THE HUMAN Y CHROMOSOME: GENE CONTENT AND CHROMOSOMAL ABNORMALITIES

by

Bruce T. Lahn

B.A. Harvard University, 1991

Submitted to the Department of Biology
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

Massachusetts Institute of Technology

October 1997

Signature of author _____
Department of Biology

Certified by _____
David C. Page
Professor, Department of Biology
Thesis supervisor

Accepted by _____
Frank Solomon
Professor, Department of Biology
Chairman, Committee on Graduate Studies

THE HUMAN Y CHROMOSOME: GENE CONTENT AND CHROMOSOMAL ABNORMALITIES

by

Bruce T. Lahn

Submitted to the Department of Biology October 1997 in partial fulfillment of the requirement for the degree of Doctor of Philosophy

ABSTRACT

This thesis is a two part report on a series of studies aimed at understanding the function, organization and evolution of the human Y chromosome.

I

The first part (Chapters 2 & 3) reports on a systematic effort to identify genes within the non-recombining region of the Y (NRY), and where applicable, their X homologs. By the method of cDNA selection, I cloned 12 novel genes, 10 with full-length cDNA sequences. These genes, combined with eight previously identified genes, are a good representation of NRY's gene content, and likely constitute the majority of genes or gene families in this region of the human genome. They fall into two distinct classes.

Genes in the first class have close homologs on the X, which were either previously identified, or cloned by us. They are referred to collectively as X/Y homologous genes, and share the following features: 1) they typically have ubiquitous expression; 2) for each pair of X/Y homologous genes, expressions from X and Y copies are typically comparable; and 3) X copies escape X-inactivation. These genes likely have housekeeping functions and meet certain predictions for genes involved in Turner syndrome, a human condition caused by complete or partial sex chromosome monosomy, or XO. Theorists have long argued that the Y is a degenerate copy of the X. X/Y homologous genes therefore represent evolutionary vestiges of the ancestral homology between the two sex chromosomes that has undergone extensive degeneration on the Y.

The second class of genes lack X-homologs and are referred to as male-specific genes. They share the following features: 1) they typically exist as multi-copy, closely related gene families; and 2) their expression is restricted to the testis. These genes are involved in male-specific biology. Given their testis-limited expression, they likely function in spermatogenesis and may account for infertility in men with Y deletions. As a chromosome with male-restricted transmission, the Y may be uniquely suitable to carry such genes that function exclusively in males.

II

The second part (Chapter 4) reports on a study of a frequently occurring chromosomal abnormality of the Y - microscopically detectable deletions of the long arm, or Yq-. Study subjects were ten unrelated Yq- men with widely varying phenotypes ranging from infertility in some to mental retardation in others. Molecular analyses revealed in three individuals with the most severe phenotypes - mental retardation and severe developmental delay - the presence of a small portion of distal Xq on the long arm of their Yq- chromosome. This arrangement, which resulted in partial X disomy, apparently arose from an aberrant Xq/Yq exchange in the paternal germline. A representative gene on distal Xq, *G6PD*, was present on the Yq- chromosome in all three severely affected individuals. The enzyme it encoded, glucose-6-phosphate dehydrogenase, displayed twice-normal activity in patients as compared to their normal parents. We postulate that functional disomy of the portion of the Xq present on the Yq- chromosome in these three individuals are responsible for their severe phenotypes. These results underscore the developmental importance of maintaining functional monosomy of the X chromosome via X-inactivation.

Thesis Advisor: David C. Page
Title: Professor, Department of Biology

Acknowledgments

I thank the following people, who among others, contributed to my graduate experience.

My advisor David Page.

Members of my committee: David Housman, Philip Sharp, Robert Weinberg, and Louis Kunkel.

Members of the lab: Raaji Alagappan, Laura Brown, Mary Goodheart, Marta Velez-Stringer, Karin Jegalian, Judith Seligman, Helen Skaletsky, Michael Rosenberg, Renee Reijo, Tomoko Kawaguchi, Chao Sun, Charles Tilford, Richa Saxena, Doug Menke (go Doug!), Julie Markwardt, Alex Bortvin, Mark Belke, Paul Bain, Steve Rosen, and others.

Members of the Whitehead Institute: Fran Lewitter, and others.

Outside collaborators: Nancy Ma, Roy Breg, Urvashi Surti, and others.

TABLE OF CONTENTS

CHAPTER 1. INTRODUCTION

Section i: Background

Sexual dimorphism is commonly observed in life and is near universal in multi-cellular eukaryotes. Over much of human history, however, the question of how an organism acquires one of two sexually distinctive morphologies has only been the subject of creative speculations.

Around the turn of the century, the bivalent nature of chromosome pairs was established through a series of microscopic studies of meioses (Paulmier, 1899, Montgomery, 1901, Wilson, 1906). In some of these studies, investigators noticed the presence of an unpaired "heterotropic" chromosome in the spermatogonia of a few insect species. T.H. Montgomery referred to it as "chromosome *x*" (Montgomery, 1901). Similar studies of other species at the time also turned up a very small "idiochromosome", seen alongside the unpaired "chromosome *x*" in spermatogonia but not in oogonia (Wilson, 1906). This small chromosome would later be known as the Y chromosome. In these species where the two sexes had different chromosome compositions, a genetic mechanism was thought to be responsible for sexual dimorphism.

Nearly two decades later, T.S. Painter, while examining mitosis and meiosis in human testes, observed one poorly matched or heteromorphic pair of chromosomes in addition to evenly matched autosome pairs. Adopting the convention established by Montgomery, Painter referred to the larger of the two heteromorphic chromosomes as "X", and the smaller as "Y".

Since then, the XX:XY system of sex determination - females being homogametic or XX, and males heterogametic or XY - has been established for a large number of species. Yet this system is far from being universal. In many species, including *C. elegans*, no Y exists, so females have two X's and males have only one X. Birds use a different ZZ:ZW system of sex determination where the heterogametic sex ZW is the female. Even more

exotic schemes of chromosomal sex determination have been described (Kallman, 1968, Orzack *et al.*, 1980). In fact, in many species, the sex of an organism is not even genetically pre-determined. It relies instead on environmental cues. Environmental sex determination is so widespread in nature that it is perhaps one of the rules rather than an exception. This topic is reviewed extensively by H. Korpelainen (1990).

In species with the XX:XY system of sex determination, the role of the Y is also not consistent. In humans as in most other mammals, the Y dominantly triggers male differentiation. In *Drosophila*, it is the autosome to X ratio that triggers sex differentiation - a ratio of one or less results in females, a ratio of two or more results in males. In this system, the presence or absence of the Y does not affect sex.

Despite the variability of sex-determining mechanisms, X and Y chromosomes are found in a great number of distantly related species. Apparently, the two heteromorphic sex chromosomes have evolved into existence in multiple independent occasions (see next section for further discussion). In each case, the Y has converged upon a set of common features: 1) its transmission is restricted to males; 2) it pairs, albeit unevenly, with the X during meiosis; 3) it is much smaller than the X; and 4) it is genetically impoverished, with far fewer genes compared to autosomal regions of comparable size. Theorists now believe that due to male-restricted transmission, Y chromosomes in diverse species travel similar evolutionary paths, converging upon a set of common features. As will become obvious in Chapters 2 & 3, the theory of Y evolution proves to be a useful framework in which the biological roles of the chromosome are more readily interpreted.

Section ii: Evolution of the Y as a male-specific chromosome

In species with the XX:XY system of sex determination, the Y is the only chromosome that is transmitted exclusively through males. Theorists have argued that consequently, the Y travels a unique and predictable path of evolution.

H.J. Muller is credited for first postulating the idea (1914) that the two heteromorphic sex chromosomes X and Y, which differ so much in morphology and gene content, have nevertheless evolved from a once homomorphic pair of chromosomes. This idea had actually been suggested much earlier by E.B. Wilson (1906) when he first described the two heteromorphic sex chromosomes. This notion is now widely accepted. According to current, more sophisticated views, the XX:XY system of sex determination in a diploid organism is first established when one of a pair of autosomes acquires the sex-determining locus (*SDL*), which is sufficient to trigger the male pathway of sexual differentiation. The chromosome bearing *SDL* becomes the Y. Its pairing partner becomes the X. Individuals with two X's develop as females by a default pathway. Individuals with an X and a Y develop as males, dominantly controlled by *SDL* on the Y.

At an early stage, regions outside of *SDL* on the Y is thought to recombine freely with corresponding, homologous regions of the X. Due to close linkage, however, sequences immediately adjacent to *SDL* are rarely exchanged from Y to X. With time, these sequences diverge from their corresponding homologous regions on the X. Sequence divergence further inhibits recombination, causing sequences even further away from *SDL* to diverge. This process is thought to progress until 1) the Y becomes non-recombining over much of its length, and 2) sequences on the Y becomes male-specific.

More recently, J. A. Marshall Graves and colleagues proposed a modified view of Y chromosome evolution (Graves and Watson, 1991, Graves, 1995). This so called addition/attrition hypothesis states that autosomal regions are continually transposed onto sex chromosomes. These newly acquired regions initially pair and recombine between X and Y, but eventually proceed down the path of sequence divergence and suppression of recombination. Autosome to sex chromosome translocations have been reported for mammals [Graves, 1991 #260] and flies [Lucchesi, 1978 #289]. Another means by which the Y could acquire anew homology with the X is through direct X to Y transpositions. One such event was shown to have taken place within the primate lineage (Schwartz *et al.*,).

What happens to genes that the Y originally shared with the X? The popular belief is that these genes have a tendency to "degenerate" (a term used by theorists to mean the loss of function through either slow means such as reductions in gene expression, or more dramatic means such as frameshifts and deletions). This belief is largely based on the observation that Y chromosomes in many species appear to have a severe paucity of genes.

Theories have been put forward to account for Y's degeneration. The first, proposed by Wilson (1906), and later independently by Muller (1914, 1918), states that since these genes never cross over to the X, any recessive deleterious mutations they carry would be perpetually sheltered by their X homologs. Homozygozing of recessive mutations, which is an effective means for their elimination, is unavailable for these Y-linked genes. Accumulation of recessive mutations would lead eventually to their degeneration.

This theory was criticized by R.A. Fisher (1935) and B. Charlesworth (1978) on the following grounds. Genes on the X should acquire recessive deleterious mutations at about the same rate as their Y homologs. Therefore, Y-linked genes carrying recessive mutations would be eliminated in the same fashion when they encounter X homologs carrying similar mutations. Mathematical treatment of this problem by M. Nei (1970) showed that even though recessive mutations on the Y are fixed at a slightly higher rate compared to their X homologs, the effect is rather insignificant, unless the effective population size is very small. Moreover, the assumption that most deleterious mutations are truly recessive with no effect on fitness in heterozygotes may be too simplistic. The fact that many organisms acquire dosage compensation for their X chromosomes - a phenomenon where X-linked genes are differentially regulated in males and females such that they are expressed at comparable levels despite the two fold difference in copy numbers - argues that perhaps most mutations leading to defective proteins are not entirely recessive, though their effect on fitness in heterozygotes may be subtle.

An alternative theory, dubbed Muller's ratchet and first proposed by Muller (1964) to explain the evolutionary advantage of recombination, was re-engineered by Charlesworth

(1978) to account for the degeneration of Y-homologs of X-linked genes. In essence, the theory states that with recombination, an intact allele could be regenerated through crossovers between two mutated alleles each carrying a mutation at a different site within the locus. Without recombination, back-mutations are the only means to restore function to deleterious mutations. Yet, back-mutations are considered rare. For this reason, Y-linked genes would degenerate at a much greater rate compared to genes in the rest of the recombining genome, a situation that leads eventually to the obliteration of these genes.

W.R. Rice proposed a supplement to Muller's ratchet, called genetic hitchhiking (Rice, 1987). The term genetic hitchhiking was first used by J. Maynard Smith and J. Haigh (1974) to mean that a mutation negatively affecting fitness may persist in a population if it is closely linked to a locus with an allele that positively affects fitness. Rice argued that genetic hitchhiking could operate on the Y in conjunction with Muller's ratchet. In essence, when combinations of genes - either in intact allele forms or carrying mutations - can be shuffled by recombination, the combination with the least mutation load would be continually regenerated. Less favorable combinations would decrease in frequency as a result of reduced fitness. Through this process, deleterious mutations can be effectively weeded out. This process of regenerating gene combinations with the least mutation load is unavailable to the Y, as the allele of one gene is permanently stuck with the allele of another gene on the same chromosome. As a consequence, deleterious mutations on the Y are not weeded out as effectively. The genetic hitchhiking theory is an extension of Muller's ratchet in that it considers recombination - not only within a genetic locus but between loci - essential for the regeneration of a contiguous DNA fragment carrying the least mutation load.

As the degeneration of Y homologs of X-linked genes progresses, there is a need for the up-regulation of these X-linked genes in males (or down-regulation in females) to maintain a comparable level of expression between the sexes. A mechanism of dosage compensation may have evolved concurrently with the Y's degeneration. As genes on the Y

become completely non-functional, their X homologs would need to be expressed at twice the level from each copy in males as compared to females. In mammals, this is accomplished by X-inactivation, the silencing of an entire X chromosome in each female cell. The subject of X-inactivation is reviewed extensively by B.R. Migeon (1994).

Theorists have thus far focused on the degeneration of Y homologs of X-linked genes, and have given less attention to another class of genes on the Y - those without X-homologs, namely male-specific genes. In species with genetic, as opposed to environmental sex determination, males and females can be viewed as two genetic variants of the same organism. The controlling mechanism of their developmental differences can be addressed by the classical genetic method of phenotype-genotype correlation. Since males and females have identical autosomal content, the control of sexual dimorphism must lie in the difference of their sex chromosome composition. Sexual differentiation can be controlled either by the number of X chromosomes, as in *Drosophila*, or by the presence or absence of the Y, as in most mammals. In humans, it is unlikely that sexual distinctions are controlled by Y homologs of X-linked genes, since they are thought to complement their X counterparts, rather than providing novel functions. In contrast, male-specific genes are unique on the Y. Their presence in males and absence in females are perhaps the crucial genetic difference between the two sexes that results in distinct sexual phenotypes. Of course, this does not exclude the involvement of non-Y genes in sexually distinctive developmental processes.

Unlike X-homologous genes on the Y, the evolution of male-specific genes has not been rigorously modeled at a theoretical level. One pertinent idea was put forth by R.A. Fisher (1931). He recognized that there may exist genes or alleles of genes that favor the fitness of one sex but are inconsequential or even detrimental to the other, a situation he termed sexual antagonism. For sexually antagonistic genes that benefit males, selective pressure favors their accumulation in male-specific regions, namely the Y, such that females would never be exposed to their negative effects.

One apparent hole in this theory, as J.J. Bull pointed out (1983), is that female antagonistic genes could simply be shut off in female cells where their expression is unfavorable. Given the ease of gene regulation, which is readily accomplished for every gene that is differentially regulated during development, either spatially or temporally, the selective pressure for the accumulation of female antagonistic genes on the Y due a failure to silence them in females appears rather insignificant. Here I propose an alternative idea, which I refer to as selective furlough. Assume that a gene benefits males but is of no consequence to females. If this gene resides on an autosome, there would be no selective pressure when it is transmitted through females. A female carrying a mutated allele of this gene would suffer no consequence. This gene is therefore in a "selective furlough" during female transmission, which is half of its evolutionary history. If however, this gene resides on the Y, it would be under constant selection as it is transmitted from male to male. This amounts to elevated selective pressure, either to maintain the gene in its original form or to push it toward forms that impart greater fitness. A rigorous mathematical treatment may reveal the extent to which this elevated selective pressure would drive male beneficial genes either to move from autosomes to the Y, or to evolve *de novo* more readily on the Y than on autosomes.

Section iii: Structure and gene content of the human Y chromosome

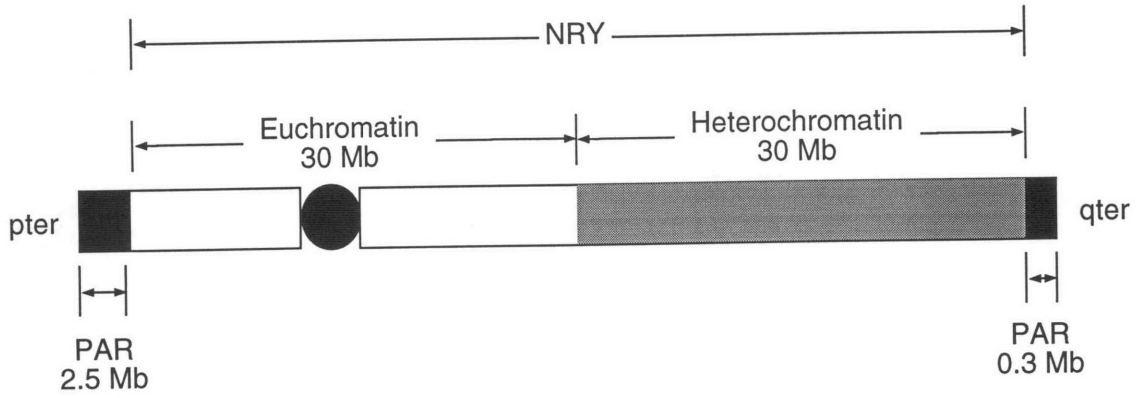
Theoretical treatments of the evolution and behavior of the Y chromosome apply to a wide range of species. Experimentally however, the human Y is by far the best studied.

The human Y was first described in 1923 by T.S. Painter (1923). Even though the association of sex with sex chromosome composition was firmly established by this time, it was not until 1959 that the dominant role of the human Y in triggering male differentiation was unveiled by reports of XO females and XXY males (Jacobs and Strong, 1959, Ford *et al.*, 1959). Subsequent decades would see an explosion of biological techniques, bringing

the knowledge of human chromosomes, including the Y, from visual observations to the molecular level.

The human Y is one of the smallest among the 24 chromosomes, measuring roughly 60 megabases, or 2% of the haploid genome. Fig. 1 is a schematic drawing of the chromosome. Table 1 lists all previously identified genes or pseudogenes.

The human Y can be divided into pseudoautosomal regions and the non-recombining region. The pseudoautosomal regions, or PARs, exist on both X and Y. The two sex chromosomes pair and freely recombine within PARs. As a result PAR sequences are identical between X and Y, and follow rules of genetic recombination and linkage like the autosome, hence the name *pseudo-autosomal.* G.A. Rappold has reviewed the human PARs comprehensively (1993). The human X and Y each have two PARs, situated on either end of each chromosome. A larger 2.5 megabase PAR is at the tip of Xp and Yp (Cooke *et al.*, 1985, Simmler *et al.*, 1985), and a smaller 0.3 megabase PAR is at the tip of Xq and Yq (Freije *et al.*, 1992). Genetic markers from PARs showed that the rate of recombination between PARs is significantly higher in males than in females. This difference is about 10-fold in the Xp/Yp PAR (Henke *et al.*, 1993), and about 5-fold in the Xq/Yq PAR (Freije *et al.*, 1992). Perhaps the formation of chiasma during meiosis is required between X and Y to ensure proper pairing and segregation of the two sex chromosomes (Henke *et al.*, 1993). The idea of obligatory chiasma between X and Y is consistent with the greatly elevated recombination rate between PARs in males. Moreover, disruptions in the mouse PAR lead to spermatogenic failure, consistent with a need for homologous pairing during meiosis (Burgoyne *et al.*, 1992). As reviewed in the previous section, theories of Y chromosome evolution suggest that crossover suppression spreads outward from *TDF*, until the entire chromosome stops recombining with the X. The presence of two PARs in humans is therefore an exception. The requirement of homologous pairing and crossover may preserve the PARs by counteracting crossover suppression. However, this requirement is far from general and should be taken with caution as an explanation for the persistence of



NRY: Non-recombining region of the Y
PAR: Pseudoautosomal region

Fig. 1. The human Y chromosome.

Table 1. Previously itentified Y genes and pseudogenes

Gene name	Reference
Genes in 1st PAR	
CSF2RA	Gough et al., 1990
IL3RA (hIL-3Ra)	Milatovich et al., 1993
ANT3 (T2)	Slim et al., 1993
ASMT (HIOMT)	Yi et al., 1993
XE7	Ellison et al., 1992
MIC2	Goodfellow 1986
SHOX	Rao et al., 1997
Genes in NRY	
XGPY: pseudogene	Weller et al., 1995
SRY	Sinclair et al., 1990
RPS4Y	Fisher et al., 1990
ZFY	Page et al., 1987
TSPY	Arnemann et al., 1991
PKY: uncharacterized	Klink et al., 1995
AMELY	Salido et al., 1992
KALP: pseudogene	del Castillo et al., 1992
STSP: pseudogene	Yen et al., 1988
HYA (SMCY)	Agulnik et al., 1994
RBM (YRRM)	Ma et al., 1993
DAZ	Reijo et al., 1995
Gene in 2nd PAR	
SYBL1	D'Esposito et al., 1996
IL9R	Kermouni et al., 1995

Note: PAR: pseudoautosomal region;
NRY: non-recombining region of the Y

PARs. Marsupial X and Y for example, are thought to lack PARs. They do line up at the tips during meiosis, but do not appear to undergo recombination (Sharp, 1982).

Even though PARs constitute only 5% of the Y, they contain as many genes thus far identified as the rest of the chromosome (Table 1). These genes appear as a typical sampling of the genome, with varying patterns of expression and diverse functions.

In humans, the two PARs flank a central, non-recombining region of the Y, or NRY. The NRY can be divided about equally into a euchromatic half and a heterochromatic half. The heterochromatic region, the largest contiguous stretch of repetitive DNA in the genome, gives the Y its characteristic bright tail during metaphase when stained with quinacrine. No gene has been found in the heterochromatin.

The euchromatic portion of the NRY has been extensively studied. It was one of the first large chromosomal regions for which an STS map was constructed based on naturally occurring Y deletions in people and a complete YAC contig (Vollrath *et al.*, 1992, Foote *et al.*, 1992). As anticipated, the STS map and the YAC contig have been of tremendous benefit for subsequent studies. They also provided the basis for dividing the euchromatic portion of the NRY into three types of regions: Y-specific single copy regions, Y-specific repeats, and X homologous regions (Foote *et al.*, 1992).

Y-specific single copy regions represent only a minority of NRY sequences, but contain the majority of genes thus far identified within the NRY.

Y-specific repeats, as the name suggests, are specific to the Y, but are present in multiple, closely related copies. The prevalence of this type of sequences is unique to the Y. Theorists have argued that the lack of recombination can result in a unique tendency for repetitive elements to accumulate on the Y. Since the NRY does not recombine, rearrangements (*i.e.*, duplication and inversion) within this region, which on autosomes cause meiotic non-disjunction, are faithfully transmitted through meioses. Consequently, repetitive sequences in shuffled arrangements can accumulate. Genes found so far in these repetitive sequences, *TSPY*, *RBM* and *DAZ* are themselves repeated.

X homologous regions share a high degree of homology with the X. The best studied is the so called DXYS1 like region. It shares over 99% sequence homology with the X, and was shown to have transposed from X to Y within the primate lineage (Schwartz *et al.*,). No genes have yet been found within X homologous regions. Any genes that may be present are expected to be highly homologous to their X counterparts. X homologous regions should not be confused with Y copies of X/Y homologous genes. The former constitute very large stretches, often over megabases of high homology with the X; the latter are Y homologs of X-linked genes, residing in regions that are otherwise Y-specific.

Eight genes were previously cloned from the NRY (Table 1), of which *SRY* is the long sought *TDF*, the master switch for male differentiation. In addition, there are four X/Y homologous genes, namely, genes with close homologs on the X. They are *RPS4Y*, *ZFY*, *AMELY* and *SMCY*. There are three male-specific genes (genes with no X homologs): *TSPY* and *RBM*, which do not have close autosomal homologs, and *DAZ* which has a close homolog on chromosome 3 (Saxena *et al.*, 1996).

The four Y copies of X/Y homologous genes represent vestiges of the ancient Y chromosome, and are exceptions to the rule of Y degeneration. Their persistence can be accounted for in two ways: 1) they may resist degeneration due to the need for double gene dosage; or 2) they were recent acquisitions on the Y. This question will be addressed in detail in Chapter 3, where examples of both appear to exist.

The two male-specific genes, *TSPY* and *RBM*, are both expressed in the adult testis, which is consistent with their postulated male-specific functions. *DAZ* may represent a recent acquisition within the primate lineage by the Y of an autosomal gene with male-specific function (Saxena *et al.*, 1996). The testis-limited expression of both *DAZ* and its chromosome 3 homolog *DAZL* is consistent with their male-specific role. Moreover, *DAZ* is implicated in male infertility. A substantial fraction of men who suffer spermatogenic failure are deleted *de novo* for a region of the Y that contains *DAZ*, suggesting that *DAZ* and perhaps *DAZL* function in spermatogenesis (Reijo *et al.*, 1995). The role of Y-linked genes

in male fertility will be treated in depth in the next section. Despite the lack of rigorous theoretical treatments of how genes controlling male-specific biology may emerge and evolve on the Y, the idea of sexual antagonism proposed by R.A. Fisher and the idea of selective furlough I proposed in the previous section at least raise the argument that perhaps male-specific genes are attracted to the Y to take advantage of its male-restricted transmission. It was recently suggested that *TSPY* and *RBM* may also have been acquired by the Y from autosomes, but evidence is less compelling as these genes have diverged a great deal from their putative autosomal homologs (Delbridge *et al.*, 1997).

Section iv: The role of the human Y in spermatogenesis

As discussed in Section ii of this chapter, male-specific genes on the Y are responsible for controlling sexually distinctive features. Like most sexually reproducing species, human males and females share virtually all developmental processes with some well-defined exceptions. Gonads and their supporting structures are vastly different between men and women. The remaining somatic differences are a collection of sexual features, some obvious (e.g., mammary glands, bone structure, muscle mass, voice and body hair), and some subtle and perhaps even controversial (e.g., temperament, aggressiveness, and other psychological traits). These secondary characteristics are modulated by hormones. In contrast to somatic features, the development of germ cells - spermatogenesis in men and oogenesis in women - are highly dissimilar.

In mammals, *SRY* is believed to be responsible for most if not all somatic sexual distinctions. In humans for example, there are 46,XX males who carry a small fragment of Y DNA (including *SRY*) on one of their X's. These men are infertile, but are otherwise normal (de la Chapelle *et al.*, 1984, Andersson *et al.*, 1986, Page *et al.*, 1987a, Petit *et al.*, 1987). In fact, many of them first seek medical attention for their infertility. In the mouse, 40,XX animals which carry an SRY transgene develop into somatically normal but infertile males (Koopman *et al.*, 1991). Apparently *SRY* alone is sufficient to trigger most if not all

aspects of somatic male differentiation. By the process of elimination, other male-specific genes on the Y are most likely involved in male-specific processes in the germline.

Experimental evidence for Y's role in spermatogenesis was first uncovered by L. Tiepolo and O. Zuffardi (1976) when they karyotyped infertile men and saw microscopic deletions of distal Yq in some It would be nearly two decades later when higher-resolution molecular evidence defined at least three distinct regions on the Y where deletions resulted in spermatogenic failure (Ma *et al.*, 1993, Reijo *et al.*, 1995, Vogt *et al.*, 1996, Pryor *et al.*, 1997). Putative genes responsible for the phenotype were called *Azoospermia Factors,* or *AZF*. This term is somewhat misleading, since azoospermia means "spermless" yet many men with deletions have reduced but non-zero sperm counts, a condition more correctly referred to as oligospermia. Two genes, *RBM* and *DAZ*, have thus far been proposed as *AZF* candidates (Ma *et al.*, 1993, Reijo *et al.*, 1995). *RBM* exists in multiple copies dispersed on the Y. It has no close homolog outside of Y. Like *RBM*, *DAZ* also exists in multiple copies, but in a local cluster. It has a close homolog *DAZL* on chromosome 3, and as discussed in the previous section, is believed to have transposed onto the Y and subsequently amplified, perhaps as a result of its important role in spermatogenesis (Saxena *et al.*, 1996).

Perhaps not coincidentally, the importance of the Y in spermatogenesis is not limited to humans. Mouse Y is essential for male fertility (Levy and Burgoyne, 1986, Burgoyne *et al.*, 1992, Conway *et al.*, 1994), as is the Y in *Drosophila* (Hardy *et al.*, 1981).

Of course, Y genes with X homologs cannot be ruled out as *AZF* candidates. In fact, it is even possible that X/Y homologous genes with essential roles in spermatogenesis are more resistant to degeneration. Perhaps the unique biology of male germ cells requires the maintenance of two functional copies of some X/Y homologous genes. The germline is where competition for survival occurs not only among organisms but among germ cells within a single organism. Upon meiosis I, germ cells each assume a different haplotype. Even though germ cells form syncytia, with a shared cytoplasm (Braun *et al.*, 1989), cellular

resources in a syncytium may not be evenly distributed, and a germ cell's haplotype may affect its transmission. In the extreme, germ cells carrying a certain haplotype display meiotic drive - they out-compete others by a large margin, resulting in distortion of Mendelian segregation (Lyttle, 1991). A well-studied case of meiotic drive in male germ cells is the mouse autosomal *T* locus. Male mice with the so called *t* allele and a wild type allele of the *T* locus (*t*/+) transmit the *t* allele to >99% of their offspring (Silver, 1993). If at least one copy of some X/Y homologous genes is required in the haplo-stage of spermatogenesis, there would be strong selection against the degeneration of the Y copy.

Section v: The human Y and Turner syndrome

Turner syndrome (TS), described by H.H. Turner (1938), is a congenital condition confined to females. Classic TS features include very poor *in utero* viability (estimated at 1%), short stature, ovarian dysgenesis, failure to develop secondary sexual characteristics, and a number of anatomical abnormalities. Two decades after Turner's description of the syndrome, cytogenetic studies by C.E. Ford and colleagues (1959), and later by others, associated TS with complete or partial monosomy of sex chromosomes. Abnormal karyotypes typically include 45,XO; 46,X,derivative(X); or 46,X,derivative(Y), found either in all the cells of a patient (nonmosaic), or some fraction of the cells (mosaic). Partial monosomy of sex chromosomes has since been an integral part of the diagnostic definition of TS. Occasionally, there are cases that present all or some of TS features, but with an apparently normal karyotype. The molecular nature of TS has been reviewed by A.R. Zinn and colleagues (1993).

Despite poor *in utero* viability, 45,XO is the only human monosomy that occasionally survives beyond the embryonic stage. This is very likely due to X-inactivation - the silencing of one of the two X chromosomes in female cells - which leads to dosage equality of X-linked genes at the level of expression among males, females and 45,XO individuals. However, the fact that 45,XO individuals are not completely normal, but rather

develop TS features, poses a paradox for X-inactivation. If silencing of one X in females were complete, 46,XX and 45,XO should be functionally equivalent. Even though one can argue that disruption of proper meiotic pairing in 45,XO individuals may lead to oogenic failure, as some experiments had suggested (Burgoyne *et al.*, 1992), it can not account for somatic features of TS. One has to postulate that X-inactivation is incomplete, and that genes implicated in TS escape X-inactivation. A caveat is that males, who only have one X, develop normally. To express an equal dosage in males as in females for the genes that escape X-inactivation, the Y chromosome has to carry the very same genes. These so called Turner genes - escaping X-inactivation in females and having functional Y homologs in males - were proposed by M.A. Ferguson-Smith (1965) long before their presence was experimentally confirmed. Based on his model, expression from two functional copies of Turner genes is essential for development. TS is a consequence of haploinsufficiency of these genes - the half-normal expression from only one functional copy.

All genes residing in PARs are potential Turner genes: they are on both X and Y and typically escape X-inactivation in females (Goodfellow *et al.*, 1986, Ellison *et al.*, 1992b, Slim *et al.*, 1993). Recently, a new gene, *SHOX* was identified on Xp/Yp PAR. Interstitial deletions of regions encompassing *SHOX* resulted in short stature. *SHOX* was postulated as a Turner gene, responsible at least partially for short stature, which is one of TS features (Rao *et al.*, 1997). There could still be other Turner genes in PARs.

The NRY is also implicated in TS. There are sex-reversed XY females with point mutations in *SRY* that presumably inactivate the protein. They have none of the somatic features of TS. XY females who have interstitial deletions that delete *SRY*, however, almost always show one or more somatic features of TS (Blagowidow *et al.*, 1989, Levilliers *et al.*, 1989, Page, Unpublished data). Deletions in these individuals apparently include Turner genes in addition to *SRY*. In the NRY, there indeed exist a class of genes that fit the criteria for Turner genes. X/Y homologous genes have close and presumably functionally comparable homologs on X and Y. X copies, where tested, escape X inactivation

(Schneider-Gädicke *et al.*, 1989, Fisher *et al.*, 1990, Agulnik *et al.*, 1994, Jones *et al.*, 1996). But none of the X/Y homologous genes are yet definitively correlated with TS.

Section vi: Chromosomal abnormalities of the human Y

Chromosomal abnormalities of the Y are frequently observed in the human population, perhaps because the Y contains few genes essential for survival, . There are two major classes: 1) re-arrangements involving the Y and another chromosome; and 2) re-arrangements involving only the Y.

The most common re-arrangements involving the Y and another chromosome are translocations between X and Y. Like autosome pairs, X and Y pair and undergo recombination within PARs. Their close proximity during meiosis, though necessary for proper segregation, may also facilitate a high frequency of aberrant crossovers between the two sex chromosomes. A large fraction of sex reversed individuals - XX males and XY females (Levilliers *et al.*, 1989) - arise from aberrant X/Y crossovers. These aberrant crossovers typically occur between short stretches of homologous sequences. These short stretches can be repetitive elements like Alu's (Rouyer *et al.*, 1987), or more unique sequences (Weil *et al.*, 1994). In the case of XX males, the terminal portion of Xp of one of the X's is replaced by the terminal portion of Yp which carries *SRY*, apparently through a Yp to Xp translocation (de la Chapelle *et al.*, 1984, Andersson *et al.*, 1986, Page *et al.*, 1987a, Petit *et al.*, 1987). Reciprocal products of this aberrant exchange are frequently found in XY females. Their terminal Yp, which normally carries *SRY*, is replaced by a translocated fragment of terminal Xp (Weil *et al.*, 1994).

Re-arrangements involving only the Y may occur between two sister Y chromosomes during cell division. One common class of abnormal Y, the so called isodicentric Y, are a result of aberrant crossovers between two Y chromosomes. Isodicentric chromosomes can be viewed as mirror-image chromosome fragments attached at the point of symmetry. Both isodicentric Yp and Yq chromosomes have been observed. For an

isodicentric Yp chromosome for example, the linear structure is as follows: pter-centromere-[a portion of Yq]-[point of symmetry]-[a portion of Yq]-centromere-pter. These chromosomes carry two centromeres, hence the term *iso-di-centric*. Isodicentric Y is found most commonly in Turner syndrome (TS) patients. Almost invariably, isodicentric Y cells exist in mosaicism with XO cells, for which the simplest explanation is that a chromosome with two centromeres is unstable during cell division and is often lost. Another type of abnormal Y found in some TS patients is the so called ring Y. They most likely arise from aberrant crossovers between two points of the same Y chromosome, one on the long arm and one on the short arm. Like isodicentric Y, ring Y cells frequently exist in mosaicism with XO cells. The ring structure perhaps inhibits proper segregation during cell division and results in frequent loss of the chromosome.

Terminal deletions are very common on the Y. Deletions of the long arm (Yq-) are much more common than deletions of the short arm (Yp-). In fact, 46,XYq- karyotype is among the most common microscopically detectable chromosome disorders, with a rate of occurrence estimated at 0.1% (Hamerton *et al.*, 1975). Since most terminal deletions of the Y are detected with low resolution microscopy, one cannot rule out the possibility that apparent Yq- or Yp- chromosomes could carry small translocations of other chromosomes. Moreover, a chromosome suffering from a true terminal deletion lacks a telomere, which is essential for its stability. This argues that either there is a yet to be characterized mechanism for telomere regeneration, or that the apparent terminal deletion is in fact 1) a translocation with another chromosome; or 2) an interstitial deletion with one breakpoint near the end of the chromosome. These possibilities are thoroughly explored in Chapter 4.

Chromosomal abnormalities of the Y have also played a crucial role in continuous efforts to build molecular maps of the chromosome, which culminated in the construction of an STS-based map with 43 intervals across the euchromatic portion of the NRY (Vergnaud *et al.*, 1986, Affara *et al.*, 1986, Oosthuizen *et al.*, 1990, Nakahori *et al.*, 1991, Kotecki *et al.*,

1991, Bardoni *et al.*, 1991, Vollrath *et al.*, 1992). Re-arrangements of the Y used in the studies included isodicentric Y, ring Y, Yq-, autosome/Y translocation, and X/Y translocation.

With the advent of Y maps based on molecular markers (Vollrath *et al.*, 1992, Foote *et al.*, 1992), large deletions are now routinely detected either by hybridization markers, or more commonly, STS based PCR markers. These techniques also allow for the detection of smaller, interstitial deletions of the Y that are undetectable by microscopy. Molecular detection of small deletions on the Y have uncovered at least three distinct regions essential for spermatogenesis (Ma *et al.*, 1993, Reijo *et al.*, 1995, Vogt *et al.*, 1996, Pryor *et al.*, 1997).

In summary, chromosomal abnormalities are essential tools for deciphering biological functions of the Y, especially in sex determination and spermatogenesis. With the ongoing discovery of genes and the construction of higher resolution maps, abnormalities of the Y will continue to be an important tool.

References

Affara, N. A., Florentin, L., Morrison, N., Kwok, K., Mitchell, M., Cook, A., Jamieson, D., Glasgow, L., Meredith, L., Boyd, E. and 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.

Agulnik, A. I., Mitchell, M. J., Mattei, M. G., Borsani, G., Avner, P. A., Lerner, J. L. and Bishop, C. E. (1994). A novel X gene with a widely transcribed Y-linked homologue escapes X-inactivation in mouse and human. *Hum Mol Genet* **3:** 879-884.

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.

Arnemann, J., Jakubiczka, S., Thuring, S. and Schmidtke, J. (1991). Cloning and sequence analysis of a human Y-chromosome-derived, testicular cDNA, TSPY. *Genomics* **11:** 108-114.

Bardoni, B., Zuffardi, O., Guioli, S., Ballabio, A., Simi, P., Cavalli, P., Grimoldi, M. G., Fraccaro, M. and 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.

Blagowidow, N., Page, D. C., Huff, D. and Mennuti, M. T. (1989). Ullrich-Turner syndrome in an XY female fetus with deletion of the sex-determining portion of the Y chromosome. *Am J Med Genet* **34:** 159-162.

Braun, R. E., Behringer, R. R., Peschon, J. J., Brinster, R. L. and Palmiter, R. D. (1989). Genetically haploid spermatids are phenotypically diploid. *Nature* **337:** 373-376.

Bull, J. J. (1983). Evolution of Sex Determining Mechanisms: Benjamin Cummings, Menlo Park, CA.

Burgoyne, P. S., Mahadevaiah, S. K., Sutcliffe, M. J. and Palmer, S. J. (1992). Fertility in mice requires X-Y pairing and a Y-chromosomal "spermiogenesis" gene mapping to the long arm. *Cell* **71:** 391-398.

Charlesworth, B. (1978). Model for evolution of Y chromosomes and dosage compensation. *Proc Natl Acad Sci U S A* **75:** 5618-5622.

Conway, S. J., Mahadevaiah, S. K., Darling, S. M., Capel, B., Rattigan, A. M. and Burgoyne, P. S. (1994). Y353/B: a candidate multiple-copy spermiogenesis gene on the mouse Y chromosome. *Mamm Genome* **5:** 203-210.

Cooke, H. J., Brown, W. R. and Rappold, G. A. (1985). Hypervariable telomeric sequences from the human sex chromosomes are pseudoautosomal. *Nature* **317:** 687-692.

D'Esposito, M., Ciccodicola, A., Gianfrancesco, F., Esposito, T., Flagiello, L., Mazzarella, R., Schlessinger, D. and D'Urso, M. (1996). A synaptobrevin-like gene in the Xq28 pseudoautosomal region undergoes X inactivation. *Nat Genet* **13:** 227-229.

de la Chapelle, A., Tippett, P. A., Wetterstrand, G. and Page, D. (1984). Genetic evidence of X-Y interchange in a human XX male. *Nature* **307:** 170-171.

del Castillo, I., Cohen-Salmon, M., Blanchard, S., Lutfalla, G. and Petit, C. (1992). Structure of the X-linked Kallmann syndrome gene and its homologous pseudogene on the Y chromosome. *Nat Genet* **2:** 305-310.

Delbridge, M. L., Harry, J. L., Toder, R., O'Neill, R. J., Ma, K., Chandley, A. C. and Graves, J. A. (1997). A human candidate spermatogenesis gene, RBM1, is conserved and amplified on the marsupial Y chromosome. *Nat. Genet.* **15:** 131-136.

Ellison, J., Passage, M., Yu, L. C., Yen, P., Mohandas, T. K. and Shapiro, L. (1992a). Directed isolation of human genes that escape X inactivation. *Somat Cell Mol Genet* **18:** 259-268.

Ellison, J. W., Ramos, C., Yen, P. H. and Shapiro, L. J. (1992b). Structure and expression of the human pseudoautosomal gene XE7. *Hum Mol Genet* **1:** 691-696.

Ferguson-Smith, M. A. (1965). Karyotype-phenotype correlations in gonadal dysgenesis and their bearing on the pathogenesis of malformations. *J. Med. Genet.* **2:** 142-155.

Fisher, E. M., Beer-Romero, P., Brown, L. G., Ridley, A., McNeil, J. A., Lawrence, J. B., Willard, H. F., Bieber, F. R. and Page, D. C. (1990). Homologous ribosomal protein genes on the human X and Y chromosomes: escape from X inactivation and possible implications for Turner syndrome. *Cell* **63:** 1205-1218.

Fisher, R. A. (1931). The evolution of dominance. *Biol Rev* **6:** 345-368.

Fisher, R. A. (1935). The sheltering of lethals. *American Naturalist* **69:** 446-455.

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.

Ford, C. E., Jones, K. W., Polani, P. E., Almida, J. C. and Briggs, J. H. (1959). A sex-chromosome anomaly in the case of gonadal dysgenesis (Turner's syndrome). *Lancet* **i:** 711-713.

Freije, D., Helms, C., Watson, M. S. and Donis-Keller, H. (1992). Identification of a second pseudoautosomal region near the Xq and Yq telomeres. *Science* **258:** 1784-1787.

Goodfellow, P. J., Darling, S. M., Thomas, N. S. and Goodfellow, P. N. (1986). A pseudoautosomal gene in man. *Science* **234:** 740-743.

Gough, N. M., Gearing, D. P., Nicola, N. A., Baker, E., Pritchard, M., Callen, D. F. and Sutherland, G. R. (1990). Localization of the human GM-CSF receptor gene to the X-Y pseudoautosomal region. *Nature* **345:** 734-736.

Graves, J. A. (1995). The origin and function of the mammalian Y chromosome and Y-borne genes--an evolving understanding. *Bioessays* **17:** 311-320.

Graves, J. A. and Watson, J. M. (1991). Mammalian sex chromosomes: evolution of organization and function. *Chromosoma* **101:** 63-68.

Hamerton, J. L., Canning, N., Ray, M. and Smith, S. (1975). A cytogenetic survey of 14,069 newborn infants. I. Incidence of chromosome abnormalities. *Clin Genet* **8:** 223-243.

Hardy, R. W., Tokuyasu, K. T. and Lindsley, D. L. (1981). Analysis of spermatogenesis in Drosophila melanogaster bearing deletions for Y-chromosome fertility genes. *Chromosoma* **83:** 593-617.

Henke, A., Fischer, C. and Rappold, G. A. (1993). Genetic map of the human pseudoautosomal region reveals a high rate of recombination in female meiosis at the Xp telomere. *Genomics* **18:** 478-485.

Jacobs, P. A. and Strong, J. A. (1959). A case of human intersexuality having a possible XXY sex-determining mechanism. *Nature* **183:** 302.

Jones, M. H., Furlong, R. A., Burkin, H., Chalmers, J., Brown, G. M., Khwaja, O. and Affara, N. A. (1996). The Drosophila developmental gene fat facets has a human homologue in Xp11.4 which escapes X-inactivation and has related sequences on Yq11.2. *Hum Mol Genet* **5:** 1695-1701.

Kallman, K. D. (1968). Evidence for the existence of transformer genes for sex in the teleost Xiphophorus maculatus. *Genetics* **60:** 811-828.

Kermouni, A., Van Roost, E., Arden, K. C., Vermeesch, J. R., Weiss, S., Godelaine, D., Flint, J., Lurquin, C., Szikora, J. P., Higgs, D. R. and et al. (1995). The IL-9 receptor gene (IL9R): genomic structure, chromosomal localization in the pseudoautosomal region of the long arm of the sex chromosomes, and identification of IL9R pseudogenes at 9qter, 10pter, 16pter, and 18pter. *Genomics* **29:** 371-382.

Klink, A., Schiebel, K., Winkelmann, M., Rao, E., Horsthemke, B., Ludecke, H. J., Claussen, U., Scherer, G. and 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.

Koopman, P., Gubbay, J., Vivian, N., Goodfellow, P. and Lovell-Badge, R. (1991). Male development of chromosomally female mice transgenic for Sry. *Nature* **351:** 117-121.

Korpelainen, H. (1990). Sex ratios and conditions required for environmental sex determination in animals. *Biol Rev Camb Philos Soc* **65:** 147-184.

Kotecki, M., Jaruzelska, J., Skowronska, M. and Fichna, P. (1991). Deletion mapping of interval 6 of the human Y chromosome. *Hum Genet* **87:** 234-236.

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 U S A* **86:** 2296-2300.

Levy, E. R. and Burgoyne, P. S. (1986). The fate of XO germ cells in the testes of XO/XY and XO/XY/XYY mouse mosaics: evidence for a spermatogenesis gene on the mouse Y chromosome. *Cytogenet Cell Genet* **42:** 208-213.

Lyttle, T. W. (1991). Segregation distorters. *Annu Rev Genet* **25:** 511-557.

Ma, K., Inglis, J. D., Sharkey, A., Bickmore, W. A., Hill, R. E., Prosser, E. J., Speed, R. M., Thomson, E. J., Jobling, M., Taylor, K. and 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.

Maynard Smith, J. and Haigh, J. (1974). The hitchhiking effect of a favorable gene. *Genetic Research* **23:** 23-35.

Migeon, B. R. (1994). X-chromosome inactivation: molecular mechanisms and genetic consequences. *Trends Genet* **10:** 230-235.

Milatovich, A., Kitamura, T., Miyajima, A. and Francke, U. (1993). Gene for the alpha-subunit of the human interleukin-3 receptor (IL3RA) localized to the X-Y pseudoautosomal region. *Am J Hum Genet* **53:** 1146-1153.

Montgomery, T. H. (1901). A study of the chromosomes of the germ-cells of *Metazoa. Transcripts of the American Philosophical Society.*

Muller, H. J. (1914). A gene for the fourth chromosome of Drosophila. *J Exp Zool* **17:** 325-336.

Muller, H. J. (1918). Genetic variability, twin hybrids and constant hybrids, in a case of balanced lethal factors. *Genetics* **3:** 422-499.

Muller, H. J. (1964). The relation of recombination to mutational advance. *Mutation Research* **1:** 2-9.

Nakahori, Y., Tamura, T., Nagafuchi, S., Fujieda, K., Minowada, S., Fukutani, K., Fuse, H., Hayashi, K., Kuroki, Y., Fukushima, Y. and et al. (1991). Molecular cloning and mapping of 10 new probes on the human Y chromosome. *Genomics* **9:** 765-769.

Nei, M. (1970). Accumulation of nonfunctional genes on sheltered chromosomes. *American Naturalist* **104:** 311-322.

Oosthuizen, C. J., Herbert, J. S., Vermaak, L. K., Brusnicky, J., Fricke, J., du Plessis, L. and Retief, A. E. (1990). Deletion mapping of 39 random isolated Y-chromosome DNA fragments. *Hum Genet* **85:** 205-210.

Orzack, S. H., Sohn, J. J., Kallman, K. D., A., L. S. and Johnston, R. (1980). Maintenance of three sex chromosome polymorphism in the platyfish, *Xiphophorus maculatus. Evolution* **34:** 663-672.

Page, D. C. (Unpublished data). .

Page, D. C., Brown, L. G. and de la Chapelle, A. (1987a). Exchange of terminal portions of X- and Y-chromosomal short arms in human XX males. *Nature* **328:** 437-440.

Page, D. C., Mosher, R., Simpson, E. M., Fisher, E. M., Mardon, G., Pollack, J., McGillivray, B., de la Chapelle, A. and Brown, L. G. (1987b). The sex-determining region of the human Y chromosome encodes a finger protein. *Cell* **51:** 1091-1104.

Painter, T. S. (1923). Studies in mammalian spermatogenesis. *J Exp Zool* **37:** 291-321.

Paulmier, F. C. (1899). The spermatogenesis of *Anasa tristis. Journal of Morphology.*

Petit, C., de, l. C. 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.

Pryor, J. L., Kent-First, M., Muallem, A., Van Bergen, A. H., Nolten, W. E., Meisner, L. and Roberts, K. P. (1997). Microdeletions in the Y chromosome of infertile men. *New England J. Med.* **336:** 534-539.

Rao, E., Weiss, B., Fukami, M., Rump, A., Niesler, B., Mertz, A., Muroya, K., Binder, G., Kirsch, S., Winkelmann, M., Nordsiek, G., Heinrich, U., Breuning, M. H., Ranke, M. B., Rosenthal, A., Ogata, T. and Rappold, G. A. (1997). Pseudoautosomal deletions encompassing a novel homeobox gene cause growth failure in idiopathic short stature and Turner syndrome. *Nature Genetics* **16:** 54-63.

Rappold, G. A. (1993). The pseudoautosomal regions of the human sex chromosomes. *Hum Genet* **92:** 315-324.

Reijo, R., Lee, T. Y., Salo, P., Alagappan, R., Brown, L. G., Rosenberg, M., Rozen, S., Jaffe, T., Straus, D., Hovatta, O. and 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.

Rice, W. R. (1987). Genetic hitchhiking and the evolution of reduced genetic activity of the Y sex chromosome. *Genetics* **116:** 161-167.

Rouyer, F., Simmler, M. C., Page, D. C. and Weissenbach, J. (1987). A sex chromosome rearrangement in a human XX male caused by Alu-Alu recombination. *Cell* **51:** 417-425.

Salido, E. C., Yen, P. H., Koprivnikar, K., Yu, L. C. and Shapiro, L. J. (1992). The human enamel protein gene amelogenin is expressed from both the X and the Y chromosomes. *Am J Hum Genet* **50:** 303-316.

Saxena, R., Brown, L. G., Hawkins, T., Alagappan, R. K., Skaletsky, H., Reeve, M. P., Reijo, R., Rozen, S., Dinulos, M. B., Disteche, C. M. and Page, D. C. (1996). The DAZ gene cluster on the human Y chromosome arose from an autosomal gene that was transposed, repeatedly amplified and pruned. *Nat Genet* **14:** 292-299.

Schneider-Gädicke, A., Beer-Romero, P., Brown, L. G., Nussbaum, R. and Page, D. C. (1989). ZFX has a gene structure similar to ZFY, the putative human sex determinant, and escapes X inactivation. *Cell* **57:** 1247-1258.

Schwartz, A., Chan, D., Brown, L. G., Alagappan, R., Pettay, D., Disteche, C., McGillivray, B., de la Chapelle, A. and Page, D. C. Reconstructing hominid Y evolution: X-homologous block, created by X-Y transposition, was disrupted by Yp inversion through LINE-LINE recombination. *Submitted.*

Sharp, P. (1982). Sex chromosome pairing during male meiosis in marsupials. *Chromosoma* **86:** 27-47.

Silver, L. M. (1993). The peculiar journey of a selfish chromosome: mouse t haplotypes and meiotic drive. *Trends Genet* **9:** 250-254.

Simmler, M. C., Rouyer, F., Vergnaud, G., Nystrom-Lahti, M., Ngo, K. Y., de la Chapelle, A. and Weissenbach, J. (1985). Pseudoautosomal DNA sequences in the pairing region of the human sex chromosomes. *Nature* **317:** 692-697.

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. and Goodfellow, P. N. (1990). A gene from the

human sex-determining region encodes a protein with homology to a conserved DNA-binding motif. *Nature* **346:** 240-244.

Slim, R., Levilliers, J., Ludecke, H. J., Claussen, U., Nguyen, V. C., Gough, N. M., Horsthemke, B. and Petit, C. (1993). A human pseudoautosomal gene encodes the ANT3 ADP/ATP translocase and escapes X-inactivation. *Genomics* **16:** 26-33.

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.

Turner, H. H. (1938). A syndrome of infantilism, congenital webbed neck, and cubitus valgus. *Endocrinology* **23:** 566-574.

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.

Vogt, P. H., Edelmann, A., Kirsch, S., Henegariu, O., Hirschmann, P., Kiesewetter, F., Köhn, F. M., Schill, W. B., Farah, S., Ramos, C., Hartmann, M., Hartschuh, W., Meschede, D., Behre, H. M., Castel, A., Nieschlag, E., Weidner, W., Gröne, H.-J., Jung, A., Engel, W. and Haidl, G. (1996). Human Y chromosome azoospermia factor (AZF) mapped to different subregions in Yq11. *Hum Mol Genet* **5:** 933-943.

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.

Wang, W., Meadows, L. R., den Haan, J. M., Sherman, N. E., Chen, Y., Blokland, E., Shabanowitz, J., Agulnik, A. I., Hendrickson, R. C., Bishop, C. E. and et al. (1995). Human H-Y: a male-specific histocompatibility antigen derived from the SMCY protein. *Science* **269:** 1588-1590.

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.

Weller, P. A., Critcher, R., Goodfellow, P. N., German, J. and Ellis, N. A. (1995). The human Y chromosome homologue of XG: transcription of a naturally truncated gene. *Hum Mol Genet* **4:** 859-868.

Wilson, E. B. (1906). Studies on chromosomes. III. The sexual difference of chromosome-groups in Hemiptera, with some consideration on the determination and inheritance of sex. *J Exp Zool* **2:** 507-545.

Yen, P. H., Marsh, B., Allen, E., Tsai, S. P., Ellison, J., Connolly, L., Neiswanger, K. and Shapiro, L. J. (1988). The human X-linked steroid sulfatase gene and a Y-encoded pseudogene: evidence for an inversion of the Y chromosome during primate evolution. *Cell* **55:** 1123-1135.

Yi, H., Donohue, S. J., Klein, D. C. and McBride, O. W. (1993). Localization of the hydroxyindole-O-methyltransferase gene to the pseudoautosomal region: implications for mapping of psychiatric disorders. *Hum Mol Genet* **2:** 127-131.

Zinn, A. R., Page, D. C. and Fisher, E. M. (1993). Turner syndrome: the case of the missing sex chromosome. *Trends Genet* **9:** 90-93.

CHAPTER 2: FUNCTIONAL COHERENCE OF THE HUMAN Y CHROMOSOME

(This chapter is adapted from a manuscript for publication, with *Science* format.)

Original title of manuscript:

Functional Coherence of the Human Y Chromosome

Bruce T. Lahn and David C. Page

Howard Hughes Medical Institute, Whitehead Institute, and Department of Biology, Massachusetts Institute of Technology, 9 Cambridge Center, Cambridge, Massachusetts 02142, USA

A systematic search of the non-recombining region of the human Y chromosome (NRY) has identified 12 novel genes or families, 10 with full-length cDNA sequences. All 12 genes, and six of eight NRY genes or families previously isolated by less systematic means, fell into two classes. Genes in the first group were expressed in many organs; these housekeeping genes have X homologs that escape X inactivation. The second group, consisting of Y-chromosomal gene families expressed specifically in testes, may account for infertility among men with Y deletions. The coherence of the NRY's gene content contrasts with the apparently haphazard content of most eukaryotic chromosomes.

Functional or developmental themes have rarely been ascribed to whole chromosomes in eukaryotes. Instead, individual chromosomes appear to contain motley assortments of genes, with extremely heterogeneous patterns of developmentally regulated expression. We speculated that the human Y chromosome might represent a functionally coherent exception, at least in its nonrecombining portion (the NRY), which comprises 95% of its length (*1*). It is known to differ from all other nuclear human chromosomes by the absence of recombination, its presence in males only, its common ancestry and persistent meiotic relationship with the X chromosome, and the tendency of its genes to degenerate during evolution (*2*).

From the 1950's to the present day, many biologists have assumed that the Y chromosome is a functional wasteland, despite the discovery of several NRY genes during this period. Studies of human pedigrees had identified many traits exhibiting autosomal or X-linked inheritance, but no convincing cases of Y-linked inheritance (*3*). In 1959, reports of XO females and XXY males established the existence of a sex-determining gene on the human Y chromosome (*4*), but this was perceived as a special case on a generally desolate chromosome. The wasteland model has been revised only during the past decade, when eight NRY transcription units (or families of closely related transcription units) were identified, mostly during regionally focused, positional cloning experiments (*5-8*). Even in recent years, it has been argued that the NRY's gene content is essentially limited to random, disintegrating vestiges of its common ancestry with the X (*9*). The Y-

specific repetitive sequences that are so plentiful in the euchromatic regions (*10, 11*) have often been assumed to be functionally inert (*12*). Realizing that these wasteland theories were based on limited, anecdotal data as to the NRY's gene content, we decided to embark on a broad, systematic gene hunt that could uncover previously unrecognized functional patterns.

A complete description of the NRY's gene content cannot be obtained using current research methods, short of sequencing the entire NRY. However, it should be possible to obtain a broad, representative sampling of NRY genes that could enable us to make comprehensive generalizations. We searched for this sampling in sequences transcribed in a single, complex tissue, the testis. To assess the suitability of the testis, and of a "cDNA selection" protocol (*13*), for this project, we first sought to crudely measure what fraction of human genes, regardless of developmental regulation, are detectably transcribed there. We did this by testing whether previously identified pseudoautosomal genes (*1*), whose diversity in developmentally regulated expression is like that of autosomal genes, could be found among testis transcripts. The nine known pseudoautosomal genes were previously identified using mRNA sources as specialized as liver, pineal gland, and skeletal muscle. The extent to which we recovered the nine known pseudoautosomal genes from sampling of testis cDNA would provide a measure of this tissue's adequacy in representing a broad array of genes.

In fact, we recovered testis cDNAs for all nine known pseudoautosomal genes, suggesting that the testis as a single source would be sufficient to provide nearly comprehensive access to NRY genes. From primary, uncloned testis cDNA, we selected and determined the nucleotide sequence of 2539 fragments that hybridized to Y chromosomal DNA. We anticipated that these sequence fragments would represent a redundant sampling of a much smaller set of genes. Nucleotide sequence analysis revealed that 579 fragments corresponded to known Y genes, including all nine pseudoautosomal genes previously reported, and seven of eight known NRY genes. (The one previously reported NRY gene that we failed to recover was *AMELY*, which is expressed only in developing tooth buds.) After further analysis, both to eliminate human repetitive sequences and to assemble overlapping fragments into contigs, novel sequences were hybridized to Southern blots of

human genomic DNAs. Sequences that detected at least one prominent male-specific fragment were judged likely to derive from the NRY, and for each we attempted to isolate cDNA clones from a human testis library (*13*). Nucleotide sequencing of cDNA clones, and rescreening of libraries as necessary, yielded full-length cDNA sequences for ten novel NRY genes or families, and partial cDNA sequences for two additional ones (Table 1). We localized all 12 novel genes on the Y chromosome (Fig. 1) (*14*) and assessed their expression in diverse human tissues by Northern blotting (Fig. 2). The novel genes encode an assortment of proteins (Table 1) and are dispersed throughout the euchromatic portions of the NRY (Fig. 1).

Although our gene hunt was systematic, it is likely that some NRY genes in addition to *AMELY* escaped detection; this could have resulted from failure to select corresponding cDNAs or from discarding them during subsequent screening steps. Like *AMELY*, other NRY genes may not have been recovered because they are not transcribed at sufficient levels in the testis. Our screening criteria may have discriminated against NRY genes located in regions of exceptionally high sequence similarity to the X chromosome. In particular we may have overlooked genes located in a 4-Mb region of the NRY characterized by 99% sequence identity to the X (*15*). Nonetheless, we suspect that the gene hunt was sufficiently comprehensive for us to form meaningful generalizations about the NRY's gene content.

The 12 novel genes readily sort into two discrete classes (Table 1). The first group, five novel NRY genes, share several features. Each has a homolog on the X chromosome encoding a very similar but non identical protein isoform; every gene is expressed in a wide range of human tissues; and each gene appears to exist in a single copy on the NRY. The other seven novel NRY genes constitute the second group, which share quite different traits. They appear to be expressed specifically in testes. They also seem to exist in multiple copies on the NRY, as judged by i) the number and intensity of hybridizing fragments on genomic Southern blots (not shown) or ii) multiple map locations on the Y. The two classes of genes suggested by our NRY-wide search also accommodate six of eight NRY genes previously identified by less systematic means (*5-8, 16*), confirming the validity of this bipartite classification.

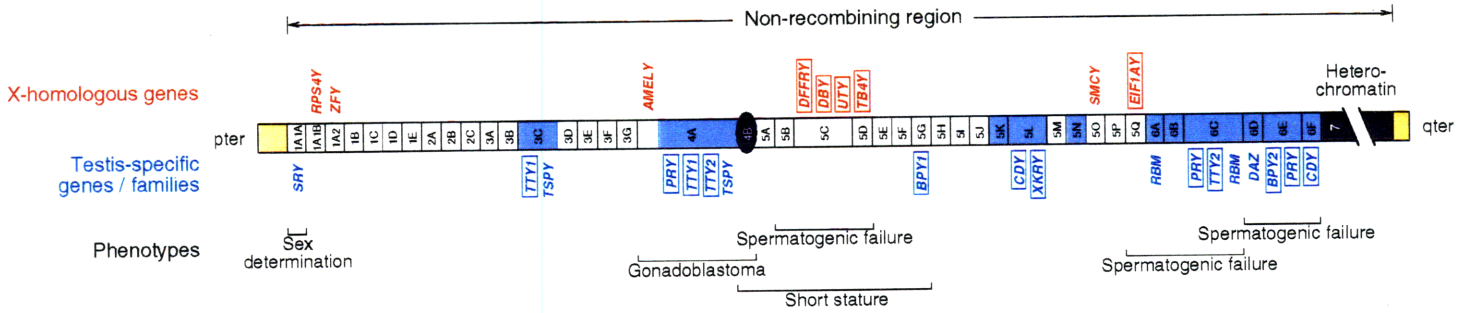


Fig. 1. Gene map of NRY. The Y chromosome consists of a large non-recombining region (NRY; euchromatin plus heterochromatin) flanked by pseudoautosomal regions (yellow). Pter, short arm telomere; qter, long arm telomere. The NRY is shown divided into 43 ordered intervals (1A1A through 7) defined by naturally occurring deletions; deletion intervals previously shown to contain Y-specific repeats are shaded blue (10, 11). Listed immediately above chromosome are nine NRY genes with functional X homologs (red); novel genes are boxed. Immediately below chromosome are 11 testis-specific genes or families (blue), some with multiple locations. Within deletion intervals, genes have not been ordered. Some testis-specific families probably have members in additional deletion intervals; indicated locations are representative but not necessarily exhaustive. At bottom are shown NRY regions implicated, by deletion mapping, in sex determination, germ cell tumorigenesis (gonadoblastoma), stature, and spermatogenic failure (7, 8, 28, 31). For *DFFRY*, previously thought to be a pseudogene, these mapping studies confirm published findings (19).

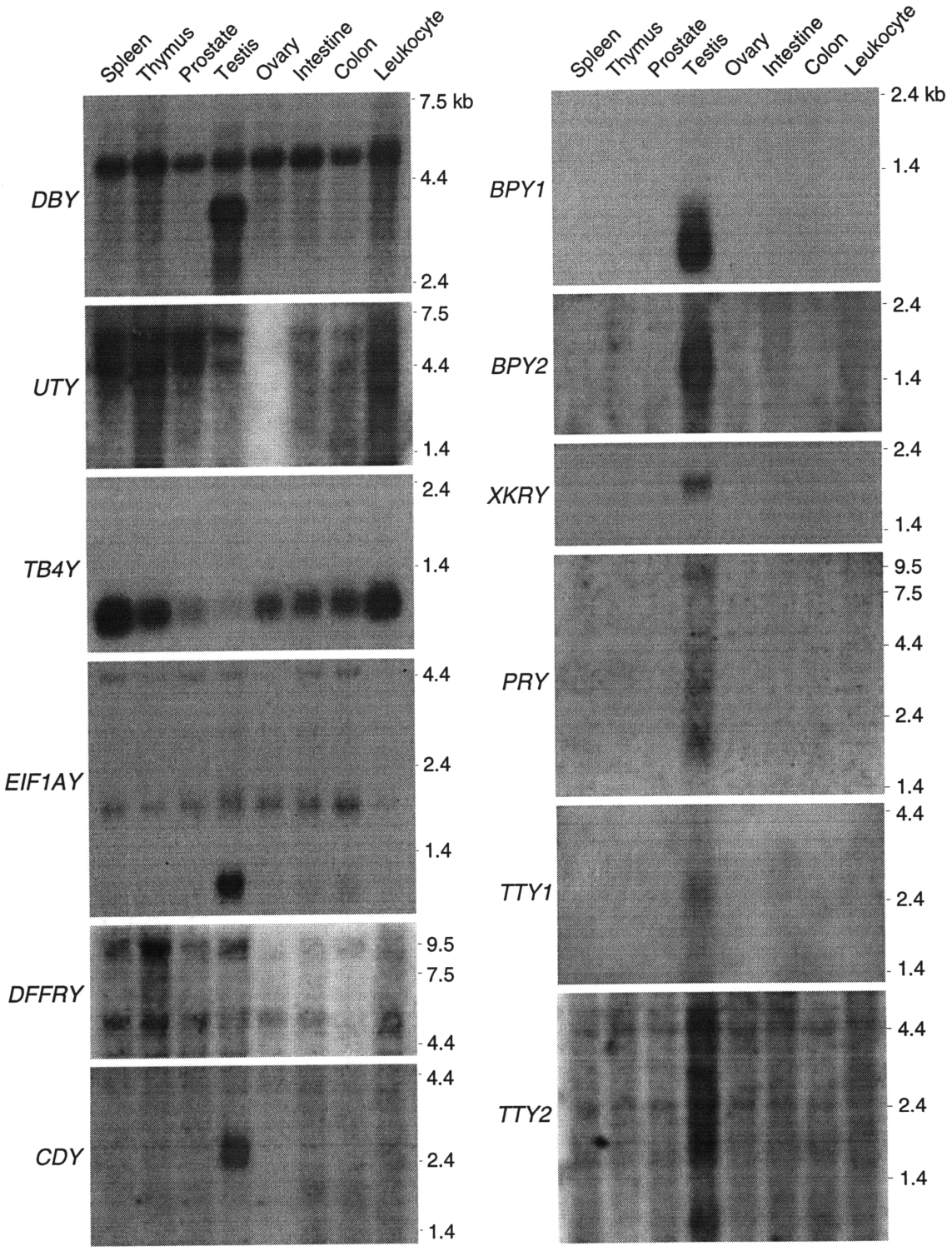


Fig. 2. Transcription of 12 novel NRY genes in human tissues. Autoradiograms produced by hybridizing 32P-labeled cDNA probes to Northern blots of poly(A)+ RNAs (2 µg/lane) from human tissues (Clontech, Palo Alto, CA). Probes employed were cDNA clones, full-length (most genes) or partial (*DBY*, nucleotides 1476-2319 of GenBank AF000985; *UTY*, nucleotides 861-1768 of GenBank AF000996; *DFFRY*, nucleotides 8604-9878 of GenBank AF000986). Hybridization at 65°C in 0.5 M NaiPO4 pH7.5, 7% SDS; washing at 65°C in 1X SSC, 0.1% SDS. *DBY*, *TB4Y*, *EIF1AY* and *DFFRY* probes cross-hybridize to transcripts derived from their X homologs. For all five X-homologous genes (*DBY*, *PRY*, *TB4Y*, *EIF1AY* and *DFFRY*), expression was tested and confirmed in three male tissues (brain, prostate and testis) by RT-PCR using Y-specific primers (not shown). For *DFFRY*, previously thought to be a transcribed pseudogene, these expression studies confirm published findings (19) .

Table 1: 12 novel genes and gene families in the NRY

NRY Genes and Gene Families						Functional X Homologs			
Gene symbol	Gene name	Comments	GenBank #	Tissue Expression	Multi-copy on Y	Gene symbol	GenBank # (ref.)	X-Y amino acid sequence identity	Escape X inactivation
DBY	*Dead Box Y*	Novel protein; "DEAD box" motif suggests this may be an RNA helicase (32)	AF000985 AF000984	ubiquitous	no	*DBX*	AF000983 AF000982	91%	yes
TB4Y	*Thymosin ß4, Y isoform*	X homolog sequesters actin (17)	AF000989	ubiquitous	no	*TB4X*	(17)	93%	yes
EIF1AY	*Translation Initiation Factor 1A, Y isoform*	X homolog is an essential initiation factor (18)	AF000987	ubiquitous	no	*EIF1AX*	(18)	97%	yes
UTY	*Ubiquitous TPR motif Y*	Mouse Y homolog recently shown to encode an H-Y antigen; contains 10 tandem "TPR" motifs implicated in protein-protein interaction (33); differential splicing may generate isoforms differing at carboxy termini	AF000996 AF000995 AF000994	ubiquitous	no	*UTX*	AF000992 AF000993	85%	yes
DFFRY	*Drosophila Fat Facets Related Y*	X homolog recently described; Y previously thought to carry a transcribed pseudogene; homologous to Drosophila deubiquinating enzyme required for eye development and oogenesis (19, 34)	AF000986	ubiquitous	no	*DFFRX*	(19)	91%	yes
CDY	*Chromodomain Y*	Novel protein with "chromodomain" (35) and putative catalytic domain (36); might modify DNA or chromosomal proteins during spermatogenesis	AF000981	testis	yes				
BPY1	*Basic Protein Y 1*	Novel, 125-residue protein rich in S, K, R, P; calculated isoelectric point (pI) 9.4; Southern blotting reveals X homolog, but no X-derived cDNA clones identified to date	AF000979	testis	yes				
BPY2	*Basic Protein Y 2*	Novel, 106-residue protein; calculated pI 10.0	AF000980	testis	yes				
XKRY	*XK Related Y*	Novel protein with similarity to XK, a putative membrane transport protein (37)	AF000997	testis	yes				
PRY	*PTP-BL Related Y*	Novel protein with some similarity to PTP-BL, a putative protein tyrosine phosphatase (38)	AF000988	testis	yes				
TTY1	*Testis Transcript Y 1*	No significant open reading frame identified	AF000990	testis	yes				
TTY2	*Testis Transcript Y 2*	No significant open reading frame identified	AF000991	testis	yes				

Many of the X-homologous genes appear to be involved in cellular housekeeping, as suggested by their ubiquitous expression and by the functions of the encoded proteins, which are well established in three cases. *TB4Y* encodes a Y isoform of thymosin ß4, which functions in actin sequestration (*17*) and which we found to be encoded by the X chromosome. *EIF1AY* encodes a Y isoform of eIF-1A, an essential translation initiation factor (*18*). *RPS4Y* encodes a Y isoform of an essential ribosomal protein (*6*).

By contrast with these single-copy, X-homologous housekeeping genes, the multi-copy NRY gene families appear to encode proteins with more specialized functions. All appear to be expressed specifically in the testis. Our study identified full-length cDNA clones for five of these gene families, which were all found to encode proteins not previously characterized (Table 1). Several of the testis-specific gene families may encode DNA or RNA-binding proteins, including two small, unrelated basic proteins: BPY1 and BPY2 (Table 1); two putative RNA-binding proteins: RBM and DAZ (*7, 8*); and CDY, which contains a "chromodomain" (a chromatin binding motif; Table 1, Fig. 3), and a catalytic domain (Fig. 3 & 4).

We postulate that the NRY's evolution was dominated by two strategies. The first strategy favored conservation of particular X-Y gene pairs to maintain comparable expression of certain housekeeping functions in males and females. This strategy is at odds with the general behavior of X-Y gene pairs during mammalian evolution. The mammalian X and Y chromosomes evolved from autosomes; most ancestral gene functions were retained on the nascent X chromosome but deteriorated on the non-recombining portion of the emerging Y chromosome (*2*). This resulted in females having two copies but males having only one copy of many genes, an inequality predominantly addressed in mammals by transcriptional silencing, or inactivation, of one X chromosome in females. Our findings on X-homologous NRY genes, together with previous studies, suggest the importance in human evolution of an additional solution: preservation of homologous genes on both NRY and X, with male and female cells expressing two copies of such genes. A critical prediction of this model is that the X homologs should escape X inactivation. This is the case for all widely expressed X-linked genes with known NRY homologs, including the X

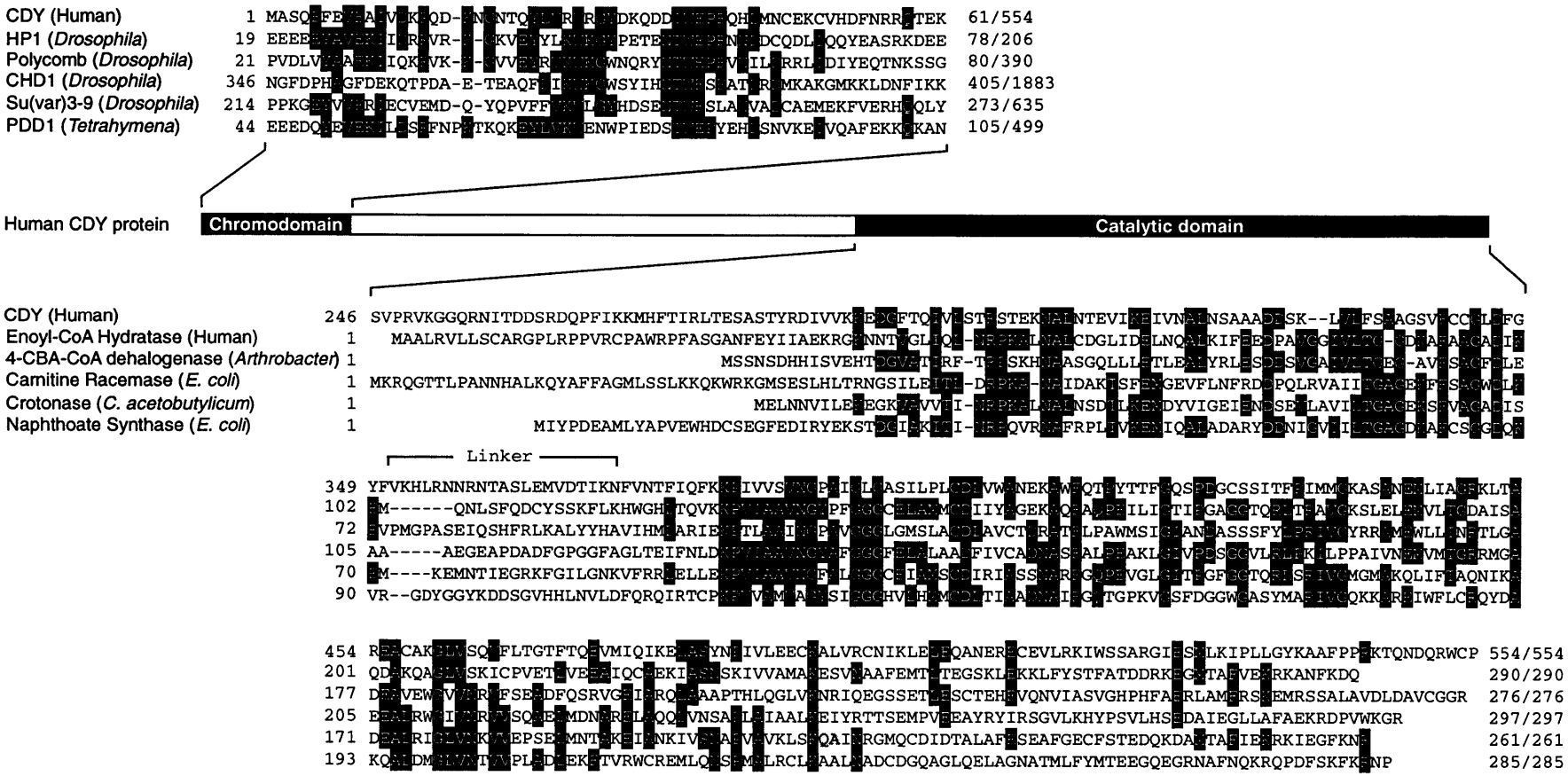


Fig. 3. Amino acid sequence alignments of human CDY's chromodomain and putative catalytic domain with their respective homologs (*35, 36*). The carboxy-terminal half of CDY shows striking amino acid similarity, over a region of more than 200 residues, to nearly the full length of several prokaryotic and eukaryotic enzymes, all of which act on substrates linked to coenzyme A. Amino acid identities in three out of six sequences are indicated by black shading. For each protein, first and last residues shown are numbered (with respect to the initiator methionine), and the protein's total length is indicated.

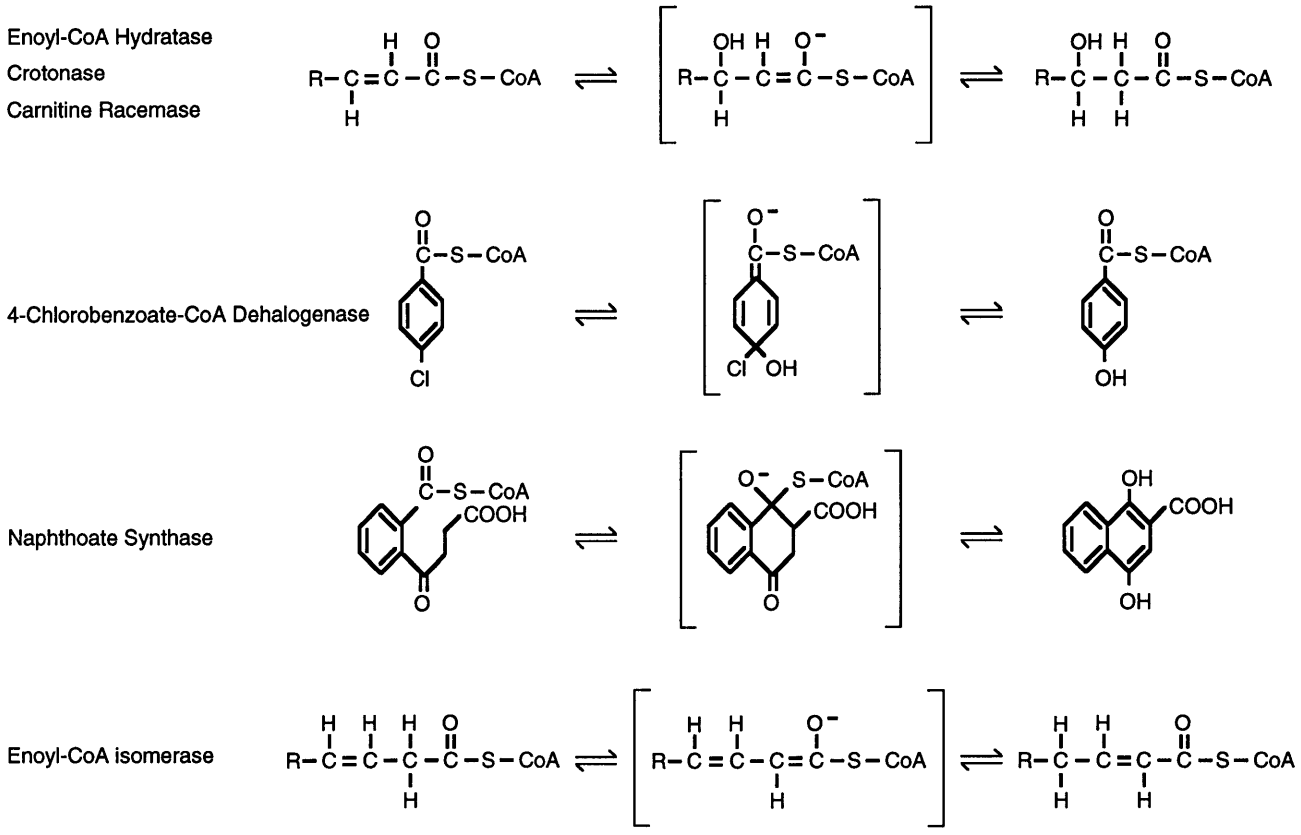


Fig. 4. Reactions catalyzed by enzymes homologous to the catalytic domain of CDY.

homologs of the five novel NRY genes reported here (*6, 19-21*). A second prediction is that the X and Y encoded proteins should be functionally interchangeable despite considerable divergence of their genes' nucleotide sequences. Indeed, each of the eight known X-NRY gene pairs encode closely related isoforms, with 85 to 97% amino acid identity throughout their lengths; functional interchangeability has been demonstrated in the one case tested to date (*22*).

These dosage compensation strategies may be relevant to Turner syndrome, classically associated with an XO sex chromosome constitution. The Turner phenotype may be due to inadequate expression of certain X-Y common genes that escape X inactivation (*23*). Given that several X-NRY genes appear to be involved in cellular housekeeping, we speculate that some features of the XO phenotype (such as poor fetal viability) reflect inadequate expression of particular housekeeping functions. The X-homologous NRY genes (Fig. 1) should be investigated as Turner candidates (*24*).

In addition to the strategy for conserving certain X-Y gene pairs, a second strategy probably shaped the NRY's evolution. This strategy favored the acquisition of testis-specific families, perhaps through selectively retaining and amplifying genes that enhance male reproductive fitness. Animal genomes may contain genes or alleles that enhance male reproductive fitness but are inconsequential or even detrimental to females, as first appreciated by R.A. Fisher (*25*). Fisher recognized that selective pressures would favor the accumulation of such genes in male-specific regions of genomes. Of course, male reproductive fitness depends critically on sperm production, the task of the adult testis. As the only male-specific portion of the mammalian genome, the NRY should have a unique tendency to accumulate male-benefit genes during evolution. Consider the human NRY's *DAZ* gene cluster, *de novo* deletions of which are associated with severe spermatogenic defects (*8*). The *DAZ* cluster on the human Y chromosome arose during primate evolution by transposition and amplification of an autosomal gene. Similarly, two other testis-specific NRY gene families — *RBM* and *TSPY* — may also be the result of the Y's having acquired and amplified autosomal genes (*26*). We speculate that the selective advantage conferred by the NRY's retaining and amplifying male fertility factors (from throughout the genome) accounts for

the multitude of testis-specific gene families there. These activities may have been preeminent in shaping the NRY's gene repertoire, since it appears that the great majority of NRY transcription units are members of testis-specific families (*27*). We suspect that the majority of the NRY's transcription units do not date from the Y's common ancestry with the X chromosome, but instead represent more recent acquisitions.

The importance of the human Y chromosome in fertility has been underscored by recent genetic studies. Many men with spermatogenic failure, while otherwise healthy, lack portions of the NRY (*7, 8, 28*). These findings have suggested the existence of NRY genes that play critical roles in male germ cell development but are not required elsewhere in the body. Previous deletion mapping studies have implicated four regions of the NRY in either spermatogenic failure or germ cell tumorigenesis, and in each of the four regions we now report novel candidate genes expressed specifically, or most abundantly (*29*), in testes (Figs. 1,2)

Although X-homologous and testis-specific genes are somewhat intermingled within the NRY, clustering is evident (Fig. 1). The geographic distribution of the two classes correlates well with previously identified sequence domains within the euchromatic NRY (*10, 11*). Ten of the 11 known testis-specific families map to previously identified regions of Y-specific repetitive sequences (*30*). Indeed, one or more testis-specific gene families are found in nearly all known regions of euchromatic Y repeats (Fig. 1). Ironically, it had been widely assumed, partly on theoretical grounds, that these domains consisted of "junk" DNA (*12*). To the contrary, our results argue that these Y-specific repetitive regions are gene-rich, containing most of the NRY's transcription units (*27*). We speculate that these were regions of rampant gene amplification during mammalian evolution. By contrast, none of the eight X-homologous genes map to the Y-repeat domains; they all map to regions previously identified as consisting largely of single-copy (or in some cases X-homologous) sequences. We postulate that, earlier in mammalian evolution, these regions of the NRY shared extensive nucleotide sequence identity with the X chromosome.

Although more genes likely remain to be discovered, the 20 genes and families shown in Fig. 1 may constitute the majority of NRY genes, and full-length cDNA sequences are available for

18 of them. The stage is now set for systematic evolutionary, biochemical and cell biological studies of this distinctive segment of the human genome.

REFERENCES AND NOTES

1. The remaining 5% of the chromosome is composed of two pseudoautosomal regions that maintain sequence identity with the X chromosome by meiotic recombination [H. J. Cooke, W. R. Brown, G. A. Rappold, *Nature* **317**, 687 (1985); M. C. Simmler, et al., *Nature* **317**, 692 (1985); D. Freije, C. Helms, M. S. Watson, H. Donis-Keller, *Science* **258**, 1784 (1992); N. A. Affara, et al., *Cytogenet Cell Genet* **73**, 33 (1996); E. Rao et al., *Nat Genet* **16**, 54 (1997)].
2. J. J. Bull, *Evolution of Sex Determining Mechanisms* (Benjamin Cummings, Menlo Park, CA, 1983); J. A. Graves, *Annu. Rev. Genet.* **30**, 233 (1996); B. Charlesworth, *Curr Biol* **6**, 149 (1996); W. R. Rice, *BioScience* **46**, 331 (1996).
3. C. Stern, *Am J Hum Genet* **9**, 147 (1957).
4. P. A. Jacobs, J. A. Strong, *Nature* **183**, 302 (1959); C. E. Ford, K. W. Jones, P. E. Polani, J. C. Almida, J. H. Briggs, *Lancet* **i**, 711 (1959).
5. D. C. Page, et al., *Cell* **51**, 1091 (1987); A. H. Sinclair, et al., *Nature* **346**, 240-4 (1990); J. Arnemann, S. Jakubiczka, S. Thuring, J. Schmidtke, *Genomics* **11**, 108 (1991); E. C. Salido, P. H. Yen, K. Koprivnikar, L. C. Yu, L. J. Shapiro, *Am J Hum Genet* **50**, 303 (1992); A. I. Agulnik, M. J. Mitchell, J. L. Lerner, D. R. Woods, C. E. Bishop, *Hum Mol Genet* **3**, 873 (1994).
6. E. M. Fisher, et al., *Cell* **63**, 1205 (1990).
7. K. Ma, et al., *Cell* **75**, 1287 (1993).
8. R. Reijo, et al., *Nat Genet* **10**, 383 (1995).
9. J. A. Graves, *Bioessays* **17**, 311 (1995).
10. D. Vollrath, et al., *Science* **258**, 52 (1992).
11. S. Foote, D. Vollrath, A. Hilton, D. C. Page, *Science* **258**, 60 (1992).
12. B. Charlesworth, *Science* **251**, 1030 (1991); E. Seboun, et al., *Cold Spring Harb Symp Quant Biol* **1**, 237 (1986).
13. We performed "cDNA selection" [M. Lovett, J. Kere, L. M. Hinton, *Proc Natl Acad Sci U S A* **88**, 9628 (1991)] with bulk cDNA from human adult testes (Clontech, Palo Alto, CA) and, as

selector, a cosmid library prepared from flow-sorted Y chromosomes (Lawrence Livermore National Laboratory: LL0YNC03). 3600 random cosmids, providing nearly five-fold coverage of the 30-Mb euchromatic region, were used to generate 150 pools of selector DNA. (Theoretically, these 3600 cosmids are expected to contain 99% of all euchromatic Y sequences. It is our experience that this cosmid library contains >95% of Y-DNA STSs tested.) With each selector pool, we carried out four successive rounds of cDNA selection, followed by two rounds of subtraction with human COT-1 DNA (Gibco BRL, Gaithersburg, MD) to remove highly repetitive sequences. A plasmid library was prepared from each of the 150 resulting pools of selected cDNA fragments, and 24 clones from each library were sequenced from one end. Of the 3600 sequences generated, about 600 were of poor technical quality and about 500 were found to derive from cloning vector or *E. coli* host, leaving 2539 sequences for further analysis. Of the 2539 sequence fragments, 536 corresponded to previously reported NRY genes (487 to *TSPY*, 15 to *RBM*, 14 to *RPS4Y*, 9 to *SMCY*, 5 to *DAZ*, 3 to *SRY*, 3 to *ZFY*) and 43 corresponded to previously reported pseudoautosomal genes (15 to *XE7*, 11 to *CSF2RA*, 4 to *IL3RA*, 3 to *ASMT*, 3 to *IL9R*, 2 to *ANT3*, 2 to *MIC2*, 2 to *SHOX*, 1 to *SYBL1*). Electronic analysis of the roughly 2000 remaining sequences revealed that about 200 contained known repetitive elements, and these were not pursued. By electronically identifying redundancies and sequence overlaps, the remaining sequences were reduced to 1093 sequence contigs. Sequences representing these 1093 contigs were individually hybridized to dot-blotted yeast genomic DNAs of 60 YACs comprising most of the Y's euchromatic region (*11*). 181 sequences that hybridized to the great majority of the YACs were judged likely to contain highly repeated elements and were not pursued, leaving 912 sequences for further analysis. These sequences were individually hybridized to Southern blots of EcoRI-digested human 46,XX female and 49,XYYYY male genomic DNAs; hybridization at 65°C in 0.5 M Na_iPO_4 pH7.5, 7% SDS; washing at 65°C in 1X SSC, 0.1% SDS. 832 hybridizations yielded interpretable results. Many sequences appeared to contain highly repeated elements common to males and females, or failed to detect an unambiguously Y-specific restriction fragment, and

these were not pursued. (This eliminated sequences derived from the pseudoautosomal regions or other regions of extremely high X-Y nucleotide similarity.) By contrast, 308 sequences hybridized to at least one prominent fragment present in 49,XYYYY but absent in 46,XX, suggesting that these sequences derived from the NRY. Each of these 308 sequences was individually used to screen, by hybridization, about 2 million plaques from a λ phage library of human adult testis cDNA (Clontech, Palo Alto, CA).

14. Genes were localized on a previously reported NRY deletion map by PCR testing of individuals carrying partial Y chromosomes (*10*). Most genes were localized to a single deletion interval. Some genes could not be unambiguously placed because copies exist in multiple locations in the NRY. In such cases, genes were localized by PCR testing of YACs encompassing the euchromatic region (*11*). Homologs of *DBY*, *TB4Y*, *EIF1AY*, and *UTY* were mapped to the human X chromosome by PCR testing a panel of human/rodent somatic hybrid cell lines (Research Genetics, Huntsville, AL). PCR conditions and primer sequences have been deposited at GenBank, where accession numbers are as follows: *DBY*, G34990; *TB4Y*, G34981; *EIF1AY*, G34991; *UTY*, G34977; *DFFRY*, G34983; *CDY*, G34975; *BPY1*, G34985; *BPY2*, G34986; *XKRY*, G34987; *PRY*, G34984; *TTY1*, G34978; *TTY2*, G34980; *DBX*, G34988; *TB4X*, G34979; *EIF1AX*, G34989; *UTX*, G34976; *DFFRX*, G34982. *TB4X* primers were designed from an unreported intron; all other PCR primers were chosen from cDNA sequences.
15. S. Mumm, B. Molini, J. Terrell, A. Srivastava, D. Schlessinger, *Genome Research* **7**, 307 (1997); A. Schwartz, et al., *Hum Molec Genet*, in press.
16. Two previously identified NRY genes do not fit neatly into either of the two classes: *SRY* is expressed in testes but exists in only one copy in the NRY. *AMELY* and its X-linked homolog *AMELX* are expressed only in tooth buds. The X inactivation status of *AMELX* is unknown. Thus, not all testis-specific NRY genes are multi-copy, and not all X-homologous NRY genes are ubiquitously expressed.

17. H. Gondo, et al., *J Immunol* **139**, 3840 (1987); D. Safer, M. Elzinga, V. T. Nachmias, *J Biol Chem* **266**, 4029 (1991).

18. T. E. Dever, et al., *J Biol Chem* **269**, 3212 (1994); J. W. Hershey, *Annu Rev Biochem* **60**, 717 (1991).

19. M. H. Jones, et al., *Hum Mol Genet* **5**, 1695 (1996).

20. A. Schneider-Gädicke, P. Beer-Romero, L. G. Brown, R. Nussbaum, D. C. Page, *Cell* **57**, 1247 (1989); A. I. Agulnik, et al., *Hum Mol Genet* **3**, 879 (1994).

21. For each of the previously untested X-linked genes (*DBX*, *UTX*, *TB4X*, *EIF1AX*), we assayed X inactivation status by two methods: i) RT-PCR on human-rodent hybrids retaining inactive human X chromosomes and ii) CpG methylation studies in which male and female genomic DNAs digested with methylation sensitive restriction endonucleases were used as templates for PCR [S.-W. Luoh, et al., *Genomics* **29**, 353 (1995)]. As judged by both assays, each of these genes escapes X inactivation (B.L. and D.C.P., unpublished results).

22. M. Watanabe, A. R. Zinn, D. C. Page, T. Nishimoto, *Nat Genet* **4**, 268 (1993).

23. M. A. Ferguson-Smith, *J. Med. Genet.* **2**, 142 (1965); A. R. Zinn, D. C. Page, E. M. Fisher, *Trends Genet* **9**, 90 (1993).

24. The potential role of *RPS4Y* and *RPS4X* in Turner syndrome is controversial (*6*) [W. Just, C. Geerkens, K. R. Held, W. Vogel, *Hum Genet* **89**, 240 (1992)]. At least one Turner gene maps to the Xp-Yp pseudoautosomal region (*1*).

25. R. A. Fisher, *Biol Rev* **6**, 345 (1931).

26. R. Saxena, et al., *Nat Genet* **14**, 292 (1996); M. L. Delbridge, et al., *Nat. Genet.* **15**, 131 (1997).

27. Each NRY testis-specific gene family has multiple members, 20 to 40 copies in the case of *TSPY* [E. Manz, F. Schnieders, A. M. Brechlin, J. Schmidtke, *Genomics* **17**, 726 (1993)], and 20 or more copies in the case of *RBM* (*7*). All together, the various Y-specific gene families may include several hundred genes or copies. Though it is not known how many of these are functional, it seems likely that Y-specific, testis-specific gene families comprise the great majority of NRY transcription units.

28. P. H. Vogt, et al., *Hum Mol Genet* **5**, 933 (1996); J. L. Pryor, et al., *New England J. Med.* **336**, 534 (1997).

29. One region implicated in spermatogenic failure (intervals 5B-5D) contains an X-homologous gene, *DBY*, with an abundant, testis-specific transcript in addition to a ubiquitous transcript (Fig. 2). Expression of genes important in male germ cell development is not necessarily restricted to the testis.

30. The single exception is *BPY1*, which cross-hybridizes to the X chromosome and maps to a previously recognized region of X homology (*11*).

31. K. Tsuchiya, R. Reijo, D. C. Page, C. M. Disteche, *Am J Hum Genet* **57**, 1400 (1995); P. Salo, et al., *Genes Chromosomes Cancer* **14**, 210 (1995); P. Salo, H. Kaariainen, D. C. Page, A. de la Chapelle, *Hum Genet* **95**, 283 (1995).

32. P. Linder, et al., *Nature* **337**, 121 (1989); R.-Y. Chuang, P. L. Weaver, Z. Liu, T.-H. Chang, *Science* **275**, 1468 (1997).

33. A. Greenfield, et al., *Nat Genet* **14**, 474 (1996); R. S. Sikorski, M. S. Boguski, M. Goebl, P. Hieter, *Cell* **60**, 307 (1990); D. Tzamarias, K. Struhl, *Genes Dev* **9**, 821 (1995).

34. J. A. Fischer-Vize, G. M. Rubin, R. Lehmann, *Development* **116**, 985 (1992); Y. Huang, R. T. Baker, J. A. Fischer-Vize, *Science* **270**, 1828 (1995).

35. T. C. James, S. C. Elgin, *Mol Cell Biol* **6**, 3862 (1986); R. Paro, D. S. Hogness, *Proc Natl Acad Sci U S A* **88**, 263 (1991); B. Tschiersch, et al., *EMBO J* **13**, 3822 (1994); M. T. Madireddi, et al., *Cell* **87**, 75 (1996); D. G. Stokes, K. D. Tartof, R. P. Perry, *Proc Natl Acad Sci U S A* **93**, 7137 (1996).

36. P. M. Palosaari, et al., *J Biol Chem* **266**, 10750 (1991); A. Schmitz, K. H. Gartemann, J. Fiedler, E. Grund, R. Eichenlaub, *Appl Environ Microbiol* **58**, 4068 (1992); V. Sharma, K. Suvarna, R. Meganathan, M. E. Hudspeth, *J Bacteriol* **174**, 5057 (1992); M. Kanazawa, et al., *Enzyme Protein* **47**, 9 (1993); Z. L. Boynton, G. N. Bennet, F. B. Rudolph, *J Bacteriol* **178**, 3015 (1996).

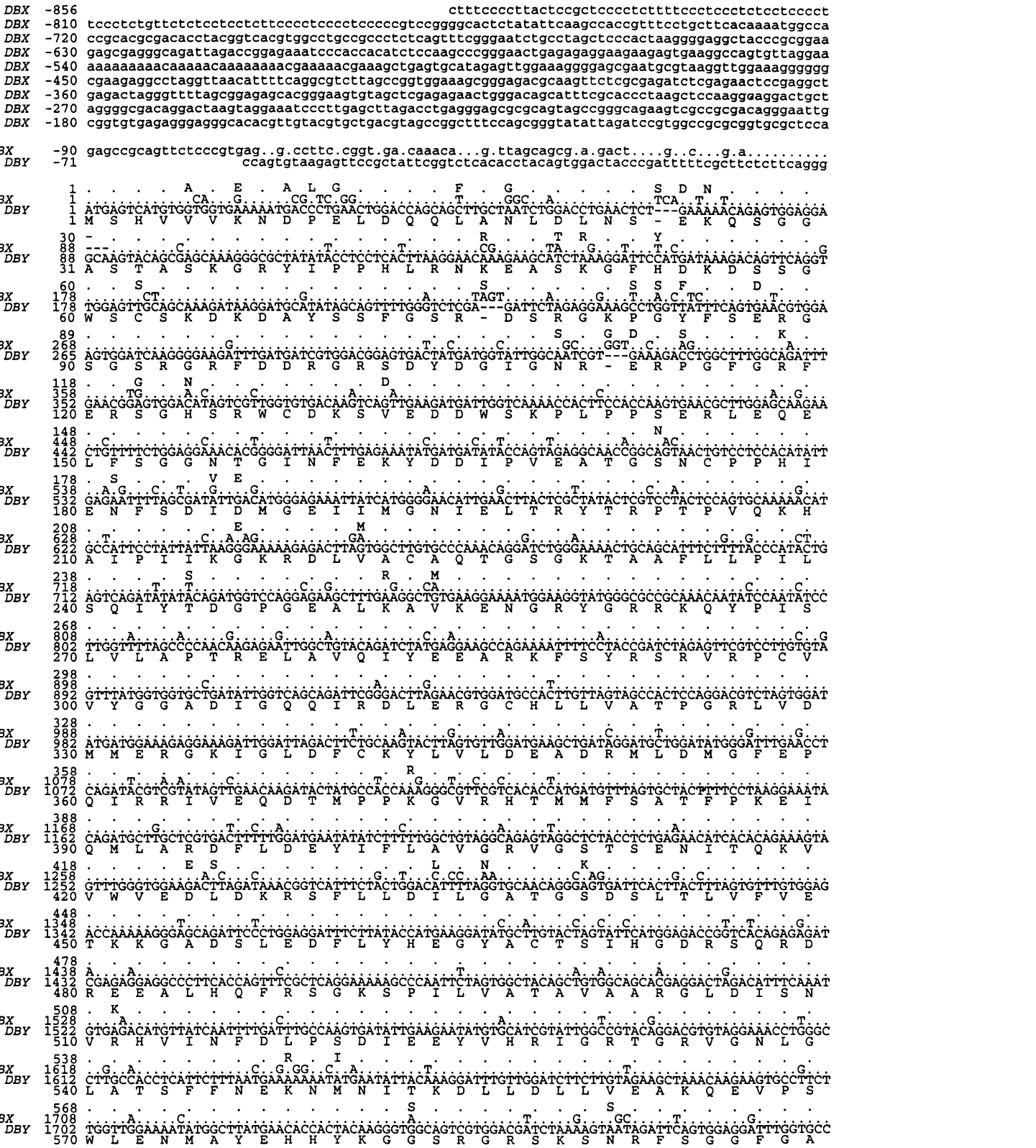
37. M. Ho, et al., *Cell* **77**, 869 (1994).

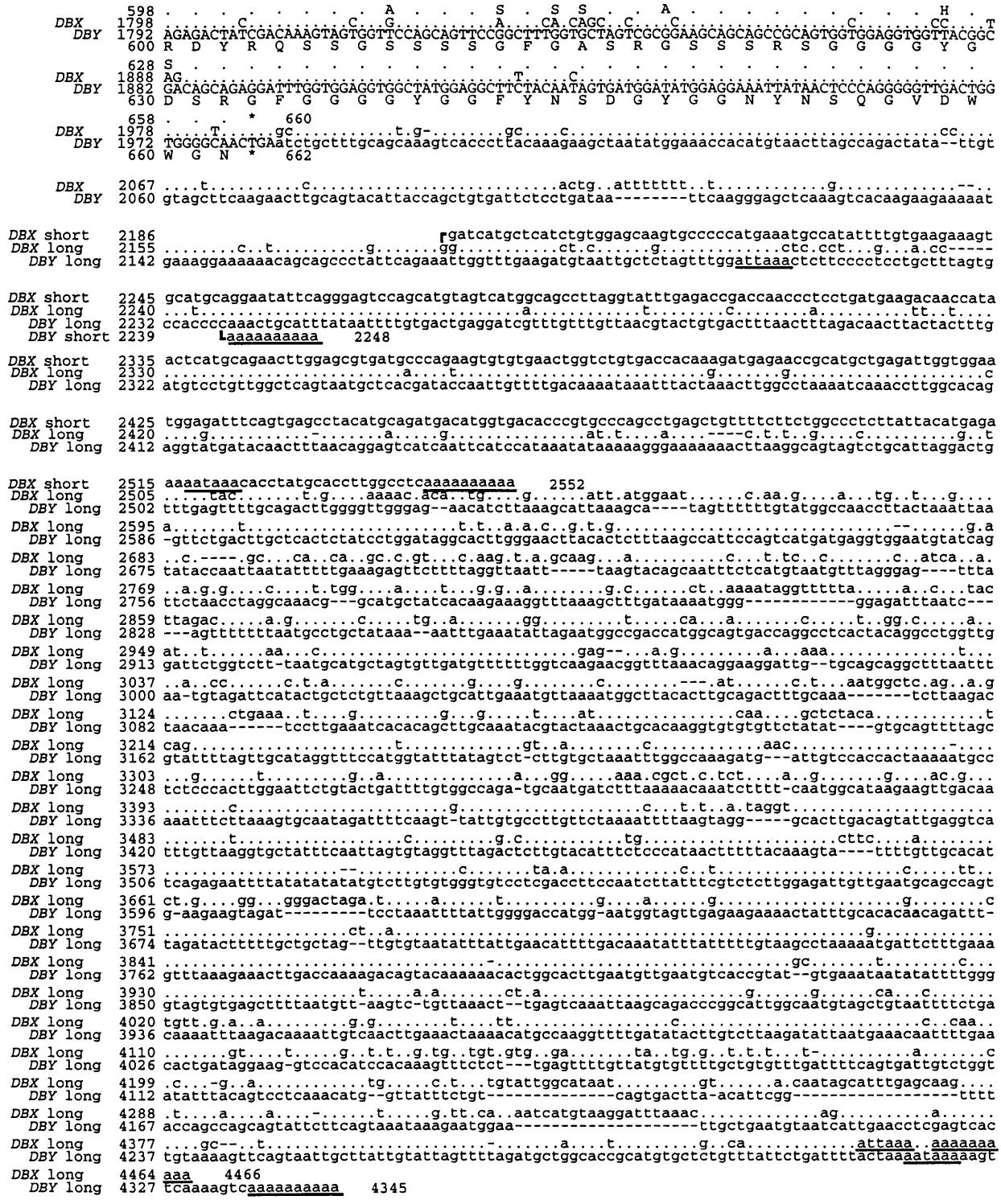
38. W. Hendriks, et al., *J Cell Biochem* **59**, 418 (1995).

39. We thank H. Skaletsky and F. Lewitter for help with sequence analysis, Lawrence Livermore National Laboratory for the flow-sorted Y cosmid library, and P. Bain, A. Bortvin, A. de la Chapelle, G. Fink, K. Jegalian, T. Kawaguchi, E. Lander, H. Lodish, P. Matsudaira, D. Menke, U. RajBhandary, R. Reijo, S. Rozen, A. Schwartz, C. Sun, and C. Tilford for comments on the manuscript. Supported by National Institutes of Health.

Appendix. Sequences of novel genes on human Y and X

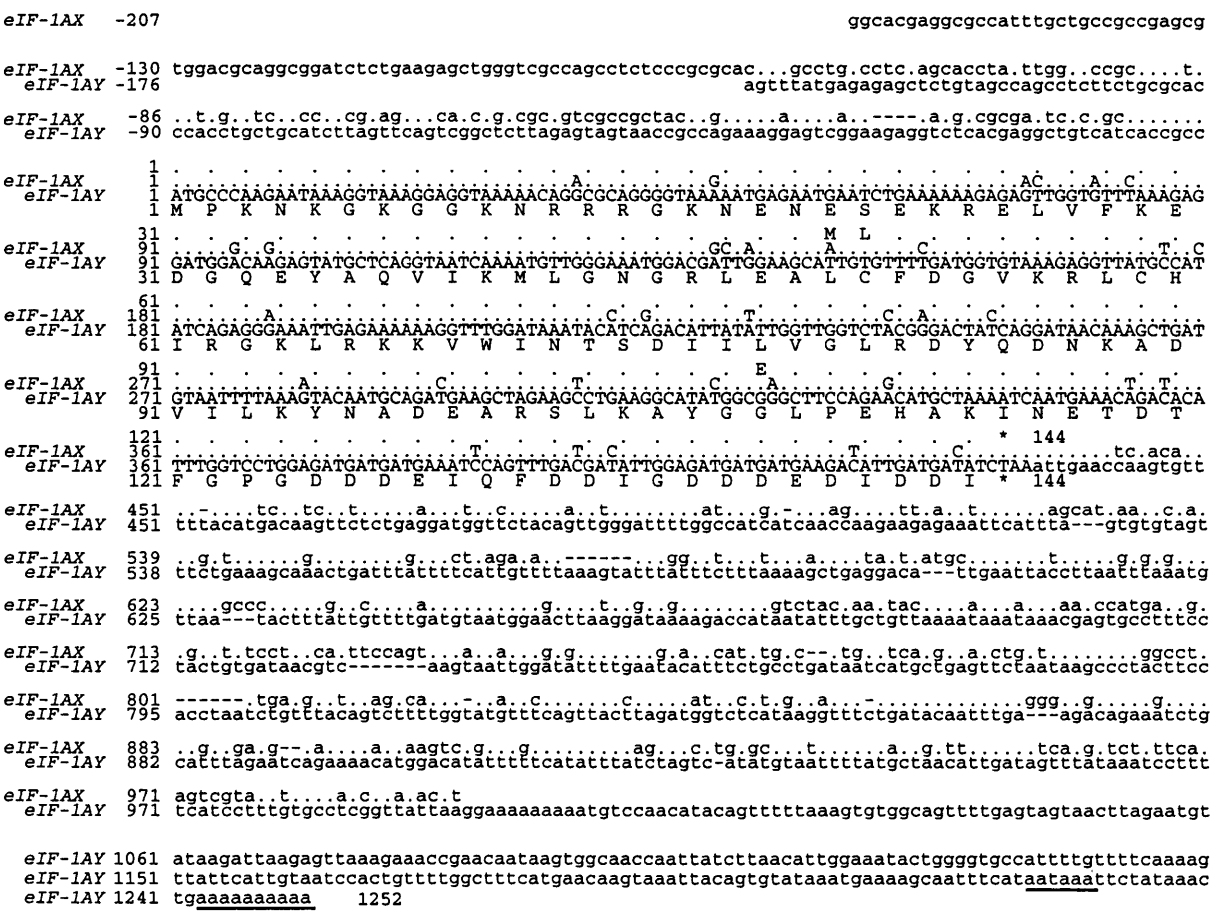
DBX & *DBY*
short and long transcripts

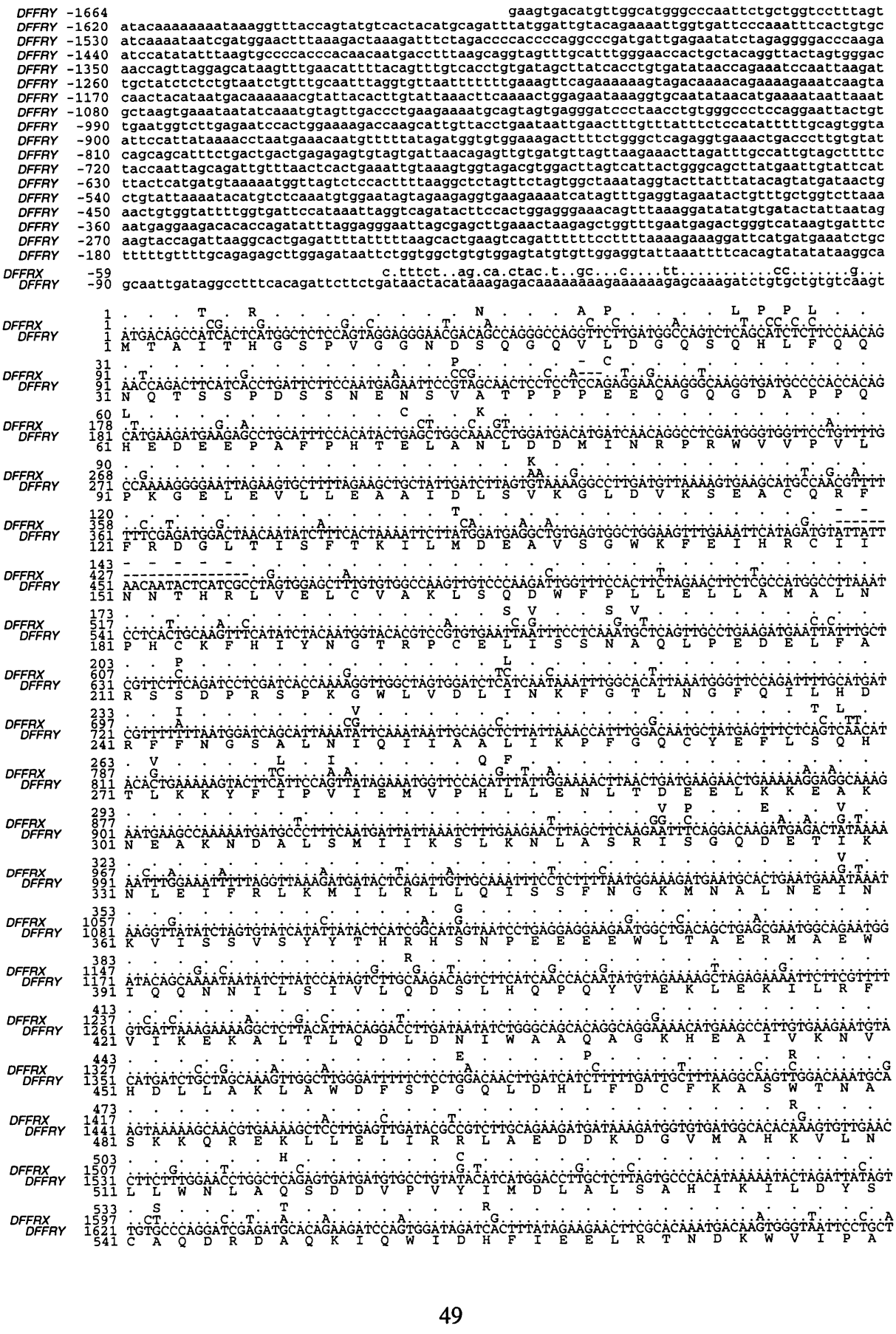


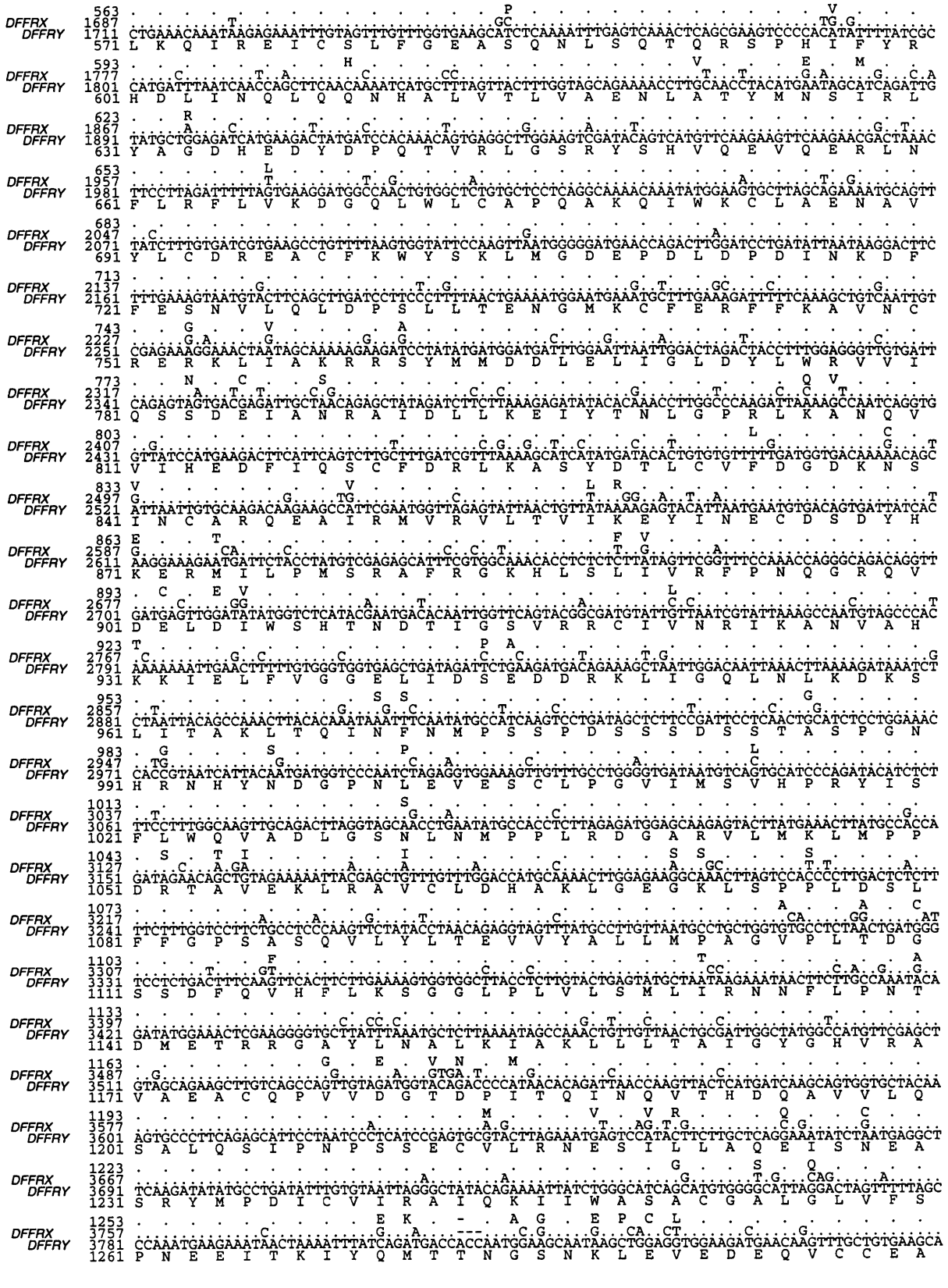


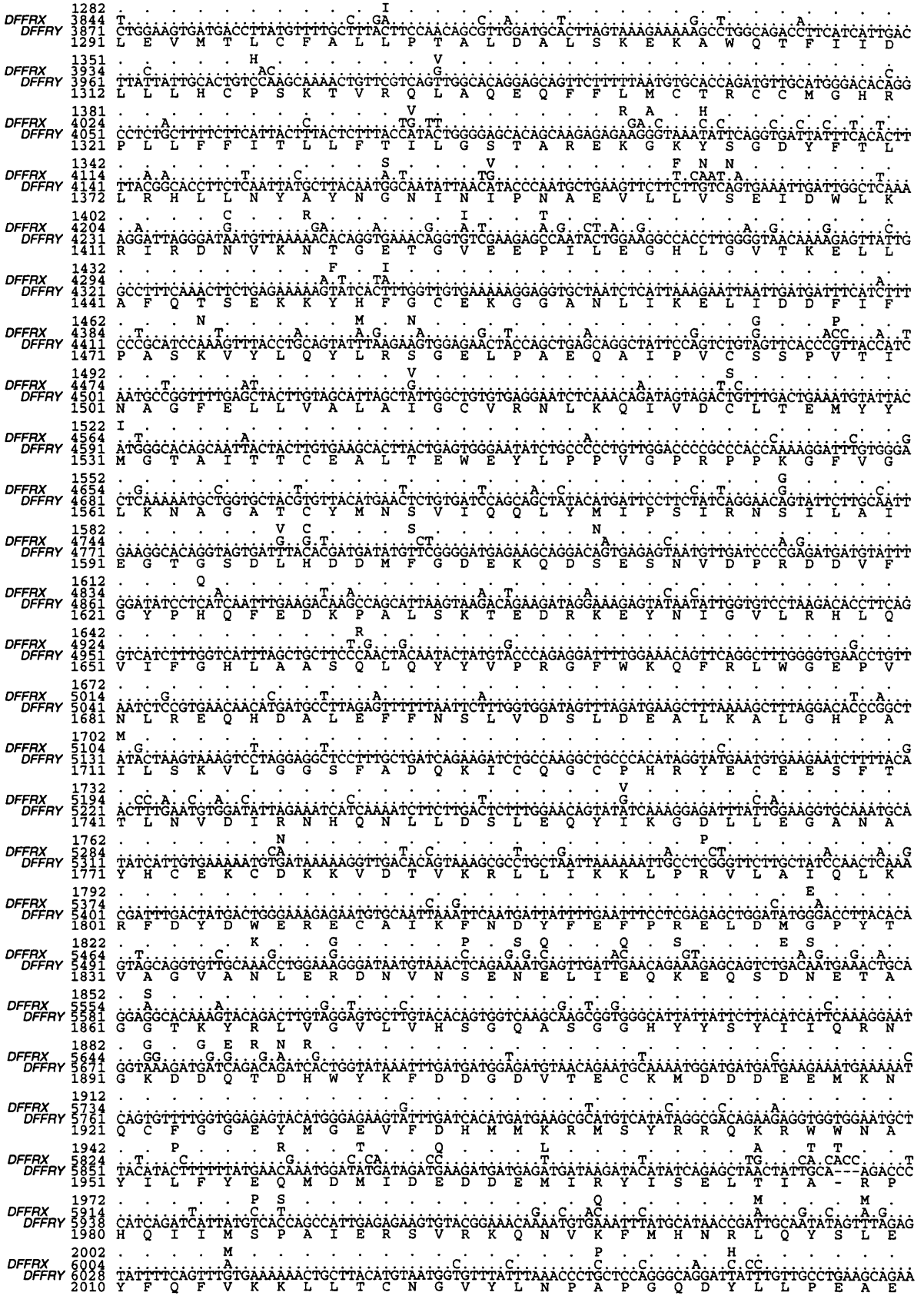
TB4X & TB4Y

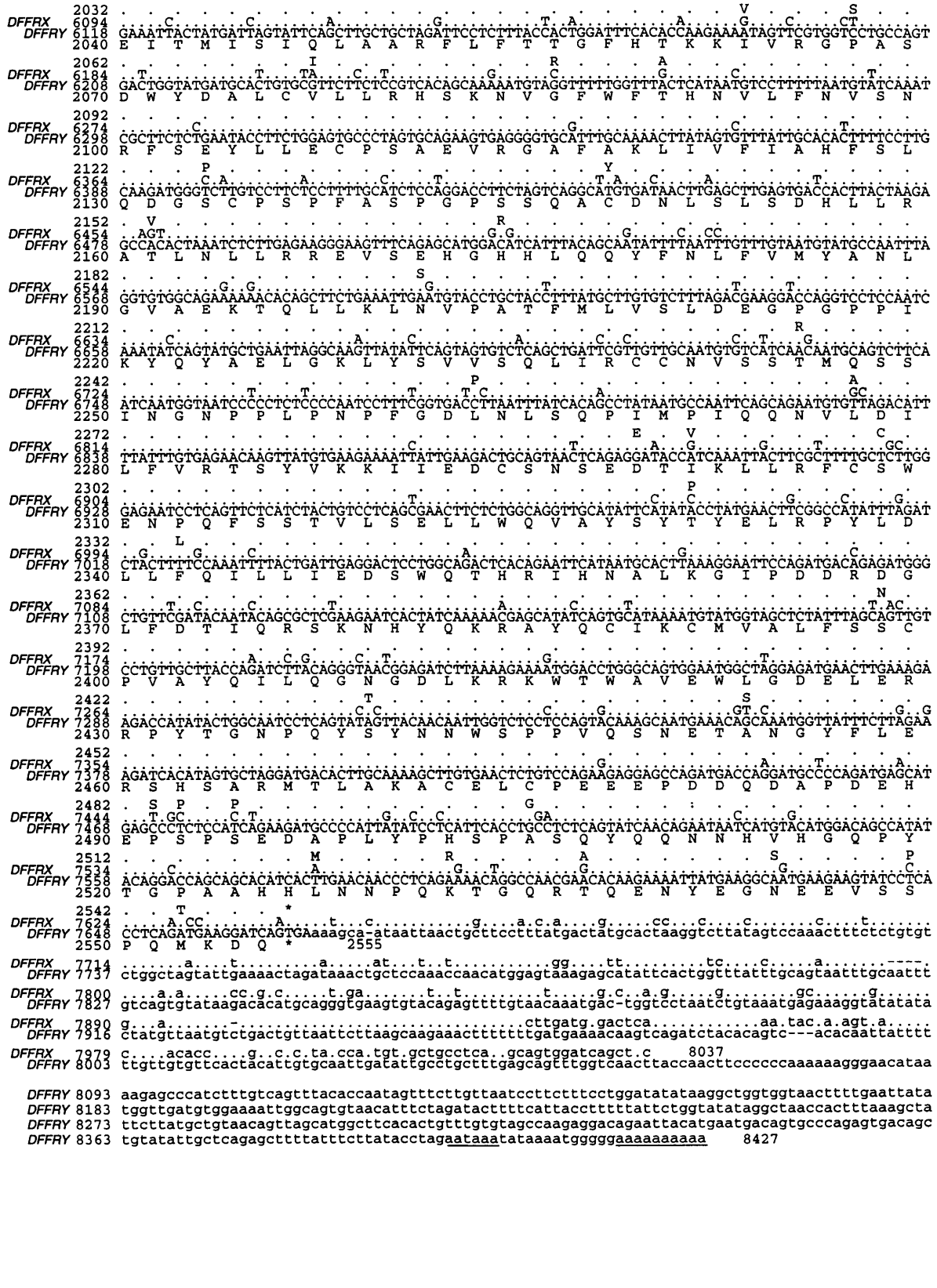
EIF1AX & EIF1AY






DFFRX
DFFRY





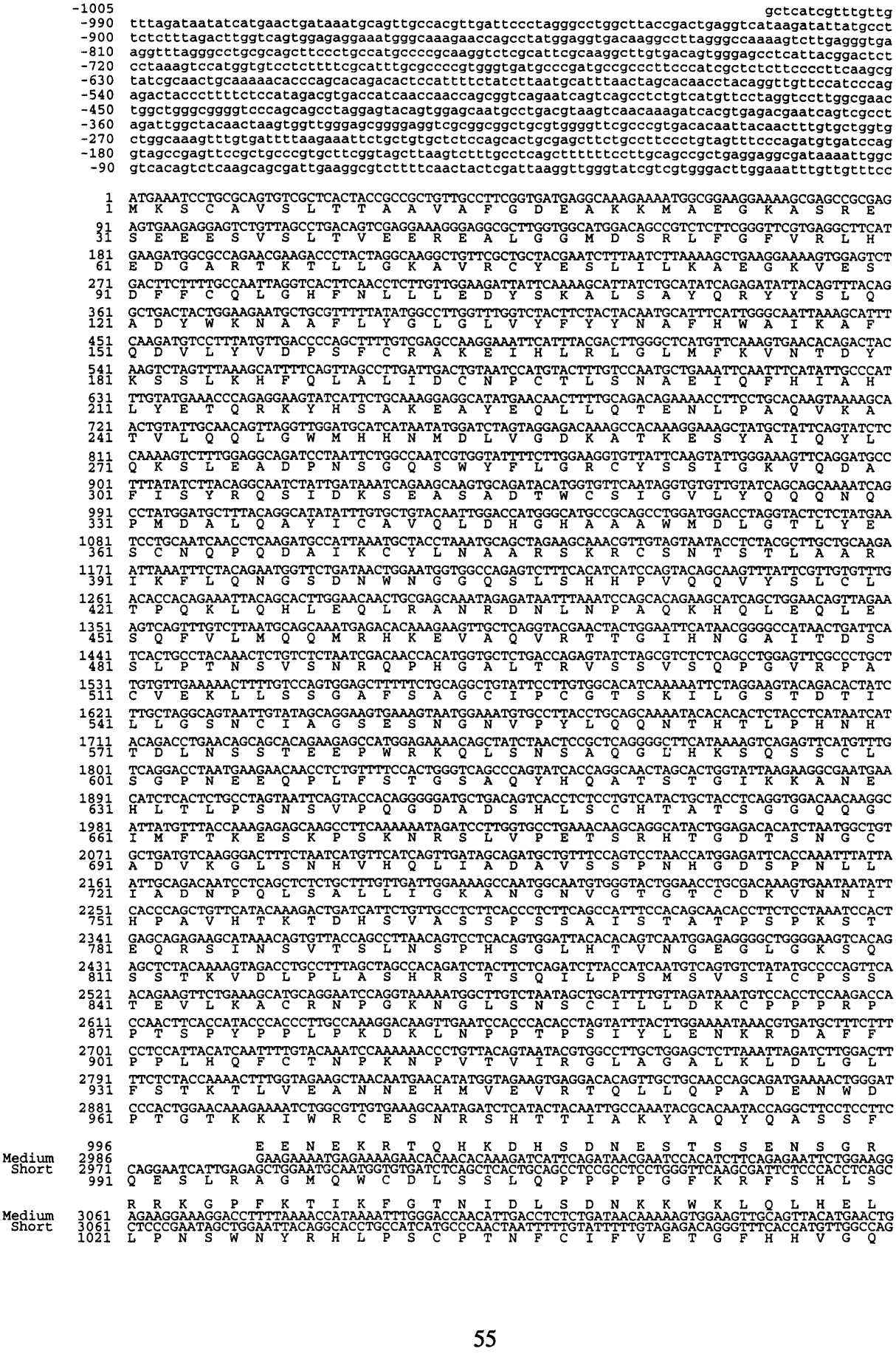
UTX
short and long transcripts

```
-26                                      aaagcaaaagaattcgctgcgtttcc

1     ATGAAATCCTGCGGAGTGTCGCTCGCTACCGCCGCCGCTGCCGCCGCCGCTTTCGGTGATGAGGAAAAGAAAATGGCGGCGGGAAAAGCG
1     M  K  S  C  G  V  S  L  A  T  A  A  A  A  A  A  A  F  G  D  E  E  K  K  M  A  A  G  K  A

91    AGCGGCGAGAGCGAGGAGGCGTCCCCCAGCCTGACAGCCGAGGAGAGGGAGGCGCTCGGCGGACTGGACAGCCGCCTCTTTGGGTTCGTG
31    S  G  E  S  E  E  A  S  P  S  L  T  A  E  E  R  E  A  L  G  G  L  D  S  R  L  F  G  F  V

181   AGATTTCATGAAGATGGCGCCAGGACGAAGGCCCTACTGGGCAAGGCTGTTCGCTGCTATGAATCTCTAATCTTAAAAGCTGAAGGAAAA
61    R  F  H  E  D  G  A  R  T  K  A  L  L  G  K  A  V  R  C  Y  E  S  L  I  L  K  A  E  G  K

271   GTGGAGTCTGATTTCTTTTGTCAATTAGGTCACTTCAACCTCTTATTGGAAGATTATCCAAAAGCATTATCTGCATACCAGAGGTACTAC
91    V  E  S  D  F  F  C  Q  L  G  H  F  N  L  L  L  E  D  Y  P  K  A  L  S  A  Y  Q  R  Y  Y

361   AGTTTACAGTCTGACTACTGGAAGAATGCTGCCTTTTTATATGGTCTTGGTTTGGTCTACTTCCATTATAATGCATTTCAGTGGGCAATT
121   S  L  Q  S  D  Y  W  K  N  A  A  F  L  Y  G  L  G  L  V  Y  F  H  Y  N  A  F  Q  W  A  I

451   AAAGCATTTCAGGAGGTGCTTTATGTTGATCCCAGCTTTTGTCGAGCCAAGGAAATTCATTTACGAGTTGGGCTTATGTTCAAAGTGAAC
151   K  A  F  Q  E  V  L  Y  V  D  P  S  F  C  R  A  K  E  I  H  L  R  V  G  L  M  F  K  V  N

541   ACAGACTATGAGTCTAGTTTAAAGCATTTTCAGTTAGCTTTGGTTGACTGTAATCCCTGCACTTTGTCCAATGCTGAAATTCAATTTCAC
181   T  D  Y  E  S  S  L  K  H  F  Q  L  A  L  V  D  C  N  P  C  T  L  S  N  A  E  I  Q  F  H

631   ATTGCCCACTTATATGAAACCCAGAGGAAATATCATTCTGCAAAAGAAGCTTATGAACAACTTTTGCAGACAGAGAATCTTTCTGCACAA
211   I  A  H  L  Y  E  T  Q  R  K  Y  H  S  A  K  E  A  Y  E  Q  L  L  Q  T  E  N  L  S  A  Q

721   GTAAAAGCAACTGTCTTACAACAGTTAGGTTGGATGCATCACACTGTAGATCTCCTGGGAGATAAAGCCACCAAGGAAAGCTATGCTATT
241   V  K  A  T  V  L  Q  Q  L  G  W  M  H  H  T  V  D  L  L  G  D  K  A  T  K  E  S  Y  A  I

811   CAGTATCTCCAAAAGTCCTTGGAAGCAGATCCTAATTCTGGCCAGTCCTGGTATTTCCTCGGAAGGTGCTATTCAAGTATTGGGAAAGTT
271   Q  Y  L  Q  K  S  L  E  A  D  P  N  S  G  Q  S  W  Y  F  L  G  R  C  Y  S  S  I  G  K  V

901   CAGGATGCCTTTATATCTTACAGGCAGTCTATTGATAAATCAGAAGCAAGTGCAGATACATGGTGTTCAATAGGTGTGCTATATCAGCAG
301   Q  D  A  F  I  S  Y  R  Q  S  I  D  K  S  E  A  S  A  D  T  W  C  S  I  G  V  L  Y  Q  Q

991   CAAAATCAGCCCATGGATGCTTTACAGGCCTATATTTGTGCTGTACAATTGGACCATGGCCATGCTGCAGCCTGGATGGACCTAGGCACT
331   Q  N  Q  P  M  D  A  L  Q  A  Y  I  C  A  V  Q  L  D  H  G  H  A  A  A  W  M  D  L  G  T

1081  CTCTATGAATCCTGCAACCAGCCTCAGGATGCCATTAAATGCTACTTAAATGCAACTAGAAGCAAAAGTTGTAGTAATACCTCTGCACTT
361   L  Y  E  S  C  N  Q  P  Q  D  A  I  K  C  Y  L  N  A  T  R  S  K  S  C  S  N  T  S  A  L

1171  GCAGCACGAATTAAGTATTTACAGGCTCAGTTGTGTAACCTTCCACAAGGTAGTCTACAGAATAAAACTAAATTACTTCCTAGTATTGAG
391   A  A  R  I  K  Y  L  Q  A  Q  L  C  N  L  P  Q  G  S  L  Q  N  K  T  K  L  L  P  S  I  E

1261  GAGGCGTGGAGCCTACCAATTCCCGCAGAGCTTACCTCCAGGCAGGGTGCCATGAACACAGCACAGCAGAATACTTCTGACAATTGGAGT
421   E  A  W  S  L  P  I  P  A  E  L  T  S  R  Q  G  A  M  N  T  A  Q  Q  N  T  S  D  N  W  S

1351  GGTGGACATGCTGTGTCACATCCTCCAGTACAGCAACAAGCTCATTCATGGTGTTTGACACCACAGAAATTACAGCATTTGGAACAGCTC
451   G  G  H  A  V  S  H  P  P  V  Q  Q  Q  A  H  S  W  C  L  T  P  Q  K  L  Q  H  L  E  Q  L

1441  CGCGCAAATAGAAATAATTTAAATCCAGCACAGAAACTGATGCTGGAACAGCTGGAAAGTCAGTTTGTCTTAATGCAACAACACCAAATG
481   R  A  N  R  N  N  L  N  P  A  Q  K  L  M  L  E  Q  L  E  S  Q  F  V  L  M  Q  Q  H  Q  M

1531  AGACCAACAGGAGTTGCACAGGTACGATCTACTGGAATTCCTAATGGGCCAACAGCTGACTCATCACTGCCTACAAACTCAGTCTCTGGC
511   R  P  T  G  V  A  Q  V  R  S  T  G  I  P  N  G  P  T  A  D  S  S  L  P  T  N  S  V  S  G

1621  CAGCAGCCACAGCTTGCTCTGACCAGAGTGCCTAGCGTCTCTCAGCCTGGAGTCCGTCCTGCCTGCCCTGGGCAGCCTTTGGCCAATGGA
541   Q  Q  P  Q  L  A  L  T  R  V  P  S  V  S  Q  P  G  V  R  P  A  C  P  G  Q  P  L  A  N  G

1711  CCCTTTTCTGCAGGCCATGTTCCCTGTAGCACATCAAGAACGCGGGGAAGTACAGACACTATTTTGATAGGCAATAATCATATAACAGGA
571   P  F  S  A  G  H  V  P  C  S  T  S  R  T  R  G  S  T  D  T  I  L  I  G  N  N  H  I  T  G

1801  AATGGAAGTAATGGAAACGTGCCTTACCTGCAGCGAAACGCACTCACTCTACCTCATAACCGCACAAACCTGACCAGCAGCGCAAAGGAG
601   N  G  S  N  G  N  V  P  Y  L  Q  R  N  A  L  T  L  P  H  N  R  T  N  L  T  S  S  A  K  E

1891  CCGTGGAAAAACCAACTATCTAACTCCACTCAGGGGCTTCACAAAGGTCAGAGTTCACATTCGGCAGGTCCTAATGGTGAACGACCTCTC
631   P  W  K  N  Q  L  S  N  S  T  Q  G  L  H  K  G  Q  S  S  H  S  A  G  P  N  G  E  R  P  L

1981  TCTTCCACTGGGCCTTCCCAGCATCTCCAGGCAGCTGGCTCTGGTATTCAGAATCAGAACGGACATCCCACCCTGCCTAGCAATTCAGTA
661   S  S  T  G  P  S  Q  H  L  Q  A  A  G  S  G  I  Q  N  Q  N  G  H  P  T  L  P  S  N  S  V

2071  ACACAGGGGGCTGCTCTCAATCACCTCTCCTCTCACACTGCTACCTCAGGTGGACAACAAGGCATTACCTTAACCAAAGAGAGCAAGCCT
691   T  Q  G  A  A  L  N  H  L  S  S  H  T  A  T  S  G  G  Q  Q  G  I  T  L  T  K  E  S  K  P

2161  TCAGGAAACATATTGACGGTGCCTGAAACAAGCAGGCACACTGGAGAGACACCTAACAGCACTGCCAGTGTCGAGGGACTTCCTAATCAT
721   S  G  N  I  L  T  V  P  E  T  S  R  H  T  G  E  T  P  N  S  T  A  S  V  E  G  L  P  N  H

2251  GTCCATCAGATGACGGCAGATGCTGTTTGCAGTCCTAGCCATGGAGATTCTAAGTCACCAGGTTTACTAAGTTCAGACAATCCTCAGCTC
751   V  H  Q  M  T  A  D  A  V  C  S  P  S  H  G  D  S  K  S  P  G  L  L  S  S  D  N  P  Q  L

2341  TCTGCCTTGTTGATGGGAAAAGCCAATAACAATGTGGGTACTGGAACCTGTGACAAAGTCAATAACATCCACCCAGCTGTTCATACAAAG
781   S  A  L  L  M  G  K  A  N  N  N  V  G  T  G  T  C  D  K  V  N  N  I  H  P  A  V  H  T  K

2431  ACTGATAACTCTGTTGCCTCTTCACCATCTTCAGCCATTTCAACAGCAACACCTTCTCCAAAATCCACTGAGCAGACAACCACAAACAGT
811   T  D  N  S  V  A  S  S  P  S  S  A  I  S  T  A  T  P  S  P  K  S  T  E  Q  T  T  T  N  S

2521  GTTACCAGCCTTAACAGCCCTCACAGTGGGCTACACACAATTAATGGAGAAGGGATGGAAGAATCTCAGAGCCCCATGAAAACAGATCTG
841   V  T  S  L  N  S  P  H  S  G  L  H  T  I  N  G  E  G  M  E  E  S  Q  S  P  M  K  T  D  L

2611  CTTCTGGTTAACCACAAACCTAGTCCACAGATCATACCATCAATGTCTGTGTCCATATACCCCAGCTCAGCAGAAGTTCTGAAGGCATGC
871   L  L  V  N  H  K  P  S  P  Q  I  I  P  S  M  S  V  S  I  Y  P  S  S  A  E  V  L  K  A  C

2701  AGGAATCTAGGTAAAAATGGCTTATCTAACAGTAGCATTTTGTTGGATAAATGTCCACCTCCAAGACCACCATCTTCACCATACCCTCCC
901   R  N  L  G  K  N  G  L  S  N  S  S  I  L  L  D  K  C  P  P  P  R  P  P  S  S  P  Y  P  P

2791  TTGCCAAAGGACAAGTTGAATCCACCTACACCTAGTATTTACTTGGAAAATAAACGTGATGCTTTCTTTCCTCCATTACATCAATTTTGT
931   L  P  K  D  K  L  N  P  P  T  P  S  I  Y  L  E  N  K  R  D  A  F  F  P  P  L  H  Q  F  C

2881  ACAAATCCGAACAACCCTGTTACAGTAATACGTGGCCTTGCTGGAGCTCTTAAGTTAGACCTGGGACTTTTCTCTACTAAAACTTTGGTG
961   T  N  P  N  N  P  V  T  V  I  R  G  L  A  G  A  L  K  L  D  L  G  L  F  S  T  K  T  L  V

2971  GAAGCTAACAATGAACATATGGTAGAAGTGAGGACACAGTTGTTGCAGCCAGCAGATGAAAACTGGGATCCCACTGGAACAAAGAAAATC
991   E  A  N  N  E  H  M  V  E  V  R  T  Q  L  L  Q  P  A  D  E  N  W  D  P  T  G  T  K  K  I

3061  TGGCATTGTGAAAGTAATAGATCTCATACTACAATTGCTAAATATGCACAGTACCAGGCCTCCTCATTCCAGGAATCATTGAGAGAAGAA
1021  W  H  C  E  S  N  R  S  H  T  T  I  A  K  Y  A  Q  Y  Q  A  S  S  F  Q  E  S  L  R  E  E

3151  AATGAAAAAAGAAGTCATCATAAAGACCACTCAGATAGTGAATCTACATCGTCAGATAATTCTGGGAGGAGGAGGAAAGGACCCTTTAAA
1051  N  E  K  R  S  H  H  K  D  H  S  D  S  E  S  T  S  S  D  N  S  G  R  R  R  K  G  P  F  K

3241  ACCATAAAGTTTGGGACCAATATTGACCTATCTGATGACAAAAAGTGGAAGTTGCAGCTACATGAGCTGACTAAACTTCCTGCTTTTGTG
1081  T  I  K  F  G  T  N  I  D  L  S  D  D  K  K  W  K  L  Q  L  H  E  L  T  K  L  P  A  F  V

3331  CGTGTCGTATCAGCAGGAAATCTTCTAAGCCATGTTGGTCATACCATATTGGGCATGAACACAGTTCAACTATACATGAAAGTTCCAGGG
1111  R  V  V  S  A  G  N  L  L  S  H  V  G  H  T  I  L  G  M  N  T  V  Q  L  Y  M  K  V  P  G

3421  AGCAGAACACCAGGTCATCAGGAAAATAACAACTTCTGTTCAGTTAACATAAATATTGGCCCAGGTGACTGTGAATGGTTTGTTGTTCCT
1141  S  R  T  P  G  H  Q  E  N  N  N  F  C  S  V  N  I  N  I  G  P  G  D  C  E  W  F  V  V  P

3511  GAAGGTTACTGGGGTGTTTTGAATGACTTCTGTGAAAAAAATAATTTGAATTTCCTAATGGGTTCTTGGTGGCCCAATCTTGAAGATCTT
1171  E  G  Y  W  G  V  L  N  D  F  C  E  K  N  N  L  N  F  L  M  G  S  W  W  P  N  L  E  D  L

3601  TATGAAGCAAATGTTCCAGTGTATAGGTTTATTCAGCGACCTGGAGATTTGGTCTGGATAAATGCAGGCACTGTTCATTGGGTTCAGGCT
1201  Y  E  A  N  V  P  V  Y  R  F  I  Q  R  P  G  D  L  V  W  I  N  A  G  T  V  H  W  V  Q  A

3691  ATTGGCTGGTGCAACAACATTGCTTGGAATGTTGGTCCACTTACAGCCTGCCAGTATAAATTGGCAGTGGAACGGTACGAATGGAACAAA
1231  I  G  W  C  N  N  I  A  W  N  V  G  P  L  T  A  C  Q  Y  K  L  A  V  E  R  Y  E  W  N  K

3781  TTGCAAAGTGTGAAGTCAATAGTACCCATGGTTCATCTTTCCTGGAATATGGCACGAAATATCAAGGTCTCAGATCCAAAGCTTTTTGAA
1261  L  Q  S  V  K  S  I  V  P  M  V  H  L  S  W  N  M  A  R  N  I  K  V  S  D  P  K  L  F  E

3871  ATGATTAAGTATTGTCTTCTAAGAACTCTGAAGCAATGTCAGACATTGAGGGAAGCTCTCATTGCTGCAGGAAAAGAGATTATATGGCAT
1291  M  I  K  Y  C  L  L  R  T  L  K  Q  C  Q  T  L  R  E  A  L  I  A  A  G  K  E  I  I  W  H
```

```
        3961   GGGCGGACAAAAGAAGAACCAGCTCATTACTGTAGCATTTGTGAAGTGGAGGTTTTTGATCTGCTTTTTGTCACTAATGAGAGTAATTCA
        1321   G  R  T  K  E  E  P  A  H  Y  C  S  I  C  E  V  E  V  F  D  L  L  F  V  T  N  E  S  N  S

        4051   CGAAAGACCTACATAGTACATTGCCAAGATTGTGCACGAAAAACAAGCGGAAACTTGGAAAACTTTGTGGTGCTAGAACAGTACAAAATG
        1351   R  K  T  Y  I  V  H  C  Q  D  C  A  R  K  T  S  G  N  L  E  N  F  V  V  L  E  Q  Y  K  M

        4141   GAGGACCTGATGCAAGTCTATGACCAATTTACATTAGCTCCTCCATTACCATCCGCCTCATCTTGAtattgttccatggacattaaatga
        1381   E  D  L  M  Q  V  Y  D  Q  F  T  L  A  P  P  L  P  S  A  S  S  *      1401

        4231   gacctttctgctattcaggaaataacccagttctgcaccactggtttttgtagctatctcgtaaggctgctggctgaaaactgtgtcta
        4321   tgcaaccttccaagtgcggagtgtcaaccaactggacgggagagagtactgctcctactccaggactctcacaaagctgatgagctgtac
        4411   ttcagaaaaaaaataataatttccatgttttgtatatatctgacaaaactggcaacatcttacagactactgacttgaagacaacctcttt
        4501   tatatttctctatttctgggctgatgaatttgttttcatctgtcttttccccccttcagaattttccttggaaaaaaaatactagcctagc
        4591   tggtcatttctttgtaaggtagttagcaattttaagtctttctttggtcaacttttttttaatgtgaaaagttaggtaagacacttttttt
        4681   actgcttttatgtttttctgtcttgttttgagaccatgatggttacacttttggttcctaaataaaatttaaaaaattaacagccaagtc
short   4771   acaaaggtaatggattgcacatagactaaggaataaacttcagatttgtgaaaaaaaaaa     4831
long                                                    └attttgtttctaatcttgatgtaaatttacactattata
long    4861   atacatatttattgcttgaaaatatttgtgaatggaatgctgttatttttttccagatttacctgccattgaaatttaaggagttctgta
long    4951   atttcaaacactactcctattacatttttctatgtgtaaataaaactgcttagcattgtacagaaacttttattaaaattgtttaatgttt
long    5041   aaagagtttctattgtttgagtttaaaaaagactttatgtacagtgcccagttttgttcattttgaaatctgataaatatatttatatat
long    5131   acttatgtatgtatatataatatatatagaaatctggatatatatgtataaatctttagaacttaaatttttctcgtttagttcacatct
long    5221   atggtagatttttgaggtgtctactgtaaagtattgcttacaaaaagtatgattatttttaaagaaatatatatggtatgtatcctcaag
long    5311   acctaaaatgtcagactggtttattgttaagttgcaattactgcaatgacagaccaataaacaattgctgccaaaaaaaaaa    5392
```

UTY

short, medium and long transcripts



```
        1051  T  K  L  P  A  F  A  R  V  V  S  A  G  N  L  L  T  H  V  G  H  T  I  L  G  M  N  T  V  Q
Medium  3151  ACTAAACTTCCTGCTTTTGCGCGTGTGGTGTCAGCAGGAAATCTTCTAACCCATGTTGGGCATACCATTCTGGGCATGAATACAGTACAA
 Short  3151  GCTTGTCTTGAACTCCTGACCTCAGGTGGTCTGCTTGCCTCAGCATCCCAAAGTGCTGGGATTACAGGTGTGAGCCACCATGCCCGGTAA
        1051  A  C  L  E  L  L  T  S  G  G  L  L  A  S  A  S  Q  S  A  G  I  T  G  V  S  H  H  A  R  *   1079

        1081  L  Y  M  K  V  P  G  S  R  T  P  G  H  Q  E  N  N  N  F  C  S  V  N  I  N  I  G  P  G  D
Medium  3241  CTGTATATGAAAGTTCCAGGGAGTCGGACACCAGGTCACCAAGAAAATAACAACTTCTGCTCTGTTAACATAAATATTGGTCCAGGAGAT
 Short  3241  acttttaaaaatgtaagcaaaattacagtatgtaaaacacacattgctaatggagaaataaagttcctacttttacatctaaaaaaaaaa 3330

        1111  C  E  W  F  V  V  P  E  D  Y  W  G  V  L  N  D  F  C  E  K  N  N  L  N  F  L  M  S  S  W
Medium  3331  TGTGAATGGTTTGTTGTACCTGAAGATTATTGGGGTGTTCTGAATGACTTCTGTGAAAAAAATAATTTGAATTTTTTAATGAGTTCTTGG

        1141  W  P  N  L  E  D  L  Y  E  A  N  V  P  V  Y  R  F  I  Q  R  P  G  D  L  V  W  I  N  A  G
Medium  3421  TGGCCCAACCTTGAAGATCTTTATGAAGCAAATGTCCCTGTGTATAGATTTATTCAGCGACCTGGAGATTTGGTCTGGATAAATGCAGGC

        1171  T  V  H  W  V  Q  T  V  G  W  C  N  N  I  A  W  N  V  G  P  L  T  A  C  Q  Y  K  L  A  V
Medium  3511  ACTGTGCATTGGGTTCAAACTGTTGGCTGGTGCAATAACATTGCCTGGAATGTTGGTCCACTTACAGCCTGCCAGTATAAATTGGCAGTG

        1201  E  R  Y  E  W  N  K  L  K  S  V  K  S  P  V  P  M  V  H  L  S  W  N  M  A  R  N  I  K  V
Medium  3601  GAACGGTATGAATGGAACAAATTGAAAAGTGTGAAGTCACCAGTACCCATGGTGCATCTTTCCTGGAATATGGCACGAAATATCAAAGTC

        1231  S  D  P  K  L  F  E  M  I  K  *    1240
Medium  3691  TCAGATCCAAAGCTTTTTGAAATGATTAAGTAAgtgccttctgaaactgctgcagtttctctttgggggtattggtagccattcagtatt
 Long   3723                                   TTGTCTTTTGAAAATTCTGAAGCAATATCAGACATTGAGAGAAGCTCTTGTTGCAGCA
        1241                                Y  C  L  L  K  I  L  K  Q  Y  Q  T  L  R  E  A  L  V  A  A

Medium  3781  tttttcaaaagaattctgttgacattaaatgatatcagcagtccagaagtcttggcaaaatgtaataagatgtaaataatcttatatatt
 Long   3781  GGAAAAGAGGTTATATGGCATGGGCGGACAAATGATGAACCAGCTCATTACTGTAGCATTTGTGAGGTGGAGGTTTTTAATCTGCTTTTT
        1261  G  K  E  V  I  W  H  G  R  T  N  D  E  P  A  H  Y  C  S  I  C  E  V  E  V  F  N  L  L  F

Medium  3871  cataagtgttataaaatctcataagattaaaatattgccttcccttaaaaaaaaaa    3926
 Long   3871  GTCACTAATGAAAGCAATACTCAAAAAACCTACATAGTACATTGCCATGATTGTGCACGAAAAACAAGCAAAAGTTTGGAAAATTTTGTG
        1291  V  T  N  E  S  N  T  Q  K  T  Y  I  V  H  C  H  D  C  A  R  K  T  S  K  S  L  E  N  F  V

 Long   3961  GTGCTCGAACAGTACAAAATGGAGGACCTAATCCAAGTTTATGATCAATTTACACTAGCTCTTTCATTATCATCCTCATCTTGAtatagt
        1321  V  L  E  Q  Y  K  M  E  D  L  I  Q  V  Y  D  Q  F  T  L  A  L  S  L  S  S  S  S  *    1347

 Long   4051  tccatgaatattaaatgagattatttctgctcttcaggaaatttctgcaccactggttttgtagctgtttcataaaactgttgactaaaa
 Long   4141  gctatgtctatgcaaccttccaagaatagtatgtcaagcaactggacacagtgctgcctctgcttcaggacttaacatgctgatccagct
 Long   4231  gtacttcagaaaaataatattaatcatatgttttgtgtacgtatgacaaactgtcaaagtgacacagaatactgatttgaagatagcctt
 Long   4321  ttttatgtttctctatttctgggctgatgaattaatattcatttgtattttaaccctgcagaatttttccttagttaaaaacactttccta
 Long   4411  gctggtcatttcttcataagatagcaaatttaaatctctcctcgatcagcttttaaaaaatgtgtactattatctgaggaagttttttac
 Long   4501  tgctttatgtttttgtgtgttttgaggccatgatgattacatttgtggttccaaaataatttttttaaatattaatagcccatatacaaa
 Long   4591  gataatggattgcacatagacaaagaaataaacttcagatttgtgatttttgtttctaaacttgatacagatttacactatttataaata
 Long   4681  cgtatttattgcctgaaaatatttgtgaatggaatgttgtttttttccagacgtaactgccattaaatactaaggagttctgtagtttta
 Long   4771  aacactactcctattacattttatatgtgtagataaaactgcttagtattatacagaaatttttattaaaattgttaaatgtttaaaggg
 Long   4861  tttcccaatgtttgagtttaaaaaagactttctgaaaaaatccacttttttgttcattttcaaacctaatgattatatgtatttttatatgt
 Long   4951  gtgtgtatgtgtacacacatgtataatatatacagaaacctcgatatataattgtatagatttttaaaagttttatttttttacatctatgg
 Long   5041  tagtttttgaggtgcctattatataaagtattacggaagtttgctgtttttaaagtaaatgtctttttagtgtgatttattaagttgtagtca
 Long   5131  ccatagtgatagcccataaataattgctggaaaattgtattttataacagtagaaaacatatagtcagtgaagtaaatattttaaaggaa
 Long   5221  acattatatagatttgataaatgttgtttataattaagagtttcttatggaaaagagattcagaatgataacctcttttagagaacaaat
 Long   5311  aagtgacttatttttttaaagctagatgactttgaaatgctatactgtcctgcttgtacaacatggtttggggtgaaggggaggaaagta
 Long   5401  ttaaaaaatctatatcgctagtaaattgtaataagttctattaaaacttgtatttcatatgaaaaaaaaaa     5471
```

CDY

```
-281                                                                ctgtggattta
-270 gctactctcacctgaggctactgagcaagttgtcatgcaccatgagacaaagcccaagctgtcccaccaggcagtaagtatggagaggtt
-180 caggcacatggcatagctgctatttcgcacaattttcactacaccagtggtgacaaaatagaagaggttcatccatacacagaacctggt
 -90 gaagagctggaggcagaaagaagtgtctatgtggagacgcaactgaaacaaaggtggcacagcaactgttccaatcccgtgtctttcctc

   1 ATGGCTTCCCAGGAGTTTGAGGTTGAAGCTATTGTTGACAAAAGACAGGATAAAAATGGGAATACACAGTATTTGGTTCGGTGGAAAGGT
   1 M  A  S  Q  E  F  E  V  E  A  I  V  D  K  R  Q  D  K  N  G  N  T  Q  Y  L  V  R  W  K  G

  91 TATGACAAACAGGATGACACTTGGGAACCAGAGCAGCACCTCATGAACTGTGAAAAATGTGTACATGATTTTAATAGACGACAGACTGAA
  31 Y  D  K  Q  D  D  T  W  E  P  E  Q  H  L  M  N  C  E  K  C  V  H  D  F  N  R  R  Q  T  E

 181 AAACAGAAAAAACTGACATGGACTACAACCAGTAGAATTTTTTCAAACAATGCCAGAAGAAGAACTTCCAGATCTACAAAAGCAAACTAT
  61 K  Q  K  K  L  T  W  T  T  T  S  R  I  F  S  N  N  A  R  R  R  T  S  R  S  T  K  A  N  Y

 271 TCTAAGAACTCTCCTAAAACGCCAGTGACTGATAAACACCACAGGTCCAAAAACCGCAAGTTATTTGCTGCCAGCAAGAACGTTAGGAGA
  91 S  K  N  S  P  K  T  P  V  T  D  K  H  H  R  S  K  N  R  K  L  F  A  A  S  K  N  V  R  R

 361 AAGGCAGCTTCAATTCTCTCCGACACAAAGAATATGGAGATAATAAATTCAACTATTGAGACCCTTGCACCTGACAGCCCCTTTGACCAC
 121 K  A  A  S  I  L  S  D  T  K  N  M  E  I  I  N  S  T  I  E  T  L  A  P  D  S  P  F  D  H

 451 AAAACTGTGAGTGGCTTTCAGAAACTTGAGAAACTGAACCCTATTGCAGCAGATCAGCAGGACACGGTGGTCTTCAAGGTGACAGAAGGG
 151 K  T  V  S  G  F  Q  K  L  E  K  L  N  P  I  A  A  D  Q  Q  D  T  V  V  F  K  V  T  E  G

 541 AAACTCCTCCGGGACCCTTTGTCACGTCCTGGTGCAGAACAGACTGGAATACAGAACAAGACTCAGATACACCCACTAATGTCGCAGATG
 181 K  L  L  R  D  P  L  S  R  P  G  A  E  Q  T  G  I  Q  N  K  T  Q  I  H  P  L  M  S  Q  M

 631 TCTGGCTCAGTTACTGCTTCTATGGCCACAGGTTCAGCTACCCGAAAGGGTATAGTGGTATTAATAGACCCATTAGCAGCCAATGGGACA
 211 S  G  S  V  T  A  S  M  A  T  G  S  A  T  R  K  G  I  V  V  L  I  D  P  L  A  A  N  G  T

 721 ACAGACATGCATACCTCAGTTCCAAGAGTGAAAGGTGGGCAAAGAAATATTACTGATGACAGCAGAGACCAGCCTTTTATCAAGAAGATG
 241 T  D  M  H  T  S  V  P  R  V  K  G  G  Q  R  N  I  T  D  D  S  R  D  Q  P  F  I  K  K  M

 811 CACTTCACCATAAGGCTAACAGAAAGTGCCAGCACATACAGAGACATTGTAGTGAAGAAAGAGGATGGATTCACCCAGATAGTGCTATCA
 271 H  F  T  I  R  L  T  E  S  A  S  T  Y  R  D  I  V  V  K  K  E  D  G  F  T  Q  I  V  L  S

 901 ACTAGATCGACAGAAAAAAATGCACTGAATACAGAAGTAATTAAAGAAATAGTTAATGCTCTGAATAGCGCTGCTGCAGATGACAGCAAG
 301 T  R  S  T  E  K  N  A  L  N  T  E  V  I  K  E  I  V  N  A  L  N  S  A  A  A  D  D  S  K

 991 CTCGTGCTGTTCAGTGCAGCTGGAAGTGTCTTTTGCTGCGGTCTTGATTTTGGGTACTTTGTGAAGCACTTAAGGAATAACAGAAACACA
 331 L  V  L  F  S  A  A  G  S  V  F  C  C  G  L  D  F  G  Y  F  V  K  H  L  R  N  N  R  N  T

1081 GCAAGCCTTGAAATGGTGGACACCATCAAGAACTTTGTGAATACTTTTATTCAATTTAAAAAGCCTATTGTTGTATCAGTCAATGGCCCT
 361 A  S  L  E  M  V  D  T  I  K  N  F  V  N  T  F  I  Q  F  K  K  P  I  V  V  S  V  N  G  P

1171 GCGATTGGACTAGGTGCATCCATCCTGCCTCTTTGTGATCTCGTGTGGGCTAATGAAAAGGCTTGGTTCCAAACCCCTTATACGACCTTT
 391 A  I  G  L  G  A  S  I  L  P  L  C  D  L  V  W  A  N  E  K  A  W  F  Q  T  P  Y  T  T  F

1261 GGACAGAGTCCAGATGGCTGTTCTTCTATTACATTCCCCATAATGATGGGTAAAGCATCTGCCAATGAAATGTTAATTGCTGGGCGAAAG
 421 G  Q  S  P  D  G  C  S  S  I  T  F  P  I  M  M  G  K  A  S  A  N  E  M  L  I  A  G  R  K

1351 CTGACAGCAAGGGAGGCATGCGCCAAAGGCCTGGTCTCTCAGGTATTTTTGACTGGAACTTTCACCCAAGAGGTTATGATTCAAATTAAG
 451 L  T  A  R  E  A  C  A  K  G  L  V  S  Q  V  F  L  T  G  T  F  T  Q  E  V  M  I  Q  I  K

1441 GAGCTTGCCTCATACAATCCAATTGTACTGGAAGAATGTAAGGCCCTCGTTCGCTGTAATATTAAGTTGGAGTTGGAACAGGCCAATGAG
 481 E  L  A  S  Y  N  P  I  V  L  E  E  C  K  A  L  V  R  C  N  I  K  L  E  L  E  Q  A  N  E

1531 AGAGAGTGTGAGGTGCTGAGGAAGATCTGGAGCTCAGCCCGAGGGATAGAATCCATGTTAAAAATACCTCTGTTGGGATATAAAGCAGCC
 511 R  E  C  E  V  L  R  K  I  W  S  S  A  R  G  I  E  S  M  L  K  I  P  L  L  G  Y  K  A  A

1621 TTCCCTCCCAGAAAGACACAGAATGATCAGAGATGGTGCCCTTGActttatagtggcacaaacgcttcagagacacacaattataagaga
 541 F  P  P  R  K  T  Q  N  D  Q  R  W  C  P  *      554

1711 cttatcttttagcataaatacttatggctcaaaatccactgacgatcattctcctaaactgaacacatgactagaattggtggtgagata
1801 tcgcttgattttcttttccttttataaaatgtctagttcttacccagttaacaaaagaaaactttatcgctctaaagtaaaacttgttacac
1891 cacaaaaaaaaaa      1903
```

BPY1

```
-72                    gagaggggtatacacagggaggccaggcagcctggagttagtcgaccgttgcgagacgttgagctgcggcag

  1 ATGAGTCCAAAGCCGAGAGCCTCGGGACCTCCGGCCAAGGCCAAGGAGACAGGAAAGAGGAAGTCCTCCTCTCAGCCGAGCCCCAGTGGC
  1 M  S  P  K  P  R  A  S  G  P  P  A  K  A  K  E  T  G  K  R  K  S  S  S  Q  P  S  P  S  G

 91 CCGAAGAAGAAGACTACCAAGGTGGCCGAGAAGGGAGAAGCAGTTCGTGGAGGGAGACGCGGGAAGAAAGGGGCTGCGACAAAGATGGCG
 31 P  K  K  K  T  T  K  V  A  E  K  G  E  A  V  R  G  G  R  R  G  K  K  G  A  A  T  K  M  A

181 GCCGTGACGGCACCTGAGGCGGAGAGCGGGCCAGCGGCACCCGGCCCCAGCGACCAGCCCAGCCAGGAGCTCCCTCAGCACGAGCTGCCG
 61 A  V  T  A  P  E  A  E  S  G  P  A  A  P  G  P  S  D  Q  P  S  Q  E  L  P  Q  H  E  L  P

271 CCGGAGGAGCCAGTGAGCGAGGGGACCCAGCACGACCCCCTGAGTCAGGAGAGCGAGCTGGAGGAACCACTGAGTAAGGGGCGCCCATCT
 91 P  E  E  P  V  S  E  G  T  Q  H  D  P  L  S  Q  E  S  E  L  E  E  P  L  S  K  G  R  P  S

361 ACTCCCCTATCTCCCTGAgcagcaactaagtttaggcccagctgccagacctcagagatctcaccagcagggtgcttcccatgttgatga
121 T  P  L  S  P  *   125

451 caataaaatgaatgtgttgcaaaaaaaaaa  480
                     94
```

BPY2

```
-332                                  aatatctcaggacccaggaccatgtgatatgggcccaacacctggatgatgttactcttctg
-270 cctaggtcatgcgtaaagagggaattagggcatattgcttggcccagtcccgtaatgatatgactctcctgcttgtgccagagccacaga
-180 agtgtgcttggtgacataatctttgaggctgtcacatcaccaagattatattgtatcactggaccagcataaagctgacacttctgacta
 -90 tgcccagccttcaaataatactacactgtataattggctcaacacccaggtgatattgttccatttacctgagaccagataaaaagccta

  1 ATGATGACGCTTGTCCCCAGAGCCAGGACACGTGCAGGACAGGATCATTACTCTCATCCCTGCCCCAGATTTTCACAGGTGCTGCTTACA
  1 M  M  T  L  V  P  R  A  R  T  R  A  G  Q  D  H  Y  S  H  P  C  P  R  F  S  Q  V  L  L  T

 91 GAGGGCATCATGACATATTGCTTGACAAAGAACCTAAGTGATGTTAATATTCTGCATAGGTTGCTAAAAAATGGGAATGTGAGAAATACC
 31 E  G  I  M  T  Y  C  L  T  K  N  L  S  D  V  N  I  L  H  R  L  L  K  N  G  N  V  R  N  T

181 TTGCTTCAGTCCAAAGTGGGCTTGCTGACATATTATGTGAAACTGTACCCGGGTGAAGTGACTCTTCTGACTAGGCCCAGCATACAAATG
 61 L  L  Q  S  K  V  G  L  L  T  Y  Y  V  K  L  Y  P  G  E  V  T  L  L  T  R  P  S  I  Q  M

271 AGATTATGCTGTATCACTGGCTCAGTGTCGAAGCCCAGATCACAGAAGTAAttgtgccatatgtggaacaagcagctaagcaatagataa
 91 R  L  C  C  I  T  G  S  V  S  K  P  R  S  Q  K  *  106

361 catccatcgtggctctgccttcaaagggaaattttacatatgtcactgggaccatcacccagatgatgtcctgcccactaaaagaattgt
451 gacataacgctgactgcaaaaactgggtaatgcaactctcctctttattctggagtctgccaaaacaagggattatcacatattgcggag
541 tccagcacccaggtaaaattttgtcatatacccagcttcagataccatgcaatgatacaactatcatacctggacccaaagaggagagat
631 attttgattctcattgccattcttatggccacaagcaaagtaatggttctcatagtggtataaagttcacacagtattatgacactccca
721 gcgtatcatagaaaatgtgagtagtacaatgagtgttataacagggaacagcaaaccaatgctattgtgattattggattcacacccagc
811 tgacgcgactatcattctctcacaagaacagaacctgcaaataaagtactaaatctcaccaaaaaaaaaa   880
```

XKRY

```
-663                                                                  attaaaaacttctgataaaattacctaagtaca
-630 cacaaacaaaaacatgcccacacaaatcacttaatttctaaaacttttaatttttctgcttctctagtaccttgtattccatcacacagc
-540 aaaatctggcagctccacttccagaatttacttgaactccacagcttatttccgatttcctgttatcaccagagtctaaaacacagttta
-450 tattgcattcacctcctattttacaccgtaatttcctactttacactctaactttatataaaaaagaaactacctttttcaagatctaatt
-360 cacgcaattttatttgttcttaattgagacttctttctaggtgctgtcacaccttgtaacgtcagatacaaatgtctctatccaatttca
-270 tgagttccagttatttttattttaagggaatgtgtatatacatttataaatttgtgtatgtgtgtattcacttattctttatttttatatgt
-180 tttgcatgcatatattcactaaatccctgataatagaaagataacaaatctttttttttctttctttttttgtatgtaaattattttccga
 -90 aggaggtgggttgggagaaatatatcttaacttggcaagtttaaaagagaaagtggccattactaatgaaaattattctctagcattttc

  1 ATGTTTATCTTTAATAGCATTGCTGATGACATATTCCCTCTTATCAGTTGTGTAGGTGCCATTCACTGCAATATACTGGCCATCCGCACT
  1 M  F  I  F  N  S  I  A  D  D  I  F  P  L  I  S  C  V  G  A  I  H  C  N  I  L  A  I  R  T

 91 GGCAACGACTTTGCTGCCATTAAGCTACAGGTGATAAAATTGATCTATCTCATGATATGGCATTCGTTGGTGATTATCTCACCTGTAGTG
 31 G  N  D  F  A  A  I  K  L  Q  V  I  K  L  I  Y  L  M  I  W  H  S  L  V  I  I  S  P  V  V

181 ACTCTGGCATTCTTCCCTGCATCTCTGAAACAGGGGAGCTTACACTTTCTATTAATCATATATTTTGTATTATTGTTGACACCATGGCTG
 61 T  L  A  F  F  P  A  S  L  K  Q  G  S  L  H  F  L  L  I  I  Y  F  V  L  L  L  T  P  W  L

271 GAGTTTTCGAAAAGTGGAACTCATCTTCCTAGCAACACAAAAATAATTCCAGCATGGTGGGTAAGTATGGATGCTTATCTTAATCATGCT
 91 E  F  S  K  S  G  T  H  L  P  S  N  T  K  I  I  P  A  W  W  V  S  M  D  A  Y  L  N  H  A

361 AGTATATGCTGCCATCAATTCTCCTGCTTGTCAGCAGTGAAACTGCAGCTGTCAAATGAGGAATTGATAAGAGACACGAGGTGGGACATA
121 S  I  C  C  H  Q  F  S  C  L  S  A  V  K  L  Q  L  S  N  E  E  L  I  R  D  T  R  W  D  I

451 CAATCCTACACTACAGATTTCAGTTTTTAGaaaatgtgataataatattgatatttagtttctttggagggaacgttttaccgaagtgtt
151 Q  S  Y  T  T  D  F  S  F  *   159

541 gtgactcaataattgccgtgtagttcatcaaaacctacatattagccctttggctttaagctccgcttctgtcagtatttgcaaccaaggt
631 ggtcgggcaaagtattgccaggagatactgaaaatcatccagaagcactgtgatattgtgtaagcatctggagaaaattcagttaaaaga
721 ataaaagtaagcagctgaggaattactatcactcatggagaagggtaggatattttcaataagtgagtatgcaatatccatatatacttt
811 cacagaacaaagagtaaagaggctgagtgtgactttataaagatactcatgaaaaatataaacaacaaaaccttggaagtagtttctaat
901 aaaattgatttttctaaaaaaaaaa    925
```

PRY

```
-182                                                                                               aa
-180 gaagaggagcacaccacaccagaaacagacatcttgcagtgtttcactgtctcaaccttatctgcacagtccgaggtcagtctgagagag
 -90 cttctgagagacccaggatgaagggatgcagtgaggtcaagagcccaaccttctttcactgacacccacctctaaggactcagaagagac

   1 ATGAATAAAATGGGCCTCAACAATCCCAAGAAGAACCACTCAAGGACAATGGGAGCCACTGGGCTTGGCTTCCTACTTCCCTGGAAACAA
   1 M  N  K  M  G  L  N  N  P  K  K  N  H  S  R  T  M  G  A  T  G  L  G  F  L  L  P  W  K  Q
  91 GACAATTTGAATGGCACTGACTGCCAGGGATGCAATATTTTATACTTCTCTGAGACTACGGGGAGCATGTGTTCTGAACTTTCCCTGAAC
  31 D  N  L  N  G  T  D  C  Q  G  C  N  I  L  Y  F  S  E  T  T  G  S  M  C  S  E  L  S  L  N
 181 AGAGGTCTTGAGGCCAGAAGGAAGAAGGATCTTAAAGACTCATTTCTCTGGAGATATGGGAAGGTTGGCTGTATCTCACTTCCACTTCGT
  61 R  G  L  E  A  R  R  K  K  D  L  K  D  S  F  L  W  R  Y  G  K  V  G  C  I  S  L  P  L  R
 271 GAGATGACCGCCTGGATTAACCCACCCCAAATTTCAGAGATTTTCCAAGGCTACCACCAGAGGGTGCACGGAGCTGATGCACTGAGCCTG
  91 E  M  T  A  W  I  N  P  P  Q  I  S  E  I  F  Q  G  Y  H  Q  R  V  H  G  A  D  A  L  S  L
 361 CAAACCAACTCTCTGAGAAGCAGGTTATCTTCACAGTGCCTCGGACAGAGCTTCCTTCTCAGGACACTCGAGAGAGCCGTGGTTTCAGGG
 121 Q  T  N  S  L  R  S  R  L  S  S  Q  C  L  G  Q  S  F  L  L  R  T  L  E  R  A  V  V  S  G
 451 CACTTGGGGACATCTGTGGCCACGTTCATGAAGAAGACTAAGCCTACTTCATCTCAGGACCCGCCCAAGAGTGGCCGCGGCTTTGGGACA
 151 H  L  G  T  S  V  A  T  F  M  K  K  T  K  P  T  S  S  Q  D  P  P  K  S  G  R  G  F  G  T
 541 CCTGCGGTCGGGTCCACCATGAGGATAAAACCTCCTTCTCTTCTGGACATGTCCAGGAGTGGCCGTTGCTACAAGTCACCTGGTGCTACG
 181 P  A  V  G  S  T  M  R  I  K  P  P  S  L  L  D  M  S  R  S  G  R  C  Y  K  S  P  G  A  T
 631 ACCAGGGTGAGAATAAAGACGTCTCCTCAGGACCCTCCCAGGAGAGTACATGGCATTGAGACATCTGGCGGCCAAGTGAGGAAAAGACAC
 211 T  R  V  R  I  K  T  S  P  Q  D  P  P  R  R  V  H  G  I  E  T  S  G  G  Q  V  R  K  R  H
 721 CCTGTCTGCAGCACCCAGAACTGAggaggggcactgccctgggccttacttcccagccctggcctccaattctgaccttacaaaagtgtc
 241 P  V  C  S  T  Q  N  *  247
 811 ccttgagtgaggcagtgaccacgcattgtcacagctaccaaagtgtggtttgcagatgatctgggcttgtttctggcagagattctggta
 901 cagagaaaggagaggcgttgagtggaaccacgatgggctgaggccaggggagacatcacaacctccaacaacactttttttcatgcttta
 991 ataaatcatttttcttagagaactaaagtagttgaaacaatatagaaacatttttaagtaggcataaaaaaaaaa   1066
```

TTY1

```
tgtctgtcagagctgtcagcctgcttaagcagagtaaaatggtacaggcagtgcagcctggtagcgagaaaaaaggctgcctgtgaaatc
ccactgtgggaccataagtggggacctcagggccccttcatggcatctccatggccatgtcatgctggagaaggaggcgtttcaagaatg
tgagctgatcgctggaaactgctcatctgactccagtctcaaaagaggctatgtgcaagaatcgggtgaagttgtgagaccccatccacc
cctcacaagattgtatccccaccctgtctgaccttactgctgctcaaactatctgtccaaggatgaaaacccaggacaaaggaggagtaa
ccctcatgatgtgaagcacgtgttcacctgtgaatataacctgaggatcatgagactatctgtggatttcacagagaagacagacgagaa
gacaccgctgacacttctccacggaggtctccttttccaccaagatgcagatgcttcttgcaaggactatcctgtgaatcccacagagaa
gacaggtgtggttccaatgccggtgcacctccagggaattctccttctctaccaagctccaggccttctgccatgatcatgagactattt
gtggatttcacagagaagataggtgaaggtacagcatggcatccacccctcaccagaggggtatccccacccctatctgaccttattacc
ttattgctgttcaaagtctctatcccagactgaaatcccaagacaatggagaagttccccctgatgatgtgaagcaccaactcctctggg
aatcaaattcgaggtaaatttaataggcccggtagagatgaatgatagtgtctctccttggattggctgaaagacaattaaacactggta
tatttctgttaaaaaaaaaa
```

TTY2

```
aggcttgccatcaccacagatggcctctgagacactgtttgaaccacatctgcacctgtgagaggccagtttgaggtatgagaacactgt
ttcaatttggacttgcctttgtcttggttcctgcttttcccagatggcacctacccaacccaggatgaatgagtgcagagaggtcaagtg
ccaggccatcttttgctgacacccttttctggtatttcaggtataagtccatcatccaaagactgctcaacatctcaccagaatatattt
caatcctcatggggcatgattctttcacaaaacccctttcaggaatggagtcagaagagtagtttccagagacaacctcacagtcttgga
acggctctgcctcccatgtgatctgaccatggagatggcatataagggccctaagtttgagacttttagggtactgcaatgcgttatcac
aggcagcctttatcctgataccaagccagctctgcctgtaccatttcctctgcttaggcaggctgacagccctgacaccctggtgctcc
agtttgagtcactatatgtggatgtgctagtcttggggcaatggacctgagctgtgagctgtagctagtgtcacaatgaatgccagcttt
gctagtaacaattccctttggcttggtagagaaggggacctctgtggaggtacaatggtggtgcactgtcacctgtcttctctgtgggat
ccatgggacagttccatgatcctaggagagggtagatgtgagccagcctgaagaaatgtcaagcagagccccaggaatgaagcacaaaat
cactacagatccaaaaggatctgcagaatttgtcaggcctgcctagacattgtaggggttagtcttattgaaatgtgtcccactgtaatt
tccaacttcagccttcctgtgttcccagcagtttctctctcccaggtggggctttctgcagaatgacacagcctcagaagctactgggct
gtgtgttactgtgggagtgttgcgagtgttggatgtcagcatgtgtgtgtggctttgtgtgtgtgtgtaggtgtctgtgtgtgtgtatgt
aaatgaattctgtggatcaggaatcagcaatgactagttaagctgtctgtgaccagccgggttccccatcgtctgcccctgccaaaaaaa
caggtactcttctacaaagaagaggagagcaccacacccaagaacagacatctcccagtgttgcattataaagcagccaacccacagaca
ctagcactctggtctgcatagcccctttaatttacctagaattcagttcccagccaagtaggtgcttcatgtcctgagggtgcaatcctc
catcatcttgagatttcatgctggtacagagagtgtgacagcaataaggtcagataggggtgagtatacaacctggtgaagggtggatgg
ggtcccgtaccttcaccagcaaaaagggtgaaaatagatgacacagaatgtgcttccaactccatccccacattcccataattgcaaaat
cagtcaacaacatggcctggtgtttaggtgggagtactccaacctgcaggaagaatttggagtgcaaattgtggccaatctggaaaactc
ctggtttgagggttttaatacctgtagtcaaatggaagtggaatagattgatgctgggtgggttgtggcctccacatttgtgtcctcttt
tactgacttccattgtcctcattggtgtagggcttcctggatctggctcaacatcttccacactaaactcttccctgttcacagaagacc
atcataaaaatgcattgtagaggccctgcaaggaccaggatgaagggagacagtgaggtcaagagcccagccatctttcactgacaccca
cttctgggttctcaggctggctgacaggtctgacagcccatcacgaaagcctgcatactcttagacacaaggactgagctatgggctcca
gctagcatcacaatgaaggccaccattgcctagggataagtccctgtgactttgtggataagaactccgtggagccaatccaaggagaga
cactatcattcatctctaccgggcctattaaatttacctcgaatttgattcccagaggagttggtgcttcacatcatcaggggaacttc
tccattgtcttgggatttcagtctgggatagagactttgaacagcaataagtttccttggtctggctcaacgtcttctaaactcaacatt
ccccagttcatggaaaatgatcctcatgggattctattgtaaaagctgcatgattcctggagaatgttttattgtcctggagtcaaccca
aaaatcaacactaccagtcatgttgcagagctccttgaattaaccttgaattcagtttccagctgagcagctgcttcacgttgtgagggg
gcaatcctccatcatcctgggatttcattctgggacacagagtttgagcagcaataaggctgggtccacattgcccctcaacagcattag
tggacatgattgtcagacttgcaatttccgcagacaccttctgtgaacatttttcaacatcatctacatgagtgagagacccgttcgaca
tgtaagaatactgcttgactttggacctgcctttgtcgtggttcctgcctttctcatagatcccctgccaggcccaggatgataggaggc
aatgaagtcaagggccgagccccattcattgaaagctgactctggggtctcgggtataattccatcacataaaatcccctcaacaactca
ccagactatattccaatctccatggaacctgattcttgcacacagcctctttcaggaatggagtcagaagagcagtcttcagagaccacc
tcagtttggaaaagcctcctccttcagtggttcccagccatggagtcatcgtgaaggggctccatggtcaataatttacggtactgcac
ttggttatcacagacagactttttcatgatagcatgccatctctgtctatatcattttcctctgcttaggcaggctgacaactctgacag
ccaggggcccgaatctacctggcaaatgtgcatgctctactctcagtgcaaaaggcctgtttgggagttctgactagtgtcacaataaat
gccgccattgcctagtgaaaaaaaaaa
```

CHAPTER 3: X/Y HOMOLOGOUS GENES

(This chapter is adapted from a manuscript for publication, with *Genomics* format.)

Original title of the manuscript:

Persistence of Y Copies of X/Y Homologous Genes in Humans Due to Functional Importance or Recent Acquisition

Bruce T. Lahn and David C. Page

Howard Hughes Medical Institute, Whitehead Institute, and Department of Biology, Massachusetts Institute of Technology, 9 Cambridge Center, Cambridge, Massachusetts 02142, USA

The X and Y sex chromosomes in mammals, highly dissimilar in size and DNA content, nevertheless share many genes in common. These genes are either X/Y identical, found in pseudoautosmal regions, or X/Y homologous but non-identical, found in regions restricted to either sex chromosome. The two heteromorphic sex chromosomes are thought to have evolved from a pair of autosomes. Due to lack of recombination, however, genes the Y originally shared with the X have mostly degenerated. Occasionally, the Y can acquire regions of homology with the X anew through transpositions. But genes within such regions would similarly degenerate in the absence of recombination. X/Y homologous genes therefore, are vestiges of much greater homology between ancestral X and Y. Nine pairs of functional X/Y homologous genes are known in humans. We carried out a systematic comparison among them, and had the following observations: 1) map order of Y copies was dissimilar with that of X copies; 2) gene pairs may differ in the degree of divergence between X and Y copies; 3) where tested, all X copies escaped X-inactivation; and 4) X and Y copies were typically expressed at comparable levels, but with clear exceptions. Further synthesis of these observations suggested that the persistence of Y copies of X/Y homologous genes could be accounted for in two ways. Either, the gene had an important housekeeping function, and was conserved on the Y to provide sufficient dosage in males; or it was a recent acquisition on the Y, with insufficient time to degenerate. These findings extended our understanding of sex chromosome evolution. They were also relevant to understanding Turner syndrome, a human condition closely related to X/Y homologous genes.

INTRODUCTION

Mammals have the XX:XY system of sex determination, where the Y chromosome dominantly triggers male differentiation. X/Y homologous genes found in many mammalian species are thought to reflect the common ancestry of the two heteromorphic sex chromosomes. According to prevailing views (Bull, 1983, Charlesworth, 1991, Graves, 1995, Graves, 1996), the two highly dissimilar sex chromosomes have evolved from a pair of autosomes. Sex chromosomes initially come about when one of a pair of autosomes acquires the male-determining locus (popularly referred to in mammals as the *Testis Determining Factor*, or *TDF*). The chromosome carrying *TDF* becomes the Y. Its original pairing partner becomes the X. Due to tight linkage, regions close to *TDF* are rarely exchanged from Y to X. Crossover suppression causes these regions to gradually diverge from the X. Sequence divergence further inhibits recombination, which leads to the divergence of regions even further away from *TDF*. In essence, crossover suppression around *TDF* has a tendency to spread outward. Since recombination is essential in weeding out deleterious mutations, once a region of the Y ceases to recombine with the X, its gene content tends to degenerate (Muller, 1914, Charlesworth, 1978, Rice, 1987, Rice, 1994). As crossover suppression approaches completion, the Y becomes a non-recombining chromosome with little homology to the X. Genes it originally shared with the X have largely degenerated. While the Y tends to become increasingly male-specific, it can occasionally acquire homology with the X anew through translocations (Graves, 1995, Schwartz *et al.*, 1997). But due to crossover suppression, the newly acquired region would again diverge from the X and its gene content would similarly degenerate. Regardless of when X/Y homologies were acquired during sex chromosome evolution, X/Y homologous genes are their remaining vestiges.

In humans, X/Y homologous genes have important implications for Turner syndrome (TS), a condition caused by partial or complete sex chromosome monosomy (*i.e.*, 45,XO karyotype). Due to X-inactivation, genes on the X are typically expressed from one copy per female cell, resulting in equal dosage in females, males, and coincidentally, 45,XO individuals. X copies of X/Y homologous genes appear to be an exception. In the few cases tested, they escape X-inactivation

(Schneider-Gädicke *et al.*, 1989, Fisher *et al.*, 1990, Agulnik *et al.*, 1994, Jones *et al.*, 1996). These genes are therefore expressed from both X's in females, and from X and Y in males. This poses a problem for 45,XO individuals as the dosage of their X/Y homologous genes is reduced in comparison to either females or males. If for some X/Y homologous genes, proper dosage is essential for development, haploinsufficiency of these genes in 45,XO individuals would lead to developmental anomalies found in TS. The molecular etiology of TS is reviewed by A.R. Zinn and colleagues (1993).

Nine pairs of functional X/Y homologous genes are known in humans. They are *ZFX/Y*, *RPS4X/Y*, *AMELX/Y*, *SMCX/Y*, *DFFRX/Y*, *DBX/Y*, *UTX/Y*, *TB4X/Y* and *EIF1AX/Y* (Page *et al.*, 1987, Fisher *et al.*, 1990, Salido *et al.*, 1992, Agulnik *et al.*, 1994, Jones *et al.*, 1996, Lahn and Page, 1997). These genes have survived, perhaps not coincidentally, the degeneration of the Y. Prior studies of X/Y homologous genes were focused on the particulars of their biology. Here, we systematically compared these nine pairs of genes with respect to the following questions: 1) are relative map orders similar between X and Y? 2) is sequence divergence between X and Y copies comparable for all pairs? 3) do all X copies escape X-inactivation? and 4) for each pair, how do levels of expression compare between X and Y copies. With these comparisons, we wish to uncover coherent themes among these genes that might have allowed them to resist the tendency of the non-recombining Y to degenerate. Our results may also bear relevance on Turner syndrome.

MATERIALS AND METHODS

Mapping by radiation hybrids. X copies of X/Y homologous genes were mapped by PCR, using DNA samples from the Genebridge 4 radiation hybrid panel (Walter *et al.*, 1994, Hudson *et al.*, 1995). For each gene, a series of positive or negative PCR results were obtained, and compared to reference markers previously mapped using the same method. Genes were localized near reference markers that shared the greatest concordance of PCR results. Map distances of genes from these reference markers were calculated (Cox *et al.*, 1990). All PCR assays consisted of 30 cycles of the following conditions: 1 min denaturing at 94°C, 45 sec annealing at 60°C, and 45 sec extension at

72°C. *TB4X* primers were designed from unreported intron sequence (unpublished data). *RPS4X* primers were also designed from unreported intron sequence (Kawaguchu *et al.*, unpublished data). All other primers were designed from published sequences. PCR primers were as follows:

Gene	Left primer	Right primer
DBX	CTACATGCAGATGACATGGTG	GGCCAAGGTGCATAGGTG
UTX	CATGTTCCCTGTAGCACATC	CGTTTCCATTACTTCCATTTCCTG
TB4X	CCCGCCCTTTCATCATCC	GCTCCCCAAAGTAGCCTTC
EIF1AX	CACGAGGCGCCATTTGCTG	CTGGAGGCCAGGCAACGTG
DFFRX	CCTCCACCTGAAGATGCC	CTGAGATCCAGGTGAATGG
ZFX	CAGAACGTTTGTAGAGATATTGG	GCATGTTGCCAGTTTTCCCT
SMCX	GAGATACTGCGTCTTCTCC	GGTTCAGGTCTAGCTTCTCT
AMELX	CTGCTGCTTCTCTGGTTGG	ACTGGTGAGAAACAGAGAGG
RPS4X	TGAGATGGATTGAATGTGGC	TTAAAGAGGGTGCCCAGGTA

Electronic sequence comparisons. Each pair of X/Y homologous genes were aligned and their sequence divergence analyzed using MegAlign (DNASTAR Inc., Madison, WI).

Assay of gene activity by RT-PCR. cDNA was made from tissues or cultured cells using two kits: MicroScale mRNA Isolation kit (Pharmacia, Uppsala, Sweden), and Marathon cDNA Synthesis kit (Clontech, Palo Alto, CA). PCR was carried out on cDNA using the same conditions as those used for mapping. PCR primers for *EIF1AX* were the same as those used for mapping. PCR primers for the rest of the genes were as follows:

Gene	Left primer	Right primer
DBX	CTTCCCTCTGTTCTCTCCTC	GTGAAGCAGGAAACGGTGG
UTX	CTAGCCTAGCTGGTCATTTC	CCATCATGGTCTCAAAACAAG
DFFRX	CACAATAGACCAAGATGATGAGT	TAATCTGATGAGGTCTGGTGG
TB4X	GCAGGGAAGGAAAGAACTTGC	CTTCCACCCCACTTCTTCC

ALD	CAAAGTCTACCCCTTGGTG	GGCACTGTTGACGAAGGTA

Assay for DNA methylation by restriction digest and PCR. To assay the presence/absence of CpG methylation within a DNA fragment, genomic DNA was digested with methylation-sensitive restriction enzymes that cut within the fragment only when cut sites were unmethylated. Following digest, standard PCR was carried out with primers flanking the fragment. The presence/absence of PCR product was interpreted as the presence/absence of methylation. Primers used for assaying methylation status for *DBX*, *TB4X* and *EIF1AX* were identical to those used for mapping. Primers for *ALD* (Jegalian and Page, unpublished data) and *ZFX* (Luoh *et al.*, 1995) were previously developed. They were as follows:

Gene	Left primer	Right primer
ALD	GTGACATGCCGGTGCTCTCCA	CGCTGCAGGAATACCCGGTTCAT
ZFX	CTACCCTTCCGCATTTTCCT	GAGCTCGGAGCTGACAAAAA

Compare relative levels of expression between X and Y copies of X/Y homologous genes. For each pair of genes, primers were designed within sequences shared identically between X and Y copies, such that they would amplify from these two copies with the same efficiency. PCR was carried out on cDNA made from male tissues. PCR conditions were the same as those used for mapping. For each PCR reaction, two forms of amplification products were made, one from the X copy, and one from the Y copy. To minimize the amount of DNA heterodimers with one strand from the X isoform and the other from the Y isoform, PCR products were diluted 20X with fresh PCR mix and re-amplified 3 more rounds. PCR products were purified by gel-filtration with Sephadex G-50. DNA was digested with enzymes that would recognize only one form of amplification product, either X-derived or Y-derived, exploiting sequence differences between the two forms. Following digest, DNA was separated on agarose gel. The amount of DNA in each band was assayed by densitometry. Primers used for amplifying the cDNA of each gene pair, and restriction enzymes

used to digest either the X-derived (listed as X enzyme) or the Y-derived (listed as Y enzyme) amplification product were as follows:

Gene	Primers	X enzyme	Y enzyme
RPS4X/Y	GCTGGATTCATGGATGTCATC	BsrG I	Ear I
	GGCAAAGCTGTTGCCATTGG		
ZFX/Y	GTTTGCATTGCGACCACAAG	Pme I	Afl II
	GAACTGAGAGAATATGGCGAC		
SMCX/Y	GATCTTGGACCTCTACAG	Hha I	Nla III
	GAAGGCTGCACAGACTGTC		
UTX/Y	CAATGTGGGTACTGGAACCTG	Ase I	Bcl I
	GTGGTCTTGGAGGTGGACAT		
DFFRX/Y	CCCTAGTGCAGAAGTGAGG	Hind III	
	CTGCCACACCTAAATTGGC		
DFFRX/Y	GGAGAATCCTCAGTTCTCATC		Bgl II
	GCCAGGAGTCCTCAATCAG		
DBX/Y	GTGTTTGTGGAGACCAAAAAGG	Nco I	
	ATTGGGCTTTTTCCTGAGCG		
DBX/Y	GGCTTGTGCCCAAACAGG		Hind III
	CAAGGACGAACTCTAGATCG		
EIF1AX/Y	CCAAGAATAAAGGTAAAGGAGG	Taq I	
	CATCATCATCTCCAGGACC		
EIF1AX/Y	GCCATGCCCAAGAATAAAGG		Hha I
	CCCAACATTTTGATTACCTGAG		
TB4X/Y	GGCTGAGATCGAGAAATTCG	Hha I	
	TTTCCTTCCCTGCCAGCC		
TB4X/Y	GTCGAAACTGAAGAAGACAG		Taq I
	TTTCCTTCCCTGCCAGCC		

RESULTS

Map Order of X/Y Homologous Genes Not Conserved Between X and Y

Based on a deletion map of the Y with 43 intervals (Vollrath *et al.*, 1992), Y copies of X/Y homologous genes were previously mapped, each to a single interval (Lahn and Page, 1997). Three genes, *DFFRY*, *DBY* and *UTY* fell into the same interval 5C. The construction of a contig across this region resolved their ordering as follows: centromere-*DFFRY-DBY-UTY*-qter (Sun *et al.*, unpublished data).

Regional localization of X copies was previously carried out for five out of the nine genes, but each with a different method (Schneider-Gädicke *et al.*, 1989, Fisher *et al.*, 1990, Salido *et al.*, 1992, Agulnik *et al.*, 1994, Jones *et al.*, 1996). To put them all on a single map defined by the same set of reference markers, we carried out radiation hybrid mapping using the Genebridge 4 radiation hybrid panel (Walter *et al.*, 1994, Hudson *et al.*, 1995). Results are shown in Figure 1. Also shown in the figure are map locations of Y copies of X/Y homologous genes in the non-recombining region of the Y (Lahn and Page, 1997). With the exception of *RPS4X*, all X copies are located on the short arm. In contrast, Y copies are scattered about on both arms. Clearly, map orders of X/Y homologous genes are poorly conserved between X and Y, with one exception. *DFFRX*, *DBX* and *UTX* are clustered on the X. Their Y homologs *DFFRY*, *DBY* and *UTY* are also clustered, and appear to be in the same order.

Sequence Divergence Between X and Y Copies May Reflect Their Evolutionary Distance

For each pair of genes, X and Y copies have apparently evolved from a common ancestor. To assess the relative time points when X and Y copies began to diverge, we carried out sequence comparisons for each pair of genes at the DNA and protein levels. For DNA comparisons, the open reading frame (ORF), 5' untranslated region (5' UTR) and 3' untranslated region (3' UTR) were compared separately. Results are summed up in Table 1. Genes are listed in the order of descending degree of nucleotide conservation in the ORF. With this arrangement, X/Y homologous

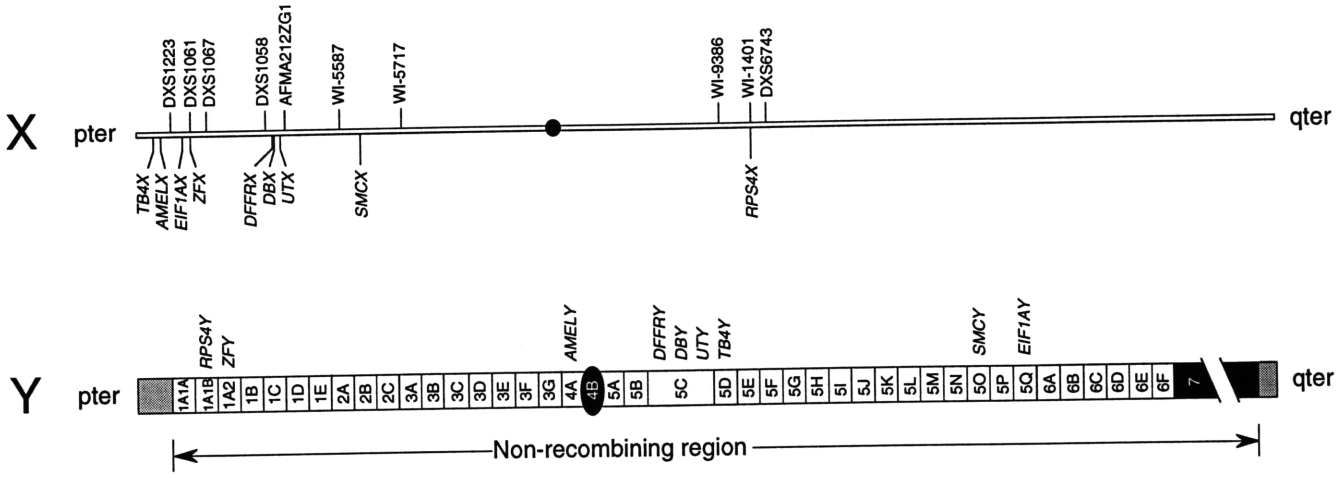


FIG. 1. Radiation hybrid map of X copies of X/Y homologous genes. Reference markers are listed above the chromosome; genes are listed below. Markers and genes are drawn to reflect their distances from each other in units of centiRay (Hudson *et al.*, 1995). The total length of the X chromosome is 521 centiRays. Also shown are positions of Y copies of X/Y homologous genes, previously mapped (Lahn and Page, 1997) based on naturally occurring Y chromosome deletions in people (Vollrath, *et al.*, 1992).

TABLE 1

Sequence Identity Between X and Y Copies of X/Y Homologous Genes

		ORF	Protein	5' UTR	3'UTR	Function
Group I	*ZFX/Y*	93% in 2403 nt	93%	86% in 219 nt	80% in 262 nt	Cell growth and proliferation
	AMELX/Y	93% in 576 nt	88%	81% in 68 nt	88% in 155 nt	Tooth enamal formation
Group II	*TB4X/Y*	93% in 132 nt	93%	N/H in 77 nt	82% in 279 nt	Actin sequestering
	EIF1AX/Y	91% in 432 nt	98%	N/H in 132 nt	67% in 526 nt	Translation initiation factor 1A
	DFFRX/Y	89% in 7641 nt	91%	N/H in 59 nt	75% in 393 nt	Unknown
	DBX/Y	88% in 1986 nt	91%	N/H in 72 nt	84% in 2477 nt	Unknown
	UTX/Y	88% in 4041 nt	85%	N/H in 84 nt	81% in 1051 nt	Unknown
Group III	*SMCX/Y*	83% in 4680 nt	85%	N/H in 276 nt	N/H in 581 nt	Unknown
	RPS4X/Y	82% in 789 nt	92%	N/H in 12 nt	N/H in 56 nt	Ribosomal protein S4

Note. Percentages rounded off to the nearest 1%; N/H: no homology detectable.

genes appear to fall into three groups. Group I genes are most conserved in the ORF between X and Y copies. They are also conserved in both 5' and 3' UTRs. They include *ZFX/Y* and *AMELX/Y*. Group II are not conserved in the 5' UTR, but are conserved in the 3' UTR. They are somewhat less conserved in the ORF as compared to group I genes, even though distinctions are subtle in a few cases. They include the next five entries in the table. Group III are not conserved at either 5' or 3' UTRs and are also least conserved in the ORF among the three groups. They include *SMCX/Y* and *RPS4X/Y*.

In most cases, the degree of conservation in the ORF at the DNA level is comparable to that at the protein level. There are exceptions: *AMELX/Y* are much more conserved at the DNA than at the protein level, whereas *EIF1AX/Y* and *RPS4X/Y* are much more conserved at the protein level.

X Copies of X/Y Homologous Genes Escape X-Inactivation

X-inactivation maintains comparable levels of expression of X-linked genes in males and females. For X/Y homologous genes however, if both X and Y copies are expressed in males, there is theoretically no need to inactivate one of the two X copies in females. Indeed, escape from X-inactivation was demonstrated in the four cases previously tested: *ZFX*, *RPS4X*, *SMCX* and *DFFRX* (Schneider-Gädicke *et al.*, 1989, Fisher *et al.*, 1990, Agulnik *et al.*, 1994, Jones *et al.*, 1996). We investigated the X-inactivation status for four additional genes: *DBX*, *UTX*, *TB4X* and *EIF1AX* by two assays: 1) the conventional method of RT-PCR on human-rodent hybrids retaining either the inactive or the active human X chromosomes (Fig. 2); and 2) the less conventional CpG methylation studies in which male and female genomic DNA digested with methylation sensitive restriction endonucleases were used as templates for PCR (Fig. 3). The RT-PCR assay clearly demonstrated that all four genes were transcribed from both hybrid cell lines - the one containing only the active human X and the one containing only the inactive human X. This is consistent with their escape from X-inactivation. The methylation assay takes advantage of the fact that the silencing of gene expression on the inactive X is typically associated with the methylation of the gene's 5' CpG rich region (Tribioli *et al.*, 1992). Conversely, if an X-linked gene escapes X-inactivation, its 5' region

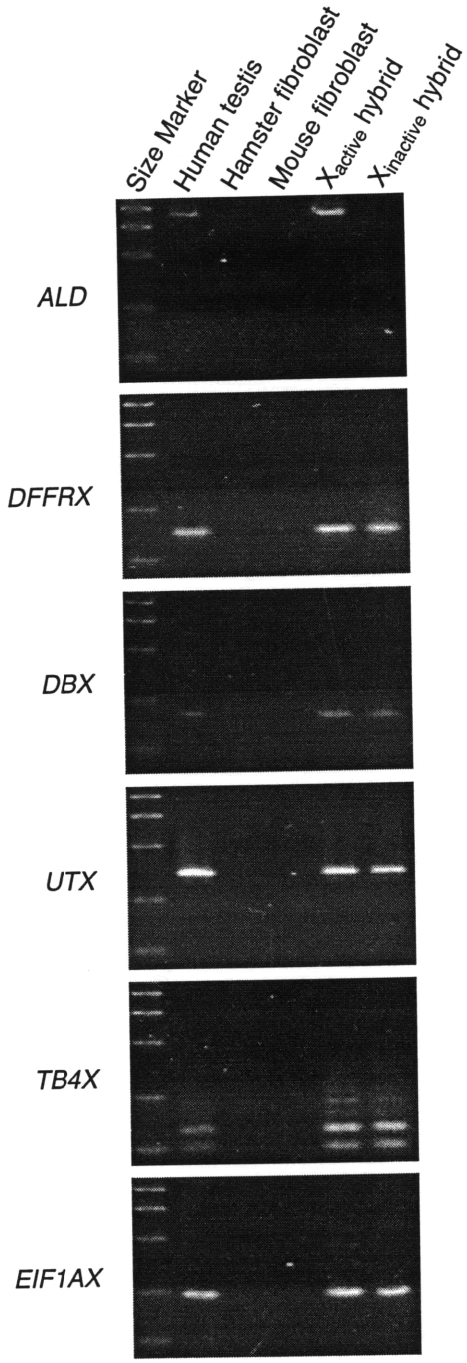


FIG. 2. RT-PCR demonstrated that *DBX*, *UTX*, *TB4X* and *EIF1AX* were expressed from both the active X and the inactive X. Two genes were included as control: 1) *ALD*, known to undergo X-inactivation (Migeon, et al., 1981) , and 2) *DFFRX*, known to escape X-inactivation (Jones, et al., 1996) . cDNA was made from tissues or cell lines. PCR was carried out on cDNAs to test for the presence/absence of transcripts. Human testis, where all these genes should be expressed, was a positive control for the RT-PCR assay. Hamster and mouse fibroblasts were negative controls where primers based on human sequences should not cross-amplify any product. The two experimental lanes were the human-hamster hybrid retaining the active human X (WHT2281) indicated as "Xactive hybrid" (Lahn, et al., 1994) , and human-mouse hybrid retaining the inactive human X (37-26R-D) indicated as "Xinactive hybrid" (Mohandas, et al., 1981) . Results for *ALD* and *DFFRX* were as expected. *ALD* could be amplified from the hybrid containing the active human X, but not the hybrid containing the inactive human X. *DFFRX* could be amplified from both hybrids. Like *DFFRX* which escaped X-inactivation, *DBX*, *UTX*, *TB4X* and *EIF1AX* all could be amplified from both hybrids, consistent with their escape from X-inactivation.

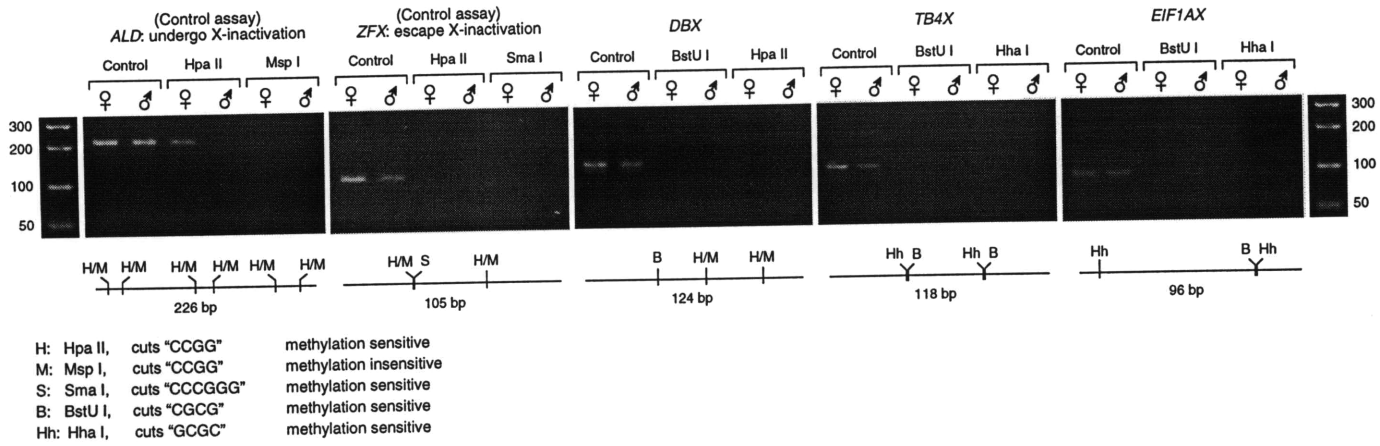


FIG. 3. Methylation status of *DBX*, *TB4X* and *EIF1AX* consistent with their escaping X-inactivation in females. PCR assays were designed within 5' CpG island regions of each gene. PCR was performed on female and male DNA samples after they underwent restriction digest. For each gene, samples in control lanes were digested with an enzyme that lacked sites within the PCR amplified region, while samples in experimental lanes were digested with enzymes that cut within the PCR amplified region (cut-sites are shown below gel pictures). Two control genes were included in the assay: 1) X-linked *ALD* which undergoes X-inactivation (Migeon, et al., 1981) ; 2) X-linked *ZFX* which escapes X-inactivation (Luoh, et al., 1995) . For *ALD*, the presence of PCR product in the female lane following digest with Hpa II (methylation sensitive) indicated that one copy is X-inactivated via methylation (by contrast, digest with the methylation-insensitive isoschizomer Msp I obliterates PCR product). For *ZFX*, *DBX*, *TB4X* and *EIF1AX*, digest with any of two methylation sensitive enzymes obliterated PCR products in females, indicating the unmethylated, active state of both copies of these genes.

would not be methylated on either the active or the inactive X (Luoh *et al.*, 1995). As expected, the methylation assay (*UTX* not tested) showed that the 5' region of each gene is demethylated on both the active and the inactive X in females, reflecting their escape from X-inactivation.

Eight out of the nine known X-linked genes with functional Y homologs were thus shown to escape X-inactivation. The one remaining gene for which X-inactivation status is unknown is *AMELX*. As previously mentioned, *AMELX/Y* are expressed exclusively from tooth buds, a place where X-inactivation status can not be readily assayed.

Comparable Levels of Expression From X and Y Copies

Y copies of X/Y homologous genes are often seen as functional supplements of X copies. Yet, for most of them, the question of how much activity is contributed by either copy has not been addressed. Here, we assayed the relative contribution to mRNA from X and Y copies for eight of the nine pairs (*AMELX/Y* not assayed).

We employed RT-PCR, using three male tissue sources: brain, prostate and testis. For each pair, we used primers shared in exact sequence by both X and Y copies. Following RT-PCR, DNA was cut with an enzyme that recognizes only one form of amplification product, either from the X copy or the Y copy. DNA fragments were separated on agarose gel, and the amount of DNA was measured by densitometry (Table 2). Two conclusions can be drawn from this data. First, with the exception of *TB4Y* which contributes 5% or less to total *TB4X/Y* mRNA, the Y isoforms typically account for between a third to half of the total transcript. This means that most of the Y copies are active at a comparable level as their X homologs, albeit they contribute somewhat less than half in several cases. Second, the relative contributions from X and Y copies stay roughly the same, at least across the three tissues tested. Noticeable exceptions are *ZFY* and *DBY* which have elevated expression in the testis. The significance of this elevation is unclear.

We did not assay the expression of *AMELX/Y*. It was previously shown that *AMELY* accounts for about 10% of total *AMELX/Y* mRNA in tooth buds (Salido *et al.*, 1992). This ratio is similar to that of *TB4X/Y*.

TABLE 2

Contribution From the Y Copy to the Total Expression of Each X/Y Homologous Gene Pair

	Brain	Prostate	Testis
RPS4Y	32±0%	37±0%	36.5±1.5%
ZFY	32.5±0.5%	48.5±1.5%	70±1%
DFFRY	31±0%	33.5±0.5%	20±0.5%
DBY	28±1%	27.5±0.5%	50±1%
UTY	54±0.5%	55.5±0.5%	54±1%
TB4Y	<<5%	<<5%	4.5±0.5%
SMCY	50±1%	48±0%	51.5±0.5%
EIF1AY	34±3%	45±2%	37.5±2.5%

Note. Percentages are given as the average of two experiments, ± their spread.

DISCUSSION

The presence of multiple pairs of X/Y homologous genes reflects the common ancestry of the two heteromorphic sex chromosomes. Y copies of these genes have managed to stay on while most other genes the Y originally shared with the X degenerated. Two factors may have contributed to their persistence: 1) Y copies are functionally required; and 2) X and Y copies began to diverge relatively recently.

If for some X/Y homologous genes, double dosage is essential for development, their Y copies would resist degeneration. For these genes, there are four predictions: 1) X and Y copies should have comparable functions; 2) their functions might be so essential that a reduction in dosage could conceivably be deleterious; 3) X copies should escape X-inactivation; and 4) X and Y copies should be expressed at comparable levels.

These predictions are mostly in agreement with observations. First, as inferred from the high degree of sequence similarity between X and Y copies, protein isoforms they encode likely have equivalent functions. For *RPS4X/Y*, the only case tested to date, functional equivalence was indeed demonstrated (Watanabe *et al.*, 1993). Second, the ubiquitous expression of all but one X/Y homologous genes (*AMELX/Y* is expressed in tooth buds) suggests essential housekeeping function. For the few of them where some functional knowledge is available, housekeeping indeed appears to be the theme - *RPS4X/Y* encode ribosome subunit S4 (Fisher *et al.*, 1990); *EIF1AX/Y* encode translation initiation factor 1A (Hershey, 1991, Dever *et al.*, 1994, Lahn and Page, 1997). *AMELX/Y* which function only in tooth development are a clear exception. Third, in all but one cases (*AMELX* has not been tested), escape from X-inactivation was demonstrated. And finally, with the exception of *TB4X/Y* and *AMELX/Y* (for which the Y isoform accounts for 10% or less of total mRNA from each gene pair) (Salido *et al.*, 1992), X and Y copies are expressed at comparable levels, typically within a factor of two.

Based on the degree of sequence divergence between X and Y copies, the nine known X/Y homologous genes can be broken down into three groups (Table 1). Such a breakdown is perhaps not coincidental, rather it may reflect the different time points at which Y copies of X/Y homologous

genes began to diverge from their X homologs. If this view is correct, Y copies of group I genes are most recent acquisitions on the Y whereas Y copies of group III genes are most ancient. Several independent lines of evidence are consistent with this view. *RPS4X/Y* and *SMCX/Y* were shown to be ancient X/Y homologous gene pairs in the mammalian lineage, existing on X and Y in both eutherian (placental) and metatherian (marsupial) mammals (Jegalian & Page, unpublished data; Graves, personal communication). In contrast, *ZFX/Y* in group I are apparently a recent member in the family of X/Y homologous genes in eutherian mammals. In marsupials, instead of being on sex chromosomes, the ortholog of *ZFX/Y* exists on an autosome (Graves and Watson, 1991). Furthermore, three pairs of genes, all in group II, *DFFRX/Y*, *DBX/Y* and *UTX/Y* are closely clustered on both X and Y, and in the same order, consistent with their Y copies being acquired through a single transposition event.

The two factors, functional importance and time of acquisition, can both affect the persistence of Y copies of X/Y homologous genes. *AMELY* in group I, for example, is present on the Y likely due to its recent acquisition rather than functional importance. *AMELX/Y* are involved in tooth development, apparently not a housekeeping function. The high degree of nucleotide homology suggests that *AMELX* and *AMELY* began diverging relatively recently. Yet, they are much less conserved at the protein level, which indicates relaxed evolutionary constraint. In fact, *AMELY* has already partially degenerated, accounting for only 10% of total *AMELX/Y* mRNA in tooth buds (Salido *et al.*, 1992). On the other hand, *RPS4Y* in group III is preserved despite being one of the most ancient genes on the Y. It is much more conserved with *RPS4X* at the protein than at the DNA level. The preservation of *RPS4Y* is in keeping with its essential housekeeping function. Among the nine pairs of X/Y homologous genes, *EIF1AX/Y* in group II are the most conserved at the protein level. For *EIF1AY*, both functional importance and a relatively short history may have contributed to its preservation; *EIF1AY* encodes an essential translation factor; and its history on the Y appears to be more recent than that of *SMCY* and *RPS4Y* in group III.

J. Graves and colleagues argued that an autosome to sex chromosome translocation occurred some time after the divergence of placental mammals and marsupials (Graves and Watson,

1991, Graves, 1995). This new addition on sex chromosomes, which was originally pseudoautosomal (namely it recombined between X and Y), stopped recombining at some point. Subsequently, Y-linked genes in this region were subject to degeneration. Our data is largely consistent with this hypothesis. X copies of group I and II genes which were relatively recent additions to the family of X/Y homologous gene, all mapped to distal Xp, the region postulated by Graves to have been acquired by sex chromosomes through a translocation. In contrast, the two group III genes SMCX and RPS4X, which have been on sex chromosomes even before the divergence of placental mammals and marsupials (Jegalian & Page, unpublished data; Graves, personal communication), mapped more towards the long arm (Fig. 1). Map order of Y copies is dissimilar with that of X copies, presumably due to shuffling by repeated inversions. Such re-arrangements are tolerated on the Y because they do not result in meiotic nondisjunctions as they might on autosomes. Three pairs of group II genes, *DFFRX/Y*, *DBX/Y* and *UTX/Y* are clustered on both X and Y, and in the same order. Perhaps the tight linkage of these genes prevented them from being shuffled apart on the Y. If group I and II genes were indeed acquired by sex chromosomes from an autosome in a single translocation event, the distinction between the two groups as presented in Table 2 may not be meaningful.

Even though there are many theories to account for the degeneration of the Y, it is not clear what steps are involved in this process. Our observations are consistent with a stepwise process that begins with reduction in expression, followed by disruption of the ORF, and finally, the X homolog becomes X-inactivated. *AMELY* and *TB4Y* are expressed at 10% or less the level of their X homologs. Yet, both genes have intact ORFs. This suggests that reductions in expression may proceed disruptions of the ORF. While X-inactivation status of *AMELX* is not known, *TB4X* clearly escapes X-inactivation. This suggests that X copies of X/Y homologous genes may escape X-inactivation long after Y copies have declined in activity. In fact, escape from X-inactivation of the X copy can persist even after the ORF of the Y copy is disrupted, as in the case of X-linked gene *STS* with a silenced nonfunctional Y homolog containing frameshifts and premature stops (Yen *et al.*, 1988).

Ferguson-Smith postulated that Turner syndrome (TS) is due to the haploinsufficiency of the so called Turner genes: those expressed from both X and Y in males and escape X-inactivation in females (Ferguson-Smith, 1965, Zinn *et al.*, 1993). With the exception of *AMELX/Y* which only affect tooth growth, X/Y homologous are all potential Turner candidates. No definitive link has yet been established between any one pair of X/Y homologous genes and Turner features. For those genes whose Y copies are expressed at very low levels, their role in Turner syndrome can be called into question. Of course, one should not absolutely rule out genes as Turner candidates based on levels of expression, as a minute reduction in dosage for a crucial gene can still result in developmental anomalies.

In conclusion, our systematic comparison of X/Y homologous genes has provided additional information on how these genes have evolved and how they might be implicated in development.

REFERENCES

Agulnik, A. I., Mitchell, M. J., Mattei, M. G., Borsani, G., Avner, P. A., Lerner, J. L. and Bishop, C. E. (1994). A novel X gene with a widely transcribed Y-linked homologue escapes X-inactivation in mouse and human. *Hum Mol Genet* **3:** 879-884.

Bull, J. J. (1983). Evolution of Sex Determining Mechanisms: Benjamin Cummings, Menlo Park, CA.

Charlesworth, B. (1978). Model for evolution of Y chromosomes and dosage compensation. *Proc Natl Acad Sci U S A* **75:** 5618-5622.

Charlesworth, B. (1991). The evolution of sex chromosomes. *Science* **251:** 1030-1033.

Cox, D. R., Burmeister, M., Price, E. R., Kim, S. and Myers, R. M. (1990). Radiation hybrid mapping: a somatic cell genetic method for constructing high-resolution maps of mammalian chromosomes. *Science* **250:** 245-250.

Dever, T. E., Wei, C. L., Benkowski, L. A., Browning, K., Merrick, W. C. and Hershey, J. W. (1994). Determination of the amino acid sequence of rabbit, human, and wheat germ protein synthesis factor eIF-4C by cloning and chemical sequencing. *J Biol Chem* **269:** 3212-3218.

Ferguson-Smith, M. A. (1965). Karyotype-phenotype correlations in gonadal dysgenesis and their bearing on the pathogenesis of malformations. *J. Med. Genet.* **2:** 142-155.

Fisher, E. M., Beer-Romero, P., Brown, L. G., Ridley, A., McNeil, J. A., Lawrence, J. B., Willard, H. F., Bieber, F. R. and Page, D. C. (1990). Homologous ribosomal protein genes on the human X and Y chromosomes: escape from X inactivation and possible implications for Turner syndrome. *Cell* **63:** 1205-1218.

Graves, J. A. (1995). The origin and function of the mammalian Y chromosome and Y-borne genes--an evolving understanding. *Bioessays* **17:** 311-320.

Graves, J. A. (1996). Mammals that break the rules: genetics of marsupials and monotremes. *Annu. Rev. Genet.* **30:** 233-260.

Graves, J. A. and Watson, J. M. (1991). Mammalian sex chromosomes: evolution of organization and function. *Chromosoma* **101:** 63-68.

Hershey, J. W. (1991). Translational control in mammalian cells. *Annu Rev Biochem* **60:** 717-755.

Hudson, T. J., Stein, L. D., Gerety, S. S., Ma, J., Castle, A. B., Silva, J., Slonim, D. K., Baptista, R., Kruglyak, L., Xu, S. H. and et al. (1995). An STS-based map of the human genome. *Science* **270:** 1945-1954.

Jones, M. H., Furlong, R. A., Burkin, H., Chalmers, J., Brown, G. M., Khwaja, O. and Affara, N. A. (1996). The Drosophila developmental gene fat facets has a human homologue in Xp11.4 which escapes X-inactivation and has related sequences on Yq11.2. *Hum Mol Genet* **5:** 1695-1701.

Lahn, B. T. and Page, D. C. (1997). Functional coherence of the human Y chromosome. *Submitted.*

Luoh, S. W., Jegalian, K., Lee, A., Chen, E. Y., Ridley, A. and Page, D. C. (1995). CpG islands in human ZFX and ZFY and mouse Zfx genes: sequence similarities and methylation differences. *Genomics* **29:** 353-363.

Muller, H. J. (1914). A gene for the fourth chromosome of Drosophila. *J Exp Zool* **17:** 325-336.

Page, D. C., Mosher, R., Simpson, E. M., Fisher, E. M., Mardon, G., Pollack, J., McGillivray, B., de la Chapelle, A. and Brown, L. G. (1987). The sex-determining region of the human Y chromosome encodes a finger protein. *Cell* **51:** 1091-1104.

Rice, W. R. (1987). Genetic hitchhiking and the evolution of reduced genetic activity of the Y sex chromosome. *Genetics* **116:** 161-167.

Rice, W. R. (1994). Degeneration of a nonrecombining chromosome. *Science* **263:** 230-232.

Salido, E. C., Yen, P. H., Koprivnikar, K., Yu, L. C. and Shapiro, L. J. (1992). The human enamel protein gene amelogenin is expressed from both the X and the Y chromosomes. *Am J Hum Genet* **50:** 303-316.

Schneider-Gädicke, A., Beer-Romero, P., Brown, L. G., Nussbaum, R. and Page, D. C. (1989). ZFX has a gene structure similar to ZFY, the putative human sex determinant, and escapes X inactivation. *Cell* **57:** 1247-1258.

Schwartz, A., Chan, D., Brown, L. G., Alagappan, R., Pettay, D., Disteche, C., McGillivray, B., de la Chapelle, A. and Page, D. C. (1997). Reconstructing hominid Y evolution: X-homologous block,

created by X-Y transposition, was disrupted by Yp inversion through LINE-LINE recombination. *Submitted.*

Tribioli, C., Tamanini, F., Patrosso, C., Milanesi, L., Villa, A., Pergolizzi, R., Maestrini, E., Rivella, S., Bione, S., Mancini, M. and et al. (1992). Methylation and sequence analysis around EagI sites: identification of 28 new CpG islands in XQ24-XQ28. *Nucleic Acids Res* **20:** 727-733.

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.

Walter, M. A., Spillett, D. J., Thomas, P., Weissenbach, J. and Goodfellow, P. N. (1994). A method for constructing radiation hybrid maps of whole genomes. *Nat Genet* **7:** 22-28.

Watanabe, M., Zinn, A. R., Page, D. C. and Nishimoto, T. (1993). Functional equivalence of human X- and Y-encoded isoforms of ribosomal protein S4 consistent with a role in Turner syndrome. *Nat Genet* **4:** 268-271.

Yen, P. H., Marsh, B., Allen, E., Tsai, S. P., Ellison, J., Connolly, L., Neiswanger, K. and Shapiro, L. J. (1988). The human X-linked steroid sulfatase gene and a Y-encoded pseudogene: evidence for an inversion of the Y chromosome during primate evolution. *Cell* **55:** 1123-1135.

Zinn, A. R., Page, D. C. and Fisher, E. M. (1993). Turner syndrome: the case of the missing sex chromosome. *Trends Genet* **9:** 90-93.

CHAPTER 4: Xq-Yq INTERCHANGE RESULTS IN RETARDATION

(This chapter is adapted from a published paper, *Nature Genetics* (1994) **8**, 243-250.)

Original title of the paper:

Xq-Yq interchange resulting in supernormal X-linked gene expression in severely retarded males with 46,XYq- karyotype

Bruce T. Lahn[1], Nancy Ma[2], W. Roy Breg[3], Robert Stratton[4], Urvashi Surti[5] & David C. Page[1]

[1]Howard Hughes Research Laboratories at Whitehead Institute, and Department of Biology, Massachusetts Institute of Technology, 9 Cambridge Center, Cambridge, Massachusetts 02142, USA

[2]Department of Human and Molecular Genetics, Collaborative Research, Inc., Waltham, Massachusetts 02154, USA

[3]Department of Human Genetics, Yale University School of Medicine, New Haven, Connecticut 06510, USA

[4]South Texas Genetics Center, 7922 Ewing Halsell Drive, San Antonio, Texas 78229, USA

[5]Departments of Pathology and Genetics, Magee Women's Hospital, University of Pittsburgh, Pittsburgh, Pennsylvania 15213, USA

Xq–Yq interchange resulting in supernormal X-linked gene expression in severely retarded males with 46,XYq- karyotype

Bruce T. Lahn[1], Nancy Ma[2], W. Roy Breg[3], Robert Stratton[4], Urvashi Surti[5] & David C. Page[1]

The critical importance of dosage compensation is underscored by a novel human syndrome ("XY_{Xq} syndrome") in which we have detected partial X disomy, demonstrated supernormal gene expression resulting from the absence of X inactivation, and correlated this overexpression with its phenotypic consequences. Studies of three unrelated boys with 46,XYq- karyotypes and anomalous phenotypes (severe mental retardation, generalized hypotonia and microcephaly) show the presence of a small portion of distal Xq on the long arm of the Y derivative. Cells from these boys exhibit twice-normal activity of glucose-6-phosphate dehydrogenase, a representative Xq28 gene product. In all three cases, the presence of Xq DNA on a truncated Y chromosome resulted from an aberrant Xq–Yq interchange occurring in the father's germline.

[1]*Howard Hughes Research Laboratories at Whitehead Institute, and Department of Biology, Massachusetts Institute of Technology, 9 Cambridge Center, Cambridge, Massachusetts 02142, USA*
[2]*Department of Human and Molecular Genetics, Collaborative Research, Inc., Waltham, Massachusetts 02154, USA*
[3]*Departments of Genetics and Pediatrics, Yale University School of Medicine, New Haven, Connecticut 06510, USA*
[4]*South Texas Genetics Center, 7922 Ewing Halsell Drive, San Antonio, Texas 78229, USA*
[5]*Departments of Pathology and Genetics, Magee Women's Hospital, University of Pittsburgh, Pittsburgh, Pennsylvania 15213, USA*

Correspondence should be addressed to D C.P.

It is widely accepted that X inactivation occurs in humans to ensure that X-linked gene expression in females is equal to rather than twice the level found in males. It follows, as many have argued, that failure to dosage-compensate X-linked genes could be the cause of certain anomalous human phenotypes[1–7]. If so, one should be able to demonstrate supernormal X-linked gene expression in association with these phenotypes. We have had an opportunity to test this prediction while exploring the origins of apparent terminal deletions of the Y chromosome long arm (that is, the 46,XYq- karyotype), which are among the most common chromosomal disorders in human populations.

At least one in 1,000 males lacks about half of the Y chromosome, including the quinacrine-bright, heterochromatic region[8]. The 46,XYq- karyotype can be associated with short stature and azoospermia[9,10], but more severe phenotypes have been reported, including profound mental retardation, hypotonia and dysmorphic features[11,12]. While exploring a possible genetic basis for the phenotypic variability observed in XYq- males, we have discovered that a subset of such males delineate a syndrome, which we refer to as XY_{Xq} syndrome. We show here, at the levels of gene expression and organismal phenotype, the consequences of failing to dosage-compensate X-linked genes that are present in two copies per cell. We have also discovered Xq–Yq counterparts to the aberrant Xp–Yp exchanges that have been so thoroughly studied in human XX males and XY females[13–19]. Finally, we consider the factors that may predispose Xq and Yq to recombine aberrantly in the paternal germline.

XYq– phenotypes and deletion breakpoints

Our study focused on ten males with 46,XYq- karyotypes. Samples from these individuals were received from various medical centers for Y chromosome analysis. All patients had been ascertained postnatally and karyotyped previously because of phenotypic abnormalities.

The phenotypes of these ten males varied dramatically (Table 1). Three individuals ("Group I") were severely mentally retarded, microcephalic boys with little or no ability to speak or comprehend words. They had very poor muscle tone, were unable to stand without assistance, and had suffered nonfebrile seizures. The three Group II individuals were school-age boys with mild to moderate learning disabilities and delays in speech and motor development. Both Group I and Group II boys had mild facial dysmorphism, and four of the six boys had undescended testes. The four Group III individuals had no history of delayed cognitive or motor development. The two youngest Group III males had mild facial dysmorphism. All ten individuals were short in stature, and the two adults (both Group III) were infertile (azoospermic).

All ten individuals lacked the distal long arm of the Y chromosome, as initially revealed by cytogenetic analyses conducted at the referring medical centers. Eight of the ten patients had been tested previously by PCR for the presence of specific Y-linked loci, resulting in localization of their Yq breakpoints[20]. We extended these studies by testing all ten individuals for 80 loci distributed across the euchromatic portion of the Y. The results allowed us to rule out interstitial deletions within the portions of the Y retained. The positions of the Yq breakpoints among these ten individuals are quite heterogeneous, with at least

eight distinct breakpoints detectable (Fig. 1).

Although the results of Y chromosome DNA analysis were consistent with simple terminal deletions of Yq in all ten cases, there were reasons to suspect that the chromosomal anomalies in at least some of these patients might be more complex. There was no correlation between the size of the deletion and the severity of the phenotype (Fig. 1). One of the smallest deletions (WHT2277) was found in a severely affected (Group I) patient whereas two of the largest deletions (LGL658 and WHT1157) were found in the mildly affected (Group III) individuals. This suggested that much of the phenotype could not be attributed to loss or disruption of specific Yq genes. Indeed, deletion of Y chromosomal genes would appear unlikely to cause severe mental retardation, hypotonia and microcephaly, as these traits are uncommon in 45,X individuals, who for purposes of this argument can be seen as having lost the entire Y chromosome. Also, it is probable that Yq- chromosomes would be stable only if the Yq telomere were retained or replaced by another telomere. Further, distal Yq and Xq form synaptonemal complexes and recombine during normal male meiosis[21,22], providing a possible opportunity for aberrant Xq–Yq recombination, perhaps producing what appears to be a Yq- chromosome. In this respect, the behaviour of distal Yq and Xq might resemble that of distal Yp and Xp, where aberrant, grossly misaligned X–Y recombination occasionally occurs, giving rise to XX males and XY females[13–19]. Of particular relevance here are the rare females whose karyotypes were originally described[23–25] as 46,XYp- (with terminal deletions of the short arm), but whose Y derivatives were subsequently found to be products of aberrant Xp–Yp interchange (ref. 18; D.C.P. *et al.*, unpublished results). By analogy, we speculated that

Table 1 Phenotypes of ten unrelated 46,XYq– males

	Patients	Age at last examination	Cognitive development	Motor development	Seizures	Height[a]	Other characteristics
Group I Severely affected	WHT1278	8 years	Severely retarded; unable to speak	Hypotonia w/ hyper-extensible joints; unable to stand without assistance	Partial complex seizures (confirmed by EEG)	119 cm (5–10%)	Microcephaly; prominent ears; hypoplastic midface; high, narrow palate; small feet; prominent keloids from surgery; undescended testes
	WHT1373	8 years	Severely retarded; unable to speak	Hypotonia; unable to stand without assistance	Repeated seizures	80 cm (<<2%)	Microcephaly; widely spaced eyes; beaked nose; high, narrow palate, small jaw; broad thumbs; undescended left testis; small right testis; repeated unexplained fevers
	WHT2277	5.5 years	Severely retarded, unable to speak	Hypotonia w/ hyper-extensible elbows; unable to stand without assistance	One prolonged grand mal seizure	100 cm (<5%)	Microcephaly with flat occiput; low posterior hairline; small upturned nose; narrow mouth, short neck; small feet; bilateral camptodactyly of 4th fingers; livedo reticularis
Group II Moderately affected	WHT1829	7.5 years	Mild but persistent speech delay; marked learning disability	Mild gross and fine motor clumsiness	None	119 cm (10–25%)	Prominent ears; unilateral amblyopia; small atrial septal defect
	WHT1832	10.5 years	Speech delay corrected by therapy	Mild fine motor clumsiness	None	125 cm (<5%)	Atypical facies with prominent low-set ears, partial ptosis; mild pulmonic stenosis; undescended testes
	WHT1876	10 years	Moderate speech and learning disabilities	Mild hypotonia and motor clumsiness	None	95 cm at 5 yrs (<5%) (before growth hormone treatment)	Nasal bridge slightly broadened; widely spaced nipples; undescended right testis
Group II Mildly affected	WHT1983	5 7 years	Normal	Normal	None	99 cm (<5%)	Microcephaly[b]; prominent ears, long eyelashes; bushy eyebrows w/mid-fusion; thin lips, small feet; hypospadias
	LGL658[c]	17 years	Normal	Normal	None	160 cm (<5%)	High-arched palate, small mandible; small teeth; mild hypothyroidism
	WHT1157	28 years	Normal	Normal	None	159 cm (<5%)	Azoospermia
	WHT2168	32 years	Normal	Normal	None	165 cm (5%)	Azoospermia

[a]Heights are expressed in cm and as age-adjusted percentiles.
[b]WHT1983's head circumference was proportionately greater than his height; given normal cognitive and motor development, his "microcephaly" may simply reflect a generalized growth delay.
[c]LGL658's clinical features are as reported by P. Salo, *et al.* (ref. 55)

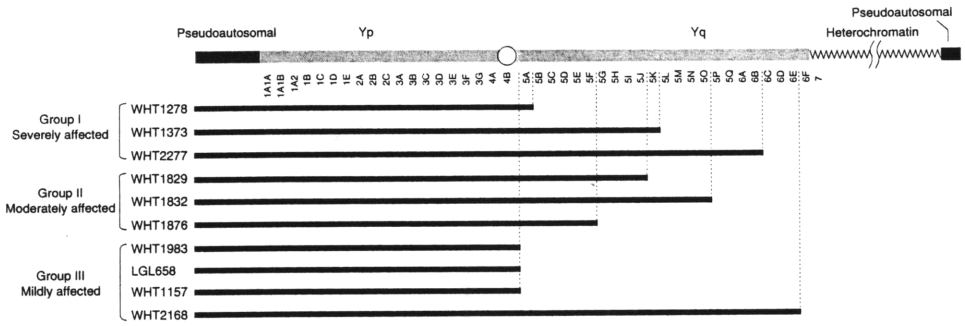


Fig. 1 Poor correlation of Yq breakpoint with phenotypic severity in ten unrelated XYq- males. Schematic representation of the Y chromosome, with Yp and Yq pseudoautosomal regions and heterochromatic region labelled. Immediately below are listed 43 deletion intervals (1A1A through 7), as defined[20]. The black bars below indicate the portions of the Y chromosome found to be present in the XYq- males by testing for 80 Y-specific STSs. Results for all patients except WHT1983 and WHT2277 were reported previously[20].

some XYq- males might be the result of grossly misaligned exchanges between Yq and Xq.

Aberrant Xq–Yq interchange

To test this possibility we typed all ten patients and, if available, their parents, for two genetic markers mapping near the Xq telomere. *DXS1108*, a strictly X-linked marker, lies about 60 kilobases (kb) proximal to the long-arm pseudoautosomal region[22,26]. *DXYS154* , a long-arm pseudoautosomal marker, is located about 140 kb from the Xq/Yq telomere[22,26].

In three patients the results suggested aberrant Xq–Yq interchange in the paternal germline. WHT1278, WHT1373 and WHT2277 exhibited two alleles at both loci, and in each case, comparison with parental genotypes revealed that one allele was maternally derived and the other paternally derived (Fig. 2). Thus, these XYq- males inherited not only one maternal X chromosome but also the distal long arm of the paternal X chromosome.

A fourth patient, WHT1829, inherited a paternal and a maternal allele for the pseudoautosomal marker *DXYS154* but only a maternal allele for the more proximal, strictly X-linked marker *DXS1108* (Fig. 2*b*). These findings suggest either an interstitial deletion on the Y chromosome or an aberrant Xq–Yq interchange with the X breakpoint falling between *DXYS154* and *DXS1108*.

By contrast, the six remaining patients showed no evidence of Xq–Yq exchange. WHT1832 and WHT1876 exhibited single maternal alleles for both markers. Single alleles for both markers were also observed in WHT1983, LGL658, WHT1157 and WHT2168; these are likely to be of maternal origin, but in the absence of parental samples, we cannot exclude the possibility that identical alleles were transmitted from both parents. (In these four cases, further evidence against the presence of a second copy of Xq28 — the most distal band — was obtained by typing for highly polymorphic markers at the Factor VIII and *GABRA3* loci. In no case were two alleles observed; data not shown.)

For the three individuals in whom analysis of Xq markers strongly suggested aberrant Xq–Yq interchange (WHT1278, WHT1373 and WHT2277), we searched for more evidence by generating human-hamster somatic

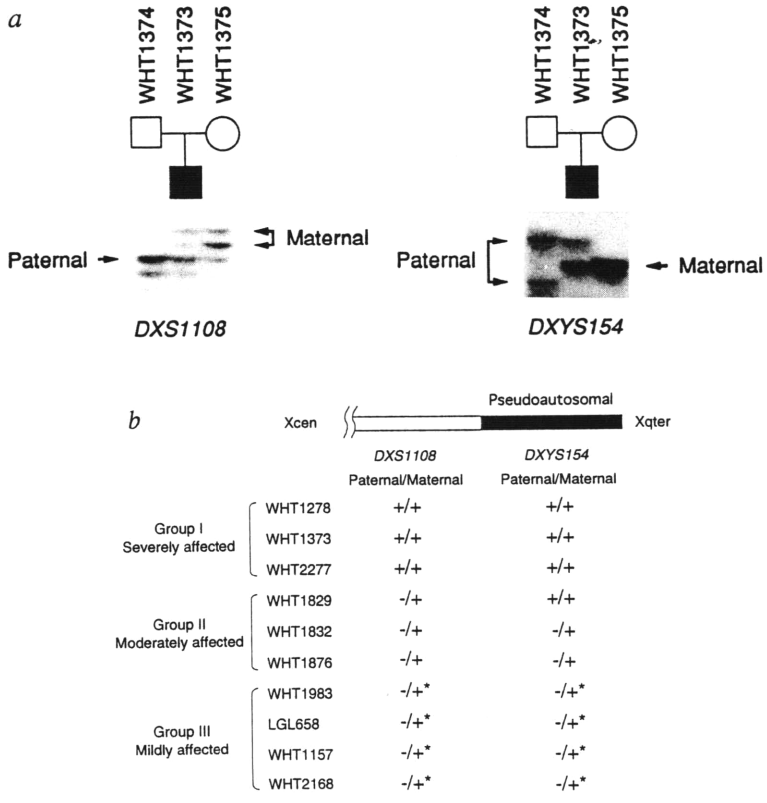


Fig. 2 *a*, Inheritance of CA-dinucleotide repeat polymorphisms at *DXS1108* (left) and *DXYS154* (right) in the family of XYq- male WHT1373. WHT1373 inherited a paternal (WHT1374) and a maternal (WHT1375) allele at each of the two loci. At *DXS1108*, an Xq28-specific locus, WHT1373 inherited the single paternal allele and maternal upper allele (indicated by upper arrow; lower arrow denotes second allele, not transmitted to WHT1373). At *DXYS154*, an Xq/Yq pseudoautosomal locus, WHT1373 inherited the paternal upper allele and the allele for which mother is homozygous. *b*, Transmission to XYq- males of paternal alleles at *DXS1108* and *DXYS154* . +/+, XYq- male inherited a paternal allele and a maternal allele; –/+, XYq- male inherited no paternal allele (only a maternal allele); –/+*, XYq- male exhibited only one allele (presumably of maternal origin), but parents were not available for testing.

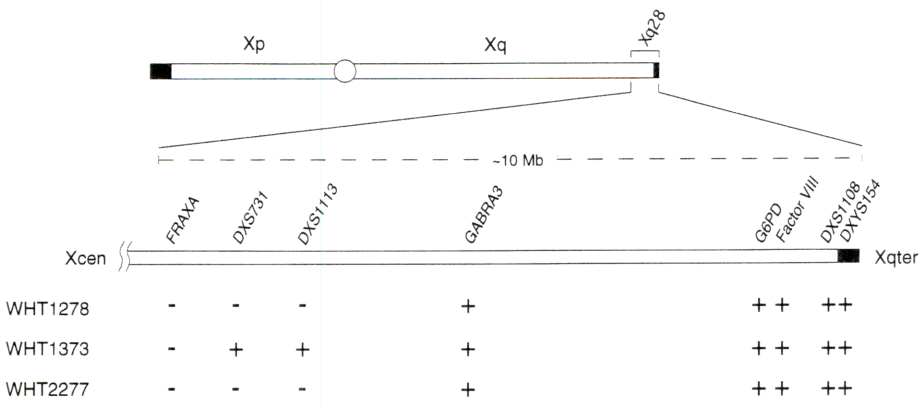
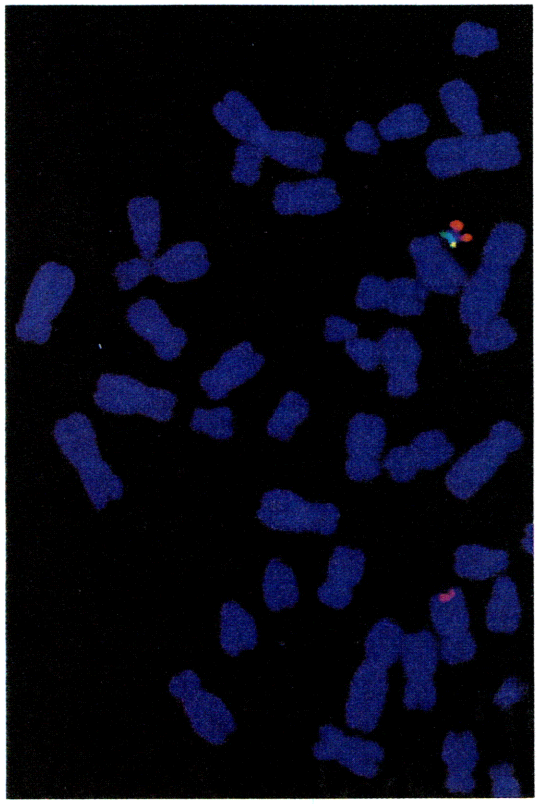


Fig. 3 Terminal portions of Xq for which the three Group I XYq- males are disomic. For each patient, somatic cell hybrids retaining the derivative Y but lacking the intact X were tested for presence (+) or absence (–) of seven X-specific (*FRAXA* through *DXS1108*) loci and one Xq–Yq pseudoautosomal (*DXYS154*) locus. Based on physical mapping of Xq (ref. 53), we estimate the size of the disomic regions to be 5–7 Mb in WHT1278 and WHT2277 and 8–10 Mb in WHT1373.

cell hybrids retaining the human Yq- chromosomes but lacking human X chromosomes. The Xq–Yq interchange model predicts that such Yq- hybrids would carry paternally derived alleles for *DXYS154* and *DXS1108*. This was the case for Yq- hybrids from all three patients (data not shown). Testing of these hybrids with other Xq DNA loci (Fig. 3) revealed the presence of variably sized terminal portions of the chromosome, with breakpoints in Xq28 (between *DXS1113* and *GABRA3* in WHT1278 and WHT2277 and between *FRAXA* and *DXS731* in WHT1373), again consistent with Xq–Yq interchange in the paternal germline.



To assess directly whether distal Xq DNA had been transferred to the truncated long arm of the Yq- chromosomes in WHT1278, WHT1373 and WHT2277, we performed fluorescence *in situ* hybridization (FISH) to metaphase chromosomes from all three patients. The FISH probe was derived from human Factor VIII, an Xq28 gene we had detected in Yq- hybrids prepared from all three individuals (Fig. 3). In normal males and females, the Factor VIII probe hybridized *in situ* only to Xq28 (data not shown). In each of the three patients, the Factor VIII probe hybridized both to the distal long arm of the intact X chromosome and to one end of the Yq- chromosome (Fig. 4). Thus, all three of these XYq- males have two copies of the Factor VIII gene, the second copy being located on the derivative Y. In each case, a second probe, specific to Yp and labelled with a different fluorescent dye, hybridized to the opposite end of the Yq- chromosome (Fig. 4). Thus, it was the truncated long arm of the derivative Y to which Xq28 DNA had been transferred in all three individuals.

In conclusion, genetic marker and FISH studies provided strong evidence that, in three unrelated XYq- males, aberrant Xq–Yq interchange in the father's germline had produced "Y_{Xq}" chromosomes that lost a terminal portion of Yq in exchange for a terminal portion of Xq (Fig. 5). There is a striking correlation with phenotype: among ten XYq- males studied, the most severe phenotypes were found in the three males with unequivocal evidence of Xq–Yq interchange. These three patients form a coherent set that is chromosomally and phenotypically distinct from the other XYq- males studied. We will refer to these three individuals as "XY_{Xq} males".

Fig. 4 *In situ* hybridization of *Factor VIII* probe to derivative Y chromosome of XYq- male WHT1278. Factor VIII plasmid p482.6 (red) and Yp-specific plasmid pDP1335 (green) localized to opposite ends of derivative Y chromosome. Factor VIII probe also localized to distal long arm of intact X. Similar results obtained with XYq- males WHT1373 and WHT2277 (not shown).

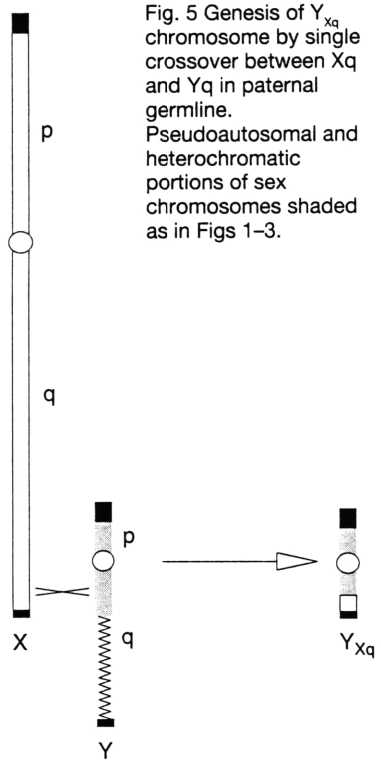


Fig. 5 Genesis of Y_{Xq} chromosome by single crossover between Xq and Yq in paternal germline. Pseudoautosomal and heterochromatic portions of sex chromosomes shaded as in Figs 1–3.

Supernormal expression of *G6PD*

The correlation between chromosomal constitution and phenotype strongly suggested a cause-and-effect relationship. As argued earlier, the severe mental retardation, hypotonia and microcephaly of the XY_{Xq} males are unlikely to be the result of the absence of specific Yq genes. Neither are the XY_{Xq} phenotypes likely to be the result of truncation, fusion or other rearrangement of specific Yq genes, since these phenotypically similar males display widely different Yq breakpoints (Fig. 1).

If the phenotypes are not readily accounted for by Yq deletion *per se*, then perhaps they can be explained by the presence of Xq28 DNA on the Y_{Xq} chromosomes. Specifically, the severe phenotypes might be the result of twice-normal expression of Xq28 genes that are normally X-inactivated when present in two copies (in normal females) but that fail to be dosage-compensated here. The X inactivation centre has been mapped to Xq13, and it must be present in two (or more) copies per cell for X inactivation to occur[27,28]. As the duplicated portion of Xq (that is, the portion of the X present on the Y_{Xq} chromosome) does not include Xq13, the cells of the XY_{Xq} males carry only one X inactivation centre and X inactivation should not occur. Thus, Xq28 genes present in two copies should not be dosage-compensated and supernormal expression of these genes might result in the severe phenotypes observed.

To examine these questions, we studied expression of the glucose-6-phosphate-dehydrogenase gene (*G6PD*), a representative Xq28 gene normally dosage-compensated via X inactivation[29], in cells from the XY_{Xq} males. *G6PD* is present on the Y_{Xq} chromosomes in all three of these individuals (Fig. 3). We quantitated G6PD enzymatic activity in lysates of cultured lymphoblastoid cells from the patients and their parents by spectrophotometry (Fig. 6). The G6PD activities in the five parents tested fell within a narrow range, while each of the three XY_{Xq} males exhibited G6PD activity approximately twice that of his parent(s).

The spectrophotometric results are consistent with the presence in the XY_{Xq} males of two actively expressed *G6PD* genes per cell. To test directly whether G6PD is expressed from the derivative Y, we assayed human G6PD activity in human-hamster somatic cell hybrids retaining Y_{Xq} but lacking intact human X chromosomes. Since human and rodent G6PD proteins have different gel mobilities, we could detect the human isoform by staining for enzymatic activity after non-denaturing electrophoresis of total protein. On non-denaturing gels, G6PD protein exists as a dimer. As shown in Fig. 7, a control hybrid containing an inactive human X chromosome expressed only the rodent isoform, while a control hybrid retaining an active human X chromosome exhibited three G6PD bands, corresponding to human homodimer, human-rodent heterodimer and rodent homodimer. (The heterodimer reflects the synthesis of human and rodent isoforms within the same cell.) Hybrids retaining Y_{Xq} chromosomes from WHT1278, WHT1373 or WHT2277 exhibited the three G6PD bands observed in the active-X hybrid. Since these Y_{Xq} hybrids contain no human X chromosome, the Y_{Xq} chromosomes must be the source of the human G6PD activity. We conclude that, in each of the three XY_{Xq} males, G6PD is expressed from both the intact X and Y_{Xq} chromosomes, resulting in twice-normal levels of expression.

Discussion

Origin of XY_{Xq} males. The most distal portions of Xp and Yp are extraordinarily recombinogenic during male meiosis. As a result, the nucleotide sequences of these regions are indistinguishable, and their inheritance is "pseudoautosomal" rather than strictly sex-linked[30,31]. Most human XX males and some human XY females are

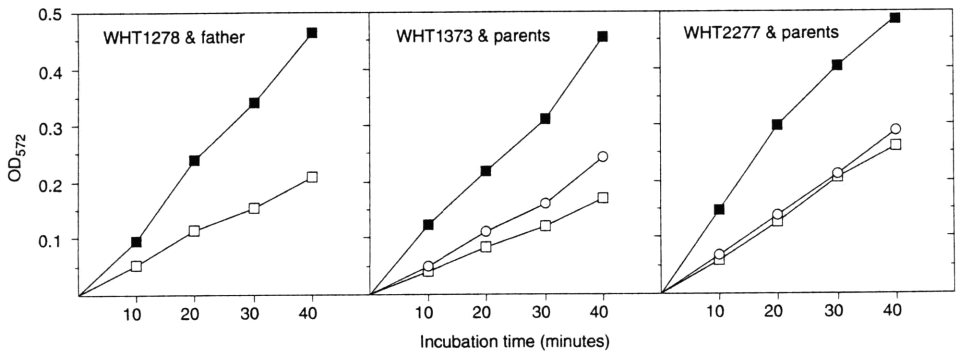


Fig. 6 Doubled G6PD activity in XY_{Xq} males compared with their parents. Conversion of glucose 6-phosphate to 6-phosphogluconolactone by lymphoblastoid extracts was measured as an increase in OD_{572} (see Methodology). Substrate was not limiting, as demonstrated by linearity of product present after 10, 20, 30 and 40 minutes of incubation. —■—, Patient; —□—, father; —○—, mother.

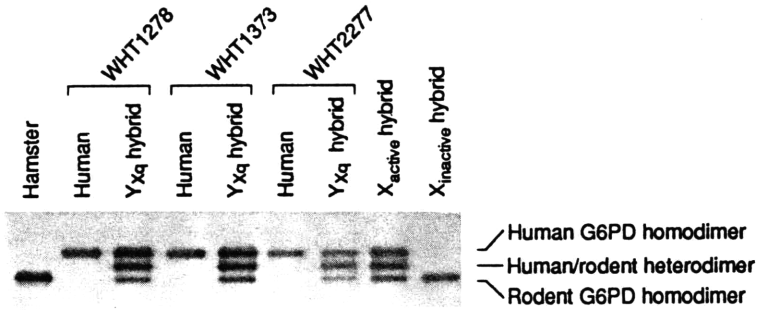


Fig. 7 Human G6PD enzymatic activity in somatic cell hybrids retaining Y_{Xq} chromosomes. Gel stained for G6PD activity after non-denaturing electrophoresis of total protein from the following cultured cells (left to right): RJK (hamster fibroblast)[50], WHT1278 (XY_{Xq} male lymphoblastoid), WHT2656 (human–hamster hybrid retaining Y_{Xq} chromosome but not intact X chromosome from WHT1278), WHT1373 (XY_{Xq} male lymphoblastoid), WHT2666 (human–hamster hybrid retaining Y_{Xq} from WHT1373), WHT2277 (XY_{Xq} male lymphoblastoid), WHT2667 (human–hamster hybrid retaining Y_{Xq} from WHT2277), WHT2660 (human–hamster hybrid retaining active X chromosome from WHT2281), 37-26R-D (human–mouse hybrid retaining inactive human X)[54]. Similar results were obtained with other Y_{Xq} hybrids prepared from WHT1278, WHT1373 and WHT2277.

the result of aberrant exchanges of terminal portions of Xp and Yp in the paternal germline[13–19]. These Xp–Yp exchanges may represent aberrant byproducts of the highly recombinogenic pairing of distal Xp and Yp during male meiosis.

During male meiosis, synaptonemal complexes are formed not only between the distal portions of Xp and Yp but also between the most distal portions of Xq and Yq (ref. 21). The nucleotide sequences of the most distal 320 kb of Xq and Yq are indistinguishable[26], and their inheritance, like that of distal Xp and Yp, is "pseudoautosomal"[22]. We speculate that the Xq–Yq exchanges giving rise to XY_{Xq} males (Fig. 5) represent aberrant byproducts of the recombinogenic pairing of distal Xq and Yq during male meiosis. (We cannot exclude the possibility that the aberrant Xq–Yq exchanges giving rise to XY_{Xq} males occurred during mitosis rather than meiosis in the paternal germline. This is also true for the aberrant Xp–Yp exchanges giving rise to XX males and XY females.) In the case of human XX males, Xp–Yp exchange can be the result of a single crossing-over between grossly misaligned X and Y chromosomes, with homologous recombination occurring at sites of local sequence similarity between Xp and Yp (refs 17,19). If some XY_{Xq} males result from a single crossing-over between Xq and Yq, then that crossing-over must also occur between grossly misaligned X and Y chromosomes (as depicted in Fig. 5). It will be of interest to learn whether aberrant Xq–Yq exchanges, like the aberrant Xp–Yp exchanges, occur at sites of local X–Y sequence similarity.

Thus, the relationship between the long-arm pseudoautosomal region and XY_{Xq} males may be analogous to that between the short arm pseudoautosomal region and XX males (and some XY females). The Xp–Yp exchange products found in such XY females are roughly reciprocal to those found in XX males[13-19,32]. By analogy, we might predict the existence in human populations of Xq–Yq exchange products reciprocal to those found in XY_{Xq} males. Indeed, three females with 46,X,der(X)t(X;Y)(q;q) karyotypes have been reported[33–35]. Though these cases, to our knowledge, have not been studied with genetic markers, the available information is consistent with their X derivatives having resulted from aberrant Xq–Yq interchange in the paternal germline.

It will be of interest to determine the nature of the derivative Y chromosomes in the seven XYq- males in whom we did not demonstrate Xq–Yq exchange, especially since these seven males display considerable phenotypic diversity (Table 1). Perhaps some of these males will prove to be the result of Xp;Yq translocation, as recently demonstrated by Bardoni and colleagues[7], or of Yq; autosome translocations

Critical importance of dosage compensation. It is generally agreed, in a teleological sense, that the purpose of X inactivation is to ensure that expression of X-linked genes in females is equal to rather than twice that in males. It is implicitly understood that failure to inactivate X-linked genes present in two copies per cell would be harmful to the organism. Indeed, some aberrant human phenotypes have been interpreted as likely resulting from supernormal expression of X-linked genes that would normally be dosage-compensated via X inactivation. These phenotypes, all associated with partial X disomy (and monosomy for the X inactivation centre) include: (i) gonadal sex reversal in 46,XY individuals with partial Xp duplications[1–3], (ii) mental retardation and dysmorphic features atypical of Turner syndrome in females with mosaic 45,X/46,X,r(X) karyotypes[4,5], (iii) severe phenotypes associated with X;autosome translocations[6] and (iv) psychomotor retardation and dysmorphic features in three males with 46,X,der(Y)t(X;Y)(p21.3 or p22.1;q11) karyotypes[7]. In each of these cases, supernormal X-linked gene expression is among the most likely explanations for the abnormal phenotype. However, to our knowledge, supernormal expression has not been directly demonstrated in any of these cases. (Three reports have described single patients with tandem Xq duplications and twice-normal activity of an X-linked enzyme[36–38], but in those cases it remains unclear whether supernormal expression of X-linked genes caused the patients' phenotypic abnormalities).

We have described three unrelated boys in whom DNA marker and FISH studies revealed similar 46,X,der(Y)t(X;Y)(q28; q11) karyotypes. All three boys exhibited severe mental retardation, generalized hypotonia and microcephaly, and these phenotypes are almost certainly due to overexpression of certain genes in the portions of Xq28 for which these boys are disomic. We have demonstrated twice-normal activity of a representative Xq28 gene product, G6PD, in cells from each of the three boys, and it is reasonable to suppose that other Xq28 products would show similar behaviour. The portions of the X for which these boys are disomic measure 5 to 10 megabases (Mb) and are likely to contain more than a hundred genes. The phenotypically critical, dosage-sensitive genes probably constitute a subset of these and may or may not include *G6PD*.

Turner syndrome (typically 45,X) and Klinefelter syndrome (47,XXY) are better-known disorders associated with abnormal sex chromosome dosage. As in the XY_{Xq} syndrome described here, the Turner and Klinefelter phenotypes are due, at least in part, to quantitatively abnormal expression of sex chromosomal genes. However, the dosage-sensitive genes critical to the XY_{Xq} syndrome should differ systematically and fundamentally from the dosage-sensitive genes critical to the Turner or Klinefelter syndromes. The Turner phenotype probably results, at least in part, from haploinsufficiency of genes that are common to the X and Y chromosomes and that escape X inactivation[39,40]. The Klinefelter phenotype probably results, at least in part, from overexpression of genes that escape X inactivation. (These "Klinefelter genes" may or may not have Y counterparts). By contrast, we predict that the dosage-sensitive genes critical to the XY_{Xq} syndrome are strictly X-linked, have no functional equivalents on the Y chromosome, and normally are subject to X inactivation.

The perplexing range of XYq- phenotypes. Males with 46,XYq- karyotypes exhibit a wide range of phenotypes, extending from normal development and normal fertility[41] to short stature and azoospermia[9,10] to severe mental retardation and dysmorphic features[11,12]. The unpredictability of the phenotype creates real dilemmas for families and those counseling them.

Our studies represent a step toward resolution of this perplexing situation. Among the ten patients with 46,XYq- karyotypes studied, we found evidence of Xq–Yq exchange in the three most severely affected individuals. We note two previously reported XYq- boys[11,12] whose phenotypes are quite similar to those of the XY_{Xq} males described here. We would not be surprised if these two boys were found to carry products of aberrant Xq–Yq interchange. One might predict that XYq- fetuses with evidence of Xq–Yq exchange (for example, from FISH studies or assay of G6PD activity) would develop severe phenotypes like those seen in our three XY_{Xq} patients. However, there is undoubtedly bias toward referral of severely affected individuals to our laboratory, and this may result in our set of XYq- patients constituting a skewed sample.

Among XY_{Xq} males, one might expect phenotypic severity to increase with the extent of Xq disomy. Among our three XY_{Xq} boys, WHT1373 displayed both the largest region of Xq disomy (Fig. 3) and the most profound growth retardation (height 80 cm, weight 35 lbs at 8 years). Conversely, we note that one XYq- male in whom an Xq;Yq translocation has been detected (cited as unpublished result in ref. 7) displays a much milder phenotype than seen in our patients (see description of case 7 in ref. 7); perhaps he is disomic for a smaller region of Xq than our patients. Further phenotypic and genotypic comparisons of these and similar cases will be of value.

Methodology

PCR analysis of DNA markers. Patient DNAs were tested for the presence of Y-chromosomal sequence-tagged sites; oligonucleotide primers, agarose gel electrophoresis, and ethidium-bromide detection were as described[20]. Thermocycling conditions: 3 min at 94 °C followed by 30 cycles of 1 min at 94 °C, 1 min at 58 °C, and 1 min at 72 °C. Similar methods were used to test human-rodent somatic cell hybrid DNAs for the presence of the following pseudoautosomal or X-linked loci:

Locus	Primers	Ref.
DXYS154	5'–GGCCTGAATTCATTTATTATTCTAATAG–3'	22
	5'–GAACAGGCAAAGATGCCCACTCTC–3'	
DXS1108	5'–ACTAGGCGACTAATACAGTGGTGC–3'	22
	5'–GTGAATTCATCATATGTGATTTCC–3'	
Factor VIII	5'–TGCATCACTGTACATATGTATCTT–3'	42
	5'–CCAAATTACATATGAATAAGCC–3'	
G6PD	5'–CCTCTATGTGGAGAATGAGAG–3'	43
	5'–CACTGCTGGTGGAAGATGTCG–3'	
GABRA3	5'–TCCTGAGGGCAGGGTCTCTGATT–3'	44
	5'–GGGTTCAGGAGACTGCACAGCAA–3'	
DXS1113	5'–ACCTGTGGAGGATAGTAGTCTGACT–3'	45
	5'–GGGAGCTTTAGAGATTTTGGTAAAC–3'	
DXS731	5'–CTCACCATTGGGTCTTCATACA–3'	46
	5'–TATGATAGGCATGAATTGTGTCTG–3'	
FRAXAC2	5'–GACTGCTCCGGAAGTTGAATCCTCA–3'	47
	5'–CTAGGTGACAGAGTGAGATCCTGTC–3'	

Patients and parents were typed for CA-dinucleotide repeat polymorphisms using radioactively labelled primers (listed above) and polyacrylamide gel electrophoresis, as described[48].

Construction of human–hamster somatic cell hybrids. Hybrids were generated as described[49] by fusing human lymphoblastoid lines with RJK (thymidine-kinase-deficient) hamster fibroblasts[50] in the presence of polyethylene glycol-4000; subsequent culture in HAT medium supplemented with glycine. After isolation and propagation of adherent clones, DNA was extracted and tested for the presence of human X-specific and Y-specific sequence tagged sites.

Fluorescence in situ hybridization (FISH). Chromosome spreads were prepared from lymphoblastoid cell lines cultured in the presence of 0.1 μg ml^{-1} colcemid for 60 min. *In situ* suppression hybridization and two-colour fluorescent detection were performed as described[51]. In brief, plasmid p482.6 (ATCC #57202), whose 9.6-kb insert derives from intron 22 of the human Factor VIII gene[52], was labelled by nick translation with biotin-14-dATP and visualized using avidin-Texas Red. Plasmid pDP1335, whose 19-kb insert derives from interval 1A1A on distal Yp (ref. 20) was labelled with digoxygenin-11-dUTP and visualized using fluorescein-conjugated antibody. Slides were counterstained with DAPI. FISH images were captured using a CCD camera and electronically processed.

Quantitation of G6PD activity. Pelleted lymphoblastoid cells were lysed in 10 volumes of water by vortexing. After microcentifugation, total protein concentration in the supernatant was measured by Lowry reaction and adjusted to 0.15 mg ml^{-1}. We then added 10 μl of adjusted supernatant to 100 μl of G6PD staining solution (0.1 M Tris pH8, 1 mg ml^{-1} Na_2-glucose-6-phosphate, 80 μg ml^{-1} $NADP^+$, 100 μg ml^{-1} Methylthiazolium tetrazolium [MTT], 40 μg ml^{-1} phenazine methosulfate[PMS]) at room temperature. Reactions were stopped after 0, 10, 20, 30, or 40 min by adding 10 μl of 10% SDS and 380 μl water. G6PD catalyzes conversion of glucose-6-phosphate to 6-phosphogluconolactone and the coupled reduction of $NADP^+$ to NADPH. In the presence of PMS, oxidation of NADPH drives conversion of MTT to formazan, whose increasing concentration was followed by monitoring optical density at 572 nm.

Electrophoretic detection of G6PD activity. Cultured cells were pelleted and lysed in an equal volume of 50 mg ml^{-1} DTT, 5 mg ml^{-1} $NADP^+$. After microcentrifugation, 10 μl of supernatant from each sample was subjected to electrophoresis on a non-denaturing 4% polyacrylamide gel. The gel was treated with G6PD staining solution for 10 min at room temperature and photographed.

Acknowledgements

We are grateful to the patients, their families, and the following colleagues for providing samples and clinical information: L. Amesse, R. Bernstein, A. de la Chapelle, A. Escoll, M. Eswara, M. Frank, L. Holmes, D. Manchester, B. Pober, C. Rudlin, M. Sandberg, P. Salo, S. Shekhter-Levin, D. Solomon and V. Wilson. We thank J. Seligman, L. Brown, R. Alagappan, A. Schwartz, P. Beer-Romero, D. Pettay, P. Bain, K. Jegalian and R. Reijo for their contributions to this work. Supported by the Howard Hughes Medical Institute and the National Institutes of Health.

Received 15 August; accepted 21 September 1994.

1 Bernstein, R. *et al.* Female phenotype and multiple abnormalities in sibs with a Y chromosome and partial X chromosome duplication H-Y antigen and Xg blood group findings. *J med. Genet* **17**, 291–300 (1980)
2. Arn, P *et al.* SRVX, a sex reversing locus in Xp21 2–p22.11. *Hum. Genet.* **93**, 389–393 (1994).
3. Bardoni, B. *et al.* A dosage sensitive locus at chromosome Xp21 is involved in male to female sex reversal. *Nature Genet* **7**, 497–501 (1994)
4 Van Dyke, D.L. *et al.* Ullrich-Turner syndrome with a small ring X chromosome and presence of mental retardation *Am J. med. Genet.* **43**, 996–1005 (1992).
5. Wolff, D J *et al.* Small marker X chromosomes lack the X inactivation center: implications for karyotype/phenotype correlations. *Am J. hum. Genet.* **55**, 87–95 (1994).
6 Schmidt, M. & Du Sart, D. Functional disomies of the X chromosome influence the cell selection and hence the X inactivation pattern in females with balanced X-autosome translocations: a review of 122 cases *Am. J med. Genet.* **42**, 161–169 (1992).
7 Bardoni, B *et al.* Functional disomy of Xp22–pter in three males carrying a portion of Xp translocated to Yq *Hum. Genet.* **91**, 333–338 (1993)
8 Hamerton, J.L., Canning, N., Ray, M. & Smith, S. A cytogenetic survey of 14,069 newborn infants. I. Incidence of chromosome abnormalities *Clin Genet.* **8**, 223–243 (1975).
9. Tiepolo, L. & Zuffardi, O. Localization of factors controlling spermatogenesis in the nonfluorescent portion of the human Y chromosome long arm *Hum Genet* **34**, 119–124 (1976)
10 Yunis, E., Garcia-Conti, F L., de Caballero, O.M. & Giraldo, A. Yq deletion, aspermia, and short stature *Hum Genet.* **39**, 117–122 (1977)
11. Nakagome, Y , Sasaki, M., Matsui, I , Kawazura, M. & Fukuyama, Y A mentally retarded boy with a minute Y chromosome. *J Pediatr.* 1163–1167 (1965).
12. Podruch, P.E & Yen, F S. Yq- in a child with livedo reticularis, snub nose, microcephaly, and profound mental retardation *J. med. Genet.* **19**, 377–380 (1982).
13. de la Chapelle, A., Tippett, P A., Wetterstrand, G. & Page, D. Genetic evidence of X–Y interchange in a human XX male. *Nature* **307**, 170–171 (1984).
14 Andersson, M , Page, D.C. & de la Chapelle, A. Chromosome Y-specific DNA is transferred to the short arm of X chromosome in human XX males. *Science* **233**, 786–788 (1986)
15 Page, D C., Brown, L.G. & de la Chapelle, A Exchange of terminal portions of X- and Y-chromosomal short arms in human XX males. *Nature* **328**, 437–440 (1987).
16. Petit, C. *et al.* An abnormal terminal X–Y interchange accounts for most but not all cases of human XX maleness *Cell* **49**, 595–602 (1987)
17 Rouyer, F , Simmler, M C., Page, D.C & Weissenbach, J A sex chromosome rearrangement in a human XX male caused by Alu-Alu recombination *Cell* **51**, 417–425 (1987)
18 Levilliers, J., Quack, B., Weissenbach, J. & Petit, C Exchange of terminal portions of X- and Y-chromosomal short arms in human XY females *Proc. natn. Acad. Sci. U.S.A.* **86**, 2296–2300 (1989).
19 Weil, D *et al.* Highly homologous loci on the X and Y chromosomes are hot-spots for ectopic recombinations leading to XX maleness *Nature Genet.* **7**, 414–419 (1994)
20. Vollrath, D. *et al.* The human Y chromosome a 43-interval map based on naturally occurring deletions. *Science* **258**, 52–59 (1992)
21. Speed, R.M. & Chandley, A.C. Prophase of meiosis in human spermatocytes analysed by EM microspreading in infertile men and their controls and comparisons with human oocytes *Hum Genet.* **84**, 547–554 (1990).
22 Freije, D., Helms, C., Watson, M.S. & Donis-Keller, H Identification of a second pseudoautosomal region near the Xq and Yq telomeres *Science* **258**, 1784–1787 (1992)
23. Rosenfeld, R G. *et al.* Sexual and somatic determinants of the human Y chromosome: studies in a 46,XYp- phenotypic female. *Am. J. hum Genet* **31**, 458–468 (1979).
24. Magenis, R.E. *et al* Turner syndrome resulting from partial deletion of Y chromosome short arm: localization of male determinants. *J. Pediatr.* **105**, 916–919 (1984).
25. Disteche, C M. *et al.* Small deletions of the short arm of the Y chromosome in 46,XY females. *Proc. natn. Acad. Sci. U.S.A.* **83**, 7841–7844 (1986).
26 Kvaloy, K , Galvagni, F & Brown, W R A. The sequence organization of the long arm pseudoautosomal region of the human sex chromosomes *Hum molec. Genet.* **3**, 771–778 (1994)
27 Brown, C.J. *et al.* Localization of the X inactivation centre on the human X chromosome in Xq13. *Nature* **349**, 82–84 (1991).
28. Kay, G F , Barton, S.C., Surani, M.A. & Rastan, S Imprinting and X chromosome counting mechanisms determine Xist expression in early mouse development *Cell* **77**, 639–650 (1994)
29 Migeon, B R Glucose-6-phosphate dehydrogenase as a probe for the study of X-chromosome inactivation in human females *Isozymes Curr Top. Biol med. Res.* **9**, 189–200 (1983).
30. Cooke, H J , Brown, W R & Rappold, G.A Hypervariable telomeric sequences from the human sex chromosomes are pseudoautosomal *Nature* **317**, 687–692 (1985)
31 Simmler, M C *et al.* Pseudoautosomal DNA sequences in the pairing region of the human sex chromosomes. *Nature* **317**, 692–697 (1985)
32 Page, D C. Sex reversal: deletion mapping the male-determining function of the human Y chromosome *Cold Spring Harbor Symposia on Quantitative Biology* **51**, 229–235 (1986)
33. Koo, G C. *et al.* Mapping the locus of the H-Y gene on the human Y chromosome *Science* **198**, 940–942 (1977)
34. Cameron, I T , Buckton, K E & Baird, D T X–Y translocation A case report *Hum. Genet* **67**, 457–459 (1984)
35 Kelly, T E *et al* X, Y translocation in a female with streak gonads, H-Y-phenotype, and some features of Turner's syndrome. *Cytogenet cell Genet* **38**, 122–126 (1984)
36. Schwartz, S *et al* Inherited X-chromosome inverted tandem duplication in a male traced to a grandparental mitotic error *Am J. hum Genet* **38**, 741–750 (1986).
37 Thode, A. *et al.* A new syndrome with mental retardation, short stature and an Xq duplication *Am. J med. Genet.* **30**, 239–250 (1988)
38. Muscatelli, F , Lena, D , Mettei, M G. & Fontes, M A male with two contiguous inactivation centers on a single X chromosome: study of X inactivation and XIST expression *Hum molec Genet.* **1**, 115–119 (1992)
39. Ferguson-Smith, M A. Karyotype-phenotype correlations in gonadal dysgenesis and their bearing on the pathogenesis of malformations *J med Genet.* **2**, 142–155 (1965)
40. Zinn, A R , Page, D C & Fisher, E.M. Turner syndrome the case of the missing sex chromosome *Trends Genet.* **9**, 90–93 (1993)
41 Meisner, L.F. & Inhorn, S.L. Normal male development with Y chromosome long arm deletion (Yq-). *J med. Genet.* **9**, 373–377 (1972)
42. Lalloz, M R , McVey, J H , Pattinson, J K. & Tuddenham, E G. Haemophilia A diagnosis by analysis of a hypervariable dinucleotide repeat within the factor VIII gene *Lancet* **338**, 207–211 (1991)
43. Freije, D & Schlessinger, D A 1.6-Mb contig of yeast artificial chromosomes around the human factor VIII gene reveals three regions homologous to probes for the DXS115 locus and two for the DXYS64 locus. *Am J hum Genet.* **51**, 66–80 (1992)
44 Hicks, A.A , Johnson, K.J , Barnard, E.A. & Darlison, M.G. Dinucleotide repeat polymorphism in the human X-linked GABRA receptor alpha 3-subunit gene *Nucl. Acids Res.* **19**, 4016 (1991)
45 Weber, C *et al* Dinucleotide repeat polymorphism close to IDS gene in Xq27.3–q28 (DXS1113). *Hum molec. Genet.* **2**, 612 (1993)
46. Hudson, T J. *et al.* Isolation and chromosomal assignment of 100 highly informative human simple sequence repeat polymorphisms *Genomics* **13**, 622–629 (1992)
47 Richards, R I *et al.* Fragile X syndrome. genetic localisation by linkage mapping of two microsatellite repeats FRAXAC1 and FRAXAC2 which immediately flank the fragile site. *J. med. Genet.* **28**, 818–823 (1991)
48 Dietrich, W F *et al* A genetic map of the mouse with 4006 simple sequence length polymorphisms. *Nature Genet* **7**, 220–245 (1994)
49. Nussbaum, R L , Airhart, S D & Ledbetter, D.H Expression of the fragile (X) chromosome in an interspecific somatic cell hybrid *Hum. Genet* **64**, 148–150 (1983).
50 Mohandas, T *et al* Regional localization of human gene loci on chromosome 9: studies of somatic cell hybrids containing human translocations. *Am J hum. Genet.* **31**, 586–600 (1979)
51. Trask, B J., Massa, H., Kenwrick, S & Gitschier, J Mapping of human chromosome Xq28 by two-color fluorescence in situ hybridization of DNA sequences to interphase cell nuclei. *Am. J. hum. Genet* **48**, 1–15 (1991)
52. Wion, K.L., Tuddenham, E.G. & Lawn, R M A new polymorphism in the factor VIII gene for prenatal diagnosis of hemophilia A *Nucl. Acids Res* **14**, 4535–4542 (1986)
53 Willard, H.L., Mandel, J.-L., Monaco, A P , Nelson, D L & Schlessinger, D Report of the committee on genetic constitution of the X chromosome In *Human Gene Mapping 1993* 656–719
54 Mohandas, T , Sparkes, R.S. & Shapiro, L J. Reactivation of an inactive human X chromosome: evidence for X inactivation by DNA methylation *Science* **211**, 393–396 (1981)
55 Salo, P , Kraaralainen, H , page, D C & de la Chapelle, A Deletion mapping of stature determinants on the long arm of the Y chromosome *Hum Genet* (in the press)

CHAPTER 5 - DISCUSSION AND CONCLUSIONS

This thesis reports on a number of projects aimed at understanding the function, organization and evolution of the human Y chromosome, and to a lesser extent, the human X. Emerging from a large body of data generated through these projects are biological insights that cover a broad range of topics.

i. Two classes of genes in the human NRY: their evolution and organization

Entire chromosomes may have properties that set them apart from the rest of the genome - unusual gene density, atypical recombination rate, etc.. But a chromosome's gene content almost always appears to be a random sampling of the genome, with no discernible theme. The non-recombining region of the human Y (NRY) is a clear exception. Most NRY genes fall into two distinct classes, each with unique properties.

Class I genes have very similar but non-identical homologs on the X. These genes, along with their X-homologs, are typically expressed in all tissues. They participate in biological processes common between males and females. Patterns of expression and functional knowledge of these genes suggest that a fair fraction of them are housekeeping. Class II genes do not have X-homologs (they may however have related sequences on autosomes). They are expressed predominantly in the adult testis (with the exception of the male-determining gene *SRY*), suggesting their roles in germline biology. The bipartite nature of NRY's gene content is a close reflection of the its evolutionary history. The Y has once been a chromosome with extensive homology with the X. This homology has diverged over time, and replaced by non-recombining, male-specific sequences (Graves, 1995, Charlesworth, 1996). Y-encoded homologs of X-linked genes have mostly degenerated in the process. Once a gene on the Y is completely silenced, its X homolog which originally escaped, would now be subject to X-inactivation. Those that persist on the Y do so either because they are required to provide double dosage for essential functions in

males, or that they are recent acquisitions by the Y and did not have sufficient time to degenerate. While genes with X homologs have a tendency to degenerate, a second class of genes - those that function exclusively in male biology, especially in spermatogenesis - are often acquired by the NRY to take advantage of its male-restricted transmission. Unlike degeneration of Class I genes on the Y which has been theorized extensively, the tendency of class II genes to accumulate in the NRY is a phenomenon that deserves more rigorous theoretical treatment.

The two classes of genes are different in their copy numbers. Class I X-homologous genes are all single copy, whereas class II male-specific genes are present in multiple copies on the Y (*SRY* is an exception). An extreme case is *TSPY* which exists in dozens of copies (Manz *et al.*, 1993). Some multi-copy genes (*e.g.*, *TSPY* and *DAZ*, *BPY1*, *BPY2*) seem to exist in local clusters, presumably due to repeated amplifications (Manz *et al.*, 1993, Reijo *et al.*, 1995). Others (*e.g.*, *RBM* and *CDY*) appear to be present in multiple, interspersed sites (Ma *et al.*, 1993). For these genes, following their local amplification, repeated inversions may have dispersed them. Each member of Class II genes are therefore a family of very closely related genes. It is unknown whether all copies of a given class II gene are active. One can speculate why class II genes exist in multiple copies: 1) to increase the level of expression; 2) to buffer the effect of deleterious mutations of any given copy; and 3) to have a greater repertoire of genes from which functional variants can evolve. Regardless of what force drives class II genes toward greater copy numbers, the Y chromosome, unrestricted by meiotic recombination, is a fertile ground for chromosomal rearrangement that can lead to gene amplification.

The origin of class I genes is well understood. The origin of class II genes is still a subject of speculation. There are several possibilities. Perhaps they were once homologs of X-encoded genes, but have diverged greatly in sequence and acquired male-specific functions. A possible example is mouse *Zfy* (Mardon and Page, 1989). Relative to other genes with X homologs, *Zfy* is distinct in several ways: 1) it shows greater than usual divergence from its X homolog *Zfx*; 2) it is duplicated, with two near-identical copies; and 3) it is expressed only in the testis, whereas *Zfx*

expression is ubiquitously. With these features, *Zfy* is more like a male-specific gene rather than a X-homologous gene. Perhaps *Zfy* has evolved away from its function in both somatic and germ cells toward an exclusive and tailored role in the male germline. Another scenario is that the Y, by being a male chromosome, attracts genes with male-specific functions from the rest of the genome. One possible example is *DAZ*, a human gene on the Y with a putative role in spermatogenesis (Reijo *et al.*, 1995). *DAZ* apparently transposed from an autosomal location to the Y, and was subsequently amplified (Saxena *et al.*, 1996).

The human NRY is genetically impoverished, encoding far fewer genes (each gene family on the Y counts as one gene) than the genome's average. Given the repetitive nature of class II genes on the Y, the actual transcriptional units could be significantly higher than the number of unrelated genes. Nevertheless, NRY's gene density is still severely below average. Similar situations have been observed on the Y in a number of other species. Again, this may be an inevitable consequence of male-restricted transmission and lack of recombination of the Y.

ii. Male fertility and the Y: from the *azoospermia factor* to an entire chromosome

The human Y had long been thought to function solely in sex determination, until studies associated large deletions of the Y to male infertility (Tiepolo and Zuffardi, 1976). In the ensuing years, researchers have looked all across the NRY for the so called *azoospermia factor* (*AZF*), a gene or genes whose disruption leads to spermatogenic failure. Yet their efforts have suffered predicaments typically unseen in conventional positional cloning. Since NRY is the only region in the genome that does not offer genetic linkage, the hunt for *AZF* relies heavily on detecting *de novo* deletions in infertile men. These deletions, though frequent, tend to be very large. This problem, combined with the repetitive nature of many genes on the Y, makes detecting the disruption of a single gene - touchstone for implicating a gene in a phenotype - somewhat untenable.

The growing understanding of the unique biology of the Y suggests a different approach. Even though the Y could potentially function in any aspect of male-specific biology, evidence suggests that besides its role in sex determination, germ cell development

is perhaps the only other male-specific biology the Y is involved in. Class II male-specific genes with testis-limited expression are very likely involved in spermatogenesis, and for these genes, the correlation of mutations of a single gene to infertility is perhaps not the absolute requirement for proving their biological roles.

The systematic gene identification reported in Chapter 2 is a major step toward creating a complete gene catalog for the NRY. Even though the project is not aimed at cloning genes responsible for specific human diseases, it nevertheless uncovered a number of *AZF* candidates. If the majority of NRY genes have been uncovered, as our data indeed suggest, a fair fraction of presently known NRY genes may prove to be *AZF*.

iii. Technical feasibility of whole-chromosome gene identification

Never before has a systematic gene search on a whole-chromosome of a mammalian species been carried out to the same degree of completion as the work reported in Chapter 2. Even though the NRY encodes far fewer genes than the genome's average, the methodology could still be applied to other chromosomes or large chromosomal regions. General steps applicable to similar exercises are as follows.

- Obtain clones from the chromosome or region of interest. If clones have already been ordered, a minimum contig is sufficient. But if clones are unordered, a certain degree of redundancy is necessary to ensure near complete coverage of the region. For the work reported in Chapter 2, 3600 unordered Y-specific cosmids representing a 5-fold redundancy of the Y were used.

- Clones are divided into pools and used to isolate representative fragments of genes, by either cDNA selection or exon trapping. cDNA selection has the following advantages: 1) not as labor intensive; 2) easy to scale up; and 3) can identify genes with very few introns. But it suffers one major limitation. If a gene is not expressed in the source tissue of the cDNA, it can not be identified. To overcome this problem, different source tissues can be pooled. Exon trapping on the other hand is not dependent on the expression pattern

of genes. But it is laborious, hard to scale up and depends on the presence of at least two introns for a gene to be recovered. Overall, cDNA selection is perhaps a better approach for the systematic cloning of genes from very large regions.

- Fragments isolated from either cDNA selection of exon trapping are cloned into plasmids from each pool. Plasmid clones are randomly picked for sequencing. The number of sequenced clones should be at least 100X the number of expected genes. This is again to ensure a near-complete coverage.

- Sequenced fragments are electronically analyzed to 1) eliminate repetitive sequences, 2) eliminate previously identified genes, and 3) reduce redundancy. After this step, the number of eligible clones can drop by a factor of 3 or more.

- Eligible clones are mapped back to the region, either by hybridization or PCR. Again, only a fraction of clones may prove to originate from the right chromosome or region.

- Phage cDNA libraries are screened with eligible clones.

- To obtain full length sequences of genes, phage clones are analyzed and where needed, phage cDNA libraries are rescreened.

The last two steps are the most time consuming. They amounted to over half the total length of the project reported in Chapter 2. Assuming an average gene density of one in every 30kb, with multiplexing and automation, a few researchers in a few years should be able to identify most genes in regions of several megabases.

iv. The human Y as a model for studying chromosomal rearrangements

Chromosomal rearrangements involving the Y are frequently observed in the human population. Microscopically detectable terminal deletions of Yq for example, occur in about 0.1% of newborn males (Hamerton *et al.*, 1975). Interstitial deletions occur at a rate of greater than one in 10,000 newborn males and account for about 10% of cases of spermatogenic failure in men (Reijo *et al.*, 1995). Two types of rearrangements less often

observed for other chromosomes are also frequently seen for the Y: isodicentric Y and ring Y. Cells carrying these derivative Y chromosomes typically exist in mosaicism with XO cells. Affected patients can have various features of Turner syndrome, presumably due to sex chromosome monosomy of their XO cells. The abundance and variety of Y rearrangements make it a good model for understanding chromosomal abnormalities in general, which are one of the leading causes of genetic defects in humans,

It is now understood that rearrangements are often the result of aberrant crossovers between two regions of the genome that share some degree of local homology. Yet, little is known about the circumstances - the particular stages of gametogenesis, the health and age of the parent, etc. - under which chromosomal rearrangements are most likely to occur. Another puzzle is terminal deletions. If they are truly terminal, two questions need to be resolved: 1) unlike most chromosomal rearrangments, they must arise through events that do not involve aberrant crossovers; and 2) their telomeres need to be repaired *de novo* by a yet uncharacterized mechanism.

The study reported in Chapter 4 has produced insight on some of the above questions. By using molecular markers, three out of ten individuals with Yq deletions (Yq-) were found to carry a minute region of Xq via aberrant crossovers. This suggests that perhaps a significant fraction, if not all apparent terminal deletions are in fact chromosomes that carry microscopically undetectable translocations. In addition, the result of the study is strongly suggestive that meiotic pairing promotes the occurrence of aberrant crossovers.

v. The importance of proper gene dosage

One fortuitous discovery of the study reported in Chapter 4 was the importance of X-inactivation to maintain proper gene dosage. For three patients, their Yq- chromosomes were shown to each carry a minute Xq translocation. These same three patients were severely retarded. Given that they all had different X and Y breakpoints on their derivative Y chromosomes, the most likely cause of their developmental retardation was the functional

disomy of genes in the duplicated region of Xq. Indeed, we were able to demonstrate twice-normal activity of a representative gene in the region, *G6PD*. Even though the importance of X-inactivation in equalizing gene dosage in males and females is widely appreciated on a theoretical level, the work reported in Chapter 4 was the first clear demonstration of the severe consequences of the failure to X-inactivate even a minute region of the chromosome.

vi. Impact on the field

Works reported in this thesis is a significant addition to our current knowledge of the human Y chromosome, and to a lesser extent, the human X. Known genes in the NRY were more then doubled. Several X homologs of NRY genes were also uncovered. A number of disease candidate genes were identified. Understanding of the Y's evolutionary history in relation to the X, and its genomic organization was deepened. The exact molecular nature of apparent terminal deletions of the Y was partially solved. The developmental importance of dosage equalization by X-inactivation was clearly demonstrated.

Results from these works could potentially serve as branching points for a number of promising studies aimed at addressing 1) specific biological processes, especially that of spermatogenesis, 2) evolution of the biology and organization of sex chromosomes, 3) disease processes related to genes on the Y and to a less extent on the X, and 4) the clinical implications of chromosomal abnormalities of the Y.

References

Charlesworth, B. (1996). The evolution of chromosomal sex determination and dosage compensation. *Curr Biol* **6:** 149-162.

Graves, J. A. (1995). The origin and function of the mammalian Y chromosome and Y-borne genes--an evolving understanding. *Bioessays* **17:** 311-320.

Hamerton, J. L., Canning, N., Ray, M. and Smith, S. (1975). A cytogenetic survey of 14,069 newborn infants. I. Incidence of chromosome abnormalities. *Clin Genet* **8:** 223-243.

Lahn, B. T. and Page, D. C. (1997). Functional coherence of the human Y chromosome. *Submitted.*

Ma, K., Inglis, J. D., Sharkey, A., Bickmore, W. A., Hill, R. E., Prosser, E. J., Speed, R. M., Thomson, E. J., Jobling, M., Taylor, K. and 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.

Manz, E., Schnieders, F., Brechlin, A. M. and Schmidtke, J. (1993). 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.

Mardon, G. and Page, D. C. (1989). The sex-determining region of the mouse Y chromosome encodes a protein with a highly acidic domain and 13 zinc fingers. *Cell* **56:** 765-770.

Reijo, R., Lee, T. Y., Salo, P., Alagappan, R., Brown, L. G., Rosenberg, M., Rozen, S., Jaffe, T., Straus, D., Hovatta, O. and 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.

Saxena, R., Brown, L. G., Hawkins, T., Alagappan, R. K., Skaletsky, H., Reeve, M. P., Reijo, R., Rozen, S., Dinulos, M. B., Disteche, C. M. and Page, D. C. (1996). The DAZ gene cluster on the human Y chromosome arose from an autosomal gene that was transposed, repeatedly amplified and pruned. *Nat Genet* **14:** 292-299.

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.

CHAPTER 6 - FUTURE DIRECTIONS

The large body of data reported in previous chapters have led to significant biological insights, but more importantly, they have opened up many possibilities for further studies. Listed below, are a few projects made immediately obvious by the added knowledge of the human sex chromosomes.

i. Further molecular characterization of novel genes on X and Y

As with the cloning of any gene, the 12 novel Y genes and 2 novel X genes reported in Chapter 2 (for which only cDNA sequences are currently available) require further molecular characterization of their genomic loci. To this end, work is already underway in the Page lab to characterize the intron/exon structure or even to sequence entire genomic loci of a subset of the novel genes. For male-specific gene families on the Y, the presence of multiple loci for each gene poses a number of technical challenges. First, it is difficult to know the precise copy number. Second, it is difficult to know which copies are functional and which are not. To address these questions, a combination of techniques are required: 1) mapping and sequencing of genomic loci, 2) identification of more cDNA clones, and 3) *in situ* hybridization to genomic DNA. Answers to these questions are important for studies that seek to correlate deletions or disruptions of sequences on the Y with phenotypic anomalies.

ii. Expression studies at the protein level

While functional knowledge is available for two of the novel Y genes, *EIF1AY* (encoding a translation factor) and *TB4Y* (encoding an actin sequester), virtually little is known about the functions of the rest of novel genes. Of particular interest are *CDY* (encoding a protein with a chromatin binding motif and a catalytic motif), *BPY1* and *BPY2* (both encoding small, highly basic proteins that are potentially involved in nucleic acid

binding). All three genes are expressed exclusively in the testis. With antibodies against these genes, one can further localize the proteins to specific cell types and even sub-cellular structures. Expression studies may open doors for further functional analyses, either *in vivo* or *in vitro*.

iii. Functional analyses of gene products

For *EIF1AY* and *TB4Y*, whose functions are well understood, it is useful to demonstrate the functional interchangeability with their X homologs. Since these genes are essential housekeeping genes, they are most likely required for cell viability. One can introduce these genes into cultured XO cells, than attempt to disrupt their homologs on the single X chromosome. Functional interchangeability can be demonstrated by recovering viable clones that carry either *EIF1AY* or *TB4Y* transgene, but are disrupted for their X homologs. Alternatively, these two human genes can be used to rescue yeast which are disrupted for their nascent homologs.

For the other novel genes whose functions are poorly understood, introducing transgenes into cultured cell lines may also provide insights into function. Over-expression of the transgene, or partial knockout of nascent transcripts by antisense may produce interpretable cellular phenotypes.

iv. Biochemical analyses of *CDY*

The protein encoded by *CDY*, with its chromatin binding motif and catalytic motif, is suitable for *in vitro* biochemical studies. CDY may bind directly to DNA, in which case, it is possible to fish out the binding substrate from either human sequences or synthesized random sequences. Alternatively, CDY may not bind directly to DNA but instead to chromosomal protein, in which case, it is possible to isolate the substrate by co-purification or by the two-hybrid system. The catalytic motif of CDY is homologous to a number of metabolic enzymes. These enzymes catalyze different reactions that all involve covalent

modifications through nucleophilic attacks (Fig. 4 in Chapter 2). Cofactor A is apparently required in all these reactions. CDY can be tested *in vitro* for its catalytic activity, using substrates of these well characterized biochemical reactions.

v. Searching for *AZF*

At least three distinct regions on the Y have been implicated in spermatogenesis. These are the so called *AZF* (*Azoospermia Factor*) regions. Men deleted in these regions show various degrees of spermatogenic failure. A subset of the 12 novel genes fall within these regions, and are thus *AZF* candidates. The definitive way to prove if any of these genes are responsible for spermatogenic failure in men is to detect either micro-deletions or point mutations of individual genes. To find micro-deletions, PCR assays can be designed to cover the entirety of each gene, with some additional 5' and 3' sequences. These PCRs can be performed on infertile men (at present, there is a collection of over 300 in the Page lab). The other approach is to detect point mutations by either SSCP or heteroduplex analysis. The detection of micro-deletions and point mutations may suffer complications for genes with multiple, closely related copies on the Y.

vi. Searching for homologs in other species and constructing animal models

Many difficulties in studying human subjects can be overcome by the construction of animal models. If homologs can be found in other species, a series of molecular, biochemical, cellular and genetic studies would become available. Knockout mice are one of the most valuable means to decipher biological functions of genes. Once they are made, human genes can be used to attempt rescues of their phenotypes. Given the difficulty in correlating phenotypes in humans to multi-copy genes, the construction of animal models is is of greater value.