The fragility of fertility

Pauline Yen

Department of Pediatrics, Harbor-UCLA Research & Education Institute, Torrance, California 90502, USA. e-mail: pyen@rei.edu

Among apparently healthy men who nonetheless produce few or no sperm, about ten percent have microdeletions on their Y chromosome. Most of these *de novo* mutations occur in the *AZFc* region on the long arm of the chromosome. The high frequency of deletion is explained by the extraordinary structure of the region, which consists almost entirely of very long repeat units.

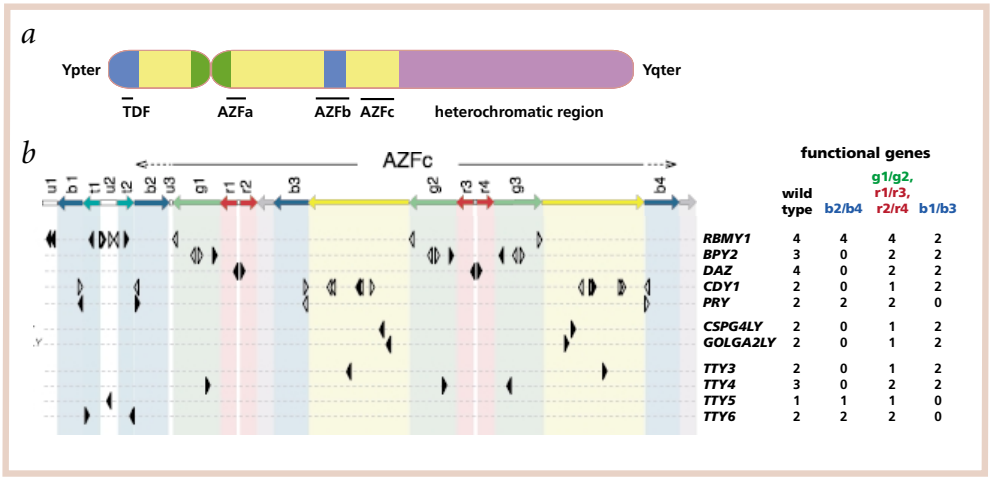
With fewer than 50 genes or gene families[1]—many expressed only in the testis—the Y chromosome is not essential for life. Its main task is to ensure that men can make and deliver sperm for the continuation of the species. The sex-determining role of the Y chromosome has been known for a long time. Its role in spermatogenesis, however, was recognized only a quarter-century ago, when Tiepolo and Zuffardi[2] found cytogenetically detectable Yq deletions in six subfertile men. An *AZF* (for 'azoospermia factor') locus was thus assigned to the long arm of the Y chromosome, near the heterochromatic region. Molecular probing later identified Yq microdeletions in about ten percent of males with sperm deficiencies[3]. Most of the deletions fall into three nonoverlapping regions, designated *AZFa*, *AZFb* and *AZFc* (see figure)[4], with *AZFc* being the most frequently deleted. Nearly all such deletions arise *de novo*, indicating a high deletion rate of about 3×10^{-4}. Most large, recurrent deletions in the human genome are caused by recombination between long direct repeats[5]. Two papers published last year showed that most *AZFa* deletions arise from recombination between two 10-kb direct repeats that are 800 kb apart[6,7]. On page 279, Tomoko Kuroda-Kawaguchi, David Page and colleagues[8] present strong evidence that most *AZFc* deletions involve a 3.5-Mb segment bounded by two 229-kb direct repeats. This is by far the largest well defined recurrent deletion in the human genome.

A thicket of repeats

What distinguishes this work from other deletion-mapping projects is both its scale and its difficulty. *AZFc* is in a region of the Y chromosome that is made up almost entirely of long direct and inverted repeats (called amplicons; see figure)[8]. These amplicons range in size from 100 to 700 kb, and members of each amplicon family are 99.8% identical, making it very difficult to tell them apart. Understandably, earlier attempts to construct physical maps of the region by conventional methods, such as mapping DNA markers using naturally occurring deletions and assembling YAC clones based on their restriction maps and marker contents, failed miserably, and the literature of male infertility has become cluttered with disparate sequence-tagged-site (STS) maps[9–13].

Thanks to the commitment and sheer determination of Kuroda-Kawaguchi and colleagues[8], we now have a meaningful map of the region, together with the sequence. The authors solved the problem by sequencing highly overlapping BAC clones derived from a single man. The strategy avoided person-to-person polymorphisms and thus allowed identification of amplicons from different families. Together with the Genome Sequencing Center at Washington University in St. Louis, the authors sequenced a total of 48 BACs with 3 Mb of overlap to assemble a 4.5-Mb sequence contig spanning the entire *AZFc* region. The contig contains three palindromes made from six distinct families of amplicons. It also contains 11 gene families—5 known and 6 new—with a total of 27 potentially functional genes (see figure). All of the genes in the region are expressed exclusively or predominantly in the testis.

Kuroda-Kawaguchi and colleagues[8] next mapped the breakpoints in men with *AZFc* deletions. Because of the region's extraordinary amplicon structure, they were unable to isolate and sequence the junction fragments, which is usually required for the identification of the deletion breakpoints. Nonetheless, their mapping data show convincingly that deletions in 47 of 48 patients were caused by recombination between two direct amplicons, b2 and b4, that are 3.5 Mb apart. Why did they observe recombination events between b2



AZFc* annotation**. ***a, The human Y chromosome showing loci with male-specific functions. ***b***, Amplicon and transcription maps of the *AZFc* region (see Fig. 2 on page 281). There are six families of amplicons (or massive repeat units) in the region, shown by different colors. Recombination could occur between any pair of amplicons that are present in the same orientation, resulting in the deletion of the segment in-between. The numbers of potential functional genes (solid triangles) that will remain on the Y chromosome after the recombination events are shown in the table at the right.

BOB CRIMI

and b4 only, when there are five pairs of direct amplicons in the contig? The answer is probably that recombination between other pairs of amplicons does not affect spermatogenesis. As shown in the figure, recombination between b2 and b4 eliminates seven gene families completely, whereas recombination involving g1/g2, r1/r3 or r2/r4 leaves some gene families intact or only partially deleted. Recombination between b1 and b3 does remove three gene families; however, b1/b3 recombination may occur less frequently than b2/b4 recombination because b1 and b3 share a shorter segment of homology. It is also possible that the gene families deleted by b1/b3 recombination do not have significant roles in spermatogenesis, so that sperm counts are largely unaffected.

Fertile ground

An electronic file of the 4.5-Mb sequence of the *AZFc* region, fully annotated with STSs, amplicons and genes, are posted on both the *Nature Genetics* web site and that of the Page lab (http://staffa.wi.mit.edu/page/Y/azfc). Investigators who have analyzed *AZFc* deletions in infertile men can now reassess their data using the new map. The availability of new STSs that distinguish between amplicons will allow investigators to redefine deletion breakpoints and identify rare deletions not caused by b2/b4 recombination. These rare deletions may allow us to evaluate the contribution of each gene family to male infertility.

How accurate are the assembled sequence and the map of the *AZFc* region? Despite the authors' painstaking checking and rechecking, the formal possibility of omissions or errors in the sequence cannot be ruled out. So, it is desirable to verify the map by other means, such as pulsed-field gel electrophoresis. Southern hybridization with a given probe should show fragments of the expected number and size. Unfortunately, no cell line or genomic DNA is available from the donor, RPC1-11, from whom all the BAC clones were derived. Use of DNA from any other man runs the risk of identifying differences due to polymorphisms. The abundance of amplicons suggests that the *AZFc* region may be highly polymorphic. For example, there are reports of restriction fragment length polymorphisms and different copy numbers of the *DAZ* genes in the region[14,15]. So, even though the map presented by Kuroda-Kawaguchi and colleagues[8] could become the 'official' map of the region, it will not be the last to be published. □

1. Lahn, B.T. & Page, D.C. *Science* **278**, 675–680 (1997).
2. Tiepolo, L. & Zuffardi, O. *Hum. Genet.* **34**, 119–124 (1976).
3. McElreavey, K. & Krausz, C. *Am. J. Hum. Genet.* **64**, 928–933 (1999).
4. Vogt, P.H, *et al. Hum. Mol. Genet.* **5**, 933–943 (1996).
5. Lupski, J.R. *Trends Genet.* **14**, 417–422 (1998).
6. Sun, C. *et al. Hum. Mol. Genet.* **9**, 2291–2296 (2000).
7. Camp, C. *et al. Hum. Mol. Genet.* **9**, 2563–2572 (2000).
8. Kuroda-Kawaguchi, T. *et al. Nature Genet.* **29**, 279–286 (2001).
9. Jones, M.H. *et al. Genomics* **24**, 266–275 (1994).
10. Yen, P.H. *Genomics* **54**, 5–12 (1998).
11. Stuppia, L. *et al. Cytogenet. Cell Genet.* **72**, 155–158 (1996).
12. Kent-First, M.G. *et al. Mol. Hum. Reprod.* **2**, 943–950 (1996).
13. Foresta, C. *et al. J. Clin. Endocrinol. Metab.* **82**, 1075–1080 (1997).
14. Jobling, M.A. *Hum. Mol. Genet.* **3**, 107–114 (1994).
15. Glaser, B. *et al. Chromosome Res.* **5**, 23–30 (1997).

The replication requirement

Veronica J. Vieland

Department of Biostatistics, Division of Statistical Genetics, University of Iowa College of Public Health and Department of Psychiatry, University of Iowa College of Medicine, Iowa City, Iowa 52242-1008, USA. e-mail: veronica-vieland@uiowa.edu

Human geneticists withhold judgment on any report of linkage or allelic association until it is independently replicated. But the independent replication criterion has become harder to meet. Does this signal the demise of the reverse genetics paradigm or a problem with the replication requirement itself?

Human geneticists are becoming inured to nonreplication of reported allelic associations for complex diseases, just as we have come to anticipate nonreplication of positive linkage findings in follow-up studies. On page 306, John Ioannidis and colleagues[1] use meta-analyses to confirm that initial reports of strong association indeed tend to be followed by studies showing little or no association; associations that initially reach statistical significance tend to lose significance when all of the follow-up data are considered. This finding has immediate implications for gene mapping that is based on disequilibrium testing. It also provides a fitting backdrop for reconsidering the independent-replication requirement in human gene-mapping studies.

Rationale for replication

The International Human Gene Mapping Workshops of the 1970s[2,3] first codified the notion of a 'confirmed' map assignment as one supported both by a lod score greater than or equal to 3 in one study and by corroborating evidence from an independent study. At that time, a primary purpose of requiring confirmation was to ensure a sound basis for clinical-genetic risk calculations. Now, however, confirmation is primarily used to determine allocation of resources for follow-up molecular work *en route* to cloning. This drift in function has led to a divergence in form. Whereas an independent replication criterion is still widely accepted in the scientific literature, this has come to mean something rather different over time. (By contrast, OMIM[4] still adheres closely to the original standards.)

Originally, confirmation precluded any overlap in personnel across the two studies. This was intended to provide protection against artifacts of the genotyping and computational technology available at that time. Today, we trust that the technology is reliable enough not to require independent validation. We tend to tolerate overlap in laboratories, computer programs and even personnel between initial and confirmatory studies. Diagnostic reliability, however, remains a concern.

It was also originally understood that corroboration could take many forms. Early localizations were often confirmed