# Positional cloning in Xp22 : towards the isolation of the gene involved in X-linked retinoschisis

Vosse, E. van de

Cover Page

# Universiteit Leiden

## Leiden University Repository

The handle http://hdl.handle.net/1887/28328  holds various files of this Leiden University dissertation.

**Author**: Vosse, Esther van de
**Title**: Positional Cloning in Xp22 : towards the isolation of the gene involved in X-linked retinoschisis
**Issue Date**: 1998-01-07

**Positional cloning in Xp22:
towards the isolation of the gene involved in
X-linked retinoschisis**

**Positionele klonering in Xp22:
richting de isolatie van het gen betrokken bij
X-gebonden retinoschisis**

(met een samenvatting in het Nederlands)

# Positional cloning in Xp22:
## Towards the isolation of the gene involved in X-linked retinoschisis

## Proefschrift

ter verkrijging van de graad van Doctor
aan de Rijksuniversiteit te Leiden,
op gezag van de Rector Magnificus Dr. W.A. Wagenaar,
hoogleraar in de faculteit der Sociale Wetenschappen,
volgens besluit van het College van Dekanen
te verdedigen op woensdag 7 januari 1998
te klokke 16.15 uur

door

## Esther van de Vosse

geboren te 's Gravenhage in 1965

**Promotiecommissie**

Promotor:        Prof. Dr. G.-J.B. van Ommen
Co-promotor:   Dr. J.T. den Dunnen
Referent:         Dr. F.P.M. Cremers (Katholieke Universiteit Nijmegen)

Overige leden:  Prof. Dr. R.R. Frants
                        Prof. Dr. C.J. Cornelisse
                        Prof. Dr.  P.T.V.M. de Jong (Interuniversitair
                                  Oogheelkundig Instituut, Amsterdam)

Voor mijn vader

# TABLE OF CONTENTS

## Chapter 2  Positional cloning in Xp22

### Appendix to chapter 2:

## Chapter 3 Cosmid-based exon trapping

## Chapter 4 Isolation of retinoschisis candidate genes.

## Chapter 5 Testing retinoschisis candidate genes

# CHAPTER 1

# INTRODUCTION

## CHAPTER 1 INTRODUCTION

### 1.1 X chromosome

The X chromosome is about 160 Mb in length (1) and contains an estimated 2500-5000 genes. The X chromosome has many special features that distinguishes it from the autosomes. The most obvious is that it is one of the sex determining chromosomes; XX individuals are female and XY individuals are male. All other chromosomes (the autosomes) are always present in two identical copies but the sex chromosomes differ greatly from each other, not only in size and morphology but also in gene content. Homologies and differences between the sex chromosomes are discussed in 1.1.1. Since the X chromosome is present in either one or two copies, unequal dosage of transcripts of X chromosomal genes in males and females would occur if not X inactivation would compensate for this. Dosage compensation in XX individuals is provided by transcriptional inactivation of a large fraction of the genes on one X chromosome; this is discussed in 1.1.2.

### 1.1.1   X/Y homology

The Y chromosome is 50 Mb (2) in length and based on its size could contain an estimated 750-1500 genes, however, this amount is an overestimation since the Y chromosome is gene poor (see below). The Y chromosome has two main functions: it is required for the male phenotype and provides a pairing partner for the X chromosome during male meiosis. The gene (or genes) required for initiating male development is called the testis determining factor (TDF). Only one gene has been identified that is a candidate for TDF; sex-determining region Y (SRY) (3). Very few other Y-specific genes have been isolated thus far (2) (the Deleted in Azoospermia gene cluster (DAZ) gene family (4,5), Spermatogenesis gene on the Y (SPGY) (6), testis-specific protein, Y-encoded (TSPY) gene family (7), Ribonucleic acid binding motifs (RBM) gene family (8)).

Only a small portion of the genes on the X and Y chromosome are shared, consequently most genes on the X chromosome are haploid in males. The shared genes are transcribed from both X and Y chromosome and are located in the pseudoautosomal regions (PAR) which effectively behave as autosomal genes. The PAR1 region on the Xp and Yp telomeres is a region of 2.6 Mb (2), delimited by the pseudoautosomal boundary (PAB). The PAR2 region (9) on subtelomeric Xq and Yq is much smaller (320 kb). So far only two genes have been cloned from

11

this region: the Interleukin 9 Receptor (IL9R) (10) and a Synaptobrevin-like gene (SYBL1) (11). The genes identified in PAR1 and PAR2 are indicated in Figure 1. In contrast to the genes in PAR1, SYBL1 in PAR2 is subject to X inactivation. Even more remarkable is that the Y copy is also inactive. The (in)activation state of IL9R is not known and if this gene shows the same pattern as SYBL1, a position effect caused by the heterochromatin on Yq may play a role in this. Another explanation is that these -although located in the PAR- are pseudo-genes.
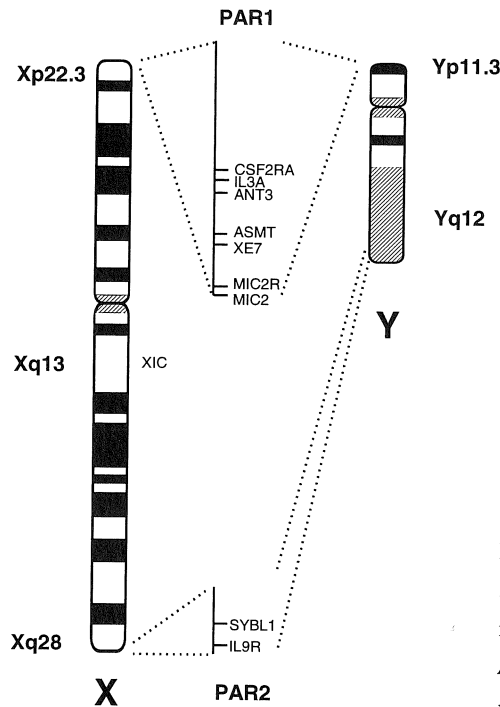


Figure 1: The pseudoautosomal regions. PAR1 has a size of 2.6 Mb, PAR2 has a size of 320 kb. The pseudoautosomal regions are identical in X and Y. The border of PAR1 is an *Alu* repeat, the border of PAR2 is a LINE repeat. XIC localised in Xq13 is discussed in 1.1.2.

Pairing between the X and Y chromosome during male meiosis seems to involve only part of the short arm of the X and Y and includes an obligatory cross-over in PAR1. However, recent reports have been published in which pairing at PAR2 is described (9,12,13). The rate of recombination in PAR2 is 168 kb/cM (14), in PAR1 it is 55 kb/cM (15). These rates are very high compared to the average for the genome (1 Mb/cM), due to the fact that these regions, while

small, have an obligate cross-over (at least in PAR1).

Other regions of homology between X and Y mainly consist of pseudo-genes on the Y, and are believed to have arisen through non-homologous pairing between the X and Y (16) followed by inversions (2). These regions are hotspots for illegitimate recombination and are located in Xp22.3/Yq11.21, involving recombination between the KAL-X and KAL-Y gene (17), in Xp22.3/Yp11, involving recombination between PKX1 and PKY1 (18), and in Xp22.3/Yp11 in a region just proximal from the PAB, involving recombination between sequences (1/3 in repeats) that have a high homology (96-98%) between X and Y (19,20).

## 1.1.2 X inactivation

To compensate for the unequal dose of X genes in males and females, one of the X chromosomes is inactivated at an early stage of embryogenesis in all somatic tissues in the female. This phenomenon is called X inactivation or lyonisation after Mary Lyon who first described it in 1961 (21). X inactivation takes place at the time of uterine implantation (22), occurs in all cells except the germ cells and is random and maintained during all further cell divisions. This mechanism of dosage compensation is unique to mammals (23).

Basically, all but very few genes on the inactivated chromosome ($X_i$) are thought to be inactive except for the genes in PAR1 and the gene(s) at the X inactivation center (XIC) itself. On the active chromosome ($X_a$) all genes are active except for the gene(s) of the X inactivation center. Amongst the few genes outside PAR1 that escape X inactivation are for example; SMCX (24), SB1.8 (25) and ubiquitin-activating enzyme (UBE1) (26). These genes do not have detectable Y homologues, so their transcription levels are likely to be different in male and female, the functional significance of this difference is not yet known. The existence of functional homologues on the Y chromosome with widely divergent or different sequence can not be ruled out.

The X inactivation center (XIC) has been localised to Xq13 (Figure 1) (27). X inactivation spreads from the XIC across the chromosome. The XIST gene (specifying an X-inactive specific transcript) maps to the XIC and is only transcribed from the $X_i$ (28). X inactivation is preceded by XIST expression (29) and as no other genes have been found in the XIC it is assumed that XIST is the gene causing X inactivation. XIST does not code for a protein and only inactivates the X it is expressed from (in cis) . Recent studies using XIST knockout mice have proven that XIST is essential for X inactivation (30).

The X*i* (also known as the Barr body) replicates late in S phase and is visible as condensed heterochromatin in the nucleus, even in G1. Although the mechanism is not completely clear, inactivation seems to be maintained by methylation of the 5'-region of genes (31). Experiments with patient-derived cell lines with supernumery X chromosomes have shown that XIC also is involved as part of a counting mechanism to ensure the appropriate activity state of X-linked genes by allowing only one active X per two sets of autosomes (32).

When one of the X chromosomes harbours a gene that through a mutation is deleterious to cells in a specific tissue, a skewed X inactivation is observed. This results in the presence of -in all or most of the cells in this tissue- the non-mutant X as the active chromosome. This is not caused by a change in the activity state of the X chromosomes but by a selection against the cells that contain the X chromosome with the mutant gene as the X*a* (for example in Incontinentia Pigmenti (IP) (33)). The opposite effect is often found when larger deletions of the X chromosome or X/autosome translocations are present, in those cases the normal X chromosome is inactivated. X/autosome translocations (34) have been found in, for example, patients with Duchenne muscular dystrophy (DMD) (35), magnesium-dependent hypocalcemia (HSH) (36), and Hunter disease (37).

### 1.1.3 Evolutionary origin of the sex chromosomes

In mammals XX individuals are female and XY individuals are male, but in birds this is the other way around (38). In the much more distantly related *D. melanogaster* XX are females and XY males, while in *C. elegans* XX are hermaphrodites and XO are males (39). Three different forms of sex chromosomes are found in fish; some fish have almost identical X and Y chromosomes, others have an X and Y which hardly recombine and finally there are fish that have lost the Y completely (38). Obviously, many species have developed sex chromosomes independently during evolution so there must be a strong evolutionary force pushing all these species to solutions of similar nature but different in endpoint.

The most likely reason for the initial cosexual species (hermaphrodites) to have favoured the evolution of separate sexes, is that self-fertilisation is more likely to produce unfit progeny than sexual reproduction (40). Cosexual species have all genes required for both 'male' and 'female' reproductive organs located on their autosomes. A simple form of acquiring a difference between sexes is seen in *C. elegans*, where missing one of these chromosomes causes male development instead of hermaphroditic. This X-monosomy however causes non-disjunction

during meiosis, resulting in non viable embryos in part of the progeny.

In general, X and Y chromosome are believed to have evolved from an autosomal pair of chromosomes (two 'pre-XY' chromosomes). The first difference between these two chromosomes may have been the occurrence of a large deletion or inversion on one of them, disturbing homologous recombination locally. Once homologous recombination was disturbed, more and more of this mutated (Y) chromosome was lost because it became prone to rearrangements and steady loss and inactivation of genes (41). In addition, a range of pseudo-genes originating from autosomal genes have accumulated on the human Y chromosome, probably through retrotransposition (16). These processes have generated a chromosome which can only have retained and/or accumulated genes that would enhance male fitness, and will otherwise only have been selected for appropriate size for efficient meiotic segregation.

Whether the evolution of a dosage compensation system was required before the degeneration of the Y chromosome could start (42) or whether it evolved as a consequence has not been proven. In *C. elegans*, transcription of both X's in hermaphrodites is reduced to ~50%, regulated by at least 8 genes and depending on X:A ratio (43). This suggests that dosage compensation was already available, independent of the presence of a Y chromosome. Dosage compensation in Drosophila is mainly obtained through increased expression of genes on the male X (regulated by male-specific lethal genes on the Y), although there is also evidence for a parallel dosage compensation pathway that down regulates some genes on the X in females. In mammals dosage compensation involves inactivation of most genes on one X in females (see 1.1.2). In many species unique dosage compensation systems have evolved that allowed development of separate sexes and thereby opened the way to evolution into a 'higher' order of species.

**The mammalian sex chromosomes**   The size of the X chromosome has been strongly conserved amongst eutherian ('placental') mammals, being 5% of the haploid genome, and it is also strongly conserved in gene content (44). Many genes on the human X chromosome have also been localised to the X chromosome in a wide variety of other eutherian mammals. Comparison with other therian mammals, the metatheria (marsupials) and prototheria (monotremes), shows that Xq is part of the X in all therians. Human Xp genes, in contrast, are located autosomally in both marsupials and monotremes. For instance, in monotremes the human Xp-linked genes Synapsin 1 (SYN1), DNA polymerase α (POLA) and Ornithine transcarbamylase (OTC) are

located in one block on chromosome 1, while the Xp-linked genes Dystrophin (DMD), Synapsin1 (SYN1), Cytochrome b heavy chain (CYBB) and Monoamine Oxidase A (MAOA) are located in one block on chromosome 2 (45) (Figure 2). In marsupials, MAOA, ZFY, OTC, DMD, STS, POLA, SYN1 and OATL1 have also been shown to be autosomal (46). This means either that this region was lost from an ancestral X chromosome in the marsupial and monotreme lineages or was acquired by an ancestral X in the eutherian lineage. Not much is known yet about the gene content of the Y chromosome in other therians.
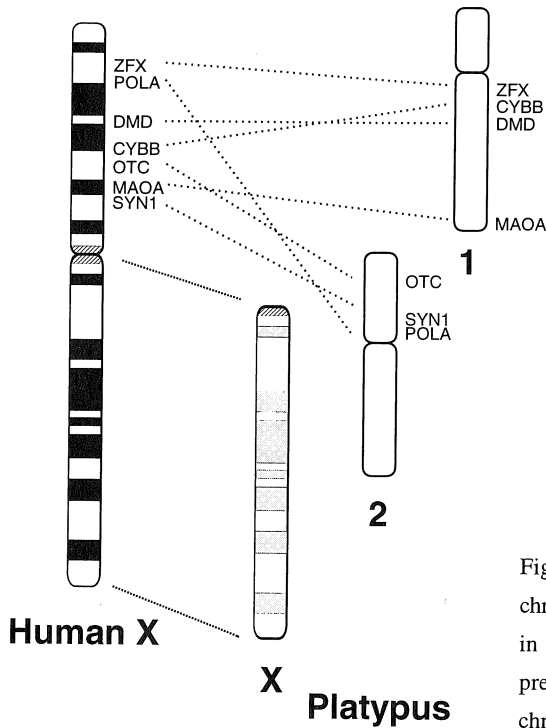


Figure 2. Assembly of the therian X chromosome. Human Xq genes are found on X in prototheria as well, human Xp genes are present in two blocks on prototheria chromosomes 1 and 2.

Not only the gene content of the X chromosome is highly conserved among eutherian mammals, so is the order. Although blocks of genes have been rearranged, the order of genes within these blocks is conserved. These rearrangements (through several inversions) are typical for each subclass. As an example the mouse X is compared to the human X in Figure 3.

| Xp22.3 | | |
| Xp22.2 | | 1 |
| Xp22.1 | | 2 |
| Xp21.3 | | |
| Xp21.2 | | 3 |
| Xp21.1 | | 4 |
| Xp11.4 | | |
| Xp11.3 | | 5 |
| Xp11.2 | | 6 |
| Xp11.1 | | 7 |
| Xq11 | | |
| Xq12 | | |
| Xq13 | | |
| Xq21.1 | | |
| Xq21.2 | | 8 |
| Xq21.3 | | |
| Xq22 | | |
| Xq23 | | |
| Xq24 | | |
| Xq25 | | |
| Xq26 | | 9 |
| Xq27 | | |
| Xq28 | | |
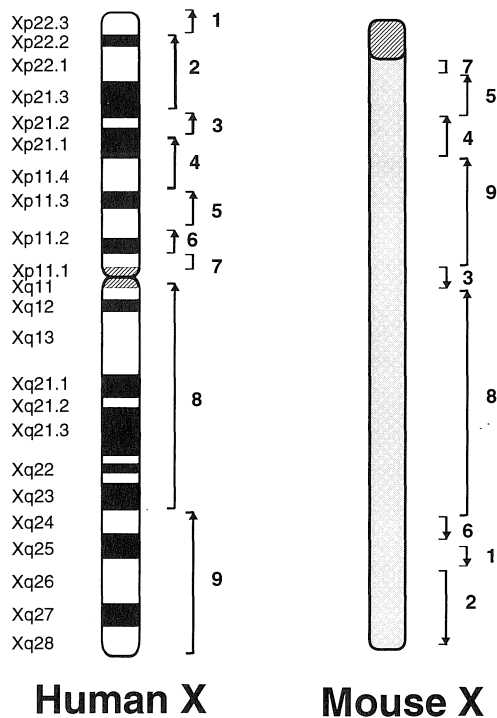
**Human X**          **Mouse X**

Figure 3: Comparison of the human and mouse X chromosomes. Through several inversions in both mouse and human X the order of the genes became different but the 9 conserved blocks can still be recognised. Note: mouse chromosomes do not show banding

## 1.1.4 Deletions in Xp, contiguous deletion syndromes

Deletions and rearrangements of chromosomal regions can greatly facilitate the mapping of disease genes. Comparison of the deletions or phenotypes in patients with contiguous deletion syndromes can be used to assign disease genes to a distinct region. For the X chromosome, deletion mapping has been very useful for the characterisation of several genomic regions, for example Xp22.3 (47-49) and Xp21 (50-52) (Figure 4). In Xp22.3 amongst others, the identification of the genes for Kallmann syndrome (53,54) and X-linked ichthyosis (STS)(55) have been facilitated by the available patients with deletions and contiguous deletion syndromes. Many of these deletions are thought to be a result of aberrant recombination between the X and Y chromosome (see 1.2.1). In Xp21, for instance the genes for Duchenne muscular dystrophy (DMD) (56), McLeod syndrome (XK) (57), and X-linked chronic granulomatous disease (CYBB) (58) have been identified using deletion mapping.

17
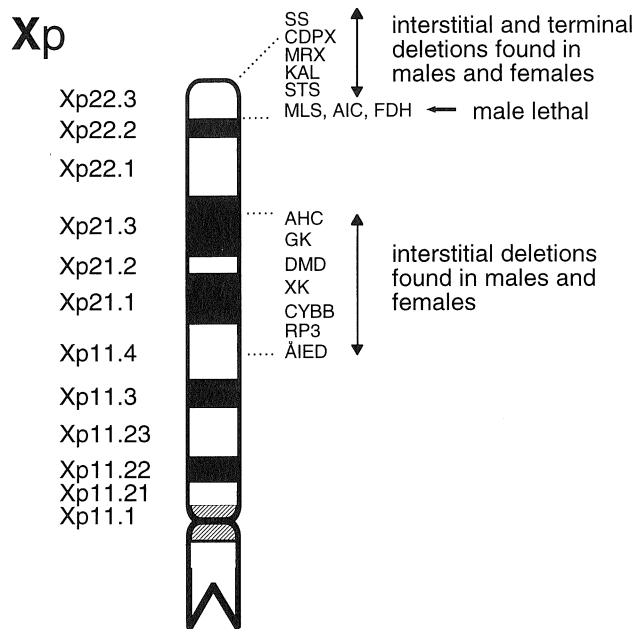
Figure 4. Contiguous deletion syndromes on Xp.

Contiguous deletion syndromes on Xp have been found in Xp22.3 and in Xp21 (extending into Xp11.4). In Xp22.3 contiguous deletion syndromes are either interstitial or terminal deletions that can involve combinations of short stature (SS), chondrodysplasia punctata (CDPX), Kallmann syndrome (KAL), mental retardation (MRX) and X-linked ichthyosis (STS) in both males and females. Microphthalmia with linear skin defects (MLS), Aicardi syndrome (AIC) and focal dermal hypoplasia (FDH, also known as Goltz syndrome) are male lethal and therefore almost exclusively found in females. The phenotypes of these three syndromes overlap so they probably result from a defect in the same gene (64) or are due to a contiguous deletion syndrome (65). In Xp21 contiguous deletion syndromes are interstitial deletions that can involve combinations of Duchenne muscular dystrophy (DMD), chronic granulomatous disease (CYBB), McLeod syndrome (XK), retinitis pigmentosa (RP), mental retardation (MRX, not indicated in figure since location is still unclear), glycerol kinase deficiency (GK), adrenal hypoplasia congenita (AHC) and Åland island eye disease (ÅIED). No contiguous deletion syndromes have been found in Xp22.1-p22.2.

18

In contrast, in the stretch of DNA between these two regions (Xp22.1-p22.31), deletions are rare. Deletions found in this region are in general due to inheritance of an X/autosome translocation and are only found in females where the phenotypic effect is either generated by spreading of X inactivation onto the autosome, nullisomy of the missing autosomal region, or by inactivation of the normal X, causing functional nullisomy of the deleted region (35,59). No large terminal or interstitial deletions (other than through a translocation event) of this region have been found. The apparent lack of large deletions in the Xp22.1-p22.31 suggests that one or more genes may be present in this region that, when present in single copy in female, or absent in male, would be lethal. Consistently, three syndromes in Xp22.31 have been found, microphthalmia with linear skin defects (MLS), Aicardi syndrome (AIC) and focal dermal hypoplasia (FDH, also known as Goltz syndrome), that appear to be male lethal.

Mutations in the genes that have been isolated so far from the Xp22.1-p22.2 region are seldomly due to deletions and when deletions are detected these are small. In the *PEX* gene, mutated in X-linked hypophosphatemic rickets (HYP), in only 4 patients out of 150 families tested, deletions were detected that ranged in size from less than 1 kb to over 55 kb (60). In *PHKA2* (phosphorylase kinase liver α-subunit), the gene mutated in X-linked liver glycogenosis type I and II (XLG) initial studies showed 1 deletion (of 3 bp) out of 2 XLG I families studied and 1 deletion (of 3 bp) out of 4 XLG II families studied (61,62). In *RSK2*, the gene mutated in Coffin-Lowry syndrome (CLS) an initial screen of 76 families revealed two deletions of 118 bp and ~2kb respectively (63).

**1.1.5 Disease genes in Xp22.1-p22.2**

Several disease genes have been localised in the Xp22.1-p22.2 region (see Figure 5), some of which were recently found. X-linked glycogenosis type I and II (XLGI and II, MIM 306000) are caused by mutations in *PHKA2* (61,62,66), X-linked hypophosphatemic rickets (HYP, MIM 307800) by mutations in *PEX* (60), Coffin-Lowry syndrome (CLS, MIM 303600) by mutations in *RSK2* (63). We have focused on RS and KFSD which are discussed below.

**X-linked juvenile retinoschisis**

X-linked juvenile retinoschisis (RS, MIM 31270) is an eye disease that causes acuity reduction and peripheral visual field loss, typically beginning early in life. The first report of what is now known as RS was by Haas in 1898 (77) who reported the simultaneous findings of changes in retina and choroid and already suggested hereditary degeneration as possible cause. The term retinoschisis was introduced by Wilczek in 1935 (78). The first suggestion of sex-linked inheritance however was by Sorsby in 1951 (79). The frequency is about 1:10.000 (80). Most patients are diagnosed at school age, although pathological changes are probably already present at birth and progression is in general slow. Folding and splitting of the macula (simulating cysts) cause the visual acuity loss (81), intra retinal splitting through the nerve fiber layer causes the peripheral visual field loss. Severity can range from mild acuity reduction to total blindness at an early age due to complete retinal detachment (82,83).The RS disease gene has a high penetrance, with variable expression between families but little variation within a family, this phenotypic variation may be due to different mutations in one gene. Other explanations for the phenotypic variation are differences in expression, modifying genes, or environmental factors. No evidence for genetic heterogeneity has been found (84,85).

Figure 5. Disease gene regions in Xp22.1-p22.2.
The markers and scale (in Mb) are according to the 6th X Chromosome Workshop (1). Markers are indicated above the bar, known genes are indicated under the bar. SEDL= spondylo-epiphyseal dysplasia (MIM 313400) (67), NHS= Nance-Horan syndrome (MIM 302350) (68), RP15= X-linked cone-rod degeneration (MIM 300029) (69), DFN6= sensorineural deafness (DFN6, MIM 300066) (70), PRTS= X-linked mental retardation with dystonic movements of the hands (MIM 309510) (71), MRX= non-specific X-linked mental retardation (MIM 309540) (72-74), RS= X-linked juvenile retinoschisis (MIM 312700) (75), KFSD= keratosis follicularis spinulosa decalvans (KFSD, MIM 308800) (76), HSH= hypomagnesemia with hypocalcemia (HSH, MIM 307600)(36).
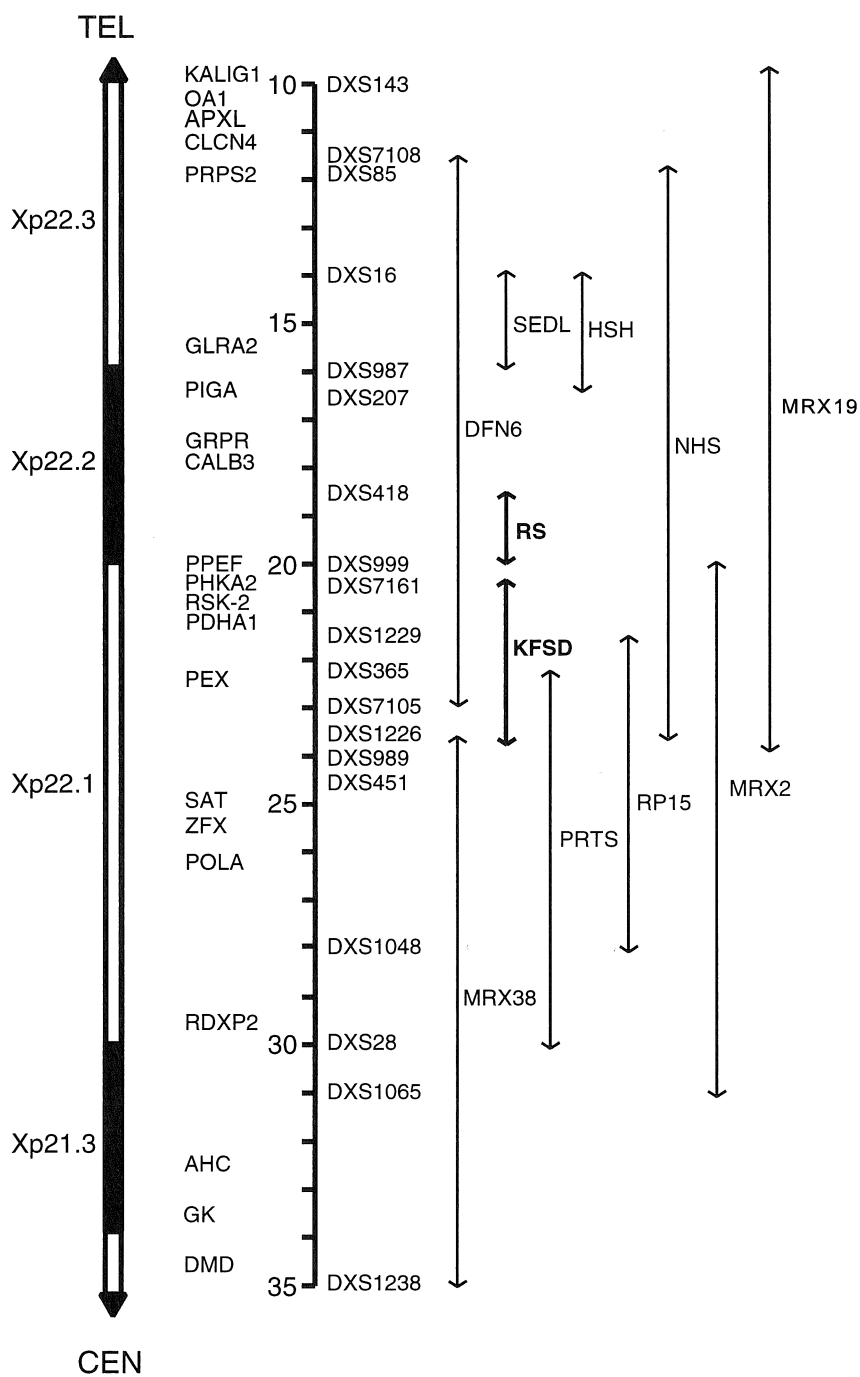
Except in one case of a female with RS, who was probably a homozygote due to a consanguineous marriage (86), all reported patients are male. Vision in female carriers is usually normal. Although publications have stated for a long time that female carriers do not have any symptoms of the disease (81,87-90) closer examination showed abnormal cone-rod interactions in some of the carriers (91) and peripheral lesions of the retina in 4 of 5 carriers (92). Features as reported in these carriers however can also be found in the normal population (93).

The most characteristic clinical finding in RS are macular changes, consisting of any of the following: splitting, radial folds, pigment dissemination and development of macular scars (81). Other findings include white areas in the peripheral retina, hyperopia, liquefaction of the vitreous body, vitreous strands, peripheral retinoschisis (in 50% of cases (94)), constricted nasal visual field, subnormal ERG, and a range of rarer findings (81).

The biochemical defect of RS is unknown, but histopathologic and electrophysiologic studies suggest a defect in the Müller cell (93-96) possibly an inability of these cells to remove the extracellular potassium ions resulting from exposure to light (93,97). In normal eye development the Müller cell has a function as a migration determinant for retinal development. Another theory proposes that the retinoschisis arises from delayed development of the retinal and choroidal vasculature, causing the retina to outgrow its blood supply (98) but this does not explain the foveal schisis. No treatment is available, although surgical intervention is sometimes performed with varying success rates and often leading to complete retinal detachment and other complications (83,99,100).

Table 1. RS candidate region
Genetic analyses in the RS disease gene region. A= Wieacker *et al.* 1983 (102), B= Alitalo *et al.* 1987 (103), C= Gellert *et al.* 1988 (90), D= Dahl *et al.* 1988 (90), E= Alitalo *et al.* 1988 (89), F= Sieving *et al.* 1990 (84), G= Alitalo *et al.* 1991 (104), H= Kaplan *et al.* 1991(92), I= Oudet *et al.* 1992(105), K= Bergen *et al.* 1993 (106), L= Biancalana *et al.* 1994 (107), M= George *et al.* 1994 (85), N= Bergen *et al.* 1995 (80), O= Pawar *et al.* 1995 (82), P= Shastry *et al.* 1996 (108), Q= Van de Vosse *et al.* 1996 (109). Marker order is according to the 6th international workshop on X chromosome mapping (110). ✳ indicates a marker used in linkage analysis, ● indicates the marker with the highest lod score. ▲ and ▼ indicate a recombination between the marker and the RS disease gene. Based on the recombinations in column M and Q the candidate region for RS is located between DXS418 and DXS999 (hatched region). The recombinants identified in earlier studies may provide a valuable further refinement of the region when analysed with markers that have become available more recently between DXS418 and DXS999.

22

| Marker | A | B | C | D | E | F | G | H | I | K | L | M | N | O | P | Q |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Xg | * | | | | | | | | | | | | | | | |
| DXS89 | | | | | * | | | | | * | | | | | | |
| DXS143 | | | | | | | | | | * | | | | | | |
| DXS85 | | * | | *▼ | * | *▼▼ | | | | * | | | | | | |
| DXS16 | | * | *▼ | * | * | * | * | * | * | * | | | * | | | |
| DXS9 | *● | | *●▼ | | * | *● | | | | * | | | | | | |
| DXS987 | | | | | | | | | | | | | | * | * | |
| DXS207 | | | | | *▼▼ | | | * | *● | | * | | *▼ | *●▼ | *● | |
| DXS1053 | | | | | | | | | | | * | ▼ | | | | |
| DXS197 | | | | | * | | *● | | | | | | | | | |
| DXS43 | | | * | *● | *▼ | *▼▼ | | *● | * | * | * | | | | | |
| DXS418 | | | | | | | | | | | | | | | | ▼ |
| DXS257 | | | | | | | | | | | * | | | | | |
| DXS999 | | | | | | | | | | | * | ▲ | * | * | * | |
| DXS7161 | | | | | | | | | | | * | | | ▲ | | ▲ |
| DXS443 | | | | | | | | | | | * | | * | | * | |
| DXS1229 | | | | | | | | | | | * | | | | | |
| DXS365 | | | | | | | | | | | * | | * | | * | |
| DXS1052 | | | | | | | | | | | * | | | ▲ | | |
| DXS274 | | | | | | | *▲▲▲ | *▲▲▲ | *▲▲ | | | | | | * | |
| DXS92 | | | | | | | * | | | | | | | | | |
| DXS1226 | | | | | | | | | | | | | | ▲ | | |
| DXS41 | | *● | * | * | *▲▲ | *▲▲ | | *▲▲ | * | *● | * | | | | | |
| DXS451 | | | | | | | | | | | | | * | | | |
| ZFX | | | | | | | * | | | | | | | | | |
| DXS208 | | | | * | | | | | | | | | | | | |
| DXS207 | | | | * | | | | | | | | | | | | |
| DXS28 | | | | * | | | | | | | | | | | | |
| DXS164 | | | | *▲ | * | | | *▲▲▲ | | | | | | | | |
| DXS206 | | | | *▲ | | | | | | | | | | | | |
| DXS84 | | | | *▲ | | | | | | | | | | | | |
| OTC | | | | *▲ | | | | | | | | | | | | |

The first linkage studies and recombination events using RFLP markers placed RS in Xp22 between DXS43 and DXS41. Further studies using microsatellite markers placed the RS locus in decreasing intervals (see Table 1), until the present localisation of RS in a 1 Mb interval between DXS418 and DXS999 (see also Chapter 2).

**Candidate genes** Recently, several candidate genes for retinoschisis have been cloned by members of the Retinoschisis Consortium (see note). The first two, *PPEF* (111) and *Txp3* (112) have been excluded as the genes mutated in RS (see Chapter 5). Two others; *SCML1* (113) and *Txp7* (114) are still being tested. Recently, the complete RS candidate region has been sequenced by the Sanger Centre based on clones provided by the Retinoschisis Consortium. Analysis of this sequence will reveal many novel genes present in the region that can be tested as candidates for RS.

Note: The Retinoschisis Consortium consists of the following groups:
B.Franco, A. Ballabio in Milan, Italy.
T.Alitalo, A. De la Chapelle in Helsinki, Finland.
D.Trump, J.R.W.Yates in Cambridge, United Kingdom.
W. Berger, H.H. Ropers in Berlin, Germany.
A.A.B. Bergen in Amsterdam, the Netherlands.
T.E. Darga, P.A. Sieving, Michigan, U.S.A.
E. Van de Vosse, J.T. Den Dunnen in Leiden, the Netherlands.

Note: other forms of hereditary retinoschisis are; autosomal recessive, autosomal dominant and some unclear hereditary forms of retinoschisis. Acquired forms of retinoschisis; degenerative retinoschisis, also called senile retinoschisis and secondary retinoschisis associated with various diseases, of which diabetic retinopathy is the most common (101).

**Keratosis follicularis spinulosa decalvans**
Keratosis follicularis spinulosa decalvans (KFSD, MIM 308800) is an extremely rare disorder affecting skin and eyes. Patients show hyperkeratosis (thickening) of the skin of the neck, ears, palms and soles, loss of eyebrows, eyelashes and beard, thickening of the eyelids with blepharitis and ectropion, corneal degeneration, photophobia and baldness (alopecia) in winding streaks. The symptoms diminish with age. KFSD was first described by Lameris in 1905 (115). The name

24

KFSD was given by Siemens (116) and the disorder has also been called Siemens syndrome. Siemens (117) described KFSD in 1925 as the first dominant sex-linked disease, however, only about half of the carriers show (mild) clinical symptoms, which is more suggestive of skewed X inactivation than of KFSD being a dominant sex-linked disease.

Essentially five families have been described thus far, located in Germany (117), the Netherlands (76), France (118), Finland (119) and the UK (120). Linkage analysis using RFLPs placed the gene for KFSD in Xp22 between DXS16 and DXS269 (121), analysis of recombinants using microsatellite markers further refined the region to between DXS7161 and DXS1226 (76) (see also Chapter 2). KFSD has been reported to show genetic heterogeneity (122,123) but since few families are available for research this may just reflect a variation in phenotype between families.

Since there are so few families with KFSD available, no systematic efforts had been undertaken yet to specifically clone the KFSD gene prior to this study. However, because the KFSD candidate region overlaps with other disease candidate regions, several genes have been cloned that can be tested as KFSD candidate genes purely based on location. The biochemical defect in KFSD is still unknown.

## 1.2 Positional cloning of disease genes

To identify the molecular mechanism underlying a hereditary disease, the mutant gene needs to be identified. If a cellular defect resulting from the mutation is identified, cloning by functional complementation is possible. If the (defective) protein is known, its identification can lead to the cloning of the corresponding gene. In most hereditary diseases however, neither protein nor cellular function are known and in those cases positional cloning is used to identify the disease gene. In the early days of gene identification, functional cloning was the only way of gene identification. Since the mid-80's positional cloning has rapidly taken over simply because techniques became available that allowed the analysis of larger regions. The ideal approach is when both functional and positional information can be used to identify a gene.

Positional cloning is usually done following a strategy (illustrated in Fig. 6) that narrows the search from the complete genome to a small region, preferably a single gene. The first step is genetic mapping: the region on a chromosome where the disease gene is localised is defined by linkage and recombinant analysis (discussed in 1.2.1). The second step is physical mapping; identification and isolation of genomic clones (e.g. YAC, P1, BAC, cosmid) that are located in

the disease gene region, construction of contigs and assembly of a restriction map (discussed in 1.2.2). The final stage is isolation of transcripts for candidate genes amongst which the disease gene may be present (discussed in 1.3).

Linkage analysis   Isolation of clones,   Digestions,   Gene
and recombinants   contig construction   hybridisations   identification

Xp22.2
Xp22.1

NotI

BssHII

BssHII
NruI
NotI

BssHII

NotI

EagI

SfiI

NruI

SfiI
NruI
NotI

**X**   Genetic map   Physical map   Restriction map   Transcript map
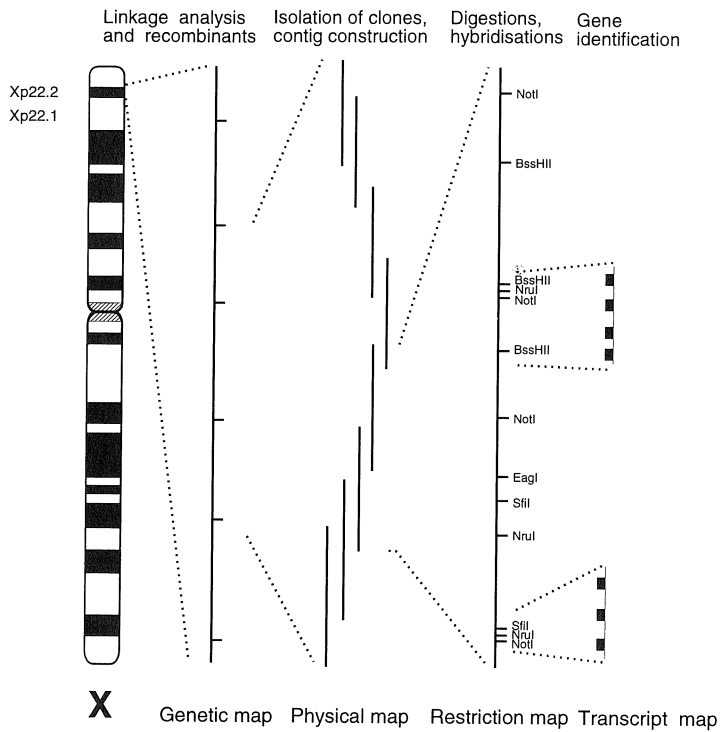
Figure 6: From chromosome to gene.

## 1.2.1  Genetic mapping

In order to isolate a gene through positional cloning, the genetic location of the gene needs to be known. Cytogenetically visible chromosomal aberrations, such as translocations or large deletions, may give a direct indication of the region. When these are not present in the patients, systematical scanning of the genome with polymorphic markers (linkage analysis) in a subset of families is sufficient to acquire an approximate chromosomal location. Once an approximate

chromosomal localisation has been established, more refined linkage analysis is used to detect markers close to a disease gene by measuring whether certain marker alleles are statistically more often inherited together with the disease than the frequency of the allele in the general population would suggest. The further two markers are apart, the more likely it becomes that a cross-over between the two markers occurs during meiosis. The general rule for such cross-overs is that when, amongst every 100 meioses, typically 1 recombination occurs (1% recombination): this is an interval of 1 centiMorgan (cM). This 1 cM interval on average represents a physical region of around 1 Mb, but this can differ greatly between regions, due to local differences in recombination rate along our chromosomes.

In recombinant analysis, usually employed for finer localisation, individual meioses are analysed to see whether a recombination has taken place between any of the markers and the disease gene. Identification of recombinations located either distal or proximal to the gene is used to reduce the candidate region.

The interval to which a disease gene can be localised using genetic mapping is limited (124). In principle, these limitations are set by the number and distribution of the available polymorphic markers, the distance (in cM) between these markers (affecting the chance of detecting cross-overs), the heterozygosity of the markers and the number of available patients. In practice, however, through the large abundance of genetic markers currently available, the genetic mapping is mainly limited by the number of patients and hence the number of cross-overs in the candidate region. The point at which the genetic interval in which the disease gene is localised is small enough to start physical mapping is hard to define. Starting physical mapping with a too large genetic interval is a waste of time and energy, while continuing genetic mapping for too long may not provide the increased refinement of localisation due to lack of informative recombinants. However, one should always remain alert to new patients and family extension, i.e. classical advances, since, if a new recombinant appears, this often greatly limits the scan region.

### 1.2.2 Physical mapping

**Isolation of clones** Once a sufficiently small interval has been established by genetic mapping or by chromosomal aberrations, physical mapping is initiated. One physical method is the isolation of YAC clones as these contain inserts of ~1 Mb, are easy to handle and are readily available. Markers located in the region can be used to isolate YAC clones either by hybridisation

of gridded YAC libraries (125), PCR screening of YAC pools (126), or -since about 1993- by screening databases (127). In regions where markers are sparse, chromosome walking can be done using end-clones (128), jumping and linking libraries (129), *Alu*-PCR products of previously isolated YACs (130) or of radiation-hybrids (131). Many whole-genome or chromosome-specific YAC libraries are available (125,132-134).

Positive genomic clones must be rescreened and their marker content established. Because a relative high percentage of YAC clones are derived from ligation of DNA fragments from different genomic regions, chimerism should be checked. This is done either by fluorescent *in situ* hybridisation (FISH) of the whole YAC -which will at the same time confirm its chromosomal localisation-, by mapping YAC end clones using FISH, or by hybridisation to panels of hybrid cell-lines. The disadvantage of using entire YACs in a FISH experiment is that small chimeric regions may not be detected. The additional advantage of generating end clones is that these can be used as markers in subsequent experiments. The length of the YACs is determined by pulsed-field gel electrophoresis (PFGE).

**Contig assembly** A contig is assembled based on marker content, on fingerprinting of shared restriction or PCR fragments or on a combination of the two. The oldest fingerprinting method is based on comparison of the hybridisation patterns after restriction digestion of the clones and hybridisation with a repetitive element (e.g. *Alu*, Line-1, THE) (135-137). The PCR based methods are based on radioactive PCR using *Alu* specific or random primers on the clones and analysing the PCR products after electrophoresis on a sequencing gel (138,139). All approaches generate a unique pattern of bands for each clone that can be analysed using computer programs.

Because YAC clones frequently show deletions, rearrangements and chimerism (132,140) it is important to analyse several coverages of the whole contig rather than a minimum tiling path of YACs. Better still is to have a contig cloned in different cloning systems (e.g. P1-clones and BACs) to analyse clones from independent sources. An additional advantage of constructing a contig from different cloning systems is that when certain genomic regions are unclonable or unstable in one system they may be obtained from another system.

**Restriction mapping** YACs in general have inserts too large to construct detailed restriction and transcript maps. A frequently used step to improve the resolution of a contig and to allow the construction of a restriction map is the isolation of clones that are an order of magnitude smaller

than the YAC clones in the original contig. The isolation of smaller clones can be performed by screening of P1- (up to 100 kb), BAC- (up to 100 kb) or cosmid (up to 40 kb) libraries (or subcloning of the YACs into one of these vectors), or by an alternative approach, YAC fragmentation.
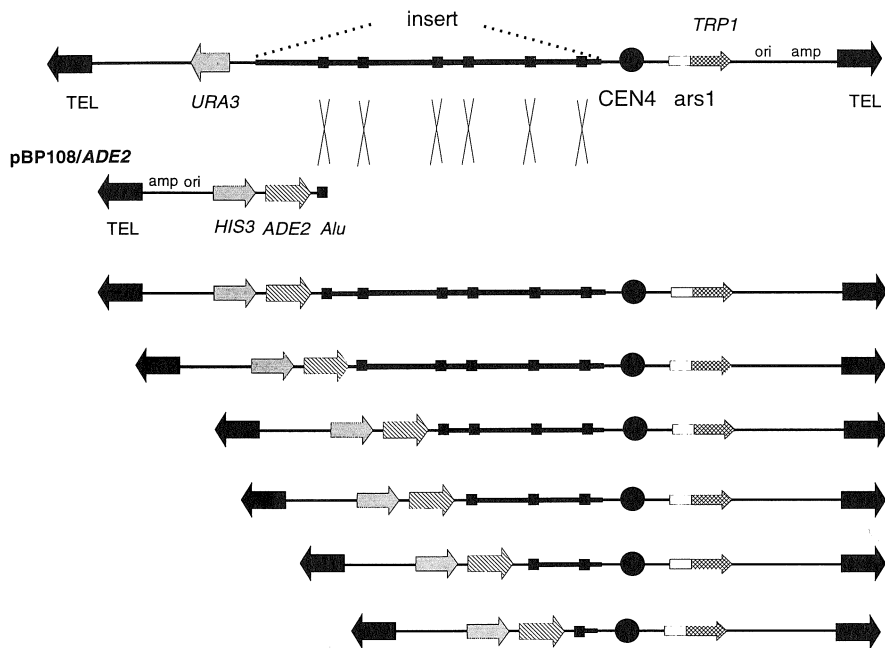


Figure 7: Principle of YAC fragmentation.
Upon transformation of yeast containing a YAC with plasmid pBP108/*ADE2*, homologous recombination between an *Alu* in the YAC and the *Alu* in pBP108/*ADE2* will occur in part of the yeast cells. Growing of the yeast on medium lacking tryptophan and adenine allows selection of fragmented YACs (which contain both *ADE2* and *TRP1*). A panel of fragmented YACs with various insert sizes is thus generated.

**YAC fragmentation** Since YAC fragmentation was first described by Pavan *et al.* in 1990 (141), several improvements of the YAC fragmentation vectors (142,143) made it a rapid and simple way of generating a panel of clones of decreasing size that can be used for clustering of markers and clones to defined, consecutive regions ('binning') and restriction mapping. YAC

fragmentation is based on the homologous recombination between a repeat in the YAC insert and a repeat in a YAC-vector arm containing a selectable marker not present in the original vector arms. After the recombination the replaced vector arm plus part of the insert is lost and a smaller YAC is obtained (Figure 7). A panel of fragmented YACs can be used to generate a restriction map without having to use partial digestions of YACs that are usually difficult to interpret (144). Panels of fragmented YACs have also been used to delimit a duplicated chromosomal region (145) and to refine translocation breakpoints (146).

### 1.3 Identification of transcripts

Methods to identify transcripts can roughly be divided in transcript dependent (cDNA based) and independent (genomic DNA based) techniques. Not one technique is capable of identifying all genes in a region, so two or more complementary techniques are required to construct a complete transcription map.

The advantage of cDNA based techniques is that when a cDNA is identified this will immediately tell something about the tissue and stage it is expressed in and it is proof that the region is transcribed. The quality of the cDNA is very important, the presence of genomic DNA, incompletely processed RNA and rRNA should be avoided (by polyA$^+$ selection).

The advantage of genomic DNA based techniques is that they are independent of the time and tissue of transcription, thus enabling the isolation of genes expressed only transiently, in a specific subset of cells, or at extremely low levels just as well as genes that are expressed ubiquitously and/or at high levels.

### 1.3.1 cDNA based gene identification

To identify a gene based on cDNA can be done following three different approaches; screening of cDNA libraries, cDNA selection and transcript sequencing. The choice of approach mainly depends on the goal.

**Screening cDNA libraries** Screening of cDNA libraries is used to identify a specific gene. The screening is usually done with a specific probe, for instance a genomic fragment deleted in patients or evolutionary conserved. The technique is simple, as it requires only hybridisation of a probe to cDNA filters. Many (gridded) cDNA libraries are available and these have been generated from a range of different tissues and developmental stages. Alternatively, one may

generate a new cDNA library from mRNA of any desired tissue. A complementary approach can be used to identify genes from a more complex source. This involves the hybridisation of radiolabeled cDNAs (from oligo(dT) primed RNA) to arrays of genomic clones to identify the clones that contain genes (147) and use those for further analysis.

**cDNA selection** cDNA selection is the more common approach to isolate transcripts from a large region and is often used to generate a transcript map from a contig. The first two articles on cDNA selection were published simultaneously by Lovett *et al.* (148) and Parimoo *et al.* (149) in 1991. Several variations have been published since (150,151) but all are based on hybridisation of cDNA to immobilised DNA, elution, amplification and subsequent rounds of hybridisation to enrich for specifically binding cDNA (Figure 8). The genomic target DNA can be derived from YAC, P1, BAC and cosmid clones. Clones propagated in bacteria have the
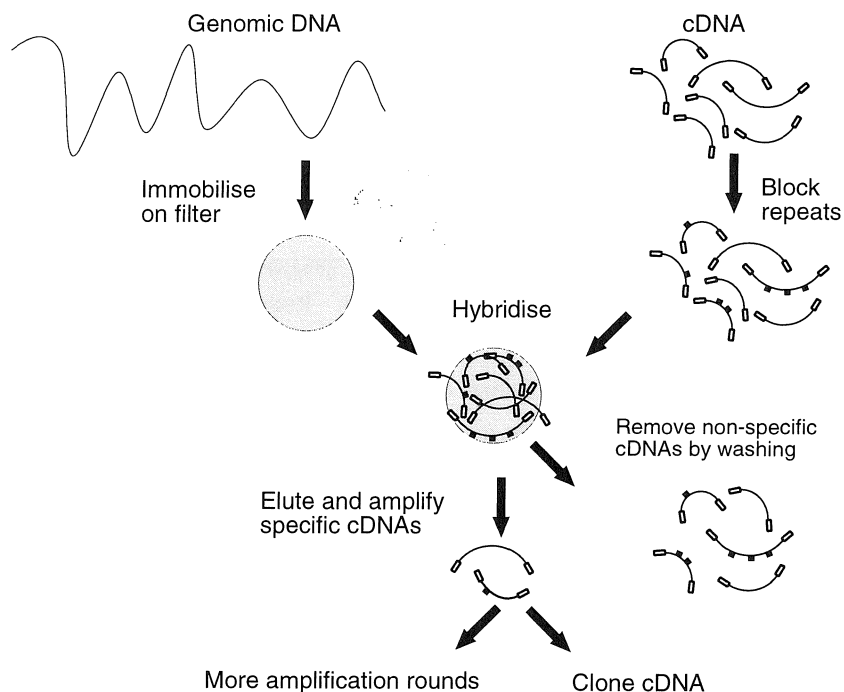


Figure 8: cDNA selection

31

advantage of generating less background than YAC clones. cDNA selection has led to the identification of a variety of novel genes amongst which the disease genes for glycerol kinase deficiency (GKD) (150), hereditary breast and ovarian cancer (BRCA1) (152) and Wiskott-Aldrich syndrome (WAS) (153).

The disadvantage of cDNA selection is that during the selection not only genuine transcripts but also pseudo-genes and homologous genes will be isolated that are located in a different region (usually on another chromosome). On the other hand, these 'artefacts' can be used to specifically isolate members of a gene family (or the 'parental' gene to a pseudo-gene). cDNA selection is further limited by the abundance of a transcript in a cDNA library, transcripts that are present at less than 0.01% are unlikely to be selected for (154).

**Transcript sequencing** Sequencing random transcripts is not an approach to isolate genes in a specific region but an approach to isolate all genes present in the genome and one of the major goals of the Human Genome Project (155). From both the 5' and 3' ends of each transcript a sequence (200-400 bp) is generated, called an expressed sequence tag (EST). All ESTs are deposited in a specific database; dbEST, and can thus be screened using sequences from the region of interest. ESTs that have already been assembled into contigs are present in a separate database called Unigene. These *in silico* cloned genes of course need to be verified as to whether they are derived from the region of interest (and not a homologue on another chromosome) and whether they are not constructed from two separate genes that happen to share a domain and are thus 'software-merged' into an overlapping transcript. Recently, many genes have been identified by *in silico* cloning, i.e. defined by comparative software analysis, based on homology to a gene in another species, like the human thymic shared Ag-1/stem cell Ag-2 gene (TSA-1/SCA-2) that was identified based on the mouse homologue (156), and two human peroxisome biogenesis disorder genes (PXR1 and PXAAA1) as the yeast PAS8 and PAS5 (peroxisome assembly genes) homologs (157). At least one group has started to systematically compare all known phenotype-causing genes in one species (*D. melanogaster*) to human ESTs, in order to define *in silico* all homologues and thus to identify potential candidate diseases for these genes based on their genomic localisation and a potential correspondence of association between phenotypes of -in this case- *Drosophila* and human (158).

### 1.3.2 Genomic DNA based gene identification

The four DNA based techniques that can be used to identify genes are evolutionary conservation, isolation of CpG islands, exon trapping and genomic sequencing.

**Evolutionary conservation**  Functionally important regions in the genome (for instance exons and regulatory sequences) are conserved through evolution. Thus evolutionary conserved sequences are likely to be an element of a gene. Evolutionary conservation is detected by hybridisation of DNA of different species to one another.

Analysis of evolutionary conservation is frequently used  to test the presence of a gene by hybridisation to so-called 'zoo-blots'. Zoo-blots are blots containing DNA from a range of species, typically DNA of mammals (for instance human, ape, bovine, rodent), birds, fish, etc. Hybridisation of a genomic fragment to such a blot gives an indication of the extent to which the fragment is conserved. Although a gene like the DMD-gene was discovered using this approach, it is a laborious method and is less suitable for analysis of large regions.

A suitable approach for larger scale analyses is the comparison with only one other species using a protocol similar to cDNA selection. Several rounds of hybridisation and amplification of genomic DNA from another species to immobilised or biotinylated genomic DNA of the region of interest will enrich for the conserved sequences (159). Unlike the zoo-blot method this does not only give an indication of conservation but also provides the homologous region as an actual clone for further analysis.

**Isolation of CpG islands**  A CpG island is a relative short stretch of a G+C rich region (up to 2 kb) in which the frequency of (unmethylated) CpG nucleotides is significantly higher than elsewhere in genomic DNA. About 60% of human genes are associated with CpG islands. They are typically located at the 5'-ends of genes in, or close to the promoter region and often include the first exon of a gene. CpG islands can be identified by restriction enzymes that recognise stretches of C and G nucleotides and at least one CpG (these restriction enzymes are known as rare cutter enzymes). A CpG island contains a cluster of these restriction sites while normally in genomic DNA these restriction sites are widely spaced (10s or 100s of kb apart). CpG islands are also called HTF islands because the enzyme which revealed these sequences, *Hpa*II, produced *Hpa*II tiny fragments (HTF) (160). There are an estimated 45,000 CpG islands present in the human genome (161).

Isolation of CpG islands is applicable to both large and small regions and different techniques can be used. The easiest involves the subcloning of any source DNA using one rare cutter enzyme and one frequent cutter enzyme and ligating these in a plasmid vector. In this way a transcription map has been succesfully generated from the Huntington disease gene region (162) where 24 out of 42 clones contained putative exons and three novel genes were isolated. A slightly different method involves digestion of the source DNA with a rare cutter enzyme and ligating linkers to the digested DNA. PCR using one primer directed at the linker and one primer directed at *Alu*-repeats will allow the amplification of the CpG islands (163).

A second, more elaborate, approach involves denaturing gradient gel electrophoresis (DGGE) and is based on the difference in melting temperature between regions with a normal and a high G+C content. In a denaturing gradient gel, fragments that are G+C rich will melt later than G+C poor fragments and will therefore have a higher electrophoretic mobility. These faster fragments can than be isolated from the gel and cloned. The source DNA needs to be digested using several enzymes (in this case: *Mse*I, *Tsp*509I, *Nla*III and *Bfa*I) to provide fragments of appropriate size prior to DGGE (164).

A third approach involves the digestion of the source DNA with a frequent cutter that leaves CpG islands in tact (*Mse*I) after which the CpG islands are isolated using a column that specifically binds methylated DNA (165).

**Exon trapping** *In vivo* identification of splice acceptor and splice donor sites, or 'exon trapping' as it was first described by Auch and Reth in 1990 (166), has been used for both small and large scale gene identification. Genomic fragments are cloned into a vector containing an exon trap cassette (a gene preceded by a strong promoter and with a multiple cloning site introduced in one of its introns) and subsequently transfected into a cell line. Transcription of the exon trap cassette gene will incorporate exons present in the genomic fragment, which will then (in principle) be included in the subsequent splicing (Figure 9). Reverse transcription (RT) and PCR on the RNA isolated from the cell-line will reveal the trapped exons which can be used for further analysis. Depending on the vector chosen and RT-PCR protocol applied, one can isolate internal, 3'-terminal or 5'-terminal exons. Internal exons can be trapped using several vectors (166-171). 3'-terminal exons can be trapped using pTAG4 (172). One system allows both internal and 3'-exon trapping; pETV-SD2 (173,174).

The major disadvantage of exon trapping is that it is sensitive to artefacts: spliced

products, involving cryptic splice sites and sequences with fortuitous homology to splice sites, as well as non-specific polyT/polyT primed products (in 3'-terminal exon trapping). The exon trap vectors mentioned above can only contain plasmid size inserts, thereby allowing at best one or only a few exons to be trapped in one product. Moreover, the loss of the genomic context through the small insert size gives rise to the isolation of sequences which are recognised as exons although in nature they are intronic or never even transcribed. Furthermore, the order in which the trapped exons are present in the genome is lost, which makes further analysis tedious. Since an average internal exon has a length of 137 bp (175) the products tend to be small and not especially suitable for screening cDNA libraries or databases.

To overcome these problems, two vectors have recently been developed that can contain larger inserts and will thus allow the simultaneous trapping of multiple exons, leaving the order intact. The exon trap products generated with these vectors make further analysis easier. These vectors are the sCOGH-vectors (176) allowing the isolation of internal, 3'-terminal and 5'-terminal exons and pTAG5 (177) suitable for 3'-terminal exon trapping. The most well known gene that has been identified using exon trapping is IT15, the Huntington disease gene (178).
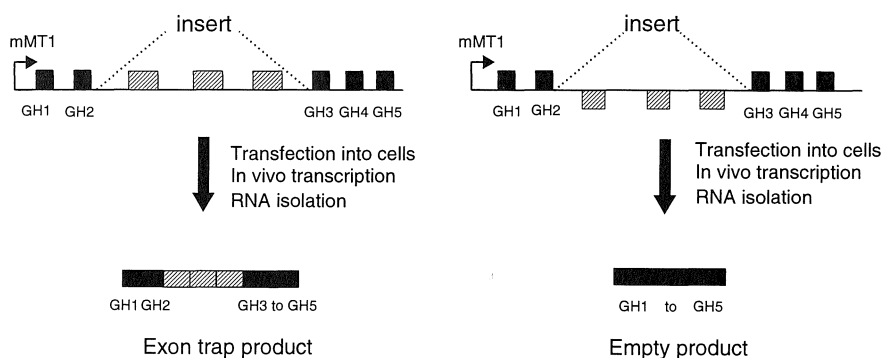


Figure 9. Exon trapping using sCOGH2.

Inserts are cloned in a multiple cloning site in intron 2 of the human growth hormone gene (GH). After transfection of DNA from the clone into a cell line, *in vivo* transcription of the GH gene will incorporate exons from the insert present in the same orientation as the growth hormone gene into the exon trap product. When no exons (or exons in the wrong orientation) are present in the insert an empty product will result. mMT1= mouse metallothioneine gene promoter. GH1-5= human growth hormone exons.

35

**Genomic sequencing** The most detailed information of any region is obtained by sequencing it completely. First one can perform database searches to find homologies with known genes or ESTs and in addition one can use computer programs to predict the location of genes. Large scale genomic sequencing has been undertaken to analyse large regions of DNA that are covered with contigs (179) and ultimately the complete human genome will be sequenced as part of the Human Genome Project.

Before searching databases with long genomic sequences, repeat masking is essential to prevent the analysis of large series of repetitive sequences. Subsequent database searches can involve comparison with all known nucleotide or protein sequences, or subsets that are for instance species-specific, contain only functional motifs, or contain only new sequences (180). A selection of these database search programs and their application is presented in Table 2. A range of programs is available to further analyse the output of these searches, many of which have been adapted for specific projects. Most programs are accessible through email or via the World Wide Web (WWW) (181).

Further analysis of the sequence can involve a number of computer programs. Programs predicting exons, open reading frames (ORFs), promotors, and assembling potential genes are called gene structure prediction programs. An overview of available gene structure prediction programs is given in Table 3.

Table 2: Database search programs

[a] = sequences are translated into 6 reading frames before searching.

[b] = Blitz is also known as SSEARCH or as the Smith-Waterman method.

[c] = Nucleotide sequence databases: Genbank, dbEST, Unigene, EMBL, DBBJ, HTGS, dbSTS, or a locally generated database.

[d] = Protein sequence databases: PIR, SWISS-Prot, GenPept, PDB.

[e] = Pattern databases: EC pattern, PIMA, PROBMIN, BLOCKS, PRINTS, PIR-ALN, FSSP, PROSITE, PRODOM, Sbase.

[f] = Many options are available to search species-specific sequences or only new entries.

| Function | Program | Input sequence | Database sequence | Database use/Notes |
|----------|---------|----------------|-------------------|---------------------|
| repeat masking | XBLAST | nucleotide | nucleotide | REPBASE database |
| | REPEATMASKER | nucleotide | nucleotide | REPEATMASKER database |
| general database search[f] | FASTA | nucleotide | nucleotide | [c], improvement of FASTP |
| | BLASTN | nucleotide | nucleotide | [c] |
| | BLASTX | translated nucleotide[a] | protein | [d], post-processing: BEAUTY |
| | BLASTP | protein | protein | [d], post-processing: BEAUTY |
| | TBLASTN | protein | translated nucleotide[a] | [c] |
| | TBLASTX | nucleotide[a] | translated nucleotide[a] | [c] |
| | Blitz[b] | protein | protein | [d] |
| | Automat | nucleotide/protein | nucleotide/protein | [c] |
| | Staden | nucleotide/protein | nucleotide/protein | [c,d] |
| search new entries in databases | XREFdb | nucleotide | nucleotide[c] | monthly automatic search |
| | FastAlert | nucleotide | nucleotide[c] | regular automatic search |
| search pattern database for protein motifs | FASTA-PAT | protein | protein motif | [e] |
| | FASTASWAP | protein | protein motif | [e] |
| | PROTOMAT | protein/nucleotide[a] | protein motif | BLOCKS & PROSITE database |
| | ProfileScan | protein | protein motif | [e] |
| | MOTIFIND | protein | protein motif | [e] |
| | MacPattern | protein | protein motif | PROSITE database |
| generate 2D/3D model of protein structure | Swiss-Model | protein | proteins of known structure | ExPDB,ExNRL-3D databases post-processing: ProMod |
| | ProteinPredict | protein | protein | SWISS-Prot database |

| Program | year | training set: species | genes | spl. | predictions: ORF | gene | dbs[g] | sequence length: max (shown) | notes |
|---|---|---|---|---|---|---|---|---|---|
| Testcode | 1982 | diverse | 570[d] | - | + | - | - | n.s. | |
| GeneModeler | 1990 | C. elegans | 5 | + | + | + | - | no (50 kb) | integrated approach[a,b] |
| Gelfand method | 1990 | mammalian | 9 | + | + | + | - | n.s. (6.5 kb) | |
| NetGene | 1991 | human | 95 | + | + | - | - | n.s. | |
| GRAIL I | 1991 | human | 18 | - | + | - | - | 100 kb | better on long (>100 bp) exons |
| GeneID | 1992 | vertebrate | 169 | + | + | + | - | 20 kb | integrated approach[a,b,f] |
| SORFIND | 1992 | human | 116 | + | + | - | - | 32 kb | predicts internal exons only[c,f] |
| Geneparser | 1993 | human | 56 | + | + | - | - | no | predicts internal exons only[c,f] |
| Genviewer | 1993 | vertebrate | | | | | | | |
| GeneMark | 1993 | prokaryote | | - | + | - | - | | |
| SITEVIDEO | 1993 | human | n.s. | + | - | - | - | n.s. (22 kb) | |
| GREAT | 1993 | vertebrate | | | | | | | |
| GenLang | 1994 | diverse | 32 | + | + | + | - | n.s. (20 kb) | integrated approach[a,b,f] |
| GRAIL II | 1994 | human | n.s. | + | + | - | - | 100 kb | output used in GAP3[f] |
| GAP3 | 1994 | human | n.s. | - | + | + | - | | uses output of GRAILII |
| FGENEH | 1994 | human | 461 | + | + | + | - | n.s. | output is assembled gene only[b,f] |
| Xpound | 1994 | human | | + | + | - | - | | [f] |
| Geneparser 3 | 1995 | human | 59 | + | + | + | + | no | integrated approach[a,b,f] |
| PromoterScan | 1995 | primate | 167 | - | - | - | - | n.s. | predicts only polymeraseII promoters |
| GeneID+ | 1996 | vertebrate | | + | + | + | + | 8 kb[e] | [f] |

Table 3: Gene structure prediction programs

[a] = Integrated approach includes analysis of initiation signal, stopcodon, polyA signal (AATAAA) and promoter (TATA box).

[b] = This approach is not especially suitable when more than one gene is present, when overlapping genes are present or to detect alternative splicing.

[c] = Approach especially suitable for partial sequences or when more than one gene is present in the sequence.

[d] = Short sequences, not genes (321 coding, 249 non-coding).

[e] = May have improved.

[f] = Used in comparison by Burset *et al.*(182) see text.

[g] = Database searches are used to compare ORF with existing proteins, output not shown.

 

 

Extensive comparison of a subset of the gene structure prediction programs (indicated in Table 3) by Burset and Guigó (182), showed that 33 - 51% of exons are predicted perfect (with exact splice boundaries), 22 - 36% of exons are totally missed, 13 - 27% of predicted exons are completely wrong. A slightly different evaluation method looks at overlap between actual and predicted exons, this ranges from 62 - 71%. Programs that also predict an amino acid sequence, generate proteins that show 52 - 62% similarity to the actual protein sequence.

The accuracies of the predictions were lower using only new sequences than when using sequences that were partly available in the databases at the time the programs were trained. Furthermore, the programs seem to perform worse on long stretches than on short stretches which will be a problem when large-scale sequence analysis is needed (183).

It is important to realise that the programs have different, complementary strengths and the choice of programs depends on emphasis (sensitivity or specificity) and desired features. However, these programs generate predictions which must be verified as no prediction program so far is capable of predicting all exons of a gene accurately, and all have a significant false positive rate. The programs develop rapidly however, and are likely to improve constantly, but will always stay one or more steps behind of the evolving needs of genomic research.

**1.4 Testing candidate genes**

Once a gene has been identified it can be tested as a candidate gene for a specific disorder. The techniques that are available to test a candidate gene are once again complementary, no single technique can identify all mutations in a given gene. These techniques can be either DNA based or RNA based. Once it has been proven that mutations in the candidate gene cause the disease, the same techniques can be used for further mutation analysis. The choice of technique for mutation analysis depends on the mutation spectrum of the gene, the size of the gene and the number and size of the exons. The most common techniques are described below, other less frequently used techniques have been described by R.G.H. Cotton, 1997 (184).

**Hybridisation** Hybridisation of a gene - or, in an earlier stage of the identification of the gene, of a genomic fragment- to Southern blots with digested DNA of patients, will reveal genetic rearrangements caused by deletions, duplications, inversions, or nucleotide changes that alter restriction sites. Aberrant fragments involving larger genomic regions can best be identified using pulsed-field gel electrophoresis (185).

**Sequencing** Sequencing is sometimes chosen, especially for smaller genes, in an initial stage -when no information is available about the type of mutations to expect- to identify the first mutations. To reduce the work load, sequencing of a candidate gene in a few patients is usually performed on RNA-derived material but can also be performed on genomic DNA when a small gene is involved.

**Single Strand Conformation Analysis (SSCA)** SSCA is based on the conformational change of denatured DNA induced by a mutation. In short, DNA fragments (typically 200-400 bp) are amplified using specific primers in a PCR and these fragments are denatured and run on a polyacrylamide gel. Missing or smaller products may indicate a deletion of (part of) the fragment or a mutation at one of the primer sites. The mobility changes are caused by single nucleotide changes or small deletions. Since analysis of the full length of the genomic DNA of a given gene is laborious, SSCA primers are usually designed in intron sequences flanking exons to amplify specifically the latter. A major disadvantage of this is that many mutations in introns that alter splice sites will be missed (186). Also, rearrangements will be missed in which the exons are

present, but no longer in the right place or orientation. This was found for the factor VIII gene, which is inverted in a large fraction (> 40%) of severe hemophilia A patients (187).

**Denaturing Gradient Gel Electrophoresis (DGGE)** In short, (double-stranded) PCR products are separated on a polyacrylamide gel with an increasing temperature or an increasing concentration of denaturant (urea/formamide). When the temperature or concentration of denaturant in the gel has been reached at which the low-temperature melting domain will become single-stranded, the electrophoretic mobility of the product is greatly reduced. The precise conditions in which this happens and thus the precise position in the gel are highly dependent on the specific nucleotide sequence. Any change in this by a mutation is likely to cause an altered migration. Due to the PCR-process, when generating the fragments in heterozygous samples, besides the normal and mutant strands also heteroduplexes are generated. These are even less stable and their gel position tends to differ even between the heteroduplexes with normal and mutant sequence in the two different strands. However, only mutations in the low-temperature melting domain of a PCR product can be detected. In order to analyse the original high-temperature domain as a low-temperature domain, a new high-temperature domain is created by addition of a 'GC-clamp' (by a GC-rich tail on one of the primers) that will alter the melting characteristics of the product (194). Single base mutations and small deletions will cause an increase or decrease in the melting temperature that can be detected as a product that runs higher or lower in the gel than the wild type product.

**RT-PCR and protein truncation test** The Protein Truncation Test (PTT) (188) is a technique based on the analysis of the encoded protein. cDNA is generated by reverse transcription (RT) of patient-derived RNA. The cDNA is amplified using PCR to generate stretches of 1-2 kb. One of the primers used for the PCR contains a transcription initiation signal and a T7 promotor (189) to facilitate transcription and the subsequent translation (Figure 10). The first check for aberrations is done by running the PCR products on an agarose gel.

*In vitro* translated products, analysed on SDS/PAGE gel, reveal mutations that directly or indirectly (through a frame-shift) cause a premature translation termination. This technique has been successfully used to identify a large number of mutations in hereditary cancers, e.g. the gene for hereditary breast and ovarian cancer (BRCA1) (190) and adenomatous polyposis coli (APC) (191). Also, PTT has allowed to find the first protein truncating mutation in CBP in

Rubinstein-Taybi syndrome (RTS) patients and thus to unambiguously implicate CBP in causing RTS (192).

One of the requirements for the application of this technique is that the gene is transcribed in the tissue that is used for RNA isolation (usually lymphocytes). However, illegitimate transcription (193) of many genes normally not expressed in lymphocytes often provide enough transcripts to produce a PCR product.
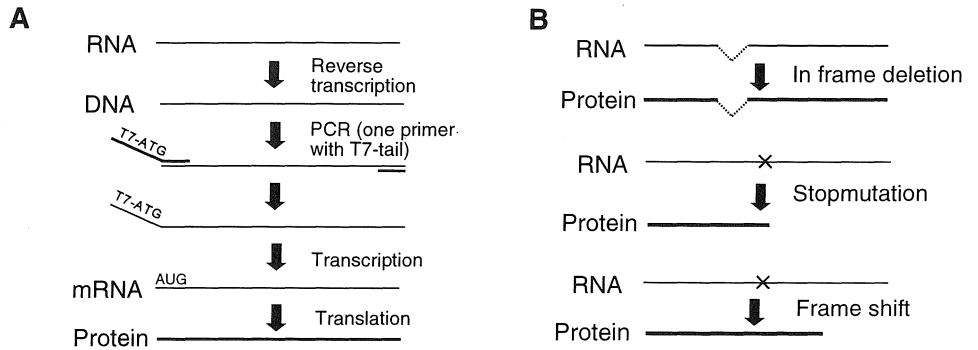


Figure 10. PTT principle.

A. RNA is reverse transcribed into cDNA. The cDNA is then amplified using a primer containing a T7-tail and a translation initiation signal. *In vitro* transcription and translation results in a protein that can be analysed on a SDS/PAGE gel. B. In frame deletions in the RNA will result in a decreased protein size, the decrease proportional to the size of the deletion. Stop mutations will result in a smaller protein, the size of the protein depending on the position of the stop mutation. Frame shift mutations usually result in a premature stop downstream thus also generating a smaller protein.

## 1.5 DISCUSSION

Since the beginning of the 90s, major improvements have been made to existing gene identification techniques but no revolutionary new techniques have been developed to identify genes in the human genome. Most groups are using a combination of the available techniques, usually including either exon trapping or cDNA selection and gratefully make use of the rapidly expanding EST database, which is of enormous assistance in the isolation of disease genes (195). New genes are found and published every week. In 1996, in the monthly journal Nature Genetics alone, up to 12 new genes were published per month (in total 105), on average 6-7 monthly.

The Human Genome Project (HGP), started in February 1988 with the National Research Council (NRC) report: 'Mapping and Sequencing of the Human Genome' (155). The three main objectives for the years 1990-2005 of the HGP were: 1) to improve the research infrastructure of human genetics, 2) to help establish DNA sequence as the primary interface between knowledge of human biology and knowledge of the biology of modelorganisms 3) to launch an open-ended effort to improve the analytical biochemistry of DNA.

To reach these goals the HGP aimed to develop genetic and physical maps of the human genome and to sequence the human genome by the year 2005. In addition genetic and physical maps of mice, worm, flies and yeast would be developed as these are valuable model organisms for studying development, diseases and treatments. In the early years there was considerable skepticism about whether the available technology would be adequate. Technical advances however, especially in PCR, FISH and YAC cloning, have been so great that the speed of genetic and physical mapping has rapidly gone up (196).

Since the start of the HGP, many human genome maps have been published based on genetic or physical data or an integration of both (127,137,197-204). Systematic sequencing of contigs has already generated many megabases of human sequence (44 Mb at 28/2/97, 170 Mb predicted at 28/2/98 (205)) and it will not be long before the complete sequence of the human genome will be available. That is not however, the end of the human genome analysis, but merely a step along the way toward understanding of the genes and their functions.

In conclusion, the time when the genome sequence of many organisms will be publicly available to boost biological research, is not far away. In a few years therefore, in what tends to be called the 'post-genome era', the focus of many groups will transcend from building contigs and transcript maps to the (large scale) functional analysis of the genes that have been identified

43

*in silico*. Developmental expression patterns, differentially spliced products, protein folding and processing, protein-protein interactions, biochemical pathways and regulatory networks need to be analysed.

The impact of these developments is already fundamentally altering positional cloning. Genome sequence based approaches will more and more replace parallel and serial transcript mapping on constructed contigs. Thus, the elucidation of the gene causing RS seems not far off. While two novel RS candidate genes have recently been cloned 'the old way', meanwhile, using a clone contig provided by the Retinoschisis Consortium, the candidate region for RS has recently been rapidly sequenced by the Sanger Centre (Hinxton, UK) and made available on the internet, thus providing the kick-off for using novel *in silico* approaches. So it seems only a matter of time to prove which of the genes cloned or predicted in the region is mutated in RS patients.

The candidate region for KFSD has not been sequenced yet. Several genes in the region have been cloned that have not even been tested yet as candidates. The identification of the KFSD gene will take longer primarily because it will be harder to prove which gene causes the disorder since only so few families are available .

Once found, the functional analysis of the RS and KFSD genes will lead to an understanding of the mechanisms underlying these diseases. As RS only involves the eye, there is good hope for a potential therapy once developed, as the retina may be more easily accessible for delivery of normal genes through viral vectors (206). KFSD is mainly a skin disorder, which also makes it more easily accessible than the organs affected in several other severe genetic diseases, like skeletal muscle in muscular dystrophies or the immune system in immunodeficiencies. However, much biological research is still required and it is even possible that the outcome of the genome-based research will allow the development of pharmacological rather than genetic means of intervention.

## 1.6 OUTLINE OF THE THESIS

In this study we aimed at the positional cloning of disease genes in Xp22.1-p22.2. To this end we constructed a YAC contig covering this region including the markers from DXS414 to DXS451. The contig enabled us to order new markers in the region and to refine the localisation of the genes for X-linked juvenile retinoschisis (RS) and for keratosis follicularis spinulosa decalvans (KFSD) (Chapter 2.1). One of the key YACs in the RS candidate region was used in YAC fragmentation experiments to generate a panel of fragmented YACs used for 'binning' clones and to construct a 2.5 Mb restriction map of this region (Chapter 2.2).

To identify candidate genes for the diseases localised in this region we have applied exon trapping. Initially, exon trapping was performed with cosmids subcloned in the vector pSPL3, isolating -among others- exons from a known gene: the liver $\alpha$-subunit of phosphorylase kinase (PHKA2). To analyse larger genomic regions more efficiently, we have designed a new exon trap vector (sCOGH2) which allows the direct analysis of cosmid-size clones (Chapter 3). We have subcloned key YACs from the region into the sCOGH vectors and used these in exon trap experiments. We have isolated several novel transcripts from the region, of which one was analysed in detail (Chapter 4.1).

Two genes, *PPEF* and *Txp3*, which have been isolated by members of the Retinoschisis Consortium were tested as candidate genes for RS (Chapter 5) but no mutations could be detected thus far.

## 1.7 REFERENCES

1    Nelson, D.L., Ballabio, A., Cremers, F., Monaco, A.P., and Schlessinger, D. Report of the sixth international workshop on X chromosome mapping 1995. *Cytogenet.Cell Genet.* 71:308-342 (1995).

2    Affara, N., Bishop, C., Brown, W., Cooke, H., Davey, P., Ellis, N., Graves, J.M., Jones, M., Mitchell, M., Rappold, G., Tyler-Smith, C., Yen, P., and Lau, Y.P. Report of the second international workshop on Y chromosome mapping 1995. *Cytogenet.Cell Genet.* 73:33-76 (1996).

3    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. A gene from the human sex-determining region encodes a protein with homology to a conserved DNA-binding motif. *Nature* 346:240-244 (1990).

4    Reijo, R., Lee, T.Y., Salo, P., Alagappan, R., Brown, L.G., Rosenberg, M., Rozen, S., Jaffe, T., Straus, D., Hovatta, O., De la Chapelle, A., Silber, S., and Page, D.C. Diverse spermatogenic defects in humans caused by Y chromosome deletions encompassing a novel RNA-binding protein gene. *Nat.Genet.* 10:383-393 (1995).

5    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. 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 (1996).

6    Henegariu, O., Hirschmann, P., Kilian, K., Kirsch, S., Lengauer, C., Maiwald, R., Mielke, K., and Vogt, P. Rapid screening of the Y chromosome in idiopathic sterile men, diagnostic for deletions in AZF, a genetic Y factor expressed during spermatogenesis. *Andrologia* 26:97-106 (1994).

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

8    Chandley, A.C. and Cooke, H.J. Human male infertility--Y-linked genes and spermatogenesis. *Hum.Mol.Genet.* 3:1449-1452 (1994).

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

10   Renauld, J.C., Duez, C., Kermouni, A., Houssiau, F., Uyttenhove, C., Van Roost, E., and Van Snick, J. Expression cloning of the murine and human interleukin 9 receptor cDNAs. *Proc.Natl.Acad.Sci.USA* 89:5690-5694 (1992).

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

12   Speed, R.M. and Chandley, A.C. Prophase of meisosis in human spermatocytes analysed by EM microspreading in infertile men and their controls and comparisons with human oocytes. *Hum.Genet.* 84:547(1990).

13   Schmitt, K., Lazzeroni, L.C., Foote, S., Vollrath, D., Fisher, E.M., Goriada, T.M., Lange, K., Page, D.C., and Arnheim, N. Multipoint linkage map of the human pseudoautosomal region, based on single-sperm typing: do double crossovers occur during male meisosis. *Am.J.Hum.Genet.* 55:423-430 (1994).

14  Li, L. and Hamer, D.H. Recombination and allelic association in the Xq/Yq homology region. *Hum.Mol.Genet.* 4:2013-2016 (1995).

15  Rouyer, F., Simmler, M.C., Johnsson, C., Vergnaud, G., Cooke, H.J., and Weissenbach, J. A gradient of sex linkage in the pseudoautosomal region of the human sex chromosomes. *Nature* 319:291-295 (1986).

16  O'Reilly, A.J., Affara, N.A., Simpson, E., Chandler, P., Goulmy, E., and Ferguson-Smith, M.A. A molecular deletion map of the Y chromosome long arm defining X and autosomal homologous regions and the localisation of the HYA locus to the proximal region of the Yq euchromatin. *Hum.Mol.Genet.* 1:379-385 (1992).

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

18  Klink, A., Schiebel, K., Winkelmann, M., Rao, E., Horsthemke, B., Ludecke, H.J., Claussen, U., Scherer, G., and Rappold, G. 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 (1995).

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

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

21  Lyon, M.F. Gene action in the X-chromosome of the mouse (mus musculus l.). *Nature* 190:372-373 (1961).

22  Riggs, A.D. and Pfeifer, G.P. X-chromosome inactivation and cell memory. *Trends Genet.* 8:169-174 (1992).

23  Latham, K.E. X chromosome imprinting and inactivation in the early mammalian embryo. *Trends Genet.* 12:134-138 (1996).

24  Wu, J., Ellison, J., Salido, E., Yen, P., Mohandas, T., and Shapiro, L.J. Isolation and characterization of XE169, a novel human gene that escapes X-inactivation. *Hum.Mol.Genet.* 3:153-160 (1994).

25  Brown, C.J., Miller, A.P., Carrel, L., Rupert, J.L., Davies, K.E., and Willard, H.F. The DXS423E gene in Xp11.21 escapes X chromosome inactivation. *Hum.Mol.Genet.* 4:251-255 (1995).

26  Coleman, M.P., Ambrose, H.J., Carrel, L., Nemeth, A.H., Willard, H.F., and Davies, K.E. A novel gene, DXS8237E, lies within 20 kb upstream of UBE1 in Xp11.23 and has a different X inactivation status. *Genomics* 31:135-138 (1996).

27  Brown, C.J., Lafreniere, R.G., Powers, V.E., Sebastio, G., Ballabio, A., Pettigrew, A.L., Ledbetter, D.H., Levy, E., Craig, I.W., and Willard, H.F. Localization of the X inactivation centre on the human X chromosome in Xq13. *Nature* 349:82-84 (1991).

28  Lyon, M.F. X-chromosome inactivation. Pinpointing the centre. *Nature.* 379:116-117 (1996).

29  Willard, H.F. X chromosome inactivation, XIST, and pursuit of the X-inactivation center. *Cell* 86:5-7 (1996).

30  Marahrens, Y., Panning, B., Dausman, J., Strauss, W., and Jaenisch, R. Xist-deficient mice are defective in dosage compensation but not spermatogenesis. *Genes & Dev.* 11:156-166 (1997).

31  Lyon, M.F. The X inactivation centre and X chromosome imprinting. *Eur.J.Hum.Genet.* 2:255-261 (1994).

32  Brown, C.J., Ballabio, A., Rupert, J.L., Lafreniere, R.G., Grompe, M., Tonlorenzi, R., and Willard, H.F. A gene from the region of the human X inactivation centre is expressed exclusively from the inactive X chromosome. *Nature* 349:38-44 (1991).

33  Harris, A., Collins, J., Vetrie, D., Cole, C., and Bobrow, M. X inactivation as a mechanism of selection against lethal alleles: further investigation of incontinentia pigmenti and X linked lymphoproliferative disease. *J.Med.Genet.* 29:608-614 (1992).

34  Mattei, M.G., Mattei, J.F., Ayme, S., and Giraud, F. X-autosome translocations: cytogenetic characteristics and their consequences. *Hum Genet* 61:295-309 (1982).

35  Bodrug, S.E., Holden, J.J., Ray, P.N., and Worton, R.G. Molecular analysis of X-autosome translocations in females with Duchenne muscular dystrophy. *EMBO J.* 10:3931-3939 (1991).

36  Chery, M., Biancalana, V., Philippe, C., Malpuech, G., Carla, H., Gilgenkrantz, S., Mandel, J.-L., and Hanauer, A. Hypomagnesemia with secondary hypocalcemia in a female with balanced X;9 translocation: Mapping of the Xp22 chromosome breakpoint. *Hum.Genet.* 93:587-591 (1994).

37  Clarke, J.T.R., Wilson, P.J., Morris, C.P., Hopwood, J.J., Richards, R.I., Sutherland, G.R., and Ray, P.N. Characterization of a deletion at Xq27-q28 associated with unbalanced inactivation of the nonmutant X chromosome. *Am.J.Hum.Genet.* 51:316-322 (1992).

38  Morell, V. Rise and fall of the Y chromosome. *Science* 263:171-172 (1994).

39  Wilkins, A.S. Moving up the hierarchy: a hypothesis on the evolution of a genetic sex determination pathway. *Bioessays* 17:71-77 (1995).

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

41  Spurdle, A.B. and Jenkins, T. The Y chromosome as a tool for studying human evolution. *Curr.Opin.Genet.Dev.* 2:487-491 (1992).

42  Bone, J.R. and Kuroda, M.I. Dosage compensation regulatory proteins and the evolution of sex chromosomes in Drosophila. *Genetics* 144:705-713 (1996).

43  Kelley, R.L. and Kuroda, M.I. Equality for X chromosomes. *Science* 270:1607-1610 (1995).

44  McKay, L., Watson, J.M., and Marshall Graves, J.A. Mapping human X-linked genes in the phalangerid marsupial Trichosurus vulpecula. *Genomics* 14:302-308 (1992).

45  Watson, J.M., Spencer, J.A., Riggs, A.D., and Marshall Graves, J.A. Sex chromosome evolution: Platypus gene mapping suggests that part of the human X chromosome was originally autosomal. *Proc.Natl.Acad.Sci.USA* 88:11256-11260 (1991).

46  Spencer, J.A., Sinclair, A.H., Watson, J.M., and Graves, J.A. Genes on the short arm of the human X chromosome are not shared with the marsupial X. *Genomics* 11:339-345 (1991).

47  Ballabio, A., Bardoni, B., Carrozzo, R., Andria.G., Bick, D., Campbell, L., Hamel, B., Ferguson-Smith, M.A., Gimelli, G., Fraccaro, M., Maraschio, P., Zuffardi, O., Guioli, S., and Camerino, G. Contiguous gene syndromes due to deletions in the distal short arm of the human X chromosome. *Proc.Natl.Acad.Sci.USA* 86:10001-10005 (1989).

48

48    Schaefer, L., Ferrero, G.B., Grillo, A., Bassi, M.T., Roth, E.J., Wapenaar, M.C., Van Ommen, G.J.B., Mohandas, T.K., Rocchi, M., Zoghbi, H.Y., and Ballabio, A. A high resolution deletion map of human chromosome Xp22. *Nature Genet.* 4:272-279 (1993).

49    Schnur, R.E., Wick, P.A., Sosnoski, D.N., Bick, D., and Nussbaum, R.L. Deletion mapping and a highly reduced radiation hybrid in the Xp22.3-p22.2 region. *Genomics* 15:500-506 (1993).

50    Bertelson, C.J., Pogo, A.O., Chaudhuri, A., Marsh, W.L., Redman, C.M., Banerjee, D., Symmans, W.A., Simon, T., Frey, D., and Kunkel, L.M. Localization of the McLeod locus (XK) within Xp21 by deletion analysis. *Am.J.Hum.Genet.* 42:703-711 (1988).

51    De Saint-Basile, G., Bohler, M.C., Fischer, A., Cartron, J., Dufier, J.L., Griscelli, C., and Orkin, S.H. Xp21 DNA microdeletion in a patient with chronic granulomatous disease, retinitis pigmentosa, and McLeod phenotype. *Hum.Genet.* 80:85-89 (1988).

52    Francke, U., Ochs, H.D., De Martinville, B., Giacalone, J., Lindgren, V., Disteche, C., Pagon, R.A., Hofker, M.H., Van Ommen, G.J., Pearson, P.L., and Wedgwood, R.J. Minor Xp21 chromosome deletion in a male associated with expression of Duchenne muscular dystrophy, chronic granulomatous disease, retinitis pigmentosa, and McLeod syndrome. *Am.J.Hum.Genet.* 37:250-267 (1985).

53    Franco, B., Guioli, S., Pragliola, A., Incerti, B., Bardoni, B., Tonlorenzi, R., Carrozzo, R., Maestrini, E., Pieretti, M., Taillon-Miller, P., Brown, C.J., Willard, H.F., Lawrence, C., Persico, M.G., Camerino, G., and Ballabio, A. A gene deleted in Kallmann's syndrome shares homology with neural cell adhesion and axonal path-finding molecules. *Nature* 353:529-536 (1991).

54    Legouis, R., Hardelin, J.-P., Levilliers, J., Claverie, J.-M., Compain, S., Wunderle, V., Millasseau, P., Le Paslier, D., Cohen, D., Caterina, D., Bougueleret, L., Delemarre-Van de Waal, H., Lutfalla, G., Weissenbach, J., and Petit, C. The candidate gene for the X-linked Kallmann syndrome encodes a protein related to adhesion molecules. *Cell* 67:423-435 (1991).

55    Ballabio, A., Parenti, G., Carrozzo, R., Sebastio, G., Andria, G., Buckle, V., Fraser, N., Craig, I., Rocchi, M., Romeo, G., Jobsis, A.C., and Persico, M.G. Isolation and characterization of a steroid sulfatase cDNA clone: genomic deletions in patients with X-chromosome-linked ichthyosis. *Proc.Natl.Acad.Sci.USA* 84:4519-4523 (1987).

56    Monaco, A.P., Neve, R.L., Coletti-Feener, C., Bertelson, C.J., Kurnit, D.M., and Kunkel, L.M. Isolation of candidate cDNAs for portions of the Duchenne muscular dystrophy gene. *Nature* 323:646-650 (1986).

57    Ho, M., Chelly, J., Carter, N., Danek, A., Crocker, P., and Monaco, A.P. Isolation of the gene for McLeod syndrome that encodes a novel membrane transport protein. *Cell* 77:869-880 (1994).

58    Royer-Pokora, B., Kunkel, L.M., Monaco, A.P., Goff, S.C., Newburger, P.E., Baehner, R.L., Cole, F.S., Curnutte, J.T., and Orkin, S.H. Cloning the gene for an inherited human disorder -chronic granulomatous disease- on the basis of its chromosomal location. *Nature* 322:32-38 (1986).

59    Robinson, D.O., Boyd, Y., Cockburn, D., Collinson, M.N., Craig, I., and Jacobs, P.A. The parental origin of de novo X-utosome translocations in females with Duchenne muscular dystrophy revealed by M27 beta methylation analysis. *Genet.Res.* 56:135-140 (1990).

60    Francis, F., Hennig, S., Korn, B., Reinhardt, R., De Jong, P., Poustka, A., Lehrach, H., Rowe, P.S.N., Goulding, J.N., Summerfield, T., Mountford, R., Read, A.P., Popowska, E., Pronicka, E.,

Davies, K.E., O'Riordan, J.L.H., Econs, M.J., Nesbitt, T., Drezner, M.K., Oudet, C., Pannetier, S., Hanauer, A., Strom, T.M., and Meindl, A. A gene (*PEX*) with homologies to endopeptidases is mutated in patients with X̌-linked hypophosphatemic rickets. *Nature Genet.* 11:130-136 (1995).

61    Van den Berg, I.E.T., Van Beurden, E.A.C.M., Malingré, H.E.M., Ploos Van Amstel, H.K., Poll-The, B.T., Smeitink, J.A.M., Lamers, W.H., and Berger, R. X-linked liver phosphorylase kinase deficiency is associated with mutations in the human liver phosphorylase kinase α subunit. *Am.J.Hum.Genet.* 56:381-387 (1995).

62    Hendrickx, J., Dams, E., Coucke, P., Fernandes, J., and Willems, P.J. X-linked liver glycogenosis type II (XLG II) is caused by mutations in PHKA2, the gene encoding the liver α subunit of phosphorylase kinase. *Hum.Mol.Genet.* 5:649-652 (1996).

63    Trivier, E., De Cesare, D., Jacquot, S., Pannetier, S., Zackai, E., Young, I., Mandel, J.-L., Sassone-Corsi, P., and Hanauer, A. Mutations in the kinase Rsk-2 associated with Coffin-Lowry syndrome. *Nature* 384:567-570 (1996).

64    Lindsay, E.A., Grillo, A., Ferrero, G.B., Roth, E.J., Magenis, E., Grompe, M., Hultén, M., Gould, C., Baldini, A., Zoghbi, H.Y., and Ballabio, A. Microphthalmia with linear skin defects (MLS) syndrome: clinicak, cytogenetic, and molecular characterization. *Am.J.Med.Genet.* 49:229-234 (1994).

65    Naritomi, K., Izumikawa, Y., Nagataki, S., Fukushima, Y., Wakui, K., Niikawa, N., and Hirayama, K. Combined Goltz and Aicardi syndromes in a terminal Xp deletion: are they a contiguous syndrome? *Am.J.Med.Genet.* 43:839-843 (1992).

66    Hendrickx, J., Coucke, P., Dams, E., Lee, P., Odievre, M., Corbeel, L., Fernandes, J.F., and Willems, P.J. Mutations in the phosphorylase kinase gene PHKA2 are responsible for X-linked liver glycogen storage disease. *Hum.Mol.Genet.* 4:77-83 (1995).

67    Heuertz, S., Smahi, A., Wilkie, A.O.M., Le Merrer, M., Maroteaux, P., and Hors-Cayla, M.C. Genetic mapping of Xp22.12-p22.31, with a refined localization for spondyloepiphyseal dysplasia (SEDL). *Hum.Genet.* 96:407-410 (1995).

68    Toutain, A., Ronce, N., Dessay, B., Robb, L., Francannet, C., Le Merrer, M., Briard, M.L., Kaplan, J., and Moraine, C. Nance-Horan syndrome: linkage analysis in 4 families refines localization in Xp22.31-p22.13 region. *Hum.Genet.* 99:256-261 (1997).

69    McGuire, R.E., Sullivan, L.S., Blanton, S.H., Church, M.W., Heckenlively, J.R., and Daiger, S.P. X-linked dominant cone-rod degeneration: linkage mapping of a new locus for Retinitis Pigmentosa (RP15) to Xp22.13-p22.11. *Am.J.Hum.Genet.* 57:87-94 (1995).

70    Del Castillo, I., Villamar, M., Sarduy, M., Romero, L., Herraiz, C., Hernandez, F.J., Rodriguez, M., Borras, I., Montero, A., Bellon, J., Cruz Tapia, M., and Moreno, F. A novel locus for non-syndromic sensorineural deafness (DFN6) maps to chromosome Xp22. *Hum.Mol.Genet.* 5:1383-1387 (1996).

71    Gedeon, A., Kerr, B., Mulley, J., and Turner, G. Pericentromeric genes for non-specific X-linked mental retardation (MRX). *Am.J.Med.Genet.* 51:553-564 (1994).

72    Donnelly, A.J., Andy Choo, K.H., Kozman, H.M., Gedeon, A.K., Danks, D.M., and Mulley, J.C. Regional localisation of a non-specific X-linked mental retardation gene (MRX19) to Xp22. *Am.J.Med.Genet.* 51:581-585 (1994).

73    Hu, L.-J., Blumenfeld-Heyberger, S., Hanauer, A., Weissenbach, J., and Mandel, J.-L. Non-specific

X-linked mental retardation: Linkage analysis in MRX2 and MRX4 families revisited. *Am.J.Med.Genet.* 51:569-574 (1994).

74   Schutz, C.K., Ives, E.J., Chalifoux, M., MacLaren, L., Farrell, S., Robinson, P.D., White, B.N., and Holden, J.J.A. Regional localization of an X-linked mental retardation gene to Xp21.1-Xp22.13 (MRX38). *Am.J.Med.Genet.* 64:89-96. (1996).

75   Van de Vosse, E., Bergen, A.A.B., Meershoek, E.J., Oosterwijk, J.C., Gregory, S., Bakker, B., Weissenbach, J., Coffey, A.J., Van Ommen, G.J.B., and Den Dunnen, J.T. A Xp22.1-p22.2 YAC contig encompassing the disease loci for RS, KFSD, CLS, HYP and RP15; refined localization of RS. *Eur.J.Hum.Genet.* 4:101-104 (1996).

76   Oosterwijk, J.C., Van der Wielen, M.J.R., Van de Vosse, E., Voorhoeve, E., and Bakker, E. Refinement of the localisation of X-linked keratosis follicularis spinulosa decalvans (KFSD) gene in Xp22.1-p22.2. *J.Med.Genet.* 32:736-739 (1995).

77   Haas, J. Über das zusammenvorkommen von Veränderungen der Retina und Chorioidea. *Arch.Augenheilkd.* 37:343-348 (1898).

78   Wilczek, M. Ein fall der netzhautspaltung (Retinoschisis) mit einer öffnung. *Ztschr.f.A.* 85:108-116 (1996).

79   Sorsby, A., Klein, M., Hurndall Gann, J., and Siggins, G. Unusual retinal detachment, possibly sex-linked. *Brit.J.Ophthalmol.* 35:1-10 (1951).

80   Bergen, A.A.B., Ten Brink, J.B., and Van Schooneveld, M.J. Efficient DNA carrier detection in X linked juvenile retinoschisis. *Br.J.Ophthalmol.* 79:683-686 (1995).

81   Lisch, W. Sex-linked Juvenile Retinoschisis. In: Straub, W. (ed) Hereditary Vitroretinal Degenerations. Karger, Basel,New York: 19-32.(1983).

82   Pawar, H., Bingham, E.L., Lunetta, K.L., Segal, M., Richards, J.E., Boehnke, M., and Sieving, P.A. Refined genetic mapping of juvenile X-linked retinoschisis. *Hum.Hered.* 45:206-210 (1995).

83   George, N.D., Yates, J.R., and Moore, A.T. Clinical features in affected males with X-linked retinoschisis. *Arch.Ophthalmol.* 114:274-280 (1996).

84   Sieving, P.A., Bingham, E.L., Roth, M.S., Young, M.R., Boehnke, M., Kuo, C.-Y., and Ginsburg, D. Linkage relationship of X-linked juvenile retinoschisis with Xp22.1-p22.3 probes. *Am.J.Hum.Genet.* 47:616-621 (1990).

85   George, N.D.L., Payne, S.J., Barton, D.E., Moore, A.T., and Yates, J.R.W. Genetic mapping of X-linked Retinoschisis. *Cytogenet.Cell Genet.* 67:354(1994).

86   Forsius, H., Eriksson, A., and Vainio-Mattila, B. Geshlechtsgebundene, erbliche Retinoschisis in zwei Familien in Finnland. *Klin.Mbl.Augenheilk.* 143:806-816 (1963).

87   Condon, G.P., Brownstein, S., Wang, N.-S., Kearns, A.F., and Ewing, C.E. Congenital hereditary (juvenile X-linked) retinoschisis. *Arch.Ophthalmol.* 104:576-583 (1986).

88   Gellert, G., Petersen, J., Krawczak, M., and Zoll, B. Linkage relationship between retinoschisis and four marker loci. *Hum.Genet.* 79:382-384 (1988).

89   Alitalo, T., Forsius, H., Kärnä, J., Frants, R.R., Eriksson, A.W., Wood, S., Kruse, T.A., and De la Chapelle, A. Linkage relationships and gene order around the locus for X-linked Retinoschisis. *Am.J.Hum.Genet.* 43:476-483 (1988).

90   Dahl, N., Goonewardena, P., Chotal, J., Anvret, M., and Petterson, U. DNA linkage analysis of

X-linked retinoschisis. *Hum.Genet.* 78:228-232 (1988).

91   Arden, G.B., Gorin, M.B., Polkinghorne, P.J., Jay, M., and Bird, A.C. Detection of the carrier state of X-linked Retinoschisis. *Am J Ophthalmol* 105:590-595 (1988).

92   Kaplan, J., Pelet, A., Hentati, H., Jeanpierre, M., Briard, M.L., Journel, H., Munnich, A., and Dufier, J.L. Contribution to carrier detection and genetic counselling in X linked retinoschisis. *J.Med.Genet.* 28:383-388 (1991).

93   George, N.D.L., Yates, J.R.W., and Moore, A.T. X linked retinoschisis. *Br.J.Ophthalmol.* 79:697-702 (1995).

94   Manschot, W.A. Pathology of hereditary juvenile retinoschisis. *Arch.Ophthalmol.* 88:131-138 (1972).

95   Yanoff, M., Kertesz Rahn, E., and Zimmerman, L.E. Histopathology of juvenile retinoschisis. *Arch.Ophthalmol.* 88:131-138 (1972).

96   Peachey, N.S., Fishman, G.A., Derlacki, D.J., and Brigell, M.G. Psychophysical and electroretinographic findings in X-linked juvenile retinoschisis. *Arch.Ophthalmol.* 105:513-516 (1987).

97   De Jong, P.T.V.M., Zrenner, E., Van Meel, G.J., Keunen, J.E.E., and Van Norren, D. Mizuo phenomenon in X-linked retinoschisis. *Arch.Ophthalmol.* 109:1104-1108 (1991).

98   Ewing, C.C. and Cullen, A.P. Fluorescein angiography in X-chromosomal maculopathy with retinoschisis (juvenile hereditary retinoschisis). *Can.J.Ophthalmol.* 7:19-28 (1972).

99   Brockhurst, R.J. Photocoagulation in congenital retinoschisis. *Arch.Ophthalmol.* 84:158-165 (1970).

100  Regillo, C.D., Tasman, W.S., and Brown, G.C. Surgical management of complications associated with X-linked retinoschisis. *Arch.Ophthalmol.* 111:1080-1086 (1993).

101  Madjarov, B., Hilton, G.F., Brinton, D.A., and Lee, S.S. A new classification of the retinoschises. *Retina.* 15:282-285 (1995).

102  Wieacker, P., Wienker, T.F., Dallapiccola, B., Bender, K., Davies, K.E., and Ropers, H.H. Linkage relationships between retinoschisis, Xg, and a cloned DNA sequence from the distal short arm of the X chromosome. *Hum.Genet.* 64:143-145 (1983).

103  Alitalo, T., Kärna, J., Forsius, H., and De la Chapelle, A. X-linked retinoschisis is closely linked to DXS41 and DXS16 but not DXS85. *Clin.Genet.* 32:192-195 (1987).

104  Alitalo, T., Kruse, T.A., and De la Chapelle, A. Refined localization of the gene causing X-linked juvenile retinoschisis. *Genomics* 9:505-510 (1991).

105  Oudet, C., Weber, C., Kaplan, J., Segues, B., Croquette, M.-F., Ollagnon Roman, E., and Hanauer, A. Characterisation of a highly polymorphic microsatellite at the DXS207 locus: confirmation of very close linkage to the retinoschisis disease gene. *J.Med.Genet.* 30:300-303 (1992).

106  Bergen, A.A.B., Van Schooneveld, M.J., Orth, U., Bleeker-Wagemakers, E.M., and Gal, A. Multipoint linkage analysis in X-linked juvenile retinoschisis. *Clin.Gen.* 43:113-116 (1993).

107  Biancalana, V., Trivier, E., Weber, C., Weissenbach, J., Rowe, P.S.N., O'Riordan, J.L.H., Partington, M.W., Heyberger, S., Oudet, C., and Hanauer, A. Construction of a high-resolution linkage map for Xp22.1-p22.2 and refinement of the genetic localization of the Coffin-Lowry syndrome gene. *Genomics* 22:617-625 (1994).

108  Shastry, B.S., Hejtmancik, F.J., Margherio, R.T., and Trese, M.T. Linkage mapping of new X-linked

juvenile retinoschisis kindreds using microsatellite markers. *Biochem.Biophys.Res.Commun.* 220:824-827 (1996).

109 Fickett, J.W. The gene identification problem: an overview for developers. *Computers.And.Chemistry.* 20:103-118 (1996).

110 Francis, F., Benham, F., Gee See, C., Fox, M., Ishikawa-Brush, Y., Monaco, A.P., Weiss, B., Rappold, G., Hamvas, R.M.J., and Lehrach, H. Identification of YAC and cosmid clones encompassing the ZFX-POLA region using irradiation hybrid cell lines. *Genomics* 20:75-83 (1994).

111 Montini, E., Rugarli, E.I., Van de Vosse, E., Andolfi, G., Puca, A.A., Den Dunnen, J.T., Ballabio, A., and Franco, B. A human homolog of the Drosophila retinal degeneration C (rdgc) gene encodes a novel serine-threonine phosphatase selectively expressed in sensory neurons of neural crest origin. *Hum.Mol.Genet.* 6:1137-1145 (1997).

112 Franco, B. et al. *In preparation* (1997).

113 Van de Vosse, E., Walpole, S.M., Nicolaou,A., Van der Bent, P., Cahn, A., Vaudin, M., Ross, M.T., Durham, J., Pavitt, R., Wilkinson, J., Grafham, D., Bergen, A.A.B., Van Ommen, G.J.B., Yates, J.R.W., Den Dunnen, J.T., and Trump, D. Characterization of a new developmental gene, *SCML1*, in Xp22. *Submitted* (1997).

114 Franco, B. et al. *In preparation* (1997).

115 Lameris, H.J. Ichthyosis follicularia. *Ned.Tijdschr.Geneeskd.* 2:1524(1905).

116 Siemens, H.W. Keratosis follicularis spinulosa decalvans. *Arch.Derm.Syph.* 151:384-386 (1926).

117 Siemens, H.W. Über einen, in der menschlichen Pathologie noch nicht beobachteten Vererbungsmodus; dominant geschlechtsgebundene Vererbung. *Arch.f.Rass.u.Ges.Biol.* 17:47-61 (1925).

118 Guillet, G., Labouche, F., Cambazard, F., Plantin, P., Gall, Y., Le Jollec, A., Zagnoli, A., and Parent, P. Keratosis follicularis spinulosa decalvans: nosological discussion of Siemens' disease. Apropos of 3 cases. *Pediatrie* 42:437-440 (1987).

119 Kuokkanen, K. Keratosis follicularis spinulosa decalvans in a family from northern Finland. *Acta Derm. Venereol.* 51:146-150 (1971).

120 Herd, R.M. and Benton, E.C. Keratosis follicularis spinulosa decalvans: Report of a new pedigree. *Br.J.Dermatol.* 134:138-142 (1996).

121 Oosterwijk, J.C., Nelen, M., van Zandvoort, P.M., van Osch, L.D.M., Oranje, A.P., Wittebol-Post, D., and Van Oost, B.A. Confirmation of X-linked inheritance and provisional mapping of the keratosis follicularis spinulosa decalvans gene on Xp in a large Dutch pedigree. *Ophthalm.Paediatr.Genet.* 13:27-30 (1991).

122 Oranje, A.P., Van Osch, L.D., and Oosterwijk, J.C. Keratosis Pilaris Atrophicans. One heterogeneous disease or a symptom in different clinical entities? *Arch.Dermatol.* 130:500-502 (1994).

23 Oosterwijk, J.C., Richard, G., Van der Wielen, M.J.R., Van de Vosse, E., Harth, W., Sandkuyl, L.A., Bakker, E. and Van Ommen, G.J.B. Molecular genetic analysis of two families with keratosis follicularis spinulosa decalvans (KFSD): refinement of gene localization and evidence for genetic heterogeneity. *Hum.Genet.* 100:520-524 (1997).

24 Boehnke, M. Limits of resolution of genetic linkage studies: implications for the positional cloning

of human disease genes. *Am.J.Hum.Genet.* 55:379-390 (1994).

125 Brownstein, B.H., Silverman, G.A., Little, R.D., Burke, D.T., Korsmeyer, S.J., Schlessinger, D., and Olson, M.V. Isolation of single-copy human genes from a library of yeast artificial chromosome clones. *Science* 244:1348-1351 (1989).

126 Green, E.D. and Olson, M.V. Systematic screening of yeast artificial-chromosome libraries by use of the polymerase chain reaction. *Proc.Natl.Acad.Sci.USA* 87:1213-1217 (1990).

127 Cohen, D., Chumakov, I., and Weissenbach, J. A first generation physical map of the human genome. *Nature* 366:698-701 (1993).

128 Nelson, D.L., Ledbetter, S.A., Corbo, L., Victoria, M.F., Ramirez-Solis, R., Webster, T.D., Ledbetter, D.H., and Caskey, C.T. Alu polymerase chain reaction: a method for rapid isolation of human-specific sequences from complex DNA sources. *Proc.Natl.Acad.Sci.USA* 86:6686-6690 (1989).

129 Poustka, A. and Lehrach, H. Jumping libraries and linking libraries: the next generation molecular tools in mammalian genetics. *Trends Genet.* 2:174-179 (1986).

130 Butler, R., Ogilvie, D.J., Elvin, P., Riley, J.H., Finniear, R.S., Slynn, G., Morten, J.E.N., Markham, A.F., and Anand, R. Walking, cloning, and mapping with yeast artificial chromosomes: a contig encompassing D21S13 and D21S16. *Genomics* 12:42-51 (1992).

131 Muscatelli, F., Monaco, A.P., Goodfellow, P.N., Hors-Cayla, M.C., Lehrach, H., and Fontes, M. Isolation of new probes from Xq12-Q13: an example of the screening of reference libraries with Alu-PCR products from radiation hybrids. *Cytogenet.Cell Genet.* 61:109-113 (1992).

132 Albertsen, H.M., Abderrahim, H., Cann, H.M., Dausset, J., Le Paslier, D., and Cohen, D. Construction and characterization of a yeast artificial chromosome library containing seven haploid human genome equivalents. *Proc.Natl.Acad.Sci.USA* 87:4256-4260 (1990).

133 Anand, R., Riley, J.H., Butler, R., Smith, J.C., and Markham, A.F. A 3.5 genome equivalent multi access YAC library: construction, characterisation, screening and storage. *Nucleic Acids Res.* 18:1951-1956 (1990).

134 Larin, Z., Monaco, A.P., and Lehrach, H. Yeast artificial chromosome libraries containing large inserts from mouse and human DNA. *Proc.Natl.Acad.Sci.USA* 88:4123-4127 (1991).

135 Driesen, M.S., Dauwerse, J.G., Wapenaar, M.C., Meershoek, E.J., Mollevanger, P., Chen, K.L., Fischbeck, K.H., and Van Ommen, G.J.B. Generation and fluorescent in situ hybridization mapping of yeast artificial chromosomes of 1p, 17p, 17q, and 19q from a hybrid cell line by high-density screening of an amplified library. *Genomics* 11:1079-1087 (1991).

136 Bellanné-Chantelot, C., Lacroix, B., Ougen, P., Billault, A., Beaufils, S., Bertrand, S., Georges, I., Glibert, F., Gros, I., Lucotte, G., Susini, L., Codani, J.-J., Gesnouin, P., Pook, S., Vaysseix, G., Lu-Kuo, J., Ried, T., Ward, D., Chumakov, I., Le Paslier, D., Barillot, E., and Cohen, D. Mapping the whole human genome by fingerprinting yeast artificial chromosomes. *Cell* 70:1059-1068 (1992).

137 Chumakov, I.M., Rigault, P., Le Gall, I., Bellanné-Chantelot, C., Billault, A., Guillou, S., Soularue, P., Guasconi, G., Poullier, E., Gros, I., Belova, M., Sambucy, J.L., Susini, L., Gervy, P., Glibert, F., Beaufils, S., Bui, H., Massart, C., De Tand, M.F., Dukasz, F., Lecoulant, S., Ougen, P., Perrot, V., and Saumler, M. A YAC contig map of the human genome. *Nature* 377 Suppl.175-183 (1995).

138 Vos, P., Hogers, R., Bleeker, M., Reijans, M., Van de Lee, T., Hornes, M., Frijters, A., Pot, J.,

Peleman, J., Kuiper, M., and Zabeau, M. AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Res.* 23:4407-4414 (1995).

139  Coffey, A., Gregory, S., and Cole, C.G. Alu-PCR fingerprinting of YACs. *Methods.Mol.Biol.* 54:97-114 (1996).

140  Nagaraja, R., Kere, J., MacMillan, S., Masisi, M.W.J., Johnson, D., Molini, B.J., Halley, G.R., Wein, K., Trusgnich, M., Eble, B., Railey, B., Brownstein, B.H., and Schlessinger, D. Characterization of four human YAC libraries for clone size, chimerism and X chromosome sequence representation. *Nucleic Acids Res.* 22:3406-3411 (1994).

141  Pavan, W.J., Hieter, P., and Reeves, R.H. Generation of deletion derivatives by targeted transformation of human-derived yeast artificial chromosomes. *Proc.Natl.Acad.Sci.USA* 87:1300-1304 (1990).

142  Pavan, W.J., Hieter, P., Sears, D., Burkhoff, A., and Reeves, R.H. High-efficiency yeast artificial chromosome fragmentation vectors. *Gene* 106:125-127 (1991).

143  Heus, J.J., De Winther, M.P.J., Van de Vosse, E., Van Ommen, G.J.B., and Den Dunnen, J.T. Centromeric and non-centromeric *ADE2*-selectable fragmentation vectors for YACs in AB1380. *Genome Res.* 7:657-660 (1997).

144  Van de Vosse, E., Van der Bent, P., Heus, J.J., Van Ommen, G.J.B., and Den Dunnen, J.T. High resolution mapping by YAC fragmentation of a 2.5 Mb Xp22 region containing the human RS, KFSD and CLS disease genes. *Mamm.Genome* 8:497-501 (1997).

145  Potier, M.-C., Dutriaux, A., and Reeves, R. Use of YAC fragmentation to delimit a duplicated region on human chromosome 21. *Mamm.Genome* 7:85-88 (1996).

146  Datson, N.A., Semina, E., Van Staalduinen, A.A.A., Dauwerse, H.G., Meershoek, E.J., Heus, J.J., Frants, R.R., Den Dunnen, J.T., Murray, J.C., and Van Ommen, G.-J.B. Closing in on the Rieger syndrome gene on 4q25: mapping translocation breakpoints within a 50 kb region. *Am.J.Hum.Genet.* 59:1297-1305 (1996).

147  Hochgeschwender, U., Sutcliffe, J.G., and Brennan, M.B. Construction and screening of a genomic library specific for mouse chromosome 16. *Proc.Natl.Acad.Sci.USA* 86:8482-8486 (1989).

148  Lovett, M., Kere, J., and Hinton, L.M. Direct selection: a method for the isolation of cDNAs encoded by large genomic regions. *Proc.Natl.Acad.Sci.USA* 88:9628-9632 (1991).

149  Parimoo, S., Patanjali, S.R., Shukla, H., Chaplin, D.D., and Weissman, S.M. cDNA selection: efficient PCR approach for the selection of cDNAs encoded in large chromosomal DNA fragments. *Proc.Natl.Acad.Sci.USA* 88:9623-9627 (1991).

150  Guo, W., Worley, K., Adams, V., Mason, J., Sylvester-Jackson, D., Zhang, Y.-H., Towbin, J.A., Fogt, D.D., Madu, S., Wheeler, D.A., and McCabe, E.R.B. Genomic scanning for expressed sequences in Xp21 identifies the glycerol kinase gene. *Nature Genet.* 4:367-371 (1993).

151  Rommens, J.M., Lin, B., Hutchinson, G.B., Andrew, S.E., Goldberg, Y.P., Glaves, M.L., Graham, R., Lai, V., McArthur, J., Nasir, J., Theilmann, J., McDonald, H., Kalchmann, M., Clarke, L.A., Schappert, K., and Hayden, M.R. A transcription map of the region containing the Huntington disease gene. *Hum.Mol.Genet.* 2:901-907 (1993).

152  Miki, Y., Swensen, J., Shattuck-Eidens, D., Futreal, P.A., Harshman, K., Tavtigian, S., Liu, Q., Cochran, C., Bennett, L.M., Ding, W., Bell, R., Rosenthal, J., Hussey, C., Tran, T., McClure, M.,

Frye, C., Hattier, T., Phelps, R., Haugen-Strano, A., Katcher, H., Yakumo, K., Gholami, Z., Shaffer, D., Stone, S., Bayer, S., Wray, C., Bogden, R., Dayananth, P., Ward, J., Tonin, P., Narod, S., Bristow, P.K., Norris, F.H., Helvering, L., Morrison, P., Rosteck, P., Lai, M., Barrett, J.C., Lewis, C., Neuhausen, S., Cannon-Albright, L., Goldgar, D., Wiseman, R., Kamb, A., and Skolnick, M.H. A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. *Science* 266:66-71 (1994).

153  Derry, J.M., Ochs, H.D., and Francke, U. Isolation of a novel gene mutated in Wiskott-Aldrich syndrome. *Cell* 78:635-644 (1994).

154  Brennan, M.B. and Hochgeschwender, U. So many needles, so much hay. *Hum.Mol.Genet.* 4:153-156 (1995).

155  National Research Council. Mapping and sequencing the Human Genome. Natl.Acad.Press, Washington,DC (1988).

156  Capone, M.C., Gorman, D.M., Ching, E.P., and Zlotnik, A. Identification through bioinformatics of cDNAs encoding human thymic shared Ag-1/stem cell Ag-2. A new member of the human Ly-6 family. *Journal of Immunology* 157:969-973 (1996).

157  Dodt, G., Braverman, N., Valle, D., and Gould, S.J. From expressed sequence tags to peroxisome biogenesis disorder genes. *Ann.N.Y.Acad.Sci.* 804:516-523 (1996).

158  Banfi, S., Borsani, G., Rossi, E., Bernard, L., Guffanti, A., Rubboli, F., Marchitiello, A., Giglio, S., Coluccia, E., Zollo, M., Zuffardi, O., and Ballabio, A. Identification and mapping of human cDNAs homologous to Drosophila mutant genes through EST database searching. *Nat.Genet.* 13:167-174 (1996).

159  Sedlacek, Z., Konecki, D.S., Siebenhaar, R., Kioschis, P., and Poustka, A. Direct selection of DNA sequences conserved between species. *Nucleic Acids Res.* 21:3419-3425 (1993).

160  Bird, A.P. CpG-rich islands and the function of DNA methylation. *Nature* 321:209-213 (1986).

161  Antequera, F. and Bird, A. Number of CpG islands and genes in human and mouse. *Proc.Natl.Acad.Sci.USA* 90:11995-11999 (1993).

162  John, R.M., Robbins, C.A., and Myers, R.M. Identification of genes within CpG-enriched DNA from human chromosome 4p16.3. *Hum.Mol.Genet.* 3:1611-1616 (1994).

163  Patel, K., Cox, R., Shipley, J., Kiely, F., Frazer, K., Cox, D.R., Lehrach, H., and Sheer, D. A novel and rapid method for isolating sequences adjacent to rare cutting sites and their use in physical mapping. *Nucl.Acids Res.* 19:4371-4375 (1991).

164  Shiraishi, M., Lerman, L.S., and Sekiya, T. Preferential isolation of DNA fragments associated with CpG islands. *Proc.Natl.Acad.Sci.USA* 92:4229-4233 (1995).

165  Cross, S.H., Charlton, J.A., Nan, X., and Bird, A.P. Purification of CpG islands using a methylated DNA binding column. *Nat.Genet.* 6:236-244 (1994).

166  Auch, D. and Reth, M. Exon trap cloning: using PCR to rapidly detect and clone exons from genomic DNA fragments. *Nucleic.Acids.Res.* 18:6743-6744 (1990).

167  Buckler, A.J., Chang, D.D., Graw, S.L., Brook, J.D., Haber, D.A., Sharp, P.A., and Housman, D.E. Exon amplification: A strategy to isolate mammalian genes based on RNA splicing. *Proc.Natl.Acad.Sci.USA* 88:4005-4009 (1991).

168  Church, D.M., Stotler, C.J., Rutter, J.L., Murrell, J.R., Trofatter, J.A., and Buckler, A.J. Isolation

of genes from complex sources of mammalian genomic DNA using exon amplification. *Nature Genet.* 6:98-105 (1994).

169 Nehls, M., Pfeifer, D., and Boehm, T. Exon amplification from complete libraries of genomic DNA using a novel phage vector with automatic plasmid excision facility: application to themouse neurofibromatosis-1 locus. *Oncogene* 9:2169-2175 (1994).

170 Kreissig, S., Schuddekopf, K., Dear, N., and Boehm, T. Expression of peptides encoded by exons in cloned mammalian DNA. *Nucl.Acids Res.* 24:4358-4359 (1996).

171 Båtshake, B. and Sundelin, J. The mouse genes for the EP1 prostanoid receptor and the PKN protein kinase overlap. *Biochem.Biophys.Res.Commun.* 227:70-76 (1996).

172 Krizman, D.B. and Berget, S.M. Efficient selection of 3'-terminal exons from vertebrate DNA. *Nucleic Acids Res.* 21:5198-5202 (1993).

173 Datson, N.A., Duyk, G.M., Blonden, L.A.J., Van Ommen, G.-J.B., and Den Dunnen, J.T. An exon trapping system providing size selection of spliced clones and facilitating direct cloning. In: Hochgeschwender, U., Gardiner, K. (eds) Identification of Transcribed Sequences. Plenum Press, New York: 169-181.(1994).

174 Datson, N.A., Duyk, G.M., Van Ommen, G.J.B., and Den Dunnen, J.T. Specific isolation of 3'-terminal exons of human genes by exon trapping. *Nucleic Acids Res.* 22:4148-4153 (1994).

175 Hawkins, J.D. A survey on intron and exon lengths. *Nucl.Acids Res.* 16:9893-9908 (1988).

176 Datson, N.A., Van de Vosse, E., Dauwerse, H.G., Bout, M., Van Ommen, G.J.B., and Den Dunnen, J.T. Scanning for genes in large genomic regions: cosmid-based exon trapping of multiple exons in a single product. *Nucleic Acids Res.* 24:1105-1111 (1996).

177 Lau, C. *International Workshop on the Identification of Transcribed Sequences* 6:(Abstract) (1996).

178 The Huntington's Disease Collaborative Research Group A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. *Cell* 72:971-983 (1993).

179 Chen, E.Y., Zollo, M., Mazzarella, R., Ciccodicola, A., Chen, C., Zuo, L., Heiner, C., Burough, F., Repetto, M., Schlessinger, D., and D'Urso, M. Long-range sequence analysis in Xq28: thirteen known and six candidate genes in 219.4 kb of high GC DNA between the RCP/GCP and G6PD loci. *Hum.Mol.Genet.* 5:659-668 (1996).

180 Fickett, J.W. Finding genes by computer: the state of the art. *Trends.Genet.* 12:316-320 (1996).

181 Brenner, S.E. BLAST, Blitz, BLOCKS and BEAUTY: sequence comparison on the net. *Trends Genet.* 11:330-331 (1995).

182 Burset, M. and Guigo, R. Evaluation of gene structure prediction programs. *Genomics.* 34:353-367 (1996).

183 Lopez, R.S., Larsen, F., and Prydz, H. Evaluation of the exons predictions of the GRAIL software. *Genomics* 24:133-136 (1994).

184 Cotton, R.G.H. Mutation detection. Oxford University Press, Oxford: (1997).

185 Den Dunnen, J.T., Liang, P., Van Ommen, G.J.B., and Van Broeckhoven, C. Mutation detection and diagnosis using PFGE. In: Davies, K.E. (ed) Human genetic disease analysis: a practical approach. 3rd edn. Oxford University Press, Oxford: 45-67.(1993).

186 Datson, N.A., Schaap, C., Bakker, E., Semina, E., Murray, J.C., Van Ommen, G.-J.B., Frants, R.R., and Den Dunnen, J.T. *De novo* splice mutation in the homeobox gene *RIEG* involved in Rieger

syndrome. *In preparation* (1997).

187 Antonarakis, S.E., Rossiter, J.P., Young, M., Horst, J., De Moerloose, P., Sommer, S.S., Ketterling, R.P., Kazazian, H.H.j., Négrier, C., Vinciguerra, C., Gitschier, J., Goossens, M., Girodon, E., Ghanem, N., Plassa, F., Lavergne, J.M., Vidaud, M., Costa, J.M., Laurian, Y., Lin, S.-W., Lin, S.-R., Shen, M.-C., Lillicrap, D., Taylor, S.A.M., Windsor, S., Valleix, S.V., Nafa, K., Sultan, Y., Delpech, M., Vnencak-Jones, C.L., Phillips, J.A.I., Ljung, R.C.R., Koumbarelis, E., Gileraki, A., Mandalaki, T., Jenkins, P.V., Collins, P.W., Pasi, K.J., Goodeve, A., Peake, I., Preston, F.E., Schwartz, M., Scheibel, E., Ingerslev, J., Cooper, D.N., Millar, D.S., Kakkar, V.V., Giannelli, F., Naylor, J.A., Tizzano, E.F., Baiget, M., Domenech, M., Altisent, C., Tusell, J., Beneyto, M., Lorenzo, J.I., Gaucher, C., Mazurier, C., Peerlinck, K., Matthijs, G., Cassiman, J.J., Vermylen, J., Mori, P.G., Acquila, M., Caprino, D., and Inaba, H. Factor VIII gene inversions in severe hemophilia A: results of an international consortium study. *Blood* 86:2206-2212 (1995).

188 Roest, P.A.M., Roberts, R.G., Sugino, S., Van Ommen, G.J.B., and Den Dunnen, J.T. Protein truncation test (PTT) for rapid detection of translation-terminating mutations. *Hum.Mol.Genet.* 10:1719-1721 (1993).

189 Sarkar, G. and Sommer, S.S. Access to a messenger RNA sequence or its protein product is not limited by tissue or species specificity. *Science* 244:331-334 (1989).

190 Hogervorst, F.B.L., Cornelis, R.S., Bout, M., Van Vliet, M., Oosterwijk, J.C., Olmer, R., Bakker, B., Klijn, J.G.M., Vasen, H.F.A., Meijers-Heijboer, H., Menko, F.H., Cornelisse, C.J., Den Dunnen, J.T., Devilee, P., and Van Ommen, G.-J.B. Rapid detection of BRCA1 mutations by the protein truncation test. *Nature Genet.* 10:208-212 (1995).

191 Van der Luijt, R., Khan, P.M., Vasen, H., Van Leeuwen, C., Tops, C., Roest, P., Den Dunnen, J., and Fodde, R. Rapid detection of translation-terminating mutations at the adenomatous polyposis coli (APC) gene by direct protein truncation test. *Genomics* 20:1-4 (1994).

192 Petrij, F., Giles, R.H., Dauwerse, H.G., Saris, J.J., Hennekam, R.C., Masuno, M., Tommerup, N., Van Ommen, G.J., Goodman, R.H., Peters, D.J., and Breuning, M.H. Rubinstein-Taybi syndrome caused by mutations in the transcriptional co-activator CBP. *Nature* 376:348-351 (1995).

193 Kaplan, J.C., Kahn, A., and Chelly, J. Illegitimate transcription: its use in the study of inherited disease. *Hum.Mutat.* 1:357-360 (1992).

194 Cariello, N.F. and Skopek, T.R. Mutational analysis using denaturing gradient gel electrophoresis and PCR. *Mutat.Res.* 288:103-112 (1993).

195 Mural, R. and Gardiner, K. Toward understanding genome function: workshop in France focuses on transcribed sequences. *Hum.Genome News* 7(5):(1996).

196 Olson, M.V. The Human Genome Project. *Proc.Natl.Acad.Sci.USA* 90:4338-4344 (1993).

197 Cann, H.M. CEPH maps. *Curr.Opin.Genet.Dev.* 2:393-399 (1992).

198 Weissenbach, J., Gyapay, G., Dib, C., Vignal, A., Morisette, J., Millasseau, P., Vaysseix, G., and Lathrop, M. A second-generation linkage map of the human genome. *Nature* 359:794-801 (1992).

199 Gyapay, G., Morisette, J., Vignal, A., Dib, C., Fizames, C., Millasseau, P., Marc, S., Bernardi, G., Lathrop, M., and Weissenbach, J. The 1993-94 Généthon human genetic linkage map. *Nature Genet.* 7:246-249 (1994).

200 Adamson, D., Albertsen, H., Ballard, L., Bradley, P., Carlson, M., Cartwright, P., Council, C.,

Elsner, T., Fuhrman, D., Gerken, S., Harris, L., Holik, P.R., Kimball, A., Knell, J., Lawrence, E., Lu, J., Marks, A., Matsunami, N., Melis, R., Milner, B., Moore, M., Nelson, L., Odelberg, S., and Peters, G. A collection of ordered tetranucleotide-repeat markers from the human genome. *Am.J.Hum.Genet.* 57:619-628 (1995).

201 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., Hu, X., Colbert, A.M.E., Rosenberg, C., Reeve-Daly, M.P., Rozen, S., Hui, L., Wu, X., Vestergaard, C., Wilson, K.M., Bae, J.S., Maitra, S., Ganiatsas, S., Evans, C.A., DeAngelis, M.M., Ingalls, K.A., Nahf, R.W., Horton, L.T.j., Oskin Anderson, M., Collymore, A.J., Ye, W., Kouyoumjian, V., Zemsteva, I.S., Tam, J., Devine, R., Courtney, D.F., Turner Renaud, M., Nguyen, H., O'Connor, T.J., Fizames, C., Fauré, S., Gyapay, G., Dib, C., Morissette, J., Orlin, J.B., Birren, B.W., Goodman, N., Weissenbach, J., Hawkins, T.L., Foote, S., Page, D.C., and Lander, E.S. An STS-based map of the human genome. *Science* 270:1945-1954 (1995).

202 Schuler, G.D., Boguski, M.S., Hudson, T.J., Hui, L., Castle, A.B., Wu, X., Silva, J., Nusbaum, H.C., Birren, B.B., Slonim, D.K., Rozen, S., Stein, L.D., Page, D., Lander, E.S., Stewart, E.A., Aggarwal, A., Bajorek, E., Brady, S., Chu, S., Fang, N., Hadley, D., Harris, M., Hussain, S., and Hudson, J.R.j. Genome maps 7. The human transcript map. Wall Chart. *Science* 274:547-562 (1996).

203 Gyapay, G., Schmitt, K., Fizames, C., Jones, H., Vega-Czarny, N., Spillett, D., Muselet, D., Prud'Homme, J.-F., Dib, C., Auffray, C., Morisette, J., Weissenbach, J., and Goodfellow, P. A radiation hybrid map of the human genome. *Hum.Mol.Genet.* 5:339-346 (1996).

204 Dib, C., Faure, S., Fizames, C., Samson, D., Drouot, N., Vignal, A., Millasseau, P., Marc, S., Hazan, J., Seboun, E., Gyapay, G., Morissette, J., and Weissenbach, J. A comprehensive genetic map of the human genome based on 5,264 microsatellites. *Nature* 380:152-154 (1996).

205 Anonymous. Second International Strategy Meeting on Human Genome Sequencing. *Unpublished* (1997).

206 Ali, R.R., Reichel, M.B., Thrasher, A.J., Levinsky, R.J., Kinnon, C., Kanuga, N., Hunt, D.M., and Bhattacharya, S.S. Gene transfer into the mouse retina mediated by an adeno-associated viral vector. *Hum.Mol.Genet.* 5:591-594 (1996).

59

# CHAPTER 2.1

**An Xp22.1-p22.2 YAC contig encompassing the disease loci for RS, KFSD, CLS, HYP and RP15; refined localization of RS and KFSD**

Van de Vosse, E., Van der Wielen, M.J.R., Meershoek, E.J., Oosterwijk, J.C., Gregory, S., Bakker, B., Weissenbach, J., Coffey, A.J., Van Ommen, G.J.B., and Den Dunnen, J.T.

# An Xp22.1-p22.2 YAC contig encompassing the disease loci for RS, KFSD, CLS, HYP and RP15; refined localization of RS and KFSD.

Esther van de Vosse, Michiel J.R. van der Wielen, Arthur A.B. Bergen, Eric J. Meershoek, Jan C. Oosterwijk, Simon Gregory, Bert Bakker, Jean Weissenbach, Alison J. Coffey, Gert-Jan B. van Ommen and Johan T. den Dunnen.

## ABSTRACT

Genetic linkage studies have mapped several diseases, including retinoschisis (RS), keratosis follicularis spinulosa decalvans (KFSD), Coffin-Lowry syndrome (CLS), X-linked hypophosphatemic rickets (XLH, locus name HYP) and X-linked dominant cone-rod degeneration (locus name RP15) to the Xp22-region. To facilitate the positional cloning of the genes involved, we have extended the molecular map of the region. Screening of several YAC-libraries allowed us to identify 63 Xp22 YACs, 52 of which localize between markers DXS414 (P90) and DXS451 (kQST80H1). Analysis of their marker content facilitated the construction of a YAC contig partially overlapping the existing map from the region and extending 1.5 Mb centromeric. The markers were ordered as follows: DXS414 - DXS987 - DXS207 - DXS1053 - DXS197 - DXS43 - DXS1195 - DXS418 - DXS999 - PDHA1 - DXS7161 - DXS443 - DXS7592 - DXS1229 - DXS365 - DXS7101 - DXS7593 - DXS1052 - DXS274 -DXS989 - DXS451. The region between DXS414 and DXS451 covers about 4.5 to 5 Mb, 1.5 Mb of which is in the newly mapped DXS1229-DXS451 region. Three additional markers (PDHA1, DXS7161 and DXS7592) were placed in the previously mapped region, thereby increasing the genetic resolution. Considering the known genetic distances, this region shows a significantly increased recombination frequency, of 0.2 Mb per cM. Using the deduced marker order, the analysis of key recombinants in families segregating RS allowed us to refine the critical region for RS to 0.6 Mb, between DXS418 and DXS7161. Similarly, the candidate region for KFSD could be limited to a 1 Mb region between DXS7161 and DXS1226.

## INTRODUCTION

To facilitate the isolation and study of disease genes, one has set out to completely characterize the human genome. This involves three main stages; mapping (both genetic and physical), cloning and sequencing. The first two steps are carried out in parallel and global physical and genetic maps have been published [1,2]. These crude maps subsequently require verification and refinement by detailed characterization of each region specifically. The human X chromosome is one of the best mapped human chromosomes and, at present, has the highest proportion of known markers and genes. Nonetheless, the Xp22.1-p22.2 region has a relative shortage of markers and is consequently bare in cloned genes [3]. On the other hand, linkage analysis mapped a number of disease loci to the Xp22.1-p22.2 region [4], including spondylo-epiphyseal dysplasia (SEDL, MIM 313400), retinoschisis (RS, MIM 312700), keratosis follicularis spinulosa decalvans (KFSD, MIM 308800), Coffin-Lowry syndrome (CLS, MIM 303600), X-linked hypophosphatemic rickets (XLH, locus name HYP, MIM 307800) and X-linked dominant cone-rod degeneration (locus name RP15, MIM 268000) [5]. Our group has a special interest in RS and KFSD. Linkage studies have placed the gene for RS, a rare hereditary vitroretinal degeneration, between DXS43 and DXS365 [6,7]. Recently, we have shown that KFSD, a rare X-linked disorder characterized by follicular hyperkeratosis of the skin, scarring alopecia of the

63

scalp, absence of eyebrows and corneal degeneration, to be located between DXS16 and DXS269 [8]. Contiguous gene syndromes and large deletions, which have greatly contributed to the unravelling of other regions of the X chromosome, have not been reported in the Xp22.1-p22.2 region.

As a first step to resolve this lack of knowledge and towards identification of RS and KFSD, we have used known STS markers, *Alu*-PCR products of YACs isolated in the course of the project and new Généthon markers to screen YAC libraries and isolate YACs from the Xp22-region. An *Alu*-PCR-based fingerprinting method ([9], and Coffey *et al.* unpublished data) was used to assemble crude contigs and to determine overlaps between the YACs. Subsequently, we have used PCR and hybridization analysis to refine the contig, order the markers and construct a physical map. The deduced map spans about 4.5 to 5 Mb and includes the loci for RS, KFSD, CLS and HYP. While this work was ongoing, Alitalo *et al.* [10] published an Xp22-contig, which overlaps ours from DXS414 to DXS1229 and which is in complete agreement with the map we will present here. We place three additional markers (PDHA1, DXS7161 and DXS7592) in the overlapping region and our map extends at least 1.5 Mb proximally, spanning eight other markers. The mapping data were used to refine the localization of RS by linkage analysis to 0.6 Mb between DXS418 and DXS7161, and the localization of KFSD to 1 Mb between DXS7161 and DXS1226 [11].

## MATERIALS AND METHODS
### YAC library screening
YAC library screening was performed by hybridization. High-density gridded filters from the ICI library [12] and the X-specific subset of the CEPH megabase library ([13] and unpublished) were provided by the EU-sponsored YAC Screening Center Leiden (The Netherlands). Gridded *Alu*-PCR product filters from ICRF-library 900 [14] were provided by the ICRF (London, U.K.).

Filters were hybridized in 0.5 M Sodiumphosphate pH 7.2, 7% SDS, 1% BSA, 1 mM EDTA for 16 hours at 65°C. Filters were washed twice at RT in 40 mM Sodiumphosphate pH7.2, 0.1% SDS before scanning using a PhosphorImager (Molecular Dynamics). *Alu*-PCR products and cosmid probes were prehybridized with 100 µg sheared human placental DNA (Sigma) and 25 µg *Alu*-dimer DNA.

### Probes
Probes used for the YAC library screening included unique probes and cosmids containing markers known to be located in Xp22 and *Alu*-PCR products derived from positive YACs isolated in the course of the project. *Alu*-PCR products were generated using primer PDJ34 [15]. Probes used for detailed YAC analysis included *PvuII/BamHI* fragments of pBR322 to identify the pYAC4 vector arms, CRI-L1391 (DXS274) [16], P122 (DXS418) [17], pD2 (DXS43) [18] and PCR products derived from the markers listed in Table 1. Probes for the analysis of recombinant patient samples included, in addition to the markers used for YAC analysis, 782 (DXS85) [19], pXUT23 (DXS16) [20], C7 (DXS28) [21] and p99-6 (DXS41) [18].

### YAC analysis
Yeast clones were colony purified and high molecular weight DNA was isolated in LMT agarose plugs (Seaplaque, FMC) as described [22]. Initial characterization of YACs was performed by PFGE, FISH and PCR. *Hin*dIII-digested YACs were electrophoresed on 0.8% SeaKem LE

agarose gels.

## Pulsed Field Gel Electrophoresis (PFGE)

PFGE-analysis, using a CHEF-system, was performed under standard conditions [22]. Agarose gels (1%) in 45 mM Tris, 45 mM Boric acid, 0.5 mM EDTA, pH 8.3 were electrophoresed for 24 h at 180 V, with pulse times ranging from 20 to 70 s. Lambda oligomers and AB1380 yeast genomic DNA were used as size standards. After electrophoresis, gels were stained, photographed, blotted on Hybond-N⁺ [22] and fixed by UV crosslinking. Filters were washed to 1x SSC at 65°C before autoradiography.

## Fluorescent In Situ Hybridization (FISH)

FISH-experiments, including two-colour FISH, were performed according to Dauwerse *et al.* [23]. Total yeast DNA was labelled by nick translation in the presence of Dig-11-UTP and biotin-14-dATP and detected by FITC (green) and TRITC (red) respectively.

## Polymerase Chain Reaction (PCR)

PCR was carried out in a Perkin-Elmer Cetus thermal cycler in 30 µl reactions containing 50 mM Tris-HCl pH 9.0, 50 mM KCl, 1.5 mM MgCl$_2$, 0.01% gelatin, 0.1% Triton X-100, 0.2 mM each dNTP, 0.2 mg/ml BSA, 25 pmol of each primer and 0.25 U *Taq* polymerase (Super*Taq*, HT Biotechnology) with 50 µl mineral oil overlay. An initial denaturation step of 5 min 94°C was followed by 30 cycles of 94°C 1 min, 50-65°C (Table 1) 1 min, 72°C 1 min. Products were separated on a 2% agarose gel.

PCR for linkage analysis in the RS families was performed in 15 µl reactions containing 150-210 ng DNA, 1.5 mM MgCl$_2$, 10 mM Tris-HCl pH 9.0, 50 mM KCl, 0.01% gelatin, 0.1% Triton X-100, 0.2 mM dATP, 0.2 mM dGTP, 0.2 mM dTTP, 0.025 mM dCTP, 0.6 U Amplitaq™ (Cetus Inc), 0.27 pmol α$^{32}$P-dCTP, 5 pmol of each primer with 50 µl mineral oil overlay. An initial denaturation step of 5 min 94°C was followed by 30 cycles of 94°C 1 min, 50°C 1 min (or according to Table 1), 72°C 2 min and a final step of 10 min 72°C. Products were separated on a sequencing gel.

## *Alu*-PCR fingerprint analysis

Primary *Alu*-PCR was performed as described previously [9] using primers ALE1 and ALE3 together in a combined reaction [24]. A dilution of the primary PCR (Fig.1a) was used in a secondary PCR for 10 cycles in the presence of end-labelled ALE1 and ALE3. The products were run on a gel (Fig.1b) as described previously [25]. After autoradiography, the fingerprints were analyzed using the semi-automated gel scanning and analysis system used for the C.*elegans* cosmid fingerprinting project [26]. The positions and intensities of the products are compared and the probability of overlap between YACs is analyzed. In the resulting contig the length of a clone is depicted as the number of products it contains (Fig.1c).

## Recombinants analysis

For RS, linkage analysis has been carried out with at least six Xp22.2 markers in 21 families, 15 of which have been published previously [6,7]. From these studies, two families with key-recombination events were selected. DNA samples of 13 individuals from these two families were analyzed with 10 (family P 22.337) or 11 (family P 24.130) polymorphic markers.

| Locus | Marker | Product size (bp) | $T_a$ | Reference |
|---|---|---|---|---|
| DXS414 | P90 | 330 | 55 | [27] |
| DXS987 | AFM120xa9 | 206-224 | 55 | [27] |
| DXS207 | pPA4B | 500 | 52 | [27] |
| DXS1053 | AFM164zd4 | 194-206 | 60 | [2] |
| DXS197 | pTS247 | 250 | 52 | [27] |
| DXS43 | pD2 | 86-130 | 55 | [28] |
| DXS1195 | AFM207zd6 | 235-239 | 56 | [2] |
| DXS418 | P122 | 140-158 | 58 | [29] |
| DXS999 | AFM234yf12 | 260-276 | 50 | [27] |
| PDHA1 | PDHA1 | 125 | 58 | [27] |
| DXS7161 | AFM291wf5 | 240-254 | 55 | [4] |
| DXS443 | pRX-324 | 204-210 | 50 | [30] |
| DXS7592 | AFMa244zg1 | 225-233 | 55 | J.W. |
| DXS1229 | AFM337wd5 | 202-230 | 56 | [2] |
| DXS365 | pRX-314 | 201-217 | 50 | [30] |
| DXS7101 | AFMa176zb1 | 156-164 | 55 | J.W. |
| DXS7593 | AFMa346zc1 | 209-223 | 55 | J.W. |
| DXS1052 | AFM163yh2 | 143-159 | 62 | [31] |
| DXS989 | AFM135xe7 | 173-199 | 50 | [27] |
| DXS451 | kQST80H1 | 182-204 | 55 | [30] |

**Table 1**. PCR markers used in the analysis.
$T_a$ = annealing temperature. J.W.= Jean Weissenbach (unpublished).
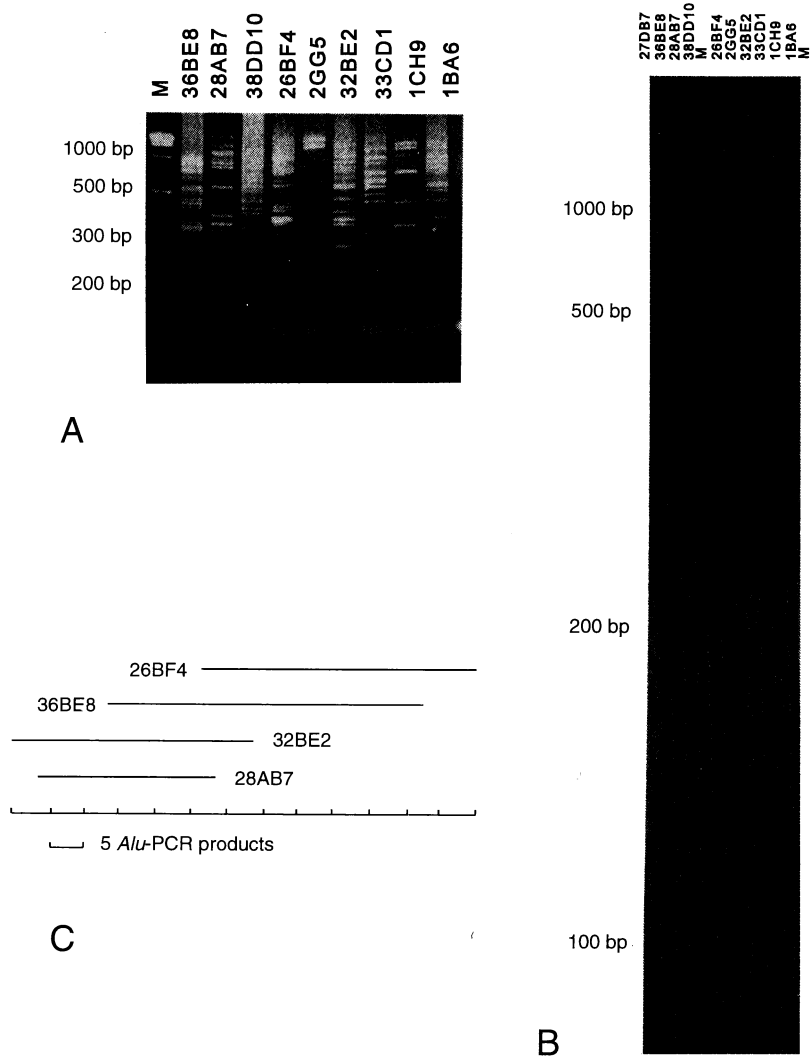
**Figure 1.** *Alu*-PCR fingerprinting. (**A**) 5 µl of primary *Alu*-PCR products of 9 YAC clones run on a 2.5 % agarose minigel. Marker (M) is a 1 kb ladder. (**B**) Autoradiograph showing the *Alu*-PCR fingerprint of the 9 YAC clones. Marker (M) is a *Sau*3AI digested ^{35}S-labelled lambda DNA marker. Sizes are indicated on the left of the autoradiograph. (**C**) Contig constructed after analysing the *Alu*-PCR bands using software originally written by J.Sulston for the *C.elegans* project [26].

## RESULTS

### YAC Library Screening

The CEPH [32], ICI [12] and ICRF [14,33] YAC libraries were screened with 21 probes which resulted in the isolation of 156 potential positive clones. To rapidly obtain a rough physical map, including potential overlap data, all 156 YACs were first *Alu*-PCR fingerprinted. An example of the *Alu*-PCR-based fingerprinting is shown in Fig.1. YACs 36BE8, 28AB7, 26BF4 and 32BE2 share several bands (Fig.1a and b) and therefore clearly overlap and form a contig (Fig.1c). Analysis of all 156 YACs revealed the presence of ten *Alu*-PCR-based contigs, containing a total of 52 YACs. Four of these contigs, 13, 23, 25 and 82 (Fig.2) were located in the Xp22.1-22.2 region as described below. The remaining 6 contigs were in other regions of the X chromosome and not further analyzed. Due to screening of the YAC library with *Alu*-PCR products of a chimeric YAC, contig 25 originally contained YACs that were located on chromosome 4 or 5 as deduced from FISH analysis.
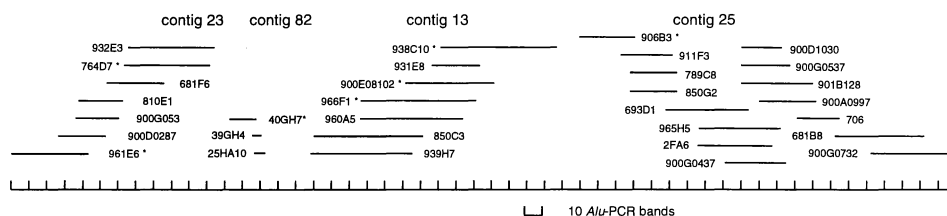


**Figure 2**. *Alu*-PCR fingerprinting contigs 13, 23, 25 and 82. The size of the clones is determined by the number of *Alu*-PCR bands it contains. An asterisk indicates YACs that give FISH-signals on Xp22 as well as on other chromosomes. Not all YACs in these contigs have been analyzed in detail.

### YAC analysis

Guided by the contigs derived from the fingerprinting, the YACs were tested for the presence of Xp22 markers. Out of the 156 YACs, 63 were positive for at least one marker, and 23 were positive for two or more markers. Since the density of YACs was high enough to construct a good contig, we only analyzed YACs containing two or more markers in greater detail and we did not analyze the 40 YACs containing only one marker. The remaining set of 93 YACs which were not positive for any of the markers, includes both false positives and YACs which, while located in an *Alu*-PCR contig, did not contain a reference marker.

The 23 YACs analyzed in detail ranged in size from 210 to over 1400 kb with an average size of 1050 kb (Table 2). The 19 CEPH YACs ranged in size from 590 to 1400 kb (1150 kb average), the 2 ICRF YACs were 580 and 1180/1300 kb (1020 kb average), and the 2 ICI YACs were 210 kb and 500 kb (380 kb average). These sizes do not differ significantly from those published (900 kb average for the CEPH library [32], 620 kb average for the ICRF library [14] and 350 kb average for the ICI library [12]). Sizes for specific YACs (Table 2) were usually consistent with available data (that is Généthon database [1] and [10]). One exceptional case was 939H7, which we find to be 1300 kb while it is reported as 270 kb in the database [1]. Since we performed a colony purification of the YAC clones it is possible that we have isolated the minor

component of a mixed YAC population. However, our data are consistent with those of Alitalo *et al.* [10] who report a length of 650 to 1500 kb and state that this YAC is unstable.

| YAC | size (kb) | FISH |
|---|---|---|
| CEPH 961E6 | 1350 + >1400 (1680) | Xp22 + 4q + 16p |
| CEPH 810E1 | 1050   (1100) | Xp22 |
| CEPH 681F6 | 950   (1060) | Xp22 |
| CEPH 743A8 | 830   (700) | Xp22 |
| ICI 40GH7 | 420 + 210 | Xp22 + 6qcen |
| CEPH 764D7 | N.A. (1420) | Xp22 + 8q1/2 |
| CEPH 932E3 | 1280 (900) | Xp22 |
| ICI 25HA10 | 430 | Xp22 |
| CEPH 811D11 | 1450 (800) | Xp22 |
| CEPH 911F3 | 1150 (950) | Xp22 |
| ICRF 900H0623 | 1180+1300 | Xp22 |
| CEPH 939H7 | 1300 (270) | Xp22 |
| ICRF 900E08102 | 580 | Xp22 + ( 3p ) |
| CEPH 742H9 | 580 + 850 (1250) | Xp22 |
| CEPH 966F1 | 950   (N.A.) | Xp22 + 19qtel |
| CEPH 960A5 | 1300  (1310) | Xp22 |
| CEPH 850C3 | 590  (N.A.) | Xp22 |
| CEPH 789C8 | >1200  (1430) | Xp22 |
| CEPH 850G2 | 1300  (1000) | Xp22 |
| CEPH 693D1 | 1400  (1720) | Xp22 |
| CEPH 965H5 | 1150  (1200) | Xp22 |
| CEPH 681B8 | 930   (980) | Xp22 |
| CEPH 933D5 | 900 (890) + 1450 | Xp22 + 9ptel |

**Table 2**. Detailed characterization of the YACs.
The YAC lengths as reported in the Généthon database are given between brackets. All YACs hybridized to Xp22, six YACs hybridized to other chromosomes as well. 900E08102 does not hybridize to #3 in all metaphases. YACs 742H9 and 960A5 were found to be non-chimeric by FISH, but have been found to be chimeric by Alitalo *et al.* [10] by analysis of YAC endclones. N.A.= not analyzed.

**FISH analysis**
To verify the chromosomal localization of the YACs and to assess potential chimerism, we performed FISH analysis of all 23 YACs. Seventeen YACs showed a hybridization signal on Xp22 only, while six YACs gave more than one hybridization signal (Table 2). The CEPH data

of hybridizations of YAC-derived *Alu*-PCR products to gridded somatic cell hybrids confirm these data; 966F1 hybridizes to X and 18, 764D7 hybridizes to X and 8, 961E6 hybridizes to X, 10, 14, 16 and 18 and 933D5 hybridizes to X, 1 and 9. Two of the six, 900E08102 and 966F1, contained a single YAC as determined by PFGE and are thus most likely chimeric. Four clones contained more than one YAC. Since we performed a colony purification, they probably result from co-cloning of two different YACs in one yeast cell. Four YACs hybridized to Xp22 as well as to the heterochromatic regions on chromosome 1, 9 and 16. These YACs were not analyzed in detail because they did not contain more than one reference marker. However, this observation indicates that there is probably homology between Xp22 and these heterochromatic regions. Our data are not sufficient to localize this homology to a specific region within the contig.

For the ICI and ICRF YACs too few were studied to draw any significant conclusion about the chimerism frequency (ICRF less than 25% according to [34,35]). From the CEPH mega YACs only 2 out of 19 (11%) were chimeric. Alitalo *et al.* [10] analyzed YAC endclones and report chimerism in two more YACs (742H9 and 960A5). Small regions in YACs originating from other chromosomal regions are unlikely to be detected by FISH analysis which explains the differences observed.
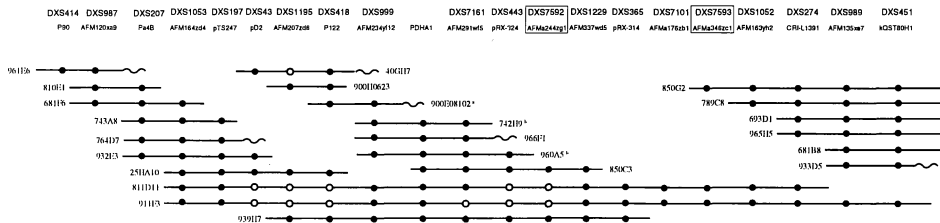


**Figure 3**. Marker based contig of the YACs. Closed circles indicate presence of the marker (after PCR or hybridization analysis), open circles indicate absence of the marker. Wavy lines indicate chimeric YACs. Multiplicity of YACs consistent with this order. Boxed markers are reagents additional to those previously located in contigs covering the region.

## YAC contig and marker order

Figure 3 shows the marker based contig of the YACs which could be constructed. YACs 961E6, 681F6, 25HA10, 939H7, 811D11 and 850G2 together span the entire region. Three YACs, 40GH7, 811D11 and 911F3, appear to contain internal deletions, since several markers were negative. In order to increase the marker density in the region and to locate markers which could not be ordered by genetic mapping, we tried to localize 18 new Généthon markers within the Xp22 region (Table 3). A first rough localization was achieved by testing on the somatic cell hybrids TG2sc1 [36] and AM445x393 [37]. TG2sc1 is a hamster hybrid containing Xp21.3-pter as the only human material, while AM445x393 contains Xp22.2-qter in a mouse background. DXS1223 and AFMa282vd1 were localized to Xp22.3 and fell outside our target region. Twelve markers located in the Xp21.3-p22.2 region, i.e. positive in both cell lines, were tested on the YACs from the contig. Five markers did not give discernable PCR products or gave inconclusive

data. We were able to localize DXS1195 between DXS418 and DXS43, DXS7161 between DXS443 and PDHA1, DXS1229 between DXS443 and DXS365, DXS7592 between DXS443 and DXS1229, DXS7101 between DXS365 and DXS1052, DXS1052 between DXS365 and DXS274, and DXS7593 between DXS7101 and DXS1052. DXS7105 is probably located between DXS365 and DXS7593 but only one YAC was positive with this marker (Table 3).

| Marker | Total human | TG2sc1 | AM445 x393 | Localization | YAC results |
|--------|-------------|--------|------------|--------------|-------------|
| DXS1028 | + | - | - | ? (not on X) | N.A. |
| DXS1043 | + | + | + | Xp21.3-p22.2 | no positives |
| DXS1052 | + | + | + | Xp21.3-p22.2 | see Figure 3 |
| DXS1061 | + | + | + | Xp21.3-p22.2 | no positives |
| DXS1065 | smear | - | + | (Xqtel-Xp21.3) | N.A. |
| DXS1195 | + | + | + | Xp21.3-p22.2 | see Figure 3 |
| DXS1202 | + | + | + | Xp21.3-p22.2 | yeast also positive |
| DXS1223 | + | + | - | Xp22.3 | N.A. |
| DXS1224 | - | - | - | ? | N.A. |
| DXS1229 | + | + | + | Xp21.3-p22.2 | see Figure 3 |
| DXS1233 | - | - | - | ? | N.A. |
| DXS7101 | + | + | + | Xp21.3-p22.2 | see Figure 3 |
| DXS7105 | + | + | + | Xp21.3-p22.2 | 811D11 |
| DXS7161 | + | + | + | Xp21.3-p22.2 | see Figure 3 |
| AFMa152xf1 | - | 400 bp | 400 bp | ? | N.A. |
| DXS7592 | + | + | + | Xp21.3-p22.2 | 939H7, 850C3 |
| DXS7593 | + | + | + | Xp21.3-p22.2 | see Figure 3 |
| AFMa282vd1 | + | + | - | Xp22.3 | N.A. |

**Table 3**. Localization of the new Généthon markers. TG2sc1 contains Xp21.3-pter as the only human material in a hamster background. AM445x393 contains Xqter-Xp22.2 in a mouse background. '+' = a product of the expected length was detected, '-' = no product of the expected length was detected. Between brackets; additional products or products with unexpected sizes amplified by the PCR.

**Candidate regions for KFSD and RS**
The new data on the marker order generated was used to refine the localization of the candidate regions for RS and KFSD. For RS we have carried out linkage analysis with at least six Xp22.2 markers in 21 RS families, 15 of which have been published previously [6,7]. From these studies, two key-recombination events were identified, further refining the RS-gene candidate region (Fig.4). In family P 22.337 the recombination between RS and DXS418, observed in patient III-3, places the disease gene proximal to DXS418. In family P 24.130, the recombination between RS and DXS7161 observed in patient III-3, places the disease gene distal to DXS7161. Unfortunately, DXS999 was not informative in this family.
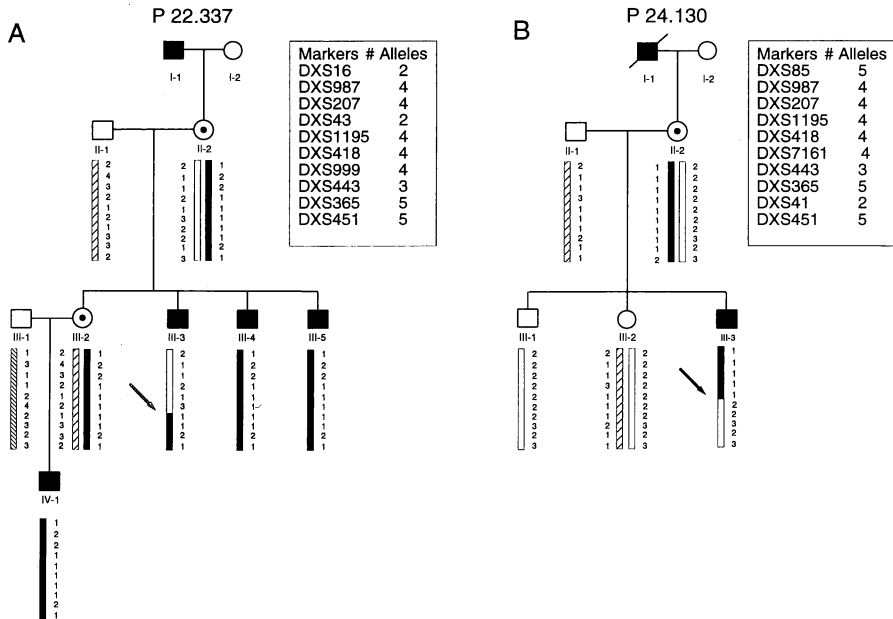
**Figure 4.** Refined critical region for RS. (A) In family P 22.337 the RS disease locus cosegregates, in general, with the haplotype 1 (DXS16) - 2 (DXS987) - 2 (DXS207) - 1 (DXS43) - 1 (DXS1195) - 1 (DXS418) - 1 (DXS999) - 1 (DXS443) - 2 (DXS365) - 1 (DXS451). The recombination observed with patient III-3, between RS and DXS16, DXS987, DXS207, DXS43, DXS1195 and DXS418, places the disease gene proximal to these loci. (B) In family P 24.130 the phase of the markers could be obtained through the analysis of the haplotypes of male III-1 and female III-2. Healthy male III-1 inherited the X-chromosome; 2 (DXS85) - 2 (DXS987) - 2 (DXS207) - 2 (DXS1195) - 2 (DXS418) - 2 (DXS7161) - 2 (DXS443) - 3 (DXS365) - 3 (DXS451) - 2 (DXS41) from his mother (II-2). The phase of the maternal X chromosomal alleles of female III-2 is also 2-2-2-2-2-2-2-2-3-2, since she inherited the X chromosome characterized by 2-1-1-3-1-1-1-2-1-1 from her father. Thus, patient III-3 is most likely recombinant for 2 (DXS7161) - 2 (DXS443) - 3 ((DXS365) - 2 (DXS41) - 3 (DXS451). The analysis of the recombination breakpoints in this family suggest the order Xpter-(DXS85, DXS987, DXS207, DXS1195, DXS418, RS) - (DXS7161, DXS443, DXS365, DXS41, DXS451) -Xcen. DXS999 was not informative.

For KFSD, recombinants were selected from a large Dutch pedigree Oosterwijk 1992[11]. Twenty polymorphic markers, 19 of which were informative, were tested (Fig. 5). Two individuals, VII-10 and VI-13 show key recombinations that refine the localization of the disease gene. VII-10, an affected male, determines the proximal border since the alleles of the affected chromosome recombine proximal to DXS257/DXS7161 with the healthy chromosome. VI-13, an affected male, determines the distal border by a recombination between DXS365 and DXS1229 [11].

| | | VI-13 | VII-10 | |
|---|---|:---:|:---:|---|
| TEL | | | | |
| DXS85 | 782 | ● | ○ | |
| DXS16 | pXUT23 | — | ○ | |
| DXS987 | AFM120xa9 | ● | ○ | |
| DXS43 | pD2 | ● | — | |
| DXS418 | P122 | ● | ○ | |
| DXS999 | AFM234yf12 | ● | — | |
| DXS257 | pQST1H3 | ● | ○ | |
| DXS7161 | AFM291wf5 | — | ○ | |
| DXS443 | pRX-324 | ● | — | |
| DXS1229 | AFM337wd5 | — | — | KFSD |
| DXS365 | pRX-314 | ● | ● | |
| DXS1226 | AFM316yf5 | ○ | ● | |
| DXS1052 | AFMa163yh2 | — | ● | |
| DXS274 | CRI L1391 | ○ | — | |
| DXS989 | AFM135xe7 | ○ | ● | |
| DXS451 | kQST80H1 | ○ | ● | |
| DXS41 | 99.6 | — | — | |
| DXS67 | B24 | — | — | |
| DXS28 | C7 | — | — | |
| DXS1235 | STR50 | — | ● | |
| DXS1236 | STR49 | ○ | ● | |
| DXS269 | P20 | ○ | ● | |
| CEN | | | | |

**Figure 5**. Refinement of the localization of the critical region for KFSD. Family members, with a recombination in the Xp22-region, were selected from a large Dutch pedigree [11].VI-13 and VII-10 are affected males.

## DISCUSSION

We have generated a contig map of Xp22.1-p22.2, thereby evaluating the use of two complementary methods: *Alu*-PCR fingerprinting and direct STS-content analysis of individual, overlapping YACs. Ultimately, the most informative data were obtained by testing specific markers directly on the individual YACs. However, the *Alu*-PCR fingerprinting method was found to be a fast and simple initial step to decrease the total workload allowing to divide the YACs into groups with overlapping sequences. The contigs derived by *Alu*-PCR fingerprinting are consistent with the contig based on marker content of the YACs. The four contigs generated on the basis of the fingerprinting alone were found to be all part of a single contig based on the STS content of the YACs. Similar results were obtained in a large chromosome 22 contig by

73

Coffey *et al.* (unpublished data). Chimerism of the YACs at the extremities of the contigs seems to play an important role. The gaps between contigs 23/82 and contigs 13/25 are both bordered by chimeric YACs, 764D7/40GH7 and 938C10/906B3 respectively. Furthermore, an overlap between a YAC that contains many *Alu*-PCR products and a YAC that contains only a few, will not be scored by the ContigC software used, since a 40% cut-off value of shared bands was used to determine overlap. This may explain why there is a gap between contigs 82 and 13, flanked by 25HA10 (producing 6 PCR-products) and 939H7 (producing 58 products).

Alu sequences are known to have a non-random distribution. Giemsa-positive bands -like Xp22.1- have a low *Alu*-repeat content [38]. The average number of *Alu*-PCR products of the YACs in our contig was 3.4 per 100 kb, varying from 1.2 to 10. Although our contig is supposed to span the Xp22.1/Xp22.2 border, we did not observe any significant differences in the number of *Alu*-PCR products derived from the YACs throughout the contig. However, we did observe a significantly decreased frequency of chimerism for the CEPH YACs isolated; 11% compared with the 40-50% chimerism reported [1]. Since a major cause of chimerism is probably recombination during the cloning step at homologous, mainly repetitive DNA sequences [39] the low frequency of chimerism in this region also indicates that it contains a lower number of highly repetitive DNA sequences.

**Marker order and physical distances**

The marker order that was obtained, DXS414 - DXS987 - DXS207 - DXS1053 - DXS197 - DXS43 - DXS1195 - DXS418 - DXS999 - PDHA1 - DXS7161 - DXS443 - DXS7592 - DXS1229 - DXS365 - DXS7101 - DXS7593 - DXS1052 - DXS274 - DXS989 - DXS451 is in agreement with the consensus map of the 5th X-chromosome Workshop [4] and as published by Alitalo *et al.*[10]. Markers DXS443 and PDHA1, not separated previously [4], could be physically ordered by YACs 742H9 and 966F1, which both contain PDHA1 and DXS999, but not DXS443. Consequently, PDHA1 is distal to DXS443. Markers DXS7161, DXS7592, DXS7593 and DXS7101 were not physically mapped before.

Given that YACs are deletion prone and may rearrange, we have tested for the criterium that two or more YACs needed to be consistent with the derived marker position. This criterium applies for the entire contig (2-7 YACs multiplicity, average 4) except for the distal end and the interval DXS43-DXS1195 which was mapped independently by Alitalo *et al.* [10]. Combining the YAC marker content (Fig.3) and YAC length (Table 2) allows the construction of a physical map of the region (Fig 6). For example, YAC 25HA10 is relatively small (430 kb), indicating that the markers DXS1053-DXS418 map close to each other. 932E3 overlaps with 25HA10 and measures 1280 kb. Since only DXS1053-DXS43 are shared, 932E3 extends at least 850 kb to the telomeric side, a region which contains DXS207 (present in 932E3, absent in 25HA10). Similarly, the 830 kb YAC 743A8 extends in the same direction, although it misses DXS43. Combining the data of 810E1 and 681F6 leads to the conclusion that DXS987 is separated by about 900 kb from the start of a 400 kb cluster containing DXS1053, DXS197, DXS43, DXS1195 and DXS418, while DXS207 maps somewhere in between.

The combined data shows that the region between DXS414 and DXS451 measures 4.5 to 5 Mb (Fig.6). We were not able to determine the distance between DXS414 and DXS987, since only one chimeric YAC (961E6) was analyzed and since no markers telomeric of DXS414 were tested. The physical distances reported by Francis *et al.* [40] and Alitalo *et al.* [10], are in full agreement with the map presented here.

74

The ratio between physical and genetic distance is roughly 1 Mb/cM for the entire human genome. The available genetic data indicate that the Xp22.1-p22.2 region analyzed here has a highly increased recombination frequency (Fig.6). The estimated physical distances, in combination with the published genetic distances [31,41], give an average ratio of approximately 0.2 Mb/cM for the region between DXS987 and DXS989. This effect is more pronounced proximal to DXS1053 and reaches a peak value of 8 cM over 900 kb, about 1 cM per 100 kb, between DXS7161 and DXS1052.
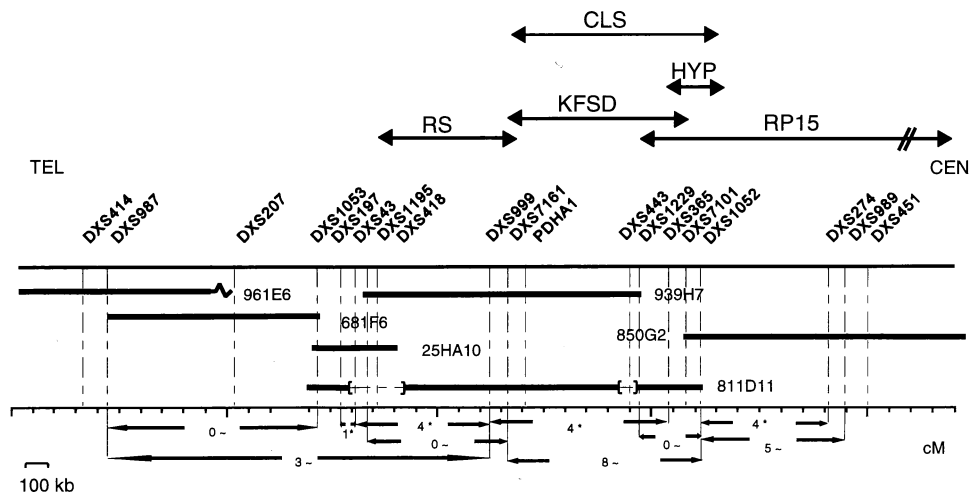


**Figure 6.** Physical map of the Xp22.1-p22.2 region. Physical map of the region obtained after combining marker content and length of all 23 YACs analyzed in detail. Only the YACs that form the minimal contig covering the entire region are shown. The indicated disease gene candidate regions are according to the 5<sup>th</sup> X-chromosome Workshop [4] and updated with the results published in this paper. The genetic distances (in cM) indicated at the bottom of the figure are according to ~Weissenbach *et al.* [41] and *Francis *et al.* [40]

## Disease gene candidate regions

The candidate regions for the HYP, CLS, RS, KFSD and RP15 disease genes are indicated in the physical map (Fig.6). The candidate region for HYP was localized between DXS365 and DXS274 by Francis *et al.* [40], a localization which was refined recently to between DXS365 and DXS1683 (a new marker between DXS365 and DXS1052) [4]. In our contig, this region measures less than 0.3 Mb. The CLS candidate region was first localized between DXS7161 and DXS1052 [42], a localization refined recently to between DXS7161 and DXS1683 [4], in our contig a region of approximately 1 Mb. The RP15 candidate region was localized between DXS1048 and DXS1229 by McGuire *et al.* [5].

The results presented here, combined with linkage data of Oosterwijk *et al.* [11], refine the localization of KFSD to less than 1 Mb between DXS7161 and DXS1226. The localization

75

of RS is refined to 0.6 Mb between DXS418 and DXS7161. Strikingly, and illustrating the relative underdevelopment of this region in terms of reagents, the limiting factor for each of these candidate regions to further refine the localization of the disease genes still is the lack of informative polymorphic markers rather than the absence of recombinants. The contig constructed allowed us to outline the specific requirements for a marker which will further narrow down the disease candidate regions. For instance in the case of KFSD, an informative marker between DXS7161 and DXS365 would be especially valuable and could potentially, decrease the candidate region from 1 Mb to less than 200 kb. Similarly, the candidate region for RS now spans a 600 kb region devoid of any informative marker. Development of new polymorphic markers, based on the clones in the contig presented, is therefore a rewarding effort which we are currently undertaking. On the other hand, the candidate regions for the diseases mentioned have now become small enough to consider establishing a transcriptional map in order to isolate the gene(s) involved in these diseases.

## ACKNOWLEDGEMENTS

## ABBREVIATIONS

KFSD, keratosis follicularis spinulosa decalvans; RS, retinoschisis; CLS, Coffin-Lowry syndrome; HYP, hypophosphatemic rickets; SEDL, spondylo-epiphyseal dysplasia; RP15, X-linked dominant cone-rod degeneration (retinitis pigmentosa-15).

## REFERENCES

1     Cohen D, Chumakov I, Weissenbach J. A first generation physical map of the human genome. Nature 1993;366:698-701.

2     Gyapay G, Morisette J, Vignal A, Dib C, Fizames C, Millasseau P, Marc S, Bernardi G, Lathrop M, Weissenbach J. The 1993-94 Généthon human genetic linkage map. Nature Genet    1994;7:246-9.

3     Schlessinger D, Mandel J-L, Monaco AP, Nelson DL, Willard HF. Report and abstracts of the fourth international workshop on human X chromosome mapping 1993. Cytogenet Cell Genet    1993;64:147-94.

4     Poustka AM, Schlessinger D. Report of the Fifth International Workshop on Human X Chromosome Mapping 1994. Heidelberg, Germany; April 24-27. Cytogenet Cell Genet 1994;67:295-358.

5     McGuire RE, Sullivan LS, Blanton SH, Church MW, Heckenlively JR, Daiger SP.

X-linked dominant cone-rod degeneration: linkage mapping of a new locus for Retinitis Pigmentosa (RP15) to Xp22.13-p22.11. Am J Hum Genet 1995;57:87-94.

6    Bergen AAB, Van Schooneveld MJ, Orth U, Bleeker-Wagemakers EM, Gal A. Multipoint linkage analysis in X-linked juvenile retinoschisis. Clin Gen 1993;43:113-6.

7    Bergen AAB, Ten Brink JB, Bleeker-Wagemakers LM, Van Schooneveld MJ. Refinement of the chromosomal position of the X-linked juvenile retinoschisis disease gene. J Med Genet 1994;31:972-5.

8    Oosterwijk JC, Nelen M, van Zandvoort PM, van Osch LDM, Oranje AP, Wittebol-Post D, Van Oost BA. Linkage analysis of Keratosis Follicularis Spinulosa Decalvans, and regional assignment to human chromosome Xp21.2-p22.2. Am J Hum Genet 1992;50:801-7.

9    Coffey AJ, Roberts RG, Green ED, Cole CG, Butler R, Anand R, Gianelli F, Bentley DR. Construction of a 2.6-Mb contig in yeast artificial chromosomes spanning the human dystrophin gene using an STS-based approach. Genomics 1992;12:474-84.

10   Alitalo T, Francis F, Kere J, Lehrach H, Schlessinger D, Willard HF. A 6-Mb YAC contig in Xp22.1-p22.2 spanning the DXS69E, XE59, GLRA2, PIGA, GRPR, CALB3 and PHKA2 genes. Genomics 1995;25:691-700.

11   Oosterwijk JC, Van der Wielen MJR, Van de Vosse E, Voorhoeve E, Bakker E. Refinement of the localisation of X-linked keratosis follicularis spinulosa decalvans (KFSD) gene in Xp22.1-p22.2. J Med Genet 1995;32:736-9.

12   Anand R, Riley JH, Butler R, Smith JC, Markham AF. A 3.5 genome equivalent multi access YAC library: construction, characterisation, screening and storage. Nucleic Acids Res 1990;18:1951-6.

13   Mohandas TK, Speed RM, Passage MB, Yen PH, Chandley AC, Shapiro LJ. Role of the pseudoautosomal region in sex-chromosome pairing during male meiosis: meiotic studies in a man with a deletion of distal Xp. Am J Hum Genet 1992;51:526-33.

14   Larin Z, Monaco AP, Lehrach H. Yeast artificial chromosome libraries containing large inserts from mouse and human DNA. Proc Natl Acad Sci U S A 1991;88:4123-7.

15   Breukel C, Wijnen J, Tops C, Van der Klift H, Dauwerse H, Meera Khan P. Vector-Alu PCR: a rapid step in mapping cosmids and YACs. Nucleic Acids Res 1990;18:3097

16   Donis-Keller H, Green P, Helms C, Cartinhour S, Weiffenbach B, Stephens K, Keith TP, Bowden DW, Smith DR, Lander ES, *et al*. A genetic linkage map of the human genome. Cell 1987;51:319-37.

17   Wapenaar MC, Petit C, Basler E, Ballabio A, Henke A, Rappold GA, Van Paassen HMB, Blonden LAJ, Van Ommen GJB. Physical mapping of 14 new DNA markers isolated from the human distal Xp region. Genomics 1992;13:167-75.

18   Aldridge J, Kunkel L, Bruns G, Tantravahl U, Lalande M, Brewster T, Moreau E, Wilson M, Bromley W, Roderick T, *et al*. A strategy to reveal high-frequency RFLPs along the human X chromosome. Am J Hum Genet 1984;36:546-64.

19   Stambolian D, Lewis RA, Buetow K, Bond A, Nussbaum R. Nance-Horan Syndrome: localization within the region Xp21.1-Xp22.3 by linkage analysis. Am J Hum Genet 1990;47:13-9.

20   Willard HF, Schmeckpeper BJ, Smith KD, Goss SJ. High resolution regional mapping of the human X chromosome by radiation induced segregation analysis of X-linked DNA probes. Cytogenet Cell Genet 1984;37:609-10.

21    Camerino G, Koenig M, Moisan JP, Oberle I, Morlé F, Jaye M, De la Salle H, Weil D, Hellkuhl B, Grzeschik KH, *et al.* Regional mapping of coagulation factor IX gene and of several unique DNA sequences on the human X chromosome. Cytogenet Cell Genet 1984;37:431-2.

22    Den Dunnen JT, Van Ommen G-JB. ; Mathew CG, editors.Methods in Molecular Biology. Vol.9: Protocols in Human Molecular Genetics. Clifton, New Jersey: Humana Press Inc. 1991; 17, Pulsed-Field Gel Electrophoresis. p. 169-82.

23    Dauwerse JG, Jumelet EA, Wessels JW, Saris JJ, Hagemeijer A, Beverstock GC, Van Ommen GJB. Extensive cross-homology between the long and the short arm of chromosome 16 may explain leukemic inversions and translocations. Blood 1992;79:1299-304.

24    Cole CG, Goodfellow PN, Bobrow M, Bentley DR. Generation of novel sequence tagged sites (STSs) from discrete chromosomal regions using Alu-PCR. Genomics 1991;10:816-26.

25    Holland J, Coffey AJ, Gianelli F, Bentley DR. Vertical integration of cosmid and YAC resources for interval mapping on the X-chromosome. Genomics 1993;15:297-304.

26    Condon GP, Brownstein S, Wang N-S, Kearns AF, Ewing CE. Congenital hereditary (juvenile X-linked) retinoschisis. Arch Ophthalmol 1986;104:576-83.

27    Schaefer L, Ferrero GB, Grillo A, Bassi MT, Roth EJ, Wapenaar MC, Van Ommen GJB, Mohandas TK, Rocchi M, Zoghbi HY, *et al.* A high resolution deletion map of human chromosome Xp22. Nature Genet 1993;4:272-9.

28    Chang Y-PC, Smith KD, Dover GJ. Dinucleotide repeat polymorphisms at the DXS85, DXS16 and DXS43 loci. Hum Mol Genet 1994;3:1029

29    Van de Vosse E, Booms PFM, Vossen RHAM, Wapenaar MC, Van Ommen GJB, Den Dunnen JT. A CA-repeat polymorphism near DXS418 (P122). Hum Mol Genet 1993;2:1202

30    Browne D, Barker D, Litt M. Dinucleotide repeat polymorphisms at the DXS365, DXS443 and DXS451 loci. Hum Mol Genet 1992;1:213

31    Rowe PSN, Goulding J, Read A, Lehrach H, Francis F, Hanauer A, Oudet C, Biancalana V, Kooh SW, Davies KE, *et al.* Refining the genetic map for the region flanking the X-linked hypophosphataemic rickets locus (Xp22.1-22.2). Hum Genet 1994;93:291-4.

32    Chumakov IM, Le Gall I, Billault A, Ougen P, Soularue P, Guillou S, Rigault P, Bui H, De Tand M-F, Barillot E, *et al.* Isolation of chromosome 21-specific yeast artificial chromosomes from a total human genome library. Nature Genet 1992;1:222-5.

33    Lehrach H, Drmanac R, Hoheisel J, *et al.* Davies KE, Tilghman SM, editors.Genome analysis vol 1: Genetic and physical mapping. Cold Spring Harbor: Cold Spring Harbor Laboratory Press, 1990;p. 39-81.

34    Monaco AP, Walker AP, Millwood I, Larin Z, Lehrach H. A yeast artificial chromosome contig containing the complete Duchenne muscular dystrophy gene. Genomics 1992;12:465-73.

35    Baldini A, Ross M, Nizetic D, Vatcheva R, Lindsay EA, Lehrach H, Siniscalco M. Chromosomal assignment of human YAC clones by fluorescence in situ hybridization: use of single-yeast-colony PCR and multiple labeling. Genomics 1992;14:181-4.

36    Wapenaar MC, Kievits T, Meera Khan P, Pearson PL, Van Ommen GJB. Isolation and characterization of cell hybrids containing human Xp-chromosome fragments. Cytogenet

Cell Genet 1990;54:10-4.

37 Wieacker P, Davies KE, Cooke HJ, Pearson PL, Williamson R, Bhattacharya S, Zimmer J, Ropers H-H. Toward a complete linkage map of the human X chromosome: regional assignment of 16 cloned single-copy DNA sequences employing a panel of somatic cell hybrids. Am J Hum Genet 1984;36:265-76.

38 Baldini A, Ward DC. In situ hybridization banding of human chromosomes with Alu-PCR products: a simultaneous karyotype for gene mapping studies. Genomics 1991;9:770-4.

39 Green ED, Riethman HC, Dutchik JE, Olson MV. Detection and characterization of chimeric yeast artificial-chromosome clones. Genomics 1991;11:658-69.

40 Francis F, Rowe PSN, Econs MJ, Gee See C, Benham F, O'Riordan JLH, Drezner MK, Hamvas RMJ, Lehrach H. A YAC contig spanning the hypophosphatemic rickets disease gene (HYP) candidate region. Genomics 1994;21:229-37.

41 Weissenbach J, Gyapay G, Dib C, Vignal A, Morisette J, Millasseau P, Vaysseix G, Lathrop M. A second-generation linkage map of the human genome. Nature 1992;359:794-801.

42 Biancalana V, Trivier E, Weber C, Weissenbach J, Rowe PSN, O'Riordan JLH, Partington MW, Heyberger S, Oudet C, Hanauer A. Construction of a high-resolution linkage map for Xp22.1-p22.2 and refinement of the genetic localization of the Coffin-Lowry syndrome gene. Genomics 1994;22:617-25.

**CHAPTER 2.2**


**High resolution mapping by YAC fragmentation of a 2.5 Mb Xp22 region containing**
**the human RS, KFSD and CLS disease genes**


Van de Vosse, E., Van der Bent, P., Heus, J.J., Van Ommen, G.J.B., and Den Dunnen, J.T.

# High-resolution mapping by YAC fragmentation of a 2.5-Mb Xp22 region containing the human RS, KFSD and CLS disease genes

**Esther Van de Vosse, Paola Van der Bent, Joris J. Heus, Gert-Jan B. Van Ommen, Johan T. Den Dunnen**

MGC-Department of Human Genetics, Leiden University, Wassenaarseweg 72, 2333 AL Leiden, The Netherlands

**Abstract.** The disease loci for X-linked Retinoschisis (RS), Keratosis follicularis spinulosa decalvans (KFSD), and Coffin-Lowry syndrome (CLS) have been localized to the same, small region in Xp22 on the human X Chromosome (Chr). To generate a high-resolution map of the available contig in this area, we have used the YAC fragmentation vectors pBP108/*ADE2* and pBP109/*ADE2* and generated fragmented YACs from a 2.5-Mb YAC (y939H7) spanning the mentioned disease gene candidate regions. Forty-seven fragmented YACs were generated and analyzed, ranging in size from 170 kb to over 2400 kb. The resulting YAC fragmentation panel was used to construct a detailed restriction map of the region and has been used to bin clones and markers. As a deletion panel, it will present a valuable resource for further mapping.

## Introduction

Nearly the complete human genome has been cloned in YAC contigs (Chumakov et al. 1995), and several whole genome maps have been published, based on radiation hybrids (Gyapay et al. 1996), genetic linkage (Dib et al. 1996), or physical data (Chumakov et al. 1995). These maps, however, have insufficient resolution to be directly useful for the next stage: the localization of transcripts. A frequently used subsequent approach to construct more detailed physical maps is to screen genomic libraries of smaller clones (P1, BACs, cosmids) and to order and characterize these. Compared with these bacterial systems, YACs are considered to be difficult to handle, to give low DNA yields, and to contain too large inserts to efficiently construct detailed transcript and restriction maps. However, the large insert does give an attractive starting point for the construction of detailed physical maps, and YAC fragmentation provides an attractive and versatile approach for this task. It is a technique by which the insert DNA of the YAC is reduced in size by homologous recombination directed at repetitive elements in the insert, introducing a new YAC telomere replacing the original one. Existing YAC fragmentation vectors were not suitable for directly fragmenting YACs in AB1380, the standard host strain of most YAC libraries. Recently an improvement of the BP108/109 fragmentation vectors has been made, the addition of the selectable marker *ADE2* (Heus et al. 1996), which greatly facilitates the generation of a fragmented YAC panel.

We are interested in the Xp22.1-p22.2 region, which contains the disease loci for RS (MIM 312700; Van de Vosse et al. 1996); KFSD (MIM 308800; Oosterwijk et al. 1995); CLS (MIM 303600; Bird et al. 1995); and NHS (MIM 302350; Bergen et al. 1994). Several YAC contigs have been published covering parts of this region (Alitalo et al. 1995; Ferrero et al. 1995; Trump et al. 1996; Van de Vosse et al. 1996). YAC y939H7 spans the entire candidate

regions of RS, CLS, and KFSD. We have used this 2.5-Mb YAC to generate a fragmentation panel spanning the entire region and used this to localize cosmids and markers and to construct a long-range restriction map.

## Materials and methods

*YAC fragmentation.* y939H7, propagated in *S. cerevisiae* strain AB1380 [*MATa, ura3-52, trp1, ade2, his5, lys2, ile, thr, can1-100* (Cohen et al. 1993)], was cultured overnight in 100 ml YPD [2% glucose, 2% bacto peptone, 1% yeast extract (Riley et al. 1992)] at 30°C to $2 \times 10^6$–$2 \times 10^7$ cells/ml. Yeast cells were transformed using an alkali cation yeast transformation kit (BIO 101) with 5 µg *Sal*I-digested pBP108/*ADE2* or pBP109/*ADE2* plasmid (Heus et al. 1996). Transformants were plated on medium lacking adenine and tryptophan. Loss of the URA3-arm of pYAC4 was checked by replica plating the colonies on medium lacking uracil, adenine, and tryptophan.

*Analysis of YACs.* Plugs containing yeast DNA were prepared as described (Den Dunnen and Van Ommen 1991) in low melting temperature agarose (Seaplaque, FMC). YACs were analyzed directly on pulsed-field gel (1% Seakem LE in 45 mM Tris, 45 mM boric acid, 0.5 mM EDTA, pH 8.3) at 180 Volt, 30 h, in 4 cycles with a switch time increase from 40 to 70 s, or at 95 Volt, 48 h, in 12 cycles with a switch time increase from 50 to 550 s.

Yeast DNA in plugs was digested overnight with 15 Units *Sfi*I, *Nru*I, *Not*I, or *Bss*HII, under the conditions recommended by the manufacturers, and analyzed by pulsed-field gel electrophoresis (PFGE) at 160 Volt, 20 h, in 4 cycles with a switch time increase from 20 to 40 s. Gels were Southern blotted on Hybond-N⁺ (Amersham) as described (Den Dunnen and Van Ommen 1991). Hybridization of Southern blots was performed in 0.5 M Na₂HPO₄, pH7.4, 7% SDS, 1 mM EDTA at 65°C. Probes were labeled with [³²P]dCTP and a multiprime labeling kit (Amersham); membranes were washed at 65°C, twice with 2*SSC,0.1% SDS, twice with 1*SSC, 0.1% SDS, and twice with 0.3*SSC,0.1% SDS. Autoradiography was performed with a PhosphorImager (Molecular Dynamics). Probes used include pBlur8, *ADE2*, pBR322 *Pvu*I/*Bam*HI-fragments (to identify the pYAC4-derived vector arms) and specific probes from the region (RDGC, PHKA2, P122).

*Marker content analysis.* PCR was carried out in a Perkin-Elmer Cetus thermal cycler; 1 µl of an agarose plug was used in a 30-µl PCR reaction (75 mM Tris-HCl, 20 mM (NH₄)₂SO₄, 0.01% Tween 20, 1.5 mM MgCl₂, 200 µM dNTPs, 0.25 units Goldstar polymerase, 100 ng/µl BSA, 25 pmol of each primer) with 50 µl mineral oil overlay. Cycles; 5 min. 94°C, followed by 30 cycles of 1 min, 94°C, 1 min, $T_a$ (as recommended), 1 min. 72°C (Table 1). Products were separated on a 2% agarose gel.

*Construction and analysis of sCOGH2-cosmids.* sCOGH2 (Datson et al. 1996b) was digested with *Xba*I, dephosphorylated and digested with *Bam*HI. DNA plugs of y939H7 were *Mbo*I digested, size fractionated, and ligated in the *Bam*HI site of sCOGH2. The ligated material was packaged with Gigapack II Plus Packaging Extract (Stratagene) and transfected into

**Table 1.** Primers used in the analysis of the fragmentation YACs. T$_a$ = annealing temperature; product lengths are given in base pairs.

| Locus | Name | T$_a$ Reference | Product length (bp) |
|---|---|---|---|
| *DXS1317* | sWXD170 | 50 Kere et al. 1992 | 67 |
| *DXS1195* | AFM207zd6 | 56 Dib et al. 1996 | 235–239 |
| *DXS418* | P122 | 56 Van de Vosse et al. 1993 | 142–158 |
| *DXS8019* | AFMa152xf1 | 50 Dib et al. 1996 | 156–174 |
| | HYAT1 | 58 Huopaniemi et al. (in press) | 160–186 |
| *DXS6762* | 12691/12692 | 50 Ferrero et al. 1995 | 114 |
| *DXS7593* | AFMa346zc1 | 50 Dib et al. 1996 | 209–231 |
| *DXS6760* | 12655/12656 | 50 Ferrero et al. 1995 | 144 |
| *DXS7176* | GM11 | 55 Ferrero et al. 1995 | 224 |
| *DXS999* | AFM234yf12 | 50 Schaefer et al. 1993 | 260–276 |
| *DXS7161* | AFM291wf5 | 55 Dib et al. 1996 | 240–254 |
| | PDHA1 | 58 Schaefer et al. 1993 | 125 |
| *DXS7177* | TG2-675 | 50 Ferrero et al. 1995 | 192 |
| *DXS443* | pRX-324 | 50 Schaefer et al. 1993 | 204–210 |
| *DXS365* | RX-314 | 50 Browne et al. 1992 | 201–217 |

*E. coli* 1046. Clones containing a human insert were selected by hybridization to pBlur8 (Deininger et al. 1993), total human DNA, and *Alu*-PCR products of YACs from the region. Cosmid DNA was isolated as described (Maniatis et al. 1989). *Eco*RI-digested cosmid DNA was analyzed on 0.8% agarose gels to determine insert sizes. Binning of the cosmids was done by hybridization to Southern blots of the YAC fragmentation panel.

## Results

Of YAC y939H7, various sizes between 2300 kb and 300 kb have been reported, suggesting instability of the clone leading to differently sized isolates (Chumakov et al. 1995; Alitalo et al. 1995; Van de Vosse et al. 1996; Ferrero et al. 1995). We have analyzed 24 individual y939H7 colonies from the original CEPH library and obtained one clone that appeared to contain the complete region, that is, 36 probes tested were positive. The YAC in this clone has a reproducible length of 2500 kb and is stable during propagation.

To construct an overlapping panel of YACs with smaller sizes, we performed a YAC fragmentation experiment. Over 200 transformants were obtained after transformation with vectors pBP108/*ADE2* and pBP109/*ADE2* respectively. Upon recombination, these vectors exchange the pYAC4-derived *URA3*-arm with a new pBP108/109-derived arm containing *ADE2*. Selection on plates

lacking uracil, tryptophan, and adenine showed that 90% of the YAC transformants had the correct phenotype, that is, containing *ADE2* and *TRP1* and lacking *URA3*. Forty-seven YACs were analyzed further by PFGE: 46 of these contained a fragmented YAC (98%). The fragmentation-YACs varied in size from over 2400 to 170 kb (Fig. 1). To verify the STS content of the YACs, the panel was hybridized with specific probes and subjected to PCR with specific primers (Fig. 2a).

To obtain a detailed physical map, DNA of the fragmentation YACs was digested with *Sfi*I, *Nru*I, *Not*I, or *Bss*HII and electrophoresed on pulsed-field gels. Southern blots were hybridized with an *ADE2* probe to identify the end fragment that contains the fragmentation vector (Fig. 3a) and with a pBR322 *Pvu*I/*Bam*HI-fragment to identify the opposite pYAC4-derived centromeric vector arm. In parallel, pBlur8 hybridization (Deininger et al. 1993) was used to identify the human DNA-containing fragments (Fig. 3b). As an example of this double analysis, we have indicated in Fig. 3a and 3b a 170-kb *Bss*HII fragment (arrows). This fragment appears as an end fragment, recombined with the fragmentation vector, in YACs 4, 5, and 6 by hybridization to both *Alu* and *ADE2*. Moreover, the full-sized fragment is present in fragmentation YACs 7 and higher and hybridizes to *Alu* but no longer to *ADE2*. Other examples of newly appearing end fragments (Fig. 3a) turning into full-size fragments (Fig. 3b) are visible for YACs 1–3



- 2300 kb
- 1250 kb
- 940/970 kb
- 820 kb
- 730 kb
- 680 kb
- 600 kb
- 440 kb
- 360 kb
- 280 kb
- 230 kb
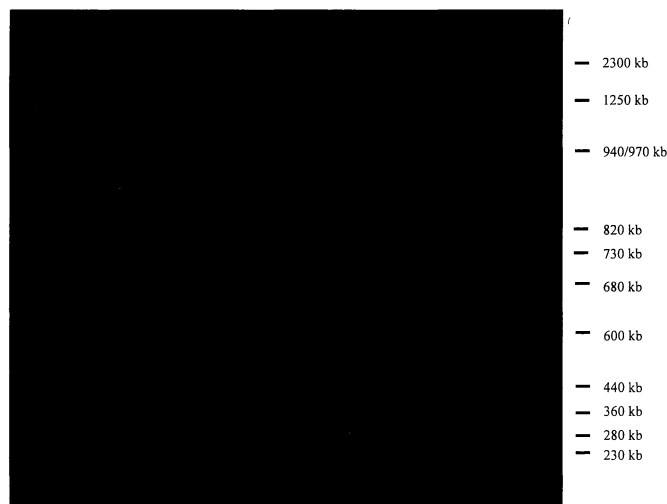
**Fig. 1.** Southern blot of pulsed-field gel (180 Volt, 30 h, 40–70 s, 4 cycles) of fragmented YACs hybridized with pBlur8 to determine the size of the YACs. The sizes of the AB1380 yeast chromosomes, indicated on the right (in kb), were used as size standards.
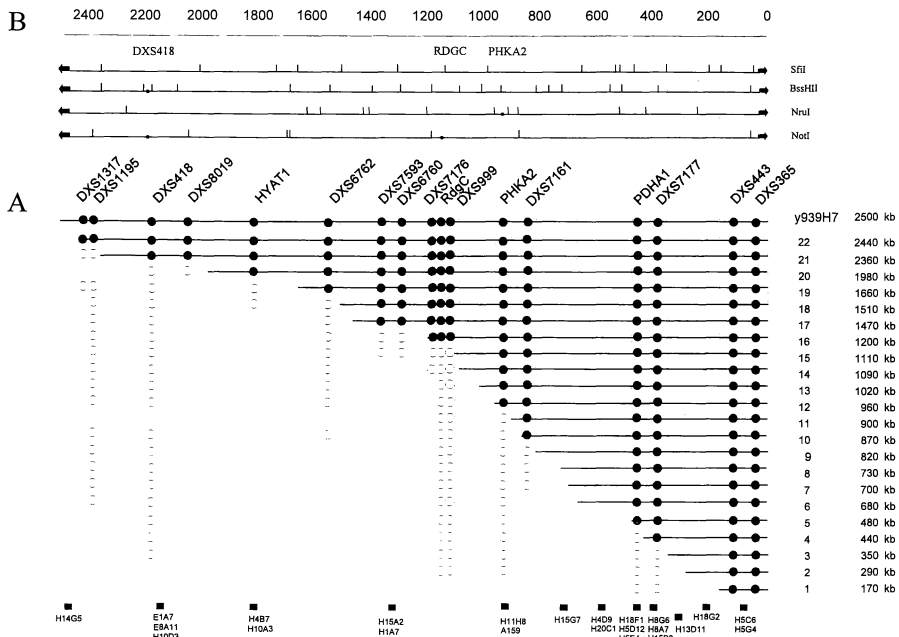
**Fig. 2. A.** Physical map of y939H7. The length of each YAC (in kb) is indicated on the right. ● depicts a positive result in a marker analysis, ○ depicts a negative result in a marker analysis. ■ indicates a positive hybridization result of a cosmid. **B.** *Bss*HII, *Sfi*I, *Not*I, and *Nru*I restriction maps of y939H7. The arrowhead indicates the YAC vector arm. All sizes indicated are in kb. The boundaries of the bins are indicated by dotted vertical lines.
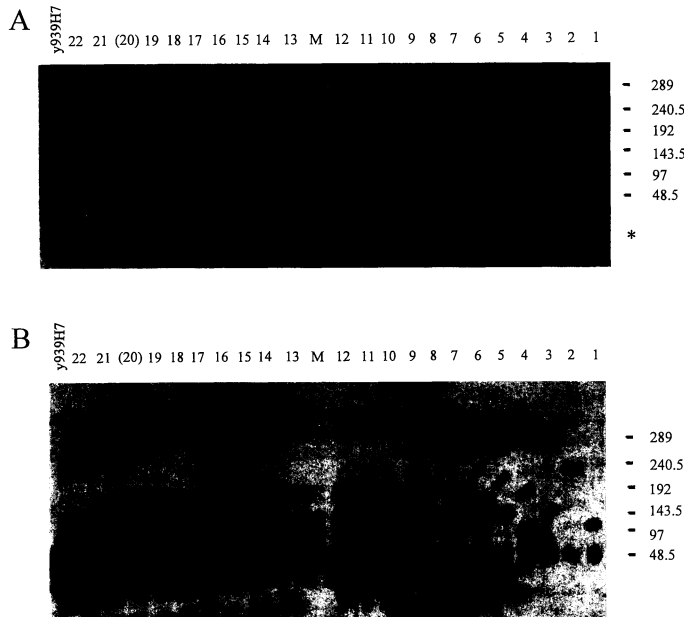


**Fig. 3.** *Bss*HII-digested fragmentation YACs hybridized with *ADE2* (panel A) or pBlur8 (panel B). **A.** The constant band marked with an asterisk is the fragment containing the yeast *ADE* gene. The variable band is the fragment containing the new vector arm. Markers used are bacteriophage lambda DNA PFGE marker (Pharmacia) and a mixture of lambda DNA digested with *Hind*III and *Pst*I. All sizes indicated are in kb. YAC 20 was not fully digested.

85

and 7–22. This parallel mapping yields a 2.5-Mb restriction map of y939H7 in one single, easy-to-interpret experiment (Fig. 2b).

To obtain a cosmid clone resource facilitating further detailed analysis of the region, we have subcloned y939H7 in cosmid vector sCOGH2. Approximately 2000 cosmid clones were picked, gridded, and screened with an *Alu* probe and *Alu*-PCR products from y939H7; 100 positive clones were isolated and analyzed further. The localization of 25 cosmids was determined by hybridization to the YAC fragmentation panel (Fig. 2a).

## Discussion

Although published in 1990 (Pavan et al. 1990) as an attractive tool for detailed physical analysis of YACs, YAC-fragmentation has not yet fulfilled its promise by becoming a frequently applied research tool. The main reason for this is that the 'first generation' YAC fragmentation vectors contain *HIS3* as a selectable marker. This gene is intact in AB1380, the standard host strain of most YAC libraries. The *HIS3* gene could, therefore, not be used directly as a selectable marker in AB1380. Transferring YACs from AB1380 to a suitable host strain to select for YAC fragmentation [for example, YPH925; *MATα ura3-52 trp1Δ63 ade2-101 cyh2^R lys2-801 his3Δ200 leu2Δ1 kar1Δ15;* (Spencer et al. 1994)] is a tedious procedure. The YAC fragmentation vectors we have used, pBP108/*ADE2* and pBP109/*ADE2,* obviate this tedious step. Presently, in our hands, YAC fragmentation has become a fast and simple method for mapping large YACs, of up to 2.5 Mb.

The YAC fragmentation panel we have constructed from y939H7 contains YACs that range in size from 170 kb to 2440 kb (Fig. 1). The panel was used to map specific markers and cosmid clones and to construct a long-range physical map for *Not*I, *Bss*HII, *Sfi*I and *Nru*I (Fig. 2). No aberrant bands were observed, which indicates that the fragmentation YACs are stable and that the fragmentation panel is a faithful representation of the insert in y939H7. In parallel reports, YAC fragmentation panels have been employed for high-resolution bin mapping panels (Potier et al. 1996; Datson et al. 1996a), but they have not previously been used to construct restriction maps of large YACs. The typical problems with the construction of restriction maps of large YACs with partial digests—for example, limited resolution, some sites of the YAC may be difficult to digest, interpretation difficulties of the complex hybridization patterns—are obviated by the approach we have used. This is mainly because each datapoint is sampled in multiple parallel lanes, providing great robustness by data redundancy. Furthermore, a fragmentation panel gives a more detailed map of the region, since it adds resolution within a restriction fragment when this fragment has been targeted by the fragmentation vector, thereby reducing its size.

It is not clear why only one out of 24 isolates of y939H7 was stable in size; it might contain a stabilizing mutation. However, we have never observed any difference between the genomic map and the y939H7-derived map, indicating that, if present, this mutation is rather small. The order of markers as determined in our YAC fragmentation panel is in agreement with those published by others (Alitalo et al. 1995; Ferrero et al. 1995; Trump et al. 1996). The only discrepancy is the reversed order of the markers *DXS418–DXS1195–DXS1317* relative to Ferrero's map. This order in that map, however, could potentially be inverted without creating discrepancies in their contig. We were able to localize two markers that have not been localized before: *HYAT1* and *DXS8019*. One marker, *DXS7593,* which was localized between *DXS7101* and *DXS1052* previously (Van de Vosse et al. 1996) owing to a deletion in y900E08102 and a previous, unstable isolate of y939H7, could now be positioned between *DXS6762* and *DXS6760*. Because of the data presented here, we conclude that the size of the RS candidate region measures 1 Mb.

86

## References

Alitalo T, Francis F, Kere J, Lehrach H, Schlessinger D, Willard HF (1995) A 6-Mb YAC contig in Xp22.1-p22.2 spanning the DXS69E, XE59, GLRA2, PIGA, GRPR, CALB3 and PHKA2 genes. Genomics 25, 691–700

Bergen AA, Ten Brink J, Schuurman EJ, Bleeker-Wagemakers EM (1994) Nance-Horan syndrome: linkage analysis in a family from the Netherlands. Genomics 21, 238–240

Bird H, Collins AL, Oley C, Lindsay S (1995) Crossover analysis in a British family suggests that Coffin-Lowry syndrome maps to a 3.4-cM interval in Xp22. Am J Med Genet 59, 512–516

Browne D, Barker D, Litt M. (1992) Dinucleotide repeat polymorphisms at the DXS365, DXS443 and DXS451 loci. Hum Mol Genet 1, 213

Chumakov IM, Rigault P, Le Gall I, Bellanné-Chantelot C, Billault A, Guillou S, Soularue P, Guasconi G, Poullier E, Gros I, Belova M, Sambucy JL, Susini L, Gervy P, Glibert F, Beaufils S, Bui H, Massart C, De Tand MF, Dukasz F, Lecoulant S, Ougen P, Perrot V, Saumler M, (1995) A YAC contig map of the human genome. Nature 377 Suppl, 175–183

Cohen D, Chumakov I, Weissenbach J. (1993) A first generation physical map of the human genome. Nature 366, 698–701

Datson NA, Semina E, Van Staalduinen AAA, Dauwerse HG, Meershoek EJ, Heus JJ, Frants RR, Den Dunnen JT, Murray JC, Van Ommen GJB (1996a) Closing in on the Rieger syndrome gene on 4q25: mapping translocation breakpoints within a 50-kb region. Am J Hum Genet 59, 1297–1305

Datson NA, Van de Vosse E, Dauwerse HG, Bout M, Van Ommen GJB, Den Dunnen JT (1996b) Scanning for genes in large genomic regions: cosmid-based exon trapping of multiple exons in a single product. Nucleic Acids Res. 24, 1105–1111

Deininger PL, Jolly DJ, Rubin CM, Friedmann T, Schmid CW (1993) Base sequence studies of 300 nucleotide renatured repeated human DNA clones. J Mol Biol 151, 17–33

Den Dunnen JT, Van Ommen G-JB (1991) Pulsed-field gel electrophoresis. *Methods Mol Biol* 9, 169–182

Dib C, Faure S, Fizames C, Samson D, Drouot N, Vignal A, Millasseau P, Marc S, Hazan J, Seboun E, Gyapay G, Morissette J, Weissenbach J (1996) A comprehensive genetic map of the human genome based on 5,264 microsatellites. Nature 380, 152–154

Ferrero GB, Franco B, Roth EJ, Firulli BA, Borsani G, Delmas-Mata J, Weissenbach J, Halley G, Schlessinger D, Chinault AC, Zoghbi HY, Nelson DL, Ballabio A (1995) An integrated physical and genetic map of a 35 Mb region on chromosome Xp22.3-Xp21.3. Hum Mol Genet 4, 1821–1827

Gyapay G, Schmitt K, Fizames C, Jones H, Vega-Czarny N, Spillett D, Muselet D, Prud'Homme J-F, Dib C, Auffray C, Morisette J, Weissenbach J, Goodfellow P (1996) A radiation hybrid map of the human genome. Hum Mol Genet 5, 339–346

Heus JJ, De Winther MPJ, Van de Vosse E, Van Ommen GJB, Den Dunnen JT (1996) Centromeric and non-centromeric ADE2-selectable fragmentation vectors for YACs in AB1380. Genome Research (in press)

Huopaniemi L, Rantala A, Tahvanainen E, De la Chapelle A, Alitalo T (1997) Linkage disequilibrium and physical mapping of X-linked juvenile retinoschisis. Am J Hum Genet, in press.

Kere J, Nagaraja R, Mumm S, Ciccodicola A, D'Urso M, Schlessinger D (1992) Mapping human chromosomes by walking with Sequence-Tagged Sites from end fragments of Yeast Artificial Chromosome inserts. Genomics 14, 241–248

Maniatis T, Fritsch E, Sambrook J (1989) *Molecular Cloning: A Laboratory Manual.* (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press)

Oosterwijk JC, Van der Wielen MJR, Van de Vosse E, Voorhoeve E, Bakker E (1995) Refinement of the localisation of X-linked keratosis follicularis spinulosa decalvans (KFSD) gene in Xp22.1-p22.2. J Med Genet 32, 736–739

Pavan WJ, Hieter P, Reeves RH (1990) Generation of deletion derivatives by targeted transformation of human-derived yeast artificial chromosomes. Proc Natl Acad Sci USA. 87, 1300–1304

Potier M-C, Dutriaux A, Reeves R (1996) Use of YAC fragmentation to delimit a duplicated region on human chromosome 21. Mamm Genome 7, 85–88

Riley JH, Ogilvie D, Anand R (1992) Construction, characterization and

screening of YAC libraries. In: *Techniques for the Analysis of Complex Genomes,* R. Anand, ed. (London: Academic Press) pp 59–79

Schaefer L, Ferrero GB Grillo A, Bassi MT, Roth EJ, Wapenaar MC, Van Ommen GJB, Mohandas TK, Rocchi M, Zoghbi HY, Ballabio A (1993) A high resolution deletion map of human chromosome Xp22. Nature Genet 4, 272–279

Spencer F, Hugerat Y, Simchen G, Hurko O, Connelly C, Hieter P (1994) Yeast karl mutants provide an effective method for YAC transfer to new hosts. Genomics 22, 118–126

Trump D, Pilia G, Dixon PH, Wooding C, Thakrar R, Leigh SEA, Nagaraja R, Whyte MP, Schlessinger D, Thakker RV (1996) Construction of a YAC contig and an STS map spanning 3.6 megabase pairs in Xp22.1 Hum Genet 97, 60–68

Van de Vosse E, Booms PFM, Vossen RHAM, Wapenaar MC, Van Ommen GJB, Den Dunnen JT (1993) A CA-repeat polymorphism near DXS418 (P122). Hum Mol Genet 2, 1202

Van de Vosse E, Bergen AAB, Meershoek EJ, Oosterwijk JC, Gregory S, Bakker B, Weissenbach J, Coffey AJ, Van Ommen GJB, Den Dunnen JT (1996) A Xp22.1-p22.2 YAC contig encompassing the disease loci for RS, KFSD, CLS, HYP and RP15; refined localization of RS. Eur J Hum Genet 4, 101–104

# APPENDIX TO CHAPTER 2

## A CA-repeat polymorphism near DXS418 (P122)

Van de Vosse, E., Booms, P.F.M., Vossen, R.H.A.M., Wapenaar, M.C., Van Ommen, G.J.B., and Den Dunnen, J.T.

# A CA-repeat polymorphism near DXS418 (P122)

E. Van de Vosse, P.F.M. Booms, R.H.A.M. Vossen, M.C. Wapenaar, G.J.B. Van Ommen and J.T. Den Dunnen

Department of Human Genetics, Leiden University, Wassenaarseweg 72, 2333 AL Leiden, The Netherlands

*Source/Description*: A human genomic cosmid library was screened with probe P122 (DXS418) (1). Cosmid cHP122.4 contained a CA-repeat which was isolated and sequenced (2). The predicted length of the amplified fragment was 162 bp and contained a polymorphic $(GC)_6CGG(CA)_{31}$ repeat.

*Primer sequences:*
5'-TGTGAGGTTTTGTTCCCTCC-3'  (AC strand)
5'-GACTGTTGAGTTTCCTCACAGC-3'  (GT strand)

*Allele frequency*: Estimated from 52 unrelated female individuals. The observed heterozygosity was 0.83. The 162 bp allele present in cosmid cHP122.4 was not present in the females analysed.

| Allele | (bp) | Frequency | Allele | (bp) | Frequency |
|--------|------|-----------|--------|------|-----------|
| A1 | 158 | .03 | A7 | 148 | .21 |
| A2 | 157 | .01 | A8 | 146 | .18 |
| A3 | 156 | .03 | A9 | 144 | .05 |
| A4 | 154 | .08 | A10 | 142 | .03 |
| A5 | 152 | .13 | A11 | 140 | .05 |
| A6 | 150 | .20 | | | |

CEPH: 1329 202 (152, 150), 1329 311 (152, 140).

*Chromosomal localization*: Locus DXS418, probe P122, has been localized to Xp22.1 between patient deletion breakpoints BA236 and BA162/BA140. This region also contains DXS9, DXS43, DXS69, DXS197, DXS207, DXS414 and the CEPH marker DXS987 (3).

*Mendelian Inheritance*: Co-dominant segregation was observed in two informative families.

*PCR Conditions*: PCR amplification was performed in 15 µl reaction volume containing 10 mM Tris-HCl (pH 9.0), 50 mM Kcl, 0.01% gelatine, 1.5 mM $MgCl_2$, 0.1 % Triton X-100, 0.06 U Supertaq (HT Biotechnology Ltd.), 200 µM each dATP, dGTP and dTTP, 2.5 µM dCTP, 0.75 µCi $\alpha^{32}PdCTP$ (at 3000 Ci/mmol), 80 ng of each primer and 200 ng genomic DNA template. 28 cycles of denaturation at 94°C for 30 sec, annealing at 58°C for 40 sec and extension at 72°C for 1 min were carried out with an initial denaturation step of 5 min at 94°C and a final extension step of 5 min at 72°C. The amplified products were run on a 6% polyacrylamide/bisacrylamide (19:1), 7 M urea sequencing gel at 40 mA and 65 Watts for approximately 3 hours with 0.5 x TBE-buffer in the upper buffer compartment and 5/6 x TBE/0.5 M NaAc in the lower compartment.

*References*: 1) Wapenaar, M.C., Petit, C., Basler, E., *et al.* (1992) *Genomics* **13**, 167-175. 2) Booms, P., Wapenaar, M.C., Van Ommen, G.J.B., and Den Dunnen, J.T. (1993) EMBL DNA sequence data base, Z22550. 3) Schaefer, L., Ferrero, G.B., Grillo, A. *et al.* (1993) *Nature Genet.* **4**, 272-279.

# CHAPTER 3

**Scanning for genes in large genomic regions: cosmid-based exon trapping of multiple exons in a single product**

Datson, N.A., Van de Vosse, E., Dauwerse, H.G., Bout, M., Van Ommen, G.J.B., and Den Dunnen, J.T.

# Scanning for genes in large genomic regions: cosmid-based exon trapping of multiple exons in a single product

Nicole A. Datson, Esther van de Vosse, Hans G. Dauwerse, Mattie Bout, Gert-Jan B. van Ommen and Johan T. den Dunnen*

MGC-Department of Human Genetics, Leiden University, Wassenaarseweg 72, 2333 Al Leiden, The Netherlands

## ABSTRACT

**To facilitate the scanning of large genomic regions for the presence of exonic gene segments we have constructed a cosmid-based exon trap vector. The vector serves a dual purpose since it is also suitable for contig construction and physical mapping. The exon trap cassette of vector sCOGH1 consists of the human growth hormone gene driven by the mouse metallothionein-1 promoter. Inserts are cloned in the multicloning site located in intron 2 of the hGH gene. The efficiency of the system is demonstrated with cosmids containing multiple exons of the Duchenne Muscular Dystrophy gene. All exons present in the inserts were successfully retrieved and no cryptic products were detected. Up to seven exons were isolated simultaneously in a single spliced product. The system has been extended by a transcription–translation-test protocol to determine the presence of large open reading frames in the trapped products, using a combination of tailed PCR primers directing protein synthesis in three different reading frames, followed by *in vitro* transcription–translation. Having larger stretches of coding sequence in a single exon trap product rather than small single exons greatly facilitates further analysis of potential genes and offers new possibilities for direct mutation analysis of exon trap material.**

## INTRODUCTION

Exon trapping has become a widely used method which is generally acknowledged as a versatile tissue-independent approach to detect genes in cloned DNA. In contrast to RNA-based methods, such as cDNA selection and direct screening of cDNA libraries, exon trapping is independent of tissue-specific gene expression. It uses cloned DNA directly to select sequences surrounded by functional splice sites (1–3). Original exon trapping protocols have been improved with respect to speed and efficiency and improvements have been made to reduce the background consisting of cryptically spliced

products and products arising from vector–vector splicing (4). However, some of the limitations of the original systems have remained unaddressed.

A major limitation of current systems is the need for subcloning of the region of interest in a vector with a capacity for inserts typically measuring 1–2 kb. This has several consequences: (i) due to the small insert size after subcloning, multiple exons will only rarely be present in one insert, resulting in exon trap clones containing only a single exon. Consequently, many of the exon trap probes derived are small (~80–150 bp) and frequently give poor signals or a high signal-to-noise ratio in subsequent experiments, e.g. the screening of cDNA-libraries or probing of Northern blots. Furthermore, since the individually trapped exons require the use of cDNA libraries in the next step to further define the gene, the initial advantage of working with an expression independent system is to a large extent lost in the subsequent step. (ii) Due to subcloning into plasmid-based exon trap vectors the gene(s) present are scattered into many separate, disconnected pieces. Any exons thus obtained have to be aligned to reconstruct their original order. Reconstruction of the gene from individually trapped exons requires a significant amount of time and effort and implies a major loss of information originally contained within the input material prior to subcloning. (iii) Subcloning disrupts the genomic context around the exons. Cloning of regions which are never transcribed or of intronic sequences without their naturally flanking exons often results in activation of cryptic splice sites, leading to recognition of false exons and a background of false positives. On the other hand, genuine exons will be missed due to poor recognition of the host system or due to unfavourable factors resulting from the cloning (e.g. spacing of restriction sites). (iv) Current exon trapping systems can only be used in combination with specific cell lines (e.g. COS cells), since they require a system of replication in the host cell, commonly based on the SV40 origin of replication (2). It is imaginable that some exons of genes with a highly tissue-specific expression pattern will not be included in the mature transcripts generated in a completely different cell type (5).

Although the 3′ exon trapping recently described (6,7) has some advantages in that it allows larger exons to be trapped, specifically identifies the end of a gene, and selects exons based

on two independent criteria i.e. splicing and polyadenylation; it does not, however, address the other limitations of small-insert exon trapping.

We have designed a large-insert exon trapping vector capable of scanning 25–40 kb genomic regions for exons. The vector has a dual use: as cosmid vector for contig construction and physical mapping, and as exon trap vector for isolation of coding sequences. In the vector, inserts are cloned into intron 2 of the human growth hormone gene (hGH) and transcription is driven by a mouse metallothionein-1 promoter (mMT-1). This is a strong, ubiquitously expressed promoter which allows many different cell types to be used, thus obviating the restriction to COS cells applying to the SV40-based systems used so far. During exon trapping the genomic context is maintained over the entire 25–40 kb region, reducing the false positive rate while yielding processed transcripts with multiple exons spliced together in the correct order. The efficiency of the system is demonstrated using cosmids containing up to seven exons of the duchenne muscular dystrophy gene (DMD). We believe that the system should greatly increase the speed and reliability of gene isolation by exon trapping by offering a solution for most major limitations of current exon trapping systems.

## MATERIALS AND METHODS

### Vector construction

sCOGH1, schematically drawn in Figure 1, was constructed as follows: cosmid vector sCos1(8) was digested with *Eco*RI and the 7.9 kb vector fragment was separated from the *Eco*RI linker by agarose gel electrophoresis and elution. Similarly, plasmid pXGH5(9) was digested with *Eco*RI and the 4 kb fragment containing the mouse metallothionein-1 promoter (mMT-1) and the human growth hormone gene (hGH) was isolated by gel-purification. Both fragments were combined by ligation, resulting in the isolation of sCOGH0a and sCOGH0b, differing in the orientation of the mMT1/hGH-insert in sCos1. Subsequently a linker composed of two complementary oligonucleotides (5′-AGCGGCCGCGAATTCGGATCCGGCGGCCGC-3′ and 5′-CTGCGGCCGCCGGATCCGAATTCGCGGCCG-3′) was synthesized containing *Not*I, *Bam*HI and *Eco*RI sites as well as *Acc*I sticky ends, and introduced into intron 2 of the hGH gene by digestion of sCOGH0b with *Acc*I and ligation. The resulting vector was designated sCOGH1.

Subsequently specific variants of sCOGH1 were constructed (Fig. 1). To facilitate screening of human positive cosmids sCOGH2, a sCOGH1 variant without the *Alu* sequence in the 3′-UTR of hGH was constructed. Three sCOGH1 fragments were ligated: fragment 1 was produced by digesting sCOGH1 with *Eco*RI. The sticky ends were filled in using T4 DNA polymerase. After *Cla*I digestion the 7.9 kb vector fragment was isolated from a gel. Fragment 2 was the gel-purified 2.4 kb *Cla*I–*Bam*HI fragment of sCOGH1 containing the 5′ part of hGH. To obtain the third fragment the 1.6 kb *Eco*RI fragment of sCOGH1, containing the 3′ part of hGH, was first subcloned in pGEM7zf (Promega). The resulting clone, pGHE1.6, was digested with *Bam*HI and *Ssp*I and the 0.9 kb 3′ hGH fragment was gel purified. Combination of the three fragments resulted in sCOGH2.

sCOGH3 is a promoterless variant of sCOGH1, constructed for trapping promoter/first exon regions of tissue-specific genes. sCOGH2 was digested with *Eco*RI and ligated, resulting in

96
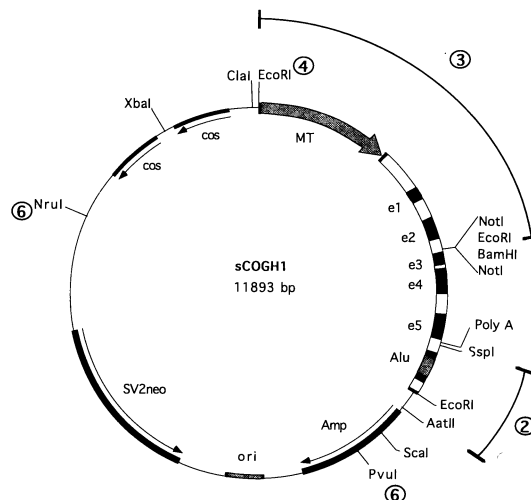


**Figure 1.** Vector sCOGH1 (11893 bp) contains the entire hGH exon trap cassette driven by a mMT-1 promoter. The MCS located in intron 2 of hGH contains sites for *Eco*RI and *Bam*HI flanked by two *Not*I sites, which can be used to release the cosmid insert. Several variants of sCOGH1 were constructed, marked in the outer ring of sCOGH1 by a corresponding number. In sCOGH2, sCOGH3, sCOGH4 and sCOGH6 the *Alu* repeat in the 3′-UTR of hGH has been removed (2). sCOGH3 lacks the mMT-1 promoter and exons 1 and 2 of hGH (3). sCOGH4 and sCOGH6 lack the *Eco*RI site at the 5′ end of hGH (4). In sCOGH6 the SVneo selectable marker has been removed between the *Pvu*I and *Nru*I sites and has been replaced by a fragment from pUC19 (6).

sCOGH3 lacking the 2.4 kb 5′ hGH fragment. A variant lacking the *Eco*RI site at the 5′ end of the mMT-1 promoter, sCOGH4, was constructed to allow use of the *Eco*RI site in the polylinker for cloning of cosmid inserts. sCOGH2 was partially digested with *Eco*RI followed by filling in of the ends with T4 DNA polymerase. The linear band of 11.3 kb was isolated from a gel and ligated, resulting in sCOGH4. In another variant, sCOGH6, the SV2neo marker was removed to allow cloning of larger cosmid inserts. For this purpose sCOGH4 was digested with *Nru*I and *Pvu*I, the 6.6 kb vector fragment was isolated from gel and ligated to the 1.3 kb *Pvu*I–*Pvu*II fragment containing the ampicillin resistance gene and *E.coli* origin of replication of pUC19 which was gel purified.

### Construction of cosmid libraries in sCOGH1

All sCOGH-derivatives were propagated in *E.coli* strain HB10B (kindly provided by Pieter de Jong). For cosmid cloning, vector DNA was linearised with *Xba*I, dephosphorylated and subsequently digested with *Bam*HI. Agarose plugs containing genomic yeast DNA and YAC DNA of yDMD(0-25)C, containing the human DMD-gene from 100 kb upstream of the brain exon 1 to 100 kb downstream of exon 79 (10), were partially digested with *Mbo*I, size fractionated and ligated into the *Bam*HI-site of sCOGH1. The ligated material was packaged using Gigapack II Plus Packaging Extract (Stratagene) and used to infect *E.coli* 1046. Cosmids containing specific regions of the DMD-gene were isolated and analysed using standard protocols

(11) by hybridization with specific DMD cDNA sequences (10). The exon content was established by PCR with exon primers and by hybridisation of the *Hind*III-digested cosmids with the DMD cDNA. The inserts of screened cosmids were reversed by *Not*I-excision of the insert, religation and transformation to *E.coli*. The orientation of the insert was determined by restriction digestion.

### Cell culture and electroporation

Initially COS-1 cells were used for transfection experiments. We found, however, that exon trapping results strongly improved using hamster V79 cells. Higher yields were obtained of full length PCR fragments. Therefore later experiments were performed with this cell line. We explain the improvement by a lower degree of homology between the hGH-primers and the endogenous hamster growth hormone gene compared to the corresponding sequences in COS-1 cells. The cells were cultured in DMEM with 10% inactivated fetal calf serum (Gibco-BRL).

Cosmid DNA was introduced by electroporation: actively growing cells were collected by centrifugation, washed in cold PBS (without bivalent cations) and resuspended in cold PBS at a density of $2 \times 10^7$ cells/ml. Cell suspension (0.5 ml) was added to 20 µl of PBS containing 10 µg cesium-chloride purified cosmid DNA and placed in a pre-chilled electroporation cuvette (0.4 cm chamber, BioRad). After 5 min on ice, the cells were electroporated in a BioRad Gene Pulser [300 V (750 V/cm); 960 µF], and placed on ice again. After 5 min the cells were transferred gently to a 100 mm tissue culture dish containing 10 ml of pre-warmed, equilibrated DMEM + 10% FCS. Transfection efficiency was monitored by assaying the hGH concentration in 100 µl of the culture medium of cells transfected in parallel with pXGH5 using the Allégro hGH Transient Gene Assay kit (Nichols Institute, San Juan Capistrano, USA) (9).

### RNA isolation, RNA-PCR and product analysis

48–72 h after transfection, the cells were harvested and total RNA was isolated using RNazolB (CINNA/BIOTECX). First-strand cDNA synthesis was performed by adding 50 pmol of primer hGHf to 2 µg total RNA in a volume of 16 µl. The mixture was incubated at 65°C for 10 min and chilled on ice. 14 µl of a mix containing 3 µl 0.1 M DTT, 3 µl 10 mM dNTPs, 0.5 µl RNasin (40 U/µl; Promega), 6 µl 5× RT buffer (250 mM Tris–HCl pH 8.3, 375 mM KCl, 15 mM MgCl$_2$; Gibco-BRL) and 150 U SuperScript Reverse Transcriptase (Gibco-BRL) were added to a final volume of 30 µl, and incubated at 42°C for 1 h. Subsequently, the solution was heated to 95°C for 5 min and chilled on ice. RNase H (2.25 U; Promega) was added and the solution was incubated at 37°C for 20 min. An aliquot of the solution (10 µl) was used in a PCR reaction containing 12.5 pmol of primer hGHe, 50 mM KCl, 1.5 mM MgCl$_2$, 10 mM Tris–HCl pH 8.0, 0.2 mM dNTPs, 0.2 mg/ml BSA and 0.25 U SuperTaq (HT Biotechnology Ltd) in a reaction volume of 50 µl, followed by an initial denaturation step of 5 min at 94°C, 30 cycles of amplification (1 min at 94°C, 1 min at 60°C and 2 min at 72°C) and a final extension of 10 min at 72°C. No additional hGHf primer was added in the PCR reaction. Nested PCR, using either internal hGH primers or combinations of a hGH primer and a DMD primer, was performed on 1 µl of the primary PCR material with 12.5 pmol of each primer and PCR conditions identical to the

first PCR. The internal hGH primers used were hGHa and hGHb. When RNA-PCR products were used for *in vitro* transcription–translation, primer hGHa was replaced by hGHORF1, hGHORF2 or hGHORF3. Direct sequencing of PCR products was performed using the Sequenase™ PCR Product Sequencing kit (USB).

### Oligonucleotides and hybridisation probes

hGHa: 5′-CGGGATCCTAATACGACTCACTATAGGCGTCTG-CACCAGCTGGCCTTTGAC-3′
hGHb: 5′-CGGGATCCCGTCTAGAGGGTTCTGCAGGAATG-AATACTT-3′
hGHe: 5′-ACGCTATGCTCCGCGCCCATCGT-3′
hGHf : 5′-ACAGAGGGAGGTCTGGGGGTTCT-3′
D69F1: 5′-GCCATAAAAATGCACTATCCA-3′
D72F1: 5′-CCTCAGCTTTCACACGATGA-3′
D72R1: 5′-TCATCGTGTGAAAGCTGAGG-3′
D73R1: 5′-ATCCATTGCTGTTTTCCATTTC-3′
D74R1: 5′-GCAGGACTACGAGGCTGG-3′
polyT-REP: 5′-GGATCCGTCGACATCGATGAATTC(T)$^{25}$-3′
hGHORF1: 5′-CGGGATCCTAATACGACTCACTATAGGAC-GACCACC<u>ATG</u>CAGCTGGCCTTTGACACCTACCAGG-AG-3′
hGHORF2: 5′-CGGGATCCTAATACGACTCACTATAGGAC-AGACCACC<u>ATGG</u>CAGCTGGCCTTT<u>C</u>ACACCTACCAG-GAG-3′
hGHORF3: 5′-CGGGATCCTAATACGACTCACTATAGGAC-AGACCACC<u>ATGGG</u>CAGCTGGCCTTTGACACCTACCA-GGAG-3′.
hGHUTR1: 5′-CAGGAGAGGCACTGGGGA-3′

hGH primers were designed from the sequence M13438 and DMD primers from the sequence M18533 (EMBL sequence database). cDNA probe 63-1/3 is a subclone of the DMD cDNA and was used to screen for cosmids containing specific regions of the DMD gene. Probe 63-1/3 contains exons 65–74. Probe P20 contains exon 45 and part of intron 44 of the DMD gene (12).

### *In vitro* transcription–translation

Modified primers, containing a T7 promoter and an eukaryotic translation initiation sequence, were used to generate PCR products suitable for *in vitro* transcription–translation. T7-PCR product (200–400 ng) was added to the TnT/T7 coupled reticulocyte lysate system (Promega). The synthesized protein products were separated on a 15% SDS-polyacrylamide minigel system. Fluorography was obtained by washing the gels in DMSO/PPO. Dried gels were exposed 16–40 h for autoradiography.

## RESULTS

### Outline of cosmid-based exon trapping procedure

Vector sCOGH1 contains all the essential elements of a cosmid vector, i.e. origin of replication, antibiotic resistance marker (ampicillin and neomycin) and two *cos* sites (Fig. 1). In addition it contains an exon trap cassette consisting of a mMT-1 promoter driving expression of the hGH gene, containing a multicloning site (MCS) located between exons 2 and 3 (see Materials and Methods for details of vector construction). The ubiquitous mMT-1 promoter allows the use of many cell types. The vector is
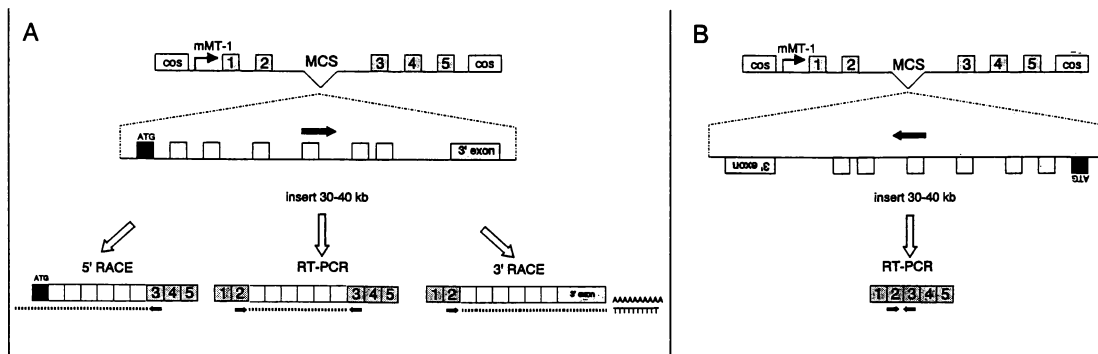
**Figure 2.** Entire cosmids constructed in vector sCOGH1 are transfected into V79 cells in which the introduced DNA is transcribed and processed. Spliced transcripts are amplified from isolated RNA. (**A**) Internal exons are amplified using 2 vector-derived primers. 5′ first exons can be amplified in a 5′ RACE using a single vector-derived primer. Similarly, 3′ terminal exons can be isolated in a 3′ RACE. (**B**) Trapping of cosmids containing inserts in the antisense orientation with respect to hGH will not result in formation of a chimeric hGH/unknown gene product. Instead, an 'empty' hGH transcript will be amplified by RT–PCR.

constructed such that the inserts can easily be excised and religated to obtain the opposite transcriptional orientations.

Cosmids are introduced into the cell type of choice by electroporation. We have tested and compared various cell lines and found V79 Chinese hamster lung cells to be a very efficient general host cell type. Upon expression, the hGH-initiated transcript will incorporate putative exons from the insert, cloned between exons 2 and 3 of the hGH gene, thus giving a chimeric product. After processing of the primary transcript, the putative RNA containing the exons to be trapped is amplified by RT–PCR using flanking vector-derived primers (Fig. 2A). In the specific event of 5′ and 3′ ends of genes being present in the insert, these will be skipped by the processing system or lead to alternatively initiated or terminated transcripts. They can be detected in the same mixture by 5′ or 3′ RACE, using opposite vector-derived primers separately (13). Gene inserts cloned in an antisense orientation will not be trapped, resulting in amplification of hGH sequences only (Fig. 2B).

RT–PCR analysis of V79 cells transfected with the sCOGH1 and sCOGH2 vectors yielded the expected 'empty' products (data not shown). A PCR with primers hGHa and hGHb gave the spliced 132 bp product containing hGH exons 2 and 3. Similarly, a 3′ RACE using primers hGHa and polyT-REP yielded a spliced 720 bp product starting in hGH exon 2 and ending in the poly(A) tail of the transcript. These experiments show that the insertion of the MCS into hGH intron 2 does not effect splice site selection.

**Trapping of multiple DMD-exons in a single spliced product**

To demonstrate the ability of the present method to isolate exonic gene segments from eukaryotic mammalian DNA, we subcloned YAC yDMD(0–25)C known to contain the human DMD gene (10). Two cosmids were isolated and used for exon trapping: cDMD2 and cDMD3. cDMD2 contains exons 72–76 and cDMD3 exons 68–74 (Fig. 3). RNA was isolated from V79 cells transfected with the cosmids and vector-derived transcripts were amplified by RT–PCR using either two vector-derived primers (hGHa and hGHb; Fig. 3, lane A,) or a combination of a

vector-derived and DMD exon primers (Fig. 3; lanes B and C). In all cases, RNA-PCR analysis yielded products containing the expected exonic DMD-segments.

In the case of cDMD2, RNA-PCR analysis using two vector-derived primers produced a product of 0.85 kb containing DMD exons 72–76 (Fig. 3). Direct sequencing confirmed that the PCR products contained the expected DMD exons, spliced together between growth hormone exons. All exon–exon transitions between DMD exons were perfect and DMD exon 72 was spliced correctly to hGH exon 2. DMD exon 76 was not spliced to hGH exon 3, but instead to a unknown sequence of 18 bp preceding the cloning site, resulting in insertion of an extra 61 bp.

Analysis of cDMD3 yielded a product of 0.88 kb containing DMD exons 68–74. DMD exon 68 was spliced correctly to hGH and all DMD exon–exon transitions were correct. Sequence analysis showed that not the hGH exon 3 splice acceptor site was used but a new site directly at the *Bam*HI cloning site. This results in a 43 bp insert of MCS and hGH intron 2 between DMD exon 74 and hGH exon 3.

**Inserts in antisense orientation**

Exon trapping of cDMD2r, containing the exonic DMD segments in the antisense orientation, gave no insert-derived products. The only product amplified was the 132 bp empty hGH exon 2/exon 3 product (Fig. 4, lane 1). This shows that, using this system, the false-positive rate of an entire 30 kb insert is effectively zero. Exon trapping of cDMD3r resulted in a PCR product of ~0.25 kb (Fig. 4, lane 4), either corresponding to a cryptic product or to an unknown exon derived from the antisense strand. Hybridisation of this product to a *Hin*dIII-digest of cDMD3 showed that it mapped to the cosmid and was spliced. The 339 bp product visible in lanes 2 (cDMD2r) and 3 (cDMD3) represents unspliced hamster growth hormone and results from traces of contamination of V79 genomic DNA in the RNA preparation.

*In vitro* **transcription–translation**

The 0.73 kb RT–PCR product of cDMD3 (Fig. 3B, lane B) was reamplified, replacing the hGH exon 2 forward primer with three

98
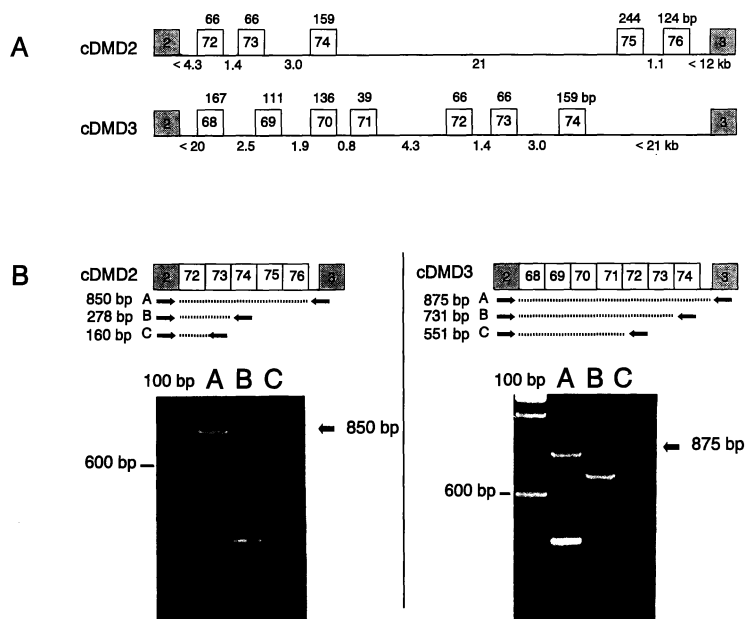
**Figure 3.** Exon trapping using DMD cosmids. (**A**) The genomic content of cosmids cDMD2 and cDMD3. The boxes represent exons and the lines introns. The shaded boxes represent hGH exons 2 and 3 and the open boxes DMD exons. The exon size in bp is indicated above the exons, the intron size in kb below the line. (**B**) RT–PCR of trapped exons. Lane A shows the RT–PCR product amplified using vector primers hGHa and hGHb. Lanes B and C of cDMD2 show the product amplified using primers hGHa in combination with D74R1 and hGHa with D73R1 respectively. Similarly lanes B and C of cDMD3 represent the product obtained with primers hGHa with D74R1 and hGHa with D72R1 respectively. The length of the products obtained is indicated. The 339 bp product visible in lane A of cDMD3 results from DNA contamination in the RNA sample and represents the endogenous growth hormone gene. Marker, 100 bp ladder (Gibco-BRL).
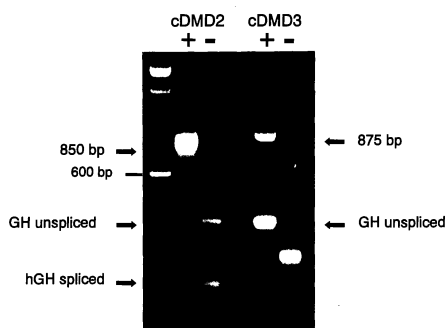


**Figure 4.** The effect of orientation of the cosmid insert. +, cosmid insert in sense orientation; –, cosmid insert in antisense orientation. Marker, 100 bp ladder (Gibco-BRL).

different primers, hGHORF1–3, containing a T7 promoter, a translation initiation sequence and either no, one or two additional nucleotides inserted between the ATG translation initiation codon and the hGH-sequence. The resulting RT–PCR products, each introducing a different reading frame, were used in an *in vitro* transcription–translation assay to scan for the presence of an open reading frame (ORF). As a control, a 0.6 kb PCR product of the hGH gene was synthesized in the three reading frames, using the same forward primers in combination with primer hGHUTR1, located in the 3′-UTR of the hGH gene. The control hGH product synthesized using primer hGHORF1 (Fig. 5), was predicted to contain an ORF of 172 amino acids and yielded the expected peptide of ~20 kDa, while no product was obtained in the two other reading frames. Similarly, *in vitro* transcription–translation of the cDMD3-derived hGHORF1 RT–PCR product yielded a peptide slightly over 30 kDa, as expected for the 230 amino acid ORF. (Fig. 5). This system is based on our earlier published 'protein truncation test' (PTT) system for the detection of open reading frames by *in vitro* transcription–translation (14).

## DISCUSSION

The cosmid-based exon trapping method described in this paper copes with several limitations of currently available exon trapping methods. Using large genomic inserts of 30 kb and larger, we isolated all exons present as a complete set, eliminating the need of subcloning and reordering of individually isolated exons and verification of their continuity from isolated cDNAs. If the cosmid inserts were in the antisense orientation, either nothing or a small product (i.e. cDMD3r) was trapped. The relevance of the latter product is still unclear; it either contains several cryptic exons or is part of a newly identified transcription unit. We did not trap any

99
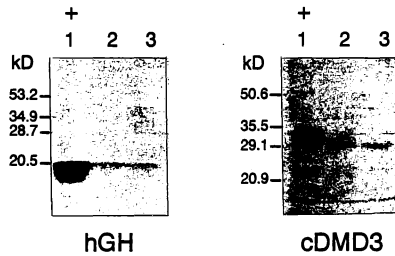
Figure 5. *In vitro* transcription–translation of RT–PCR products. The left side of the figure shows the translated peptide obtained with hGH and the right side the translated peptide obtained with the cDMD3 hGH/DMD chimeric transcript. Lanes 1–3, products amplified with primer hGHORF1, 2 and 3 respectively. Marker, pre-stained SDS–PAGE Low Range Marker (BioRad). The + indicates which of primers hGHORF1–3 should result in an in frame product.

false exons and the false positive rate obtained was in fact zero. Splicing was perfect between all DMD exon–exon transitions and hGH exon 2 and DMD exons. In both cosmids analysed the last DMD exon was not spliced to the splice acceptor of hGH exon 3, but to a site directly upstream of or in the multiple cloning site. This indicates the existence of *cis*-active 'higher order' effects in splicing, further underscoring the advantage of concerted trapping of a series of unknown exons, selected during evolution to cooperate in their parent gene. When separate exons are inserted in an 'alien' context this fine-tuning will be lost, which is probably the explanation for the differences in trapping efficiency of different exons using current systems. Alternatively, but not mutually exclusive, the selection of cryptic splice sites could be related to the maintenance of an open reading frame which has recently been shown to be an important factor influencing splice site selection (15).

We have constructed several variants of sCOGH1. In sCOGH2 the *Alu* repeat in the 3'-UTR of hGH has been removed, facilitating the screening of human positive cosmids with radiolabelled human DNA after subcloning of, for example, YACs from a mixture of YAC and total yeast DNA. In sCOGH6 a 4.7 kb fragment containing the SV2neo selectable marker has been removed, facilitating cloning of larger inserts. sCOGH3 differs from sCOGH1 by a deletion of the mMT1/hGH-exons 1 to 2 region (i.e. the promoter and 5' end of the gene). Due to the removal of the promoter and 5' end of the hGH-gene no RNA will be produced unless an insert contains a 5'-first exonic gene segment and a promoter which is active in the chosen cell line. These 5'-exonic sequences can be isolated efficiently from the RNA using a 5' RACE protocol (13).

RNA production can be boosted by super-inducing the mMT-1 promoter with heavy metal ions such as $Zn^{2+}$ and $Cd^{2+}$ (16). Neomycin selection can be used to select transfected clones specifically, but the system as described works so efficiently that we have never applied this selection, and in fact removed the neo gene in the sCOGH6 vector to generate 4.7 kb more space. The vectors used do not require a specific system for replication in the host cell and can be used in combination with any *in vivo* or *in vitro* system able to produce correctly processed RNA. In particular, due to the use of a strong ubiquitously expressed promoter, the necessity to use COS-1 cells for the initiation of transcription from the SV40 promoter is eliminated. The

sCOGH-system allows one to use other cell types (e.g. hamster V79), opening up several possibilities including targeting of tissue specific genes, e.g. in combination with sCOGH3, and functional complementation in specific cell types.

The results described in this paper deal with single cosmids containing part of a large gene. In gene rich regions more than one gene might be present in the cosmid insert and it is unclear what would happen in such an event. Most likely, transcripts initiated at the strong mMT-1 promoter will overrun cloned promoters, a situation similar to known genes with multiple promoters (17). We expect that the presence of cloned 3' exons will usually cause transcription termination. Still, examples are known where genes have multiple 3' exons, often expressed in a tissue specific manner, indicating that transcription can proceed and trap downstream sequences. Since with this system two RT–PCR reactions are standard, one with hGH exon 2 and 3 primers and one 3' RACE reaction, in most cases where multiple genes are cloned one should at least trap sequences from the most upstream gene. The identification of all genes from gene rich regions will depend on the use of a highly redundant cosmid contig covering the region. To scan large regions with the sCOGH-system, one has two possibilities: perform one experiment with a mixture of cosmids or use every cosmid in a single experiment. The feasibility of using complex mixtures remains to be tested. However, the situation will not differ significantly from that using small-insert vectors, where the high complexity of the input material introduces several technical problems. First, a large proportion of the clones will be empty and produce a PCR-favoured small product. Secondly, a wide range of products will be trapped with large size differences making it difficult to recognise the individual products. Consequently, PCR conditions should be chosen carefully to allow amplification of a wide size-range of RT–PCR products, especially for the cosmid-based system e.g. by using long-range PCR protocols. Since each cosmid contains a 25–40 kb insert, covering extensive regions with a manageable number of clones should be possible. Therefore, we would opt to use multiple cosmids simultaneously but in a miniaturised exon trap experiment where the cosmids are not mixed.

As demonstrated using RT–PCR and *in vitro* transcription–translation of products synthesized in all three possible reading frames derived from the hGH control and cDMD3, the exon trapping system can be coupled with a direct transcription–translation test (TTT) to detect the presence of large ORFs in the isolated sequences. This TTT approach provides an efficient tool to discriminate *bona fide* coding sequences from false positives. At the same time, this assay facilitates the identification of mutations by comparison of translated products derived from different sources of input genomic DNA, e.g. normal versus patient samples. Recently, we have shown that such a test can be performed even when only limited parts of a newly identified coding sequence have been elucidated (18). Since the proper connection of adjacent exons provides for correct translation, any disturbance in patient samples will become immediately apparent and highlight the area to be sequenced. In this way we could identify the CBP gene as the gene involved in Rubinstein-Taybi by the detection of translation terminating mutations in some patient-derived products.

## ACKNOWLEDGEMENTS

## REFERENCES

1  Duyk,G.M., Kim,S., Myers,R.M. and Cox,D.R. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 8995–8999.

2  Buckler,A.J., Chang,D.D., Graw,S.L., Brook, J.D., Haber, D.A., Sharp, P.A. and Housman,D.E. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 4005–4009.

3  Auch,D. and Reth,M. (1990) *Nucleic Acids Res.* **18**, 6743–6744.

4  Church,D.M., Stotler,C.J., Rutter,J.L., Murrell,J.R., Trofatter,J.A. and Buckler,A.J. (1994) *Nature Genet.* **6**, 98–105.

5  Andreadis,A., Nisson,P.E., Kosik,K.S. and Watkins,P.C. (1993) *Nucleic Acids Res.* **21**, 2217–2221.

6  Datson,N.A., Duyk,G.M., Van Ommen,G.J.B. and Den Dunnen,J.T. (1994) *Nucleic Acids Res.* **22**, 4148–4153.

7  Krizman,D.B. and Berget,S.M. (1993) *Nucleic Acids Res.* **21**, 5198–5202.

8  Evans,G.A., Lewis,K. and Rothenberg,B.E. (1989) *Gene* **79**, 9–20.

9  Selden,R.F., Howie,K.B., Rowe,M.E., Goodman,H.M. and Moore,D.D. (1986) *Mol. Cell. Biol.* **6**, 3173–3179.

10  Den Dunnen,J.T., Grootscholten,P.M., Dauwerse,J.D., Monaco,A.P., Walker,A.P., Butler,R., Anand,R., Coffey,A.J., Bentley,D.R., Steensma, H.Y. *et al.* (1992) *Hum. Mol. Genet.* **1**, 19–28.

11  Maniatis,T., Fritsch,E.F. and Sambrook,J. (1989) *Molecular Cloning: A Laboratory Manual.* Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

12  Wapenaar,M.C., Kievits,T., Blonden,L.A.J., Bakker,E., Den Dunnen,J.T., Van Ommen,G.J.B. and Pearson,P.L. (1987) *Cytogenet. Cell Genet.* **46**, 711.

13  Frohman,M.A., Dush,M.K. and Martin,G.R. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 8998–9002.

14  Roest,P.A.M., Roberts,R.G., Sugino,S., Van Ommen,G.J.B. and Den Dunnen,J.T. (1993) *Hum. Mol. Genet.* **2**, 1719–1721.

15  Dietz,H.C. and Kendzior,R.J.,Jr. (1994) *Nature Genet.* **8**, 183–188.

16  Searle,P.F., Stuart,G.W. and Palmiter,R.D. (1985) *Mol. Cell. Biol.* **5**, 1480–1488.

17  Ahn,A.H. and Kunkel,L.M. (1993) *Nature Genet.* **3**, 283–291.

18  Petrij,F., Giles,R.H., Dauwerse,J.G., Saris,J.J., Hennekam,R.C.M., Masuno.M. Tommerup,N., Van Ommen,G.J.B., Goodman,R.H., Peters,D.J.M. *et al.* (1995) *Nature* **376**, 348–351.

# CHAPTER 4.1

## Characterization of a new developmental gene, *SCML1*, in Xp22

Van de Vosse, E., Walpole, S.M., Nicolaou, A., Van der Bent, P., Cahn, A., Vaudin, M., Ross, M.T., Durham, J., Pavitt, R., Wilkinson, J., Grafham, D., Bergen, A.A.B., Van Ommen, G.J.B., Yates, J.R.W., Den Dunnen, J.T., and Trump, D.

# Characterization of a new developmental gene, *SCML1*, in Xp22

Esther van de Vosse[1], Susannah M. Walpole[2], Antonia Nicolaou[2], Paola van der Bent[1], Anthony Cahn[3], Mark Vaudin[3], Mark T. Ross[3], Jillian Durham[3], Rebecca Pavitt[3], Jane Wilkinson[3], Darren Grafham[3], Mark Ross[3], Arthur A.B. Bergen[4,*], Gert-Jan B. van Ommen[1,*], John R.W. Yates[2,*], Johan T. den Dunnen[1,*], and Dorothy Trump[2,*].

[1]MGC-Department of Human Genetics, Leiden University, Wassenaarseweg 72, 2333 AL Leiden, The Netherlands. [2]Departments of Medical Genetics and Pathology, Box 238, Lab Block level 3, Addenbrooke's Hospital, Hills Road, Cambridge CB2 2QQ, United Kingdom. [3]The Sanger Centre, Wellcome Trust Genome Campus, Hinxton CB10 1SA, United Kingdom, [4]The Netherlands Ophthalmic Research Institute, P.O.Box 12141, 1100 AC Amsterdam, The Netherlands.

*Note: A.A.B. Bergen, G.-J.B. Van Ommen, J.R.W. Yates, J.T. Den Dunnen and D. Trump are members of the Retinoschisis Consortium, others members are: B. Franco, A. Ballabio (TIGEM, Milan, Italy), W. Berger, H.H. Ropers (Berlin, Germany), T.E. Darga, P.A. Sieving (Michigan, USA), T.Alitalo, A. De la Chapelle (Helsinki, Finland).

## ABSTRACT

We have identified a new human developmental gene in Xp22 through our positional cloning studies of X-linked juvenile retinoschisis (RS). Using exon trap products from PAC and YAC clones we have isolated a set of exons which hybridize to a 3 kb mRNA. Expression of this RNA is detectable in a range of tissues but is most pronounced in skeletal muscle and heart. The gene, designated 'Sex comb on midleg like-1' *(SCML1)*, maps 14 kb centromeric of marker DXS418, between DXS418 and DXS7994, with its transcriptional orientation from telomere to centromere. *SCML1* spans 18 kb of genomic DNA, consists of 6 exons and the transcript contains a 624 bp open reading frame. The predicted 27 kiloDalton *SCML1* protein contains two domains which each have a high homology with two *Drosophila* transcriptional repressors of the Polycomb group (PcG) genes and their homologs in mouse and human. PcG genes are known to be involved in the regulation of homeotic genes and the mammalian homologs of the PcG genes repress the expression of *Hox* genes. *SCML1* is a new human member of this gene family and is likely to play an important role in the control of embryonal development.

## INTRODUCTION

A number of disease genes have been mapped to Xp22, including X-linked juvenile retinoschisis (RS, MIM 312700) (1), Nance Horan syndrome (NHS, MIM 302350) (2), sensorineural deafness (DFN6, MIM 300066) (3), non-specific X-linked mental retardation (MRX19) (4) and Fried syndrome (5). We are pursuing positional cloning studies to identify the gene causing X-linked juvenile retinoschisis (RS), the commonest cause of juvenile macular degeneration in males (1).

The RS locus has been mapped to Xp22.1-p22.2 by linkage analysis, with no evidence of genetic heterogeneity (6) and more recently, has been localised to the region between DXS418 and DXS999 (7,8), a physical distance of approximately 1 Mb (9,10). Haplotype conservation

and linkage analysis in Finnish RS patients suggests a location between DXS418 and DXS9911 (11). In a significant proportion of diseases localized to Xp, extensive deletions have been found in patients which helped guide the search to the gene(s) involved (12). Intriguingly, in spite of extensive searches, deletions have not yet been identified in the RS-gene candidate region.

Xp22 has been extensively mapped and several groups have published yeast artificial chromosome (YAC) contigs, P1-derived artificial chromosome (PAC) contigs and restriction maps (7,10,13,14) of this region. Clones form these contigs are currently being used for gene isolation although to date only one gene has been identified in this region, *PPEF* (Protein Phosphatase with EF hand motifs) (15). Although this gene was the human homolog of the *retinal degeneration gene* C (*rdg*C) of *Drosophila*, no mutations could be found in RS patients, rendering it unlikely that *PPEF* is involved in RS (15,16).

We have employed exon trapping to identify new genes in the RS region. PACs and YACs were selected from the contigs spanning DXS418 to DXS999 (7,10) and subcloned in the pSPL3b (17) and sCOGH2 (18) exon trapping vectors. A range of products containing putative exons were isolated and analyzed in further detail. Here we report the identification, characterization and analysis of a new gene from this region. It is highly homologous to the *Drosophila Scm* gene, a member of the Polycomb group (PcG) genes which are involved in the regulation of segmentation (19) and we have therefore designated it 'Sex comb on midleg like-1' (*SCML1*). We have found no mutations in *SCML1* in RS patients but this developmental gene remains a candidate for other diseases in Xp22.

## MATERIALS AND METHODS

### Exon trap experiments
YAC clones y939H7 (CEPH library) and y900E08102 (ICRF library) and PAC clone dJ389A20 (de Jong PAC library) were obtained from the Sanger Centre (UK).

*Alu*I partial digests were performed on PAC clone dJ389A20. 2-6 kb fragments were size-selected for subcloning into *Eco*RV digested, pSPL3b (Integrated Genetics) (17,20) prior to transformation into XL1-blue *E.coli* cells by electroporation. The resulting DNA was extracted and transfected into COS-7 cells by lipofection. RNA isolation and cDNA synthesis were performed on the incubated COS-7 cells followed by PCR amplification, *Bst*XI digestion, and ligation into pAMP10 (Gibco). Exon trap products were transformed into XL1-blue cells and sequenced using the AmpliCycle™ Kit (PerkinElmer).

YACs y939H7 and y900E08102 were partially digested with *Mbo*I, ligated in the *Bam*HI site of sCOGH2 (18), packaged using Gigapack II Plus Packaging Extract (Stratagene) and used to infect *E.coli* DH5α. Subclones containing human insert DNA were selected by hybridization. sCOGH exon trap experiments were performed as described by Datson *et al.* (18). Exon trap products were subcloned into pGEM-T (Promega) and sequenced using the AmpliCycle™ Kit (Perkin Elmer).

### Analysis of cDNAs, Northern blot hybridization
cDNAs za14f05, yq67g02 and zd45e08 were obtained from the HGMP Resource Centre and the Sanger Centre (Hinxton, UK). Aliquots of a fetal retinal cDNA library (Stratagene #937203) were used as template for PCR. All cDNAs were sequenced using the AmpliCycle™ Kit (Perkin Elmer). Human multiple tissue Northern blots (#7760-1 and #7756-1) were purchased from

Clontech and hybridized with cDNAs za14f05 and zd45e08 according to manufacturer's recommendations.

## RNA analysis

Human testis RNA was purchased from Clontech and RNA was isolated from human fetal brain using RNAzolB. cDNA was amplified from these using 1 µl random primer (Promega) which was annealed to 2 µg RNA at 65°C for 10 minutes. Reverse transcription was completed in 25 mM Tris-HCl (pH 8.3), 37.5 mM KCl, 1.5 mM MgCl$_2$, 10 mM DTT, 1 mM dNTPs, 40 U RNAsin (Promega), 600 U SuperScript$^{TM}$II at 42°C for 60 minutes. RNA from lymphoblastoid cell lines was extracted using Trizol reagent (Gibco BRL #15596-026) and cDNA amplified using the Superscript Preamplification System (Gibco BRL #18089-011) with both oligo-dT and random hexamer primers according to manufacturer's instructions.

**Table 1** Primer sequences

| Primer | Direction | Localization | 5' => 3' sequence |
|--------|-----------|--------------|-------------------|
| SCML-AF4 | for | exon A1 | TTTCCGAAGCGTCGAGTG |
| SCML-AR1 | rev | exon A1 | GCACGCGAGACCAGTGAT |
| SCML-AR2 | rev | exon A1 | TTCGGAAAGGTCCTGGCAC |
| SCML-A2F | for | exon A2 | CACAAATAAACCCTCCAGCA |
| SCML-A3F | for | exon A3 | AAACAAAACCTGAATTTGTCATAAA |
| SCML-BF | for | exon B | CAGGAACCGAATATTGTATCTG |
| SCML-BR | rev | exon B | GGCAATGAATAAGGACATCATC |
| SCML-CR | rev | exon C | GATTTGTCCACAGGGATCTCG |
| SCML-DF2 | for | exon D | GAGCAACCTTCCAAGGCCATCC |
| SCML-DF | for | exon D | AGGACCCGATCCTCAGCCGC |
| SCML-DR2 | rev | exon D | CTCGGAGTGCGGCTGAGGATC |
| SCML-EF | for | exon E | GTTACAAGGTCACCAGTTG |
| SCML-ER | rev | exon E | CAGGTTGAAGGGTGCTTAGTG |
| SCML-FF2 | for | exon F | ATTGACCGACTTAAACAAGG |
| SCML-FR4 | rev | exon F | GCCCAATCTAAATTTGCACAAGG |
| SCML-FR5 | rev | exon F | TTCTAATACTATCTAGCAG |
| SCML-FR2 | rev | exon F | TGGGTACAGCATCTTCATACAAAC |
| SCML-FR1 | rev | exon F | ACACCTGAGGACTGTTCAAGTGG |

## PCR

PCRs and nested PCRs were performed on 1 µl RT product in 67 mM Tris-HCl (pH8.8), 16.6 mM (NH$_4$)$_2$SO$_4$, 6.7 mM MgCl$_2$, 100 µM dNTPs and 1 U Red hot polymerase or in 10 mM Tris-HCl, 50 mM KCl, 1.5 mM MgCl$_2$, 0.01% gelatin, 0.1% Triton X-100, 200 µM dNTPs, 0.1 mg/ml BSA and 1 U Amplitaq polymerase in a total volume of 30-50 µl, with 10 pmol primer and an annealing temperature of 60°C. Primers are indicated in Table 1. 5'-RACE using the

human fetal brain Marathon-Ready™ cDNA (Clontech) was performed according to manufacturer's recommendations.

**DNA sequence analysis**
During the analysis of the gene, genomic sequence from PAC dJ389A20 became available. Exon prediction programs (GENSCAN, Fexh, Fgeneh, Xpound, Grail3, Hexon) were used to identify putative exons in the same orientation as *SCML1*. Blast was used to search for sequence homologies and further analysis was performed using the Wisconsin GCG package.

**Patient samples and mutation analysis**
Patient DNA was isolated from peripheral blood lymphocytes as described (21) and from lymphoblastoid cell lines as above. For sequencing, PCR products were used as a template and purified using QIAquick (QIAGEN #28104). Sequencing was performed on an ABI373 sequencer using the AmpliCycle™ Kit (Perkin Elmer).

## RESULTS

To identify genes in the RS gene candidate region in Xp22 we have subcloned PAC dJ389A20 and YACs y939H7 and y900E08102 into exon trap vectors pSPL3b and sCOGH2 respectively. A range of exon trap products was obtained, cloned and sequenced. Using database comparisons, eight products identified ESTs (Genbank: R98881, R98971, N68481, N91325, W69543, W69459). Of the 3 cDNAs corresponding to these ESTs, two were derived from a fetal liver and spleen library (yq67g02 and za14f05) and one from a fetal heart library (zd45e08). The exon trap products and cDNAs are indicated in Figure 1B. Sequence analysis of the cDNA clones showed that they were largely overlapping, producing a single transcript. Furthermore, sequencing of a PCR product (generated using primers SCML-AF4/CR) from a fetal retinal cDNA library contained identical 5' sequence (indicated as FRL937202.1 in Fig.1C). The localization of the transcript, close to and centromeric of DXS418, was determined by hybridization to Southern blots of a YAC fragmentation panel derived from y939H7 (9), by hybridization to PACs from across the region and by PCR using clones from the region (10) as templates (Fig.1). The gene maps between DXS418 and DXS7994.

The genomic sequence revealed a stretch of 13 A residues at the position where the 3' end of the cDNAs contained a polyA tail (Fig.2, position 1334). Since no polyadenylation signal was present upstream of this site we reasoned that these clones might be generated by internal oligo-dT priming during cDNA-synthesis, and to isolate the remainder of the full length transcript, we designed primers from the cDNA sequence, other exon trap products isolated from the region, computer-predicted exons and ESTs localized up or downstream. These sequences were based on the genomic sequence of PAC dJ389A20 which became available during the course of the project (determined by the Sanger Centre (Hinxton, UK) in collaboration with the RS-consortium). Several combinations of these primers were used in RT-PCR analysis performed on RNA isolated from fetal brain and normal lymphoblastoid cell lines. Primers were designed using EST H04958, derived from cDNA yj51e12, localized 3' to the cDNA contig (Table 1). Using primer SCML-DF (exon D) in combination with SCML-FR1 and SCML-FR2 (exon F) a 2 kb PCR product was obtained containing part of exon D, E and F. Moreover, EST H04958 contains a AATAAA-polyadenylation signal 22 bp upstream from the polyA stretch at
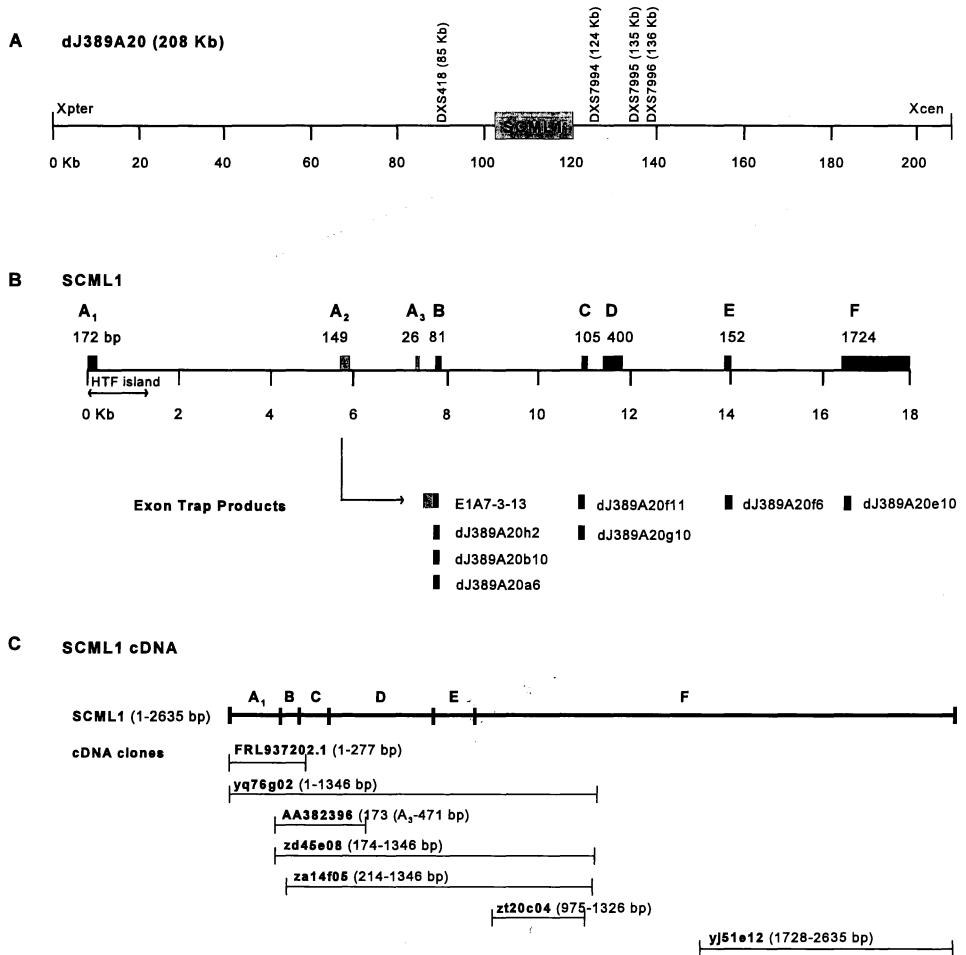
**Figure 1**. Localization of *SCML1* in the Xp22.1-p22.2 region.

**A**. PAC dJ389A20. *SCML1* maps between the markers DXS418 and DXS7994. The RS candidate region has been localized between the markers DXS418 (telomeric) and DXS999 (centromeric, not indicated in this figure).

**B**. *SCML1* gene. The genomic structure of *SCML1* is shown with its 6 exons (A1, B, C, D, E, F) varying in size from 81 bp to 1724 bp. The two potential alternate exons A2 and A3 are indicated. Exon trap product E1A7-13, which contains A2 spliced to B, was derived from cosmid E1A7. The other exon trap products dJ389A20b10, dJ389A20e10, dJ389A20a6, dJ389A20f6, dJ389A20h2, dJ389A20f11 and dJ389A20g10 were derived from PAC dJ389A20. Exon A2 (gagtagaggtttaccactcttaggtgactaagcagtatcacaaataaaccctccagcaagtttaaaaatattaggtccaactcagaggaagtggagtttc tcctgttgcacaaaaatgatgtctaacagctccagtgaaatcgatgtg) and A3 (ctcgtgc-aancaaaacctgaatttgtcaataaa) are described in the text.

C. *SCML1* cDNA. The entire *SCML1* cDNA is shown with the cDNA clones indicated below the region of *SCML1* they contain. The cDNAs yq67g02, za14f05, zd45e08, yj51e12 and zt20c04 correspond to GenBank accession numbers: R98881 and R98971, N68481 and N91325, W69543 and W69459, H04958 and H04957, and AA287232 respectively.

```
       exon A    < SCML-AR1                              < SCML-AR2        SCML-AF4 >
1      ATCACTGGTCTCGCGTGGGCGTGACCAGGCCCGGTTTCCGGTGCCAGGACCTTTCCGAAGCGTCGAGTGGCCTAACGGTCACAGCTGTCGCCCATCGGAGAGGCAGGACTACTGCGAGCA
                                                         ▼exon B   SCML-BF >
121    GTTTTACCGCGACCTCCGGAGGCCGGCGTGACAGGCTCTGTCACTAAAATAGGAACCGAATATTGTATCTGACGCATCCTGTAATACTGAAGAGCAACTGAAGACAGTTGATGATGTCCT
       < SCML-BR   ▼exon C                                                                              < SCML-CR  exon D
241    TATTCATTGCCAGGTTATATATGATGCTCTGCAAAACCTGGATAAGAAGATTGATGTGATTCGTAGAAAGGTTTCAAAAATCCAACGTTTCCATGCGAGATCCCTGTGGACAAATCATAA

361    GCGATATGGATATAAAAAGCATTCTTACCGGCTTGTTAAAAAGCTTAAACTCCAGAAAATGAAGAAAAATGAGGTTTACGAGACAATTCTCCTACCCTGAAAGTTACAGCCCCACTTTACC
                                                        M--K--K--N--E--V--Y--E--T--F--S--Y--P--E--S--Y--S--P--T--L--P-
                        SCML-DF2 >                                          SCML-DF >           < SCML-DR2
481    AGTGTCAAGGCGTGAGAATAATTCCCCGAGCAACCTTCCAAGGCCATCCTTTGCATGGAAGAATACCAGCGAGCTGAGCTGGAGGAGGACCCGATCCTCAGCCGCACTCCGAGTCCAGT
       -V--S--R--R--E--N--N--S--P--S--N--L--P--R--P--S--F--C--M--E--E--Y--Q--R--A--E--L--E--E--D--P--I--L--S--R--T--P--S--P--V-
601    GCATCCCTCAGATTTCTCTGAGCATAATTGTCAGCCGTATTATGCATCTGATGGTGCAAGGTATGGTTCTTCTTCAGGGCTCTGCCTTGGCAACCCTCGGGCTGACAGCATCCACAACAC
       -H--P--S--D--F--S--E--H--N--C--Q--P--Y--Y--A--S--D--G--A--T--Y--G--S--S--S--G--L--C--L--G--N--P--R--A--D--S--I--H--N--T-
                                                    ▼exon E   SCML-EF >                                   < SCML-ER
721    TTACTCAACTGACCATGCTTCTGCAGCACCACCTTCAGTTACAAGGTCACCAGTTGAAAATGACGGTTACATAGAGGAAGGAAGCATCACTAAGCACCCTTCAACCTGGTCGGTGGAAGC
       -Y--S--T--D--H--A--S--A--A--P--P--S--V--T--R--S--P--V--E--N--D--G--Y--I--E--E--G--S--I--T--K--H--P--S--T--W--S--V--E--A-
                                                               ▼exon F
841    AGTGGTCCTATTTCTAAAACAAACAGATCCTCTTGCATTATGCCCTCTTGTCGACCTCTTCAGAAGCCAIGAAATTGACGGGAAGGCTCTGCTCCTACTCACGAGTGACGTGTTGCTGAA
       -V--V--L--F--L--K--Q--T--D--P--L--A--L--C--P--L--V--D--L--F--R--R--S--H--E--I--D--G--K--A--L--L--L--L--T--S--D--V--L--L--K-
                                                      SCML-FF2 >                                     < SCML-FR4
961    GCACTTGGGGGTGAAGCTGGGAACGGCTGTGAAGCTATGCTACTACATTGACCGACTTAAACAAGGAAAATGCTTTGAAAATTGAAAAAATCCTTGTGCAAATTTAGATTGGGCCAACTT
       -H--L--G--V--K--L--G--T--A--V--K--L--C--Y--Y--I--D--R--L--K--Q--G--K--C--F--E--N-- *

1081   CTAGAGGCACCAATGCCTTCTTAGTGTGGAATCATTTTTCTGCCCTTTAGTGGTTTTTGTTTTGTAGAAAGTATCTCTCAAAATATATTATAGCTAGAATTGTAGAACTATGTTATAGTC
                            < SCML-FR5
1201   CAGTCTACTTCTTTAAAAACCATTTAAACTGCTAGATAGTATTAGAATAGTCCAATAGAAAAATTCATTCTTTATAGGTCTTTAAAAATTACTTTTATTATATTGTTTACAAATATATTTC
1321   ATGCAAGAAACAGAAAAAAAAAAAAACCCTTTGATTCTGGTTCATCTCGATACAGAGAACCAAAACAGCTAAGAGAGGTATTATCAGGGTTGACAACTCCTATGATTGAATCTATGGGAA
1441   TTATTCCTCAGAAGAGAATTTAAAGGTGTACCCATATATATCTCTTTCTGGAGTATTTTATCTGTCTGATGTTGCAGTATTCTACAAGTTTCCAGAAAGAGAATAGCCATATAAATTATT
1561   TTCCTTTCTGCTATTATTTCTCTATATGTTTTATTTATTCAGATTTAGAGTAAAAAATAAGCATATAAACTTTTATTATGTGCTCTTAACAGTTTTAAGATAAACTATAGGATAGATAGA
1681   ATGGTTATTTTATGCAAGAAATATTGTACCGCAAGGGTGGTTTGGATGAAGTCTGACTACTTTTTTTCAAACAAACTATTATATTAAAACTGTCATATTTTGGCTAAGTTTGGACCTATA
1801   ACTACACTTTCATTGTTTGCATCTCTCTATGAAGATACGTCTGTCCAAACTTTTAAAAGGCATAACTGTATTTTATGTGTTTATTCTTTATATAGATAGTATTTTTATATTTTATTCTCAC
1921   CCGAAGTATTCACACAATCTTTTTAAAAAAAATTTGAAATGGCATTTTGTATTGCCACAGAGGTAGGATGAGCCATATATTAGTGAAATGTTTTATTTTGTAAAATATAAATGGATTATT
2041   TGCCATCATTAGTACCTCTCAACTTACTTTTTAGAGGACAAGAAACAATCTGTAGATTGGTTTCCATACAGGGAAGTTCTCCGTCCTATGCAATGTTTCTAATTAATTTGCTTAATTCTG
2161   AGCCATTAATCCTGCTACACTTTGAATGATACATTAATTCAGACTAATCTTTGGGGGCTTTATTTTGTAAGTTAGAACTTTCAAGGGAAACATGTTCAACACTATTATTTTGTTATAAAT
2281   TTATAACTTTGTTATTACATTGTGTAACAAATATAAGGTTTACGAGCTATGAGAATTGGTGCTATCACCATTAGCTATTTGCTGTAATGTCAAGAAAATGTTCACCAGATGCAAGAATGT
2401   ACCTTTTCTTTTTAGAAAGCCAAATGTACTTTAGACATGAATGCAACTATTTAAAGAATAGCTTCATCAATGTTATTCCTTACATGTCATAAGATTCTTACTTAAACTTGGTCTTCTTTC
                   < SCML-ER2                    < SCML-ER1
2521   AAAATTGTTTGTATGAAGATGCTGTACCCACTTGAACAGTCCTCAGGTGTTTACATAAATACTATGTTTTACAGTTTTCATATTTTAAAATATTAATAAAGTTAAATCACAATAGT
```

**Figure 2** The nucleotide and deduced amino acid sequence of the *SCML1* gene. The exon boundaries are indicated by triangles. The 13 genomic As at position 1334 and the polyadenylation signal are underlined. Primer sequences are boxed.

its 3' end (Fig.2). No products were obtained from primers designed from 11 predicted exons and 3 exon trap products upstream of the transcript and orientated in the same direction or from 5' RACE experiments on fetal brain cDNA using primers SCML-AR1, SCML-AR2 and SCML-BR.

## Organizaton of the *SCML1* gene

To determine the exon/intron boundaries, we subcloned and sequenced genomic fragments of cosmid E1A7 and analyzed the genomic sequence of PAC dJ389A20. The 2635 bp cDNA sequence comprises 6 exons ranging in size from 81 to 1724 bp (Table 2). The complete gene spans 18 kb of genomic DNA and is transcribed from telomere to centromere.

Northern blots, containing poly(A)⁺ RNA of 8 adult and 4 fetal tissues, were hybridized

with cDNAs za14f05 and zd45e08. This revealed a major transcript of ~3 kb in all tissues. Expression in adult tissues was most prominent in skeletal muscle and heart, while liver and placenta also gave strong signals. Expression is weaker but clearly detectable in pancreas, kidney, lung and brain (Fig.3A). Four tested fetal tissues, i.e. brain, lung, liver and kidney also showed strong expression (Fig.3B). A weak transcript of ~12 kb was detected in the four fetal tissues and adult pancreas, placenta and skeletal muscle which may be due to cross-hybridization to a homolog or to alternative splicing of the transcript (Fig 3A and 3B).

**Table 2** Splice sites in *SCML1*. Exonic sequences are capitalized.

| exon | 5' splice donor | intron length | 3' splice acceptor | exon |
|------|------------------|----------------|---------------------|--------|
| exon A1 | ACTAAAATAGgtct | 8337 bp | gcagGAACCGAATA | exon B |
| exon B | TCATTGCCAGgtat | 3301 bp | atagGTTATATATG | exon C |
| exon C | GACAAATCATgtaa | 389 bp | tcagAAGCGATATG | exon D |
| exon D | CCACCTTCAGgtat | 1521 bp | gaagTTACAAGGTC | exon E |
| exon E | CAGAAGCCATgtaa | 1291 bp | atagGAAATTGACG | exon F |

We obtained evidence of alternative splicing upstream of exon B. Clone yq76g02 from a fetal liver and spleen cDNA library and the PCR product from th fetal retinal library FRL937202.1 consisted of exon A1 spliced to exon B. However, exon trap product E1A7-3-13, contained 149 bp of sequence derived from intron A (designated exon A2, Fig.1B) spliced to exon B and in addition, cDNA AA382396, a 332 bp testis derived cDNA, contained exon B, C, part of exon D and 26 bp of sequence derived from intron A (designated exon A3, Fig.1B). PCR of fetal brain cDNA using a forward primer in exon A1, A2 or A3 in combination with a reverse primer in exon C yielded only products in combination with primers in exon A1 and exon C. Lymphoblastoid cDNA, however gave a product witha primer from exon A1 and exon C and also a weaker product with a primer from exon A3 and exon C (confirmed by sequencing). These results indicate that A2 and A3 may be exons present in alternate trancripts of *SCML1*.

### *SCML1* protein

The largest open reading frame found consists of 208 amino acids. The translation initiation site lies at position 419 in exon D (Fig.2). The sequence around the ATG initiation codon has a score of 63 (on a scale from 0-100) compared to the Kozak consensus sequence (CCCGCCGCCACCATGG) (23,24). Database searches revealed homology with several other proteins (Fig.4). The highest homology (57 % identity, 88 % similarity in a 42 amino acid C-terminal domain) is found with the *Scm*-protein (Sex comb on midleg) of *Drosophila melanogaster* (19). The ORF extends 246 bp upstream of the translation initiation site at position 419 but no other translation initiation site is present and furthermore, no homologies could be found between the predicted amino acid sequence and the PcG genes.

## Analysis of RS patients

To screen for deletions and mutations in genomic DNA we hybridized cDNA za14f05 to Southern blots containing DNA of unrelated RS patients. DNA of 23 RS patients was digested with *Msp*I, *Bam*HI, *Eco*RV and *Eco*RI and DNA of another 49 unrelated RS patients was digested with *Hind*III. No aberrant fragments were detected.

The complete coding region was amplified from 3 RS patients' lymphoblastoid cell RNA (using primers SCML-BF/-DR2, -EF/-FR4, -ER/-DF2, and -FF2/-FR5) and sequenced. No sequence variation was found.
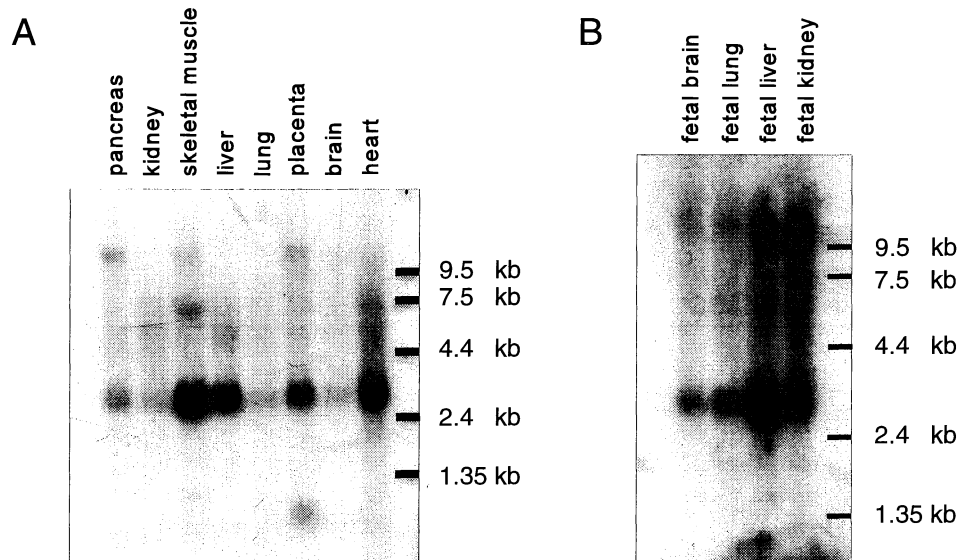


**Figure 3**. Northern blot analysis of *SCML1*
A  Multiple tissue Northern blot (Clontech, # 7760-1) hybridized with cDNA za14f05.
B  Multiple tissue fetal Northern blot (Clontech, #7756-1) hybridized with cDNA zd45e08.
A major transcript of ~3 kb was detected in all tissues and a fainter transcript of 12 kb is present in all 4 fetal tissues, and adult pancreas, placenta and skeletal muscle.

## DISCUSSION

Using pSPL3b and sCOGH2 exon trapping on clones from the RS candidate region we have isolated a new gene, designated *SCML1* for Sex comb on midleg like-1. The gene is localized close to marker DXS418, consists of 6 exons, has a transcriptional orientation from telomere to centromere and encodes a 3 kb transcript. Based on Northern blot hybridization and cDNA library clones, expression of *SCML1* was found in six fetal tissues analyzed (brain, lung, liver, kidney, retina and heart) and 10 different adult tissues, predominantly muscle and heart. Mutation screening of the gene in RS patients using hybridization analysis and sequencing did not reveal any mutations, rendering it unlikely that this gene is involved in retinoschisis.

The size of the Northern signal (approximately 3 kb) is in accordance with the size of the

cDNA isolated (2,635 bp), excluding a polyA tail. Exon A lies in an HTF island and contains 4 *Hpa*II sites. Another 8 *Hpa*II sites are present within the first 2 kb of intron A. No TATA-box and CCAAT-box sequences were found immediately 5' of exon A but the region does contain seven consensus Sp1 binding site sequences. These features are indicative of a promoter located in this region. The ORF can be extended 246 bp upstream of the translation initiation site through exon C to the beginning of exon B. No other initiation site is present within this sequence.

### *SCML1* protein

The largest open reading frame found consists of 208 amino acids and encodes a predicted 27 kiloDalton protein. Database searches revealed homology with several other proteins (Fig.4). The highest homology was found with the *Scm*-protein of *Drosophila melanogaster* (19). *Scm* is a member of the Polycomb group (PcG) of genes in *Drosophila*. High homologies were also found with another PcG gene, the polyhomeotic gene (*Ph*) in *Drosophila* (25), and three mammalian PcG homologs, a transcriptional repressor (*Rae28/Mph1*) in mouse (26) and two polyhomeotic homologs (*HPH1*, *HPH2*) in human (27) (Fig.4).

Although the *SCML1* transcript contains a large 3' untranslated region (1,593 bp), the translational stop codon lies at the same position as that in the homologous proteins (see Fig.4), directly after the SPM domain. This domain is thus found exclusively at the extreme C-terminus of these proteins which is likely to reflect an important feature of its function. The SPM domain (19), has been named after three *Drosophila* proteins in which it was found first, <u>S</u>cm, <u>P</u>h, and *l(3)<u>m</u>bt* (28), and is also known as the SEP (yeast <u>s</u>terility, <u>E</u>ts related, <u>P</u>cG proteins) domain. It has been found in several cytoplasmic proteins and members of the Ets family of transcription factors in yeast, nematode and *Dictyostelium* (26). The second domain of high homology we identified, designated the SP domain (after <u>S</u>cm and <u>P</u>h) has not been reported before and is not present in the *l(3)mbt* protein.

Alternate splicing, of which there is some evidence, could extend this ORF further and lead to the encoding of a larger protein. However, searches or alternate spliced products using 5'-RACE and RT-PCR with fetal brain RNA and lymphoblastoid RNA respectively failed to give products. It may be that alternate splice products are present in other tissues.

*SCML1* is smaller than *Scm* but contains two conserved domains SPM and SP which are likely to be functional. However, it lacks the potential zinc finger domains and the mbt repeats present in *Scm* (Fig. 4). The 5' potential extension of the ORF present in *SCML1* has no further homology with *Scm*, but products of tissue specific alternate splicing might encode a second larger protein.

**Figure 4**.
**A** Domains in *SCML1* and related genes from the Polycomb group (PcG). *Scm* = Sex combs on midleg (*Drosophila*), *Mph1* = mouse homolog of the polyhomeotic gene, *HPH1* and *HPH2* are the human homologs of the polyhomeotic gene, *Ph* = Polyhomeotic gene (*Drosophila*). The SPM and SP domains are present in all sequences. Domains shaded with downward slopes represent potential zinc fingers. The two horizontal striped domains in *Scm* are mbt repeats and the other domains indicated are of unknown function.
**B** Amino acid sequence alignment of *SCML1* and homologous regions in *Scm*, *Mph1*, *HPH1*, *HPH2* and *Ph*. Dark shading indicates amino acid residues that are present in 50% of the aligned sequences. Light shading indicates similar amino acid residues in 50% of the aligned sequences. The bottom line is the consensus sequence: * indicates identity in all 6 aligned sequences, • indicates similarity or identity in

at least 50% of aligned sequences. Sequences were aligned using the PILEUP program from the Wisconsin GCG package, shading was performed by the BOXSHADE server.

**A**

| | | | |
|---|---|---|---|
| | | SP | SPM |
| SCML1 | | | 208 aa |
| Scm | | | 877 aa |
| Mph1 | | | 1012 aa |
| HPH1 | | | 1004 aa |
| HPH2 | | | 434 aa |
| Ph | | | 1589 aa |

**B**

```
SCML1      1  MKKNEVYETFSYPES-YSPTLEVSRRENNSPSNIERPSFCMEEYQRAEIEED-EILSRTPSPVHPSDFSEHNCQPYYASDGATYGSSSGLCLGNP
Scm      719  ----------SYYK---SPTELSSSASLPTSVSTEFTGCQSASSTALAACGVTAAKAATAPAGAA-----------------------------
Mph1     896  ----------SYCE-ALSPTSEGELSVRAGECERDLGNTITTPSTP-EICGINEVF-------------------------------------
HPH1     888  ----------SYCE-ALSPTSEGELSVRAGECERDLGNPNTAPETP-EIECINEVF------------------------------------
HPH2     321  ----------SYCE-ELSPISASSSTSAGDEASGTWSCPTCICCTWWAWDTTSCQ---------------------------------------
Ph      1475  ----------SPLSVAL-PTLA-PLSVV-TSCAAEKSSEVN--GIDRPP--------------------------------------------
consensus 1901       *... ..**... ... .... . ..  .   . .. .. .

SCML1     94  RADSIHNTYSTDHASAAPPSVTRSPVENDGYIEEGSITKEPSTWSVEAVVIELKQTD-EIAL P-LVDLFRSEEIDGKALLLITSDVILKHLGVK
Scm      771  ---------ATAGASPSYTAITSPVSTPTSALANSHPSGEIDVTIEEVIQYIESNDNSLAVHG---DLFRKEEIDGKALLLINSEMMKYNGIK
Mph1     940  ---------------------------------------LSSQPSQWSVEEVYEFI----ASICGCQEIAEFRSQEIDGQALLLLKEEHLMSAMNIK
HPH1     932  ---------------------------------------LSSQPSRWSVEEVYEFI----ASICGCQEIAEFRSQEIDGQAELLLKEEHLMSAMNIK
HPH2     365  ---------------------------------------VSHQ--VNVEEVYEFI----RSIEGCQEIAEEFAQEIDGQALLLLKEEHLMSVMNIK
Ph      1509  ---------------------------------------LSS----WSVDEVSNFI----REIEGCQDYVDLEIQQEIDGQALLLLKEHEVNANGMK
consensus 1996                                          ....... ..........*.    .* ...........* *****.*.**********. .**

SCML1    187  LGTAYKGCYYLDRLKQGKCFEN*-------
Scm      854  LGPALKICNLVNKVNGRRNNLAL*------
Mph1     995  LGPALKICAKINVLKET*------------
HPH1     987  LGPALKICAKINVLKET*------------
HPH2     417  LGPALKIYARISYLKDS*------------
Ph      1560  LGPALKIVAKVESIKEVPPPGEAKDPGAQ*
consensus 2091  **.**..*............ .
```

114

The PcG genes encode transcriptional repressors in *Drosophila* essential for proper spatial expression of homeotic genes and thus appropriate development. Murine homologs of PcG genes that have recently been isolated (*Bmi1*, *Mel18*, *Mph1*) seem likewise to be involved in transcriptional repression of the vertebrate counterparts of the homeotic genes, the *Hox* genes (26). Mice mutated in the *Bmi1* or *Mel18* genes display axial skeletal malformations (with knock-out mutations causing posterior tranformation of the skeleton and overexpression leading to anterior transformation) associated with other abnormalities including large bowel obstruction and thymus hypoplasia (29-31). In keeping with this, the expression of *Hox* genes has been shown to be altered in these mutants.

The *SCML1* gene is therefore likely to be involved in transcriptional repression of *Hox* genes. Mutations in this gene may cause developmental malformations rather than a retinal degeneration as is found in RS. Other diseases localized in this region for which *SCML1* remains a candidate are Nance-Horan syndrome (NHS) (2), sensorineural deafness (DFN6) (3), non-specific X-linked mental retardation (MRX19) (4), and Fried syndrome (5), although none of these conditions have a pattern of abnormality similar to that seen in the mouse *Bmi1* or *Mel18* mutants. Two patients have been described with a combination of vertebral abnormalities, bowel atresia and thymus abnormalities (32), and a pair of brothers have been described with vertebral abnormalities and Hirschsprung disease in one and vertebral abnormalities and anal atresia in the other (33). Mutations in *SCML1* could be a cause of such syndromes. Alternatively, one could hypothesize that deletions in this gene are lethal due to disturbed embryonal development, perhaps explaining the absence of the identification of any deletion in this region of the X-chromosome in either males or females. Further studies are needed to elucidate the role of this gene in development, its relation with other human genes of the Polycomb group and the response of developmental genes such as *Hox* genes to *SCML1* mutations.

## ACKNOWLEDGEMENTS

## REFERENCES

1. George, N.D.L., Yates, J.R.W. and Moore, A.T. (1995) X linked retinoschisis. *Br.J.Ophthalmol.* **79,** 697-702.
2. Toutain, A., Ronce, N., Dessay, B., Robb, L., Francannet, C., Le Merrer, M., Briard, M.L., Kaplan, J. and Moraine, C. (1997) Nance-Horan syndrome: linkage analysis in 4 families refines localization in Xp22.31-p22.13 region. *Hum.Genet.* **99,** 256-261.
3. Del Castillo, I., Villamar, M., Sarduy, M., Romero, L., Herraiz, C., Hernandez, F.J., Rodriguez, M., Borras, I., Montero, A., Bellon, J., Cruz Tapia, M. and Moreno, F. (1996)

115

A novel locus for non-syndromic sensorineural deafness (DFN6) maps to chromosome Xp22. *Hum.Mol.Genet.* **5,** 1383-1387.

4.     Donnelly, A.J., Andy Choo, K.H., Kozman, H.M., Gedeon, A.K., Danks, D.M. and Mulley, J.C. (1994) Regional localisation of a non-specific X-linked mental retardation gene (MRX19) to Xp22. *Am.J.Med.Genet.* **51,** 581-585.

5.     Strain, L., Wright, A.F. and Bonthron, D.T. (1997) Fried syndrome is a distinct X linked mental retardation syndrome mapping to Xp22. *J.Med.Genet.* **34,** 535-540.

6.     Sieving, P.A., Bingham, E.L., Roth, M.S., Young, M.R., Boehnke, M., Kuo, C.-Y. and Ginsburg, D. (1990) Linkage relationship of X-linked juvenile retinoschisis with Xp22.1-p22.3 probes. *Am.J.Hum.Genet.* **47,** 616-621.

7.     Van de Vosse, E., Bergen, A.A.B., Meershoek, E.J., Oosterwijk, J.C., Gregory, S., Bakker, B., Weissenbach, J., Coffey, A.J., Van Ommen, G.J.B. and Den Dunnen, J.T. (1996) A Xp22.1-p22.2 YAC contig encompassing the disease loci for RS, KFSD, CLS, HYP and RP15; refined localization of RS. *Eur.J.Hum.Genet.* **4,** 101-104.

8.     George, N.D.L., Payne, S.J., Barton, D.E., Moore, A.T. and Yates, J.R.W. (1994) Genetic mapping of X-linked Retinoschisis. *Cytogenet.Cell Genet.* **67,** 354

9.     Van de Vosse, E., Van der Bent, P., Heus, J.J., Van Ommen, G.J.B. and Den Dunnen, J.T. (1997) High resolution mapping by YAC fragmentation of a 2.5 Mb Xp22 region containing the human RS, KFSD and CLS disease genes. *Mamm.Genome,* **8,** 497-501.

10.    Walpole, S.M., Nicolaou, A., Howell, G.R., Whittaker, A., Bentley, D.R., Ross, M.T., Yates, J.R.W. and Trump, D. (1997) High resolution physical map of the X-linked retinoschisis interval in Xp22. *Genomics,* (in press)

11.    Huopaniemi, L., Rantala, A., Tahvanainen, E., De la Chapelle, A. and Alitalo, T. (1997) Linkage disequilibrium and physical mapping of X-linked juvenile Retinoschisis. *Am.J.Hum.Genet.* **60,** 1139-1149.

12.    Ballabio, A. and Andria, G. (1992) Deletions and translocations involving the distal short arm of the human X-chromosome: review and hypotheses. *Hum.Mol.Genet.* **1,** 221-227.

13.    Alitalo, T., Francis, F., Kere, J., Lehrach, H., Schlessinger, D. and Willard, H.F. (1995) A 6-Mb YAC contig in Xp22.1-p22.2 spanning the DXS69E, XE59, GLRA2, PIGA, GRPR, CALB3 and PHKA2 genes. *Genomics,* **25,** 691-700.

14.    Ferrero, G.B., Franco, B., Roth, E.J., Firulli, B.A., Borsani, G., Delmas-Mata, J., Weissenbach, J., Halley, G., Schlessinger, D., Chinault, A.C., Zoghbi, H.Y., Nelson, D.L. and Ballabio, A. (1995) An integrated physical and genetic map of a 35 Mb region on chromosome Xp22.3-Xp21.3. *Hum.Mol.Genet.* **4,** 1821-1827.

15.    Montini, E., Rugarli, E.I., Van de Vosse, E., Andolfi, G., Puca, A.A., Den Dunnen, J.T., Ballabio, A. and Franco, B. (1997) A human homolog of the Drosophila retinal degeneration C (rdgc) gene encodes a novel serine-threonine phosphatase selectively expressed in sensory neurons of neural crest origin. *Hum.Mol.Genet.* **6,** 1137-1145.

16.    Van de Vosse, E., Franco, B., Van der Bent, P., Montini, E., Orth, U., Hanauer, A., Tijmes, N., Van Ommen, G.J.B., Ballabio, A., Den Dunnen, J.T. and Bergen, A.A.B. (1997) Exclusion of *PPEF* as the gene causing X-linked juvenile retinoschisis. *Hum.Genet.* (in press).

17.    Church, D.M., Stotler, C.J., Rutter, J.L., Murrell, J.R., Trofatter, J.A. and Buckler, A.J. (1994) Isolation of genes from complex sources of mammalian genomic DNA using exon amplification. *Nature Genet.* **6,** 98-105.

18. Datson, N.A., Van de Vosse, E., Dauwerse, H.G., Bout, M., Van Ommen, G.J.B. and Den Dunnen, J.T. (1996) Scanning for genes in large genomic regions: cosmid-based exon trapping of multiple exons in a single product. *Nucleic Acids Res.* **24,** 1105-1111.

19. Bornemann, D., Miller, E. and Simon, J. (1996) The Drosophila Polycomb group gene *Sex comb on midleg* (Scm) encodes a zinc finger protein with similarity to polyhomeotic protein. *Development,* **122,** 1621-1630.

20. Buckler, A.J., Chang, D.D., Graw, S.L., Brook, J.D., Haber, D.A., Sharp, P.A. and Housman, D.E. (1991) Exon amplification: A strategy to isolate mammalian genes based on RNA splicing. *Proc.Natl.Acad.Sci.USA,* **88,** 4005-4009.

21. Maniatis, T., Fritsch, E. and Sambrook, J. (1989) *Molecular Cloning: A Laboratory Manual.* Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.

22. Orita, M., Suzuki, Y., Sekiya, T. and Hayashi, K. (1989) Rapid and sensitive detection of point mutations and DNA polymorphisms using the polymerase chain reaction. *Genomics,* **5,** 874-879.

23. Kozak, M. (1987) An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs. *Nucleic Acids Res.* **15,** 8125-8148.

24. Shapiro, M.B. and Senapathy, P. (1987) RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. *Nucleic.Acids.Res.* **15,** 7155-7174.

25. DeCamillis, M., Cheng, N.S., Pierre, D. and Brock, H.W. (1992) The polyhomeotic gene of Drosophila encodes a chromatin protein that shares polytene chromosome-binding sites with Polycomb. *Genes Dev.* **6,** 223-232.

26. Alkema, M.J., Bronk, M., Verhoeven, E., Otte, A., Van 't Veer, L.J., Berns, A. and Van Lohuizen, M. (1997) Identification of Bmi1-interacting proteins as constituents of a multimeric mammalian Polycomb complex. *Genes & Dev.* **11,** 226-240.

27. Gunster, M.J., Satijn, D.P.E., Hamer, K.M., Den Blaauwen, J.L., De Bruijn, D., Alkema, M.J., Van Lohuizen, M., Van Driel, R. and Otte, A.P. (1997) Identification and characterization of interactions between the vertebrate polycomb-group protein BMI1 and human homologs of polyhomeotic. *Mol.Cell.Biol.* **17,** 2326-2335.

28. Wismar, J., Loffler, T., Habtemichael, N., Vef, O., Geissen, M., Zirwes, R., Altmeyer, W., Sass, H. and Gateff, E. (1995) The *Drosophila melanogaster* tumor suppressor gene lethal(3)malignant brain tumor encodes a proline-rich protein with a novel zinc finger. *Mech.Dev.* **53,** 141-154.

29. Alkema, M.J., Van der Lugt, N.M.T., Bobeldijk, R.C., Berns, A. and Van Lohuizen, M. (1995) Transformation of axial skeleton due to overexpression of bmi-1 in transgenic mice. *Nature,* **374,** 724-727.

30. Van der Lugt, N.M.T., Domen, J., Linders, K., Van Roon, M., Robanus-Maandag, E., Te Riele, H., Van der Valk, M., Deschamps, J., Sofroniew, M., Van Lohuizen, M. and Berns, A. (1994) Posterior transformation, neurological abnormalities, and severe hematopoietic defects in mice with a targeted deletion of the *bmi-1* proto-oncogene. *Genes & Dev.* **8,** 757-769.

31. Akasaka, T., Kanno, M., Balling, R., Mieza, M.A., Taniguchi, M. and Koseki, H. (1996) A role for mel-18, a Polycomb group-related vertebrate gene, during the anterioposterior specification of the axial skeleton. *Development,* **122,** 1513-1522.

32. Urioste, M., Lorda-Sánchez, I., Blanco, M., Burón, E., Aparicio, P. and Martínez-Frías,

M.L. (1996) Severe congenital limb deficiencies, vertebral hypersegmentation, absent thymus and mirror polydactyly: a defect expression of a developmental control gene? *Hum Genet,* **97,** 214-217.

33.    Melhem, R.E. and Fahl, M. (1985) Fifteen dorsal vertebrae and rib pairs in two siblings. *Pediatr.Radiol.* **15,** 61-62.

**CHAPTER 4.2**

**A novel human serine-threonine phosphatase related to**
**the *Drosophila retinal degeneration C* (*rdg*C) gene is selectively expressed**
**in sensory neurons of neural crest origin**

Montini, E., Rugarli, E.I., Van de Vosse, E., Andolfi, G., Mariani, M., Puca, A.A., Consalez, G.G., Den Dunnen, J.T., Ballabio, A., and Franco, B.

# A novel human serine-threonine phosphatase related to the *Drosophila retinal degeneration* C (*rdgC*) gene is selectively expressed in sensory neurons of neural crest origin

E. Montini[1,+], E. I. Rugarli[1,+], E. Van de Vosse[3], G. Andolfi[1], M. Mariani[2], A. A. Puca[1], G. G. Consalez[2], J. T. den Dunnen[3], A. Ballabio[1] and B. Franco[1,*]

[1]Telethon Institute of Genetics and Medicine (TIGEM) and [2]Department of Biological Technological Research (DIBIT), San Raffaele Biomedical Science Park, Milan, Italy and [3]MGC, Department of Human Genetics, Leiden University, Leiden, the Netherlands

Through our transcriptional mapping effort in the Xp22 region, we have isolated by exon trapping a new transcript highly homologous to the *Drosophila retinal degeneration* C (*rdgC*) gene. *rdgC* encodes a serine/threonine phosphatase protein and is required in *Drosophila* to prevent light-induced retinal degeneration. This human gene is the first mammalian member of the serine-threonine phosphatase with EF hand motif gene family, and was thus named *PPEF* (Protein Phosphatase with EF calcium-binding domain). The expression pattern of the mouse *Ppef* gene was studied by RNA *in situ* hybridization on embryonic tissue sections. While *rdgC* is expressed in the visual system of the fly, as well as in the mushroom bodies of the central brain, we found that *Ppef* is highly expressed in sensory neurons of the dorsal root ganglia (DRG) and neural crest-derived cranial ganglia. The selective pattern of expression makes *PPEF* an important marker for sensory neuron differentiation and suggests a role for serine-threonine phosphatases in mammalian development.

## INTRODUCTION

Our group is involved in the construction of a transcription map of the human Xp22 region. To achieve this aim, we have constructed a detailed physical map of a 35 Mb region spanning human chromosome Xp22.3–Xp21.3. The backbone of the map is represented by a single contiguous stretch of 585 overlapping yeast artificial chromosome (YAC) clones covering the entire region (1). Several disease loci have been mapped in this region including Retinoschisis (RS), Nance–Horan syndrome (NHS), Coffin–Lowry syndrome (CLS), and Keratosis Follicularis Spinulosa Decalvans (KFSD) (1). As a first step toward building a transcription map of this region, we decided to concentrate our

efforts on YAC clone 939H7 which spans the entire critical region for RS and partially spans the critical regions for NHS, CLS and KFSD.

This effort led us to the isolation of a gene highly homologous to the *Drosophila retinal degeneration* C (*rdgC*) gene. The *rdgC* gene encodes a serine/threonine phosphatase protein (2) and is required in *Drosophila* to prevent light-induced retinal degeneration (3). *rdgC* is expressed in the visual system of the fly, as well as in the mushroom bodies of the central brain. We named this new gene *PPEF*, for Protein Phosphatase with EF hand motif.

To test the involvement of *PPEF* in the pathogenesis of RS, we isolated the full-length cDNA, established the genomic structure, and searched for mutations in RS patients by PTT (protein truncation test) and SSCP (single strand conformation polymorphism) analyses. Identification of the mouse homolog of this gene allowed us to perform RNA *in situ* hybridization studies on mouse embryo tissue sections, revealing a very specific pattern of expression localized in sensory neurons of cranial and dorsal root ganglia.

## RESULTS

### Identification and characterization of the PPEF cDNA

One hundred cosmid clones were selected by screening an X-specific cosmid library with *Alu* PCR products derived from YAC 939H7. Cosmid clones were grouped and used for exon trapping experiments. Several exon trapping products were sequenced and used to search non-redundant DNA and protein databases through the BLAST-X, BLAST-P and TBLAST-N algorithms. One of them (clone 3pn1D2) was found to be identical to EST H18854, which shows significant homology to the *Drosophila rdgC* protein (accession no. M89628). Clone 3pn1D2 mapped back to YAC939H7 in the RS critical region and to cosmid 44C1 containing marker *DXS999* (Fig. 1). The same transcript was subsequently identified in our laboratory, using a bioinformatic approach aimed at the identification of human homologs of *Drosophila* genes involved in the generation of
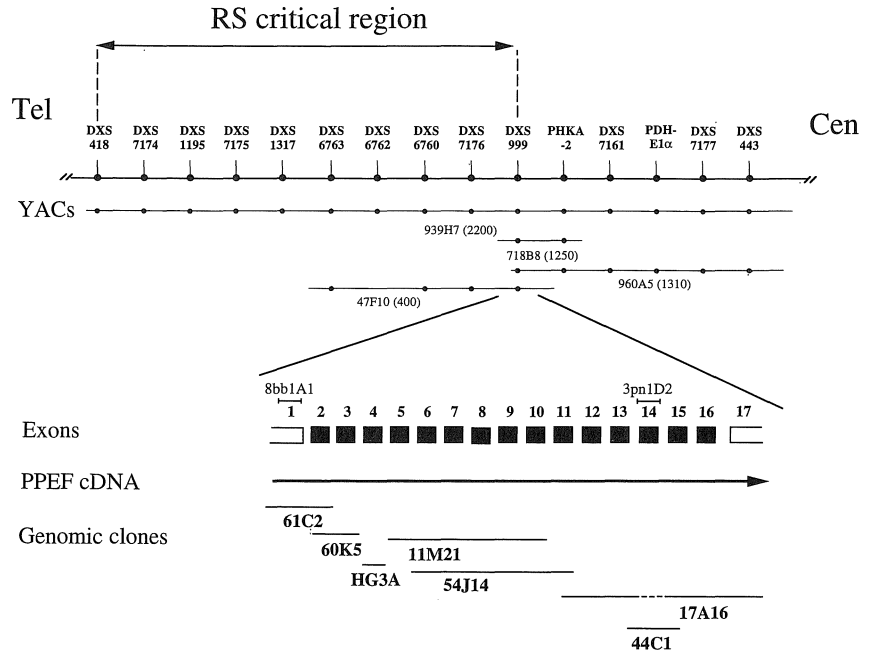
**Figure 1.** Map of the RS critical region. The map displays the order of markers and the YACs positive for two or more markers. The RS critical region spans from *DXS418* distally to *DXS999* proximally. YAC clones are indicated by thin bars. Cosmid and phage clones are indicated by thick bars. The number and position of the exons are shown.

mutant phenotypes, and found to correspond to DRES10 (*Drosophila* related expressed sequence #10) (4).

The I.M.A.G.E. cDNA clone 51064 corresponding to EST H18854 was used to screen both an infant (5) and a fetal (Clontech) brain cDNA library. Six different cDNA clones were isolated and characterized by end-sequencing, restriction mapping and PCR, using specific primers in combination with vector primers. Subsequently, an additional cDNA (nt19) was isolated by screening a teratocarcinoma/neuron (mature hNT neuron, Stratagene 937233) cDNA library with a different exon trapping product (clone 8bb1a1). Characterization of each of the cDNA clones allowed us to establish a cDNA contig of 2872 bp (base pairs) (accession no. X97867). The authenticity of the 5′ end of the cDNA was validated by sequencing the corresponding genomic region. The putative initiation codon was identified at position 484 and is located within a nucleotide sequence that fulfills Kozak's criteria for an initiation codon (6). The first in-frame stop codon (TAA) was identified at nucleotide 2443, predicting a protein product of 653 amino acids. Sequence analysis of the predicted protein product revealed the presence of two putative functional domains. The first domain comprises amino acids 150–438 and shares sequence similarity with the catalytic domain of phosphoprotein phosphatases of the serine-threonine kind (Fig. 2A). The second domain is present at the carboxyl end of the predicted protein (amino acids 566–643) and contains potential $Ca^{++}$-binding sites as defined by the EF hand motif (7) (Fig. 2A and B). On the basis of these homologies, we therefore named the gene *PPEF*, for Phosphoprotein Phosphatase with EF hand motif.

Sequence analysis and comparison of *PPEF* with previously identified phosphoprotein phosphatases revealed conservation of several invariant residues including Asp59, Asp88, His61 and His139. Site-directed-mutagenesis experiments revealed that substitution of these residues results in abolishment of phosphatase activity (8,9). These data suggest that PPEF might be a functional phosphatase.

A BLASTX search with PPEF disclosed the highest homology with the *Drosophila rdgC* gene, which also encodes for a serine-threonine phosphatase with EF motif and, when mutated, causes light-induced retinal degeneration in the fly. *PPEF* and *rdgc* are 60% identical at the nucleotide level, and 62% similar and 42% identical at the protein level. To the best of our knowledge, the *rdgC* and PPEF proteins are the only two molecules in which a phosphoprotein phosphatase domain coexists with EF hand motifs. Furthermore, within the phosphoprotein phosphatase domain, PPEF displays much higher homology with *rdgC* than with other members of the phosphoprotein phosphatase gene family (data not shown), indicating that PPEF and *rdgC* belong to a distinct subfamily of serine-threonine phosphatases.

### *PPEF* genomic structure and RS patient analysis

The complete genomic structure of the *PPEF* gene was determined using the available cosmid clones. Genomic clones corresponding to exon 4 were obtained by screening a total phage genomic library using a specific PCR probe. *Eco*RI and *Hind*III

122

**Figure 2.** Sequence analysis of the PPEF predicted protein. (A) Amino acid sequence comparison between PPEF (top) and *Drosophila rdg*C (bottom) proteins. The region of homology shared with other members of the protein phosphatase family is underlined. The EF calcium-binding domain is indicated by a dashed line. (B) Amino acid identity within the EF hand motifs between PPEF *rdg*C, Troponin C and Plastin. The positions of the $Ca^{++}$ chelating side chains are labeled X–Z.

cosmid and phage fragments hybridizing to the cDNA were subcloned in pBluescript and sequenced using primers designed from the cDNA sequence. Seventeen exons were identified and the sequences of all intron–exon boundaries were determined. Exon sizes and splice junction sequences are shown in Table 1. Junction sequences are in agreement with 5′ and 3′ splice site consensus motifs, with the exception of the exon 15 5′ splice site which shows the sequence GCAAGTG, instead of the consensus 5′ splice site sequence $G_{100}T_{100}A_{62}A_{68}G_{84}T_{63}$. Differences in the GT dinucleotide have been reported in 0.13% of splice site sequences (10). The exon trapping products 8bb1A1 and 3pn1D2 were found to correspond to the first untranslated exon and to exon 14, respectively (Fig. 1). Figure 1 shows the map of the genomic clones and the position of the exons.

Sequence homology and mapping data identified the *PPEF* gene as a good candidate for retinoschisis and thus, a mutation study in RS patients was undertaken. Thirty-seven unrelated male patients with clinical features of RS, but with no reported chromosomal abnormality involving the Xp22 region, were tested for small rearrangements or point mutations in the *PPEF* gene by Southern and SSCP analyses (data not shown). Furthermore, PTT analysis was carried out on nine different RS patients (data not shown). No anomalies were found with either technique, thus suggesting that *PPEF* is not involved in the pathogenesis of RS.

**PPEF is alternatively spliced**

RT-PCR experiments, using nested primers, on RNA isolated from lymphocytes of RS patients and normal controls revealed several different size products (Fig. 3B). Similar results were obtained using RT-PCR on RNA isolated from seven other tissues (data not shown). Sequence analysis of these different products and comparison with the cDNA sequence confirmed that they were the result of alternative splicing. Figure 3A shows a schematic representation of all the splice variants detected. RT-PCR1a corresponds to a product in which exon 5 was spliced out, resulting in a truncated protein of 600 amino acids. RT-PCR1b is the result of using a cryptic splice site within exon 11. This product lacks 84 bp and, therefore, encodes a protein deleted of 28 amino acids. RT-PCR2 corresponds to a product in which exons 12 and 13 were spliced out. This splicing event causes the premature termination of the protein. Finally, RT-PCR3 corresponds to a product in which only exon 12 was spliced out. This alternative spliced variant results in an in-frame deletion of 186 bp and, thus, in a protein which is missing 86 amino acids. These variant forms have never been identified during cDNA library screening, and may not have any functional significance.

**Table 1.** Splice junction sequences of the PPEF gene

| Exon number | Splice junctions 3′ splice site | 5′ splice site | Exon size (bp) |
|---|---|---|---|
| 1 | ttttctttcagCCCTCTG | AAACAGgtaatatgtt | 103 |
| 2 | tcctcattagAATCTAT | ACACATgtgagtactg | 168 |
| 3 | gggtctgcagCACTGA | ATGCAGgtctgttttg | 128 |
| 4 | ttttgtccagTTATCC | AGCTAGgtaagtaaaa | 61 |
| 5 | cctttccagAATTAA | CAACAGgtaagtggaa | 161 |
| 6 | ccgctcacagATACTT | TCTGTGgtaagtttca | 115 |
| 7 | tgcactgcagGTGATT | TACAAGgtaaatgatg | 47 |
| 8 | cttccacagAATGGT | TCTGAGgtaacccag | 167 |
| 9 | tttttttttagGTATGG | TATAAGgtaagacatg | 37 |
| 10 | ttacttacagCTACAT | AACAAGgtaagaagta | 150 |
| 11 | tgttttatagATGAAA | GAACAGgtaggtaatc | 153 |
| 12 | ttgttccaagATTATT | GGGAAGgtaagctaaa | 86 |
| 13 | ctttaaccagGTGGTG | CCAAAGgtgtgtatac | 143 |
| 4 | ttttctttgcagAGTGG | AATCAGgtaacaaatt | 107 |
| 15 | ttgttgttagGAAAAC | CAAGAGgcaagtgaaa | 164 |
| 16 | tttattttagGCTCAT | ACTCAGgtaaataaat | 85 |
| last | attcttttagGCCTGA | – | |

## Expression studies

The expression pattern of the *PPEF* transcript was determined by both Northern analysis on human tissues and RNA *in situ* hybridization on embryonic mouse tissue sections. Northern blot analysis on both human adult and fetal tissues detected two transcripts (2.7 and 4.3 kb, respectively) selectively expressed in the brain. These transcripts were found to be strongly up-regulated during fetal life (Fig. 4).

Previous zoo blot experiments indicated evolutionary conservation of the *PPEF* gene in several species including hamster, rat, mouse, pig, chicken and cow (data not shown). To obtain a probe for RNA *in situ* hybridization studies on mouse tissue sections, we screened a mouse embryonic E11.5 day cDNA library (Clontech) using probe c14-c22 obtained by PCR with primers c14 and c22. This PCR product corresponds to the region with the highest sequence homology (63.2% at the nucleotide level) between the human *PPEF* and the *Drosophila rdgC* genes. This screening led to the isolation of a partial cDNA clone (EM800; accession no. Y08234). Sequence analysis revealed 82.4% identity at the nucleotide level, and 83.1% similarity and 76.4% identity at the protein level between mouse and human (data not shown). The homology between the mouse and *Drosophila* proteins was revealed to be 65.3% similarity and 43.1% identity.

RNA *in situ* hybridization was performed on mouse embryonic tissue sections from embryonic day 10.5 (E10.5) to embryonic day 16.5 (E16.5). These experiments revealed an expression pattern overtly different from that displayed by the *rdgC* gene in the fly. *Ppef* is almost exclusively expressed in the peripheral nervous system, within sensory neurons of neural crest origin. *Ppef* expression is up-regulated at E12.5 in dorsal root ganglia
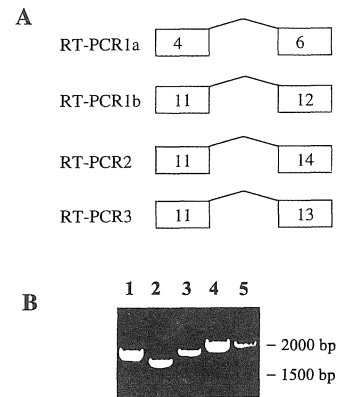


**Figure 3.** (**A**) Schematic representation of different splice variants identified by RT-PCR experiments, using nested primers. (**B**) RT-PCR analysis on RNA isolated from lymphocytes of RS patients and normal controls. Lane 1 corresponds to the products RT-PCR1a and RT-PCR1b. Lane 2 to RT-PCR2, lane 3 to RT-PCR3. Lane 4 corresponds to a product identical to the cDNA described in the text. Lane 5 contains a 100 bp marker.

(DRG), and in some sensory cranial ganglia (Fig. 5A). Sensory neurons of the vertebrate peripheral nervous system have two distinct embryological origins. Several studies, mainly in the chick, have shown that neurons of the DRG, of the dorso-medial part of the trigeminal ganglion, and of the superior ganglia of cranial nerves IX and X (jugular ganglion) are derived from the neural crest, while neurons of the ventrolateral part of the trigeminal ganglion, the geniculate, vestibuloacustic, petrosal and nodose ganglia, are derived from ectodermal placodes, i.e., thickening of the surface ectoderm (11). *Ppef* expression was found to be restricted to neuronal populations of neural crest-derived sensory ganglia. In fact, *Ppef* is highly expressed in neurons of the DRG (Fig. 5C and D), in distinct neuronal populations of the trigeminal ganglion (Fig. 5B), and in the superior ganglia of the IX and X cranial nerves. No expression was observed in the geniculate, vestibuloacustic (Fig. 5B), petrosal, and nodose ganglia.

Although sensory ganglia are already formed in the mouse well before E12.5, this stage is believed to correspond with the start of neurogenesis in these structures (12). An enlarged view of *Ppef* expression within DRG and the trigeminal ganglion (Fig. 5B) clearly demonstrated that most, but not all, neurons express *Ppef*. A slightly decreased expression could still be detected in sensory ganglia at E16.5. We do not know if *Ppef* expression in sensory neurons persists in adult life. The discrepancy in the strength of the hybridization signals on Northern blot in adult and fetal brain is probably due to the inclusion of the cranial sensory ganglia in the brain sample of fetal origin. The only sites of *Ppef* expression, outside sensory neurons, were found in the inner ear (Fig. 5E) and in a small group of neurons located at the midbrain/pons junction (Fig. 5F). We did not detect any expression signals above background level in the embryonic and adult mouse retina (data not shown).
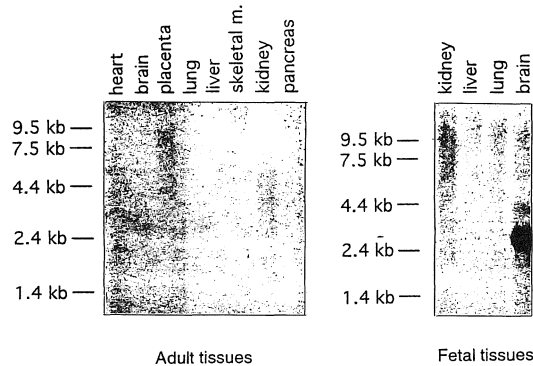
**Figure 4.** Northern blot analysis of the PPEF gene. PolyA⁺ RNA from multiple adult and fetal tissues hybridized to cDNA corresponding to the PPEF gene. A 2.7 kb transcript is detected in both adult and fetal brain, while a less abundant 4.3 kb band is detected only in fetal brain.

### Linkage mapping of *Ppef* in mouse

The finding in the N2 progeny of the BSS backcross (13) of male individuals carrying only B alleles (hemizygotes) led to the unequivocal assignment of *Ppef* to the murine X chromosome. The analysis of the strain distribution pattern (SDP) observed in the same progeny permitted the localization of *Ppef* to the distal third of the chromosome (Fig. 6). *Ppef* is in linkage with DXXrf132 (Spencer *et al.*, in preparation) (θ = 1.41; LOD = 19.1), and was mapped using a probe containing dbEST 122118 (accession no. F06456) with homology to the *Saccharomyces cerevisiae* GTR1 gene. In the BSS map, *Ppef* is located telomeric to DXXrf132. *Ppef* maps ~1.5 cM centromeric to *Grpr*, the gastrin releasing peptide receptor gene, identified using DXMit20 primers (Korobova and Arnheim, unpublished) and *Piga* (θ = 1.37; LOD = 19.7). The locus order defined in mouse between the *Amel* (Korobova and Arnheim, unpublished) and *Ppef* loci correlates with the physical map established in human (1). However, the position of the *Oa1* gene with respect to the *Clc4-1-Amel-(Piga-Grpr)-Ppef* linkage group confirms the prior notion of a rearrangement (inversion) within this region of human–mouse synteny (14).

## DISCUSSION

### PPEF is the first mammalian member of the protein phosphatase with EF hand motif gene family

A wide variety of cellular functions, including cell signalling, gene expression, membrane transport and secretion and cell division, are regulated by the reversible phosphorylation of proteins on serine and threonine residues (15). The phosphatases that catalyze the dephosphorylation of these amino acids are a crucial component of this regulation. The serine-threonine phosphatases belong to a rapidly expanding gene family, and six different mammalian members (PP1, PP2A, PP2B, PP2C, PP4 and PP5) have been described so far. They can be distinguished depending on their ability to dephosphorylase either the α- or the β-subunit of phosphorylase kinase, and on their sensitivity to specific inhibitors (16). These phosphatases have been highly conserved during evolution from yeasts to vertebrates (17).

In *Drosophila*, the *rdgC* gene encodes a serine/threonine phosphatase. This phosphatase shares 30% homology with the catalytic domain of PP1, PP2A and PP2B, but is unique due to the presence of five Ca⁺⁺-binding sites, as defined by the EF hand motif in the C-terminus. Owing to its particular features, *rdgC* is likely to be a member of a novel subfamily of protein phosphatases, characterized by the coexistence of the catalytic phosphatase domain and Ca⁺⁺-binding sites. Very little is known about the function of this phosphatase and no evidence for vertebrate homologs has been produced so far. The *rdgC* gene is required in *Drosophila* to prevent light-induced degeneration of the retina. *Drosophila rdgC* mutants show normal retinal morphology and photoreceptor physiology at a young age. The retina of one-day-old *rdgC* mutants has wild type structure, but by three days, the photoreceptors R1–R6 begin degenerating. By five days, degeneration of photoreceptors R1–R6 is complete and photoreceptors R7 and R8 begin showing signs of degeneration (3). The *rdgC* gene is thought to be involved in the regeneration of rhodopsin and is expressed in the retina, ocelli, optic lobes and in the mushroom bodies of the central brain (2).

We report here the isolation of the first mammalian member of the serine-threonine phosphatase with EF hand motif gene family. The new gene was named *PPEF* (Protein Phosphatase with EF calcium-binding domain) and is highly homologous to the *Drosophila rdgC* gene. In order to study the expression pattern of PPEF during embryonic development, we isolated a partial murine homologous cDNA. This cDNA shows 82.4% identity at the nucleotide level and 83.1% similarity and 76.4% identity at the protein level with the human *PPEF* gene. Linkage mapping experiments, performed in the mouse, established that this gene is located on the murine X chromosome in a region synthenic with human Xp22 (18). These data strongly indicate that the murine *Ppef* gene may be considered a *bona fide* ortholog for the human *PPEF*.

### *PPEF* expression is restricted to sensory neurons of neural crest origin

In the fly, *rdgC* is highly expressed in the compound eye as well as in the ocelli, the other photoreceptor-containing organ, and is thought to participate in a rhodopsin-initiated pathway that regulates photoreceptor membrane renewal. The tissue
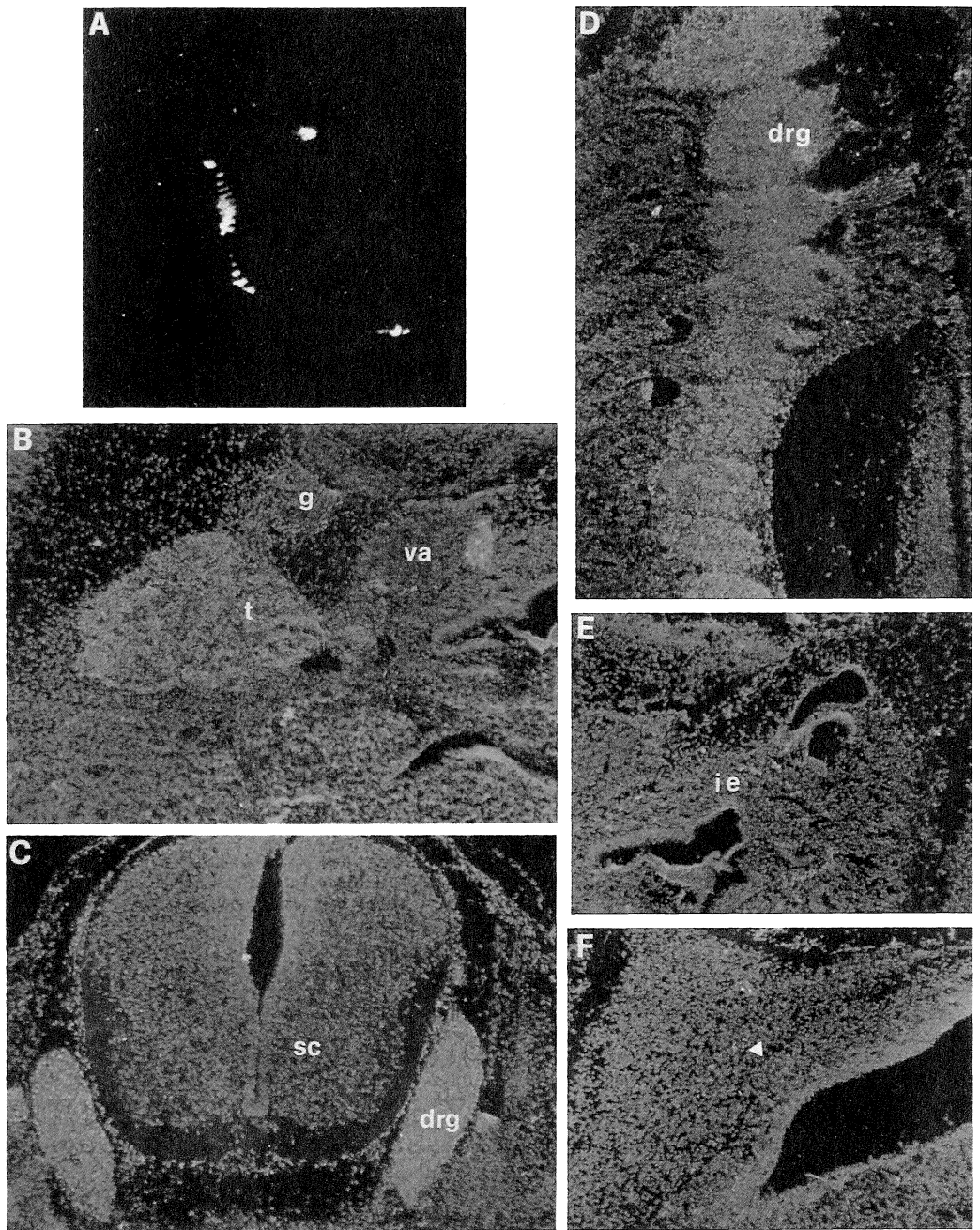
**Figure 5.** Expression pattern of PPEF, as revealed by *in situ* hybridization in a 13.5 day old mouse embryo. (A) Autoradiography of a sagittal section shows expression restricted to the trigeminal ganglion and the dorsal root ganglia. Sagittal sections through the head show PPEF expression in the trigeminal ganglion (B). A strong hybridization signal is present in the dorsal root ganglia, as displayed in a sagittal section (C) and in a transverse section (D). Ppef positive signal is also present in the inner ear (E) and in a group of neurons located at the pons/midbrain junction (arrow in F). Abbreviations: tg, trigeminal ganglion; g, geniculate ganglion; va, vestibuloacoustic ganglion; drg, dorsal root ganglia; sc, spinal cord; ie, inner ear.

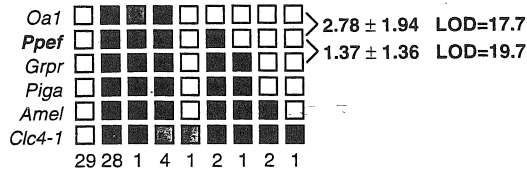| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Oa1 | ☐ | ■ | ■ | ☐ | ☐ | ☐ | ☐ | ☐ | |
| Ppef | ☐ | ■ | ■ | ☐ | ■ | ☐ | ☐ | ☐ | |
| Grpr | ☐ | ■ | ■ | ☐ | ■ | ■ | ☐ | ☐ | |
| Piga | ☐ | ■ | ■ | ■ | ■ | ■ | ☐ | ☐ | |
| Amel | ☐ | ■ | ■ | ☐ | ■ | ■ | ■ | ☐ | |
| Clc4-1 | ☐ | ■ | ■ | ▨ | ▨ | ■ | ■ | ■ | |

2.78 ± 1.94  LOD=17.7
1.37 ± 1.36  LOD=19.7

29 28 1 4 1 2 1 2 1

**Figure 6.** Haplotype and linkage analysis of *Ppef* and flanking loci in the murine X chromosome through the analysis of the BSS backcross (Jackson Laboratory, Bar Harbor, ME). Empty squares indicate the *Mus spretus* allele; solid squares indicate the C57BL/6J allele; stippled squares indicate genotype not determined. Numbers to the right, between rows, indicate recombination fractions ± standard error, and LOD scores. Columns represent different haplotypes observed on the X chromosome. Numbers below columns define the number of individuals in the progeny sharing each haplotype.

localization of the mouse *Ppef* gene was studied by RNA *in situ* hybridization on embryonic tissue sections. In contrast with the expression pattern of the *Drosophila rdgC* gene, we found that *Ppef* is selectively expressed in neurons of the DRG, in distinct neuronal populations of the trigeminal ganglion, and in the superior ganglia of the IX and X cranial nerves. No expression was observed in the geniculate, vestibuloacustic, petrosal and nodose ganglia. This selective pattern of expression correlates with embryological origin, with Ppef being a marker specific for sensory neurons of neural crest origin. Several genes have been found to be expressed in sensory neurons and RNA *in situ* hybridization studies have allowed the study of the distribution of specific transcripts within sensory ganglia. In addition, immunocytochemical and histochemical techniques have been used extensively to biochemically define distinct neuronal populations in sensory ganglia by the presence of peptides, enzymes and specific carbohydrate groups (19,20). So far, however, the only other genes that show an expression pattern restricted to sensory neurons are the neurotrophin receptor genes (21).

For their survival, sensory neurons of the developing peripheral nervous system rely upon specific members of the Nerve Growth Factor (NGF) family of neurotrophins, which are secreted by their targets. It is believed that different neuronal populations in DRG are responsive to different neurotrophins and express different neurotrophin receptors (22,23). *In vitro* studies of neurons from cranial sensory ganglia have shown that there is a difference in the response of placode-derived and neural crest-derived neurons to neurotrophins; neural crest-derived neurons are responsive to NGF, while placode-derived neurons respond to Brain-Derived Neurotrophic Factor (BDNF) (24,25).

The selective expression of *Ppef* in cranial sensory ganglia of neural crest-origin and the lack of expression in placode-derived ganglia suggest that this gene is expressed by NGF-responsive neurons. NGF exerts its effect by eliciting a phosphorylation cascade through activation of the TrkA tyrosine-kinase receptor (26). Although the phosphorylation events mediated by the binding of NGF to the TrkA receptor have been extensively studied, very little is known about dephosphorylation pathways and the phosphatases involved. It is an appealing hypothesis that PPEF might be involved in the specific signalling pathway initiated by NGF in sensory neurons. Consistently, up-regulation of *Ppef* expression coincides with the time at which neurogenesis and TrkA receptor expression begins in sensory ganglia.

In conclusion, our study provides evidence for the presence of a mammalian protein phosphatase with EF hand motifs. Although this phosphatase is highly homologous to the *Drosophila rdgC* gene, expression studies seem to suggest that it represents a related gene with a different function. We do not exclude the possibility that additional members of this protein phosphatase gene family will be identified in the near future. Alternatively, *PPEF* might be the evolutionary equivalent of the *Drosophila rdgC* gene, but we must assume that the two genes have diverged in terms of function. Since mapping data placed *PPEF* within the critical region for RS, we have excluded its involvement in the pathogenesis of X-linked juvenile RS by searching mutations in RS patients. Future experiments, including the identification of sub-populations of sensory neurons expressing PPEF phosphatase and of the natural substrate of the protein, will greatly contribute to the understanding of the biological role of the *PPEF* gene in mammalian development and will allow testing for its possible involvement in NGF signal transduction.

## MATERIALS AND METHODS

### cDNA identification

YAC clone 939H7 was converted into cosmid clones by hybridization of long range *Alu*-PCR product obtained with a variety of human-specific *Alu* primers. PCR amplification was performed according to Gu *et al.* (27). Cosmid clones were grouped (10 clones per group), digested with *Bam*HI/*Bgl*II, cloned in the pSPL3 vector, and used for the exon amplification experiments as described previously (28) and by Montini *et al.* (manuscript in preparation). In order to identify the *PPEF* full-length transcript, three human cDNA libraries were screened: a teratocarcinoma/neuron cDNA library (mature hNT neuron, Stratagene 937233), a fetal brain cDNA library (Clontech HL3003a), and the Bento Soares infant brain 1NIB arrayed cDNA library. For the isolation of a partial murine cDNA clone, an 11.5 day embryo (Clontech ML3003a) mouse cDNA library was used. Plating, hybridization and washing conditions were performed as previously described (29). Primers used to obtain the probe used to screen the mouse cDNA library were as follows:

c14, AAGTCCTGAAGCAAATGCCG;
c22, GCCATACCTCAGATTCATC.

### cDNA sequence analysis

cDNA sequence analysis and nucleotide and protein database searches were performed as previously described (4). Data on similarity/identity were obtained using the Bestfit program of the GCG software package, version 8.1. The multiple alignment analyses were generated using the PileUp program of the Wisconsin GCG software package, ver. 8.1.

### Expression studies

Commercial Northern blots (Clontech) containing human RNA from fetal and adult tissues were hybridized and washed using the conditions recommended by the manufacturer. Mouse embryo tissue sections were prepared and RNA *in situ* hybridization experiments were performed as previously described (30). cDNA clone EM800 was linearized with appropriate restriction enzymes to transcribe either sense or antisense ^{35}S-labelled riboprobes. Slides were exposed for 10 days. Micrographs are double

127

exposures: red represents the *in situ* hybridization signal, and blue shows the nuclei stained with Hoechst 33258 dye.

## RNA isolation and RT-PCR

RNA isolation from blood and RT-PCR experiments were carried out as previously described (31). RT-PCR was carried out with nested primers. Primers 1F+1R were used for the first round, the resulting PCR products were then reamplified using primers 2F+2R. Primers used for RT-PCR:

1F, 5'-TGAAGGCCAGACAACACTATG-3';
1R, 5'-ACTAGGTCCAACCCAGTTTCT-3';
2F, 5'-TAATACGACTCACTATAGGAACAGACCACCAT GGAATATGCTGATGAACAAGGC-3'
2R, 5'-CTTTTCTCTGCTACTGACTATGAA-3'.

Primers for sequencing:

909, 5'-CTTGGAAGAATTCTATGCCTGG-3';
947, 5'-ATTGTACCGATTGGGAGCCA-3';
1528, 5'-CCATAGTATCCACTCTTTGGCGA-3';
1501, 5'-CCTCTTCGCCAAAGAGTGGATA-3'.

## Linkage mapping in mouse

Genetic mapping was achieved utilizing a (C57BL/6j × SPRET/Ei)F1 × SPRET/Ei (BSS) backcross generated and distributed by the Jackson Laboratory (Bar Harbor, ME) (13). An *Msp*I RFLP was identified by hybridization of C57BL/6j and SPRET/Ei parental DNAs cut with each of the six restriction enzymes (*Eco*RI, *Eco*RV, *Kpn*I, *Msp*I, *Taq*I and *Xba*I). Four Southern panels containing *Msp*I-cut parental DNAs and N2 progeny ($n = 94$) DNAs were hybridized with a *Ppef* cDNA probe. The resulting strain distribution pattern (SDP) was analyzed with the Map Manager 2.6 program (32).

## ACKNOWLEDGEMENTS

## REFERENCES

1. Ferrero, G.B., Franco, B., Roth, E.J., Firulli, B.A., Borsani, G., Delmas-Mata, J., Weissenbach, J., Halley, G., Schlessinger, D., Chinault, A.C., Zoghbi, H.Y., L., N.D. and Ballabio, A. (1995) An integrated physical and genetic map of a 35 Mb region on chromosome Xp22.3–Xp21.3. *Hum. Mol. Genet.* **4**, 1821–1827.
2. Steele, F.R., Washburn, T., Rieger, R. and O'Tousa, J.E. (1992) Drosophila *retinal degeneration C (rdgC)* encodes a novel serine/threonine protein phosphatase. *Cell* **69**, 669–676.
3. Steele, F. and O'Tousa, J.E. (1990) Rhodopsin activation causes retinal degeneration in drosophila *rdgC* mutant. *Neuron* **4**, 883–890.
4. Banfi, S., Borsani, G., Rossi, E., Bernard, L., Guffanti, A., Rubboli, F., Marchitiello, A., Giglio, S., Coluccia, E., Zollo, M., Zuffardi, O. and Ballabio, A. (1996) Identification and mapping of human cDNAs homologous to *Drosophila* mutant genes through EST database searching. *Nature Genet.* **13**, 167–174.

5. Soares, M.B., Bonaldo, M.F., Jelene, P., Su, L., Lawton, L. and Efstratiadid, A. (1994) Construction and characterization of a normalized cDNA library. *Proc. Natl. Acad. Sci. USA* **91**, 9228–9232.
6. Kozak, M. (1984) Compilation and analysis of sequences upstream from the translational start site in eukaryotic mRNAs. *Nucleic Acids Res.* **12**, 857–873.
7. Kretsinger, R.H. (1980) Structure and evolution of calcium-modulated proteins. *CRC Crit. Rev. Biochem.* **8**, 119–174.
8. Barton, G.J., Cohen, P.T.W. and Barford, D. (1994) Conservation analysis and structure prediction of the protein serine/threonine phosphatases: Sequence similarity with diadenosine tetraphosphatase from *Escherichia coli* suggests homology to the protein phosphatases. *Eur. J. Biochem.* **220**, 225–237.
9. Barford, D. (1996) Molecular mechanisms of the protein serine/threonine phosphatases. *Trends Biochem. Sci.* **251**, 407–412.
10. Shapiro, M.B. and Senapathy, P. (1987) RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. *Nucleic Acids Res.* **15**, 7155–7174.
11. Le Douarin, N.M., Kalcheim, C. and Teillet, M.-A. (1992). The cellular and molecular basis of early sensory ganglion development. In Scott, S.A. (ed.), *Sensory Neurons: Diversity, Development, and Plasticity.* Oxford University Press, New York, pp. 143–170.
12. Altman, J. and Bayer, S.A. (1984) The development of the rat spinal cord. *Adv. Anat. Embryol. Cell. Biol.* **85**, 1–164.
13. Rowe, L.B., Nadeau, J.H., Turner, R., Frankel, W.N., Letts, V.A., Eppig, J.T., Ko, M.S.H., Thurston, S.J. and Birkenmeier, E.H. (1994) Maps from two interspecific backcross DNA panels available as a community genetic mapping resource. *Mamm. Genome* **5**, 253–274.
14. Dinulos, M.B., Bassi, M.T., Rugarli, E.I., Chapman, V., Ballabio, A. and Disteche, C.M. (1996) A new region of conservation is defined between human and mouse X chromosomes. *Genomics* **35**, 244–247.
15. Fischer, E.H. and Krebs, E.G. (1990) *Biochim. Biophys. Acta* **1000**, 297–301.
16. Cohen, P.T.W., Brewis, N.D., Hughes, V. and Mann, D.J. (1990) Protein serine/threonine phosphatases; an expanding family. *FEBS Lett.* **268**, 355–359.
17. Chen, M.X., McPartlin, A.E., Brown, L., Chen, Y.H., Barker, H.M. and Cohen, P.T.W. (1994) A novel human protein serine/threonine phosphatase, which possesses four tetratricopeptide repeat motifs and localizes to the nucleus. *EMBO J.* **13**, 4278–4290.
18. Herman, G.E., Boyd, Y., Brown, S.D.M., de Gouyon, B., Haynes, A. and Quaderi, N. (1996) Encyclopedia of the mouse genome V. Mouse X chromosome. *Mamm. Genome* **6**, S317–S330.
19. Lawson, S.N. (1992). Morphological and biochemical cell types of sensory neurons. In Scott, S.A. (ed.), *Sensory Neurons: Diversity, Development, and Plasticity.* Oxford University Press, New York, pp. 27–59.
20. Hunt, S.P., Mantyh, P.W. and Priestley, J.V. (1992) The organization of biochemically characterized sensory neurons. In Scott, S.A. (ed.), *Sensory Neurons: Diversity, Development, and Plasticity.* Oxford University Press, New York, pp. 60–76.
21. Martin-Zanca, D., Barbacid, M. and Parada, L.F. (1990) Expression of the *trk* proto-oncogene is restricted to the sensory cranial and spinal ganglia of neural crest origin in mouse development. *Genes Dev.* **4**, 683–694.
22. McMahon, S.B., Armanini, M.P., Ling, L.H. and Phillips, H.S. (1994) Expression and coexpression of Trk receptors in subpopulations of adult primary sensory neurons projecting to identified peripheral targets. *Neuron* **12**, 1161–1171.
23. Snider, W.D. and Wright, D.E. (1996) Neurotrophins cause a new sensation. *Neuron* **16**, 229–232.
24. Lindsay, R.M., Thoenen, H. and Barde, Y.-A. (1985) Placode and neural crest-derived sensory neurons are responsive at early developmental stages to brain-derived neurotrophic factor. *Dev. Biol.* **112**, 319–328.
25. Davies, A.M. and Lindsay, R.M. (1985) The cranial sensory ganglia in culture: differences in the response of placode-derived and neural crest-derived neurons to nerve growth factor. *Dev. Biol.* **111**, 62–72.
26. Bothwell, M. (1995) Functional interactions of neurotrophins and neurotrophin receptors. *Annu. Rev. Neurosci.* **18**, 223–253.
27. Gu, Y., Ferrero, G.B., Schofield, T., Franco, B., Lee, C.C., Graves, M., Arenson, A., Ballabio, A., Chinault, A.C. and Nelson, D.L. (1995) Towards a cosmid-based map of the distal short arm of the human X chromosome. Abstract presented at the 6th X Chromosome Workshop, Banff, Alberta, Canada, pp. B3.
28. Church, D.M., Stotler, C.J., Rutter, J.L., Murrell, J.R., Trofatter, J.A. and Buckler, A.J. (1994) Isolation of genes from complex sources of mammalian genomic DNA using exon amplification. *Nature Genet.* **6**, 98–105.

29. Franco, B., Guioli, S., Pragliola, A., Incerti, B., Bardoni, B., Tonlorenzi, R., Carrozzo, R., Maestrini, E., Pieretti, M., Taillon-Miller, P., Brown, C.J., Willard, H.F., Lawrence, C., Persico, M.G., Camerino, G. and Ballabio, A. (1991) A gene deleted in Kallmann's syndrome shares homology with neural cell adhesion and axonal path-finding molecules. *Nature* **353**, 529–536.

30. Rugarli, E.I., Lutz, B., Kuratani, S.C., Wawersik, S., Borsani, G., Ballabio, A. and Eichele, G. (1993) Expression pattern of the Kallmann syndrome gene in the olfactory system suggests a role in neuronal targeting. *Nature Genet.* **4**, 19–25.

31. den Dunnen, J.T., Roest, P.A.M., Van der Luijt, R.B. and Hogervorst, F.B.L. (1996). The Protein Truncation Test (PTT) for rapid detection of translation-terminating mutations. In Pfeifer, G.P. (ed.), *Technologies for Detection of DNA Damage and Mutations*. Plenum Press, New York, pp. 323–341.

32. Manly, K.F. and Elliott, R.W. (1991) RI manager, a microcomputer program for the analysis of data from recombinant inbred strains. *Mamm. Genome* **1**, 123–126.

# CHAPTER 5.1

**Exclusion of *PPEF* as the gene causing X-linked juvenile retinoschisis**

Van de Vosse, E., Franco, B., Van der Bent, P., Montini, E., Orth, U., Hanauer, A., Van Ommen, G.J.B., Ballabio, A., Den Dunnen, J.T., and Bergen, A.A.B.

# Exclusion of *PPEF* as the gene causing X-linked juvenile retinoschisis

**Abstract** X-linked juvenile retinoschisis (RS) is a progressive vitreoretinal degeneration localised in Xp22.1-p22.2. A human homologue of the retinal degeneration gene C (*rdgC*), a gene that in *Drosophila melanogaster* prevents light-induced retinal degeneration, was localised in the RS obligate gene region. We have tested the gene, designated *PPEF* in humans, as a candidate gene in RS patients using RT-PCR and the protein truncation test on RNA and SSCP on DNA. No mutations were identified, making it highly unlikely that *PPEF* is the gene implicated in RS. The data presented facilitate mutation analysis of the *PPEF*-gene in other diseases which have been or will be localised to this region.

## Introduction

X-linked juvenile retinoschisis (RS, MIM 312700) is a progressive vitreoretinal degeneration, with a frequency of about 1 in 10 000 (Bergen et al.1995). Symptoms vary from mild loss of visual acuity and peripheral field defects to total blindness due to complete retinal detachment; the age of onset is variable

E. van de Vosse · P. van der Bent · G.-J.B. van Ommen
J.T. den Dunnen (✉)
MGC-Department of Human Genetics, Leiden University,
Wassenaarseweg 72, NL-2333 AL Leiden, The Netherlands.
Tel.: +31-71-5276105; Fax: +31-71-5276075
email: ddunnen@ruly46.MedFac.Leidenuniv.NL

B. Franco · E. Montini · A. Ballabio
Telethon Institute of Genetics and Medicine,
Via Olgettina 58, I-20132 Milano, Italy.

U. Orth
Institut für Humangenetik, UKE, Butenfeld 42,
D-22529 Hamburg, Germany.

A. Hanauer
Institut de Genetique et de Biologie Moleculaire et Cellulaire,
CNRS, INSERM, 67404 Illkirch, France.

N. Tijmes · A.A.B. Bergen
The Netherlands Ophthalmic Research Institute, P.O.Box 12141,
NL-1100 AC Amsterdam, The Netherlands.

(George et al.1996). The gene causing RS has been localised by extensive linkage analysis to a region on Xp22 between markers DXS418 and DXS999 (George et al.1994; Van de Vosse et al.1996; Huopaniemi et al.1997). Several YAC contigs have been constructed spanning the 1 Mb obligate gene region (Ferrero et al.1995; Alitalo et al.1995; Van de Vosse et al.1997). Gene identification techniques are currently used to isolate candidate genes for RS and to test these in mutation analysis of patients derived samples.

Exon trapping experiments carried out on YAC clone y939H7, covering the RS-gene candidate region, yield several products. Two of the identified exon trapping products corresponded to a novel human transcript (*PPEF*, Montini et al.1997) which was highly homologous to the retinal degeneration gene C (*rdgC*, M89628). In *Drosophila melanogaster*, this gene is required to prevent light-induced retinal degeneration (Steele and O'Tousa 1995; Steele et al.1992). *PPEF* was mapped back to YAC clone y939H7 close to DXS999 and resides therefore in the RS critical region (Fig 1). The localisation in the RS obligate gene region and the phenotype observed in the fly, made *PPEF* an attractive candidate for RS. *PPEF*, Protein Phosphatase with EF hand motifs, encodes a serine/threonine protein phosphatase (Montini et al.1997). The gene consists of 17 exons, with a coding region
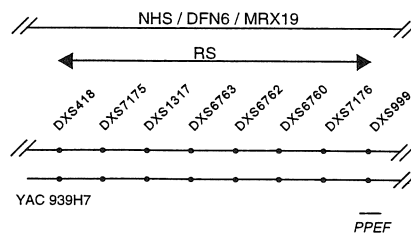


**Fig. 1** Localisation of the *PPEF* gene in the retinoschisis (*RS*), Nance-Horan syndrome (*NHS*), DFN6 and MRX19 candidate regions.

of 1962 bp having a 61.7 % similarity on protein level with *rdgC*. Here we report testing of the *PPEF* gene as a candidate gene for RS.

## Materials and methods

Patient samples

Patient DNA was isolated from blood lymphocytes as described (Maniatis et al.1989). Patient RNA was isolated from blood lymphocytes as described (Den Dunnen et al.1996) using RNAzolB (Campro Scientific). All tissue specific RNAs were obtained from Clontech.

Hybridisation

Southern blots containing *Eco*RI, *Msp*I, *Bam*HI, *Eco*RV, *Hind*III digested DNA of RS patients were made using standard techniques (Maniatis et al.1989). A 25-ng aliquot of DNA was labelled with 5 µl [α-$^{32}$P]-dCTP using the Prime-it II random primer labelling kit (Stratagene). Hybridisation was performed with 10$^6$ cpm hybridisation mix (0.125 M NaHPO$_4$, 0.25 M NaCl, 1 mM EDTA, 7% SDS, 10% PEG-6.000) at 65°C. Filters were washed once for 5 min at room temperature in 2 x SSC, 0.2% SDS and twice for 30 min at 65°C in 2 x SSC, 0.2% SDS. Autoradiography was done by overnight exposure of Kodak X-Omat AR film.

SSCP analysis

SSCP analysis on patient DNA was performed as described by Orita (Orita et al.1989) and adapted for use of non-radioactive samples (Renieri et al.1994) using the primers and conditions given in Table 1.

RT, PCR and protein truncation test (PTT)

Reverse transcription (using 1-3 µg RNA) and two rounds of PCR and

PTT analysis were essentially done according to Den Dunnen et al. (1996). Cycling conditions for the first PCR round: 3 min at 93°C, followed by 30 cycles of 1 min at 93°C, 1 min at 58°C, 4 min at 72°C, and finally one cycle of 7 min at 72°C. Cycling conditions for the second PCR round were 3 min at 93°C, then 32 times 1 min at 93°C, 1 min at 59°C, 3 min at 72°C and finally once for 4 min at 72°C. Approximately 250-500 ng PCR-product was used in a coupled *in vitro* transcription/translation reaction. *PPEF* primers used for RT-PCR (*PPEF-1F, -1R, -2F, -2R*) are indicated in Table 1. Primer *PPEF-2F* has a tail containing a T7 promoter sequence and a eukaryotic translation initiation signal, facilitating subsequent analysis using *in vitro* transcription and translation (Sarkar and Sommer 1989). The primers for the RT-PCR experiments were designed on the sequence available at the time, this sequence did not include the 5' end of the sequence that was revealed later (Montini et al.1997).

Sequence analysis

PCR products were sequenced using an AmpliCycle sequence kit (Perkin Elmer) according to the manufacturers instructions.

## Results and Discussion

In most X-linked diseases, DNA deletions of various sizes form a significant fraction of the mutations found. We carried out a scan for the presence of deletions and rearrangements in genomic DNA by hybridisation of the *PPEF* cDNAs (Montini et al.1997) to Southern blots containing DNA of 60 unrelated RS patients. DNA was digested with *Msp*I, *Bam*HI, *Eco*RV (23 samples), *Hind*III (37 samples) and *Eco*RI (all samples). Neither deletions nor aberrant fragments could be detected with the cDNAs, thereby excluding the presence of large deletions in the *PPEF* gene in these RS patients (data not shown).

For a more detailed mutation analysis of the *PPEF*-gene, we performed SSCP analysis on genomic

**Table 1** Sequences of the primers and conditions used in RT-PCR and SSCP analysis. Primer *PPEF-2F* has a tail containing a T7 promoter sequence and a eukaryotic translation initiation signal (*= GCTAATACGACTCACTATAGGAACAGACCACCATG)

| Primer set | Ta | Product size (bp) | MgCl$_2$ (mM) | Sequence of forward primer | Sequence of reverse primer |
|---|---|---|---|---|---|
| 1 | 53 | 239 | 1.5 | CAGAAGTTGAATTCATGAAC | GTAGTTTCCTATGCTACTC |
| 2 | 55 | 238 | 2.5 | GAAGCACCTACTTCTCCTAAC | CCTCGAGGTCGACGGTATC |
| 3 | 55 | 182 | 1.5 | TTGTCACAGTAGCTGTTTGG | GCTCTTGATGAAGACAATTG |
| 4 | 54 | 318 | 2.5 | AGTGCCTTACATGGGCTAG | GGGCATCTGTTATGTACAAG |
| 5 | 55 | 259 | 1.5 | ACGATGTAGGACCAAGAGG | GCTTGCTCCACCTTTACAG |
| 6 | 56 | 159 | 1.5 | GGGCATTGCATCTTGTTCTC | TATCTGCCCTAAGACTGCCC |
| 7 | 55 | 289 | 1.5 | ACACGGCCTGACTTTAAAAG | CAGCATTTTCCAGAGTGCG |
| 8 | 55 | 185⁻ | 1.5 | TGCATGACTCATGGAAGTAG | AATCTGGTCTTTCTTGGCTC |
| 9 | 51 | 252 | 2.5 | TTCCCTTCTAAATCCCTGAG | CAATAAACTGAACCTGTCAG |
| 10 | 55 | 255 | 1.5 | GAATAAGCAGAGGGTTGGAC | CCCTGTTGTACGTGCGATC |
| 11 | 55 | 281 | 1.5 | CTCACTTGTAAGTTACAGCG | TGTGCTTAGGGGAAGGATC |
| 12 | 52 | 240 | 2.5 | TTTGAGAACTAATGTTACGTG | GTGATACCGTGATACCAG |
| 13 | 52 | 215 | 2.5 | AAATGAAACACAACAGGATG | ATGTAACTTGGTGTGTTAAG |
| 14 | 57 | 275 | 1.5 | TCCCAAGAGGTTGCATTC | CACCCTGGCTAGGTTTTAG |
| 15 | 46 | 202 | 2.5 | AATATGTTCTAACACTTAG | TCAAAGTGTACTCATTTTG |
| 16 | 54 | 312 | 1.5 | ACCCTTGCCTTAGGTGGGTC | TAGCTGTTTCAGGGAGCCTG |
| PPEF-1 | 58 | 2122 | 6.7 | TGAAGGCCAGA CAACACTATG | ACTAGGTCCAACCCAGTTTCT |
| PPEF-2 | 59 | 1960 | 6.7 | *GAATATGCTGATGAACAAGGC | CTTTTCTCTGCTACTGACTATGAA |

DNA of 37 patients. Sixteen primersets were designed to amplify 16 of the 17 exons of the *PPEF* gene, excluding exon 1 containing 5'UTR sequences only. No deletions or aberrant fragments could be detected and no polymorphic fragments were found that could be used as RFLPs (Table 1).

To detect mutations either altering the promotor or splice sites of the gene or causing frame shift mutations leading to a premature stop in the open reading frame, we analysed the gene on RNA level using RT-PCR and PTT (Roest et al.1993). Analysis of RT-PCR products made on RNA isolated from lymphocytes of nine unrelated RS patients revealed products ranging in size from 800 to 1900 bp. *In vitro* transcription and translation of the PCR products showed that the size of the translation products corresponded to open reading frames in accordance with the length of all PCR products. Consequently, no mutations affecting transcription of the *PPEF* gene nor mutations causing premature translation termination could be identified.

Based on the extensive mutation analysis in both DNA (hybridisation and SSCP) and RNA (RT-PCR and PTT), we conclude that *PPEF* is not likely to be the gene implicated in RS. Experiments using RNA *in situ* hybridisation of *PPEF* on mouse embryo tissue sections revealed expression of the gene in the brain and basal ganglia but not in the developing eye (Montini et al.1997). The latter suggests a different function for the mammalian homologue of the *Drosophila rdgC* gene. Given the accumulating evidence from SSCP and *in situ* hybridisation experiments, we decided not to complete the PTT analysis, which did not include the 363 bp at the 5' end of the translated region of the gene.

While our data provisionally exclude the *PPEF* gene as the gene involved in RS, its map position (Fig. 1) renders it a candidate gene for other diseases localised in this region, such as Nance-Horan syndrome (Bergen et al.1994), MRX19 (Donnelly et al.1994) and DFN6 (Del Castillo et al.1996). The expression and mutation analysis data presented here, should provide the tools to test the involvement of *PPEF* in these diseases, or others which in the future might be mapped to this region.

## References

Alitalo T, Francis F, Kere J, Lehrach H, Schlessinger D, Willard HF (1995) A 6-Mb YAC contig in Xp22.1-p22.2 spanning the DXS69E, XE59, GLRA2, PIGA, GRPR, CALB3 and PHKA2 genes. Genomics 25:691-700

Bergen AA, Ten Brink J, Schuurman EJ, Bleeker-Wagemakers EM (1994) Nance-Horan syndrome: linkage analysis in a family from the Netherlands. Genomics 21:238-240

Bergen AAB, Ten Brink JB, Van Schooneveld MJ (1995) Efficient DNA carrier detection in X linked juvenile retinoschisis. Br J Ophthalmol 79:683-686

Del Castillo I, Villamar M, Sarduy M, Romero L, Herraiz C, Hernandez FJ, Rodriguez M, Borras I, Montero A, Bellon J, Cruz Tapia M, Moreno F (1996) A novel locus for non-syndromic sensorineural deafness (DFN6) maps to chromosome Xp22. Hum Mol Genet 5:1383-1387

Den Dunnen JT, Roest PAM, Van der Luijt RB, Hogervorst FBL (1996) The Protein Truncation Test (PTT) for rapid detection of translation-terminating mutations. In: Pfeifer GP (ed) Technologies for detection of DNA damage and mutations. Plenum Press, New York: 323-341

Donnelly AJ, Andy Choo KH, Kozman HM, Gedeon AK, Danks DM, Mulley JC (1994) Regional localisation of a non-specific X-linked mental retardation gene (MRX19) to Xp22. Am J Med Genet 51:581-585

Ferrero GB, Franco B, Roth EJ, Firulli BA, Borsani G, Delmas-Mata J, Weissenbach J, Halley G, Schlessinger D, Chinault AC, Zoghbi HY, Nelson DL, Ballabio A (1995) An integrated physical and genetic map of a 35 Mb region on chromosome Xp22.3-Xp21.3. Hum Mol Genet 4:1821-1827

George ND, Yates JR, Moore AT (1996) Clinical features in affected males with X-linked retinoschisis. Arch Ophthalmol 114:274-280

George NDL, Payne SJ, Barton DE, Moore AT, Yates JRW (1994) Genetic mapping of X-linked retinoschisis. Cytogenet Cell Genet 67:354

Huopaniemi L, Rantala A, Tahvanainen E, De la Chapelle A, Alitalo T (1997) Linkage disequilibrium and physical mapping of X-linked juvenile Retinoschisis. Am J Hum Genet 60:1139-1149

Maniatis T, Fritsch E, Sambrook J (1989) Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY

Montini E, Rugarli EI, Van de Vosse E, Andolfi G, Mariani, M., Puca AA, Den Dunnen JT, Ballabio A, Franco B (1997) A novel human serine-threonine phosphatase related to the *Drosophila retinal degeneration C (rdgC)* gene is selectively expressed in sensory neurons of neural crest origin. Hum Mol Genet 6:1137-1145

Orita M, Suzuki Y, Sekiya T, Hayashi K (1989) Rapid and sensitive detection of point mutations and DNA polymorphisms using the polymerase chain reaction. Genomics 5:874-879

Renieri A, Galli L, De Marchi M, Li Volti S, Mollica F, Lupo A, Maschio G, Peissel B, Rossetti S, Pignatti P, et al (1994) Single base pair deletions in exons 39 and 42 of the COL4A5 gene in Alport syndrome. Hum Mol Genet 3:201-202

Roest PAM, Roberts RG, Sugino S, Van Ommen GJB, Den Dunnen JT (1993) Protein truncation test (PTT) for rapid detection of translation-terminating mutations. Hum Mol Genet 10:1719-1721

Sarkar G, Sommer SS (1989) Access to a messenger RNA sequence or its protein product is not limited by tissue or species specificity. Science 244:331-334

Steele F, O'Tousa JE (1995) Rhodopsin activation causes retinal degeneration drosophila rdgC mutant. Neuron 4:883-890

Steele FR, Washburn T, Rieger R, O'Tousa JE (1992) Drosophila retinal degeneration C (rdgC) encodes a novel serine/threonine protein phosphatase. Cell 69:669-676

Van de Vosse E, Bergen AAB, Meershoek EJ, Oosterwijk JC, Gregory S, Bakker B, Weissenbach J, Coffey AJ, Van Ommen GJB, Den Dunnen JT (1996) A Xp22.1-p22.2 YAC contig encompassing the disease loci for RS, KFSD, CLS, HYP and RP15; refined localization of RS. Eur J Hum Genet 4:101-104

Van de Vosse E, Van der Bent P. Heus JJ, Van Ommen GJB, Den Dunnen JT (1997) High resolution mapping by YAC fragmentation of a 2.5 Mb Xp22 region containing the human RS, KFSD and CLS disease genes. Mamm Genome 8:497-501

135

# CHAPTER 5.2

## Exclusion of the *Txp3* gene as the gene causing X-linked juvenile retinoschisis

Van de Vosse, E., Franco, B., Bergen, A.A.B., Van Ommen, G.J.B., Ballabio, A., and Den Dunnen, J.T.

In preparation

# Exclusion of the *Txp3* gene as the gene causing X-linked juvenile retinoschisis

E. van de Vosse, B. Franco, A.A.B. Bergen, G.J.B. van Ommen, A. Ballabio and J.T. den Dunnen.

## Introduction

Using pSPL3 exon trapping on subclones of YAC y939H7, two exons were isolated that were used as probes on a human cDNA library. A cDNA was isolated, that consisted of 7 exons and which had a 2364 bp open reading frame (Figure1). Hybridisation on Northern blots showed strong hybridisation to a product of 9.5 kb in brain and weaker hybridisation to a product of the same length in placenta, lung, kidney and pancreas (Franco and *et al.* 1997). The full length gene, designated *Txp3*, has not been isolated yet. *Txp3* maps close to DXS999, making it a good candidate for X-linked juvenile retinoschisis. We tested *Txp3* as a candidate gene in 9 retinoschisis patients using RT-PCR, a Protein Truncation Test (PTT) and SSCP.

## Materials and methods

### RNA

Patient RNA was isolated from blood lymphocytes basically as described in (Den Dunnen *et al.* 1996) using RNAzolB (Campro Scientific). All tissue specific RNAs were obtained from Clontech.

**Table 1** *Txp3* primers used in RT-PCR.

| Primer name | Primer sequence  (5' -> 3') |
| --- | --- |
| *Txp3-1F* | AGAAGTGGGGGACTCGGC |
| *Txp3-1R* | TTGCCGGCTTGGTTTCTATT |
| *Txp3-2F* | CGGATCCTAATACGACTCACTATAGGAACAGACCACCATG GGAAAGTTCCCACCAACCAGTGA |
| *Txp3-2R* | CGGATCCTTGAAATGTAGGGTGATTCAAA |
| *Txp3-3F* | GAAGGTGCTAGGACCACTTCC |
| *Txp3-3R* | AGTCATCTCTGGAGGGAGCT |
| *Txp3-4F* | CGGATCCTAATACGACTCACTATAGGAACAGACCACCATG GTTAACCATCCTCAGTCC TTGGAAA |
| *Txp3-4R* | CGGATCCAAGAAAAAGATTCGTGAGGTGC |

Primers *Txp3-2F* and *Txp3-4F* have a tail containing a *Bam*HI restriction site, a T7 promoter sequence and a eukaryotic translation initiation signal, facilitating subsequent analysis using *in vitro* transcription and translation (CGGATCCTAATACGACTCACTATAGGAACAGACCACCATG). Primers *Txp3-2R* and *Txp3-4R* have a tail containing a *Bam*HI restriction site.

**Figure 1**
Nucleotide and protein sequence of the *Txp3* gene. Primer binding sites are indicated as boxes, primer names and exons are indicated above the sequence. The change in the nucleotide sequence (from G to TT at 1849) after the sequence was updated is indicated with an arrow.

```
      exon 1                                                    Txp3-1F                                              Txp3-2F
1     gcacgagcggcgacggcgtcctcaggagctgtgggggtcccctgctagaagtgggggactcggcggggggagtcatttaatacttcatgattagaacaaatatgtgaaagttcccaccaacc

121   agtgagaatttcttccttcagacggtttttggatcttactgcacagctttctgagaagttcttttggtgccatgttttgtggcttgcatcaaaagaggagtttgtcttcatgaagattcct

241   aacattggtaatgtgATGAATAAATTTGAGATCCTTGGGGTTGTAGGTGAAGGAGCCTATGGAGTTGTACTTAAATGCAGACACAAGGAAACACATGAAATTGTGGCGATCAAGAAATTC
                     M--N--K--F--E--I--L--G--V--V--G--E--G--A--Y--G--V--V--V--L--K--C--R--H--K--E--T--H--E--I--V--A--I--K--K--F--

361   AAGGACAGTGAAGAAAATGAAGAAGTCAAAGAAACGACTTTACGAGAGCTTAAAATGCTTCGGACTCTCAAGCAGGAAAACATTGTGGAGTTGAAGGAAGCATTTCGTCGGAGGGGAAAG
      K--D--S--E--E--N--E--E--V--K--E--T--T--L--R--E--L--K--M--L--R--T--L--K--Q--E--N--I--V--E--L--K--E--A--F--R--R--R--G--K--
                                                                                                          exon 2
481   TTGTACTTGGTGTTTGAGTATGTTGAAAAAAATATGCTCGAATTGCTGGAAGAAATGCCAAATGGAGTTCCACCTGAGAAAGTAAAAAGCTACATCTATCAGCTAATCAAGGCTATTCAC
      L--Y--L--V--F--E--Y--V--E--K--N--M--L--E--L--L--E--E--M--P--N--G--V--P--P--E--K--V--K--S--Y--I--Y--Q--L--I--K--A--I--H--
                                            exon 3                                                      exon 4
601   TGGTGCCATAAGAATGATATTGTCCATCGAGATATAAAACCAGAAAATCTCTTAATCAGCCACAATGATGTCCTAAAACTGTGTGACTTTGGTTTTGCTCGTAATCTGTCAGAAGGCAAT
      W--C--H--K--N--D--I--V--H--R--D--I--K--P--E--N--L--L--I--S--H--N--D--V--L--K--L--C--D--F--G--F--A--R--N--L--S--E--G--N--
                                                                                   exon 5
721   AATGCTAATTACACAGAGTACGTTGCCACCAGATGGTATCGGTCCCCAGAACTCTTACTTGGCGCTCCCTATGGAAAGTCCGTGGACATGTGGTCGGTGGGCTGTATTCTTGGGGAGCTT
      N--A--N--Y--T--E--Y--V--A--T--R--W--Y--R--S--P--E--L--L--L--G--A--P--Y--G--K--S--V--D--M--W--S--V--G--C--I--L--G--E--L--
                                                                  Txp3-3F
841   AGCGATGGACAGCCTTTATTTCCTGGAGAAAGTGAAATTGACCAACTTTTTACTATTCAGAAGGTGCTAGGACCACTTCCATCTGAGCAGATGAAGCTTTTCTACAGTAATCCTCGCTTC
      S--D--G--Q--P--L--F--P--G--E--S--E--I--D--Q--L--F--T--I--Q--K--V--L--G--P--L--P--S--E--Q--M--K--L--F--Y--S--N--P--R--F--
               exon 6          Txp3-4F
961   CATGGGCTCCGGTTTCCAGCTGTTAACCATCCTCAGTCCTTGGAAAGAAGATACCTTGGAATTTTGAATAGTGTTCTACTTGACCTAATGAAGAATTTACTGAAGTTGGACCCAGCTGAC
      H--G--L--R--F--P--P--A--V--N--H--P--Q--S--L--E--R--R--Y--L--G--I--L--N--S--V--L--L--D--L--M--K--N--L--L--K--L--D--P--A--D--
                            Txp3-2R
1081  AGATACTTGACAGAACAGTCTTTGAATCACCCTACATTTCAAACCCAGAGACTTCTGGATCGTTCTCCTTCAAGGTCAGCAAAAAGAAAACCTTACCATGTGGAAAGCAGCACATTGTCT
      R--Y--L--T--E--Q--S--L--N--H--P--T--F--Q--T--Q--R--L--L--D--R--S--P--S--R--S--A--K--R--K--P--Y--H--V--E--S--S--T--L--S--
              exon 7      Txp3-1R
1201  AATAGAAACCAAGCCGGCAAAAGTACTGCTTTGCAGTCTCACCACAGATCTAACAGCAAGGACATCCAGAACCTGAGTGTAGGCCTGCCCCGGGCTGACGAAGGTCTCCCTGCCAATGAA
      N--N--R--N--Q--A--G--K--S--T--A--L--Q--S--H--H--R--S--N--S--K--D--I--Q--N--L--S--V--G--L--P--R--A--D--E--G--L--P--A--N--E--

1321  AGCTTCCTAAATGGAAACCTTGCTGGAGCTAGTCTTAGTCCACTGCACACCAAAACCTACCAAGCAAGCAGCCAGCCTGGGTCTACCAGCAAAGATCTCACCAACAACAACATACCACAC
      S--F--L--N--G--N--L--A--G--A--S--L--S--P--L--H--T--K--T--Y--Q--A--S--S--Q--P--G--S--T--S--K--D--L--T--N--N--N--I--P--H--

1441  CTTCTTAGCCCAAAAGAAGCCAAGTCAAAAACAGAGTTTGATTTTAATATTGACCCAAAGCCTTCAGAAGGCCCAGGGACAAAGTACCTCAAGTCAAACAGCAGATCTCAGCAGAACCGC
      L--L--S--P--K--E--A--K--S--K--T--E--F--D--F--N--I--D--P--K--P--S--E--G--P--G--T--K--Y--L--K--S--N--S--R--S--Q--Q--N--R--

1561  CACTCATTCATGGAAAGCTCTCAAAGCAAAGCTGGGACACTGCAGCCCAATGAAAAGCAGAGTCGGCATAGCTATATTGACACAATTCCCCAGTCCTCTAGGAGTCCCTCCTACAGGACC
      H--S--F--M--E--S--S--Q--S--K--A--G--T--L--Q--P--N--E--K--Q--S--R--H--S--Y--I--D--T--I--P--Q--S--S--R--S--P--S--Y--R--T--

1681  AAGGCCAAAAGCCATGGGGCACTGAGTGACTCCAAGTCTGTGAGCAACCTTTCTGAAGCCAGGGCCCAAATTGCGGAGCCCAGTACCAGTAGGTACTTCCCATCTAGCTGCTTAGACTTG
      K--A--K--S--H--G--A--L--S--D--S--K--S--V--S--N--L--S--E--A--R--A--Q--I--A--E--P--S--T--S--R--Y--F--P--S--S--C--L--D--L--

1801  AATTCTCCCACCAGCCCAACCCCCACCAGACACAGTGACACGAGAACTGGCTCAGCCCTTCTGGAAGAAATAACCGAAATGAGGGAACGCTGGACTCACGTCGAACCACAACCAGACATT
      N--S--P--T--S--P--T--S--P--T--R--H--S--D--T--R--T--G--S--A--A--L--L--E--E--I--T--E--M--R--E--R--W--T--H--V--E--P--Q--P--D--I--
                                                                                                   Txp3-4R
1921  CTAAGACGATGGAGGAATTGAAGCTGCCGGAGCACATGGACAGTAGCCATTCCCATTCACTGTCTGCACCTCACGAATCTTTTTCTTATGGACTGGGCTACACCAGCCCCTTTTCTTCCC
      L--R--R--R--W--R--N--*--

2041  AGCAACGTCCTCATAGGCATTCTATGTATGTGACCCGTGACAAAGTGAGAGCCAAGGGCTTGGATGGAAGCTTGAGCATAGGGCAAGGGATGGCAGCTAGAGCCAACAGCCTGCAACTCT
                            Txp3-3R
2161  TGTCACCCCAGCCTGGAGAACAGCTCCCTCCAGAGATGACTGTGGCAAGATCTTCGGTCAAAGAGACCTCCAGAGAAGGCACCTCTTCCTTCCATACACGCCAGAAGTCTGAGGGTGGAG

2281  TGTATCATGACCCACACTCTGATGATGGCACAGCCCCCAAAGAAAATAGACACCTATACAATGATCCTGTGCCAAGGAGAGTTGGTAGCTTTTACACACTGCCATCTCCACGTCCAGACA

2401  ATTCTTTCCATGAAAATAATGTGTCAACTAGAGTTTCTTCTCTACCATCAGAGAGCAGTTCTGGAACCAACCACTCAAAAAGACAACCAGCATTCGATCCATGGAAAAGTCCTGAAAATA

2521  TTAGTCATTCAGAGCAACTCAAGGAAAAAAGAGAAGCAAGGATTTTTCAGGTCAATGAAAAAGAAAAAAGAAGAAATCTCAAACAACCTGGA
```

**Primers**
Fragment A is obtained by a nested PCR, with primers *Txp3-1F* and *Txp3-1R* in a first PCR and *Txp3-2F* and *Txp3-2R* in a second PCR. Fragment B is obtained by a nested PCR, with primers *Txp3-3F* and *Txp3-3R* in a first PCR and *Txp3-4F* and *Txp3-4R* in a second PCR. Primer sequences are indicated in Table 1.
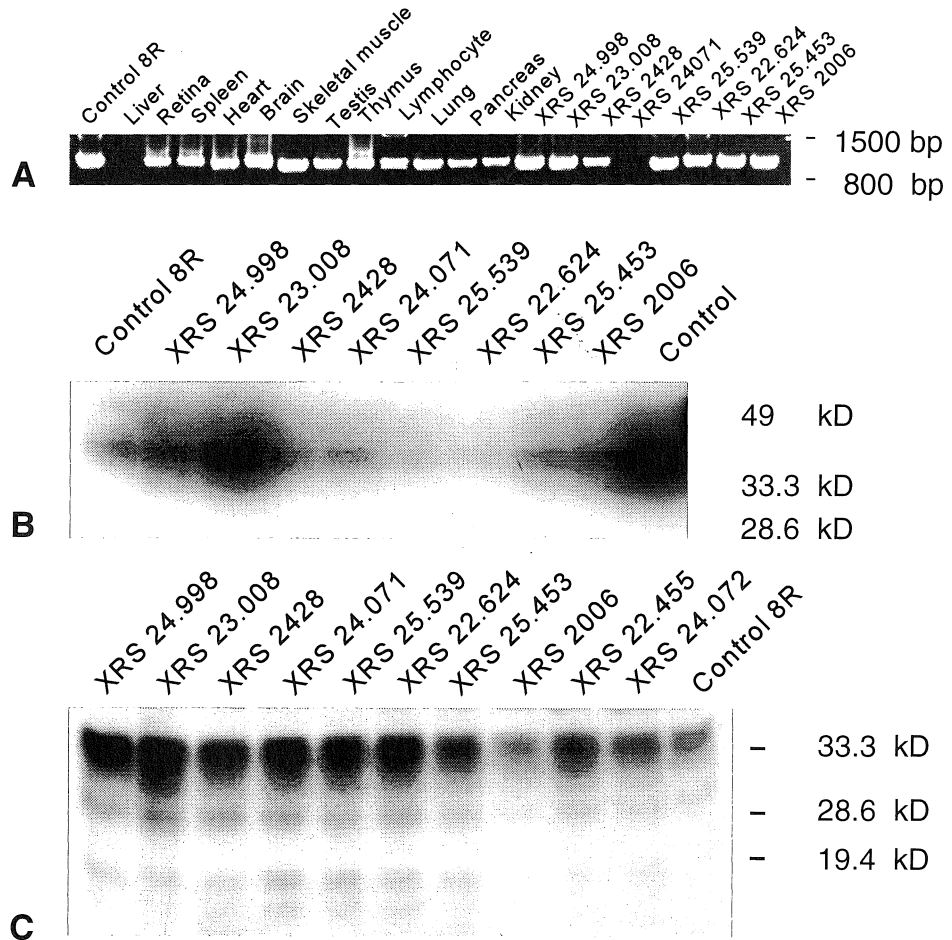


**Figure 2 A**. RT-PCR products of fragment A generated on RNA from 12 different tissues and RNA isolated from lymphocytes of 8 unrelated RS patients and one normal individual (control 8R). The product from XRS 24.071 (not visible on this photograph) was generated in a higher quantity in a duplo experiment. **B**. SDS/PAGE analysis of the translation products of fragment A. **C**. SDS/PAGE analysis of the translation products of fragment B.
The marker used is a prestained SDS-PAGE standard (Biorad).

**Mutation detection**

All procedures, RT, PCR, *in vitro* transcription/translation and SDS-PAGE analysis were performed as described in (Van de Vosse *et al.* 1997). SSCP analysis was performed as described by Orita (Orita *et al.* 1989) using the RT-PCR products digested with *Rsa*I (fragment A and B) or *Hae*III (fragment B).

**Results and Discussion**

Four primersets (set 1 and 3 for initial PCR, set 2 and 4 for a nested PCR) were designed on *Txp3* in order to generate two overlapping products of 1019 bp (fragment A) and 1025 bp (fragment B) respectively. Using these primers, transcripts could be generated on RNA of retina, spleen, heart, brain, skeletal muscle, testis, thymus, lymphocyte, lung, pancreas and kidney. All products had the expected length. No transcripts could be detected in liver.

To scan the *Txp3* gene for mutations, we generated RT-PCR products on RNA isolated from lymphocytes from 9 unrelated RS patients and one normal individual. The expected fragments were generated with both primer sets in all individuals (Figure 2a). Subsequent transcription and translation of the RT-PCR products, followed by analysis of the obtained proteins on an SDS/PAGE gel did not produce aberrant fragments. Consequently, no mutations affecting transcription (splice site or promoter mutations) of the *Txp3* gene nor mutations causing premature translation termination could be identified (Fig. 2B - 2C)

To identify possible non-translation terminating mutations (single base pair changes or triplet deletions) the RT-PCR products were digested with *Rsa*I (fragment A and B) or *Hae*III (fragment B) and analysed with SSCP (data not shown). Again, no aberrant fragments could be detected.

The sequence we have used to design the primers for PTT analysis of *Txp3* contained a difference from the sequence indicated in Figure 1. Base 1849 G was later found to be TT. The change results in a frameshift which opens the reading frame over the original stopcodon (position 1940) to beyond the sequence currently known (2610 bp). Since our 3' nested primer anneals at position 2007 we did not analyse the entire C-terminal region of *Txp3*. Hence we have missed mutations if they were present between base pairs 2007 and 2610.

In parallel experiments (by B. Franco, T. Alitalo and D. Trump, personal communications) in which *Txp3* exons were amplified from genomic DNA of patients and analysed using SSCP. No mutations could be detected.

Based on the RT-PCR, PTT and SSCP analysis it was concluded that *Txp3* is not likely to be the gene causing RS.

**References**

Den Dunnen, J.T., Roest, P.A.M., Van der Luijt, R.B. and Hogervorst, F.B.L. The Protein Truncation Test (PTT) for rapid detection of translation-terminating mutations. In Pfeifer, G.P. (Ed.), Technologies for detection of DNA damage and mutations. Plenum Press, New York, 1996, pp.323-341.

Franco, B. *et al.* In preparation (1997)

Orita, M., Suzuki, Y., Sekiya, T. and Hayashi, K. Rapid and sensitive detection of point mutations and DNA polymorphisms using the polymerase chain reaction. Genomics 5 (1989) 874-879.

Van de Vosse, E., Franco, B., Van der Bent, P., Montini, E., Orth, U., Hanauer, A., Tijmes, N., Van Ommen, G.J.B., Ballabio, A., Den Dunnen, J.T. and Bergen, A.A.B. Exclusion of *PPEF* as the gene causing X-linked juvenile retinoschisis. Hum.Genet. *in press* (1997)

**SUMMARY**

The mammalian X chromosomes have several interesting features, including that they are one of the sex determining chromosomes, contain regions of homology with the Y chromosome (for instance in the pseudoautosomal regions), and display X inactivation. The human X chromosome is also known to contain many disease genes. The region we have focused on (Xp22.1-p22.2) contains amongst others the genes for X-linked juvenile retinoschisis (RS) and keratosis follicularis spinulosa decalvans (KFSD). RS is a slowly progressive retinal degeneration that causes a decrease in acuity and visual field and can result in total blindness. KFSD is a rare disorder that causes a range of symptoms of which hyperkeratosis of specific skin areas and absence of facial hair are its most striking clinical features.

In order to construct a physical map of the Xp22.1-p22.2 region, we screened YAC libraries with markers and clones from this region and assembled the obtained YACs into a contig (overlapping set) based on *Alu* PCR fingerprinting and marker content. This YAC contig spans a region of 4.5 to 5 Mb from the marker DXS451 to DXS414 (Chapter 2.1). Using this YAC contig we ordered several new markers that we subsequently used to refine the candidate regions for RS and KFSD by analysis of recombinants. Based on the length of the YACs in the YAC contig we estimated that the candidate region for KFSD -between the markers DXS7161 and DXS1226- was 1 Mb and the candidate region for RS -between the markers DXS418 and DXS999- was 600 kb. However, when more markers became available between DXS418 and DXS999 it appeared that the key YAC in the region (y939H7, 1.3 Mb in length) had a large deletion between these markers. To obtain the full-length YAC, we analysed many individual colonies from the original culture and found a stable clone that contains all known markers between DXS418 and DXS999 and has a length of 2.5 Mb. With this stable clone we performed YAC fragmentation. The yeast colony was transformed with a plasmid (pBP108/*ADE2*) that contained a yeast telomere, the *ADE2* gene and an *Alu* repeat sequence. Homologous recombination between the *Alu*'s in the YAC and the *Alu* in the plasmid generated fragmented YACs that could be selected for through the *ADE2* gene, now present in the YACs. The experiments resulted in a panel of fragmented YACs ranging in size from 170 kb to 2.4 Mb. This panel facilitated the construction of a large-scale restriction map and allowed binning of clones in the region (Chapter 2.2). The RS candidate region could then be estimated to be 1 Mb in size, almost twice the size of the original estimation.

145

*Summary*

We chose exon trapping as method to isolate candidate genes for the diseases localised to this region and to construct a transcriptional map. We began with the (plasmid-based) pSPL3 exon trap vector and the first transcripts we isolated were part of *PHKA2*, a known gene in the region. To improve the speed and efficiency of exon trapping, we constructed a new, cosmid-based, exon trap vector (sCOGH) to analyse larger stretches of DNA (up to 40 kb), thereby leaving more of the genomic context in tact, and allowing the isolation of multiple exons in one product. We first tested the system by exon trapping cosmids that contained up to seven exons of a known gene, DMD. Analysis of the exon trap products proved that the exons were correctly spliced (Chapter 3).

We subcloned key YACs from the Xp22.1-p22.2 region in the sCOGH exon trap vector and used the resulting cosmids in exon trap experiments. We generated many exon trap products, one of which was characterised extensively and proved to contain part of the *SCML1* gene (Chapter 4.1), a gene that was simultaneously isolated by D. Trump (Cambridge, U.K) using pSPL3 exon trapping. We also contributed to the analysis of *PPEF*, another gene in the RS candidate region that was found using pSPL3 exon trapping (Chapter 4.2). This gene has a homologue in *Drosophila* (*rdgC*) that plays a role in the prevention of light induced retinal degeneration and therefore seemed a strong candidate for RS. However, *PPEF* is not expressed in the mammalian eye, suggesting a different function for the human homologue.

We tested two candidate genes, *PPEF* and *Txp3*, in RS patients to search for mutations causing the disease, but no mutations were found using several techniques (Chapter 5). This makes it unlikely that either of these two genes is involved in RS. The newly identified *SCML1* gene has not been extensively tested for mutations in RS patients. Based on PAC clones provided by the Retinoschisis Consortium the entire RS candidate region has been sequenced, allowing *in silico* identification of putative genes from the region. The identification of the RS gene therefore seems not far off. Although several genes have been isolated in the KFSD candidate region recently, none of these genes have been tested in KFSD patients so far.

## SAMENVATTING VOOR LEKEN

DNA (desoxyribonucleïnezuur) is een molecuul dat bestaat uit vier verschillende bouwstenen (de nucleotiden of basen); guanine (G), adenine (A), cytosine (C) en thymine (T), die in lange ketens aan elkaar gekoppeld zijn. Hoewel in zoogdieren in 98% van het DNA deze Gs, As, Cs en Ts in schijnbaar willekeurige volgorde aan elkaar gekoppeld zijn, bevat het DNA ook zo'n 50.000 tot 100.000 stukken (de genen), waar de volgorde van de nucleotiden heel precies de informatie bevat die vertaald kan worden in eiwitten. Door een exacte regulering van het moment, de plaats in het lichaam, en de hoeveelheid van vertaling van elk gen in een specifiek eiwit, worden alle processen in elk organisme van eicel tot volwassen stadium geregeld. Het DNA is aanwezig in (vrijwel) alle cellen in het lichaam. Bij de voortplanting geeft elk van de ouders de helft van zijn of haar DNA door aan het nageslacht, het DNA is zo de drager van het erfelijk materiaal (het genoom).

Het erfelijk materiaal van de mens is verdeeld over 23 paren chromosomen, hiervan zijn er 22 identieke paren, van groot naar klein genummerd chromosoom 1 tot en met 22, en één paar geslachts chromosomen. Deze geslachts chromosomen bestaan uit 2 X chromosomen (bij de vrouw) of uit een X en een Y chromosoom (bij de man). Het X chromosoom heeft een grootte vergelijkbaar met chromosoom 7 of 8 (160 miljoen basen) en bevat ongeveer 2500 tot 5000 genen. Het Y chromosoom is het kleinste chromosoom (50 miljoen basen) en bevat waarschijnlijk weinig genen. Wanneer bij de man een fout aanwezig is in een gen op het X chromosoom resulteert dit in een X-gebonden aandoening. Afhankelijk van welk gen het is kan dit bijvoorbeeld een spierziekte zijn -zoals bij Duchenne spier dystrofie waar een spiereiwit niet goed wordt aangemaakt-, een combinatie van afwijkingen in de ontwikkeling, -zoals bij het syndroom van Hunter waar een eiwit benodigd voor de stofwisseling defect is-, of een onschuldige aandoening -zoals de veel voorkomende kleurenblindheid-. X-gebonden aandoeningen komen minder vaak bij vrouwen voor doordat ze bijna altijd op hun tweede X chromosoom een gezonde kopie van het aangedane gen meedragen. Vrouwen kunnen het afwijkende gen echter wel doorgeven aan hun dochters en zonen.

Het onderzoek beschreven in dit proefschrift was erop gericht een deel van het X chromosoom (namelijk Xp22.1-p22.2) in kaart te brengen om vervolgens de genen te kunnen vinden die in dit gebied bij verschillende erfelijke aandoeningen betrokken zijn. In figuur 5 is aangegeven van welke aandoeningen bekend is dat ze in dit gebied liggen. Het onderzoek werd

na verloop van tijd gericht op het vinden van twee specifieke genen; het gen dat in patiënten met X-gebonden juveniele retinoschisis (RS) aangedaan is en het gen dat in patiënten met keratosis follicularis spinulosa decalvans (KFSD) aangedaan is. RS komt voor bij 1 op de 10 à 20.000 mensen en is een langzaam voortschrijdende oogziekte waarbij de gezichtsscherpte afneemt en het gezichtsveld nauwer wordt. De ziekte wordt meestal op jonge leeftijd ontdekt, vaak doordat het schoolbord slecht leesbaar blijkt, en kan op latere leeftijd uiteindelijk leiden tot blindheid. De oorzaak van de afname in gezichtsscherpte is het ontstaan van vouwen en scheuren in de gele vlek (de macula, waar normaal het scherpste beeld verkregen wordt), de oorzaak van de vernauwing van het gezichtsveld is het splijten van het netvlies (retina) waarbij in de ergste gevallen het netvlies geheel los kan laten. Er is momenteel geen behandeling beschikbaar om RS te genezen of de symptomen te bestrijden. KFSD is een zeer zeldzame afwijking waarbij verdikking van de huid van de nek, oren, handpalmen en voetzolen, verlies van wenkbrauwen, wimpers en baardgroei, kaalheid in draaiende strepen, ontsteking van de oogleden, degeneratie van het hoornvlies, en overgevoeligheid voor licht voorkomen. De symptomen van KFSD nemen na verloop van tijd af.

Het in kaart brengen van het Xp22.1-p22.2 gebied is gedaan met kunstmatige chromosomen in gist (yeast artificial chromosomes = YACs) die elk een stukje van het menselijk X chromosoom bevatten. Een selectie van overlappende YACs (een contig) werd gemaakt door van bekende stukjes DNA (zogenaamde markers) de aanwezigheid en volgorde in de YACs te bepalen. Met deze markers werd tevens in DNA van RS en KFSD patiënten en hun familieleden door middel van overervingsonderzoek bepaald waar deze genen ongeveer liggen. Zo konden we bepalen dat het RS gen gelegen is tussen de markers DXS418 en DXS999 en het KFSD gen tussen de markers DXS7161 en DXS1226 (zie hoofdstuk 2.1). Een van deze YACs (y939H7), die het gebied tussen DXS418 en DXS999 geheel bevat, hebben we gebruikt in experimenten die de YAC stukken kleiner maakt (YAC fragmentatie) zodat deze eenvoudiger te analyseren werd. Met behulp van de gefragmenteerde YACs hebben we een zeer gedetailleerde kaart van het gebied kunnen maken en de grootte van het gebied kunnen bepalen op ongeveer 1 miljoen basen (zie hoofdstuk 2.2).

Om (stukken van) genen te kunnen isoleren uit deze kandidaat gebieden hebben we de 'exon trap' techniek gekozen. Na het aflezen van een gen, waarbij een kopie (het mRNA) van het gen ontstaat die in plaats van thymine uracil (U) bevat, worden niet coderende stukken (intronen) tussen de coderende stukken (exonen) uitgesplitst. Exon trappen is erop gebaseerd dit uitsplitsen

148

te forceren in een experimentele opzet zodat de exonen gevonden kunnen worden. Naast een bestaand exon trap systeem (pSPL3) hebben we een nieuw exon trap systeem (sCOGH) gebruikt. De door ons ontwikkelde exon trap vector, sCOGH, hebben we eerst uitgetest op een bekend gen; DMD, het gen dat aangedaan is bij Duchenne spier dystrofie (zie hoofdstuk 3). Slechts één van de genen die met behulp van het sCOGH systeem gevonden werd, *SCML1*, is tot nu toe verder geanalyseerd (zie hoofdstuk 4.1). Van de genen gevonden met het pSPL3 systeem was het eerste een reeds bekend gen; *PHKA2*. We hebben wegens het succes van het sCOGH systeem vervolgens -behalve aan de gezamenlijke analyse van *PPEF* (zie hoofdstuk 4.2)- weinig moeite gestoken in de analyse van de pSPL3 producten.

Twee genen uit het RS kandidaat gebied (*PPEF* en *Txp3*) zijn met verschillende technieken getest in RS patiënten, in geen van beide werden echter afwijkingen gevonden (zie hoofdstuk 5). Het is daardoor onwaarschijnlijk dat een van deze twee genen een rol speelt in het onstaan van RS, of ze een rol spelen bij een van de andere ziekten die in dit gebied liggen is nog niet getest. Er zijn nog geen genen getest in KFSD patiënten, hoewel er al wel genen in het kandidaat gebied beschikbaar zijn. Het isoleren en testen van meer nieuwe genen in de RS en KFSD kandidaat gebieden zal uiteindelijk leiden tot de identificatie van de genen die in RS en KFSD patiënten aangedaan zijn. Het vinden van deze genen zal inzicht geven in de functie ervan en zo mogelijk leiden tot een vorm van therapie.

## ADDENDUM

In the discussion it was predicted that the gene involved in X-linked juvenile retinoschisis would be identified soon. The cloning of the RS gene by Sauer *et al.*\* was indeed published in the October issue of Nature Genetics, just before this thesis was printed.

The gene, designated *XLRS1* (for *X-linked retinoschisis-1*) was found using a combination of approaches, including exon trapping, cDNA selection and analysis of the sequences from the PACs spanning the RS region, generated by the Sanger Centre in collaboration with the Retinoschisis Consortium. The *XLRS1* gene consists of 6 exons, contains 2 possible poly-adenylation sites (at positions 847 and 4020 respectively) and detects transcripts on Northern blots of 1.1 and 3.1 kb, suggesting alternative poly-adenylation. The transcript is highly abundant in retina and not expressed in other tissues. *XLRS1* has an 687 bp ORF and encodes a 244 amino acid precursor protein with a 23 amino acid residue leader sequence. The predicted mature protein has a length of 201 amino acids, with a mass of 23 kD.

The mature protein shows a region of high homology to discoidin. This domain is evolutionary conserved and shared by several other proteins implicated in phospholipid binding and cell-cell interactions on membrane surfaces. *XLRS1* could be necessary for establishing cell-specific synaptic contacts. A defective protein may thus interfere with the formation of interneuronal connections during embryonal development thereby causing RS. Further analysis is required to elucidate the exact function of *XLRS1* and its role in the developing retina.

The expression profile, being retina specific, led to the testing for mutations in RS patients. One nonsense, one frameshift, one splice acceptor and six missense mutations were identified in nine unrelated RS families. These mutations prove that the gene, when mutated, indeed causes RS.

Based on the mutation spectrum observed so far, its potential function and the lack of symptoms in carriers, it is likely that loss of function (rather than a dominant negative effect) is the cause of the pathology. This gives room for speculation on possible methods for therapy of RS patients. As RS is a slowly progressive disease, gene therapy at a very early age may prevent total blindness later in life.

\* C.G. Sauer, A. Gehrig, R. Warneke-Wittstock, A. Marquardt, C.C. Ewing, A. Gibson, B. Lorenz, B. Jurklies, and B.H.F. Weber. (1997) Positional cloning of the gene associated with X-linked juvenile retinoschisis. *Nature Genet.* 17:164-170.

## LIST OF ABBREVIATIONS

| | |
|---|---|
| bp | base pair |
| CEPH | Centre d'Étude du Polymorphisme Humain |
| CLS | Coffin-Lowry syndrome |
| cM | centiMorgan |
| DGGE | denaturing gradient gel electrophoresis |
| DNA | deoxyribonucleic acid |
| FISH | fluorescence *in situ* hybridisation |
| HSH | hypomagnesemia with hypocalcemia |
| HTF | *Hpa*II tiny fragments |
| kb | kilo base |
| kDa | kilo Dalton |
| KFSD | keratosis follicularis spinulosa decalvans |
| Mb | Mega base |
| NHS | Nance Horan syndrome |
| OMIM | on line Mendelian inheritance in man |
| PAB | pseudoautosomal boundary |
| PAR | pseudoautosomal region |
| PCR | polymerase chain reaction |
| PFGE | pulsed-field gel electrophoresis |
| PPEF | protein phosphatase with EF calcium-binding domain |
| PTT | protein truncation test |
| RFLP | restriction fragment length polymorphism |
| RNA | ribonucleic acid |
| RP15 | X-linked cone rod degeneration (retinitis pigmentosa 15) |
| RS | X-linked juvenile retinoschisis |
| RT | reverse transcription |
| SCML1 | sex comb on midleg-like 1 |
| SEDL | spondyloepiphyseal dysplasia of the late type |
| SSCA | single-strand conformation analysis |
| XIC | X inactivation center |
| YAC | yeast artificial chromosome |

**PUBLICATIONS**

Skraastad, M.I., Van de Vosse, E., Belfroid, R., Höld, K., Vegter-Van der Vlis, M., Sandkuijl, L.A., Bakker, E., and Van Ommen, G.J.B. Significant linkage disequilibrium between the Huntington disease gene and the loci D4S10 and D4S95 in the Dutch population. *Am.J.Hum.Genet.* 51:730-735 (1992).

Van de Vosse, E., Booms, P.F.M., Vossen, R.H.A.M., Wapenaar, M.C., Van Ommen, G.J.B., and Den Dunnen, J.T. A CA-repeat polymorphism near DXS418 (P122). *Hum.Mol.Genet.* 2:2202 (1993).

Oosterwijk, J.C., Van der Wielen, M.J.R., Van de Vosse, E., Voorhoeve, E., and Bakker, E. Refinement of the localisation of X-linked keratosis follicularis spinulosa decalvans (KFSD) gene in Xp22.1-p22.2. *J.Med.Genet.* 32:736-739 (1995).

Roepman, R., Bauer, D., Rosenberg, T., Van Duijnhoven, G., Van de Vosse, E., Platzer, M., Rosenthal, A., Ropers, H.-H., Cremers, F.P.M., and Berger, W. Identification of a gene disrupted by a microdeletion in a patient with X-linked retinitis pigmentosa (XLRP). *Hum.Mol.Genet.* 5:821-826 (1996).

Datson, N.A., Van de Vosse, E., Dauwerse, H.G., Bout, M., Van Ommen, G.J.B., and Den Dunnen, J.T. Scanning for genes in large genomic regions: cosmid-based exon trapping of multiple exons in a single product. *Nucleic Acids Res.* 24:1105-1111 (1996).

Van de Vosse, E., Bergen, A.A.B., Meershoek, E.J., Oosterwijk, J.C., Gregory, S., Bakker, B., Weissenbach, J., Coffey, A.J., Van Ommen, G.J.B., and Den Dunnen, J.T. A Xp22.1-p22.2 YAC contig encompassing the disease loci for RS, KFSD, CLS, HYP and RP15; refined localization of RS. *Eur.J.Hum.Genet.* 4:101-104 (1996).

Van de Vosse, E., Van der Bent, P., Heus, J.J., Van Ommen, G.J.B., and Den Dunnen, J.T. High resolution mapping by YAC fragmentation of a 2.5 Mb Xp22 region containing the human RS, KFSD, CLS and NHS disease genes. *Mamm.Genome* 8:497-501 (1997).

152

Heus, J.J., De Winther, M.P.J., Van de Vosse, E., Van Ommen, G.J.B., and Den Dunnen, J.T. Centromeric and non-centromeric *ADE2*-selectable fragmentation vectors for YACs in AB1380. *Genome Res.* 7:657-660 (1997).

Montini, E., Rugarli, E.I., Van de Vosse, E., Andolfi, Mariani, M., G., Puca, A.A., Consalez, G.G., Den Dunnen, J.T., Ballabio, A., and Franco, B. A novel human serine-threonine phosphatase related to the *Drosophila retinal degeneration C* (*rdg*C) gene is selectively expressed in sensory neurons of neural crest origin. *Hum.Mol.Genet.* 6:1137-1145 (1997).

Oosterwijk, J.C., Richard, G., Van der Wielen, M.J.R., Van de Vosse, E., Harth, W., Sandkuyl, L.A., Bakker, E. and Van Ommen, G.J.B. Molecular genetic analysis of two families with keratosis follicularis spinulosa decalvans (KFSD): refinement of gene localization and evidence for genetic heterogeneity. *Hum. Genet.* 100: 520-524 (1997).

Van de Vosse, E., Franco, B., Van der Bent, P., Montini, E., Orth, U., Hanauer, A., Tijmes, N., Van Ommen, G.J.B., Ballabio, A., Den Dunnen, J.T., and Bergen, A.A.B. Exclusion of *PPEF* as the gene causing X-linked juvenile retinoschisis. *Hum.Genet.* (in press) (1997).

Van de Vosse, E., Walpole, S.M., Nicolaou,A., Van der Bent, P., Cahn, A., Vaudin, M., Ross, M.T., Durham, J., Pavitt, R., Wilkinson, J., Grafham, D., Bergen, A.A.B., Van Ommen, G.J.B., Yates, J.R.W., Den Dunnen, J.T., and Trump, D. Characterization of a new developmental gene, *SCML1*, in Xp22. *Submitted* (1997).

Kraayenbrink, T., Van de Vosse, E., Tijmes, N.T., Van Schooneveld, M.J., Ten Brink, J.B., Schuurman, E.J.M., Van Ommen, G.-J.B., Den Dunnen, J.T., and Bergen, A.A.B. Haplotype sharing among non-familial Dutch RS patients, and linkage analysis reduces the obligate RS gene region to a genomic interval less than 250 kb. *Submitted* (1997)

**CURRICULUM VITAE**

The author of this Ph.D. thesis was born in The Hague on October 6, 1965. In June 1985, she graduated from high school, and in August of the same year she started her Biomedical Sciences study at the Medical Faculty of the University of Leiden. During this study, she performed undergraduate research projects in the Embryology and Teratology section of the Department of Anatomy (Dr. Christel Vermeij-Keers), in the department Experimental Hematology (Dr. Wim Fibbe) and in the Department of Human Genetics (Dr. Mette Skraastad). She graduated in August 1991 with a project characterising clones from the Huntington disease gene region. From October 1991 to October 1997 she was a Ph.D. student in the Department of Human Genetics of Leiden University, where, under supervision of Professor Gert-Jan van Ommen and Dr. Johan den Dunnen, she performed the research described in this thesis. In February 1998 she will start working at the Howard Hughes Medical Institute of the University of Colorado in Boulder where she will be working with Professor Robert Boswell on the function of genes in the embryogenesis of *Drosophila melanogaster*.

De auteur van dit proefschrift werd geboren op 6 oktober 1965 te Den Haag. In juni 1985 behaalde zij haar VWO-diploma aan de Louise de Coligny Scholengemeenschap te Leiden. In augustus 1985 begon zij aan haar studie Biomedische Wetenschappen aan de Rijksuniversiteit te Leiden. Tijdens haar studie werden onderzoeksstages verricht binnen de afdeling Embryologie en Teratologie van de vakgroep Anatomie (Dr. Christel Vermeij-Keers), binnen de vakgroep Experimentele Hematologie (Dr. Wim Fibbe) en binnen de vakgroep Anthropogenetica (Dr. Mette Skraastad). Het doctoraal Biomedische Wetenschappen werd door haar gehaald in augustus 1991. Van october 1991 tot october 1997 was zij achtereenvolgens aangesteld als onderzoeker-in-opleiding en wetenschappelijk medewerker binnen de afdeling Anthropogenetica van de Rijksuniversiteit te Leiden, waar zij onder leiding van Professor Gert-Jan van Ommen en Dr. Johan den Dunnen het in dit proefschrift beschreven promotie-onderzoek verrichtte.
Vanaf februari 1998 zal zij aangesteld zijn aan het Howard Hughes Medical Institute van de University of Colorado in Boulder waar zij in de groep van Professor Robert Boswell onderzoek zal doen naar de functie van genen tijdens de embryonale ontwikkeling van de fruitvlieg.