Copy number variation on the human Y chromosome

Mark A. Jobling

Department of Genetics, University of Leicester, University Road, Leicester LE1 7RH, UK

Address for correspondence and reprints: Prof Mark A. Jobling, Department of Genetics, University of Leicester, University Road, Leicester LE1 7RH, UK
Tel.: +44 (0)116 252 3427. Fax: +44 (0)116 252 3378. Email: maj4@leicester.ac.uk

Keywords: copy number variation; Y chromosome; phylogeny; AZF; STR

Abstract

The Y chromosome is unusual in being constitutively haploid and escaping recombination for most of its length. This has led to a correspondingly unusual genomic landscape, rich in segmental duplications, which provide a rich environment for the generation of copy number variation (CNV). Interest in the chromosome comes from diverse fields, including infertility research, population genetics, forensics and genealogy. Together with inclusion in more systematic surveys, this has led to the ascertainment of a variety of CNVs. Assessment in the context of the well-resolved Y phylogeny allows their mutational history to be deciphered, and estimation of mutation rate. The functional consequences of variants are moderated by the specialisation of the chromosome and the presence of functionally equivalent X-chromosomal homologues for some genes. However, deletions of the *AZFa*, *b*, and *c* regions cause impaired spermatogenesis, while partial deletions and duplications within these regions, and deletions and duplications elsewhere, may be selectively neutral or have subtle phenotypes.

Introduction

The Y chromosome is, both literally and metaphorically, the most singular of human chromosomes. Specialised in sex determination, inessential for viability, and (for most of its length) constitutively haploid, it has evolved a bizarre genomic landscape and a correspondingly unusual repertoire of copy number variants (CNVs). In marked contrast to the functional specialisation of the Y chromosome itself is the eclectic nature of the research fields that encompass the study of Y diversity - including molecular reproductive genetics, genome dynamics, population genetics and forensics – and this has led to the ascertainment of CNVs in many different ways. Compared to many autosomal regions, ascertainment is therefore high, although, as explained below, probably biased.

Because of its haploidy, the Y chromosome is unconstrained by pairing along its entire length, and this has led to the accumulation of a high proportion of segmental duplications (~35% (Skaletsky et al., 2003), compared to a genome average of only ~5% (Bailey et al., 2002)) that can act as substrates for the generation of CNVs through non-allelic homologous recombination (NAHR). Many Y-CNVs are certainly driven by NAHR, and show the recurrence expected for such variants. However, the high ascertainment of variation in some regions means that some of the variants detected represent sporadic and non-recurrent rearrangements with very low mutation rates.

The Y chromosome's unusual evolutionary history provides some unique opportunities to study the rates and processes of CNV generation, but at the same time its sequence organisation poses some particular practical challenges. This review will describe the genomic and phylogenetic context for Y-chromosomal CNV (Y-CNV) studies, and summarise what is currently known about Y-CNVs, how they have been discovered, and what their functional significance might be. The pseudoautosomal regions, in which recombination with the X chromosome occurs, are not considered here, and the focus is on the non-recombining region of the Y chromosome (NRY), also known as the male-specific region of the Y (MSY).

The Y phylogeny as a framework for interpreting CNVs

The power to detect recurrence or otherwise of variants comes from the fact that all sequences on a given Y chromosome share a single evolutionary history due to the lack of recombination (Jobling and Tyler-Smith, 2003). We can therefore employ a phylogenetic approach to understanding mutation history and the dynamics of CNV formation.

The Y phylogeny is based on binary markers (mostly single nucleotide polymorphisms [SNPs]) that have low mutation rates, and therefore can be regarded

largely as unique events in human history. Haplotypes constructed using such markers are known as haplogroups, and can be arranged in a robust tree that provides a framework for considering CNVs and other variants, and asking questions about recurrence and rate. Figure 1a shows the major branches of the most recently published version of the Y phylogeny; in detail, the full tree is constructed from 599 mutations defining 311 distinct haplogroups (Karafet et al., 2008), but its resolution will increase enormously as new and reliable Y-chromosomal sequences become available.

The finding of a seemingly identical CNV in all members of a haplogroup (a monophyletic set) implies that the sharing is by descent (Figure 1b). Of course, if the CNV has a high mutation rate it might be expected to revert so that some haplogroup members no longer carry the variant, but we can nonetheless reconstruct the CNV status of the common ancestor assuming the minimum number of mutational events (a maximum parsimony approach). Conversely, the finding of such a CNV distributed among different branches of the phylogeny, where the likely ancestral state is absence of the CNV, indicates multiple recurrence.

This approach allows the minimum number of independent events to be counted for a given Y-CNV. The rate of CNV generation could then be deduced if the number of generations encompassed by the sampled chromosomes were known, but estimation of this parameter is not trivial. Estimates of the time to most recent common ancestor (TMRCA) for the major branches of the phylogeny are available, based on coalescent analysis (Hammer and Zegura, 2002) or other methods (Karafet et al., 2008), but rely upon markers that are subject to ascertainment bias. A less biased (though more labour intensive) approach is to resequence segments in all sampled chromosomes (Repping et al., 2006) and estimate the time encompassed in the tree resulting from newly ascertained SNPs by comparison with the chimpanzee orthologous sequence, assuming a date for the human-chimpanzee divergence.

Surveys of the population distributions of Y-chromosomal haplogroups have shown that they are highly geographically differentiated (Underhill and Kivisild, 2007), with particular populations carrying their own characteristic sets of lineages. This is due to the powerful influence of genetic drift on the Y, driven by its low effective population size, and the high reproductive variance of males (Jobling and Tyler-Smith, 2003). High geographical differentiation is important in the ascertainment of CNVs, since the inclusion of only limited sets of populations in CNV studies could mean that many variants will be missed. Furthermore, drift can elevate the frequency of a CNV-carrying haplogroup in some regions or populations (some examples are given below), so high observed frequency does not necessarily mean a high rate of the CNV-generating process.

The reference sequence, and others

The starting point for the consideration of any Y-chromosomal variant is the reference sequence. The completion of a reliable sequence of this chromosome (Skaletsky et al., 2003) was a particularly impressive achievement. Assembly was complicated because of the high levels of very similar intrachromosomal segmental duplications and XY-homologous sequences, and these problems had to be surmounted by sequencing to high coverage mainly in a library derived from the DNA of one individual man. The majority of the Y-chromosomal reference sequence thus derives from a single chromosome belonging to haplogroup R, with one 0.8-Mb segment on the long arm (the *AZFa* region) deriving from a different man (Sun et al., 1999) and belonging to haplogroup G. This contrasts with the rest of the genome, where libraries were constructed from the DNAs of many different individuals, so that the final sequence constitutes a complex and artificial mosaic.

As with all chromosomes, the Y sequence is incomplete over the repetitive centromeric region. Most of the euchromatin, however, is complete, although some other large repetitive regions remain as gaps. These include the DYZ5 repeat array on Yp (containing the *TSPY1* genes, discussed below), and the large and highly polymorphic heterochromatic region on distal Yq.

The reference sequence is powerful because it allows an unambiguous description of sequence organisation and the identification of duplicated and repeated sequences - in particular, segmentally duplicated and often extremely (>99.98%) similar sequences that have been dubbed 'ampliconic' (Skaletsky et al., 2003). This provides an invaluable framework for considering any new variant, and the possible mutational mechanisms underlying it. At the same time, however, the reference sequence is constraining, because other Y chromosomes from other lineages are likely to be different in organisation and sequence copy number, and might even contain sequences that are entirely absent from the reference. Proposed mutational mechanisms that are based on the reference might not apply.

High-throughput application of traditional capillary DNA sequencing, and the advent of new sequencing technologies (Bentley, 2006) has yielded two new Y-chromosome sequences to date, deriving from Craig Venter (Levy et al., 2007) and James D. Watson (Wheeler et al., 2008). It is unfortunate that both belong to the same branch of the Y phylogeny as the reference sequence (hgR-M269), so that limited new information is provided. The current 1000 Genomes Project (www.1000genomes.org) will include fathers of HapMap trios whose Y chromosomes belong to haplogroups E and I, and the general availability of the new technologies should allow large-scale sequencing of chromosomes from many different parts of the tree. However, in

addition to the issue of choice of haplogroup, the assemblies are likely to be unreliable, which compromises ability to diagnose CNVs, at least at the medium to large scale. Indeed, all the new sequencing technologies rely upon mapping against a reference sequence, so may have limited power to characterise novel structural organisations of Y chromosomes.

How are Y-CNVs found?

The discovery of Y-CNVs has arisen from several fields of investigation, all of which involve biases due to the regions of the chromosome surveyed, and/or the populations analysed. As discussed above, population bias will lead to haplogroup bias due to the strong population differentiation of Y variation, and hence the identification of an unrepresentative sample of CNVs.

Early polymorphism surveys in evolutionary studies revealed a number of CNVs (e.g. (Jobling, 1994; Jobling et al., 1996; Oakey and Tyler-Smith, 1990; Page et al., 1982)), some of which were of value as markers in the era before widely available SNPs and STRs. Many of these variants were properly defined only later, when the sequence became available (though some remain to be defined), and the history of this field will not be detailed here.

Molecular reproductive genetics

In infertility research, CNVs are functional candidates for effects on sperm count or other phenotypes. These studies have focused on the three regions on Yq (two of which overlap) that have been associated with infertility when they are deleted (*AZFa*, *b*, and *c* (Vogt et al., 1996)). Studies of men with reduced sperm count or specific spermatogenic abnormalities, as well as control groups, has revealed variation in copy number, particularly in the *AZFc* region, a remarkable segment of DNA composed of very large and highly similar repeat units (Kuroda-Kawaguchi et al., 2001). Specific CNVs involving the *AZF* regions will be further discussed in the next section. Detection of CNVs in infertility studies is usually via the presence or absence of sequence-tagged sites (STSs), many of which are unique in the reference sequence. Over 1200 of these have been defined (Skaletsky et al., 2003; Tilford et al., 2001; Vollrath et al., 1992), and locating relevant STSs for a particular interval has recently been facilitated by the availability of an online database, MSY Breakpoint Mapper (Lange et al., 2008) http://breakpointmapper.wi.mit.edu/. Some studies have reinforced STS-based findings by carrying out quantitative PCR or FISH-based analyses. The use of presence or absence of particular paralogous sequence variants (PSVs – known in some studies as sequence family variants) in repeated sequences to diagnose deletions (Fernandes et al., 2004b) has caused some controversy (Fernandes

et al., 2004a; Repping et al., 2004), in part because gene conversion between copies could lead to misinterpretation.

Forensic and population studies

CNVs detected in forensic and population studies can be serendipitous by-products of the practical exploitation of short tandem repeat markers (STRs), emerging from the observation of anomalous STR haplotypes (Figure 2) (Balaresque et al., 2008a; Bosch and Jobling, 2003; Butler et al., 2005; King et al., 2005). When a normally single-copy STR within a haplotype is absent, and this is not due to a primer-site mutation, it indicates a deletion variant. Conversely, when it gives two peaks in an electropherogram it can signal a duplication variant, though this needs to be distinguished from mosaicism of STR alleles differing through somatic mutation, particularly if the DNA source is a cell-line (Balaresque et al., 2008b; Banchs et al., 1994). There is underascertainment of duplications by this method, because their detection relies upon an STR mutation having led to a detectable size difference between the duplicated alleles; when the alleles are identical in length, they will be detectable only by quantitative methods, which are not routinely employed. Some STRs are present in more than one copy in the reference sequence, and in these cases other variations in copy number can be observed. For example, the normally bilocal DYS385 can show three, or sometimes four, alleles (Butler et al., 2005). If the putative deletion or duplication involves two or more STRs, this can give an indication of the physical extent of the CNV.

Although the number of chromosomes surveyed with STRs can be large (for example, release 23 of the Y chromosome Haplotype Reference Database (Willuweit et al., 2007) ~[http://www.yhrd.org/] contains ~55,000 9-STR haplotypes), only a limited proportion of the chromosome is covered, because the number of the available >200 STRs (Kayser et al., 2004) used is usually small, and they are non-randomly distributed (Hanson and Ballantyne, 2006). The number of STRs included in forensic studies is at least 9, and a current commercial multiplex, Y-Filer (Applied Biosystems), contains 17; the number used in population studies can vary from 6 (e.g. (Bowden et al., 2007)) to as many as 61 (Xue et al., 2005). There is a notable bias in population studies against use of multi-locus STRs, which can lead to the underrepresentation of regions (such as *AZFc*) in which multi-copy sequences are the norm.

Targeted and systematic CNV surveys

Following the description of the genomic architecture of the Y, a structural survey was undertaken of 47 chromosomes belonging to different branches of the

phylogeny (Repping et al., 2006). Attention was focused on already known or suspected variants, including the 'ampliconic' regions, the *TSPY1* array length variation, a short arm paracentric inversion, and length variation of the distal Yq heterochromatin. A combination of STS content analysis, pulsed-field gel electrophoresis, metaphase and interphase FISH, and cytogenetic length measurement (for the heterochromatin) were used. A number of novel CNVs of the *AZFc* region were discovered, and the phylogenetic context allowed an estimate of rates of mutation of various CNVs and other variants. The careful approach to characterising CNVs is a strength of this study, but its targeted nature based on previously known variants and the reference sequence organisation constitutes a bias.

Inclusion of the Y chromosome in systematic genomewide CNV surveys (Locke et al., 2006; Perry et al., 2008; Redon et al., 2006; Sharp et al., 2005), plus the availability of whole Y chromosome resequencing data (Levy et al., 2007), has provided a more objective picture of Y-CNVs. As a result, the Database of Genomic Variants (Iafrate et al., 2004) (http://projects.tcag.ca/variation/; April 18 2008 update) contains 51 Y variants, of which 40 lie in the non-recombining region. These vary over a wide range of scales from a few tens of base-pairs to hundreds of kilobases, and have been detected in different samples, using different methods with differing sensitivities.

The largest scale study (Redon et al., 2006) surveyed the HapMap samples – 270 individuals with ancestry in Africa, Europe, and East Asia, carrying 104 distinct Y chromosomes. Y-CNVs were detected using a Whole Genome TilePath array of BAC clones; for other chromosomes, CNVs could be validated by comparative analysis of hybridization intensities on Affymetrix GeneChip 500K SNP arrays, but the lack of Y-SNPs on these arrays meant that this was not possible for the Y.

Limited published information is available on the distribution of the HapMap Y chromosomes in the Y phylogeny (International HapMap Consortium, 2005). The African chromosomes are of low diversity, with all lying within hg E (and 29/30 in one sublineage, E3a); all but one of the European chromosomes belong to two lineages, R1 and I; the Asian chromosomes have higher diversity, but with predominance in hgs D, NO, and a poorly defined group including hg F, H or K. Many major lineages are thus absent from the survey, including the basal haplogroups A and B.

The variation found in this study (Figure 3) largely corresponds to already known variants including *TSPY1* and the *AZFc* region, though some other variants deserve further investigation. Resolution of the study is limited by the size of the clones in the tiling array, so that only CNVs >50 kb were called efficiently, and

combining calls from individuals tends to inflate CNV size. An independent study has developed an X- and Y-chromosome-specific BAC tiling path array that allows the mapping and characterisation of XY translocations and other sex-chromosomal rearrangements, as well as Y-CNVs (Karcanias et al., 2007).

Examples of known Y CNVs

Much of the current interest in CNVs focuses on their likely functional significance in terms of effects on gene content or expression (Hurles et al., 2008). However, there are a number of reasons to suspect that functional effects of such variation on the Y chromosome might be slight, or at least difficult to interpret. The great majority of the ~0.1% of men who carry an additional Y chromosome (47,XYY) are not identified (Abramsky and Chapple, 1997), because the extra Y has relatively mild phenotypic effects, and even carrying one or two more is tolerated (Shanske et al., 1998). This suggests that extra doses of individual genes may not have serious consequences, unless the stoichiometry of Y-specific gene doses is crucial. In fact, many Y-specific genes are present in multiple copies in the reference sequence (Skaletsky et al., 2003), which implies that subtle alterations in copy number might not have noticeable effects.

However, careful follow-up of 47,XYY males and comparison to controls (Higgins et al., 2007; Ratcliffe, 1999) does show a higher incidence of reduced intelligence, delayed speech development, increased stature, and significantly elevated mortality from a number of causes, in particular epilepsy. Furthermore, absence of the Y, and presence of only one X (45,X), gives Turner syndrome, with a suite of abnormalities (Elsheikh et al., 2002) indicating that the dosage of some XY-homologous genes is important. Although some of these genes are pseudoautosomal, some lie on the NRY, and haploinsufficiency through CNV could lead to expression of some Turner syndrome features. Finally, since many genes involved in spermatogenesis have accumulated on the NRY (Skaletsky et al., 2003), there could be subtle effects on male fertility.

Interpretation of the functional effects of Y-CNVs depends on how they were identified. Clearly, those identified by workers in the field of male infertility research can be linked to appropriate data on sperm parameters and testicular histology; however, when Y-CNVs are identified in forensic or population studies, phenotypic information on the individuals carrying the variant chromosomes is not usually available, so inferences on their effects are often indirect. However, the populations surveyed frequently include examples from parts of the world where andrologists do

not normally venture, so there is the potential to identify interesting and important variants that may throw light on the complex subject of spermatogenic genes.

A particular difficulty in interpreting the effects of Y-CNVs is due to the absence of recombination. A phenotype might be associated with a particular Y chromosome type, and while this could be due to a CNV of interest, it could also be due to some other unknown Y-variant to which the CNV is permanently linked. In contrast, on an autosome, recombination allows the study of the effect a given CNV on various different chromosomal backgrounds.

AZFc deletions, duplications and complex rearrangements

Deletions in the *AZFc* region (Figure 4) are the commonest known cause of Y-linked male infertility (Vogt et al., 1996), and the availability of the reference sequence allowed the definition of the remarkable structure of the region (Kuroda-Kawaguchi et al., 2001), which spans several megabases and is made up almost entirely of large paralogous repeats ('amplicons') that are highly similar (>99.9%) in sequence due to gene conversion (Rozen et al., 2003), and contain a set of repeated testis-specific protein-coding and untranslated genes. Most *AZFc* deletions associated with spermatogenic failure are caused by NAHR between the repeats b2 and b4, removing 3.5Mb of DNA including all copies of the genes *DAZ* and *BPY2*. Emphasising the complexity of interpreting the phenotypic effects of rearrangements, however, the sperm phenotypes are heterogeneous, and *AZFc* deletions can even be found in males who have fathered offspring (Chang et al., 1999; Kühnert et al., 2004). Phenotypic complexity has caused controversy about the effects of Y-CNVs in the region (McElreavey et al., 2006), and the lack of simple genotype-phenotype correlations has led to the resurrection of an earlier idea that the effects of deletions might be mediated through meiotic disruption (Vogt et al., 2008), rather than specific gene loss.

Given the complex repetitive structure of the region, it is no surprise that many other rearrangements can and do occur – some examples are shown in Figure 4b. Indeed, in a thorough description (Repping et al., 2006) of the possible architectures, following putative recombination events, of the *AZFc* region based on the reference sequence, there are 9 possible structures following single recombination events, a further 57 following double recombination events, and a further 799 following triple events. These include inversions that can give rise to novel substrates for deletions or duplications. The known and well-characterised NAHR-mediated rearrangements are considerably fewer than this, but include examples that are apparently fixed in particular branches of the Y phylogeny. One example is the gr/gr deletion following b2/b3 inversion that is fixed in haplogroup

N (Fernandes et al., 2004b), and makes up >50% of Y chromosomes in such populations as the Finns (Rosser et al., 2000).

When ascertainment is good (Balaresque et al., 2008a; Repping et al., 2006), rare variants as well as recurrent NAHR-mediated variants are found; these often do not map to direct repeats, and (at least based on the reference sequence organisation) are likely to be mediated by non-recurrent, non-homologous processes. Significant over-representation of independent CNVs involving DYS448 deletion in two haplogroups (C and G) and under-representation in haplogroup R, to which the reference sequence belongs, indicates that there may be structural predisposition or susceptibility in particular lineages (Balaresque et al., 2008a).

AZFa deletions and duplications

CNV at the *AZFa* region was originally defined by the observation of rare deletions associated with spermatogenic failure (Vogt et al., 1996), in particular Sertoli-cell only syndrome, in which post-meiotic germ cells are absent. The ~790-kb deleted region contains two protein-coding genes, *USP9Y* and *DDX3Y* (formerly *DBY*), both with functional X-linked homologues. Although early observations had suggested a key role for *USP9Y* (Sun et al., 1999), individuals with partial deletions affecting only this gene are fertile (Krausz et al., 2006), suggesting that *DDX3Y* is critical.

The mechanism underlying *AZFa* deletion (Figure 5a) is NAHR between a direct pair of ~10-kb human endogenous retroviral sequences (HERVs) (Blanco et al., 2000; Kamp et al., 2000; Sun et al., 2000), which also engage in directional gene conversion (Bosch et al., 2004). Two independent examples of the reciprocal duplication were identified after investigation of chromosomes showing STR duplications (Bosch and Jobling, 2003), and shown to be due to HERV-mediated NAHR. No direct information is available about the phenotypic effects of such duplications, although fertility is unlikely to be seriously impaired, since the chromosomes can persist in the population. Examination of the YHRD shows 18 other examples of STR duplication that probably represent *AZFa* duplication chromosomes. Notably, rearrangements involving *AZFa* were not identified in any of the systematic surveys (Redon et al., 2006; Repping et al., 2006), reflecting their low frequency in the population.

AZFb deletions

The *AZFb* region was originally defined as one of the three non-overlapping intervals in which deletions could impair spermatogenesis (Vogt et al., 1996), though larger deletions characterised as *AZFb+c* were also observed. Molecular

characterisation showed that the *AZFb* region actually overlaps the *AZFc* region (Repping et al., 2002); most *AZFb* deletions are caused by NAHR between repeats within palindromes P5 and proximal P1 (Figure 5b) encompassing 6.2Mb and removing 19 protein-coding genes. In most *AZFb+c* deletions NAHR between P5 and distal P1 removes 7.7Mb of DNA, containing 25 protein-coding genes (Repping et al., 2002). Other deletions in the region seem not to be homology-mediated (Repping et al., 2002; Vinci et al., 2005). Reciprocal P5/P1-mediated duplications have not been reported, though they should be readily recognisable through STR allele duplications if they persist in populations for sufficiently long.

AMELY deletions and duplications

Interstitial deletions of Yp have been detected due to the very widespread use of the *AMELY*/*AMELX* sex test incorporated in all commercial autosomal forensic profiling kits, in which a normal male is signalled by Y- and X-specific PCR products that differ in size (Sullivan et al., 1993). When an individual lacks the Y product, but other evidence indicates a male and small-scale primer site mutation can be excluded, a so-called *AMELY* deletion is diagnosed (Santos et al., 1998). The very high ascertainment means that some are rare sporadic cases with apparent non-homologous mechanisms, but the majority are 3.0-3.8-Mb deletions mediated by NAHR between the proximal array of *TSPY1* repeats (DYZ5) and a single distal *TSPY1* repeat (Jobling et al., 2007). Consideration of the phylogenetic context indicates that at least seven such independent deletions were represented in a set of 45 deletion chromosomes, though one founder deletion, within haplogroup J2e1* has risen to high frequency in some populations (Cadenas et al., 2006; Jobling et al., 2007), reaching ~2% in India. The phenotypic consequences of the deletion (which removes the genes *PRKY* and *TBL1Y*, as well as *AMELY*) are not clear, but seem unlikely to be severe.

The reciprocal *TSPY1*-mediated duplication of the region has also been observed, in a pair of brothers (Murphy et al., 2007). Ascertainment was initially through an elevated *AMELY*:*AMELX* ratio in PCR in one of the brothers who was being investigated because of non-Hodgkin lymphoma, and was confirmed using CGH. The brothers were otherwise apparently normal.

As in the case of the *AZFa* duplications, *AMELY*-region CNVs were not detected in systematic surveys (Redon et al., 2006; Repping et al., 2006).

TSPY1 copy number variation

The *TSPY1* genes encode a protein that is a member of a superfamily including the protooncogene SET, found in the cytoplasm of spermatogonia

(Schnieders et al., 1996). The genes are unusual in being arranged in a tandem array of 20.4-kb repeat units on proximal Yp, with a single active copy located more distally. Array length varies through NAHR, but maintenance of a minimum copy number through selection is suggested by the evolutionary conservation of multiple copies of the gene on the Y chromosomes of other mammals (Guttenbach et al., 1992; Jakubiczka et al., 1993; Murphy et al., 2006; Raudsepp et al., 2004), and the limited degree of copy number polymorphism observed in two studies of human Y chromosomes: one study finds a median number of 29 copies, with a range of 18-47 in a sample of 89 chromosomes (Mathias et al., 1994); the other finds a median of 32, with range 23-64 in 47 chromosomes (Repping et al., 2006). Increased *TSPY1* copy number has been reported in infertile males (Vodicka et al., 2007), but the phenotype of reduced copy number is unknown.

Other candidate CNVs

Duplications of other STRs, including DYS19, DYS390, DYS391, DYS393 and DYS385 suggest the existence of other Y-CNVs [http://www.yhrd.org/]. In the case of DYS19 duplications, the repetitive sequence context makes the underlying mechanism difficult to discern (Balaresque et al., 2008b), though again the high ascertainment appears to be identifying rare CNVs that may have arisen by non-homologous mechanisms. Some are identical by descent, and have reached high frequency in particular populations (Balaresque et al., 2008b; Capelli et al., 2007) through drift and social selection.

The future

As CNV typing and new sequencing methods advance, they should provide means for a relatively unbiased assessment of CNV on the Y. Most CNVs are small, but we currently know very little about small Y-CNVs, so much remains to be discovered. The 1000 Genomes Project aims to detect variants down to 1% frequency in European, East Asian and African populations, and will detect some of the small CNVs, although the size range accessible to current technologies remains unclear. It is to be hoped that additional surveys can examine other populations – South Asian populations are currently conspicuous by their absence - and that the Y CNVs discovered are included in commercially available CNV typing platforms – there is coverage on the Illumina Human1M-Duo BeadChip, which is promising.

It would certainly be a pity if the Y chromosome were ignored in future CNV analyses. After all, the sex chromosomes constitute the major copy number polymorphism, with half of our species carrying one X and one Y chromosome, and the other two Xs and no Y chromosome; in addition, one in a thousand men carry

extra Y chromosomes without knowing it. Gene dosage effects are modulated through X inactivation, XY homology, and the functional specialisation of the Y, but nonetheless we still have much to learn about the variation of sex-chromosomal gene dosage and its influence on inter-individual variation.

Acknowledgements

I thank Chris Tyler-Smith for helpful comments. My work is supported by a Wellcome Trust Senior Fellowship in Basic Biomedical Science (grant no. 057559).

Figure legends

Figure 1: Phylogenetic framework for the study of Y-CNVs.
a) The Y phylogeny, constructed using binary markers, and its major branches labelled A-T. b) Pie charts showing the frequencies of three hypothetical Y-CNVs in various branches. Based on parsimony, we can estimate minimum numbers of events. CNV1: one event - unique occurrence in the founder of superhaplogroup DE; CNV2: minimum four events - occurrence in the founder of superhaplogroup NO, in hgF1 and S, plus at least one reversion within NO; CNV3: eight events in independent haplogroups.

Figure 2: Discovering Y-CNVs through anomalous STR haplotypes.
Following (in this hypothetical schematic case) NAHR-mediated deletion, absence of one STR in a haplotype signals the event; following the reciprocal duplication, peak height can signal the event, but it is more usually inferred once mutation has given two STR alleles of different length that are readily distinguishable.

Figure 3: Regions rich in Y-CNVs found in a systematic survey.
Above the chromosomal idiogram are shown the approximate positions of palindromes 1-8 (P1-8) and the inverted repeats IR2 and IR3 (Skaletsky et al., 2003). Below the idiogram is shown a representation of the log2 ratios from comparative genomic hybridisation (CGH) to BAC clones spanning the Y euchromatin for the HapMap individuals (Redon et al., 2006). Log2 ratios greater than zero are seen as green, and those lower as red, above and below the yellow line. The most dynamic regions correspond to the *TSPY1* array and the *AZFc* region. PAR: pseudoautosomal region.

Figure 4: Ampliconic organisation of the *AZFc* region, and examples of CNVs.

a) Reference sequence architecture, with highly similar large paralogous repeat units indicated by coloured arrows (Kuroda-Kawaguchi et al., 2001). Above this are indicated widely used STRs (prefixed 'DYS') in which missing or extra alleles can indicate CNVs, and below are the protein-coding genes in the region.
b) Examples of known structures generated by NAHR, with grey rectangles indicating the amplicons within which recombination occurs – exact position is uncertain. The b2/b4 deletion is associated with spermatogenic failure, but the phenotypic effects of the others are controversial; they may be selectively

neutral. Some inversions are shown because they act as intermediates in duplication/deletion formation.

Figure 5: Organisation of the *AZFa* and *b* regions, and the region around *AMELY*. In each case widely used STRs are indicated above the lines, and protein-coding gene names below.

a) The *AZFa* region. The flanking HERVs (red arrows) sponsor deletion and duplication through NAHR, indicated by the curved grey arrow.
b) The *AZFb* region. Two classes of common deletions are sponsored by repeats in P1 and P5 (orange arrows) through NAHR, indicated by the curved grey arrows.
c) The region around *AMELY*. A common and recurrent class of deletions is sponsored by NAHR between the single distal copy of *TSPY1* and the proximal array of *TSPY1* repeats (pink triangles).
d) The *TSPY1* array varies in ~20-kb repeat copy number between ~20 and ~40.

References

Abramsky L, Chapple J: 47,XXY (Klinefelter syndrome) and 47,XYY: estimated rates of and indication for postnatal diagnosis with implications for prenatal counselling. Prenat Diagn 17:363-368 (1997).

Bailey JA, Gu Z, Clark RA, Reinert K, Samonte RV, Schwartz S, Adams MD, Myers EW, Li PW, Eichler EE: Recent segmental duplications in the human genome. Science 297:1003-1007 (2002).

Balaresque P, Bowden GR, Parkin EJ, Omran GA, Heyer E, Quintana-Murci L, Roewer L, Stoneking M, Nasidze I, Carvalho-Silva DR, Tyler-Smith C, de Knijff P, Jobling MA: Dynamic nature of the proximal AZFc region of the human Y chromosome: multiple independent deletion and duplication events revealed by microsatellite analysis. Hum Mut in press, DOI: 10.1002/humu.20757 (2008a).

Balaresque P, Parkin EJ, Roewer L, Carvalho-Silva DR, Mitchell RJ, van Oorschot RAH, Henke J, Stoneking M, Nasidze I, Wetton J, de Knijff P, Tyler-Smith C, Jobling MA: Genomic complexity of the Y-STR DYS19: inversions, deletions and founder lineages carrying duplications. Int J Legal Med in press (2008b).

Banchs I, Bosch A, Guimera J, Lazaro C, Puig A, Estivill X: New alleles at microsatellite loci in CEPH families mainly arise from somatic mutations in the lymphoblastoid cell-lines. Hum Mut 3:365-372 (1994).

Bentley DR: Whole-genome re-sequencing. Curr Opin Genet Dev 16:545-552 (2006).

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

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

Bosch E, Jobling MA: Duplications of the *AZFa* region of the human Y chromosome are mediated by homologous recombination between HERVs and are compatible with male fertility. Hum Mol Genet 12:341-347 (2003).

Bowden GR, Balaresque P, King TE, Hansen Z, Lee AC, Pergl-Wilson G, Hurley E, Roberts SJ, Waite P, Jesch J, Jones AL, Thomas MG, Harding SE, Jobling MA: Excavating past population structures by surname-based sampling: the genetic legacy of the Vikings in northwest England. Mol Biol Evol 25:301-309 (2007).

Butler JM, Decker AE, Kline MC, Vallone PM: Chromosomal duplications along the Y-chromosome and their potential impact on Y-STR interpretation. J Forensic Sci 50:853-859 (2005).

Cadenas AM, Regueiro M, Gayden T, Singh N, Zhivotovsky LA, Underhill PA, Herrera RJ: Male amelogenin dropouts: phylogenetic context, origins and implications. Forensic Sci Int 166:155-163 (2006).

Capelli C, Brisighelli F, Scarnicci F, Blanco-Verea A, Brion M, Pascali V: Phylogenetic evidence for multiple independent duplication events at the DYS19 locus. Forensic Sci Int Genet 1:287-290 (2007).
Chang PL, Sauer MV, Brown S: Y chromosome microdeletion in a father and his four infertile sons. Hum Reproduct 14:2689-2694 (1999).
Elsheikh M, Dunger DB, Conway GS, Wass JA: Turner's syndrome in adulthood. Endocr Rev 23:120-140 (2002).
Fernandes S, Paracchini S, L.H. M, Floridia G, Tyler-Smith C, Vogt PH: Are sequence family variants useful for identifying deletions in the human Y chromosome? - Reply. Am J Hum Genet 75:517-519 (2004a).
Fernandes S, Paracchini S, Meyer LH, Floridia G, Tyler-Smith C, Vogt PH: A large AZFc deletion removes DAZ3/DAZ4 and nearby genes from men in Y haplogroup N. Am J Hum Genet 74:180-187 (2004b).
Guttenbach M, Müller U, Schmid M: A human moderately repeated Y-specific DNA sequence is evolutionarily conserved in the Y chromosome of the great apes. Genomics 13:363-367 (1992).
Hammer MF, Zegura SL: The human Y chromosome haplogroup tree: nomenclature and phylogeny of its major divisions. Annu Rev Anthropol 31:303-321 (2002).
Hanson EK, Ballantyne J: Comprehensive annotated STR physical map of the human Y chromosome: Forensic implications. Leg Med (Tokyo) 8:110-120 (2006).
Higgins CD, Swerdlow AJ, Schoemaker MJ, Wright AF, Jacobs PA: Mortality and cancer incidence in males with Y polysomy in Britain: a cohort study. Hum Genet 121:691-696 (2007).
Hurles ME, Dermitzakis ET, Tyler-Smith C: The functional impact of structural variation in humans. Trends Genet 24:238-245 (2008).
Iafrate AJ, Feuk L, Rivera MN, Listewnik ML, Donahoe PK, Qi Y, Scherer SW, Lee C: Detection of large-scale variation in the human genome. Nat Genet 36:949-951 (2004).
International HapMap Consortium: A haplotype map of the human genome. Nature 437:1299-1320 (2005).
Jakubiczka S, Schnieders F, Schmidtke J: A bovine homologue of the human TSPY gene. Genomics 17:732-735 (1993).
Jobling MA: A survey of long-range DNA polymorphisms on the human Y chromosome. Hum Mol Genet 3:107-114 (1994).
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: Structural variation on the short arm of the human Y chromosome: recurrent multigene deletions encompassing *Amelogenin Y*. Hum Mol Genet 16:307-316 (2007).

Jobling MA, Samara V, Pandya A, Fretwell N, Bernasconi B, Mitchell RJ, Gerelsaikhan T, Dashnyam B, Sajantila A, Salo PJ, Nakahori Y, Disteche CM, Thangaraj K, Singh L, Crawford MH, Tyler-Smith C: Recurrent duplication and deletion polymorphisms on the long arm of the Y chromosome in normal males. Hum Mol Genet 5:1767-1775 (1996).
Jobling MA, Tyler-Smith C: The human Y chromosome: an evolutionary marker comes of age. Nat Rev Genet 4:598-612 (2003).
Kamp C, Hirschmann P, Voss H, Huellen K, Vogt PH: 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 (2000).
Karafet TM, Mendez FL, Meilerman M, Underhill PA, Zegura SL, Hammer MF: New binary polymorphisms reshape and increase resolution of the human Y-chromosomal haplogroup tree. Genome Res 18:830-838 (2008).
Karcanias AC, Ichimura K, Mitchell MJ, Sargent CA, Affara NA: Analysis of sex chromosome abnormalities using X and Y chromosome DNA tiling path arrays. J Med Genet 44:429-436 (2007).
Kayser M, Kittler R, Erler A, Hedman M, Lee AC, Mohyuddin A, Mehdi SQ, Rosser Z, Stoneking M, Jobling MA, Sajantila A, Tyler-Smith C: A comprehensive survey of human Y-chromosomal microsatellites. Am J Hum Genet 74:1183-1197 (2004).
King TE, Bosch E, Adams SM, Parkin EJ, Rosser ZH, Jobling MA: Inadvertent diagnosis of male infertility through genealogical DNA testing. J Med Genet 42:366-368 (2005).
Krausz C, Degl'Innocenti S, Nuti F, Morelli A, Felici F, Sansone M, Varriale G, Forti G: Natural transmission of USP9Y gene mutations: a new perspective on the role of AZFa genes in male fertility. Hum Mol Genet 15:2673-2681 (2006).
Kühnert B, Gromoll J, Kostova E, Tschanter P, Luetjens CM, Simoni M, Nieschlag E: Case report: natural transmission of an AZFc Y-chromosomal microdeletion from father to his sons. Hum Reprod 19:886-888 (2004).
Kuroda-Kawaguchi T, Skaletsky H, Brown LG, Minx PJ, Cordum HS, Waterston RH, Wilson RK, Silber S, Oates R, Rozen S, Page DC: The AZFc region of the Y chromosome features massive palindromes and uniform recurrent deletions in infertile men. Nat Genet 29:279-286 (2001).
Lange J, Skaletsky H, Bell GW, Page DC: MSY Breakpoint Mapper, a database of sequence-tagged sites useful in defining naturally occurring deletions in the human Y chromosome. Nucleic Acids Res 36:D809-814 (2008).
Levy S, Sutton G, Ng PC, Feuk L, Halpern AL, Walenz BP, Axelrod N, Huang J, Kirkness EF, Denisov G, Lin Y, MacDonald JR, Pang AW, Shago M, Stockwell TB, Tsiamouri A, Bafna V, Bansal V, Kravitz SA, Busam DA, Beeson KY, McIntosh TC,

Remington KA, Abril JF, Gill J, Borman J, Rogers YH, Frazier ME, Scherer SW, Strausberg RL, Venter JC: The diploid genome sequence of an individual human. PLoS Biol 5:e254 (2007).
Locke DP, Sharp AJ, McCarroll SA, McGrath SD, Newman TL, Cheng Z, Schwartz S, Albertson DG, Pinkel D, Altshuler DM, Eichler EE: Linkage disequilibrium and heritability of copy-number polymorphisms within duplicated regions of the human genome. Am J Hum Genet 79:275-290 (2006).
Mathias N, Bayés M, Tyler-Smith C: Highly informative compound haplotypes for the human Y chromosome. Hum Mol Genet 3:115-123 (1994).
McElreavey K, Ravel C, Chantot-Bastaraud S, Siffroi JP: Y chromosome variants and male reproductive function. Int J Androl 29:298-303 (2006).
Murphy KM, Cohen JS, Goodrich A, Long PP, Griffin CA: Constitutional duplication of a region of chromosome Yp encoding AMELY, PRKY, and TBL1Y: implications for sex chromosome analysis and bone marrow engraftment analysis. J Mol Diagn 9:408-413 (2007).
Murphy WJ, Pearks Wilkerson AJ, Raudsepp T, Agarwala R, Schaffer AA, Stanyon R, Chowdhary BP: Novel gene acquisition on carnivore Y chromosomes. PLoS Genet 2:e43 (2006).
Oakey R, Tyler-Smith C: Y chromosome DNA haplotyping suggests that most European and Asian men are descended from one of two males. Genomics 7:325-330 (1990).
Page DC, de Martinville B, Barker D, Wyman A, White R, Francke U, Botstein D: Single-copy sequence hybridizes to polymorphic and homologous loci on human X and Y chromosomes. Proc Natl Acad Sci USA 79:5352-5356 (1982).
Perry GH, Ben-Dor A, Tsalenko A, Sampas N, Rodriguez-Revenga L, Tran CW, Scheffer A, Steinfeld I, Tsang P, Yamada NA, Park HS, Kim JI, Seo JS, Yakhini Z, Laderman S, Bruhn L, Lee C: The fine-scale and complex architecture of human copy-number variation. Am J Hum Genet 82:685-695 (2008).
Ratcliffe S: Long-term outcome in children of sex chromosome abnormalities. Arch Dis Child 80:192-195 (1999).
Raudsepp T, Santani A, Wallner B, Kata SR, Ren C, Zhang HB, Womack JE, Skow LC, Chowdhary BP: A detailed physical map of the horse Y chromosome. Proc Natl Acad Sci USA 101:9321-9326 (2004).
Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD, Fiegler H, Shapero MH, Carson AR, Chen W, Cho EK, Dallaire S, Freeman JL, Gonzalez JR, Gratacos M, Huang J, Kalaitzopoulos D, Komura D, MacDonald JR, Marshall CR, Mei R, Montgomery L, Nishimura K, Okamura K, Shen F, Somerville MJ, Tchinda J, Valsesia A, Woodwark C, Yang F, Zhang J, Zerjal T, Armengol L, Conrad DF, Estivill

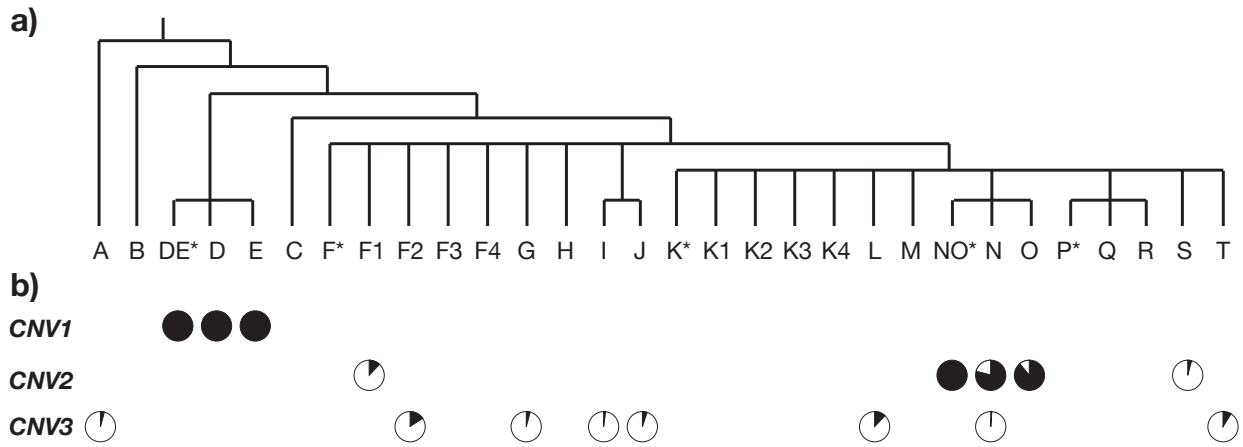
X, Tyler-Smith C, Carter NP, Aburatani H, Lee C, Jones KW, Scherer SW, Hurles ME: Global variation in copy number in the human genome. Nature 444:444-454 (2006).
Repping S, Korver CM, Oates RD, Silber S, van der Veen F, Page DC, Rozen S: Are sequence family variants useful for identifying deletions in the human Y chromosome? Am J Hum Genet 75:514-517 (2004).
Repping S, Skaletsky H, Lange J, Silber S, Van Der Veen F, Oates RD, Page DC, Rozen S: Recombination between palindromes P5 and P1 on the human Y chromosome causes massive deletions and spermatogenic failure. Am J Hum Genet 71:906-922 (2002).
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: High mutation rates have driven extensive structural polymorphism among human Y chromosomes. Nat Genet 38:463-467 (2006).
Rosser ZH, Zerjal T, Hurles ME, Adojaan M, Alavantic D, Amorim A, Amos W, Armenteros M, Arroyo E, Barbujani G, Beckman G, Beckman L, Bertranpetit J, Bosch E, Bradley DG, Brede G, Cooper G, Côrte-Real HBSM, de Knijff P, Decorte R, Dubrova YE, Evgrafov O, Gilissen A, Glisic S, Gölge M, Hill EW, Jeziorowska A, Kalaydjieva L, Kayser M, Kivisild T, Kravchenko SA, Krumina A, Kucinskas V, Lavinha J, Livshits LA, Malaspina P, Maria S, McElreavey K, Meitinger TA, Mikelsaar A-V, Nafa K, Mitchell RJ, Nicholson J, Nørby S, Pandya A, Parik J, Patsalis PC, Pereira L, Peterlin B, Pielberg G, Prata MJ, Previderé C, Roewer L, Rootsi S, Rubinsztein DC, Saillard J, Santos FR, Stefanescu G, Sykes BC, Tolun A, Villems R, Tyler-Smith C, Jobling MA: Y-chromosomal diversity within Europe is clinal and influenced primarily by geography, rather than by language. Am J Hum Genet 67:1526-1543 (2000).
Rozen S, Skaletsky H, Marszalek JD, Minx PJ, Cordum HS, Waterston RH, Wilson RK, Page DC: Abundant gene conversion between arms of massive palindromes in human and ape Y chromosomes. Nature 423:873-876 (2003).
Santos FR, Pandya A, Tyler-Smith C: Reliability of DNA-based sex tests. Nat Genet 18:103 (1998).
Schnieders F, Dörk T, Arnemann J, Vogel T, Werner M, Schmidtke J: Testis-specific protein, Y-encoded (TSPY) expression in testicular tissues. Hum Mol Genet 5:1801-1807 (1996).
Shanske A, Sachmechi I, Patel DK, Bishnoi A, Rosner F: An adult with 49,XYYYY karyotype: case report and endocrine studies. Am J Med Genet 80:103-106 (1998).
Sharp AJ, Locke DP, McGrath SD, Cheng Z, Bailey JA, Vallente RU, Pertz LM, Clark RA, Schwartz S, Segraves R, Oseroff VV, Albertson DG, Pinkel D, Eichler EE:

Segmental duplications and copy-number variation in the human genome. Am J Hum Genet 77:78-88 (2005).
Skaletsky H, Kuroda-Kawaguchi T, Minx PJ, Cordum HS, Hillier L, Brown LG, Repping S, Pyntikova R, Ali J, Bieri T, Chinwalla A, Delehaunty A, Delehaunty K, Du H, Fewell G, Fulton L, Fulton R, Graves T, Hou S-F, Latrielle P, Leonard S, Mardis E, Maupin R, McPherson J, Miner T, Nash W, Nguyen C, Ozersky P, Pepin K, Rock S, Rohlfing T, Scott K, Schultz B, Strong C, Tin-Wollam A, Yang S-P, Waterston RH, Wilson RK, Rozen S, Page DC: The male-specific region of the human Y chromosome: a mosaic of discrete sequence classes. Nature 423:825-837 (2003).
Sullivan KM, Mannucci A, Kimpton CP, Gill P: A rapid and quantitative DNA sex test - fluorescence-based PCR analysis of X-Y homologous gene amelogenin. Biotechniques 15:636 (1993).
Sun C, Skaletsky H, Birren B, Devon K, Tang Z, Silber S, Oates R, Page DC: An azoospermic man with a *de novo* point mutation in the Y-chromosomal gene *USP9Y*. Nat Genet 23:429-432 (1999).
Sun C, Skaletsky H, Rozen S, Gromoll J, Nieschlag E, Oates R, Page DC: Deletion of *azoospermia factor a* (*AZFa*) region of human Y chromosome caused by recombination between HERV15 proviruses. Hum Mol Genet 9:2291-2296 (2000).
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: A physical map of the human Y chromosome. Nature 409:943-945 (2001).
Underhill PA, Kivisild T: Use of Y chromosome and mitochondrial DNA population structure in tracing human migrations. Annu Rev Genet 41:539-564 (2007).
Vinci G, Raicu F, Popa L, Popa O, Cocos R, McElreavey K: A deletion of a novel heat shock gene on the Y chromosome associated with azoospermia. Mol Hum Reprod 11:295-298 (2005).
Vodicka R, Vrtel R, Dusek L, Singh AR, Krizova K, Svacinova V, Horinova V, Dostal J, Oborna I, Brezinova J, Sobek A, Santavy J: TSPY gene copy number as a potential new risk factor for male infertility. Reprod Biomed Online 14:579-587 (2007).
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 H-J, Jung A, Engel W, Haidl G: Human Y chromosome azoospermia factors (AZF) mapped to different subregions in Yq11. Hum Mol Genet 5:933-943 (1996).
Vogt PH, Falcao CL, Hanstein R, Zimmer J: The AZF proteins. Int J Androl (2008).
Vollrath D, Foote S, Hilton A, Brown LG, Beer-Romero P, Bogan JS, Page DC: The human Y chromosome - a 43-interval map based on naturally-occurring deletions. Science 258:52-59 (1992).

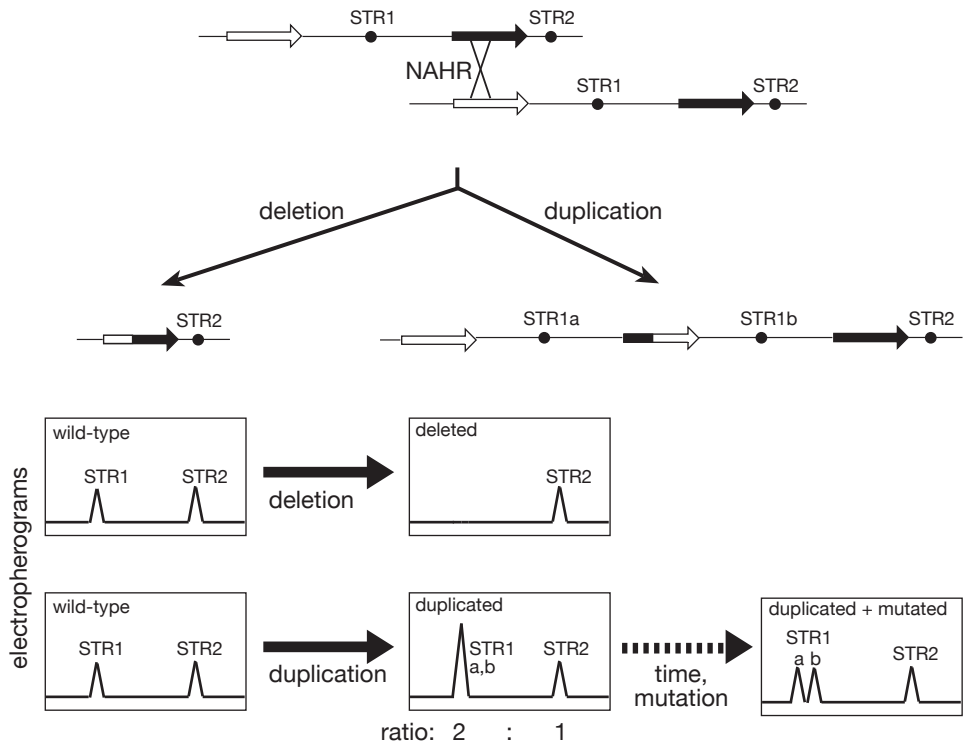
Wheeler DA, Srinivasan M, Egholm M, Shen Y, Chen L, McGuire A, He W, Chen YJ, Makhijani V, Roth GT, Gomes X, Tartaro K, Niazi F, Turcotte CL, Irzyk GP, Lupski JR, Chinault C, Song XZ, Liu Y, Yuan Y, Nazareth L, Qin X, Muzny DM, Margulies M, Weinstock GM, Gibbs RA, Rothberg JM: The complete genome of an individual by massively parallel DNA sequencing. Nature 452:872-876 (2008).

Willuweit S, Roewer L, on behalf of the International Forensic Y Chromosome User Group: Y chromosome haplotype reference database (YHRD): Update. Forensic Sci Int Genet 1:83-87 (2007).

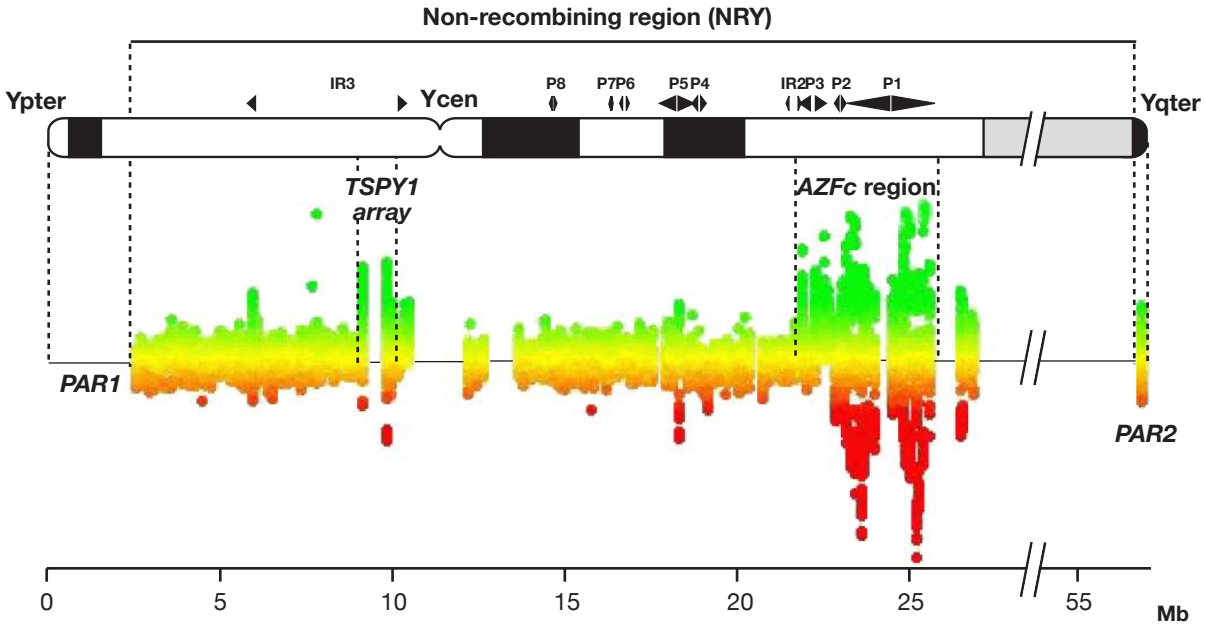
Xue Y, Zerjal T, Bao W, Zhu S, Lim SK, Shu Q, Xu J, Du R, Fu S, Li P, Yang H, Tyler-Smith C: Recent spread of a Y-chromosomal lineage in northern China and Mongolia. Am J Hum Genet 77:1112-1116 (2005).



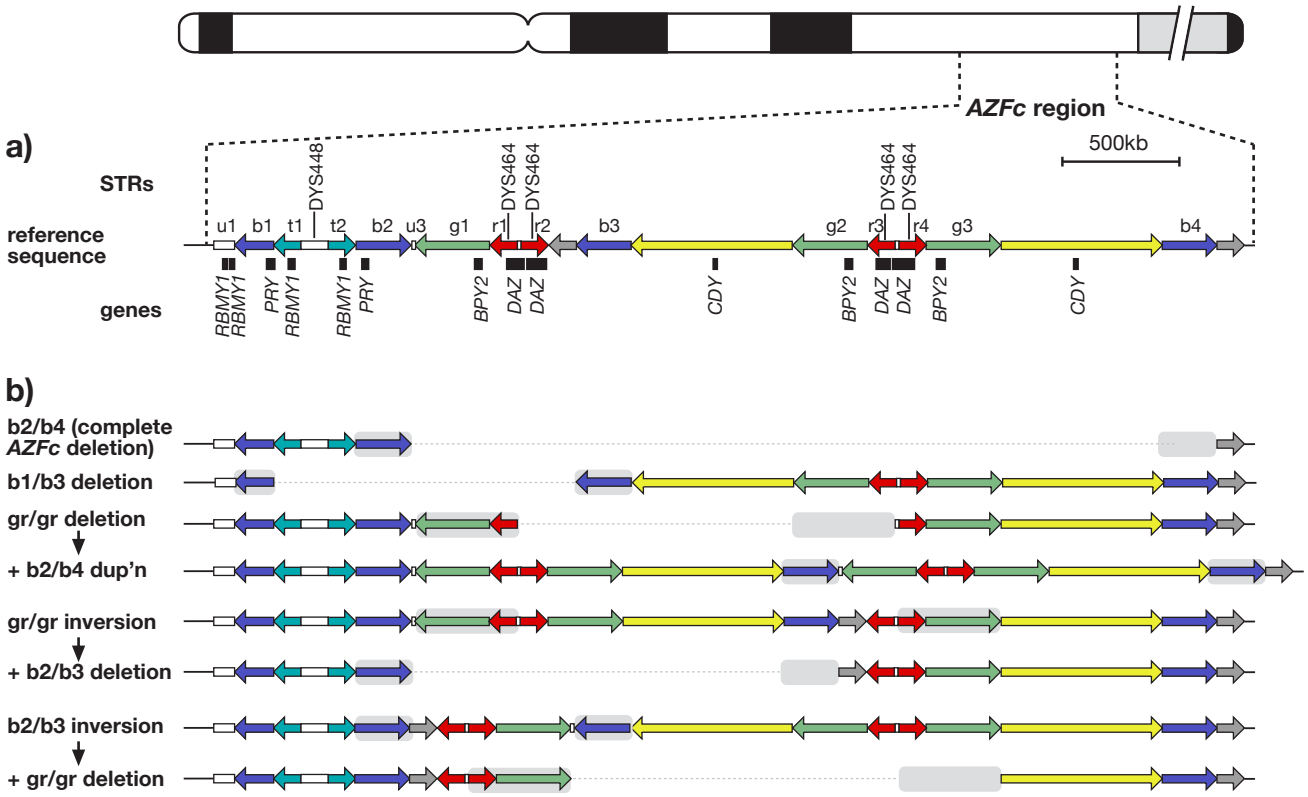
Jobling, Figure 1



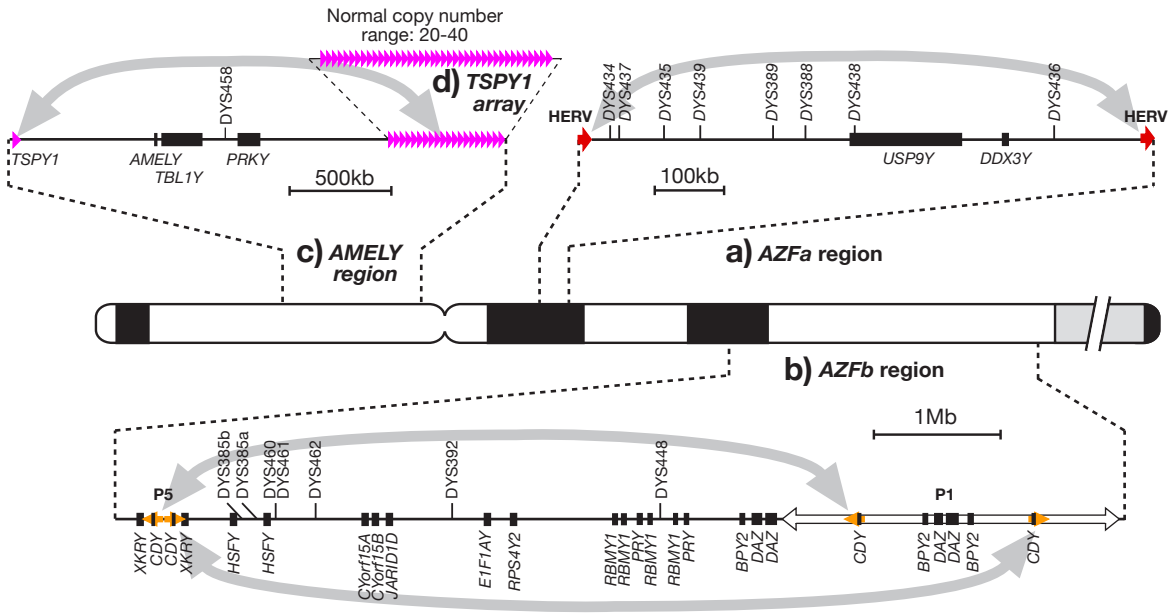
Jobling, Figure 2



Jobling, Figure 3



Jobling, Figure 4



Jobling, Figure 5