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

Julian Lange[1,2], Helen Skaletsky[1,2], George W. Bell[1] and David C. Page[1,2,*]

[1]Whitehead Institute, 9 Cambridge Center, Cambridge, MA 02142, USA and [2]Howard Hughes Medical Institute and Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02142, USA

ABSTRACT

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 <14 kb 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 male-specific region (MSY) (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 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

*To whom correspondence should be addressed. Tel: +1 617 258 5203; Fax: +1 617 258 5578; Email: dcpage@wi.mit.edu

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 analysing 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: (i) 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 (ii) 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 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 MSYs 23 Mb of euchromatic sequence. This corresponds to an average spacing of <14 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; 70–0.5°C/cycle for 40 s (i.e. a ramp from 70 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 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 (Figure 1) that includes nucleotide sequence coordinates, landmark STSs, protein-coding genes and major structural features of the MSY.

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

Sequence type	Number of PCR assays (STSs)			Number of loci amplified			
	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 MSYs 23 Mb of euchromatic sequence.
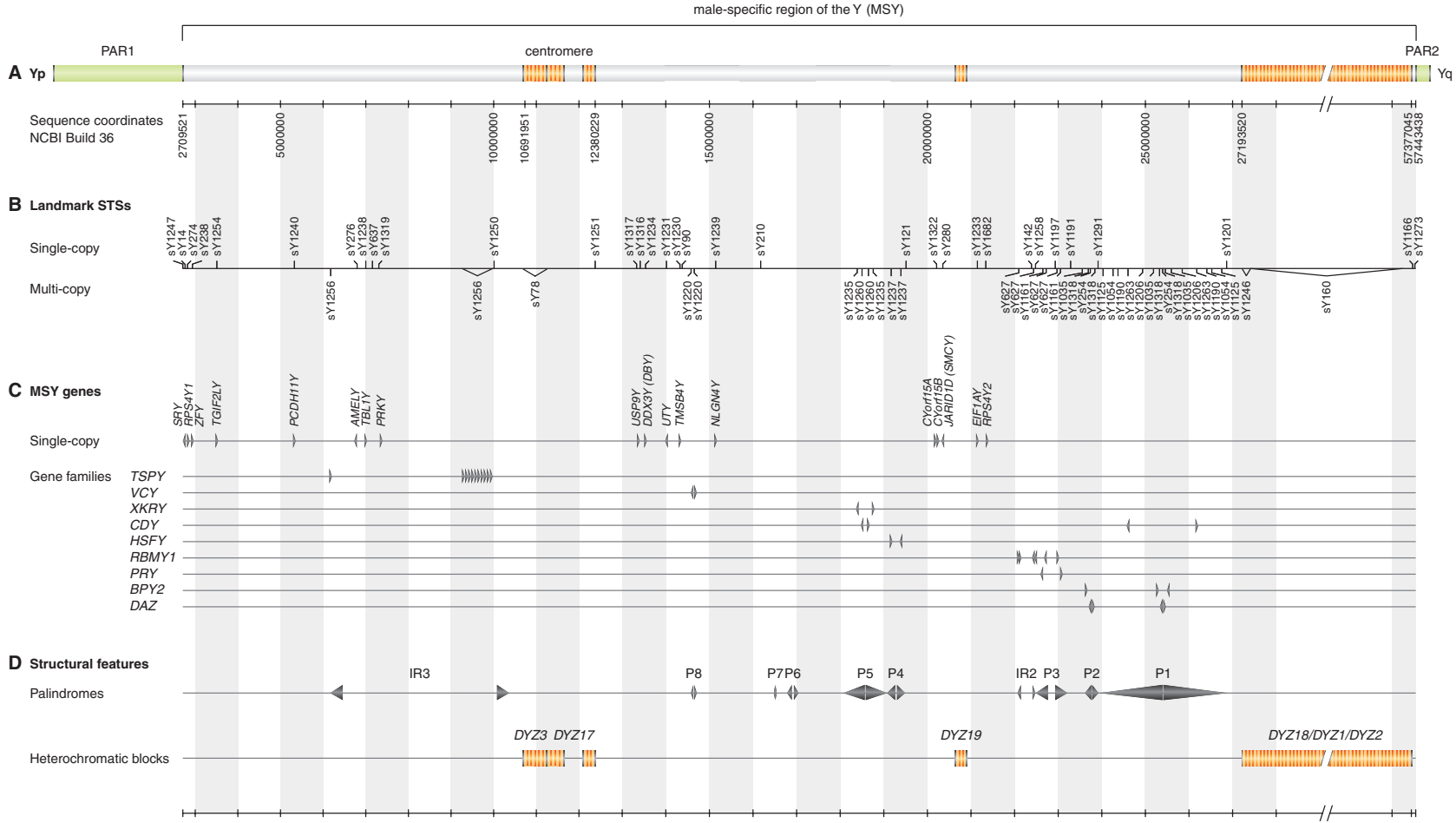
Figure 1. The human MSY. (**A**) Schematic of the entire Y chromosome, with pseudoautosomal regions (PAR1 and PAR2) in green, euchromatic regions of MSY in gray and heterochromatic regions of MSY in orange. Immediately below are Y chromosome sequence coordinates based on NCBI Build 36. (**B**) Locations of 51 landmark STSs. (**C**) MSY protein-coding genes and gene families; arrowheads indicate 5′-to-3′ orientation. (**D**) Major structural features.

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 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.

ACKNOWLEDGEMENTS

Conflict of interest statement. None declared.

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.

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 14 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.

REFERENCES

1. Skaletsky,H., Kuroda-Kawaguchi,T., Minx,P.J., Cordum,H.S., Hillier,L., Brown,L.G., Repping,S., Pyntikova,T., Ali,J. *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,P.A. and Ross,A. (1966) Structural abnormalities of the Y chromosome in man. *Nature*, **210**, 352–354.
3. Tiepolo,L. and Zuffardi,O. (1976) Localization of factors controlling spermatogenesis in the nonfluorescent portion of the human Y chromosome long arm. *Hum. Genet.*, **34**, 119–124.
4. Guellaen,G., Casanova,M., Bishop,C., Geldwerth,D., Andre,G., Fellous,M. and Weissenbach,J. (1984) Human XX males with Y single-copy DNA fragments. *Nature*, **307**, 172–173.
5. Page,D.C., de la Chapelle,A. and Weissenbach,J. (1985) Chromosome Y-specific DNA in related human XX males. *Nature*, **315**, 224–226.
6. Vergnaud,G., Page,D.C., Simmler,M.C., Brown,L., Rouyer,F., Noel,B., Botstein,D., de la Chapelle,A. and Weissenbach,J. (1986) A deletion map of the human Y chromosome based on DNA hybridization. Am. *J. Hum. Genet.*, **38**, 109–124.
7. Affara,N.A., Florentin,L., Morrison,N., Kwok,K., Mitchell,M., Cook,A., Jamieson,D., Glasgow,L., Meredith,L. *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,N.A., Ferguson-Smith,M.A., Tolmie,J., Kwok,K., Mitchell,M., Jamieson,D., Cooke,A. and Florentin,L. (1986) Variable transfer of Y-specific sequences in XX males. *Nucleic Acids Res.*, **14**, 5375–5387.

9. Andersson,M., Page,D.C. and 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,C.M., Casanova,M., Saal,H., Friedman,C., Sybert,V., Graham,J., Thuline,H., Page,D.C. and 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. and 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,D.C., Brown,L.G. and 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. and 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,L.G., Beer-Romero,P., Bogan,J.S. and Page,D.C. (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. and 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,B.T., Ma,N., Breg,W.R., Stratton,R., Surti,U. and Page,D.C. (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,L.Y. (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,T.Y., Salo,P., Alagappan,R., Brown,L.G., Rosenberg,M., Rozen,S., Jaffe,T., Straus,D. *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,D.C. and Disteche,C.M. (1995) Gonadoblastoma: molecular definition of the susceptibility region on the Y chromosome. *Am. J. Hum. Genet.*, **57**, 1400–1407.

20. Vogt,P.H., Edelmann,A., Kirsch,S., Henegariu,O., Hirschmann,P., Kiesewetter,F., Kohn,F.M., Schill,W.B., Farah,S. *et al.* (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,D.C., Weil,D., Petit,C. and Rappold,G.A. (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. and Page,D.C. (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. and Vogt,P.H. (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,C.A., Jobling,M.A., Affara,N. and Hurles,M.E. (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,L.G., Minx,P.J., Cordum,H.S., Waterston,R.H., Wilson,R.K., Silber,S., Oates,R. *et al.* (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,R.D., Page,D.C. and 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,S.K., Korver,C.M., Pyntikova,T., Kuroda-Kawaguchi,T., de Vries,J.W., Oates,R.D. *et al.* (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,L.H., Floridia,G., Tyler-Smith,C. and Vogt,P.H. (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,S.K., Korver,C.M., Brown,L.G., Marszalek,J.D., Gianotten,J., Oates,R.D., Silber,S., van der Veen,F. *et al.* (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,M.A., Lo,I.C., Turner,D.J., Bowden,G.R., Lee,A.C., Xue,Y., Carvalho-Silva,D., Hurles,M.E., Adams,S.M. *et al.* (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,A.H., Berta,P., Palmer,M.S., Hawkins,J.R., Griffiths,B.L., Smith,M.J., Foster,J.W., Frischauf,A.M., Lovell-Badge,R. *et al.* (1990) A gene from the human sex-determining region encodes a protein with homology to a conserved DNA-binding motif. *Nature*, **346**, 240–244.

32. Page,D.C. (1987) Hypothesis: a Y-chromosomal gene causes gonadoblastoma in dysgenetic gonads. *Development (Cambridge, U.K.)*, **101**(Suppl.), 151–155.

33. Nathanson,K.L., Kanetsky,P.A., Hawes,R., Vaughn,D.J., Letrero,R., Tucker,K., Friedlander,M., Phillips,K.A., Hogg,D. *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. and Page,D.C. (1992) The human Y chromosome: overlapping DNA clones spanning the euchromatic region. *Science*, **258**, 60–66.

35. Tilford,C.A., Kuroda-Kawaguchi,T., Skaletsky,H., Rozen,S., Brown,L.G., Rosenberg,M., McPherson,J.D., Wylie,K., Sekhon,M. *et al.* (2001) A physical map of the human Y chromosome. *Nature*, **409**, 943–945.

36. Repping,S., van Daalen,S.K., Brown,L.G., Korver,C.M., Lange,J., Marszalek,J.D., Pyntikova,T., van der Veen,F., Skaletsky,H. *et al.* (2006) High mutation rates have driven extensive structural polymorphism among human Y chromosomes. *Nat. Genet.*, **38**, 463–467.

37. Rozen,S. and 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,T.L., Waterston,R.H., Clark,R.A., Eichler,E.E., Munch,C., Schempp,W. and Rappold,G. (2005) Interchromosomal segmental duplications of the pericentromeric region on the human Y chromosome. *Genome Res.*, **15**, 195–204.