
27193520
57377045
57443438
B
Landmark STSs
Single-copy
Multi-copy
sY1247
sY14
sY274
sY238
sY1254
sY1240
sY276
sY1238
sY637
sY1319
sY1250
sY1251
sY1317
sY1316
sY1234
sY1231
sY1230
sY90
sY1239
sY210
sY121
sY1322
sY280
sY1233
sY1682
sY142
sY1258
sY1197
sY1191
sY1291
sY1201
sY1166
sY1273
sY1256
sY1256
sY78
sY1220
sY1220
sY1235
sY1260
sY1260
sY1235
sY1237
sY1237
sY627
sY627
sY1161
sY627
sY627
sY1161
sY1035
sY1318
sY254
sY1318
sY1125
sY1054
sY1190
sY1206
sY1035
sY1318
sY254
sY1318
sY1035
sY1206
sY1263
sY1190
sY1054
sY1125
sY1246
sY160
C
MSY genes
Single-copy
Gene families
SRY
RPS4Y1
ZFY
TGIF2LY
PCDH11Y
AMELY
TBL1Y
PRKY
USP9Y
DDX3Y (DBY)
UTY
TMSB4Y
NLGN4Y
CYorf15A
CYorf15B
JARID1D (SMCY)
EIF1AY
RPS4Y2
TSPY
VCY
XKRY
CDY
HSFY
RBMY1
PRY
BPY2
DAZ
D
Structural features
Palindromes
IR3
P8
P7
P6
P5
P4
IR2
P3
P2
P1
Heterochromatic blocks
DYZ3
DYZ17
DYZ19
DYZ18/DYZ1/DYZ2

sex reversal syndromes (e.g., XX males and XY females)[4,5,8-13,15,21], the identification of the sex-determining gene *SRY*[31], the emergence of MSY deletions as the most common known genetic cause of spermatogenic failure[18,20,22-27], and the implication of MSY genes in the etiology of gonadoblastoma and testis cancer[19,32,33]. The first comprehensive maps of the human Y chromosome were based on naturally occurring deletions: Y-specific restriction fragments and, subsequently, Y-specific sequence-tagged sites (STSs) were ordered by testing for their presence or absence in genomic DNAs of series of individuals with partial Y chromosomes[6,7,14]. In turn, these deletion-mapped STSs were used to build scaffolds of overlapping recombinant DNA clones[34], ultimately leading to a tiling path of BAC clones for the sequencing of the MSY[1,35].

With the availability of the nucleotide sequence of the MSY[1], we and other investigators have employed a growing catalog of STSs in an expanded effort to identify MSY deletions and to understand their phenotypic effects and modes of origin. For example, several recurrent interstitial deletions associated with spermatogenic failure have recently been defined at the DNA-sequence level[22-27]. Such studies have underscored the roles of many MSY genes in reproduction and have identified ectopic homologous recombination as a prominent mechanism of MSY deletion.

Although a few such classes of MSY deletions have been examined at the DNA-sequence level, many others have yet to be thoroughly explored. These include translocations of the Y chromosome to the X chromosome or to an autosome, Y isochromosomes, and ring chromosomes[2,14,16,17]. Such structural abnormalities are associated with a wide array of phenotypes, including spermatogenic failure, manifestations of Turner syndrome, ambiguous external genitalia, gonadoblastoma, and delay in development and growth.

Detailed DNA-sequence analysis of uncharacterized MSY deletions is most readily accomplished with Y-specific STS assays, which employ the polymerase chain reaction (PCR). Each such STS assay provides a straightforward means of determining the presence or absence, in a sample of human genomic DNA, of a specific point along the length of the Y chromosome[14]. In the course of analyzing normal and aberrant Y chromosomes over the past two decades, we and our colleagues have generated 1287 robust,

Y-specific STSs. Here we present MSY Breakpoint Mapper (http://breakpointmapper.wi.mit.edu/), a database of these STSs and an interface for use in examining MSY deletions.

Database of STSs

The MSY Breakpoint Mapper database contains 1287 Y-specific STSs generated and tested during the past two decades[1,14,18,22,26,35,36]. Among these are 695 previously unreported STSs that our laboratory has recently deposited in GenBank. Each of the 1287 archived STSs is operationally defined by a PCR assay. In most cases, PCR primers were selected using Primer3[37]. Each of the 1287 PCR assays was empirically demonstrated to yield the expected product when normal male genomic DNAs were employed as templates, but not when normal female genomic DNAs were employed. Of these 1287 PCR assays, 1277 amplify single-copy or low-copy-number sequences found in the euchromatic portions of the MSY. The remaining 10 PCR assays amplify highly repetitive heterochromatic sequences found only in the Y chromosome (Table 1).

For 992 of the 1277 euchromatic STSs, the PCR assay amplifies a single, unique site in the MSY. Each of these 992 single-copy STSs is therefore useful in assaying the presence or absence of a unique point along the length of the Y chromosome. It is noteworthy that, for several dozen of these single-copy STSs, the unique target site exhibits sequence similarity to one or more other points along the Y chromosome. For each of these STSs, the corresponding PCR assay is nonetheless specific to the targeted locus, as shown using negative control templates of two kinds: 1) genomic DNAs from individuals with Y chromosomes known to be deleted for the STS target site, but also known to carry the sequence-similar locus, and 2) BAC clones lacking the STS target site, but carrying the sequence-similar locus. In each case, the PCR assay proved to be specific to the STS target site.

The remaining 285 (of 1277) euchromatic STSs are not single-copy but multi-copy. The MSY contains many lengthy, dispersed repeats or "amplicons" that exhibit >99.9% sequence identity. A PCR assay for such sequences will necessarily amplify multiple loci. For each of the 285 multi-copy STSs, the

Table 1. Numbers of PCR assays (STSs) and corresponding amplified loci archived in the database.

	Number of PCR assays (STSs)			Number of loci amplified			
Sequence type	Single-copy	Multi-copy	Total	Single-copy	Multi-copy	Total	Density
Euchromatin[a]	992	285	1277	992	706	1698	1 per 13.6 kb[c]
Heterochromatin[b]	NA	NA	10	NA	NA	NA	NA

[a]STSs that span euchromatin-heterochromatin boundaries are categorized here as deriving from euchromatin.

[b]Due to the repetitive nature of MSY heterochromatic sequences, it is not possible to determine the copy number of STSs that derive from such sequences or the total number of loci amplified.

[c]Calculation based on the MSY's 23 Mb of euchromatic sequence.

PCR assay amplifies identical or nearly identical sequences at two or more points along the length of the Y chromosome. Together, these 285 multi-copy PCR assays identify a total of 706 MSY loci.

In total, the PCR assays for 1277 euchromatic STSs – 992 single-copy and 285 multi-copy – amplify 1698 loci distributed across the MSY's 23 megabases (Mb) of euchromatic sequence. This corresponds to an average spacing of less than 14 kilobases (kb) between PCR-amplified loci (Table 1).

Most PCR assays cataloged in the database can be performed using one of two protocols:

1) Standard PCR

94°C for 3 min

35 cycles of: 94°C for 1 min; 61°C for 1 min; 72°C for 1 min

72°C for 5 min

2) Touchdown PCR

94°C for 1 min

20 cycles of: 92°C for 30 s; 70°C – 0.5°C/cycle for 40 s (i.e., a ramp from 70°C to 60.5°C with temperature decrements of 0.5°C per cycle)

20 cycles of: 92°C for 30 s; 60°C for 40 s + 1 s/cycle (i.e., a ramp from 40 s to 59 s with time increments of 1 s per cycle)

Several PCR assays can also be performed using alternative protocols, as described in the associated NCBI GenBank entries.

Querying the database

The query input webpage (http://breakpointmapper.wi.mit.edu/mapper.html) features a schematic of the human Y chromosome that includes nucleotide sequence coordinates, landmark STSs, protein-coding genes, and major structural features of the MSY (Figure 1). Sequence coordinates (Figure 1A) are based on NCBI Build 36, an assembly of the human genome sequence that incorporates all known human MSY sequence[1,38]. The 51 landmark STSs (Figure 1B) consist of sY1247 and sY1273, located at the boundaries

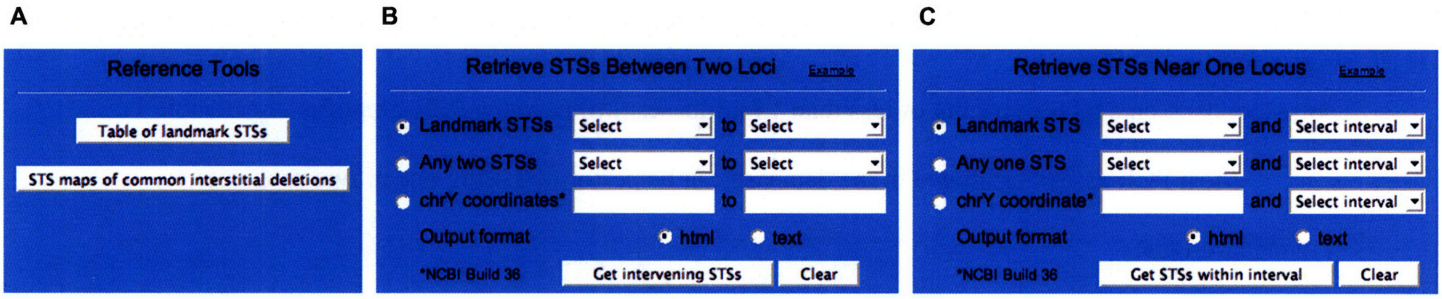


Figure 2. Querying the database.

(**A**) Reference Tools box containing links to a table of 51 landmark STSs and to deletion maps of common interstitial deletions.

(**B**) Query box for generating a catalog of STSs between two landmark STSs, or between any two STSs in the database, or between any two Y chromosome sequence coordinates.

(**C**) Query box for generating a catalog of STSs in the vicinity of a particular STS or sequence coordinate.

between MSY and pseudoautosomal sequences on Yp and Yq, respectively, and a previously published panel of 49 STSs that our laboratory has employed in screening for MSY deletions (see Table SM-2 in Supplementary Methods of ref.[36]). A table of these 51 landmark STSs is available on the query input webpage under Reference Tools (Figure 2A). MSY genes (Figure 1C) and structural features (Figure 1D) are annotated as previously described[1].

Two modes of querying the database are available to the user. The user may elect to generate a catalog of STSs between two loci (Figure 2B). They can be two landmark STSs, as listed in pull-down menus; any two STSs in the database, as listed in pull-down menus; or any two Y chromosome sequence coordinates, as entered by the user. Alternatively, the user may elect to generate a catalog of STSs in the vicinity of a particular STS or sequence coordinate (Figure 2C). In this case, the user selects the STS from a pull-down menu or enters the sequence coordinate, and then selects the size of the surrounding interval, again from a pull-down menu. Links to illustrations of these two modes of query are provided within the input boxes. In addition, a link to schematics of common MSY deletions, including predicted results for the 51 landmark STSs, is provided under Reference Tools (Figure 2A).

Regardless of the mode of query, the resulting catalog of STSs is displayed in two ways: as a table and in a custom Y chromosome browser (Figure 3). For each STS, the table indicates start and end sequence coordinates, PCR product length, primer sequences, amplification conditions, and GenBank accession number. Multi-copy STSs are identified as such, and the sequence coordinates of each additional locus of amplification are tabulated. Below the table, a custom Y chromosome browser displays the region encompassing the cataloged STSs. The position of each STS and of each protein-coding gene within the interval is indicated. The user can also interrogate the Y chromosome browser independent of the initial STS database query.

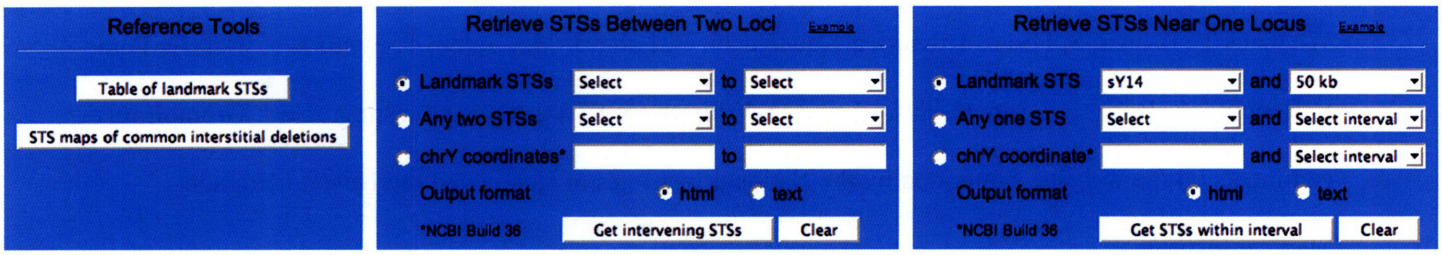
Figure 3. An example of query results.

(**A**) Querying the database for STSs in the 50-kb interval surrounding sY14.

(**B**) The queried interval contains fourteen STSs. For each STS, the table lists sequence coordinates, PCR product length, primers, PCR conditions (as a link to a pop-up), GenBank accession number (as a link to the NCBI entry), and in the case of a multi-copy STS, sequence coordinates of co-amplified loci.

(**C**) The custom Y chromosome browser indicates the location of the queried interval, and the identities and positions of all STSs and protein-coding genes within the interval.

A



B

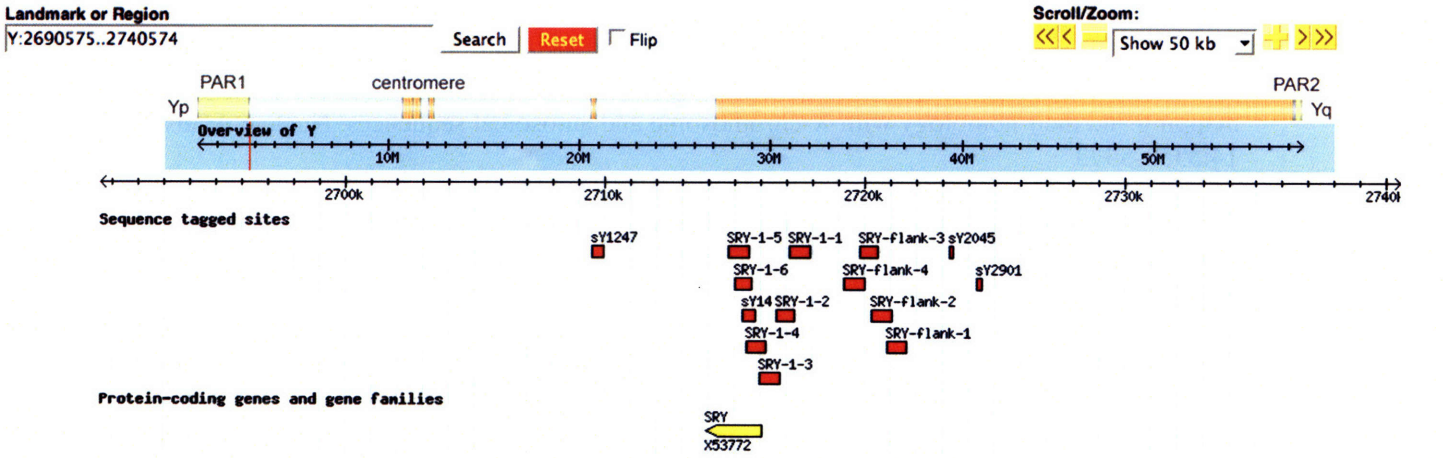
The following STSs are within chrY:2690575-2740574:

STS Identifier	Multi-copy	STS Start	STS End	STS Length	Primer 1	Primer 2	PCR Conditions	GenBank	Co-amplified Loci
sY1247		2709521	2709952	432	GAACTCTGCAAACCTCCTGG	TTTTGAGGCGGAGTCTCG	Standard PCR	G75493	
SRY-1-5		2714797	2715594	798	CAGTCCAGCTGTGCAAGAGA	ATAGGCAGGCTCACTTCTGG	Touchdown PCR	BV679099	
SRY-1-6		2715010	2715695	686	TTTCGAACTCTGGCACCTT	GCCAATGTTACCCGATTGTC	Touchdown PCR	BV679100	
sY14		2715341	2715810	470	GAATATTCCCGCTCTCCGGA	GCTGGTGCTCCATTCTTGAG	Standard PCR	G38356	
SRY-1-4		2715455	2716244	790	AGCCATCCTAGAAGTTGGGC	TGGGTCGCTTCACTCTATCC	Touchdown PCR	BV679098	
SRY-1-3		2715969	2716764	796	AGGAATTGTGGCATCATTCAG	TCACAAAACGAGAGGACACAA	Touchdown PCR	BV679097	
SRY-1-2		2716624	2717313	690	TCCCCACAACCTCTTTCATC	AAATGGGAGGAAAGTCCCC	Touchdown PCR	BV679096	
SRY-1-1		2717134	2717907	774	CGCTGTACCTCTCCATAGCC	AAATGACACAAGGCACCACA	Touchdown PCR	BV679095	
SRY-flank-4		2719239	2720033	795	CCCAACATCTTGAAAGGACA	GGGCTTAACCAACAGCAAAA	Touchdown PCR	BV679094	
SRY-flank-3		2719848	2720513	666	TGAAACCTTGTTTTTACAGGCA	GCAGGTCACCAAAGACACAA	Touchdown PCR	BV679093	
SRY-flank-2		2720286	2721050	765	GATTTGACCAAGCCTTTGGA	ATGAAGTAGGGTGCAGTGGG	Touchdown PCR	BV679092	
SRY-flank-1		2720869	2721618	750	CCAAAGCTATGAGAGAGGCAG	CAGCAGCTGTTTGACCAAGA	Touchdown PCR	BV679091	
sY2045		2723295	2723399	105	AGCTTAAACTGGTGTGCT	GAACCTAAAGTGCAGCAT	Standard PCR	G66044	
sY2901		2724300	2724522	223	GATTATGGGCCGAAATAGCA	GAATACAGCCCCTTTGGTCA	Standard PCR	BV703768	

C

Showing 50 kb chrY:2690575-2740574

Instructions: Search using an STS identifer, sequence coordinates, or gene symbol. Gene symbols may require a final asterisk wildcard to indicate a gene copy. You can also specify a gene (Gene:BPY2*) or an STS (STS:BPY2*).
Examples: Y, Y:20000000..20100000, sY14, sY34, sY1229, AMELY, AMELY-4, DAZ*, sY*, Gene:BPY2*, STS:BPY2*, Gene:*.



Acknowledgements

We thank L. Brown, S. Foote, T. Kuroda-Kawaguchi, B. Lahn, J. Marszalek, T. Pyntikova, R. Reijo Pera, S. Repping, S. Rozen, C. Sun, C. Tilford, and D. Vollrath for STS design and testing; and J. Gromoll for comments on the website. Supported by the National Institutes of Health and the Howard Hughes Medical Institute.

References

1. Skaletsky H, Kuroda-Kawaguchi T, Minx PJ, Cordum HS, Hillier L, Brown LG, Repping S, et al. (2003) The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. Nature 423:825-837

2. Jacobs PA, Ross A (1966) Structural abnormalities of the Y chromosome in man. Nature 210:352-354

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

4. Guellaen G, Casanova M, Bishop C, Geldwerth D, Andre G, Fellous M, Weissenbach J (1984) Human XX males with Y single-copy DNA fragments. Nature 307:172-173

5. Page DC, de la Chapelle A, Weissenbach J (1985) Chromosome Y-specific DNA in related human XX males. Nature 315:224-226

6. Vergnaud G, Page DC, Simmler MC, Brown L, Rouyer F, Noel B, Botstein D, de la Chapelle A, Weissenbach J (1986) A deletion map of the human Y chromosome based on DNA hybridization. Am J Hum Genet 38:109-124

7. Affara NA, Florentin L, Morrison N, Kwok K, Mitchell M, Cook A, Jamieson D, Glasgow L, Meredith L, Boyd E, et al. (1986) Regional assignment of Y-linked DNA probes by deletion mapping and their homology with X-chromosome and autosomal sequences. Nucleic Acids Res 14:5353-5373

8. Affara NA, Ferguson-Smith MA, Tolmie J, Kwok K, Mitchell M, Jamieson D, Cooke A, Florentin L (1986) Variable transfer of Y-specific sequences in XX males. Nucleic Acids Res 14:5375-5387

9. Andersson M, Page DC, de la Chapelle A (1986) Chromosome Y-specific DNA is transferred to the short arm of X chromosome in human XX males. Science 233:786-788

10. Disteche CM, Casanova M, Saal H, Friedman C, Sybert V, Graham J, Thuline H, Page DC, Fellous M (1986) Small deletions of the short arm of the Y chromosome in 46,XY females. Proc Natl Acad Sci USA 83:7841-7844

11. Petit C, de la Chapelle A, Levilliers J, Castillo S, Noel B, Weissenbach J (1987) An abnormal terminal X-Y interchange accounts for most but not all cases of human XX maleness. Cell 49:595-602

12. Page DC, Brown LG, de la Chapelle A (1987) Exchange of terminal portions of X- and Y-chromosomal short arms in human XX males. Nature 328:437-440

13. Levilliers J, Quack B, Weissenbach J, Petit C (1989) Exchange of terminal portions of X- and Y-chromosomal short arms in human XY females. Proc Natl Acad Sci USA 86:2296-2300

14. Vollrath D, Foote S, Hilton A, Brown LG, Beer-Romero P, Bogan JS, Page DC (1992) The human Y chromosome: a 43-interval map based on naturally occurring deletions. Science 258:52-59

15. Weil D, Wang I, Dietrich A, Poustka A, Weissenbach J, Petit C (1994) Highly homologous loci on the X and Y chromosomes are hot-spots for ectopic recombinations leading to XX maleness. Nat Genet 7:414-419

16. Lahn BT, Ma N, Breg WR, Stratton R, Surti U, Page DC (1994) Xq-Yq interchange resulting in supernormal X-linked gene expression in severely retarded males with 46,XYq- karyotype. Nat Genet 8:243-250

17. Hsu LY (1994) Phenotype/karyotype correlations of Y chromosome aneuploidy with emphasis on structural aberrations in postnatally diagnosed cases. Am J Med Genet 53:108-140

18. Reijo R, Lee TY, Salo P, Alagappan R, Brown LG, Rosenberg M, Rozen S, Jaffe T, Straus D, Hovatta O, et al. (1995) Diverse spermatogenic defects in humans caused by Y chromosome deletions encompassing a novel RNA-binding protein gene. Nat Genet 10:383-393

19. Tsuchiya K, Reijo R, Page DC, Disteche CM (1995) Gonadoblastoma: molecular definition of the susceptibility region on the Y chromosome. Am J Hum Genet 57:1400-1407

20. Vogt PH, Edelmann A, Kirsch S, Henegariu O, Hirschmann P, Kiesewetter F, Kohn FM, Schill WB, Farah S, Ramos C, Hartmann M, Hartschuh W, Meschede D, Behre HM, Castel A, Nieschlag E, Weidner W, Grone 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

21. Schiebel K, Winkelmann M, Mertz A, Xu X, Page DC, Weil D, Petit C, Rappold GA (1997) Abnormal XY interchange between a novel isolated protein kinase gene, PRKY, and its homologue, PRKX, accounts for one third of all (Y+)XX males and (Y-)XY females. Hum Mol Genet 6:1985-1989

22. 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

23. 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

24. 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

25. 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

26. Repping S, Skaletsky H, Lange J, Silber S, Van Der Veen F, Oates RD, Page DC, Rozen S (2002) Recombination between palindromes P5 and P1 on the human Y chromosome causes massive deletions and spermatogenic failure. Am J Hum Genet 71:906-922

27. Repping S, Skaletsky H, Brown L, van Daalen SK, Korver CM, Pyntikova T, Kuroda-Kawaguchi T, de Vries JW, Oates RD, Silber S, van der Veen F, Page DC, Rozen S (2003) Polymorphism for a 1.6-Mb deletion of the human Y chromosome persists through balance between recurrent mutation and haploid selection. Nat Genet 35:247-251

28. Fernandes S, Paracchini S, Meyer LH, Floridia G, Tyler-Smith C, Vogt PH (2004) A large AZFc deletion removes DAZ3/DAZ4 and nearby genes from men in Y haplogroup N. Am J Hum Genet 74:180-187

29. Repping S, van Daalen SK, Korver CM, Brown LG, Marszalek JD, Gianotten J, Oates RD, Silber S, van der Veen F, Page DC, Rozen S (2004) A family of human Y chromosomes has dispersed throughout northern Eurasia despite a 1.8-Mb deletion in the azoospermia factor c region. Genomics 83:1046-1052

30. Jobling MA, Lo IC, Turner DJ, Bowden GR, Lee AC, Xue Y, Carvalho-Silva D, Hurles ME, Adams SM, Chang YM, Kraaijenbrink T, Henke J, Guanti G, McKeown B, van Oorschot RA, Mitchell RJ, de Knijff P, Tyler-Smith C, Parkin EJ (2007) Structural variation on the short arm of the human Y chromosome: recurrent multigene deletions encompassing Amelogenin Y. Hum Mol Genet 16:307-316

31. Sinclair AH, Berta P, Palmer MS, Hawkins JR, Griffiths BL, Smith MJ, Foster JW, Frischauf AM, Lovell-Badge R, Goodfellow PN (1990) A gene from the human sex-determining region encodes a protein with homology to a conserved DNA-binding motif. Nature 346:240-244

32. Page DC (1987) Hypothesis: a Y-chromosomal gene causes gonadoblastoma in dysgenetic gonads. Development 101 Suppl:151-155

33. Nathanson KL, Kanetsky PA, Hawes R, Vaughn DJ, Letrero R, Tucker K, Friedlander M, et al. (2005) The Y deletion gr/gr and susceptibility to testicular germ cell tumor. Am J Hum Genet 77:1034-1043

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

35. Tilford CA, Kuroda-Kawaguchi T, Skaletsky H, Rozen S, Brown LG, Rosenberg M, McPherson JD, Wylie K, Sekhon M, Kucaba TA, Waterston RH, Page DC (2001) A physical map of the human Y chromosome. Nature 409:943-945

36. Repping S, van Daalen SK, Brown LG, Korver CM, Lange J, Marszalek JD, Pyntikova T, van der Veen F, Skaletsky H, Page DC, Rozen S (2006) High mutation rates have driven extensive structural polymorphism among human Y chromosomes. Nat Genet 38:463-467

37. Rozen S, Skaletsky H (2000) Primer3 on the WWW for general users and for biologist programmers. Methods Mol Biol 132:365-386

38. Kirsch S, Weiss B, Miner TL, Waterston RH, Clark RA, Eichler EE, Munch C, Schempp W, Rappold G (2005) Interchromosomal segmental duplications of the pericentromeric region on the human Y chromosome. Genome Res 15:195-204

Chapter 3

Isochromosomes and associated reproductive disorders are byproducts of the homologous recombination that maintains human Y-chromosomal palindromes

Julian Lange, Helen Skaletsky, Saskia K.M. van Daalen, Cindy M. Korver, Laura G. Brown, Robert D. Oates, Sherman Silber, Sjoerd Repping, David C. Page

Author contributions

J. Lange performed analysis of Y-chromosome content in patients, designed and performed FISH experiments. H. Skaletsky participated in all aspects of study design, generated PCR assays for STSs and for FISH probes, performed all statistical analyses. S.K.M. van Daalen and C.M. Korver assisted with FISH experiments and analysis of FISH results. L.G. Brown curated the patient database. R.D. Oates and S. Silber provided patient samples. S. Repping participated in the design of FISH assays. D.C. Page provided helpful discussions and direction throughout the project.

J. Lange and D.C. Page wrote the manuscript.

Summary

Massive palindromes in the human Y chromosome harbor mirror-image gene pairs essential for spermatogenesis. During human evolution, these gene pairs have been maintained by intrapalindrome, arm-to-arm recombination. These findings raise questions about the mechanism of this recombination, and the risk of harmful effects. We report 67 unrelated human individuals with Y isochromosomes formed by homologous crossing over between opposing arms of palindromes on sister chromatids. These ectopic recombination events occur at nearly all Y-linked palindromes, giving rise to male infertility, Turner syndrome, and sex reversal. Based on our findings, we propose that intrapalindrome sequence identity is maintained via the noncrossover pathway of homologous recombination. The DNA double-strand breaks that normally initiate this pathway can be alternatively resolved by crossing over between sister chromatids to form isochromosomes. Our observations imply that both crossover and noncrossover pathways are currently active in nearly all Y-linked palindromes, exposing an Achilles' heel in the mechanism of preserving palindrome-borne genes.

Introduction

Since the emergence of the mammalian X-Y sex determination system about 300 million years ago, functional specialization of the evolving Y chromosome has driven the stepwise suppression of sexual recombination with the X chromosome, resulting in the differentiation of what was once an ordinary pair of recombining autosomes[1]. However, such increasingly limited meiotic exchange has also fostered the decay of the Y chromosome[2]. Whereas the pseudoautosomal regions (PARs) of the present-day human sex chromosomes continue to pair and recombine, the remaining 95% of the Y chromosome, the male-specific region of the Y (MSY; Figure 1A), does not undergo any form of productive recombination with the X chromosome and has retained only a handful of its ancestral genes[3].

To counter further evolutionary decay, massive palindromes (Figure 1B, and Figure S1 in Supplemental Data) that harbor a suite of critical spermatogenesis genes have evolved in the MSY and

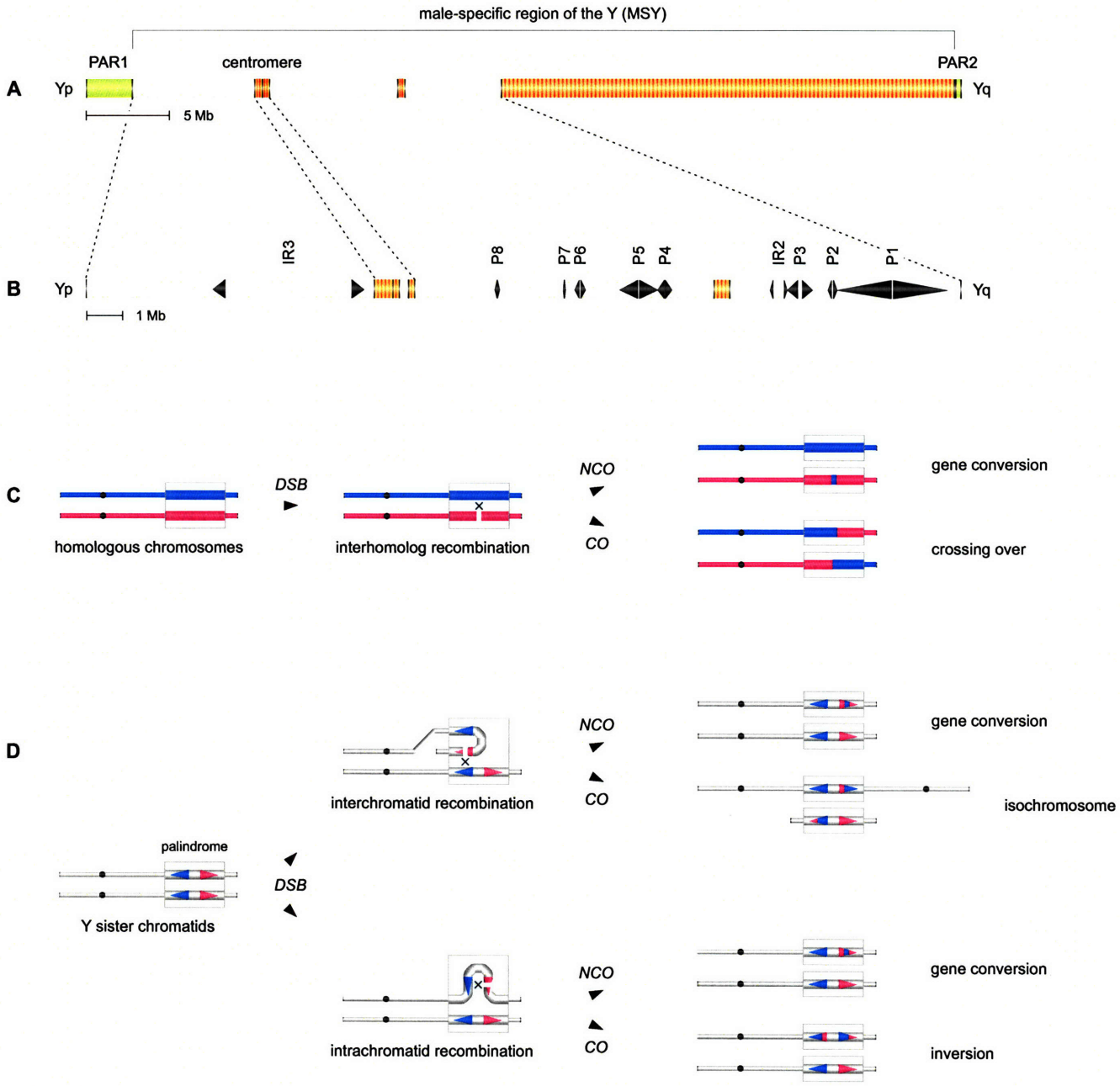
Figure 1. Hypothesized mechanism of Y isochromosome formation through homologous recombination in an MSY palindrome.

(A) Schematic representation of human Y chromosome. The male-specific region of the chromosome, or MSY, is flanked by two pseudoautosomal regions, PAR1 and PAR2 (green), and contains blocks of heterochromatin (orange).

(B) Expanded view of MSY euchromatin, indicating locations of eight palindromes (P1 through P8) and two inverted repeats (IR2 and IR3), all to scale.

(C) Conventional meiotic recombination between homologous chromosomes: DNA double-strand break (DSB) resolution by noncrossover (NCO) and crossover (CO) pathways, yielding, respectively, gene conversion and crossing over.

(D) A model of homology-dependent recombination in MSY palindromes: DSB resolution by NCO pathways yields gene conversion. Resolution by crossing over within a chromatid produces an inversion, while crossing over between sister chromatids produces an isochromosome.


male-specific region of the Y (MSY)
PAR1
centromere
PAR2
A
Yp
Yq
5 Mb
B
IR3
P8
P7
P6
P5
P4
IR2
P3
P2
P1
1 Mb
C
DSB
homologous chromosomes
interhomolog recombination
NCO
CO
gene conversion
crossing over
D
interchromatid recombination
isochromosome
palindrome
Y sister chromatids
intrachromatid recombination
inversion

have been maintained by gene conversion. Each palindrome is composed of two large arms separated by a small, unique "spacer" sequence at the center. Ranging in wing-span from 30 kilobases (kb) to 2.9 megabases (Mb), the palindromes collectively comprise 25% of the euchromatic MSY[3]. The palindromes predate the divergence of the human and chimpanzee lineages, yet they exhibit arm-to-arm sequence identities of > 99.9%[4]. Evidence in modern human lineages of sequence homogenization in MSY palindromes indicates that gene conversion there remains a frequent process[4,5].

As the mechanistic basis of the gene conversion that maintains MSY palindromes has yet to be elucidated, we considered possible parallels to a better understood and potentially related process. Meiotic recombination between homologous chromosomes frequently results in nonreciprocal exchange, or gene conversion, with no crossing over. In meiotic recombination between homologs, such gene conversion is the result of DNA double-strand break (DSB) formation followed by noncrossover (NCO) resolution of those breaks (Figure 1C)[6-9]. Similarly, we hypothesized that arm-to-arm sequence identity within palindromes might be maintained by DSB formation and NCO resolution, either between sister chromatids or within one chromatid (Figure 1D). Since DSB formation can also lead to crossing over in the case of homologous chromosomes (Figure 1C), we further hypothesized that analogous options exist in the case of MSY palindrome recombination (Figure 1D). A testable prediction of this hypothesis is that interchromatid crossover (CO) resolution of DSBs formed in MSY palindromes might give rise to Y isochromosomes (Figure 1D). We examined a large collection of human individuals, many known to carry structurally aberrant Y chromosomes, for evidence of such events. We identified 67 unrelated individuals with Y isochromosomes that evidently formed by homology-mediated crossing over between palindromes on sister chromatids, and that are associated with a wide range of sex-linked reproductive disorders including male infertility, Turner syndrome, and sex reversal.

Results

To test our hypothesis that DSBs formed in Y-chromosomal palindromes are sometimes resolved by crossing over between sister chromatids (Figure 1D), we set out to identify and molecularly characterize Y isochromosomes arising spontaneously in human populations. Y isochromosomes ascertained through cytogenetic (light microscopic) studies have been reported in the literature[10,11], but their molecular basis and mode of origin remained unknown. Moreover, we suspected that Y isochromosomes might sometimes go undetected or be misidentified in conventional cytogenetic studies, due to the Y chromosome's relatively small size and its paucity of microscopic landmarks. Thus, we decided to conduct a broad screen for potential Y isochromosomes by molecular testing, using Y-DNA markers, to capture a large sample and obtain insight into mechanisms of origin.

We conducted this screen for Y isochromosomes by reanalyzing DNA samples from 2,380 individuals that had been studied in our laboratory during the past 25 years[12-17]. These individuals had been ascertained by one of two means. In 830 of these individuals, light microscopic examination (karyotyping) had revealed a structurally anomalous Y chromosome, or a discrepancy between sex chromosome constitution and sex phenotype ("sex reversal"). The remaining 1,550 individuals were men with spermatogenic failure (< 5 million spermatozoa per ml semen; normal > 20 million spermatozoa per ml semen), a phenotype often associated with anomalies of the Y chromosome.

Employing a series of precisely mapped, Y-specific sequence-tagged sites (STSs) as markers[13,18], we screened genomic DNA samples from all 2,380 individuals for molecular evidence of Y isochromosomes. Most of the MSY's palindromes are located on the long arm (Yq; Figure 1B), where recombination between opposing palindrome arms on sister chromatids should generate "isoYp" chromosomes, bearing two copies of the entire short arm (Yp) but lacking the most distal portion of Yq (Figure 1D). Accordingly, we searched specifically for evidence of isoYp chromosomes.

Fifty unrelated individuals with potential isoYp chromosomes ascertained by PCR mapping of breakpoints to Yq palindromes

Our initial screen focused on three STS markers whose presence or absence is readily scored by PCR: sY14, which marks the most distal male-specific portion of Yp; sY78, which marks the centromere; and sY1273, which marks the most distal male-specific portion of Yq (Figure 2A). Individuals bearing isoYp chromosomes should carry the Yp and centromeric markers, but lack the Yq marker. Among the 2,380 individuals tested, we identified 100 cases whose STS signatures were consistent with an isoYp chromosome: sY14 and sY78 were present, but sY1273 was absent.

We inferred that, in each of these 100 cases, the Y chromosome was broken somewhere between the centromeric marker sY78 and the Yq marker sY1273. To confirm this inference, and to fine-map each of these 100 breakpoints, we proceeded to assay additional landmark STSs in Yq and the centromeric region, including STSs at the boundaries of each palindrome (Figures 2B and 2C). In each of these 100 cases, we localized a single breakpoint on Yq; all examined STSs proximal to the breakpoint were found to be present, while all those distal to the breakpoint were absent (Figure 2D).

If any of these 100 individuals with Yq breakpoints carries an isoYp chromosome generated by crossing over between opposing palindrome arms on sister chromatids, then their breakpoint should lie within one of Yq's palindromes. In fact, in 58 of these 100 cases, the breakpoint was localized to within a palindrome or other large inverted repeat. That is, in each of these 58 cases, the STS straddling the proximal boundary of a particular palindrome was present while the STS straddling the distal boundary of that palindrome was absent (breakpoints within blue regions in Figure 2D).

A further prediction of our model of isochromosome formation is that an intact spacer at the center of the targeted palindrome should be retained in the resulting isochromosome (Figure 2E). Apart from our model, one might expect deletion breakpoints to map with equal frequency to the palindromes' proximal and distal arms, which display 99.9% sequence identity; the spacer would be preserved only in cases with distal-arm breakpoints. We therefore tested spacer-flanking STSs in each of the 58 cases with breakpoints in palindromes to determine whether the spacer was present or absent. In 50 of 58 cases,

STSs flanking the spacer of the affected palindrome were present, demonstrating the spacer's integrity (Figure 2D; see also Table S1A). In seven cases, the spacer of the affected palindrome was absent (Table S1B). (One case displayed a breakpoint within the spacer of IR2 [Table S1C].) This distribution – 50 cases with spacers intact, and seven cases with spacers absent – is inconsistent with deletion breakpoints arising with equal frequency in the proximal and distal arms of palindromes (chi square, $p < 0.0001$), but it is consistent with most of the cases having arisen via our proposed mechanism.

In summary, our PCR-based screening of 2,380 individuals identified 50 unrelated cases that satisfied two predictions of our model of isochromosome formation (Figure 1D): 1) the absence of all Y-chromosomal material distal to a particular palindrome, as assayed by STSs straddling palindrome boundaries, and 2) the retention of the spacer of the affected palindrome (Figure 2E). Interestingly, of the nine palindromes and inverted repeats in Yq, all but one was targeted in at least one of these 50 cases (Figure 2D, Table 1).

IsoYp chromosomes demonstrated by metaphase fluorescence *in situ* hybridization

Our model of isochromosome formation by crossing over between opposing palindrome arms on sister chromatids predicts not only the loss of Y-chromosomal material, but also a mirror-image duplication of the retained portion of the chromosome (Figure 1D). In each of the 50 individuals identified in our PCR-based screen, the sets of Y-chromosomal loci retained and deleted were consistent with an isoYp chromosome formed by this model. However, neither the copy number of loci that were retained nor their arrangement could be surmised from these PCR assays. To cytogenetically investigate the structure of the Y chromosome in a sampling of these individuals, we turned to fluorescence *in situ* hybridization (FISH) on metaphase spreads from lymphoblastoid cell lines.

An isoYp chromosome generated by this model should bear a mirror-image duplication of the entirety of Yp, the centromere, and the segment of Yq proximal to the affected palindrome (Figure 1D). The more distal the affected palindrome, the larger the size of the Yq segment retained. For example, the distance from the centromere to palindrome P8 is 2.2 Mb, while the distance from the centromere to

Figure 2. STS content of structurally anomalous Y chromosomes in 100 human individuals with deletion breakpoints in the long arm (Yq) or centromere.

(A) STSs employed in the initial screen for Yq or centromeric breakpoints. Genomic DNAs were screened for presence of distal Yp STS sY14 and centromeric STS sY78, and absence of distal Yq STS sY1273.

(B) Expanded view of centromere and Yq, indicating locations of eight palindromes (P1 through P8), one inverted repeat (IR2), and five heterochromatic regions containing highly repetitive elements (*DYZ1-3* and *DYZ17-19*).

(C) STSs employed in fine mapping of deletion breakpoints. These STSs include boundary and spacer-flanking markers for each palindrome and inverted repeat, and markers around each of five blocks of heterochromatin.

(D) Results of testing genomic DNAs from 100 individuals for presence or absence of STSs. Solid black bars encompass STSs found to be present. Gray bars indicate breakpoint intervals that could not be further narrowed due to cross-amplification at other loci. See Figure S2 for identifiers of the 100 individuals tested.

(E) Predicted molecular signature of an isoYp chromosome formed by homologous recombination between sister chromatids at palindrome P5. Thirteen individuals displaying this molecular signature are shown in Figure 2D.

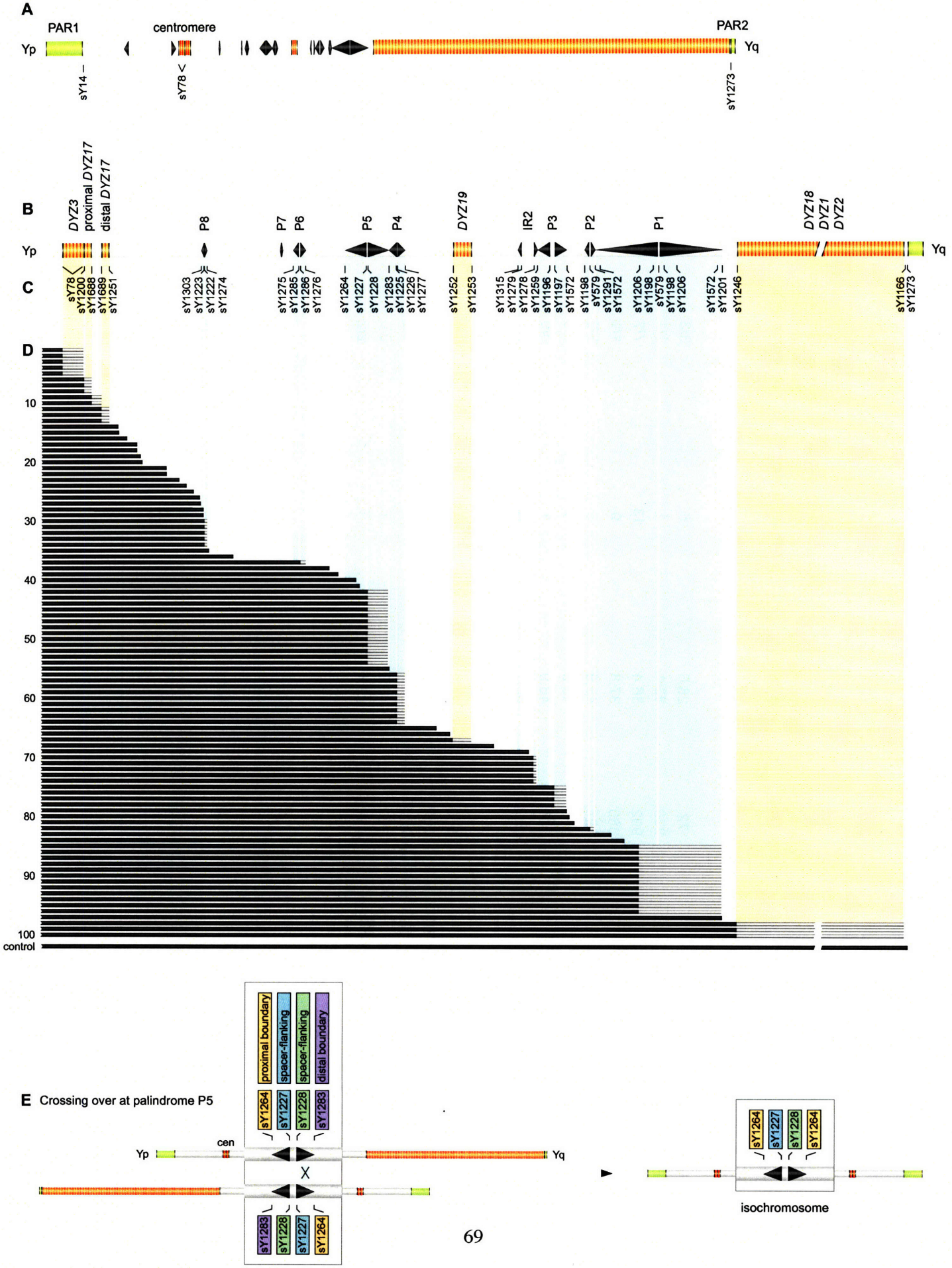

A
PAR1
centromere
PAR2
Yp
Yq
sY14
sY78
sY1273
B
DYZ3
proximal DYZ17
distal DYZ17
P8
P7
P6
P5
P4
DYZ19
IR2
P3
P2
P1
DYZ18
DYZ1
DYZ2
Yp
Yq
C
sY78
sY1200
sY1688
sY1689
sY1251
sY1303
sY1223
sY1222
sY1274
sY1275
sY1285
sY1286
sY1276
sY1264
sY1227
sY1228
sY1283
sY1225
sY1226
sY1277
sY1252
sY1253
sY1315
sY1279
sY1278
sY1259
sY1196
sY1197
sY1572
sY1198
sY579
sY1291
sY1572
sY1206
sY1198
sY579
sY1198
sY1206
sY1572
sY1201
sY1246
sY1166
sY1273
D
10
20
30
40
50
60
70
80
90
100
control
E Crossing over at palindrome P5
proximal boundary
spacer-flanking
spacer-flanking
distal boundary
sY1264
sY1227
sY1228
sY1283
cen
Yp
Yq
X
sY1283
sY1228
sY1227
sY1264
sY1264
sY1227
sY1228
sY1264
isochromosome

Table 1. Ectopic homologous recombination between palindromes on sister chromatids results in the formation of Y isochromosomes.

Region	Targeted repeat	Total target size (kb)[a,b]	Isochromosome size (Mb)[b]	Number of cases		
				STS mapping results consistent with proposed event	General FISH assays isochromosomes observed/tested	Specific FISH assays isochromosomes via mechanism/tested
IsoYp chromosomes						
Yq euchromatin	P8	72	28.6	5	3/3	3/3
	P6	220	33.1	1	1/1	1/1
	P5	992	36.4	13	7/7	6/6
	P4	380	37.8	9	1/1	1/1
	IR2	124	43.9	5	3/3	3/3
	P3	566	45.0	4	3/3	3/3
	P2	244	46.9	1	0/0	0/0
	P1	2,900	50.1	12	7/7	6/6[d]
Total				50	25/25[c]	23/23
Centromeric and Yq heterochromatin	*DYZ3*	500	21.9	5	2/2	2/2
	proximal *DYZ17*	200	22.6	3	1/1	1/1
	distal *DYZ17*	200	23.9	3	0/0	0/0
	DYZ19	330	40.9	1	0/0	0/0
	DYZ18/DYZ1/DYZ2	31,700	85.4	3	2/2	2/2
Total				15	5/5	5/5
IsoYq chromosomes						
Yp euchromatin	IR3	596	101.4	2	1/1	1/1
Total				2	1/1	1/1
Grand total				67	31/31	29/29

:r can take place in either the left or right arm of a palindrome, therefore the total target size is twice the size of a single arm.

t sizes and isochromosome sizes are based on: estimated total Y chromosome, 59 Mb[19], euchromatic MSY sequence[3] (except *TSPY* a

R1, 2.71 Mb (NCBI human genome assembly Build 36), *DYZ3* array, 500 kb[3]; proximal *DYZ17* array, 200 kb; pericentromeric

449 kb[21]; distal *DYZ17* array, 200 kb[21]; *DYZ19* array, 330 kb (our unpublished data); *DYZ18/DYZ1/DYZ2* array, 31.7 Mb; euchromat

veen distal boundary of *DYZ18/DYZ1/DYZ2* and PAR2, 100 kb[3]; PAR2, 335 kb (NCBI human genome assembly Build 36).

chromosomal material was detected by PCR, no signal was observed in hybridizations to metaphase nuclei in two individuals, one w

P4, the other with a breakpoint in P2. Previous karyotype information indicated mosaicism for an XO cell line in each of these two c

me that cell lines for each contain a high percentage of XO cells. We report here results for the 25 cases of 27 tested for which we de

signal.

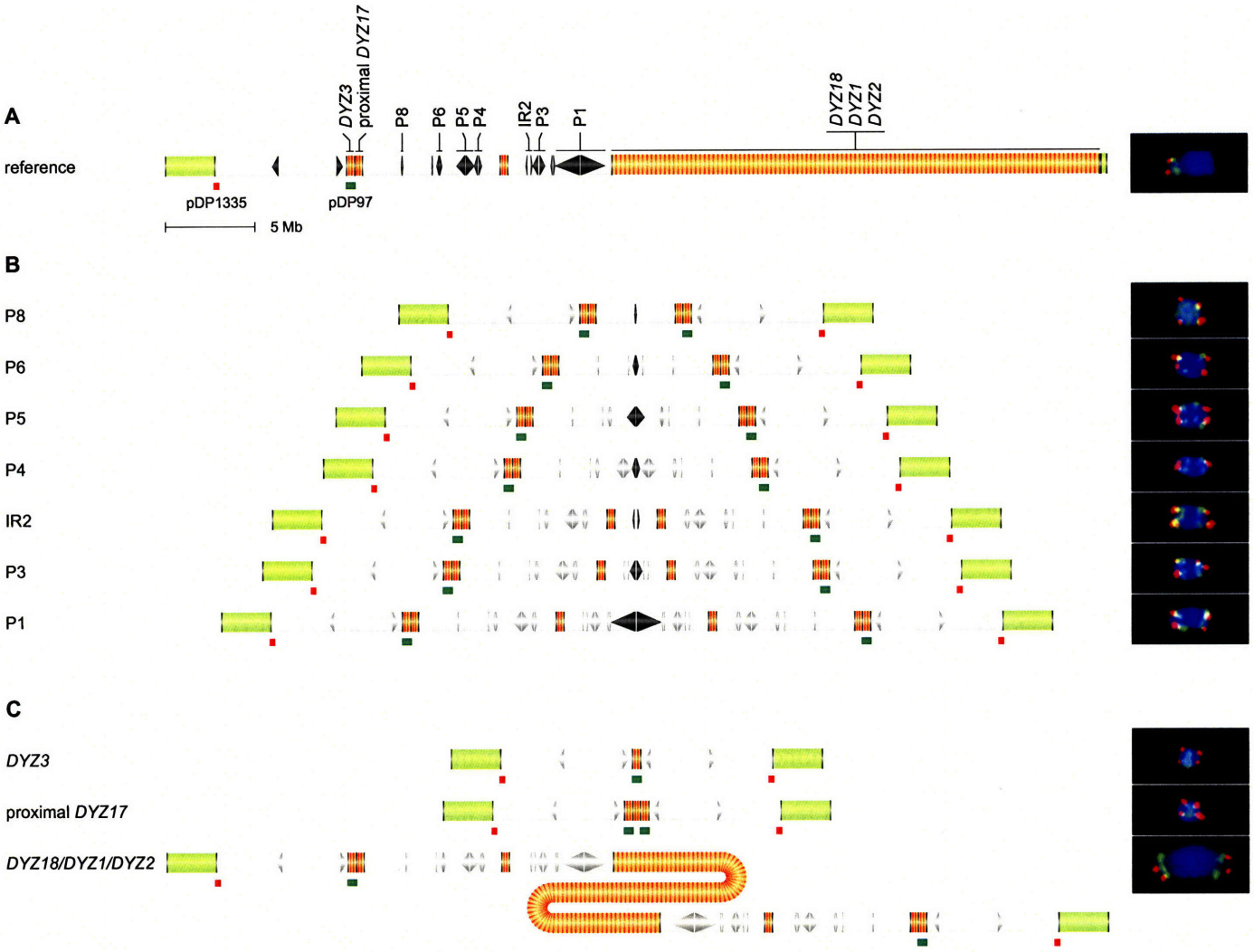
mbination events involving inverted repeats in P1, P2, P3, and P5 are predicted to yield similar STS mapping results. Specific FISH a

the isoYp chromosome in one case, WHT3189, had likely arisen by homologous recombination between, on one chromatid, palindrc

ıer chromatid, a 1,116-kb region of 99.9% sequence identity encompassing palindrome P2. The predicted size of this isoYp chromos

palindrome P1 is 11.5 Mb (Figure 3A). Accordingly, the distance between the two centromeres in resulting isochromosomes should increase as more distal palindromes are targeted (see schematics in Figure 3B). To test these predictions, we designed a two-color FISH assay, employing probes that, in a reference Y chromosome, hybridized to the most distal male-specific portion of Yp (probe pDP1335) and to *DYZ3* sequences in the centromere (probe pDP97; normal control male shown in Figure 3A). This enabled us to determine the copy number and spatial relationship of the distal Yp locus and the centromere. We selected 27 of the 50 cases for study by FISH, including at least one representative for each of the eight targeted palindromes.

Figure 3. IsoYp chromosomes confirmed by metaphase fluorescence *in situ* hybridization.
(A) Hybridization sites for FISH probes pDP1335 (to distal Yp), in red, and pDP97 (to centromeric *DYZ3* repeats), in green, are shown below a schematic representation of the Y chromosome. At right: Co-hybridization of pDP1335 and pDP97 to the Y chromosome of a normal male control produced the expected pattern of red and green signals.
(B) Co-hybridization of pDP1335 and pDP97 to metaphase chromosomes from individuals displaying breakpoints at Yq palindromes P8, P6, P5, P4, P3 and P1, or at IR2 inverted repeat, and therefore presumed to carry isoYp chromosomes. Hybridization sites are shown below a schematic of each predicted isoYp chromosome; the targeted palindrome or inverted repeat lies at the center of each schematic. At right: Two-color FISH produced the predicted red-green-green-red hybridization patterns.
(C) Co-hybridization of pDP1335 and pDP97 to metaphase chromosomes from individuals displaying breakpoints in Y heterochromatin and presumed to carry the isoYp chromosomes shown schematically. At right: Two-color FISH produced the predicted hybridization patterns.


A
reference
DYZ3
proximal DYZ17
P8
P6
P5
P4
IR2
P3
P1
DYZ18
DYZ1
DYZ2
pDP1335
pDP97
5 Mb
B
P8
P6
P5
P4
IR2
P3
P1
C
DYZ3
proximal DYZ17
DYZ18/DYZ1/DYZ2

In 25 of the 27 cases examined, we detected hybridization signals for both pDP1335 and pDP97 on metaphase spreads. In each of these 25 cases, we observed duplication of both Yp and the centromere, and the predicted pattern of two centromeres flanked by two Yp arms (Figure 3B, Table 1). In addition, a comparison of FISH results across all 25 cases demonstrated an increase in the distance between centromeric signals as more distal palindromes were targeted (Figure 3B). In the remaining two of 27 cases (including the only case with a breakpoint in palindrome P2), we observed no hybridization signal for either pDP1335 or pDP97, despite both loci having tested positive by PCR. These discordant results may be due to high-grade mosaicism for a 45,X ("XO") cell line. In sum, we microscopically demonstrated the presence of an isoYp chromosome in all 25 scorable cases, distributed among all but one of the eight targeted palindromes and inverted repeats on Yq.

Proposed mechanism of isoYp chromosome formation validated by interphase FISH

Our results to this point supported the hypothesis of isochromosome formation via homology-mediated crossing over between opposing palindrome arms on sister chromatids (Figure 1D). These results included retention of the proximal boundary and entire spacer of a targeted palindrome but deletion of the distal boundary, all as judged by PCR (in 50 cases), and mirror-image duplication of Yp and the centromere, as assayed by FISH (confirmed in all 25 scorable cases examined). However, these results could also be explained by recombination between, on one chromatid, a Yq palindrome and, on its sister chromatid, sequences closer to the centromere. These two models yield different predictions regarding the copy number of DNA sequences located immediately proximal to the targeted palindrome (Figure 4A). Whereas our proposed model predicts the duplication of these sequences – with copies originating from both chromatids – the alternative explanation predicts that these sequences would not be duplicated.

We designed FISH assays to examine this question empirically. For each targeted palindrome, we isolated a genomic DNA clone or prepared a long-range PCR product that would hybridize specifically to sequences just proximal to that palindrome. For example, probe 15485/15486 hybridizes to a site 44 kb proximal to palindrome P5, while probe 15469/15470 hybridizes to a site 208 kb proximal to IR2 (normal

control male shown in Figure 4B). Anticipating that the resolution offered by metaphase FISH would be insufficient for these experiments, we performed FISH on interphase spreads prepared from lymphoblastoid cell lines bearing isoYp chromosomes. We tested 23 of the 25 cases for which we had confirmed an isochromosome on metaphase spreads, including at least one representative for each of seven targeted palindromes. (In the remaining two cases, metaphase FISH had revealed a very low percentage of cells with an isoYp chromosome, likely reflecting mosaicism for a 45,X cell line, and making statistical analysis of interphase nuclei difficult.) To determine the number of probe hybridization sites, and thereby differentiate the two models, we scored at least 100 nuclei in each of the 23 cases examined. In 22 of the 23 cases, we found that the sequence just proximal to the targeted palindrome was duplicated (Figures 4C, 4D, S3, and S4; Table 1), as predicted by our model of isoYp chromosome formation.

These findings eliminated the alternative model in all but one case, WHT3189. Further analysis of this seemingly exceptional case confirmed our model of isochromosome formation. WHT3189 possesses a breakpoint in palindrome P1, large blocks of which are repeated in palindromes P2, P3, and P5 (Figure S4). We postulated that the isoYp chromosome in WHT3189 might be due to homologous crossing over between sister chromatids, but with only one of the two recombination points in palindrome P1, and the other in P2, P3, or P5. Additional interphase FISH studies indicated that this isoYp chromosome had likely arisen by homologous recombination between, on one chromatid, palindrome P1, and on the other chromatid, a block of 99.9% sequence identity encompassing palindrome P2 (Figure S4). In sum, in all 23 cases examined by interphase FISH, the results validated the proposed mechanism of isochromosome formation (Table 1).

IsoYp chromosomes arising from recombination in heterochromatic repeats

As described above, PCR-based screening of 2,380 selected individuals had identified 100 cases with breakpoints between the centromeric marker sY78 and the distal Yq marker sY1273 (Figure 2D). In 58 of these 100 individuals, additional PCR testing revealed breakpoints in Yq palindromes. We will now turn

Figure 4. Duplication in isoYp chromosomes of sequences immediately proximal to targeted palindromes.

(A) Prior results (STS mapping and metaphase FISH) were consistent with two models of isochromosome formation. These two models can be distinguished by determining, by FISH, the copy number of sequences located immediately proximal to the targeted palindrome.

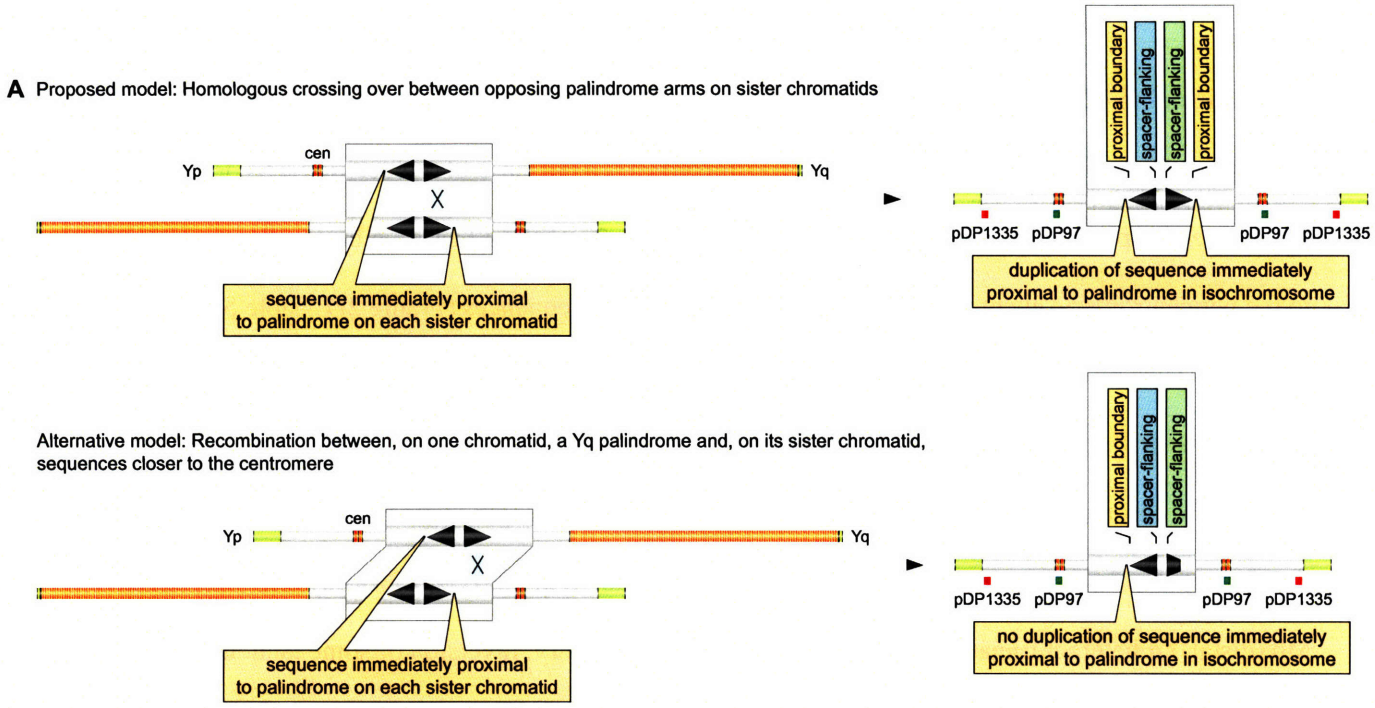
(B) Hybridization sites for FISH probes 15485/15486 (proximal to P5), 15469/15470 (proximal to IR2), 1136L22 (at the proximal boundary of *DYZ18/DYZ1/DYZ2* heterochromatin), and 242E13 (to *DYZ1* repeats) are shown below a schematic representation of the Y chromosome. Hybridizations to the Y chromosome of a normal male control produced the expected signals: 15485/15486 and 15469/15470 on interphase spreads, and co-hybridization of 1136L22 and 242E13 on metaphase spreads.

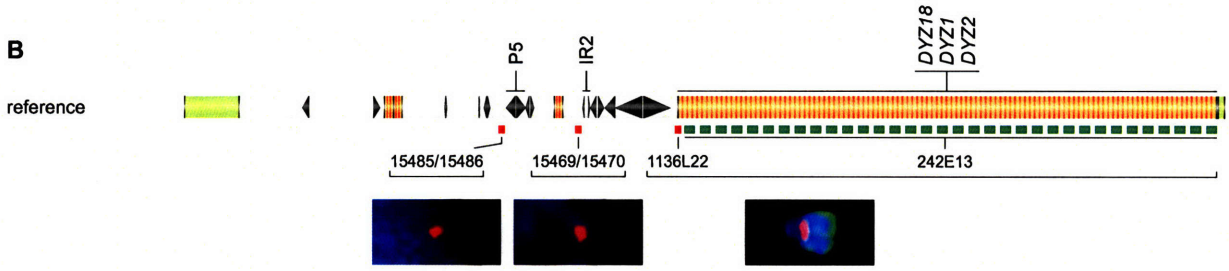
(C) Interphase FISH demonstrates the duplication of 15485/15486 on an isoYp chromosome with a breakpoint in palindrome P5.

(D) Interphase FISH demonstrates the duplication of 15469/15470 on an isoYp chromosome with a breakpoint in inverted repeat IR2.

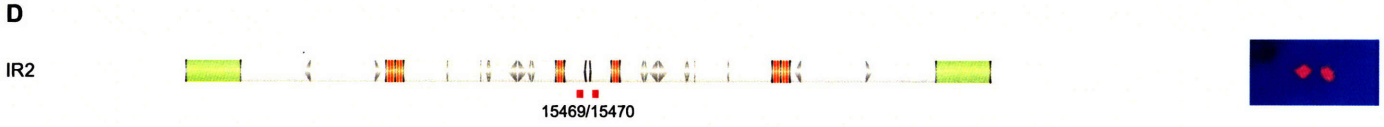
(E) Metaphase FISH demonstrates the duplication of 1136L22 on an isoYp chromosome with a breakpoint in *DYZ18/DYZ1/DYZ2* heterochromatin. Co-hybridization of 1136L22, in red, with 242E13, in green, produced the predicted red-green-red pattern.

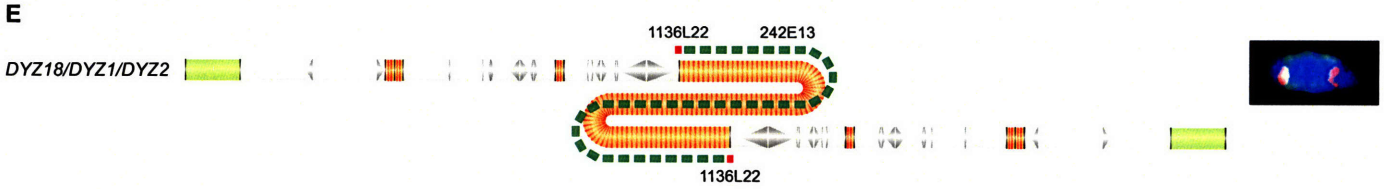
See Figures S3, S4, and S5 for additional data.


A Proposed model: Homologous crossing over between opposing palindrome arms on sister chromatids
cen
Yp
Yq
X
sequence immediately proximal to palindrome on each sister chromatid
proximal boundary
spacer-flanking
spacer-flanking
proximal boundary
pDP1335
pDP97
pDP97
pDP1335
duplication of sequence immediately proximal to palindrome in isochromosome
Alternative model: Recombination between, on one chromatid, a Yq palindrome and, on its sister chromatid, sequences closer to the centromere
cen
Yp
Yq
X
sequence immediately proximal to palindrome on each sister chromatid
proximal boundary
spacer-flanking
spacer-flanking
pDP1335
pDP97
pDP97
pDP1335
no duplication of sequence immediately proximal to palindrome in isochromosome


B
reference
P5
IR2
DYZ18
DYZ1
DYZ2
15485/15486
15469/15470
1136L22
242E13


C
P5
15485/15486


D
IR2
15469/15470


E
DYZ18/DYZ1/DYZ2
1136L22
242E13
1136L22

to the remaining 42 cases, which displayed breakpoints outside palindromes. We employed MSY Breakpoint Mapper (http://breakpointmapper.wi.mit.edu/)[18], a database of Y-specific STSs, to guide PCR-based fine-mapping of the breakpoints in these 42 cases. In 27 of the 42 cases, a breakpoint was localized to the pericentromeric euchromatin or to non-palindromic Yq sequences (Figure 2D, Table S1D).

In each of the 15 remaining cases, we mapped a breakpoint to one of five large blocks of centromeric or Yq heterochromatin (breakpoints within orange regions in Figure 2D; Table S1E). In the reference Y chromosome, these heterochromatic regions contain lengthy tandem arrays of short (5- to 171-basepair) sequence units, and they range in size from about 200 kb (*DYZ17* arrays) to more than 30 Mb (distal Yq heterochromatin; *DYZ18/DYZ1/DYZ2*)[3,21]. In theory, inversion of some tandemly repeated units within one of these heterochromatic regions could generate a target for isoYp chromosome formation by our proposed mechanism. We speculated that some (or all) Y chromosomes might carry such intraheterochromatic inversions, and thus we examined the possibility that the 15 cases with breakpoints in centromeric or Yq heterochromatin carry isoYp chromosomes.

We selected three cases with breakpoints in the centromere (two in *DYZ3* and one in *DYZ17*) and two cases with breakpoints in distal Yq heterochromatin for further study by FISH. Co-hybridization of probes pDP1335 (to distal Yp) and pDP97 (to the centromere) to metaphase spreads of these five cases yielded the predicted patterns (Figure 3C, Table 1). In each of the three cases with breakpoints in the centromere, we observed two Yp signals flanking one centromeric signal. In each of the two cases with breakpoints in distal Yq heterochromatin, we observed two Yp signals flanking two distinct centromeric signals. Thus, isoYp chromosomes are present in all five scrutinized cases with breakpoints in centromeric or Yq heterochromatin.

Our proposed model of isochromosome formation predicts the duplication of all retained sequences apart from the targeted heterochromatic regions. We designed additional FISH assays to explore this matter in each of the five cases. In the two cases with breakpoints in distal Yq heterochromatin, we found that the proximal boundary of this heterochromatic region was duplicated, as predicted (Figure 4E, Table 1). Similarly, in the three cases with centromeric breakpoints, we observed

the duplication of a Yp sequence located 280 kb from the centromere (Figure S5, Table 1). In sum, these FISH studies are consistent with our proposed mechanism. Future studies may identify inversions within heterochromatic regions that would facilitate isochromosome formation by this mechanism. Alternatively, Darlington's hypothesis of isochromosome formation via centromere misdivision[22] could account for the cases with breakpoints in centromeric heterochromatin.

IsoYq chromosomes generated by recombination at an inverted repeat on Yp

While most of the MSY's palindromes and inverted repeats are located on Yq, the 298-kb arms of the inverted repeat IR3 are located on Yp (Figures 1B and 5A). The 3.6-Mb spacer bounded by IR3's arms has been inverted repeatedly during human history[3,20,23-26]. Such inversion might be due to DSB formation in one IR3 arm and intrachromatid resolution of that DSB by crossing over with the other IR3 arm (see Figure 1D). Further, we postulated that interchromatid resolution of such a DSB could generate an isoYq chromosome, bearing two copies of the centromere and Yq but lacking the portion of Yp distal to IR3. We employed a targeted, PCR-based strategy to screen our collection of 2,380 individuals for such isoYq chromosomes.

Our screen made use of eight STSs. The first three of these, sY14, sY78, and sY1273, mark distal Yp, the centromere, and distal Yq, respectively (Figure 2A). IsoYq chromosomes should lack the distal Yp marker but retain the centromeric and distal Yq markers. The other five STSs were as follows: sY1241 and sY1244 straddle the distal and proximal boundaries of IR3, sY1242 and sY1243 flank its 3.6-Mb spacer, and sY1280 marks the Yp pericentromeric euchromatin (Figure 5B). In an isoYq chromosome formed by crossing over between opposing IR3 arms on sister chromatids, the distal boundary STS should be deleted, but the proximal boundary STS, both spacer-flanking STSs, and the pericentromeric STS should be retained. Among 2,380 individuals tested, we identified two cases in which the STS data were consistent with an isoYq chromosome formed by this model at IR3 (Figure 5C, Table S1F). (This screen would also have identified isoYq chromosomes with breakpoints in centromeric *DYZ3* repeats, but we found none.)

Figure 5. IsoYq chromosomes formed by crossing over between sister chromatids at a Yp inverted repeat: STS content and FISH analyses.

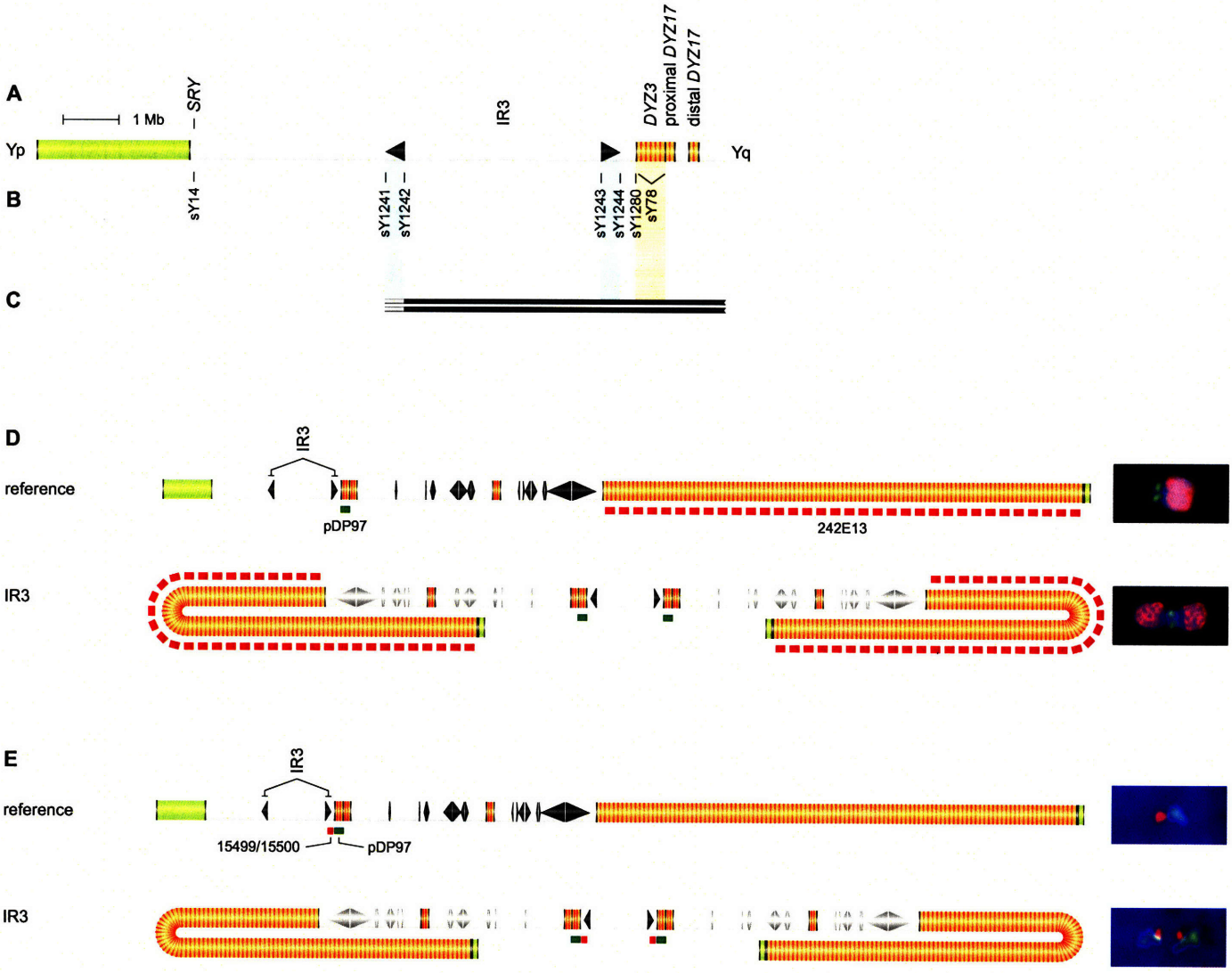
(A) Schematic representation of Yp and centromeric region of Y chromosome, indicating locations of IR3 inverted repeat and male-determining gene *SRY*.

(B) STSs employed in fine mapping of deletion breakpoints in two individuals in whom sY14 was absent, and sY78 and sY1273 were present. These STSs include boundary and spacer-flanking markers for the IR3 inverted repeat, and a marker in Yp pericentromeric euchromatin.

(C) Results of testing genomic DNAs from two individuals for presence or absence of STSs. Solid black bars encompass STSs found to be present. Gray bars indicate breakpoint intervals within IR3's distal arm that could not be further narrowed due to cross-amplification of identical sequences in IR3's proximal arm. See Figure S6 for identifiers of the two individuals tested.

(D) Co-hybridization of FISH probes pDP97 (to centromeric *DYZ3* repeats), in green, and 242E13 (to *DYZ1* repeats in distal Yq heterochromatin), in red, to the metaphase Y chromosome of a normal male (above) and to the metaphase isoYq chromosome of an individual displaying a breakpoint in IR3 (below).

(E) Interphase FISH demonstrates duplication of sequences proximal to IR3 in an isoYq chromosome. Co-hybridization of probes 15499/15500 (proximal to IR3), in red, and pDP97 (to centromeric *DYZ3* repeats), in green, produced the predicted patterns of red-green on the Y chromosome of a normal male control (above), and green-red-red-green on an isoYq chromosome with a breakpoint in IR3 (below). See Figure S7 for additional data.


A
1 Mb
SRY
IR3
DYZ3
proximal DYZ17
distal DYZ17
Yp
Yq
B
sY14
sY1241
sY1242
sY1243
sY1244
sY1280
sY78
C
D
reference
IR3
pDP97
242E13
IR3
E
reference
IR3
15499/15500
pDP97
IR3

An isoYq chromosome generated by this model should bear a mirror-image duplication of the entirety of Yq, the centromere, and the segment of Yp immediately proximal to IR3. To test these predictions, we designed two-color FISH assays and applied them to one of the two cases identified by STS screening. Co-hybridizing centromeric and distal Yq heterochromatin probes to metaphase chromosomes, we observed the predicted duplication of Yq (Figure 5D). With a specific interphase FISH assay, we observed mirror-image duplication of the centromere and of sequences located 28 kb proximal to IR3, all as predicted by our model of isoYq chromosome formation at IR3 (Figures 5E and S7).

Sex phenotype correlates with intercentromeric distance in isoYp chromosomes

In summary, our screen of 2,380 individuals identified 67 cases whose STS content was consistent with the presence of a Y isochromosome formed by the proposed model: 65 isoYp chromosomes with recombination points in palindromes or heterochromatic regions in Yq or the centromere, and two isoYq chromosomes with recombination points in IR3, a Yp inverted repeat. Given the Y chromosome's pivotal role in masculinizing the fetus, we speculated that phenotypic analysis of these 67 individuals might provide retrospective insight into the behavior of isochromosomes *in vivo*, during human development. Specifically, we hypothesized that, among the 65 individuals with isoYp chromosomes, there would exist an inverse correlation between isochromosome size and the likelihood or frequency of male anatomic development.

In isoYq chromosomes formed at IR3, the male-determining gene *SRY*, located on distal Yp, is absent (Figure 5A). As expected, both individuals carrying such isoYq chromosomes were phenotypic females. In contrast, two copies of *SRY*, assayed with sY14, are present on isoYp chromosomes. Despite this, 15 of the 65 individuals with isoYp chromosomes had female external genitalia and had been reared as females. (Of the 50 remaining cases, 49 were phenotypic males, and one individual's phenotype was not recorded.)

In humans, gender assignment is based on visual examination of the external genitalia. In the embryonic gonad, the transient expression of *SRY* in the supporting cell lineage directs the

masculinization of external genitalia. We postulated that the apparent discordance between genotypic sex and phenotypic gender could be explained by a combination of 1) loss of the isoYp chromosome in some cell lineages, including the gonadal supporting cells that express *SRY* in the developing embryo, and 2) retention of the isoYp chromosome in other cell lineages, including lymphocytes from which the cell lines analyzed in this study derive. Such individuals would have female external genitalia and would be mosaic for an isoYp chromosome. (Our PCR-based screen would have precluded further study of individuals displaying the inverse – loss of the isoYp chromosome in lymphocytes and retention in gonadal supporting cells – since this screen required detection of Y-chromosome sequences in DNA extracted from blood cells.)

Dicentric chromosomes are mitotically unstable unless the two centromeres are sufficiently close together[27]. In such cases, stability may be achieved either because the two centromeres function as a single centromere, or because their close proximity constrains them physically from attaching to opposite spindle poles. Dicentric chromosomes with large intercentromeric distances remain unstable until one centromere becomes functionally inactivated[28-30]. We therefore predicted a positive correlation between isochromosome size and the likelihood of a female phenotype.

We compared the distributions of intercentromeric distances in 15 females and 49 males with isoYp chromosomes (Figure 6). Intercentromeric distances were on average significantly higher in females than in males (25.5 Mb, 95% CI 21.5–31.5 Mb versus 13.2 Mb, 95% CI 9.9–17.0 Mb; 95% CI determined by bootstrapping method). Overall, the distributions in females and males were significantly different (Wilcoxon two-sample test, $p < 0.00001$). Whereas recombination points in seven of 15 females were in P1, the most distal Yq palindrome, the majority of recombination points in males were in the centromere and proximal Yq palindromes P5 and P4. All 11 cases with recombination points in the centromere, and therefore bearing monocentric isoYp chromosomes, were male. In sum, the analysis of isochromosome size confirmed the prediction of a correlation between intercentromeric distance and phenotypic gender.

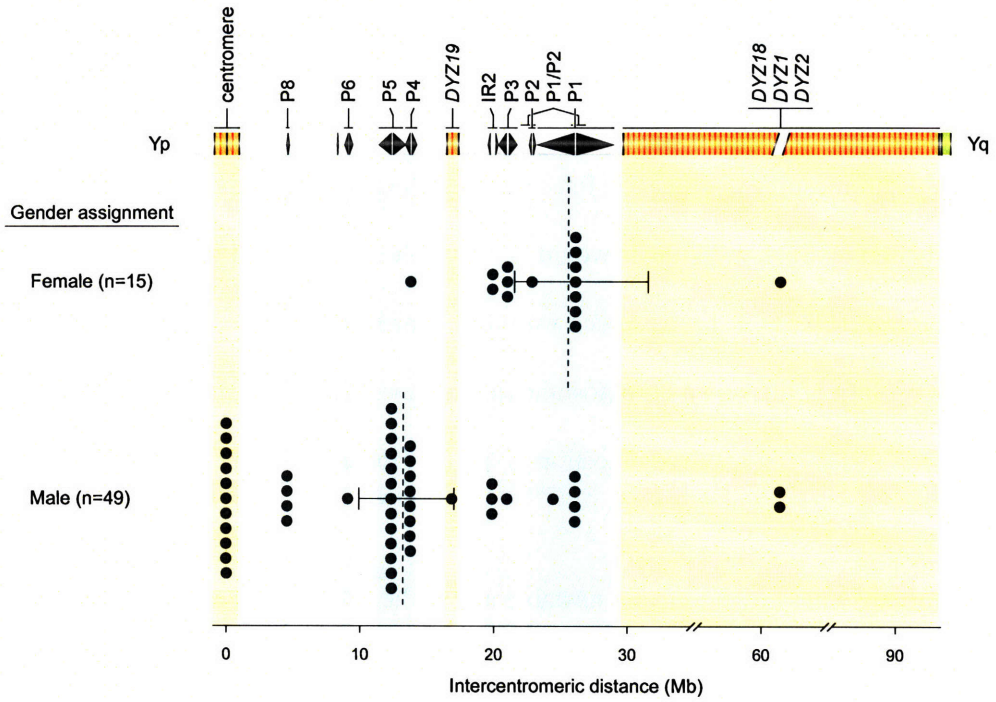


Figure 6. Distributions of intercentromeric distances in females and males carrying isoYp chromosomes. Of the 65 individuals carrying isoYp chromosomes, 15 are female and 49 are male. (The gender of one individual is unknown). This plot displays the distance between the two centromeres in each of these 15 females (above) and 49 males (below). Dotted line indicates the mean intercentromeric distance for each gender. 95% CI calculated by bootstrapping method. The difference in the distributions is significant by Wilcoxon two-sample test.

Discussion

Our findings, combined with previous results, reveal fundamental, mechanistic parallels between recombination within the human MSY and meiotic recombination between homologous chromosomes as studied in yeast and other experimental systems. The resulting model of MSY recombination provides a broad framework for understanding much of the biology of the Y chromosome in health and disease.

A model of MSY recombination

According to our model, isochromosome formation and palindrome maintenance in the MSY are alternative products of a common mechanism that is reminiscent of conventional meiotic recombination between homologous chromosomes (Figure 1C). That common mechanism involves 1) the generation of a DNA double-strand break (DSB) within one arm of an MSY palindrome, and 2) homologous repair of that DSB using the opposing arm of the same palindrome as template. As in conventional meiotic recombination between homologs, MSY recombination intermediates generated during DSB repair can be resolved with or without crossing over. Resolution with crossing over (CO) yields an isochromosome (when recombination occurs between sister chromatids) or an inversion within the palindrome (when recombination occurs within a chromatid). Alternatively, resolution without crossing over (NCO) results in gene conversion and palindrome maintenance. This model provides an economical explanation for all of the following: 1) our present evidence of Y isochromosome formation by crossing over between opposing arms of palindromes, 2) previous evidence of gene conversion at these same palindromes[4,5], and 3) previous evidence of repeated inversion, during human history, of the 3.6-Mb spacer of the IR3 inverted repeats on Yp[3,20,23-26]. As reported here, isochromosomes appear to have arisen via this mechanism at nine of the 10 large palindromes and inverted repeats identified on the human Y chromosome, suggesting the generality of the model.

An even more inclusive model (Figure 7) stems from the realization that, on the human Y chromosome, many palindromic sequences are repeated elsewhere on the chromosome in direct (as

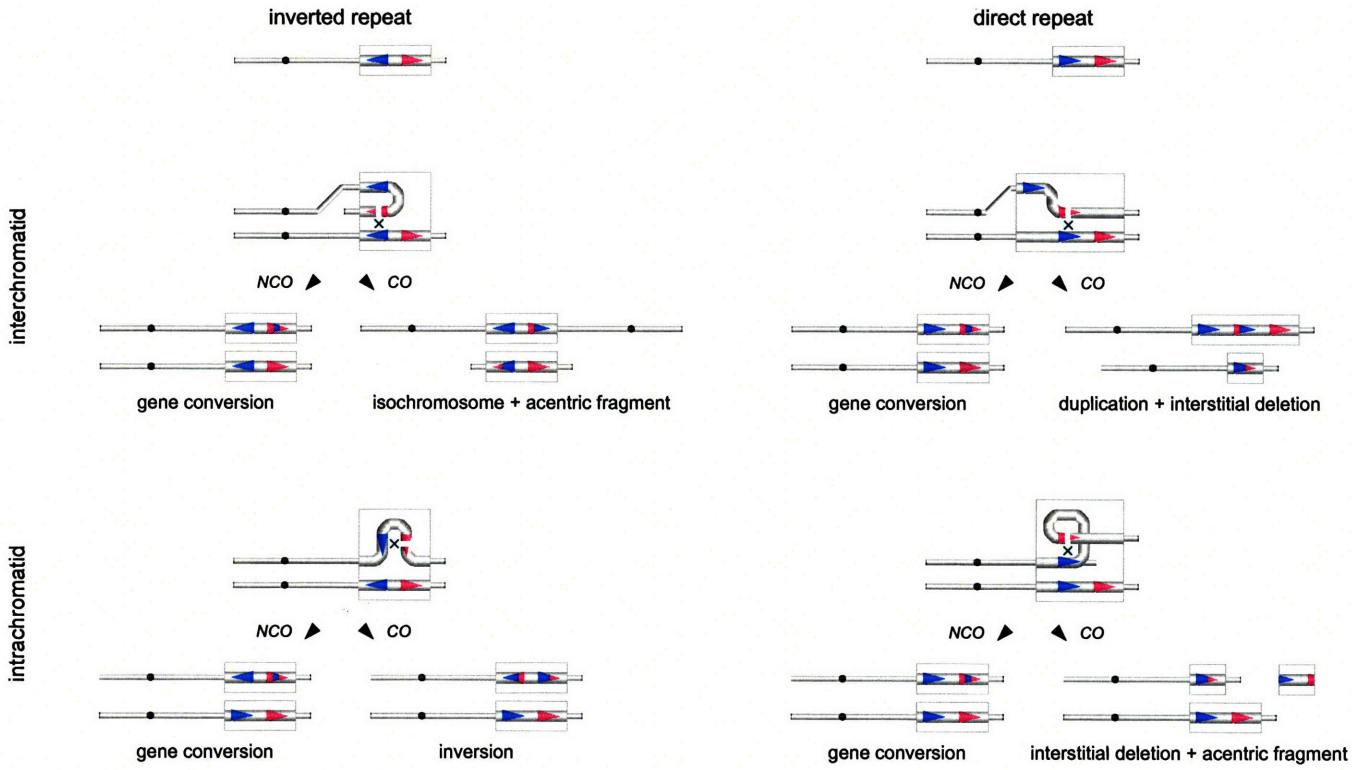


Figure 7. Model of double-strand break (DSB) resolution in MSY direct repeats and palindromes (inverted repeats) by noncrossover (NCO) and crossover (CO) pathways.

Sequence identity between direct repeats, and between palindrome arms, is maintained by NCO resolution of DSBs; such gene conversion could occur within a single chromatid or between sister chromatids. In the case of direct repeats, CO resolution of DSBs results in interstitial deletions; when such events involve sister chromatids, tandem duplications may also result. In the case of palindromes or inverted repeats, CO resolution within a chromatid results in inversion, while CO resolution between sister chromatids results in isochromosome formation, as reported here.

opposed to inverted) orientation[3,16]. What if a DSB arising within one arm of a palindrome were homologously repaired using the more geographically distant, direct repeat as template? Resolution with crossing over (CO) would then yield an interstitially deleted Y chromosome lacking the region between the direct repeats – precisely as observed in many men with spermatogenic failure[16,17,31-34]. Alternatively, resolution without crossing over (NCO) would result in gene conversion between the direct repeats. Such NCO events could explain the strikingly high sequence identity observed between geographically distant direct repeats on the human Y chromosome[3]. In sum, many aspects of human Y chromosome structure, behavior, and evolution are economically explained by DSBs arising within palindrome arms. Homologous repair of these DSBs yields diverse classes of outcomes, these being determined by three binary variables in the process of repair: 1) NCO versus CO pathways, 2) inverted versus direct-repeat templates, and 3) interchromatid versus sister chromatid templates (Figure 7).

Medical consequences of Y isochromosome formation

In addition to sex reversal, the Y isochromosomes we describe in this study are associated with other reproductive disorders, including male infertility and Turner syndrome. Of 49 men with isoYp chromosomes, 20 were ascertained on the basis of spermatogenic failure. A decrease in sperm production may be due to several factors that are not mutually exclusive: loss of genes, effects on chromosome pairing during meiosis, or insufficient expression of MSY genes in the testis. Most isoYp chromosomes lack many genes (Table S2), and in almost all cases, the deleted segments encompass interstitial deletions known to severely affect spermatogenic fitness[15-17,31-35]. Spermatogenesis may be affected by a disruption of normal meiotic pairing between X and Y chromosomes due to duplication of one pseudoautosomal region (PAR) and loss of the other PAR disrupts normal meiotic pairing between X and Y chromosomes[36]. This may explain why men with isoYp chromosomes formed at P1 are infertile despite carrying a full complement of MSY genes. Alternatively, mitotic instability could lead to isochromosome loss in male germ cells and an overall decrease in the expression of genes required for spermatogenesis.

Five of 15 phenotypic females and one phenotypic male exhibited features of Turner syndrome, which is most frequently associated with an "XO" karyotype. Isochromosome instability may have led to mosaicism for XO cells, and the diversity of Turner features among these six individuals is likely due to differences in the lineage distribution of Y isochromosome-bearing cells. Maternal X dominance of XO karyotypes in Turner syndrome patients above the expected two-thirds supports the hypothesis that pre-zygotic Y isochromosome formation followed by post-zygotic loss account for a subset of XO individuals[37]. The frequency of Y isochromosome formation may be higher than observed.

Future study of isochromosomes

Now that we are beginning to understand the importance of Y isochromosomes and the mechanisms by which they form, it will be useful to integrate them into studies of sex-linked phenotypes. A previous study[38] has demonstrated that some Y isochromosomes remain undetected. In our study, the disparate sensitivities of PCR-based and cytogenetic techniques were apparent in several cases with high-grade mosaicism for a 45,X cell line. The presence of Y isochromosomes may henceforth be identified with a series of landmark STS assays (Figure 8, Table S3) and FISH assays, as we have described here. Our findings also offer a mechanistic basis for other human isochromosomes. As research into segmental duplications uncovers palindromes throughout the human genome, their involvement in isochromosome formation should be investigated.

Figure 8. Molecular signatures of 21 isoYp chromosomes and two isoYq chromosomes predicted to form by 23 distinct ectopic homologous recombination events involving MSY palindromes, inverted repeats, and highly repetitive heterochromatic regions.

(A) Predicted STS results.

(B) Predicted deletion maps. See Figure S4 for isoYp chromosomes with breakpoints in palindrome P1.

	sY14	sY1241	sY1242	sY1243	sY1244	sY1280	sY78	sY1200	sY1688	sY1689	sY1251	sY1303	sY1223	sY1222	sY1274	sY1312	sY1310	sY1311	sY1304	sY1275	sY1285	sY1286	sY1276	sY1264	sY1227	sY1228	sY1283	sY1225	sY1226	sY1277	sY1252	sY1253	sY1315	sY1279	sY1278	sY1259	sY1196	sY1197	sY1572	sY1198	sY579	sY1291	sY1206	sY1201	sY1246	sY1166	sY1273
IsoYp chromosomes																																																	
DYZ3	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
proximal *DYZ17*	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
distal *DYZ17*	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
P8	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
P7	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
P6	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
P5	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
P4	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
DYZ19	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
IR2	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
P3	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
P2	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
b3/b2 - proximal P1/P3	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
IR5 - proximal P1/P5	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
P1.2	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
rg/gr - P1/P2	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
P1	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
P1.1	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
IR5 - distal P1/P5	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
b4/b1 - distal P1/P3	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
DYZ18/DYZ1/DYZ2	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
IsoYq chromosomes																																															
IR3	-	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
DYZ3	-	-	-	-	-	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+

B

Yp
Yq
sY14
IR3
DYZ3
proximal DYZ17
distal DYZ17
P8
P7
P6
P5
P4
DYZ19
IR2
P3
P2
P1
DYZ18
DYZ1
DYZ2
sY78
sY1200
sY1688
sY1689
sY1251
sY1303
sY1223
sY1222
sY1274
sY1312
sY1310
sY1311
sY1304
sY1275
sY1285
sY1286
sY1276
sY1264
sY1227
sY1228
sY1283
sY1225
sY1226
sY1277
sY1252
sY1253
sY1315
sY1279
sY1278
sY1259
sY1196
sY1197
sY1572
sY1198
sY579
sY12991
sY1572
sY1206
sY1198
sY579
sY1198
sY1206
sY1572
sY1201
sY1246
sY1166
sY1273
IsoYp chromosomes
DYZ3
proximal DYZ17
distal DYZ17
P8
P7
P6
P5
P4
DYZ19
IR2
P3
P2
b3/b2 - proximal P1/P3
IR5 - proximal P1/P5
P1.2
rg/gr - P1/P2
P1
P1.1
IR5 - distal P1/P5
b4/b1 - distal P1/P3
DYZ18/DYZ1/DYZ2
IsoYq chromosomes
IR3
DYZ3
sY14
sY1241
sY1242
sY1243
sY1244
sY1280
sY78
sY1273

Methods

Y-DNA Marker Analysis

Mapping of Y-chromosome deletion breakpoints was carried out using MSY Breakpoint Mapper, an online database of Y-specific STSs[18]. Table S4 lists all STSs employed in this study, together with their GenBank accession numbers.

Fluorescence *in situ* hybridization

Metaphase and interphase nuclei were hybridized with labeled DNA probes by methods previously described[39]. In each interphase experiment that involved the counting of sites of hybridization, ≥100 nuclei were scored. Table S5 provides information on all FISH assays and probes.

Acknowledgements

We thank the following individuals for human DNA or blood samples, or cell lines:
M. Alamares, G. Beverstock, H. Bode, J. Bodurtha, A. Brothman, B. Clark, J. Crawford, A. Daniel, A. de la Chapelle, A. Donnenfeld, P. Donohoue, S. Fallet, G. Feole, L.M. Frank, S. Hahm, M. Hajianpour, K. Harper, D. Hoar, A. Hughes, J. Jacobs, M. Jahmedor, L. Jenkins, L. Layman, W. Long, D. Manchester, P. McDonough, B. McGillivray, R. McVie, M. Mennuti, W. Miller, A. Milunsky, P. Murphy, C. Palmer, S. Patil, S. Peacock, M. Phelan, J. Priest, J.B. Ravnan, C. Rudlin, L. Shapiro, J. Shulkin, J. Siegel, E. Simpson, D. Solomon, C. Stevens, P. Storto, R. Stratton, M. Summers, U. Surti, J. Tollis, J. Vigorita, N. Wang, M. Watson, L. Weiss, V. Wilson, P. Wyatt, T. Yang-Feng, J. Zenger Hain, and J. Zunich.
Supported by the National Institutes of Health, the Howard Hughes Medical Institute, the Netherlands Organization for Scientific Research, and the Academic Medical Center. J. Lange was the recipient of a Boehringer Ingelheim fellowship.

References

1. Lahn BT, Page DC (1999) Four evolutionary strata on the human X chromosome. Science 286:964-967

2. Charlesworth B, Charlesworth D (2000) The degeneration of Y chromosomes. Philos Trans R Soc Lond B Biol Sci 355:1563-1572

3. Skaletsky H, Kuroda-Kawaguchi T, Minx PJ, Cordum HS, Hillier L, Brown LG, Repping S, et al. (2003) The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. Nature 423:825-837

4. Rozen S, Skaletsky H, Marszalek JD, Minx PJ, Cordum HS, Waterston RH, Wilson RK, Page DC (2003) Abundant gene conversion between arms of palindromes in human and ape Y chromosomes. Nature 423:873-876

5. Adams SM, King TE, Bosch E, Jobling MA (2006) The case of the unreliable SNP: recurrent back-mutation of Y-chromosomal marker P25 through gene conversion. Forensic Sci Int 159:14-20

6. Szostak JW, Orr-Weaver TL, Rothstein RJ, Stahl FW (1983) The double-strand-break repair model for recombination. Cell 33:25-35

7. Allers T, Lichten M (2001) Differential timing and control of noncrossover and crossover recombination during meiosis. Cell 106:47-57

8. Börner GV, Kleckner N, Hunter N (2004) Crossover/noncrossover differentiation, synaptonemal complex formation, and regulatory surveillance at the leptotene/zygotene transition of meiosis. Cell 117:29-45

9. Guillon H, Baudat F, Grey C, Liskay RM, de Massy B (2005) Crossover and noncrossover pathways in mouse meiosis. Mol Cell 20:563-573

10. Jacobs PA, Ross A (1966) Structural abnormalities of the Y chromosome in man. Nature 210:352-354

11. Hsu LY (1994) Phenotype/karyotype correlations of Y chromosome aneuploidy with emphasis on structural aberrations in postnatally diagnosed cases. Am J Med Genet 53:108-140

12. Vergnaud G, Page DC, Simmler MC, Brown L, Rouyer F, Noel B, Botstein D, de la Chapelle A, Weissenbach J (1986) A deletion map of the human Y chromosome based on DNA hybridization. Am J Hum Genet 38:109-124

13. Vollrath D, Foote S, Hilton A, Brown LG, Beer-Romero P, Bogan JS, Page DC (1992) The human Y chromosome: a 43-interval map based on naturally occurring deletions. Science 258:52-59

14. Lahn BT, Ma N, Breg WR, Stratton R, Surti U, Page DC (1994) Xq-Yq interchange resulting in supernormal X-linked gene expression in severely retarded males with 46,XYq- karyotype. Nat Genet 8:243-250

15. Reijo R, Lee TY, Salo P, Alagappan R, Brown LG, Rosenberg M, Rozen S, Jaffe T, Straus D, Hovatta O, et al. (1995) Diverse spermatogenic defects in humans caused by Y chromosome deletions encompassing a novel RNA-binding protein gene. Nat Genet 10:383-393

16. 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

17. Repping S, Skaletsky H, Lange J, Silber S, Van Der Veen F, Oates RD, Page DC, Rozen S (2002) Recombination between palindromes P5 and P1 on the human Y chromosome causes massive deletions and spermatogenic failure. Am J Hum Genet 71:906-922

18. Lange J, Skaletsky H, Bell GW, Page DC (2007) MSY Breakpoint Mapper, a database of sequence-tagged sites useful in defining naturally occurring deletions in the human Y chromosome. Nucleic Acids Res published online October 26

19. Morton NE (1991) Parameters of the human genome. Proc Natl Acad Sci USA 88:7474-7476

20. Repping S, van Daalen SK, Brown LG, Korver CM, Lange J, Marszalek JD, Pyntikova T, van der Veen F, Skaletsky H, Page DC, Rozen S (2006) High mutation rates have driven extensive structural polymorphism among human Y chromosomes. Nat Genet 38:463-467

21. Kirsch S, Weiss B, Miner TL, Waterston RH, Clark RA, Eichler EE, Munch C, Schempp W, Rappold G (2005) Interchromosomal segmental duplications of the pericentromeric region on the human Y chromosome. Genome Res 15:195-204

22. Darlington CD (1939) Misdivision and the genetics of the centromere. J Genet 37:341-364

23. Page DC (1986) Sex reversal: deletion mapping the male-determining function of the human Y chromosome. Cold Spring Harb Symp Quant Biol 51 Pt 1:229-235

24. Affara NA, Ferguson-Smith MA, Tolmie J, Kwok K, Mitchell M, Jamieson D, Cooke A, Florentin L (1986) Variable transfer of Y-specific sequences in XX males. Nucleic Acids Res 14:5375-5387

25. Jobling MA, Williams GA, Schiebel GA, Pandya GA, McElreavey GA, Salas GA, Rappold GA, Affara NA, Tyler-Smith C (1998) A selective difference between human Y-chromosomal DNA haplotypes. Curr Biol 8:1391-1394

26. Tilford CA, Kuroda-Kawaguchi T, Skaletsky H, Rozen S, Brown LG, Rosenberg M, McPherson JD, Wylie K, Sekhon M, Kucaba TA, Waterston RH, Page DC (2001) A physical map of the human Y chromosome. Nature 409:943-945

27. Therman E, Trunca C, Kuhn EM, Sarto GE (1986) Dicentric chromosomes and the inactivation of the centromere. Hum Genet 72:191-195

28. Hair JB (1953) The origin of new chromosomes in Agropyron. Heredity 6 (Suppl.):215-233

29. Koshland D, Rutledge L, Fitzgerald-Hayes M, Hartwell LH (1987) A genetic analysis of dicentric minichromosomes in Saccharomyces cerevisiae. Cell 48:801-812

30. Sullivan BA, Willard HF (1998) Stable dicentric X chromosomes with two functional centromeres. Nat Genet 20:227-228

31. 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

32. 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

33. 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

34. Repping S, Skaletsky H, Brown L, van Daalen SK, Korver CM, Pyntikova T, Kuroda-Kawaguchi T, de Vries JW, Oates RD, Silber S, van der Veen F, Page DC, Rozen S (2003) Polymorphism for a 1.6-Mb deletion of the human Y chromosome persists through balance between recurrent mutation and haploid selection. Nat Genet 35:247-251

35. Vogt PH, Edelmann A, Kirsch S, Henegariu O, Hirschmann P, Kiesewetter F, Kohn FM, Schill WB, Farah S, Ramos C, Hartmann M, Hartschuh W, Meschede D, Behre HM, Castel A, Nieschlag E, Weidner W, Grone 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

36. Mohandas TK, Speed RM, Passage MB, Yen PH, Chandley AC, Shapiro LJ (1992) Role of the pseudoautosomal region in sex-chromosome pairing during male meiosis: meiotic studies in a man with a deletion of distal Xp. Am J Hum Genet 51:526-533

37. Uematsu A, Yorifuji T, Muroi J, Kawai M, Mamada M, Kaji M, Yamanaka C, Momoi T, Nakahata T (2002) Parental origin of normal X chromosomes in Turner syndrome patients with various karyotypes: implications for the mechanism leading to generation of a 45,X karyotype. Am J Med Genet 111:134-139

38. Kirsch S, Weiss B, De Rosa M, Ogata T, Lombardi G, Rappold GA (2000) FISH deletion mapping defines a single location for the Y chromosome stature gene, GCY. J Med Genet 37:593-599

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

Supplemental Data

Figures

Figure S1. Triangular dot plot of the MSY reference sequence shows intrachromosomal sequence similarities.

Figure S2. STS content of structurally anomalous Y chromosomes in 100 human individuals with deletion breakpoints in the long arm (Yq) or centromere.

Figure S3. Comprehensive results of testing for duplication, in isoYp chromosomes, of sequences immediately proximal to targeted palindromes.

Figure S4. Comprehensive results of testing the copy number, in isoYp chromosomes with breakpoints in palindrome P1, of sequences located between P1 and P5.

Figure S5. Comprehensive results of testing for duplication, in isoYp chromosomes, of sequences immediately proximal to targeted heterochromatic repeats.

Figure S6. STS content of structurally anomalous Y chromosomes in two human individuals with deletion breakpoints in the short arm (Yp).

Figure S7. Comprehensive results of testing for duplication, in an isoYq chromosome, of sequences immediately proximal to Yp inverted repeat IR3.

Tables

Table S1. Deletion breakpoint intervals of 100 human individuals with breakpoints in the long arm (Yq) or centromere and two human individuals with breakpoints in the short arm (Yp) of the MSY.

Table S2. Y chromosome gene content and copy number of 21 isoYp chromosomes and two isoYq chromosomes predicted to form by 23 distinct ectopic homologous recombination events involving palindromes, inverted repeats, and highly repetitive heterochromatic regions.

Table S3. Target size, isochromosome size, and molecular signatures of 21 isoYp chromosomes and two isoYq chromosomes predicted to form by 23 distinct ectopic homologous recombination events involving palindromes, inverted repeats, and highly repetitive heterochromatic regions.

Table S4. STSs used in the analysis of Y chromosome content.

Table S5. Metaphase and interphase FISH assays of Y chromosome structure in individuals with breakpoints in the MSY.

Figure S1. Triangular dot plot of the MSY reference sequence shows intrachromosomal sequence similarities.
Triangular intra-arm dot plot of the MSY reference sequence (adapted from ref.[3]). In this plot, each dot represents a match of >98% within a window of 1000bp. Repeat elements have been masked. Euchromatic repeats are shown in black, heterochromatic repeat arrays in red. Inverted repeats such as palindromes appear as vertical lines, direct repeats as horizontal lines.

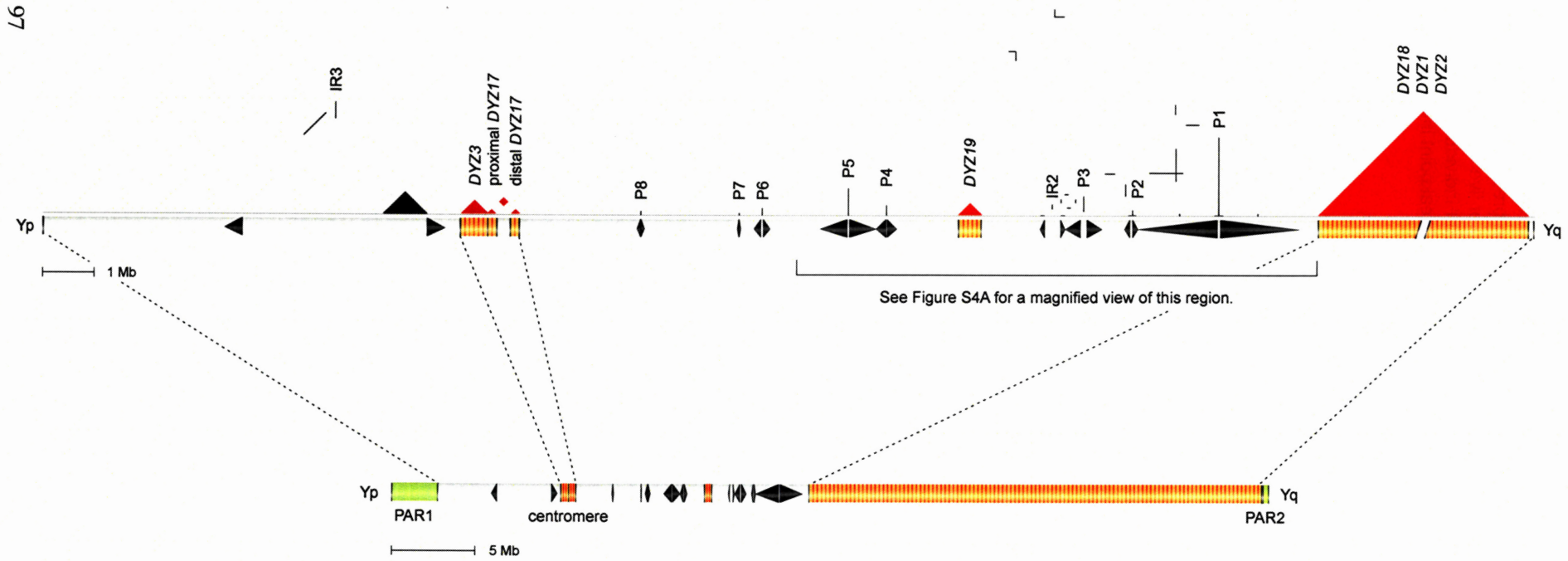


Figure S2. STS content of structurally anomalous Y chromosomes in 100 human individuals with deletion breakpoints in the long arm (Yq) or centromere.

(A) Expanded view of centromere and Yq, indicating locations of eight palindromes (P1 through P8), one inverted repeat (IR2), and five heterochromatic regions containing highly repetitive elements (*DYZ1-3* and *DYZ17-19*).

(B) STSs employed in fine mapping of deletion breakpoints. These STSs include boundary and spacer-flanking markers for each palindrome and inverted repeat, and markers around each of five blocks of heterochromatin.

(C) Results of testing genomic DNAs from 100 individuals with identifiers for presence or absence of STSs. Solid black bars encompass STSs found to be present. Gray bars indicate breakpoint intervals that could not be further narrowed due to cross-amplification at other loci.

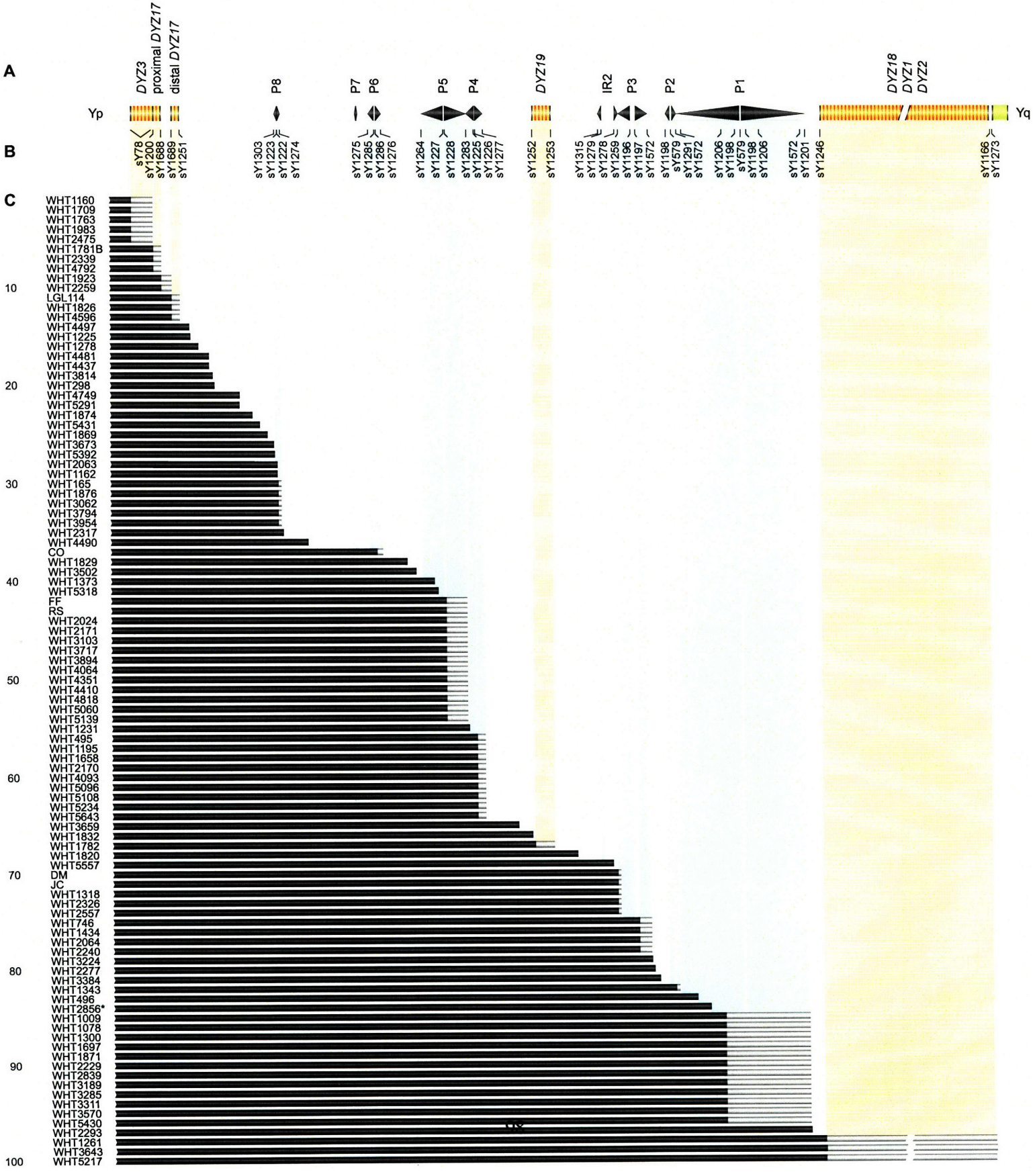


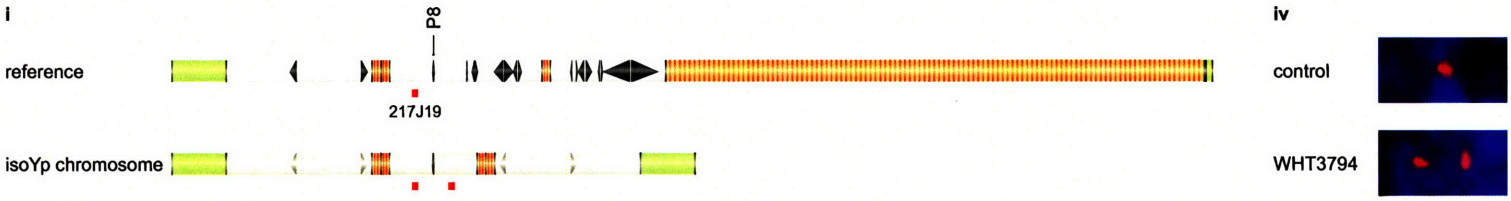
Figure S3. Comprehensive results of testing for duplication, in isoYp chromosomes, of sequences immediately proximal to targeted palindromes.

(A-F) For palindromes P8 (A), P6 (B), P5 (C), P4 (D), IR2 (E), and P3 (F) that, based on STS mapping, were hypothesized to have been targeted by ectopic homologous recombination in the formation of isoYp chromosomes, we designed specific FISH assays to determine, in individuals with isoYp chromosomes, the copy number of sequences located immediately proximal to targeted palindromes.

In each series of figures, we show: (i) schematics of a reference chromosome and of an isoYp chromosome formed by the proposed mechanism at each palindrome, and the locations of probes hybridized to interphase spreads; (ii) counting results tabulated for a control male and for each case tested with the probe indicated; (iii) counting results summarized graphically for each case; and (iv) representative FISH images for a control male and for an individual with an isoYp chromosome to demonstrate common FISH patterns.

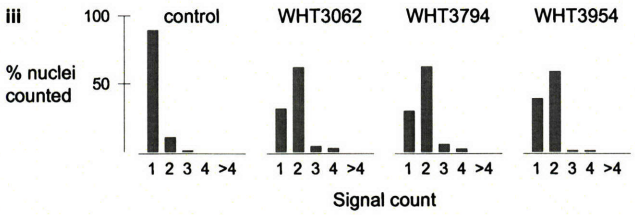
See Figure S4 for isoYp chromosomes formed at inverted repeats in P1.

A P8

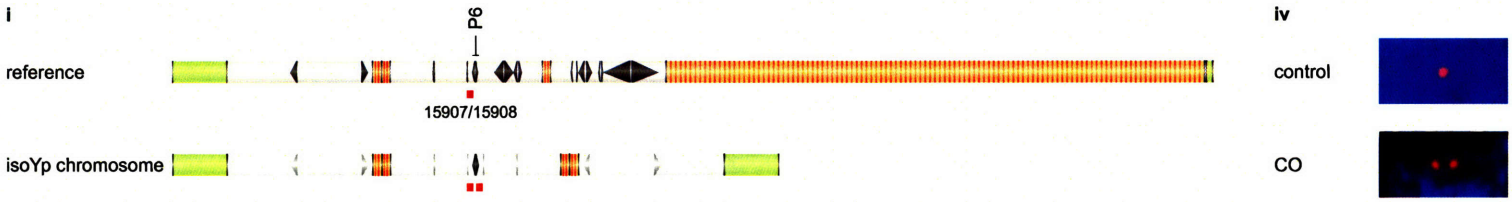


ii 217J19

	Number of nuclei with signal count of 1	2	3	4	>4	Number of sites of hybridization
control	179	21	0	0	0	1
WHT3062	63	124	8	5	0	2
WHT3794	60	125	11	4	0	2
WHT3954	78	118	2	2	0	2



B P6



ii 15907/15908

	Number of nuclei with signal count of 1	2	3	4	>4	Number of sites of hybridization
control	117	32	1	0	0	1
CO	69	50	1	0	0	2

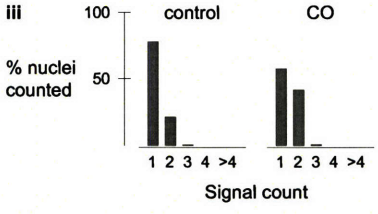
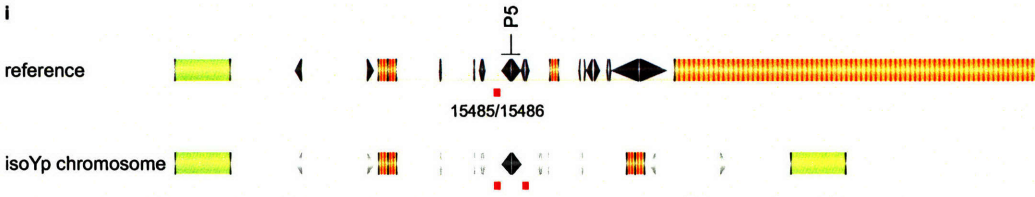
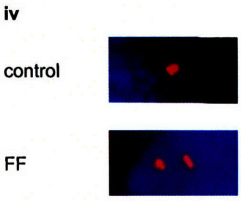


Figure S3 continued. Comprehensive results of testing for duplication, in isoYp chromosomes, of sequences immediately proximal to targeted palindromes.

C P5

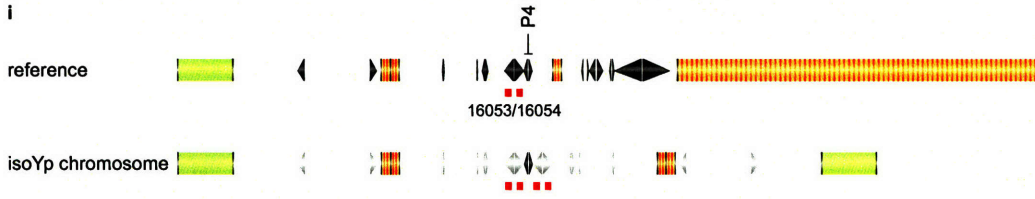


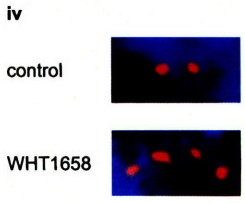


ii 15485/15486

	Number of nuclei with signal count of 1	2	3	4	>4	Number of sites of hybridization
control	187	12	1	0	0	1
FF	78	119	3	0	0	2
RS	81	115	4	0	0	2
WHT2024	102	94	4	0	0	2
WHT2171	79	116	5	0	0	2
WHT3717	100	99	1	0	0	2
WHT4064	81	113	4	2	0	2

D P4





ii 16053/16054

	Number of nuclei with signal count of 1	2	3	4	>4	Number of sites of hybridization
control	103	97	0	0	0	2
WHT1658	14	60	72	54	0	4

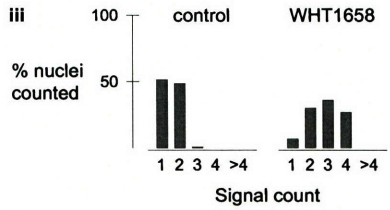
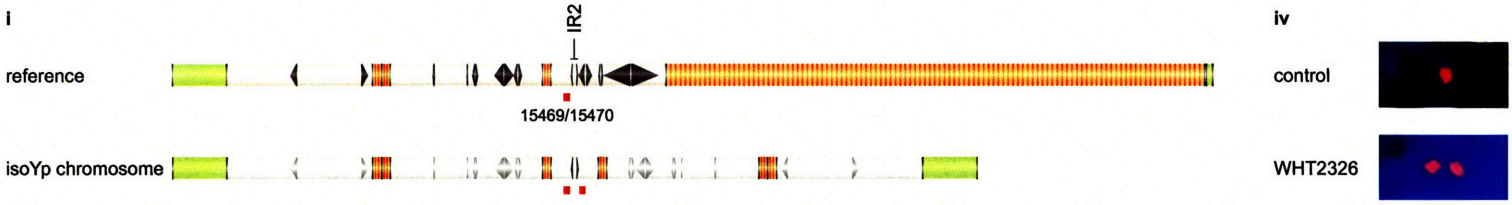


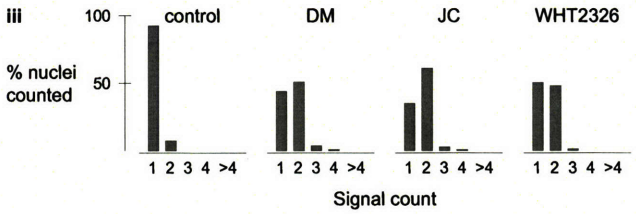
Figure S3 continued. Comprehensive results of testing for duplication, in isoYp chromosomes, of sequences immediately proximal to targeted palindromes.

E IR2

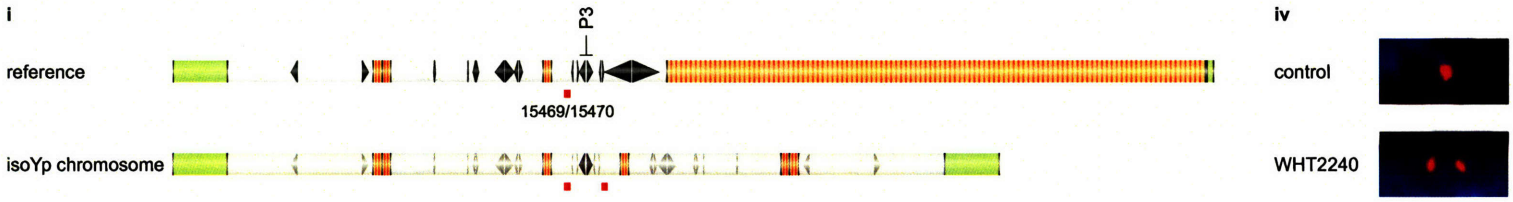


ii 15469/15470

	Number of nuclei with signal count of 1	2	3	4	>4	Number of sites of hybridization
control	185	15	0	0	0	1
DM	88	102	8	2	0	2
JC	70	122	6	2	0	2
WHT2326	101	96	3	0	0	2



F P3



ii 15469/15470

	Number of nuclei with signal count of 1	2	3	4	>4	Number of sites of hybridization
control	185	15	0	0	0	1
WHT1434	68	120	8	4	0	2
WHT2064	55	129	13	3	0	2
WHT2240	55	137	6	2	0	2

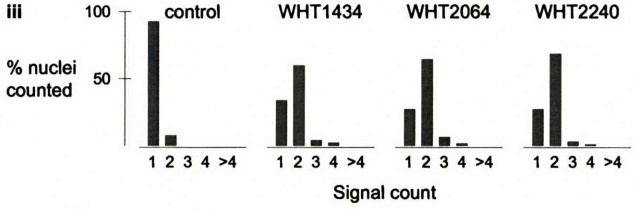


Figure S4. Comprehensive results of testing the copy number, in isoYp chromosomes with breakpoints in palindrome P1, of sequences located between P1 and P5.

Large blocks of palindrome P1 are repeated in palindromes P2, P3, and P5. Like opposing arms of palindromes, these inverted repeats could serve as targets for the proposed model. In each case displaying an isoYp chromosome with a breakpoint in P1, the inverted repeats that were targeted in the formation of the isoYp chromosome can be determined by STS mapping and specific FISH assays.

(A) Triangular dot plot of the region of the MSY long arm spanned by palindromes P5 and P1, which features many amplicons in both inverted and direct orientation (plot adapted from ref.[3]). Color segmentation of the palindromes highlights the complex ampliconic structure of this region. In this plot, each dot represents a match of >98% within a window of 1000bp. Repeat elements have been masked. Euchromatic repeats are shown in black, heterochromatic repeat arrays in red. Inverted repeats including palindromes appear as vertical lines, direct repeats as horizontal lines.

(B) Ampliconic structure of the region spanned by P5 and P1 in the reference MSY and in isoYp chromosomes predicted to form by eight distinct ectopic homologous recombination events between inverted repeats in this region. Shown are (i) the reference sequence spanning P5 and P1, and the target sequences on sister chromatids and ampliconic structures of isoYp chromosomes predicted to form at the following inverted repeats: (ii) b3/b2 - proximal P1/P3; (iii) IR5 - proximal P1/P5; (iv) P1.2; (v) rg/gr - P1/P2; (vi) P1; (vii) P1.1; (viii) IR5 - distal P1/P5; and (ix) b4/b1 - distal P1/P3. Predicted isoYp chromosomes can be distinguished by the results of three STSs tested on genomic DNA - sY1291, sY1206, and sY1201 - and three FISH probes hybridized to interphase spreads - 1325K3, 18E8, and 100J21.

(C) Summary of predicted STS mapping and FISH results for each proposed isoYp chromosome shown in (B).

(D-F) Six individuals with isoYp chromosomes that, by mapping of STSs including sY1291, sY1206, and sY1201, displayed breakpoints in P1, were tested with three FISH probes to determine which inverted repeat had been targeted in each case. For each of three FISH probes 1325K3 (D), 18E8 (E), and 100J21 (F), we show: (i) counting results tabulated for a control male and for each of six cases tested; (ii) counting results summarized graphically for each case; and (iii) representative FISH images for a control and two individuals with isoYp chromosomes to demonstrate common FISH patterns.

(G) Summary of observed STS mapping and FISH results in (D-F). By comparing observed results with predicted results in (C), we inferred which inverted repeat was targeted in the formation of each isoYp chromosome. Five of the six isoYp chromosomes had formed between opposing arms of P1, and one had formed between inverted repeats rg/gr in P1 and P2.

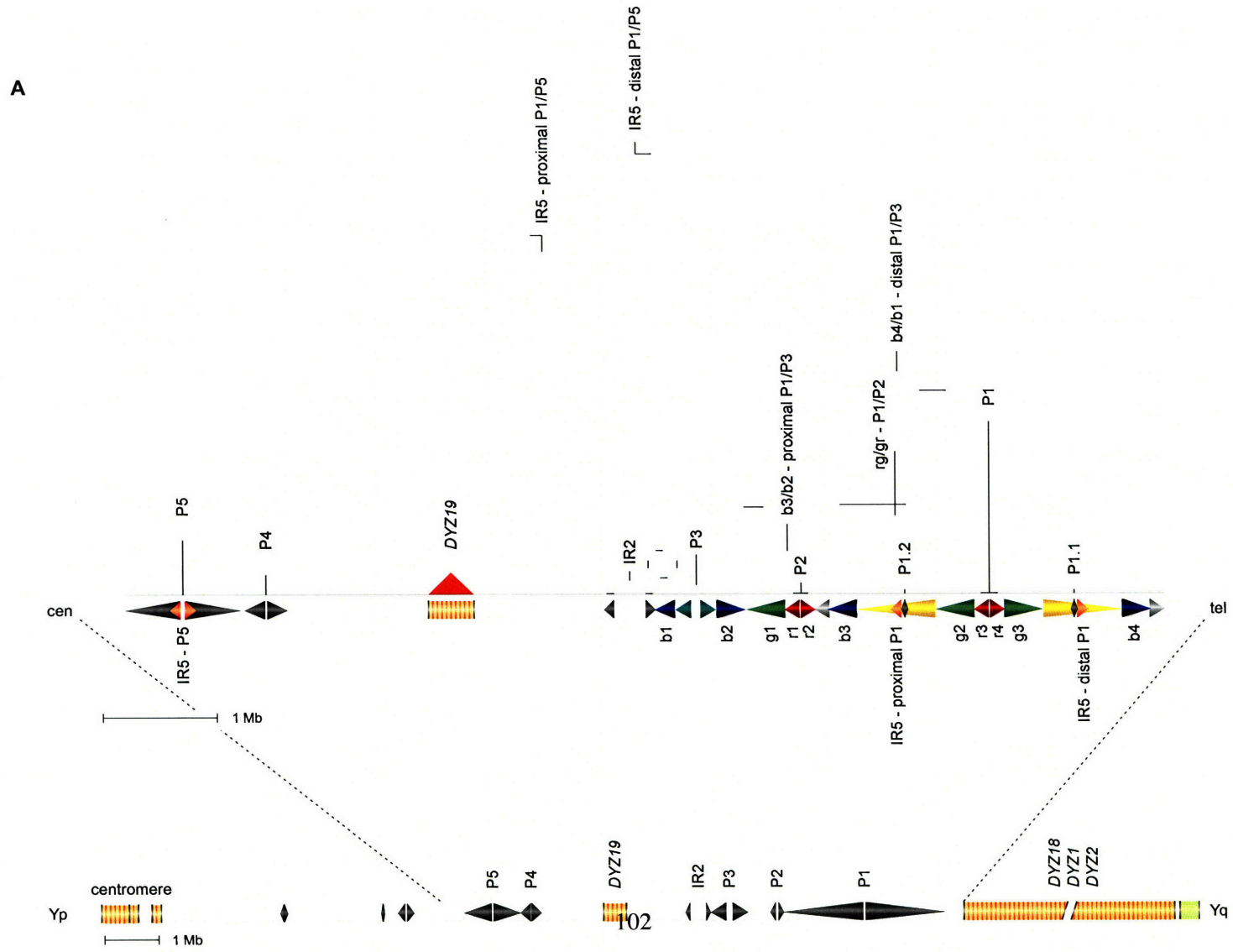
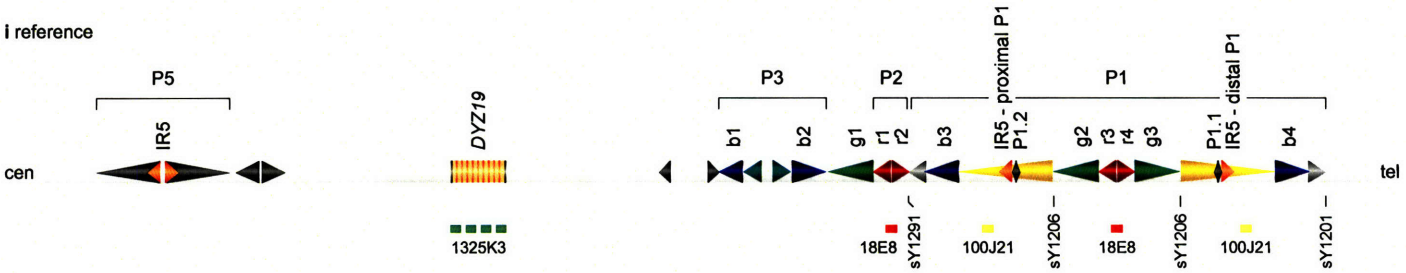
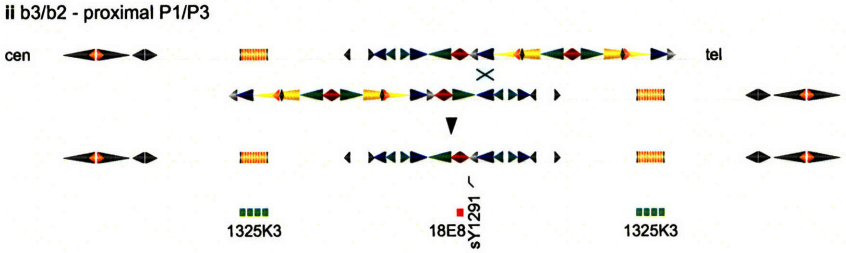
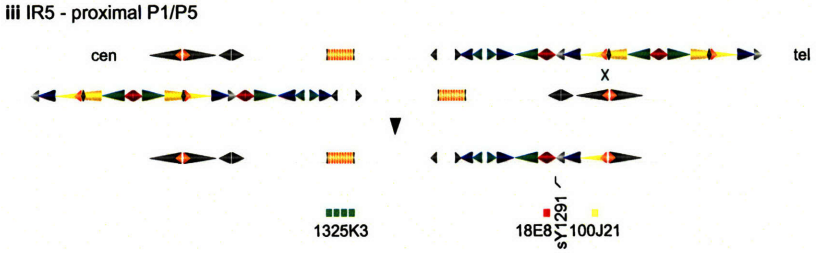


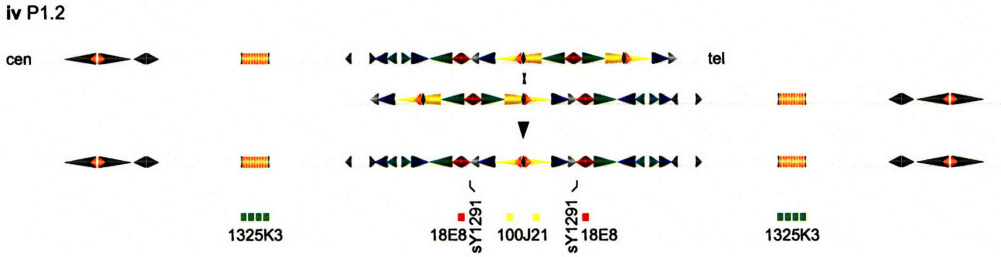
Figure S4 continued. Comprehensive results of testing the copy number, in isoYp chromosomes with breakpoints in palindrome P1, of sequences located between P1 and P5.

B









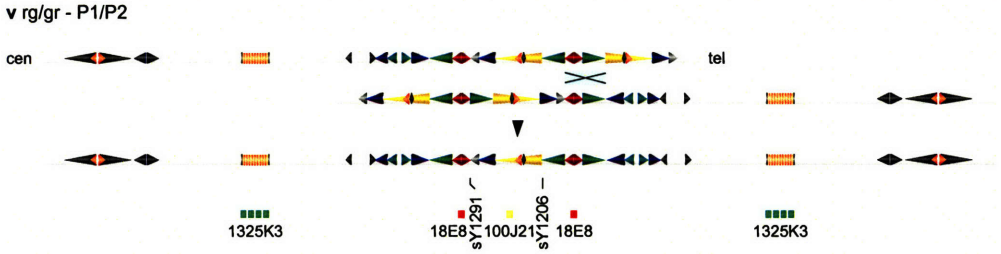
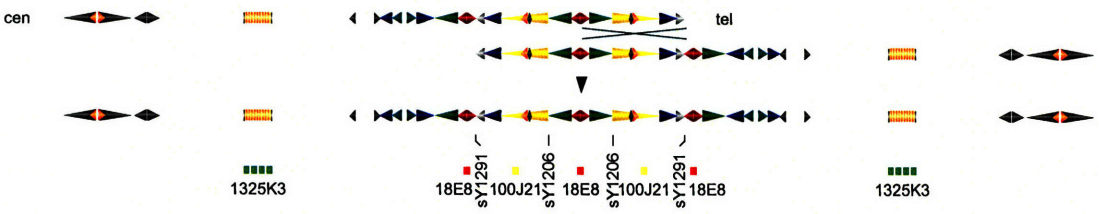


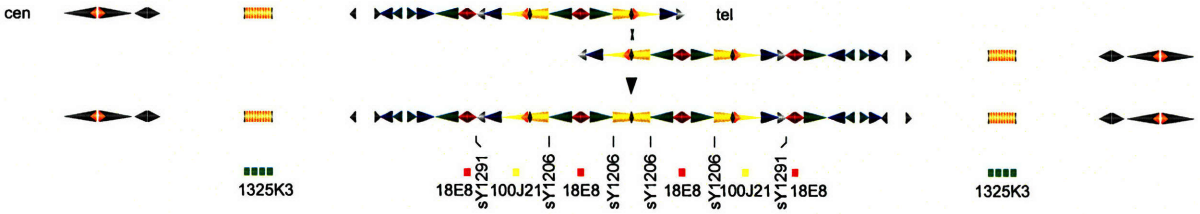
Figure S4 continued. Comprehensive results of testing the copy number, in isoYp chromosomes with breakpoints in palindrome P1, of sequences located between P1 and P5.

B continued

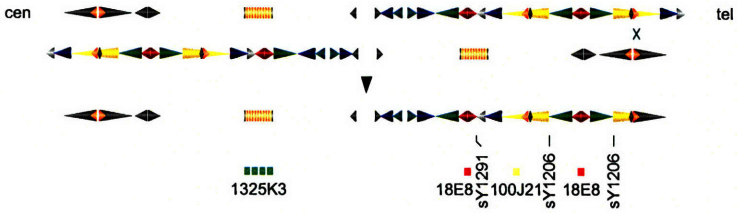
vi P1



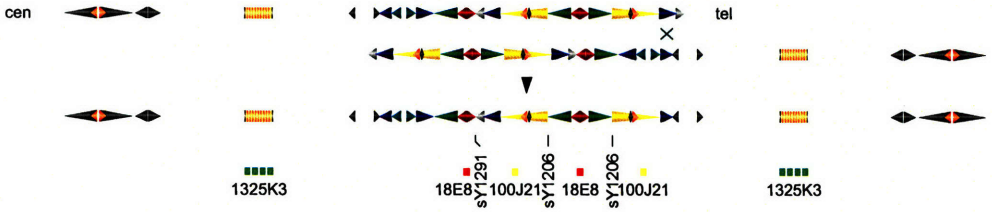
vii P1.1



viii IR5 - distal P1/P5



ix b4/b1 - distal P1/P3



C Summary of predicted STS mapping and FISH results

	Presence or absence of each STS			Number of copies present for each probe		
	sY1291	sY1206	sY1201	1325K3	18E8	100J21
reference	+	+	+	1	2	2
b3/b2 - proximal P1/P3	+	-	-	2	1	0
IR5 - proximal P1/P5	+	-	-	1	1	1
P1.2	+	-	-	2	2	2
rg/gr - P1/P2	+	+	-	2	2	1
P1	+	+	-	2	3	2
P1.1	+	+	-	2	4	2
IR5 - distal P1/P5	+	+	-	1	2	1
b4/b1 - distal P1/P3	+	+	-	2	2	2

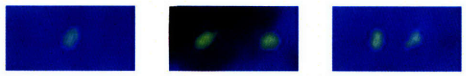
Figure S4 continued. Comprehensive results of testing the copy number, in isoYp chromosomes with breakpoints in palindrome P1, of sequences located between P1 and P5.

D 1325K3

i

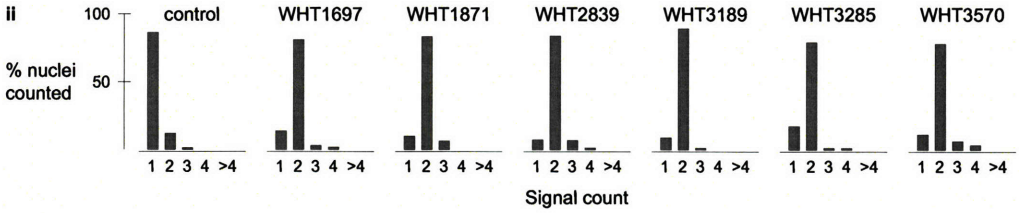
	Number of nuclei with signal count of 1	2	3	4	>4	Number of sites of hybridization
control	173	24	3	0	0	1
WHT1697	28	162	6	4	0	2
WHT1871	20	167	13	0	0	2
WHT2839	15	168	14	3	0	2
WHT3189	18	179	3	0	0	2
WHT3285	35	159	3	3	0	2
WHT3570	23	157	13	7	0	2

iii



Note: Due to high levels of XO mosaicism in several individuals, nuclei with signal counts of zero were counted but excluded in the determination of the number of sites of hybridization.

ii

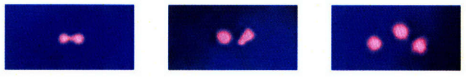


E 18E8

i

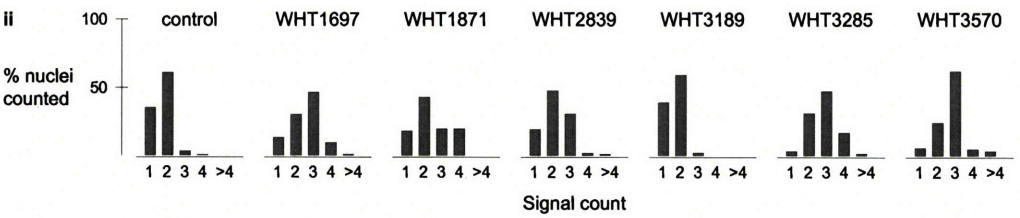
	Number of nuclei with signal count of 1	2	3	4	>4	Number of sites of hybridization
control	68	118	6	1	0	2
WHT1697	21	48	74	15	1	3
WHT1871	10	24	11	11	0	3
WHT2839	45	112	72	4	2	3
WHT3189	87	132	4	0	0	2
WHT3285	6	62	94	33	3	3
WHT3570	8	38	92	7	5	3

iii



Note: Due to high levels of XO mosaicism in several individuals, nuclei with signal counts of zero were counted but excluded in the determination of the number of sites of hybridization.

ii

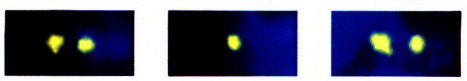


F 100J21

i

	Number of nuclei with signal count of 1	2	3	4	>4	Number of sites of hybridization
control	45	146	8	1	0	2
WHT1697	70	115	3	0	0	2
WHT1871	35	63	1	1	0	2
WHT2839	51	89	5	0	0	2
WHT3189	136	14	0	0	0	1
WHT3285	63	125	6	2	0	2
WHT3570	51	138	10	1	0	2

iii



Note: Due to high levels of XO mosaicism in several individuals, nuclei with signal counts of zero were counted but excluded in the determination of the number of sites of hybridization.

ii

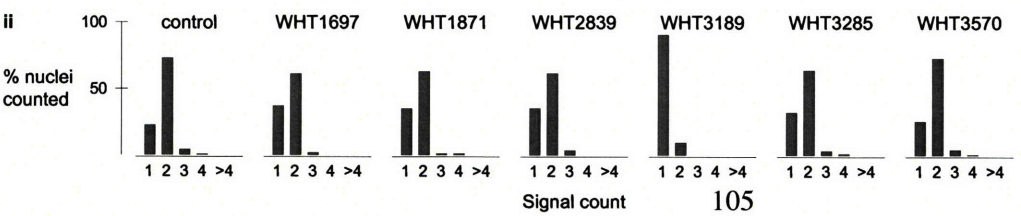


Figure S4 continued. Comprehensive results of testing the copy number, in isoYp chromosomes with breakpoints in palindrome P1, of sequences located between P1 and P5.

G Summary of observed STS mapping and FISH results

	Presence or absence of each STS			Number of copies present for each probe			
	sY1291	sY1206	sY1201	1325K3	18E8	100J21	Inferred targeted repeat
control	+	+	+	1	2	2	reference
WHT1697	+	+	-	2	3	2	P1
WHT1871	+	+	-	2	3	2	P1
WHT2839	+	+	-	2	3	2	P1
WHT3189	+	+	-	2	2	1	rg/gr - P1/P2
WHT3285	+	+	-	2	3	2	P1
WHT3570	+	+	-	2	3	2	P1

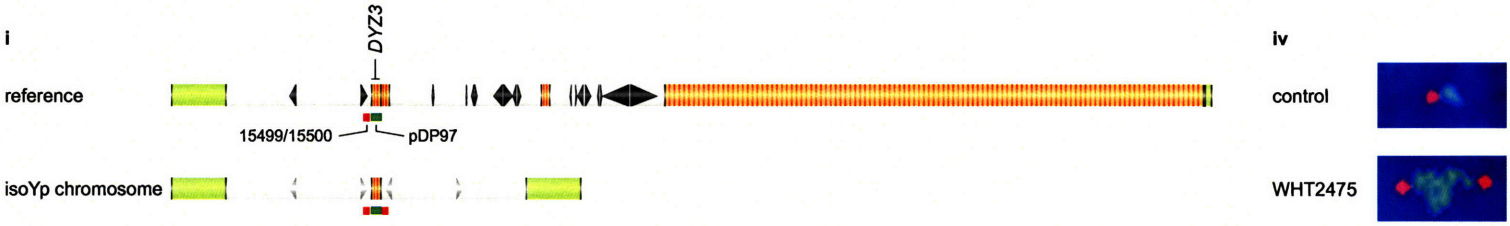
Figure S5. Comprehensive results of testing for duplication, in isoYp chromosomes, of sequences immediately proximal to targeted heterochromatic repeats.

(A-C) For heterochromatic repeats *DYZ3* (A), proximal *DYZ17* (B), and *DYZ18/DYZ1/DYZ2* (C) that, based on STS mapping, were hypothesized to have been targeted by ectopic homologous recombination in the formation of isoYp chromosomes, we designed specific FISH assays to determine, in individuals with isoYp chromosomes, the copy number of sequences located immediately proximal to targeted heterochromatic repeats.

In the series of figures (A and B), we show: (i) schematics of a reference chromosome and of an isoYp chromosome formed by the proposed mechanism at each heterochromatic repeat, and the locations of probes hybridized to interphase spreads; (ii) counting results tabulated for a control male and for each case tested with the probe indicated; (iii) counting results summarized graphically for each case; and (iv) representative FISH images for a control male and for an individual with an isoYp chromosome to demonstrate common FISH patterns.

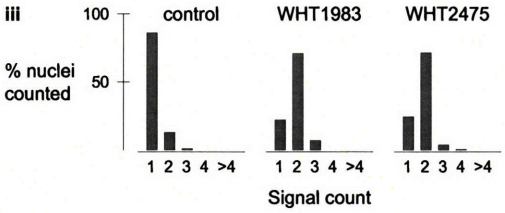
In the series of figures (C), we show: (i) schematics of a reference chromosome and of an isoYp chromosome formed by the proposed mechanism at *DYZ18/DYZ1/DYZ2*, and the locations of two probes hybridized to metaphase spreads; (ii) conclusions on the number of signals for probe 1136L22 for a control male and for each tested case; and (iii) representative FISH images for a control male and for an individual with an isoYp chromosome to demonstrate common FISH patterns.

A *DYZ3*

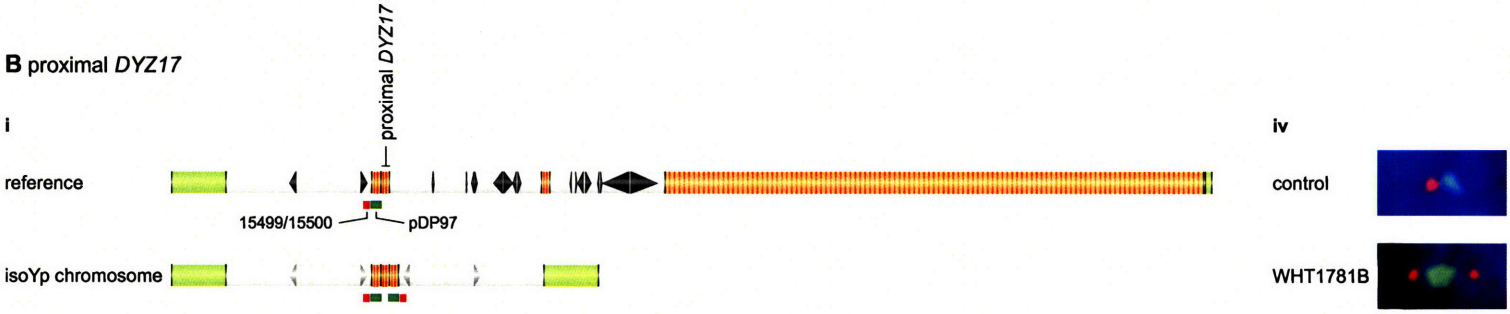


ii 15499/15500

	Number of nuclei with signal count of 1	2	3	4	>4	Number of sites of hybridization
control	172	26	2	0	0	1
WHT1983	44	142	14	0	0	2
WHT2475	49	143	7	1	0	2



B proximal *DYZ17*



ii 15499/15500

	Number of nuclei with signal count of 1	2	3	4	>4	Number of sites of hybridization
control	172	26	2	0	0	1
WHT1781B	125	73	2	0	0	2

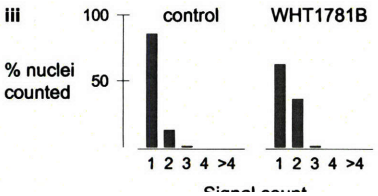
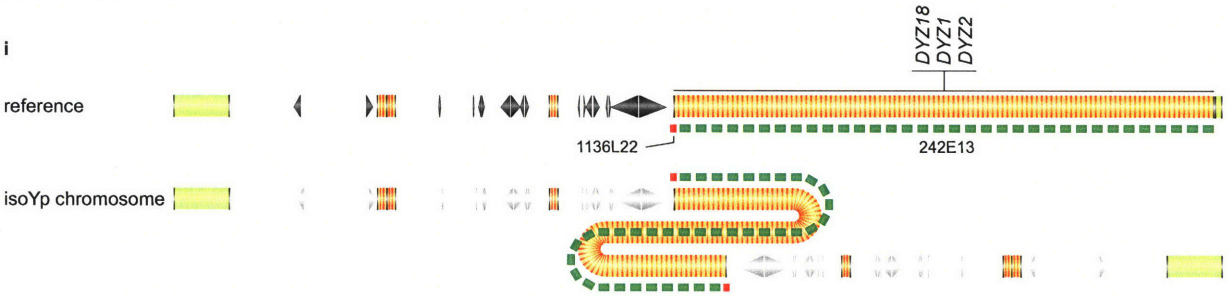


Figure S5 continued. Comprehensive results of testing for duplication, in isoYp chromosomes, of sequences immediately proximal to targeted heterochromatic repeats.

C *DYZ18/DYZ1/DYZ2*



ii 1136L22

	Number of sites of hybridization
control	1
WHT1261	2
WHT3643	2

Figure S6. STS content of structurally anomalous Y chromosomes in two human individuals with deletion breakpoints in the short arm (Yp).

(A) Schematic representation of Yp and centromeric region of Y chromosome, indicating locations of IR3 inverted repeat and male-determining gene *SRY*.

(B) STSs employed in fine mapping of deletion breakpoints in two individuals in whom sY14 was absent, and sY78 and sY1273 were present. These STSs include boundary and spacer-flanking markers for the IR3 inverted repeat, and a marker in Yp pericentromeric euchromatin.

(C) Results of testing genomic DNAs from two individuals for presence or absence of STSs. Solid black bars encompass STSs found to be present. Gray bars indicate breakpoint intervals within IR3's distal arm that could not be further narrowed due to cross-amplification of identical sequences in IR3's proximal arm.

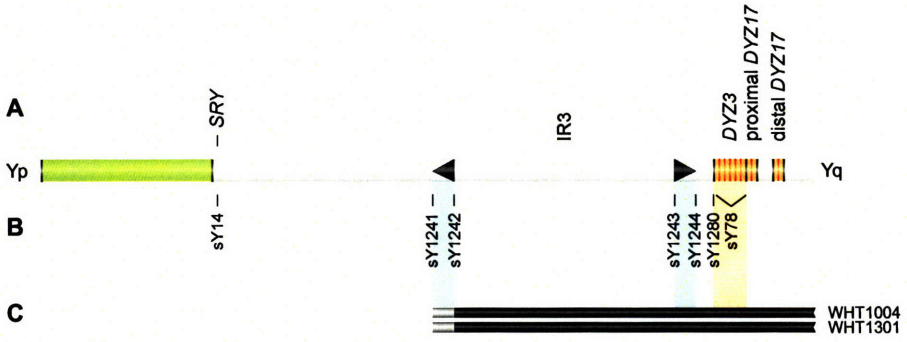
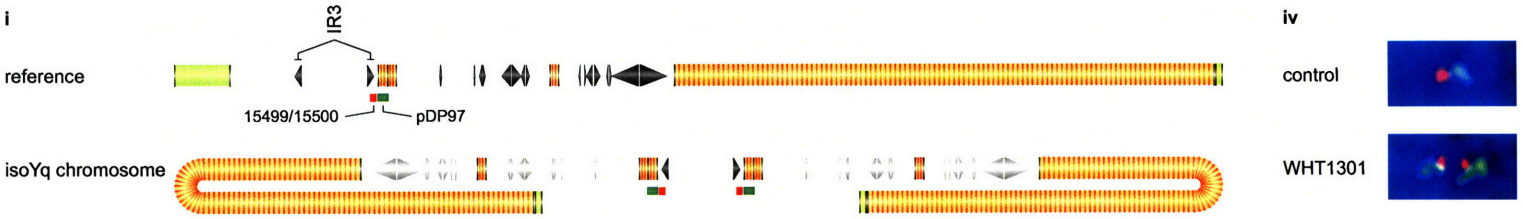


Figure S7. Comprehensive results of testing for duplication, in an isoYq chromosome, of sequences immediately proximal to Yp inverted repeat IR3.

For inverted repeat IR3 that, based on STS mapping, was hypothesized to have been targeted by ectopic homologous recombination in the formation of isoYq chromosomes, we designed a specific FISH assay to determine, in an individual with an isoYq chromosome, the copy number of sequences located immediately proximal to IR3.

In the series of figures, we show: (i) schematics of a reference chromosome and of an isoYq chromosome formed by the proposed mechanism at IR3, and the locations of probes hybridized to interphase spreads; (ii) counting results tabulated for a control male and for the case tested with the probe indicated; (iii) counting results summarized graphically; and (iv) representative FISH images for a control male and for an individual with an isoYq chromosome to demonstrate common FISH patterns.



ii 15499/15500

	Number of nuclei with signal count of					Number of sites of hybridization
	1	2	3	4	>4	
control	172	26	2	0	0	1
WHT1301	35	159	3	3	0	2

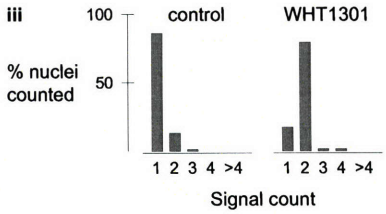


Table S1. Deletion breakpoint intervals of 100 human individuals with breakpoints in the long arm (Yq) or centromere, and of two human individuals with breakpoints in the short arm (Yp) of the MSY.
In each table, individuals are ordered by breakpoint location in the Y chromosome. For each individual, the breakpoint interval is indicated by the most distal STS that was present and the most proximal STS that was absent.

Table S1A. Deletion breakpoint intervals of 50 individuals with breakpoints in distal arms of palindromes or inverted repeats on Yq.
The 50 individuals with breakpoints in distal arms of palindromes or inverted repeats exhibit molecular signatures consistent with an isoYp chromosome formed by the proposed model. See Figures 2D and S2C for STS deletion mapping results, Table S4B for GenBank accession numbers of STSs.

Case identifier	Distal STS present	Proximal STS absent	Inferred location of breakpoint
WHT165	sY1222	sY1274	Distal arm of P8
WHT1876	sY1222	sY1274	Distal arm of P8
WHT3062	sY1222	sY1274	Distal arm of P8
WHT3794	sY1222	sY1274	Distal arm of P8
WHT3954	sY1222	sY1274	Distal arm of P8
CO	sY1286	sY1276	Distal arm of P6
FF	sY1228	sY1283	Distal arm of P5
RS	sY1228	sY1283	Distal arm of P5
WHT2024	sY1228	sY1283	Distal arm of P5
WHT2171	sY1228	sY1283	Distal arm of P5
WHT3103	sY1228	sY1283	Distal arm of P5
WHT3717	sY1228	sY1283	Distal arm of P5
WHT3894	sY1228	sY1283	Distal arm of P5
WHT4064	sY1228	sY1283	Distal arm of P5
WHT4351	sY1228	sY1283	Distal arm of P5
WHT4410	sY1228	sY1283	Distal arm of P5
WHT4818	sY1228	sY1283	Distal arm of P5
WHT5060	sY1228	sY1283	Distal arm of P5
WHT5139	sY1228	sY1283	Distal arm of P5
WHT495	sY1226	sY1277	Distal arm of P4
WHT1195	sY1226	sY1277	Distal arm of P4
WHT1658	sY1226	sY1277	Distal arm of P4
WHT2170	sY1226	sY1277	Distal arm of P4
WHT4093	sY1226	sY1277	Distal arm of P4
WHT5096	sY1226	sY1277	Distal arm of P4
WHT5108	sY1226	sY1277	Distal arm of P4
WHT5234	sY1226	sY1277	Distal arm of P4
WHT5643	sY1226	sY1277	Distal arm of P4
DM	sY1278	sY1259	Distal arm of IR2
JC	sY1278	sY1259	Distal arm of IR2
WHT1318	sY1278	sY1259	Distal arm of IR2
WHT2326	sY1278	sY1259	Distal arm of IR2
WHT2557	sY1278	sY1259	Distal arm of IR2
WHT746	sY1197	sY1572	Distal arm of P3
WHT1434	sY1197	sY1572	Distal arm of P3
WHT2064	sY1197	sY1572	Distal arm of P3
WHT2240	sY1197	sY1572	Distal arm of P3
WHT1343	sY579	sY1291	Distal arm of P2
WHT1009	sY1206	sY1201	Distal arm of P1
WHT1078	sY1206	sY1201	Distal arm of P1

Table S1A continued. Deletion breakpoint intervals of 50 individuals with breakpoints in distal arms of palindromes or inverted repeats on Yq.

Case identifier	Distal STS present	Proximal STS absent	Inferred location of breakpoint
WHT1300	sY1206	sY1201	Distal arm of P1
WHT1697	sY1206	sY1201	Distal arm of P1
WHT1871	sY1206	sY1201	Distal arm of P1
WHT2229	sY1206	sY1201	Distal arm of P1
WHT2839	sY1206	sY1201	Distal arm of P1
WHT3189	sY1206	sY1201	Distal arm of P1
WHT3285	sY1206	sY1201	Distal arm of P1
WHT3311	sY1206	sY1201	Distal arm of P1
WHT3570	sY1206	sY1201	Distal arm of P1
WHT5430	sY1206	sY1201	Distal arm of P1

Table S1B. Deletion breakpoint intervals of seven individuals with breakpoints in proximal arms of palindromes on Yq.
In seven cases with breakpoints in proximal arms of palindromes, breakpoints were fine-mapped with MSY Breakpoint Mapper, an online database of Y-specific STSs[18]. See Figures 2D and S2C for STS deletion mapping results, Table S4C for GenBank accession numbers of STSs.

Case identifier	Distal STS present	Proximal STS absent	Inferred location of breakpoint
WHT2063	sY595	sY1282	Proximal arm of P8
WHT1162	sY1774	sY1746	Proximal arm of P8
WHT1373	sY1749	sY1754	Proximal arm of P5
WHT5318	sY1235	sY828	Proximal arm of P5
WHT1231	sY1741	sY1750	Proximal arm of P4
WHT496	sY1740	sY1742	Proximal arm of P1
WHT2856[a]	sY1768	sY1770	Proximal arm of P1

[a]The deletion in WHT2856 occurs on an inverted variant of palindromes P1 and P2 (see ref.[20]).

Table S1C. Deletion breakpoint interval of one individual with a breakpoint in the spacer of Yq inverted repeat IR2.
In one case with a breakpoint in the spacer of Yq inverted repeat, the breakpoint was fine-mapped with MSY Breakpoint Mapper, an online database of Y-specific STSs[18]. See Figures 2D and S2C for STS deletion mapping results, Table S4C for GenBank accession numbers of STSs.

Case identifier	Distal STS present	Proximal STS absent	Inferred location of breakpoint
WHT5557	sY3228	sY3223	In spacer of IR2

Table S1D. Deletion breakpoint intervals of 27 individuals with breakpoints in pericentromeric euchromatin or in non-palindromic euchromatin on Yq.

In 27 cases with breakpoints in pericentromeric euchromatin or non-palindromic euchromatin on Yq, breakpoints were fine-mapped with MSY Breakpoint Mapper, an online database of Y-specific STSs[18]. See Figures 2D and S2C for STS deletion mapping results, Table S4C for GenBank accession numbers of STSs.

Case identifier	Distal STS present	Proximal STS absent	Inferred location of breakpoint
WHT1923	sY1688	sY1689	Pericentromeric euchromatin
WHT2259	sY1688	sY1689	Pericentromeric euchromatin
WHT4497	sY183	sY81	Between centromere and P8
WHT1225	sY81	sY2303	Between centromere and P8
WHT1278	sY1766	sY1765	Between centromere and P8
WHT4481	sY629	sY630	Between centromere and P8
WHT4437	sY629	sY630	Between centromere and P8
WHT3814	sY1744	sY1743	Between centromere and P8
WHT298	APXLP-5	sY2314	Between centromere and P8
WHT4749	sY1182	UTY-28	Between centromere and P8
WHT5291	sY1184	UTY-28	Between centromere and P8
WHT1874	sY1767	sY1769	Between centromere and P8
WHT5431	sY2372	sY89	Between centromere and P8
WHT1869	sY1758	sY1745	Between centromere and P8
WHT3673	sY1759	sY1762	Between centromere and P8
WHT5392	sY1899	sY1900	Between centromere and P8
WHT2317	sY2370	sY2382	Between P8 and P7
WHT4490	sY1773	NLGN4Y-2	Between P8 and P7
WHT1829	sY1747	sY1757	Between P6 and P5
WHT3502	sY1753	sY1748	Between P6 and P5
WHT3659	sY1751	sY1752	Between P4 and *DYZ19*
WHT1832	sY1760	sY1755	Between P4 and *DYZ19*
WHT1820	sY1763	sY1761	Between *DYZ19* and IR2
WHT3224	sY1572	sY1192	Between P3 and P2
WHT2277	sY1191	sY1108	Between P3 and P2
WHT3384	sY1756	sY1764	Between P3 and P2
WHT2293	sY1771	sY1772	Between P1 and *DYZ18/DYZ1/DYZ2*

Table S1E. Deletion breakpoint intervals of 15 individuals with breakpoints in one of five highly repetitive heterochromatic regions in the centromere or Yq.
See Figures 2D and S2C for STS deletion mapping results, Table S4B for GenBank accession numbers of STSs.

Case identifier	Distal STS present	Proximal STS absent	Inferred location of breakpoint
WHT1160	sY78	sY1200	*DYZ3*
WHT1709	sY78	sY1200	*DYZ3*
WHT1763	sY78	sY1200	*DYZ3*
WHT1983	sY78	sY1200	*DYZ3*
WHT2475	sY78	sY1200	*DYZ3*
WHT1781B	sY1200	sY1688	Proximal *DY17*
WHT2339	sY1200	sY1688	Proximal *DY17*
WHT4792	sY1200	sY1688	Proximal *DY17*
LGL114	sY1689	sY1251	Distal *DYZ17*
WHT1826	sY1689	sY1251	Distal *DYZ17*
WHT4596	sY1689	sY1251	Distal *DYZ17*
WHT1782	sY1252	sY1253	*DYZ19*
WHT1261	sY1246	sY1166	*DYZ18/DYZ1/DYZ2*
WHT3643	sY1246	sY1166	*DYZ18/DYZ1/DYZ2*
WHT5217	sY1246	sY1166	*DYZ18/DYZ1/DYZ2*

Table S1F. Deletion breakpoint intervals of two individuals with breakpoints in the Yp inverted repeat IR3 and exhibiting molecular signatures consistent with an isoYq chromosome formed by the proposed model.
See Figures 5C and S6C for STS deletion mapping results, Table S4D for GenBank accession numbers of STSs.

Sample identifier	Distal STS present	Proximal STS absent	Inferred location of breakpoint
WHT1004	sY1242	sY1241	Distal arm of IR3
WHT1301	sY1242	sY1241	Distal arm of IR3

Table S2. MSY gene content and copy number of 21 isoYp chromosomes and two isoYq chromosomes predicted to form by 23 distinct ectopic homologous recombination events involving palindromes, inverted repeats, and highly repetitive heterochromatic regions.

	Copy number per gene																										
Targeted repeat	*SRY*	*RPS4Y1*	*ZFY*	*TGIF2LY*	*PCDH11Y*	*TSPY*	*AMELY*	*TBL1Y*	*PRKY*	*USP9Y*	*DDX3Y*	*UTY*	*TMSB4Y*	*VCY*	*NLGN4Y*	*XKRY*	*CDY*	*HSFY*	*CYorf15A*	*CYorf15B*	*JARID1D*	*EIF1AY*	*RPS4Y2*	*RBMY1*	*PRY*	*BPY2*	*DAZ*
Reference	1	1	1	1	1	33	1	1	1	1	1	1	1	2	1	2	4	2	1	1	1	1	1	6	2	3	4
IsoYp chromosomes																											
DYZ3	2	2	2	2	2	66	2	2	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
proximal *DYZ17*	2	2	2	2	2	66	2	2	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
distal *DYZ17*	2	2	2	2	2	66	2	2	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
P8	2	2	2	2	2	66	2	2	2	2	2	2	2	2	0	0	0	0	0	0	0	0	0	0	0	0	0
P7	2	2	2	2	2	66	2	2	2	2	2	2	2	4	2	0	0	0	0	0	0	0	0	0	0	0	0
P6	2	2	2	2	2	66	2	2	2	2	2	2	2	4	2	0	0	0	0	0	0	0	0	0	0	0	0
P5	2	2	2	2	2	66	2	2	2	2	2	2	2	4	2	2	2	0	0	0	0	0	0	0	0	0	0
P4	2	2	2	2	2	66	2	2	2	2	2	2	2	4	2	4	4	2	0	0	0	0	0	0	0	0	0
DYZ19	2	2	2	2	2	66	2	2	2	2	2	2	2	4	2	4	4	4	2	2	2	0	0	0	0	0	0
IR2	2	2	2	2	2	66	2	2	2	2	2	2	2	4	2	4	4	4	2	2	2	2	2	4	0	0	0
P3	2	2	2	2	2	66	2	2	2	2	2	2	2	4	2	4	4	4	2	2	2	2	2	10	2	0	0
P2	2	2	2	2	2	66	2	2	2	2	2	2	2	4	2	4	4	4	2	2	2	2	2	12	4	2	2
b3/b2 - proximal P1/P3	2	2	2	2	2	66	2	2	2	2	2	2	2	4	2	4	4	4	2	2	2	2	2	12	4	1	2
IR5 - proximal P1/P5	2	2	2	2	2	66	2	2	2	2	2	2	2	4	2	3	4	2	1	1	1	1	1	6	2	1	2
P1.2	2	2	2	2	2	66	2	2	2	2	2	2	2	4	2	4	6	4	2	2	2	2	2	12	4	2	4
rg/gr - P1/P2	2	2	2	2	2	66	2	2	2	2	2	2	2	4	2	4	5	4	2	2	2	2	2	12	4	3	4
P1	2	2	2	2	2	66	2	2	2	2	2	2	2	4	2	4	6	4	2	2	2	2	2	12	4	4	6
P1.1	2	2	2	2	2	66	2	2	2	2	2	2	2	4	2	4	6	4	2	2	2	2	2	12	4	6	8
IR5 - distal P1/P5	2	2	2	2	2	66	2	2	2	2	2	2	2	4	2	3	4	2	1	1	1	1	1	6	2	3	4
b4/b1 - distal P1/P3	2	2	2	2	2	66	2	2	2	2	2	2	2	4	2	4	6	4	2	2	2	2	2	10	2	3	4
DYZ18/DYZ1/DYZ2	2	2	2	2	2	66	2	2	2	2	2	2	2	4	2	4	8	4	2	2	2	2	2	12	4	6	8
IsoYq chromosomes																											
IR3	0	0	0	0	0	33	1	1	1	2	2	2	2	4	2	4	8	2	2	2	2	2	2	12	4	6	8
DYZ3	0	0	0	0	0	0	0	0	0	2	2	2	2	4	2	4	8	2	2	2	2	2	2	12	4	6	8

Table S3. Target size, isochromosome size, and molecular signatures of 21 isoYp chromosomes and two isoYq chromosomes predicted to form by 23 distinct ectopic homologous recombination events involving palindromes, inverted repeats, and highly repetitive heterochromatic regions.
For each isochromosome, the breakpoint interval is indicated by the most distal STS predicted to be present and the most proximal STS predicted to be absent. See Table 1 for references used in determining isochromosome sizes, Figures 8A and 8B for predicted STS deletion mapping results.

Targeted repeat	Single target size (kb)	Total target size (kb)	Isochromosome size (Mb)	Distal STS present	Proximal STS absent
IsoYp chromosomes					
DYZ3	array	500	21.9	sY78	sY1200
proximal *DYZ17*	array	200	22.6	sY1200	sY1688
distal *DYZ17*	array	200	23.9	sY1689	sY1251
P8	36	72	28.6	sY1222	sY1274
P7	8.7	17.4	32.3	sY1311	sY1304
P6	110	220	33.1	sY1286	sY1276
P5	496	992	36.4	sY1228	sY1283
P4	190	380	37.8	sY1226	sY1277
DYZ19	array	330	40.9	sY1252	sY1253
IR2	62	124	43.9	sY1278	sY1259
P3	283	566	45.0	sY1197	sY1572
P2	122	244	46.9	sY579	sY1291
b3/b2 - proximal P1/P3	229	458	46.2	sY1291	sY1206
IR5 - proximal P1/P5	131	262	42.5	sY1291	sY1206
P1.2	9.9	18.8	48.7	sY1291	sY1206
rg/gr - P1/P2	558	1,116	48.5	sY1206	sY1201
P1	1,450	2,900	50.1	sY1206	sY1201
P1.1	9.9	18.8	51.5	sY1206	sY1201
IR5 - distal P1/P5	131	262	44.0	sY1206	sY1201
b4/b1 - distal P1/P3	168	336	48.6	sY1206	sY1201
DYZ18/DYZ1/DYZ2	array	31,700	85.4	sY1246	sY1166
IsoYq chromosomes					
IR3	298	596	101.4	sY1242	sY1241
DYZ3	array	500	96.1	sY78	sY1280

Table S4. STSs used in the analysis of Y chromosome content.
In each table, STSs are ordered by location in the Y chromosome, from distal Yp to centromere on the short arm (Yp), and centromere to distal Yq on the long arm (Yq).

Table S4A. Three STSs used in initial screening of 2,380 individuals for deletions in Yq.
See Figure 2A for locations of STSs.

STS	GenBank accession	Location in Y chromosome
sY14	G38356	Distal Yp gene *SRY*
sY78	G38359	*DYZ3* array in centromere
sY1273	G75500	Distal to *DYZ18/DYZ1/DYZ2* at boundary of PAR2

Table S4B. Landmark STSs used to determine Y chromosome content in individuals in whom distal Yp STS sY14 and centromeric STS sY78 were present, and distal Yq STS sY1273 was absent.
See Figures 2C, S2B, and S4B for locations of STSs.

STS	GenBank accession	Location in Y chromosome
sY78	G38359	*DYZ3* array in centromere
sY1200	G75479	Transition between centromeric *DYZ3* array and proximal *DYZ17* array
sY1688	BV448840	Transition between proximal *DYZ17* array and pericentromeric euchromatin
sY1689	BV448841	Transition between pericentromeric euchromatin and distal *DYZ17* array
sY1251	G75496	Transition between distal *DYZ17* and euchromatic Yq
sY1303	G75514	P8 proximal boundary
sY1223	G75485	P8 proximal flank of spacer
sY1222	G73595	P8 distal flank of spacer
sY1274	G75501	P8 distal boundary
sY1312	G73594	P7 proximal boundary
sY1310	G73592	P7 proximal flank of spacer
sY1311	G73593	P7 distal flank of spacer
sY1304	G73586	P7 distal boundary
sY1275	G75502	P6 proximal boundary
sY1285	G73584	P6 proximal flank of spacer
sY1286	G73585	P6 distal flank of spacer
sY1276	G75503	P6 distal boundary
sY1264	G72346	P5 proximal boundary
sY1227	G72343	P5 proximal flank of spacer
sY1228	G72344	P5 distal flank of spacer
sY1283	G75508	P5 distal boundary, P4 proximal boundary
sY1225	G73582	P4 proximal flank of spacer
sY1226	G73583	P4 distal flank of spacer
sY1277	G75504	P4 distal boundary
sY1252	G75497	*DYZ19* proximal boundary
sY1253	G75498	*DYZ19* distal boundary
sY1315	G75515	IR2 proximal boundary
sY1279	G75505	IR2 proximal flank of spacer
sY1278	G75504	IR2 distal flank of spacer
sY1259	BV444812	IR2 distal boundary, P3 proximal boundary
sY1196	G67167	P3 proximal flank of spacer
sY1197	G67168	P3 distal flank of spacer
sY1572	BV444813	P3 distal boundary
sY1198	G67169	Boundary between green and red amplicons in P2 and P1
sY579	G63909	Spacer of P2 and P1
sY1291	G72340	P2 distal boundary, P1 proximal boundary
sY1206	G67171	Boundary between yellow and green amplicons in P1
sY1201	G67170	P1 distal boundary
sY1246	G75492	Proximal boundary between Yq euchromatin and *DYZ18/DYZ1/DYZ2*
sY1166	G66149	Y-specific sequence distal to *DYZ18/DYZ1/DYZ2*
sY1273	G75500	Distal to *DYZ18/DYZ1/DYZ2* at boundary of PAR2

Table S4C. STSs that define breakpoint intervals in individuals with breakpoints in proximal arms of palindromes, in the spacer of inverted repeat IR2, in pericentromeric euchromatin, or in non-palindromic euchromatin on Yq.
See MSY Breakpoint Mapper[18] for locations of STSs.

STS	GenBank accession	Location in Y chromosome
sY1688	BV448840	Transition between proximal *DYZ17* array and pericentromeric euchromatin
sY1689	BV448841	Transition between pericentromeric euchromatin and distal *DYZ17* array
sY183	G66624	Between centromere and P8
sY81	G66519	Between centromere and P8
sY2303	G66185	Between centromere and P8
sY1766	BV682297	Between centromere and P8
sY1765	BV682296	Between centromere and P8
sY629	G65852	Between centromere and P8
sY630	G49201	Between centromere and P8
sY1744	BV682275	Between centromere and P8
sY1743	BV682274	Between centromere and P8
APXLP-5	BV678979	Between centromere and P8
sY2314	G66193	Between centromere and P8
sY1182	G64729	Between centromere and P8
sY1184	G64731	Between centromere and P8
UTY-28	BV979197	Between centromere and P8
sY1767	BV682298	Between centromere and P8
sY1769	BV682300	Between centromere and P8
sY2372	G66221	Between centromere and P8
sY89	G66625	Between centromere and P8
sY1758	BV682289	Between centromere and P8
sY1745	BV682276	Between centromere and P8
sY1759	BV682290	Between centromere and P8
sY1762	BV682293	Between centromere and P8
sY1899	BV694008	Between centromere and P8
sY1900	BV694009	Between centromere and P8
sY595	G34985	Arms of P8
sY1282	BV681875	Arms of P8
sY1774	BV682305	Arms of P8
sY1746	BV682277	Arms of P8
sY2370	G66219	Between P8 and P7
sY2382	G66229	Between P8 and P7
sY1773	BV682304	Between P8 and P7
NLGN4Y-2	BV679041	Between P8 and P7
sY1747	BV682278	Between P7 and P6
sY1757	BV682288	Between P7 and P6
sY1753	BV682284	Between P7 and P6
sY1748	BV682279	Between P7 and P6
sY1749	BV682280	Arms of P5
sY1754	BV682285	Arms of P5
sY1235	BV210874	Arms of P5
sY828	BV703529	Arms of P5
sY1741	BV682272	Arms of P4
sY1750	BV682281	Arms of P4
sY1751	BV682282	Between P4 and *DYZ19*

Table S4C continued. STSs that define breakpoint intervals in individuals with breakpoints in proximal arms of palindromes, in pericentromeric euchromatin, or in non-palindromic euchromatin on Yq.

STS	GenBank accession	Location in Y chromosome
sY1752	BV682283	Between P4 and *DYZ19*
sY1760	BV682291	Between P4 and *DYZ19*
sY1755	BV682286	Between P4 and *DYZ19*
sY1763	BV682294	Between *DYZ19* and IR2
sY1761	BV682292	Between *DYZ19* and IR2
sY3228	BV704093	In spacer of IR2
sY3223	BV704088	In spacer of IR2
sY1572	BV444813	P3 distal boundary
sY1192	G67166	Unique sequence between P3 and green amplicon g1
sY1191	G73809	Unique sequence between P3 and green amplicon g1
sY1108	BV681874	Green amplicons
sY1756	BV682287	Green amplicons
sY1764	BV682295	Green amplicons
sY1740	BV682271	Yellow amplicons
sY1742	BV682273	Yellow amplicons
sY1768	BV682299	Yellow amplicons
sY1770	BV682301	Yellow amplicons
sY1771	BV682302	Between P1 and *DYZ18/DYZ1/DYZ2*
sY1772	BV682303	Between P1 and *DYZ18/DYZ1/DYZ2*

Table S4D. Landmark STSs used to determine Y chromosome content in individuals in whom distal Yp STS sY14 was absent, and centromeric STS sY78 and distal Yq STS sY1273 were present.
See Figures 5B and S6B for locations of STSs.

STS	GenBank accession	Location in Y chromosome
sY14	G38356	Distal Yp in male-determining gene *SRY*
sY1241	G75487	IR3 distal boundary
sY1242	G75488	IR3 distal spacer flank
sY1243	G75489	IR3 proximal spacer flank
sY1244	G75490	IR3 proximal boundary
sY1280	G75506	Yp pericentromeric region
sY78	G38359	*DYZ3* array in centromere

Table S5. Metaphase and interphase FISH assays of Y chromosome structure in individuals with breakpoints in the MSY.

Table S5A. Metaphase FISH assays of Y chromosome structure.
See Table S5B for information on probes.

Aim	Assay	Results
IsoYp chromosome detection in individuals positive for sY14, sY78 and negative for sY1273	Visualization of chromosome structure by co-hybridization of pDP1335 and pDP97	Figure 3
IsoYq chromosome detection in individuals positive for sY78, sY1273, and negative for sY14	Visualization of chromosome structure by co-hybridization of 242E11 and pDP97	Figure 5D

Table S5B. Probes used in metaphase FISH assays of Y chromosome structure.
See Figures 3 and 5D for probe hybridization sites.

Probe	Hybridization site in reference Y chromosome	Source of DNA
pDP1335	*SRY* on distal Yp	Plasmid
pDP97	*DYZ3* elements in centromere	Plasmid
242E13	*DYZ1* elements in distal Yq heterochromatin	RPCI-11 BAC

Table S5C. Interphase and metaphase FISH assays of copy number of specific sequences in isochromosomes, to more precisely determine the sites of recombination leading to isochromosome formation.
See Table S5D for information on probes.

Targeted repeat	Assay	Results
DYZ3	Counting of 15499/15500 in co-hybridization with pDP97 on interphase spreads	Figure S5A
proximal *DYZ17*	Counting of 15499/15500 in co-hybridization with pDP97 on interphase spreads	Figure S5B
P8	Counting of 217J19 on interphase spreads	Figure S3A
P6	Counting of 15907/15908 on interphase spreads	Figure S3B
P5	Counting of 15485/15486 on interphase spreads	Figures 4C and S3C
P4	Counting of 16053/16054 on interphase spreads	Figure S3D
IR2	Counting of 15469/15470 on interphase spreads	Figures 4D and S3E
P3	Counting of 15469/15470 on interphase spreads	Figure S3F
DYZ18/DYZ1/DYZ2	Visualization of copy number of 1136L22 in co-hybridization with 242E13 on metaphase spreads	Figure 4E and S5C
IR3	Counting of 15499/15500 in co-hybridization with pDP97 on interphase spreads	Figures 5E and S7
P1	Counting of 1325K3 on interphase spreads	Figure S4D
P1	Counting of 18E8 on interphase spreads	Figure S4E
P1	Counting of 100J21 on interphase spreads	Figure S4F

Table S5D. Probes used in interphase and metaphase FISH assays of copy number of specific sequences in isochromosomes, to more precisely determine the sites of recombination leading to isochromosome formation.
See Figures 4, 5E, S3, S4, S5, and S7 for probe hybridization sites.

Probe	Hybridization site in reference Y chromosome	Source of DNA
15499/15500	Proximal Yp between centromere and proximal arm of IR3, ie. 280 kb from *DYZ3* elements in centromere, 28kb proximal to IR3	Long-range PCR[a]
pDP97	*DYZ3* elements in centromere	Plasmid
217J19	1 Mb proximal to P8	RPCI-11 BAC
15907/15908	2.5 kb proximal to P6	Long-range PCR[a]
15485/15486	44 kb proximal to P5	Long-range PCR[a]
16053/16054	9.7 kb and 477kb proximal to P4, ie. 9.7 kb from the end of each arm of palindrome P5	Long-range PCR[a]
15469/15470	208 kb proximal to IR2, 581 kb proximal to P3	Long-range PCR[a]
1136L22	Spans proximal boundary between euchromatic Yq and *DYZ18/DYZ1/DYZ2* heterochromatin	RPCI-11 BAC
242E13	*DYZ1* elements *DYZ18/DYZ1/DYZ2* heterochromatin	RPCI-11 BAC
1325K3	*DYZ19* elements in Yq	RPCI-11 BAC
18E8	Red amplicons in P1 and P2	Cosmid
100J21	Yellow amplicons in P1	RPCI-11 BAC

[a]See Table S5E for primer sequences and amplification protocol for synthesis of FISH probes by long-range PCR.

Table S5E. Primers and amplification protocol for long-range PCR production of specific FISH probes.

Left primer	Sequence	Right primer	Sequence	RPCI-11 BAC template	Product size (bp)
15499	TTTCCGCACATTCTCCTAGCAAATCTCA	15500	GGCTGACAGAAGGGAGAAAATGAAAGGA	155J5	10,395
15907	GTGTGCATCAGGATTCATGGAGACAGAA	15908	TGTTTGGTTAGACACACCCTTTGCAAATTC	455E3	10,468
15485	GCTACACCCTTTGTGGAGAATTTCGGAC	15486	AGAAGGCTAATCTTCAAGGGTGGTGCAG	537C24	10,582
16053	ACACCACCTGAGGTCAGGACTTTGAGAC	16054	AGAAAATAGGATGTGTGCAGCCAACCAG	529I21	10,324
15469	CTTTTCCTGCCATTGCTTTTGGTGTTTT	15470	CAAGGGAGCCTTGATCAGCACTTTTCTT	209I11	10,446

Long-range PCR was performed using Advantage 2 Taq polymerase (BD Biosciences) according to the manufacturer's instructions. Each primer was at 1 uM final concentration. For a 100 ul reaction, the template DNA was either 20 uL of a 1/10 dilution of an overnight BAC innoculant or 50 ng of extracted BAC DNA. Amplification conditions were 95°C for 1 minute; 30 cycles of 95°C for 30 seconds, 68°C for 10 minutes; 68°C for 10 minutes.

Chapter 4

Extensive non-allelic homologous recombination between homeologous domains of the human X and Y chromosomes

Julian Lange, Helen Skaletsky, David C. Page

Author contributions

J. Lange performed analysis of Y-chromosome content and X-Y junction sequences in patients, sequenced hotspot HSA for evidence of X-Y gene conversion, genotyped X-chromosome SNPs and analyzed HapMap data to examine signatures of allelic recombination. H. Skaletsky designed all PCR assays, generated the plot of target sites of X-Y recombination, analyzed sequence features of the human X- and Y-chromosome regions, performed comparative sequence analysis between human and chimpanzee X and Y chromosomes. D.C. Page provided helpful discussions and direction throughout the project.

J. Lange wrote the manuscript.

Summary

Suppression of recombination between sex chromosomes in the heterogametic sex and the consequent decay of the sex-specific chromosome are hallmarks of genetic sex determination systems. This is well illustrated by the human X and Y chromosomes, which co-evolved from an ordinary pair of autosomes roughly 300 million years ago. Lacking a recombination partner, the Y chromosome has suffered a dramatic reduction in both size and gene content, and retained only a few regions with sequence similarity to the X chromosome, including a handful of genes with X-linked homologs. Here we report evidence of extensive non-allelic homologous recombination between homeologous regions of the X and Y chromosomes that ceased to be allelic roughly 30 million years ago. Six hotspots of ectopic crossing over between genomic regions encompassing the ancestral gene pair *PRKX* and *PRKY*, with characteristics akin to hotspots of meiotic recombination, guided us to examine these regions more closely. In doing so, we uncovered abundant non-allelic X-Y sequence homogenization, as would be predicted from the model of meiotic homologous recombination in which a double-strand break is the common precursor to both crossover and noncrossover pathways. Further, we find that hotspots of ectopic recombination coincide with sites of allelic recombination. Our data suggest that extant sequence similarity between diverging sex chromosomes plays a role in shaping the sequence landscape of homeologous domains, by slowing the rates at which they diverge.

Introduction

Since the emergence of the X-Y sex determination system in the mammalian ancestor roughly 300 million years ago, human X and Y chromosomes have co-evolved from an ordinary pair of autosomes to become specialized chromosomes that share both allelic and homeologous (formerly allelic) domains[1]. Having gained sex-determining function, the proto-Y underwent several inversions that sequentially suppressed recombination with the proto-X[2]. The present-day sex chromosomes exhibit the molecular signatures of these events – the X chromosome in the form of a few large, contiguous regions, or strata,

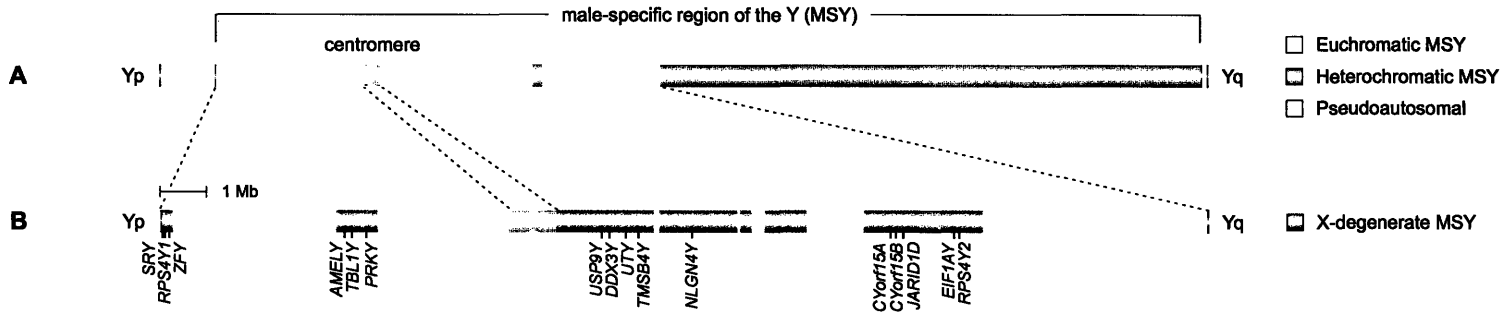


Figure 1. The male-specific region of the human Y chromosome (MSY).

(A) Schematic representation of the human Y chromosome. The MSY is flanked by two pseudoautosomal regions (green), and contains blocks of heterochromatin (dark gray) in the centromere and long arm (Yq).

(B) Expanded view of MSY euchromatin. Regions that exhibit significant sequence identity to the X chromosome are highlighted (yellow) and protein-coding genes within these regions are indicated.

that were formerly allelic with the Y chromosome[1-3], and in the Y chromosome as a scattered collection of X-degenerate sequences that contain genes with X-linked homologs (Figure 1)[1]. In contrast, pseudoautosomal regions (PARs) located at the ends of each chromosome actively participate in X-Y pairing and allelic recombination during male meiosis, and are consequently homologous.

Despite a lack of pairing with the X chromosome during meiosis, the modern human male-specific region of the Y (MSY) is not as recombinationally inert as previously thought. Whereas limited meiotic exchange has led to degeneration and gene loss along much of the Y chromosome[4], productive recombination in the form of intrachromosomal gene conversion maintains the >99.9% arm-to-arm sequence identities of massive palindromes strewn throughout the chromosome, and may be preserving members of palindrome-borne, testis-specific gene families[5]. In addition, numerous recurrent ectopic, homology-mediated crossing over events between Y-linked amplicons have been characterized and lead to a wide array of phenotypes, including spermatogenic failure, manifestations of Turner syndrome, ambiguous external genitalia, gonadoblastoma, and delay in development and growth[6-12].

The discovery that the MSY indeed engages in recombination prompted us to address whether extant X-Y sequence identity promotes interchromosomal recombination between homeologous domains. We focused on recurrent ectopic crossing over events between homeologous regions on the X and Y chromosomes, each encompassing a member of the ancestral gene pair *PRKX* and *PRKY*, that lead to X-Y translocations and XX male syndrome or XY female syndrome[13-16]. To gain insight into the mechanistic basis of these homology-mediated events, we considered parallels to the better understood process of meiotic recombination between homologous chromosomes, in which double-strand breaks (DSBs) are a common precursor to the crossover (CO) and noncrossover (NCO) pathways (Figure 2A)[17-20]. We hypothesized that ectopic X-Y crossing over events between homeologous regions encompassing *PRKX* and *PRKY* might be due to DSB formation followed by CO resolution, and further, that resolution of these DSBs may alternatively occur via the NCO pathway, resulting in sequence homogenization (Figure 2B). In examining 32 unrelated individuals carrying such X-Y translocations, we identified numerous hotspots of ectopic X-Y crossing over with characteristics reminiscent of hotspots of allelic recombination.

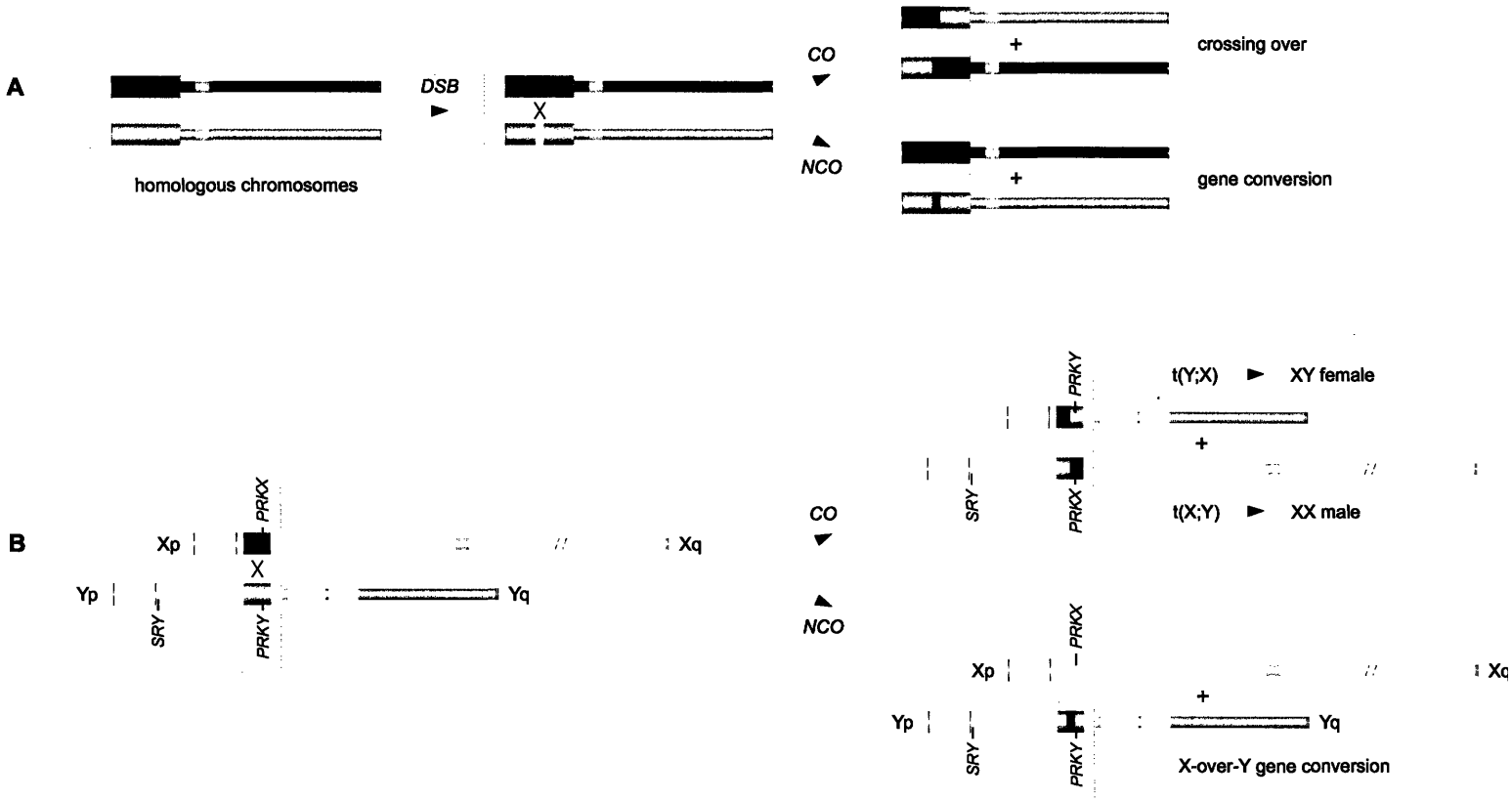


Figure 2. Hypothesized mechanism of homologous recombination between homeologous domains of the X and Y chromosomes.

(A) Conventional meiotic recombination between homologous chromosomes: double-strand break (DSB) resolution by crossover (CO) and noncrossover (NCO) pathways, yields crossing over and gene conversion, respectively.

(B) Homology-dependent recombination between homeologous domains on the X and Y chromosomes encompassing *PRKX* and *PRKY*: whereas DSB resolution by the CO pathway produces translocations found in XX males and XY females, NCO resolution results in gene conversion. Here, X sequence is shown to be overwriting Y sequence. Y-overwriting-X would likewise result in X-Y sequence homogenization.

Furthermore, through comparative sequencing, we uncovered extensive X-Y sequence homogenization between these regions in both the human and chimpanzee lineages. We posit that such exchange slows the divergence of homeologous domains of the X and Y chromosomes subsequent to the suppression of allelic recombination.

Results

Identification of hotspots of ectopic crossing over between homeologous domains of the X and Y chromosomes

We screened for ectopic X-Y crossing over events by analyzing genomic DNA from 2,380 individuals studied in our laboratory during the past 25 years[9,10,15,21-25]. These individuals fall into two broad classes of ascertainment. The first class consists of 830 individuals obtained by either karyotypic identification of a structurally aberrant Y chromosome, or discordance between sex chromosome constitution and sex phenotype ("sex reversal"). The second class consists of 1,550 men with spermatogenic failure (< 5 million spermatozoa per ml semen; normal > 20 million spermatozoa per ml semen), a phenotype of XX males.

We first sought to identify Y chromosomes with breakpoints in the short arm (Yp). Our initial screen therefore employed three Y-specific sequence-tagged sites (STSs), whose presence or absence is readily scored by PCR: sY14, which marks *SRY* on distal Yp; sY78, which marks the centromere; and sY1273, which marks the most distal male-specific portion of the long arm (Yq) (Figure 3A). Whereas XY females should lack sY14 but carry sY78 and sY1273, XX males should carry sY14 but lack sY78 and sY1273. Among the 2,380 individuals tested, we discovered 59 XX males and 16 XY females that displayed such results (Figure 3A).

We inferred that, in each of these 75 cases, the Y chromosome was broken somewhere between the Yp marker sY14 and the centromeric marker sY78. *PRKY* is located in a 3.6-megabase (Mb) segment flanked by two 298-kilobase (kb) arms of the inverted repeat IR3 (Figure 3B). This segment has been

Figure 3. Structural polymorphism in the short arm (Yp) of the Y chromosome, sequence-based map of STSs, protein-coding genes in X-degenerate sequences, X-Y sequence similarity, and deletions in the Y chromosomes of six XY females and 26 XX males due to X-Y recombination between the regions encompassing *PRKX* and *PRKY*.

(A) STSs employed in the initial screen of 2,380 individuals for Yp breakpoints. XX males that carry Y-chromosome breakpoints in the vicinity of *PRKY* are predicted to carry sY14 but lack sY78 and sY1273. XY females with breakpoints in the vicinity of *PRKY* are predicted to lack sY14 but carry sY78 and sY1273. 75 cases – 59 XX males and 16 XY females – were identified.

(B) Recurrent homology-mediated inversion between 298-kb inverted repeats IR3 (dark gray arrows) in Yp has resulted in an inversion polymorphism of the 3.6-Mb segment that contains *PRKY*.

(C) Expanded view of the region bounded by the inverted repeat IR3, in the non-reference orientation of the IR3/IR3 inversion. Shown are the three protein-coding genes in X-degenerate sequences in this region, including *PRKY*.

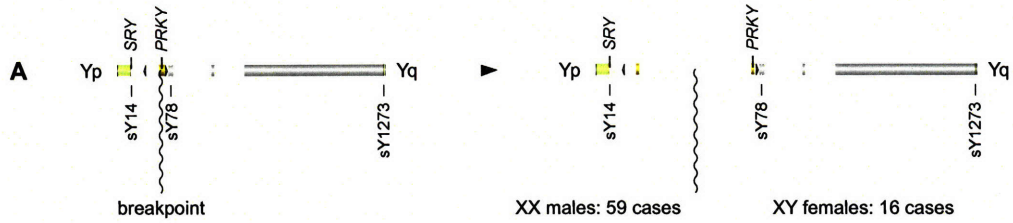
(D) Six STSs used in further analysis of 75 cases identified by initial screen, to refine breakpoints within a 537-kb region encompassing *PRKY*.

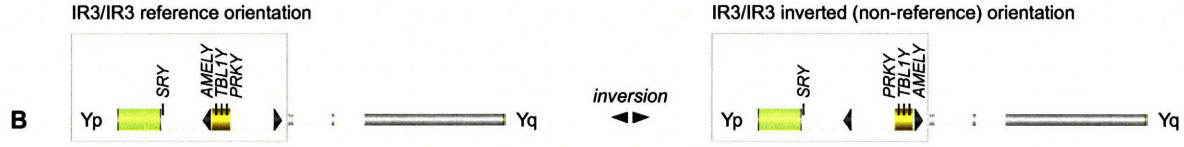
(E) Genomic structure of *PRKY* within the region bounded by STSs sY1811 and sY1320.

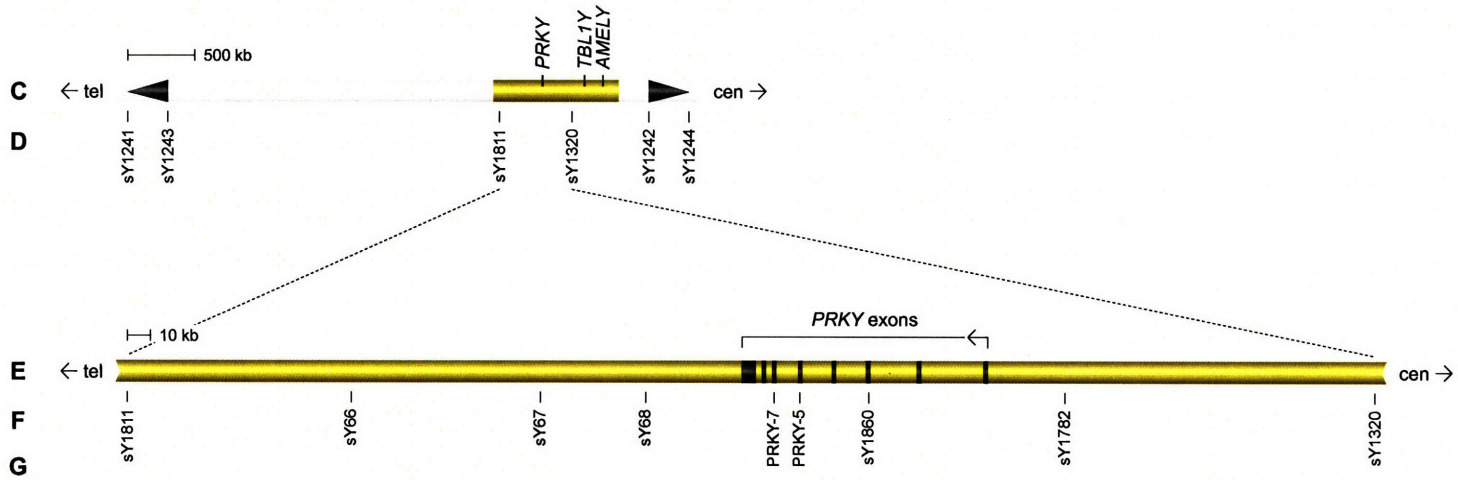
(F) STSs used in low-resolution mapping of deletion breakpoints between sY1811 and sY1320.

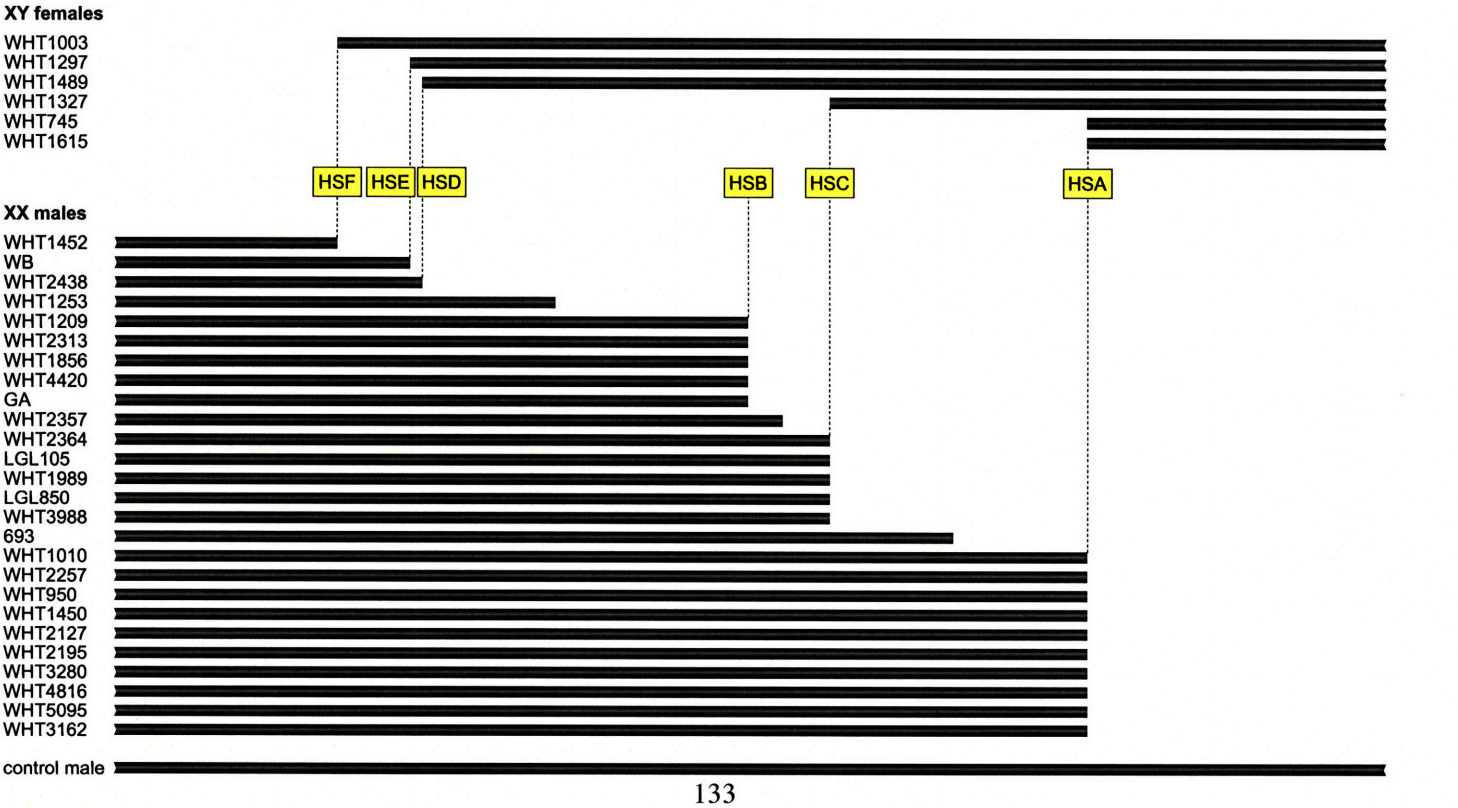
(G) Results of testing genomic DNAs from six XY females and 26 XX males for the presence or absence of STSs. Solid black bars encompass STSs that were present in low- and high-resolution screening. Six crossover hotspots (HSA through HSF) are indicated.

(H) High-resolution mapping of deletion breakpoints in six XY females and 26 XX males. STSs in bold are those used for low-resolution screening, which define the eight deletion intervals in i-viii. STS intervals are not to scale. **(i)** sY1811 to sY66, containing hotspot HSF. **(ii)** sY66 to sY67, containing hotspots HSE and HSD. **(iii)** sY67 to sY68. **(iv)** sY68 to PRKX-7, containing hotspot HSB. **(v)** PRKY-7 to PRKY-5. **(vi)** PRKY-5 to sY1860, containing hotspot HSC. **(vii)** sY1860 to sY1782. **(viii)** sY1782 to sY1320, containing hotspot HSA.

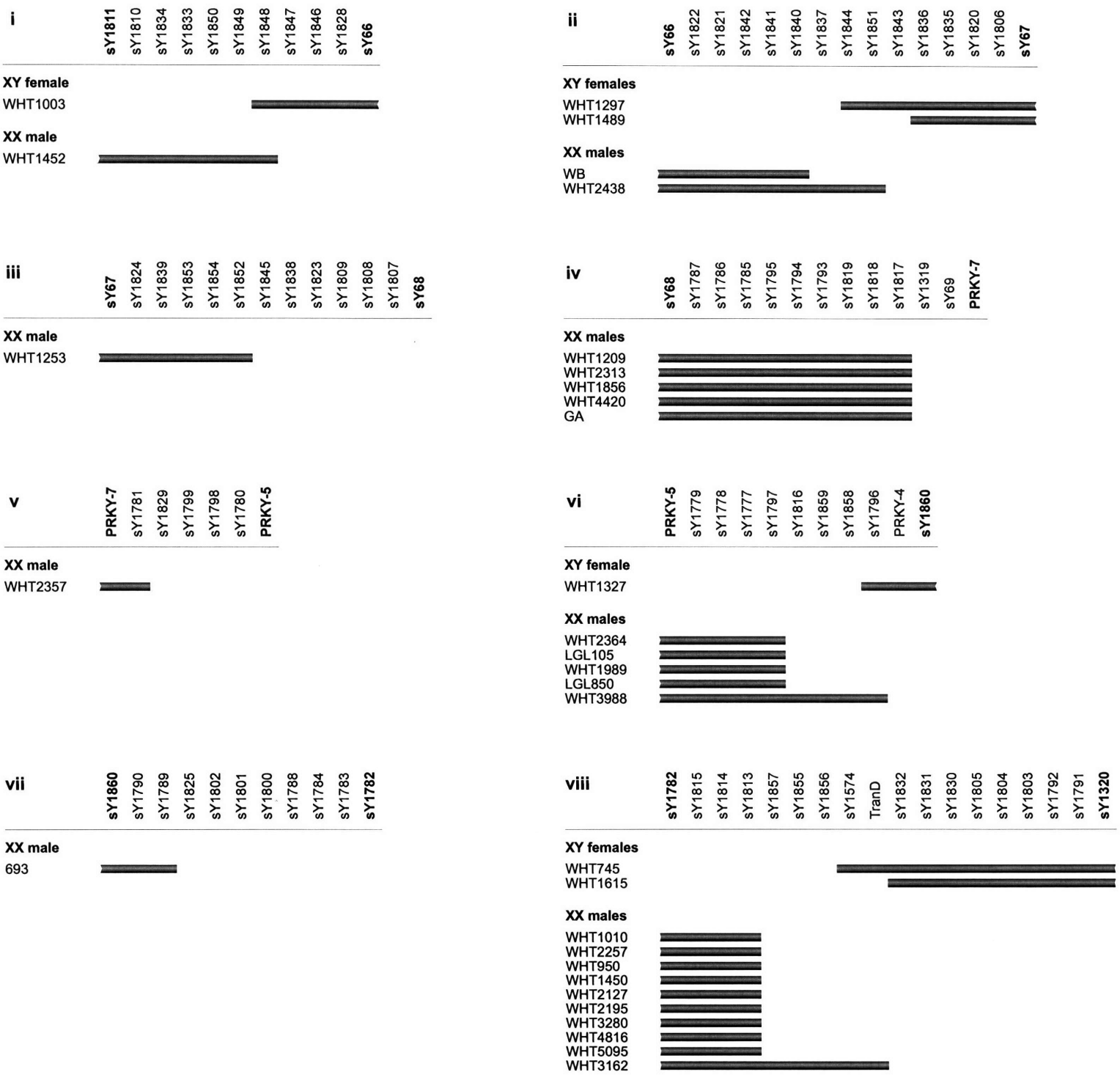

A
Yp
SRY
PRKY
sY14
sY78
Yq
sY1273
breakpoint
XX males: 59 cases
XY females: 16 cases


IR3/IR3 reference orientation
IR3/IR3 inverted (non-reference) orientation
B
Yp
SRY
AMELY
TBL1Y
PRKY
Yq
inversion


500 kb
C
tel
PRKY
TBL1Y
AMELY
cen
D
sY1241
sY1243
sY1811
sY1320
sY1242
sY1244
10 kb
PRKY exons
E
F
G
sY66
sY67
sY68
PRKY-7
PRKY-5
sY1860
sY1782


XY females
WHT1003
WHT1297
WHT1489
WHT1327
WHT745
WHT1615
HSF
HSE
HSD
HSB
HSC
HSA
XX males
WHT1452
WB
WHT2438
WHT1253
WHT1209
WHT2313
WHT1856
WHT4420
GA
WHT2357
WHT2364
LGL105
WHT1989
LGL850
WHT3988
693
WHT1010
WHT2257
WHT950
WHT1450
WHT2127
WHT2195
WHT3280
WHT4816
WHT5095
WHT3162
control male

H


i
sY1811
sY1810
sY1834
sY1833
sY1850
sY1849
sY1848
sY1847
sY1846
sY1828
sY66
XY female
WHT1003
XX male
WHT1452
ii
sY66
sY1822
sY1821
sY1842
sY1841
sY1840
sY1837
sY1844
sY1851
sY1843
sY1836
sY1835
sY1820
sY1806
sY67
XY females
WHT1297
WHT1489
XX males
WB
WHT2438
iii
sY67
sY1824
sY1839
sY1853
sY1854
sY1852
sY1845
sY1838
sY1823
sY1809
sY1808
sY1807
sY68
XX male
WHT1253
iv
sY68
sY1787
sY1786
sY1785
sY1795
sY1794
sY1793
sY1819
sY1818
sY1817
sY1319
sY69
PRKY-7
XX males
WHT1209
WHT2313
WHT1856
WHT4420
GA
v
PRKY-7
sY1781
sY1829
sY1799
sY1798
sY1780
PRKY-5
XX male
WHT2357
vi
PRKY-5
sY1779
sY1778
sY1777
sY1797
sY1816
sY1859
sY1858
sY1796
PRKY-4
sY1860
XY female
WHT1327
XX males
WHT2364
LGL105
WHT1989
LGL850
WHT3988
vii
sY1860
sY1790
sY1789
sY1825
sY1802
sY1801
sY1800
sY1788
sY1784
sY1783
sY1782
XX male
693
viii
sY1782
sY1815
sY1814
sY1813
sY1857
sY1855
sY1856
sY1574
TranD
sY1832
sY1831
sY1830
sY1805
sY1804
sY1803
sY1792
sY1791
sY1320
XY females
WHT745
WHT1615
XX males
WHT1010
WHT2257
WHT950
WHT1450
WHT2127
WHT2195
WHT3280
WHT4816
WHT5095
WHT3162

inverted repeatedly during human history, and both orientations of the inversion are widespread among extant Y chromosomes[1,16,26-29]. Ectopic crossing over events between the genomic regions encompassing *PRKX* and *PRKY* that generate XX males and XY females involve the non-reference orientation of IR3 (Figure 3C)[16,28]. To identify Y chromosomes with breakpoints in the immediate vicinity of *PRKY* and in the non-reference orientation of IR3, we tested each of the 75 cases with six additional Yp STSs: sY1241 and sY1244, which straddle the left and right boundaries of IR3; sY1243 and sY1242, which flank the 3.6-Mb spacer of IR3 in the non-reference orientation; and sY1811 and sY1320, which bound a 537-kb region that contains *PRKY* (Figure 3D). We found six XY females and 26 XX males with Y-chromosome breakpoints between sY1811 and sY1320.

In each of the 32 cases, we narrowed the breakpoint interval until we could attempt PCR amplification of an X-Y junction sequence. First, with additional low-resolution mapping using STSs sY66, sY67, sY68, PRKY-7, PRKY-5, sY1860, and sY1782 (Figures 3E and 3F), we binned the 32 cases into eight deletion intervals (Figure 3G). We then employed MSY Breakpoint Mapper, an online database of Y-specific STSs[30], to guide PCR-based, high-resolution mapping to identify the Y-chromosome breakpoint in each case (Figure 3H). Finally, we postulated that for each Y-chromosome site of crossing over, the homeologous locus within the region encompassing *PRKX* contained the X-chromosome site of crossing over. We therefore generated a plot of putative target sequences by comparing the 359-kb region encompassing *PRKX* versus the 370-kb region encompassing *PRKY* (Figure 4). With specific PCR assays, we amplified and sequenced the X-Y junction in each of the 32 individuals (Figure 5, Table 1).

Our analysis revealed the presence of six narrow hotspots of ectopic crossing over, HSA through HSF. Each hotspot is defined as the site of clustering, within a narrow region, of at least two events (Figure 4, Table 2). Two of these hotspots, HSA and HSB, had previously been identified (Table 1)[13-15], and each of the six previously reported X-Y junction sequences is at one of these two hotspots (Figure 4). The hotspots range in size from 379 base pairs (bp) to 2166 bp. 29 of the 32 events reported here took place in hotspots, and 23 occurred within the three hottest hotspots. For cases WHT950 and WHT2127,

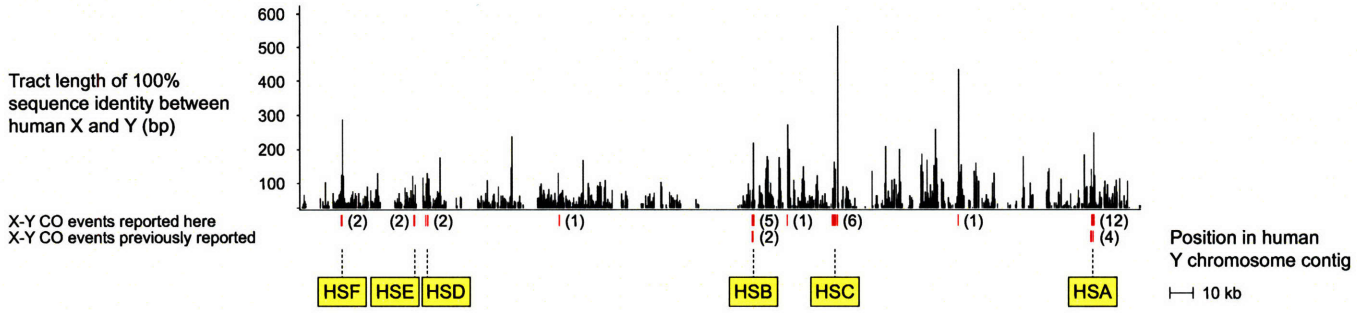


Figure 4. Hotspots of ectopic homologous crossing over between the X and Y chromosomes.

Plot of tracts of 100% sequence identity between X-Y homeologous domains of a 370-kb region encompassing *PRKY* and a 359-kb region encompassing *PRKX*, based on reference sequences. Sites of 32 ectopic crossing over events reported here, six previously reported ectopic crossing over events, and six hotspots are indicated.

Figure 5. Alignment at hotspot HSA of X-Y crossing over and gene conversion events.

Alignment of X and Y reference sequences, X-Y junction sequences for two XY females, 10 XX males reported here, and X-Y junction sequences for four previously reported XX males[13,14], and two X-overwriting-Y gene conversion events. X sequences are in pink, Y sequences in yellow. Blocks labeled with numbers indicate sites and tract lengths of 100% X-Y sequence identity, separated by X-Y nucleotide differences.


Hotspot HSA (X :1316 bp, Y: 1315 bp)
X reference
Y reference
XY females
WHT745
WHT1615
XX males
WHT1010
WHT2257
WHT950
WHT1450
WHT2127
WHT2195
WHT3280
WHT4816
WHT5095
WHT3162
Previously reported XX males
PAR475
CHM005
CHL071
MAR320
X-overwriting-Y gene conversion
GC1
GC2

Table 1. Summary of ectopic crossing over events between regions encompassing *PRKX* and *PRKY*.

Patient identifier	Length of 100% X-Y sequence identity at crossover site (bp)	Hotspot
XY females:		
WHT1003	53	HSF
WHT1297	42	HSE
WHT1489	16	HSD
WHT1327	140	HSC
WHT745	72	HSA
WHT1615	246	HSA
XX males:		
WHT1452	122	HSF
WB	44	HSE
WHT2438	75	HSD
WHT1253	32	–
WHT1209	217	HSB
WHT2313	217	HSB
WHT1856	86	HSB
WHT4420	86	HSB
GA	50	HSB
WHT2357	270	–
WHT2364	48	HSC
LGL105	10	HSC
WHT1989	161	HSC
LGL850	47	HSC
WHT3988	562	HSC
693	432	–
WHT1010	68	HSA
WHT2257	68	HSA
WHT950	246	HSA
WHT1450	246	HSA
WHT2127	246	HSA
WHT2195	246	HSA
WHT3280	246	HSA
WHT4816	246	HSA
WHT5095	246	HSA
WHT3162	40	HSA
Previously reported XX males[a]:		
NAN194	217	HSB
PAR089	86	HSB
PAR475	49	HSA
CHM005	77	HSA
CHL071	246	HSA
MAR320	246	HSA

[a]Refs.[13,14]

Table 2. Hotspots of ectopic crossing over between regions encompassing *PRKX* and *PRKY*.

Hotspot	Number of observed crossing over events[a]	Length on X (bp)[b]	Length on Y (bp)[b]
HSA	16	1316	1315
HSB	7	634	639
HSC	6	1883	2166
HSD	2	548	537
HSE	2	372	379
HSF	2	788	792

[a]Includes six previously reported junction sequences of which four are at HSA and two are at HSB (Table 1)[13,14].

[b]Length refers to size of region encompassing all observed events at each hotspot.

our results confirmed published X-Y junction sequences[15]. Remarkably, we observed reciprocal events at five of the six hotspots.

Gene conversion between homeologous domains of the X and Y chromosomes

Our observations that all events were due to homology-mediated exchange and that the narrow crossover hotspot sizes were similar to those of hotspots of allelic recombination[31] led us to investigate potential additional commonalities with the model of meiotic recombination between homologous chromosomes.

We first analyzed the X and Y hotspots of ectopic recombination for sequence features of allelic recombination hotspots. A genome-wide comparison of the sequence composition of allelic recombination hotspot and coldspot regions has shown that hotspots are enriched for THE1A and THE1B repeat elements, and in particular for those that contain the seven-nucleotide motif CCTCCCT[32]. This short motif may play a role in hotspot determination since it is similarly enriched in hotspots even after repeats are masked. Furthermore, human sperm typing studies at a 1.3-kb hotspot in the major histocompatibility (MHC) class II region have identified a polymorphism in a CCTCCCT motif within the hotspot that affects the rate of crossing over[33]. We therefore located all THE1A elements, THE1B elements, and CCTCCCT motifs within the 359 kb of X sequence and 370 kb of Y sequence (Figure 6A). In the reference X chromosome sequence, we found at least one CCTCCCT motif either within or < 500 bp away from three hotspots: HSA, HSC, and HSF. In the reference Y chromosome sequence, the same three hotspots as well as a fourth, HSD, harbor at least one copy of this seven-nucleotide motif. At HSA, there are two copies within or < 500bp away from the X hotspot sequence and three copies within or < 500 bp from the Y hotspot sequence. We did not observe an association of hotspots with THE1B elements, and there were no THE1A elements in either the X or Y chromosome sequences.

In the double-strand break repair model of meiotic recombination between homologous chromosomes, DSBs are resolved by either crossing over or NCO-associated gene conversion (Figure 2A)[17-20]. Indeed, such gene conversion is active in mammalian meiotic allelic crossover hotspots[34,35], and sites of recurrent intrachromosomal gene conversion and ectopic crossing over in the human MSY

coincide[12,36]. Could DSBs initiated in *PRKX* or *PRKY* in the male germline occasionally be processed in a manner analogous to the NCO pathway (Figure 2B)?

To determine whether the homeologous sequences at crossover hotspot HSA also engaged in non-allelic gene conversion, we examined X and Y sequences in global samples of X and Y chromosomes. We used the Y-chromosome genealogical tree to select DNAs from 92 individuals that encompass worldwide Y-chromosome diversity[29] – and are expected to carry diverse X-chromosome haplotypes around HSA – and sequenced segments encompassing HSA on the X and the Y chromosomes. We incorporated five new Y-chromosome SNPs located within HSA into the previously established Y-chromosome genealogical tree. Of these, three appeared to be due to two independent X-overwriting-Y gene conversion events – two SNPs within four nucleotides were likely the result of one such event (Figure 6B). In each case, the homeologous nucleotide on the X chromosome was monomorphic. It is highly unlikely that three of five Y substitutions mutated by chance to the homeologous X sequence ($p < 10^{-5}$). These results indicated that ectopic X-overwriting-Y gene conversion is currently active at hotspot HSA.

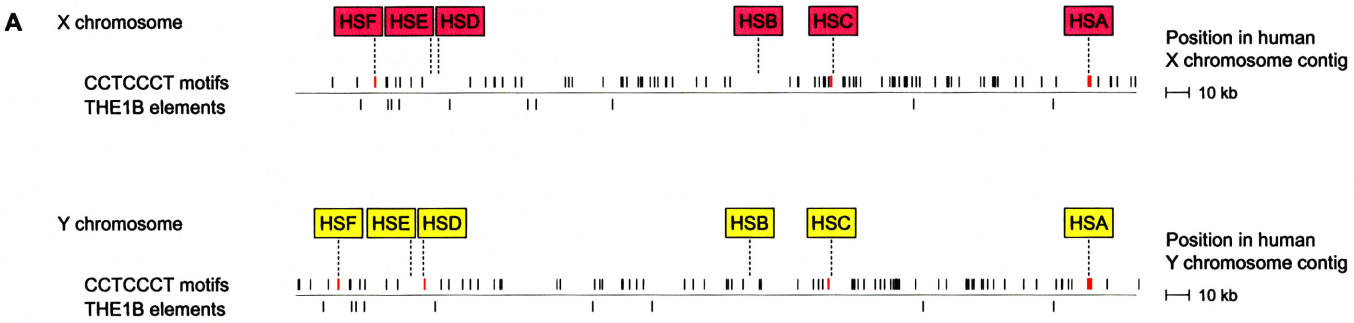
We next asked whether the homeologous domains encompassing *PRKX* and *PRKY* engaged in non-allelic gene conversion, by comparing the human X and Y sequences with orthologous sequences in the chimpanzee. Suppression of recombination (and the onset of sequence divergence) between these regions is thought to have occurred roughly 30 million years ago in the ancestral X and Y chromosomes, and therefore vastly predates the human-chimpanzee split[2,3,37]. Therefore, in the absence of X-Y sequence homogenization in either lineage, orthologous regions between species are expected to exhibit greater sequence identity than homeologous regions of the X and Y within one species, in a roughly uniform manner (Figure 6C). We searched for evidence of X-Y sequence homogenization in both the human and chimpanzee lineages, by examining the frequency with which tracts of sequence identity between the X and Y were evidently the result of X-Y gene conversion. Our results indicate that X-Y gene conversion has been extensive in both human and chimpanzee lineages.

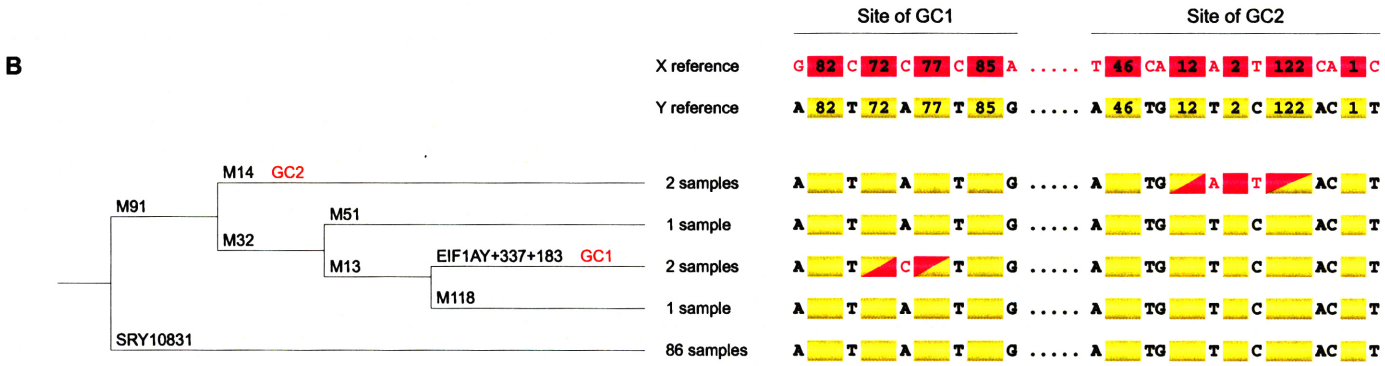
Figure 6. Commonalities of ectopic homologous recombination between the X- and Y-chromosome regions encompassing *PRKX* and *PRKY*, with allelic homologous recombination.

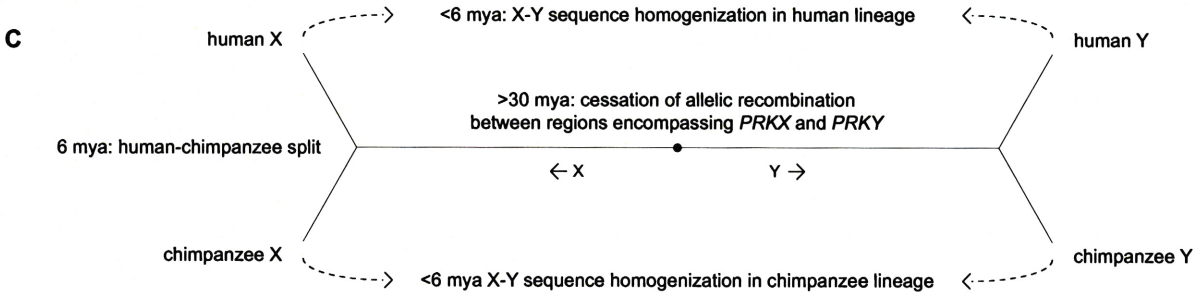
(A) Locations within these regions of two sequence features of allelic recombination hotspots. Tick marks indicate CCTCCCT motifs and THE1B elements in these regions, and are in red if they are located within or < 500 bp from a hotspot.

(B) Two X-overwriting Y gene conversion events. Left, minimal Y-chromosome genealogical tree, based on the biallelic Y-chromosome polymorphisms indicated, of 92 samples tested at hotspot HSA. X-overwriting-Y gene conversion events GC1 and GC2 are shown within the tree. At right: Alignment at GC1 and GC2 of X and Y reference sequences and 92 Y chromosomes assayed. X sequences are in pink, Y sequences in yellow. Blocks labeled with numbers indicate sites and tract lengths of 100% X-Y sequence identity, separated by X-Y nucleotide differences. See Figure 5 for full alignment of X-overwriting-Y gene conversion events in HSA.

(C) Model of sequence divergence and homogenization between these regions in human and chimpanzee.


A
X chromosome
HSF
HSE
HSD
HSB
HSC
HSA
CCTCCCT motifs
THE1B elements
Position in human
X chromosome contig
10 kb
Y chromosome
HSF
HSE
HSD
HSB
HSC
HSA
CCTCCCT motifs
THE1B elements
Position in human
Y chromosome contig
10 kb


B
Site of GC1
Site of GC2
X reference
Y reference
M14 GC2
M91
M51
M32
M13
EIF1AY+337+183 GC1
M118
SRY10831
2 samples
1 sample
2 samples
1 sample
86 samples


C
human X
<6 mya: X-Y sequence homogenization in human lineage
human Y
>30 mya: cessation of allelic recombination
between regions encompassing PRKX and PRKY
6 mya: human-chimpanzee split
← X
Y →
chimpanzee X
<6 mya X-Y sequence homogenization in chimpanzee lineage
chimpanzee Y

Coincidence of allelic and ectopic recombination hotspots

We reasoned that since modern allelic crossover hotspots established by sperm typing correlate well with polymorphism data-derived historical recombination hotspots, sites of pathological crossing over might similarly coincide with historical recombination hotspots. If the allelic crossover potential of a hotspot is a function of its sequence context, then the propensity for DSB formation during allelic (X-X) recombination might translate to a similar propensity for DSB formation and ectopic (X-Y) crossing over there during male meiosis. To gain further insight into the recombinogenic nature of the substrates of ectopic X-Y recombination, we examined the molecular signature of allelic recombination in the region encompassing *PRKX*.

We first determined whether the X-chromosome crossover hotspot HSA lies within a region of frequent allelic recombination. Twenty-four SNPs (Table 3) from dbSNP, that together cover a 24,198-bp region containing HSA (Figure 7A), were genotyped in 92 XY individuals. (None of the SNPs was within the 1315 bp of HSA.) Using Haploview[38], we determined pairwise D' values – a measure of linkage disequilibrium (LD) – and found that hotspot HSA lies within a region of historical recombination between two haplotype blocks (Figure 7B).

To examine LD over the entire X-chromosome region encompassing the six ectopic hotspots, we exploited polymorphism data generated by the International HapMap Consortium[39]. We constructed an LD plot of HapMap data with Haploview, and precisely marked the ectopic hotspots (Figures 7C-F). Our analysis showed that, in addition to HSA, four further hotspots – HSF, HSD, HSB, and HSC – lie in historically recombinogenic regions.

Table 3. Positions and minor allele frequencies of 24 SNPs around hotspot HSA in the X chromosome.

SNP	dbSNP	X chromosome position[a]	Distance from HSA (bp)[b]	Minor allele frequency[c]
Left of HSA:				
1	*rs2009624*	3653467	-11783	0.457
2	*rs4595276*	3653776	-11474	0.120
3	*rs5961994*	3657863	-7387	0.152
4	*rs5962001*	3658000	-7250	0.152
5	*rs7877157*	3661654	-3596	0.196
6	*rs6567602*	3661707	-3543	0.293
7	*rs5961450*	3664499	-751	0.293
Right of HSA:				
8	*rs1965021*	3667328	763	0.217
9	*rs5961460*	3669072	2507	0.295
10	*rs28631469*	3669112	2547	0.102
11	*rs17331195*	3671525	4690	0.065
12	*rs12116286*	3671677	5112	0.391
13	*rs12116344*	3671773	5208	0.391
14	*rs5962076*	3672017	5452	0.391
15	*rs5962077*	3672099	5534	0.391
16	*rs5962078*	3672102	5537	0.391
17	*rs1483016*	3677075	10510	0.337
18	*rs6641899*	3677084	10519	0.413
19	*rs35133862*	3677103	10538	0.424
20	*rs6641900*	3677311	10746	0.424
21	*rs5961468*	3677511	10946	0.489
22	*rs5961469*	3677515	10950	0.489
23	*rs2046294*	3677527	10962	0.217
24	*rs5961470*	3677664	11099	0.489

[a]Position in X chromosome from NCBI Build 35.

[b]Distance from left- or right-most nucleotide of HSA.

[c]Minor allele frequencies based on 92 XY genomic DNAs assayed.

Figure 7. Coincidence of ectopic and allelic recombination hotspots in the X-chromosome region encompassing *PRKX*.

Confidence bound linkage disequilibrium (LD) color scheme used in plots: white squares indicate evidence of recombination (low LD), dark squares signify high LD, gray squares are uninformative.

(A) Map of 24 SNPs in the 24,198-bp region surrounding hotspot HSA, according to NCBI Build 35.

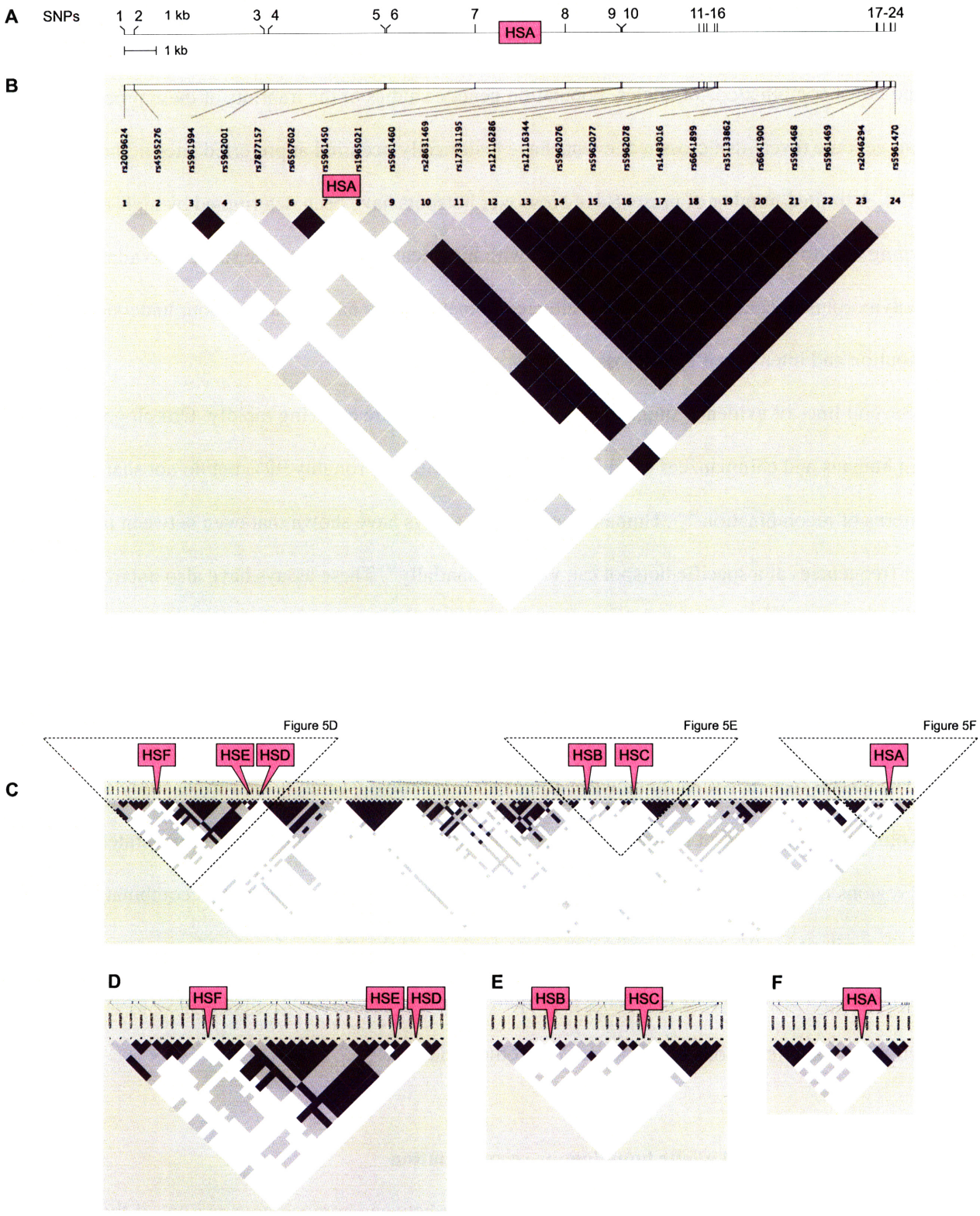
(B) Haploview-generated LD plot of 24 SNPs in this region demonstrates that hotspot HSA, located between SNPs 7 (*rs5961450*) and 8 (*rs1965021*), lies in a historically recombinogenic segment.

(C) Haploview-generated plot of LD structure of the 345-kb region encompassing all six ectopic hotspots on the X chromosome, using data available from the International HapMap Project. Five of six ectopic hotspots (all except HSE) are in regions that exhibit low LD.

(D) Magnification of regions containing hotspots HSF, HSE, and HSD.

(E) Magnification of regions containing hotspots HSB and HSC.

(F) Magnification of regions containing hotspot HSA.


A
SNPs
1 2
1 kb
3 4
5 6
7
HSA
8
9 10
11-16
17-24
1 kb
B
rs2009624
rs4595276
rs5961994
rs5962001
rs7877157
rs6567602
rs5961450
rs1965021
rs5961460
rs28631469
rs17331195
rs12116286
rs12116344
rs5962076
rs5962077
rs5962078
rs1483016
rs6641899
rs35133862
rs6641900
rs5961468
rs5961469
rs2046294
rs5961470
HSA
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24
C
Figure 5D
Figure 5E
Figure 5F
HSF
HSE
HSD
HSB
HSC
HSA
D
HSF
HSE
HSD
E
HSB
HSC
F
HSA

Discussion

Mutation and sexual recombination perpetually alter the landscape of genomic diversity by creating new alleles and novel combinations of alleles. Population genetic studies of the imprints of these processes in the human genome reveal that crossover events have historically occurred at preferred sites in the human genome[40-42]. A limited number of present-day crossover hotspots have been determined by high-resolution sperm typing and, for the most part, correlate well with historical sites[31]. With the surge in available human polymorphism data, studies of fine-scale recombination rates have broadened our understanding of the distribution and intensity of these crossover hotspots[43].

Several lines of evidence suggest that crossover hotspots are evolving rapidly. Orthologous regions in humans and chimpanzees exhibit sequence similarities of roughly 99% but do not share the same patterns of recombination[44,45]. Human sperm typing assays have shown that even between men, crossover frequencies at a specific hotspot can vary substantially[33]. These assays have also uncovered *bona fide* crossover hotspots concealed within regions of low historical recombination, evidence that emerging hotspots may not be captured in polymorphism data[46]. Taken together, these results imply significant fluidity in the evolution of hotspot sequences.

The 99.9% sequence identity between alleles affords little opportunity for further study of the recent evolution of sequences at hotspots. Ectopic homologous recombination between moderately diverged regions might serve as a more tractable surrogate, should allelic and ectopic recombination share mechanistic similarities. We have examined this assumption and the effects of recombination on the evolution of hotspot sequences by analyzing recurrent, pathological, homology-mediated crossover events between formerly allelic regions of the X and Y chromosomes.

Convergence of ectopic and allelic homologous recombination

We identified hotspots of ectopic homologous crossing over between homeologous regions of the X and Y chromosomes that respectively encompass *PRKX* and *PRKY*, members of an ancestral gene pair. These

ectopic recombination hotspots shared several features with allelic hotspots. Allelic crossover hotspots, assayed directly via sperm typing or inferred from genetic variation data, exhibit the following characteristics: (1) clustering of crossover events within narrow regions of 1-2 kb[31], (2) coincidence with sites of gene conversion, as predicted by the double-strand break repair model of meiotic recombination[17-20,34,35], and (3) genome-wide enrichment for THE1A elements, THE1B elements, and for the seven-nucleotide motif CCTCCCT[32]. The set of ectopic crossover hotspots that we discovered displays these three features: a highly punctate distribution of crossovers, coincidence with regional X-Y gene conversion, and presence of the CCTCCCT motif at four of six hotspots.

Our discovery of an overlap between hotspots of ectopic recombination and regions of allelic recombination leads us to propose that the determinants of normal allelic recombination influence the location of ectopic homologous recombination events. In the female germline, sites of allelic recombination in the X-chromosome region encompassing *PRKX* are prone to DSB formation, the initiating step of homologous recombination. Construction of LD maps from (1) genotyping of SNPs around hotspot HSA, and from (2) employing region-wide SNP data generated by the HapMap Consortium, suggests that these loci are also susceptible to DSBs in the male germline. In the absence of a second X chromosome, such breaks can apparently be processed by either the CO pathway or the NCO pathway, with homeologous sequences in the Y chromosome. Taken together, these findings suggest that the mechanism underlying ectopic homologous recombination in the etiology of XX males and XY females operate similarly to allelic recombination.

Two recent studies have demonstrated similar connections between allelic recombination and genomic disorders caused by deletions and duplications at low-copy repeats (LCRs). De Raedt and colleagues[47] examined three copies of an LCR, of which two copies are in direct orientation on chromosome 17q and one copy is on chromosome 19. Ectopic homologous recombination between the two copies on chromosome 17q generates *NF1* microdeletions that cause neurofibromatosis type 1 (NF1)[48]. The deletion breakpoints cluster to two narrow regions within the LCRs, and each region coincides with a hotspot of allelic recombination in the copy on chromosome 19. Similarly, Lindsay and

colleagues[49] investigated LCRs on chromosome 17p. Ectopic homologous crossing over between these LCRs results in microduplications and microdeletions that respectively cause Charcot-Marie-Tooth disease type 1A (CMT1A)[50] and hereditary neuropathy with pressure palsies (HNPP)[51]. In dissecting signatures of allelic recombination in these LCRs, they found that the hotspot of ectopic crossing over lay adjacent to an allelic recombination hotspot.

Dynamics of recombination in sex chromosomes

What are the forces that govern the location, frequency, and phenotypic outcomes of the ectopic homologous recombination events examined here? Meiotic pairing between PARs – located at the termini of the X and Y chromosomes (Figure 1A) – ensures that the chromosomes are in physical proximity during meiosis and could promote the occasional alignment of and recombination between formerly allelic sites. Although the orientation of the region encompassing *PRKY* (Figure 3B) might not affect the propensity for ectopic recombination with the region encompassing *PRKX*, it likely dictates the resultant phenotype of an X-Y crossing over event. Whereas the ectopic crossing over events presented here involve the non-reference orientation of the Y-chromosome region, recombination that includes the reference orientation is also possible. Indeed, each of the two X-overwriting-Y gene conversion events reported here likely involved the reference orientation (Figure 6B and ref.[29]). Such recombination requires contortion within either the X or the Y chromosome for proper alignment, as is likely the case for previously reported intra-Y-chromosome and X-Y ectopic homologous recombination events[12,52]. An absence of observed crossing over events involving the reference orientation may be due to an inviability of the resultant translocated chromosome. Alternatively, our ascertainment criteria may have precluded such individuals from this study.

At the nucleotide level, sequence identity must play a role since many of the crossing over events take place between the largest tracts of 100% X-Y identity (Figure 4). However, homology alone cannot explain their distribution since numerous prominent tracts were not targeted. We propose that additional factors – such as those that control the distribution of DSB formation in the initiation of allelic

recombination – influence the positioning of ectopic X-Y exchange. This could also account for the unevenness of X-Y sequence identity across the regions examined here, since X- and Y-chromosome hotspots of ectopic recombination should preserve their sequence identity while neighboring sites diverge. This is apparent in the X- and Y-chromosome genomic regions encompassing ancestral gene pair *KAL1* and *KALP*. Whereas the 3′ regions of these genes share high sequence identity due to X-Y gene conversion, the 5′ regions have diverged[1].

Mammalian X and Y chromosomes began diverging roughly 300 million years ago from what was an ordinary pair of recombining autosomes[2]. Sequential inversions in the proto-Y chromosome and the resultant stepwise cessation of recombination are apparent in the parsing of sequence divergence between modern X and Y chromosomes into formerly homologous regions[2,3]. The inversions have been dated based on the divergences of ancestral X-Y gene pairs[1,2] or genomic X and Y sequences[3]. The results we present here indicate that while inversion events initiate the process of divergence, X and Y sequences might differentiate more slowly than anticipated because of ectopic gene conversion.

Suppression of allelic recombination along much of the human MSY has resulted in the loss of the vast majority of its genes[4]. This finding has led to predictions that in the continued absence of productive recombination, the remaining protein-coding genes will meet similar fates over the next 10 million years, eventually leading to the extinction of the human Y chromosome[53,54]. Previously published results suggest that the human MSY is actively engaging in productive recombination. Massive palindromes that are maintained by intrachromosomal gene conversion have evolved in the MSY, perhaps to protect testis-specific genes located exclusively in the palindrome arms[1,5]. In addition, all unique genes in the human MSY have been conserved since the divergence of human and chimpanzee lineages[55]. Here we report evidence of extensive non-allelic X-Y gene conversion. It is tempting to speculate that if a Y-linked gene sustains a mutation, this mechanism could provide an X-linked ancestral gene partner for sequence homogenization.

Methods

Patient samples

We screened our collection of 2,380 individuals – obtained over 25 years by karyotypic identification of an aberrant Y chromosome, discordance between sex chromosome constitution and sex phenotype, or spermatogenic failure – for Y chromosomes with Yp breakpoints in a 370-kb region that contains *PRKY* and exhibits particularly prominent X-Y sequence identity. We analyzed genomic DNA samples from these 2,380 individuals with a series of precisely mapped, Y-specific plus/minus STSs. Screening this collection identified 32 unrelated individuals – six XY females and 26 XX males – who, on the basis of plus/minus STS analysis, were determined to have Y-chromosome breakpoints in the Yp region encompassing *PRKY*, previously identified as a site of recurrent X-Y exchange[13-15]. Seventeen of these 32 individuals had undergone prior characterization in early physical mapping of the Y chromosome and in studies of XX males and XY females (Table 4)[14,15,21-23,56]. DNA samples were available from a patrilineal relative in 10 cases and from a matrilineal relative in 12 cases. By plus/minus STS analysis, all patrilineal relatives had intact Y chromosomes and no matrilineal relative carried MSY material. The institutional review board of the Massachusetts Institute of Technology approved this study. Informed consent was obtained for all participants.

Plus/minus STS analysis and X-Y junction sequencing

Fine-mapping of Y-chromosome deletion breakpoints in 32 individuals was carried out using MSY Breakpoint Mapper, an online database of Y-specific STSs[30]. New primers for STSs employed in low- and high-resolution deletion mapping were designed using Primer3[57]. All STSs and the associated PCR conditions have been deposited in GenBank (Table 5). For each of the 32 cases, we fine-mapped the Y deletion interval to < 1 kb, then amplified and sequenced a chimeric X-Y junction fragment. We first identified the most highly similar segment within the X-chromosome region homeologous to the Y-chromosome deletion interval, as a candidate target for homology-mediated recombination. We then

Table 4. Previous characterization of 32 individuals analyzed in this study.

	Reference					
Patient identifier	Vergnaud et al., 1986[21]	Andersson et al., 1986[22]	Vollrath et al., 1992[23]	Klink et al., 1995[56]	Wang et al., 1995[14]	Schiebel et al., 1997[15]
XY females:						
WHT1003			✓			
WHT1297			✓			
WHT1489			✓			
WHT1327			✓			✓
WHT745			✓			✓
WHT1615			✓			✓
XX males:						
WHT1452						
WB			✓		✓	
WHT2438						
WHT1253			✓			
WHT1209			✓			✓
WHT2313						
WHT1856						✓
WHT4420						
GA			✓		✓	✓
WHT2357						
WHT2364						
LGL105	✓	✓	✓		✓	✓
WHT1989						
LGL850			✓			✓
WHT3988						
693				✓		✓
WHT1010			✓			✓
WHT2257						
WHT950		✓	✓			✓
WHT1450						
WHT2127			✓			✓
WHT2195						
WHT3280						
WHT4816						
WHT5095						
WHT3162						

Table 5. GenBank accession numbers of Y-chromosome STSs employed in low- and high-resolution deletion mapping.

STS	GenBank accession
PRKY-4	BV679048
PRKY-5	BV679049
PRKY-7	BV679050
sY14	G38356
sY66	G12014
sY67	G40970
sY68	G12015
sY69	G66516
sY78	G38359
sY1241	G75487
sY1242	G75488
sY1243	G75489
sY1244	G75490
sY1273	G75500
sY1319	BV210882
sY1320	G75614
sY1574	BV682948
sY1777	BV682949
sY1778	BV682950
sY1779	BV682951
sY1780	BV682952
sY1781	BV682953
sY1782	BV682954
sY1783	BV682955
sY1784	BV682956
sY1785	BV682957
sY1786	BV682958
sY1787	BV682959
sY1788	BV682960
sY1789	BV682961
sY1790	BV682962
sY1791	BV682963
sY1792	BV682964
sY1793	BV682965
sY1794	BV682966
sY1795	BV682967
sY1796	BV682968
sY1797	BV682969
sY1798	BV682970
sY1799	BV682971
sY1800	BV682972
sY1801	BV682973
sY1802	BV682974
sY1803	BV682975
sY1804	BV682976
sY1805	BV682977
sY1806	BV682978
sY1807	BV682979
sY1808	BV682980
sY1809	BV682981
sY1810	BV682982
sY1811	BV682983
sY1813	BV682985
sY1814	BV682986
sY1815	BV682987
sY1816	BV682988
sY1817	BV682989
sY1818	BV682990
sY1819	BV682991
sY1820	BV682992
sY1821	BV682993
sY1822	BV682994
sY1823	BV682995
sY1824	BV682996
sY1825	BV682997
sY1828	BV682998
sY1829	BV682999
sY1830	BV683000
sY1831	BV683001
sY1832	BV683002
sY1833	BV683003
sY1834	BV683004
sY1835	BV683005
sY1836	BV683006
sY1837	BV683007
sY1838	BV683008
sY1839	BV683009
sY1840	BV683010
sY1841	BV683011
sY1842	BV683012
sY1843	BV683013
sY1844	BV683014
sY1845	BV683015
sY1846	BV683016
sY1847	BV683017
sY1848	BV683018
sY1849	BV683019
sY1850	BV683020
sY1851	BV683021
sY1852	BV683022
sY1853	BV683023
sY1854	BV683024
sY1855	BV683025
sY1856	BV683026
sY1857	BV683027
sY1858	BV683028
sY1859	BV683029
sY1860	BV683030
TranD	Ref.[15]

designed PCR assays using Primer3[57] to amplify across the X-Y junction (Table 6). The resulting PCR products were sequenced with a BigDye Terminator kit and an ABI 3700 automated sequencer (Applied Biosystems). X-Y junction sequences have been deposited in GenBank.

Plot of target sites of X-Y recombination

Using reference sequences (UCSC Genome March 2006 Build, human X sequence coordinates: 3,328,454 to 3,709,731; human Y sequence coordinates 7,102,249 to 7,507,249), we generated a plot of recombination targets between 359-kb and 370-kb regions encompassing *PRKX* and *PRKY*, respectively. Reference sequences were first aligned using Clustal W with default parameters[58]. Employing custom Perl code, the alignment was then plotted as the nucleotide position in the Y contig (x-axis) versus the tract length of 100% sequence identity with the X contig (y-axis). This code is available upon request.

Resequencing of hotspot HSA in 92 XY genomic DNAs

To search for evidence of X-overwriting-Y sequence homogenization within the 1315 bp of hotspot HSA in the Y chromosome, we sequenced 1569 bp and 1568 bp encompassing crossover hotspot HSA in the X and Y chromosomes, respectively, of 92 unrelated males. These 92 samples were selected to represent worldwide Y-chromosome diversity, using biallelic Y-haplotype markers as previously described[29]. For each region, we designed two overlapping chromosome-specific STS assays using Primer3[57], and PCR products were sequenced with internal chromosome-specific primers (Table 7). STSs and the associated PCR conditions have been deposited in GenBank.

Comparative sequencing of human and chimpanzee X and Y chromosomes

We aligned reference human X, chimpanzee X, human Y, and chimpanzee Y chromosome sequences using Clustal W with default parameters[58]. Whereas human X, human Y, and chimpanzee Y chromosome reference sequences are genomic clone-based (UCSC Genome March 2006 Build, human X and Y sequence coordinates as above; chimpanzee Y sequence coordinates: 22,524,159 to 22,948,243),

Table 6. X-Y junction fragment sequencing and GenBank accession numbers.

	Amplification and sequencing[a]					GenBank
	Forward primer[b]		Reverse primer[b]			accession
Patient identifier	Name	5′–3′ sequence	Name	5′–3′ sequence	Protocol[c]	of junction
XY females:						
WHT1003	14304	GGTCTATTTTAAGCAGGTGGGATTA	14389	GTGGTGCTACCAGGCGCT	Touchdown	EU251153
WHT1297	14416	GGGGTTGTGGGTGCATAA	14415	TGGGAAACTGATACTACTCCAAGTT	Touchdown	EU251154
WHT1489	14360	TCCCCTCTATGTGTCCATGAA	14359	GTAAAAGCCAATACACAACACTCG	Touchdown	EU251155
WHT1327	14388	CACAGGACAAGATAGGAGGTCA	13519	CAGGAAGGAGTCCTGATCCA	Touchdown	EU251156
WHT745	14399	ATTATGTCCTTTGCAGGAAT	TranDfor[d]	CCTGCCTTTTTTAGTTTCCAGCA	Touchdown	EU251157
WHT1615	14298	GGCCCACGTTTGCCTC	13721	TTCCCGCAGGATCTATCATC	Touchdown	EU251158
XX males:						
WHT1452	14026	CTGTGGACGCCACATGGT	14078	ACAGACACTCTCAGACGCACTT	Standard	EU251159
WB	14148	TGAGTGTATCATTGGATTACTTTGC	14147	CACTTTTGTGAGATTTAGCTCATCT	Standard	EU251160
WHT2438	14144	TGCATTTCTCTCCTACAAGCTATTA	14031[b]	TCCCTGCACTGTCTCTTTCC	Standard	EU251161
WHT1253	14176	TGTAGAGGCTTCCAGTAACGTG	14015[b]	CCAGGGATACTTGGGACAGA	Standard	EU251162
WHT1209	13888	GTCAGCAAACATGAGACTTCATG	13887	CGAGATCACACCACTGCAT	Standard	EU251163
WHT2313	13888	GTCAGCAAACATGAGACTTCATG	13887	CGAGATCACACCACTGCAT	Standard	EU251164
WHT1856	13888	GTCAGCAAACATGAGACTTCATG	13887	CGAGATCACACCACTGCAT	Standard	EU251165
WHT4420	13888	GTCAGCAAACATGAGACTTCATG	13887	CGAGATCACACCACTGCAT	Standard	EU251166
GA	13890	TTCTCACCTACTGAGTTGACCGT	13889	AGTCTGGGAAACGTGGTGAG	Standard	EU251167
WHT2357	13715	AAATTAGCCAGGTGTGGTGC	13716	GTTTTGAAATATGTACCCATTGTGG	Standard	EU251168
WHT2364	13752	CACACCTTCCCAGCAGGA	13751	TCTTGTGCTGACCTCCTATCTTG	Standard	EU251169
LGL105	13752	CACACCTTCCCAGCAGGA	13751	TCTTGTGCTGACCTCCTATCTTG	Standard	EU251170
WHT1989	13752	CACACCTTCCCAGCAGGA	13751	TCTTGTGCTGACCTCCTATCTTG	Standard	EU251171
LGL850	13752	CACACCTTCCCAGCAGGA	13751	TCTTGTGCTGACCTCCTATCTTG	Standard	EU251172
WHT3988	13786	CTTCCTGGTAACATAAGTCAGGG	13785	AGCAACAGGAGCAAGGATGA	Standard	EU251173
693	14000	CAGGCAGGCTTCACAGAGTT	13999	GCTACAAAGGGGGAAAACAAG	Standard[e]	EU251174
WHT1010	13582	ATGGCCCACGTTTGCCTA	TranCX1for[d]	GGGCACAGTGGTTCACACTG	Standard	EU251175
WHT2257	13582	ATGGCCCACGTTTGCCTA	TranCX1for[d]	GGGCACAGTGGTTCACACTG	Standard	EU251176
WHT950	13582	ATGGCCCACGTTTGCCTA	TranCX1for[d]	GGGCACAGTGGTTCACACTG	Standard	EU251177
WHT1450	13582	ATGGCCCACGTTTGCCTA	TranCX1for[d]	GGGCACAGTGGTTCACACTG	Standard	EU251178
WHT2127	13582	ATGGCCCACGTTTGCCTA	TranCX1for[d]	GGGCACAGTGGTTCACACTG	Standard	EU251179
WHT2195	13582	ATGGCCCACGTTTGCCTA	TranCX1for[d]	GGGCACAGTGGTTCACACTG	Standard	EU251180
WHT3280	13582	ATGGCCCACGTTTGCCTA	TranCX1for[d]	GGGCACAGTGGTTCACACTG	Standard	EU251181
WHT4816	13582	ATGGCCCACGTTTGCCTA	TranCX1for[d]	GGGCACAGTGGTTCACACTG	Standard	EU251182
WHT5095	13582	ATGGCCCACGTTTGCCTA	TranCX1for[d]	GGGCACAGTGGTTCACACTG	Standard	EU251183
WHT3162	13784	GCTACTGTGATAGGTAGAATAATGGC	13783	GCATTTGCACGTCAAGATCA	Standard	EU251184

[a]For each PCR product, primers used for sequencing were the same as those used for amplification, except: (1) PCR product of 13888/13889 (patients WHT1209, WHT2313, WHT1856, and WHT4420) was sequenced with primer 13888 only, and (2) PCR product of 14000/13999 (patient 693) was sequenced with an internal primer 14075 (sequence: CAGGGTCTCACTCTGTCACAC) in addition to 14000 and 13999.

[b]For XY females, the forward primer is X-specific and the reverse primer is Y-specific. For XX males, the forward primer is Y-specific and the reverse primer is X-specific, except for two reverse primers, 14031 and 14015, each of which is X/Y-common.

[c]PCR protocols Standard and Touchdown as previously described[30].

[d]Ref.[15]

[e]Modified cycling conditions of Standard protocol: 94°C for 3 min; 35 cycles of 94°C for 1 min, 63°C for 1 min, 72°C for 1 min; 72°C for 5 min; hold at 4°C.

Table 7. Amplification and sequencing of hotspot HSA in human X and Y chromosomes.

		Sequencing primers			
	GenBank	Forward primer		Reverse primer	
STS	accession	Name	5′-3′ sequence	Name	5′-3′ sequence
X chromosome:					
sX3237	BV726552	14399	ATTATGTCCTTTGCAGGAAT	14400	CATGTGTTCTCATCGTTCAT
sX3236	BV726551	18800	GAGCTTGGAAACGCTGGAAG	18799	GAAAGTTTCTGTCTCTTATG
Y chromosome:					
sY1867	BV693976	14370	GAGATTATGTCCTTTGCAGGAAC	16332	CATGTGTTCTCATCGTTCAG
sY1866	BV693975	TranDrev[a]	TACTGTGATAGGTAGAATAATGGC	TranDfor[a]	CCTGCCTTTTTTAGTTTCCAGCA

[a]Ref.[15]

chimpanzee X chromosome reference sequence is whole-genome shotgun-based[1,3,55,59,60]. In order to obtain high-quality chimpanzee X chromosome sequence for comparative analysis, we generated clone-based sequence for the chimpanzee X region of interest. Using previously described methods[59], we selected and sequenced three BACs (CH1251-5042G18, GenBank accession number AC189002; CH1251-4112J18, AC189001; CH251-305M6, AC188419) and two fosmids (CH251-2O12, AC186397; CH251-534A21, AC185343) that cover all but an estimated 14 kb of the orthologous chimpanzee X-chromosome region.

SNP selection and genotyping in 92 XY genomic DNAs

From the International HapMap Project database (release 21, NCBI Build 35, dbSNP build 125)[39], we selected 24 SNPs in the 24,198-bp X-chromosome region surrounding hotspot HSA for typing in genomic DNA from 92 unrelated XY males (described above). We corrected marker position assignments for eight SNPs. For genotyping, we designed a total of eight STS assays using Primer3[57] (Table 8). Each assay amplified one or more SNP sites and was sequenced with primers as indicated. STSs and the associated PCR conditions have been deposited in GenBank.

LD plot construction with HapMap Consortium data

To construct an LD plot across the X-chromosome region encompassing *PRKX*, we directly imported, from HapMap into Haploview[38], genotype data for individuals from the Centre d'Etude du Polymorphisme Humain collection (CEU). We minimized uninformative pairwise LD comparisons between markers by selecting SNPs with minor allele frequencies of > 0.10 for incorporation into the LD plot.

Table 8. Amplification and sequencing of 24 SNPs around hotspot HSA in the X chromosome.

		Amplification		Sequencing primer	
SNP	dbSNP	STS	GenBank accession	Name	5′-3′ sequence
Left of HSA:					
1	*rs2009624*	sX3238	BV726553	14354	TGAGCCATGATCGCACTATT
2	*rs4595276*	sX3238	BV726553	14353	CAAAACAGGGTGGGAAAATG
3	*rs5961994*	sX3239	BV726554	18870	TAGAGTCTTTTCTCTTCC
4	*rs5962001*	sX3239	BV726554	18870	TAGAGTCTTTTCTCTTCC
5	*rs7877157*	sX3240	BV726555	18867	CCCAGCTAGTTGGGAGGT
6	*rs6567602*	sX3240	BV726555	18867	CCCAGCTAGTTGGGAGGT
7	*rs5961450*	sX3241	BV726556	18798	ATAGTAATTTTCACTTAACA
Right of HSA:					
8	*rs1965021*	sX3242	BV726557	14349	GCGACAGAGCAAGATTCCAT
9	*rs5961460*	sX3243	BV726558	18866	TCTCCACATTCATTTCCA
10	*rs28631469*	sX3243	BV726558	18866	TCTCCACATTCATTTCCA
11	*rs17331195*	sX3244	BV726559	18861	TGACAGGCTTCTGGTTCT
12	*rs12116286*	sX3244	BV726559	18861	TGACAGGCTTCTGGTTCT
13	*rs12116344*	sX3244	BV726559	18861	TGACAGGCTTCTGGTTCT
14	*rs5962076*	sX3244	BV726559	18862	TAGCAGCATTACTCACAA
15	*rs5962077*	sX3244	BV726559	18862	TAGCAGCATTACTCACAA
16	*rs5962078*	sX3244	BV726559	18862	TAGCAGCATTACTCACAA
17	*rs1483016*	sX3245	BV726560	18858	ACCTGTACATGTACCCCT
18	*rs6641899*	sX3245	BV726560	18858	ACCTGTACATGTACCCCT
19	*rs35133862*	sX3245	BV726560	18858	ACCTGTACATGTACCCCT
20	*rs6641900*	sX3245	BV726560	18858	ACCTGTACATGTACCCCT
21	*rs5961468*	sX3245	BV726560	18857	CAATCTTACATTCTCACC
22	*rs5961469*	sX3245	BV726560	18857	CAATCTTACATTCTCACC
23	*rs2046294*	sX3245	BV726560	18857	CAATCTTACATTCTCACC
24	*rs5961470*	sX3245	BV726560	18857	CAATCTTACATTCTCACC

Acknowledgements

We thank T. Graves and R. Wilson at the Washington University School of Medicine Genome Sequencing Center for sequencing chimpanzee X chromosome clones; and the following individuals for human DNA or blood samples, or cell lines: H. Bode, J. Bodurtha, M. Bofinger, C. Booth, C. Bruning, T. Byrnes, A. de la Chapelle, A. Chudley, M. Clarke, C. Disteche, J. Hall, J. Hersh, B. Lippe, R. McChane, B. McGillivray, R. Oates, E. Pergament, L. Randolph, I. Rosenthal, L. Shapiro, M.H.K. Shokeir, E. Simpson, and M. Willing. Supported by the National Institutes of Health and the Howard Hughes Medical Institute.

References

1. Skaletsky H, Kuroda-Kawaguchi T, Minx PJ, Cordum HS, Hillier L, Brown LG, Repping S, et al. (2003) The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. Nature 423:825-837

2. Lahn BT, Page DC (1999) Four evolutionary strata on the human X chromosome. Science 286:964-967

3. Ross MT, Grafham DV, Coffey AJ, Scherer S, McLay K, Muzny D, Platzer M, et al. (2005) The DNA sequence of the human X chromosome. Nature 434:325-337

4. Charlesworth B, Charlesworth D (2000) The degeneration of Y chromosomes. Philos Trans R Soc Lond B Biol Sci 355:1563-1572

5. Rozen S, Skaletsky H, Marszalek JD, Minx PJ, Cordum HS, Waterston RH, Wilson RK, Page DC (2003) Abundant gene conversion between arms of palindromes in human and ape Y chromosomes. Nature 423:873-876

6. 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

7. 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

8. 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

9. 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

10. Repping S, Skaletsky H, Lange J, Silber S, Van Der Veen F, Oates RD, Page DC, Rozen S (2002) Recombination between palindromes P5 and P1 on the human Y chromosome causes massive deletions and spermatogenic failure. Am J Hum Genet 71:906-922

11. Repping S, Skaletsky H, Brown L, van Daalen SK, Korver CM, Pyntikova T, Kuroda-Kawaguchi T, de Vries JW, Oates RD, Silber S, van der Veen F, Page DC, Rozen S (2003) Polymorphism for a 1.6-Mb deletion of the human Y chromosome persists through balance between recurrent mutation and haploid selection. Nat Genet 35:247-251

12. Lange J, Skaletsky H, van Daalen SKM, Korver CM, Brown LG, Oates RD, Silber S, Repping S, Page DC Isochromosomes and associated reproductive disorders are byproducts of the homologous recombination that maintains human Y-chromosomal palindromes. Manuscript in preparation

13. Weil D, Wang I, Dietrich A, Poustka A, Weissenbach J, Petit C (1994) Highly homologous loci on the X and Y chromosomes are hot-spots for ectopic recombinations leading to XX maleness. Nat Genet 7:414-419

14. Wang I, Weil D, Levilliers J, Affara NA, de la Chapelle A, Petit C (1995) Prevalence and molecular analysis of two hot spots for ectopic recombination leading to XX maleness. Genomics 28:52-58

15. Schiebel K, Winkelmann M, Mertz A, Xu X, Page DC, Weil D, Petit C, Rappold GA (1997) Abnormal XY interchange between a novel isolated protein kinase gene, PRKY, and its homologue, PRKX, accounts for one third of all (Y+)XX males and (Y-)XY females. Hum Mol Genet 6:1985-1989

16. Jobling MA, Williams GA, Schiebel GA, Pandya GA, McElreavey GA, Salas GA, Rappold GA, Affara NA, Tyler-Smith C (1998) A selective difference between human Y-chromosomal DNA haplotypes. Curr Biol 8:1391-1394

17. Szostak JW, Orr-Weaver TL, Rothstein RJ, Stahl FW (1983) The double-strand-break repair model for recombination. Cell 33:25-35

18. Allers T, Lichten M (2001) Differential timing and control of noncrossover and crossover recombination during meiosis. Cell 106:47-57

19. Börner GV, Kleckner N, Hunter N (2004) Crossover/noncrossover differentiation, synaptonemal complex formation, and regulatory surveillance at the leptotene/zygotene transition of meiosis. Cell 117:29-45

20. Guillon H, Baudat F, Grey C, Liskay RM, de Massy B (2005) Crossover and noncrossover pathways in mouse meiosis. Mol Cell 20:563-573

21. Vergnaud G, Page DC, Simmler MC, Brown L, Rouyer F, Noel B, Botstein D, de la Chapelle A, Weissenbach J (1986) A deletion map of the human Y chromosome based on DNA hybridization. Am J Hum Genet 38:109-124

22. Andersson M, Page DC, de la Chapelle A (1986) Chromosome Y-specific DNA is transferred to the short arm of X chromosome in human XX males. Science 233:786-788

23. Vollrath D, Foote S, Hilton A, Brown LG, Beer-Romero P, Bogan JS, Page DC (1992) The human Y chromosome: a 43-interval map based on naturally occurring deletions. Science 258:52-59

24. Lahn BT, Ma N, Breg WR, Stratton R, Surti U, Page DC (1994) Xq-Yq interchange resulting in supernormal X-linked gene expression in severely retarded males with 46,XYq- karyotype. Nat Genet 8:243-250

25. Reijo R, Lee TY, Salo P, Alagappan R, Brown LG, Rosenberg M, Rozen S, Jaffe T, Straus D, Hovatta O, et al. (1995) Diverse spermatogenic defects in humans caused by Y chromosome deletions encompassing a novel RNA-binding protein gene. Nat Genet 10:383-393

26. Page DC (1986) Sex reversal: deletion mapping the male-determining function of the human Y chromosome. Cold Spring Harb Symp Quant Biol 51 Pt 1:229-235

27. Affara NA, Ferguson-Smith MA, Tolmie J, Kwok K, Mitchell M, Jamieson D, Cooke A, Florentin L (1986) Variable transfer of Y-specific sequences in XX males. Nucleic Acids Res 14:5375-5387

28. Tilford CA, Kuroda-Kawaguchi T, Skaletsky H, Rozen S, Brown LG, Rosenberg M, McPherson JD, Wylie K, Sekhon M, Kucaba TA, Waterston RH, Page DC (2001) A physical map of the human Y chromosome. Nature 409:943-945

29. Repping S, van Daalen SK, Brown LG, Korver CM, Lange J, Marszalek JD, Pyntikova T, van der Veen F, Skaletsky H, Page DC, Rozen S (2006) High mutation rates have driven extensive structural polymorphism among human Y chromosomes. Nat Genet 38:463-467

30. Lange J, Skaletsky H, Bell GW, Page DC (2007) MSY Breakpoint Mapper, a database of sequence-tagged sites useful in defining naturally occurring deletions in the human Y chromosome. Nucleic Acids Res published online October 26

31. Jeffreys AJ, Kauppi L, Neumann R (2001) Intensely punctate meiotic recombination in the class II region of the major histocompatibility complex. Nat Genet 29:217-222

32. Myers S, Bottolo L, Freeman C, McVean G, Donnelly P (2005) A fine-scale map of recombination rates and hotspots across the human genome. Science 310:321-324

33. Jeffreys AJ, Neumann R (2002) Reciprocal crossover asymmetry and meiotic drive in a human recombination hot spot. Nat Genet 31:267-271

34. Guillon H, de Massy B (2002) An initiation site for meiotic crossing-over and gene conversion in the mouse. Nat Genet 32:296-299

35. Jeffreys AJ, May CA (2004) Intense and highly localized gene conversion activity in human meiotic crossover hot spots. Nat Genet 36:151-156

36. Bosch E, Hurles ME, Navarro A, Jobling MA (2004) Dynamics of a human interparalog gene conversion hotspot. Genome Res 14:835-844

37. Brunet M, Guy F, Pilbeam D, Mackaye HT, Likius A, Ahounta D, Beauvilain A, et al. (2002) A new hominid from the Upper Miocene of Chad, Central Africa. Nature 418:145-151

38. Barrett JC, Fry B, Maller J, Daly MJ (2005) Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 21:263-265

39. Frazer KA, Ballinger DG, Cox DR, Hinds DA, Stuve LL, Gibbs RA, Belmont JW, et al. (2007) A second generation human haplotype map of over 3.1 million SNPs. Nature 449:851-861

40. Reich DE, Cargill M, Bolk S, Ireland J, Sabeti PC, Richter DJ, Lavery T, Kouyoumjian R, Farhadian SF, Ward R, Lander ES (2001) Linkage disequilibrium in the human genome. Nature 411:199-204

41. Daly MJ, Rioux JD, Schaffner SF, Hudson TJ, Lander ES (2001) High-resolution haplotype structure in the human genome. Nat Genet 29:229-232

42. Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, Higgins J, DeFelice M, Lochner A, Faggart M, Liu-Cordero SN, Rotimi C, Adeyemo A, Cooper R, Ward R, Lander ES, Daly MJ, Altshuler D (2002) The structure of haplotype blocks in the human genome. Science 296:2225-2229

43. McVean GA, Myers SR, Hunt S, Deloukas P, Bentley DR, Donnelly P (2004) The fine-scale structure of recombination rate variation in the human genome. Science 304:581-584

44. Winckler W, Myers SR, Richter DJ, Onofrio RC, McDonald GJ, Bontrop RE, McVean GA, Gabriel SB, Reich D, Donnelly P, Altshuler D (2005) Comparison of fine-scale recombination rates in humans and chimpanzees. Science 308:107-111

45. Ptak SE, Hinds DA, Koehler K, Nickel B, Patil N, Ballinger DG, Przeworski M, Frazer KA, Pääbo S (2005) Fine-scale recombination patterns differ between chimpanzees and humans. Nat Genet 37:429-434

46. Jeffreys AJ, Neumann R, Panayi M, Myers S, Donnelly P (2005) Human recombination hot spots hidden in regions of strong marker association. Nat Genet 37:601-606

47. De Raedt T, Stephens M, Heyns I, Brems H, Thijs D, Messiaen L, Stephens K, Lazaro C, Wimmer K, Kehrer-Sawatzki H, Vidaud D, Kluwe L, Marynen P, Legius E (2006) Conservation of hotspots for recombination in low-copy repeats associated with the NF1 microdeletion. Nat Genet 38:1419-1423

48. Lopez Correa C, Brems H, Lazaro C, Marynen P, Legius E (2000) Unequal meiotic crossover: a frequent cause of NF1 microdeletions. Am J Hum Genet 66:1969-1974

49. Lindsay SJ, Khajavi M, Lupski JR, Hurles ME (2006) A chromosomal rearrangement hotspot can be identified from population genetic variation and is coincident with a hotspot for allelic recombination. Am J Hum Genet 79:890-902

50. Pentao L, Wise CA, Chinault AC, Patel PI, Lupski JR (1992) Charcot-Marie-Tooth type 1A duplication appears to arise from recombination at repeat sequences flanking the 1.5 Mb monomer unit. Nat Genet 2:292-300

51. Chance PF, Abbas N, Lensch MW, Pentao L, Roa BB, Patel PI, Lupski JR (1994) Two autosomal dominant neuropathies result from reciprocal DNA duplication/deletion of a region on chromosome 17. Hum Mol Genet 3:223-228

52. Guioli S, Incerti B, Zanaria E, Bardoni B, Franco B, Taylor K, Ballabio A, Camerino G (1992) Kallmann syndrome due to a translocation resulting in an X/Y fusion gene. Nat Genet 1:337-340

53. Aitken RJ, Marshall Graves JA (2002) The future of sex. Nature 415:963

54. Graves JA, Koina E, Sankovic N (2006) How the gene content of human sex chromosomes evolved. Curr Opin Genet Dev 16:219-224

55. Hughes JF, Skaletsky H, Pyntikova T, Minx PJ, Graves T, Rozen S, Wilson RK, Page DC (2005) Conservation of Y-linked genes during human evolution revealed by comparative sequencing in chimpanzee. Nature 437:100-103

56. Klink A, Schiebel K, Winkelmann M, Rao E, Horsthemke B, Ludecke HJ, Claussen U, Scherer G, Rappold G (1995) The human protein kinase gene PKX1 on Xp22.3 displays Xp/Yp homology and is a site of chromosomal instability. Hum Mol Genet 4:869-878

57. Rozen S, Skaletsky H (2000) Primer3 on the WWW for general users and for biologist programmers. Methods Mol Biol 132:365-386

58. Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22:4673-4680

59. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, et al. (2001) Initial sequencing and analysis of the human genome. Nature 409:860-921

60. Mikkelsen TS, Hillier LW, Eichler EE, Zody MC, Jaffe DB, Yang S-P, Enard W, et al. (2005) Initial sequence of the chimpanzee genome and comparison with the human genome. Nature 437:69-87

Chapter 5

Conclusions

Julian Lange

In this thesis, I explore the modes and outcomes of homologous recombination in the male-specific region of the human Y chromosome (MSY). First, I show that Y isochromosomes associated with diverse reproductive disorders, including male infertility, Turner syndrome, and sex reversal, arise via homology-mediated crossing over between opposing palindrome arms on sister chromatids. Second, I provide evidence of extensive gene conversion between formerly allelic regions of the X and Y chromosomes. Together, these findings strongly suggest that homologous recombination operates in the MSY, previously thought to be devoid of such processes.

The canonical double-strand break repair (DSBR) model of recombination between homologous chromosomes predicts that sites of crossing and gene conversion coincide[1]. Each of the MSY territories investigated in this thesis meets that prediction. Arm-to-arm sequence identity of MSY palindromes is maintained by gene conversion[2]; in Chapter 3 on the formation of Y isochromosomes by intrapalindrome crossing over, we propose that such gene conversion occurs via double-strand break formation in one arm and repair synthesis templated by homologous sequences in the other arm. Conversely, the X- and Y-chromosome regions that engage in sequence homogenization are known to undergo ectopic homologous crossing over, leading to X-Y translocations associated with sex reversal syndromes[3,4]. The coincidence of MSY sites of crossing over and gene conversion suggests that the DSBR model, well studied in terms of recombination between alleles, might similarly apply to sequences in the MSY.

Have mechanisms evolved to preserve MSY gene content at the price of an occasional evolutionary "dead end"? Intrapalindromic gene conversion may be sheltering palindrome-borne genes[2] and, as speculated in Chapter 4, X-Y gene conversion may be preserving Y-linked genes with homologs in the X chromosome. However, such gene conversion bears an Achilles' heel: DSBs formed in palindromes and X-degenerate sequences and resolved by crossing over leads to Y isochromosomes and X-Y translocations, respectively. The vast majority of individuals who carry such structural abnormalities are sterile.

Future directions

The principal findings presented in this thesis prompt additional questions and pave avenues of further exploration. Since the Y chromosome is rife with amplicons in direct and inverted orientation, what intrachromosomal homologous recombination events have yet to be discovered? If the intercentromeric distance in a dicentric isoYp chromosome – carrying two copies of the sex-determining gene *SRY* – influences phenotypic sex, what further predictions arise about centromere function on such a chromosome? Does the proposed mechanism of crossing over within palindromes operate in the formation of other human isochromosomes? Is X-Y gene conversion between formerly allelic regions in the X and Y chromosomes limited to the regions described in Chapter 4?

In each of the ectopic intra-Y-chromosome homologous recombination events characterized to date – interstitial deletions[5-13], inversions[14], and Y isochromosomes reported in Chapter 3 – both targeted amplicons are located in the same Y-chromosome arm. The Y chromosome also contains highly homologous inverted repeats comprised of one amplicon on the short arm (Yp) and one amplicon on the long arm (Yq)[15]. While intrachromatid crossing over between such inverted repeats would result in a pericentromeric inversion, among the first Y-chromosome structural abnormalities identified[16], interchromatid crossing over would result in a Y chromosome that is partially duplicated for one arm and partially deleted for the other. We are currently investigating whether these inverted repeats are targets for ectopic homologous recombination.

Of the 65 individuals with isoYp chromosomes and thus two copies of *SRY*, 15 were female. We postulated that in these 15 females, discordance between phenotypic gender and genotypic sex was due to a combination of 1) loss of the isoYp chromosome in some cell lineages, including the gonadal supporting cells that express *SRY* in the developing embryo, and 2) retention of the isoYp chromosome in other cell lineages, including lymphocytes from which the cell lines analyzed in this study derive. (Mosaicism could similarly account for the diversity of somatic phenotypes among the 65 individuals.) Chromosomes carrying two functional centromeres are mitotically unstable[17]. Thus, dicentric

chromosomes persist only if one centromere is functionally inactivated or if the two centromeres are close together (reviewed in ref.[18]). Our observation that higher intercentromeric distance in isoYp chromosomes is correlated with phenotypic females is concordant with previous findings that intercentromeric distance affects mitotic stability[19-21].

In an isoYp chromosome with a short intercentromeric distance, the two centromeres might be sufficiently close together to obviate the need for inactivation of one of the two centromeres. It would be of interest to contrast, in cell culture assays, centromere function of isoYp chromosomes formed at proximal Yq palindromes such as P8 (intercentromeric distance = 4.5 Mb) and isoYp chromosomes formed at distal Yq palindromes such as P1 (intercentromeric distance = 26.1 Mb). To test centromere function in a subset of isoYp chromosomes, we have initiated immunofluorescence experiments employing antibodies to CENP-C, a protein that is exclusively associated with functional centromeres[22]. To the extent that such *in vitro* assays recapitulate *in vivo* events, isoYp chromosomes with short intercentromeric distances might be more likely to retain two functional centromeres.

An interesting question is whether other human isochromosomes – constitutional or somatic – form via the same mechanism. The most commonly observed constitutional isochromosome contains two copies of the X-chromosome long arm and can cause Turner syndrome. Most of these isoXq chromosomes are dicentric[23] and contain sequences from the short arm[24-28], which is known to harbor palindromes[29]. In cataloging inverted repeats in the human X chromosome, Warburton and colleagues[29] suggested that X isochromosomes are generated by sister-chromatid exchange at these repeats. Similarly, dicentric iso17q chromosomes are the most common isochromosomes in human cancer[30]. In investigating isoYq chromosomes from 11 patients, Barbouti and colleagues[31] determined the reference genomic structure of the 17p breakpoint region and discovered large, palindromic, low-copy repeats there. Further studies may implicate a role for homologous recombination in the formation of these and other isochromosomes.

Our exploration of gene conversion between formerly allelic regions encompassing *PRKX* and *PRKY* was motivated by previous findings of ectopic crossing over between these domains[3,4].

Accordingly, the observation of extensive X-Y sequence homogenization at homeologous regions encompassing *KAL1* and *KALP* on the X and Y chromosome, respectively[15], should forecast ectopic crossing over there. Indeed, Guioli and colleagues[32] reported such an X-Y translocation in a sterile male presenting with short stature, mental retardation, and features of Kallmann syndrome including anosmia and hypogonadism. Among our collection of individuals with apparent terminal deletions, we have also observed X-Y translocations between *KAL1* and *KALP*.

Has gene conversion between the X and Y chromosome occurred elsewhere? In each of the two cases described above, suppression of allelic recombination has been dated at roughly 30 million years[33]. Thus, X-Y divergence is limited and gene conversion can be observed by aligning genomic DNA sequences. However, many ancestral X-Y gene pairs lie in genomic regions that cannot be informatively aligned. In such cases, other methods, such as alignment of coding sequences and measurement of synonymous substitution rates, may yield further evidence of X-Y sequence homogenization. It will be intriguing to examine the extent to which MSY genes effectively reside in the same allele pool as their X-linked homologs.

Acknowledgements

I thank J. Mueller for his comments.

References

1. Szostak JW, Orr-Weaver TL, Rothstein RJ, Stahl FW (1983) The double-strand-break repair model for recombination. Cell 33:25-35

2. Rozen S, Skaletsky H, Marszalek JD, Minx PJ, Cordum HS, Waterston RH, Wilson RK, Page DC (2003) Abundant gene conversion between arms of palindromes in human and ape Y chromosomes. Nature 423:873-876

3. Weil D, Wang I, Dietrich A, Poustka A, Weissenbach J, Petit C (1994) Highly homologous loci on the X and Y chromosomes are hot-spots for ectopic recombinations leading to XX maleness. Nat Genet 7:414-419

4. Schiebel K, Winkelmann M, Mertz A, Xu X, Page DC, Weil D, Petit C, Rappold GA (1997) Abnormal XY interchange between a novel isolated protein kinase gene, PRKY, and its homologue, PRKX, accounts for one third of all (Y+)XX males and (Y-)XY females. Hum Mol Genet 6:1985-1989

5. 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

6. 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

7. 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

8. 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

9. Repping S, Skaletsky H, Lange J, Silber S, Van Der Veen F, Oates RD, Page DC, Rozen S (2002) Recombination between palindromes P5 and P1 on the human Y chromosome causes massive deletions and spermatogenic failure. Am J Hum Genet 71:906-922

10. Repping S, Skaletsky H, Brown L, van Daalen SK, Korver CM, Pyntikova T, Kuroda-Kawaguchi T, de Vries JW, Oates RD, Silber S, van der Veen F, Page DC, Rozen S (2003) Polymorphism for a 1.6-Mb deletion of the human Y chromosome persists through balance between recurrent mutation and haploid selection. Nat Genet 35:247-251

11. Fernandes S, Paracchini S, Meyer LH, Floridia G, Tyler-Smith C, Vogt PH (2004) A large AZFc deletion removes DAZ3/DAZ4 and nearby genes from men in Y haplogroup N. Am J Hum Genet 74:180-187

12. Repping S, van Daalen SK, Korver CM, Brown LG, Marszalek JD, Gianotten J, Oates RD, Silber S, van der Veen F, Page DC, Rozen S (2004) A family of human Y chromosomes has dispersed throughout northern Eurasia despite a 1.8-Mb deletion in the azoospermia factor c region. Genomics 83:1046-1052

13. Jobling MA, Lo IC, Turner DJ, Bowden GR, Lee AC, Xue Y, Carvalho-Silva D, Hurles ME, Adams SM, Chang YM, Kraaijenbrink T, Henke J, Guanti G, McKeown B, van Oorschot RA, Mitchell RJ, de Knijff P, Tyler-Smith C, Parkin EJ (2007) Structural variation on the short arm of the human Y chromosome: recurrent multigene deletions encompassing Amelogenin Y. Hum Mol Genet 16:307-316

14. Repping S, van Daalen SK, Brown LG, Korver CM, Lange J, Marszalek JD, Pyntikova T, van der Veen F, Skaletsky H, Page DC, Rozen S (2006) High mutation rates have driven extensive structural polymorphism among human Y chromosomes. Nat Genet 38:463-467

15. Skaletsky H, Kuroda-Kawaguchi T, Minx PJ, Cordum HS, Hillier L, Brown LG, Repping S, et al. (2003) The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. Nature 423:825-837

16. Jacobs PA, Ross A (1966) Structural abnormalities of the Y chromosome in man. Nature 210:352-354

17. McClintock B (1939) The behavior in successive nuclear divisions of a chromosome broken at meiosis. Proc Natl Acad Sci USA 25:405-416

18. Therman E, Trunca C, Kuhn EM, Sarto GE (1986) Dicentric chromosomes and the inactivation of the centromere. Hum Genet 72:191-195

19. Hair JB (1953) The origin of new chromosomes in Agropyron. Heredity 6 (Suppl.):215-233

20. Koshland D, Rutledge L, Fitzgerald-Hayes M, Hartwell LH (1987) A genetic analysis of dicentric minichromosomes in Saccharomyces cerevisiae. Cell 48:801-812

21. Sullivan BA, Willard HF (1998) Stable dicentric X chromosomes with two functional centromeres. Nat Genet 20:227-228

22. Earnshaw WC, Ratrie H, 3rd, Stetten G (1989) Visualization of centromere proteins CENP-B and CENP-C on a stable dicentric chromosome in cytological spreads. Chromosoma 98:1-12

23. de la Chapelle A, Stenstrand K (1974) Dicentric human X chromosomes. Hereditas 76:259-268

24. Callen DF, Mulley JC, Baker EG, Sutherland GR (1987) Determining the origin of human X isochromosomes by use of DNA sequence polymorphisms and detection of an apparent i(Xq) with Xp sequences. Hum Genet 77:236-240

25. Sharp CB, Bedford HM, Willard HF (1990) Pericentromeric structure of human X "isochromosomes": evidence for molecular heterogeneity. Hum Genet 85:330-336

26. Lorda-Sanchez I, Binkert F, Maechler M, Schinzel A (1991) A molecular study of X isochromosomes: parental origin, centromeric structure, and mechanisms of formation. Am J Hum Genet 49:1034-1040

27. Wolff DJ, Miller AP, Van Dyke DL, Schwartz S, Willard HF (1996) Molecular definition of breakpoints associated with human Xq isochromosomes: implications for mechanisms of formation. Am J Hum Genet 58:154-160

28. James RS, Dalton P, Gustashaw K, Wolff DJ, Willard HF, Mitchell C, Jacobs PA (1997) Molecular characterization of isochromosomes of Xq. Ann Hum Genet 61:485-490

29. Warburton PE, Giordano J, Cheung F, Gelfand Y, Benson G (2004) Inverted repeat structure of the human genome: the X-chromosome contains a preponderance of large, highly homologous inverted repeats that contain testes genes. Genome Res 14:1861-1869

30. Mertens F, Johansson B, Mitelman F (1994) Isochromosomes in neoplasia. Genes Chromosomes Cancer 10:221-230

31. Barbouti A, Stankiewicz P, Nusbaum C, Cuomo C, Cook A, Hoglund M, Johansson B, Hagemeijer A, Park SS, Mitelman F, Lupski JR, Fioretos T (2004) The breakpoint region of the most common isochromosome, i(17q), in human neoplasia is characterized by a complex genomic architecture with large, palindromic, low-copy repeats. Am J Hum Genet 74:1-10

32. Guioli S, Incerti B, Zanaria E, Bardoni B, Franco B, Taylor K, Ballabio A, Camerino G (1992) Kallmann syndrome due to a translocation resulting in an X/Y fusion gene. Nat Genet 1:337-340

33. Ross MT, Grafham DV, Coffey AJ, Scherer S, McLay K, Muzny D, Platzer M, et al. (2005) The DNA sequence of the human X chromosome. Nature 434:325-337

Appendix A

Recombination between palindromes P5 and P1 on the human Y chromosome causes massive deletions and spermatogenic failure

Sjoerd Repping, Helen Skaletsky, Julian Lange, Sherman Silber, Fulco van der Veen, Robert D. Oates, David C. Page, Steve Rozen

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.

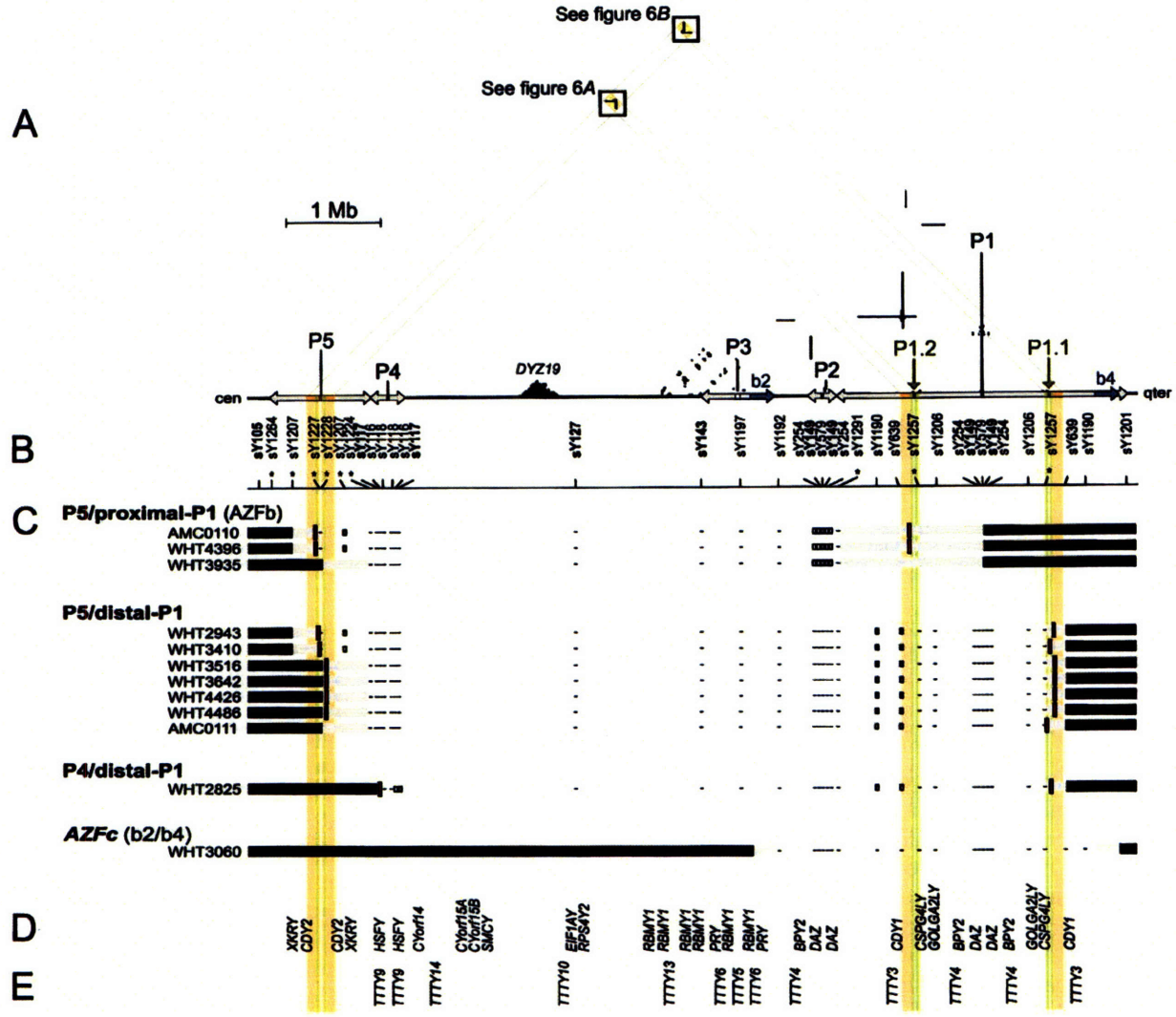


Figure 1 Deletions between P5 and P1 and between P4 and P1 in relation to the sequence of the Y chromosome. *A*, Triangular dot plot, encompassing P5, P4, and *AZFc* (which contains P1, P2, and part of P3). The baseline represents 8 Mb of Y-chromosome long-arm sequence. This triangular display avoids 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

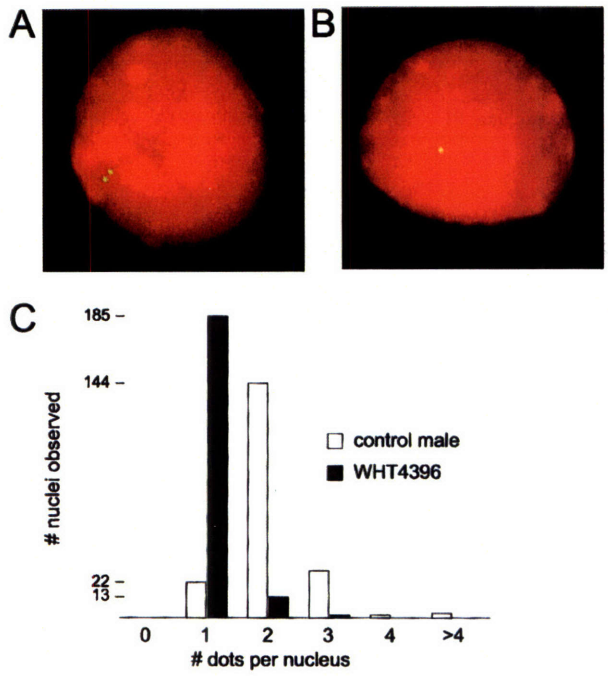


Figure 2 FISH, showing *DAZ* clusters in an unaffected man and in a man with a P5/proximal-P1 (*AZFb*) deletion. *A*, Representative interphase nucleus from a fertile, male control individual hybridized with cosmid 18E8 (*yellow*), showing two dots (corresponding to two *DAZ* clusters, each containing 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

Gene[a]	Expression Pattern[b]	No. of Copies	No. of Copies Deleted		
			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).

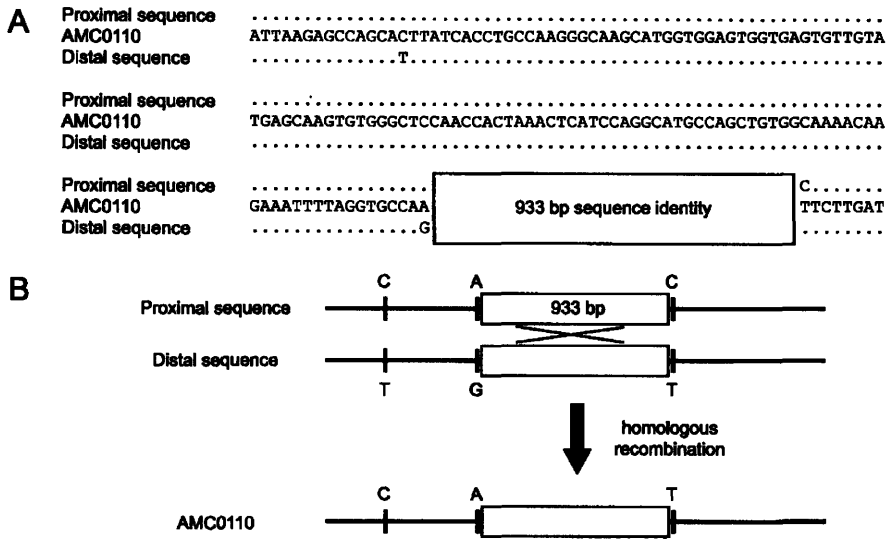
Breakpoints Cluster near Central P5, P1.1, and P1.2

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
Distal sequence
ATTAAGAGCCAGCACTTATCACCTGCCAAGGGCAAGCATGGTGGAGTGGTGAGTGTTGTA
Proximal sequence
AMC0110
Distal sequence
TGAGCAAGTGTGGGCTCCAACCACTAAACTCATCCAGGCATGCCAGCTGTGGCAAAACAA
Proximal sequence
AMC0110
Distal sequence
GAAATTTTAGGTGCCAA
933 bp sequence identity
TTCTTGAT
B
Proximal sequence
Distal sequence
C
A
C
933 bp
T
G
T
homologous recombination
AMC0110
C
A
T
C
Proximal sequence
WHT4396
Distal sequence
AGAGGCAAGGTGGGTGTAGGTACCATAATGGACAGCTGTCGCAAAGCAGCAATCAGAAGA
Proximal sequence
WHT4396
Distal sequence
GTCTGACACCTGTG
73 bp sequence identity
TTAAGTTATAGAAACAGAAA
Proximal sequence
WHT4396
Distal sequence
ACTTTGGTGTCAAATAGATGAAGAACTATTCTAAATTATAAAAACAGAACCACAGCTTCT
D
Proximal sequence
WHT2943
Distal sequence
ATAGACAAGTAACAGGCTCTGAAATTCAGGCAATAAGTAATATTGCAAAAAAAAAGTCCA
Proximal sequence
WHT2943
Distal sequence
GGACCAGATGGATTCACATCCGAATTCTACCAGAGGTACAAGGAGGAGCTGGTACCATTC
Proximal sequence
WHT2943
Distal sequence
CTTCTGAAACGAT
25 bp sequence identity
TCCTCTCTAACTCATTTTATGAG
Proximal sequence
WHT2943
Distal sequence
GCCAGCATCATCCTGATACCAAAGCCTGGCAGAGACACAGCAAAAATAGAGAATTTTAGA
Proximal sequence
WHT2943
Distal sequence
CCAAAATCCCTGATGAACATCAATGCAAGAATCGTCAATAAAAAACTGGCAAACCGAACC
E
Proximal sequence
Patients = WHT3516 / WHT3642 / WHT4426 / WHT4486
Distal sequence
CAAAGATGCGGGTACTGCTCTATTTCACAAG
Proximal sequence
Patients
Distal sequence
GCATTGATCAAGAG
933 bp sequence identity
CTGGCACCTAA
Proximal sequence
Patients
Distal sequence
AATTTCTTGTTTTGCCACAGCTGGCATGCCTGGATGAGTTTAGTGGTTGGAGCCCACACT
Proximal sequence
Patients
Distal sequence
TGCTCATACAACACTCACCACTCCACCATGCTTGCCCTTGGCAGGTGATAAATGCTGGCT

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.

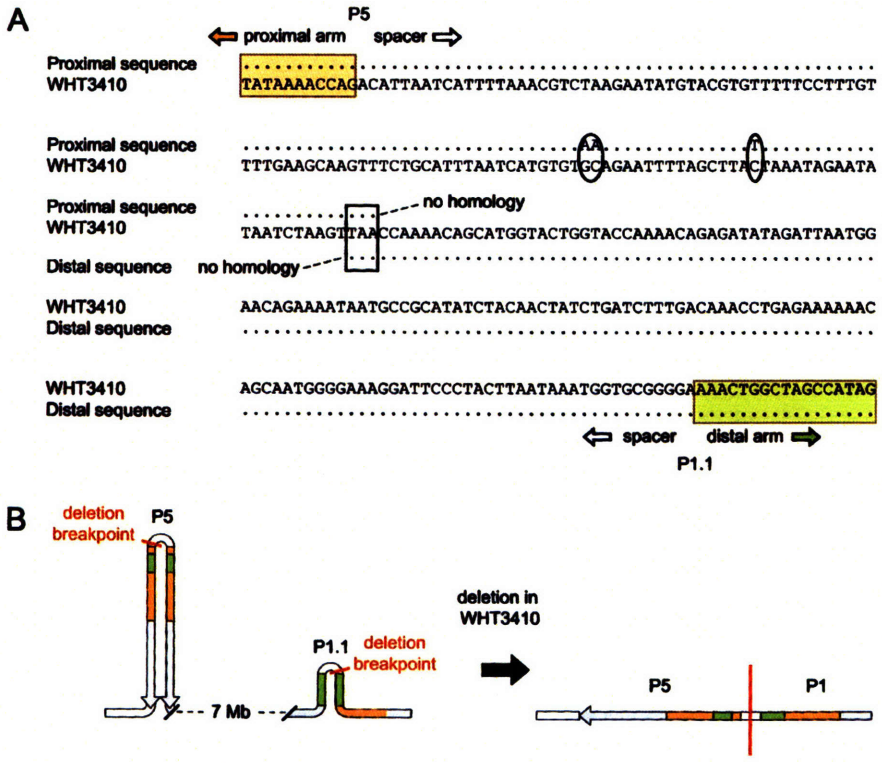


Figure 4 P5/distal-P1 deletion junction in WHT3410. *A*, Junction sequence aligned with proximal and distal Y-chromosome reference sequences. Ovals mark 3 bp that differ between WHT3410 and the Y-chromosome reference sequence. The white box marks the 3 bp that contain the deletion junction. *B*, Breakpoint 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

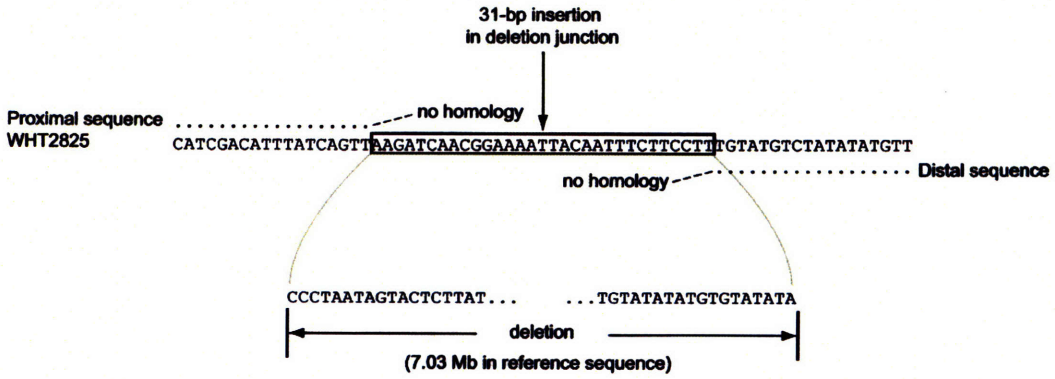


Figure 5 P4/distal-P1 deletion junction in WHT2825. Junction sequence is aligned with proximal and distal Y-chromosome reference sequence.

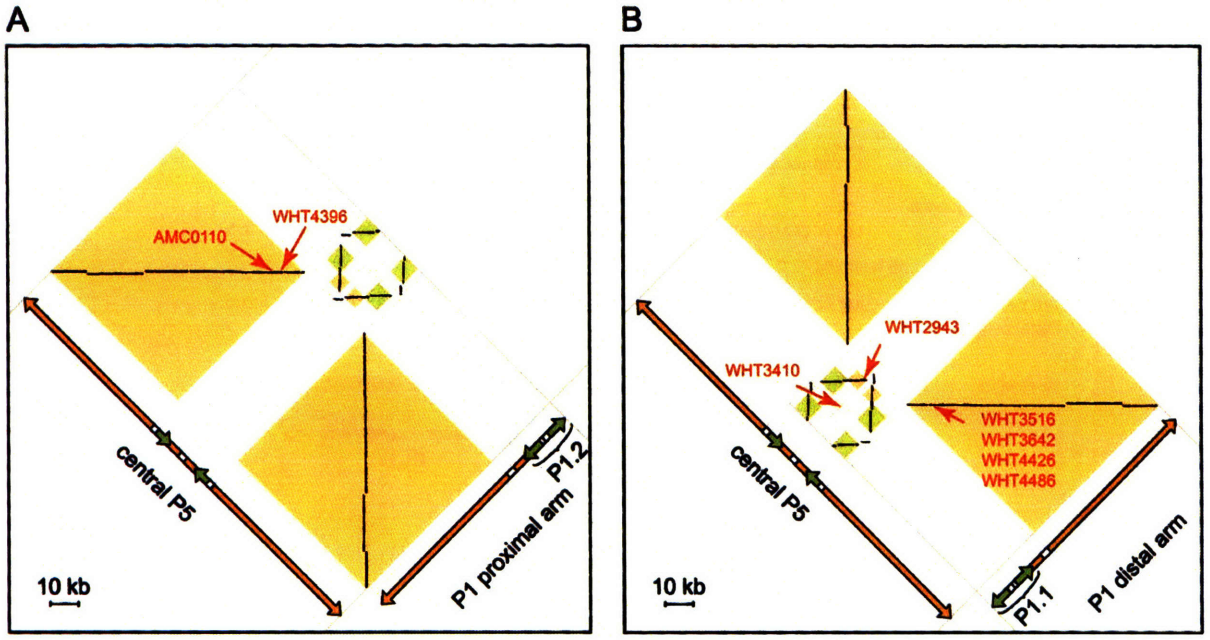


Figure 6 "High-magnification" dot plots of the two regions of high similarity between P5 and P1 (see fig. 1*A*). Each dot represents 50 bp of identity in a window of 50 bp. Orientation and orange and green shading is as in figure 1*A*. *A*, Central P5 versus P1.2 and the neighboring region of the proximal 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

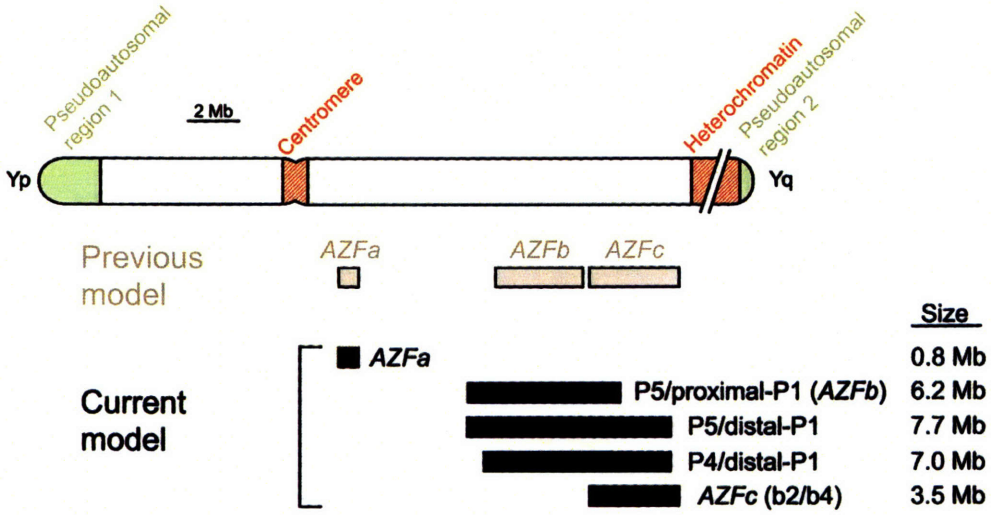


Figure 7 Previous model of recurrent, interstitial Y-chromosome deletions that cause infertility in men (Vogt et al. 1996) contrasted with current model.

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

Appendix B

High mutation rates have driven extensive structural polymorphism among human Y chromosomes

Sjoerd Repping, Saskia K.M. van Daalen, Laura G. Brown, Cindy M. Korver, Julian Lange, Janet D. Marszalek, Tatyana Pyntikova, Fulco van der Veen, Helen Skaletsky, David C. Page, and Steve Rozen

nature genetics

High mutation rates have driven extensive structural polymorphism among human Y chromosomes

Sjoerd Repping[1,2], Saskia K M van Daalen[2], Laura G Brown[1], Cindy M Korver[2], Julian Lange[1], Janet D Marszalek[1], Tatyana Pyntikova[1], Fulco van der Veen[2], Helen Skaletsky[1], David C Page[1] & Steve Rozen[1]

Although much structural polymorphism in the human genome has been catalogued[1–5], the kinetics of underlying change remain largely unexplored. Because human Y chromosomes are clonally inherited, it has been possible to capture their detailed relationships in a robust, worldwide genealogical tree[6,7]. Examination of structural variation across this tree opens avenues for investigating rates of underlying mutations. We selected one Y chromosome from each of 47 branches of this tree and searched for large-scale variation. Four chromosomal regions showed extensive variation resulting from numerous large-scale mutations. Within the tree encompassed by the studied chromosomes, the distal-Yq heterochromatin changed length ≥12 times, the *TSPY* gene array changed length ≥23 times, the 3.6-Mb IR3/IR3 region changed orientation ≥12 times and the *AZFc* region was rearranged ≥20 times. After determining the total time spanned by all branches of this tree (~1.3 million years or 52,000 generations), we converted these mutation counts to lower bounds on rates: $\geq 2.3 \times 10^{-4}$, $\geq 4.4 \times 10^{-4}$, $\geq 2.3 \times 10^{-4}$ and $\geq 3.8 \times 10^{-4}$ large-scale mutations per father-to-son Y transmission, respectively. Thus, high mutation rates have driven extensive structural polymorphism among human Y chromosomes. At the same time, we found limited variation in the copy number of Y-linked genes, which raises the possibility of selective constraints.

Recent studies point to substantial large-scale copy number variation within the human genome, and a few such studies have shown large inversions[1–5]. With the exception of one large inversion that arose once in human history[4], the mutational dynamics underlying common large-scale, structural polymorphisms have not been addressed. Are these polymorphisms usually the result of independent, recurrent mutation, or are they inherited from a single founder? How often do mutations generate structural variants? The male-specific region of the human Y chromosome (**Fig. 1**) offers unique opportunities for investigating these questions because of its clonal inheritance and the availability of a robust genealogical tree that describes in detail the relationships among extant Y chromosomes (**Fig. 2**)[6,7].

With these questions in mind, we assembled a collection of Y chromosomes representing 47 major branches of the genealogy and encompassing worldwide diversity (**Fig. 2** and **Supplementary Table 1** and **Supplementary Fig. 1** online). On the basis of analysis of published empirical evidence, we searched for nine broad categories of potential structural variation among these chromosomes (**Fig. 1a, 2, 3** and **Supplementary Methods** online). For several reasons, we focused on potential structural variation involving the segmentally duplicated, ampliconic regions of the Y chromosome. Previous results

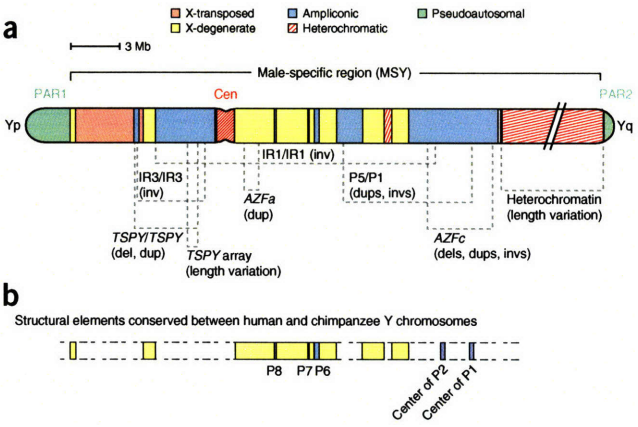


Figure 1 Overview of potential structural variation in the human Y chromosome. At top, the structure of the reference Y chromosome, including short and long arms (Yp and Yq), pseudoautosomal regions 1 and 2 (PAR1 and PAR2) and centromere (Cen). (**a**) Potential structural polymorphisms for which we assayed (details in **Supplementary Methods**). (**b**) Structural elements conserved between human and chimpanzee Y chromosomes are shown according to their position in the reference human Y chromosome. These conserved elements consist of the X-degenerate sequence, palindromes P8, P7 and P6 and the centers of palindromes P2 and P1.

[1]Howard Hughes Medical Institute, Whitehead Institute, and Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02142, USA. [2]Center for Reproductive Medicine, Department of Obstetrics and Gynecology, Academic Medical Center, Amsterdam 1105 AZ, The Netherlands. Correspondence should be addressed to D.C.P. (page_admin@wi.mit.edu).

Received 21 December 2005; accepted 31 January 2006; published online 26 February 2006; doi:10.1038/ng1754

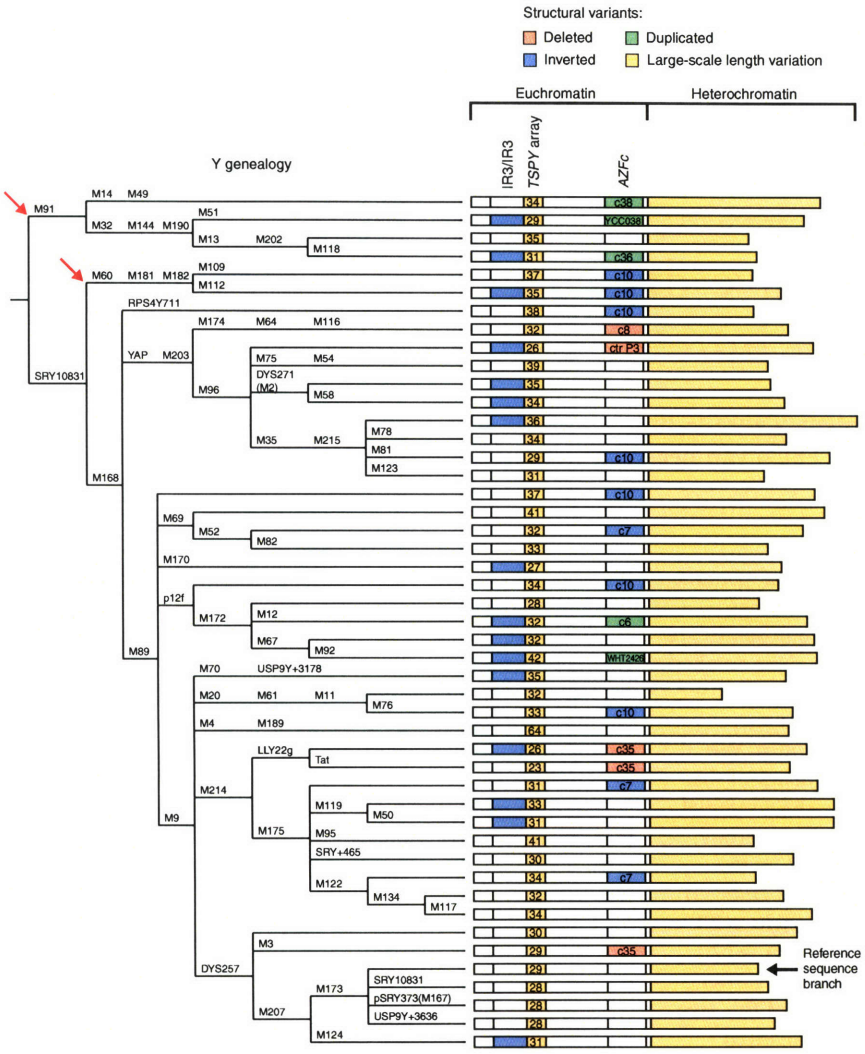


Figure 2 Y chromosome genealogical tree (left) and identified structural polymorphisms (right). Chromosomes were assigned to one of 47 branches by typing for the stable, biallelic polymorphisms indicated (for example, M91 and M60; refs. 6,7). Red arrows indicate major branches confined to Africa[6]. For each branch, the structure of the Y chromosome sampled is schematized, including, at far right, the length of distal-Yq heterochromatin. Within the euchromatin, the presence of a particular structural variant is indicated by a color-coded rectangle. Codes denoting specific *AZFc* architectures are explained in **Figure 4**, **Supplementary Table 2** and **Supplementary Figures 2–7**. See **Supplementary Figures 3** and **4** for the 'ctr P3' deletion and for 'YCC038', which contains a small deletion, but in which duplication predominates. The reference Y chromosome belongs to the indicated branch (**Supplementary Methods**), but, as no corresponding cell line exists, its heterochromatin and *TSPY* array lengths could not be determined. **Supplementary Figure 1** provides sample identifiers and Y-haplotype designations[6,7].

showed only a handful of large-scale structural rearrangements in nonampliconic portions of the human Y in the ~6.5 million years since humans and chimpanzees diverged[8,9]. Moreover, none of these differences is polymorphic among extant human Y chromosomes (data reported in **Supplementary Methods**). At the same time, available data suggest that there is little large-scale structural similarity between the ampliconic regions of the human and chimpanzee Y chromosomes, with conserved ampliconic structures confined to palindromes P6, P7 and P8 and the centers of palindromes P1 and P2 (refs. 9,10; **Fig. 1b**). Thus, among human Y chromosomes, structural polymorphisms would most likely involve ampliconic regions.

We designed assays that would be maximally informative for each of the nine categories of structural variation, which include inversions and subtle changes in copy number (**Supplementary Methods**). For example, pulsed-field DNA blots can detect subtle differences in the length of the *TSPY* array[11], metaphase FISH can detect several different kinds of potential pericentric inversions[12,13] and multicolor interphase FISH can detect other large inversions and distinguish between alternative large-scale organizations of the *AZFc* region[14,15].

Of the nine categories of potential structural variation, our search detected four, which occurred in four regions of the chromosome. Two of the regions, the distal-Yq heterochromatin and the *TSPY* array, showed large-scale length variation. The distal-Yq heterochromatin is composed of low-complexity sequences organized in tandem arrays[16]. It ranged in length from 29% to 54% of the metaphase Y chromosome, with a median of 44% (**Figs. 1**, **2**, **3a** and **5a**). The *TSPY* array is composed of highly similar 20.4-kb repeat units, each containing a copy of the *TSPY* gene and of the *CYorf16* transcription unit[11,16,17]. The *TSPY* array ranged in size from 23 to 64 units (0.47 to 1.3 Mb), with a median of 32 units (0.65 Mb; **Figs. 1**, **2**, **3b** and **5b**).

The third region, in proximal Yp, was inverted in 16 chromosomes (**Figs. 1**, **2**, **3c–f**, **5d**)[18,19]. We localized the boundaries of this 3.6-Mb inversion to within 100 kb of the IR3 repeats, strongly supporting the model that the inversions originated via ectopic homologous recombination between the IR3 repeats (**Fig. 3c–f**)[16,20].

The fourth region, *AZFc*, demonstrated abundant architectural polymorphism (**Figs. 1**, **2**, **4** and **5d**). Because this region is composed almost entirely of large, nearly identical, repeated amplicons[21], there are myriad possibilities for rearrangement via ectopic homologous recombination. We predicted *AZFc* architectures that could result from homologous recombination between amplicons ≥100 kb in length and designed combinations of two-color FISH and plus/minus STSs to detect these potential architectures (**Supplementary Table 2** and **Supplementary Fig. 2** online). These assays showed that 20 of the 47 chromosomes had variant *AZFc* architectures, the largest of which involved a duplication of ~3.5 Mb (**Figs. 2**, **4**, **5d** and **Supplementary Figs. 3–7** online).

For each of the four regions showing structural polymorphism, we determined the minimum number of independent mutation events needed to produce the distribution of variants across the genealogical tree; that is, a minimum-mutation history. The many distinct lengths that we observed in the distal-Yq heterochromatin and *TSPY* array must have been the result of multiple mutations (**Figs. 2** and **5a**,**b**). For a more complete analysis, we calculated minimum-mutation histories using methods that accommodate experimental variance in

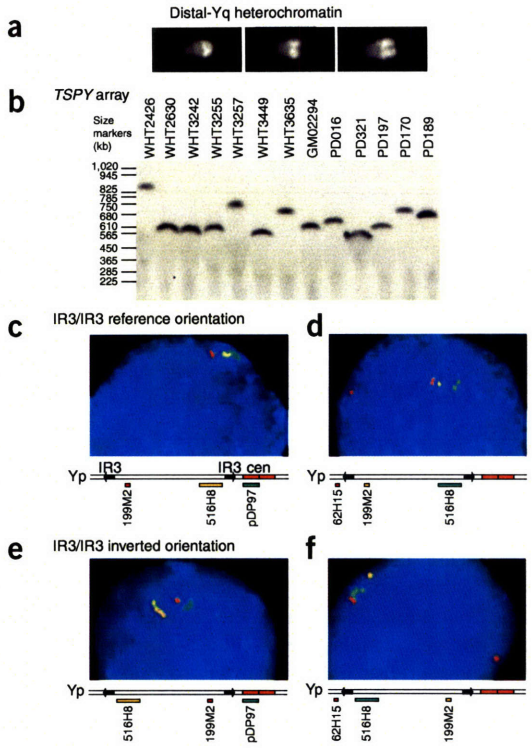


Figure 3 Assaying variation in heterochromatin length, *TSPY* array length and IR3/IR3 orientation. (**a**) Quinacrine staining of metaphase Y chromosomes with distal-Yq heterochromatin that is short (sample 4566), average (PD178) or long (PD123). (**b**) *Pme*I pulsed-field DNA blot to assay the number of *TSPY* repeats. (**c**,**d**) Three-color FISH of interphase nuclei with the reference orientation of the IR3/IR3 region (sample WHT3242). Below each nucleus is a schematic diagram of proximal Yp with IR3 repeats indicated; regions detected by each probe (199M2, 516H8, pDP97 and 62H15) are indicated in the color of the probe's stain. (**e**,**f**) Interphase nuclei with IR3/IR3 inversion (sample WHT3257). Results shown in **c** and **e** map the proximal inversion breakpoint between 516H8 and pDP97 in the reference orientation and between 199M2 and pDP97 in the inverted orientation. Results shown in **d** and **f** map the distal inversion breakpoint between 62H15 and 199M2 in the reference orientation and between 62H15 and 516H8 in the inverted orientation. 62H15 cross-hybridizes to the X chromosome, generating additional red dots at nuclear margins in **d** and **f**.

allele length (**Supplementary Methods**)[22]. For the *TSPY* array, this yielded an estimate of ≥ 23 changes in length, and for the distal-Yq heterochromatin, an estimate of ≥ 12 large changes in length. For the *TSPY* array, these changes were probably due to unequal crossing-over. Likewise, for the distal-Yq heterochromatin, large changes were probably attributable to unequal crossing-over, although small changes probably also occurred via mutational mechanisms that operate on micro- or minisatellites. For IR3/IR3, there were ≥ 12 independent inversion events, and for *AZFc*, there were ≥ 20 rearrangement events (**Supplementary Methods**). We also noted that in minimum-mutation histories of *AZFc*, inversion events were overrepresented compared with a null model of equally probable inversions, deletions and duplications ($P < 0.038$; **Supplementary Methods**). A predominance of inversion events could be caused by (i) more frequent inversion events than deletion or duplication events or (ii) natural selection against deletions and duplications but not inversions.

Having estimated lower bounds on the numbers of mutations causing structural variation, we proceeded to investigate their rates. For this, we needed to determine the denominator of the rate: that is, the total time represented by all branches in the genealogy. To estimate this, we used the total number of SNPs in the tree, the average number of SNPs on paths from the root to the leaves and the time to the last common ancestor of extant human Y chromosomes (**Supplementary Methods**). Use of previously reported SNPs in this estimate might have led to bias, if some parts of the tree were more intensively screened for SNPs than others. Therefore, we resequenced ~80 kb in the 47 chromosomes, thereby identifying 94 SNPs in an unbiased way. Using these SNPs, we estimated the total time represented by the

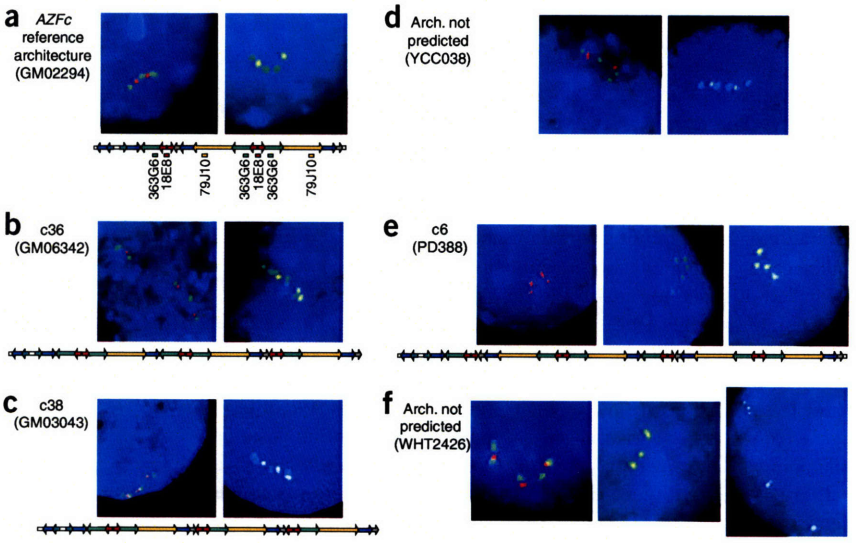


Figure 4 Detecting architectural variation in *AZFc*. Sample identifiers shown in parentheses. (**a**) Two-color FISH of interphase nuclei with *AZFc* reference architecture. Below the nuclei, *AZFc* reference architecture is depicted as a sequence of color-coded arrows representing amplicons[21]. Probes and sites of hybridization are shown; probe colors match those of detected amplicons. Left: FISH with green and red probes. Right: FISH with green and yellow probes. (**b**–**d**) Interphase nuclei with variant *AZFc* architectures probed as in **a**. Inferred amplicon organizations are shown below pairs of nuclei. (**e**) Interphase nuclei probed with 18E8 (red, left), 363G6 (green, center) and 79J10 (yellow, right), indicating four pairs of red amplicons, six green amplicons and four yellow amplicons. Of *AZFc* architectures with these counts, only c6 is likely to be generated from the reference by one recombination event, although others can be generated by two successive events (**Supplementary Figs. 2** and **5** and **Supplementary Table 2**). (**f**) Interphase nuclei from sample WHT2426. Left: probed as in left panel of **a**; center: probe 79J10; right: probe 366C6, which hybridizes to the gray amplicon in *AZFc* (see **a**) and to chromosome 1 (**Supplementary Methods**). Three closely spaced dots at the upper left arise from *AZFc* and indicate three gray amplicon copies. The two strong signals in the lower half arise from chromosome 1. No predicted *AZFc* architecture would yield this pattern of FISH results (**Supplementary Table 2**).

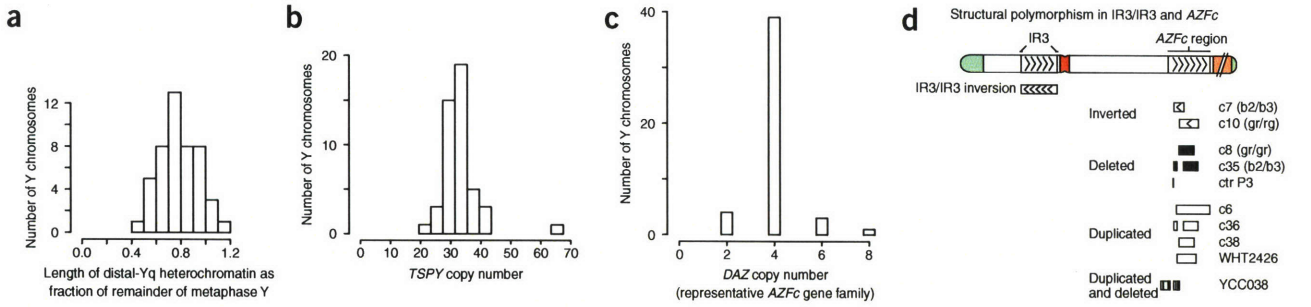


Figure 5 Summary of identified Y chromosome structural variation. (**a**–**c**) Distributions of (**a**) heterochromatin length, (**b**) number of repeat units in the *TSPY* array[16] and (**c**) *DAZ* gene copy number. See **Supplementary Figure 8** for *CDY1* and *BPY2* gene copy numbers. (**d**) IR3/IR3 inversion in Yp and variant architectures involving *AZFc*. The sizes and locations of the duplications and deletions in YCC038 and WHT2426 are estimated (**Fig. 4d,f, Supplementary Table 1, Supplementary Fig. 4** and **Supplementary Methods**).

genealogical tree at 1.3 million years, which conservatively corresponds to 52,000 generations (**Supplementary Methods**).

Using this denominator and mutation counts from the minimum-mutation histories, we inferred lower bounds on mutation rates. These are lower bounds because such histories never involve reversion or recurrence events unless essential to explain the distribution of variants. For the distal-Yq heterochromatin, ≥ 12 large changes in length over 52,000 generations correspond to a rate $\geq 2.3 \times 10^{-4}$ large-scale mutations per father-to-son transmission of a Y chromosome. For the *TSPY* array, ≥ 23 changes in length correspond to a rate $\geq 4.4 \times 10^{-4}$. For *AZFc*, ≥ 20 rearrangement events in the tree correspond to a rate $\geq 3.8 \times 10^{-4}$, a lower bound broadly consistent with the independent estimate of 2.5×10^{-4} for one particular *AZFc* mutation, the b2/b4 deletion[21]. For IR3/IR3, the minimum-mutation count of ≥ 12 inversion events corresponds to a rate of $\geq 2.3 \times 10^{-4}$. In addition, it was possible to obtain a maximum-likelihood estimate of the rate of IR3/IR3 inversion events (**Supplementary Methods**). These events seem to have resulted from a single mutational mechanism and are likely to have occurred at the same rate regardless of the orientation of the IR3/IR3 region. Thus, the analysis needed to examine only a single parameter, the rate of inversions, which showed maximum likelihood at 9.2×10^{-4} inversion events per father-to-son transmission of a Y chromosome.

How do the rates of large-scale structural mutation estimated here compare with rates of other kinds of mutations in the human genome? The rates we observed are at the low end of the range of rates among mini- and microsatellites but are ~10,000 times the average rate of single-nucleotide substitutions (**Supplementary Methods**). Considering that structural mutations of the Y chromosome often affect hundreds or thousands of kilobases and sometimes alter gene copy number, these mutations may be a major source of Y-linked phenotypic variation in human populations.

Despite the prevalence of Y chromosomes with large-scale differences from the reference sequence (**Fig. 2**), multicopy gene families showed limited variation in copy number, with pronounced modes and few excursions to low or high numbers of copies. We observed gene copy number variation only in the *TSPY* array and in *AZFc* and flanking areas. In *AZFc*, a predominance of inversions resulted in few chromosomes with gene copy numbers that differed from the reference sequence (**Figs. 2, 5c** and **Supplementary Fig. 8**). Furthermore, the *TSPY* genes, whose tandem array has undergone frequent changes in length, also showed limited variation in copy number. Indeed, the coefficient of variation of *TSPY* copy number (18.6%; s.d. as a percent of the mean) was less than that of *AZFc* gene families (*DAZ*, 24.2%; *BPY*, 27.4%; *CDY1*, 24.2%; **Fig. 5b,c** and **Supplementary Fig. 8**). Is limited variation in gene copy number consistent with the high mutation rates underlying widespread structural diversity among human Y chromosomes? As previously reported, natural selection has acted against one *AZFc* variant, the gr/gr deletion, in which several testis-specific gene families have reduced copy numbers and which confers increased risk of spermatogenic failure[14,23–27]. Thus, it is possible that natural selection had a wider role in constraining variation among human Y chromosomes by removing extremely high– or low–copy number variants from the population.

METHODS

Human samples. All assays were performed on human lymphoblastoid cell lines, cultured human fibroblasts or DNAs extracted from them. Most of these samples were obtained from the National Human Genome Research Institute/National Institute of General Medical Sciences DNA Polymorphism Discovery Resource (Coriell Cell Repositories)[28] or other public collections. To maximize coverage of the Y chromosome genealogical tree, we also studied several human cell lines from our own collections (**Supplementary Fig. 1** and **Supplementary Table 1**). **Supplementary Methods** lists availability of cell lines representing the structural variants described here.

Length of distal-Yq heterochromatin. For 46 of 47 men tested, we used quinacrine staining to measure heterochromatin length as a fraction of the total length of the metaphase Y chromosome, as previously described[29] (≥ 25 nuclei per sample, except for WHT3870 (12 nuclei)). We assessed the reproducibility of these measurements as discussed in **Supplementary Methods**. In one sample, WHT3299, the Y chromosome contained so little distal-Yq heterochromatin that it could not be measured using quinacrine staining. Instead, we used metaphase FISH (**Supplementary Methods**). The very short heterochromatin in individual WHT3299 was inherited by his son and thus was not an artifact of cell culture.

Length of *TSPY* array. We used *Pme*I pulsed-field DNA blotting to measure the length of the *TSPY* array (**Fig. 3b**). Gels were electrophoresed for 25 h at 14 °C, 6 V cm^{-1} (200 V), with a 60- to 120-s switch-time ramp. The probe was the PCR product of STS sY1256. We estimated the number of *TSPY* repeats by subtracting the lengths of non–*TSPY*-repeat flanks (10.9 kb) at the ends of the *Pme*I fragment, dividing by 20.37 kb (the size of the repeat unit[16]) and rounding. In all chromosomes, we confirmed, by sequencing, the presence of the *Pme*I site proximal to the *TSPY* array (**Supplementary Methods**). We did not sequence the *Pme*I site distal to the array, but loss of that site would increase the size of the *Pme*I fragment by only 16 kb. To further verify our findings, we assayed all samples on DNA blots based on a second restriction enzyme, *Xba*I (ref. 11) and obtained size estimates consistent with the *Pme*I-based sizes (**Supplementary Methods**).

Detecting IR3/IR3 orientation and *AZFc* architectures. One-, two- or three-color FISH was performed as described[30]. For each sample and set of probes,

≥200 nuclei were scored. Apart from plasmid pDP97 and cosmid 18E8, all clones used as probes were BACs derived from the RPCI-11 library (prefix 'RP11-'); refs. 16,20 provide map positions. The computer program that enumerated potential *AZFc* architectures is available on request (S.R.).

Rates of mutations giving rise to structural polymorphism. We determined minimum-mutation histories of the structural variants studied either manually (IR3/IR3 orientation, *AZFc* architecture) or using our implementation of Sankoff's algorithm[22] (distal-Yq heterochromatin, *TSPY* array, IR3/IR3 orientation); code is available on request (S.R.). We sequenced 237 PCR products in the 47 chromosomes to ascertain in an unbiased way the SNPs used to determine the total length of time represented by the tree. All SNPs detected, as well as the genotypes of the 47 chromosomes at these SNPs, have been submitted to dbSNP. See **Supplementary Methods** for details of the maximum likelihood analysis of the rate of IR3/IR3 inversion events.

Accession codes. GenBank: PCR product of STS sY1256, G75613; PCR products resequenced in SNP discovery, BV678971–BV679207.

Note: Supplementary information is available on the Nature Genetics website.

ACKNOWLEDGMENTS

We thank G. Farino for DNA sequencing; V. Frazzoni and G. Rogers for tissue culture; N.A. Ellis, M.F. Hammer, T. Jenkins, R.D. Oates and S. Silber for cell lines and blood samples; J. de Vries, N. Leschot and P. Underhill for technical and scientific advice; J.E. Alfoldi, A.E. Baltus, D.W. Bellott, A. Chakravarti, M.J. Daly, J.F. Hughes, L. Kruglyak, Y.-H. Lim, J.L. Mueller and D.E. Reich for comments on the manuscript and A.G. Clark for advice and guidance on studies of mutation rates. This work was supported by the US National Institutes of Health, the Howard Hughes Medical Institute, the Netherlands Organization for Scientific Research and the Academic Medical Center.

COMPETING INTERESTS STATEMENT

The authors declare that they have no competing financial interests.

Published online at http://www.nature.com/naturegenetics
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/

1. Iafrate, A.J. *et al.* Detection of large-scale variation in the human genome. *Nat. Genet.* **36**, 949–951 (2004).
2. Sebat, J. *et al.* Large-scale copy number polymorphism in the human genome. *Science* **305**, 525–528 (2004).
3. Sharp, A.J. *et al.* Segmental duplications and copy-number variation in the human genome. *Am. J. Hum. Genet.* **77**, 78–88 (2005).
4. Stefansson, H. *et al.* A common inversion under selection in Europeans. *Nat. Genet.* **37**, 129–137 (2005).
5. Tuzun, E. *et al.* Fine-scale structural variation of the human genome. *Nat. Genet.* **37**, 727–732 (2005).
6. Underhill, P.A. *et al.* Y chromosome sequence variation and the history of human populations. *Nat. Genet.* **26**, 358–361 (2000).
7. The Y Chromosome Consortium. A nomenclature system for the tree of human Y-chromosomal binary haplogroups. *Genome Res.* **12**, 339–348 (2002).
8. Vignaud, P. *et al.* Geology and palaeontology of the Upper Miocene Toros-Menalla hominid locality, Chad. *Nature* **418**, 152–155 (2002).
9. Hughes, J.F. *et al.* Conservation of Y-linked genes during human evolution revealed by comparative sequencing in chimpanzee. *Nature* **437**, 100–103 (2005).
10. Rozen, S. *et al.* Abundant gene conversion between arms of palindromes in human and ape Y chromosomes. *Nature* **423**, 873–876 (2003).
11. Tyler-Smith, C., Taylor, L. & Muller, U. Structure of a hypervariable tandemly repeated DNA sequence on the short arm of the human Y chromosome. *J. Mol. Biol.* **203**, 837–848 (1988).
12. Grace, H.J., Ally, F.E. & Paruk, M.A. 46,Xinv(Yp+q-) in four generations of an Indian family. *J. Med. Genet.* **9**, 293–297 (1972).
13. Bernstein, R., Wadee, A., Rosendorff, J., Wessels, A. & Jenkins, T. Inverted Y chromosome polymorphism in the Gujerati Muslim Indian population of South Africa. *Hum. Genet.* **74**, 223–229 (1986).
14. Repping, S. *et al.* Polymorphism for a 1.6-Mb deletion of the human Y chromosome persists through balance between recurrent mutation and haploid selection. *Nat. Genet.* **35**, 247–251 (2003).
15. Repping, S. *et al.* A family of human Y chromosomes has dispersed throughout northern Eurasia despite a 1.8-Mb deletion in the azoospermia factor c region. *Genomics* **83**, 1046–1052 (2004).
16. Skaletsky, H. *et al.* The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature* **423**, 825–837 (2003).
17. Manz, E., Schnieders, F., Muller Brechlin, A. & Schmidtke, J. *TSPY*-related sequences represent a microheterogeneous gene family organized as constitutive elements in *DYZ5* tandem repeat units on the human Y chromosome. *Genomics* **17**, 726–731 (1993).
18. Affara, N.A. *et al.* Variable transfer of Y-specific sequences in XX males. *Nucleic Acids Res.* **14**, 5375–5387 (1986).
19. Page, D.C. Sex reversal: deletion mapping the male-determining function of the human Y chromosome. *Cold Spring Harb. Symp. Quant. Biol.* **51**, 229–235 (1986).
20. Tilford, C. *et al.* A physical map of the human Y chromosome. *Nature* **409**, 943–945 (2001).
21. Kuroda-Kawaguchi, T. *et al.* The *AZFc* region of the Y chromosome features massive palindromes and uniform recurrent deletions in infertile men. *Nat. Genet.* **29**, 279–286 (2001).
22. Sankoff, D. Minimal mutation trees of sequences. *SIAM J. Appl. Math.* **28**, 35–42 (1975).
23. Machev, N. *et al.* Sequence family variant loss from the AZFc interval of the human Y chromosome, but not gene copy loss, is strongly associated with male infertility. *J. Med. Genet.* **41**, 814–825 (2004).
24. de Llanos, M., Ballesca, J.L., Gazquez, C., Margarit, E. & Oliva, R. High frequency of gr/gr chromosome Y deletions in consecutive oligospermic ICSI candidates. *Hum. Reprod.* **20**, 216–220 (2005).
25. Ferlin, A. *et al.* Association of partial *AZFc* region deletions with spermatogenic impairment and male infertility. *J. Med. Genet.* **42**, 209–213 (2005).
26. Hucklenbroich, K. *et al.* Partial deletions in the *AZFc* region of the Y chromosome occur in men with impaired as well as normal spermatogenesis. *Hum. Reprod.* **20**, 191–197 (2005).
27. Lynch, M. *et al.* The Y chromosome gr/gr subdeletion is associated with male infertility. *Mol. Hum. Reprod.* **11**, 507–512 (2005).
28. Collins, F.S., Brooks, L.D. & Chakravarti, A.A. DNA polymorphism discovery resource for research on human genetic variation. *Genome Res.* **8**, 1229–1231 (1998).
29. Schnedl, W. Flurescenzuntersuchungen ueber die langenvariabilitaet des Y-chromosoms beim menschen. *Humangenetik* **12**, 188–194 (1971).
30. Saxena, R. *et al.* Four *DAZ* genes in two clusters found in the *AZFc* region of the human Y chromosome. *Genomics* **67**, 256–267 (2000).