THE UNIVERSITY OF CALGARY

Efforts to Isolate a Gene for X-linked Congenital Stationary Night Blindness by Positional Cloning

by

Kym Marie Boycott

A DISSERTATION

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE

DEGREE OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF MEDICAL SCIENCE

CALGARY, ALBERTA

JANUARY, 1997

© Kym Marie Boycott 1997

National Library of Canada

Acquisitions and Bibliographic Services

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque nationale du Canada

Acquisitions et services bibliographiques

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence

Our file Notre référence

The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

L'auteur conserve la propriété du droit d'auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

0-612-35039-8

Canada

ABSTRACT

In an effort to isolate the gene responsible for X-linked congenital stationary night blindness (CSNB1), a positional cloning strategy has been used to reduce the minimal genetic region for CSNB1, construct a physical contig encompassing the CSNB1 locus, identify new candidate genes, and analyze one of these genes for mutations in affected individuals.

By mapping new polymorphic markers to Xp11.23-p11.22 using a panel of hybrids, the minimal genetic region containing the CSNB1 locus in Family P060 was refined to the region bounded telomerically by DXS722 and centromerically by DXS8023, an estimated 2.2 Mb. Based on a number of published crossover events that place the proximal limit to CSNB1 at DXS255, efforts to isolate candidate genes for CSNB1 have focused on the 1.5 Mb region bounded distally by DXS722 and proximally by DXS255.

To construct a physical contig between DXS722 and DXS255, YAC and cosmid clones were identified. A 2 Mb contig was constructed from the content of 44 DNA markers and spans the distal marker ZNF21 and extends proximally to include DXS255.

Overlapping genomic clones were chosen from the physical contig and used to enrich for cDNA transcripts expressed in the frontal cortex, fetal brain, retina, and placenta by direct cDNA selection. Using this technique, 44 unique expressed sequence tags (ESTs) were added to the transcription map of Xp11.23-p11.22. The EST JRL4-A1 was chosen to be investigated in detail and a unique gene, KAT1, was

identified that contains a 262 amino acid open reading frame and seven exons.

The full-length cDNA clone for KAT1 was used as a probe to scan for deletions by Southern analysis in 15 CSNB1 families and no aberrant bands were detected. Each exon of KAT1 was then directly sequenced in representative affected and unaffected control individuals from Family P060. From the results of this analysis, KAT1 was excluded as the gene responsible for CSNB1 in Family P060.

ACKNOWLEDGMENTS

I am indebted to my supervisor, Dr. Torben Bech-Hansen for his patience, support, understanding, and friendship. I would like to thank the various past and present members of my supervisory committee, Drs. Dylan Edwards, Tony Garber, Carey Johnson, Jim McGhee, and Mike Walter, and my external examiners, Drs. Derrick Rancourt and Lap-Chee Tsui, for their participation and contribution toward this effort.

The members of the Molecular Diagnostic Laboratory at the Alberta Children's Hospital and the past and present members of Dr. Bech-Hansen's laboratory were always helpful, supportive, and a pleasure with which to work. I would especially like to thank Kathy Gratton for being an excellent teacher and for her high expectations which I strived to meet, and Bob Winkfein for his helpful discussions and advice.

.A special thanks to my husband Randy, to my parents George and Susie, and to my sisters Nikki and Binny for their support and encouragement throughout not only this work but my entire life.

This work was supported by the Cook-McCann studentship from the RP Research Foundation of Canada (K.M.B.) and operating grants from the RP Research Foundation, I.D. Bebensee Foundation, and the Canadian Genome and Analysis and Technology Program (N.T.B.H.).

DEDICATION

To my husband Randy

and

In loving memory of Brigadier Gerard Hilton Boycott

1909-1994

TABLE OF CONTENTS

Approval Page........ ii
Abstract........ iii
Acknowledgments........ v
Dedication........ vi
Table of Contents........ vii
List of Tables........ xi
List of Figures........ xii
List of Abbreviations........ xiv

CHAPTER 1 - INTRODUCTION........ 1

CHAPTER 2 - LITERATURE REVIEW........ 6
- Clinical Characteristics of CSNB1........ 6
- Mapping the CSNB1 Locus........ 8
- Genetic Markers........ 13
- Somatic Cell Hybrid Mapping........ 14
- Physical Contig Construction........ 17
- Genes and ESTs in Xp11.23........ 21
- Identification of Novel Genes........ 22
  - Exon Amplification........ 24
  - Direct cDNA Selection........ 26
- Mutation Analysis........ 33
- The Positional Candidate Approach........ 39

CHAPTER 3 - MATERIALS AND METHODS........ 41
- General Materials and Methods........ 41
  - Hybrid Cell Lines........ 41
  - PCR-conditions for Mapping........ 42
  - Analysis of PCR Products by Gel Electrophoresis........ 43
  - Restriction Endonuclease Digestion........ 44
  - Agarose Gel Electrophoresis of Digestion Products........ 44
  - DNA Transfer........ 45
  - Radiolabelling of DNA and Hybridization. 45
  - DNA Quantitation........ 47
  - Rapid Small-Scale Plasmid Preparation... 47

Isolation of DNA from Low Melting Point Agarose Gels.. 48
DNA Sequencing using Sequenase®........... 49
Acrylamide Gel Electrophoresis.................. 50
Sublocalization of New Polymorphic Markers... 51
Analysis of Polymorphic Markers in CSNB1 Families.. 51
Families... 51
PCR Analysis... 53
Isolation of Large Insert DNA Clones................... 55
DNA Markers... 55
Identification of YACs...................................... 55
Initial Characterization of YACs.................. 57
Identification of Cosmids................................ 60
Initial Characterization of Cosmids........... 61
Generation of New STSs................................. 62
End-clone STSs....................................... 62
Design of ESTs from Existing Sequence.. 67
Identification of Overlaps............................. 67
Isolation of Transcripts by Direct cDNA Selection 68
Isolation of YAC Templates.......................... 68
Isolation of Cosmid Templates.................... 69
cDNA Sets.. 69
Direct cDNA Selection...................................... 71
Blocking of cDNAs.................................. 71
Nick Translation of YAC and Cosmid Templates................................. 71
Primary Selection................................... 73
Secondary Selection............................... 76
Cloning of Direct cDNA Selection Products 76
Analysis of Selected cDNA Clones.............. 79
Sequencing of Inserts.......................... 80
Development of New Region-Specific ESTs 80
Mapping of ESTs.................................. 81
Characterization of KAT1.. 81
Sequencing.. 81
Northern Analysis.. 81
PCR Expression Profiles................................. 82
Exon/Intron Boundaries................................. 83
cDNA Based Primers............................ 83
Cosmid Sequencing.............................. 84

Walking by Linker Ligation............. 85
5'RACE.. 85
Mutation Analysis of KAT1....................................... 88
Deletion Analysis... 88
Sequencing of Exons in Affected Individuals.. 89

CHAPTER 4 - RESULTS.. 92
Refining X Chromosome Crossover Events in CSNB1 Families.. 92
Sublocalization of Initial Sets of New Polymorphic Markers.................................... 93
Analysis of DXS988, DXS1000, and DXS1003 in CSNB1 Families.. 99
Generation of New Polymorphic Markers from Published Sequence and Sublocalization of Additional Polymorphic Markers.......... 101
Additional Polymorphic Marker Analysis in CSNB1 Families.. 103
Distal Limit.. 103
Proximal Limit...................................... 104
Development of a Physical Contig......................... 109
Construction of the Physical Contig........... 109
Integrity of Clones.. 113
Order of DNA Markers.................................... 133
Weak Links in the Contig............................. 134
Size of Contig, DNA Marker Density, and Coverage.. 135
Orientation of Contig on the X Chromosome 135
Physical Mapping of Genetic Markers...... 136
Physical Mapping of Genes and Expressed Sequences.. 137
Isolation of New Genes by Direct cDNA Selection 138
cDNA Sets.. 138
Selection of YAC and Cosmid Templates.. 138
Direct cDNA Selection..................................... 144
Characterization of Selection Products and Development of New ESTs.............................. 151
Proximal Placental cDNA Selection. 156
Distal Placental cDNA Selection........ 157
JRL-Group 1 cDNA Selection............. 159

JRL-Group 2 cDNA Selection............. 160
JRL-Group 3 cDNA Selection............. 161
JRL-Group 4 cDNA Selection............. 164
Verification of Primer Pairs as Region-Specific...................................... 167
Evaluation of Direct cDNA Selection Technique and Generation of an Integrated Map...... 175
Characterization of a Candidate Gene for CSNB1 186
Selection of a Candidate Gene....................... 186
Sequencing of a KAT1 cDNA Clone............ 187
Expression of KAT1.. 187
Northern Analysis................................ 187
PCR Expression Profiles...................... 188
cDNA Sequencing Project Information 193
5' RACE.. 194
Characterization of Exon/Intron Boundaries of the KAT1 Gene.. 199
cDNA Based Primers............................ 199
Cosmid Sequencing............................... 204
Walking by Linker Ligation............. 204
The KAT1 Gene.. 210
Mutation Analysis of KAT1 in CSNB1 Affected Individuals.. 217
Deletion Analysis using KAT1....................... 217
Sequence Analysis of KAT1 in Family P060.. 220

CHAPTER 5 - DISCUSSION
Refining the Genetic Limits for a CSNB1 Locus... 228
Physical Mapping.. 231
Isolation of New Genes by Direct cDNA Selection 237
Characterization of a Candidate Gene for CSNB1 248
Mutation Analysis.. 249
Future Studies.. 250

CHAPTER 6 - SUMMARY.. 253

REFERENCES.. 257

LIST OF TABLES

Table		Page
1.	Published Polymorphic CA Repeat Markers	52
2.	Cycling Conditions for Polymorphic CA Repeat Markers	54
3.	DNA Markers Used in Preliminary Construction of Contig in Xp11.23-p11.22	56
4.	Primer Sequences used End-clone Generation in pYAC4 Clones	65
5.	New Polymorphic Marker Designed from Published Sequence	94
6.	Chromosomes with Crossovers Flanking the CSNB1 Locus	100
7.	New Polymorphic Markers Designed from Published Sequence	101
8.	ESTs Designed from Published Sequences	110
9.	Eighteen New STSs Generated from Selected YAC Clones	129
10.	YAC Clones Used in this Study	130
11.	Clones Generated from cDNA Selection	150
12.	Newly Developed ESTs from Placental Selection of Proximal YACs	157
13.	Newly Developed ESTs from Placental Selection of Distal YACs	158
14.	Newly Developed ESTs from JRL-Group 1 Selection	160
15.	Newly Developed ESTs from JRL-Group 3 Selection	164
16.	Newly Developed ESTs from JRL-Group 4 Selection	168
17.	ESTs Designed from Published Sequences	185
18.	Sequencing Primers for cDNA Clone 263773	187
19.	Expression Analysis of KAT1	193
20.	Primer Sets Designed Using 263773 cDNA Sequence	200
21.	KAT1 Primer for Walking by Linker Ligation	205
22.	Exon Amplification of KAT1	221

LIST OF FIGURES

Figure		Page
1.	Positional Cloning Approach	3
2.	Gene Map of Xp11.3-p11.23	10
3.	End-cloning by Ligation-Mediated PCR	63
4.	Direct cDNA Selection	74
5.	Individuals Chosen for Mutation Analysis	90
6.	Mapping of DXS988 on Hybrid Mapping Panel	95
7.	Hybrid Mapping Panel	97
8.	Analysis of DXS8023 and DXS722 in Recombinant Individuals from Family P060	105
9.	Crossovers Defining the Location of CSNB1 in Family P060	107
10.	Probing of the Lawrence Livermore Cosmid Library with TFE3	111
11.	Determination of STS Content of YAC Clones using PCR	114
12.	Determination of DNA Marker Content of YAC Clones by Southern Hybridization	116
13.	Pulse Field Gel and Corresponding Autoradiogram of the Stable YAC yWXD5905	118
14.	Pulse Field Gel and Corresponding Autoradiogram of the Unstable YACs yWXD1909 and yWXD6799	120
15.	Products of Ligation-Mediated PCR using the YAC yWXD6129	123
16.	Sequence of EC6129R	126
17.	Region-specific Testing of EC6129R	128
18.	Physical Contig in Xp11.23-p11.22	130
19.	YACs Chosen for Placental Direct Selection	139
20.	YACs Chosen for JRL Direct Selection	142
21.	Purification of 707g5 for Direct Selection	145
22.	cDNA Enrichment for the Distal Placental Selection	148
23.	cDNA Inserts from the JRL-Group 3 cDNA Selection	152
24.	Mapping Panel for JRL-Group 4 cDNA Clones	154
25.	Mapping of JRL2-D3 cDNA Clone	162
26.	Mapping of JRL4-A1 cDNA Clone	165
27.	Partial Sequence of JRL4-A1 cDNA Clone	169
28.	Preliminary Mapping of JRL4-A1 EST	171

29. Mapping of JRL4-A1 on Hybrid Mapping Panel.......... 173
30. Sublocalization of JRL4-A1 on Physical Contig by PCR.......... 176
31. Integrated Genetic, Physical, and Transcript Map-A.. 179
32. Integrated Genetic, Physical, and Transcript Map-B.. 181
33. Integrated Genetic, Physical, and Transcript Map-C.. 183
34. Northern Blot of cDNA Clone 263773.......... 189
35. PCR Expression Profile of JRL4-A1.......... 191
36 5' RACE of KAT1.......... 195
37. Sequence of 5' RACE Products.......... 197
38. Intron/Exon Boundary of F-4/R-2 Product.......... 202
39. Ligation-Mediated PCR Product for Walking.......... 206
40. Sequence of Walking Product.......... 208
41. Nucleotide and Translation Sequence of KAT1.......... 211
42. Intron Sequence of KAT1.......... 213
43. Genomic Structure of KAT1.......... 215
44. Deletion Analysis Using KAT1.......... 218
45. Amplification of Exon 3 for Mutation Analysis.......... 222
46. Mutation Analysis of Exon 3.......... 224
47. Polymorphism in Intron C.......... 226

ABBREVIATIONS

A	adenosine residue
AIED	Åland Island Eye disease
BAC(s)	bacterial artificial chromosome(s)
bp	base pair(s)
BSA	bovine serum albumin
C	cytidine residue
°C	degrees Celsius
CHO	Chinese hamster ovary
Ci	Curie(s)
cm	centimetre
cM	centimorgan
COD1	X-linked cone dystrophy
cpm	counts per minute
CSNB1	X-linked congenital stationary night blindness
dATP	deoxyadenosine triphosphate
dCTP	deoxycytidine triphosphate
ddATP	dideoxyadenosine triphosphate
ddCTP	dideoxycytidine triphosphate
ddGTP	dideoxyguanosine triphosphate
ddH_2O	double distilled water
DTT	dithiothreitol
ddTTP	dideoxythymidine triphosphate
dGTP	deoxyguanosine triphosphate
DNA	deoxyribonucleic acid

dsDNA	double stranded deoxyribonucleic acid
dTTP	deoxythymidine triphosphate
EDTA	ethylenediaminetetraacetic acid
ERG	electroretinography
EtBr	ethidium bromide
FISH	fluorescent *in situ* hybridization
G	guanosine residue
g	gram(s)
HCl	hydrochloric acid
HEPES	*N*-2-hydroxyethylpiperazine-*N'*-2-ethanesulfonic acid
HPRT	hypoxanthine phosphoribosyl transferase
ICRF	Imperial Cancer Research Fund (London)
IPTG	isopropylthio-β-D-galactoside
kb	kilobase pair(s)
KCl	potassium chloride
lod	logarithm of the odds
M	molar (concentration in mole(s)/litre)
MAOA	monoamine oxidase A
MAOB	monoamine oxidase B
min	minute(s)
μg	microgram(s)
mg	milligram(s)
$MgCl_2$	magnesium chloride
$MgSO_4$	magnesium sulfate
μL	microlitre(s)

mL	millilitre(s)
µM	micromolar
mM	millimolar
mmol	millimole(s)
NaCl	sodium chloride
NaOH	sodium hydroxide
NDP	Norrie disease
ng	nanogram
OATL1	ornithine aminotransferase-like-1
OD	optical density
OLB	oligonucleotide labelling buffer
OTC	ornithine carbamoyltransferase
PCR	polymerase chain reaction
PIC	polymorphism information content
pmol	picomole
RFLP(s)	restriction fragment length polymorphism(s)
RNA	ribonucleic acid
RNAse	ribonuclease
RP2	retinitis pigmentosa 2
RP3	retinitis pigmentosa 3
s	seconds
SDS	sodium dodecyl sulfate
STR(s)	simple tandem repeat(s)
STS(s)	sequence tagged site(s)
SYP	synaptophysin
T	thymidine residue

TFE3	transcription factor for the enhancer μE3
T_m	melting temperature
TEMED	*N,N,N',N'*-tetramethylethylenediamine
TIMP1	tissue inhibitor of metalloproteinase 1
TK	thymidine kinase
UV	ultraviolet
V	volts
VNDR	variable number of dinucleotide repeats
VNTR	variable number of tandem repeats
v/v	volume/volume
W	watts
w/v	weight/volume
WAS	Wiscott-Aldrich syndrome
WASP	Wiscott-Aldrich syndrome protein
X-gal	5-bromo-4-chloro-3-indolyl-β-D-galactoside
YAC(s)	yeast artificial chromosome(s)
ZNF	zinc finger

CHAPTER 1 - INTRODUCTION

X-linked congenital stationary night blindness (CSNB1) is an eye disorder that includes impairment of night vision, reduced visual acuity and, in some cases, myopia and congenital nystagmus. Electroretinography reveals a marked reduction of the b-wave in affected individuals suggesting that X-linked CSNB is due to a molecular defect in the bipolar layer of the retina. In the absence of specific knowledge about the basic biochemical defect causing the clinical manifestations of CSNB1, efforts to isolate this gene must be based on a positional cloning strategy.

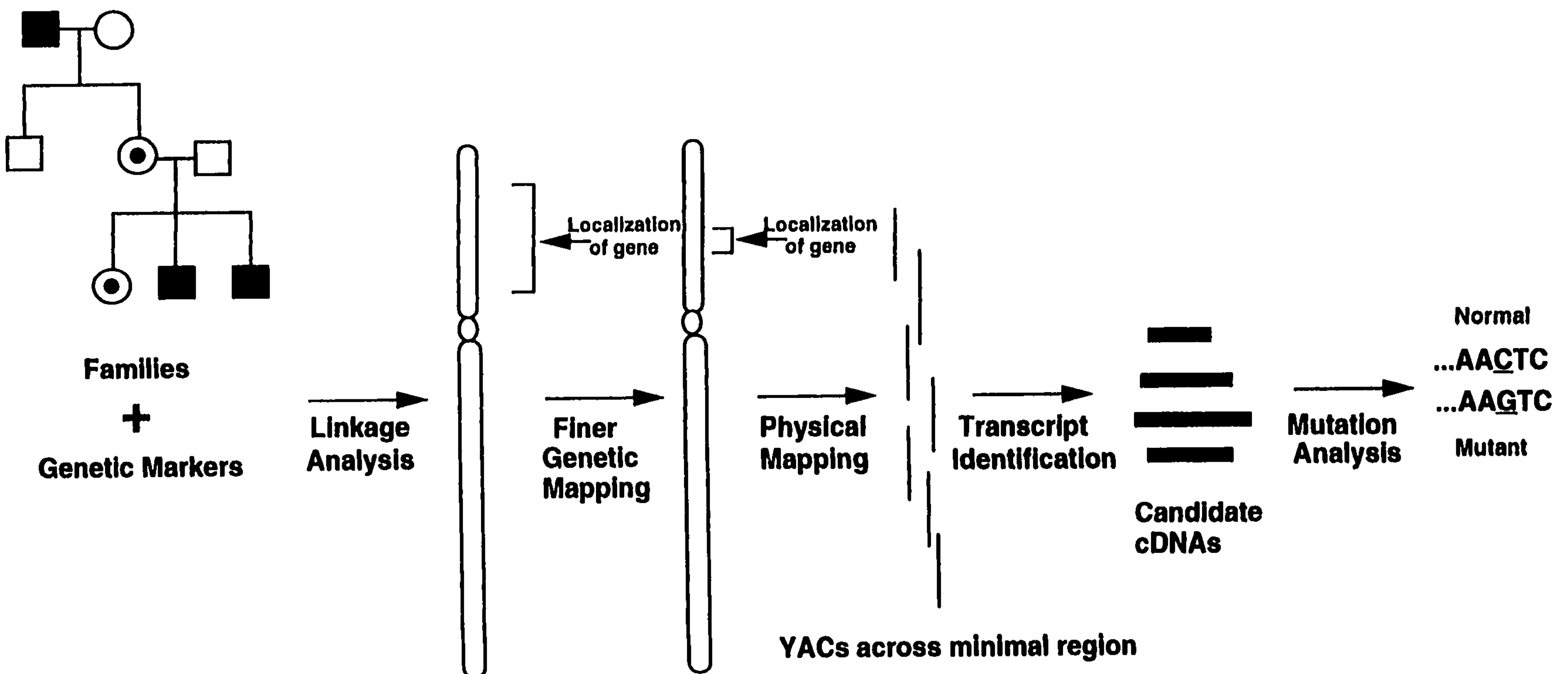
Positional cloning (see Figure 1; Collins, 1992; Collins, 1995) assumes no functional information and isolates the gene responsible for the disease manifestations solely on the basis of map position. The first step in positional cloning is usually the collection of pedigrees in which a specific disease gene is segregating. These families are studied with multiple polymorphic markers until evidence for linkage is obtained. Identification of patients who have visible cytogenetic rearrangements can greatly assist the low and high resolution mapping of the disease gene. The majority of positional cloning successes have relied on such rearrangements and many of the rest have been aided by identification of patients carrying deletions of tens or hundreds of kilobases (Collins, 1995). The recent identification of the RPGR (retinitis pigmentosa GTPase regulator) gene implicated in a form of X-linked retinitis pigmentosa, RP3, was based on the identification of a number of patients carrying

deletions of tens to hundreds of kilobases (Meindl *et al.*, 1996; Roepman *et al.*, 1996). Until 1993, only the cystic fibrosis gene (Rommens *et al.*, 1989) had been identified in the absence of rearrangements or deletions but now this number has grown to about 20% of the some 50 genes identified by positional cloning (Collins, 1995).

Once linkage has been established to markers on a specific chromosome, additional fine mapping can then be applied to narrow the location of the disease causing gene, though the resolution is limited by the number of informative meioses available within the pedigrees. In most cases it is unusual to have more than 100 meiotic events available, so fine mapping is often limited to genetic intervals of about 1 cM or approximately a million base pairs (Collins, 1995). The effort then begins to survey the minimal region for attractive candidate genes. If no such candidates are available, or the known ones have been eliminated, then novel transcripts must be identified from the region. To deal with the difficult problem of identifying transcripts from large genomic intervals, a number of new strategies, such as direct cDNA selection (Lovett *et al.*, 1991; Parimoo *et al.*, 1991) and exon trapping (Duyk *et al.*, 1990; Buckler *et al.*, 1991), have been developed over the last few years.

Once additional candidate genes have been isolated and characterized, sequence analysis in affected and unaffected individuals and the demonstration of mutations must occur to identify the gene responsible for the condition. Mutation analysis can occur in a number of ways including direct sequencing of the

Figure 1. The positional cloning approach. Briefly, positional cloning begins with the collection of families segregating the disease gene of interest which are then studied with multiple polymorphic markers until there is evidence for linkage with one or more of these. After refining the position of the locus to a small genomic region, the candidate interval is physically cloned and these clones are used for transcript identification. Sequence analysis and demonstration of mutations in a candidate gene is the final phase of this strategy (modified from Figure 2 of Collins, 1992).


Families
+
Genetic Markers
Linkage
Analysis
Localization
of gene
Finer
Genetic
Mapping
Localization
of gene
Physical
Mapping
YACs across minimal region
Transcript
Identification
Candidate
cDNAs
Mutation
Analysis
Normal
...AACTC
...AAGTC
Mutant

region or single strand conformation analysis. Though challenging, more than 50 genes have been identified to date based on their position (Collins, 1995). This dissertation describes my efforts toward adding yet another entry to the growing list of positionally cloned disease genes.

In an effort to isolate the gene responsible for CSNB1, I have used a positional cloning strategy to reduce the minimal genetic region for CSNB1, construct a physical contig encompassing the CSNB1 locus, isolate candidate genes by direct cDNA selection, and analyze one of these genes for mutations in affected individuals.

Based on previous studies of over 20 families diagnosed with X-linked CSNB, a CSNB1 locus was mapped to Xp11.23-p11.22. To refine further the location of the gene for CSNB1, additional polymorphic markers were required, in particular to define more accurately the position of the crossovers which had been identified in this set of CSNB1 families. Using a panel of radiation and conventional somatic cell hybrids, new polymorphic markers were mapped to Xp11.23-p11.22 for use in refining crossover events and to put tight distal and proximal limits on the location of the CSNB1 locus.

Once a minimal genetic region had been defined for the CSNB1 disease gene, a physical contig spanning the region was constructed from YAC and cosmid clones. The contig provided physical information about the region as well as the templates for the isolation of novel candidate genes.

A minimal set of overlapping clones were chosen from the physical contig which covered the majority of the minimal region. These clones were used to enrich for cDNA transcripts expressed in the frontal cortex, fetal brain, retina, and placenta by direct cDNA selection. Based on expression and mapping information, an EST was chosen for further characterization and analyzed as a candidate gene for CSNB1.

CHAPTER 2 - LITERATURE CITED

Clinical Characteristics of CSNB1:

Congenital stationary night blindness is a hereditary retinal disorder with autosomal dominant, autosomal recessive, and X-linked recessive patterns of inheritance (McKusick, 1992). X-linked congenital stationary night blindness (CSNB1, McKusick 310500) (McKusick, 1992), the more common of these, is a nonprogressive disorder characterized by a disturbance or absence of night vision at birth. Moderately to severely decreased visual acuity, nystagmus, strabismus, and myopia may be associated with CSNB1 (Merin *et al.*, 1970; Carr, 1974; Krill, 1977; Miyake *et al.*, 1994; Khouri *et al.*, 1988; Pearce *et al.*, 1990). Electroretinography (ERG), which measures electrical responses of the outer and middle retina, is of primary importance in the diagnosis of CSNB1. With either light or dark adaptation (i.e. photopic or scotopic), the ERG in CSNB1 patients most often shows a close to normal a-wave, and a significantly reduced or absent b-wave, resulting in a negative wave shape (Carr, 1974; Krill,

1977; Heckenlively *et al.*, 1983; Miyake *et al.*, 1994; Pearce *et al.*, 1990). Also the photopic ERG oscillatory potentials in CSNB1 patients are extremely small or absent (Heckenlively *et al.*, 1983; Lachapelle *et al.*, 1983). These results indicate the normalcy of the outer retinal layers (a-wave) and an abnormality in the bipolar cell region (b-wave). The defect in X-linked CSNB1 is believed to involve an impairment in neural transmission from the photoreceptors to the bipolar cells of the retina, although the precise nature of the defect remains unknown (Carr, 1974; Hill *et al.*, 1974). Female carriers of this X-linked disorder do not show functional disturbances by ERG and therefore cannot be identified clinically. However, a reduction of oscillatory potentials in the photopic ERG of carriers has been reported (Miyake and Kawase, 1984).

According to the suggested classification system of Miyake and coworkers (1986), there may be two types of X-linked CSNB. The complete form, which they refer to as CSNB1, is characterized by the absence of the b-wave on scotopic testing and, in most cases, is associated with myopia. The incomplete form, which he refers to as CSNB2, is accompanied with either myopia or hyperopia and a reduced scotopic b-wave, indicating some residual rod function. However, both types of X-linked CSNB have been described within one family (Khouri *et al.*, 1988; Pearce *et al.*, 1990), and these investigators contend that the differences noted between individuals and families can be interpreted by a single mutant gene hypothesis, exhibiting a wide variation in clinical expression. Whether the variable expression observed in individuals with X-linked CSNB

represents the manifestation of one or more loci (genetic heterogeneity) must await the identification of the specific gene(s) causing this eye disorder.

Mapping the CSNB1 Locus:

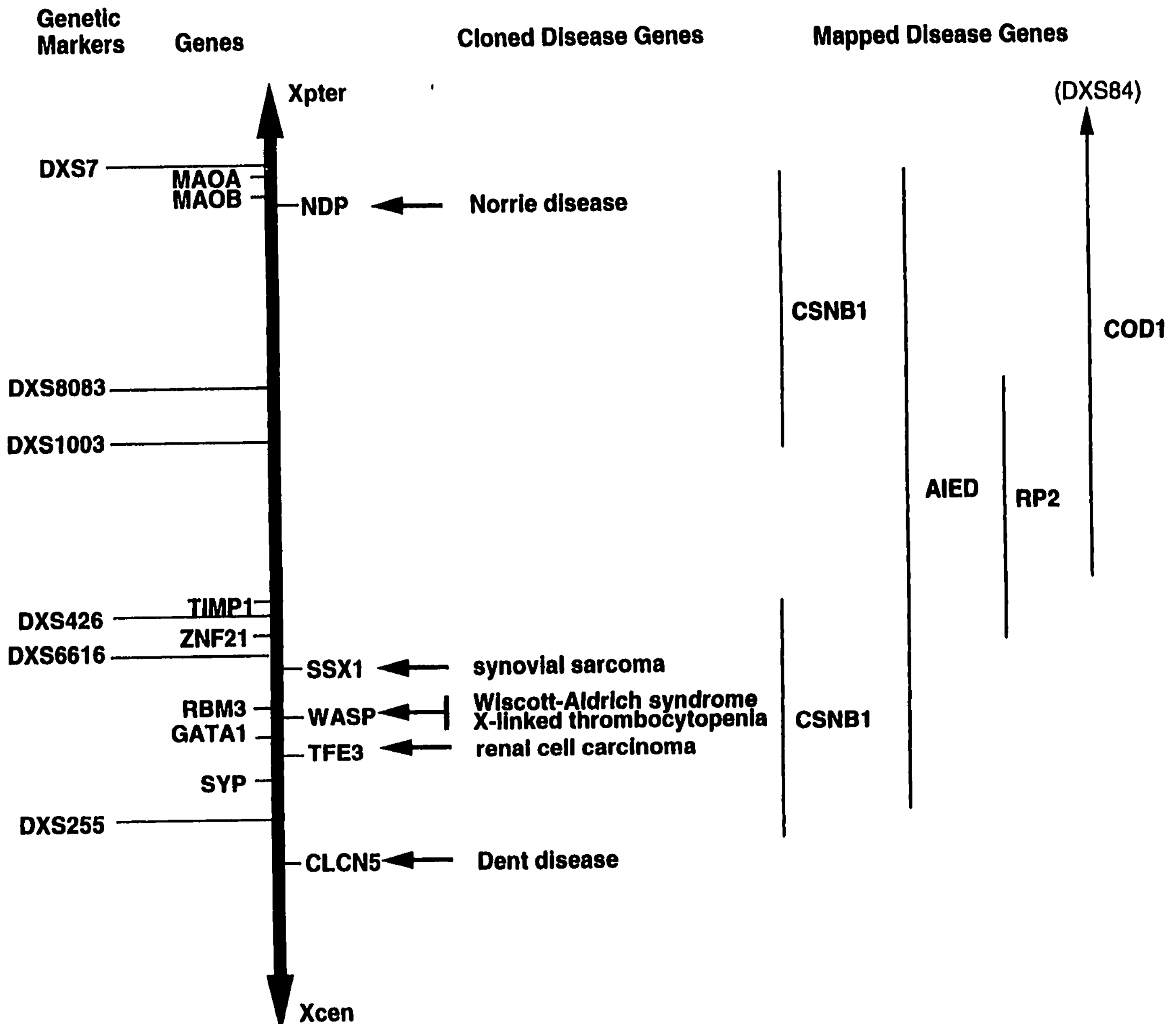
Mapping of the CSNB1 locus has involved genetic linkage analysis using various marker loci on the X chromosome. The CSNB1 disease gene has been shown to be tightly linked with the polymorphic sites at DXS7 (Xp11.3) (lod scores of 7.35 (Musarella *et al.*, 1989) and 2.02 (Gal *et al.*, 1989) at 0% recombination), and MAOB (monoamine oxidase B, Xp11.3) (lod score of >3.0 at 0% recombination (Berger *et al.*, 1995)). In one large family from Alberta (P060), linkage was detected with the marker DXS255 (Xp11.23), with a lod score of 6.73 at 6% recombination (Bech-Hansen *et al.*, 1990). Subsequent analysis of this family gave a lod score of 15.48 at 1% recombination (Bech-Hansen *et al.*, 1991). Further clinical analysis of individuals in this family led to a change in the clinical status of two individuals, previously thought to be normal, to being mildly affected. Upon recalculation, this resulted in a lod score of 15.26 at 0% recombination (Bech-Hansen *et al.*, manuscript in preparation).

Analysis of genetic recombinant chromosomes has defined distal and proximal limits to the genetic region of CSNB1 on the short arm of the X chromosome. Recombinant chromosomes have placed the CSNB1 gene proximal to DMD (Duchenne muscular dystrophy, Xp21.1) (Aldred *et al.*, 1992a; Aldred *et al.*, 1992b), proximal to the DXS7 locus (Xp11.3) (Bech-Hansen *et al.*, 1992), and proximal to MAOA/B

(Xp11.3) (Bech-Hansen *et al.*, 1991) in families whose complete/incomplete status was not described. The CSNB1 gene has also been localized proximal to OTC (ornithine transcarbamylase, Xp11.4) in a family referred to as having complete X-linked CSNB (Musarella *et al.*, 1989), and proximal to MAOB (monoamine oxidase B, Xp11.3) in a family referred to as having incomplete X-linked CSNB (Bergen *et al.*, 1994). The CSNB1 gene has been localized distal to the DXS1003 locus (Xp11.3) in a family whose complete/incomplete status was not described (Berger *et al.*, 1995), and distal to the loci for DXS255 (Xp11.22) and TIMP1 (tissue inhibitor of metalloproteinase, Xp11.3-p11.23) in families referred to as having complete X-linked CSNB (Musarella *et al.*, 1989). However, analysis of a recombinant chromosome carried by an affected daughter of a carrier mother and an affected father from the large Alberta family P060 whose complete/incomplete status has not been described, has placed the CSNB1 gene proximal to TIMP1 (Xp11.3-p11.23) (Bech-Hansen and Pearce, 1993). These results clearly raise the possibility that there may be two loci for CSNB1 on the X chromosome (Figure 2). Analysis of additional CSNB1 families and genetic markers near the CSNB1 locus (loci) will help to identify and refine the crossovers events and reduce the genetic region(s) to which the X-linked CSNB gene(s) is located.

A third CSNB locus, that cannot be clinically distinguished from the other forms of X-linked CSNB, has recently been mapped to Xp21.1, near the RP3 gene (Bergen *et al.*, 1995). Night blindness is a well known early feature of RP, and is followed by progressive

Figure 2. Gene map of Xp11.3-p11.23. Known genes in Xp11.3-p11.23 are shown to the left of the X chromosome segment illustrated, while genes implicated in genetic diseases are shown to the right of the chromosome. Candidate intervals for genes involved in ocular diseases are shown on the far right of the diagram and the defining genetic markers are indicated on the far left of the diagram.


Genetic Markers
Genes
Cloned Disease Genes
Mapped Disease Genes
Xpter
Xcen
DXS7
MAOA
MAOB
NDP
Norrie disease
DXS8083
DXS1003
TIMP1
DXS426
ZNF21
DXS6616
SSX1
synovial sarcoma
RBM3
WASP
Wiscott-Aldrich syndrome
X-linked thrombocytopenia
GATA1
TFE3
renal cell carcinoma
SYP
DXS255
CLCN5
Dent disease
CSNB1
CSNB1
AIED
RP2
COD1
(DXS84)

concentric visual field loss, pigmentary retinopathy and reduced ERG amplitudes. Some autosomal forms of CSNB and RP have been shown to be caused by different mutations in both the rhodopsin gene on 3q (Dryja *et al.*, 1993) as well as the β-subunit of rod photoreceptor cGMP-specific phosphodiesterase on 4p (Gal *et al.*, 1994). Possibly there is also a functional relationship between X-linked CSNB and RP due to different mutations in a common gene in Xp21.1. This notion is supported by the discovery of a putative missense mutation in the RP3 gene (RPGR) in a family with X-linked CSNB (Herrmann *et al.*, 1996).

In addition to CSNB1, three other eye-disease genes have been mapped to the Xp11 region including retinitis pigmentosa (RP2) (Ott *et al.*, 1990), Åland Island Eye Disease (AIED) (Alitalo *et al.*, 1991; Schwartz and Rosenberg, 1991, Glass *et al.*, 1993), and X-linked cone dystrophy (COD1) (Bergen *et al.*, 1993; Hong *et al.*, 1994). The gene for RP2 has recently been localized to the region between DXS8083 (Xp11.3) and DXS6616 (Xp11.23) (Thiselton *et al.*, 1996), placing it in an overlapping region with the CSNB1 gene that maps between DXS7/MAOA and DXS1003/TIMP1 and raising the possibility that these two diseases are also caused by mutations in the same gene. The location of AIED has been refined to the region between DXS7 (Schwartz and Rosenberg, 1991) and DXS255 (Glass *et al.*, 1993) and it has been postulated that AIED and CSNB2 are the same entity since they have been regarded by some as clinically indistinguishable (Musarella *et al.*, 1989), or that they too are also allelic. The data on the genetic localizations of AIED, RP2, and CSNB1 and CSNB2 are as

yet insufficient to clarify whether they are indeed caused by mutations in the same gene(s) due to the lack of methodological ERG standardization with respect to AIED and X-linked CSNB and the lack of informative crossover events within one pedigree. Elucidation and characterization of mutations in the genes responsible for these disorders will resolve these issues.

Genetic Markers:

Many types of sequence polymorphisms are present in the genome and can be employed in positional cloning studies. Two major types of DNA polymorphisms stem from variations in the number of repeat units, microsatellite variable number of dinucleotide repeats (VNDRs) (Weber and May, 1989) and the more complex variable number of tandem repeats (VNTRs) (Jeffreys *et al.*, 1986; Nakamura *et al.*, 1987). These polymorphisms frequently have multiple alleles in a population making them highly informative markers. Microsatellite markers have been found to be more randomly distributed on the X chromosome than the classical VNTR (Luty *et al.*, 1990) and uniform spacing would place them about once every 40 kb (Hamada *et al.*, 1984).

Another major type of DNA polymorphism comes from discrete base changes in a specific DNA sequence. These polymorphisms are the most frequent and widely distributed type of sequence variation in the genome and are usually biallelic in populations (Nickerson *et al.*, 1992). Transitional base changes appear more frequently than transversions and transitional base changes involving C and T are

favoured over those involving G and A. Many estimates on the frequency of sequence polymorphisms in the human genome have been reported and they generally range from one in every 200 bp to one in every 1,000 bp (Cooper *et al.*, 1985; Miyamoto *et al.*, 1988). In some instances, these polymorphisms provide the underlying sequence variation responsible for restriction fragment length polymorphisms (RFLPs). Although individual biallelic polymorphisms are generally not as informative as polymorphic repeats, multiple closely linked markers can be combined into haplotypes that can be as informative as a repeat polymorphism (Nickerson *et al.*, 1992). Most types of DNA polymorphisms, including simple nucleotide substitutions (Nickerson *et al.*, 1992), VNDRs (Weber and May, 1989), and complex VNTR repeats (Jeffreys *et al.*, 1986) are, or can easily be converted to, polymorphic sequence-tagged sites (STSs) which can be then be used to integrate the genetic map with the physical map of a region (Nickerson *et al.*, 1992). Conversely, the existing STS content map of the human genome potentially represents a valuable resource for the isolation of new diallelic markers and this resource has been successfully exploited to isolate new DNA sequence polymorphisms with minimal effort (Kwok *et al.*, 1996).

Somatic Cell Hybrid Mapping:

Chromosomal mapping methods based on somatic cell genetics have been central to the isolation of genetic areas of interest. Somatic cell hybrids containing several human chromosomes, a single

human chromosome, or defined fragments of human chromosomes on a rodent background, can be used both for high-resolution genetic mapping and as a resource for molecular cloning of human sequences. Hybrids containing subchromosomal fragments can bridge the limits of somatic cell genetics and molecular methods, allowing separation of DNA fragments of 0.3-30 Mb.

Hybrids with a reduced number of human chromosomes can be used in chromosomal mapping of a probe or gene and such a mapping panel is currently distributed by the NIGMS Human Genetic Mutant Cell Repository (Drwinga *et al.*, 1993). Hybrids retaining single human chromosomes are also valuable tools in human genetics because of the reduced complexity. Hybrids containing deletion and translocation derivatives of human chromosomes have the advantage that they can regionally localize DNA of interest, though the resolution of such hybrids depends upon the variation of naturally occurring chromosomal breakpoints for a given chromosome. Methods that do not rely on naturally occurring breakpoints to reduce the complexity and amount of chromosomal DNA that enters the cell, such as irradiation and fusion gene transfer, are exceptionally useful for fine mapping the chromosomal location of genes and for isolating DNA probes.

Irradiation and fusion gene transfer (IFGT) was introduced by Goss and Harris (1975) who demonstrated that chromosome fragments generated by lethal irradiation of donor human cells could be rescued by fusion to rodent recipient cells. However, initially, insufficient markers were available to exploit this technology and

IFGT methods were not widely used. Interest in IFGT was renewed when the original approach was modified by using a rodent-human somatic cell hybrid containing a single human chromosome as the donor cell (Benham *et al.*, 1989; Cox *et al.*, 1990; Goodfellow, 1991; Walter *et al.*, 1993; Walter *et al.*, 1995), with the advantage that all of the 'radiation reduced hybrids' generated were from a known chromosome.

To produce radiation reduced hybrids a donor cell is irradiated and fused with a recipient cell (Goodfellow, 1991). The irradiation step functions to deliver a lethal dose of radiation which kills the donor cell and ensures that any surviving cells are true hybrids, and also causes double-stranded breaks in chromosomal DNA. Increasing the radiation dose increases the frequency of chromosome breaks. As currently used, IFGT is not a practical method for creating maps of entire genomes, but is exceptionally useful for generating maps of portions of individual chromosomes. To map an entire chromosome between 100 and 200 hybrids are needed and, consequently, a map of the whole genome would require over 4,000 hybrids (Barrett, 1992). Walter and coworkers (1994) found a solution to this problem by reverting to the original protocol of Goss and Harris (1975). Instead of using a human-rodent hybrid as a chromosome donor, they used a diploid human fibroblast. By an extension of this work, a single set of 100 WG-RH cell lines can be used to produce a high resolution map of the entire genome. This approach is currently being used for the construction of the detailed gene map of the

human genome (Schuler *et al.*, 1996), which is described in a later section (page 35).

Panels of characterized somatic cell hybrids which may consist of hybrids produced by a variety of techniques, can facilitate the regional localization of markers assigned to a particular chromosome and provide valuable mapping information prior to or in conjunction with the application of higher resolution physical mapping techniques. Integrated maps that include polymorphic markers for genetic analysis, sequence tagged sites (STSs) for physical mapping, and expressed sequence tags (ESTs) and genes are important for any positional cloning project. The use of classical somatic cell hybrids and radiation-reduced hybrids can be a powerful tool to sublocalize markers quickly and generate maps of chromosomal regions. The use of somatic cell hybrid mapping panels complements genetic and high-resolution physical mapping, and is particularly useful in those regions of the human genome that are either difficult to clone or unstable in yeast.

Physical Contig Construction:

Construction of contiguously overlapping sets of genomic clones (contigs) spanning extended regions of the genome, defined in part by polymorphic genetic markers, can substantially facilitate the localization and identification of new disease genes as well as provide DNA templates for sequencing individual chromosomes. Olson and colleagues (1989) proposed that sequence tagged sites (STSs) be used as the landmarks to define positions on the physical map. STSs are

short tracts of single-copy DNA sequence that can be easily recovered at any time by PCR. There are a number of advantages to the STS strategy. Distribution of such DNA markers between laboratories is greatly eased since the markers are information based (DNA sequence retrievable from publication or database) rather than resource based. Secondly, polymorphic genetic markers as well as ESTs are a variation of a STS and both can be used to facilitate the integration of the genetic and transcript maps with the physical map of a particular region, respectively. Sequence information is the natural language of physical mapping and PCR and cost-effective oligonucleotide synthesis provide the technical means to root the physical map of the human genome firmly in the genomic DNA sequence itself.

Construction of a physical contig involves the ordering and spacing of DNA segments, each identified by a unique STS. The resolution of STSs required for the generation of a physical map depends on the capacity of the cloning vectors. Traditionally, cosmids and yeast artificial chromosomes (YACs) have been the primary tools for physical mapping. Present cosmid cloning systems have maximum capacities of approximately 40 kb (Feiss *et al.*, 1977; Hohn and Collins, 1980). One of the advantages of the cosmid system is the high efficiency of *in vitro* packaging, permitting the construction of high-complexity libraries from small amounts of DNA (e.g. derived from flow sorted chromosomes (Nizetic *et al.*, 1991)). Though pure cosmid DNA can be easily obtained using simple alkaline extraction procedures, the comparatively small insert size

and frequent instability of clones (Kim *et al.*, 1992) complicate the use of cosmids for physical mapping of complex genomes.

YAC vectors can be used to clone segments of several hundred kb up to 1 Mb (Burke *et al.*, 1987). STSs map with an average resolution of 100 kb (requiring 30,000 STSs to cover the whole human genome), would allow recovery for most regions of the human genome by cloning in the YAC system, but not in cosmids (Olson *et al.*, 1989). Overlapping yeast artificial chromosomes (YACs) would then be fitted together by their common content of STSs. Despite the important role YACs have played in the construction of physical maps of entire human chromosomes, such as chromosome 21 (Chumakov *et al.*, 1992), the Y chromosome (Foote *et al.*, 1992), and recently 94% of the entire human genome (Hudson *et al.*, 1995), they have the disadvantages of low transformation efficiency, inconvenient manipulation, difficult isolation of cloned DNA, and a high rate of chimerism and clonal instability (Green *et al.*, 1991; Boycott *et al.*, 1996), often restricting the reliability of YACs for mapping and sequencing purposes.

Efforts to overcome the limitations of both cosmids and YACs have resulted in several alternative cloning approaches. Fosmids (Kim *et al.*, 1992) stably propagate cosmid sized human DNA inserts in a F factor-based (*Escherichia coli* fertility plasmid) vector, but are again restricted by their small insert size. A system based on bacteriophage P1 is also in use and the P1 vector has a maximum cloning capacity of 100 kb (Sternberg, 1990; Pierce *et al.*, 1992). The P1 system has the advantage of allowing selection for recombinants

over non-recombinants, as well as the presence of two P1-derived replication mechanisms, the single-copy replicon for stable propagation of clones, and the multi-copy replicon under control of the inducible *lac* operator, for preparing clone DNA. Though an elaborate in vitro packaging system is required, an alternative T4 packaging system may allow for an insert size up to 122 kb (Rao *et al.*, 1992).

Like Fosmids, the bacterial artificial chromosome (BAC) vector also uses the *Escherichia coli* fertility plasmid (Shizuya *et al.*, 1992). Circular ligation products are introduced into bacterial cells by electroporation but, unlike the P1 system, there is no positive selection for recombinants, necessitating the identification of recombinant clones by colony hybridization with human repetitive DNA probes. The BAC system provides stable propagation and maintenance of relatively large (>300 kb) genomic DNA fragments as single-copy plasmids in recombination-deficient host strains. BACs are characterized by high stability, minimal chimerism, and ease of purification of large inserts (Schmitt *et al.*, 1996). One limitation of this system, however, is that DNA recovery from clones is relatively low owing to the replication F-factor limiting the BAC plasmid to one or two copies per cell.

Recently, the features of the P1 and F-factor systems have been combined into a new P1-derived vector for the cloning of large DNA fragments by electroporation (Ioannou *et al.*, 1994). The P1-derived artificial chromosomes (PACs) have an average insert size of 130-150 kb and are characterized by high stability and minimal chimerism.

The new P1 vector maintains all the advantages of the earlier P1 system, including control of insert copy number and the ability to select for recombinant clones, but removes the need for an elaborate *in vitro* packaging system and instead recombinant DNA can be introduced into *E. coli* by electroporation. Both PACs and BACs are now becoming more popular tools for use in physical mapping as the STS content map of the human genome continues to grow and the use of smaller insert containing clones becomes more feasible.

Almost all cf the 160 million base pairs of the X chromosome are now covered by a YAC contig that was verified and aligned by its content of STSs and other markers placed by cytogenetic, somatic cell hybrid, or linkage mapping techniques (Nelson *et al.*, 1995; Nelson *et al.*, in press). Efforts are now underway to convert the long-range YAC contigs on the X chromosome, as on other chromosomes, into higher resolution cosmid, BAC, and PAC contigs, generating sequence-ready templates. Though the interest in Xp11.23-p11.22 has promoted efforts to obtain continuous cloned coverage in YAC clones, previous efforts to this one have only been partially successful (Coleman *et al.*, 1994; Derry *et al.*, 1994; Knight *et al.*, 1994; Mazzarella and Srivastava, 1994; Kwan *et al.*, 1995a).

Genes and ESTs in Xp11.23:

Several genes have been cloned and mapped to the short arm of the human X chromosome between the zinc finger gene ZNF21 and DXS255 in Xp11.23-p11.22 (Figure 2). These include the erythroid-specific transcription factor GATA1 (Caiulo *et al.*, 1991), the

transcription factor for the enhancer μE3 (TFE3) (Beckmann *et al.*, 1990; Lafreniere *et al.*, 1991), synaptophysin (SYP) (Ozcelik *et al.*, 1990; Lafreniere *et al.*, 1991), and recently RBM3 (Derry *et al.*, 1995). In addition several new transcripts have been identified including MG21, MG44, MG61, MG81 (Geraghty *et al.*, 1993), DXS1011E (Parrish and Nelson, 1993), and IS2, IS3, IS4, and IS7 (Derry *et al.*, 1994). The p11 region of the X chromosome is of particular medical interest because of its involvement in many known genetic conditions. Recently the genes responsible for Norrie disease (NDP) (Berger *et al.*, 1992), Wiscott-Aldrich syndrome (WASP) (Derry *et al.*, 1994), X-linked thrombocytopenia (WASP) (Villa *et al.*, 1995), Dent disease (CLCN5) (Fisher *et al.*, 1995), and two genes (SSX1 and SSX2) involved in t(X;18)-positive synovial sarcomas (Clark *et al.*, 1994; de Leeuw *et al.*, 1994; Crew *et al.*, 1995; de Leeuw *et al.*, 1995) have been cloned from the Xp11.23-p11.22 region. In addition, the TFE3 gene has very recently been implicated in t(X;1)(p11.2;q21.2) papillary renal cell carcinomas (Sidhar *et al.*, 1996) (Figure 2).

Identification of Novel Genes:

For those involved in positional cloning of human disease genes, finding candidate coding regions within a genomic contig is rate-limiting. Because of the limitations of genetic analysis and linkage disequilibrium studies, the candidate genomic region cannot usually be narrowed to less than several hundred kilobases and is often a megabase or more in size, containing dozens of genes. All known genes and ESTs in the interval are evaluated as candidates, which can

often involve detailed genomic characterization in the case of ESTs, and if none are identified as the disease gene, a search for new candidates must begin. Techniques that are currently used to identify coding regions within large DNA segments fall into two general classes, expression-dependent and structure-based techniques.

Expression-dependent techniques rely on identifying cDNAs that correspond to mRNAs derived from the region of interest. Examples of this type of technique are direct cDNA selection (Lovett *et al.*, 1991; Parimoo *et al.*, 1991), the hybridization of radiolabelled cDNAs to arrayed genomic clones (Hochgeschwender *et al.*, 1989), the use of purified YAC DNA to screen directly a cDNA library (Elvin *et al.*, 1990), and techniques based on selective amplification of cDNAs that contain human sequences from somatic cell hybrids (Lui *et al.*, 1989).

Structure-based techniques select for functional elements involved in gene expression and are expression independent. Examples of the second type of technique are exon amplification (Duyk *et al.*, 1990; Buckler *et al.*, 1991), the use of CpG islands as potential sign posts for the ends of some transcription units (Lindsay and Bird, 1987), the use of zoo blots to detect cross-species conservation of genomic sequences (Rommens *et al.*, 1989), and homologous recombination (Kurnit and Seed, 1990).

Direct sequencing of a genomic region can also be considered a structure-based technique for isolating new genes. Long-range sequencing has, until now, been considered too costly and labour intensive to sustain systematic gene-finding in the human genome.

However, as more of the YAC-based physical contigs are converted to sequence-ready templates consisting of PAC, BAC, or cosmid clones, direct sequencing as a route for novel gene isolation becomes more feasible. At the present time, about 8 Mb of the 3600 Mb of the human genome have been sequenced (Gibbs, 1995), and as the Human Genome initiative turns its focus to human genomic DNA sequencing, this total will increase very rapidly. Direct sequencing of targeted regions containing disease genes will become the route to take to isolate and characterize additional candidate genes in the final stages of a positional cloning project. Recently, the cloning of the RPGR gene implicated in RP3 involved sequencing two cosmids covering the proximal part of the 75 kb minimal region (Meindl *et al.*, 1996). However, for positional cloning projects with disease genes localized to within a Mb, direct sequencing of such a region is an option only for laboratories with access to large sequencing centres. Therefore, the approach for the past several years, and possibly for the next couple of years while the large-scale human genome sequencing effort gets underway, has been to use either direct cDNA selection, exon amplification, single-pass sequencing of small regions, or a combination of all three to isolate additional candidate genes from a targeted region. Both direct cDNA selection and exon amplification have become widespread in use and been successful in isolating a number of disease genes.

Exon Amplification: Exon amplification facilitates the recovery of transcribed sequences in cloned mammalian genomic DNA through

the functional identification of cis-acting sequences required for RNA splicing (Buckler *et al.*, 1991). Exon amplification has had an increasing number of successful applications, including the identification of disease genes responsible for X-linked glycerol kinase deficiency (Walker *et al.*, 1991), neurofibromatosis type II (Trofatter *et al.*, 1993), and Menkes disease (Vulpe *et al.*, 1993). In exon amplification (Auch and Reth, 1990; Buckler *et al.*, 1991; Hamaguchi *et al.*, 1992), fragments of genomic DNA are subcloned into a vector, such as pSPL-1 (Buckler *et al.*, 1991) or pSPL-3 (Church *et al.*, 1994), containing a constitutive promoter, splice donor and acceptor sites. Constructs are transfected into COS-1 cells where transcription is initiated from the vector promoter. If the genomic fragment contains an intact exon, the transcripts are spliced so that the exon is juxtaposed with vector sequences in mature transcripts. These transcripts can then be reverse transcribed and amplified using PCR. The genomic substrate most frequently used for exon amplification is cosmids, but phage clones (Vulpe *et al.*, 1993), and PACs (Roy *et al.*, 1995), have also been used. Using YAC DNA as the genomic substrate requires additional steps either before (purifying the YAC from the yeast chromosomes), or after (removing the yeast background) exon amplification. Because of this YACs are not as frequently used in exon amplification and YAC contigs are usually converted to cosmid contigs before exon amplification begins.

Exon amplification is independent of gene expression and in effect can facilitate 'normalization' of genes in the region (Church *et al.*, 1994). Because of the independence of exon amplification to gene

expression, tissue specific and developmentally regulated genes will be isolated with the same efficiency as ubiquitously expressed genes. However, exon amplification is limited by its requirement for the presence of functional 3' and 5' splice sites flanking a target exon. Intronless or single intron genes will presently be missed by this approach (Church *et al.*, 1993). Exon skipping can also occur because different exons may splice to the flanking sequences with different efficiencies (North *et al.*, 1993).

Once a set of candidate exons have been isolated by exon amplification, it must be confirmed that they are indeed part of a gene. Exons can be cloned and the flanking sequences can be determined to see if they show good homology to the 5' or 3' splice consensus (North *et al.*, 1993). Exons can also be used to screen cDNA libraries and used as probes on Northern or Zoo blots for confirmation. Potential exons with open reading frames of greater than 100 bp can also be tested for coding potential with 'neural-net' gene prediction programs such as GRAIL (Uberbacher and Mural, 1991). Primers can be constructed for true exons and expression profiles can be derived by PCR based screening of various cDNA libraries. Full-length transcripts can be isolated by using the exon as a probe on cDNA libraries and by performing RACE PCR (Frohman *et al.*, 1988), and isolated exons can be linked by common cDNAs.

Direct cDNA selection: To overcome the limitations of more conventional approaches to the isolation of coding regions, two groups (Lovett *et al.*, 1991; Parimoo *et al.*, 1991) devised PCR-based

methods for enriching cDNAs. These methods have been called 'cDNA selection', 'direct selection', and 'direct cDNA selection' and have been used recently for the identification of the genes responsible for Wilson disease (Bull *et al.*, 1993), and Wiskott-Aldrich syndrome (Derry *et al.*, 1994).

The concept underlying direct selection is that a large genomic region can be used to 'fish out' complementary cDNAs. This is accomplished by hybridizing a solution containing the cDNAs to DNA from the genomic contig. The genomic region is the hook and the cDNAs are the fish (fishing analogy (Lovett, 1994b)). The genomic region can consist of a single cloned piece of DNA or a combination of several clones. Pools of cosmids (Bull *et al.*, 1993; Futreal *et al.*, 1994; Miki *et al.*, 1994; Simmons *et al.*, 1995; Osborne-Lawrence *et al.*, 1995) and pools of YACs (Osborne-Lawrence *et al.*, 1995) have both been used. Similarily, cDNAs from multiple tissue sources can also be pooled for selection while retaining the identity of the tissue source if amplifiable by a unique primer set. In this way, direct selection can potentially address tissue and developmental specific expression in the same round (Lovett, 1994b).

cDNAs for direct selection can be derived from conventional libraries, by amplifying a population of cDNA inserts from the library using PCR and flanking vector primers. However, conventional amplified libraries can contain unforeseen contaminants and the representation of sequences can be skewed. Libraries derived from complex tissues are unlikely to contain cDNAs that represent rare transcripts ($<1 \times 10^{-6}$). Therefore, it is preferable that complex sets

of primary uncloned cDNAs that have been ligated to adapters to permit PCR amplification be used for direct selection. These sets of cDNAs have the advantage that they circumvent the potential loss of particular transcripts as a result of cloning bias in the construction of cDNA libraries. In addition, if the genomic DNA is present in a limiting amount during the selection then some level of cDNA abundance normalization may be possible. It becomes a game of musical chairs and most copies of an abundant cDNA will be left without a 'genomic chair' and therefore will not be selected (Lovett, 1994b), while a low-abundance cDNA will be present closer to a 1:1 ratio with its 'genomic chair', and most molecules of such a cDNA species will hybridize and be selected. This should result in a net decrease in the relative abundance of the high-abundance cDNAs and an increase in the relative abundance of the low-abundance cDNAs. Some level of abundance normalization has been observed in selections where the genomic target is limiting (Morgan *et al.*, 1992).

In the original design of this method (Lovett *et al.*, 1991; Parimoo *et al.*, 1991), the genomic DNA was immobilized on a filter and after hybridization with the cDNA pools, the filter was washed extensively and the bound cDNAs eluted from the filter. A PCR amplification step was incorporated to generate sufficient eluted material for additional selection steps or for molecular cloning. The major difference between these methods was the way in which repetitive elements within the cDNAs and genomic DNA were managed. In one, repeats within the cDNAs were blocked (Lovett *et al.*, 1991), whereas in the other (Parimoo *et al.*, 1991), repeats in the genomic DNA were

blocked. Both strategies worked well and enriched the desired cDNAs several thousand-fold after two rounds of selection. The largest target used in these studies was a 550 kb YAC and it was used to enrich a cDNA, initially present at one part per million, 1,000-fold in one round of selection (Lovett *et al.*, 1991).

A problem encountered in filter-based selection was the lack of control over the hybridization, which meant it could be somewhat variable. This was attributed to the pseudo-first order hybridization kinetics of filter hybridizations (Britten and Davidson, 1985). The method was adapted so that both genomic DNA and cDNA were present in a solution hybridization reaction, which follows more easily controlled second-order kinetics. In 1992, two reports describing the use of biotin-streptavidin capture systems to facilitate solution hybridization were published (Korn *et al.*, 1992; Morgan *et al.*, 1992). Korn and coworkers (1992) reported a 80,000-fold enrichment of a particular cDNA when a single cosmid was used. Morgan and coworkers (1992) reported enrichments ranging from several thousand-fold to more than 100,000-fold of particular cDNAs when a 425 kb YAC was used. Such dramatic enrichments can be achieved even with large genomic targets, which is often what positional cloners are left with after minimizing the genetic region for a particular disease gene. Selection using 1 Mb of chromosome 5q31 (147 pooled cosmids) achieved a 20,000-fold enrichment for the reporter cDNA annexin 6 (Lovett, 1994b).

Despite potential problems in using filter-based selections, this version of cDNA enrichment has nonetheless been shown to be

effective. The gene for Wiscott-Aldrich syndrome was isolated by filter-based selections of YACs and cosmids (Derry *et al.*, 1994), and the gene for Wilson disease was isolated by filter-based selections of cosmids (Bull *et al.*, 1993). It would appear that both filter- (Bull *et al.*, 1993; Derry *et al.*, 1994; Rommens *et al.*, 1994) and solution-based hybridizations (Futreal *et al.*, 1994; Miki *et al.*, 1994; Osborne-Lawrence *et al.*, 1995; Simmons *et al.*, 1995) are presently in use and the variability of filter based hybridizations can be circumvented by carefully controlled experimental conditions.

Direct selection suffers background hybridization from a number of contaminants. To find new genes efficiently among selected products isolated by this method, it is important to either preblock and/or screen out common contaminanting sequences. When YAC DNA is used as the genomic target, ribosomal cDNAs represent a major contaminant (Lovett *et al.*, 1991; Parimoo *et al.*, 1991; Lovett, 1994a; 1994b). Ribosomal RNAs are represented in most, if not all, libraries or pools of cDNAs. Even when YACs are purified by pulsed field gel electrophoresis, they are inevitably slightly contaminated with yeast chromosomes, including yeast chromosome 12 containing over 100 copies of the yeast ribosomal locus. Ribosomal cDNAs can comprise up to 70% of selected sequences when ribosomal sequences are not blocked (Lovett, 1994a; 1994b).

cDNAs derived from polyadenylated mitochondrial transcripts are also present in all cDNA libraries. Such cDNAs can be enriched for by contaminating *E. coli* genomic DNA when DNA cloned into *E. coli* vector-host systems, such as cosmids, are used (Lovett, 1994a). In

this case, host DNA can be removed by density gradient centrifugation. The 16 kb circular mitochondrial genome runs aberrantly in pulsed field electrophoresis and can contaminate YAC DNA preparations and therefore selections that use YACs. The yeast 2µ plasmid is another circular DNA that contaminates YAC preparations. cDNA libraries often contain this DNA molecule because many investigators, as well as commercial vendors, have used yeast tRNAs, containing the 2 micron sequence, as carrier DNA during cDNA synthesis. Both of these contaminating genomes can be eliminated from YAC preparations by a pre-electrophoresis step (Lovett, 1994a).

Intermediate repetitive sequences, such as Alu repeats, can be blocked efficiently but occasionally an Alu-containing cDNA is isolated and it can be difficult to ascertain whether it is a spurious contaminant or was selected by hybridization to a stretch of single-copy sequence. If random primed cDNA libraries were used, another part of the cDNA that does not include the repeat may also have been isolated and this can often help distinguish these possibilities. These intermediate repeats may be present in as much as 10% of cDNAs isolated in some selections, depending on their density in the genomic target sequence. Such repeats can map to the contig quite specifically but a whole-genome Southern blot hybridized with a cDNA probe resolves this difficulty (Lovett, 1994a). Finally some single-copy sequences that do not map to the target genomic contig are normally selected. These contaminants may comprise up to 30% of the selected material that remains when other contaminants have

been eliminated. However, these contaminants are easily identified by mapping experiments, and the remaining 70% of clones do map to the genomic contig.

The major limitation of direct selection is that it is expression-dependent and only specific subsets of genes are present in a given tissue. A gene can potentially be missed by this procedure because it is not represented in the subsets of cDNAs undergoing selection, even though direct selection can assay a number of different tissues at once. Genes expressed only within a certain window of development would be particularly vulnerable to this limitation (Lovett, 1994b). Direct selection is also subject to limitations inherent in aspects of the procedure itself such as those that result from the use of PCR. A length bias is one such limitation in which the tendency is for ever shorter fragments to be generated at each successive round of selection and PCR amplification. Another limitation is that sequence errors are increasingly likely to be introduced with each round of PCR amplification (Lovett, 1994b). Direct selection can also be affected by factors associated with hybridizations. Certain genes that have short exons may fail to hybridize to their genomic locus efficiently. In addition certain cDNAs hybridize to non-unique sequences such as related members of a gene family and pseudogenes (Lovett, 1994a; 1994b).

Once a potential transcript has been isolated by direct cDNA selection, it must be confirmed that it is indeed part of a gene. As mentioned above, most contaminant clones are easily identified by mapping and removed from further consideration. The remaining

transcription units can be sequenced to facilitate detection of overlap between cDNA clones and for use in similarity searches of the sequence databases. In addition, primer pairs specific to the cDNAs can be constructed and initial expression profiles of the novel cDNAs can be derived from amplification of various cDNA libraries. The cDNAs can also be used as probes on Northern blots for further confirmation and sizing of the full length transcripts. Screening cDNA libraries with the isolated cDNA fragment and RACE PCR (Frohman *et al.*, 1988) can be used to isolate full length cDNAs.

Both exon amplification and direct cDNA selection have been successful in isolating specific human disease genes. A combination of these approaches can potentially isolate all genes in a specific region of interest, and should be considered in any positional cloning strategy and is a must for the generation of a complete transcription map. Whether direct selection, exon amplification, or a combination of both methods are chosen for a particular project really depends on the resources available and the characteristics of the disease cloning project. For disease genes that are almost certainly expressed in an affected tissue and are mapped to a relatively large genomic interval (>>1 Mb) covered only in YACs, direct selection would enable the tissue-specific transcripts to be isolated efficiently for analysis as candidate genes.

Mutation Analysis:

Mutations result from different types of DNA alteration. In the absence of cytogenetic rearrangements or large deletions, mutation

analysis of candidate genes can be a time consuming step for the positional cloner. Two classes of mutations result from a single nucleotide substitution: a missense mutation which changes a single amino acid but does not change the rest of the protein, and a nonsense mutation which occurs when the nucleotide substitution produces a stop codon and the translation of the protein product is terminated at that point. Frameshift mutations occur when one or more nucleotides are either inserted or deleted and if the number inserted or deleted is not a multiple of three, a change in the reading frame alters the remainder of the translation of the protein, and, in most instances, a stop signal is encountered prematurely. Another class of mutation, exon/intron splice-site mutations, result from single nucleotide substitutions or the insertion or deletion of nucleotides in the splice sites, causing abnormal inclusion or exclusion of DNA in the coding sequence and an aberrant protein. A final class of mutation occurs when a substitution, insertion or deletion in a gene's regulatory region causes a change in protein transcription.

Mutation detection can be divided into two categories: a scanning mode where a stretch of DNA is searched for unknown mutations, and a diagnostic mode, where specific tests are designed to detect known or expected types of mutations. Since the final goal of a positional cloning project is to identify previously unknown mutations, detection methods used mainly to search a stretch of DNA for such mutations will be discussed. The Southern blot offers a good first step in mutation analysis and remains one of the fastest

methods to screen for mutations (Grompe, 1993). Detailed knowledge of the structure and sequence of a gene is not required and a preliminary screen of genomic DNA from affected individuals can be performed with a probe of interest immediately after its isolation. Large deletions and insertions may be detected by the presence of junction fragments, loss of bands, or changes in band intensities in blots and point mutations may also be detectable if they alter restriction sites (Grompe, 1993).

Other detection methods for identifying previously unknown mutations fall into two classes, those based on aberrant migration of mutant molecules during electrophoresis, including single-stranded conformation (SSC) analysis (Orita *et al.*, 1989a; Orita *et al.*, 1989b), denaturing gradient gel electrophoresis (DGGE) (Sheffield *et al.*, 1989), and heteroduplex analysis (Nagamine *et al.*, 1989), and those strategies that rely on cleavage of RNA or DNA molecules prior to analysis, such as RNase based cleavage (Myers *et al.*, 1985) and chemical mismatch cleavage (Cotton *et al.*, 1988). New enzymatic cleavage methods using either mismatch repair enzymes (Wagner *et al.*, 1995) or resolvases (Marshal *et al.*, 1995) offer great promise for analysis of larger fragments of DNA in the future. The strategies relying on cleavage of RNA or DNA molecules are at this time not widely used and will not be discussed further.

The single-stranded conformation (SSC) assay (Orita *et al.*, 1989a; Orita *et al.*, 1989b) has become one of the most widely used of the scanning technologies. Wild-type and mutant target DNA are amplified by PCR, denatured and then electrophoresed side by side

through a non-denaturing polyacrylamide gel. The two single-stranded DNA molecules from each denatured PCR product assume a three-dimensional conformation which depends on the primary sequence. Sequence differences between wild-type and mutant DNA may result in differential migration of one or both of the mutant strands. Amplification products are rendered radioactive by the addition of [$\alpha-^{32}P$] dCTP to the PCR reaction, and differential migration can be detected by autoradiography. DNA sequencing can then be used to determine the exact nature of the alteration in the PCR product. SSC analysis is simple and relatively sensitive, detecting 70-95% of mutations in PCR products of 200 bp or less (Michaud *et al.*, 1992). This method does not detect all sequence changes but the increased throughput possible with SSC analysis makes it an attractive, viable alternative to direct sequencing for mutation detection on a research basis. SSC analysis will not normally detect large deletions, insertions, or duplications, which might affect one or more complete exons or even the entire gene, nor will it detect a regulatory mutation. After identification of an aberrant band by SSC analysis, the location of the sequence difference in the fragment has to be determined by sequencing.

Denaturing gradient gel electrophoresis (DGGE) relies on the differential electrophoretic migration of wild-type and mutant double-stranded DNA (generated by PCR) through an increasing gradient of denaturing agent for the detection of mutations. As the DNA migrates, the strands progressively dissociate in discrete sequence-dependent domains of low-melting temperature and this

partial "melting" of the double-stranded DNA leads to an abrupt decrease in mobility. The sensitivity of DGGE is greatly enhanced if heteroduplex DNA between wild-type and mutant sequences is used for the analysis (Sheffield *et al.*, 1989). Once the appropriate PCR primers and denaturant conditions have been developed for a specific region, single base differences can be detected with 95% accuracy in PCR products of up to 600 bp in length. Heteroduplex molecules with a single base pair variance may also show differential mobility from homoduplexes in regular polyacrylamide gels (Nagamine *et al.*, 1989). New gel matrices (Hydrolink and MDE from AT Biochem) have become available which markedly enhance the ability to detect mutation induced mobility shifts in heteroduplex molecules with a sensitivity similar to SSC analysis in fragments less than 300 bp in length. The location of the sequence difference in the fragment has to be determined by sequencing.

The necessary final step of any mutation detection method is DNA sequencing because it defines the location and nature of the change. Rapid, accurate and efficient direct sequencing represents the ideal mutation scanning technique. To sequence PCR products successfully by the conventional dideoxy termination protocol, the double-stranded PCR product must be converted into a single stranded sequencing template. Several methods have been described to achieve this including asymmetric PCR (Gyllensten and Erlich, 1988), biotinylation of one of the PCR primers (Gibbs *et al.*, 1990), and genomic amplification with transcript sequencing (GAWTS) (Stoflet *et al.*, 1988). The most recent modification of the dideoxy chain

termination sequencing technology is cycle sequencing (Ruano and Kidd, 1991). In this method as little of 5 fmol of starting template can be simultaneously amplified and sequenced by the addition of dideoxy terminators to a PCR reaction.

Dideoxy fingerprinting represents a combination of SSC analysis and direct sequencing (Sarkar *et al.*, 1992). After amplification, the PCR product is sequenced with a dideoxy terminator to generate a fingerprint of bands. The sequencing reactions from wild-type and mutant samples are then electrophoresed through a non-denaturing polyacrylamide gel. Mutations are detected as shifts of individual bands in the ladder and this also provides some information about the localization of the sequence alteration.

The relative usefulness of these techniques for the detection of single base alterations depends on a number of factors including the size of the fragment to be analyzed, the percentage of mutations detectable, whether the method provides exact location of the mutation, and whether sequence information and genome organization is known. Once a base change has been identified, three criteria are used for distinguishing potentially disruptive disease-causing mutations from polymorphisms in candidate genes and include cosegregation of the variant with the disease in the family, absence of the variant in control chromosomes, and amino acid substitution in, or truncation of, the protein encoded by the variant sequence.

The Positional Candidate Approach:

The positional cloning strategy in its purest form is already beginning to give way to a streamlined version, which has been referred to as the 'positional candidate' approach (Collins, 1995). In the ideal case, the critical region contains a 'positional candidate' whose known or inferred function relates to the pathophysiology of the disease. Examples of this type of approach would include the identification of fibrillin as the gene responsible for Marfan syndrome (Dietz *et al.*, 1991). Even in the absence of an obvious candidate gene, regional gene catalogues accelerate the search by providing a wealth of markers and transcripts. As a component of the Human Genome Project, large scale sequencing of complementary DNAs (cDNAs) from a variety of tissues is well underway through the efforts of the I.M.A.G.E. Consortium and the Wash U-Merck cDNA sequencing project (Hillier *et al.*, 1996; Lennon *et al.*, 1996). At the present time the public cDNA collection consists of more than 450,000 human sequences and it is estimated that the current EST database may represent more than half of all human genes (Schuler *et al.*, 1996). By focusing on 3' untranslated regions, Schuler and coworkers (1996) compared the sequence of 163,215 3' ESTs and 8516 3' ends of known genes from Genbank, and reduced this set to 49,625 clusters. Using two sets of whole genome radiation hybrids (WG-RH), the Genebridge4 RH panel (Gyapay *et al.*, 1996) and the G3 RH panel, and the CEPH YAC library, 16,354 gene based STSs were localized to within a few megabases of DNA (Schuler *et al.*, 1996). In recent years the value of such a human gene map has become

increasingly clear as hunts for disease genes have been rapidly accelerated by combining linkage information with partial inventories of candidate genes in the region. A comprehensive gene map, such as the one recently released (Schuler *et al.*, 1996), should theoretically allow positional cloners to proceed immediately to gene characterization bypassing the often time consuming identification of novel transcripts.

CHAPTER 3 - MATERIALS AND METHODS

General Materials and Methods:

Hybrid Cell Lines:

Radiation-reduced human-hamster hybrid cell lines were kindly provided by Dr. J. Knight (Institute of Cancer Research, St. Thomas' Hospital, London, UK) and Dr. H. Zoghbi and K. Ellison (Baylor College of Medicine, Houston, Texas). Three hybrids (H21, H99, and H151) were chosen from a panel of 226 hybrids on the basis of marker retention (J. Knight and P. Goodfellow, unpublished data). These hybrid cell lines were generated by subjecting the hybrid line C12D ($HPRT^-/TK^+$; carrying a human X chromosome in a hamster background) to 20,000 rads of radiation followed by selective fusion rescue with hamster cell line A23. Seven additional radiation-reduced hybrids (K6, K11, K16, K35, K52, K54, and K55) were chosen from a panel of 55 hybrids developed by Dr. H. Zoghbi and K. Ellison. These cell lines were produced by subjecting the hybrid line AG9.1 ($HPRT^-/TK^+$; carrying the short arm of the X chromosome in a hamster background) to 7,000 rads of radiation followed by fusion rescue with hamster cell line A23 (Ellison *et al.*, 1993).

Four conventional hybrids were also used in these studies. GM06318B, GM07298, and GM10063, were obtained from the NIGMS Human Genetic Mutant Cell Repository, and hybrid SIN176, a human-hamster hybrid with del(Xp)(pter-p22.11::p11.23-qter) (Ingle *et al.*, 1985), was kindly provided by Dr. A. Zinn (University of Texas

Southwestern Medical School, Dallas, Texas). All of the hybrids described are part of a hybrid mapping panel that has been characterized and described (Boycott *et al.*, in press).

PCR-conditions for Mapping;

Nonradioactive PCR was performed on either a Perkin-Elmer Thermocycler 480 or a MJ Research DNA Engine using the TNK buffer system (Blanchard *et al.*, 1993). One hundred ng of template was used in a 5 µL reaction containing 10 mM Tris-HCl (pH8.3), 1.5 mM $MgCl_2$, 5 mM NH_4Cl, either 25 mM KCl (TNK25), 50 mM KCl (TNK50), or 100 mM KCl (TNK100), 0.4 µM each primer, dNTP mix (100 µM of each dATP, dGTP, dCTP, and dTTP), and 0.5 U *Taq* DNA polymerase (GIBCO/BRL). The reaction was run for 35 cycles of 94°C for 30 s, 55°C for 45 s, and 72°C for 45 s, following an initial denaturation at 94°C for 7 min and finishing with a final extension step at 72°C for 7 min.

Radioactive PCR was carried out in a 20 µL reaction containing 25 to 50 ng of template, 50 mM KCl, 10 mM Tris-HCl pH8.3, 1.5 mM $MgCl_2$, 170 µg/mL BSA, 0.05% Tween-20, 0.05% Nonidet P-40, 0.25 µM of each primer, dNTP mix (20 µM of each dATP, dGTP, dCTP, and dTTP), 3-5 µCi [α-32P] dCTP at 3,000 Ci/mmol (Amersham), and 0.5 U *Taq* DNA polymerase (GIBCO/BRL). Following an initial denaturation at 94°C for 7 min, standard amplification conditions were 94°C for 30s, 55°C for 30s, and 72°C for 30s, followed by a final extension step at 72°C for 7 min.

Analysis of PCR Products by Gel Electrophoresis:

Nonradioactive PCR products were analyzed in 2% Nusieve®GTG®agarose (FMC BioProducts)/1% agarose (GIBCO/BRL) gels or 1.5% SEPARIDE (GIBCO/BRL)/1% agarose gels (Boycott, 1996), with 1X TBE (89 mM Tris, 89 mM boric acid, 2 mM EDTA) or 1X TAE (39 mM Tris, 20 mM acetic acid, 10 mM EDTA), respectively, as the running buffer. Ethidium bromide was added to the molten gel solution prior to casting, at a concentration of 0.5 µg/mL. Prior to loading, the PCR reaction was mixed with loading dye (1X, 5% glycerol and 0.035% orange G) and electrophoresis was carried out at 90-100 V for 1 hour on a 10 cm gel tray. pBluescript cut with *Hae*II was routinely used as the DNA marker standard. Products were visualized with an ultraviolet transilluminator set on 302 nm and photographed with Polaroid Type 52 film using a red filter.

Radioactive PCR products were analyzed in 2% Nusieve®GTG®agarose/1% agarose gels with 1X TBE as the running buffer. Two µL of PCR product was mixed with an equal volume of loading dye (10X, 50% glycerol and 0.35% orange G) and electrophoresis was carried out at 90-100 V for 1 hour on a 10 cm gel tray. The region of the gel containing the unincorporated nucleotides was removed and the gel was placed on several sheets of Whatman 3MM filter paper, covered in plastic wrap, and dried under vacuum at 80°C in a Bio-Rad model 483 slab drier. The gels were exposed to Kodak XAR-5 film at -70°C for one hour. pBluescript digested with *Hpa*II and end-filled with radioactive dCTP was used as the DNA marker standard. Sixty ng of digested pBluescript was

incubated for 30 min at 37°C with 1X buffer, 1 unit of the Klenow fragment of DNA polymerase I (GIBCO/BRL), and 30 µCi of [α-^{32}P] dCTP at 3,000 Ci/mmol. Prior to loading, 2 ng of marker was mixed with an equal volume of loading dye (10X, 50% glycerol and 0.35% orange G).

Restriction Endonuclease Digestion:

All restriction enzymes were purchased from GIBCO-BRL. Restriction digestions of DNA were performed using 1-4 units of enzyme per 1 µg of DNA and the manufacturer's buffer system. One µg of DNA was routinely used for the digestion of plasmid and cosmid DNA, and 5-10 µg of DNA was used for human genomic, hybrid, or YAC digests. Sterile double-distilled water (ddH$_2$O) was used to bring the reactions to the desired volume. Incubations temperatures used were as recommended by the manufacturer. Digestions were allowed to proceed for one to two hours for plasmid and cosmid DNA, and for four hours for human genomic DNA, hybrid DNA, and YAC DNA with an addition of two units of enzyme per µg of DNA after two hours of incubation. A 10X loading dye containing 50% glycerol and 0.35% bromophenol blue was added to the reactions prior to electrophoresis.

Agarose Gel Electrophoresis of Digestion Products:

Electrophoresis of DNA digests was performed in 0.8% agarose, with 1X TBE as the running buffer. Ethidium bromide, at a concentration of 0.5 µg/mL, was added to the molten gel solution

prior to casting. For optimal resolution, digests of genomic, hybrid, YAC, and cosmid DNA were electrophoresed at 50-70 V for 17-19 hours on 25 cm gel trays. Plasmid DNA digests were electrophoresed more rapidly at 90-100 V for three to four hours or one to two hours on 25 cm and 10 cm gel trays, respectively. Bacteriophage λ DNA digested with *Hind*III or *Hind*III/*Eco*RI was used as a DNA marker standard for all digests. Following electrophoresis, EtBr stained DNA was visualized and photographed.

DNA Transfer:

DNA was transferred from the agarose gels onto Hybond™-N+ nylon membrane (Amersham) using a vacuum transfer system (Tyler Research Instruments). Prior to transfer, the DNA contained in the gel was depurinated using 0.25N HCl with gentle shaking for 15 min. DNA was denatured during the transfer by flooding the gel with 0.4N NaOH. Sixty cm of water vacuum was applied to the system for 90 min. Following the transfer, the membrane was neutralized in 2X SSC (300 mM NaCl, 30 mM sodium citrate, pH 7.0), blotted briefly between layers of Whatman 3MM paper and heat-sealed in a Seal-a-Meal bag (Dazey) ready for Southern hybridization.

Radiolabelling of DNA and Hybridization:

The random hexanucleotide primer method (Feinberg and Vogelstein, 1983) (Feinberg and Vogelstein, 1984) was used to label DNA for probing membrane bound DNA. One hundred and fifty ng of DNA was denatured by boiling for five min in a total volume of 30 µL

of sterile ddH_2O, quenched on ice, and labelled in a 50 µL reaction containing 13 µL OLB (195 mM Tris pH 8.0, 780 mM HEPES, 40 mM β-mercaptoethanol, 1.87 mg/mL BSA, 19.5 mM $MgCl_2$, 109 µM of each of dATP, dGTP, and dTTP, and approximately 716 µg/mL hexanucleotides), 50 µCi [α-^{32}P] dCTP at 3,000 Ci/mmol, and two units of the Klenow fragment of DNA polymerase I. The reaction was allowed to proceed for 90 min at room temperature, followed by separation of the labelled DNA from the unincorporated nucleotides using a G50 Sephadex (medium) spin column (Pharmacia). The labelled DNA was denatured for five min prior to hybridization.

Membranes to be probed were incubated at 55°C for 90 min in a prehybridization buffer containing 0.5M sodium phosphate:1mM EDTA, 1% w/v BSA, 5% w/v SDS, and 30% v/v formamide. Sheared and denatured herring sperm DNA was used as a non-specific blocker at a concentration of 270 µg/mL in the prehybridization buffer, and 100 µg/mL in the hybridization buffer. Dextran sulphate was added to the hybridization buffer just prior to hybridization at a concentration of 50 mg/mL. After prehybridization, the buffer was removed from the bag and replaced with hybridization buffer and probe (> 3×10^6 cpm/mL). Hybridization was carried out at 55°C for 16-20 hours.

Following hybridization, membranes were washed two times for 15 min each in 2X SSC/0.1% SDS at 55°C. Membranes were blotted dry, wrapped in plastic wrap, and exposed to Kodak XAR-5 diagnostic film with Dupont Quanta III intensifying screens at -70°C. Films

were developed in a Kodak M35A X-OMAT Processor or a RG II FUJI X Ray Film Processor.

To remove the hybridized probe, the membranes were agitated in 500 mL of a boiling 0.1X SSC, 0.1% SDS solution three times for at least five min per agitation. The membrane was then tested with a Berthold counter and/or exposed to film to ensure all traces of the probe had been removed.

DNA Quantitation:

Concentration and purity of all DNA samples was ascertained by measuring optical densities of samples at 260 nm and 280 nm with a Beckman DU®-65 spectrophotometer. A unit of absorbance at OD_{260} was taken to represent 50 µg/mL of double-stranded DNA, and 34 µg/mL of single stranded oligonucleotides. A DNA preparation was considered to be of good purity if it yielded an OD_{260}/OD_{280} ratio of approximately 1.8.

Rapid Small-Scale Plasmid Preparation:

Single colonies were picked with a sterile loop and inoculated into 5 mL of LB broth with the appropriate antibiotic at 50 µg/mL and grown overnight at 37°C with vigorous shaking. Bacterial cultures were pelleted in a bench top centrifuge at 1,300g for 15 min. The supernatant was discarded and the pellets were washed by resuspending in 500 µL of sterile ddH_2O. The cells were then transferred to microfuge tubes and spun in a microfuge for two min. The supernatant was again discarded, and the pellets resuspended in

300 µL STET (8% sucrose, 0.5% Triton X-100, 50 mM EDTA, and 10 mM Tris-HCl, pH 8.0). Twenty-five µL of a freshly prepared lysozyme solution (10 mg/mL in STET) was added with mixing, and the tubes were boiled for one minute, followed immediately by a 15 minute spin at full speed in a microfuge. The pellets were removed from the tubes with sterile toothpicks, and the supernatant was extracted once with phenol:chloroform:isoamyl alcohol (25:24:1), and once with chloroform:isoamyl alcohol (24:1). The DNA was precipitated by adding an equal volume of isopropanol, and pelleting in a microfuge for 15 min at 12,000g. The pellet was dissolved in 50 µL of low TE (10 mM Tris-HCl pH 8.0, 1 mM EDTA pH8.0) with 40 µg/mL DNAse-free RNAse A, and incubated at 42°C for 40 min. DNA was recovered by ethanol precipitation, and the plasmid pellet was dissolved in low TE (16 µL for use directly in dideoxy DNA sequencing, or 50 µL for restriction digestion and PCR analyses).

Isolation of DNA from Low Melting Point Agarose Gels:

When DNA fragments required gel purification, the DNA was electrophoresed in 1% low melting point agarose (GIBCO/BRL) gels and visualized by EtBr staining. After electrophoresis, the band was cut out of the gel and 1 µL of 50X Gelase buffer was added per 50 mg of gel (1X buffer consists of 40 mM Bis-Tris-pH6.0 and 40 mM NaCl). The gel was melted at 70°C for 20 min and then equilibrated at 45°C for ten min. After equilibration, half a unit of Gelase™ enzyme (Cedarlane Laboratories) was added per 200 mg of gel and digestion

proceeded for 2-3 hours. DNA was precipitated after digestion using room temperature EtOH and 4 M ammonium acetate.

DNA Sequencing using Sequenase®:

The DNA pellet from a rapid small-scale plasmid preparation was dissolved in 16 μL of low TE and denatured using the alkali method. The DNA was incubated for five min at room temperature with 4 μL 1N NaOH, and precipitated by the addition of 2 μL of 4 M ammonium acetate, and 50 μL of 95% ethanol. The mixture was left on ice for 15 min and spun in a microfuge at 12,000g for 15 min. The pellet was washed with 70% ethanol, air dried, and dissolved in 7 μL sterile ddH_2O. One pmole of primer was added to the plasmid DNA with 1 X reaction buffer (40 mM Tris-HCl pH 7.5, 20 mM $MgCl_2$, 50 mM NaCl), in a total volume of 10 μL. The primer was annealed to the template by incubation at 65°C for two min, and cooling slowly to 30°C over 30 min in a Perkin Elmer thermocycler 480. The labelling reaction consisted of the 10 μL of template with annealed primer, 1 μL of 0.1 M DTT, 2 μL of 1X labelling mix (1.5 μM of each dGTP, dCTP, and dTTP), 12.5 μCi of [α-^{35}S] dATP at 1322 Ci/mmol, and 2 μL of Sequenase® version 2.0 (Amersham) diluted 1:8 in dilution buffer (10 mM Tris-HCl pH 7.5, 5 mM DTT, 0.5 mg/mL BSA), and was mixed and incubated at room temperature for two to five min. Three and a half μL of the labelling mix was then added to each of four tubes marked G, A, T, and C, each containing 2.5 μL of the appropriate dideoxy termination mix (all termination mixes contained 80 μM of each of dGTP, dATP, dCTP, and dTTP, 50 mM NaCl, and 8 μM of the

appropriate dideoxynucleotide). The reactions were incubated at 37°C in a heating block for five min, and terminated by adding four µL of stop solution (95% formamide, 20 mM EDTA, and 0.05% bromophenol blue, 0.05% xylene cyanol). Samples were heated to 90°C for two min in a heating block and five µL per lane was loaded in an acrylamide gel.

Acrylamide Gel Electrophoresis:

All sequencing reactions and determination of allele sizes of polymorphic DNA markers were carried out on 31.0 cm x 38.5 cm denaturing polyacrylamide gels using a model S2 (GIBCO/BRL) apparatus. A 75 mL gel solution was prepared consisting of 8 M urea, 1X TBE, and 6% acrylamide (from a 40% acrylamide:bis acrylamide (38:2) stock solution). The urea was dissolved with stirring over low heat, and the mixture was filtered through a 0.22 micron Millipore filter attached to a 60 mL syringe. Immediately prior to casting the gel, a fresh 10% solution of ammonium persulfate (dissolved in water) was added to the filtered gel solution at a final concentration of 0.1% along with 0.032% v/v TEMED. The solution was mixed and cast between clean glass plates with 0.4 mm spacers, and either a 49 well sharkstooth comb (5.7 mm wells for sequencing reactions), or a 59 well comb (3 mm wells for analysis of alleles). The gels were run for an appropriate duration to obtain maximum resolution for the region of interest at a constant power of 70 W. The gel was transferred onto Whatman 3MM filter paper, covered with plastic wrap, and dried under vacuum at 80°C in a Bio-Rad model

483 slab drier. Dried gels were exposed to Kodak XAR-5 film at room temperature.

Sublocalization of New Polymorphic Markers:

Primer sequence and allele sizes of most of the polymorphic markers were obtained from GDB (Table 1) and three new STRs were designed from sequences obtained from Genbank (see Results, Tables 5 and 7) using Primer Designer (Version 1.01, Scientific and Educational Software). PCR analysis of the retention of STRs by hybrids was performed either with or without a radioactive label at least twice in a Perkin-Elmer Thermocycler 480 or a MJ Research DNA Engine, as described in the 'General Materials and Methods' section. Murine, CHO, and human DNA were used as internal controls. PCR products were separated and analyzed as described in the 'General Materials and Methods' section.

Analysis of Polymorphic Markers in CSNB1 Families:

Families:

DNA samples from CSNB1 families were obtained from a collaboration between Dr. N.T. Bech-Hansen and Dr. W.G. Pearce, Department of Ophthalmology, University of Alberta. All samples were obtained with informed consent by the originating laboratory. Polymorphic CA repeat markers were analyzed in CSNB1 Families

Table 1. Published Polymorphic CA Repeat Markers

Marker Designation	Sequence Forward Primer (5' to 3')	Sequence Reverse Primer (5' to 3')	Size (bp)	Source[a]
DXS988	ggctcttgaacagaaaacag	gctatcatagtggatgtgaagtagt	134-148	
DXS993	ggatcctgtttacagcctgt	ctacagagcagttcactggg	292-312	
DXS1003	ttcacccatagaagccgt	ccattcctcactggcaag	169-195	
DXS556	agtttgagggcttcgtttac	tatgaagacagccaacttaga	176-192	Thiselton *et al.*,1993
DXS573	gccacccaatctaaagttgc	ggtgatgatgagtgataagg	137-145	Roustan *et al.*, 1992
DXS983	cacactgcattaaatcctcg	caagttaccctacctgcctc	173-185	
DXS1039	ctcctgttcctggtatgtga	agaagaatgcctgttngggt	83-103	
DXS1126	ttctagaaaggtgcgtgtcgtctgg	gaccattccctctcaaacacaaacg	230-252	Donnelly *et al.*, 1994
DXS1158	tcttcaccagtcctaaaacttc	cgacaccctgtccctctaa	193-?	Barker and Fain, 1993
DXS1190	atcaccagacagaatcacc	ttttatccattcagccac	210-222	
DXS1194	cacacaacttgaaactgctgag	aagtatgttgccacagaaacc	261-283	
DXS1199	ggtgactgactctgtggc	tggagtgaaatcaacatttaacata	277-295	
DXS1204	atgaacccttaactcatttagcagg	agcntgcaccaacatgcc	237-251	
DXS1213	ccatagcccacctttc	tctgtgcatttatgtttatttg	230-244	
DXS1240	tccagactggcaacagaac	gatctagccaagggcaatt	168-172	Nemeth, personal com.
DXS6940	tcacatcctggacatacacc	tgagtatgtcacaggatgtg	166-170	Kwan *et al.*, 1996b
DXS8017	agcttaccactggggtga	ttcaggacacgctcttaatg	145-159	
DXS8023	cttttgtgcaaactgttcc	tctctgcctggtacacg	130-152	
DXS8024	tctgagagcaccaaccac	tgccagttttcttatggaatc	177-187	
DXS8062	atgtctacccacctgcttgc	gatccctggcctgctg	90-98	

a If not indicated the primer sequences are from GDB.

P060, P070, P130, P140, P170, P190, P200, P230, and P240. These families were chosen for analysis because they contained an adequate number of affected and unaffected individuals to determine phasing of alleles. Families P130, P170, P190, and P230 are two generation families with several affected individuals. Families P070, P140, P200, and P240 are multigeneration families with several affected individuals. Family P060 is a large multigeneration family from Alberta that contains over 140 individuals, at least 40 of whom are affected (originally reported by (Bech-Hansen *et al.*, 1990)). Three of these 40 affected individuals are manifesting sisters, as reported previously (Bech-Hansen and Pearce, 1993).

PCR Analysis:

Allele detection was performed by amplifying 25 ng of genomic DNA from an appropriate number of family members so that allele phasing could be determined in individuals carrying critical recombinant X chromosomes. PCR was performed in a 20 μL reaction containing 50 mM KCl, 10 mM Tris-HCl pH8.3, 1.5 mM $MgCl_2$ (DXS988, DXS1000, DXS1003, DXS1039, DXS1126, DXS6940, DXS8023, DXS8024), or 1.8mM $MgCl_2$ (DXS573, DXS722, DXS1240), 170 μg/mL BSA, 0.05% Tween-20, 0.05% Nonidet P-40, 0.25 μM of each primer, dNTP mix (20 μM of each dATP, dGTP, dCTP, and dTTP), 3-5 μCi [α-32P] dCTP at 3,000 Ci/mmol, and 0.5 U *Taq* DNA polymerase. Following an initial denaturation at 94°C for 7 min, amplification

conditions were as shown in Table 2, followed by a final extension step at 72°C for 7 min.

Table 2. Cycling Conditions for Polymorphic CA Repeat Markers

Primer	Denaturation	Annealing	Extension	Cycles
DXS573	94°C for 30 s	59°C for 30 s	72°C for 2 min	30
DXS722	94°C for 30 s	60°C for 30 s	72°C for 30 s	30
DXS988	94°C for 30 s	62°C for 30 s	72°C for 30 s	30
DXS1000	94°C for 30 s	62°C for 30 s	72°C for 10 s	30
DXS1003	94°C for 30 s	59°C for 30 s	72°C for 30 s	25
DXS1039	94°C for 30 s	63°C for 30 s	72°C for 30 s	30
DXS1126	94°C for 30 s	55°C for 2 min	72°C for 1 min	27
DXS1240	94°C for 30 s	60°C for 30 s	72°C for 30 s	30
DXS6940	94°C for 1 min	55°C for 2 min	72°C for 1 min	27
DXS8023	94°C for 1 min	55°C for 2 min	72°C for 1 min	27
DXS8024	94°C for 1 min	55°C for 2 min	72°C for 1 min	27

Two μL aliquots of the PCR reactions were added to two μL of denaturing buffer (95% v/v formamide, 10 mM EDTA pH 8.0, 0.1% bromophenol blue, 0.1% w/v xylene cyanol), and heated at 90°C for 2 min. The samples were electrophoresed on 6% polyacrylamide sequencing gels and allele sizes were determined by comparison to single stranded M13 sequencing ladder labelled with (α-^{35}S) dATP

(Amersham). DXS1003 was Klenow treated (0.5 units of Klenow to 5 μL of the reaction for 20 min at 37°C) prior to denaturation and electrophoresis to reduce single base pair slippage products (Weber, 1990), a procedure that was subsequently found to be unnecessary.

Isolation of Large Insert DNA Clones:

DNA Markers:

The DNA markers from Xp11.23-p11.22 used in these studies are detailed in Tables 3, 8, and 9. Using either PCR or Southern hybridization data, all markers used in contig construction were mapped to an interval delineated by the hybrids mapping panel (see Results, Figure 7).

Identification of YACs:

YAC libraries were screened at the Center for Genetics in Medicine, Washington University School of Medicine, in St. Louis, through a collaboration with Dr. David Schlessinger. YAC libraries E, F, and I, which together contain more than 15 X-chromosome equivalents (Nagaraja *et al.*, 1994), were screened using all STSs, STRs, and ESTs from the Xp11.23-p11.22 region. Library E was derived from a 49,XXXXX cell line, library F from a Xpter-Xq27.3 hamster-human somatic hybrid cell line, and library I, derived from a 48,XXXX cell line, is the ICI YAC library (Anand *et al.*, 1990). Two other libraries, A and M, were screened with a limited set of selected DNA markers from this region. Library A was made from a human

Table 3. DNA Markers Used in Preliminary Construction of Contig in Xp11.23-p11.22

DNA Marker	Description	Reference
DXS226	anonymous probe p2bC6	Paulsen *et al.* (1986)
DXS576	STS	S. Lindsay, personal com.
DXS1331	STS	Kere *et al.* (1992)
DXS1358	STS	B. Moore, personal com.
DXS8366	STS	Boycott *et al.* (1996)
HRASP	STS	Kere *et al.* (1992)
DXS255	M27b, VNTR	Fraser *et al.* (1989)
	STS	A. Zinn, personal com.
OATL1	2 kb Eco RI fragment of HuOAT6 cDNA , RFLP	Mitchell *et al.* (1988)
DXS573	STR	Roustan *et al.* (1992)
DXS1039	STR	Gyapay *et al.* (1994)
DXS1126	STR	Donnelly *et al.* (1994)
DXS1240	STR	A. Nemeth, personal com.
DXS1470	STR	B. Moore, personal com.
DXS6940	STR	Kwan et al., 1995
DXS1007E	EST00367	Polymeropoulos *et al.* (1993)
DXS1011E	EST01029	Polymeropoulos *et al.* (1993)
DXS7469E	EST, Xp664	M. Wehnert, personal com.
GATA1	STS for GATA 1 gene	Kere *et al.* (1992)
SYP	0.8 kb Pst I/Bgl II fragment from 3' end of SYP (SYP313)	Ozcelik *et al.* (1990)
TFE3	2.5 kb cDNA probe	Beckmann *et al.* (1990)
	STS from 3' end of gene	J. Derry, personal com.
WASP	exon 2 STS	Derry *et al.* (1994)
ZNF21	STR[a], DXS6849	N. T. Bech-Hansen, personal com.

[a] Based on CA dinucleotide repeat lying in *BstNI* fragment of cosmid ICRFc104E04126 (Kwan *et al.*, 1995) identified with ZNF21 probe (KOX14) (Huebner et al., 1991).

lymphoblast cell line (46,XY) (F. Abidi, R. Mazzarella, J.-Y. Yoon, and D. Schlessinger, unpublished data), and library M refers to the Mega CEPH library (Cohen *et al.*, 1993). Screening was carried out by PCR, with robotic assistance, using a uniform temperature regimen and the TNK buffer system, as described previously in the 'General Materials and Methods' section (Blanchard *et al.*, 1991) (Blanchard *et al.*, 1993). Isolated YACs were sent from St. Louis as stabs in YAC *trp*$^-$, *ura*$^-$ selective agar ((pH 7.0), 0.7% w/v yeast nitrogen base without amino acids, w/v 2% glucose, w/v 0.005% adenine sulphate, w/v 0.005% tyrosine, w/v 1.4% vitamin assay casamino acids, and 1.5% w/v agar).

Initial Characterization of YACs:

YACs were streaked on YAC selective agar plates (*trp*$^-$, *ura*$^-$) and incubated at 30°C for 48 hours to isolate single colonies. A multi-colony sample and five clonal isolates were inoculated in 5 mL of YAC selective media (YAC selective agar without agar) and grown with shaking at 30°C for 48 hours. Fifty per cent glycerol stocks were prepared with 600 µL of each culture and stored at -70°C. The remaining 4.4 mL of culture was used to prepare intact yeast chromosomal DNA in agarose plugs. Each culture made four plugs.

To prepare plugs, the YAC culture was pelleted at 1,300g for 15 min. The supernatant was poured off and the pellet resuspended in 100 mM EDTA (pH 8.0). This suspension was pelleted again, as above, and the supernatant poured off. The pellets were then resuspended in 120 µL CPES (pH 5.6, 23 mM Na_2HPO_4, 8.4 mM citric

acid, 50 mM EDTA-pH 8.0, 0.9 M sorbitol) with 5 mM DTT. Two hundred µL of a solution containing 1:1 8 mg/mL yeast lysing enzymes (Sigma) in CPE buffer (23 mM Na_2HPO_4, 8.4 mM citric acid, 50 mM EDTA-pH 8.0): 4.0% low melting point agarose in CPE buffer was then added to the resuspended pellet, mixed, and 80 µL was dispensed into each of four chilled blocks. The formed block were allowed to solidify at -20°C for 10 min and were then pushed out into a tube containing 5 mL of CPE buffer and incubated at 37°C for one hour with occasional inverting. After the incubation, the CPE was poured off and replaced with 2 mL of Solution A (0.5 M EDTA-pH8.0, 1% sodium lauroyl sarkosine and 0.5 mg/mL of proteinase K added prior to use). The blocks were incubated with Solution A at 50°C for 48 hours. After incubation, one plug from each sample was loaded directly into a gel for pulsed-field gel electrophoresis and Solution A was removed from the remaining plugs and replaced with 0.5 M EDTA. Unused plugs were stored at 4°C.

The integrity and estimation of YAC size performed using pulsed-field gel electrophoresis in 1% agarose gels, 0.5X TBE (44.5 mM Tris, 44,5 mM boric acid, 1 mM EDTA-pH 8.0), on a CHEF DRIII (BioRad) apparatus. Plugs were loaded into the wells, covered with 1% low melting point agarose in 0.5X TBE, and the gel was loaded into the chamber. Pulse times were ramped from 3 sec to 30 sec over 28 hours (6V/cm, 120° angle) at 14°C. Size estimation of the YACs is based on co-migration of *Saccharomyces cerevisiae* marker chromosomes (strain YP149, 17 chromosomes from 90 to 1,600 kb).

Gels were stained after electrophoresis in EtBr (0.5 μg/mL) for 30 min and visualized and photographed.

The yeast chromosomal DNA was transferred from the agarose gels onto Hybond™-N+ nylon membrane using the vacuum transfer system as described in the 'General Materials and Methods' section. To detect the YACs, filters were hybridized with human genomic DNA probes. In most cases yeast DNA was prepared from a single (largest) clonal isolate of a particular YAC for further characterization.

Five mL of YAC selective media was inoculated from the glycerol stock of the clonal isolate of choice for a particular YAC and grown with shaking at 30°C for 48 hours. The cells were pelleted at 1,300g for 15 min, resuspended in 100 mM EDTA (pH 8.0), and pelleted again. The pellets were then resuspended in 1 mL CPES with 20 μg/mL yeast lysing enzymes and incubated at 37°C for one hour with occasional inverting. The cells were gently pelleted at 900g for 10 min and the supernatant poured off. The pellet was resuspended in 1 mL of DNA extraction buffer (10 mM Tris-HCl-pH 8.0, 0.1 M EDTA-pH 8.0, 0.5% SDS, and 20 μg/mL RNAse A) and incubated at 37°C for one hour. After one hour, 100 μg/mL of proteinase K was added and incubation temperature was increased to 50°C for at least 3 hours. The DNA was extracted three times with phenol:chloroform:isoamyl alcohol (25:24:1), and once with chloroform:isoamyl alcohol (24:1), and then ethanol precipitated with 10% volume of 4 M NH_4OAc and 2 volumes of 95% ethanol in a 40 minute spin at 10,000g. The yeast DNA pellet was rinsed with 70% ethanol, allowed to dry, resuspended

in 150 µL low TE, and quantitated using a spectrophotometer. This yeast DNA preparation typically yielded 75-100 µg.

Identification of Cosmids:

Two X-chromosome specific cosmid libraries were used: LL0XNC01 (Biomedical Sciences Division, Lawrence Livermore National Laboratory, Livermore, CA 94550 under the auspices of the National Laboratory Gene Library Project sponsored by the U.S. Department of Energy), and the ICRF (Nizetic *et al.*, 1991). The ICRF cosmid library was constructed from flow-sorted human X chromosomes and was arrayed on two filters, each containing 9,216 cosmids, representing four chromosome equivalents. The Lawrence Livermore cosmid library was constructed from a flow-sorted human X chromosome/hamster hybrid and 24,000 cosmids were gridded into 250 96-well plates, representing six chromosome equivalents. For purposes of hybridization, the 250 96-well plates were regridded into 63 384-well plates by Dr. N.T. Bech-Hansen at the Lawrence Livermore laboratories, and later the clones from these 63 plates were spotted onto 16 filters (4 x 384 on 15 filters and 3 x 384 on one filter).

Cosmids were identified from LL0XNC01, designated library U, by screening with selected STSs by PCR at Washington University School of Medicine, or by hybridization of the gridded filters with selected probes. The ICRF gridded X chromosome-specific cosmid library was also screened, in duplicate, with selected probes by filter hybridization. For screenings, both filter sets were pre-hybridized at

65°C for 4-6 hours in a buffer of 5X Denhardt's reagent (0.1% Ficoll-type 400, 0.1% polyvinylpyrrolidone, 0.1% BSA), 6X SSPE (0.3 M NaCl, 60 mM sodium phosphate, 6mM EDTA-pH 7.4), 0.5% SDS, and 50 µg/mL denatured, sheared herring sperm DNA. The probes were labelled as described in the 'General Materials and Methods' section. The filters were hybridized with probe and fresh buffer (as per pre-hybridization buffer without herring sperm DNA) in sealed bags at 65°C for 16-20 hours. A visible vector background was accomplished by labelling Lawrist 4 (ICRF cosmid filters) and Lawrist 16 (Lawrence Livermore cosmid filters) with [α-^{35}S] dATP. Following hybridization the filters were washed at 65°C, as describe previously, and exposed to Kodak XAR-5 diagnostic film overnight at -70°C. Positive cosmid clones identified from the ICRF cosmid library were kindly sent to us by Gunther Zehetner at ICRF as bacterial stabs in agar. Positive cosmid clones identified from the Lawrence Livermore cosmid library, by hybridization, were taken directly from the laboratory set of 384-well plates.

Initial Characterization of Cosmids:

Cosmids identified from both libraries were initially streaked on LB agar plates with kanamycin (50 µg/mL) and grown overnight at 37°C. A multi-colony sample and two single colonies were inoculated and grown overnight in LB media with kanamycin (50 µg/mL). Twenty per cent glycerol stocks were made using 850 µL of culture and were stored at -70°C . DNA was isolated from the remaining culture by the rapid small-scale plasmid preparation as described in

the 'General Materials and Methods' section. Cosmid DNA was then cut with EcoRI and electrophoresed, as described in the 'General Materials and Methods' section. The size of each cosmid insert was calculated by summing the sizes of all restriction fragments seen with EtBr staining, excluding 5 kb for the vector.

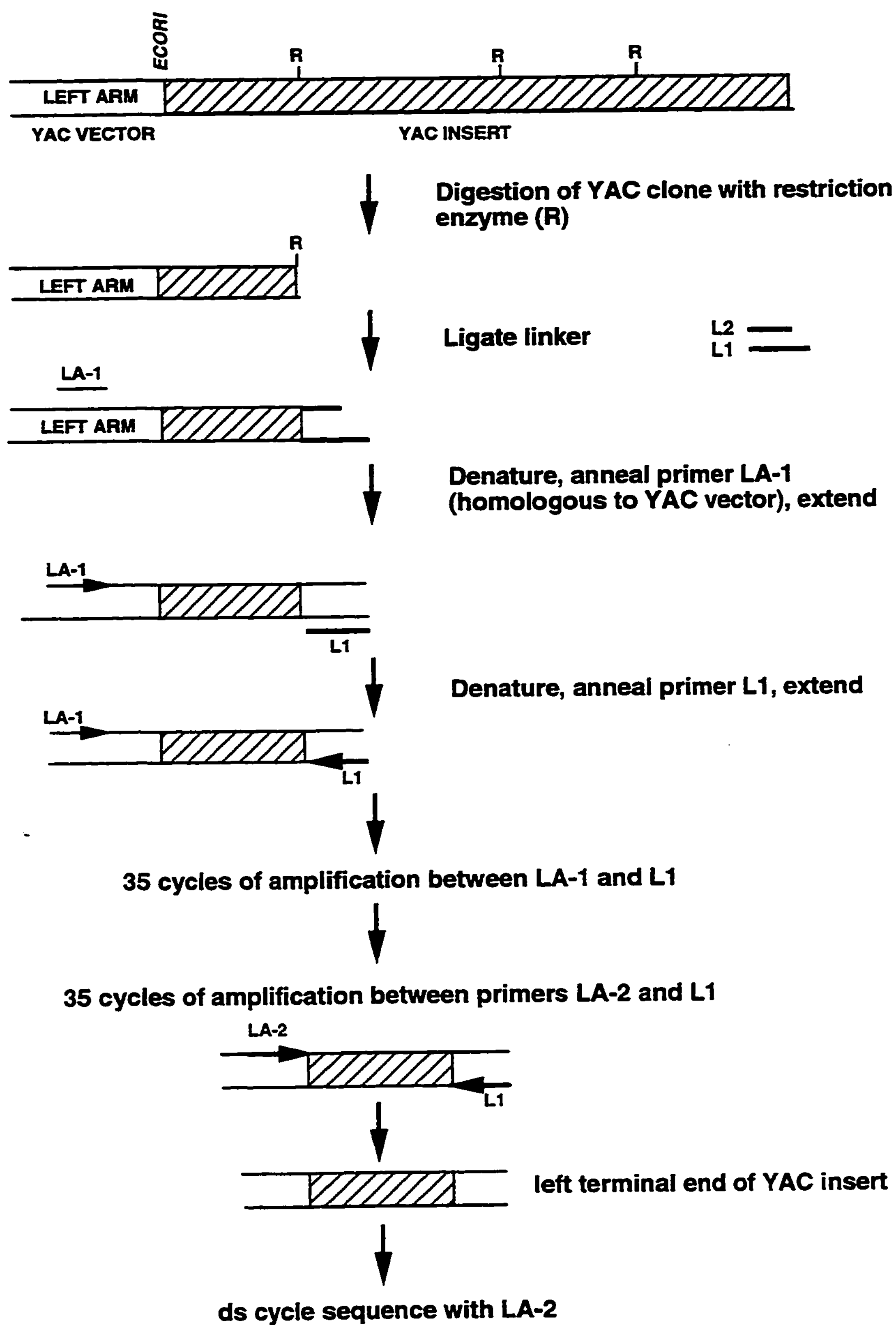
Generation of New STSs:

New STSs were characterized and mapped for contig construction.

End-clone STSs: Generation of ligation-mediated PCR products (Kere *et al.*, 1992) (Muller *et al.*, 1994) was performed as depicted in Figure 3. One hundred ng of yeast DNA containing the YAC of interest in pYAC 4 vector was digested with *Rsa*I, *Alu*I, *Pvu*II, *Eco*RV, or *Sca*I in 13 µL containing the appropriate 10X buffer and five units of the enzyme, for 2 hours at 37°C. The linker was annealed in a 100 µL-reaction containing 1 pmol/µL L1 and 1 pmol/µL L2 (Table 4) in a Perkin Elmer Thermocycler 480 by denaturing at 95°C for two min and allowing L1 and L2 to anneal by ramping to 21°C over 30 min. Annealed linker was stored at -20°C and thawed on ice only.

A blunt end ligation was carried out with 1X ligation buffer (50 mM Tris-HCl pH 7.6, 10 mM $MgCl_2$, 1 mM ATP, 1 mM DTT, 5% w/v polyethylene glycol-8000), 2 pmol of annealed linker, and 20 units of T4 DNA ligase (New England Biolabs), which were added directly to the digestion reactions and incubated at 16°C for 16 hours. The reactions were diluted with 60 µL of ddH_2O and denatured at 95°C for 10 min. Two µL of the ligated mixtures were amplified by PCR in

Figure 3. Generation of a YAC end-clone by ligation-mediated PCR. The figure depicts the generation of an end-clone from the left end of the YAC. The YAC clone is digested with restriction enzyme R and the products are ligated to the linker, made up of L1 and L2. The ligated products are then subjected to rounds of PCR with the LA-1 (vector primer left) and L1. The products of this PCR are then reamplified using an internal primer, LA-2 (vector primer left internal) and L1. The left terminal end of the YAC insert is then isolated and sequenced. The restriction enzymes (R) used in the generation of end-clones from pYAC4 have been described in the text, and L1, L2, LA-1 and LA-2 are described in Table 4.



ECORI
R
R
R
LEFT ARM
YAC VECTOR
YAC INSERT
Digestion of YAC clone with restriction enzyme (R)
R
LEFT ARM
Ligate linker
L2
L1
LA-1
LEFT ARM
Denature, anneal primer LA-1 (homologous to YAC vector), extend
LA-1
L1
Denature, anneal primer L1, extend
LA-1
L1
35 cycles of amplification between LA-1 and L1
35 cycles of amplification between primers LA-2 and L1
LA-2
L1
left terminal end of YAC insert
ds cycle sequence with LA-2

10 µL reactions overlaid with paraffin oil containing 0.05 µM L1 and 0.5 µM YAC vector arm-specific primer (Table 4), TNK50 buffer, dNTP mix (125 µM dATP, dCTP, dGTP, and dTTP), and 0.25 units of *Taq* DNA polymerase. Temperature cycling conditions were 94°C for 30 s, 65°C for 45 s, and 72°C for 1min, 30 s for 35 cycles. After cycling, the PCR reaction was diluted with 150 µL of sterile ddH_2O and 4 µL of the diluted reaction was reamplified using L1 and an internal YAC vector arm-specific primer (0.25 µM each) (Table 4) in a 20 µL reaction overlaid with paraffin oil under identical conditions. The reamplification products were purified by isopropanol precipitation and a portion was checked on a 2% Nusieve®GTG®agarose/1% agarose gel, in 1X TBE running buffer. Those with single products greater than 200 bp were sequenced.

Table 4. Primer Sequences used for End-clone Generation in pYAC 4 Clones

Primer Name	Primer Sequence 5'-3'
Vector Primer Left	CACCCGTTCTCGGAGCACTGTCCGACCGC
Vector Primer Left Internal	TCTCGGTAGCCAAGTTGGTTTAAGG
Vector Primer Right	ATATAGGCGCCAGCAACCGCACCTGTGGCG
Vector Primer Right Internal	TCGAACGCCCGATCTCAAGATTAC
Linker Primer L1	GCGGTGACCCGGGAGATCTGAATTC
Linker Primer L2	GAATTCAGATC

The dsDNA Cycle Sequencing System (GIBCO/BRL) was used to determine the DNA sequence of the ligation-mediated PCR products. The internal left or right vector primer was end-labelled with 10 μCi [γ-^{33}P] of dATP at 3,000 Ci/mmol (Amersham) in a 5 μL reaction containing 1 pmol of primer, 1X kinase buffer (70 mM Tris-HCl pH 7.6, 10mM $MgCl_2$, 5 mM dithiothreitol), and one unit of T4 polynucleotide kinase, for sequencing of the left and right end of a YAC, respectively.

The reaction was incubated at 37°C for 30 min and then terminated by incubation at 95°C for five min. Template DNA (50 fmol) in 26 μL of sterile ddH_2O, 4.5 μL of 10X *Taq* sequencing buffer (300 mM Tris-HCl pH 9.0, 50 mM $MgCl_2$, 300 mM KCl, 0.5% w/v W-1), and 0.5 μL *Taq* DNA polymerase (5 units/μL) were added to the labelled primer (total volume of 13.5 μL). Eight μl of this mix was then added to each of four tubes labelled G, A, T, and C, each containing 2 μL of the appropriate dideoxy termination mix (all termination mixes contained 50 μM each dATP, dCTP, dTTP, and 7-deaza-dGTP, as well as: Mix G-0.2 mM ddGTP; Mix A-2 mM ddATP; Mix T-2 mM ddTTP; Mix C-1 mM ddCTP). Reaction mixtures were overlaid with paraffin oil and cycled at 95°C for 30 s, 55°C for 30 s, 72°C for 60 s for 20 cycles, followed by 95°C for 30 s and 72°C for 60 s for 10 cycles. Reactions were terminated by the addition of 5 μL of stop solution (95% v/v formamide, 10 mM EDTA pH 8.0, 0.1% bromophenol blue, 0.1% w/v xylene cyanol). DNA was denatured by heating the reaction mixtures to 90°C for five min in a heating block and 6 μL was loaded into each lane of a 6% polyacrylamide gel.

Sequences were analyzed by two computer programs: BLASTn (Altschul *et al.*, 1990) was used to detect sequences homologous to common repetitive elements, and Primer Designer (Version 1.01, Scientific and Educational Software) was used to select primer pairs for STS PCR reactions. Primer pairs were verified to be region-specific STSs and assigned to a mapping interval by testing them on a set of different DNA templates, including the testing panel from the Center for Genetics in Medicine (Nagaraja *et al.*, 1994) and the panel of conventional and radiation hybrids, as described in the 'General Materials and Methods' section.

Design of ESTs from Existing Sequence: ESTs were designed from sequences obtained from Genbank. These sequences were also analyzed by the two computer programs described previously and the primer pairs were confirmed to be region-specific STSs and assigned to a mapping interval.

Identification of Overlaps:

Overlaps were characterized using STS content mapping and hybridization data. STS content mapping was carried out using PCR, as described in the 'General Materials and Methods' section, and the TNK buffer conditions optimized for library screening. Five of the markers (DXS226, DXS255, OATL1 (huOAT6), TFE3, and SYP) were used as probes in Southern hybridization analyzes of YAC clones digested with *Eco*RI, as described in the 'General Materials and Methods' section. YAC DNA was tested for each of the 44 DNA

markers from the p11.23-p11.22 region of the X chromosome. Cosmid DNA was only tested for DNA markers in the vicinity of the marker used for isolation of the clone.

Isolation of Transcripts by Direct cDNA Selection:

Isolation of YAC Templates:

YACs were purified by pulsed-field gel electrophoresis and isolated by Prep-a-Gene (BioRad) for use as templates in direct cDNA selection. At least 50 plugs for YACs chosen for use in direct cDNA selection (see Results, Figures 19 and 20) were made as described previously. These plugs were loaded into 1% rapid agarose gels (GIBCO/BRL) using a preparative comb and run for 12 hours ramping from 1-6 s for YACs less than 50 kb, 16 hours ramping from 2-12 s for YACs between 50 and 125 kb, 34 hours ramping from 10-60 s for YACs greater than 300 kb, 28 hours ramping from 15-75 s for YACs greater than 600 kb, and 28 hours ramping from 60-120 s for YACs greater than 1 Mb, to optimize resolution.

The band containing the YAC was excised from the gel after staining with EtBr, cut into 2 mm pieces, and transferred to a 50 mL Falcon tube. The volume of the gel was estimated and three times volume of binding buffer (6 M sodium perchlorate, 50 mM Tris-HCl pH8.0, 10 mM EDTA) was added. The gel was dissolved with swirling at 55°C for 15 min and 30 µL of matrix (assuming >5 µg of DNA in gel slice) was added and the tube was rotated for 15 min. The tube was then centrifuged for 30 s and the supernatant removed. The pellet

containing bound DNA was resuspended in 750 µL binding buffer and transferred to a 1.5 mL centrifuge tube which was centrifuged for 30 s and the supernatant removed. The matrix was then washed with 750 µL of wash buffer (20 mM Tris-HCl pH 7.5, 2 mM EDTA, and 50% ethanol) twice and the bound DNA eluted by resuspending the pellet in 60 µL of elution buffer (10 mM Tris-HCl pH 8.0, 1mM EDTA) and incubating the tube at 37°C for 15 min. After incubation the matrix was pelleted by centrifugation, the supernatant containing the DNA was transferred to a new tube, and the DNA was quantitated as described in the 'General Materials and Methods' section. Since the yield of the YAC DNA using this method was at the limit of detection using the available spectrophotometer, the presence of YAC DNA was confirmed by using radioactive PCR (as described in the 'General Materials and Methods' section) to test an aliquot of the supernatant for retention of a YAC-specific marker.

Isolation of Cosmid Templates:

Cosmids chosen as templates for direct cDNA selection (see Results, Figure 20) were purified by the rapid small-scale plasmid preparation as described in the 'General Materials and Methods' section.

cDNA Sets:

Placental cDNAs were kindly provided by Dr. Michael Lovett (University of Texas Southwestern, Dallas) and prepared by priming a reverse transcriptase reaction with random hexamers on RNA from

human placenta and blunt end ligating a PCR amplification cassette OLIGO1, as described (Lovett, 1994a). The retinal, frontal cortex, and fetal brain cDNA sets were kindly provided by Dr. Johanna Rommens (The Hospital for Sick Children, Toronto) and prepared by random priming with an extended 5' end oligonucleotide, RXGA'B'C'N6, to generate random mixtures from RNA that did not require the ligation of oligonucleotide linkers at any step, as described (Rommens *et al.*, 1994). These cDNA sets will be referred to collectively as, JRL. Both the OLIGO1, 5'CTGAGCGGAATTCGTGAGACC3', and the RXG, 5'CGGAATTCTCGAGATCT3', primers contain *Eco*RI restriction sites for subsequent subcloning of selected cDNA fragments.

Two μg of placental cDNA was amplified by PCR using 0.4 μM OLIGO1, 1 μL of the cDNA set, 1.5 mM $MgCl_2$, 50 mM KCl, 10 mM Tris-HCl pH 7.3, 170 μg/mL BSA, 0.05% Tween 20, 0.05% NP-40, 250 μM dATP, dCTP, dGTP, and dTTP, and 0.75 units of *Taq* DNA polymerase in a 25 μL reaction. The reactions were overlaid with paraffin oil and cycled for 30 cycles of 94°C for one minute, 59°C for one minute, and 72°C for two min in a Perkin Elmer Thermocycler 480. The PCR reactions were pooled and precipitated using 4 M ammonium acetate and 95% ethanol.

Two μg of the retinal, frontal cortex, and fetal brain (JRL) cDNAs were amplified by PCR using 2.0 μM RXG, 1 μL of the cDNA set, 1.5 mM $MgCl_2$, 50 mM KCl, 10 mM Tris-HCl pH 7.3, 170 μg/mL BSA, 0.05% Tween 20, 0.05% NP-40, 250 μM dATP, dCTP, dGTP, and dTTP, and five units of *Taq* DNA polymerase in a 25 μL reaction. The reactions were overlaid with paraffin oil and underwent 30 cycles of

94°C for one minute, 55°C for one minute, and 72°C for two min in a Perkin Elmer Thermocycler 480. Typically 5-15 PCR reactions for each cDNA set were pooled and precipitated using 4 M ammonium acetate and 95% ethanol. The retinal, frontal cortex and fetal brain cDNAs were pooled in a 1:1:1 ratio after precipitation and prior to direct cDNA selection.

Direct cDNA Selection:

Blocking of cDNAs: Suppression of repeat sequences in the cDNA pools was carried out as described (Lovett, 1994a). Two µg of each of the placental cDNA set and the JRL cDNA set was resuspended in five µL of sterile ddH_2O. Two µg of COT-1 (GIBCO/BRL), one µg of yeast DNA, and one µg of pWE vector were added to the cDNA pools in a total volume of 10 µL, overlaid with paraffin oil, and denatured at 95°C for five min. After denaturation, ten µL of 1 X hybridization buffer (0.75 M NaCl, 20 mM sodium phosphate pH 7.2, 5 mM EDTA, 5 X Denhardt's solution (0.1% Ficoll-type 400, 0.1% polyvinylpyrrolidone, and 0.1% BSA), and 0.1% SDS) was added under the oil giving a total volume of 20 µL of DNA (DNA concentration 200 ng/µL). The mixture was then incubated at 65°C for four hours with gentle shaking. These hybridization conditions corresponded to a $COT_{1/2}$ of 10 moles nucleotide litre^{-1}•s.

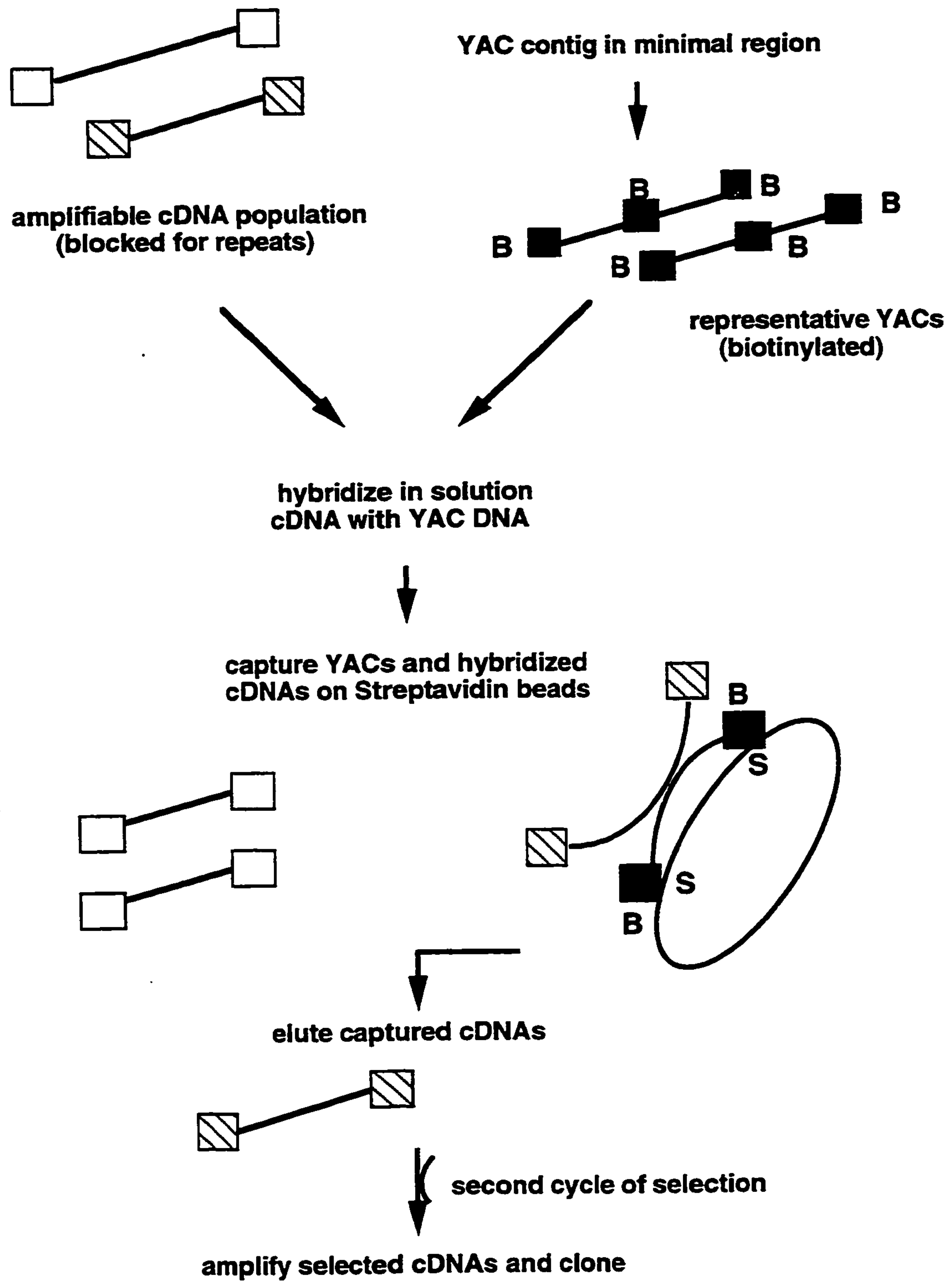
Nick Translation of YAC and Cosmid Templates: YAC or cosmid DNA was labelled by nick-translation with biotin-16-dUTP (Boehringer Mannheim). Two hundred ng of templates for a

particular selection were mixed in a molar ratio, precipitated, and dissolved in 6 µL of sterile ddH_2O. To this mix, 1X nick translation buffer, 2 µL of enzyme mix (DNA polymerase I and DNase I), 0.1 mM of each dNTP (dTTP and biotin-16-dUTP mixed in a 2:1 ratio), and 0.5 µL of [α-^{32}P] dCTP as a tracer were added to a final volume of 20 µL. The reaction was incubated at 15°C for 90 min and stopped by heating at 65°C for ten min. Unincorporated dNTPs were removed by passage through a Sephadex G-50 (medium) spin column. To confirm that biotin had indeed been incorporated, the percent incorporation of biotin relative to the percent incorporation of radioactive label was determined. Twenty percent of the elutant volume was counted, added to 50 µL of streptavidin beads (Dynabeads®, DYNAL®), and mixed for 15 min. The beads were washed once with 50 µL of binding buffer (10 mM Tris-HCl pH 7.5, 1 mM EDTA, 1 M NaCl), before being removed from solution with a magnetic separator. The wash was counted and the percent of biotin incorporation was determined from the ratio of total counts-wash counts/total counts. To continue with the direct cDNA selection, 75% of the radioactive label needed to have been bound to the beads, indicating efficient biotin incorporation. The remaining elutant was then precipitated with 4 M ammonium acetate and 95% ethanol and resuspended in sterile ddH_2O at a concentration of 20 ng/µL. One hundred nanograms of biotinylated template was overlaid with paraffin oil and denatured by heating at 95°C for five min.

Primary Selection: Selections were carried out as described (Lovett, 1994a), and as depicted in Figure 4. One hundred ng of denatured template was added to the prehybridized cDNA in 1 X hybridization solution. The hybridizations were performed at 65°C in a total volume of 20 μL for 58 hours (COT1/2 of >100 moles nucleotide litre^{-1}•s). YAC DNA was captured on 100 μL of streptavidin-coated paramagnetic beads (Dynal) by incubation at room temperature for 15 min with occasional shaking. The beads were removed from solution and resuspended in 1 mL of wash solution 1 (1 X SSC, 0.1% SDS). The beads were washed twice in wash solution 1 for 15 min each at room temperature and then washed three times at 65°C with wash solution 2 (0.1 X SSC, 0.1% SDS) for 15 min each. The hybridized cDNAs were eluted with 200 mM NaOH for ten min at room temperature followed by neutralization with 100 μL of 1 M Tris-HCl pH 7.5. This material was run through a Sephadex G-50 (medium) column and the cDNAs collected.

To evaluate the primary selection, the unselected and primary selected cDNAs inserts were amplified by PCR and electrophoresed on a 1% agarose gel, transferred, and hybridized with the respective positive reporter genes for each selection. PCR amplified GATA1 (Caiulo *et al.*, 1991) (Kere *et al.*, 1992), contained on yWXD2686, was used to monitor the efficiency of the placental distal selection, and PCR amplified SYP exon 5, contained on yWXD1908 was used to monitor the efficiency of the retinal, frontal cortex, and fetal brain selections. Presence of the positive reporter gene in only the primary selected material was evidence of enrichment. PCR was

Figure 4. Isolation of region-specific cDNAs by direct cDNA selection. Briefly, the amplified cDNA population is blocked for repeats and hybridized in solution with biotinylated YAC DNA. The YAC DNA and hybridized cDNAs are then captured on Streptavidin beads and after several washing steps (not shown), the cDNAs are eluted. The selected cDNAs are amplified and subjected to a second round of enrichment, and the resulting cDNAs from the secondary selection are amplified and cloned for further analysis. The steps have been described in greater detail in the text.


YAC contig in minimal region
B
B
B
B
B
B
amplifiable cDNA population
(blocked for repeats)
representative YACs
(biotinylated)
hybridize in solution
cDNA with YAC DNA
capture YACs and hybridized
cDNAs on Streptavidin beads
B
S
S
B
elute captured cDNAs
second cycle of selection
amplify selected cDNAs and clone

performed to yield 2 μg of primary selected cDNA, as described in the 'cDNA Set' section.

Secondary Selection: The selection procedure was repeated with an additional 100 ng of biotinylated YAC DNA and the primary selected cDNA as described for the primary selection. The efficiency of the selection procedure was monitored as for the primary selection except cDNAs from each of the starting, primary, and secondary stages of the selection, were electrophoresed, transferred, and hybridized with the positive reporter genes. Secondary selected material from the retinal, frontal cortex, and fetal brain selection was amplified by PCR, as described previously, for cloning and the secondary selected material from the placental selection was amplified, as described previously, using the modified OLIGO 1 primers described in the following section.

Cloning of Direct cDNA Selection Products:

Secondary selected placental cDNA material was cloned using CloneAmp™ pUC19 System (GIBCO/BRL) which uses Uracil DNA Glycosylase (UDG) to facilitate directional cloning of PCR products. This method relies on the incorporation of dUMP residues in place of dTMP into the 5' end of each amplification primer. After amplification, treatment with UDG renders dUMP residues abasic and unable to base-pair resulting in 3' protruding termini able to efficiently anneal to pAMP19. To clone PCR products using this system, the OLIGO 1 amplification primers were redesigned to

include deoxy-UMP residues at the 5' end of each primer: OLIGO-U forward 5'AUGGAGAUCUCUCTGAGCGGAATTCGTGAGACC3'; OLIGO-U reverse 5'ACGCGUACUAGUCTGAGCGGAATTCGTGAGACC3'. The new OLIGO 1 primers were used to amplify the secondary selected placental cDNA material as described previously. Two µL of the PCR product (10-50 ng) was annealed to 22 ng of pAMP19 vector in a reaction containing 1X Annealing buffer (20 mM Tris-HCl pH 8.4, 50 mM KCl, 1.5 mM $MgCl_2$), and one unit of Uracil DNA Glycosylase in a total volume of 20 µL. The 63 bp control insert DNA provided with the pAMP 19 kit was annealed to the vector according to the manufacturer's instructions. The reactions were incubated at 37°C for 30 min and 5 µL was used for transformation into 100 µL of Subcloning Efficiency DH5α frozen competent cells. The mixture was incubated on ice for 30 min, heat-shocked for 45 s, and iced for two min. One mL of LB was added and the cells were incubated at 37°C for one hour with shaking for recovery. The transformation was plated in aliquots of 400 µL, 300 µL, 200 µL, and 100 µL on LB-Ampicillin plates (50 µg/mL) treated with three µmoles of IPTG and 800 µg of X-gal for blue/white screening. The vector pUC19 was used as a control to monitor transformation efficiency. Untransformed competent cells were streaked on plates with and without ampicillin to confirm host cell sensitivity to the antibiotic. Plates were left to dry at room temperature, inverted and incubated at 37°C overnight.

Blue/white screening is based on the host cell encoding only the carboxy-terminal region of β-galactosidase and the vector carrying

the regulatory sequences and the coding region for the amino-terminal fragment of β-galactosidase. Neither host- nor plasmid-encoded fragments are active by themselves, but by association they form an enzymatically active protein (α-complementation). In the presence of IPTG and the chromogenic substrate X-gal, Lac^+ bacteria resulting from α-complementation form blue colonies. The insertion of a fragment of foreign DNA into the polycloning site of the vector results in production of an amino-terminal fragment that is not capable of α-complementation and therefore bacteria containing recombinant plasmids form white colonies. White colonies were picked, transferred into 96-well plates containing LB, 7.5% glycerol and 50 µg/mL ampicillin, and grown overnight before storing at -70°C.

Secondary selected retinal, frontal cortex, and fetal brain cDNAs were cleaved with *Eco*RI, and ligated to *Eco*RI-digested pBluescript vector (Strategene). One hundred µg of pBluescript was digested with 200 units of *Eco*RI (GIBCO/BRL) in a total volume of 200 µL. The vector was dephosphorylated by the addition of one unit of calf intestinal phosphatase (CIP) at 37°C for 30 min. The phosphatase was deactivated by incubation at 75°C for ten min and removed from the digested vector by two extractions with phenol:chloroform:isoamyl alcohol (25:24:1), and one extraction with chloroform:isoamyl alcohol (24:1). DNA was ethanol precipitated and the pellet dissolved to a concentration of 200 ng/µL. Secondary selected material was amplified by PCR to produce 100 µL of PCR reaction product, and extracted and precipitated. The DNA was

dissolved in 23 µL of sterile ddH_2O and digested with 40 units of *Eco*RI with React 3 in a total volume of 30 µL. The products of the digestion were electrophoresed on a 2% low melting point agarose gel and cDNAs between 300 bp and 900 bp were isolated using Gelase and EtOH precipitation. Two hundred ng of prepared vector was mixed with a three-fold molar excess of cDNAs, ethanol precipitated, and redissolved in seven µL of low TE. The insert was then ligated to the vector with ligation buffer (30 mM Tris-HCl pH 7.6, 10 mM $MgCl_2$, 1 mM ATP, 1 mM DTT, 5% w/v polyethylene glycol-8000) and one unit of T4 DNA ligase (GIBCO/BRL) in a total reaction volume of ten µL and incubated overnight at 12°C. Linearized and dephosphorylated vector DNA without insert were used as ligation controls and ten ng of each ligation and ligation control were transformed into 100 µL of Subcloning Efficiency DH5α frozen competent cells. The vectors pBluescript and pUC19 were used as controls to monitor transformation efficiency. White colonies were picked and transferred to 96-well plates for storage at -70°C.

Analysis of Selected cDNA Clones:

Recombinant plasmids were picked and DNA isolated by the rapid small-scale plasmid preparation as described in the 'General Materials and Methods' section. Inserts were obtained by cutting the plasmids with *Not*I and *Bam*HI for the placental clones or *Eco*RI for the retinal, frontal cortex and fetal brain clones, electrophoresing the products on low melting point agarose gels (GIBCO/BRL), and isolating the insert DNA using Gelase, as described in the 'General

Materials and Methods' section. Individual inserts were radiolabelled with [α-^{32}P] dCTP, as described in the 'General Materials and Methods' section, and hybridized to Southern blots containing DNA from human, as well as DNA from the hybrids GM06318B (X only), H99 (Xp11), Chinese hamster ovary, and YACs or cosmids used in the selection to confirm that the inserts were single copy and mapped to the correct chromosomal region. For the selections from the proximal region, JRL Groups 1 and 2 and proximal placental, the mapping panels also contained DNA from the hybrids H151 (breakpoint in Xp11.23) and SIN176 (Ingle *et al.*, 1985) (breakpoint in Xp11.23) to help confirm and refine the location of the respective cDNA clones.

Sequencing of Inserts. Plasmid DNA was prepared and the inserts were single pass sequenced using a modification of the Sanger dideoxy sequencing method (Sanger *et al.*, 1977) and Sequenase® Version 2, as described in the 'General Materials and Methods' section. M13F primer was used for sequencing the placental clones and the T7 primer was used for sequencing the JRL clones.

Development of New Region-Specific ESTs:

Sequences were analyzed by BLAST using the nr and dbest databases (Altschul *et al.*, 1990). Primers were selected from the sequence using Primer Designer (Version 1.01, Scientific and Educational Software) and synthesized on a Beckman Oligo 1000 DNA Synthesizer. Primer pairs were verified to be region-specific by

testing them on a set of conventional and radiation hybrids using TNK100 and the standard PCR conditions as in the 'General Materials and Methods' section.

Mapping of ESTs: New ESTs were sublocalized using the YAC and cosmid clones previously described (Boycott *et al.*, 1996). In addition, primers were designed from published EST sequence in Genbank, verified to be region-specific and positioned on the map, as described previously.

Characterization of KAT1:

Sequencing:

Plasmid DNA was prepared using the rapid plasmid small-scale method, as described in the 'General Materials and Methods' section, and the cDNA insert of clone 263773 was sequenced from both ends using T3 and T7 and Sequenanse Version 2.0, as described in the 'General Materials and Methods' section. Primers were designed from the generated sequence using Primer Designer (Version 1.01, Scientific and Education Software) and used to generate complete cDNA sequence for this clone.

Northern Analysis:

The cDNA clone 263773 was excised from pT7T3D vector using *Not*I and *Eco*RI and purified from low melting point agarose by melting the gel, quickly freezing it on a dry ice/ethanol bath, and

centrifugation. The clone was labelled using the random hexamer method and hybridized to a Clontech Multiple Tissue Northern Blot (Clontech) using ExpressHyb (Clontech). The Northern blot was prehybridized in 5 mL of ExpressHyb at 68°C for 30 min with shaking, after which time the ExpressHyb was removed from the bag. The probe was added to the blot in 5 mL of fresh ExpressHyb and hybridized at 68°C for one hour with shaking. The blot was rinsed several times at room temperature for 30 min with 2 X SSC, 0.05% SDS, followed by two washes for 20 min each at 50°C with 0.1 X SSC, 0.1% SDS. The blot was immediately wrapped in plastic wrap and exposed to film with two intensifying screens at -70°C for five hours to five days.

PCR Expression Profiles:

Primers designed to amplify JRL4-A1 crossed an intron and were therefore ideal for testing in cDNA libraries by eliminating possible positive signals due to genomic contamination. JRL4A1 was tested on 100 ng of cDNA from oligo dT primed retina, fetal retina, fetal brain, and placenta and random primed frontal cortex, fetal brain, retina, and placenta. PCR was performed using TNK100 and the standard PCR conditions described in the 'General Materials and Methods' section. PCR products were electrophoresed on a 1.5% SEPARIDE/1% agarose gel and visualized by EtBr staining.

Exon/Intron Boundaries:

cDNA Based Primers: Overlapping primer sets were designed from the cDNA sequence and tested in human genomic DNA using TNK100 and the standard PCR conditions described in the 'General Materials and Methods' section. Those primer sets that gave the expected product size were sequenced using Thermo Sequenase™ (Amersham) to confirm their origin. Those primer sets that gave a larger than expected product size were assumed to contain an intron and were sequenced from both ends to generate the intron sequence and boundary information. Products were excised in low melting point gel and the gel was digested with Gelase. Thermo Sequenase™ cycle sequencing is based on internal labelling by extension of the primers limited to a few nucleotides by using only three of the four dNTPs. Primers were designed to accommodate a labelling step with [α^{35}-S] dATP by analysis of the cDNA sequence downstream of the primer location. All primers were located within sequence so that at least two dATPs occurred before a fourth dNTP downstream of the 3' end of the primer. Approximately 50 ng of DNA was used for sequencing in a reaction that contained 0.5 pmole of primer (either forward or reverse), 12.5 µCi of [α^{35}-S] dATP, one µL of two of the three remaining dNTPs (3 µM of dCTP, 7-deaza dGTP, or dTTP), reaction buffer (29.7 mM Tris-HCl pH 9.5, 7.4 mM $MgCl_2$), and two µL of Thermo Sequenase (2 units/µL of Thermo Sequenase™, 0.006 units/µL of *Thermoplasma acidophilum* inorganic pyrophosphatase, in 50 mM Tris-HCl pH 8.0, 1 mM DTT, 0.1 mM EDTA, 0.5% Tween-20, 0.5% Nonidet P-40, 50% glycerol) in a total volume of 17.5 µL.

Reactions were mixed, overlaid with paraffin oil, and cycled for 60 cycles of 95°C for 15 s, 60°C for 30 s. After cycling, the reactions were terminated by transferring 3.5 μL of the labelling mix to each of four tubes containing 4 μL of ddATP, ddCTP, ddGTP, and ddTTP (150 μM each dATP, dCTP, 7-deaza-dGTP, dTTP, and 1.5 μM of the ddNTP). Reactions were again overlaid with paraffin oil and cycled for 60 cycles of 95°C for 30 s, 72°C for 60 s. After the termination step, the reactions were stopped by adding 4 μL of stop solution (95% formamide, 20 mM EDTA, 0.05% bromophenol blue, and 0.05% xylene cyanol FF). Samples were heated at 70°C for two min prior to loading on a sequencing gel, as described previously.

Cosmid Sequencing: A partial cosmid contig of Xp11.23 has been recently constructed by a group in Germany (Schindelhauer *et al.*, 1996). Through collaboration with Dr. A. Meindl at Ludwig-Maximilians-Universitat in Munchen, Germany, this cosmid contig was made available and KAT1 was determined to be on cosmid U138D3. This group then shotgun cloned this cosmid into M13 and arranged for it to be sequenced by Dr. A. Rosenthal at the Institut fur Molekulare Biotechnologie in Jena, Germany. The unedited primary sequence data was made available and this was aligned to the portion of the KAT1 gene where the exon/intron boundaries had not yet been determined by the 'Find' function in the word processing program, Microsoft Word 5.1.

Walking by Linker Ligation: The method for cloning of YAC ends was adapted for walking from known exon sequence into unknown intron sequence. The procedure was performed exactly as described in the 'Generation of New STSs' section with a few exceptions, including the template used, the primers used, and the restriction enzymes used. First, the cosmid DNA from U138D3 was used instead of YAC DNA to increase the copy number of the target. Secondly, gene specific primers were used instead of YAC vector specific primers. For the walk into the proximal end of intron B the primers R-6 and R-5 (see Results, Table 20), external and internal primers respectively, were used and for the walk into the distal end of intron B the primers F1-4 and F-2 (see Results, Table 20), external and internal primers respectively, were used. The restriction enzymes chosen for digestion of the cosmid had to meet two criteria, they had to be blunt cutters to facilitate the subsequent linker ligation, and they had to not cut between the primers R-6 and R-5 and the primers F1-4 and F-2. The restriction enzymes *Alu*I, *Dra*I, *Eco*RV, *Pvu*II, *Rsa*I, and *Sca*I met these criteria in both instances and all were used to increase the chance that a reasonably sized product would be generated containing unknown intron information.

5' RACE:

RACE reactions (Frohman *et al.*, 1988) were performed using human brain Marathon-ready cDNA (Clontech). Marathon-ready cDNAs are adaptor-ligated ds cDNAs ready for use as templates in both 5'-and 3'-RACE reactions based on innovations in the design of

the Marathon adaptor and suppression PCR technology. Construction of Marathon-ready cDNA begins with cDNA synthesis from polyA$^+$ RNA. First strand synthesis uses Moloney murine leukemia virus and a modified lock-docking oligo(dT) primer which contains two degenerate nucleotide positions at the 3' end, positioning the primer at the start of the poly-A tail and eliminating the 3' heterogeneity inherent with conventional oligo(dT) priming. Second strand synthesis is performed with a cocktail of *E. coli* DNA polymerase I, RNase H, and *E. coli* DNA ligase. Blunt ends are then created with T4 DNA polymerase and the ds cDNA is ligated to the Marathon cDNA adapter. The ability to perform 5'- and 3'-RACE from the same template is possible by the design of the adaptor. RACE reactions are primed with an internal gene-specific primer and the Marathon adaptor primer AP1. The adaptor-ligated cDNA does not contain a binding site for the AP1 primer until after the first round of thermal cycling, when the inner gene-specific primer is extended creating an AP1 binding site at the 5' (or 3') terminus of the cDNA. In subsequent cycles of amplification, both AP1 and the gene-specific primer can bind and the cDNA of interest can be exponentially amplified. Nonspecific products are greatly reduced because the AP1 binding site cannot be created on the general population of cDNA molecules, which also lack binding sites for the gene-specific primer. To reduce further the occurrence of non-specific products the exposed 3' end of the adaptor is blocked with an amine group to prevent the extension of the 3' end, which would create an AP1 binding site and allow nonspecific amplification. In addition, the

adaptor primer used is shorter than the adaptor itself (suppression PCR) and therefore prevents amplification of templates, where the 3' end has been extended to create an AP1 binding site, by preventing primer binding by panhandle structure formation.

RACE PCR was carried out using touchdown PCR and Clontech's Advantage Klen*Taq* Polymerase Mix (Klen *Taq*-1 DNA polymerase (an exo-minus, N-terminal deletion of *Taq* DNA polymerase) as the primary polymerase, a minor amount of an unspecified proofreading polymerase to provide 3' to 5' proofreading activity, and *Taq*Start Antibody for 'hot start' PCR), which was specifically developed for long range PCR using cDNA templates. To allow the use of touchdown PCR, gene-specific primers were chosen from the 5' end of the cDNA sequence that had a G/C content of greater than 50% and a T_m of at least 70°C. Touchdown PCR involves using an annealing/extension temperature that is several degrees higher than the T_m of the primers during the initial PCR cycles. Although primer annealing and amplification is less efficient at this higher temperature, it is more specific and only gene-specific synthesis occurs. The annealing/extension temperature is then reduced to the T_m of the AP1 primer for the remaining PCR cycles, permitting efficient, exponential amplification of the gene-specific template.

5' RACE of KAT1 was performed with the R-4 primer (see Results, Table 20) in a 50 µL reaction volume containing 1X Klen*Taq* PCR reaction buffer, 0.2 mM dNTP mix (of dATP, dCTP, dGTP, and dTTP), 1X Advantage Klen*Taq* Polymerase Mix, 5 µL of Marathon-ready cDNA, and 0.2 µM each of AP1 and R-4 primers. The reaction mix

was overlaid with paraffin oil and incubated at 95°C for 1 min followed by 5 cycles of 94°C for 30 s and 72°C for 4 min, 5 cycles of 94°C for 30 s and 70°C for 4 min, and 25 cycles of 94°C for 30 s and 68°C for 4 min on a Perkin-Elmer DNA Thermal Cycler 480. After PCR 25 μL of the reaction was run on a 1.5% SEPARIDE/1% agarose gel in 1X TAE buffer and visualized by EtBr staining. Products were run into low melting point agarose gel, excised, purified by Gelase, and reamplified using the AP2 internal adaptor primer and R-5 (see Results, Table 20), a primer internal to R-4. Products were again isolated and purified by Gelase and sequenced using Thermo Sequenase™ and the R-5 primer. Products were run on an acrylamide gel and exposed to X-ray film.

Mutation Analysis of KAT1:

-Deletion Analysis:

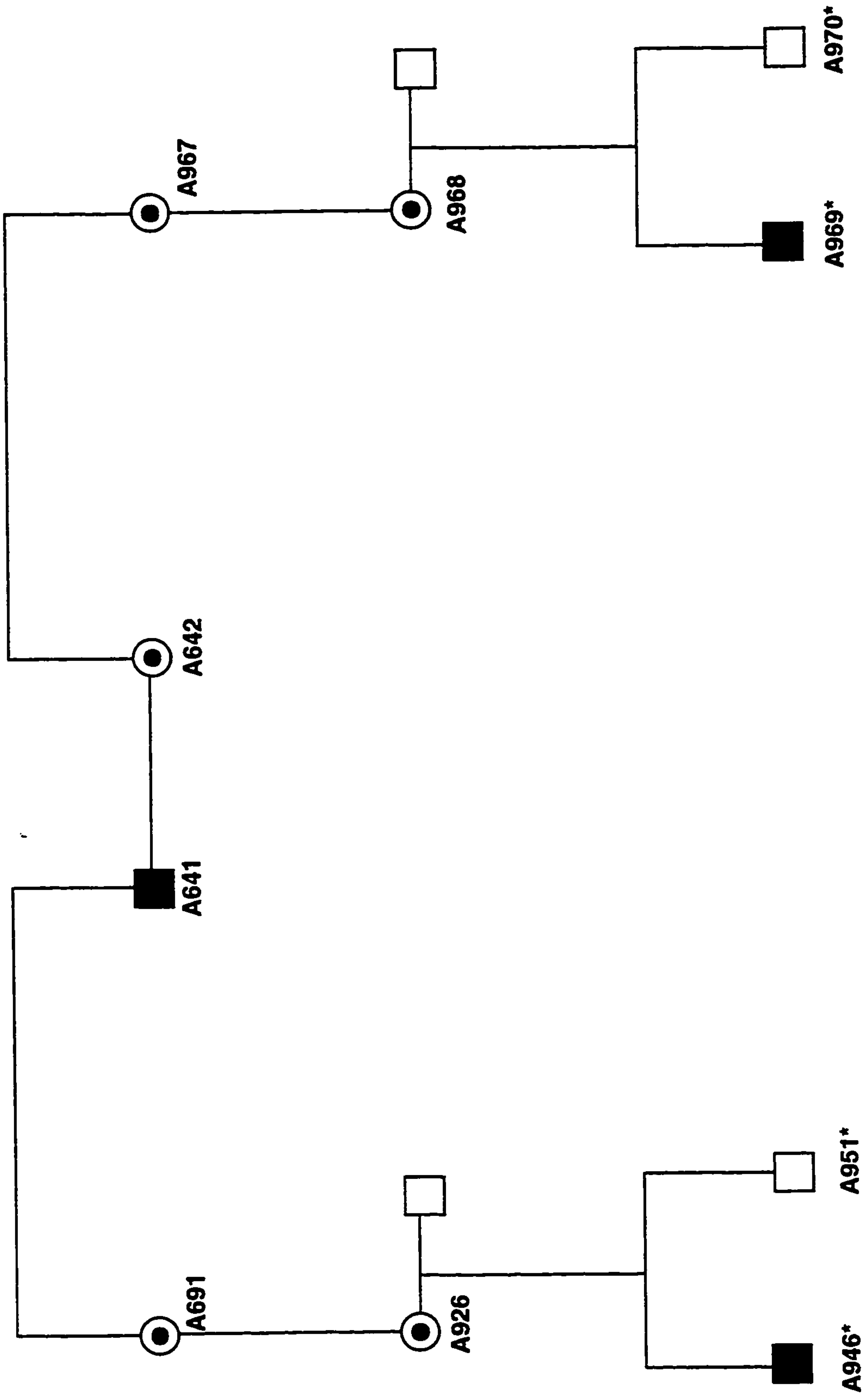
The cDNA clone 263773 was excised from pT7T3D vector using *Not*I and *Eco*RI and purified from low melting point agarose by melting the gel, quickly freezing it on a dry ice/ethanol bath, and centrifugation. Using an aliquot of the supernatant, the clone was labelled using the random hexamer method, described previously, and hybridized to filters containing genomic DNA cut with BamHI, HindIII, EcoRI, and PstI from individuals from CSNB1 Families P050, P060, P070, P080, P0100, P130, P140, P150, P170, P180, P190, P200, P230, P240, and 260. Hybridization and washes were as described in the 'General Materials and Methods' section and the

films were exposed at -70°C with amplification screens for three days.

Sequencing of Exons in Affected Individuals:

From the intron sequence generated during characterization of the exon/intron boundaries of KAT1, primers were designed using Primer Designer (Version 1.01, Scientific and Educational Software) to amplify the exons. These exons were amplified in two affected-unaffected sibling pairs from Family P060 (Figure 5) using TNK100 and the standard PCR conditions described previously. The PCR products were isolated from low melting point agarose gels using Gelase and sequenced using both the forward and reverse primer and Thermo Sequenase™, as described previously. Sequencing reactions were run on acrylamide gels and the terminations of each dNTP were run beside each other to facilitate recognition of base pair changes.

Figure 5. Relationship of the affected and control individuals from Family P060 used for mutation analysis. Squares indicate males and circles indicate females. Filled in squares are males showing clinical manifestations of CSNB1 and circles with dots in the centre indicate female carriers of CSNB1. A946 and A969 were the affected individuals and A951 and A970 were the unaffected controls used in the mutational analysis (individuals indicated by an asterisks following their 'A' designation).


A967
A968
A970*
A969*
A642
A641
A691
A926
A951*
A946*

CHAPTER 4 - RESULTS

Efforts to identify the gene for CSNB1 by a positional cloning strategy have taken place in four phases. Initially, recombinant X chromosomes within families segregating CSNB1 were investigated to narrow the candidate interval for the disease gene. Once a minimal interval had been established, the second phase involved construction of a physical contig across this minimal region. The third phase used this physical contig as a resource for the isolation of novel region-specific transcripts to serve as putative candidate genes for CSNB1. In the fourth phase, one transcript was chosen for characterization and this gene, KAT1, was excluded as the gene responsible for the clinical manifestations of CSNB1. The results of each of the four phases are described in detail in the following pages.

Refining X Chromosome Crossover Events in CSNB1 Families:

The first phase of this study was initiated to refine the location of the gene(s) for CSNB1 on the human X chromosome. For this, additional polymorphic markers were required to define more accurately the position of the crossover events which had been identified in our set of nine reasonably sized CSNB1 families (Families P060, P070, P130, P140, P170, P190, P200, P230, and P240). In six of these families, ten crossovers had been identified between MAOA-MAOB loci and DXS426 and one crossover had been identified between ZNF21 and DXS255, which placed the CSNB1 gene

proximal to MAOA-MAOB in five families and in Family P060 further refining the location of the CSNB1 gene proximal to ZNF21. Crossovers on the proximal side of the CSNB1 gene still needed to be characterized to put a tight flanking limit on the CSNB1 locus. Recently published polymorphic microsatellite markers (Table 1) that had been generally localized to Xp11 were chosen for further mapping studies using a hybrid mapping panel consisting of radiation-reduced and conventional somatic cell hybrids. Those markers that were found to map to locations within the current candidate interval were investigated within the families that had X chromosome crossover events.

Sublocalization of Initial Set of New Polymorphic Markers:

A hybrid mapping panel consisting of ten radiation-reduced hybrids and four somatic cell hybrids has been characterized in our lab to generate a physical map of human Xp11 (Boycott *et al.*, in press). New polymorphic markers were sublocalized in Xp11 based on their retention in the particular hybrid cell lines, as evaluated by PCR. The individual retention pattern of markers was compared to the somatic cell hybrid framework map and each marker was localized based on this comparison. An initial set of three markers, DXS988, DXS993, and DXS1003, were sublocalized on Xp11. Figure 6 shows the retention of DXS988 on the hybrid templates H21, H151, K6, K11, K52, K54, SIN176, and GM10063 as determined by PCR and this retention information was used to integrate DXS988 into interval 20 on the hybrid mapping panel (Figure 7). A new left primer for

an additional marker, DXS1000 (Table 5), was designed from existing sequence in Genbank because initial results with the original primer set were less than optimal and subsequent primer analysis using Primer Designer indicated that the left primer did not meet the standards set by this program. DXS1000 was also sublocalized within Xp11 and was found to map to the same interval in Xp11.22 as did DXS988 (interval 20). Both of these markers were investigated further to see if they could define a proximal limit to CSNB1 in any of the families (see next section). The retention of DXS1003 in the hybrids indicated that it maps proximal to MAOA-MAOB and distal to DXS426 (interval 12), and it was thought that DXS1003 could be helpful in resolving the crossovers in the region MAOA-MAOB to DXS426. DXS993 mapped distal to MAOA-MAOB (interval 5) and therefore did not map within the Xp11 region of interest.

Table 5. New Polymorphic Marker Designed from Published Sequence

Marker Designation	Primer Sequences 5' to 3'	Size (bp)	Genbank Accession No.
DXS1000	CTCTTTGGATTCCTGGACTT CTGTGTTATTGTACCTCTGCAT	149-155	Z17110

Figure 6. Retention of polymorphic marker DXS988 on a set of ten radiation-reduced and four somatic cell hybrids. Retention was evaluated using PCR and incorporation of a radioactive label, as described in the 'General Materials and Methods' section. The names of the hybrids are indicated along the top and control lanes include female (Human-F) and male (Human-M) human genomic DNA, mouse and CHO DNA, and ddH_2O. PCR products, of approximately 140 bp in size, were separated on 2% Nusieve® GTG® agarose/1% agarose gels and visualized by autoradiography. The two bands present in the female human genomic suggest that she has a different allele on each of her X chromosomes. Retention of DXS988 on hybrids H21, H151, K6, K11, K52, K54, SIN176, and GM10063 indicate that DXS988 belongs in interval 20 of the hybrid mapping panel. The marker is pBst cut with HpaII and labelled by an end-filling reaction. The sizes of the relevant fragments are indicated.

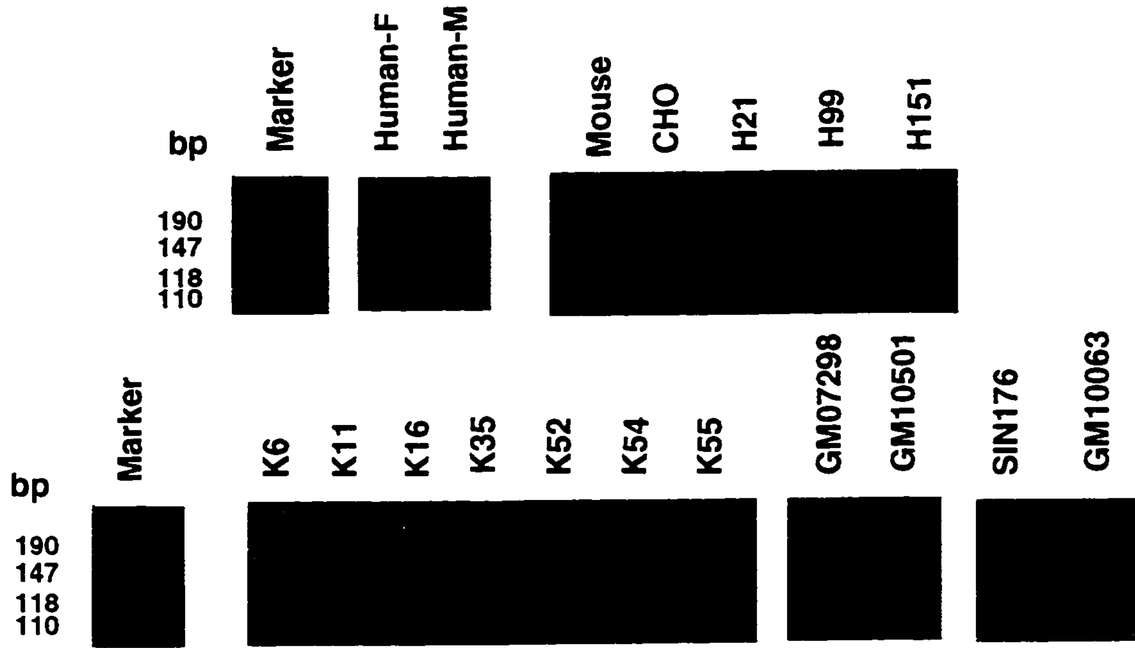
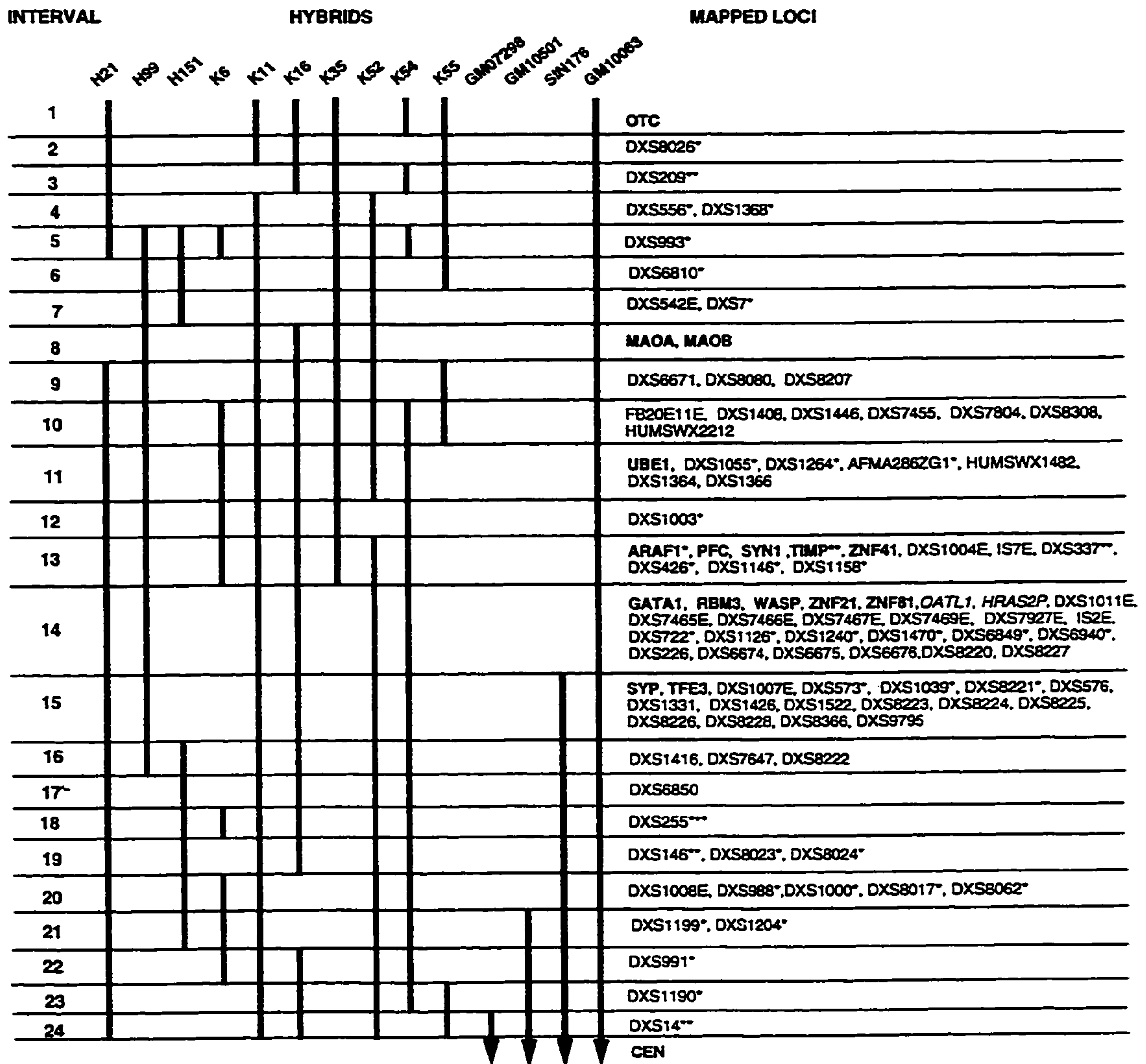

bp
Marker
Human-F
Human-M
Mouse
CHO
H21
H99
H151
190
147
118
110
bp
Marker
K6
K11
K16
K35
K52
K54
K55
GM07298
GM10501
SIN176
GM10063
190
147
118
110

Figure 7. Sublocalization of DNA markers in Xp11 using ten radiation hybrids and four somatic cell hybrids. DNA markers in boldface are genes, italics indicate pseudogenes, an 'E' after the marker name indicates EST, and ***, **, * indicates VNTR, RFLP, STR, respectively. The remaining markers are STSs. Construction of this map was the result of work by several individuals in the laboratory of Dr. N.T. Bech-Hansen and has been reported (Boycott *et al.*, in press).


INTERVAL
HYBRIDS
MAPPED LOCI
H21
H99
H151
K6
K11
K16
K35
K52
K54
K55
GM07298
GM10501
SIN176
GM10063
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17*
18
19
20
21
22
23
24
OTC
DXS8026*
DXS209**
DXS556*, DXS1368*
DXS993*
DXS6810*
DXS542E, DXS7*
MAOA, MAOB
DXS6671, DXS8080, DXS8207
FB20E11E, DXS1408, DXS1446, DXS7455, DXS7804, DXS8308, HUMSWX2212
UBE1, DXS1055*, DXS1264*, AFMA286ZG1*, HUMSWX1482, DXS1364, DXS1366
DXS1003*
ARAF1*, PFC, SYN1 ,TIMP**, ZNF41, DXS1004E, IS7E, DXS337**, DXS426*, DXS1146*, DXS1158*
GATA1, RBM3, WASP, ZNF21, ZNF81, OATL1, HRAS2P, DXS1011E, DXS7465E, DXS7466E, DXS7467E, DXS7469E, DXS7927E, IS2E, DXS722*, DXS1126*, DXS1240*, DXS1470*, DXS6849*, DXS6940*, DXS226, DXS6674, DXS6675, DXS6676, DXS8220, DXS8227
SYP, TFE3, DXS1007E, DXS573*, DXS1039*, DXS8221*, DXS576, DXS1331, DXS1426, DXS1522, DXS8223, DXS8224, DXS8225, DXS8226, DXS8228, DXS8366, DXS9795
DXS1416, DXS7647, DXS8222
DXS6850
DXS255***
DXS146**, DXS8023*, DXS8024*
DXS1008E, DXS988*, DXS1000*, DXS8017*, DXS8062*
DXS1199*, DXS1204*
DXS991*
DXS1190*
DXS14**
CEN

Analysis of DXS988, DXS1000, and DXS1003 in CSNB1 Families:

The markers DXS1003 and DXS988, were informative in most of the CSNB1 families. The crossovers identified previously in the region MAOA-MAOB to DXS426 (Table 6) were not recombined for DXS1003, refining the localization of these crossovers to the region between MAOA-MAOB and DXS1003 (Table 6), but the localization of the CSNB1 gene proximal to MAOA-MAOB (distal limit in Families P070, P130, P140, 260) remained the same. DXS1003 also confirmed a crossover in Family PO60 (D19) that was recombined for markers proximal and distal to DXS1003 and placed the CSNB1 gene proximal to ZNF21 (distal limit in Family P060) (Table 6).

In three CSNB1 families, six new crossovers were identified with DXS988 which placed CSNB1 distal to DXS988 (proximal limit) (Table 6). These crossovers were bounded distally by DXS255 and proximally by DXS988, an estimated distance of 3-4 cM (Donnelly *et al.*; 1994). The marker DXS1000 maps to the same interval as DXS988 and was uninformative in all the CSNB1 families except Family P130. In this family, individual E18 showed a crossover occurring between DXS255 and DXS988 but was not recombined for DXS1000, placing DXS1000 distal to DXS988 (Table 6). In this family, the boundaries of the crossover defining the proximal limit to the disease gene had been refined to the small interval between DXS1000 and DXS988. However, additional markers were needed to refine further the other crossover events identified and the minimal genetic region for CSNB1 in these families.

Table 6. Chromosomes with Crossovers Flanking the CSNB1 Locus

Individual	Status	Sex	MAOA	MAO B	DXS1003	DXS426	ZNF21	DXS722	DXS1126	DXS573	DXS255	DXS8023	DXS8024	DXS1000	DXS988
P060-D11	U	M	X	X	ni	ni	O	nt	nt	O	O	nt	nt	nt	O
P060-E120	A	M	nt	X	O	ni	O	nt	nt	nt	O	nt	nt	nt	O
P060-D161	NC*	F	nt	X	nt	nt	O	nt	nt	O	nt	nt	nt	nt	O
P060-D174	A*	M	nt	X	nt	nt	O	nt	nt	O	nt	nt	nt	nt	O
P060-E149	U	M	nt	X	O	nt	ni	nt	nt	O	O	nt	nt	nt	O
P070-E13	A	M	nt	X	O	O	ni	nt	nt	nt	O	nt	nt	nt	ni
P130-E16	A	M	X	X	O	O	O	nt	nt	nt	O	nt	nt	nt	O
P140-B1	A	M	X	X	O	O	O	nt	nt	nt	O	nt	nt	nt	O
P260-D49	NC*	F	X	ni	nt	nt	O	nt	nt	nt	nt	nt	nt	nt	O
P260-E25	NC*	F	ni	X	nt	nt	O	nt	nt	nt	nt	nt	nt	nt	O
P060-D19	A	F	nt	X	X	X	X	X	O	O	O	nt	nt	nt	O
P060-D3	A	M	O	O	O	ni	O	nt	nt	O	O	X	ni	ni	X
P060-E6	U	M	O	O	O	O	O	nt	nt	O	O	O	ni	ni	X
P060-E117	U	M	nt	O	O	ni	O	nt	nt	O	O	O	O	ni	X
P060-C26	C	F	nt	O	O	nt	O	nt	nt	ni	O	O	ni	ni	ni
P0130-E18	U	M	O	O	O	O	O	nt	nt	nt	O	nt	nt	O	X
P0140-D1	U	M	O	O	O	O	ni	nt	nt	nt	O	ni	ni	ni	X

Note: U - unaffected; A - affected with CSNB1; C - carrier; NC - noncarrier; * status inferred; M - male; F - female
O - allele in phase (nonrecombinant) with CSNB1 locus; X - allele not in phase (recombinant) with CSNB1 locus
ni - noninformative for locus; nt - not tested

Generation of New Polymorphic Markers from Published Sequence and Sublocalization of Additional Polymorphic Markers:

To refine the minimal genetic region for CSNB1 further, two sets of primers were designed from published sequence in Genbank that contained polymorphic CA repeats (Table 7). For the CA repeat at DXS722, primers had not yet been designed from the sequence by the investigator (Thiselton *et al.*, 1995). For the CA repeat at DXS991, the primers that had been chosen by Genethon were deemed unacceptable by what was now routine analysis of all new primers using Primer Designer, and therefore new primers were designed that would meet the criteria known to identify primers that amplify well under the standard PCR conditions in use.

Table 7. New Polymorphic Markers Designed from Published Sequence

Marker Designation	Sequence 5' to 3'	Size (bp)	Genbank Accession No.
DXS722	CTCATAGTCCTAGCTGCTGG CCTGATCCTGTTCCACTGGG	188-194	X67606
DXS991	CTCCAGCACTTCAACCACAG ATCATTTGAGCCAATTCTCC	272-296	Z16680

Both of these markers, along with 16 others (Table 1), were chosen to be sublocalized within Xp11 in efforts to refine the location of critical crossover events and reduce the minimal genetic region for CSNB1. The markers DXS983, DXS1194, and DXS1213 were localized

to Xq and would therefore not be of use in these studies. The markers DXS991 (interval 22), DXS1190 (interval 23), DXS1199 (interval 21), DXS1204 (interval 21), DXS8017 (interval 20), and DXS8062 (interval 20) were sublocalized to regions of Xp11 (Figure 7) that were at or proximal to the location of DXS988 and DXS1000 and would therefore not refine the location of the proximal limit of CSNB1 (Table 6). The marker DXS556 (interval 4, Figure 7) mapped distal to MAOA-MAOB and would not be of use in refining the location of the crossover events occurring between MAOA-MAOB and DXS1003. The marker DXS1158 (interval 13, Figure 7) mapped proximal to DXS1003 and distal to ZNF21 and would therefore not be of use in refining the location of the crossover events occurring between MAOA-MAOB and DXS1003 or the crossover event occurring between ZNF21 and DXS255.

The markers DXS8023 and DXS8024 were mapped to interval 19, based on their retention in the hybrids H21, H151, K11, K16, K52, K54, SIN176, and GM10063 (Figure 7). Both of these markers lie between DXS255 and DXS988 and had the potential to refine the location of the five DXS988 crossovers and the proximal limit of CSNB1.

The markers DXS722, DXS1126, DXS1240, and DXS6940, were all mapped to interval 14 based on their retention in the hybrids H21, H99, K11, K16, K52, K54, and GM10063 (Figure 7). The markers DXS573 and DXS1039 both mapped to interval 15 based on their retention in the above hybrids and the hybrid SIN176 (Figure 7). These six markers lie between ZNF21 and DXS255 and have the

potential to refine the crossover that occurs in this region in individual D19 from Family P060.

Additional Polymorphic Marker Analysis in CSNB1 Families:

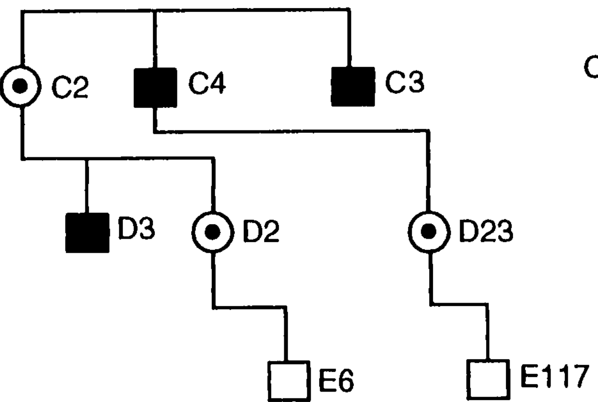
Distal Limit: To refine the location of the crossover event occurring between ZNF21 and DXS255 in individual D19 from Family P060, appropriate individuals were tested with the newly mapped markers. Initially DXS1039 was found to be uninformative in this part of Family P060, while DXS573, a more distal marker (based on physical contig data, see next section and Boycott *et al.*, 1996), was found to informative. Individual D19 was not recombined for DXS573 so the location of the crossover event was refined to the region between ZNF21 and DXS573, but the distal limit of CSNB1 in this family was still at ZNF21. The markers DXS1126 and DXS1240 could not be orientated with respect to each other but they do lie distal to DXS573 (based on physical contig data, see next section and Boycott *et al.*, 1996). These markers were tested on the family and DXS1240 was found to be uninformative. However, D19 was not recombined for the DXS1126 locus, thus refining the location of the crossover event in this individual to the region between ZNF21 and DXS1126 (Table 6; Figure 9), but the distal limit of CSNB1 in this family was still at ZNF21.

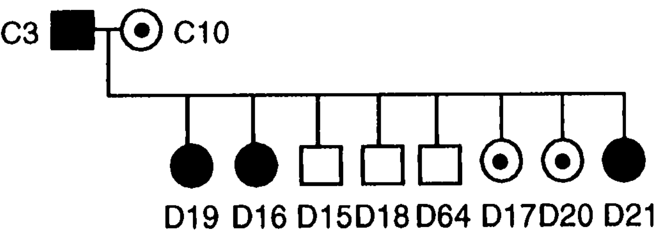
To further narrow the minimal genetic region for CSNB1, the marker DXS722, lying proximal to ZNF21 and distal to DXS1126 (based on physical contig data, see next section and Boycott *et al.*, 1996), was analyzed in this part of Family P060. The marker

DXS722 was informative and individual D19 was recombined between CSNB1 and this marker (Table 6; Figures 8 and 9), refining the location of this critical crossover event to the region between DXS722 and DXS1126, and placing the gene for CSNB1 in this family proximal to DXS722. Marker DXS6940, which maps between DXS722 and DXS1126 (based on physical contig data, see next section and Boycott *et al.*, 1996), was also analyzed in the family but was uninformative. The crossover defining the distal limit of CSNB1 in Family P060 at DXS722 occurs between the markers DXS722 and DXS1126, a distance of less than 300 kb (see Physical Contig section and Boycott *et al.*, 1996).

Proximal Limit: To refine the location of the five crossover events occurring between DXS255 and DXS988, appropriate individuals were analyzed with the markers DXS8023 and DXS8024. DXS8024 was uninformative for four of the five crossover events but did refine the location of the crossover event in individual E117 from Family P060 to the region between DXS8024 and DXS988 (Table 6), however the proximal limit of the disease gene remained at DXS988. DXS8023 was informative in Family P060 and individual D3 was recombined between CSNB1 and this marker (Table 6, Figures 8 and 9), refining the location of this proximal crossover to the region between DXS255 and DXS8023, a distance of less than 700 kb (data not shown; Nelson *et al.*, 1995), and placing the CSNB1 gene in Family P060 distal to DXS8023.

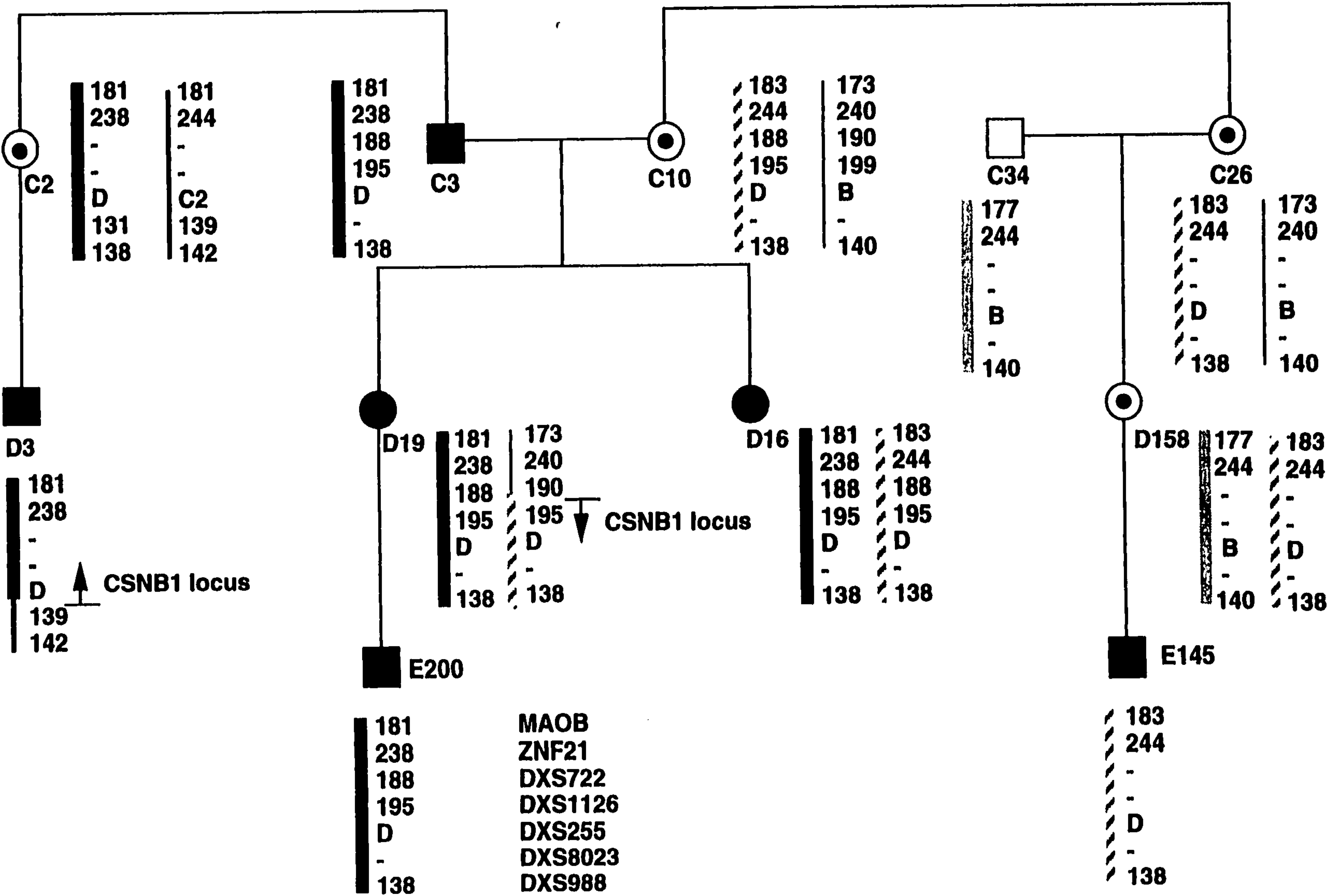
Figure 8. Refining the proximal and distal limits of CSNB1 in Family P060 with the polymorphic CA repeats DXS8023 (A) and DXS722 (B). From analyzing DXS8023 in this part of Family P060 it was shown that D3 carries the 139 bp allele while his affected uncles carry the 131 bp allele. Since D3 is affected, the gene for CSNB1 must lie distal to DXS8023. From analyzing DXS722 in another part of Family P060, the affected female D19 was shown to have the 188 bp, 190 bp genotype of her carrier sisters. As she inherited the CSNB1 chromosome from her father, the recombinant chromosome is a product of her mother's CSNB1 and normal chromosomes. Since D19 is affected, the CSNB1 gene must lie proximal to DXS722.


C2
C4
C3
D3
D2
D23
E6
E117


C3
C10
D19 D16 D15 D18 D64 D17 D20 D21

Alleles
131 139 139 131 131 133 133 131 131 143 143
DXS8023
188 188 190 188 190 188 188 190 190 190 188 190 188 190 188 188
Alleles
DXS722

Figure 9. Crossover events defining the location of the CSNB1 gene in Family P060. Individuals D3 and D19, carry chromosomes with critical crossover events which define the proximal and distal limits of the CSNB1 gene, respectively. Both the thick, solid black line and the candystriped line depict chromosomes which carry the CSNB1 gene. A dash indicates the marker was either not informative or not tested.


C2
C3
C10
C34
C26
D3
D19
D16
D158
E200
E145
CSNB1 locus
CSNB1 locus
MAOB
ZNF21
DXS722
DXS1126
DXS255
DXS8023
DXS988

In summary then, the CSNB1 gene in Family P060 is flanked distally by DXS722 and proximally by DXS8023, a region estimated to be less than 2.2 Mb (data not shown; Nelson *et al.*, 1995; Boycott *et al.*, 1996; Schindelhauer *et al.*, 1996).

Development of a Physical Contig:

Construction of the Physical Contig:

The CSNB1 minimal region in Family P060 was defined as the interval between DXS722 and DXS8023. As there are a number of crossovers that would place the CSNB1 gene distal to DXS255 (see Literature Review, page 8), the minimal region for CSNB1 could potentially be restricted to the interval between DXS722 and DXS255. To identify candidate genes from this region, a physical contig of 2 Mb was constructed (Boycott *et al.*, 1996) and is described here.

DNA markers were initially grouped into intervals based on their retention in conventional and radiation hybrids (Figure 7). The minimal genetic region for CSNB1 between DXS722 and DXS255, spans intervals 14 to 18 on the hybrid mapping panel. Initially, 22 markers (Table 3) were mapped within these intervals and YACs were isolated using these markers and grouped into preliminary contigs based on their DNA marker content. To bridge gaps, add YAC coverage to the region, and increase DNA marker density, 22 new region-specific STRs/STSs/ESTs were developed, including one STR (Table 5) and three ESTs (Table 8) designed from existing sequences

in Genbank, and 18 STSs (Table 9, 'Integrity of YAC Clones' section) designed from selected end-fragments of YACs.

Table 8. ESTs Designed from Published Sequences[a]

DXS No.	Primer Sequences 5'-3'	Size (bp)	TNK Buffer
DXS7465E	GGCCCTCTCTCTACTCCTTG GGTGGGGATTGACAGGCATG	110	100
DXS7466E	CTTCCTGACTACCTGTGCCC AGAGGAGAACCAGATTCAGT	110[b]	100
DXS7467E	AAAGGAGCTGCTAAGACTTA AAAGGCTCTGACGTGTGTGC	143	100

[a] Geraghty *et al.*, 1993.
[b] Primers designed from cDNA sequence and amplifies a 145 bp product from genomic DNA.

To fill remaining gaps and add coverage to unstable regions, cosmids were isolated by filter hybridization for the loci DXS226 and SYP from the ICRF cosmid library, and SYP and TFE3 from the Lawrence Livermore cosmid library, and cosmids were isolated for the loci EC4484L, DXS573, and EC2931L from the Lawrence Livermore cosmid library by PCR. Figure 10 shows a hybridization of a filter containing part of the Lawrence Livermore cosmid library with the probe TFE3. All positive cosmid clones were isolated as single colonies and tested for retention of surrounding DNA markers. The physical contig was assembled using 45 YACs and 6 cosmids corresponding to the 44 DNA markers from the Xp11.23-p11.22

Figure 10. Autoradiogram showing hybridization of TFE3 onto one of the 16 filters containing the Lawrence Livermore X Chromosome Cosmid library. The filter contains spotted 1536 cosmid clones which were visualized by hybridization using Lawrist 16 labelled with [α-^{35}S] dATP. The two clones positive for TFE3 are indicated with arrows.

region spanning ZNF21 and DXS255. The contig was constructed based on the DNA marker content of YAC and cosmid clones, as determined by PCR (Figure 11), or Southern hybridization (Figure 12), and information on DNA marker retention in conventional and radiation hybrids (Figure 7). The computer program SEGMAP (Green and Green, 1991) was used to confirm the order of the markers and the placement of the YAC and cosmid clones in the contig. The physical contig is presented in Figure 18.

Integrity of Clones:

The stability and sizes of the various YACs was evaluated using pulsed-field gel electrophoresis. Five clonal isolates of each YAC were run on a pulsed-field gel with the marker strain YP149 or the host strain AB1380, or both. Figure 13A is an example of such a gel with clonal isolates from the YAC yWXD5905. Pulsed-field gels were blotted and probed with total human genomic DNA to detect the YACs within the yeast chromosome background (Figure 13B). The sizes of the YACs were determined by comparing the position of the YAC with the YP149 size standard. The arrow in Figure 13 indicates the clonal isolates of the YAC yWXD5905, all of which are 150 kb in size.

YAC stability was assessed by looking at the sizes of the five clonal isolates for each YAC. The clonal isolates of the YAC yWXD5905 shown in Figure 13 were the same size indicating that this YAC is stable, in contrast to some YAC clones from this region of the X chromosome. Figure 14 shows the analysis of clonal isolates

Figure 11. STS content analysis of YAC clones using the DNA marker EC3978L by PCR performed under the standard conditions described in the 'General Materials and Methods' section. The YAC clone names are indicated along the top and are described in Table 10. The marker EC3978L is a YAC end-clone (Table 9) and is present on the parent clone yWXD3978, as well as five other YACs, yWXD4484, yWXD3986, yWXD4057, yWXD3203, and yWXD3432. Products were separated on a 1.5% SEPARIDE/1% agarose gel and visualized by EtBr staining. The DNA size standard is pBst cut with *Hae*III and the size of relevant bands are shown. The control lanes include human genomic DNA and ddH_2O. The Human* control lane is human genomic DNA derived from an affected male with CSNB1 from Family P060.

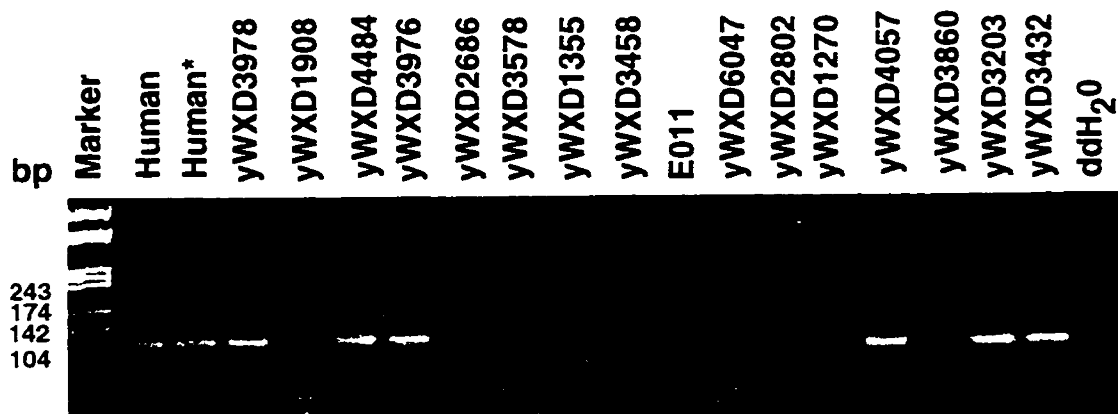
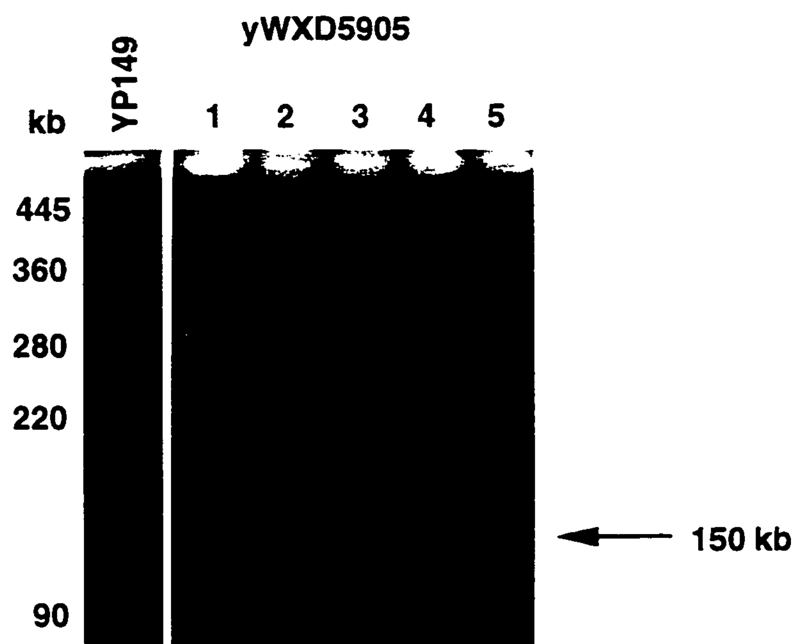

bp
Marker
Human
Human*
yWXD3978
yWXD1908
yWXD4484
yWXD3976
yWXD2686
yWXD3578
yWXD1355
yWXD3458
E011
yWXD6047
yWXD2802
yWXD1270
yWXD4057
yWXD3860
yWXD3203
yWXD3432
ddH2O
243
174
142
104

Figure 12. DNA marker content analysis of YAC clones using Southern hybridization. The YAC clone names are indicated along the top and are described in Table 10. YAC clones were digested with *Eco*RI and electrophoresed on 0.8% agarose gels, blotted, and probed with the DNA marker DXS226. The control lanes include human genomic DNA and DNA from a hybrid containing only the X chromosome (GM06318B). DXS226 is contained in the YACs yWXD2686 and yWXD3578.

yWXD1908
yWXD3986
yWXD3978
yWXD4484
yWXD4057
yWXD3860
yWXD3432
yWXD3203
yWXD6099
yWXD3225
yWXD5986
yWXD3578
yWXD2686
yWXD2687
yWXD6752
yWXD6753
C0874
X-only
Human

Figure 13. Pulsed-field gel electrophoresis of YAC yWXD5905 (A) and subsequent probing of the membrane with total human genomic DNA (B). yWXD5905 does not comigrate with any of the yeast chromosomes and its position relative to the DNA size standard YP149 can be determined simply by EtBr staining (A) and analysis of the Southern hybridization results (B). The size of this YAC has been determined to be 150 kb and it appears to be stable.

A.



B.

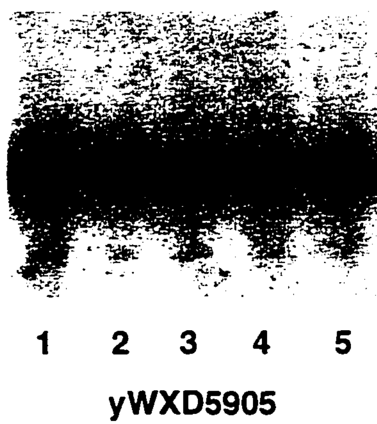
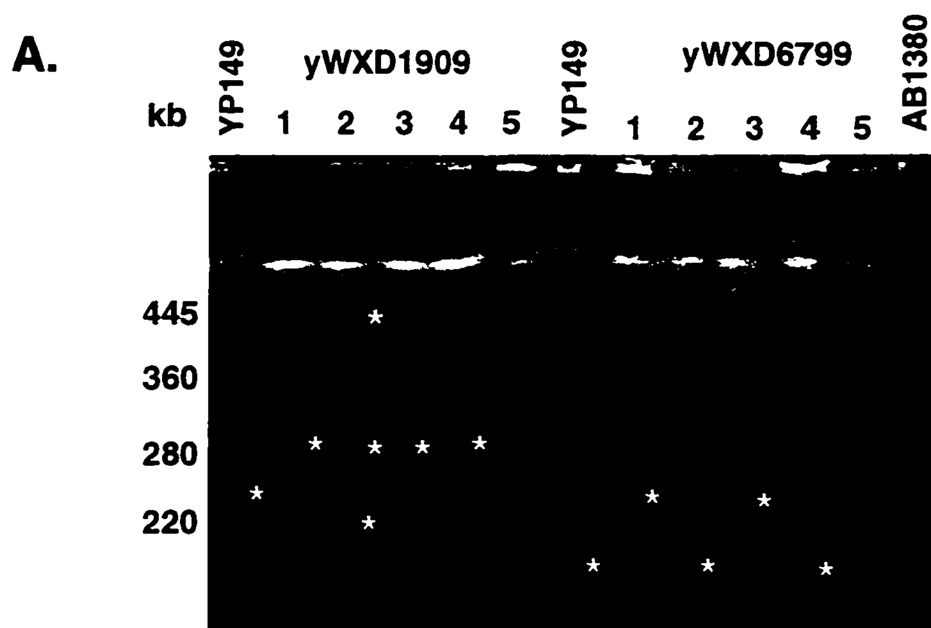
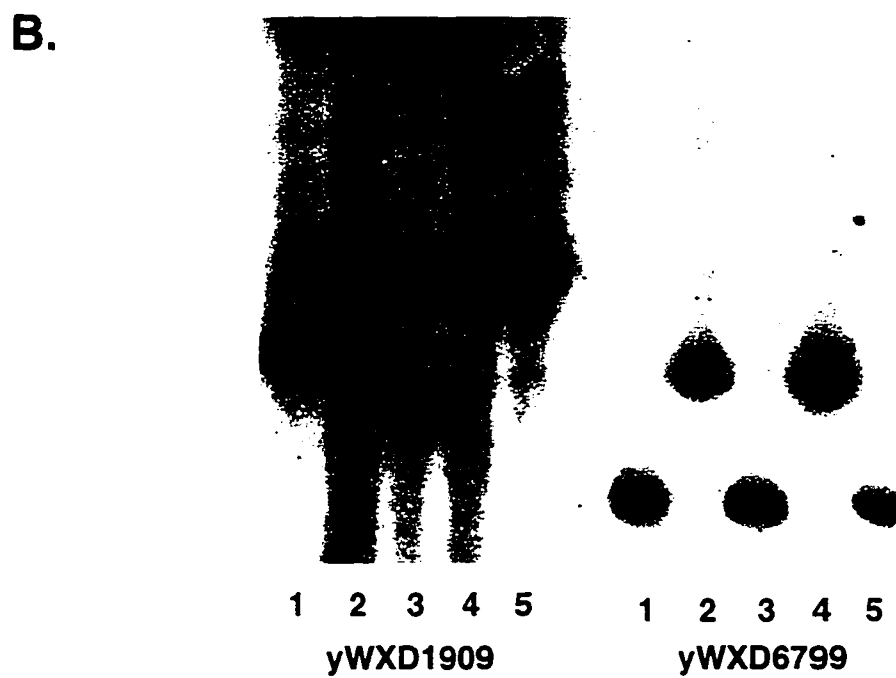


Figure 14. Pulsed-field gel electrophoresis of the YACs yWXD1909 and yWXD6799 (A) and subsequent probing of the membrane with total human genomic DNA (B). Based on both the EtBr staining and analysis of the Southern hybridization results, the YAC clones were located relative to the DNA size standard YP149 and are indicated by asterisks. Both of these YACs appear to be quite unstable. Chromosomes from the host strain AB1380 were included to aid in determining the location of YACs using EtBr visualization.


A.
kb
YP149
yWXD1909
1
2
3
4
5
YP149
yWXD6799
1
2
3
4
5
AB1380
445
360
280
220


B.
1 2 3 4 5
yWXD1909
1 2 3 4 5
yWXD6799

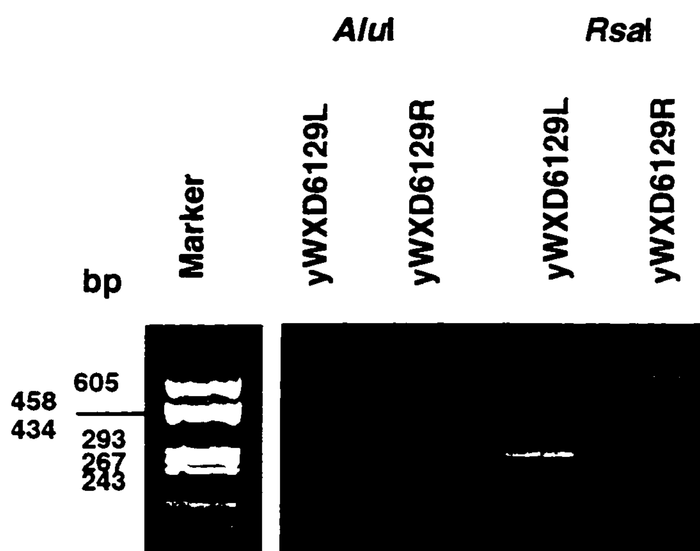
from YACs yWXD1909 and yWXD6799, both of which were found to be unstable and subsequently not used in the final construction of the physical contig. The size variation of these YACs was due to instability, and not merely the result of well-to-well contamination, because the clonal isolates were tested and found to retain a number of markers in Xp11.23 (data not shown).

Usually the largest YAC clonal isolate was chosen for further characterization. The sizes of the YACs ranged from 35 kb to 1770 kb (Table 10), the majority of which appeared to be stable at the indicated size. However, it should be noted that after 48 hours of growth, the most stable deletion derivative can dominate the culture, so that while a YAC may appear stable, it has merely assumed its most stable form and is not indicative of the true parent clone.

In addition to YACs, cosmids were isolated for a number of markers and the molecular size of the cosmids averaged 40 kb (Table 10). However, there were several cases of cosmids containing the TFE3 locus from both the Lawrence Livermore and ICRF cosmid libraries that had undergone obvious internal deletions since the size of the cosmids ranged from 10-28 kb (data not shown), an insert size impossible to package using this system (Feiss *et al.*, 1977; Hohn and Collins, 1980).

Generation of YAC ends by ligation-mediated PCR not only adds additional region specific markers to the region, but also evaluates the integrity of the YAC clone. Attempts to isolate left and right ends of YACs bordering gaps and those in regions of low marker density were quite successful. Figure 15 shows the products of ligation-

Figure 15. EtBr stained gel showing the products of a ligation-mediated PCR using the enzymes *Rsa*I and *Alu*I on the left and right ends of YAC yWXD6129. All four reactions produced single major bands and the largest generated from each end was sequenced. The DNA size standard is pBst cut with *Hae*III and the sizes of the relevant bands are shown.


AluI
RsaI
bp
Marker
yWXD6129L
yWXD6129R
yWXD6129L
yWXD6129R
605
458
434
293
267
243

mediated PCR using the YAC yWXD6129, and *Rsa*I and *Alu*I restriction enzymes. Ends generated by this method containing single products over 200 bp in size, as indicated by electrophoresis, were sequenced using the dsDNA cycle sequencing system, the sequence analyzed, and primers designed to amplify new STSs. Part of the sequence generated by the cloning of the right end of yWXD6129, including the locations of both the right and left primers, is shown in Figure 16. New end-clone STSs were tested on the panel of conventional and radiation hybrids, as well as their parent clone, to determine whether they were region-specific and could be used for chromosome walking. Amplification using the new end-clone STS, designated EC6129R (for end-clone of yWXD6129, right end) on the panel of hybrids and its parent YAC yWXD6129 is shown in Figure 17.

Of the 27 isolated end-clones, 18 were found to be region-specific STSs (Table 9) and were used to isolate additional YACs to add coverage and bridge gaps. Of the nine chimeric ends, the three derived from the I or M library were from another human chromosome and the six derived from the F library were from hamster DNA. Of a total of 20 YAC insert-ends evaluated from the F library, six were derived from hamster, indicating a 30% rate of chimerism. Chimeric YAC clones are indicated by dashed lines in Figure 18. The YAC clones, yWXD5742, yWXD4518, yWXD4057, YWXD1908, and ICRFy900E011 are also shown as dashed lines in Figure 18. These YACs are too large to fit within the constraints of the map, suggesting that these clones are also chimeric.

Figure 16. Right end-clone sequence of yWXD6129 determined by sequencing the isolated *Rsa*I end-clone product shown in Figure 15 by the dsDNA cycle sequencing system. The sequence used for the left and right primers of the EC6129R STS are indicated (Table 9). The dideoxy G, A, T, and C lanes are indicated and the termination products have been resolved on a 6% acrylamide gel.

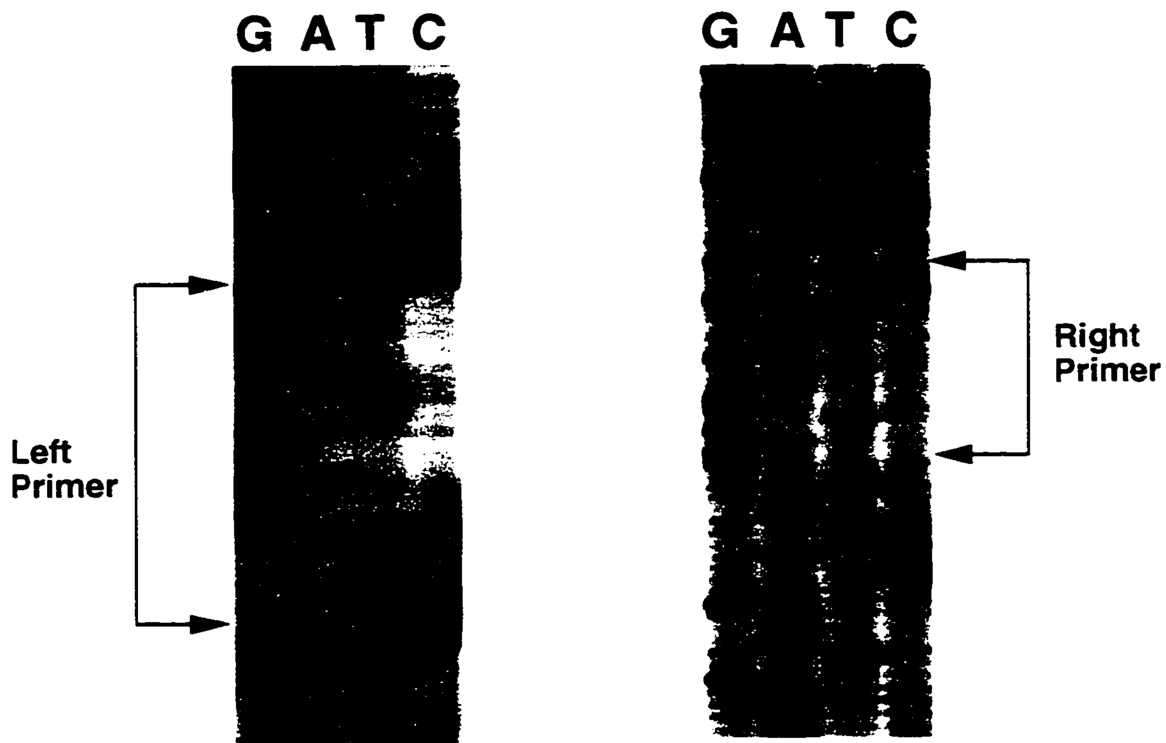

G A T C
G A T C
Left
Primer
Right
Primer

Figure 17. Testing of the new STS EC6129R on the hybrid mapping panel and the parent YAC yWXD6129. The names of the hybrids are indicated along the top and the control lanes include DNA from mouse, CHO, human and the X-only hybrid GM06318B. The EC6129R STS (DXS8222, Table 9) is retained on hybrids H21, H99, H151, K11, K16, K52, K54, SIN176, and GM10063, as well as the parent YAC, illustrating that this end-clone is region-specific (interval 16) and was derived from yWXD6129. The DNA size standard is pBst cut with *Hae*III and the sizes of the relevant bands are indicated.

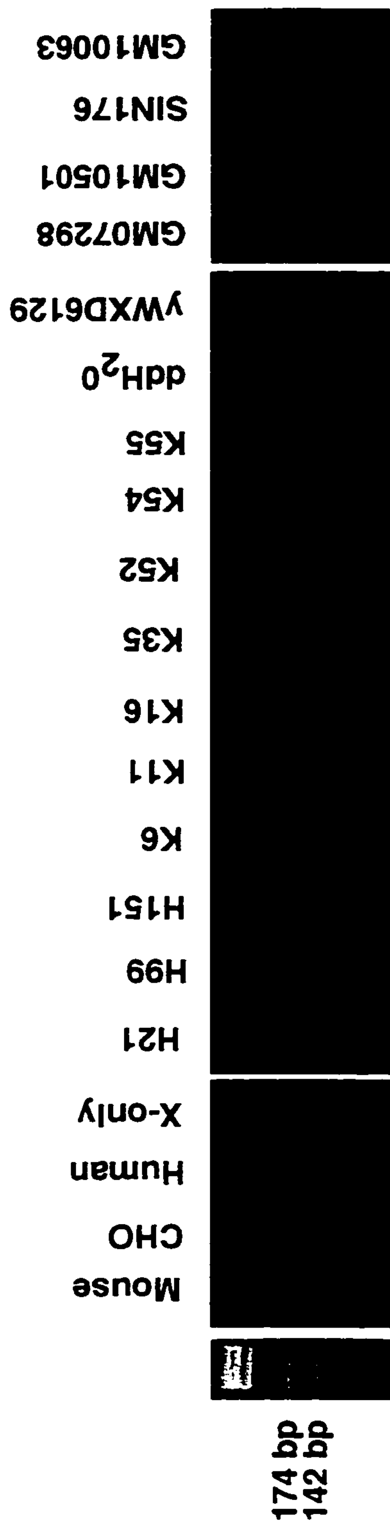
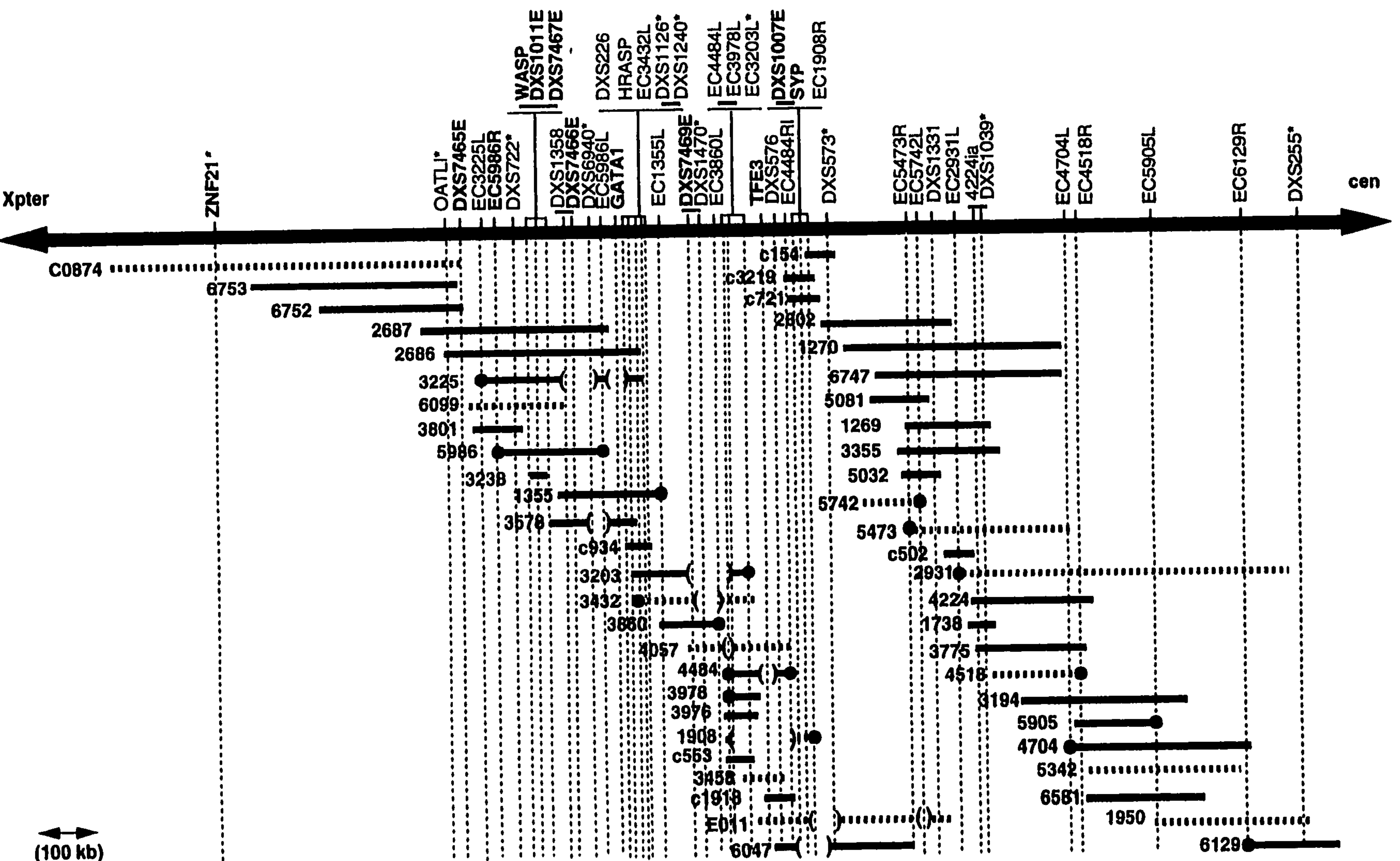

GM10063
SIN176
GM10501
GM07298
yWXD6129
ddH20
K55
K54
K52
K35
K16
K11
K6
H151
H99
H21
X-only
Human
CHO
Mouse
174 bp
142 bp

Figure 18. Physical contig and DNA marker map in human Xp11.23-p11.22 (Boycott *et al.*, 1996). The thick, dark line with bidirectional arrows indicates the chromosome with the telomere to the left and the centromere to the right. DNA markers and genes are ordered across the top of the figure, as determined by retention in conventional and radiation hybrids, YAC and cosmid content, parsimony, and SEGMAP analysis. Genes are indicated by boldface type and polymorphic markers are shown with an asterisk. A line under two or more markers indicates that they could not be ordered relative to each other. YAC and cosmid clones are represented as lines together with their names and are described in Table 10. The names of cosmid clones are superceded by a 'c'. The length of the solid line corresponding to a YAC or cosmid indicates its approximate size. Dashed lines associated with a YAC indicate that there is evidence of chimerism or, in the case of yWXD5742, yWXD4518 yWXD4057, yWXD1908, and ICRFy900E011, that the size of the YAC is too large to fit within the constraints of the map, suggesting that these clones are also chimeric. In either of these cases the length of the YAC corresponds only to the markers it contains. Deletions in clones are indicated by brackets. YAC end-clones are represented by black circles.



Xpter
cen
ZNF21*
OATLI*
DXS7465E
EC3225L
EC5986R
DXS722*
WASP
DXS1011E
DXS7467E
DXS1358
DXS7466E
DXS6940*
EC5986L
GATA1
DXS226
HRASP
EC3432L
DXS1126*
DXS1240*
EC1355L
DXS7469E
DXS1470*
EC3860L
EC4484L
EC3978L
EC3203L*
TFE3
DXS576
EC4484RI
DXS1007E
SYP
EC1908R
DXS573*
EC5473R
EC5742L
DXS1331
EC2931L
4224ia
DXS1039*
EC4704L
EC4518R
EC5905L
EC6129R
DXS255*
C0874
6753
6752
2687
2686
3225
6099
3801
5986
3238
1355
3578
c934
3203
3432
3680
4057
4484
3978
3976
1908
c553
3458
c1918
E011
6047
c154
c3219
c721
2802
1270
6747
5081
1269
3355
5032
5742
5473
c502
2931
4224
1738
3775
4518
4194
5905
4704
5342
6581
1950
6129
(100 kb)

FISH analysis performed on the YAC clone yWXD2931 indicated two chromosomal assignments, only one of which was from the expected location on the X chromosome (Dr. B. Roland, personal communication). Since EC2931L maps to the region of interest and EC6129R maps on yWXD2931, it would appear that yWXD2931 is intact from its left end to EC6129R. yWXD2931 does not contain DXS255 which would indicate that the right end of the YAC contains the chimeric insert. A chimeric fragment of approximately 1200 kb would account for the large size of this YAC (1770 kb) relative to the estimated region it spans (550 kb) (Figure 18).

Based on the most parsimonious marker order, nine of the YACs in Figure 18 appear to contain internal deletions (indicated as brackets in Figure 18). Two of the deletions occur in the distal portion of the contig between DXS1358 and DXS226. Approximately 40% of the YACs from this subregion show both size and STS content differences between clonal isolates (data not shown). There are at least three to five intact YAC clones that complement the observed deletions in this region. Seven of the deletions involve one or more of a group of markers between DXS7469E and DXS573, including the genes TFE3 and SYP. As many as 80% of the YACs in this region show both size and STS content differences between clonal isolates (data not shown). This region of the contig appears to be highly unstable in YACs.

Order of DNA Markers:

Markers could be unambiguously ordered along the chromosome except for five groups, (WASP, DXS1011E and DXS7467E), (DXS1358 and DXS7466E), (DXS1126 and DXS1240), (DXS1007E and SYP), and (4224ia and DXS1039), which cosegregated in all of the clones tested. Two additional pairs of markers, (DXS7469E and DXS1470), and (EC4484L and EC3978L), could also not be ordered on the basis of cosegregation but did demonstrate differences in deletion. The markers, DXS7469E and DXS1470, cosegregated in two YACs, yWXD3860, yWXD4057, and codeleted in one YAC, yWXD3203. In the YAC yWXD3432, DXS7469E was present but DXS1470 was deleted. If the most parsimonious deletion is assumed to have occurred in yWXD3432, then it would appear that DXS7469E maps distal to DXS1470. The second pair of markers, EC4484L and EC3978L, cosegregated in four YACs, yWXD3432, yWXD4484, yWXD3978, yWXD3976 and one cosmid, c553, but EC4484L was deleted in yWXD3203, yWXD4057, and EC3987L was deleted in yWXD1908. If the most parsimonious deletions are assumed to have occurred in these YACs, then it would appear that EC4484L maps distal to EC3978L. However, in these two cases, since the two markers cannot be ordered relative to each other based on their presence in nondeleted YACs the order has been indicated in Figure 18 as unresolved.

Weak Links in the Contig:

In the map there are two weak links, defined as two markers being held together by only one clone. Both of these weak links occur in the region between DXS7469E and DXS573, suggesting this region may be prone to deletion. The first of these weak links, between EC3860L and EC4484L/EC3978L, is held together only by yWXD4057. Upon screening the YAC libraries with EC3860L only the parent YAC, yWXD3860, and the YAC yWXD4057 were identified. Screening of the YAC libraries with EC4484L and EC3978L identified several YACs (five of which are shown in Figure 18), but only one of them, yWXD4057, was positive for EC3978L and EC3860L. Cosmid screening with EC4484L, EC3978L, and EC3203L isolated 13 cosmids, none of which contained EC3860L (one representative cosmid, c553, is depicted in Figure 18). The second weak link is between EC1908R and DXS573, and no YAC was identified that contained both of these STSs. YAC library screening with EC1908R isolated only its parent YAC and screening with DXS573 isolated three YACs that were highly unstable. A single YAC, yWXD6047, contained DXS573 and extended to include markers distal to EC1908R, but was deleted for EC1908R, and the adjacent markers SYP and DXS1007E. Cosmid library screening with DXS573 isolated 11 cosmids, one of which, c154, is the only link between EC1908R and DXS573. The region between EC1908R and DXS573 appears to be uncloneable in the yeast strain AB1380.

Though the region between ZNF21 and OATL1 (ornithine aminotransferase-like-1) appears as a weak link in Figure 18, this

region has been detailed in other studies (Coleman *et al.*, 1994; Hagemann *et al.*, 1994; Knight *et al.*, 1994; Kwan *et al.*, 1995a).

Size of Contig, DNA Marker Density, and Coverage:

The size of the contig was estimated to be 2 Mb taking into account the extent of overlap between adjacent non-chimeric, intact YACs and cosmids. In the region between OATL1 and DXS573, a physical distance estimated to be 675 kb, there are 32 DNA markers which gives an average density of one marker every 20 kb. The region between DXS573 and DXS255 is estimated to be 750 kb and contains 12 STSs which gives an average density of one STS every 60 kb. Across the 2 Mb of this contig there is an average of one STS every 50 kb. Whether the YAC clone coverage is estimated from the sum of the nonchimeric portions of YAC clones, or by calculating the average number of YACs hit per STS, the set of YAC clones presented in Figure 18 represent five-fold coverage of the region (4.8 and 5.2-fold respectively, by the two different methods). Despite the extensive screening of different libraries and coverage of the region, small gaps in the physical contig cannot be excluded.

Orientation of the Contig on the X chromosome:

The contig was unambiguously oriented on the X chromosome using the panel of hybrids which allowed us to place each of the markers into an interval defined by the various breakpoints (Figure 7). A linkage map of the human X chromosome indicated that DXS1126 is distal to DXS255 (Donnelly *et al.*, 1994), firmly anchoring

DXS255 as the most centromeric marker included in the physical map. Previous use of the SIN176 hybrid for physical mapping of markers on the human X chromosome indicated that OATL1 is distal to SYP, TFE3 and DXS255 (Lafreniere *et al.*, 1991), anchoring OATL1 as the most telomeric marker of this set. In addition, the YAC ICRFy900C0874 is in common with other published more distal contigs (Coleman *et al.*, 1994; Hagemann *et al.*, 1994; Knight *et al.*, 1994; Kwan *et al.*, 1995a) which gives evidence to indicate that it is the most distal YAC in this contig.

Physical Mapping of Genetic Markers:

Sequencing of EC3203L revealed a $(TC)_{11}$ repeat. Analysis of DNA from 50 unrelated, Caucasian females showed this repeat to be polymorphic with allele frequencies of 0.38 (155 bp product), 0.61 (159 bp product), and 0.01 (161 bp product) (data not shown). A 153 bp product was generated from the parent YAC, yWXD3203, but not found on any of the 100 chromosomes evaluated. Further testing will be required to determine whether this allele occurs as a rare variant in the population or was a mutation incurred during the cloning procedure. This polymorphic marker adds density to the genetic map of this region. Eleven genetic markers, including one RFLP, one VNTR, eight CA repeats, and one TC repeat, are contained within this contig. The order and estimated physical distance separating these markers is predicted to be: Xpter-ZNF21 (CA_n) -400kb- OATL1 (RFLP) -120kb- DXS722 (CA_n) -130kb- DXS6940 -90 kb- (DXS1126 (CA_n), DXS1240 (CA_n)) -100kb- DXS1470 (CA_n)

-80kb- EC3203L (TC_n) -150kb- DXS573 (CA_n) -270kb- DXS1039 (CA_n) -540kb- DXS255 (VNTR)-Xcen.

Physical Mapping of Genes and Expressed Sequences:

Twelve genes or expressed sequences are contained within this contig. The order and estimated physical distance separating these markers is predicted to be: Xpter-ZNF21 -420kb- DXS7465E (MG66) -60kb- DXS7927E (EC5986R, MG81) -60kb- WASP, DXS1011E, DXS7467E(MG21) -60kb- DXS7466E (MG44) -60kb- GATA1 -150kb- DXS7469E (Xp664) -120kb- TFE3 -70kb- SYP (DXS1007E)-Xcen.

Between DXS722 and DXS255 there are five ESTs, DXS1011E, DXS7467E, DXS7466E, DXS7469E, and DXS1007E, and four genes, WASP, GATA1, TFE3, and SYP. Of the four genes, SYP was a likely candidate for CSNB1 because it is expressed in both the brain and the retina and is involved in neural transmission. However, this gene has been scanned almost entirely for mutations in Family P060 and none have been detected (Dr. N.T. Bech-Hansen, personal communication). The EST DXS1007E is part of the 5' untranslated region of SYP and so not part of candidate gene that hasn't already been scanned. The ESTs DXS7467E and DXS7466E were both derived from a retinal cDNA library and are therefore possible candidates for CSNB1. However, because they lie between DXS722 and DXS1126, a region that could potentially be excluded as containing the candidate gene if the recombination event in D19 was refined further, these ESTs have not yet been evaluated as possible candidates. To enrich the region between DXS722 and DXS255 for additional transcripts,

YACs and cosmids from the physical contig have been used for direct cDNA selection of a number of cDNA sets.

Isolation of New Genes by Direct cDNA selection

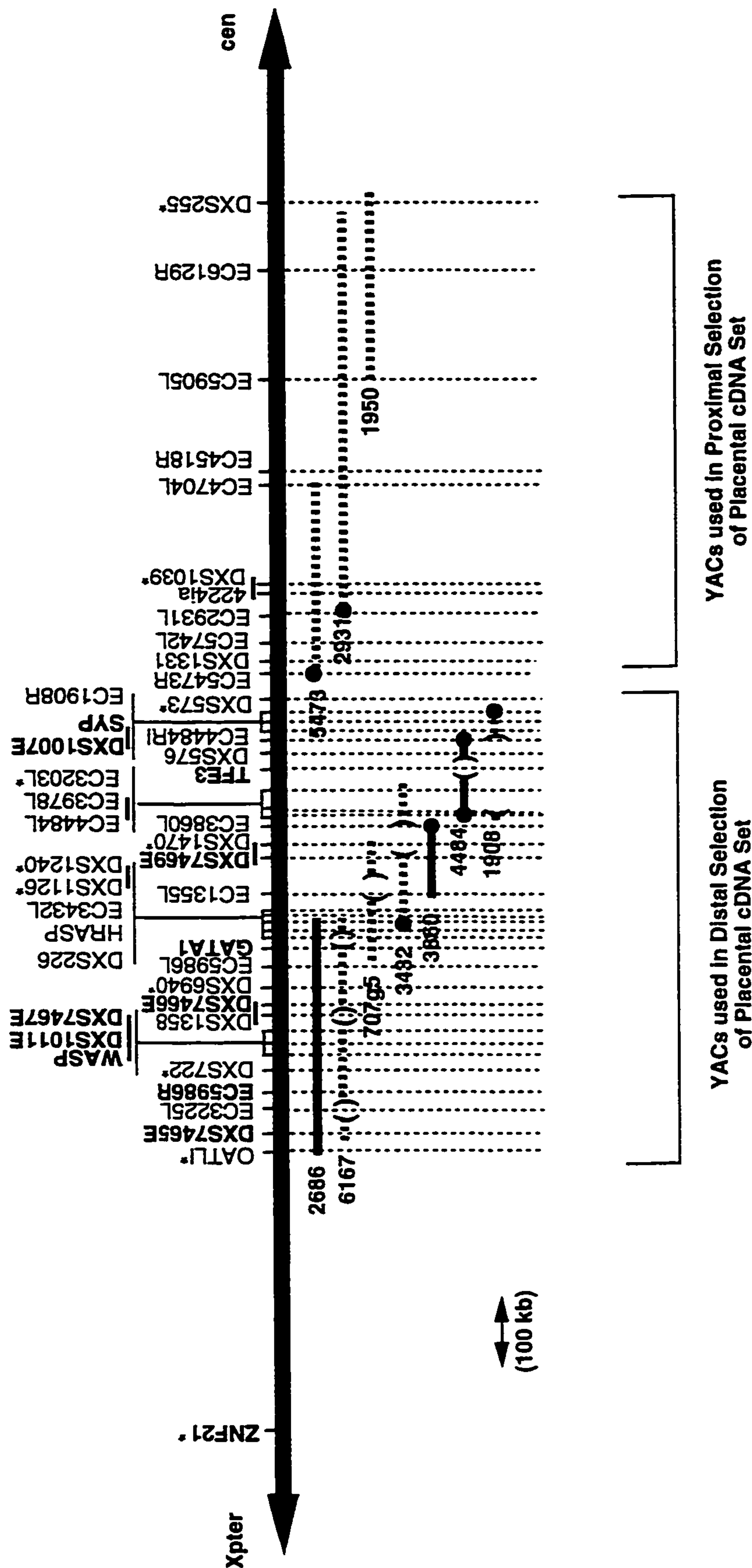
cDNA Sets:

Direct cDNA selection with four cDNA sources (placenta, retina, fetal brain, and frontal cortex) was used to isolate transcripts from the region between OATL1 and DXS255 in Xp11.23. Six selections were carried out, two with the placental cDNA set and four with pooled retina, fetal brain, and frontal cortex cDNA sets, referred to collectively as JRL. YACs and cosmids isolated and characterized during physical mapping of Xp11.23-p11.22 were used as templates to isolate region-specific cDNAs by direct cDNA selection.

Selection of YAC and Cosmid Templates:

Representative YACs from the region between OATL1 and DXS255, containing most if not all of the minimal genetic region for the CSNB1 locus, were chosen to ensure 100% coverage. For the placental cDNA selection, the region was divided into the distal and proximal portions. The distal region was represented by the YACs yWXD2686, yWXD6167, 707g5, yWXD3432, yWXD3860, yWXD4484, and yWXD1908 (Figure 19). The YACs yWXD6167 and 707g5 were not included on the physical contig shown in Figure 18. yWXD6167 was chosen for inclusion in this selection to duplicate coverage

Figure 19. YACs used in the 'distal' and 'proximal' placental direct cDNA selections. The YACs yWXD2686, yWXD6167, 707g5, yWXD3432, yWXD3860, yWXD4484, and yWXD1908 were part of the distal selection and the YACs yWXD5473, yWXD2931, and yWXD1950 were part of the proximal selection. The distances between markers in the central-proximal region of the Figure have been altered from that in Figure 18 due to additional long-range physical mapping information within Xp11.23-p11.22 that has become available recently (Schindelhauer *et al.*, 1996).

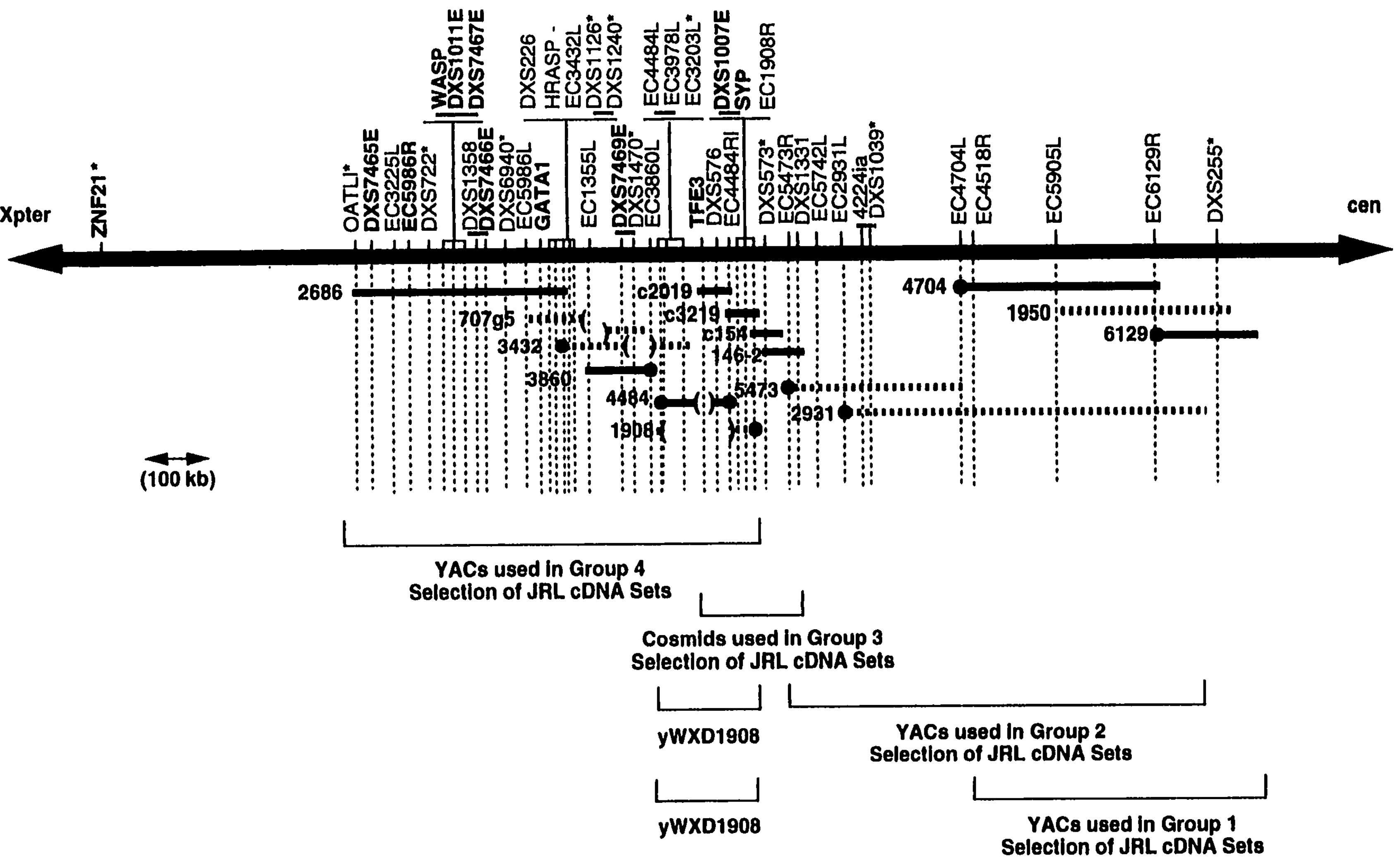


cen
Xpter
DXS255*
EC6129R
EC5905L
EC4518R
EC4704L
DXS1039*
EC2931L
EC5742L
DXS1331
EC5473R
DXS573*
EC1908R
SYP
DXS1007E
DXS576
TFE3
EC3203L*
EC3978L
EC4484L
EC3860L
DXS1470*
DXS7469E
DXS1240*
DXS1126*
EC1355L
EC3432L
HRASP
GATA1
DXS226
EC5986L
DXS6940*
DXS7466E
DXS1358
DXS7467E
DXS1011E
WASP
DXS722*
EC5986R
EC3225L
DXS7465E
OATL1*
ZNF21*
1950
2931
5473
4484
1908
3860
3432
2686
6167
(100 kb)
YACs used in Proximal Selection of Placental cDNA Set
YACs used in Distal Selection of Placental cDNA Set

already present due to yWXD2686. 707g5 is a CEPH YAC (Albertsen *et al.*, 1990) that cosegregates the markers DXS7469E and DXS1470 and was used in this selection to add coverage to this unstable region. For the placental selection, the proximal region of the contig was represented by the YACs yWXD5473, yWXD2931, and yWXD1950 (Figure 19). The YAC yWXD2686 contains the GATA1 gene which is expressed in the placenta, as determined by testing the GATA1 primers on the placenta cDNA set (data not shown). GATA1 was therefore used as the positive reporter gene for the distal placental selection. The placental distal and proximal selections were performed simultaneously, and it was assumed that if the distal selection showed enrichment for GATA1, then the proximal selection also produced enrichment for genes specific to this region.

For the JRL selection, the contig was divided into four groups. The most distal portion of the region was represented by the YACs yWXD2686, 707g5, yWXD3432, yWXD3860, yWXD4484, and yWXD1908 (Group 4) (Figure 20). The central portion of the region was represented by the cosmids c146, c154, c3219, and c2019 (Group 3). c146 and c2019 were not included on the physical contig shown in Figure 18, but are shown along with the markers they contain in Figure 20. The proximal portion of the region was represented by Groups 1 and 2, with Group 1 being the most proximal. Group 2 contained the YACs yWXD5473, yWXD2931, and yWXD1908 and Group 1 contained the YACs yWXD4704, yWXD6129, yWXD1950, and yWXD1908 (Figure 20). The synaptophysin gene, SYP, was found to be expressed in the frontal cortex, fetal brain, and

Figure 20. YACs used in the JRL-Groups 1, 2, 3, and 4 selections. The YACs yWXD2686, 707g5, yWXD3432, yWXD3860, yWXD4484, and yWXD1908 were part of Group 4, cosmids c2019, c3219, c154, and 146-2 were part of Group 3, YACs yWXD5473, yWXD2931, and yWXD1908 were part of Group 2, and YACs yWXD4704, yWXD1950, and yWXD6129 were part of Group 1. As in Figure 19, the distances between markers in the central-proximal region have been altered from that in Figure 18 due to additional long-range physical mapping information within Xp11.23-p11.22 that has become available recently (Schindelhauer *et al.*, 1996).



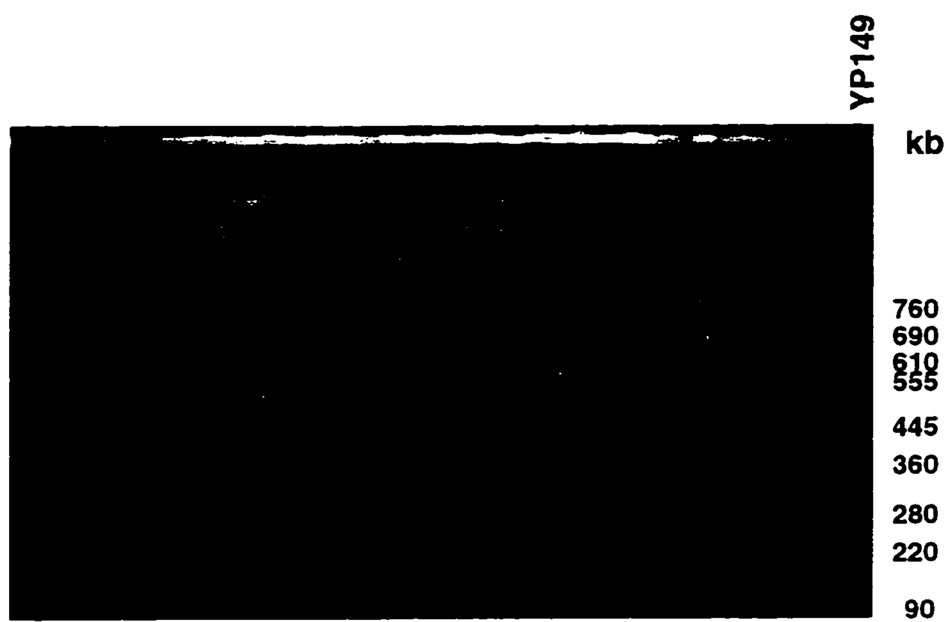
Xpter
cen
ZNF21*
OATLI*
DXS7465E
EC3225L
EC5986R
DXS722*
WASP
DXS1011E
DXS7467E
DXS1358
DXS7466E
DXS6940*
EC5986L
GATA1
DXS226
HRASP -
EC3432L
DXS1126*
DXS1240*
EC1355L
DXS7469E
DXS1470*
EC3860L
EC4484L
EC3978L
EC3203L*
TFE3
DXS576
EC4484Rl
DXS1007E
SYP
EC1908R
DXS573*
EC5473R
DXS1331
EC5742L
EC2931L
4224ia
DXS1039*
EC4704L
EC4518R
EC5905L
EC6129R
DXS255*
2686
707g5
3432
3860
4484
1908
c2019
c3219
c154
146-2
5473
2931
4704
1950
6129
(100 kb)
YACs used in Group 4
Selection of JRL cDNA Sets
Cosmids used in Group 3
Selection of JRL cDNA Sets
yWXD1908
YACs used in Group 2
Selection of JRL cDNA Sets
yWXD1908
YACs used in Group 1
Selection of JRL cDNA Sets

retinal cDNA sets by PCR with primers designed to amplify exon 5 (data not shown). The YAC yWXD1908 contains the SYP gene and was included in the Group 1, 2, and 4 selections so that enrichment for these selections could be evaluated using this gene. The cosmid c3219, used in the Group 3 selection, contains the SYP gene and therefore SYP could also be used to evaluate enrichment in this selection.

Direct cDNA Selection:

For the isolation of DNA from selected YACs, yeast chromosomes were prepared and YACs were resolved by pulsed-field gel electrophoresis using a preparative comb. Figure 21 shows the separation of the YAC 707g5 from the yeast chromosomes. However, due to its size (675 kb), 707g5 comigrated with the 690 kb yeast chromosome under the conditions used and both bands had to be extracted. Comigration with yeast chromosomes was the case for about half of the YACs. Gel excised YAC DNA was purified using Prep-a-Gene and quantitated using a spectrophotometer. Due to the low levels of recovery using this technique, the presence of YAC DNA was confirmed by PCR using a radioactive label and markers known to be contained within the particular YACs (data not shown). YAC DNA was present in every case. Cosmid DNA was considerably easier to purify and was done using the rapid plasmid small-scale preparation. Because of the large yield using this technique, cosmid DNA was easily quantitated using spectrophotometry.

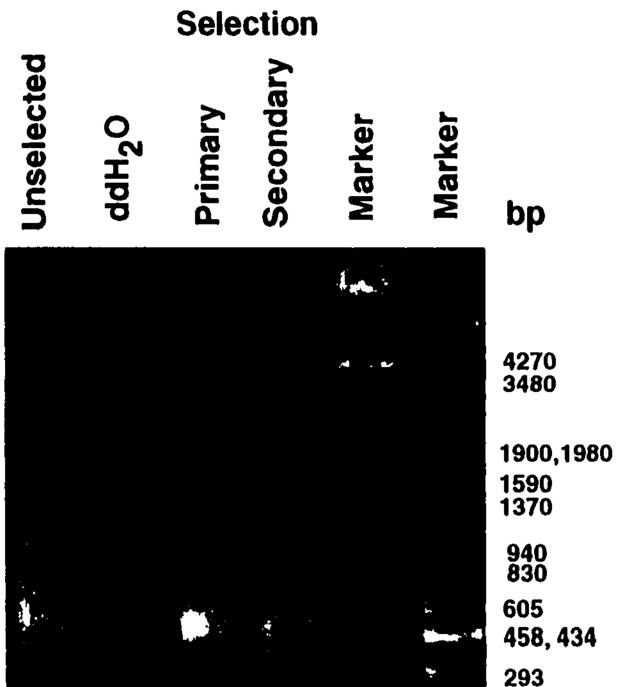
Figure 21. Ethidium bromide stained gel showing separation of 707g5 from the other yeast chromosomes using pulsed-field gel electrophoresis. The gel was ramped from 15 s to 75 s over a 28 hour period to ensure maximum separation of this 675 kb YAC from yeast chromosomal DNA in this size range. 707g5 is running with the 690 kb yeast chromosome and the brackets indicate the part of the gel that was excised for DNA purification. The marker is yeast strain YP149 and the sizes of the relevant bands are shown.


YP149
kb
760
690
610
555
445
360
280
220
90

YAC and cosmid DNA was biotinylated using nick translation and biotin-16-dUTP and the efficiency monitored using [α^{32}-P] dCTP as a tracer during the reaction. The efficiency of biotin incorporation relative to the labelled substrate was over 90% for YACs and slightly less than 75% for cosmids. cDNAs were blocked and primary and secondary selections carried out as described in the Materials and Methods. Progress of the selection of the fetal brain, frontal cortex, and retina (JRL) cDNA set was monitored with the reporter gene, synaptophysin, while progress of the distal placental selection was monitored with the reporter gene, GATA1. When these probes were hybridized to the respective original cDNA sources, the primary selected cDNAs, and the secondary selection products, enrichment was observed. Figure 22 shows the enrichment of the distal placental selection. By EtBr staining the reduction in the complexity of the cDNA material is evident after the primary and secondary selection. Southern hybridization with GATA1 generates signals only in the primary and secondary selected material and not in the unselected material, indicating enrichment has occurred for this gene. The molecular lengths of the secondary selected cDNAs ranged from 100-700 bp in both selections with an average insert length of 300 bp for the placental cDNAs and 350 bp for the JRL cDNAs.

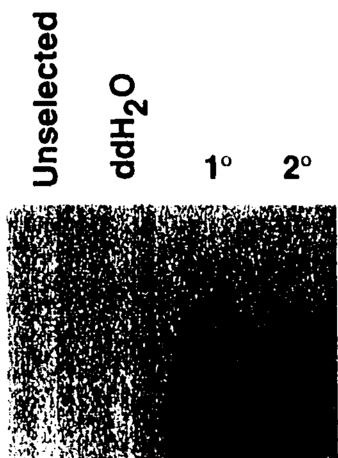
The selected placental cDNAs were amplified by PCR using the modified OLIGO1 primer and cloned using the Cloneamp™ pUC19 system for the rapid cloning of amplification products. An attempt was made to clone the selected JRL products in the same way but PCR with the modified RXG primer formed large amounts of primer

Figure 22. Unselected, and primary and secondary enrichment products for the placental cDNA selection using the distal YACs. (A) Selection products were electrophoresed on a 1.2% agarose gel and stained with EtBr. The reduction in complexity of the cDNA set is evident after the primary selection. The DNA was transferred to nylon membrane and then probed with the gene GATA1. Hybridization of GATA1 is seen only in the lanes containing the primary and secondary selected material after a one hour exposure (B) and a 48 hour exposure (C), indicating that enrichment has occurred for GATA1 during this selection.



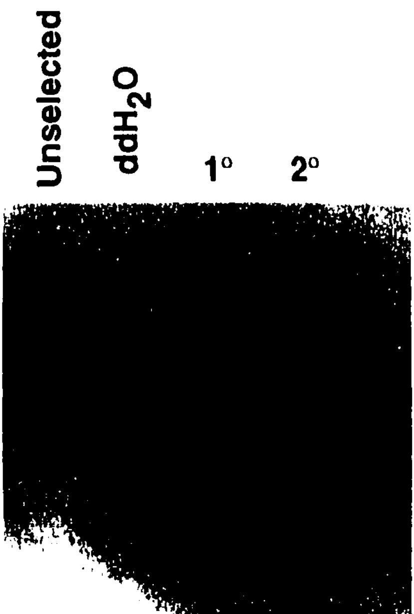
Placental - Distal YACs

A.



1 Hour exposure

B.



48 Hour exposure

C.

dimer that would have been cloneable using this system and given rise to clones containing only primer as insert. After a failed attempt with TA cloning, the JRL selected products were cut with *Eco*RI and cloned into this site in pBluescript. The transformations for both the placental and JRL selections were plated and all the colonies were either worked up directly from the plate to check for insert or inoculated into a well of a 96-well plate for analysis at a later time. The number of clones generated from the transformations of each selection are indicated in Table 11. The name of all selection products is based on the cDNA set, plac (placental) or JRL (frontal cortex, fetal brain, retina), and the number indicates the order in which the clone was inoculated from the transformation plate or its location in the 96-well plate (letter followed by a number).

Table 11. Clones Generated from cDNA Selections

Selection Designation[a]	No. of Clones
Placental-proximal	24
Placental-distal	36
JRL-Group 1	24
JRL-Group 2	20
JRL-Group 3[b]	206
JRL-Group 4	27

[a] JRL refers to the frontal cortex, fetal brain, retina selection.

[b] The genomic clones in Group 3 consisted only of cosmids.

Characterization of Selection Products and Development of New ESTs:

Insert fragments for all clones isolated from the selections were isolated from DNA recovered from the rapid small-scale plasmid procedure and excised using the appropriate enzymes. The insert bands were isolated in low melting point gel and the DNA extracted using Gelase™. This was the case for all selections except for JRL-Group 3. Due to the large number of clones isolated in this selection compared to the other selections and the relatively small genomic region covered by the cosmids used, only 14 clones were characterized. Excision and electrophoresis of 13 of these 14 clones by *Eco*RI is shown in Figure 23. In total 60 clones from the placental selections and 85 clones from the JRL selections were characterized.

Mapping panels were created for each of the selections for use in preliminary characterization of the inserts. Each mapping panel contained human genomic DNA, as well as DNA from the hybrids GM06318B (X only) and H99 (Xp11), Chinese hamster ovary (CHO), and the YACs or cosmids used in the selection. For the selections from the proximal region, JRL-Groups 1 and 2 and proximal placental, the mapping panels also contained DNA from the hybrids H151 (breakpoint in Xp11.23) and SIN176 (breakpoint in Xp11.23) to confirm and refine the location of cDNA clones. DNA was cut with EcoRI, electrophoresed in 0.8% agarose gels, and transferred to membrane. Inserts were radiolabelled and hybridized to the corresponding mapping panels. At lease four copies of each mapping

Figure 23. JRL-Group 3 clones cut with *Eco*RI, electrophoresed on a 1% low melting point agarose gel, and visualized by EtBr staining. DNA size markers are pBluescript cut with *Hae*III on the left and bacteriophage DNA cut with *Hind*III and *Eco*RI on the right. The size of the relevant bands are indicated. The 2.9 kb pBluescript vector is evident in every lane and the inserts range from 290 to 600 bp in size.

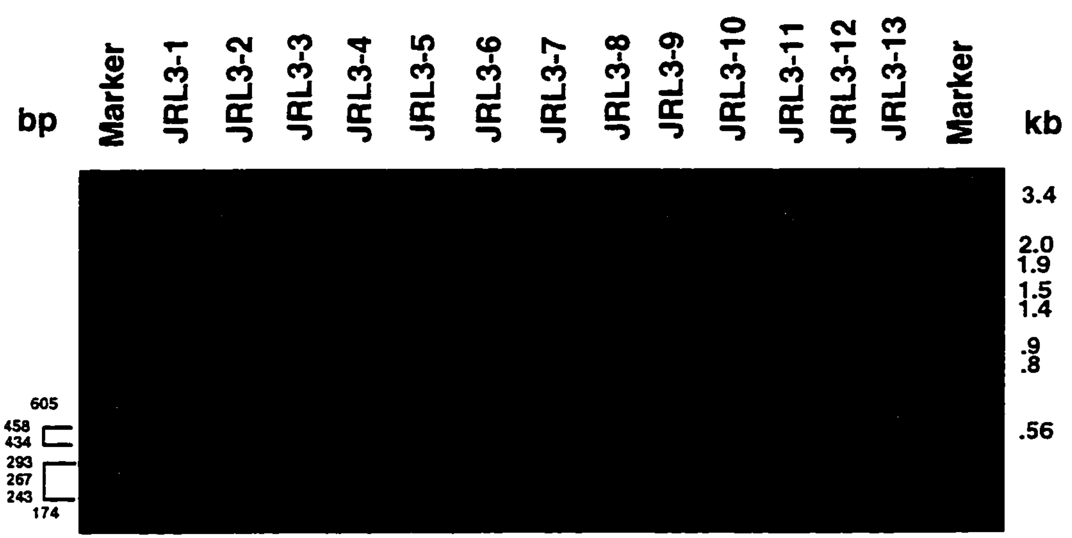
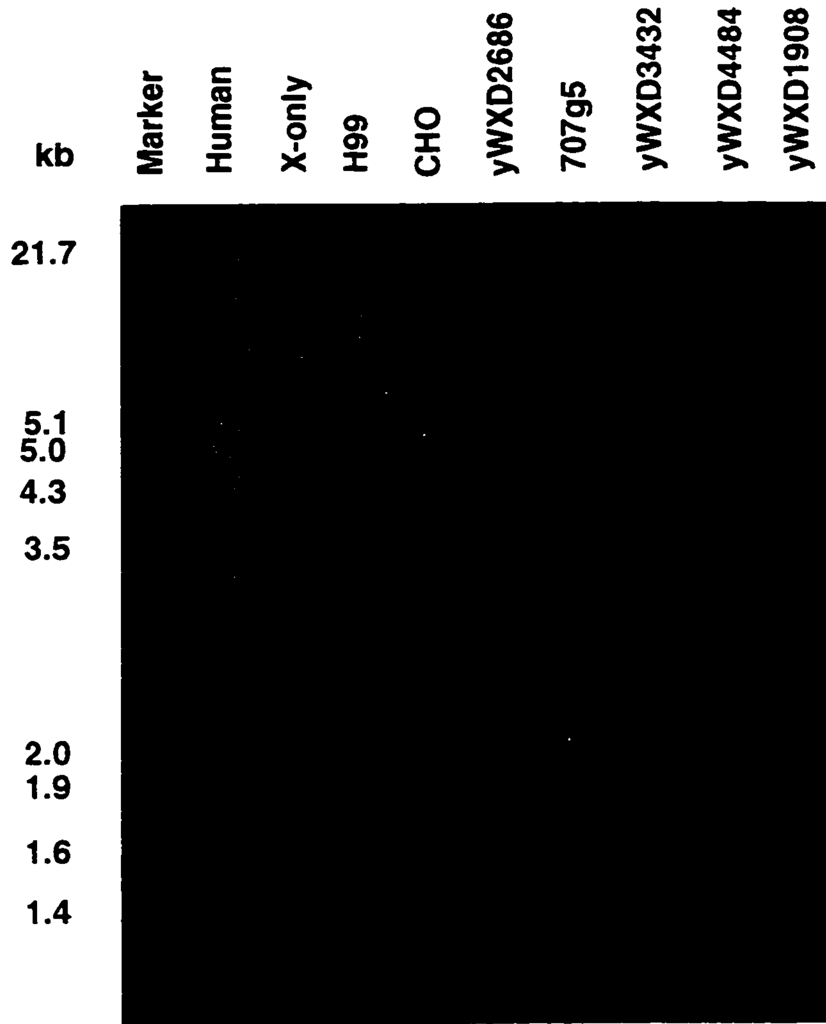

bp
Marker
JRL3-1
JRL3-2
JRL3-3
JRL3-4
JRL3-5
JRL3-6
JRL3-7
JRL3-8
JRL3-9
JRL3-10
JRL3-11
JRL3-12
JRL3-13
Marker
kb
3.4
2.0
1.9
1.5
1.4
.9
.8
.56
605
458
434
293
267
243
174

Figure 24. Ethidium bromide stained 0.8% agarose gel showing *Eco*RI digested DNA from human, CHO, hybrid, and YAC sources (indicated along the top). To create a mapping filter for JRL-Group 4 cDNA clones, the DNA was transferred from the gel to nylon membrane. The DNA marker is bacteriophage lambda DNA cut with *Hind*III and *Eco*RI and the sizes of the bands are indicated.


kb
Marker
Human
X-only
H99
CHO
yWXD2686
707g5
yWXD3432
yWXD4484
yWXD1908
21.7
5.1
5.0
4.3
3.5
2.0
1.9
1.6
1.4

panel were made to ensure rapid localization of the cDNA clones. Figure 24 is an example of a mapping panel for JRL-Group 4 clones.

Proximal Placental cDNA Selection: Of the 24 clones characterized from this selection, seven had no insert, leaving 17 clones. Because of the presence of the chimeric YAC yWXD2931 in this selection, seven of the 17 clones did not map to the X chromosome. An additional seven clones appeared to be contaminants brought through the selection, leaving only three clones that mapped to this region of the X chromosome, placB3, placB11, and placB12. The three region specific clones were sequenced and based on this primary single-pass sequence they did not appear to overlap with each other. Primers were designed to amplify each of these three cDNA clones by PCR. All primer sets amplified the expected size product (Table 12). The proximal placental cDNA selection procedure was 59% (10/17) efficient when considering the number of isolated cDNAs that map to the YACs used as templates in the selection. However, due to chimerism in YAC yWXD2931, this selection yielded only 18% (3/17) Xp11.23-p11.22 region-specific cDNAs.

Table 12. Newly Developed ESTs from Placental Selection of Proximal YACs

Lab Designation	Primer Sequences (5'-3')	Size (bp)
placB3-2[a]	CTCCACCACGATAGACCCAT AGATTCCTGGTAGCTTTGGC	223
placB11	CTTGCACAAACCCATCACTT CTGCTAGATATGAGAACAGG	111
placB12	CCATGCTTGTTCAATCTATC AGGCTTGCGTCTCAAGAACT	110

[a] placB3-2 also has the GDB designation DXS7471E

Distal Placental cDNA Selection: Of the 36 cDNAs characterized from the distal placental selection, eight had no insert, leaving 28 clones. Of these 28 clones, 21 mapped to Xp11.23, while seven clones appeared to be contaminants brought through the selection. The efficiency of this direct cDNA selection was 75% (21/28), and the yield of Xp11.23 region-specific clones was also 75% (21/28). The 21 clones were sequenced and reduced to 14 nonoverlapping sets, a redundancy of 33%. Three of these clones had been previously identified, placD4 was from part of the GATA1 gene, and placC10 and placD12 were part of the ESTs IS3 and IS4, respectively (Derry *et al.*, 1994). To amplify the remaining 11 unique clones by PCR, primers pairs were designed. Ten of the primer pairs (placC1, placC7, placC11, placC12, placD1,11, placD2,7, placD10,E5, placE2, placE3,8,9, and placE11) yielded the expected size product and were region specific (Table 13). The primers designed to amplify the placental clone, placC5,8,E10 failed to give any product and were

likely spanning an intron that was too large to be amplified under the PCR conditions used. Redesigning the primers for this clone may add an additional EST based PCR marker to this region.

Table 13. Newly Developed ESTs from Placental Selection of Distal YACs

Lab Designation	Primer Sequences (5'-3')	Size (bp)
placC1	AACAATCCCAGCAGTCACCA CGAGGCTGACACAAGATTCA	80
placC7	ATGGAGTTGGATTCCTGTGG CCACTCATCTAAGGTCTGAC	156
placC11	CACAGGGAAGCTGACCAAAG TGTAGAGCAGCAACTTGTTG	146
placC12	ACAGCCTTCTAGTCTAGCCC GCCACATGCAGCTCAGGATA	108
placD1,11	AGTGATAAGGAGGTGTGCGG CACATCCTGACACACACTTC	200
placD2,7	AACAAGCAGATGAGAGTGGA CCCTCCTTTCATCTTCCTCC	94
placD10,E5	GAAGTCCGCGTAGTATCTTT TCCACCTCTTTAACGAGCCT	124
placE2[a]	GAAGGGAACCAAGGTACCAG GCTGACCTTTGACTTCCATC	133
placE3,8,9	GGTATTTAGGGATAGATATG AAACTGTCTTATTGAAGCCC	137
placE11	CCTCACACATACCAAGCCTT TTAAGCAGAGATCAGGAGAT	122

[a] placE2 also has the GDB designation DXS7472E

JRL-Group 1 cDNA Selection: Of the 24 cDNAs characterized from this selection, ten mapped to Xp11.23-p11.22, six did not map on the X chromosome, and eight were repeats. This indicates an efficiency of 67% (16/24) for this selection, however due to chimerism in the YACs used as template for the selection only 42% (10/24) of the clones were Xp11.23-p11.22 region-specific. The region-specific cDNA clones were sequenced and reduced to nine nonoverlapping sets, a redundancy of 10%. Primers were designed to amplify the nine unique clones by PCR (JRL1-B1, JRL1-B4, JRL1-B6, JRL1-B11, JRL1-C3, JRL1-C4,5, JRL1-C7, JRL1-C9, JRL1-C10) and all yielded the expected size product size (Table 14). The clones JRL1-C3 and JRL1-C9 did not map to a specific YAC but did map to the Xp11.23 region when used as labelled probes on the mapping panels. These clones were sequenced, primers designed, and these were tested on YAC templates from the Group 1 region. These ESTs failed to amplify using any of the templates from Group 1 and were also negative on H151 and SIN176, indicating that they may be located in the Group 4 region. The ESTs were tested in this region using YAC and cosmid templates and found to map here. This could be explained by a small amount of yWXD2686 contaminating the Group 1 YACs leading to the isolation of JRL1-C3 and JRL1-C9 from a region other than that represented by Group 1.

Table 14. Newly Developed ESTs from JRL-Group 1 Selection

Lab Designation	Primer Sequences (5'-3')	Size (bp)
JRL1-B1	GTTCACCCATGCCTGACCTT TATTTCTCCCTGTCCCTAAA	215
JRL1-B4	GTCAGGCACTGAAACCCAAA AGGTTTCTGAAGCCACTCAT	101
JRL1-B6	AAACCCAAAGGAAACTCAGC CGAATTCTGAGTATCGTGGG	115
JRL1-B11	GTGAACCAATAAGAGTCAGA TGTCCATCTTGAAGCACCAT	161
JRL 1-C3	AGACATCTTCATTTCCGAGA TACTGACAGTCTGTACCAAG	195
JRL1-C4,5	CTTATGCAAAGACCTCACAA GAGTCTCTGCAAGCCTGTGT	94
JRL1-C7	CTTCTAAGACCCACCTCCTC AAGAATGGTGCTTCAAGATG	188
JRL 1-C9	CCACATTGGGTCAGAACTTT AAGACATCTTCATTTCCGAG	144
JRL1-C10	GGGCTGGCTTTCTGTCTGTC GTTTCAACACAATTTCTGCC	142

JRL-Group 2 cDNA Selection: Of the 20 cDNAs characterized from this selection, none mapped to Xp11.23-p11.22 and three clones appeared to contain low level repeats. While the efficiency of the selection was 85% (17/20), due to the use of the chimeric YAC yWXD2931, all of these clones originate from a chromosome other than the X and therefore 0% (0/20) of the clones were Xp11.23-p11.22 region-specific. Figure 25 is an example of the mapping of a cDNA that originates from a chromosomal region homologous to the chimeric portion of yWXD2931. Signals are seen only in the human

genomic DNA lane and the yWXD2931 lane but not on the X-only hybrid or on hybrids H99 or H151. No clones from this selection were characterized further.

JRL-Group 3 cDNA Selection: Of the 14 cDNAs characterized from this selection, 11 mapped to Xp11.23, while three clones appeared to contain repeat sequences. This indicates an efficiency of 79% (11/14) for this selection, and the yield of Xp11.23 region-specific clones was also 79% (11/14).. The eleven clones were sequenced and reduced to 10 nonoverlapping sets, a redundancy of 10%. One of the clones was part of the SYP gene and two were different parts of the TFE3 gene, and these three clones were not characterized further. To amplify the remaining seven unique clones by PCR, primer pairs were designed, all of which (JRL3-1, JRL3-4, JRL3-8, JRL3-10, JRL3-12, JRL3-13, and JRL3-14) yielded the expected size product (Table 15).

Figure 25. Mapping of JRL2-D3 cDNA clone on a mapping panel for the JRL-Group 2 selection. JRL2-D3 is positive only on human and yWXD2931 indicating that it is from a chromosome other than the X which presumably corresponds to the chimeric portion of yWXD2931.

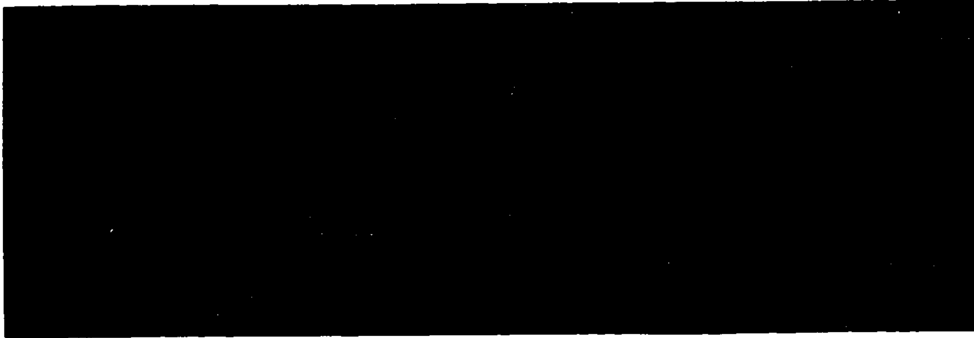

Human
X-only
H99
H151
CHO
yWXD2931
yWXD5473
yWXD1908

Table 15. Newly Developed ESTs from JRL-Group 3 Selection

Lab Designation	Primer Sequences (5'-3')	Size (bp)
JRL3-1	GTCCTCTAGAAAGCGTGGCA ACTGGGCTATACACTGATAT	147
JRL3-4	AAACTCAAACCCAGATGCAA TCACCTACTAAGCAGTTACA	176
JRL3-8	CGCCTACACTGTGGTTCACA CTTATGCAAAGACCTCACAA	124
JRL3-10	AAGAATCAGGACTAGCTGCT CACCGCCTGTGATCATTTAG	129
JRL3-12	GATGCCTGGCAAGTTCTAGA GAGCACCTACTATGTACCAG	129
JRL3-13	GTCAGGCACTGAAACCCAAA TGGCTCAGGGTAACTGCAAT	139
JRL3-14	AATGCTAGGGACATTTATTT TGCTCAGTATCTTCATCTGG	255

JRL-Group 4 cDNA Selection: Of the 27 cDNAs characterized from this selection, 22 mapped to Xp11.23, while four clones appeared to be contaminants brought through the selection. One clone did not map to the X chromosome though it did map to yWXD1908. Based on this information, the efficiency of this selection was 85% (23/27) but, due to chimerism in yWXD1908, only 81% (22/27) of clones were Xp11.23 region-specific. The mapping of one of the cDNA clones from this group, JRL4-A1, on hybrid and YAC templates by Southern hybridization is shown in Figure 26. The 22 clones were sequenced and reduced to 16 nonoverlapping sets, a redundancy of 27%. One clone, JRL4-8, was sequenced with every possible pBluescript

Figure 26. Autoradiogram of JRL4-A1 insert hybridized to a JRL-Group 4 mapping panel containing DNA from critical hybrids and the YACs used in the selection. The names of the hybrids and YACs are indicated along the top. Hybridization of JRL4-A1 to human, X-only (GM06318B), and H99 verifies that this clone maps to Xp11 and hybridization to the YAC 707g5 places it in Xp11.23.

Human
X-only
H99
CHO
yWXD2686
707g5
yWXD3432
yWXD4484
yWXD1908

sequencing primer but no sequence was generated. It would appear that this clone underwent an undefined vector mutation which resulted in loss of the sequencing primer annealing sites. The clone JRL4-B1 was part of the gene SYP and was not characterized further. To amplify the remaining 15 unique clones by PCR, primers pairs were designed, all of which (JRL4-1, JRL4-2, JRL4-3, JRL4-4, JRL4-5, JRL4-9, JRL4-A1, JRL4-A3, JRL4-A4, JRL4-A6,9, JRL4-A7,8,12,B-2,4-7, JRL4-A10, JRL4-B4, JRL4-B5, and JRL4-B6) yielded the expected size product with the exception of JRL4-A1 which crosses and intron and gives a larger product (Table 16). The primer pair location in the JRL4-A1 sequence is shown in Figure 27.

Verification of Primer Pairs as Region-Specific: All primer pairs were region-specific as tested by PCR amplification on the templates used for the construction of the mapping panels. Figure 28 is an example of this preliminary mapping by PCR with the EST JRL4-A1. In a number of instances the primer pair for a particular EST was also tested on the entire set of hybrids in the hybrid mapping panel, as shown for the EST JRL4-A1 in Figure 29.

Table 16. Newly Developed ESTs from JRL-Group 4 Selection

Lab Designation	Primer Sequences (5'-3')	Size (bp)
JRL4-1	CCACCACTCACAAGCTGTGT	173
	TTCAGTGGTTCCAGAGGTTT	
JRL4-2	GTTCTTCAGAAGTCAGCAAA	169
	CGATAGTTGAGTAGATGTGC	
JRL4-3	TCAGCAGTGGTCTTTGGAAG	176
	GGTTTCACTATGAAGGGAGG	
JRL4-4	ACCAAGACTCTGTCTCCAAA	132
	GACAAAGAGGCTGGGATTGC	
JRL4-5	ATGACAGAGGAAAGGCAAGT	133
	GTGAGTGGCCCTCCTCCTAA	
JRL4-9	GGCTTCTGGTGGAGTTAGGA	89
	CTCACTAGCCAGTCCCAAGG	
JRL4-A1	TTTCTCTCTGTCTACCTTGT	281[a]
	CTGCGGGCTCCCTTACTACTG	
JRL4-A3	AAATCAGATTGGAGATGTCC	98
	CTGATCTGATGACCTATTTG	
JRL4-A4	AAAGGAACAGGGAGGACTTT	105
	TCCTCTACCTAGAACACCCT	
JRL4-A6,9	TTTGGGAGACAGAGTCAGTG	118
-	CACCTCTTTCTGCCCTCTCA	
JRL4-A7,8,12,	GGTTCAGATCCATCCCTTGC	100
B-2,4-7	CCCAATCTAGCGGAGGTAAA	
JRL4-A10	AGAGGAACAAATGTATTTAG	159
	CAGTCCACTGGAACTACTAT	
JRL4-B4	GTGAAGGTGACAGAGGTGAG	156
	TAGTTTGACCAATCTCACGA	
JRL4-B5	GGGATAACAATAGTGCTTAC	161
	GGGTATTGAAGGATTGAATG	
JRL4-B6	AACCACCACTCACAAGCTGT	182
	CAGTTTGATTCAGTGGTTCC	

[a] The primer set JRL4-A1 crosses an intron and amplifies a genomic product of 470 bp.

Figure 27. Sequence of the cDNA clone JRL4-A1 and the location of the left and right primers used to amplify this EST (Table 16). DNA was prepared for sequencing using the rapid small-scale plasmid preparation and sequenced using Sequenase® Version 2. The dideoxy G, A, T, and C lanes are indicated along the top and the termination products have been resolved on a 6% acrylamide gel.

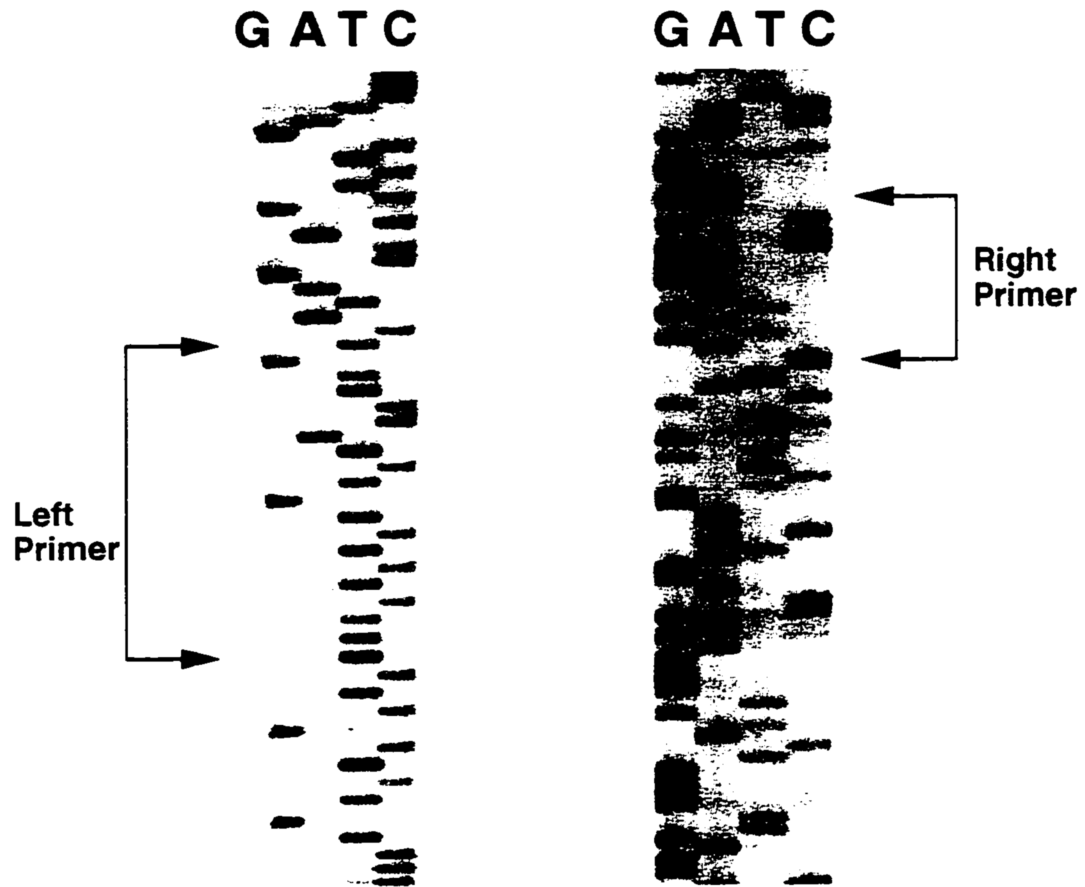
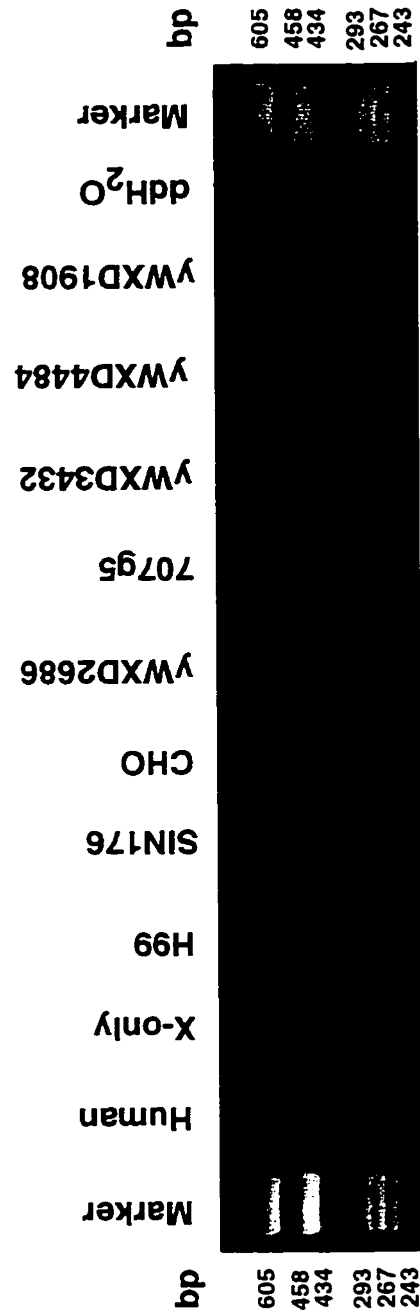
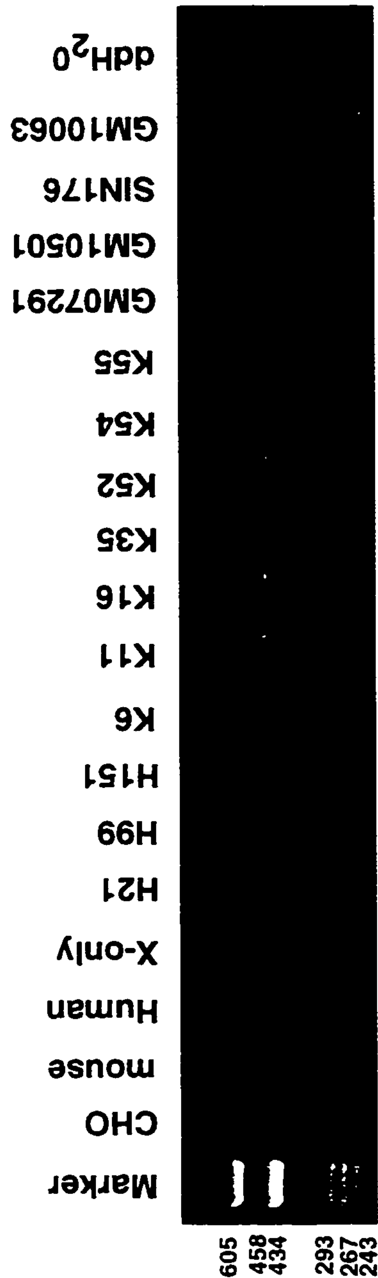

G A T C
G A T C
Right
Primer
Left
Primer

Figure 28. Preliminary mapping of the JRL4-A1 EST by PCR on the hybrids H99, and SIN176 and the YACs yWXD2686, 707g5, yWX3432, yWXD4484, and yWXD1908 used in the Group 4 selection. The positive controls are total human genomic DNA and the X-only hybrid GM06318B, and the negative controls are CHO and ddH_2O. The PCR products were electrophoresed on a 1.5% SEPARIDE/1% agarose gel and visualized by EtBr staining. The DNA marker is pBluescript cut with *Hae*III and the sizes of the bands are indicated.



bp
605
458
434
293
267
243
Marker
ddH_2O
yWXD1908
yWXD4484
yWXD3432
707g5
yWXD2686
CHO
SIN176
H99
X-only
Human
Marker
bp
605
458
434
293
267
243

Figure 29. Mapping of the JRL4-A1 EST by PCR on the entire hybrid mapping panel. The control lanes are CHO, mouse, human genomic DNA and ddH_2O. The PCR products were electrophoresed on a 1.5% SEPARIDE/1% agarose gel and visualized by EtBr staining. Lanes are as indicated across the top of the gel. The DNA marker is pBluescript cut with *Hae*III and the sizes of the bands are indicated.



Marker
CHO
mouse
Human
X-only
H21
H99
H151
K6
K11
K16
K35
K52
K54
K55
GM07291
GM10501
SIN176
GM10063
ddH2O
605
458
434
293
267
243

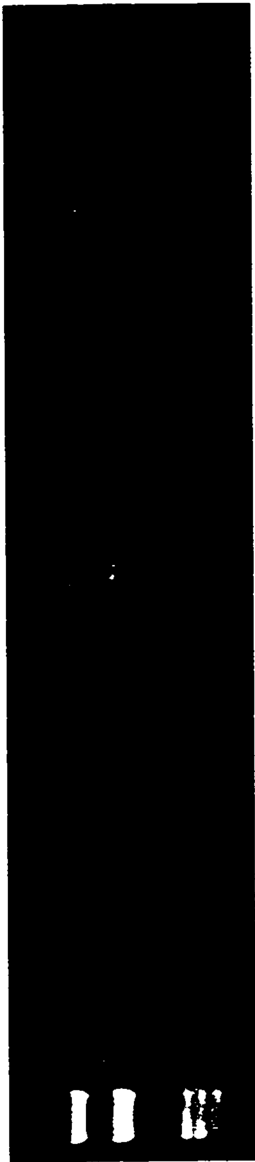
Evaluation of the Direct cDNA Selection Technique and Generation of an Integrated Map:

Since the frontal cortex, fetal brain, and retinal cDNA sets were pooled prior to selection and are all amplified by the RXG primer, it cannot be discerned which of the newly developed ESTs are derived from which tissue. Considering the placental selections together, the efficiency of the selection was 69% but only 53% of the selected cDNA clones were region-specific. Considering the JRL selections together, the efficiency of the selection was 79% but only 50% of the selected clones were region-specific.

One of the concerns in using the direct cDNA selection procedure is the potential problem of isolating nonspecific clones that are selected by repetitive sequences in the genomic template, or in the yeast background that still remains after purification of the YACs. Of the 130 clones analyzed in this study, 31 clones were contaminants. Several of these clones were sequenced and analysis revealed that pBluescript, *Alu*, Line 1, mitochondrial DNA, 18S rRNA, and 28S rRNA contaminants were among them with the repeat contaminant being the most common (data not shown).

In total then, 44 new ESTs were established for the Xp11.23-p11.22 region, 13 derived from the placental selections and 31 derived from the JRL selections. The location of these new ESTs was determined by amplification of relevant YAC and cosmid templates from the detailed physical contig using the EST primer pairs. Figure 30 shows the amplification of these templates for the EST JRL4-A1. The lanes containing the positive human genomic control, 707g5, and

Figure 30. Sublocalization of the cDNA clone JRL4-A1 using the EST primer pair to amplify the YAC and cosmid templates (indicated along the top) from Figure 18 by PCR. Control lanes are human genomic DNA and ddH_2O. DNA marker is pBluescript cut with *Hae*III and the relevant band sizes are indicated.


bp
605
458
434
293
267
243
Marker
Human
yWXD2687
yWXD2686
yWXD3225
yWXD5986
yWXD1355
707g5
yWXD3578
c934
yWXD3860
yWXD3432
yWXD3202
yWXD4057
yWXD4484

bp
605
458
434
293
267
243
Marker
yWXD3978
yWXD3976
yWXD1908
c553
c1918
E011
yWXD3458
yWXD6047
ddH2O

yWXD3860 show the expected product size of 470 bp, while the water control lane is negative. Based on the retention of JRL4-A1 only in these YACs, this clone can be sublocalized on the physical map (Figure 32). The remaining 43 ESTs were sublocalized on the physical map in a similar manner (Figures 31, 32, and 33).

As part of the effort to establish a detailed transcript map in Xp11.23, recently published cDNA and gene sequences (Derry *et al.*, 1994) were integrated into the physical map of this region. PCR primers were designed to amplify the four ESTs (IS2, IS3, IS4, and IS7) and the gene (RBM3) from the sequence published by Derry and coworkers (1994; 1995) (Table 17). Each of these ESTs were shown to be region-specific by analysis on the hybrid mapping panel. Based on this information, the EST IS7 was found to map distal to OATL1 and ZNF21 and was therefore not localized further. The remaining three ESTs and the gene RBM3 were integrated into the map and are shown in Figure 31.

The region between OATL1 and DXS255 is approximately 1700 kb (Boycott *et al.*, 1996; Schindelhauer *et al.*, 1996) and the map now contains 91 DNA markers, giving an average resolution of one marker every 18 kb. Furthermore this group of markers includes five genes, two pseudogenes, 56 ESTs, and nine simple tandem repeats (STRs), generating an integrated genetic, physical, and transcript map of this region.

Figure 31. Integrated genetic, physical and transcript map of the distal portion of Xp11.23 between the markers OATL1 and HRASP. The region contains 31 DNA markers over 400 kb with an average resolution of one marker every 13 kb. YAC end-clones have been represented by their GDB designations (Table 9). Of these 31 DNA markers, three are genes (RBM3, WASP, and GATA1), two are pseudogenes (OATL1 and HRASP), two are polymorphic markers (*), eight are published ESTs (in bold), and 12 are new ESTs developed during this effort (in bold). Of the 12 new ESTs, seven were derived from the JRL cDNA sets and five were derived from the placental cDNA set. JRL4-A7 is the EST JRL4-A7, 8, 12, B-2, 4-7 (Table 16). A line under two or more markers indicates that they could not be ordered relative to each other because they cosegregated in every physical clone tested. There is on average one EST every 20 kb throughout this region.

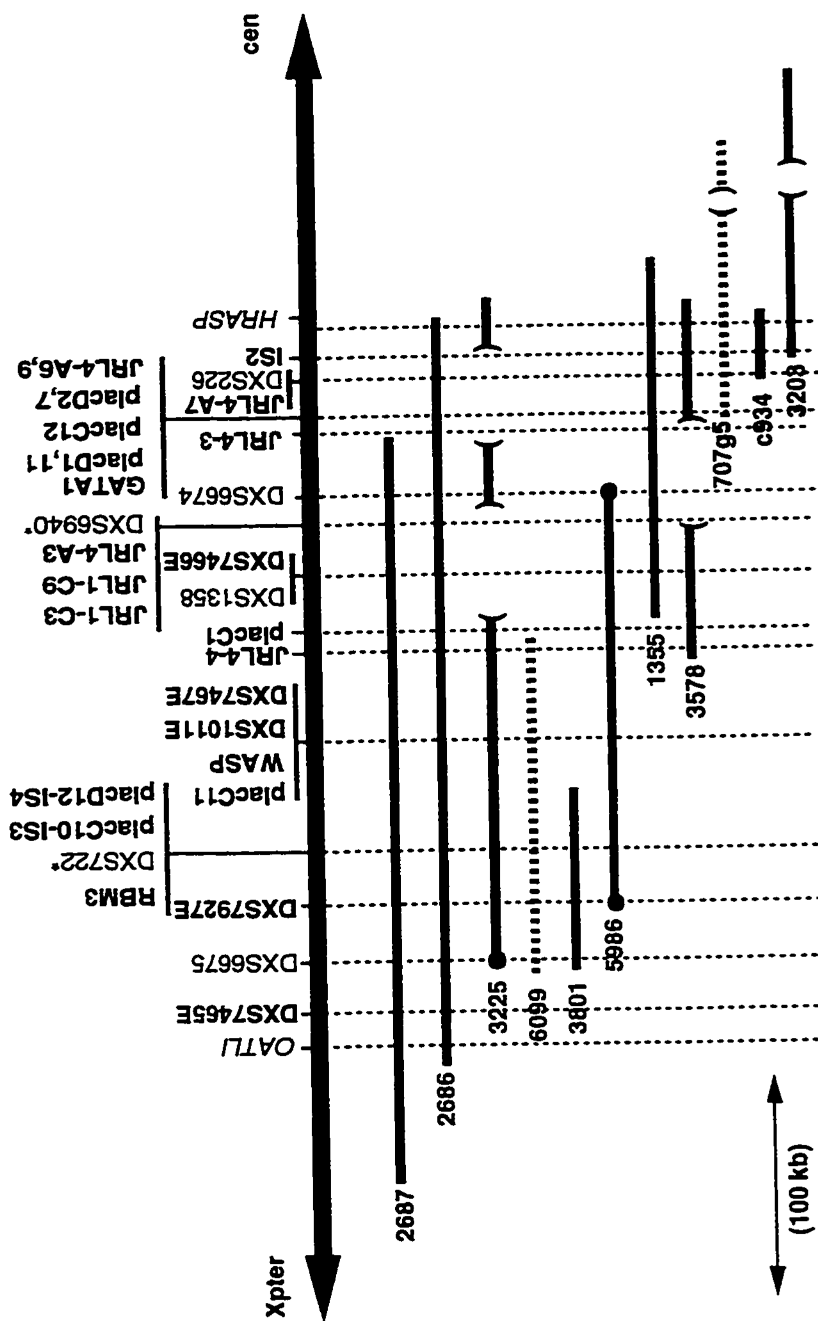

cen
Xpter
HRASP
IS2
DXS226
JRL4-A7
JRL4-3
DXS6674
DXS7466E
DXS1358
placC1
JRL4-4
DXS7467E
DXS1011E
WASP
placC11
DXS7927E
DXS6675
DXS7465E
OATL1
JRL4-A6,9
placD2,7
placC12
placD1,11
GATA1
DXS6940*
JRL4-A3
JRL1-C9
JRL1-C3
placD12-IS4
placC10-IS3
DXS722*
RBM3
2687
2686
3225
6099
3801
5986
1355
3578
707g5
c934
3208
(100 kb)

Figure 32. Integrated genetic, physical and transcript map of the central portion of Xp11.23 between the markers DXS6676 and DXS1007E. The region contains 34 DNA markers over 300 kb with an average resolution of one marker every 9 kb. YAC end-clones have been represented by their GDB designations (Table 9). Of these 34 DNA markers, one is a gene (TFE3), four are polymorphic markers (*), two are published ESTs (in bold), and 20 are new ESTs developed during this effort (in bold). Of the 20 new ESTs, 14 were derived from the JRL cDNA sets and six were derived from the placental cDNA set. A line under two or more markers indicates that they could not be ordered relative to each other because they cosegregated in every physical clone tested. There is on average one EST every 15 kb throughout this region. The order of several of the markers in this region have been altered relative to that in Figure 18 based on additional work with new cosmids and BACs in this region (data not shown), as well as new long-range mapping information within Xp11.23-p11.22 that has become available recently (Schindelhauer *et al.*, 1996).

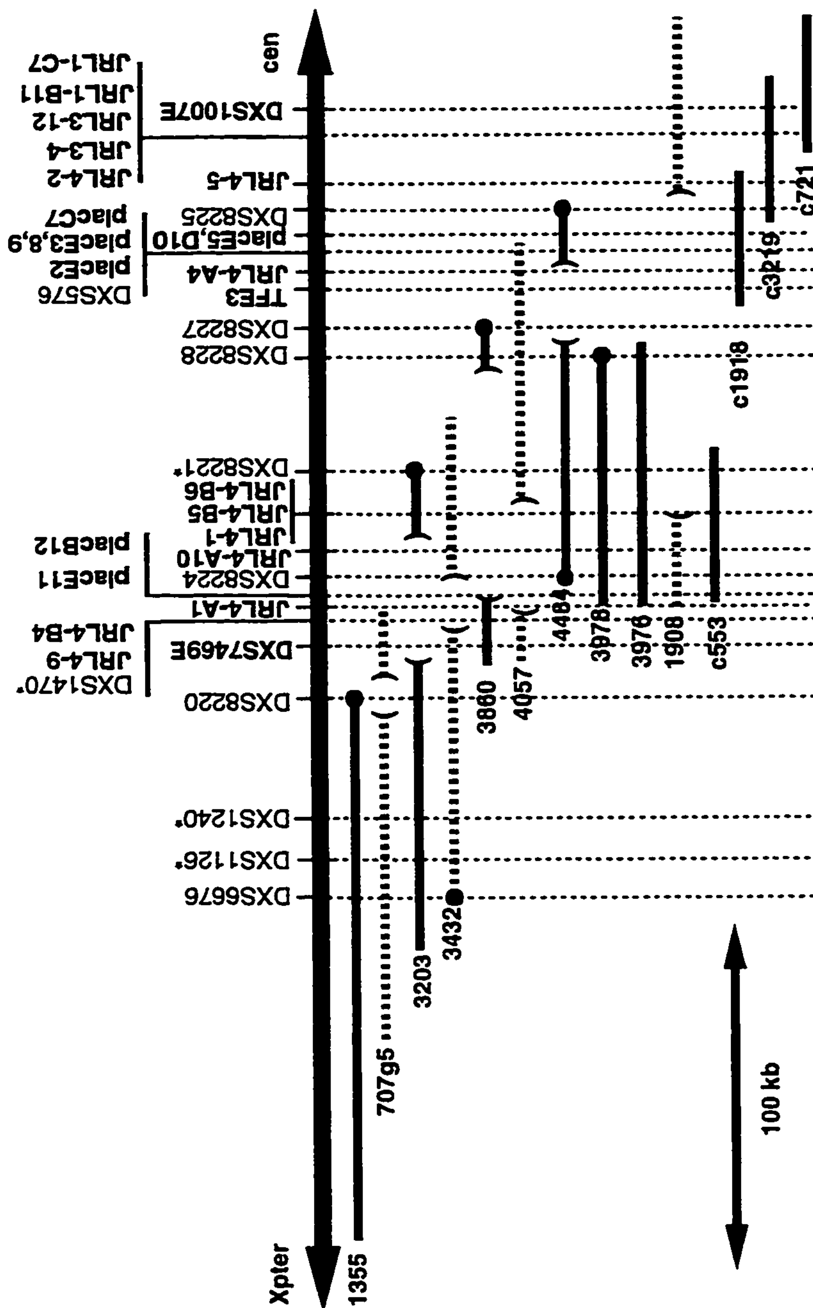

cen
Xpter
DXS1007E
JRL4-5
DXS8225
placE5,D10
JRL4-A4
TFE3
DXS8227
DXS8228
DXS8221*
JRL4-B6
JRL4-B5
JRL4-1
JRL4-A10
DXS8224
JRL4-A1
DXS7469E
DXS8220
DXS1240*
DXS1126*
DXS6676
JRL1-C7
JRL1-B11
JRL3-12
JRL3-4
JRL4-2
placC7
placE3,8,9
placE2
DXS576
placB12
placE11
JRL4-B4
JRL4-9
DXS1470*
1355
707g5
3203
3432
3860
4057
4484
3978
3976
1908
c553
c721
100 kb

Figure 33. Integrated genetic, physical and transcript map of the proximal portion of Xp11.23-p11.22 between the markers SYP and DXS255. The region contains 26 DNA markers over 750 kb with an average resolution of one marker every 30 kb. YAC end-clones have been represented by their GDB designations (Table 9). Of these 26 DNA markers, one is a gene (SYP), three are genetic markers (*), and 12 are new ESTs developed during this effort (in bold). Of the 12 new ESTs, ten were derived from the JRL cDNA sets and two were derived from the placental cDNA set. A line under two or more markers indicates that they could not be ordered relative to each other because they cosegregated in every physical clone tested. The majority of the new ESTs are concentrated in the distal portion of this map. The order of the markers DXS1331 and DXS1522 and the physical distances in the distal portion of this region have been altered relative to that in Figure 18 based on additional work with new cosmids in this region, several of which are shown in context of the map positions of the new ESTs, as well as new long-range mapping information within Xp11.23-p11.22 that has become available recently (Schindelhauer *et al.*, 1996).

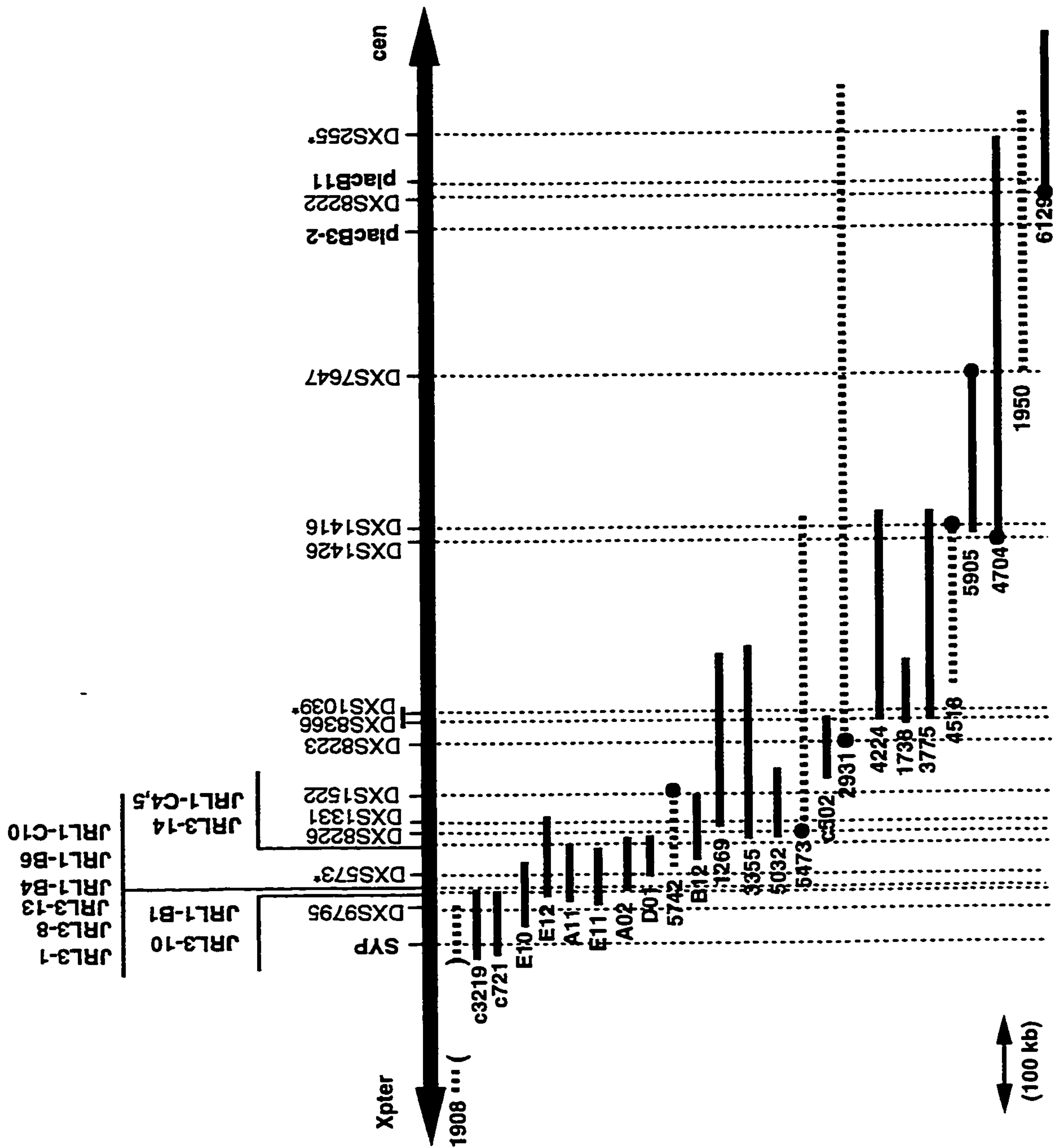

cen
Xpter
DXS255*
placB11
DXS8222
placB3-2
DXS7647
DXS1416
DXS1426
DXS1039*
DXS8366
DXS8223
DXS1522
DXS1331
DXS8226
DXS573*
DXS9795
SYP
JRL1-C4,5
JRL3-14
JRL1-B1
JRL3-10
JRL1-C10
JRL1-B6
JRL1-B4
JRL3-13
JRL3-8
JRL3-1
1908
c3219
c721
E10
E12
A11
E11
A02
D01
5742
B12
1269
3355
5032
5473
c502
2931
4224
1738
3775
4518
5905
4704
1950
6129
(100 kb)

Table 17. ESTs Designed from Published Sequences[a]

Designation	Primer Sequences (5'-3')	Size (bp)
IS2	TTATGCTGTGACACCACTGC	244[b]
	AGTAGTAACACCAGGCTGAC	
IS3-plac C10	CAGAGATGCACTGAACTCAT	140
	ATTAAATGACTCACCTGAGG	
IS4-placD12	TGTGGGCTCTCTTCCTTAGA	150
	CCTAAGTGGAATCAATGAAG	
IS7	CCATCTTGTCTCGTCGCATT	227
	TTTGGTTGCTTCAGGGAAAG	
RBM3	ACACAAGGAATAATTTCTGA	274
	AGCTCCAATAAACCTAAGAG	

[a] Derry *et al.*, 1994.

[b] Primers designed from cDNA sequence and span an intron. Product in genomic DNA is approximately 375 bp.

Characterization of a Candidate Gene for CSNB1:

Selection of a Candidate Gene:

The region between DXS722 and DXS255 at this stage of my studies contained four genes and 54 ESTs and the task was one of sifting through this transcript information to generate a prioritized list of candidate genes to be analyzed for mutations in individuals affected with CSNB1. The EST JRL4-A1 was chosen for further characterization as a candidate gene for CSNB1 in Family P060 for several reasons. JRL4-A1 had been localized proximal to the DXS1126 locus and therefore was not going to be potentially excluded as a candidate gene if the crossover event occurring between DXS722 and DXS1126 on the recombinant chromosome carried by D19 (Family P060) was refined further. Secondly, expression analysis (see below) of JRL4-A1 indicated that it was expressed both in the brain and the retina, making this EST a better candidate than one expressed only in an unaffected tissue. Lastly, JRL4-A1 was chosen for further characterization because a longer cDNA clone corresponding to this EST had been identified and initial characterization of this cDNA using a Northern Blot indicated that most of the whole gene was present in this clone (see below). The gene that corresponds to the JRL4-A1 cDNA will be referred to as KAT1.

Sequencing of a KAT1 cDNA Clone:

Database analysis with the sequence of JRL4-A1 indicated that it was part of the 5' end of a larger cDNA clone 263773 that had been partially characterized by the I.M.A.G.E. Consortium and the Wash U-Merck sequencing effort. The cDNA clone 263773 was obtained from Genome Systems Inc. St Louis, MO for a nominal charge. DNA was prepared for Sequenase® sequencing using the rapid plasmid small scale preparation. As the cDNA insert for this clone was in the vector pT3T7D, the primers T7 and T3 were used to generate sequence from the 5' and 3' end of the cDNA, respectively. Subsequently, two new primers (Table 18) were designed from the DNA sequence generated from the T7 and T3 primers and used to complete the assembly of the cDNA sequence (see later, Figure 41).

Table 18. Sequencing Primers for cDNA Clone 263773

Primer Name	Sequence 5' to 3'
773-T3-1	CCAGTCTTGGCCTCATTCCG
773-T7-1	ACTATGACGATGATCCTGTG

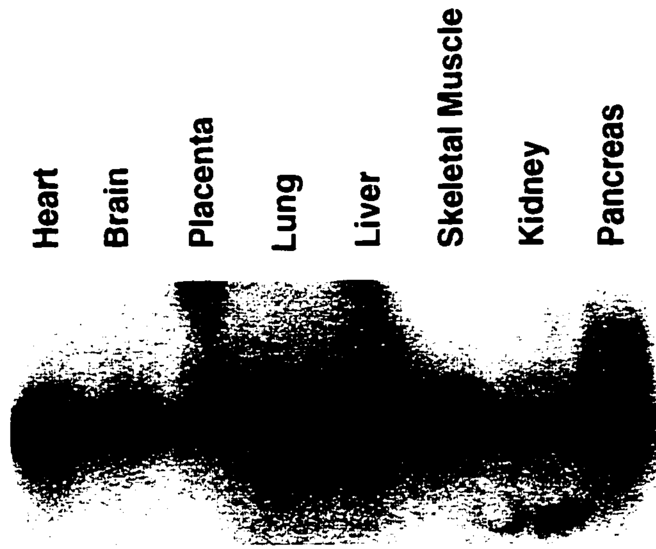
Expression of KAT1:

Northern Analysis: The cDNA 263773, corresponding to KAT1, was excised from the vector pT3T7D and the DNA isolated as described in Materials and Methods. The extracted DNA was labelled and used as a radioactive probe on a Northern Blot to analyze

expression and to size the full-length transcript. Clontech's human multiple tissue Northern blot was used for this purpose and contained mRNA from heart, brain, placenta, lung liver, skeletal muscle, kidney, and pancreas. KAT1 is abundantly expressed in all tissues that were part of the Northern blot and hybridized to a transcript of approximately 1100 bp (Figure 34).

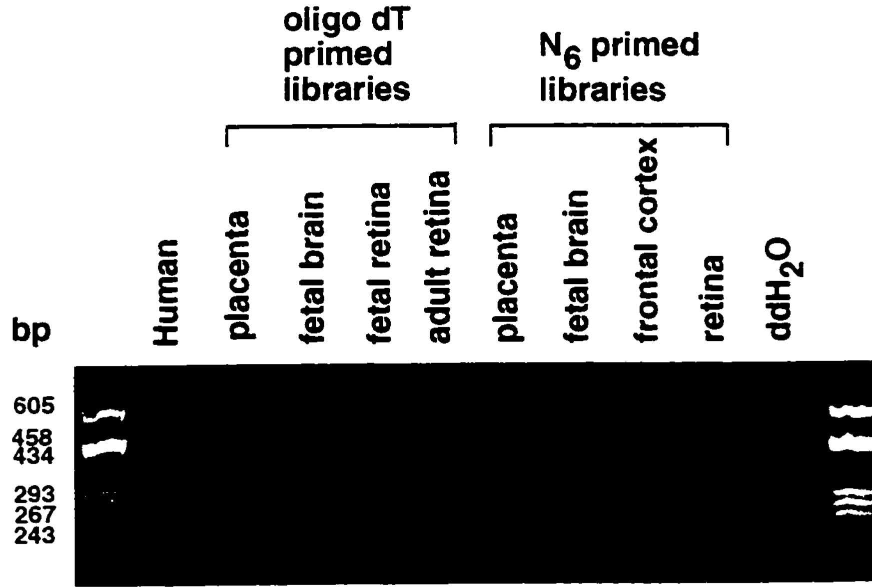
PCR Expression Profiles: PCR amplification using the JRL4-A1 EST on cDNA library aliquots is a convenient way to determine whether clones corresponding to this EST are present in a particular cDNA library, and therefore whether KAT1, by inference, is expressed in a particular tissue. Choosing a primer pair that amplifies across an intron, such as those for JRL4-A1, allows one to detect genomic contamination in either the cDNA library or the PCR reaction itself and prevents incorrect interpretation of the expression results. Amplification of JRL4-A1 on aliquots from the oligo dT primed placental, fetal brain, fetal retina, adult retina libraries, and the random primed placenta, fetal brain, frontal cortex and retina libraries indicated the presence of this gene in each of these libraries (Figure 35).

Figure 34. Autoradiogram of cDNA clone 263773 hybridized to a Clontech human multiple tissue Northern blot. Two μg of RNA has been loaded in each lane. The tissue types are indicated along the top. Comparison to the size marks on the filter indicates that this transcript is approximately 1100 bp (not shown).



Heart
Brain
Placenta
Lung
Liver
Skeletal Muscle
Kidney
Pancreas

Figure 35. PCR analysis of the EST JRL4-A1 on cDNA libraries and cDNA sets. The origin of each of the libraries or sets is indicated along the top. In human genomic DNA the primer pair encompasses an intron and hence the amplified fragment (470 bp) is larger than that seen when cDNA is used as the template (281 bp). The band in the N6 primed retina cDNA set is present, but is very faint due to the quality of this particular cDNA set (Dr. J. Rommens, personal communication). The DNA marker is pBluescript cut with *Hae*III and the fragment sizes are indicated.


oligo dT primed libraries
N_6 primed libraries
bp
Human
placenta
fetal brain
fetal retina
adult retina
placenta
fetal brain
frontal cortex
retina
ddH_2O
605
458
434
293
267
243

cDNA Sequencing Project Information: Large scale sequencing of complementary DNAs (cDNAs) from a variety of tissues is well underway through the efforts of the I.M.A.G.E. Consortium and the Wash U-Merck cDNA sequencing project (Lennon *et al.*, 1996) (Hillier *et al.*, 1996). The information generated by this effort is made available through public databases and is a resource of expression information. Analysis of the dbEST database using KAT1 sequence indicates that there are at least 16 other cDNAs that contain part of this sequence. Investigation as to the library origin of these cDNAs indicates that KAT1 is expressed in at least eight different tissues (Table 19).

Table 19. Expression Analysis of KAT1

Analysis	Tissue
Northern Blot	heart, brain, placenta, lung, liver, skeletal muscle, kidney, pancreas
PCR cDNA Library Expression Profiles	fetal brain, fetal retina, adult brain, adult retina, placenta
Sequencing Projects	fetal liver/spleen, fetal lung, melanocyte, parathyroid tumour, breast, infant brain, retina, heart

5' RACE:

To verify that the cDNA clone 263773, corresponding to KAT1, contains the 5' end of the gene, 5' RACE was performed using Clontech's Marathon-Ready brain cDNA and Klen*Taq* Polymerase Mix. The primer R-4 (Table 20, Figure 43) was chosen for RACE because of its location in the 5' end of the gene and the presence of an internal primer, R-5 (Table 20, Figure 43), that could be used if reamplification of the RACE product was required. Both the primers satisfied the criteria (G/C content of 50-70% and T_m of at least 60°C) for touch-down PCR. Amplification with R-4 and AP1 (adaptor primer) yielded two bands of approximately 450 bp (upper band) and 385 bp (lower band). These bands were excised separately from the gel, the DNA was purified using Gelase™ and reamplified using the internal primers R-5 and AP2 (internal adaptor primer). Reamplification gave only the upper (290 bp) or the lower band (225 bp) (Figure 36). These bands were excised and the purified DNA was sequenced using Thermo Sequenase™ (Figure 37). The sequences of the two bands were identical except for a 68 bp insertion in the sequence of the upper band which was considered to be alternatively spliced (Figures 37, 41, 42, and 43). Primers were designed from the RACE sequence to amplify the alternatively spliced region (Table 20) and they gave the expected product in human genomic DNA (data not shown). The sequence of both bands also indicated that the 5' end of the cDNA was intact in the original cDNA clone 263773.

Figure 36. 5' RACE products of the KAT1 gene using human brain Marathon-Ready cDNA. Amplification of the cDNA with the primers R-4 and AP1 yielded two products, an upper band of 450 bp and a lower band of about 385 bp. Reamplification of the upper and lower products in separate reactions using the internal primers R-5 and AP2 yielded a 290 bp product (from the upper band) and a 225 bp product (from the lower band). Products were run on a 1.5% SEPARIDE/1% agarose gel and visualized by EtBr staining. The DNA marker is pBluescript cut with *Hae*III and the sizes of the relevant fragments are indicated.

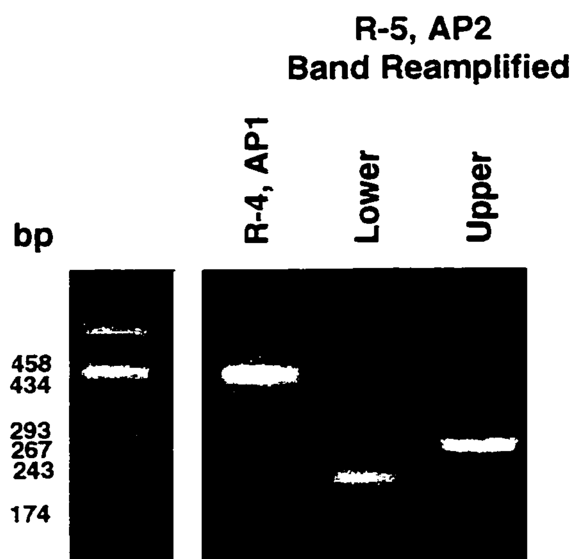
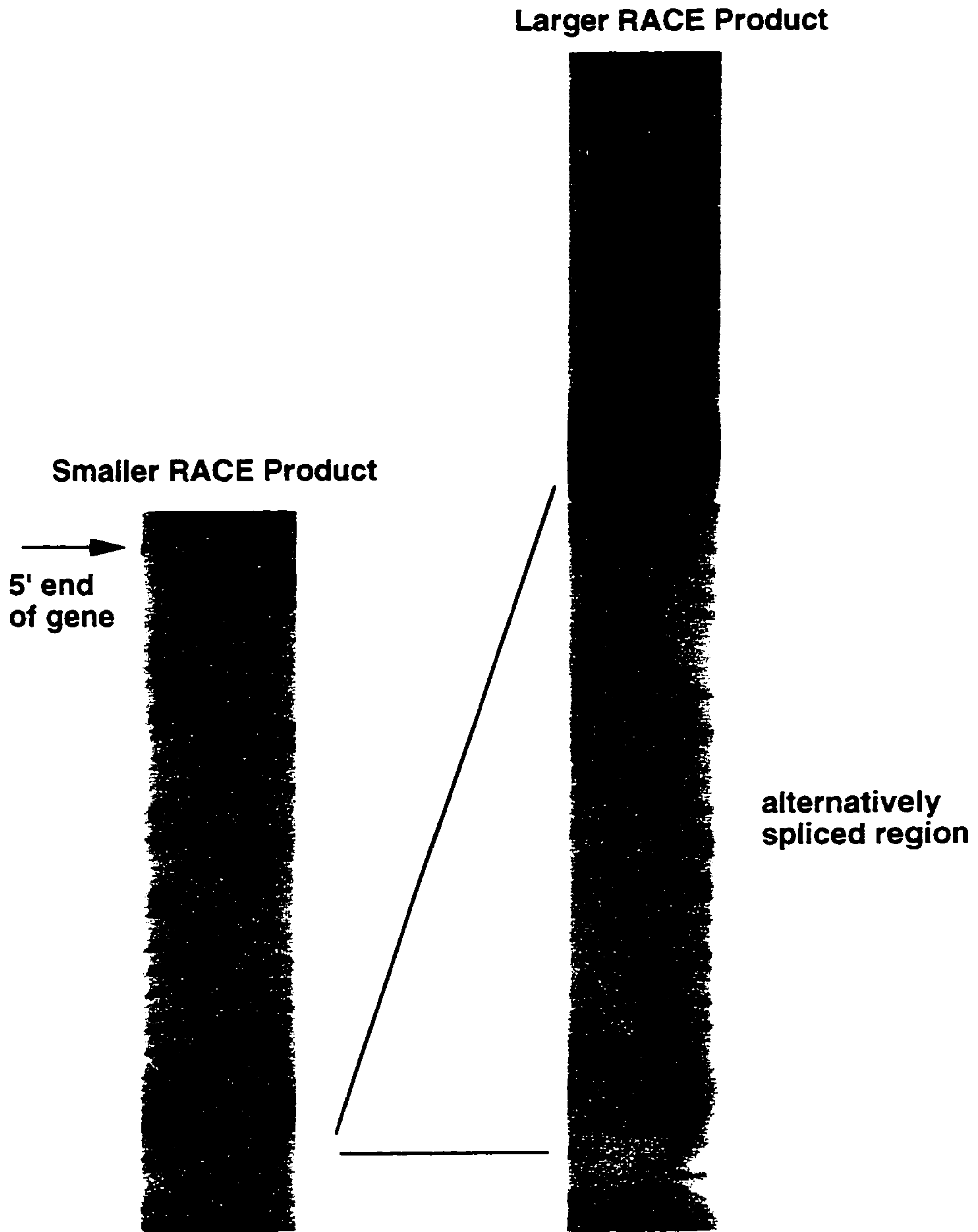

R-5, AP2
Band Reamplified
R-4, AP1
Lower
Upper
bp
458
434
293
267
243
174

Figure 37. Sequence of the upper and lower 5' RACE products amplified using R-5 and AP2 shown in Figure 36. The DNA was purifed using Gelase™ and sequenced using R-5 and Thermo Sequenase™. The dideoxy G, A, T, and C lanes are indicated along the top of the figure and the termination products have been resolved on a 6% acrylamide gel. The 68 bp corresponding to the alternatively spliced region are indicated. This sequence is detailed in Figures 41 and 42.


Larger RACE Product
Smaller RACE Product
5' end
of gene
alternatively
spliced region

Characterization of Exon/Intron Boundaries of the KAT1 Gene:

The genomic structure of KAT1 was elucidated by a combination of three different approaches. Initially, overlapping sets of primers were designed based on the cDNA sequence of clone 263773. Introns C, D, E, and F were isolated in this way. For the 5' end of KAT1 this approach was unsatisfactory because of the presence of at least one intron that was too large to be amplified by the PCR conditions in use. To elucidate the genomic structure in this region, a cosmid containing KAT1 was single-pass sequenced through a collaborative effort with the German Human Genome Project (Dr. A. Meindl, Munchen, and Dr. T. Rosenthal, Jena), and ligation-mediated PCR was used to walk into unknown intron sequence using KAT1 exon based primers. Cosmid sequencing generated partial sequence for Intron A and the region 5' to exon 1. Walking by linker ligation generated partial sequence for intron B.

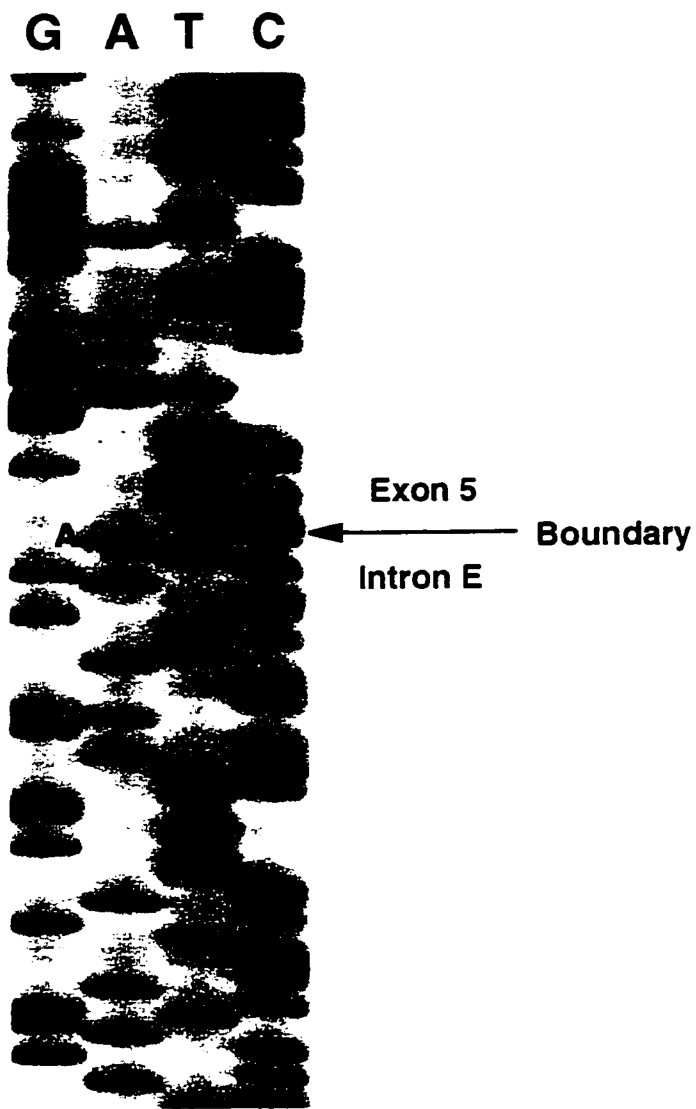
cDNA Based Primers: Five sets of overlapping primers were initially designed using the cDNA sequence of clone 263773 (Table 20, Figures 41 and 43). Amplification with three of the primer sets, F-3/R-3, F-4/R-2, F-5/R-1 under the standard PCR conditions using TNK100 buffer gave products of greater than expected size, suggesting the presence of at least one intron. These products were isolated in low melting point agarose gels, purified using Gelase™, and sequenced using Thermo Sequenase™. These sequences were compared with the cDNA sequence to identify the location and sequence of the introns. The primer set F-3/R-3 amplified Intron D,

Table 20. Primers Designed to Amplify cDNA 263773

Primer Set	Sequence 5' to 3'	Expected Size (bp)	Actual Size (bp)
AEF, AER	CGTAGCGGCGACACGAGAGA	63	63
	AATACACTCGTGCAGGTAGT		
F-1, R-5	GAAGGTCTCATTCGGTGTTT	239	none
	TCATCGTCATAGTCCTCGGC		
F-2, R-4	GGCCAAGAGAGGCATCCTCA	243	none
	AGCTTCTTGGCCGATTTGGT		
F-3, R-3	CCCTTACTACTGGAATGCAG	272	470
	CTCTCTCTGTCTACCTTGTC		
F-4, R-2	GACAAGTCTGACAGGGATCG	217	460
	AGCTAGGGTCCATGGGGTCT		
F-5, R-1	CTCCCTATCCCAAGAGCAAG	294	605
	AGGATCACCAGAAAGCTTTT		
F-2-2, IR-2	GGAAGAGATCATTGCCGAGG	289	850
	TCATTCTACTTGGCCTCTCC		
F-1-2, IR-4	CCCAGGTATCGTAGCGGCGA	354	none
	GGCTCCCTCCCTTGTGTCCC		
F-1-3, IR-4	GAGTGTATTGGTCTGTCTGC	301	270
	GGCTCCCTCCCTTGTGTCCC		
F-1-4, IR-2	GCTATGCCGCTGCCCGTTGC	992	none
	TCATTCTACTTGGCCTCTCC		
F-1-5, IR-2	GTGACAGCCTTCCACTACCT	1041	none
	TCATTCTACTTGGCCTCTCC		

F-4/R-2 amplified Intron E, and F-5/R-1 amplifed both Introns E and F (Figures 41, 42, and 43). Figure 38 indicates the location of the exon/intron boundary between Exon 5 and Intron E in the F-4/R-2 product. The sequences of the exon/intron boundaries all conform to the canonical splice donor and acceptor sites (Figures 41, 42, and 43). The primer pairs F-1/R-5 and F-2/R-4 (Table 20) did not generate any products and so the primers were redesigned. The new primer pair F-2-2/IR-2 (IR-2 is located within Intron D) gave a larger than expected product and was sequenced to determine the location of the intron C. The primer IF-5 5' CTTGGTTGTGAGGTTTAGGG 3' was designed from intron C sequence to complete the sequence of this intron. The new primer pairs F-1-2/IR-4 (IR-4 is located within Intron C) and F-1-3/IR-4 (Table 20) gave no product and a smaller product than expected, respectively, the latter of which when sequenced did not originate from this region. BLAST analysis of intron C containing IR-4 indicated that it was an *Alu* repeat, and this repeat probably resulted in a PCR amplified product from an origin other than KAT1. Two additional 5' primers, F-1-4 and F-1-5 (Table 20), were then designed to amplify genomic DNA with the primer IR-2, known to anneal specifically, but again these primers failed to amplify any product. Complete characterization of KAT1 was evidently going to require additional approaches because of the putative large intron located in the 5' portion of this gene that could not be crossed using cDNA based primers and the PCR conditions in use.

Figure 38. Sequence of the F-4/R-2 product from the KAT1 gene using R-2 as a sequencing primer and Thermo Sequenase™. The dideoxy G, A, T, and C lanes are indicated along the top of the figure and the termination products have been resolved on a 6% acrylamide gel. The exon/intron boundary between Exon 5 and Intron E is indicated along with the two bases that conform to the splice donor site consensus sequence (CA-GT).


G A T C
A
Exon 5
Boundary
Intron E

Cosmid Sequencing: A partial cosmid contig of Xp11.23 has been recently constructed by a group in Germany (Schindelhauer *et al.*, 1996) and was made available through collaboration with Dr. A. Meindl at Ludwig-Maximilians-Universitat in Munchen. KAT1 was determined to be on cosmid U138D3 of this contig. This group then shotgun cloned this cosmid into M13 and arranged for it to be sequenced by Dr. A. Rosenthal at the Institut fur Molekulare Biotechnologie in Jena, Germany. The unedited primary sequence data was made available and consisted of 541 primary reads. This data was dropped into a word processing file and the sequence was scanned using the 'Find' function and six base portions of the 5' end of KAT1 that were predicted to lie near exon/intron juctions. Forty-seven bases that were 5' to the first base of exon 1, 162 bases in the 5' region of Intron A, and 30 bases in the 3' region of Intron A were found. This intron information was used, in combination with information generated by walking (see next section), to complete the genomic structure of KAT1 (Figures 41, 42, and 43), and to amplify exons 1 and 2 using intron-based primers (Table 22, Mutation Analysis Section).

Walking by Linker Ligation: The cosmid sequence information did not elucidate the entire structure of KAT1 and Intron B remained to be defined. Therefore, walking into Intron B was required for the amplification of both Exons 2 and 3. The method used for cloning of YAC ends was adapted for walking from known exon sequence into unknown intron sequence. The procedure was performed exactly as

described for the generation of YAC ends, with the exceptions that cosmid DNA from U138D3 was used instead of YAC DNA, and gene specific primers were used instead of YAC vector specific primers. For the walk into the 3' end of Intron B, the primer R-6 (designed for this purpose, Table 21) and R-5 (Table 20, Figure 43) were chosen as the external and internal primers respectively, because they were predicted to flank the proximal end of Intron B. For the walk into the 5' end of Intron B, the primers F-1-4 and F-2 (Table 20, Figure 43) were chosen as the external and internal primers respectively, because they were predicted to flank the distal end of Intron B.

Table 21. KAT1 Primer for Walking by Linker Ligation

Primer Name	Primer Sequence 5' to 3'
R-6	AACCTGGTGGCCTCGTAGTC

For the walk into the 3' end of Intron B, an *Alu*I digestion product of 350 bp was isolated and sequenced using the R-5 primer and Thermo Sequenase™. The sequence generated aligned to the cDNA sequence for 10 bases before it deviated into Intron B sequence. For the walk into the 5' end of Intron B, an EcoRV product of 230 bp (Figure 39) was sequenced using the F-2 primer and Thermo Sequenase™ (Figure 40). The sequence generated aligned to the cDNA sequence for six bases before it deviated into Intron B

Figure 39. Ligation-mediated PCR product amplified from the ligated product produced by the digestion of U138D3 with *Eco*RV. Amplification was performed using F-1-4/L1 followed by reamplification with F-2/L1. Products were separated on a 1.5% SEPARIDE/1% agarose gel and visualized by EtBr staining. The marker is pBluescript cut with *Hae*III and the sizes of the relevant bands are indicated.

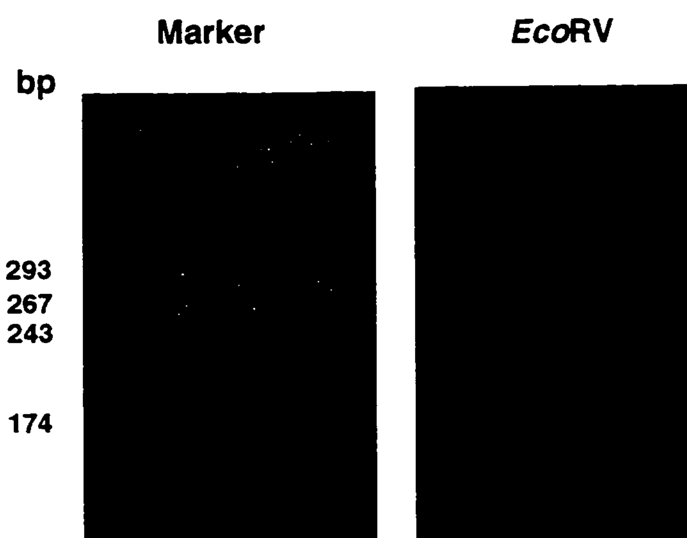
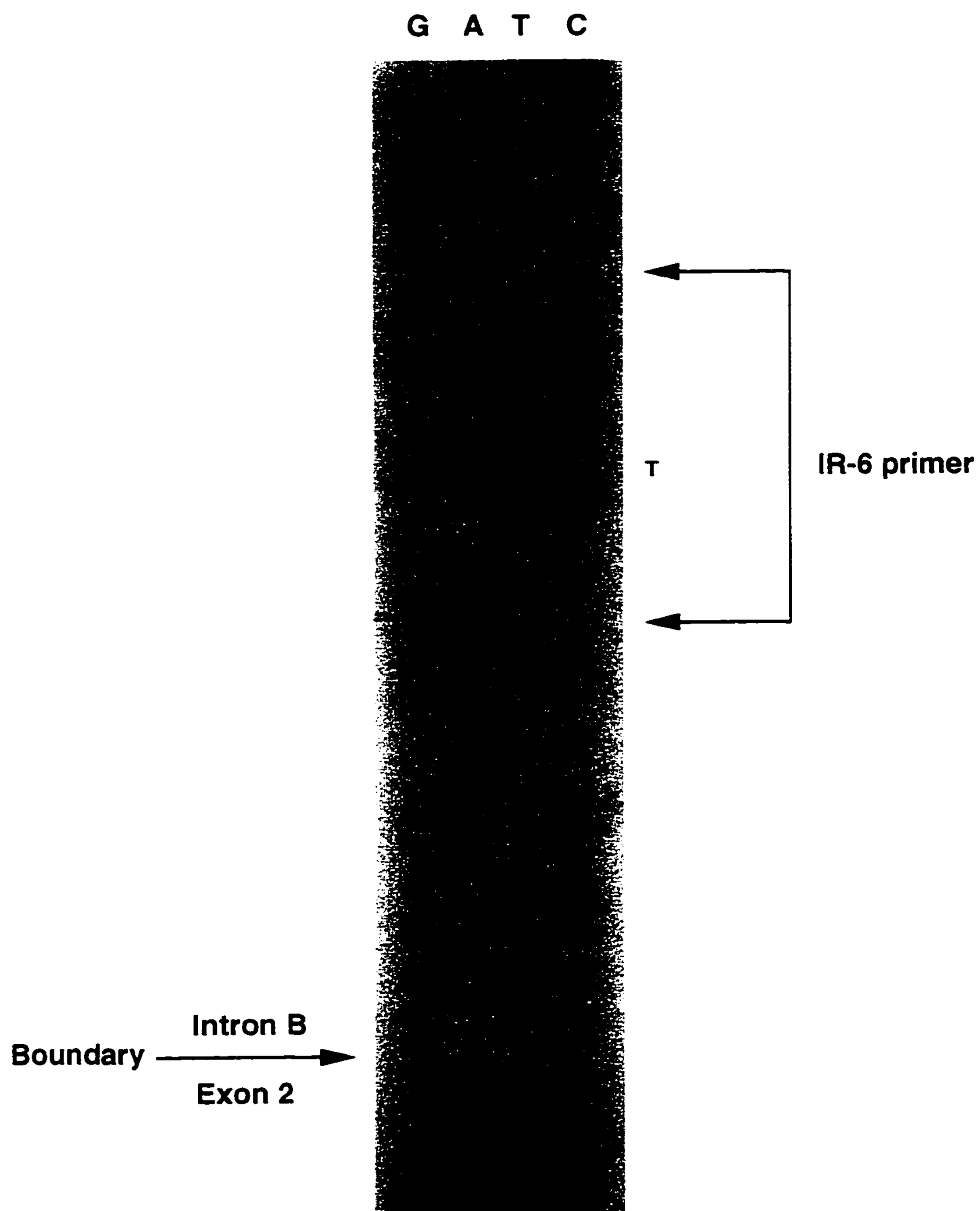

Marker
*Eco*RV
bp
293
267
243
174

Figure 40. The *Eco*RV product from Figure 39 was isolated, purified using Gelase™, and sequenced using the F-2 primer and Thermo Sequenase™. The dideoxy G, A, T, and C lanes are indicated along the top of the figure and the termination products have been resolved on a 6% acrylamide gel. The location of the exon/intron boundary between Exon 2 and Intron B is indicated along with the two bases that conform to the splice donor site consensus sequence (GT).


G A T C
T
IR-6 primer
Intron B
Boundary
Exon 2

sequence. This intron information was used, in combination with information generated by cosmid sequencing and cDNA based primer analysis, to complete the genomic structure of KAT1 (Figures 41, 42, and 43), and to amplify exons 2 and 3 using intron-based primers (Table 22, Mutation Analysis Section).

The KAT1 Gene:

The gene referred to as KAT1 is located in Xp11.23 proximal to the polymorphic makers DXS1126 and DXS1240 and distal to the polymorphic marker DXS255 (Figure 32). The gene contains seven exons and six introns (Figures 41, 42, and 43) and has an open reading frame, which begins in Exon 2 with an initiator methionine, of 262 amino acids (Figures 41 and 43). The exons range in size from 64 bp (Exon 6) to 285 bp (Exon 5) (Figures 41, 42, and 43). Five out of six introns have been defined and these range in size from 132 bp (Intron F) to 650 bp (Intron A) (Figures 42 and 43). Intron B was too large to be amplified under the PCR conditions in use, or long range PCR conditions (data not shown), and remains undefined in size. KAT1 is a relatively small gene with a short 3' untranslated region and spans at least 3 kb of genomic DNA. Two transcripts are transcribed from KAT1, one that has a 129 bp Exon 1 (Exon 1-A) and one that has a 61 bp Exon 1 (Exon 1-B) (Figures 41, 42, and 43). This alternative splicing event occurs 5' to the open reading frame. KAT1 is ubiquitously expressed and is expressed in both the brain and the retina. Both transcripts have been seen in brain (5' RACE results)

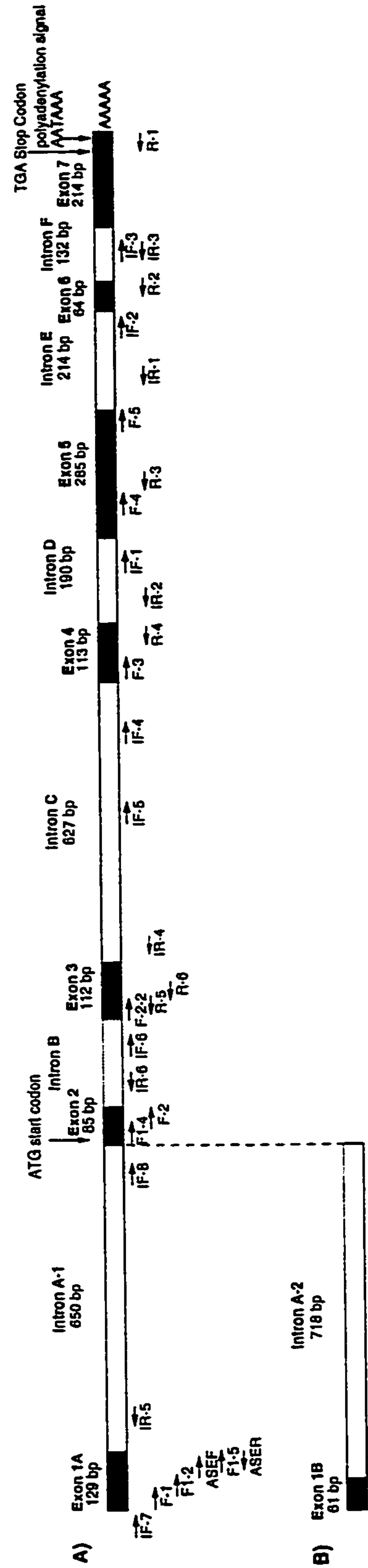
Figure 41. Nucleotide and translated amino acid sequence of the exons of KAT1. The translation of the larger transcript, which utilized Exon 1A, is shown. Splice donor and acceptor sites are shown in lowercase letters. The putative initiator methionine codon, the stop codon, and the polyadenylation signal sequence are in boldface type. Detailed intron sequence of this gene is shown in Figure 42.

GGAAGGCCTCGTTGAGAGAAGGTCTCATTCGGTGTTTTGGGAAGAGAGTCGT
GTGGGCCCAGGTATCGTAGCGGCGACACGAGAGAGACGGGCGGTGTGACAGCC
TTCCACTACCTGCACGAGTGTATTGgt...IntronA...agGTCTGTCTGCTATCAGC
TATG CCG CTG CCC GTT GCG CTG CAG ACC CGC TTG GCC AAG AGA
M P L P V A L Q T R L A K R
GGC ATC CTC AAA CAT CTG GAG CCT Ggt...IntronB...agAA CCA GAG
G I L K H L E P E P E
GAA GAG ATC ATT GCC GAG GAC TAT GAC GAT GAT CCT GTG GAC
E E I I A E D Y D D D P V D
TAC GAG GCC ACC AGG TTG GAG GGC CTA CCA CCA AGC TGG TAC AAG
Y E A T R L E G L P P S W Y K
GTG TTC GAC CCT TCC TGgt...IntronC...agC GGG CTC CCT TAC TAC TGG
V F D P S C G L P Y Y W
AAT GCA GAC ACA GAC CTT GTA TCC TGG CTC TCC CCA CAT GAC CCC
N A D T D L V S W L S P H D P
AAC TCC GTG GTT ACC AAA TCG GCC AAG AAG CTC AGA AGC AGT
N S V V T K S A K K L R S S
AAT GCA Ggt...IntronD...agAT GCT GAA GAA AAG TTG GAC CGG AGC
N A D A E E K L D R S
CAT GAC AAG TCG GAC AGG GGC CAT GAC AAG TCG GAC CGC AGC CAT
H D K S D R G H D K S D R S H
GAG AAA CTA GAC AGG GGC CAC GAC AAG TCA GAC CGG GGC CAC
E K L D R G H D K S D R G H
GAC AAG TCT GAC AGG GAT CGA GAG CGT GGC TAT GAC AAG GTA
D K S D R D R E R G Y D K V
GAC AGA GAG AGA GAG CGA GAC AGG GAA CGG GAT CGG GAC CGC
D R E R E R D R E R D R D R
GGG TAT GAC AAG GCA GAC CGG GAA GAG GGC AAA GAA CGG CGC
G Y D K A D R E E G K E R R
CAC CAT CGC CGG GAG GAG CTG GCT CCC TAT CCC AAG AGC AAG AAG
H H R R E E L A P Y P K S K K
Ggt...IntronE...agCA GTA AGC CGA AAG GAT GAA GAG TTA GAC CCC
A V S R K D E E L D P
ATG GAC CCT AGC TCA TAC TCA GAC CGG CCC CGgt...IntronF...agG GGC
M D P S S Y S D R P R G
ACG TGG TCA ACA GGA CTC CCC AAG CGG AAT GAG GCC AAG ACT GGC
T W S T G L P K R N E A K T G
GCT GAC ACC ACA GCA GCT GGG CCC CTC TTC CAG CAG CGG CCG TAT
A D T T A A G P L F Q Q R P Y
CCA TCC CCA GGG GCT GTG CTC CGG GCC AAT GCA GAG GCC TCC CGA
P S P G A V L R A N A E A S R
ACC AAG CAG CAG GAT TGAAGCTTCGCCTCCCTGGCCCTGGGTTAAAATA
T K Q Q D
AAAGCTTTCTGGTGATCCTGCCCACCAAAAAAAAAAAAAAAAA

Figure 42. Intron sequence of the KAT1 gene. Intron sequence is shown in lowercase letters and the included exon sequence is shown in uppercase letters. The 68 bases shown in uppercase letters located in Exon 1 are alternatively spliced in the two different transcripts of KAT1 that have been characterized. The bases from the 3' untranslated region of Exon 7 are shown and the polyadenylation signal is underlined.

ttcattgcctcctgagcgtantccagttactttcaggctcggggagt**GAAG...Exon1B**....GTA
TCGTAGCGGCGACACGAGAGAGACGGGCGGTGTGACAGCCTTCCACTACCTGC
ACGAGTGTATTGgtaacgttggggtgggtcactcttttggagctggaggaagngctgccctctg
ctcccccatcccagcgctttggtttctcccacccggcagtcatggggacgntggatgagcccatttctga
agtgcggaggggccgggcttcttagggngtgggcgggggа......ctgcttgtcagtttgttcgtctgtcc
ctag**GTCT...Exon2...CCTG**gtgagacagctaaaagcagatggtcctaacacgtgccagcct
cgacaggcacttttgttgttgatactgacgcttgatacctagcatgggccaactgtgcaccaagcaatg
gggtggcaggcatttttcgttgttaacaatgtggattcctgacatccatccctaccgcgtgtcaagt......
ctaacagtagaactcatcatctgaacccagacacagatttaggtaacctttttttgtaaacctcctaata
tagagagttggatttggaatgtgaaagatactcatattcttccccggcacacacacacacgtgtctttt
atgtgtggcttgtgtgaatgtgagtgcgtgtctctaggctctacatgtaggatgaggtcccctggtcctt
atctgctctcctccatccccaccttgactaccag**AACC...Exon3...CCTG**gtgagcctgggtga
gggggagctaacttctggcttcacccttcctgtgctgaccttggttgtgaggtttagggggacacaagg
gagggagcctcgggtggagggtgttggcattaggtatcagcaggactcaagttgctgcctgctgggg
cctggctcctctggggttggaagactgtcttttctctctttttgaaacggagtttcactcttgtagcccgg
gctggagtgcaatggtgtgatcacggcttatggcaacctccgcctcgtgggttcaagcgattctcctgt
ctcagcctcccaggtagctgggattacaggtgcacaccaccacacccagctaattttgtatttttgta
gagatggggtttcatcatattggtcaggctggtcccgaactcctgacctcaggtgatccgcctgcctcg
gcctcccaaagtgctgggattacaggcatgagccaccacgcccagctgacagtgttttctcttcaggag
aagagattctggggatcctcctgtggggtcctgacatggggcaggtataggcagagactggggagac
aagagttgaggatccaaggcaaggacatctgtgctgacacttctttcctgcggccccatag**CGGG...**
Exon4...GCAGgtgagttggcaggtacaagcgtgccttgagtgatcttagcagttctcacggaga
ggccaagtagaatgatagtggatcagaggggcaggtgagaccaggcgggcccagcctcaggcaag
ggaggttgttgaaggcagaggctgctgggtcctggggtggcaagaggtcacttcaagacttgtgtccc
cag**ATGC...Exon5...AAGG**gtaagctgggcagaatggggctcggtgagaccaacaaggtg
cagggtgacctgcgtgaggaagccttccctcaaaaagatgcctggacctggggctagaggagggtgc
tgtggtacatggcagccaggggcttcatttcttcttgtggggtggggctcagtgatcaggggctcctgg
tgcctctattgaagactttgccctgccacttccacag**CAGT...Exon6...CCCG**gtaagtgacaa
cccctcttgactcagtacgtggacaccatcctccggcctccttcctccattcctacttgggaccaggtgg
gcagagaccgccacatcacccatccccatcccctgactctttcaccggcag**GGGC...Exon7...A**
ATAAAAGCTTTCTGGTGATCCTGCCCACC

Figure 43. Genomic structure of the KAT1 gene. All exons and introns are drawn to scale with the exception of Intron B, the size of which has not been determined. The putative start codon is indicated as well as the termination codon and the polyadenylation signal. The location of the primers used to characterize this gene and designed for use in mutation analysis are also indicated. The F or R primers correspond to forward or reverse orientation within exon sequence, respectively, and the IF or IR primers correspond to forward or reverse orientation within intron sequence, respectively. The number following these designations refers to the order in which they were made. The ASEF and ASER primers amplify the portion of Exon 1A that is removed by alternative splicing to give Exon 1B and the two alternative transcripts for KAT1 are indicated (A and B).


A)
Exon 1A
129 bp
Intron A-1
650 bp
ATG start codon
Exon 2
85 bp
Intron B
Exon 3
112 bp
Intron C
627 bp
Exon 4
113 bp
Intron D
190 bp
Exon 5
285 bp
Intron E
214 bp
Exon 6
64 bp
Intron F
132 bp
Exon 7
214 bp
TGA Stop Codon
polyadenylation signal
AATAAA
AAAAA
IF-7
F-1
F1-2
ASEF
F1-5
ASER
IR-5
IF-8
F1-4
F-2
IR-6
IF-6
F-2-2
R-5
R-6
IR-4
IF-5
IF-4
F-3
R-4
IR-2
IF-1
F-4
R-3
F-5
IR-1
IF-2
R-2
IF-3
IR-3
R-1
B)
Exon 1B
61 bp
Intron A-2
718 bp

and fetal liver/spleen (database analysis), but only the larger transcript has been seen other tissues (database analysis).

Mutation Analysis of KAT1 in CSNB1 Affected Individuals:

Deletion Analysis using KAT1:

The availability of any large rearrangements in the CSNB1 gene, even if present in only a small subset of patients, would aid the process of gene identification. With genetic heterogeneity in X-linked CSNB being likely, it is necessary to screen affected representatives from all CSNB1 families for deletions because it is not possible at this point to determine in most affected families which CSNB1 gene is responsible. DNA from CSNB1-affected individuals from 15 different families were available for these studies. DNA from these individuals was cleaved with four six cutter enzymes, electrophoresed, and transferred to nylon membrane to generate a set of four deletion analysis panels. The 263773 cDNA insert was isolated, labelled, and hybridized to the set of deletion analysis panels. Analysis of the autoradiograms showed no anomalies with the KAT1 probe hybridized on membranes containing *Bam*HI (Figure 44), *Eco*RI, *Hind*III, or *Pst*I cut patient DNA.

Figure 44. Autoradiogram showing the hybridization pattern of the cDNA clone 263773 on patient DNA from 15 different CSNB1 families cut with *Bam*HI. The individuals' designation numbers are listed across the top. A641, A679, A915, and A644 are all from Family P060. All individuals are male and affected except for the male random control.

Random
A641
A679
A915
A644
A502
A712
A804
A1002
A1302
A1405
A1502
A1700
A1801
A1903
A2003
A2308
A2401
A2647

Sequence Analysis of KAT1 in FamilyP060:

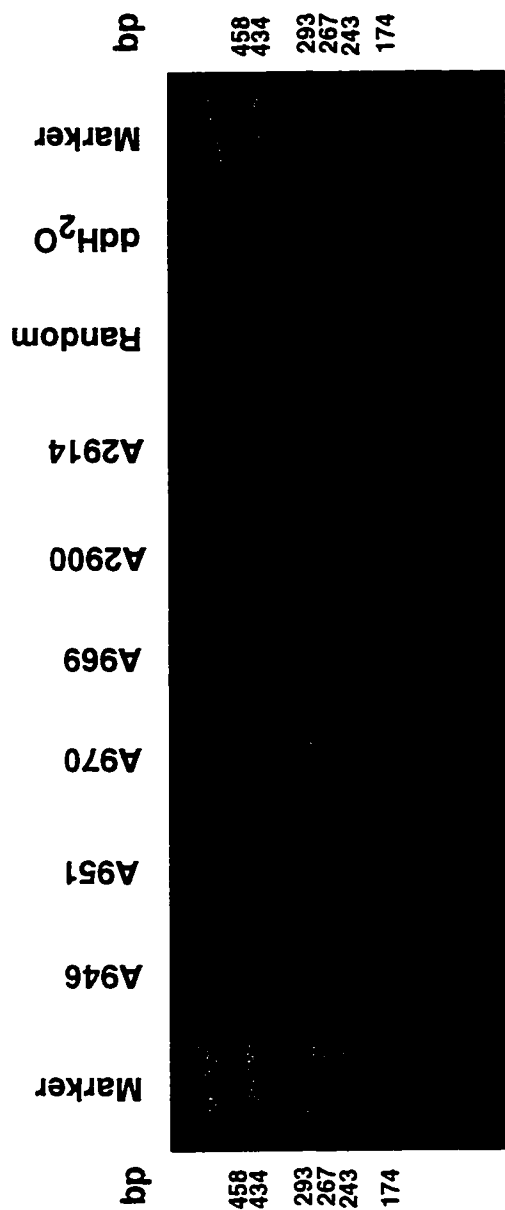
Mutation analysis of KAT1 was performed by direct sequencing, instead of other mutation detection methods such as SSC analysis, for two reasons. First, since CSNB1 is X-linked and initially only affected males would be screened for mutations, detection of base changes on a heterozygous background was not going to be a problem. Secondly, CSNB1 is likely genetically heterogeneous and therefore, initially, only representatives from Family P060 would be screened for mutations. Because of the small number of DNA samples undergoing the mutation screening, direct sequencing would be as convenient as any other detection method with the advantage of increased sensitivity. Two affected individuals (A946 and A969) and two unaffected controls (A951 and A970) were chosen from Family P060 for sequence analysis (Figure 5). Because of the possibility that AIED and CSNB1 are the same clinical entity, an affected and unaffected individual from Family 290 (a family diagnosed with AIED) were also included in the mutation analysis of KAT1 by sequencing. Individual A2900 is affected with AIED and his brother A2914 is unaffected. The primers used to characterize the exon/intron boundaries (Table 20) were initially used to amplify genomic DNA from the affected individuals and unaffected controls to screen for mutations in this gene. To screen this gene completely, primers were designed based on the intron sequence to amplify the exons and splice junctions (Table 22). These primers were used to amplify genomic DNA, the bands isolated, the DNA extracted using Gelase™ and sequenced from both ends using Thermo Sequenase™. Figure 45 shows the PCR

amplification products of Exon 3 of KAT1 using IF-6/IR-4 and Figure 46 shows the dideoxy C sequence of Exon 3 in the six individuals. After screening the exons and portions of the introns twice, only an intron polymorphism was detected in the 3' end of Intron C (Figure 47).

Table 22. Exon Amplification of KAT1

Primer Set	Primer Sequence 5' to 3'	Product (bp)
IF-7, IR-5	TCCAGTTACTTTCAGGCTCG	290
Exon 1	GCCCCTCCGCACTTCAGAAA	
IF-8, IR-6	TTGTCAGTTTGTTCGTCTGT	150
Exon 2	CTGGCACGTGTTAGGACCAT	
IF-6, IR-4	TGCGTGTCTCTAGGCTCTAC	284
Exon 3	GGCTCCCTCCCTTGTGTCCC	
IF-4, IR-2	CAAGAGTTGAGGATCCAAGG	248
Exon 4	TCATTCTACTTGGCCTCTCC	
IF-1, IR-1	CTCAGGCAAGGGAGGTTGTT	459
Exon 5	CCCAGGTCCAGGCATCTTTT	
IF-2, IR-3	GCTCCTGGTGCCTCTATTGA	181
Exon 6	CAATGAGGAATGGAGGAAGG	
IF-3, R-1	CCTTCCTCCATTCCTCATTG	287
Exon 7	AGGATCACCAGAAAGCTTTT	

Figure 45. Amplification of Exon 3 of KAT1 by PCR using the primer set IF-6/IR-4 in individuals A946, A951, A969, A970, A2900, and A2914. Products were 284 bp in size, electrophoresed on 1.5% SEPARIDE/1% agarose gels and visualized by ethidium bromide staining. Products were run into low melting point gel and isolated using Gelase™ for sequencing (see Figure 46). The DNA marker is pBluescript cut with *Hae*III and the sizes of the relevant fragments are indicated.



bp
458
434
293
267
243
174
Marker
ddH2O
Random
A2914
A2900
A969
A970
A951
A946
Marker
bp
458
434
293
267
243
174

Figure 46. Mutation screening of Exon 3 of KAT1. Exon 3 was amplified by IF-6/IR-4 (Figure 45) and sequenced using Thermo Sequenase™ and IF-6 as the sequencing primer. The sequencing reaction was run on a 6% acrylamide gel and the dideoxy Gs, As, Ts, and Cs were run together for ease of detection of base changes. Only the Cs are shown here and the exon sequence is indicated.

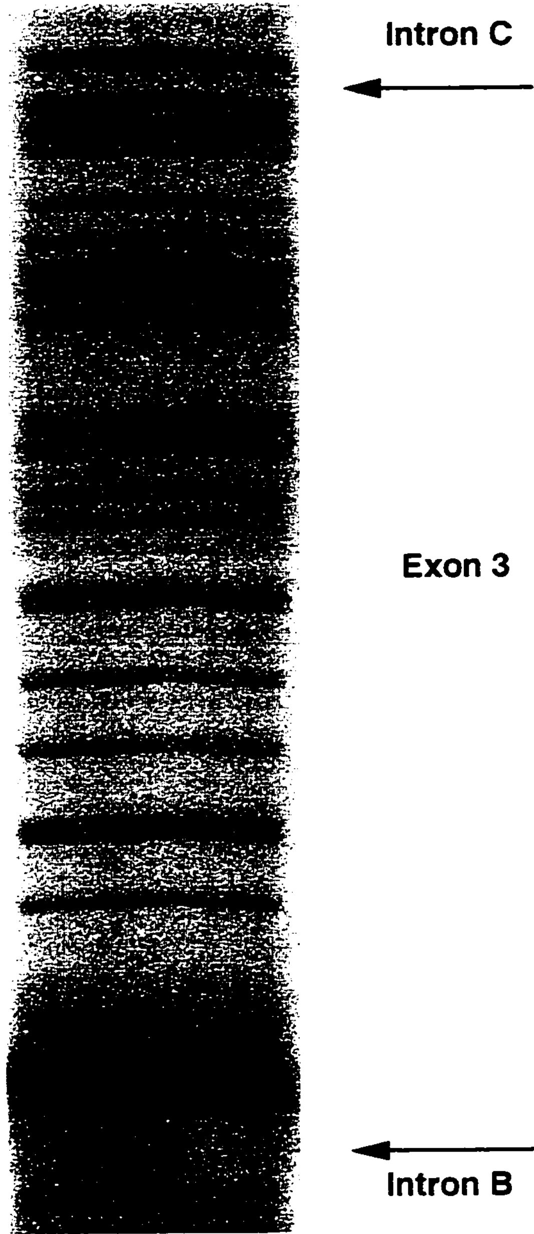
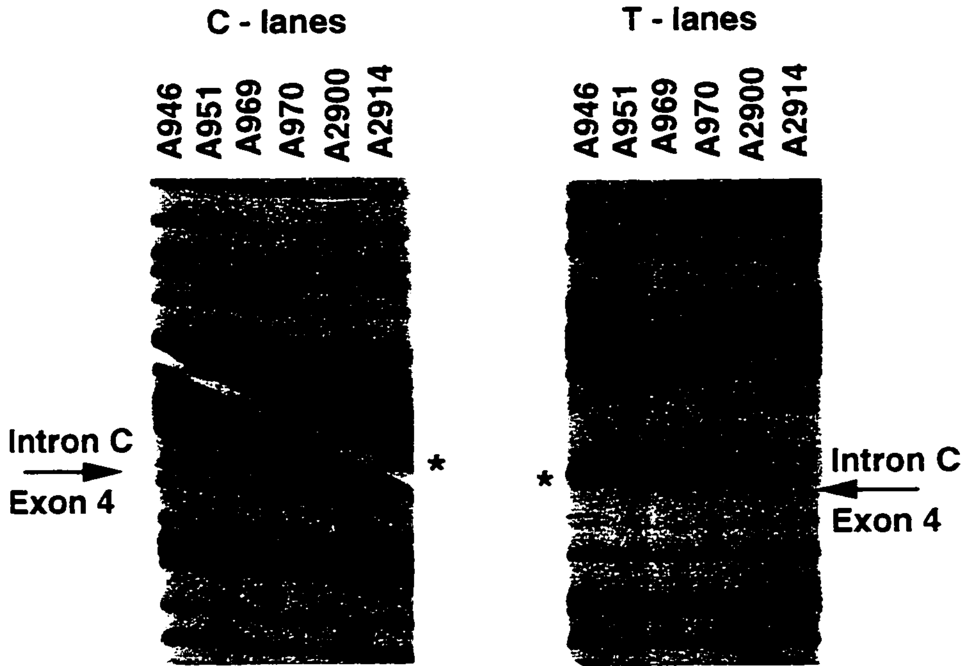

Intron C
Exon 3
Intron B

Figure 47. Polymorphism present in Intron C in two of the six individuals shown here. The polymorphism is a T to C transition. Genomic DNA was amplified with the primer set IF-4/IR-2, the products run into low melting point gel and isolated using Gelase™. The DNA was sequenced using Thermo Sequenase™ and IR-2 as the sequencing primer. The dideoxy C and T reactions are shown here resolved on a 6% acrylamide gel.


C - lanes
A946
A951
A969
A970
A2900
A2914
T - lanes
A946
A951
A969
A970
A2900
A2914
Intron C
Exon 4
*
*
Intron C
Exon 4

CHAPTER 5 - DISCUSSION

Refining the Genetic Limits for a CSNB1 Locus:

For a positional cloning project to have a reasonable chance of success, the genetic region containing the disease locus must be as refined as possible to reduce the enormous task of isolating, characterizing, and analyzing candidate genes. When this project to identify the gene(s) for X-linked CSNB began, the genetic location of the gene(s) for CSNB1 had a distal limit of MAOA/MAOB in Families P060, P070, P130, P140, and 260 and further refined to ZNF21 in Family P060, and a proximal limit still needed to be confirmed. To put tight flanking markers on the location of the CSNB1 gene(s), 23 markers were mapped on a set of radiation-reduced and conventional somatic cell hybrids. A number of these markers were refined to intervals of interest and were investigated in the families. In both families P130 and P140, the minimal genetic region containing the disease gene spanned the interval between MAOA-MAOB and DXS988, approximately 15 cM. The position of the crossovers in both of these families were reasonably well-defined and further marker analysis would not reduce the critical region significantly. Based only on this information, this interval was too large to undertake isolation and candidate gene analysis.

In Family P060, however, the genetic region containing the disease gene was refined to the region between DXS722 and DXS8023, a distance of less than 2.2 Mb (data not shown; Nelson *et*

al., 1995; Boycott *et al*., 1996; Schindelhauer *et al*., 1996). All available genetic markers were analyzed in this family and the crossover in D19 defining the distal limit occurs between DXS722 and DXS1126, a distance of 300 kb, and the crossover in D3 defining the proximal limit occurs between DXS255 and DXS8023, a distance of less than 700 kb. Attempts to refine the location of these crossovers further using microsatellite markers have been unsuccessful due to lack of informativeness and availability of markers.

Single nucleotide substitutions are the most common and widely distributed type of polymorphism in the genome. Therefore it is likely that any random STS obtained from the genome (i.e.. YAC end-clone or random clone) would have a high probability of containing one or more single nucleotide substitutions depending on the size of the STS (Nickerson *et al*., 1992). An attempt was made to refine crossover events in the two critical recombinants in Family P060 by scanning STSs generated from random clones or YAC end-clones, with suitable map positions, for nucleotide substitutions in appropriate family members (data not shown). In over 800 nucleotides assessed in this way, no single base variations were detected.

Therefore, in Family P060, the minimal genetic region for CSNB1 lies between DXS722 and DXS8023. For these studies, the portion of the minimal region in Family P060 that lies between DXS722 and DXS255, a distance of less than 1.5 Mb (Boycott *et al*., 1996; Schindelhauer *et al*., 1996) has been focused on. This effort on the minimal region distal to DXS255 is based on the crossovers with DXS255 that have been detected in other families segregating both

CSNB1 (Musarella *et al.*, 1989) and AIED (Glass *et al.*, 1993). Whether the CSNB1 disease gene segregating in these other CSNB1 families (Musarella *et al.*, 1989) is the same one as in Family P060 remains to be determined. However, the characterization of a crossover that also places AIED, a disease that has been postulated to be the same entity as incomplete CSNB1 or allelic to incomplete CSNB1, distal to DXS255 gives confidence in the choice to focus the initial positional cloning efforts on the region distal to DXS255.

Females who are known carriers of X-linked disorders show a wide variation in clinical expression, much of which can be attributed to lyonization whereby one of the two X chromosomes in each somatic cell of a female is randomly inactivated in the early embryo (Lyon, 1962). Most heterozygous (carrier) females with CSNB1 are asymptomatic, however four affected females have been reported from a five-generation family with CSNB1, most likely due to uneven X-chromosomal lyonization (Ruttum *et al.*, 1992). Individual D19 from Family P060 places the distal limit of the CSNB1 locus in this family at DXS722. This localization is dependent on D19's status as an affected female whose manifestations are a result of her carrying two copies of a mutant CSNB1 gene, rather than uneven lyonization. D19 is part of a large Mennonite family (Family P060) in which she and two of her sisters were found to have manifestations of CSNB1 similar to that seen in their affected sons (Bech-Hansen and Pearce, 1993). All of the sons of these sisters are affected, and they carry either the affected grandfather's X chromosome or the grandmother's M1 haplotype X chromosome. These observations are consistent with

the grandmother's M1 X chromosome carrying a mutant CSNB1 gene, and the grandmother being a carrier of CSNB1. Subsequent to this publication, further analysis of the grandmother's side of the family has revealed other affected males carrying her M1 affected X chromosome. The large degree of homozygosity present in the region between DXS722 and DXS255 in the manifesting females from Family P060, eight out of eight markers tested, suggests that these daughters are homozygous by descent and therefore are carrying two copies of the same mutation in the same CSNB1 gene.

Physical Mapping:

A YAC and cosmid contig was constructed to cover the putative minimal genetic region for this CSNB1 locus between DXS722 and DXS255. The physical contig spans 2 Mb and contains 44 DNA markers in Xp11.23-p11.22 with an average resolution of one marker every 50 kb. The use of available genetic markers in construction of the physical map allowed the total integration of the genetic map with the physical map of this region (Figure 18). The contig contains 45 YAC and six cosmid clones and was assembled from the DNA marker content of YAC and cosmid clones, DNA marker retention in conventional and radiation hybrids, parsimony, and SEGMAP analysis. Six pairs of markers and one set of three markers could not be ordered using the current set of clones. Further work with cosmid clones should resolve the relative order of these markers. The use of the YAC ICRFy900C0874 in construction of this

physical contig allowed for physical integration with previously published contigs (Coleman *et al.*, 1994; Hagemann *et al.*, 1994; Knight *et al.*, 1994; Kwan *et al.*, 1995a) and extends the contiguous cloned region distally to cover the interval between DXS1264 and DXS255 in Xp11.3-p11.22, a distance of approximately 2.7 Mb.

Though YAC libraries containing more than 15 X-chromosome equivalents were exhaustively screened, the physical contig presented in Figure 18 represents only five-fold coverage of the region. This apparent discrepancy can be partly accounted for by the fact that highly unstable YACs were not included in Figure 18, as they did not provide any additional information to the map. When these unstable YACs are taken into account, screening of the E, F, and I YAC libraries isolated approximately nine-fold coverage of the region (data not shown). This value is in agreement with that obtained by Nagaraja *et al.* (1994) when they looked at the number of YACs recovered for several hundred STSs from the E, F, and I libraries. While the F library, derived from a human-hamster cell line, yielded the expected number of clones, the E and I libraries, derived from cells with five and four X chromosomes respectively, yielded numbers of YACs corresponding to an apparent content of only two X chromosomes. These unexpected results led to the suggestion that only the single active X chromosome is effectively cloned and that the specialized chromatin structure retained by the inactive X chromosomes may inhibit efficient cloning (Nagaraja *et al.*, 1994).

Chimerism and apparent instability were features observed among the YACs isolated for the development of the current contig and appears to be a regional problem. In contrast to this experience, one group, using the same set of libraries for the X chromosome, reported a 3.6 Mb contig in the Xp22.1 region which only had an 11% rate of chimerism and a 6-22% rate of internal deletions (Trump *et al.*, 1995). These rates are in sharp contrast to the 30% rate of chimerism and 40-80% rate of internal deletions detected in the present study. It would seem that the central and distal parts of the ZNF21-DXS255 region studied are particularly troublesome in YAC clones. In an attempt to overcome this type of problem, work with yeast RAD strains is continuing in an effort to develop an optimal recombination-deficient strain as a YAC host which will significantly decrease the rate of YAC chimerism (Haldi *et al.*, 1994) and stabilize deletion prone YACs (Kohno *et al.*, 1994). Construction of megabase-sized human YACs in the *rad52* strain MHY5201 has demonstrated a significant decrease in the rate of chimerism (8% compared to 50% in the wildtype host AB1380) (Haldi *et al.*, 1994). In addition, a yeast host disarmed in *rad51/rad52* and *rad1* has been shown to stabilize deletion prone YACs containing DNA from the human colour vision locus of Xq28 or rDNA units (Kohno *et al.*, 1994). It remains to be seen whether such recombination-deficient systems would reduce the rate of YAC chimerism and stabilize the YACs from the subregion identified in this study.

The central region between the genes TFE3 and SYP appears to be underrepresented in the YAC libraries that were tested relative to

other loci in this region. However, this does not appear to be the case in the cosmid libraries, with 18 and 14 clones being isolated for TFE3 and SYP respectively (data not shown). Cosmid clone supplementation was critical for construction of this contig as they closed one gap, confirmed marker order, gave indication of distances between markers, and added physical coverage. Others have also found cosmids to provide most of the needed supplementation for regions troublesome in YACs (Rogner *et al.*, 1994). However, it has also been noted that regions with the highest GC and gene content may contain sequences difficult to clone in both yeast and *E.coli* cloning systems (Rogner *et al.*, 1994), and as mentioned in the Results section, several cosmid clones from this region also contained obvious internal deletions. However, through a recent collaboration with Research Genetics, a BAC contig is being constructed in the Xp11.23-p11.22 region. From analysis of over 30 BACs with some 60 DNA markers from this region (both STSs and new ESTs), no evidence for chimerism or instability in these clones has been observed (Boycott *et al.*, manuscript in preparation). From these results, it would seem then that this particular cloning system has no problems handling high GC and gene content, in contrast to the yeast cloning system and the cosmid cloning system used in this region, probably due to the maintenance of genomic fragments as single copy plasmids in a recombination-deficient host. To gain additional confidence that the entire minimal region between DXS722 and DXS255 has been cloned and that the YAC and cosmid templates chosen for subsequent direct cDNA selection are fully representative,

it will be helpful to align the YAC map with the long range map of the region.

Preliminary long-range mapping data has been generated by a number of groups for this region of the X chromosome. Derry and coworkers (1994) estimated the distance from GATA1 to TFE3 to be 150 kb. Based on the contig presented here, we estimate the distance to be 270 kb. This group also estimate the distance between SYP and TFE3 to be 400 kb (Derry *et al.*, 1994), while we estimate this distance to be less than 100 kb. Clearly, more work needs to be done in this region to clarify these results. Meindl *et al.*, (1995) also estimated the distance between SYP and DXS255 to be about 900 kb which is close to our estimate of 860 kb. Very recently the same group (Schindelhauer *et al.*, 1996) has completed, in part using results from this study, a detailed long range map and cosmid contig of the region spanning OATL1 to DXS1039. Their map includes 20 markers in common with this map (Figure 18) and the relative order of all 20 of the markers is consistent between the two maps. The only discrepancies between the two maps involves the relative distances between the markers in the region between TFE3 and DXS1331. This work predicted the distance between TFE3 and SYP to be less than 100 kb while their map would indicate the distance to be 155 kb. In addition, while this work predicted the distance between DXS573 and DXS1331 to be 180 kb, their map would indicate that it is only 30 kb. Further analysis of cosmids from this region has confirmed their findings and Figures 19, 20, 31, 32, and 33 have been redrawn relative to Figure 18 to take the smaller

distance between DXS573 and DXS1331 into account. It is possible then that the physical map generated during this project is missing approximately 55 kb of coverage between the TFE3 and SYP genes. This will be discussed further in context of the minimal set of overlapping clones used in direct cDNA selection.

Several other groups have done preliminary contig work in this region. Mazzarella and Srivastava (1994) published several contigs physically linking a number of ESTs. Their preliminary placement of DXS1011E near GATA1 (their contig 'f') is confirmed and extended here. The map presented here was not able to order DXS1126 and DXS1240 relative to each other because they cosegregated in every clone tested. However, Kwan and coworkers (1995a) described two YACs not used in this study and were able to map DXS1126 distal to DXS1240. The cloning of the gene for Wiscott-Aldrich syndrome (Derry *et al.*, 1994) generated a framework map of the Xp11.23 region, aspects of which have been confirmed and extended during this work.

The order of the DNA markers OATL1, DXS7465E, (DXS1011E, WASP), DXS1358, DXS722, GATA1, DXS226, HRASP, (DXS1126, DXS1240), (DXS7469E, DXS1470), TFE3, (SYP, DXS1007E), DXS573, DXS1039, and DXS255 established from this effort is in full agreement with the consensus map of the 6th X Chromosome Workshop (Nelson *et al.*, 1995), except for one discrepancy. The Workshop consensus map places DXS7466E (MG44) distal to DXS7467E (MG21) while this map would place DXS7467E (MG21) distal to DXS7466E (MG44). The presence of DXS7467E (MG21) and

absence of DXS7466E (MG44) on one of the more distal YACs, yWXD6099, and the presence of DXS7466E (MG44) but not DXS7467E (MG21) on the more proximal YACs, yWXD1355 and yWXD3578, is the basis for our orientation of DXS7467E (MG21) distal to DXS7466E (MG44). The report by Schindelhauer *et al.* (1996) agrees with the orientation proposed for MG21 and MG44 by this work.

The physical contig resulting from this study provides the most extensive coverage of the Xp11.23-11.22 region to date. The published physical map contains STS markers at an average resolution of 50 kb, which more than achieves the physical mapping goals of the human genome project, and includes mapping information for 14 previously known genes, pseudogenes, and ESTs. With the application of region-specific gene identification techniques, this set of YAC and cosmid clones represents molecular access to new, previously unidentified genes from this region.

Isolation of New Genes by Direct cDNA Selection:

Using a minimal set of overlapping YAC and cosmid clones spanning the putative minimal region containing CSNB1 between DXS722 and DXS255, transcripts were identified from placental, retinal, frontal cortex, and fetal brain tissues by direct cDNA selection. In total, 44 novel ESTs were added to the region between DXS722 and DXS255, 13 from the placental selection and 31 from the combined retinal, frontal cortex, and fetal brain selection (JRL

selection). The retina cDNA set that was part of the JRL cDNA selection contains a 3' three base identifier which can be used to indicate which of the new cDNAs isolated from a pooled selection were derived from the retina cDNA set. Greater than 90% of the region-specific cDNAs were assessed for this identifier and not one was found to contain it. One explanation for this is that the retina cDNA set was poor in quality and therefore was not efficiently bound to the genomic template during the selection. However, several of the JRL cDNAs isolated were known or shown to be expressed in the retina (i.e. synaptophysin and KAT-1) and therefore cDNAs expressed in the retina were apparently isolated via the frontal cortex and fetal brain selection. Since the CSNB1 gene is most likely expressed in both the brain and the retina, the lack of cDNAs derived specifically from the retina cDNA set was not thought to have compromised the efforts of these studies.

Direct cDNA selection proved to be an effective and efficient way to add to the transcript information for the Xp11.23-p11.22 region. The direct cDNA selection procedure was 75% efficient (98 clones with unique genome positions/130 clones tested) but due to the use of chimeric YACs only 52% (67 region-specific clones/130 clones tested) of the clones were Xp11.23-p11.22 region-specific. The 25% of contaminating clones isolated during the two selections were represented by 32 clones, of which eight were sequenced. Sequence analysis using BLAST indicated that within this set of contaminating clones, two were 28S rRNA, one was 18S rRNA, one was 100% homologous to pBluescript, one was of mitochondrial origin, one was

an *Alu* repeat, and two were L1 repeats. It would seem that all of the potential contaminants (Lovett, 1994a; 1994b) that plague cDNA enrichments using large genomic clones are represented in this set but at a low enough abundance that they did not interfere with the analysis.

The use of chimeric YACs, detected and undetected, decreases the efficiency of isolation of region-specific transcripts and in this case decreased the isolation of Xp11.23-specific transcripts to 50%. While aware of the chimerism of yWXD2931, the isolation of transcripts not mapping to the human X chromosome would also suggest that YAC yWXD1908 is also chimeric. Both yWXD2931 (library M) and yWXD1908 (library I) were derived from YAC libraries constructed from the whole human genome, so it is possible that yWXD1908 also contains a portion of another human chromosome. FISH analysis of this YAC would confirm if this is indeed the case.

While adding coverage to a region at the time not covered by any other YAC, the use of the chimeric YAC yWXD2931, proved to add additional work to the project in the end due to the isolation and mapping of cDNA clones to a chromosome(s) other than the X. The YAC yWXD2931 is estimated to contain 1200 kb of DNA from another, as yet undetermined, chromosome (Boycott *et al.*, 1996). In total 24 cDNA clones have been isolated from placenta, retina, fetal brain, and frontal cortex tissue that are presumed to map to this 1200 kb derived from another chromosome. Identification of the location of this 1200 kb fragment of yWXD2931 by further FISH

analysis will give a map location to these 24 cDNAs and could be given to investigators who are interested in that region.

The minimal set of overlapping YACs and cosmids used for the placental and the JRL (retina, frontal cortex, fetal brain) selections were different due to the status of the map when the two selections were performed (the placental selection was carried out almost a year earlier than the JRL selection). Evaluating the clones used in the placental selection, in light of further work (Boycott *et al.*, 1996) and information that has become available recently (Schindelhauer *et al.*, 1996), has indicated that two gaps in physical coverage, one between TFE3 and SYP (maximum of 70 kb) and one at the DXS573 locus (50 kb), would have been present. Since the primary interest was in cDNA transcripts derived from brain or retinal tissue, these gaps do not necessarily compromise the integrity of the effort of isolating candidate genes for CSNB1. With respect to the JRL selection, a maximum of 55 kb from the region between TFE3 and SYP may have been missing based on information that has become available recently (Schindelhauer *et al.*, 1996). Cosmid sequencing within this region is now occurring (Nyakatura *et al.*, 1996) and access to this sequence through a collaboration with this group will determine whether candidate genes for CSNB1 have been missed because of this 55 kb gap in coverage.

Products from the JRL secondary selection were cloned using restriction enzyme digestion and ligation into pBluescript. Products from the placental secondary selection were reamplified with OLIGO1 adapted primers and cloned using the UDG cloning system. When

this system is used to clone products originally amplified with only one primer, only 50% of the products that were reamplified with the adapted primers were cloneable. This is due to the fact that the system uses directional cloning and the modification of the OLIGO1 primer generates two primers and both new 5' ends must be present to anneal to the vector. This problem is potentially cause for concern if the yield of cloned products is hindered. However, since the redundancy of the distal placental selection was 33% this would indicate that enough molecules with compatible ends were generated using the UDG system such that a number of clones were represented more than once. The redundancy could not be evaluated in the same way for the proximal placental selection because of the large number of transcripts isolated that mapped to another chromosome (due to the use of the chimeric YAC yWXD2931) which were not sequenced to see how many of them were represented more than once.

The number of clones generated from the YAC selections using either the UDG system (placental distal and proximal sets) or pBluescript (JRL Group 1, 2, and 4) were similar, with an average of 26 clones per selection of similar coverage. However, for the one cosmid selection that was performed, JRL-Group 3, over 200 clones were isolated. Considering the relative amount of target DNA present in the JRL selections, Group 1, 0.9 Mb, Group 2, 2.3 Mb, Group 4, 1.4 Mb and Group 3, 160 kb, this result is even more pronounced. Analysis of only 14 of these clones indicated a redundancy of 10%. It was thought that a large number of the unique clones had been sampled after analysis of only this small number of clones due to this

redundancy, the fact that many of the transcripts had a common hybridization pattern, and that two genes known to map to the region, SYP and TFE3, had both been identified in the set of 14 clones evaluated. The higher number of clones generated by cDNA enrichment with cosmids compared to YACs was most likely due to the purity, accurate quantitation, and copy number of the cosmid templates which resulted in more cDNA hybridizations and therefore a higher number of isolated transcripts to be cloned from the cosmid-based selection.

Initial attempts at direct cDNA selection of this region were performed with cloned libraries by amplifying the cDNA inserts out of the phage vector using the vector primers. Selection was carried out as described for the cDNA sets later used and reported here, and enrichment was observed for the positive reporter gene synaptophysin. Visualization of both the primary and secondary selected products on EtBr stained gels showed a smear of selected products as well as a very intense band corresponding the size of the product that would be generated from nonrecombinant phage. Cloning of these products using the UDG system yielded thousands of clones. Random single-pass sequencing of about 20 of these clones indicated that they were all products of priming nonrecombinant phage (data not shown). Because of their small size, less than 150 bp depending on the phage vector used in cDNA library construction, they were cloned preferentially. After several attempts to remove this contaminant by gridding and hybridization with different positive reporter genes, and gel purification prior to cloning, this

avenue was abandoned and cDNA sets were obtained for use in direct cDNA selection. While it is possible to carry out direct cDNA selection using very high quality libraries containing almost no non-recombinants, most investigators have moved to the use of primary cDNA sets that can be amplified by PCR.

We have generated a detailed high resolution map of the Xp11.23-p11.22 region spanning DXS722 and DXS255. This map contains genetic, physical, and transcript information including 44 new ESTs developed from cDNA fragments identified using direct cDNA selection and four ESTs developed from published sequence in Genbank (Derry *et al.*, 1994) and the gene RBM3 (Derry *et al.*, 1995). The new ESTs were initially tested on the critical members of the hybrid mapping panel and the YAC and cosmid templates used in the selection. These ESTs, as well as the published ESTs, were then extensively analyzed on the full set of physical clones in the relevant region so that they could be integrated into the map with greater confidence.

The ESTs, IS2, IS3, IS4, IS7 and the gene RBM3 have previously been sublocalized within Xp11.23 (Derry *et al.*, 1994). The relative order of IS2, IS3, IS4 and RBM3 suggested by this map is in agreement with that originally put forth by Derry and coworkers (1994). For IS7, Derry and coworkers (1994) placed this cDNA proximal to ZNF21 and distal to OATL1, however our somatic cell hybrid mapping data would suggest that IS7 is in fact distal to ZNF21.

Some of the ESTs that map close to each other are likely derived from the same gene, even though their sequences do not overlap. Therefore the ESTs described here represent a maximum set of new genes added to this region. Additional work with these ESTs, including expression profiles, transcript sizes estimated from Northern blots, and isolation of longer cDNAs, will need to be done to determine the exact number of genes to which these ESTs correspond.

The 5' and 3' sequences of cDNAs deposited in dbEST by the efforts of the I.M.A.G.E. Consortium and the Wash U-Merck cDNA sequencing project (Hillier *et al.*, 1996; Lennon *et al.*, 1996) have proven to be helpful for isolating longer cDNA clones and collapsing ESTs into the same transcript. The database search with JRL4-A1 isolated a number of cDNAs from a variety of tissues, one of which, 263773, was completely sequenced to facilitate determination the genomic structure as was described for evaluation of this gene as a candidate for CSNB1. Searches with placC10 (IS3) and placC11 yielded one and three hits, respectively. PlacC10 (IS3) pulled out the cDNA clone 297709 (from fetal lung tissue) and sequencing yielded 2035 bases, analysis of which indicated that it contained both placC10 (IS2) and placD12 (IS4) (data not shown). PlacC11 pulled out the cDNA clones 196653, 282407, and 231190 (from adult brain, multiple sclerosis, and pineal gland, respectively) which were sequenced to yield the same 659 bases with minor polymorphisms. Sequence analysis indicated that these clones contained only placC11

and not any of the neighbouring ESTs. Further characterization of these transcripts will potentially add two novel genes to this region.

Database searches with sequence from the newly isolated cDNA sequences revealed two other significant hits. The first hit was with the sequence from JRL4-A7 and indicated that it was 100% identical to EST016, which has been mapped to chromosome 1 (Sanders *et al.*, 1994, Genbank submission T25299). Mapping of this cDNA fragment as both a probe and an EST puts it in this region of the X chromosome. As a probe, at least two bands are evident after Southern hybridization with human genomic DNA. The EST gives two bands using human genomic DNA and less stringent PCR conditions, one of the expected size, 100 bp, that maps to the X chromosome, and one that is slightly larger and not on the X chromosome. Whether this band maps to chromosome 1 has not been determined. Further work with this cDNA fragment will help to determine whether JRL4-A7 is part of a gene family or whether EST016, JRL4-A7, or both, are part of transcribed pseudogenes. The second hit was with the sequence from JRL4-3 and indicated that almost 40 bases were 87% identical to the WASP mRNA sequence, however this cDNA fragment shares no other identity with WASP. This cDNA fragment does not appear to be chimeric due to the fact that primers designed from the sequence amplify the expected product. In addition, the EST JRL4-3 maps at least 50 kb proximal to WASP. Further investigation will be required to determine if this is a unique gene that has sequence, protein, or functional similarity to WASP. The remainder of the new cDNA sequences revealed no significant database similarities.

Only seven per cent of the sequences generated by direct cDNA selection (3 out of 44) found matches in the database with cDNA ends sequenced through the efforts of the I.M.A.G.E. Consortium and the Wash U-Merck cDNA sequencing project (Hillier *et al.*, 1996; Lennon *et al.*, 1996). However, this result does not provide evidence against the estimation that 50-70% of all human genes are now represented in dbEST (Schuler *et al.*, 1996) because the cDNA fragments isolated during this work were derived from random primed cDNA sets. Since the sequence information in dbEST is either 3', 5', or both, unless the cDNA (gene) is small, (i.e.. <1 kb and therefore almost the entire sequence is present in the database) the fragments of cDNA sequence generated in this work have a low probability of finding a match with the end sequences.

Work has also been done to isolate additional longer cDNAs using the new ESTs as probes on retina (kindly provided by Jeremy Nathans) and fetal brain (Strategene) cDNA libraries. Longer transcripts have been isolated for both JRL3-14 (4.0 kb) and JRL4-B6 (3.0 kb) from the fetal brain cDNA library. Both of these transcripts require further sequencing and Northern Blot analysis to determine if they contain any of the neighbouring ESTs. A longer transcript (1.3 kb) has also been isolated from the retina cDNA library for JRL3-4 and has been sequenced but found not to contain any of the neighbouring ESTs.

Enrichments of several thousand-fold to more than 100,000-fold of particular reporter genes have been described using direct cDNA selection (Morgan *et al.*, 1992; Lovett, 1994). As a measure of the

enrichment efficiency achieved by the direct cDNA selection procedure described in these studies, the frequencies of the ESTs JRL3-14, JRL4-B6, and JRL3-4 in the direct cDNA selection libraries were compared to their frequencies in the fetal brain or retina cDNA libraries. The EST JRL3-14 represented 7% (1/14 clones) of the direct cDNA selection Group 3 library while it was present in the fetal brain cDNA library at a level of 0.0002% (1/500,000 plaques). This represents a practical enrichment of approximately 35,000-fold. The levels of enrichment for the ESTs JRL4-B6 and JRL3-4 were evaluated in a similar manner and were determined to be 10,000-fold and 70,000-fold, respectively. These three evaluations of enrichment indicate the efficiency of the direct cDNA selection procedure used in these studies and the results were comparable to that achieved by others (Morgan *et al.*, 1992; Lovett, 1994).

The integrated map between OATL1 and DXS255, encompassing the putative minimal genetic region for CSNB1 in Family P060, now contains 91 DNA markers including 54 ESTs, five genes, two pseudogenes, and eight STRs and is an integrated and detailed source of region-specific genetic, physical, and transcript information. This map will be a great resource for CSNB1 positional cloning efforts and will also be helpful in establishing the total genomic sequence of Xp11.23-p11.22.

Characterization of a Candidate Gene for CSNB1:

The gene KAT1, corresponding to the cDNA 263773, was chosen for analysis as a candidate gene for CSNB1 in Family P060. KAT1 is a highly expressed transcript present in all tissues tested, including the retina. The genomic structure of this gene was elucidated by genomic DNA amplification using cDNA based primers, cosmid sequencing, and walking by linker ligation. KAT1 contains seven exons and two transcripts, differing by 68 bp, are derived from this gene by alternative splicing in Exon1. The putative ORF of 262 amino acids begins with an initiator methionine in Exon 2. This ORF utilizes the first methionine and the surrounding bases conform to the Kozak consensus sequence (Kozak, 1989). Since the ORF does not begin until Exon 2, the alternative splicing event alters the 5' untranslated region of the gene. The function of the alternative splice site is unknown but may involve cellular localization of the two different transcripts and the resulting translated protein. The larger transcript was seen in all tissues in which KAT1 has been shown to be expressed, including the retina, while the smaller transcript was only seen in brain (RACE results) and fetal liver/spleen (database analysis results). This observation may reflect the lack of expression of the smaller transcript of KAT1 in the other tissues, or lower level expression of this transcript and consequently its underrepresentation in the I.M.A.G.E. Consortium cDNA libraries.

Analysis of the cDNA nucleotide sequence in the nr database by BLASTn (Altschul *et al.*, 1990) indicated that Exons 2 and 3 of KAT1

had weak sequence similarity to a calmitine=mitochondrial calcium-binding protein. Analysis of the ORF in the nr database by BLASTP and the db-links/CRSeqAnnot.fa database by BEAUTY (Worley *et al.*, 1995) indicated that amino acids at postions 100-150, located in Exon 5, had weak similarity to over 150 proteins including several calcium binding proteins. The significance of these weak similarities to the function of KAT1 remains to be determined.

Mutation Analysis:

Most successful positional cloning projects to date have relied on either cytogenetic rearrangements or the identification of patients carrying deletions of tens or hundreds of kilobases (Collins, 1995). As an efficient and easy way to quickly assess the presence of such rearrangements in CSNB1 families, DNA from representatives from each of these families was analyzed by Southern hybridization using the full length cDNA of the candidate gene KAT1. Unfortunately no such rearrangements or deletions were visible with this analysis. Deletion analysis was also carried out with a number of the cDNA transcripts isolated by the direct cDNA selection procedure including placB3-2, placE2, placC11, JRL4-1, JRL4-2, and JRL4-5 (data not shown) and no rearrangements or deletions were detected.

Because of the absence of a visible deletion or rearrangement, the presence of subtle base changes, small insertions or deletions in KAT1 in affected individuals was assessed by direct sequencing of the exons and splice junctions. Direct sequencing was chosen as the

mutation detection method because the detection of nucleotide changes in a small number of affected males for an X-linked disorder it is relatively simple and virtually 100% sensitive. Analysis of the exon and splice junction sequences of KAT1 in representative affected and unaffected control individuals from Family P060, as well as an AIED family (Family 290), detected no base changes, small insertions, or deletions. Based on this analysis, the gene KAT1 can essentially be eliminated as the gene responsible for the clinical manifestations of CSNB1 or AIED in either of these families. However, KAT1 cannot be totally excluded as the CSNB1 gene because the presence of a base change, insertion, or deletion in the gene's regulatory region that would affect protein synthesis has not been investigated.

Future Studies:

Though these efforts have not yet succeeded in isolating the gene responsible for CSNB1 in Family P060, the resources are now in place for successful and rapid completion of this positional cloning project. The ESTs isolated during this work are a valuable resource that will facilitate access to the candidate genes in the region. To accomplish this, the ESTs must first be reduced to a minimal set of genes that can be further characterized and analyzed for mutations. There are a number of ways to approach this and probably all avenues should be pursued, if possible, to ensure that the most information is gained in the shortest amount of time. Isolation of longer cDNA clones by

cDNA library or database screening with the ESTs is already underway and can reduce the ESTs into a minimal set as well as provide the cDNA clones as resources for future studies. Alongside this, PCR amplification of the cDNA libraries using primers from different ESTs will help reduce the ESTs into a minimal set quickly and highlight which ESTs should be used to isolate longer clones by eliminating some of the redundancy in the gene set. Once the longer transcripts have been isolated, Northern Blot analyses and RACE amplifications will facilitate the isolation of full-length transcripts which can serve as the starting point for characterization of the genomic structure. Complementing this strategy are the efforts of the German Human Genome Project to sequence this region of Xp11.23. In collaboration with this genome group, DNA sequence will be provided to us which we can use to quickly characterize the genes to which ESTs of interest correspond.

Several longer transcripts corresponding to different ESTs have already been isolated, including placC11, placC10 (placD12), JRL3-14, JRL4-B6, and JRL3-4, and are available for further analysis. Whether the genomic structure of these genes will be characterized as was the structure of KAT1 or by direct sequencing of the corresponding genomic clones remains to be determined. Either way, several of these genes represent good candidates for CSNB1 in Family P060 and will be given top priority for mutation analysis. In addition, all of the 44 ESTs and the available longer cDNA clones should be used as hybridization probes on Southern blots of digested patient DNA to

ensure that an obvious rearrangement is not being missed that would point out the location of the CSNB1 gene.

Future efforts by our group and others, will give this region the distinction of being among the first portions of the human genome to be sequenced and have its gene content fully detailed. Analysis of Xp11.23 so far would indicate that it is very gene rich and mutations in several of these genes have already been implicated in human genetic disease (Figure 2). Identification of the gene involved in CSNB1 in Family P060 will resolve the issues regarding genetic heterogeneity, genetic variability, and clinical diagnoses for X-linked CSNB as well as give much insight into the molecular mechanisms behind this form of human retinopathy.

CHAPTER 6 - SUMMARY

X-linked congenital stationary night blindness (CSNB1) is an eye disorder that includes impairment of night vision, reduced visual acuity and, in some cases, myopia and congenital nystagmus. Electroretinography reveals a marked reduction of the b-wave in affected individuals suggesting that X-linked CSNB is due to a molecular defect in the bipolar layer of the retina. In an effort to isolate the gene responsible for this disease, a positional cloning strategy has been used to reduce the minimal genetic region for CSNB1, construct a physical contig encompassing the CSNB1 locus, isolate transcripts by direct cDNA selection, and characterize and analyze one of these genes for mutations in affected individuals.

Based on previous studies of over 20 families diagnosed with X-linked CSNB, a CSNB1 locus was mapped to Xp11.23-p11.22. Using a panel of radiation and conventional somatic cell hybrids, a set of 13 new polymorphic markers were mapped to this region. The proximal limit of CSNB1 in Family P060 was refined to the marker DXS8023 by analysis of a crossover event in individual D3. The distal limit of CSNB1 in Family P060 was refined to the marker DXS722 by analysis of a crossover event in individual D19. In Family P060, then, the minimal genetic region containing the CSNB1 locus was refined to the region bounded telomerically by DXS722 and centromerically by DXS8023, a region estimated to be less than 2.2 Mb. A number of crossover events in other CSNB1 families have placed the proximal limit to CSNB1 at DXS255. Based on this published information,

initial efforts toward the isolation of candidate genes for CSNB1 have focused on the putative minimal region for CSNB1 between DXS722 and DXS255, a region spanning less than 1.5 Mb.

To construct a physical contig spanning this putative minimal region, YAC and cosmid clones were identified. The contig was constructed from the content of 44 DNA markers (including 20 new sequence tagged sites (STSs) and 12 genes or expressed sequences) for 45 YACs and 6 cosmids. The contig spans the distal marker ZNF21 and extends proximally to include DXS255, with an average STS marker density of one every 50 kb. Analysis of over 300 YAC and cosmid clones from this region was required to provide 5-fold coverage of what proved to be a difficult region of the X chromosome to clone.

A minimal set of overlapping clones were chosen from the physical contig to enrich for cDNA transcripts expressed in the frontal cortex, fetal brain, retina, and placenta by direct cDNA selection. One hundred and thirty cDNA clones were isolated and analyzed using this approach. Those clones that were identified as being region-specific were sequenced and 44 unique expressed sequence tags (ESTs) were added to the transcription map of the region in context of the physical contig. In addition, several new genes and ESTs recently placed in this region by other investigators were also integrated into the transcript map to generate a detailed genetic, physical, and transcript map of this portion of the Xp11.23-p11.22 region.

Through database analysis, one of the cDNA transcripts isolated by direct cDNA selection, JRL4-A1, identified a longer cDNA (263773) partially characterized by the Wash U-Merck cDNA sequencing project. The expression profile of JRL4-A1 indicated that it was expressed in many tissues, including both the retina and the brain. Because of its expression pattern and physical location between the markers DXS1126 and DXS255, this gene was chosen to be characterized in detail. Northern Blot analysis revealed a transcript of 1100 bp and 5' RACE experiments indicated that the entire 5' end of this gene (KAT1) was contained in the 263773 clone. cDNA based primers, cosmid sequencing, and walking by linker ligation were used to identify exon/intron boundaries and the gene was found to contain seven exons. Alternative splicing involving Exon 1 generates two different transcripts from this gene. KAT1 has an ORF of 262 amino acids that begins with an initiator methionine codon in Exon 2.

The 263773 clone was used as a probe to scan for deletions in the 15 CSNB1 families by Southern analysis, and no aberrant bands were detected. The exons of KAT1 were then directly sequenced twice in representative affected and unaffected control individuals from Family P060. From the results of this analysis, KAT1 was excluded as the gene responsible for CSNB1 in Family P060.

Future efforts to isolate the gene for CSNB1 in Family P060 will be facilitated by the additional resources, in the form of novel ESTs and a detailed genetic, physical, and transcript map of the region, generated by this work. Continued analysis of the ESTs in the region in an effort to reduce them into common transcripts will produce

additional candidates genes for mutation analysis in CSNB1 patients, as well as increase the density of the gene map in Xp11.23-p11.22.

REFERENCES

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

Aldred, M. A., Dry, K. L., Sharp, D. M., Van Dorp, D. B., Brown, J., Hardwick, L. J., Lester, D. H., Pryde, F. E., Teague, P. W., Jay, M., Bird, C., Jay, B. and Wright, A. F. (1992a). Linkage analysis in X-linked Congenital Stationary Night Blindness. *Genomics* **14**:99-104.

Aldred, M. A., Dry, K. L., Sharp, D. M., Van Dorp, D. B., Brown, J., Hardwick, L. J., Lester, D. H., Pryde, F. E., Teague, P. W., Jay, M., Bird, A. C., Jay, B. and Wright, A. F. (1992b). Linkage analysis in X-linked congenital stationary night blindness (erratum). *Genomics* **22**:255.

Alitalo, T., Kruse, T. A., Forsius, H., Eriksson, A. W. and de la Chapelle, A. (1991). Localization of the Aland Island Eye disease locus to the pericentromeric region of the X chromosome by linkage analysis. *Am. J. Hum. Genet* **48**:31-38.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* **215**:403-410.

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

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

Barker, D. F., and Fain, P. R. (1993). Definition and mapping of STSs at STR and RFLP loci in Xp11-Xq22. *Genomics* **18**:712-716.

Barrett, J. H. (1992). Genetic mapping based on radiation hybrid data. *Genomics* **13**:95-103.

Beeh-Hansen, N. T., Field, L. L., Schramm, A. M., Reedyk, M., Craig, I. W., Fraser, N. J. and Pearce, W. G. (1990). A locus for X-linked Congenital Stationary Night Blindness is located on the proximal portion of the short arm of the X chromosome. *Hum. Genet.* **84**:406-408.

Bech-Hansen, N. T., Marshall, K. J., Moore, B. J., Schramm, A. M., Field, L. L., Breakefield, X. O. and Pearce, W. G. (1991). Localization of CSNB1 to region Xp11.23-Xp11.22 proximal to MAOA,B. In human gene mapping II. *Cytogenet. Cell Genet.* **58**:2056.

Bech-Hansen, N. T., Moore, B., J. and Pearce, W. G. (1992). Mapping of locus for X-linked Congenital Stationary Night Blindness (CSNB1) proximal to DXS7. *Genomics* **12**:409-411.

Bech-Hansen, N. T. and Pearce, W. G. (1993). Manifestations of X-linked Congenital Stationary Night Blindness in three daughters of an affected male: Demonstration of Homozygosity. *Am. J. Hum. Genet* **52**:71-77.

Beckmann, H., Su, L.-K. and Kadesch, T. (1990). TFE3: A helix-loop-helix protein that activates transcription through the immunoglobulin enhancer uE3 motif. *Genes Dev.* **4**:167-179.

Benham, F., Hart, K., Crolla, J., Bobrow, M., Francavilla, M., and Goodfellow, P. N. (1989). A method for generating hybrids containing nonselected fragments of human chromosomes. *Genomics* **4**:509-517.

Bergen, A. A. B., Meire, F., ten Brink, J., Schuurman, E. J. M., van Ommen, G.-J. B. and Delleman, J. W. (1993). Additional evidence for a gene locus for progressive cone dystrophy with late rod involvement in Xp21.1-p11.3. *Genomics* **18**:463-464.

Bergen, A. A. B., Kestelyn, P., Leys, M. and Meire, F. (1994). Identification of a key recombinant which assigns the incomplete congential stationary night blindness gene proximal to MAOB. *Med. Genet.* **31**:580-582.

Bergen, A. A. B., Brink, J. B. t., Riemsiag, F., Schuurman, E. J. M. and Tijmes, N. (1995). Localization of a novel X-linked congenital stationary night blindness locus: Close linkage to the RP3 type retinitis pigmentosa gene region. *Hum. Molec. Genet.* **4**:931-935.

Berger, W., Meindl, A., van de Pol, T. J. R., Cremers, F. P. M., Ropers, H. H., Doerner, C., Monaco, A., Bergen, A. A. B., Lebo, R., Warburg, M., Zergollern, L., Lorenz, B., Gal, A., Bleeker-Wagemakers, E. M., and Meitinger, T. (1992). Isolation of a candidate gene for Norrie disease by positional cloning. *Nature Genet.* **1**:199-203.

Berger, W., van Duijnhoven, G., Pinckers, A., Smits, A., Ropers, H.-H. and Cremers, F. (1995). Linkage analysis in a Dutch family with X-linked recessive congenital stationary night blindness (XL-CSNB). *Hum. Genet.* **95**:67-70.

Blanchard, M. M., Sloan, D. and Nowotny, V. (1991). Advances in Lab. Automat. Robotics. "*In* Automation of the Mapping of the Human Genome" (J. R. Strimaitis, and J. N. Little, Eds.), Vol. 8. pp. 333-345, Zymark Corp., Hopkinton.

Blanchard, M. M., Taillon, M., P., Nowotny, P. and Nowotny, V. (1993). PCR buffer optimization with uniform temperature regimen to facilitate automation. *PCR Methods Appl.* **2**:234-240.

Boycott, K. M. (1996). The use of SEPARIDE gel matrix for separation of PCR products. *Focus* **18**:74.

Boycott, K. M., Halley, G. R., Schlessinger, D. and Bech-Hansen, N. T. (1996). A 2-Megabase physical contig incorporating 43 DNA markers on the human X chromosome at p11.23-11.22 from ZNF21 to DXS255. *Genomics* **33**:488-497.

Boycott, K. M., Moore, B. J., Gratton, K. J., Stoddart, K. L., Roland, B., and Bech-Hansen, N. T. Integration of 101 DNA markers across human Xp11 using a panel of somatic cell hybrids. *Cytogenet. Cell Genet.* (in press).

Britten, R. J. and Davidson, E. H. (1985). Hybridization strategy. *In* "Nucleic Acid Hybridization: A Practical Approach" (B. D. Hames, and S. J. Higgins, Eds.), pp.3-14, IRL Press, Oxford.

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.

Bull, P. C., Thomas, G. R., Rommens, J. M., Forbes, J. R. and Wilson Cox, D. (1993). The Wilson disease gene is a putative copper transporting P-type ATPase similar to the Menkes gene. *Nature Genet.* **5**:327-337.

Burke, D. T., Carle, G. F. and Olson, M. V. (1987). Cloning of large segments of exogenous DNA into yeast by means of artificial chromosome vectors. *Science* **230**:806-812.

Caiulo, A., Nicolis, S., Bianchi, P., Zuffardi, O., Bardoni, B., Maraschio, P., Ottolenghi, S., Camerino, G. and Giglioni, B. (1991). Mapping the gene encoding the human erythroid transcription factor NFE1-GF1 to Xp11.23. *Hum. Genet.* **86**:388-390.

Carr, R. E. (1974). Congenital Stationary Night Blindness. *Trans. Am. Ophthalmol. Soc.* **72**:449-487.

Chumakov, I., Rigault, P., Guillou, S., Ougen, P., Billaut, A., Guasconi, G., Gervy, P., LeGall, I., Soularue, P., Grinas, L., Bougueleret, L., Bellanne-Chantelot, C., Lacroix, B., Barillot, E., Gesnouin, P., Pook, S., Vayssiex, G., Frelat, G., Schmitz, A., Sambucy, J.-L., Bosch, A., Estivill, X., Weissenbach, J., Vignal, A., Riethman, H., Cox, D., Patterson, D., Gardiner, K., Hattori, M., Sakaki, Y., Ichikawa, H., Ohki, M., Le Paslier, D., Heilig, R., Antonarakis, S., and Cohen, D. (1992). Continuum of overlapping clones spanning the entire human chromosome 21q. *Nature* **359**:380-387.

Church, D. M., Banks, L. T., Rogers, A. C., Graw, S. L., Houseman, D. E., Gusella, J. F. and Buckler, A. J. (1993). Identification of human chromosome 9 specific genes using exon amplification. *Hum. Molec. Genet.* **2**:1915-1920.

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.

Clark, J., Rocques, P. J., Crew, A. J., Gill, S., Shipley, J., Chan, A. M.-L., Gusterson, B. A. and Cooper, C. S. (1994). Identification of novel genes, SYT and SSX, involved in the t(X;18)(p11.2;q11.2) translocation found in human synovial sarcoma. *Nature Genet.* **7**:502-508.

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

Coleman, M. P., Nemeth, A. H., Campbell, L., Raut, C. P., Weissenbach, J. and Davies, K. E. (1994). A 1.8-Mb YAC contig in Xp11.23: Identification of CpG islands and physical mapping of CA repeats in a region of high gene density. *Genomics* **21**:337-343.

Collins, F. S. (1992). Positional cloning: Let's not call it reverse anymore. *Nature Genet.* **1**:3-6.

Collins, F. S. (1995). Positional cloning moves from perditional to traditional. *Nature Genet.* **9**:347-349.

Cooper, D. N., Smith, B. A., Cooke, H. J., Niemann, S. and Schmidtke, J. (1985). An estimate of unique DNA sequence heterozygosity in the human genome. *Hum. Genet.* **69**:201-205.

Cotton, R. G. H., Rodrigues, N. R. and Campbell, R. D. (1988). Reactivity of cytosine and thymine in single-base-pair mismatches with hydroxylamine and osmium tetroxide and its application to the study of mutations. *Proc. Natl. Acad. Sci. USA* **85**:4397-4401.

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

Crew, A. J., Clark, J., Fisher, C., Gill, S., Grimer, R., Chand, A., Shipley, J., Gusterson, B. A. and Cooper, C. S. (1995). Fusion of SYT to two genes, SSX1 and SSX2, encoding proteins with homolgy to the Kruppel-associated box in human synovial sarcoma. *EMBO J.* **14**:2333-2340.

de Leeuw, B., Balemans, M., Olde Weghuis, D., Seruca, R., Janz, M., Geraghty, M. T., Gilgenkranz, S., Ropers, H. H. and Geurts van Kessel, A. (1994). Molecular cloning of the synovial sarcoma-specific translocation (X;18)(p11.2;q11.2) breakpoint. *Hum. Molec. Genet.* **3**:745-749.

de Leeuw, B., Balemans, M., Olde Weghuis, D. and Geurts van Kessel, A. (1995). Identification of two alternative fusion genes, SYT-SSX1 and SYT-SSX2, in t(X;18)(p11.2;q11.2)-positive synovial sarcomas. *Hum. Molec. Genet.* **4**:1097-1099.

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

Derry, J. M., Kerns, J. a. and Francke, U. (1995). RBM3, a novel human gene in Xp11.23 with a putative RNA-binding domain. *Hum. Molec. Genet.* **4**:2307-2311.

Dietz, H. C., Cutting, G. R., Pyeritz, R. E., Maslen, C. L., Sakai, L. Y., Corson, G. M., Puffenberger, E. G., Hamosh, A., Nanthakumar, E. J., Curristin, S. M., Stetten, G., Meyers, D. A. and Francomano, C. A. (1991). Marfan syndrome caused by a recurrent *de novo* missense mutation in the fibrillin gene. *Nature* **353**:337-339.

Donnelly, A., Kozman, H., Gedeon, A. K., Webb, S., Lynch, M., Sutherland, G. R., Richards, R. I. and Mulley, J. C. (1994). A linkage map of microsatellite markers on the human X chromosome. *Genomics* **20**:363-370.

Drwinga, H. L., Toji, L. H., Kim, C. H., Greene, A. E. and Mulivor, R. A. (1993). NIGMS human/rodent somatic cell hybrid mapping panels 1 and 2. *Genomics* **16**:311-314.

Dryja, T. P., Berson, E. L., Rao, V. R. and Oprian, D. D. (1993). Heterozygous missense mutation in the rhodopsin gene as a cause of Congential Stationary Night Blindness. *Nature Genet.* **4**:280-283.

Duyk, G. M., Kim, S., Myers, R. M. and Cox, D. R. (1990). Exon trapping: A genetic screen to identify candidate transcribed sequences in cloned mammalian genomic DNA. *Proc. Natl. Acad. Sci. USA* **87**:8995-8999.

Elenius, V., Eriksson, A. and Forsius, H. (1968). ERG in a case of X-chromosomal pigment deficiency of fundas in combination with myopia, dyschromatopsia and defective dark-adaptation. *In* "The Clinical Value of Electroretingraphy: Proceedings of the 5th ISCERG Symposium, Ghent, 1966" (J. Francois, Ed.), pp. 369-377, Karger, New York.

Ellison, K. A., Roth, J., McCabe, E. R. B., Chinault, A. C. and Zoghbi, H. Y. (1993). Isolation of a yeast artificial chromosome contig spanning the X chromosomal translocation breakpoint in a patient with Rett syndrome. *Am. J. Med. Genet.* **47**:1124-1134.

Elvin, P., Slynn, G., Black, D., Graham, A., Butler, R., Riley, J., Anand, R. and Markham, A. F. (1990). Isolation of cDNA clones using yeast artificial chromosome probes. *Nucleic Acids Res.* **18**:3913-3917.

Feinberg, A. P. and Vogelstein, B. (1983). A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. *Analytical Biochem.* **132**:6-13.

Feinberg, A. P. and Vogelstein, B. (1984). Addendum: A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. *Analytical Biochem.* **137**:266-267.

Feiss, M., Fisher, R. A., Crayton, M. A., and Egner, C. (1977). Packaging of the bacteriophage lambda chromosome: Effect of chromosome length. *Virology* **77**:281-293.

Fisher, S. E., van Bakel, I., Lloyd, S. E., Pearce, S. H. S., Thakker, R. V. and Craig, I. W. (1995). Cloning and characterization of CLCN5, the human kidney chloride channel gene implicated in Dent disease (an X-linked hereditary nephrolithiasis). *Genomics* **9**:598-606.

Foote, S., Vollrath, D., Hilton, A. and Page, D. C. (1992). The human Y chromosome: overlapping DNA clones spanning the euchromatic region. *Science* **258**:60-66.

Fraser, N. J., Boyd, Y., and Craig, I. (1989). Isolation and characterization of a human variable copy number tandem repeat at Xcen-p11.22. *Genomics* **5**:144-148.

Frohman, M. A., Dush, M. K., and Martin, G. R. (1988). Rapid production of full-length cDNAs from rare transcripts: Amplification using a single gene-specific oligonucleotide primer. *Proc. Natl. Acad. Sci. USA* **85**:8998-9002.

Futreal, P. A., Cochran, C., Rosenthal, J., Miki, Y., Swenson, J., Hobbs, M., Bennett, L. M., Haugen-Strano, A., Marks, J., Barrett, J. C., Tavtigian, S. V., Shattuck-Eidens, D., Kamb, A., Skolnick, M. and Wiseman, R. W. (1994). Isolation of a delivered homeobox gene, MOX1, from the BRCA1 region on 17q21 by solution hybrid capture. *Hum. Molec. Genet.* **3**:1359-1364.

Gal, A., Schinzel, A., Orth, U., Fraser, N. A., Mollica, F., Craig, I. W., Kruse, T., Machler, M., Neugebauer, M. and Bleeker-Wagemakers, L. M. (1989). Gene of X-chromosomal Congenital Stationary Night Blindness is closely linked to DXS7 on Xp. *Hum. Genet.* **81**:315-318.

Gal, A., Orth, U., Baehr, W., Schwinger, E. and Rosenberg, T. (1994). Heterozygous missense mutation in the rod cGMP phosphodiesterase β-subunit gene in autosomal dominant stationary night blindness. *Nature Genet.* **7**:64-67.

Geraghty, M. T., Brody, L. C., Martin, L. S., Marble, M., Kearns, W., Pearson, P., Monaco, A. P., Lehrach, H. and Valle, D. (1993). The isolation of cDNAs from OATL1 at Xp11.2 using a 480-kb YAC. *Genomics* **16**:440-446.

Gibbs, R. A. (1995). Pressing ahead with human genome sequencing. *Nature Genet.* **11**:121-125.

Glass, I. A., Good, P., Coleman, M. P., Fullwood, P., Giles, M. G., Lindsay, S., Nemeth, A. H., Davies, K. E., Willshaw, H. A., Fielder, A., Kilpatrick, M. and Farndon, P. A. (1993). Genetic mapping of a cone and rod dysfunction (Aland Island eye disease) to the proximal short arm of the human X chromosome. *Journal of Med. Genet.* **30**:1044-1050.

Goodfellow, P. N. (1991). Irradiation and fusion gene transfer. *In* "Methods in Molecular Biology" (E. J. Murray, Ed.), Vol. 7, pp. 53-61, Humana Press Inc, Clifton.

Goss, S. J. and Harris, H. (1975). New method for mapping genes in human chromosomes. *Nature* **255**:680-684.

Green, E. D. and Green, P. (1991). Sequence-tagged site (STS) content mapping of human chromosomes: Theoretical considerations and early experiences. *PCR Methods Appl.* **1**:77-90.

Green, E. D., Riethman, H. C., Dutchik, J. E. and Olson, M. V. (1991). Detection and characterization of chimeric yeast artificial-chromosome clones. *Genomics* **11**:658-669.

Grompe, M. (1993). The rapid detection of unknown mutations in nucleic acids. *Nature Genet.* **5**:111-117.

Gyapay, G., Morissette, J., Vignal, A., Dib, C., Fizames, C., Missasseau, P., Marc, S., Bernardi, G., Lathrop, M., and Weissenbach, J. (1994). The 1993-1994 Genethon human genetic linkage map. *Nature Genet* **7**:246-339.

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

Gyllensten, U. B. and Erlich, H. A. (1988). Generation of single-stranded DNA by the polymerase chain reaction and its application to direct sequencing of the HLA-DQA locus. *Proc. Natl. Acad. Sci. USA* **85**:7652-7656.

Hagemann, T., Surosky, R., Monaco, A. P., Lehrach, H., Rosen, F. S. and Kwan, S.-P. (1994). Physical mapping in a YAC contig of 11 markers on the human X chromosome in Xp11.23. *Genomics* **21**:262-265.

Haldi, M., Perrot, V., Saumier, M., Desai, T., Cohen, D., Cherif, D., Ward, D. and Lander, E. S. (1994). Large human YACs constructed in a *rad52* strain show a reduced rate of chimerism. *Genomics* **24**:478-484.

Hamada, H., Petrino, M. G., Kakunaga, T., Seidman, M. and Stollar, B. D. (1984). Characterization of genomic poly(dT-dG)•poly(dC-dA) sequences: Structure, organization, and conformation. *Mol. Cellular Biol.* **4**:2610-2621.

Hamaguchi, M., Sakamoto, H., Tsuruta, H., Sasaki, H., Muto, T., Sugimura, T. and Terada, M. (1992). Establishment of a highly sensitive and specific exon-trapping system. *Proc. Natl. Acad. Sci. USA* **89**:9779-9783.

Heckenlively, J. R., Martin, D. A. and Rosenbaum, A. L. (1983). Loss of electroretinographic oscillatory potentials, optic atrophy, and dysplasia in congenital stationary night blindness. *Am. J. Ophthal.* **96**:526-534.

Herrmann, K., Meindl, A., Apfelstedt-Sylla, E., Wissinger, B., Ciccodicola, A., Lorenz, B., Wittwer, B., Hergersberg, M., D'Urso, M. and Meitinger, T. (1996). RPGR mutation analysis in patients with retinitis pigmentosa and congenital stationary night blindness. *Am. J. Hum. Genet* **59**:A263.

Huebner, K., Druck, T., Croce, C.M., and Thiesen, H.-J. (1991). Twenty-seven nonvoerlapping zinc finger cDNAs from human T cells map to nine different chromosomes with apparent clustering. *Am J. Hum Genet.* **48**:726-740.

Hill, D. A., Arbel, K. F. and Berson, E. L. (1974). Cone electroretinograms in congenital nyctalopia with myopia. *Am. J. of Ophthal.* **78**:127-136.

Hillier, L., Lennon, G., Becker, M., Bonaldo, M. F., Chiapelli, B., Chissoe, S., Dietrich, N., DuBuque, T., Favello, A., Gish, W., Hawkins, M., Hultman, M., Kucaba, T., Lacy, M., Le, M., Le, N., Mardis, E., Moore, B., Morris, M., Parsons, J., Prange, C., Rifkin, L., Rohlfing, T., Schellenberg, K., Soares, B. M., Tan, F., Thierry-Meg, J., Trevaskis, E., Underwood, K., Wohldman, P., Waterston, R., Wilson, R., and Marra, M. (1996). Generation and analysis of 280,000 human expressed sequence tags. *Genome Res.* **6**:807-828.

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

Hohn, B., and Collins, J. (1980). A small cosmid for efficient cloning of large DNA fragments. *Gene* **11**:291-298.

Hong, H. K., Ferrell, R. E. and Gorin, M. B. (1994). Clinical diversity and chromosomal localization of X-linked cone dystrophy (COD1). *Am. J. Hum. Genet* **55**:1173-1181.

Hudson, T. J., Stein, L. D., Gerety, S. S. and al., e. (1995). An STS-based map of the human genome. *Science* **270**:1945-1954.

Ingle, C., Williamson, R., de la Chapelle, A., Herva, R. R., Haapala, K., Bates, G., Willard, H. F., Pearson, P. and Davies, K. E. (1985). Mapping DNA sequences in a human X chromosome deletion which extends across the region of the Duchenne muscular dystrophy mutation. *Am. J. Hum. Genet* **37**:451-462.

Ioannou, P. A., Amemiya, C. T., Garnes, J., Kroisel, P. M., Shizuya, H., Chen, C., Batzer, M. A. and de Jong, P. J. (1994). A new bacteriophage P1-derived vector for the propagation of large human DNA fragments. *Nature Genet.* **6**:84-89.

Jeffreys, A. J., Wilson, V., Thein, S. L., Weatherall, D. J. and Ponder, B. A. J. (1986). DNA "fingerprints" and segregation analysis of multiple markers in human pedigrees. *Am. J. Hum. Genet* **39**:25-37.

Kere, J., Nagaraja, R., Mumm, S., Ciccodicola, A., D'Urso, M. and Schlessinger, D. (1992). Mapping human chromosomes by walking with sequence-tagged sites from end fragments of yeast artificial chromosome inserts. *Genomics* **14**:241-248.

Khouri, G., Mets, M. B., Smith, V. C., Wendell, M. and Pass, A. S. (1988). X-linked Congenital Stationary Night Blindness: Review and report of a family with hyperopia. *Arch. Ophthalmol.* **106**:1417-1422.

Kim, U.-J., Shizuya, H., de Jong, P., Birren, B. and Simon, M. I. (1992). Stable propagation of cosmid sized human DNA inserts in an F factor-based vector. *Nucleic Acids Res.* **20**:1083-1085.

Knight, J. C., Grimaldi, G., Thiesen, H.-J., Bech-Hansen, N. T., Fletcher, C. D. M. and Coleman, M. P. (1994). Clustered organization of Kruppel Zinc-Finger genes at Xp11.23, flanking a translocation brekpoint at OATL1: a physical map with locus assignments for ZNF21, ZNF41, ZNF81, and ELK1. *Genomics* **21**:180-187.

Kohno, K., Wada, M., Schlessinger, D., D'Urso, M., Tanabe, S., Oshiro, T. and Imamoto, F. (1994). Stability of YACs containing ribosomal or RCP/GCP locus DNA in wild-type *S. cerevisiae* and RAD mutant strains. *DNA Res.* **1**:191-199.

Korn, B., Sedlacek, Z., Manca, A., Kioschis, P., Konecki, D., Lehrach, H. and Poustka, A. (1992). A strategy for the selection of transcribed sequences in the Xq28 region. *Hum. Molec. Genet.* **1**:235-242.

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

Krill, A. E. (1977). Congenital stationary night blindness. *In* "Krill's Hereditary Retinal and Choroidal Diseases" (A. E. Krill, Ed.), Vol. 2, pp. 391-420, Harper and Row, Hagerstown.

Kurnit, D. M. and Seed, B. (1990). Improved genetic selection for screening bacteriophage libraries by homologous recombination in vivo. *Proc. Natl. Acad. Sci. USA* **87**:3166-3169.

Kwan, S.-P., Hagemann, T. L., Blaese, R. M. and Rosen, F. S. (1995a). A high-resolution map of genes, microsatellite markers, and new dinucleotide repeats from UBE1 to the GATA locus in the region Xp11.23. *Genomics* **29**:247-252.

Kwan, S.-P., Hagemann, T. L., Radtke, B. E., Blaese, R. M., and Rosen, F. S. (1995b). Identification of mutations in the Wiscott-Aldrich syndrome gene and characterization of a polymorphic dinucleotide repeat at DXS6940, adjacent to the disease gene. *Proc. Natl. Acad. Sci. USA* **92**:4706-4710.

Kwok, P.-Y., Deng, Q., Zakeri, H., Taylor, S., L. and Nickerson, D. A. (1996). Increasing the information content of STS-based genome maps: Identifying polymorphisms in mapped STSs. *Genomics* **31**:123-126.

Lachapelle, P., Little, J. M. and Polomeno, R. C. (1983). The photopic electroretinogram in Congenital Stationary Night Blindness with myopia. *Invest. Ophthal. Visual Science* **24**:442-450.

Lafreniere, R. G., Brown, C. J., Powers, V. E., Carrel, L., Davies, K. E., Barker, D. F. and Willard, H. F. (1991). Physical mapping of 60 DNA markers in the p21.1>q21.3 region of the human X chromosome. *Genomics* **11**:352-363.

Lennon, G., Auffray, C., Polymeropoulos, M. and Soares, M. B. (1996). The I.M.A.G.E. consortium: An integrated molecular analysis of genomes and their expression. *Genomics* **33**:151-152.

Lindsay, S. and Bird, A. P. (1987). Use of restriction enzymes to detect potential gene sequences in mammalian DNA. *Nature* **327**:336-338.

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

Lovett, M. (1994a). Direct selection of cDNAs using genomic contigs. *In* "Current Protocols in Human Genetics" (J. Seidman, Ed.), pp.6.3.1-6.3.15, Wiley Interscience, New York.

Lovett, M. (1994b). Fishing for complements: Finding genes by direct selection. *Trends Genet.* **10**:352-357.

Lui, P., Legerski, R. and Siciliano, J. (1989). Isolation of human transcribed sequences from human-rodent somatic cell hybrids. *Science* **236**:813-815.

Luty, J. A., Guo, Z., Willard, H. F., Ledbetter, D. H., Ledbetter, S. and Litt, M. (1990). Five polymorphic microsatellite VNTRs on the human X chromosome. *Am. J. Hum. Genet* **46**:776-783.

Lyon, M. F. (1962). Sex chromatin and gene action in the mammalian X chromosome. *Am. J. Hum. Genet* **14**:135-148.

Marshal, R. D., Koontz, J. and Sklar, J. (1995). Detection of mutations by cleavage of DNA heteroduplexes with bacteriophage resolvases. *Nature Genet.* **9**:177-183.

Mazzarella, R. and Srivastava, A. K. (1994). Physical linkage of expressed sequence tags (ESTs) to polymorphic markers on the X chromosome. *Hum. Molec. Genet.* **3**:1095-1101.

McKusick, V. A. (1992). "Mendelian inheritance in man Vol. 2", John Hopkins University Press, Baltimore.

Meindl, A., de Carvalho, M. R. S., Schindelhauer, D., Herrmann, K., Grimm, L., Wehnert, M., Ross, M. T. and Meitinger, T. (1995). Two novel genes mapped to cosmid contigs in Xp21.1 and Xp11.23. In: Report of the Sixth International Workshop on Human X Chromosome Mapping 1995. *Cytogenet. Cell Genet.* **71**:307-342.

Meindl, A., Dry, K., Herrmann, K., Manson, F., Ciccodicola, A., Edgar, A., Carvalho, M. R. S., Achatz, H., Hellebrand, H., Lennon, A., Migliaccio, C., Porter, K., Zrenner, E., Bird, A., Jay, M., Lorenz, B., Wittwer, B., D'Urso, M., Meitinger, T. and Wright, A. (1996). A gene (RPGR) with homology to the RCC1 guanine nucleotide exchange factor is mutated in X-linked retinitis pigmentosa (RP3). *Nature Genet.* **13**:35-42.

Merin, S., Rowe, H., Auerbach, E. and Landau, J. (1970). Syndrome of congenital high myopia with nyctalopia. *Am. J. Ophthal.* **70**:541-547.

Michaud, J., Brody, L. C., Steel, G., Fontaine, G., Martin, L. S., Valle, D., and Grant, M. (1992). Strand-separating conformational polymorphism analysis: Efficacy of detection of point mutations in the human δ ornithine aminotransferase gene. *Genomics* **13**: 389-394.

Miki, Y., Swensen, J., Donna, S.-E., Futreal, 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. (1994). A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. *Science* **266**:66-71.

Mitchell, G. A., Looney, J. E., Brody, L. C., Steel, G., Suchanek, M., Engelhard, J. F., Willard, H. F., and Valle, D. (1988). Human ornithine-d-aminotransferase. *J. Biol. Chem.* **263**:14288-14295.

Miyake, Y. and Kawase, Y. (1984). Reduced amplitude of oscillatory potentials in female carriers of x-linked recessive congenital stationary night blindness. *Am. J. Ophthalmol.* **98**:208-215.

Miyake, Y., Yagasaki, K., Horiguchi, M., Kawase, Y. and Kanda, T. (1986). Congenital Stationary Night Blindness with negative electroretinogram. *Arch. Ophthalmol.* **104**:1013-1020.

Miyake, Y., Horiguchi, M., Terasaki, H. and Kondo, M. (1994). Scotopic threshold response in complete and incomplete types of congenital stationary night blindness. *Invest. Ophthal. Visual Science* **35**:3770-3775.

Miyamoto, M. M., Koop, B. F., Slightom, J. L., Goodman, M. and Tennant, M. R. (1988). Molecular systematics in higher primates: genealogical relations and classification. *Proc. Natl. Acad. Sci. USA* **85**:7627-7631.

Morgan, J. G., Dolganov, G. M., Robbins, S. E., Hinton, L. M. and Lovett, M. (1992). The selective isolation of novel cDNAs encoded by the regions surrounding the human interleukin 4 and 5 genes. *Nucleic Acids Res.* **20**:5173-5179.

Muller, U., Graeber, M. B., Haberhausen, G. and Kohler, A. (1994). Molecular basis and diagnosis of neurogenetic disorders. *J. Neurol. Sciences* **124**:119-140.

Musarella, M. A., Weleber, R. G., Murphey, W. H., Young, R. S. L., Anson-Cartwright, L., Mets, M., Kraft, S. P., Polemeno, R., Litt, M. and Worton, R. G. (1989). Assignment of the gene for complete X-linked Congenital Stationary Night Blindness (CSNB1) to Xp11.3. *Genomics* **5**:727-737.

Myers, R. M., Larin, Z. and Maniatis, T. (1985). Detection of single base substitutions by ribonuclease cleavage at mismatches in RNA: DNA duplexes. *Science* **239**:1242-1246.

Nagamine, C. M., Chan, K. and Lau, Y.-F. C. (1989). A PCR artifact: generation of heteroduplexes. *Am. J. Hum. Genet* **45**:337-339.

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. (1994). Characterization of four human YAC libraries for clone size, chimerism and X chromosome sequence representation. *Nucleic Acids Res.* **22**:3406-3411.

Nakamura, Y., Leppert, M., O'Connell, P., Wolff, R., Holm, T., Culver, M., Martin, C., Fujimoto, E., Hoff, M., Kumlin, E. and White, R. (1987). Variable number of tandem repeat (VNTR) markers for human gene mapping. *Science* **235**:1616-1622.

Nelson, D. L., Ballabio, A., Cremers, F., Monaco, A. P. and Schlessinger, D. (1995). Report of the sixth international workshop on X chromosome mapping 1995. *Cytogenet. Cell Genet.* **71**:307-342.

Nelson, D. L., Ballabio, A., Cremers, F., Monaco, A. P. and Schlessinger, D. (1997). Report of the seventh international workshop on X chromosome mapping 1996. *Cytogenet. Cell Genet.* (in press).

Nickerson, D. A., Whitehurst, C., Boyson, C., Charmley, P., Kaiser, R. and Hood, L. (1992). Identification of clusters of biallelic polymorphic sequence-tagged sites (pSTSs) that generate highly informative and automatable markers for genetic linkage mapping. *Genomics* **12**:377-387.

Nizetic, D., Zehetner, G., Monaco, A. P., Gellen, L., Young, B. D. and Lehrach, H. (1991). Construction, arraying, and high-density screening of large insert libraries of human chromosomes X and 21: their potential as reference libraries. *Proc. Natl. Acad. Sci. USA* **88**:3233-3237.

North, M. A., Sansiau, P., Buckler, A. J., Church, D., Jackson, A., Patel, K., Trowsdale, J. and Lehrach, H. (1993). Efficiency and specificity of gene isolation by exon amplification. *Mammalian Genome* **4**:466-474.

Nyakatura, G., Platzer, M., Drescher, B., Hellebrand, H., Strom, T., Meindl, A., and Rosenthal, A. (1996). Xp11.23: Sequencing and analysis of 300 kb around the synaptophysin (SYP) gene. *In* "Report of the Seventh International Workshop on Human X Chromosome Mapping 1996" (D. L. Nelson, Ed.), (in press).

Olson, M., Hood, L., Cantor, C. and Botstein, D. (1989). A common language for physical mapping of the human genome. *Science* **245**:1434-1435.

Orita, M., Iwahana, H., Kanazawa, H., Hayashi, K., and Takao, S. (1989a). Detection of polymorphisms of human DNA by gel electrophoresis as single-strand conformation polymorphisms. *Proc. Natl. Acad. Sci. USA* **86**:2766-2770.

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

Osborne-Lawrence, S., Welcsh, P. L., Spillman, M., Chandrasekharappa, S. C., Gallardo, T. D., Lovett, M. and Bowcock, A. M. (1995). Direct selection of expressed sequences within a 1-Mb region flanking BRCA1 on human chromosome 17q21. *Genomics* **25**:248-255.

Ott, J., Bhattacharya, S., Chen, J. D., Denton, M. J., Donald, J., Dubay, C., Farrar, G. J., Fishman, G. A., Frey, D., Gal, A., Humphries, P., Jay, B., Jay, M., Litt, M., Machler, M., Musarella, M., Neugebauer, M., Nussbaum, R. L., Terwilliger, J. D., Weleber, R. G., Wirth, B., Wong, F., Worton, R. G. and Wright, A. F. (1990). Localizing multiple X chromosome-linked retinitis pigmentosa loci using multilocus homogeneity tests. *Proc. Natl. Acad. Sci. USA* **87**:701-704.

Ozcelik, T., Lafreniere, R. G., Archer, B. T., Johnston, P. A., Willard, H. F., Francke, U. and Sudhof, T. C. (1990). Synaptophysin: Structure of the human gene and assignment to the X chromosome in man and mouse. *Am. J. Hum. Genet.* **47**:551-561.

Palmieri, G., Romano, G., Ciccodicola, A., Casamassimi, A., Campanile, C., Esposito, T., Cappa, V., Lania, A., Johnson, S., Reinbold, R., Poustka, A., Schlessinger, D. and D'Urso, M. (1994). YAC contig organization and CpG island analysis in Xp28. *Genomics* **24**:149-158.

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

Parrish, J. E. and Nelson, D. L. (1993). Regional assignment of 19 X-linked ESTs. *Hum. Molec. Genet.* **2**:1901-1905.

Paulsen, K., Forrest, S., Scherer, G., Ropers, H.-H., and Davies, K. (1986). Regional localization of X chromosome short arm probes. *Hum. Genet.* **74**:155-159.

Pearce, W. G., Reedyk, M. and Coupland, S. G. (1990). Variable expressivity in X-linked Congenital Stationary Night Blindness. *Can. J. Ophthal.* **25**:3-10.

Pierce, J. C., Sauer, B. and Sternberg, N. (1992). A positive selection vector for cloning high molecular weight DNA by the bacteriophage P1 system: Improved cloning efficacy. *Proc. Natl. Acad. Sci. USA* **89**:2056-2060.

Polymeropoulos, M. H., Xiao, H., Sikela, J. M., Adams, M., Venter, J. C., and Merril, C. R. (1993). Chromosomal distribution of 320 genes from a brain cDNA library. *Nature Genet* **4**: 381-386.

Rao, V. B., Thaker, V. and Black, L. W. (1992). A phage T4 in vitro packaging system for cloning long DNA molecules. *Gene* **113**:25-33.

Roepman, R., van Duijnhoven, G., Rosenberg, T., Pinkers, A. J. L. G., Bleeker-Wagemakers, L. M., Bergen, A. A. B., Post, J., Beck, A., Reinhardt, R., Ropers, H.-H., Cremers, F. P. M. and Berger, W. (1996). Positional cloning of the gene for X-linked retinitis pigmentosa 3: Homology with the guanine-nucleotide-exchange factor RCC1. *Hum. Molec. Genet.* **5**:1035-1041.

Rogner, U. C., Kioschis, P., Wilke, K., Gong, W., Pick, E., Dietrich, A., Lehrach, H. and Poustka, A. (1994). A YAC clone map spanning 7.5 megabases of human chromosome band Xq28. *Hum. Molec. Genet.* **3**:2137-2146.

Rommens, J. M., Iannuzzi, M. C., Kerem, B.-S., Drumm, M. L., Melmer, G., Dean, M., Rozmahel, R., Cole, J. L., Kennedy, D., Hidaka, N., Zsiga, M., Buchwald, M., Riordan, J. R., Tsui, L.-C. and Collins, F. S. (1989). Identification of the cystic fibrosis gene: chromosome walking and jumping. *Science* **245**:1059-1065.

Rommens, J. M., Mar, L., McArthur, J., Tsui, L.-C. and Scherer, S. W. (1994). Towards a transcriptional map of the q21-q22 region of chromosome 7. *In* "Identification of Transcribed Sequences" (U. Hochgeschwender, and K. Gardiner, Eds.), pp. 65-79, Plenum Press, New York.

Roustan, P., Curtis, A. R. J., Kamakari, S., Thiselton, D., Lindsay, S., and Bhattacharya, S. S. (1992). Dinucleotide repeat polymorphism at the DXS573 locus. *Hum. Mol. Genet.* **2**:92.

Roy, N., Mahadevan, M. S., McLean, M., Shutler, G., Yaraghi, Z., Farahani, R., Baird, S., Besner-Johnston, A., Lefebvre, C., Kang, X., Salih, M., Aubry, H., Tamai, K., Guan, X., Ioannou, P., Crawford, T. O., de Jong, P. J., Surh, L., Ikeda, J.-E., Kroneluk, R. G. and MacKenzie, A. (1995). The gene for neuronal apoptosis inhibitory protein is partially deleted in individuals with spinal muscular atrophy. *Cell* **80**:167-178.

Ruano, G. and Kidd, K. K. (1991). Coupled amplification and sequencing of genomic DNA. *Proc. Natl. Acad. Sci. USA* **88**:2815-2819.

Ruttum, M. S., Lewandowski, M. F. and Bateman, J. B. (1992). Affected females in X-linked congential stationary night blindness. *Ophthal.* **99**:747-752.

Sanger, F., Nicklen, S. and Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74**:5463-5467.

Sarkar, G., Yoon, H.-S. and Sommer, S. S. (1992). Dideoxy fingerprinting (ddF): A rapid and efficient screen for the presence of mutations. *Genomics* **13**:441-443.

Schindelhauer, D., Hellebrand, H., Grimm, L., Bader, I., Meitinger, T., Wehnert, M., Ross, M. and Meindl, A. (1996). Long range map of 3.5 Mb region in Xp11.23-22 with a sequence ready map from a 1.1 Mb gene-rich interval. *Genome Res.* **6**:1056-1069.

Schmitt, H., Kim, U.-J., Slepak, T., Blin, N., Simon, M. I. and Shizuya, H. (1996). Framework for a physical map of the human 22q13 region using bacterial artificial chromosomes (BACs). *Genomics* **33**:9-20.

Schuler, G. D., Boguski, M. S., Stewart, E. A., Stein, L. D., Gyapay, G., Rice, K., White, R. E., Rodriquez-Tome, P., Aggarwal, A., Bajorek, E., Bentolila, S., Birren, B. B., Butler, A., Castle, A. B., Chiannilkulchai, N., Chu, A., Clee, C., Cowles, S., Day, P. J. R., Kibling, T., Drouot, N., Dunham, I., Duprat, S., East, C., Edwards, C., Fan, F.-B., Fang, N., Fizames, C., Garrett, C., Green, L., Hadley, D., Harris, M., Harrison, P., Brady, S., Hicks, A., Holloway, E., Hui, L., Hussain, S., Louis-Dit-Sully, C., Ma, J., MacGilvery, A., Mader, C., Maratukulam, A., Matise, T. C., McKusick, K. B., Morissette, J., Mungall, A., Muselet, D., Nusbaum, H. C., Page, D. C., Peck, A., Perkins, S., Piercy, M., Qin, F., Quackenbush, J., Ranby, S., Reif, T., Rozen, S., Sanders, C., She, X., Silva, J., Slonim, D. K., Soderlund, C., Sun, W.-L., Tabar, P., Thangarajah, T., Vega-Czarny, N., Vollrath, D., Voyticky, S., Wilmer, T., Wu, X., Adams, M. D., Auffray, C., Walter, N. A. R., Brandon, R., Dehejia, A., Goodfellow, P. N., Houlgatte, R., Hudson Jr., J. R., Ide, S. E., Iorio, D. R., Lee, W. Y., Seki, N., Nagase, T., Ishikawa, K., Nomura, N., Phillips, C., Polymeropoulos, M. H., Sandusky, M., Schmitt, K., Berry, R., Swanson, K., Torres, R., Venter, J. C., Sikela, J. M., Beckmann, j. S., Weissenbach, J., Myers, R. M., Cox, D. R., James, M. R., Bentley, D., Deloukas, P., Lander, E. S., and Hudson, T. J. (1996). A gene map of the human genome. *Science* **274**:540-546.

Schwartz, M. and Rosenberg, T. (1991). Aland Eye Disease: Linkage data. *Genomics* **10**:327-332.

Sheffield, V. C., Cox, D. R., Lerman, L. S. and Myers, R. M. (1989). Attachment of a 40-base-pair G+C-rich sequence (GC-clamp) to genomic DNA fragments by the polymerase chain reaction results in improved detection of single-base changes. *Proc. Natl. Acad. Sci. USA* **86**:232-236.

Shizuya, H., Birren, B., Kim, U.-J., Mancino, V., Slepak, T., Tachiiri, Y. and Simon, M. (1992). Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in Escherichia-coli using an F-factor-based vector. *Proc. Natl. Acad. Sci. USA* **89**:8794-8797.

Sidhar, S. K., Clark, J., Gill, S., Hamoudi, R., Crew, A. J., Gwilliam, R., Ross, M., Linehan, W. M., Birdsall, S., Shipley, J. and Cooper, C. S. (1996). The t(X:1)(p11.2;q21.2) translocation in papillary renal cell carcinoma fuses a novel gene PRCC to the TFE3 transcription factor gene. *Hum. Molec. Genet.* **5**:1333-1338.

Simmons, A. D., Goodart, S. A., Gallardo, T. D., Overhauser, J. and Lovett, M. (1995). Five novel genes from the cri-du-chat critical region isolated by direct selection. *Hum. Molec. Genet.* **4**:295-302.

Sternberg, N. (1990). Bacteriophage P1 cloning system for the isolation, amplification, and recovery of DNA fragments as large as 100 kilobase pairs. *Proc. Natl. Acad. Sci. USA* **87**:103-107.

Stoflet, E. S., Koeberl, D. D., Sarkar, G. and Sommer, S. S. (1988). Genomic amplification with transcript sequencing. *Science* **239**:491-494.

Thiselton, D. L., Kamakari, S., Curtis, A. R. J., Coleman, M. P., Davies, K. E., Lindsay, S., and Bhattacharya, S. S. (1993). Dinucleotide repeat polymorphism at the DXS556 locus. *Hum. Molec. Genet.* **2**:613.

Thiselton, D. L., Lindsay, S., Kamakari, S., Hardcastle, A. J., Roustan, R. and Bhattacharya, S. S. (1995). Genetic and physical mapping of five novel microsatellite markers on human Xp21.1-p11.22. *Genomics* **25**:279-281.

Thiselton, D. L., Hampson, R. M., Nayudu, M., Van Maldergem, L., Wolf, M. L., Saha, B. K., Bhattacharya, S. S., and Hardcastle, A. J. (1996). Mapping the RP2 locus for X-linked Retinitis Pigmentosa on proximal Xp: A genetically defined 5-cM critical region and exclusion of candidate genes by physical mapping. *Genome Res.* **6**:1093-1102.

Trofatter, J. A., MacCollin, M. M., Rutter, J. L., Murrell, J. R., Duyao, M. P., Parry, D. M., Eldridge, R., Kley, N., Menon, A. G., Pulaski, K., Haase, V. H., Ambrose, C. M., Munroe, D., Bove, C., Haines, J. L., Martuza, R. L., MacDonald, M. E., Seizinger, B. R., Short, M. P., Buckler, A. J. and Gusella, J. F. (1993). A novel moesin-, ezrin-, radixin-like gene is a candidate for the neurofibromatosis 2 tumour suppressor. *Cell* **72**:791-800.

Trump, D., Pilia, G., Dixon, P. H., Wooding, C., Thakrar, R., Leigh, S. E. A., Nagaraja, R., Whyte, M. P., Schlessinger, D. and Thakker, R. V. (1995). Construction of a YAC contig and an STS map spanning 3.6 megabase pairs in Xp22.1. *Cytogenet. Cell Genet.*

Uberbacher, E. C. and Mural, R. J. (1991). Locating protein coding regions in human DNA sequences by a multiple-sensor-neural network approach. *Proc. Natl. Acad. Sci. USA* **88**:11261-11265.

Villa, A., Notarangelo, L., Macchi, P., Mantuano, E., Cavagni, G., Brugnoni, D., Strina, D., Patrosso, M. C., Ramenghi, U., Sacco, M. G., Ugazio, A. and Vezzoni, P. (1995). X-linked thrombocytopenia and Wiscott-Aldrich syndrome are allelic diseases with mutations in the WASP gene. *Nature Genet.* **9**:414-417.

Vulpe, C., Levinson, B., Whitney, S., Packman, S. and Gitschier, J. (1993). Isolation of a candidate gene for Menkes disease and evidence that it encodes a copper-transporting ATPase. *Nature Genet.* **3**:7-13.

Waardenberg, P. J. (1970). Some notes on publications of Professor Arnold Sorsby and on Aland eye disease (Forsius-Erikson syndrome). *J. Med. Genet.* **7**:194-199.

Wagner, R., Debbie, P. and Radman, M. (1995). Mutation detection using immobilized mismatch binding protein (MutS). *Nucleic Acids Res.* **23**:3944-3948.

Walker, C. L., Cargile, C. B., Floy, K. M., Delannoy, M. and Migeon, B. R. (1991). The Barr body is a looped X chromosome formed by telomere association. *Proc. Natl. Acad. Sci. USA* **88**:6191-6195.

Walter, M. A., and Goodfellow, P. N. (1993). Radiation hybrids: Irradiation and fusion gene transfer. *Trends Genet.* **9**:352-356.

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

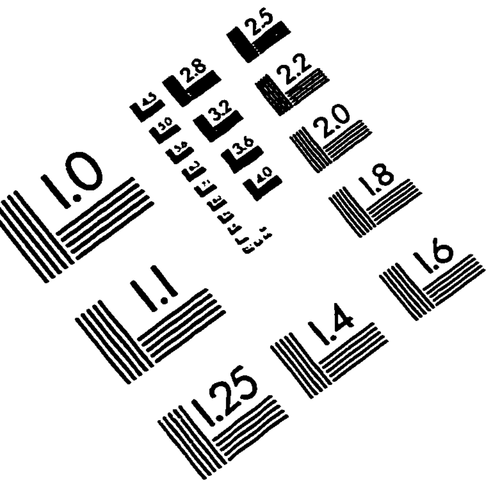
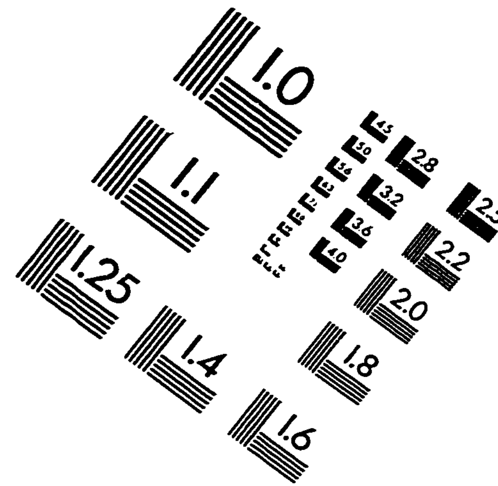
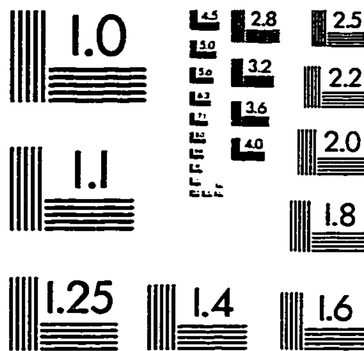
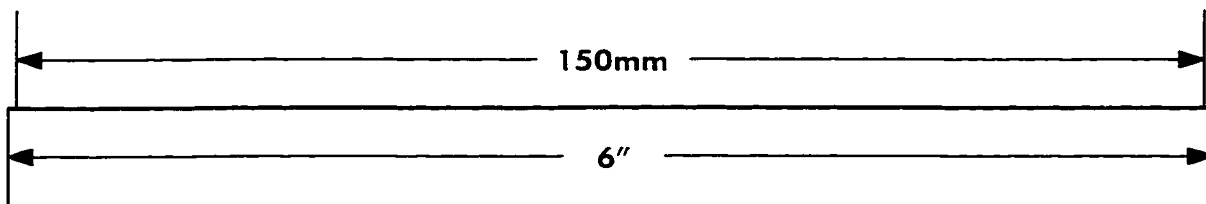
Walter, M. A., and Goodfellow, P. N. (1995). Irradiation and fusion gene transfer. *Molec. Biotech.* **3**:117-128.

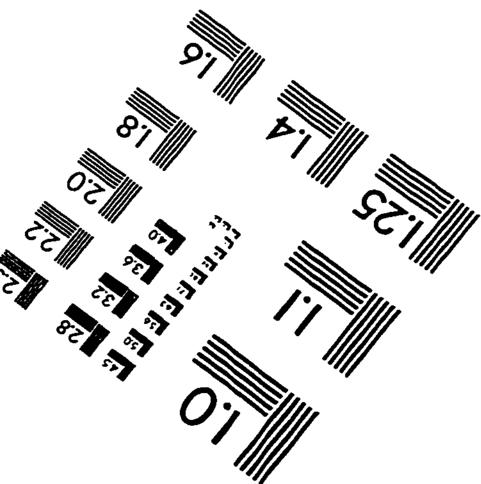
Weber, J. L. and May, P. E. (1989). Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. *Am. J. Hum. Genet* **44**:388-396.

Weber, J. L. (1990). Human DNA polymorphisms based on length variations in simple-sequence tandem repeats. *In* "Genetic and Physical Mapping" (K. E. Davies, S. M. Tilgham, Eds.), Vol. 1, pp. 159-181, Cold Spring Harbor Laboratory, New York.

Worley, K.C., Wiese, B. A., and Smith, R. F. (1995). BEAUTY: An enhanced BLAST-based search tool that integrates multiple biological information resources into sequence similarity search results. *Genome Res.* **5**:173-184.

IMAGE EVALUATION TEST TARGET (QA-3)



APPLIED IMAGE . Inc
1653 East Main Street
Rochester, NY 14609 USA
Phone: 716/482-0300
Fax: 716/288-5989

© 1993, Applied Image, Inc., All Rights Reserved