# Positional Cloning of an X;8 Translocation (p22.13;q22.1) associated with Multiple Exostoses and Autism

Dissertation submitted to the University of Oxford
for the degree of Doctor of Philosophy

by

Yumiko Ishikawa-Brush

Faculty of Biological Sciences
Department of Biochemistry
University of Oxford
The Wellcome Trust Centre
Oxford

and

Trinity College
Oxford

SCIENCE LIBRARY OXFORD
DEPOSITED
19 JAN 1998
THESIS

Trinity Term 1997

Dedicated in admiration to:

**Yoshio and Katsuko Ishikawa**
**and**
**Dr. Gerome Brush II**

You are the energy of my life.

# PREFACE

This dissertation is the result of my own work and submitted to the University of Oxford only. Any other work conducted by a collaborator is clearly stated.

Yumiko Ishikawa-Brush
Trinity Term 1997
Oxford

# ABSTRACT

Autism is characterized by qualitative impairments in communication and reciprocal social interaction. About 3 in 10,000 in the general population suffer from this neurodevelopmental disorder. The majority of patients also manifest mental retardation and about 20-45% epilepsy.

Multiple exostoses is the commonest form of all skeletal dysplasias, affecting 1 in 50,000 live births. The condition, characterized by cartilaginous protuberances at the ends of the diaphyses, affects the extremities causing skeletal deformities and short stature.

Autism and multiple exostoses are considered to be inherited disorders, but the underlying biochemical defects of the disorders are unknown. Both of these conditions allow for survival but considerably diminish the quality of life.

An X;8 translocation was identified in a female patient, ML, with autism and multiple exostoses. Her phenotypic manifestations are likely due to the chromosomal abnormality.

A positional candidate cloning strategy was used to investigate the genes involved in the translocation. The translocation breakpoint was first isolated in Yeast Artificial Chromosomes (YACs), then in cosmid and plasmid clones. The translocation was reciprocal within a 5'-GGCA-3' sequence found on both X and 8 chromosomes without gain or loss of a single nucleotide. The translocation breakpoint on the X chromosome occurred in the first intron of the gastrin releasing peptide receptor (GRPR) gene and on

chromosome 8 approximately 30 kb distal to the 3' end of the Syndecan-2 gene (SDC2), also known as human heparan sulfate proteoglycan or fibroglycan.

Although the GRPR gene was shown to escape X-inactivation and the coding region of the SDC2 gene was not disrupted, a dosage effect of the GRPR gene and a position effect of the SDC2 gene may, however, have contributed to the phenotype observed in this patient. The orientation of these genes with respect to the translocation was incompatible with the formation of a fusion gene.

The GRPR and SDC2 genes may provide insight into the biochemical nature of autism and multiple exostoses. Investigation of mutations in these two genes in unrelated patients with either autism or multiple exostoses as well as linkage and association studies are needed to validate them as candidate genes.

# ACKNOWLEDGEMENTS

There is a way where there is a will and a chance. Prof. Ian W Craig was the first to give me the chance to study genetics and to whom I am most indebted.

A great many thanks also to Dr. John F Powell whose patience, guidance and enthusiasm I trusted completely. This exciting project was given by him and his collaborator, Dr. Patrick Bolton. Without their support and encouragement this thesis would not have been possible.

I am deeply grateful to my supervisor, Prof. Anthony P Monaco, who taught me the joy of science and scientific skills. His love of science has inspired me and his energy and commitment to science has given me an immense incentive. It is beyond description in either Japanese or English to express my gratitude for his invaluable advice and support during the preparation of this thesis.

I would also like to thank my husband, Dr. Gerome Brush II, for his truly steady support over a long and sometimes uncertain course. His logical thinking has stimulated the way of my own thoughts.

I am very grateful to Drs. Veronica J Buckle, Katrina A Rack, Elaine R Levy, Caroline Stark, Alain Hovnanian, Christophe Philippe, Jose Mejia, Elena Maestrini, Iona Millwood, Andrea Nemeth, and Eimear Dunne for useful discussions, and in particular to Dr. Ann P Walker for introducing me to the ways of the laboratory. Thanks also to Katrina Hobbs and Alain Goldsmith for synthesizing oligonucleotides, Emma Jones for ordering supplies, and Iris Randall for help with references.

All good things have come to me through the warm support and encouragement of these good friends, Drs. Caroline Garey, Hiro Kato, June Davies, Ying Lui, Peng Yeong Woon, and Bai-Ping Zhang.

# CONTENTS


**PRELUDE**

| | | |
|---|---|---|
| Title | | i |
| Dedication | | ii |
| Preface | | iii |
| Abstract | | iv-v |
| Acknowledgements | | vi |
| Contents | | vii-x |
| List of Figures | | xi-xiii |
| List of Tables | | xiv |
| Abbreviations | | xv-xviii |
| **CHAPTER I--INTRODUCTION** | | **1** |
| 1.1 | PATIENT, ML | 3 |
| 1.1.1 | Developmental History of Patient, ML | 3 |
| 1.1.2 | Physical Examination | 5 |
| 1.1.3 | Language and Social Behaviour | 7 |
| 1.1.4 | Family | 7 |
| 1.2 | AUTISM | **8** |
| 1.2.1 | Diagnosis | 10 |
| 1.2.2 | Epidemiology | 11 |
| 1.2.3 | Neuroanatomy | 15 |
| 1.2.4 | Neurophysiology | 17 |
| 1.2.5 | Neurochemistry | 18 |
| 1.2.6 | Genetic Factors | 20 |
| 1.2.7 | Genetic Heterogeneity | 22 |
| 1.2.8 | Chromosomal Abnormalities | 25 |
| 1.3 | MULTIPLE EXOSTOSES | **26** |
| 1.3.1 | Diagnosis | 26 |
| 1.3.2 | Epidemiology | 26 |
| 1.3.3 | Genetic Disorder | 28 |
| 1.3.4 | Genetic Heterogeneity | 28 |
| 1.3.5 | Malignant Transformation | 30 |
| 1.3.6 | Association with other disorders | 31 |
| 1.4 | CLONING STRATEGIES | **31** |
| 1.4.1 | Functional Cloning | 34 |
| 1.4.2 | Positional Cloning | 36 |
| 1.4.2.1a | Chromosomal Abnormalities | 37 |
| 1.4.2.1b | Genetic Linkage Analysis | 40 |
| 1.4.2.1c | Affected Sib-Pair Analysis | 43 |
| 1.4.2.1d | Allelic Association Study & Linkage Disequilibrium | 44 |
| 1.4.2.2 | Physical Mapping | 46 |
| 1.4.2.3 | Transcript Mapping | 49 |
| 1.4.3 | Positional Candidate Cloning | 54 |
| 1.4.4 | Computer Analysis of DNA Sequence | 55 |
| 1.5 | AIM OF THESIS | **57** |

| | | |
|---|---|---|
| **CHAPTER II--MATERIALS AND METHODS** | | **59** |
| 2. 1 | PREPARATION OF DNA FROM RECOMBINANT CLONES | 60 |
| 2. 1. 1 | Purification of plasmid and cosmid DNA | 60 |
| 2. 1. 2 | Purification of Bacteriophage DNA | 61 |
| 2. 1. 3 | Preparation of YAC DNA | 63 |
| 2. 2 | PREPARATION OF DNA BY PCR AMPLIFICATION | **65** |
| 2. 2. 1 | PCR Templates | 66 |
| 2. 2. 2 | Designing for Primers for polymerase chain reactions (PCR) | 66 |
| 2. 2. 3 | Primer Preparation | 67 |
| 2. 2. 4 | Amplification Condition for Standard PCR | 67 |
| 2. 2. 5 | Preparation of Long DNA Fragments using PCR | 71 |
| 2. 2. 6 | Interspersed repeat element PCR (Alu-PCR Amplification) | 71 |
| 2. 2. 7 | Preparation of Di-and Tri-nucleotide Oligo-probes | 72 |
| 2. 2. 8 | RT PCR and 5′ Rapid Amplification of cDNA Ends (5′ RACE) | 72 |
| 2. 3 | ISOLATION OF INSERTS FROM CLONED DNA | **74** |
| 2. 3. 1 | Cloned DNA Digestion by Restriction Endonuclease | 75 |
| 2. 3. 2 | DNA Fragment Size Separation using Gel Electrophoresis | 75 |
| 2. 3. 3 | YAC DNA Isolation using Pulsed Field Gel Electrophoresis | 78 |
| 2. 3. 4 | Visualization of DNA with Ethidium bromide staining | 79 |
| 2. 3. 5 | Purification of Insert DNA and PCR Products | 79 |
| 2. 4 | MEASUREMENT OF DNA AND CELL CONCENTRATION | **79** |
| 2. 4. 1 | DyNA Quant 200 Fluorometer (Hoefer) | 79 |
| 2. 4. 2 | Spectrophotometer with Optical Density (OD) | 80 |
| 2. 4. 3 | Cell Counts with Haemocytometer | 81 |
| 2. 5 | NUCLEIC ACID TRANSFER | 81 |
| 2. 5. 1 | Southern Blots | 81 |
| 2. 5. 2 | Patient Genomic DNA Blots | 82 |
| 2. 5. 3 | Patient Northern Blots | 82 |
| 2. 6 | CLONING, LIBRARY CONSTRUCTION AND FILTER LIFTS | **83** |
| 2. 6. 1 | Existing Library Filters | 83 |
| 2. 6. 2 | Constructed Cosmid Library Filters | 84 |
| 2. 6. 3 | Phage Library Filters | 87 |
| 2. 6. 4 | Plasmid Subclone Filters | 90 |
| 2. 6. 5 | Cosmid End-Cloning | 91 |
| 2. 7 | PROBE CONSTRUCTION WITH ISOTOPES | **92** |
| 2. 7. 1 | Double-Stranded DNA | 92 |
| 2. 7. 2 | High Molecular Weight dsDNA | 93 |
| 2. 7. 3 | Single-Stranded DNA | 93 |
| 2. 8 | PROBE CONSTRUCTION WITH NON-ISOTOPES | **94** |
| 2. 8. 1 | Double-Stranded DNA | 94 |
| 2. 8. 2 | High Molecular Weight dsDNA | 94 |
| 2. 8. 3 | Single-Stranded DNA | 94 |
| 2. 8. 4 | Non-Isotopic Labeling via PCR | 94 |
| 2. 9 | HYBRIDIZATION, POST-WASH & DETECTION | **95** |
| 2. 9. 1 | Isotopic Probes | 95 |
| 2. 9. 2 | Non-Isotopic Probes | 97 |
| 2. 10. | EXON-AMPLIFICATION | **98** |
| 2. 10. 1 | Transformation | 99 |

| | | |
|---|---|---|
| 2. 10. 2 | Transfection | 99 |
| 2. 10. 3 | RNA Extraction and cDNA PCR | 99 |
| 2. 10. 4 | Isolation of Exons | 99 |
| 2. 11 | SEQUENCING | **100** |
| 2. 11. 1 | Automated DNA Cycle Sequencing | 100 |
| 2. 11. 2 | Sequence Analysis | 101 |
| 2. 12 | PCR LIBRARY SCREENING | **102** |
| 2. 12. 1 | PCR YAC Library | 102 |
| 2. 12. 2 | PCR cDNA Library | 103 |
| 2. 13 | POLYMORPHISM ANALYSIS | **103** |
| 2. 13. 1 | Conformation Single-Stranded Gel Electrophoresis | 103 |
| 2. 13. 2 | Sequence Confirmation | 104 |
| 2. 14 | PANELS FOR X-INACTIVATION STUDY | **104** |
| 2. 15 | CYTOGENETIC ANALYSIS | **105** |
| 2. 15. 1 | X-Inactivation | 105 |
| 2. 15. 2 | Karyotype | 106 |
| 2. 16 | FLUORESCENCE *IN SITU* HYBRIDIZATION | **106** |
| 2. 16. 1 | Probe Preparation by Nick Translation | 108 |
| 2. 16. 2 | Slide Preparation | 109 |
| 2. 16. 3 | Hybridization and Post-Hybridization Wash | 110 |
| 2. 16. 4 | Detection | 111 |
| 2. 17 | CHEMICAL AND MATERIAL SUPPLIERS | **112** |
| **CHAPTER III IDENTIFICATION OF THE TRANSLOCATION BREAKPOINT** | | **114** |
| 3. 1 | CYTOGENETIC ANALYSIS | 114 |
| 3. 1. 1 | X-Inactivation Karyotype | 114 |
| 3. 1. 2 | Karyotype | 116 |
| 3. 2 | FLANKING MARKERS AND YAC ISOLATION | 116 |
| 3. 2. 1 | FISH Probes | 116 |
| 3. 2. 2 | FISH Analysis of YACs | 119 |
| 3. 3 | COSMID ISOLATION | 122 |
| 3. 3. 1 | Whole YAC Screening of the ICRF FSCX Cosmid Library | 122 |
| 3. 3. 2 | Search for Polymorphic Di- and Tri-repeats | 132 |
| 3. 3. 3 | Construction of the Cosmid Libraries | 134 |
| 3. 3. 4 | Construction of the Cosmid Contig | 136 |
| 3. 4 | SUMMARY | 137 |
| **CHAPTER IV-TRANSLOCATION BREAKPOINT WITHIN THE GRPR GENE** | | **140** |
| 4. 1 | IDENTIFICATION OF THE TRANSLOCATION BREAKPOINT | 140 |
| 4. 1. 1 | Isolation of GRPR cDNA | 140 |
| 4. 1. 2 | Physical mapping in the GRPR gene region | 141 |
| 4. 2 | The GRPR Gene | 146 |
| 4. 2. 1 | Polymorphism in Exon-2 of the GRPR Gene | 146 |

4.2.2 The GRPR gene and Escape from X-Inactivation 149
4.2.3 Tissue Specific Expression of GRPR 151

4.3 JUNCTION FRAGMENT 151
4.3.1 Cloning the Junction Fragment 154
4.3.2 Construction of a Restriction Map 157
4.3.3 Sequencing the Junction Fragments 161

4.4 SUMMARY 166

**CHAPTER V--CHROMOSOME 8 AND THE SYNDECAN-2 (SDC2) GENE 167**

5.1 EXAMINATION OF EXISTING CHROMOSOME 8 COSMIDS 168

5.2 PHYSICAL MAPPING OF CHROMOSOME 8q22.1 170
5.2.1 Isolation of the Chromosome-8 Specific STSs 170
5.2.2 Isolation of YACs 174
5.2.3 Construction of Cosmid Libraries 174

5.3 IDENTIFICATION OF THE SYNDECAN-2 (SDC2) GENE 174
5.3.1 Testis cDNA Screening 178
5.3.2 Construction of a Cosmid Contig 178
5.3.3 The SDC2 gene 180

5.4 SEARCH FOR GENES OTHER THAN SDC2 185
5.4.1 Homology Studies of Chromosome 8 STS fragments 185
5.4.2 Exon Trapping in the 8q22 region 189
5.4.3 Homology Studies of Chromosome 8 *Hind*III fragments 190
5.4.4 Expression of the SDC2 gene 190

5.5 BALANCED RECIPROCAL TRANSLOCATION 192

5.6 SUMMARY 192

**CHAPTER VI--DISCUSSION 196**

6.1 SUMMARY OF THE RESULTS 196

6.2 CHROMOSOMAL TRANSLOCATION 199

6.3 GASTRIN RELEASING PEPTIDE RECEPTOR (GRPR) GENE 200

6.4 GRPR AND AUTISM 210

6.5 SYNDECAN-2 (SDC2) GENE 219

6.6 SDC2 AND MULTIPLE EXOSTOSES 230

6.7 FUTURE WORK 242

**REFERENCES 248**

**PUBLICATIONS 268**

# LIST OF FIGURES


| | | |
|---|---|---|
| 1. 1A | Patient ML (front) at age 26 | 4 |
| 1. 1B | Patient ML (portrait) at age 26 | 4 |
| 1. 2 | The hands of patient ML showing the evidence of multiple exostoses | 6 |
| 1. 3 | Concepts behind phenotypic indices of autism from Bolton | 9 |
| 1. 4 | Diagram showing the bone before and after the growth period | 27 |
| 2. 1 | Cubic polynomial regression of log standard length | 77 |
| 2. 2 | Signal amplification of probe-hybrids | 107 |
| 3. 1 | ML's karyotype is 46,X,t(X;8)(p22.13;q22.1) | 115 |
| 3. 2 | Schematic representation of ML's reciprocal X;8 translocation | 117 |
| 3. 3 | Location of probes used in screening the ICRF YAC library and the number of YACs isolated by each probe | 118 |
| 3. 4 | FISH result of the YAC-57 spanning the translocation breakpoint | 121 |
| 3. 5 | Positive cosmids from the ICRF flow-sorted X chromosome cosmid library | 123 |
| 3. 6 | Contig of the Lawrist cosmids with two gaps | 125 |
| 3. 7 | FISH positive of Cos-17 | 127 |
| 3. 8 | Cos-17 hybridization to normal female and patient (ML) genomic DNA | 128 |
| 3. 9 | Identification of Lawrist cosmids overlapping Cos-17 | 129 |
| 3. 10 | Isolation of 18 unique fragments from Cos-17-*Hind*III subclones | 131 |
| 3. 11 | Hybridization of a Cos-17 unique sequence subclone to patient genomic DNA | 132 |
| 3. 12 | Hybridization of (CA)n probes onto Cos-17 *Hind*III subclones | 133 |
| 3. 13 | Hybridization of a GRPR exon-1 probe onto a cosmid replica filter | 135 |
| 3. 14 | Location of GRPR primers | 136 |
| 3. 15 | The Alu- PCR products: Alu244, Alu245, Alu246 and Alu247 | 138 |
| 3. 16 | Cosmid contig covering the gastrin releasing peptide receptor (GRPR) gene | 139 |
| 4. 1 | G2 cosmid identified the junction fragment in the patient | 142 |
| 4. 2. | Primers for GRPR mutation and splice site analysis | 143 |

| | | |
|---|---|---|
| 4. 3 | Detection of the translocation junction fragment in patient genomic DNA digested with *Pst*I, *Eco*RV, *Bgl*II, and *Eco*RI | 145 |
| 4. 4 | Relationships between the translocation breakpoint and the SuperCos X chromosome cosmid subclones, 7a and 7b | 147 |
| 4. 5 | Position of the translocation in the SuperCos cosmid contig and the translocation breakpoint occurred within the first intron of GRPR. | 148 |
| 4. 6 | Sequence analyses of GRPR indicating a base substitution mutation in exon-2 | 150 |
| 4. 7A | Primers used for the X-inactivation analysis | 151 |
| 4. 7B | Escape from X inactivation of GRPR | 152 |
| 4. 8 | Northern blot hybridization analysis of GRPR expression | 153 |
| 4. 9 | Schematic representation, summarizing 3 separate hybridizations of (1) 7a and 7b, (2) 7a only, and (3) 7b only, to a patient phage library | 154 |
| 4. 10A | EtBr stained gel photo containing the patient genomic DNA digested with *Bgl*II (B) or *Bgl*II/*Sal*I (B/S). | 155 |
| 4. 10B | Hybridization of SuperCos cosmid subclones 7a and 7b to patient genomic DNA cloned in phage and digested with *Bgl*II (B) or *Bgl*II/*Sal*I (B/S). | 156 |
| 4. 11A | Hybridization of 7a to multiple digests of patient genomic DNA | 158 |
| 4. 11B | Hybridization of 7b to multiple digests of patient genomic DNA | 159 |
| 4. 12 | Relationships of subclones 7a and 7b derived from SuperCos G2, and restriction maps of the junction fragment subclones of der(8) (9-2) and der(X) (14-16), and the normal X chromosome (7-6) | 160 |
| 4. 13A | Confirmation of the translocation breakpoint by PCR amplification | 162 |
| 4. 13B | Summary of PCR amplification | 162 |
| 4. 14 | Long PCR amplifications for orienting the junction fragments with respect to the GRPR gene | 163 |
| 4. 15 | Relationships of subclones 7a and 7b derived from SuperCos G2, and restriction maps of the junction fragment subclones | 164 |
| 4. 16 | Identification and orientation of the two derivative chromosomes with respect to the *Bgl*II junction fragments | 165 |
| 5. 1. | Relationship between the chromosome 8 markers and the translocation breakpoint of the Japanese patient (Yoshiura et al. 1994). | 169 |
| 5. 2 | Schematic diagram of the chromosome 8 specific primers 33 and 67 | 171 |
| 5. 3 | Confirmation that primers 33 and 67 are specific to chromosome 8 | 172 |
| 5. 4 | Detection of the translocation junction fragment in patient genomic DNA digested with *Eco*RI and *Pst*I and hybridized with probe STS3367 | 173 |

| | | |
|---|---|---|
| 5. 5A | Isolation of chromosome 8 YACs 2B, 13B and 38B by PCR screening of the ICI YAC library using primers 33 and 67 | 175 |
| 5. 5B | Size determination of chromosome 8 YACs, 2B, 13B, and 38B. | 176 |
| 5. 5C | Confirmation that YAC 2B, 13B and 38B are specific to chromosome 8 | 176 |
| 5. 6 | FISH confirmation of non-chimeric YAC, 2B on the chromosomal metaphase spread from a normal female | 177 |
| 5. 7 | Position of cDNA clones with respect to the SDC2 gene | 179 |
| 5. 8 | Examples of isolation of the cosmid end-clones | 179 |
| 5. 9 | Cosmid contig generated by the 'clone-to-well' computer program | 181 |
| 5. 10 | Relative positions of the four identified SDC2 introns | 184 |
| 5. 11 | Location of the SDC2 gene and ML's translocation breakpoint | 186 |
| 5. 12 | Orientation of SDC2 with respect to the translocation breakpoint | 187 |
| 5. 13 | Incompatibility for a fusion gene between the GRPR and SDC2 genes | 188 |
| 5. 14 | Consensus splice "AG-GT" sites | 191 |
| 5. 15 | Northern blot hybridization analysis of SDC2 expression | 193 |
| 5. 16 | The 5'-GGCA-3' sequence where the translocation occurred | 194 |
| 6. 1 | Intracellular signalling mechanism of G-protein | 202 |
| 6. 2A&B | Distribution of bombesin in rat tissue and brain | 205 |
| 6. 3 | Diagram of the limbic system | 207 |
| 6. 4 | Amygdala's role in learning about and remembering fear | 209 |
| 6. 5 | Distribution of the human and mouse genes and inactivation status on the human X chromosome. | 215 |
| 6. 6 | Syndecan, CD44, and Glypican proteoglycans | 220 |
| 6. 7 | Core protein of proteoglycans | 221 |
| 6. 8 | Core proteins of the four known human Syndecans | 225 |
| 6. 9 | Possible functions of the syndecans at the cell surface | 227 |
| 6. 10A | Human haemoglobin developmental switches | 234 |
| 6. 10B | Gene expression in the beta-globin gene cluster controlled by locus control regions | 234 |
| 6. 11 | Genes in the 8q22-24 region and murine chromosomal homologues | 239 |

# LIST OF TABLES


| | | |
|---|---|---|
| 1. 1 | Major nucleic acid and protein sequence databases | 55 |
| 1. 2 | The FASTA and BLAST programs for sequence comparisons | 56 |
| 1. 3 | Major databases of relevance to the Human Genome Project | 57 |
| 2. 1A | PCR primers and conditions of amplification | 68-69 |
| 2. 1B | PCR primers and conditions of amplification | 70 |
| 2. 2 | Standard lambda *Hind*III fragment length and distance travelled on a gel | 76 |
| 2. 3 | The equation estimates log of fragment size | 78 |
| 2. 4 | Hybridization conditions for three kinds of probes | 109 |
| 3. 1. | Results of the FISH analyses of the X-chromosome ICRF YACs | 120 |
| 3. 2 | FISH results of 21 Lawrist cosmids | 126 |
| 3. 3 | The order of four Lawrist cosmids: Cos-17, Cos-22, Cos-31, and Cos-35 | 129 |
| 3. 4 | Subclones of Cos-17 containing unique sequences from human genomic DNA | 130 |
| 3. 5 | Cos-17 CA repeat subclone alleles | 134 |
| 5. 1 | Translocation junction fragment sizes detected by hybridizing probe STS3367 | 172 |
| 5. 2 | Sequencing analysis of the *Hind*III subclones | 182 & 183 |
| 5. 3 | Homology searches around the chromosome 8 translocation breakpoint | 189 |
| 5. 4 | Homology searches of fragments generated from Exon Trapping | 190 |
| 6. 1 | Bombesin superfamily and GRP | 201 |
| 6. 2 | Human bombesin-like receptors and ligand | 203 |
| 6. 3 | Neural substrates for early symptoms of autism | 211 |
| 6. 4 | Legend to Fig. 6. 5 | 216 |
| 6. 5 | Dosage effects of hypoxanthine adenine phosphoribosyl transferase on phenotype | 217 |
| 6. 6 | Comparison of some human proteoglycans | 222 |
| 6. 7 | Myc family of cellular oncogenes | 230 |
| 6. 8 | Two cytological states, euchromatin and heterochromatin of the eukaryotic genome. | 231 |
| 6. 9 | Genes in the 8q22-24 region | 240 |

# ABBREVIATIONS AND ACRONYMS USED

| | |
|---|---|
| 2-ME | 2-mercaptoethanol |
| $^{32}P$ | phosphorus-32 |
| aa | amino acid |
| AA | Atypical Autism |
| Ab | antibody |
| ACH | achondroplasia |
| ADI | Autism Diagnostic Interview |
| ADOS | Autism Diagnostic Observation Schedule |
| Amp | ampicillin |
| ANT3 | adenine nucleotide translocator 3 |
| AR | androgen receptor |
| ARS | yeast origin of replication |
| AS | Asperger's syndrome |
| ASMT | encodes acetylserotoninN-methyl transferase |
| ATP | adenosine 5'-triphosphate |
| BAC | bacterial artificial chromosomes |
| BLAST | Basic Local Alignment Search Tool |
| bp | base pair |
| BSA | bovine serum albumin |
| BrdU | bromodeoxyuridine |
| CA 1, 2, and 3 | carbonic isoenzymes (MIM 114800) |
| CBFAT1 | core-binding factor alpha subunit 2 (MIM 133435) |
| CCD | Cleidocranial dysplasia (MIM 119600) |
| CDD | Childhood Disintegrative Disorder |
| cDNA | complementary DNA |
| CEN | yeast centromere |
| cen | centromere |
| CF | cystic fibrosis (MIM 219700) |
| $CHCL_3$ | chloroform |
| CHEF | contour-clamped homogeneous electric field gel |
| CIAP | calf intestinal alkaline phosphatase |
| cM | centi-Morgan |
| COH1 | Cohen Syndrome (MIM 216550) |
| c.p.m. | counts per minute |
| CSGE | conformation-sensitive gel electrophoresis |
| CSF2RA | colony-stimulating factor receptor |
| CT | computerized tomographic scanning |
| DAPI | 4',6-diamidine-2-phenylindole dihydrochloride |
| dATP | deoxyadenosine triphosphate |
| dbEST | database of expressed sequence tags |
| dCTP | deoxycytidine triphosphate |
| ddATP | di-deoxyadenosine triphosphate |
| ddCTP | di-deoxycytidine triphosphate |
| ddGTP | di-deoxyguanosine triphosphate |
| ddTTP | di-deoxythymidine triphosphate |

| | |
|---|---|
| DEPC | diethyl pyrocarbonate |
| DEFECT | Interstitial **D**eletions of 11p and EXT2, **E**nlarged partial **F**oramina, Multiple **E**xostoses, **C**raniofacial dysostosis, mental re**T**ardation (MIM 601224) |
| dGTP | deoxyguanosine triphosphate |
| DIG | digoxigenin |
| DMD | Duchenne muscular dystrophy (MIM 310200) |
| DSM-III R | Diagnostic and Statistical Manual |
| DNA | deoxyribonucleic acid |
| DNase | deoxyribonuclease |
| $dH_2O$ | distilled water |
| ds | double-stranded |
| DTT | dithiothreitol |
| dTTP | deoxythymidine triphosphate |
| EBV | Epstein-Barr virus |
| *E. coli* | *Escherichia coli* |
| EDTA | ethylene diaminetetracetate |
| EEGs | electroencephalograms |
| EST | expressed sequence tag |
| EtBr | ethidum bromide |
| EtOH | ethanol |
| EXT1 | Multiple exostoses Type I (MIM 133700) |
| FCS | fetus calf serum |
| FISH | fluorescence *in-situ* hybridization |
| FITC | fluorescein isothiocyanate |
| $FGFR_3$ | fibroblast growth factor receptor |
| FRAXA | fragile site A (X-chromosome) |
| FRAXE | fragile site E (X-chromosome) |
| FRAXF | fragile site F (X-chromosome) |
| FSCX | flow-sorted chromosome X (cosmid library) |
| g | gram |
| GAP | Genome Assembly Program |
| G6PD | glucose-6-phosphate dehydrogenase |
| GCG | University of Wisconsin Genetics Computer Group |
| GDB | genome database |
| GE | genome equivalent |
| GLA | galactosidase, alpha |
| GRAIL | gene recognition analysis internet link |
| GSIX | encodes Xg blood group antigen |
| GTS | Gilles de la Tourette Syndrome (MIM 137580) |
| h | hour |
| HD | Huntington's disease (MIM 143100) |
| HLA | human leukocyte antigens |
| HME | Hereditary Multiple Exostoses (MIM 133700) |
| HPRT | hypoxanthine guanine phosphoribosyl transferase |
| HRR | haplotype relative risk test |
| IBD | identical by descent |
| ICD-10 | International Classification of Diseases |
| ICRF | Imperial Cancer Research Fund |

| | |
|---|---|
| ILRA | interleukin-3 receptor |
| IPTG | isopropyl-β-D-thiogalactopyranoside |
| KAL1 | Kallman syndrome gene |
| kb | kilobase |
| KFS | Klippel-Feil syndrome (MIM 148900) |
| LGS | Langer-Giedion syndrome (MIM 150230) |
| LGS & MR | LGS and mental retardation (MIM 150230) |
| LiCl | lithium chloride |
| LINES | long interspersed repeat elements |
| LMP | low melting point agarose |
| Mb | megabase |
| µg | microgram |
| µl | microlitre |
| mg | milligram |
| MIC2 | encodes cell surface antigen CD99 |
| min | minute |
| ml | millilitre |
| ML | patient ML |
| mM | millimolar |
| MNK | Menkes disease |
| MOPS | 3-N-morpholino-propanesulphonic acid |
| MRI | magnetic resonance imaging scanning |
| mRNA | messenger RNA |
| MYC | Myc family of cellular oncogenes (MIM 133700) |
| MYBL1 | avian myeloblastosis viral oncogene homologue-like 1 |
| NCBI | National Center for Biotechnology Information |
| NF1 | Neurofibromatosis type 1 (MIM 16220) |
| OD | optical density |
| ODF1 | outer dense fiber of sperm tails 1 (MIM 182878) |
| oligo (dT) | oligodeoxythymidylic acid |
| OMIM | on-line Mendelian Inheritance in Man |
| O/N | overnight |
| ORF | open reading frame |
| ORNL | Oak Ridge National Laboratory |
| OTC | ornithine transcarbamylase |
| P1 | bacteriophage artificial chromosomes |
| PAC | P1-derived artificial chromosomes |
| PAGE | polyacrylamide gel electrophoresis |
| PBS | phosphate buffered saline |
| PCR | polymerase chain reaction |
| PDD | Pervasive Developmental Disorders |
| PDHA1 | pyruvate dehydrogenase E1 alpha polypeptide |
| PEG | polyethylene glycol |
| PET | positron emission tomography scanning |
| PFGE | pulsed field gel electrophoresis |
| p. f. u. | plaque forming units |
| PGK1 | phosphoglycerate kinase 1 |
| PHA | phytohaemagglutinin |
| PhOH | phenol |

| | |
|---|---|
| PI | propidium iodide |
| PKU | Phenylketonuria (MIM 261600) |
| POLA | polymerase (DNA directed), alpha |
| PON | paraoxonase enzyme |
| RSP4X | encodes ribosomal protein S4 on X chromosome |
| RF | recombination fraction |
| RFLP | restriction fragment length polymorphism |
| RNA | ribonucleic acid |
| RNase | ribonuclease |
| rpm | revolutions per minute |
| RT | reverse transcriptase |
| SDC2 | Syndecan-2 (MIM 142460) |
| SDS | sodium dodecyl sulphate |
| sec | second |
| SMCX | encodes H-Y antigen |
| ss | single-stranded |
| SSC | standard saline citrate |
| STS | sequence tagged site |
| STS | Steroid sulfatase gene |
| TAE | Tris-acetate |
| TB | terrific broth |
| TBE | Tris-borate-EDTA buffer |
| TdT | terminal deoxynucleotidyl transferase |
| TE | Tris-EDTA buffer |
| TEL | telomere |
| TEMED | N,N,N',N',-tetramethyl-1,2-diaminoethane |
| TIMP1 | tissue inhibitor of metalloproteinase 1 |
| TDT | transmission disequilibrium test |
| TNA | trinitriloacetic acid |
| TR | Texas red |
| Tris | tris (hydroxymethyl) amino methane |
| TRPS1 | Trichorhinophalangeal syndrome (MIM 190350) |
| TSC | Tuberose Sclerosis (MIM 191100) |
| UBE1 | ubiquitin activating enzyme E1 |
| UV | ultraviolet |
| VNTR | variable number of tandem repeats |
| v/v | volume for volume |
| WAGR | **W**ilms tumour, **A**niridia, **G**enital anomalies, and mental **R**etardation (MIM 194070) |
| w/v | weight for volume |
| w/w | weight for weight |
| WWW | world-wide web |
| X-gal | 5-bromo-4-chloro-3-indolyl-$\beta$–D-galactopyranoside |
| XE7 | encodes XE7 antigen: homology to CD99 |
| XG | encodes Xg blood group regulator |
| XIC | X chromosome inactivation center |
| XIST | inactive X specific transcript |
| YAC | yeast artificial chromosome |
| ZFX | zinc finger protein of unknown function |

# CHAPTER I

## INTRODUCTION

A year old infant points to an object with slurping noises to share his interest with mother, and at the age of two, peekaboo becomes a favourite form of play. Before 3, a normal child engages in make-believe and pretends to feed his teddy-bear with an empty spoon. In contrast to these normal behaviours, patient ML (1. 1), points to an object only when she wants to have it and has little interest in playing with toys or friends. She twiddles and flicks her spoon repetitively, and keeps the same shampoo bottle by her side. As a young adult, aged 29, she is still absorbed in these activities. She is diagnosed as autistic. She also suffers from multiple exostoses which are present in her wrists, knees, ankles, and left clavicle.

Autism (1. 2) and multiple exostoses (1. 3) are considered to be hereditary, but the biological basis for these two disorders is currently unclear. The positional cloning strategy (Collins, 1992) (1. 4. 2), initiated by either identification of chromosomal abnormalities (1. 4. 2. 1a) or genetic linkage analysis (1. 4. 2. 1b), has been successful in identifying genes for such disorders, without any prior insights into causation, simply by localizing disease genes on chromosomes. Linkage analysis attempts to determine an initial localization of a disease gene by looking at the segregation pattern between an identifiable disease phenotype and a known marker in a pedigree.

Linkage analysis, however, may not be suitable for traits such as autism. Autistic patients rarely marry and reproduce, and it is unusual to find more than one affected person in a family. Moreover autism may be polygenic in which multiple loci contribute small effects towards the occurrence of the disease (Bolton and Rutter, 1990).

These problems have not been found in multiple exostoses. The symptoms are clinically visible, but not lethal, making it easier to collect family data. Disease genes in multiple exostoses are heterogeneous and two genes have so far been isolated. Both loci can be involved in contiguous deletion syndromes (1. 3. 6) with other disorders, making it a more serious disorder. Other multiple exostosis loci are still to be discovered. However, most reported cases are sporadic, and mild cases might not be systematically reported.

The availability of chromosomal rearrangements (1. 4. 2. 1a) that disrupt a disease gene, such as deletions, duplications, insertions, or translocations, provide an opportunity to precede directly with positional cloning, without the initial step of linkage analysis. The breakpoint of a translocation immediately targets the region of a chromosome in which to search for a disease locus. Such translocations and deletions were very important in cloning the genes for Duchenne Muscular Dystrophy (Kunkel *et al.*, 1985; Ray *et al.*, 1985), Retinoblastoma (Friend *et al.*, 1986), Neurofibromatosis Type I (Cawthon *et al.*, 1990; Wallace, 1990), Choroidermia (Cremers *et al.*, 1990), Aniridia (Ton *et al.*, 1991), Menkes disease (Chelly *et al.*, 1993; Tumer *et al.*, 1992) and many others listed in reviews by Collins, 1992

and 1995 (Collins, 1992; Collins, 1995).

ML is rare among patients suffering from autism and multiple exostoses for she also has a reciprocal translocation between her X and 8 chromosomes. The translocation has opened the way for positional cloning of candidate genes for these diseases. In this thesis the method was first undertaken to identify the translocation breakpoint on the X chromosome. Part of the DNA in this region was sequenced and a gene disrupted by the translocation was identified. The same procedure was carried out on chromosome 8 and a gene was isolated. These genes may be candidates for contributing directly to the phenotype of this patient.

## 1.1 Patient, ML

A 29 year old female patient, ML (Fig. 1. 1A and 1B), has a karyotype of a balanced translocation 46,X,t(X;8)(p22.13;q22.1) and manifests multiple exostoses, short stature, mental retardation, epilepsy and autism. Her developmental history and current level of function, outlined in the following, were obtained by interviews with her mother by Bolton *et al.* (Bolton *et al.*, 1995).

### 1.1.1 Developmental History of Patient, ML

ML was born in 1967 as the first of two children. The second child was a son. Her mother was 37 years old at the time of birth. The delivery was taken at home, but because of a 10-minute delay before her first cry, she was taken to the local special care unit. Although she was initially hypotonic, breathed irregularly, and developed abdominal distension, she subsequently improved. She sat at 15 months without aid and walked at 20 months.

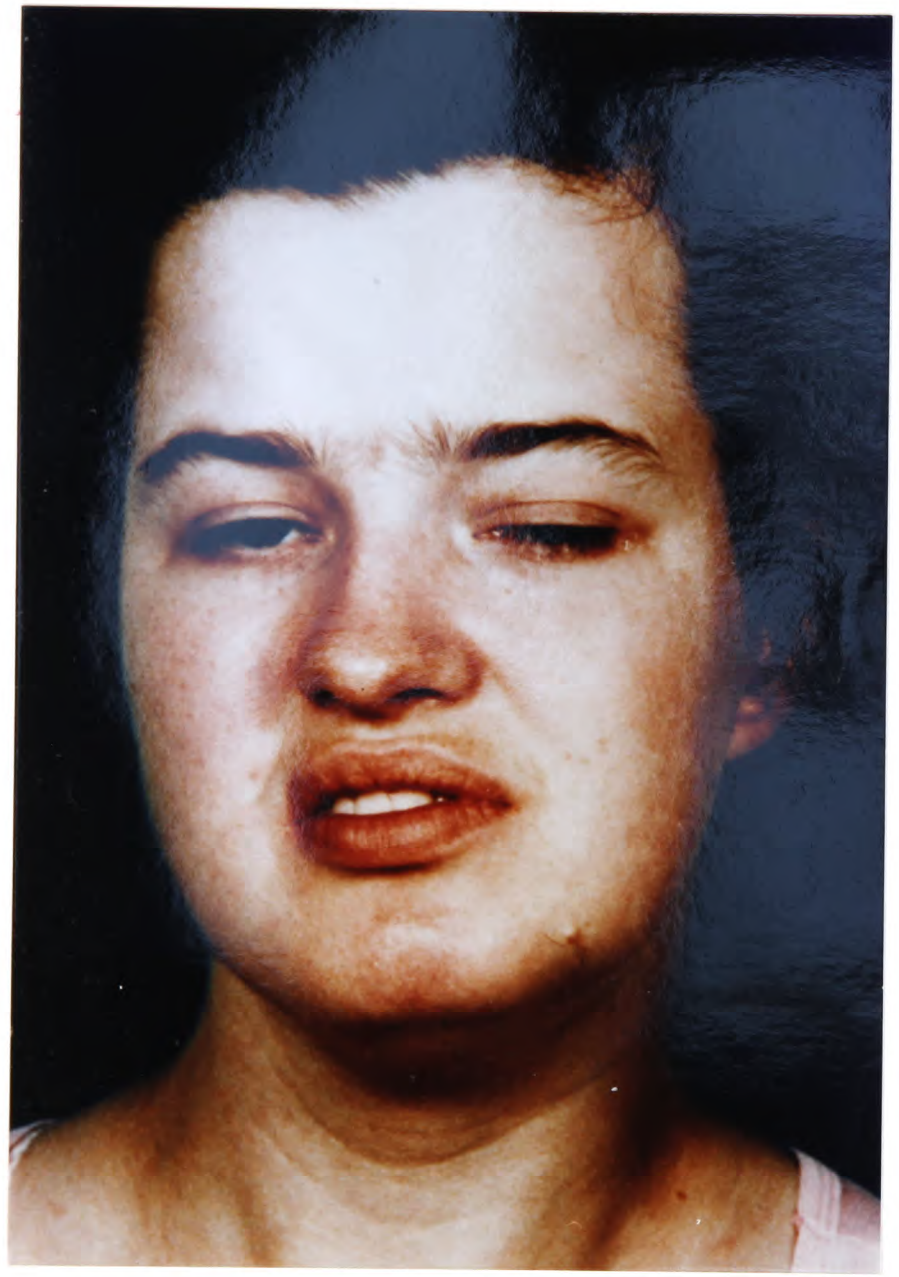




Fig. 1. 1A. and 1B. Patient ML at age 26.

During early childhood she had 3 febrile convulsions and during adolescence developed grand mal epilepsy which has been controlled since by anti-epilepsy (phenobarbitone) medication.

### 1. 1. 2 ML's Physical Examination

ML is now a young adult. She is short in stature (151.5 cm, 3rd centile). Her span is 145 cm and her hands are 15.8 cm in length (less than 3rd centile) (Fig. 1. 2). The 4th and 5th metacarpals are very short. ML has mild brachycephaly, but her head circumference is near the average at 55.9 cm.

Her facial appearance, with eye brows thicker medially, led to a misdiagnosis of Langer-Giedion Syndrome (LGS), but in October, 1993, ML was clinically re-examined and re-diagnosed as suffering from multiple exostoses (EXT: see section 1. 3). LGS, also known as Trichorhynophalangeal Syndrome Type II [LGS/TRPS2: MIM 150230], results from a contiguous deletion of genes related to Trichorhynophalangeal Syndrome Type I [TRPS1: MIM 190350] and EXT, and manifests characteristics of both (Hou *et al.*, 1995). TRPSI is characterized by distinct facies (large anteverted ears, a bulbous nose, elongated philtrum, thing upper lips, sparse hair, receding mandible, malpositioned teeth, and medially bushy eyebrows). Multiple exostoses is characterized by cone shaped epiphyses. However, according to P. Bolton and Y. R. W. Yates (Bolton *et al.*, 1995), ML's ears are not large (5.9 cm) and her nose, philtrum and lips are normal. It was not possible to determine if there was epiphysial coning since her metacarpal epiphyses had already fused. ML, however, lacks joint laxity and multiple exostoses are present around both ankles, knees, wrists and left clavicle. She walks with a wide-based gait is



Fig. 1. 2. The hands of patient ML showing the evidence of multiple exostoses around her metacarpals.

poorly coordinated, and her shoulders are remarkedly sloped. Thus, the observations more closely corresponded to a diagnosis of multiple exostoses than LGS.

### 1. 1. 3 ML's Language and Social Behaviours

At the age of two she began to use single words and only by age of six did she begin to use phrased speech. Her speech is currently abnormal with delayed echolalia. She also fails to use gestures and facial expressions for her responses. Her understanding of speech is limited. Her IQ is 35 (Merrill Palmer Assessment) (Stutsman, 1948). ML shows no attachment to and no interest in others. She seeks neither social reciprocity nor comfort. She absorbs herself in listening to music and musical boxes, watching things spin and looking at picture books, and she sometimes rocks or claps her hands repetitively. She dislikes objects being moved or her routine altered. She has been periodically aggressive, leading to self injury. Her diagnosis of autism was based on the DSM-III R (Diagnostic and Statistical Manual, American Psychiatric Association, 1993) and ICD-10 (International Classification of Diseases, 10th ed., World Health Organization 1993) criteria for autism (see Section 1. 3) and administered by P. Bolton.

### 1. 1. 4 ML's Family

ML is from the socio-economic middleclass. Her mother is a retired midwife. Her father was an optician who died of stomach cancer at the age of 53, so genetic information on him was not available. The karyotypes of her mother and brother were observed to be normal. Her brother was hospitalized frequently during his infancy because of an undiagnosed illness characterized

by persistent vomiting, diarrhoea and sweating. His coordination was remarkedly poor and he was sent to a school for children with cerebral palsy. In his childhood, he had great difficulty in learning to read and write, but managed to continue his education at a polytechnic college where he finished and obtained an engineering degree.

There is no family history of multiple exostoses, epilepsy or mental retardation. The chromosomal translocation was not seen in either her mother or brother. Her father is deceased and his karyotype was not available.

## 1. 2. AUTISM

Autism was first identified and described independently by Leo Kanner (1943) (Kanner, 1943) and Hans Asperger (1944) (Asperger, 1944). Both chose to use the word 'autistic', introduced by psychiatrist Ernest Bleuler in 1911. Bleuler coined 'autism' from the Greek word *autos* meaning 'self', to describe a withdrawal characteristic in schizophrenia (another term coined by Bleuler). Both Kanner, working in Baltimore, and Asperger, working in Vienna, came to the same definition for 'autism' by observing children, but there was a difference. Asperger included a subgroup of autistic children with a wider range of symptoms, including nearly normal children, while all of the 11 children studied by Kanner had the three 'nuclear' features of autism. Kanner defined this as (1) a preference for being alone--*'autistic aloneness'*, (2) an insistence on unchanging surroundings--*'obsessive insistence on sameness,'* and (3) a delay in language development. The current definition of autism (Fig. 1. 3) overlaps with Kanner's three cardinal features and clinical

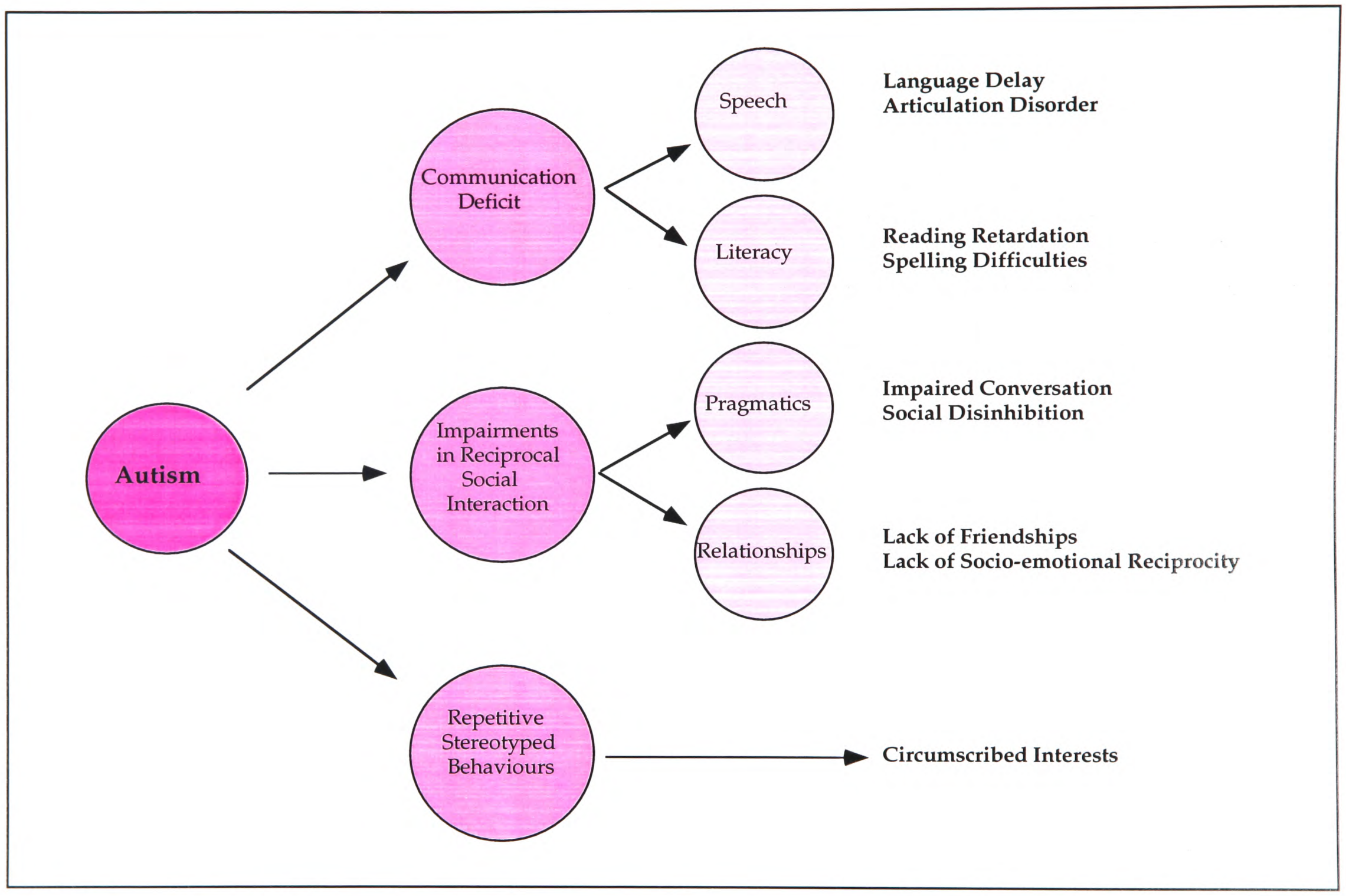



Fig. 1. 3. Concepts behind phenotypic indices of autism from Bolton *et al*, 1991

diagnostic criteria specified under the following headings:

1. Qualitative impairment in reciprocal social interaction.
2. Qualitative impairment in verbal and non-verbal communication, and in imaginative activity.
3. Markedly restricted repertoire of activities and interests.

For clinical convenience, currently, Asperger's syndrome (AS) is defined as a less severe form of autism. The patient can be of normal intelligence and language capability, but lacks social and communicative skills. AS is classified as a subgroup within the spectrum of autistic disorders. Other subgroups are Childhood Disintegrative Disorder (CDD), and Atypical Autism (AA). CDD individuals develop normally until 3 years of age and then gradually manifest a severe autistic-like disability accompanied by increasing psychomotor decline and mental retardation that deteriorates with age. AA is a term for an individual at either end of the autistic spectrum with either a very high or low function. These three, AS, CDD, and AA, are denoted as Pervasive Developmental Disorders (PDD). The disorders on the autism spectrum are not discrete and diagnostic complications arise between AS and AA and in particular between severe autism and severe/profound mental retardation.

### 1. 2. 1 Diagnosis

Autism is diagnosed as a behavioural disorder solely from interview and observation. There is no biochemical test. Clinical diagnostic criteria have been well established since the late 1980s. Two are in common use, the Diagnostic and Statistical Manual (DSM-IV) of the American Psychiatric Association, 1994, and the International Classification of Diseases (ICD-10) criteria of the World Health Organization. Both criteria cover a spectrum of

autistic disorders.

For patient ML, the ICD-10 set of criteria for diagnosis was used. The criteria are based on the Autism Diagnostic Interview (ADI) (Le Couteur *et al.*, 1989) and the Autism Diagnostic Observation Schedule (ADOS). ADI and ADOS are based on series of observations with pre-defined rankings on which the investigators rate the patient. The ADI is an interview with one of the parents (usually the mother) and ADOS is based on interviews and observations of the patient. The interview covers details of obstetric history in pre-, mid- and post-pregnancy. Also covered are details of the patient's behaviour and social communication, based on cognitive abilities (language and grammar), mannerisms (facial expression, body movement and gestures), social interactions (reciprocal communication and cooperation), interests and everyday skills. Autistic diagnoses and symptoms are evaluated according to the mental age rather than the age of the child. All are videotaped and are assessed by multiple investigators for reliability of diagnosis.

### 1. 2. 2 Epidemiology

The first epidemiological study of autism was reported by Lotter (1966) (Lotter, 1966). He found 35 autistic patients out of 78,000 children in the county of Middlesex, giving an incidence rate of 4.5 per 10,000 for children aged 8 to 10. He divided the 35 autistic children into two subgroups, as to whether they exhibited Kanner's three nuclear features or not. There were 15 children out of 78,000 who had all Kanner's nuclear features, giving an incidence of 2 per 10,000. There was an excess of boys in both groups, 2.8 boys to 1 girl in the nuclear group, an 2.4:1 in the non-nuclear group (Lotter, 1966). This sex

difference is a common observation. In a large-scale population study in London by Wing (1981), the ratio of boys to girls was 15:1 at the highest ability level, 2:1 at the lowest ability level (Wing, 1981). He concluded that the sex ratio increases with increased ability. In 1982, Lord *et al.* found 384 autistic boys and 91 autistic girls, aged 3 to 8 years at the University Clinic of North Carolina (Lord *et al.*, 1982). The ratio of boys to girls was 5:1 at the higher end of the ability range and 3:1 at the lower end. The autistic girls were, on average, more seriously impaired than the autistic boys. Bryson *et al.* (1988) studied 20,800 Canadian children between 6 and 14 years and found 21 to be autistic, corresponding to an incidence of 10 per 10,000. The excess of boys was 2.5:1 (Bryson *et al.*, 1988). This estimate of the incidence is rather high and in general it is considered to be 2-4 per 10,000 (Rutter, 1970; Smalley *et al.*, 1988). The excess of autistic boys over girls is consistent, as was originally noted by both Kanner and Asperger.

Kanner and Asperger had not noted an association of autism with mental retardation although they observed there were individuals with some degree of mental deficiency. However, all epidemiological studies have indicated a strong association between the two disorders. Rutter and Lockyer in London, 1967, found 40% among 63 autistic children were severely retarded with IQ of less than 50, 30% moderately retarded with IQ band between 50 and 70, and 30% normal with an IQ above 70. IQ appeared to be stable. The IQ scores of the same autistic individuals remained the same at a second testing ten years later (Rutter and Lockyer, 1967). Three-quarters of the Middlesex (Lotter, 1966) and Canadian (Bryson *et al.*, 1988) populations of

autistic children were mentally retarded. Only 16% of 475 autistic children in the North Carolina samples had IQ over 70 (Lord *et al.*, 1982). The shape of the distribution of the IQ test scores for both autistic populations was the same as for the normal population except for shifting the mean values to 50 and 70 respectively. Autistic children unlike normal children, never reach extremely high levels of IQ (Goldstein and Lancy, 1985).

Both Kanner and Asperger were impressed by the unusual skills of some autistic children. Despite generally low measured intelligence, there can be exceptionally high performance on certain isolated skills involving construction or space. Kanner coined the term *'islets of abilities'* while others call them *'splinter skills'*. Autistic children are good at 'Block Design' in the Wechsler Scales of Intelligence Test, copying an abstract pattern with little cubes within a time limit. They almost always score better on precise information subtests than on comprehension subtests. In contrast, it is often the other way round for retarded non-autistic children. Some autistic patients are good at 'rote memory' and can report the day of the week for any given date, even in the past and future. Some can read aloud fluently without comprehension (Frith, 1993).

Although the Middlesex study and Kanner's samples showed that autism occurred more often in families of superior socio-economic status, in general, it is considered there is no link between the two (Gillberg, 1984; Schopler *et al.*, 1979). In a recent paper, however, instead of linking autism to financial status, a relationship was demonstrated between autism and occupations of parents. The incidence of autistic children whose fathers and

grandfathers engaged in engineering was twice higher than those in other occupational classes. Occupational preference is often dependent on individual skills and capability. The authors concluded that engineering is thought to make the least requirement of the skills of social relationships and communication, which autistic individuals lack the most. The paper suggests that autistic-like elements may exist in the personalities of the parents of autistic patients, and these may have contributed to the choice of career and occupation of these parents (Baron-Cohen *et al.*, 1997).

The onset of autism was investigated by Ornitz, 1977. He interviewed over 100 parents of autistic and normal children under the age of four. Only half of the parents had any suspicion of abnormality in the first year of life (Ornitz *et al.*, 1977). In another survey in Britain, only 12 out of 93 parents with an autistic child noticed any oddities before their child's first birthday (Newson *et al.*, 1984). Distinctive abnormal patterns in the loops and whorls on fingertips and toes have been reported in some autistic children (Walker, 1978). Fingerprints are considered a physical indicator of foetal development and might be a clue for the biological onset of autism.

A disproportionally larger number of autistic patients have been observed from mothers immigrating from tropical countries. 22 against an expected 5 for the size of the population was reported in Caribbean-born English (Wing, 1979). A similar excess was found in Sweden (Gillberg *et al.*, 1987). On this basis, autism has been suggested to be due to a viral infection and these mothers from their new homes in northern countries before they had developed immunity. Possible viral causes of autism may be

retroviruses, herpes, or cytomegalovirus which can remain dormant for years before their activation. Autism may result if the central nervous system becomes infected before or after birth. A well-documented association with an infectious disease is that between autism and congenital rubella (Chess, 1971; Gillberg *et al.*, 1987). Virus infection, however, is not likely to account for more than the rare case with modern immunization programs (Lord *et al.*, 1994).

In summary, autism is a rare behavioral disorder, with an incidence of 2-4 per 10,000 children and with the skewed sex ratio of 3-4 male to 1 female (Smalley *et al.*, 1988). 75% of autistic patients were mentally retarded (Rutter, 1970; Smalley *et al.*, 1988), but some patients have exceptional skills in 'rote memory' and in performing complicated tasks if these did not required social understanding (Frith, 1993). The appearance of symptoms after 1 year of life may suggest that autistic children suffer from a deficiency in a particular mental capacity that in normal development does not mature until the end of infancy. Autism is not a progressive disease, but persists. It cannot be corrected and affects the whole of development from infancy onwards. The viral cause of autism is negligible. Autism occurs in all kinds of families and cultures (Gillberg, 1984; Schopler *et al.*, 1979).

### 1. 2. 3 Neuroanatomy

For many years, it was thought that a child could become autistic because of psychologically threatening experiences, such as deprivation or rejection by the mother. Many cases, however, remained unexplained and efforts were made to discover a biological cause of autism.

A defective structure in the brain was suggested and Bauman and Kemper (1985) found an increased cell density in post-mortem examinations of several autistic patients. They explained that brain cells at first proliferate and then 'die back' in normal development, but in autism the instruction to switch off the nerve cell growth occurs too late, resulting in increased cell density (Bauman and Kemper, 1985). However, they later (1991) reported no gross lesions found in the brains of autistic patients at autopsy (Bauman, 1991; Ciaranello *et al.*, 1982). Availability of brain autopsies are few and the 'live brain' has been studied using computerized tomographic (CT) scanning, magnetic resonance imaging (MRI), and positron emission tomography (PET) scanning. CT scans make brain sections visible laminographically by processing computer images. MRI allows visualization of neural tissue for the detection of lesions. PET measures the rate of glucose utilization of specific brain areas. Delong (1978) identified enlargement of the left ventricles using CT scans. The left hemisphere is known to be associated with language development and language delay is one of the cardinal features of autism (Delong, 1978). A number of neuropsychological studies seemed to confirm this hypothesis, but any such lesion in autistic children may simply be secondary to the underlying cause (Fein *et al.*, 1984). In an MRI study, Courchesne *et al.* (1988) identified an unusual malformation of a small part of the cerebellum in a large number of autistic people with and without mental retardation (Courchesne *et al.*, 1988). They (1994) and Hashimoto *et al.* (1994) narrowed the abnormal regions to the cerebellar vermian lobules VI and VII in autistic patients (Courchesne *et al.*, 1994; Hashimoto *et al.*, 1994). 86% of the patients

had vermian hypoplasia and 16% of these had vermian lobules smaller than the mean area of normal subjects. 12% of the autistic patients had hyperplasia of which 34% had larger than average vermis. Autopsy analyses also found severe Purkinje neuron loss in the posterior vermis (lobules VI and VII and VIII to X) (Courchesne *et al.*, 1994). Abnormalities of the temporal lobes in autism were identified in an MRI study (Piven *et al.*, 1996) and also in a PET study (Chugani *et al.*, 1996). The infant monkeys with neonatal damage specific to the medial temporal lobe (amygdalohippocampal complex) displayed autistic-like behaviours (Bachevalier, 1996). The relationship between these abnormalities and behavioural symptoms is unclear, but these neuroanatomical studies identified not gross but visible biological abnormalities in three specific brain regions, the frontal lobes, cerebellum, and temporal lobes.

### 1. 2. 4 Neurophysiology

Autism was initially considered solely a behavioural disorder and was diagnosed on the basis of behavioural symptoms. However, Ornitz (1978) was the first to report its association with an organic factor, epilepsy (Ornitz, 1978). This epilepsy is not a periodical seizure commonly seen, but unusual with one peak of incidence during early childhood and second peak during adolescence (Gillberg *et al.*, 1987). The incidence occurs in 25-40% of autistic patients (Rutter *et al.*, 1990). The different pattern was observed among profoundly retarded autistic patients with the first peak in the pre-school period (Rutter *et al.*, 1990).

Autonomic measures of respiratory responses, heart rate and skin

conductance, as well as electroencephalograms (EEGs) have been studied in autistic patients and all results have demonstrated severe developmental immaturity in autistic patients (James and Barry, 1980). Habituation of the orienting response instinctively occurs in any animals to protect themselves because they remember the alert stimulus occurred before, but in autistic children this response was abnormally slow (James *et al.*, 1980). This may relate to cognitive deficit in autism in which repeated stimuli are not recognized as new, a kind of abnormality in information processing (Frith, 1993).

### 1. 2. 5 Neurochemistry

A large number of biochemical studies have also been done. Demasio and Maurer (1978) observed autistic behaviours in adults with impairments in the dopamine system. Dopamine, a monoamine, acts as a neurotransmitter in the central nervous system. Animals with disruptions of the dopamine pathway manifested neurological symptoms such as poor voice control, lack of expression, flapping limb movements, repetitious actions, and lack of spontaneity (Demasio and Maurer, 1978; Poncet and Habib, 1994). Ridley and Baker (1983) demonstrated overactivity and highly stereotyped behaviours in marmosets treated with amphetamine which induces dopamine activity by blocking some of the functions of the amygdala (Ridley and Baker, 1983). Dopamine antagonists have been shown to be effective in improving some symptoms of autism, including motor stereotypes, but their use was associated with a number of adverse effects such as Parkinsonian symptoms (McDougle *et al.*, 1994). Dopamine receptor supersensitivity may

be related to stereotypes and self-injury (Lewis *et al.*, 1995), but unrelated to other behavioural symptoms of autism (Coleman and Gillberg, 1985; McDougle *et al.*, 1996).

Another neurotransmitter, serotonin, was found to be significantly elevated in the blood in 6 of 23 children (26%) with autism (Schain and Freedman, 1961). Hyperserotonemia was also observed in other autistic children, levels 30% to 50% above normal. The abnormality occurred in the uptake and storage of serotonin by blood platelets (Anderson and Hoshino, 1987). Elevated platelet serotonin was observed in 40% of autistic children and the elevation was correlated with that of their mothers, fathers, and siblings (Rolf *et al.*, 1993). However, some children (Hoshino *et al.*, 1984) and adults (McBride *et al.*, 1989) with autism had decreased serotonin levels. Dietary depletion of tryptophan, a precursor of serotonin, resulted in a depressive relapse in 65% of 17 autistic patients and a worsening of global behaviours such as whirling, flapping, pacing, hitting oneself, rocking, and toe walking, but the authors warned that this short-term depletion of the serotonin precursor may only be aggravating the autistic condition rather than being related directly to the original cause of autism (McDougle *et al.*, 1996).

The bimodal distribution of serotonin, hyper- and hyposerotonemia, were more closely investigated. A continuum of serotonin levels was found, highest in the severely retarded without autism, high in the severely retarded with autism, and relatively normal in higher functioning autism without retardation (Comings and Coimings, 1988). This suggested that the serotonin

level was more likely a marker for mental retardation than autism (Comings and Comings, 1991). Herault, *et al.* (1997) found the most severely affected children with neurological syndrome had the highest serotonin levels in whole blood, suggesting a hyperserotonemia in whole blood, as a marker of neurological syndromes rather than of autism (Herault *et al.*, 1996). They also found that autistic patients had high urinary serotonin, but low whole blood serotonin. Furthermore they tested markers of the serotonin receptor and tryptophan hydroxylase, but there was no significant differences in allele and genotype frequencies in their autistic population compared with an age-matched healthy population (Herault *et al.*, 1996). However, a short variant in the promoter of the serotonin transporter gene was preferentially transmitted among 86 autistic nuclear families, suggesting a possible susceptibility locus for autism (Cook *et al.*, 1997).

### 1. 2. 6 Genetic Factors

Pre- and perinatal hazards such as birth asphyxia, respiratory distress, resuscitation, delayed cry, and apnoea in autism is generally high (Folstein and Rutter, 1988). The incidence was reported 37% in autism and 12% in schizophrenia (Kolvin *et al.*, 1971). These obstetric complications, however, have been considered to follow from the disorder rather than be its cause (Goodman, 1990; Rutter, 1991) and have been inconsistently related to autism (Bailey *et al.*, 1991). Many believe that these complications are determined not only from environmental factors but from a genetic disorder (Bailey *et al.*, 1995; Bolton *et al.*, 1991; Folstein and Rutter, 1977; Folstein *et al.*, 1988; Rutter, 1991; Rutter *et al.*, 1990; Smalley *et al.*, 1988). Regarding the obstetric

complications of ML's birth, the same conclusion was reached in which Bolton (1995) states 'Obstetric complications associated with autism are a consequence of the underlying genetic abnormality, rather than a cause of the condition (Bolton *et al.*, 1994).

The first systematic twin study of autism reported a concordance of 36% in 11 MZ twins and no concordance in 10 DZ twins, ascertained by a nationwide search in the UK for same-sexed pairs (Folstein *et al.*, 1977). Very similar results were reported by Bailey again in a nationwide UK sample (Bailey *et al.*, 1991). In Scandinavia, Steffenburg reported concordance rates of 91% in MZ twins and 0% in DZ twins (Steffenburg *et al.*, 1989). This consistently high concordance in MZ in comparison with DZ indicates a genetic factor in autism, which estimates the heritibility as >90% (Bailey *et al.*, 1995).

Another important parameter in assessing the genetic contribution to autism is the relative risk. Large values of the parameter also predict the relative ease of genetic mapping. The relative risk for siblings, Lambda S, is defined as recurrence risk for a sibling of an affected individual divided by the risk for the general population (Risch, 1990). Lambda S is 500 for cystic fibrosis, 15 for type I diabetes, 8.6 for schizophrenia and 3.5 for type II diabetes (Risch, 1990). In autism, the incidence in the general population is estimated to be about 2-4 in 10,000 children (1. 2. 2) and the sibling risk to be 3% (Smalley *et al.*, 1988). The resulting Lambda S of 75-100 indicates a relatively strong familial clustering in autism and predicts that attempts to map these genes should be successful considering the strong heritibility of autism.

In spite of this strong concordance among MZ twins and the high lambda S, there is a sharp "falling off" in the rate of autism between first and second degree relatives, and between second and third degree relatives of an autistic patient. This pattern suggests that autistic phenotypes arise from a combination of different autistic loci that interact epistatically. The high concordance in MZ twins results from sharing the same combination of autistic loci. Siblings would rarely share the same combination of multiple loci and on average share 50% of their genes. An observation which is consistent with a number of epistatic autistic loci is the association between severe autism and a familial loading for a broader spectrum of autistic-like phenotypes (retardation, severe language delay, severe reading and spelling difficulties, impaired social reciprocity and repetitive stereotyped behaviours) (Bolton *et al.*, 1991; Rutter, 1991) (Fig. 1. 3). Autism shows a gradient with qualitative impairments in behaviours and may be a polygenic disorder determined by the combined action of a number of genetic loci, each with a small effect (Bailey *et al.*, 1995; Rutter, 1991).

### 1. 2. 7 Genetic Heterogeneity

Autism is often associated with other genetic disorders more often than expected by chance. Traits closely studied are Fragile X syndrome, Phenylketonuria [PKU: MIM 261600], Tuberose Sclerosis [TSC: MIM 191100], and Neurofibromatosis [NF1: MIM 16220].

Approximately 2.5% of autism is associated with Fragile X syndrome (Bailey *et al.*, 1995). The rate of the association was once reported to be much higher, as high as 22.5% (Fish *et al.*, 1986; Gillberg and Wahlstrom, 1985), but

this estimate has been under review over the last decade, since the diagnosis became well defined with the identification of the FRAXA, FRAXE, and FRAXF genes for the Fragile X syndrome. Abnormalities of language delay, echolalia speech, and voice oddities exist in both patients of Fragile X and autism. However, Fragile X patients are noticeable in showing gaze avoidance, social anxiety, and a dislike of being touched. These symptoms are absent in autistic patients. The overlapping similarities between the two were only at the level of some individual symptoms or the shared behavioural similarity due to the mental retardation.

PKU is an inborn error of metabolism inherited as an autosomal recessive disorder and is caused by the deficiency of the liver enzyme, phenylalanine hydroxylase, which converts phenylalanine into tyrosine. Tyrosine is a precursor of dopamine, noradrenaline and epinephrine. PKU patients suffer from the accumulation of toxic phenylalanine and deficiency of tyrosine and develop autistic behaviours (Friedman, 1969). Patients are characteristically mute, mentally retarded, and display some autistic behaviours such as unfriendliness and repetitiveness. PKU can be treated with a phenylalanine-free diet. Most children with untreated PKU are autistic (Friedman, 1969), with the association of about 20%, but autistic symptoms frequently disappeared after the dietary treatment (cited references by Friedman, 1969: Bickel *et al.*, 1954; Woolf, Griffiths, and Moncrieff, 1955; Armstrong and Tyler, 1955; Blainy and Guillford, 1956; Armstrong, Low, and Bosma, 1957; Berry, Sutherland, Guest, and Umbarger, 1958; Lewis, 1959). However, Friedman found no corresponding improvement in IQ after the

dietary treatment and concluded that the autism and mental retardation in PKU patients have different mechanisms (Friedman, 1969).

TSC is an autosomal dominant disorder characterized by multiple and widespread tubers affecting the skin, heart, kidney, eyes, and brain. TSC-determining loci have been mapped to chromosomes 9q34 (TSC1) and 16p13 (TSC2) (van Slegtenhorst *et al.*, 1997). Mental retardation, autism, and epilepsy have known to be associated with TSC although epilepsy is indirectly associated with TSC through its association with autism (Riikonen and Amnell, 1981; Taft and Cohen, 1971). Approximately 2% of autistic patients also suffer from TSC (Bailey *et al.*, 1995; Blomquist *et al.*, 1985; Hunt and Dennis, 1987). Autistic patients with normal IQ rarely manifest TSC (Lord *et al.*, 1994). The risk of autism is 200 to 1000 times higher in TSC patients than in the general population and the risk of TSC is 100 to 300 times greater in autistic patients (Gillberg *et al.*, 1994; Wing, 1979). The association between TSC and autism could be due to the location of cortical tubers (Curatolo *et al.*, 1991; Williamson and Bolton, 1995), and is strongly related to the presence of tubers in the temporal lobes (Bolton and Griffiths, 1997). This finding is consistent with neuroanatomical analysis of MRI (Piven *et al.*, 1996) and the PET scans (Chugani *et al.*, 1996) as described in section 1. 2. 3. The risk of mental retardation was shown to increase with the number of tubers present (Williamson *et al.*, 1995).

NF1 is an autosomal dominant disorder, characterized by tumours of the cranial and peripheral nerves and skin. Gillberg (1984) estimated a 6% association between autism and NF1 (Gillberg, 1984), but a follow-up survey is

needed for confirmation (Folstein *et al.*, 1988).

Both NF1 and TSC are autosomal dominant and following the expected sex ratio would be 1 male to 1 female. However, the sex ratio in autism is 3-4 times higher in males than females (Lord *et al.*, 1982). This discrepancy may be explained by the sex ratio among severely retarded autistic children is less skewed to males, narrowing to equal chance (Lord *et al.*, 1982) as described in section 1. 2. 2. NF1 and TSC disorders, therefore, may be associated with mental retardation rather than with Kanner's cardinal autism (Folstein *et al.*, 1988).

### 1. 2. 8 Chromosomal Abnormalities

Case reports have shown links between autism and gross chromosomal abnormalities including monosomy and trisomy of large chromosome regions, suggesting a variety of autistic loci. They are monosomy of 9p24-pter and trisomy of 6p21-pter (Eden *et al.*, 1985); inversion/duplication of chromosome 3p and 16p, 5p+ and 17p- (Mariner *et al.*, 1986) trisomy of chromosome 6p23-pter and monosomy of 2q37-qter (Burd *et al.*, 1988); interstitial deletion of chromosome 17 and translocation of chromosome 5 (Gillberg and Coleman, 1992); 9p-deletion syndrome (Tully *et al.*, 1993); abnormalities of chromosome 18 (Ghaziuddin *et al.*, 1993); duplication of Xp22.1-pter (Rao *et al.*, 1994); deletion of chromosome 7p11.2 (Vostanis *et al.*, 1994); monosomy of chromosome 5p15.3-pter (Vostanis *et al.*, 1994), and two cases of duplication in 15q11-13 (Baker *et al.*, 1994; Cook, 1997). The involvement of autism with various single gene disorders also indicates genetic heterogeneity.

## 1. 3 MULTIPLE EXOSTOSES

Exostoses occur in patients either singly (solitary exostoses) or multiply (multiple exostoses). Both appear to be identical by histological examination and are characterized by cartilaginous protuberances at the ends of the diaphyses of the extremities. They can arise as dysplastic outgrowths from any growth plate. Exostoses are rarely present at birth, but gradually develop from growth plates between the epiphysis and metaphysis (Fig. 1. 4) until the growth plate closes at the end of puberty. They increase in size and undergo ossification. Other bones that can be involved are the pelvic girdle, ribs, and less frequently, vertebrae, the sternum, skull, and carpal and tarsal bones.

In multiple exostosis, long bone metaphyses thicken. Bone growth is slow and irregular, often giving rise to skeletal deformities and short stature (Schmale *et al.*, 1994). The physical presence of exostoses in multiple exostoses hinders the range of movements such as arm rotation and joint flexibility. In the severely affected, exostoses strain neighbouring tissues, giving rise to complications such as compression or irritation of adjacent nerves, vessels, and tendons (Bhojraj and Panjwani, 1993; Eder *et al.*, 1993).

### 1. 3 . 1 Diagnosis

The symptoms of multiple exostoses are detected by palpable bony swellings and gait disorders. Over 80% are diagnosed during the first decade. The analysis of the morphology is examined by X-ray, CT, ultrasound, and MRI scans.

### 1. 3 . 2 Epidemiology

The frequency of solitary exostosis is high, about 1 in 200 children. Multiple

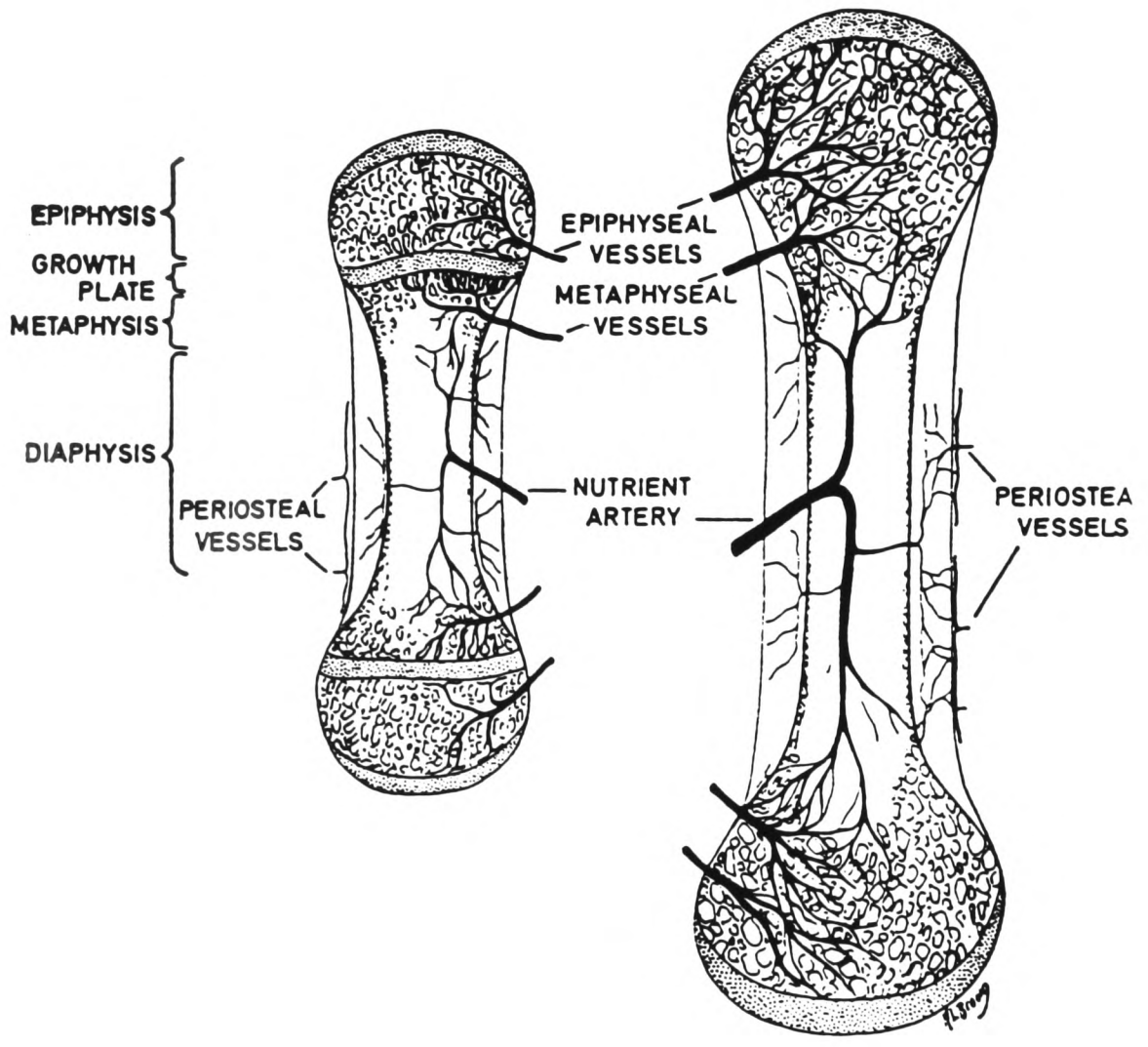



Fig. 1. 4. Diagram showing the bone before (left) and after the growth period (right). The growth plate isolates the epiphysis from the metaphysis during the growing period, but they anastomose after growth has ceased.

exostoses are the commonest of all skeletal dysplasias with a prevalence of 1 in 50,000 in the west (Cook *et al.*, 1993; Schmale *et al.*, 1994). In Chamorros village of Guam Island, the frequency of multiple exostoses was reported as high as 1 in 1,000 (Krooth *et al.*, 1961).

In Chamorros it was also reported that the disease was more frequent and more severe in males. The entire affected sample of an American population surveyed by Wicklund *et al.* (1995), also indicated an excess of males (104:76) (Wicklund *et al.*, 1995). However this skewed ratio was not observed in nuclear families (probands, affected sibs, and parents). The excess of males appeared not to be of any primary genetic origin and to be related to males having more severe and more frequent complications in the disease (Wicklund *et al.*, 1995).

### 1.3.3 Genetic Disorder

The presence of solitary exostosis occurs sporadically in the population and is commonly treated with surgical excision. The presence of multiple exostoses (EXT) is often familial (80%) and designated as Hereditary Multiple Exostoses [HME: MIM 133700 and MIM 133701] [Mckusick, 1992 #116] or Diaphyseal Aclasis. HME is transmitted in an autosomal dominant fashion (Solomon, 1964). The expression of symptoms varies. Solomon (1964) observed under radiographic examination small exostoses in even mildly affected individuals and considered that penetrance is 100% (Solomon, 1964). Full penetrance was also reported in recent years (Hennekam, 1991; Wicklund *et al.*, 1995).

### 1.3.4 Genetic Heterogeneity

HME is genetically heterogeneous with three loci (EXT1, EXT2 and EXT3)

mapped to 8q24.1, 11p11, and 19p, respectively. Buhler and Malik (1984) first observed cytologically a deletion in the region of 8q24.1 in a patient with EXT (Buhler and Malik, 1984). Cook *et al.* (1993) demonstrated strong evidence for linkage of EXT to chromosome 8q24.11-q24.13. They also suggested heterogeneity because 30% of families in their sample did not map to this region (Cook *et al.*, 1993). Subsequently two additional EXT loci were reported: a locus mapping to the pericentromeric region of chromosome 11 (EXT2) by Wu *et al.* (1994) (Wu *et al.*, 1994) and a locus on 19p (EXT3) by Le Merrer *et al.* (1994) (Le Merrer *et al.*, 1994). EXT-like (EXT-L) sequences have been identified by computer homology searches of databases of expressed sequence tags (dbEST) and mapped to chromosome 1p36.1 (Stickens *et al.*, 1996).

The distribution of the three types of EXT genes among patients is not clear, although EXT1 appears to be of higher frequency. In 21 EXT families, Le Merrer *et al.* (1994) estimated linkage in 65% of families to 8q and 35% to 19p (Le Merrer *et al.*, 1994). Another study of 29 families found linkage in 44% of families to chromosome 8, 28% to chromosome 11, and 28% to chromosome 19 (Legeai-Mallet *et al.*, 1995). Blanton *et al.* (1996) observed in 12 U. S. families with EXT six showing linkage with chromosome 8 and six with chromosome 11. None of their families mapped to chromosome 19 (Blanton *et al.*, 1996). By screening both EXT1 and EXT2 genes in 17 families not previously mapped, mutations were found in 12 families (70%), 41% in EXT1 and 29% in EXT2 (Philippe *et al.*, 1997).

The EXT1 and EXT2 genes have been cloned and sequenced (Ahn *et al.*,

1995; Stickens *et al.*, 1996; Wuyts *et al.*, 1996), but EXT3 remains to be confirmed by positional cloning. The coding region of EXT1 is about 2.2 kb with 11 exons. EXT2 is about 3.5 kb with 16 exons. Both genes are significantly homologous with each other, suggesting derivation from a common ancestral gene that gave rise to an EXT multigene family (Stickens *et al.*, 1996). The age of onset and phenotypic variation of EXT could be the result of different allelic combinations of these EXT genes.

Further heterogeneity of EXT may occur on chromosome 8. Yoshiura *et al.* reported an 8q23 translocation in association with multiple exostoses (Yoshiura *et al.*, 1994), and the patient studied in this thesis has a translocation in the region of 8q22.1 (Bolton *et al.*, 1995). Dr. D. Zaletayev and Dr.Q-G Ruan found Russian and Chinese patients linked to chromosome 8q, but not to the EXT1, 8q24.1 region (personal communication, 1996).

### 1.3.5 Malignant Transformation

The functions of the EXT genes are unknown, but sequence and mutation analyses suggest they may act as tumour-suppressor genes (Hecht *et al.*, 1995; Stickens *et al.*, 1996). An exostosis can degenerate to a chondrosarcoma or an osteosarcoma. The risk of this malignant change has been estimated to be 1 to 4% for a solitary exostosis and 20-50% in HME (Matsuno *et al.*, 1988). Hecht *et al.* (1995) lowered the latter estimate to 2-5% after using an age-adjusted incidence rate of 1/100,000 for all bone cancer. They observed loss of heterozygosity in these tumours for chromosome 8 and 11 markers and suggested that EXT genes may be tumour-suppressor genes involved in a multistep model of tumour development (Hecht *et al.*, 1995).

### 1. 3 . 6 Association with Other Disorders

Both EXT1 and EXT2 are associated with contiguous deletion syndromes (reviewed in Ballabio, 1991 (Ballabio, 1991)). EXT1 has been reported in association with Langer-Giedion syndrome, also known as trichorhynophalangeal syndrome Type II [LGS/TRPS2: MIM 150230] (McKusick, 1994). The clinical features of LGS considerably overlap with trichorhynophalangeal syndrome Type I [TRPS1: MIM 190350] (McKusick, 1994). They include sparse scalp hair, bushy eyebrows, a bulbous nose, long philtrum, short stature and cone-shaped epiphyses. The lack of multiple exostoses distinguishes TRPS I from LGS. LGS is now considered to be a contiguous gene syndrome of the TRPS I and EXT1 genes (Ballabio, 1991; Ludecke *et al.*, 1995).

EXT2 is associated with the WAGR syndrome (**W**ilms tumour, **A**niridia, **G**enital anomalies, and mental **R**etardation: MIM 194070) (McGaughran *et al.*, 1995) and the DEFECT 11 microdeletion syndrome (Interstitial **D**eletions of 11p and EXT2, **E**nlarged partial **F**oramina, Multiple **E**xostoses, **C**raniofacial dysostosis, mental re**T**ardation: MIM 601224) (Wuyts *et al.*, 1996).

## 1. 4 CLONING STRATEGIES

Gene cloning strategies have evolved along with technical advances in both molecular biology and world-wide informatics. Genomic DNA fragments can be cloned in a variety of sizes, ranging from 100 kb to 2000 kb in yeast artificial chromosomes (YACs), 75-95 kb in bacteriophages (P1s), 75-300 kb in bacterial artificial chromosomes (BACs) and P1-derived artificial chromosomes (PACs),

35-45 kb in cosmids, and less than 10 kb in plasmids. Genomic DNA fragments can be also amplified *in vitro* by the polymerase chain reaction (PCR) using specific oligonucleotides to prime DNA synthesis. PCR amplifies only short genomic DNA fragments within the range of 100 bp to 10 kb, but can amplify from minute amount of any sizes or kinds of target or template DNA sequence such as whole genomic DNA, cloned DNA, or pieces of DNA fragments. PCR also purifies the desired DNA such that only DNA required for analysis remains, and this DNA is amplified to large quantities for further analysis in a short time of 2-3 h.

Radiation hybrid mapping facilitates the mapping of human DNA sequence to a specific subchromosomal region. In this type of somatic cell hybrid, X-rays are used to generate fragments of the chromosomes of donor cells and these are allowed to fuse to the chromosomes of a recipient cell (Walter and Goodfellow, 1993). Breakpoints occur randomly and this results in a collection of hybrids. Markers close together will tend to occur on the same fragment and the frequency of breakage between two markers provides a linear map order for the DNA clones (Cox *et al.*, 1990). Reasonably high resolution maps of the entire human genome are available with a single panel of 100-200 hybrids (Schuler *et al.*, 1996; Walter *et al.*, 1994).

cDNA cloning has also contributed to gene identification. Searching through genomic DNA for a specific gene is inefficient. Not only are there about 100,000 genes in which to search, but also these genes are coded by only about 5% of the total genomic DNA. Gene cloning attempts to identify mRNA and then sequence the coding cDNA. mRNA is an unstable single

stranded molecule, but cDNA is a complement to mRNA and can be cloned as a stable double stranded DNA, synthesized from mRNA using reverse transcriptase. cDNA therefore represents only functional, coding sequences, and can be used to identify genes in the genome. Collections of cDNA from various tissues from fetus and adults and also from different developmental stages are constructed as cDNA libraries and stored for later use in gene identification. The search for coding regions has improved by the expressed sequence tags (ESTs) which are generated by sequencing part of mRNA usually the 3′ end. The sequences of several hundred thousand ESTs are available in publicly accessible databases which has facilitated gene search by PCR. There are currently about 16,000 ESTs mapped using whole genome radiation hybrids at an average interval of about 0.5 Mb (Schuler *et al.*, 1996).

Accumulation of highly polymorphic markers and the ESTs in databases and computer software for genetic linkage analysis, affected sib-pair analysis, protein and DNA sequence homology searches, and exon prediction, have expedited the search for disease genes.

The earliest cloning strategy is known as 'functional cloning' (1. 4. 1), beginning with identifying defective protein related to a disease. The protein is used to identify the mRNA and DNA that codes it. However, functional insight into the majority of disorders is not known. By 1993 most genes with identifiable biochemical defects had been cloned. As the productivity of functional cloning declined, another method, known as, 'positional cloning' (1. 4. 2) had emerged and was becoming well established when this project began. This method is DNA site-oriented and does not require knowledge of

the defective protein. It begins with mapping a gene to a subchromosomal region by observing cytological abnormalities such as deletions, translocations, or inversion breakpoints (1. 4. 2. 1a), if these are not available, by pre-mapped markers, linked to a gene, searched by genetic linkage analysis (1. 4. 2. 1b), affected sib-pair analysis (1. 4. 2. 1c), and/or allelic association study (1. 4. 2. 1d). Once the initial localization has been determined, physical mapping (1. 4. 2. 2) will be carried out to clone the region of interest into genomic DNA fragments, beginning with large fragments which in turn are cloned into smaller ones using a variety of cloning vectors. At each stage, the presence of coding regions is searched in the fragments by constructing a transcript map (1. 4. 2. 3) until the coding region is found which upon mutation gives rise to the disease. Because of the increasing numbers of ESTs available, a candidate gene may be one of these ESTs. A new strategy, 'positional candidate gene cloning' (1. 4. 3) searches a candidate gene from the ESTs found in the refined physical map by 'positional cloning' technique. This approach has expedited the gene-search process and has become a prime method of identifying human disease genes. This will be applied to identifying genes involved in complex polygenic disorders in the future.

### 1. 4. 1 Functional Cloning

Functional cloning is based on the observation that hereditary diseases caused by defective proteins are translated from mRNA derived from mutated DNA sequences. A partially identified amino acid sequence of the known protein is used to determine its possible DNA coding sequences (resulting from the redundancy in the DNA code) from which are generated a variety of

oligonucleotide DNA probes. These are used as hybridization probes to identify specific cDNA clones in cDNA libraries, and thus the gene encoding the protein. Alternatively, antibodies to the protein can be used to identify mRNA directly through identification of polysomes through antibody binding to the growing peptide. Polysomes, however, have been superseded by antibody screening of expression libraries. These libraries are constructed by cloning cDNA into an expression-vector. Inserts within the recombinant DNA clones are expected to be expressed within the host cell to produce foreign polypeptides. Labeled antibodies can be reacted with colony filters to identify a bacterial colony which synthesized the desired polypeptide.

In some cases, a functional complementation assay has been used to identify human genes which have specific biochemical functions. The conserved human protein may be able to complement a yeast mutant defective in the corresponding protein. Normal phenotypes may be rescued by transforming the mutant yeast cells with a specific human DNA fragment. Besides yeast, mammalian cell lines that are deficient in DNA repair have been also generated and can be transformed by normal human DNA fragments to produce a repair-competent phenotype.

Well known examples of 'functional cloning' are the cloning of genes for blood clotting factor VIII, phenylketonuria, sickle cell anaemia and blood group antigens such as Rhesus, Kell and LW. The first two examples are described briefly in the following.

The haemophilia A gene coding for the blood clotting factor, factor VIII, was identified in 1984. Although factor VIII was present in too low

quantities in human plasma for easy isolation, enough could be isolated from pig blood by protein fractionation techniques using a combination of gel filtration, ion-exchange and/or affinity chromatography. The purified product was cleaved into peptides and for some peptides amino acid sequences were determined. From these sequences, degenerate antisense oligonucleotide probes were prepared and used as primary hybridization probes to screen a pig cDNA library. The pig factor VIII cDNA clone was then used to screen human cDNA libraries and the human haemophilia A gene was identified (Gitschier *et al.*, 1984).

Phenylketonuria [PKU: MIM 261600] was known to be caused by a lack of the enzyme phenylalanine hydroxylase which catalyses the oxidation of phenylalanine to tyrosine. The gene for phenylketonuria was identified using specific antibodies to the enzyme. Phenylalanine hydroxylase was purified from rat liver, where it was known to be expressed. There is a high structural and sequence homology of the enzyme between human and rat. Polysomes containing ribosomes and the growing phenylalanine hydroxylase peptide and its mRNA were precipitated by the antibodies. The purified mRNA was converted to cDNA, which was used to screen a human liver cDNA library to identify the corresponding human gene (Robson *et al.*, 1982).

### 1. 4. 2 Positional Cloning

Positional cloning begins with knowledge of a disease gene's subchromosomal location (Collins, 1992), provided by chromosomal abnormalities (1. 4. 2. 1a), linkage analysis (1. 4. 2. 1b), or in complex diseases, affected sib-pair analysis (1. 4. 2. 1c). This is followed by a refinement of the

chromosome position by more genetic analysis and physical mapping (4. 2. 2. 2) to identify the location of a gene within a genomic clone. Using genetically linked and/or physically mapped molecular markers, genomic DNA clones are isolated by chromosome walking towards the gene. To ensure the placement of a gene in one of these clones, the candidate region is covered with a clone contig, overlapping genomic DNA clones isolated by chromosome walking or sequence tagged site (STS) content. A transcript map (1. 4. 2. 3) is then constructed within the contig. Various methods such as Zoo blotting, CpG island isolation, direct cDNA selection, cDNA screening and/or exon-trapping can be used for the identification of coding regions. In addition, genes or expressed sequence tags (ESTs) previously mapped to the region can be tested as candidate genes.

1. 4. 2. 1a Chromosomal Abnormalities

Chromosomal abnormalities, such as interstitial deletions, inversions, and translocations have cytogenetically provided a candidate region without linkage analysis and greatly facilitated identification of disease genes. These abnormalities can be instantly visualized by fluorescence *in situ* hybridization (FISH) by hybridizing fluorescently labeled YACs or other genomic clones to metaphase chromosome spreads from a patient and/or detected by hybridizing radioactively labeled genomic probes to Southern blots containing patient's genomic DNA digested with restriction endonuclease and subjected to gel electrophoresis. The breakpoint in such chromosomal abnormalities can be identified by patient-specific restriction fragments (junction fragments) in gel size-fractionation. Many successful disease gene

identifications by positional cloning have relied upon the availability of such visible cytogenetic rearrangements (Collins, 1992). Some examples have already been introduced at the beginning of this section (page 2).

Duchenne Muscular Dystrophy [DMD: MIM 310200] is one of the most successful examples of gene identification using interstitial deletions in a patient. DMD is the largest known gene of about 2.3 Mb, containing 79 exons. One patient with DMD and three other X-linked disorders had a cytogenetically visible Xp21 deletion. The clones corresponding to the sequences deleted in the patient were isolated by a subtraction cloning technique. This technique uses two DNA populations, a test and driver DNA. The test DNA from a normal individual was digested with *Mbo*I and the driver DNA prepared from the DMD patient was randomly sonicated. Two DNA populations were mixed and allowed to denature and re-anneal. Because of the deletion, the sequences in the test DNA corresponding to the deleted sequences in the driver DNA cannot form heteroduplexes and instead form homoduplexes. These homoduplexes, having sticky ends generated with *Mbo*I digestion at both strands, are selectively ligated to the *Bam*HI cloning site of the vector since *Mbo*I and *Bam*HI are isoschizomers. These clones were enriched for sequences missing in the DMD patient (Kunkel *et al.*, 1985).

Chromosomal translocation studies have resulted in many successful cases of gene identification. Additional confirmation of the DMD gene localization came from studies of rare affected females with balanced X-autosome translocations (Ray *et al.*, 1985). In each case, the X chromosome breakpoint disrupted the DMD gene although the autosomal breakpoints are

different. Two other examples given here are for Neurofibromatosis Type I (NF1) and Menkes disease (MNK). NF1 was identified by mapping translocation breakpoints to 17q11.2 found in two NF1 patients. The DNA clones from a chromosome 17 specific DNA library was used as hybridization probes against total genomic DNA from patient and normal individuals, which had been digested with restriction endonucleases and size fractionated by pulsed field gel electrophoresis (PFGE). One of the clones detected patient-specific fragments and later it was shown that both translocation breakpoints occurred within the NF1 gene (Fountain *et al.*, 1989).

MNK is a neurodegenerative disorder caused by copper transport defects leading to death in infancy or early childhood. One patient had a balanced X;2 translocation with a breakpoint at Xq13.2-q13.3. In another patient, the long arm segment Xq13.3-q21.2 was inserted into the short arm of the X chromosome. These chromosomal rearrangements suggested that MNK mapped to Xq13.3. The position of the translocation and insertion breakpoints were identified in three YACs derived from this region by fluorescence *in situ* hybridization (FISH) to metaphase chromosome spreads from the patient (Tumer *et al.*, 1992). The YAC clones spanning the translocation breakpoint provided easy entry to physical mapping and isolation of the disease gene (Chelly *et al.*, 1993; Mercer *et al.*, 1993; Vulpe *et al.*, 1993).

Large-scale expansion of specific trinucleotide repeats can cause human disease (Fu *et al.*, 1991). This additional chromosomal abberation was first observed in the patient with a moderate degree of mental retardation (IQs

about 50), carrying an X chromosome that has a cytogenetic fragile site at the band Xq27. Warren *et al.* (1990), taking advantage of X chromosome flanking gene (hypoxanthine guanine phosphoribosyl transferase and glucose 6-phosphate dehydrogenase) for selection and screening respectively, established somatic cell hybrids from the fragile site region. The fragility often resulted in chromosome breakage and both proximal and distal parts to the breakage were retained in the separate hybrids (Warren *et al.*, 1990). These panels of hybrids defined the location of markers around the fragile site. Using physical mapping techniques (1. 4. 2. 2), more markers were generated, radioactively labeled, used as probes to isolate YAC clones, and restriction maps were constructed using PFGE. Aberrantly large fragments were found in the patient due to the aberrant methylation of a CpG island. This suggested the possibility of a methylation anomaly in the region, leading to the inability of these methylation-sensitive enzymes to cleave. The intriguing problem of the findings of length variation within fragile X families and the severity of mental retardation in proportion to increasing generations were resolved by the discovery of disease causation by CGG trinucleotide instability found in the FMR1 gene. Unstable expanding trinucleotide repeats have been important to identify several other disease genes such as Huntington's disease, Spino-cerebellar ataxia, Machado-Joseph disease, Myotonic dystrophy, and others.

1. 4. 2. 1b Genetic Linkage Analysis

Many diseases are not associated with chromosomal abnormalities. In such cases, linkage analysis provides an initial candidate region by analyzing

genetic markers which are coinherited with the disease phenotype because of physical closeness between the marker and the disease locus. A marker can be a protein or DNA sequence derived from a single chromosomal location. It must be mendelian and polymorphic to increase the chance that a randomly selected person will be heterozygous. Before 1970, available markers were blood groups and HLA tissue types, but since then DNA markers have become available.

DNA variations can occur naturally as single nucleotide substitutions or as different numbers of short repeat units, generated by unequal crossing over at homologous recombination sites or replication slippage. Some single nucleotide substitutions can be detected by digesting genomic DNA with a restriction enzyme, generating restriction fragment length polymorphisms (RFLPs). RFLPs usually have two alleles, the presence or absence of the restriction site. Although useful, a marker with two alleles limits the information in a pedigree for linkage analysis because of the higher expected levels of homozygosity. Around 1985, highly informative multi-allelic markers became available with the discovery of tandem repeats, which are often transcriptionally inactive and vary in number from person to person. These markers have been classified according to mean length of repeat, either with a mean length of 15 bp (minisatellites) or of 3 bp (microsatellites). The former is also called variable number of tandem repeats (VNTRs) because polymorphism occurs in the copy number of any type of tandem repeat. The units of VNTRs vary from 10 to 1000 (0.1 to 20 kb long) arrays of short tandem repeats (Jeffreys *et al.*, 1987). This length is unfortunately too long for their

amplification *in vitro* by the polymerase chain reaction, PCR. To accommodate PCR, microsatellite markers are used in place of minisatellites. Their much smaller arrays of tandem repeats, of di-, tri-, and tetra-nucleotides, are easily amplified and have become standard tools for linkage analysis. Unlike the VNTRs which are clustered at telomeres, microsatellites are dispersed throughout the human genome and are useful as random genetic markers. Using several markers simultaneously as haplotypes in multipoint mapping allows the identification of markers flanking the disease locus. Proximal and distal makers lie in centromeric and telomeric positions with respect to the disease locus.

The degree of linkage between markers and a disease gene is commonly measured using linkage analysis. This analysis uses recombination events between the marker and the disease locus in the meioses of a family pedigree. The closer a marker is to the disease gene, the less likely the occurrence of recombination between the two at meiosis, and the two are more likely to be inherited together. This likelihood is calculated using a logarithmic function of the odds ratio (lod score) that loci are linked (with recombination fraction, $\theta$) rather than unlinked (recombination fraction 0.5). Lod scores are calculated for values of $\theta$ between 0 and 0.5. The largest lod identifies the most likely recombination fraction between the marker and disease locus (Morton, 1955). Often one family with a disease is not large enough to provide statistical significance of linkage, but lod scores, being logarithms of likelihoods, can be added across families. A lod score greater than +3 is used as evidence of linkage, and less than -2 as evidence

against linkage.

1. 4. 2. 1c Affected Sib-Pair Analysis

Linkage analysis requires specification of the mode of inheritance, gene frequencies in the population, and the level of penetrance of each genotype. For many diseases, however, this information is difficult to accurately quantify. The problem is particularly difficult in complex diseases. Such diseases can be caused by mutations in two or more unlinked genes, giving rise to the same phenotype (locus heterogeneity), or by the combined effect of a number of genetic loci, each with a small effect (polygenic disease). These diseases often arise from multifactorial sources that are both genetic and environmental. The degree to which genetic factors contribute to a disease can be investigated by twin studies which compare concordance differences between monozygotic and dizygotic twins. When a disease is found to segregate in families, the degree of familial clustering of the disease can be examined by calculating the lambda S value, the recurrence risk for siblings of an affected proband compared with the population risk. The larger the value, the stronger the familial clustering.

Affected sib-pair analysis may identify candidate susceptibility regions for a disease by the use of the shared chromosome segment method which ignores unaffected people and looks for chromosomal segments that are shared by affected individuals, identical by descent (IBD).

In an affected sib-pair study, siblings diagnosed with the disease, are typed for markers and the markers can be tested for linkage to a disease or phenotypic trait by measuring the allele or haplotype sharing in the affected

sibs. Alleles in sibs that are identical when inherited from one parent imply linkage to a disease locus if the sharing is above the frequency predicted by random segregation.

The example of Diabetes Type I illustrates the scale and power of the sib pair method (Davies *et al.*, 1994). Initially 96 pairs of affected sibs with 290 markers spread across the genome were typed. Several markers were run in the same gel lane without producing overlapping bands, using paired microsatellite markers with different allele sizes. This multiplex PCR reaction allowed the simultaneous analysis of a great many polymorphic microsatellite markers. Fluorescently labeled markers in several colours and the use of automated gels have greatly accelerated the typing of individuals in families. Ten different chromosome regions showed significant sharing ($p<0.005$) among affected sib-pairs. Two were in the HLA region and the insulin region which agreed with previous studies of these two susceptibility loci. To confirm the other eight regions two other independent sets of 102 and 84 affected sib-pairs were typed. As a result two more confident candidate susceptibility regions were identified.

1. 4. 2. 1d Allelic Association Studies and Linkage Disequilibrium

The susceptibility regions defined by affected sib-pair analysis will usually be too large to positionally clone the disease gene. In linkage analysis, when a marker is closely linked to a disease locus, the frequency of recombinants decreases, and eventually the linkage analysis becomes uninformative in fine resolution. Analysis of linkage disequilibrium can be used to examine associations in the general population between the disease and alleles of

candidate genes or markers within the susceptibility region. When a disease mutation first arises in an individual, it will be associated with a particular set of marker alleles (haplotype) on the chromosome where the mutation occurred. For very closely linked markers, the chance of recombination is very small, so that the descendants of this original mutated chromosome will tend to have the same marker haplotype. There will be a difference in the distribution or frequency of the haplotype between normal and disease-carrying chromosomes. This linkage disequilibrium can be studied more reliably using the haplotype relative risk (HRR) test (Khoury, 1994) and the transmission disequilibrium test (TDT) (Spielman *et al.*, 1993). In previous association studies, the selection of control groups was problematic and may have been drawn from genetically distinct subpopulations. However, this is resolved in both the HRR and TDT tests using internal controls. In HRR, the 'control' is not a real individual, but consists of the two parental alleles which are not transmitted to the affected person. The TDT begins with families with one or more affected offspring and at least one parent which is heterozygous at the marker locus under examination. One of the two marker alleles of each heterozygous parent is transmitted to each affected offspring and one is not. The test compares the frequency of alleles among the transmitted and non-transmitted alleles. The significance of the association in both HRR and TDT tests is tested by a $X^2$ test.

The power of linked markers is well illustrated in many studies. Two examples are given here of for gene identification initiated from genetic mapping and affected sib-pair analysis.

Huntington's Disease [HD: MIM 143100] is an autosomal dominant neurological disorder with late onset. The penetrance is age-related. A significant linkage was found on the short arm of chromosome 4 (Gusella, 1983) with a recombination fraction of 4%. The marker was then mapped to 4p16.3 by the somatic cell hybridization technique. HD is a disease with a low rate of new mutations, retaining markers close to the disease gene. Linkage disequilibrium analysis was therefore employed to refine the genetic localization of HD. The segment around the markers where linkage disequilibrium was most pronounced was used for exon trapping. The HD gene mutation was identified to result from an expanded trinucleotide (CAG) repeat.

No chromosomal rearrangements were found in cystic fibrosis (CF: MIM 219700) patients. Studies of affected sib-pairs showed that the gene for CF was linked to a protein polymorphism of the paraoxonase enzyme (PON) (Eiberg and Mohr, 1985). The chromosomal location of PON was not known at that time, but rapidly mapped to chromosome 7 using linkage analysis (Eiberg and Mohr, 1985). Many more markers were used to improve the genetic analysis of CF families. Four markers were found to be in linkage disequilibrium with CF and at this stage the positional cloning strategy was employed to close the gap between the closest linked markers, and the CF gene was isolated (Knowlton and Cohen-Haguenauer, 1985; Tsui and Buchwald, 1985; Wainwright and Scambler, 1985; White and Woodward, 1985).

1. 4. 2. 2 Physical Mapping

To define a candidate region suitable for positional cloning, linkage analysis

and affected sib-pair study must map a disease locus in a region defined between proximal and distal markers of about 2 cM or 2 x $10^6$ base pairs (2 megabases) apart. If this distance is not achieved with existing markers, more markers are generated by screening microsatellite markers from yeast artificial chromosomes (YACs) isolated from the candidate region and the process is repeated. YACs are vital for positional cloning and described in the following.

Within a critical region determined through linkage analysis, coding regions should be identified in clones of genomic DNA inserts if these clones overlap to cover the whole region between the proximal and distal markers. A primary task of physical mapping is to assemble such a series of DNA clones with overlapping inserts (clone contig). Critical regions, however, of about 2 Mb in size are very large and covering the whole region is laborious if insert DNA sizes are too small. Yeast artificial chromosomes (YACs) have greatly facilitated positional cloning because they allow the cloning of human DNA fragments up to 2 Mb in length (Chumakov *et al.*, 1992; Larin *et al.*, 1991). Relatively few YAC clones are needed to cover the critical region. The YAC's capacity to clone such large fragments provides a transition from 'genetic linkage mapping' to 'physical mapping'.

A YAC vector is only about 10 kb in size. It contains a yeast centromere (CEN), two telomeres (TEL), and a yeast origin of replication (ARS). These three components are required in all eukaryotes for linear chromosomal replication and segregation. CEN is required for disjunction of sister chromatids in mitosis and of homologous chromosomes at the first meiotic

division, TEL for complete replication of linear molecules and for protection of the ends of chromosomes from nuclease attack, and ARS for replication origins (Murray and Szostak, 1983). A YAC vector becomes stable with a suitably sized foreign DNA fragment inserted to give a linear DNA molecule in which the TEL sequences are correctly positioned at the termini. Recombinants can be selected by vector complementation of defective uracil and tryptophan genes in the yeast host, AB1380. Insertion of the foreign DNA is indicated by the disruption of an ochre tRNA suppressor gene at the cloning site giving rise to the red *ade*2 phenotype in recombinants (Burke *et al.*, 1987).

Since the emergence of YACs, other cloning vectors have become available, such as bacteriophage, P1, bacterial artificial chromosomes, BACs , and P1-derived artificial chromosomes, PACs, whose insert sizes range 75-300 kb. The insert size to be cloned in these systems is much smaller than that of YAC vectors, but their higher cloning efficiencies, lower chimeric and rearranged risks, and simpler insert purification have made them attractive alternatives to YACs (Monaco and Larin, 1994).

Methods of constructing a clone contig and identifying coding DNA regions in YACs also apply to P1s, BACs, and PACs. To construct a clone contig, YAC clones are isolated by hybridizing labeled proximal and distal markers as probes to a YAC library or by PCR screening of DNA pools. These YAC clones become fixed starting points for assembling the clone contig by bidirectional chromosome walking from both ends. At each step the end of the insert DNA from these YACs are purified, labeled, and used as 'end-

probes' to screen the YAC library. The positive YAC clones should contain partly overlapping sequences. These positive YAC clones can then be used to generate new, distal end-probes for further rounds of screening of the YAC library (Riley *et al.*, 1990).

End-probes can be readily prepared by PCR amplification of uncharacterized sequences adjacent to the known YAC vector sequence using a bubble linker (vectorette-PCR) (Riley *et al.*, 1990). This linker is a double-stranded oligonucleotide with an imperfect complementation in the middle that forms a bubble. The YAC clones are digested with a frequently cutting restriction enzyme and mixed with the bubble linkers for ligation. The linker ensures that PCR amplification occurs only from the YAC vector-specific primer and not the bubble linker primer. The bubble linker primer cannot prime DNA synthesis initially because complementary strands are required for annealing. However, the first strand synthesized by the vector-specific primer contains sequences complementary to the bubble linker. The second round of PCR amplifies uncharacterized sequence flanking the vector sequence from the bubble linker primer. The PCR product can be excised, purified, and used as an end-probe for chromosome walking, or can be sequenced to generate an STS for PCR screening of the library.

1. 4. 2. 3 Transcript Mapping

There are a variety of methods of identifying coding DNA regions in genomic clones. Sequences of regulatory and coding DNA are strongly conserved in evolution unlike that of non-coding DNA because altering the gene sequence (mutation) will often be disadvantageous for cellular function and are

eliminated by natural selection. This conservation can be seen by hybridizing small genomic fragments against a Southern blot of genomic DNA samples from various animal species ('Zoo blots') (Monaco *et al.*, 1986). The hybridization requires special conditions to suppress hybridization signals from DNA sequences in the probe which are highly repetitive in the human genome.

Another method of identify coding DNA regions is to construct a restriction enzyme map and identify CpG (cytidine connected covalently by a 3'-5' phosphodiester bond to a guanosine) islands in genomic clones. This dinucleotide enrichment often stretches 1-2 kb in length and 56% of these islands in vertebrates are associated with the 5' end of genes (Cross and Bird, 1995). The rest of the DNA in the vertebrate genome is markedly deficient in CpG island (Antequera and Bird, 1993). These islands can be identified in a restriction enzyme map.

The restriction map is constructed by a series of restriction digestions of YACs or other genomic clones in the contig with 'rare-cutting' restriction enzymes (enzymes with recognition sites 8 bp long, such as, *Sac*II, *Eag*I, *Bss*HII, or *Not*I) (Dracopoli, 1995). The number and sizes of the DNA fragments, produced by each restriction endonuclease cleavage, must be determined by gel electrophoresis, followed by comparison with size markers. This information must be supplemented by a series of double digestions, in which the YAC DNA is digested by two restriction endonucleases at once. Comparing the results of single and double digests will allow many of the restriction sites to be mapped. When it is ambiguous how to position some

restriction sites in a map, partial digestion can be carried out under conditions that will result in cleavage of only a limited number of the restriction sites on YAC DNA molecule. This can be achieved either by a short incubation period so that the enzyme does not have time to cut all the restriction sites, lower concentrations of enzyme, or by incubation at a lower temperature of 4°C which will limit the activity of the enzyme. The restriction map can be also useful to refine a YAC contig by identifying overlapping YAC DNA clones by aligning the restriction sites among YACs. The restriction endonucleases, *Sac*II, *Eag*I, *Bss*HII, or *Not*I, recognize a CG sequence. The cluster of these restriction sites therefore indicate the association of the CpG island and possible location of a gene (Cross and Bird, 1995).

Direct cDNA selection is an alternative method to identify coding DNA regions. The cDNA is obtained from heteroduplexes between genomic inserts and cDNA derived from mRNA via reverse transcriptase (Lovett *et al.*, 1991; Parimoo *et al.*, 1991). Genomic clones are digested with frequently cutting restriction nucleases (enzyme with recognition sites 4 bp long) and ligated to labeled linkers with biotin. The fragments are amplified by PCR using biotin-labeled linker-specific primers and hybridized with the cDNA inserts. Prior to the hybridization, repeat sequences in genomic DNA must be suppressed by prehybridizing with human placental DNA. Hybridization products with a biotin group will bind to streptavidin-coated paramagnetic beads. All biotin-labeled fragments from the genomic inserts are eliminated by a magnet and the cDNA is eluted and amplified using PCR primers specific for the vector sequence flanking the cDNA. The cDNA obtained is enriched

by submitting the PCR products to further hybridization cycles.

YACs containing possible coding DNA regions identified by these methods can be used as probes to screen cosmid libraries, or they can be subcloned into cosmids to construct a cosmid library. Placing a disease gene or coding DNA regions in cosmids with 45 kb in length is relevant prior to the ultimate goal to subcloning into plasmids to clone both introns and exons, which cDNA clones do not have. Genomic clones are useful for the analysis of splice-site mutations.

Construction of a cosmid contig covering the whole YACs of interest will ensure identification of genes or coding DNA regions within the contig. The cosmids isolated by screening cosmid libraries or by cosmid libraries constructed from the YACs can be ordered by cosmid walking. The vectorette-PCR method used to isolate end-probes in YACs is optional, because entire cosmid inserts can be used as probes for hybridization to a Southern blots containing cosmid DNA fragments. These fragments were derived from a single restriction enzyme, for instance, *Hind*III. This whole cosmid-cosmid hybridization is carried out under conditions which suppress hybridization signals from DNA sequences in the probe which are highly repetitive in the human genome. Cosmids can be classified in a cluster. To refine the cosmid order in a cosmid contig, end-probes can be isolated simply by digestion and ligation. Cosmids are digested with a frequent cutting restriction nuclease and ligated in the same tube. Only the ends of cosmid inserts attached to a vector can be transformed and amplified successfully in host bacteria. These end-products can be used as hybridization probes for cosmid walking to

construct a cosmid contig covering the whole YAC of interest.

Coding regions in these cosmids can be identified by searching for evolutionary conserved DNA sequences, CpG islands, and/or using cDNA selection described above. These methods, however, have their limitations. Like all expression-based systems, the method depends on appropriate levels of gene expression and genes containing very short exons may be missed because of unstable heteroduplex formation. cDNAs may bind to pseudogenes, or some genes may not be associated with a CpG island.

A cosmid is about one tenth of a YAC in size. It can be more easily handled and there are more methods for identifying coding DNA region. The cosmids of interest hybridizes to mRNA from the tissues where the mutant phenotype is expressed. These cosmids can be used as probes to hybridize to cDNA libraries constructed from the same tissue from which the mutant phenotype is expressed.

Alternatively, coding DNA regions from these cosmids can be identified using the 'exon trapping' technique (Buckler *et al.*, 1991). This method is based on the observation that mammalian genes usually contain multiple exons which need to be spliced together at the RNA level. Exons in a cosmid can be trapped using a cellular RNA splicing assay which detects sequences that are capable of splicing to exons within a specially designed vector.

Cosmids of interest can be subcloned into plasmids. The coding sequences can be detected by hybridizing to northern blots of RNA isolated from fetal and adult human tissues. Inserts in plasmid clones can be

sequenced using plasmid vector-specific primers. The likelihood that the sequence represents part of a gene is analyzed by computer software which performs tasks such as 'homology searching' and 'exon-prediction' (1. 4. 4). The former compares the insert sequence against all available sequences in sequence databases at nucleotide, amino acid, and protein level. 'Exon-prediction' scans the insert sequence to identify the locations of possible exons by screening for conserved sequences found at exon-intron boundaries and by examining presence or absence of comparatively long open reading frames without termination codons.

### 1. 4. 3 Positional Candidate Cloning

This strategy identifies the gene responsible for a disease by mapping the mutant gene to a specific chromosomal region and then search for an appropriate candidate among the genes already localized in the region. The candidate gene can be confirmed by tests for mutation in the individuals suffering from the disease. Positional candidate methods have dominated disease gene identification in recent years because of the rate of the EST locations. The following is an example of a positional candidate cloning strategy.

A gene encoding a fibroblast growth factor receptor (FGFR3) was discovered during a chromosome walk toward the Huntington disease gene (1. 4. 2. 1d). Later, a gene for achondroplasia (ACH), a hereditary dwarfism due to retarded growth of the long bones, was also mapped to the same chromosomal region. Missense mutations in FGFR3 were then found in the ACH patients (Shiang *et al.*, 1994). This confirmed the FGFR3 gene was

responsible for the growth retardation characterising the ACH disease.

### 1. 4. 4 Computer Analysis of DNA Sequence

All cloning strategies have been facilitated by developments in computing and informatics. Nucleic acid and protein sequences are stored in the databases shown in Table 1. 1. These can be accessed through the University of Wisconsin Genetics Computer Group (GCG) sequence analysis software package either via the server at the Pathology department, University of Oxford or the world-wide web (WWW) server at the National Center for

| Database | Description | Electronic Addresses |
|---|---|---|
| GenBank | DNA and protein sequences | http://www.ncbi.nlm.nih.gov |
| EMBL | DNA sequences | http://www.ebi.ac.uk |
| DDBJ | DNA and protein sequences | http://www.nig.oc.jp |
| PIR | Protein sequences | http://www.ncbi.nlm.nih.gov |
| SWISS-PROT | Protein sequences | http://www.ncbi.nlm.nih.gov |

Table 1. 1. Major nucleic acid and protein sequence databases (http://www.ncbi.nlm.nih.gov/baxevani/csh/table 1,html)

Biotechnology Information (NCBI) server, using WWW browser such as NETSCAPE NAVIGATOR. GenBank listed a total of about 500 Mb of nucleotides sequence (1997), which is about 35% of the total sequences in the human genome.

Sequences obtained by an ABI 373A DNA Sequencer (Applied Biosystems) can be edited using the Factura and Sequencing Navigator software (Applied Biosystems). The nucleotide sequence and the inferred amino acid sequences for all six reading frames (including reverse sequences)

are submitted to computer software which determines whether the sequence is likely to be derived from part of a gene. A comparison is performed against all available sequences of a known gene, cDNA, protein, or gene-associated sequence of human or non-human origin. Commonly used homology searching software are summarized in Table 1. 2. These can be accessed through either GCG, the WWW, or via electronic mail.

The probability that a sequence contains an exon is performed by GRAIL service, accessed from the Oak Ridge National Laboratory (ORNL) server by electronic mail. This program will identify the locations of likely exons by screening for conserved sequences found at exon/intron junctions and the existence of a splice branch

| Program | Compares |
|---|---|
| FASTA | A nucleotide sequence against a nucleotide sequence database or an amino acid sequence against a protein sequence database |
| TFASTA | An amino acid sequence against nucleotide sequence database |
| BLASTN | A nucleotide sequence against a nucleotide sequence database |
| BLASTX | A translation of nucleotide sequence into amino acid sequence against a protein sequence database |
| TBLASTN | A protein sequence against a nucleotide sequence database translated in all reading frames |

Table 1. 2. The programs for sequence comparisons from http://www.ncbi.nlm.nih.gov/baxevani/csh/table 1,html.

site (Uberbacher and Mural, 1991). In this way candidate genes can be identified from high homology scores or hypothesized through lower homology scores with genes from other regions of the genome or other species.

Other major databases relevant to gene cloning strategies are listed in Table 1. 3. The Genome Database (GDB) contain genetic disease information in text form with interacting manager programs including locus, polymorphism, probe and citation. The on-line Mendelian Inheritance in Man (OMIM) is an electronic catalog of McKusick's Mendelian Inheritance in Man.

| Database | Description | Electronic Access |
|---|---|---|
| GDB | Human mapping database | http://gdbwww.gdb.org |
| dbEST | Collection of partial sequences from cDNA clones | http://www.ncbi.nlm.nih.gov/dbEST/index.html |
| OMIM | Catalog of V. McKusick's Mendelian Inheritance in Man | http://www3.ncbi.nlm.nih.gov/omim |
| Généthon | Human genetic maps based on linked $(CA)_n$ repeat markers | http://www.genethon.fr |
| CEPH | YAC-based physical maps of the human genome | http://www.cephb.fr |
| Whitehead Institute | YAC-based physical maps of the human genome | http://www-genome.wi.mit.edu |
| HGMP-RC (Human Gene Mapping Project-Resource Centre) | Centre to provide biological materials, on line access to databases, and the tools to analyze them | http://www.hgmp.mrc.ac.uk |

Table 1. 3. Selected molecular biology email servers (http://www.ncbi.nlm.nih.gov/baxevani/csh/table 1,html)

## 1. 5 AIM OF THESIS

Patient ML manifests a variety of behavioural and physical abnormalities. She is impaired in communication and reciprocal social interaction and her interests and gesticulations are limited to repetitive patterns. She is mentally retarded and epileptic. This constitutes a wide range of symptoms, but all are common to the diagnostic criteria for autism. Her short stature and bone

protuberances around wrists, knees, ankles, and left clavicle were diagnosed as multiple exostoses.

There is currently no specific knowledge of the biochemical dysfunctions underlying autism and multiple exostoses. They are considered to be genetically complex or heterogeneous disorders, which has deterred the identification of genes, although progress has been made recently in multiple exostoses. The translocation identified in patient ML provides an opportunity to directly investigate the genetics of these two diseases, assuming that the translocation contributes directly to her phenotype. In this thesis, molecular genetic techniques were used to determine which genes have been disrupted by the translocation. Knowledge of these genes and their biological roles will be used to evaluate them as candidate genes for autism and multiple exostoses.

The positional and positional-candidate cloning strategies were the main approaches used to identify the genes, by resolving, one by one, the following specific aims:

1. confirm the activation of the translocated X chromosome.
2. identify YACs, cosmids, and plasmids spanning the translocation breakpoint.
3. isolate and clone the translocation junction fragments.
5. identify genes disrupted by the translocation breakpoint.
6. characterize the genes with respect to ML's phenotype.
7. determine the orientation of the genes and the possible formation of a fusion gene.
8. investigate expression of the disrupted gene on the active or inactive X chromosome.

# CHAPTER II

## MATERIALS AND METHODS

The approach taken was to identify ever smaller DNA fragments containing the translocation breakpoint, until the location of the breakpoint was identified. The general strategy was to construct probes from genomic DNA which can be radioactively or fluorescently detected and to generate membrane filters containing a collection of the cloned DNA fragments. The contigs of overlapping genomic clones and fragments of interest close to the translocation breakpoint were selected by hybridization forming heteroduplexes between the detectable probes and the target-fragments.

This Chapter is a description of (a) the preparation of DNA probes amplified by either, recombinant DNA cloning (2. 1 and 2. 3) or cell-free DNA amplification using the polymerase chain reaction (PCR) (2. 2), (b) the measurement of DNA concentration and cell counts (2. 4), (c) the transfer of nucleic acids to blots (2. 5) and membrane filters containing collections of DNA fragments in YACs (2. 6), cosmids, phages and plasmids, and DNA labeling (2. 7 and 2. 8). Hybridization (2. 9) was detected and confirmed using either autoradiographic or fluorescence illuminating methods. Exon-Amplification (2. 10) and Exon-Prediction using computer analysis (2. 11) were employed to search for expressed DNA sequences. PCR library screening, polymorphism analysis, X-inactivation panel, karyotype study, and fluorescent *in situ* hybridization (FISH) is separately described in 2. 12, 2. 13, 2.

14, 2. 15, and 2. 16 respectively.

## 2. 1 PREPARATION OF DNA BY RECOMBINANT CLONING

Insert DNA integrated in vector DNA (plasmid, cosmid) was amplified in host-bacterial cells and purified using an alkaline lysis method. YAC DNA in yeast cells was prepared in solution using either guanidium-hydrochloride extraction or glass-bead methods. The lithium method (Anand and Southern, 1990; Anand *et al.*, 1989) was used in the preparation of YAC DNA imbedded in agarose blocks.

### 2. 1. 1 Purification of Plasmid and Cosmid DNA

The preparation of plasmid and cosmid DNA from a culture of bacterial cells required four steps.

(1) Transformed bacteria were propagated in 15 ml of Luria broth (LB) aerated by shaking at 250 rpm on a rotary platform overnight. On the following day, 1.5 ml of culture was mixed with 160 µl of 10 x HMFM and stored at -70°C for future use as a permanent stock. The rest of the culture was harvested by centrifugation at 3,000 rpm for 10 min.

(2) DNA was purified from the cell contents using the three-stage-alkaline lysis method (Sambrook *et al.*, 1989). The bacterial pellet was resuspended in 300 µl of ALI solution (50 mM Tris-HCl, pH 7.5, 10 mM EDTA). This solution removed magnesium ions and destabilized the overall structure of the cell envelope. A final concentration of 100 µg/ml RNase A was added to the ALI solution to eliminate RNA. 300 µl of ALII solution (0.2 M NaOH, 1% (w/v) SDS) was then added to break the lipid molecules of the cell membranes which prompted cell lysis and denatured DNA and protein.

(3) The lysed cell components were insoluble in 300 µl of ALIII solution (3 M sodium acetate, pH 4.8) and were pelleted by centrifugation at 13,000 rpm for 30 min while the circular and supercoiled plasmid DNA remained in solution.

(4) DNA in the aqueous phase was transferred to a fresh tube and precipitated with 2 volumes of absolute ethanol in the presence of monovalent cations of NaCl (at a final concentration of 0.1 M NaCl). Excess salt was removed using a 70% (v/v) ethanol wash. The resulting DNA was resuspended in a suitable volume of 1 x TE (10 mM Tris-HCl, 1 mM EDTA, pH 7.5).

On occasion, prior to the ethanol precipitation, phenol-chloroform extraction was required for a further purification from protein contamination. The pellet at step (3) was resuspended with 500 µl of 1 x TE and twice extracted with an equal volume of 1:1 (v/v) phenol/chloroform and then once with an equal volume of chloroform. Plasmid DNA used in sequencing was extracted using DNA purification columns from Qiagen Inc.

| | |
|---|---|
| Luria broth (LB)<br>(1 litre) | 10.0 g bacto tryptone<br>5.0 g Bacto yeast extract<br>10.0 g NaCl<br>Adjust pH to 7.5 with NaOH<br>For plate, add 15 g bacto agar before autoclaving. |
| 10 x HMFM<br>(1 litre) | 0.9 g $MgSO_4.7H_2O$<br>4.5 g $Na_3$citrate<br>9.0 g $(NH_4)_2SO_4$<br>18.0 g $KH_2PO_4$<br>63.0 g $K_2HPO_4$ (anhydrous)<br>440.0 g glycerol<br>Make up to 1 litre volume with $dH_2O$, autoclave |
| ALI solution | 100 µg/ml RNase A<br>50 mM glucose<br>50 mM Tris-HCl, pH 8.0<br>10 mM EDTA |
| ALII solution | 200 mM NaOH<br>1% (w/v) SDS |
| ALIII solution | 3.0 M $NaOAc.3H_{(2)}0$, pH 4.8 or<br>3.0 M Kac, pH 5.5 |
| TE | 10 mM Tris-HCl, pH 8.0<br>1 mM EDTA, pH 8.0 |

### 2. 1. 2 Purification of Bacteriophage DNA

An agar plug containing a single plaque of bacteriophage was picked using a

fine glass pipette into 100 µl SM solution, which was left for 2 h to allow the phage to diffuse out of the agar plug. This diffusate (approximately 3 x $10^6$ bacteriophages) was incubated with the appropriate strain of bacteria at 37°C for 15 min and then inoculated in 25 ml of NZCYM culture medium. The mix was incubated overnight at 37°C with vigorous shaking to promote bacteriophage infection and bacterial cell lysis. On the following day, a drop of chloroform and 1.4 g NaCl were added to dissociate the bacteriophage particles from the bacterial debris. Another 15-min-incubation promoted further lysis. The bacterial debris was pelleted by centrifugation at 3,500 rpm for 30 min. The supernatant was incubated for 45 min with 1 µg/ml DNase and 1 µg/ml RNase A to digest bacterial DNA and RNA respectively. Solid polyethylene glycol (PEG 6000) to reach a final concentration of 7% was added to precipitate bacteriophage particles in the cold (4°C) for several hours to overnight. After centrifugation, the pellet was resuspended in the phage diluent and treated with 0.5% (w/v) SDS, 0.25 mg/ml Proteinase K (PK), and 20 mM EDTA, which degraded the bacteriophage particles, releasing DNA. The bacteriophage DNA was obtained after phenol-chloroform extraction and ethanol precipitation, and was analyzed by electrophoresis in 1% (w/v) agarose (SeaKem, F. M. C.) gel (2. 3. 3).

| | |
|---|---|
| NZCYM medium<br>(1 litre) | 10.0 g NZ amine<br>5. 0 g NaCl<br>5.0 g Bactoyeast extracts<br>1.0 g Casamino acids<br>2.0 g $MgSO_4.7H_2O$ and then 950 ml $dH_2O$<br>Adjust pH to 7.0 with 200 µl 5 M NaOH |
| Phage Diluent | 100 mM NaCl<br>10 mM Tris-HCl, pH 8.0<br>100 mM $MgCl_2$ |

### 2. 1. 3 Preparation of YAC DNA

YAC DNA in yeast cells (AB1380 strain of *Saccharomyces cerevisiae* (Burke *et al.*, 1987)) was prepared in two forms (1) in solution for use with Fluorescence *in situ* Hybridization (FISH) and PCR experiments, and (2) in agarose blocks for use with FISH and Pulsed Field Gel Electrophoresis (PFGE) to isolate YAC DNA. Yeast clones were streaked onto the selective media, -U, -T, or AHC plates containing 50 μg/ml ampicillin and grown for three days at 30°C. Colonies were inoculated individually in 15 ml of the -U, -T or AHC media for two days. Glycerol stocks were made from pellets formed from 1.5 ml of culture with 1.5 ml of YPD containing 20% glycerol and stored at -70°C.

| | |
|---|---|
| -U, -T solution | 2% glucose<br>0.67% yeast nitrogen base without amino acids<br>1 x amino acids (Difco) |
| 10 x amino acids<br>(1 litre) | 300 mg of tyrosine dissolved in 500 ml $dH_2O$ and mixed with 30 ml each of adenine, arginine, histidine and methionine (7.5 mg/ml)<br>40 ml of 7.5 mg/ml lysine<br>70 ml of 7.5 mg/ml phenylalanine<br>80 ml of 7.5 mg/ml leucine<br>80 ml of 7.5 mg/ml valine<br>40 ml of 7.5 mg/ml isoleucine<br>Make up to volume to 1 l $dH_2O$ & autoclave |
| AHC<br>(1 litre) | 500 ml $dH_2O$<br>20 g dextrose (2%)<br>10 g casein acid hydrolysate<br>10 ml adenine hemisulphate (3 mg/ml)<br>Add $dH_2O$ up to 900 ml, and autoclave<br>Add 100 ml 10 x yeast nitrogen base (Difco)<br>Add ampicillin (50 μg/ml)<br>For plates 20 g bactoagar was added. |
| YPD<br>(1 litre) | 10 g yeast extract<br>20 g Peptone<br>Add 1 l $dH_2O$, 17 g of bacto-agar & autoclave<br>Add 50 ml of filter-sterilized 40% glucose solution |

#### 2. 1. 3a Preparation of YAC DNA in solution

YAC DNA in solution was prepared in two ways. In the first, yeast cells from

50 ml of culture were pelleted at 3,000 rpm for 10 min and resuspended in 5 ml of 0.9 M sorbitol, 20 mM EDTA, 14 mM 2-mercaptoethanol (2-ME). The mixture was incubated at 37°C for 1 h to spheroplast by disrupting the yeast cell walls with a final concentration of 4 mg/ml of Novozyme (Novo Biolabs) or 2 U/ml Zymolyase (ICN Biomedicals Inc.) and precipitated at 1,000 rpm for 10 min. The pellet was resuspended and incubated in 5 ml Guanidium-hydrochloride solution at 65°C for 10 min to promote lysis, which was subsequently cooled to room temperature and precipitated with an equal volume of ethanol. The resuspended pellet in 2 ml of 1 x TE was incubated at 37°C for 1 h with a final concentration of 100 µg/ml RNase A to remove RNA. Proteins were digested by incubation at 65°C for 1 h with a final concentration of 200 µg/ml PK. RNA and protein were then removed by a phenol-chloroform extraction and resuspended in 250 µl 1 x TE. In general the yield was 100 µg per 50 ml culture.

| | |
|---|---|
| Guanidium-hydrochloride solution | 4.5 M GuHCl<br>0.1 M EDTA<br>0.15 M NaCl<br>0.05% Sarkosyl, pH 8.0 |

The second method of preparing YAC DNA in solution used 180 micron-glass beads (Sigma). Yeast cells from 50 ml of culture were pelleted as described in 2. 1. 3a, washed with 500 µl $dH_2O$, pelleted again, resuspended in 200 µl GDIS (Tris-HCl, pH 7.4, 1 mM EDTA), and added to 0.35 g of acid washed glass beads and 200 µl of phenol. This mixture was vortexed continuously for 5 min. The suspension, after spinning for 4 min at 13,000 rpm in a microcentrifuge, was transferred to a fresh Eppendorf tube, incubated at 37°C with 50 µg RNase A (Sigma, DNase free), extracted once more with phenol, then once with

phenol:chloroform:isoamyl alcohol (25:24:1), and precipitated with 0.1 x sodium acetate and 2 x absolute ethanol. The pellet was rinsed with 70% (v/v) ethanol, air-dried, and resuspended in 20 µl $dH_2O$. The average yield from a 50 ml culture was 200 µg.

2. 1. 3b Preparation of yeast and YAC DNA blocks

Intact chromosomal YAC DNA was generated by the Lithium method (Anand *et al.*, 1990) in which whole yeast cells are embedded in agarose blocks before spheroplasting, lysis and deproteinization. 2% (w/v) low melting gel (SeaPlaque, F.M.C.) in 1 M sorbitol and 20 mM EDTA was dissolved by boiling, and cooled to 50°C in a water bath, and 14 mM of 2-ME was added. Block moulds (Pharmacia) were laid out on a glass plate placed on ice. Yeast cells from 50 ml culture were pelleted at 1,000 rpm for 10 min. The pellet was resuspended in 400 µl of Solution I (1 M sorbitol, 20 mM EDTA, 14 mM 2-ME) and mixed with 500 µl of the molten agarose and dispensed into the slots of the block mould on ice. The set agarose inserts were released in a 50 ml Falcon tube containing 5 ml of 1 M sorbitol, 20 mM EDTA, 14 mM 2-ME, 10 mM Tris-HCl, pH 7.5, 2 U/ml Zymolyase and incubated at 37°C for 2 h. The agarose inserts were incubated in the lysis solution (1% lithium dodecyl sulphate, 100 mM EDTA, 10 mM Tris-HCl, pH 8.0) at 37°C for 1 h and stored in $TE_{50}$ at 4°C until loading on a gel for sizing and purifying YACs (2. 3. 1).

## 2. 2 PREPARATION OF DNA BY PCR AMPLIFICATION

When the DNA fragments of interest were less than 10 kb in length and flanked by known oligonucleotide primer sequences (18-25 nucleotides long)

on both sides, they were amplified *in vitro* from their DNA templates using the polymerase chain (PCR) reaction. This cell-free nucleic acid amplification was achieved in 2-4 h by multiple cycles of DNA polymerization, each followed by a heat denaturation, annealing with primer, and extension with thermo-stable DNA polymerase.

Standard PCR and Long PCR were used to amplify fragments of about 0.1-5.0 kb and 5.0-10.0 kb in size, respectively. The templates used were genomic DNA, cDNA, YAC and cosmid colonies. Purified PCR products were used directly as hybridization probes or as templates for sequencing. Exon-amplified products were cloned in plasmids for long-term storage (2. 9).

### 2. 2. 1 PCR Templates

PCR amplification of DNA fragments was performed in a total PCR reaction buffer of 50 µl, using as template approximately 100 ng of genomic DNA or 80 ng of cDNA. When YAC or cosmid DNA was the required template, a colony of yeast or bacteria from the agar plate was picked and mixed in 200 µl $dH_2O$, boiled for 10 min, and spun for 15 min. 3 µl of the supernatant was directly mixed with the PCR reaction buffer and used as a template.

### 2. 2. 2 Designing PCR Primers

The specificity and efficiency of the PCR amplification was maximized by designing primers according to the following criteria:

(1) The base composition of a primer is about 50% GC and AT.
(2) A G or C is placed at a 5′ and 3′ ends of the primers for stability.
(3) The length of the primer is between 18-22 bp.
(4) The annealing/melting temperature (Tm) is between 56-60°C, calculated using the formula:
Tm=[(A +T) x 2°C] + [(G + C) x 4°C] (Thein and Hinton, 1993).
(5) Internal palindromes are avoided.

(6) Four or more consecutive bases are avoided.

### 2. 2. 3 Primer Preparation

All primers/oligonucleotides were synthsized at Clare Hall, ICRF. They were resuspended in 200μl of 0.3 M NaOAc, 10 mM $MgCl_2$, and precipitated in 3 volumes of absolute ethanol. The pellet was resuspended in 100 μl of 1 x TE. The concentration was calculated using a spectrophotometer (2. 4. 2). Two final concentrations of 500 ng/μl and 50 ng/μl were prepared for each primer for use in PCR and sequencing, respectively.

### 2. 2. 4 Amplification Condition for Standard PCR

PCR amplifications were performed using 50-100 ng of template DNA and 100 ng of each primer in a 25 μl or 100 μl reaction buffer (10 mM Tris-HCl, pH 7.5, 50 mM KCl and 1 mM $MgCl_2$) containing 0.1 mM of each dCTP, dGTP, dATP and dTTP with 2 U of Taq DNA polymerase (Boehringer-Mannheim). Depending on the primer sets the concentration of $MgCl_2$ was adjusted (Table 2. 1) and optimized for a high yield of product. A negative control with no template and a positive control of known size was performed in each PCR experiment. Amplification was carried out in a Perkin Elmer Cetus Thermal Cycler. The standard PCR cycle conditions were 25 cycles of denaturation (30 sec at 94°C), annealing (1 min at a temperature shown in Table 2. 1) and extension (1.5 min at 72°C) preceded by 3 min of denaturation at 94°C and followed by a final extension at 72°C for 10 min. The PCR products were analyzed on 2% (w/v) agarose gels with a ϕx174RF/*Hae*III marker (2. 3. 2). The desired bands of the PCR products were purified using either the excision technique from the gel described in Section 2. 3. 5. or the QIAquick PCR

| Locus | Name | Primer Sequences | Anneal Temp. | $MgCl_2$ Conc. |
|---|---|---|---|---|
| Xp * | 7at3out-L | CACTTCTTTCTTCATGCTCTG | 50.3°C | 1.5 mM |
| Xp * | 7at3in-R | GACAGAAAACGTAATAAAGATGTG | 50.4°C | 1.5 mM |
| Xp * | 7at7out-R | GTGGCATTACATCTGATGCAG | 52.3°C | 1.5 mM |
| Xp * | 7at3IN-L | CTAGAATTGCAATATATAATAACTG | 50.3°C | 1.5 mM |
| Xp * | 7Bt3out-R | GCCATCATTTTTTCCTCACTGTAC | 53.9°C | 1.5 mM |
| Xp * | 9-2t7in-R | GCATGTAGCTGAAAAGGAGC | 51.7°C | 1.5 mM |
| Xp * | 7bLEFT | GATAACCAAAGGGGAAAACAC | 50.3°C | 1.5 mM |
| Xp * | 385bbLEFT | CATATAAATAGAATAACTACAG | 43.5°C | 1.5 mM |
| Xp * | 386bbRIGHT | CAGAACTCACTTCTCTCACTC | 52.3°C | 1.5 mM |
| Xp * | TRANS-RR | GGTAAGCCTAATATGTAGAAC | 48.4°C | 1.5 mM |
| Xp * | TRANS-LT | GAAAGGGTAAGCCTCAGTG | 51.7°C | 1.5 mM |
| Xp * | NXT7WK | GAGGTGGTGAGAAGCATCTG | 53.7°C | 1.5 mM |
| Xp * | CH8-LT | CTATTGACTCTGAACTCAGC | 49.6°C | 1.5 mM |
| Xp * | NXT731WK | CCTCATTTCACCATCTTAAAAC | 49.1°C | 1.5 mM |
| Xp * | 638TOE3 | GTGAACTCCTCTATTGCCC | 51.0°C | 1.5 mM |
| Xp * | 639ib1-con | CCTAAGCTAGGGAGGTGTTG | 53.7°C | 1.5 mM |
| Xp * | 639ib1-con | CCTAAGCTAGGGAGGTGTTG | 53.7°C | 2.25 mM |
| Xp * | 641NXT3out | CCAACCTAGAATTCTATGCC | 49.6°C | 2.25 mM |
| Xp * | 642NXT3out | CCAACCTAGAATTCTATGCC | 49.6°C | 2.25 mM |
| Xp * | 643i2mmout | CTGAAGGATGCCTTCTCTTC | 51.7°C | 2.25 mM |
| Xp * | 950NXT3RR | GTGGATCTTCCTTCAGTATTTC | 51.0°C | 2.25 mM |
| Xp * | 234TOE1 | CAATATGTTACACTCAAGAAGTC | 49.8°C | 1.5 mM |
| Xp * | E1-1a | GAAAGACGCTGTGGGAAAATAG | 52.9°C | 1.5 mM |
| Xp * | E1-1b | CACCTAAAAGTTCCCTTAGATG | 51.0°C | 1.5 mM |
| Xp * | E1-2a | CATAGATCTTATCTTCATCTTCAC | 50.4°C | 1.5 mM |
| Xp * | E1-2b | GCATCCACTGGAGCACACG | 55.3°C | 1.5 mM |
| Xp * | E1-3a | GTTCATTTCCAGTCTGGCTTG | 52.3°C | 1.5 mM |
| Xp * | E1-3b | CAATATGTTACACTCAAGAAGTC | 49.8°C | 1.5 mM |
| Xp * | E2-I2mm | CCTCCTTAAATATTGTACGCC | 52.0°C | 1.5 mM |
| Xp * | E2-I1mm | TGCAGATTCAAGTGTCATGGC | 50.3°C | 1.5 mM |
| Xp * | NX1 | GGAGACTCAGACTAGAATGG | 60.0°C | 1.5 mM |
| Xp * | NX2 | GAGCGCCGTGAGTGTGAAG | 51.0°C | 1.5 mM |
| Xp * | Ia1 | CTTCACACTCACGGCGCTC | 50.4°C | 1.5 mM |
| Xp * | M1 | CAATGGATATCCAGGCCTCC | 55.3°C | 1.5 mM |
| Xp * | M2 | GGAAGATTGTAAGCACTCTGG | 52.3°C | 1.5 mM |
| Xp * | Ia2 | GGAGGCCTGGATATCCATTG | 49.8°C | 1.5 mM |
| Xp * | Ib1 | CCAGAGTGCTTACAATCTTCC | 52.0°C | 1.5 mM |
| Xp * | B2 | CACCTACACCACTCAGGAG | 55.3°C | 1.5 mM |
| Xp * | B1 | CGACTTGCCAAGACAGTGC | 52.3°C | 1.5 mM |
| Xp * | Ib2 | GCACTGTCTTGGCAAGTCG | 49.8°C | 1.5 mM |
| Xp * | E3-L | CTGAAGGCTCTTTGGAGGG | 53.1°C | 1.5 mM |
| 8q * | 9-2STS-R | GGATTTTCCCTCTTTGCTCAG | 52.3°C | 1.5 mM |
| 8q * | 9-2STS-L | GAGTTAGCATTTCTGTCATTTGC | 51.6°C | 1.5 mM |
| 8q * | 390bbRIGHT | CTCTATCACCTAGTTTGG | 45.6°C | 1.5 mM |
| 8q * | 526WK | GGATTACAGATGTGAGCCAC | 51.7°C | 1.5 mM |

Table 2. 1 (A). PCR primers and conditions of amplification, published or (*) designed by Y. Ishikawa-Brush and synthesized by ICRF, Clare Hall.

| Locus | Name | Primer Sequences | Anneal Temp. | $MgCl_2$ Conc. |
|---|---|---|---|---|
| 8q * | BK33RR | GTCTCAGTTCCTTCCTGTAG | 51.7°C | 1.5 mM |
| 8q * | CH8T3-35WK | CTGCCCTGAGTTGGAACAG | 53.1°C | 1.5 mM |
| 8q * | CH8-67BACK | GCTGAGTTCAGAGTCAATAG | 49.6°C | 1.5 mM |
| 8q * | C8-27WKLT | GGAAGAGAGCAGGGGAGG | 54.7°C | 1.5 mM |
| 8q * | C8-28RRWK | CTGGCCATCACTTTTCTCATC | 52.3°C | 1.5 mM |
| 8q * | C8-29LT | CTACAGGAAGGAACTGAGAC | 51.7°C | 1.5 mM |
| 8q * | C8-30WIDE | CATCTCTGTATGTTGCCCAGG | 52.2°C | 1.5 mM |
| 8q * | Synd-1 | GGGAGCCAGAGGAAAAGAAG | 53.7°C | 1.5 mM |
| 8q * | Synd-2 | CGGGGGCGCTCGCTTCCTC | 61.7°C | 1.5 mM |
| 8q * | Synd-3R | CTGATGGCAATATTAAGACAGAC | 51.6°C | 1.5 mM |
| 8q * | Synd-4L | CATGACAGACTTTATTTTGGCAC | 51.6°C | 1.5 mM |
| 8q * | Synd-NST | CTTCTTTTCCTCTGGCTCCCTC | 56.6°C | 1.5 mM |
| 8q * | Synd-RACE | GAGCTCCGGGGTATCTGGAG | 57.8°C | 1.5 mM |
| 8q * | G4T7WK1 | CTGTCATGGTGCAATTTGTC | 59.6°C | 1.5 mM |
| 8q * | G4T3WK1 | CACTTAGAAATCAAACCCACG | 50.3°C | 1.5 mM |
| 8q * | G16T3WK1 | CATGGACCTATGATAACAATG | 48.4°C | 1.5 mM |
| 8q * | G16T7WK1 | CGGGGAGAACACACTCTC | 52.5°C | 1.5 mM |
| 8q * | G1T7WK1 | CCTAGCCTCTCTGGACTG | 52.5°C | 1.5 mM |
| 8q * | G1T3WK1 | CTGGTACTTTCTGTACTAAAC | 48.4°C | 1.5 mM |
| 8q * | G1T3XWK2 | CATAGAATAACAACTGCTGG | 47.6°C | 1.5 mM |
| 8q * | G11T7WK1 | GGGTATATACCCAGTAATGG | 49.6°C | 1.5 mM |
| 8q * | G11T3WK1 | CTCTTCTAAATGTATCTAATGG | 47.3°C | 1.5 mM |
| 8q * | G14T3WK1 | GTACATAATTGTGTGCCATGC | 50.3°C | 1.5 mM |
| 8q * | G10T7WK1 | GTGGGGTTATCAGAAAAGGG | 51.7°C | 1.5 mM |
| 8q * | G10T3WK1 | CCAGCCTGGTCTCAAACTC | 53.1°C | 1.5 mM |
| 8q * | G37T3WK1 | GGCTTAATTTGAAAACCCCC | 49.6°C | 1.5 mM |
| 8q * | G34T7WK1 | CTCACTGTACACAACAGTCC | 51.7°C | 1.5 mM |
| 8q * | G34T3WK1 | GGTTACCTATTAATGGGATGG | 50.3°C | 1.5 mM |
| 8q * | G23T7WK1 | CCTGTTGTCATCACCAGGG | 53.1°C | 1.5 mM |
| 8q * | G23T3WK1 | CTCCAGTTTGCCCTTACAAC | 51.7°C | 1.5 mM |
| 8q * | TOBAM-1R | GTTTTGCCCTGGTTGCAAGC | 53.7°C | 1.5 mM |
| 8q * | TOBAM-1L | CGAGTTGTGTCAGCTCTGG | 53.7°C | 1.5 mM |
| 8q * | SPGII-3R | GCCGAAGAGGATACAAATGTG | 52.3°C | 1.5 mM |
| 8q * | SPGII-3L | CACATTTGAATCCTCTTCGGC | 52.3°C | 1.5 mM |
| 8q * | SPGII-XL | CTTCATCCTTCTTTCTCATGC | 50.3°C | 1.5 mM |
| 8q * | G37T3WK2 | CGTTCTTAACACTTTGGGCAG | 52.3°C | 1.5 mM |
| 8q * | F2.2H-R | CCACGCTCACTGTTCTTTAG | 51.7°C | 1.5 mM |
| 8q * | F2.2H-L | CTCTCACCACTGAGTTGAG | 51.0°C | 1.5 mM |
| 8q * | H3L | GTAGTCATCGTCATCAATAGG | 50.3°C | 1.5 mM |
| 8q * | BAMAWAY | CCAGAGCTGACAACAACTCG | 53.7°C | 1.5 mM |
| 8q * | G223T7WK1 | GCTACTGTTATCAATGTTTCC | 48.4°C | 1.5 mM |
| 8q * | G10T7WK2 | CATTATGGGGACTTTCCTCC | 51.7°C | 1.5 mM |
| 8q * | G10T3WK2 | GAAGCACTTTATGGCTCTTC | 49.6°C | 1.5 mM |
| 8q * | G1T7WK2 | CATCTCTGAAACACCAAAGC | 49.6°C | 1.5 mM |
| 8q * | G1T3WK2 | CACTTTTAGTTGCAAGTGCG | 49.6°C | 1.5 mM |

Table 2. 1 (A). Continued

| LocusName | | Primer Sequences | Anneal Temp. | $MgCl_2$ Conc. |
|---|---|---|---|---|
| 8q * | H3OVER | CTGAAGCTTCTTCAATGGAG | 49.6°C | 1.5 mM |
| 8q * | SPG3LWK1 | CCACTCTTCTAACTGCAGG | 51.0°C | 1.5 mM |
| 8q * | G11T7WK2 | CAGGAAAACCTTCCTAAATAAG | 49.1°C | 1.5 mM |
| 8q * | G11T3WK2 | CTTTACACTGTCTCTGCCTG | 51.7°C | 1.5 mM |
| 8q * | G5G18T3WK1 | CAACAGTCCAGGTGCAGG | 52.5°C | 1.5 mM |
| 8q * | H3R | CCTATTGATGACGATGACTAC | 50.3°C | 1.5 mM |
| 8q * | G1T7BACK | CTCAGGCTGATGGAGTAG | 50.2°C | 1.5 mM |
| 8q * | G16T7BACK | CAATGTTGCTGCACAATTCG | 49.6°C | 1.5 mM |
| 8q * | G16T3BACK | GACAATATGTAGAGGTAGAAG | 48.4°C | 1.5 mM |
| 8q * | G14T7WK1 | CAGTGTCACTTTGTGTCTGG | 51.7°C | 1.5 mM |
| 8q * | G16T3WK2 | GTCTGTTGCTGACTGAAATAG | 50.3°C | 1.5 mM |
| 8q * | INTROH3LL | CTCACCATCTCCCTCTGAC | 53.1°C | 1.5 mM |
| 8q * | G34T3BACK | GACTTTGCAATCAGTGCAGG | 51.7°C | 1.5 mM |
| 8q * | H3OVWK1 | CTCTTTGTCACTCCCTTACTG | 52.3°C | 1.5 mM |
| 8q * | ATGBACK | CAAGCCCAAGGTGAGCAGG | 55.3°C | 1.5 mM |
| CAG-tri | | CAGCAGCAGCAGCAGCAGCAGCAGCAG | 69.7°C | 1.5 mM |

Table 2. 1 (A). Continued

| LocusName | Primer Sequences | Anneal Temp. | $MgCl_2$ Conc. |
|---|---|---|---|
| DMD PbF1 | GAAGATCTATATTTTACAACGCAGAAATATGG | 55.0°C | 1.5 mM |
| DMD PbR | CTTCCATGCCAGCTGTTTTTCTGTCACTC | 55.0°C | 1.5 mM |
| DMD PmF | GAAGATCTAGACAGTGGATACATAACAAATGCATG | 55.0°C | 1.5 mM |
| DMD PmR | TTCTCCGAAGGTAATTGCCTCCCAGATCTGAGTCC | 55.0°C | 1.5 mM |
| DMD 3F | TCATCCATCATCTTCGGCAGATTAA | 55.0°C | 1.5 mM |
| DMD 3R | CAGGCGGTAGAGTATGCCAAATGAAAATCA | 55.0°C | 1.5 mM |
| SD6 | TCTGAGTCACCTGGACAACC | 53.7°C | 1.5 mM |
| SA2 | ATCTCAGTGGTATTTGTGAGC | 50.3°C | 1.5 mM |
| SD2 | CUACUACUACUAGTGAACTGCACTGTGACAAGC | 62.8°C | 1.5 mM |
| SA4 | CUACUACUACUACACCTGAGGAGTGAATTGGTCG | 64.3°C | 1.5 mM |
| CUA-SD5 | CUACUACUACUACCCTCGAGGTCGACCCAGC | 67.2°C | 1.5 mM |
| CAU-SA5 | CAUCAUCAUCAUCTAGAACTAGTGGATCTCCAGG | 62.9°C | 1.5 mM |
| Bluescript T3 | ATTAACCCTCACTAAAG | 64.3°C | 1.5 mM |
| Bluescript T7 | AATACGACTCACTATAG | 64.3°C | 1.5 mM |
| SP6 Promoter | CATACGATTTAGGTGACACTATAG | 64.3°C | 1.5 mM |
| M13 Forward | GTAAAACGACGGCCAGT | 64.3°C | 1.5 mM |
| M13 Reverse | CAGGAAACAGCTATGAC | 64.3°C | 1.5 mM |
| gt10 Forward | AGCAAGTTCAGCCTGGTTAAG | 64.3°C | 1.5 mM |
| gt10 Reverse | CTTATGAGTATTTCTTCCAGGGTA | 64.3°C | 1.5 mM |
| gt11Forward | GGTGGCTACGACTCCTGGAGCCCG | 64.3°C | 1.5 mM |
| gt11Reverse | TTGACACCAGACCAACTGGTAATG | 64.3°C | 1.5 mM |

Table 2. 1 (B). PCR primers and conditions of amplification, published and synthesized by ICRF, Clare Hall.

purification kit (Qiagen Inc.)

### 2. 2. 5 Preparation of Long DNA Fragments using PCR

The Expanded Template System (Boehringer-Mannheim) was performed in parallel with regular PCR when the expected size was unknown. This system uses *Pwo* (*Pyrococcus woesei*) DNA polymerase which possesses 3′ to 5′ exonuclease activity (3′ exo+) associated with 'proof reading', but which does not have 5′ to 3′ exonuclease activity (5′ exo-). Because of its 5′ exo- activity, lesser amounts of the dNTPs were required. This facilitated removal of excess incorporated dNTPs and the phenol-chloroform procedure was not required. This system provides three different buffers depending on the amplification length. Buffer II, designed for about 10 kb amplification and containing 50 mM Tris-HCl, pH 9.2, 16 mM $(NH_4)_2SO_4$ and 2.25 mM $MgCl_2$ was most often used. The amplification began with a 'hot start', essential for avoiding primer degradation by 3′ exo+. The amplification was performed in a Perkin Elmer GenAmp 9600 Thermal Cycler using the following procedure.

| | |
|---|---|
| 1 cycle | Denature template 2 min at 94°C. |
| 10 cycles | Denature at 94°C for 10 sec. |
| | Annealing temperature for 30 sec. |
| | Elongation temperature at 68°C for 10 min. |
| 20 cycles | Denature at 94°C for 10 sec. |
| | Annealing temperature for 30 sec. |
| | Elongation temperature at 68°C for 10 min |
| | with cycle elongation of 20 sec for each cycle. |
| 1 cycle | Elongation at 68°C for 10 min. |

The PCR products were analyzed on 0.8% (w/v) agarose gels with a λ*Hind*III marker (2. 3. 2). The PCR products were purified as described above (2. 3. 5).

### 2. 2. 6 Interspersed repeat element PCR (Alu-PCR Amplification)

Probes were newly generated using Alu-PCR Amplification. YACs spanning

the breakpoint were used as templates for Alu repeat primed PCR using various combinations of two of either, Collins-5, Collins-3 (Tagle and Collins, 1992), 2729 and 3144 (Monaco *et al.*, 1991). The PCR was performed in a reaction buffer containing 10 mM Tris-HCl, pH 8.3, 50 mM KCl, 1.5 mM $MgCl_2$, 250 μM each dNTP, and 500 ng YAC DNA. An initial denaturation of 4 minutes at 94°C was followed by 30 cycles with conditions suitable to each set of primers (Table 2. 1A & B).

### 2. 2. 7 Preparation of Di-and Tri-nucleotide Oligo-probes

Primers with a length of 24 nt of 10 possible trinucleotide repeats (AAC/GTT, AAG/CTT, AAT/ATT, ACC/GGT, ACG/CGT, ACT/AGT, AGG/CCT, ATC/GAT, CAG/CTG, CCG/CGG) and di-nucleotide repeats (CA/GT) were synthesized at Clare Hall, ICRF, prepared as described in (2. 2. 7), and used as oligo-probes (2. 5. 3 and 2. 6. 4).

### 2. 2. 8 5′ Rapid Amplification of cDNA Ends (5′ RACE)

To obtain 5′ flanking sequences from mRNA templates, the 5′-RACE technique was employed. The procedure involved 7 stages: (1) design an initial antisense primer outwards from the 5′-end region of the SDC2 gene, (2) synthesize the first strand of cDNA from a total RNA template using the initial antisense primer and reverse transcriptase (RT), (3) incorporate a homopolymeric tail (a dC-tail) at the 3′-end of this cDNA strand using terminal deoxynucleotidyl transferase (TdT), (4) degrade the template with RNase H, (5) anneal a sense primer(poly dG)containing an extended (anchor) sequence, (6) synthesize the second strand using an internal (nested) primer to produce a complementary copy of the anchor sequence, and (7) amplify by

PCR using the nested and anchor sequence primers. The result is an amplified sequence that extends beyond the 5'-end of the known sequences.

2. 2. 8a Reverse transcriptase PCR (RT-PCR)

Total cellular RNA was isolated from an EBV transformed lymphoblastoid cell line using the TRIzol reagent, following the recommended protocol (Gibco, BRL). 2.5 μg of RNA was reverse transcribed using random hexamers (Pharmacia)andSuperScript II reverse transcriptase (Gibco, BRL). 0.4 ng of each cDNA was amplified with each set of primers for 30 cycles.

2. 2. 8b First-strand cDNA synthesis

A volume of 14 μl of diethylpyrocarbonate (DEPC)-treated water containing 1 μg of RNA and 200 ng of initial primer was incubated for 10 min at 70°C to denature the RNA, then chilled on ice. To this was added 4 μl of the Synthesis-buffer and then incubated for 2 min at 42°C. 200 U of SuperScript reverse transcriptase (Gibco, BRL) was added and the mixture incubated for 30 min at 42°C and again for 5 min at 55°C. 2U of RNase H was added to remove RNA strands and then incubated for 10 min at 55°C.

| | |
|---|---|
| Synthesis-buffer | 20 mM Tris-HCl pH 8.4<br>50 mM KCl<br>2.5 mM $MgCl_2$<br>100 μg/ml BSA<br>10 mM DTT<br>100 nM cDNA initial primer<br>500 μM each dATP, dCTP, dGTP, dTTP |

2. 2. 8c TdT tailing of cDNA

Purified first strand cDNA using a GlassMAX DNA Isolation (Gibco, BRL) was resuspended with DEPC-treated water to a final volume of 16 μl, which was incubated for 10 min at 70°C and chilled for 1 min on ice. 4μl of TdT

synthesis buffer (0.5 x final concentration) was added and incubated 10 min at 37°C. The TdT was heat-inactivated for 10 min at 68°C.

| | |
|---|---|
| TdT synthesis buffer | 10 mM Tris-HCl pH 8.4<br>25 mM KCl<br>1.25 mM $MgCl_2$<br>50 µg/ml BSA<br>10 mM DTT<br>0.5 U/µl TdT |

2. 2. 8d dC tailed cDNA amplification

A 5 µl aliquot of the TdT mixture was amplified in dCTP synthesis buffer by PCR. A single PCR of 30 cycles did not generate a sufficient amount of PCR product and a second round of PCR was performed. The products were analyzed by 2% (w/v) agarose gel electrophoresis and cloned in plasmid *p*AMP 1 vector (2. 7. 5).

| | |
|---|---|
| dCTP synthesis buffer | 20 mM Tris-HCl pH 8.4<br>50 mM KCl<br>2.5 mM $MgCl_2$<br>100 µg/ml BSA<br>10 mM DTT<br>400 nM Anchor primer<br>200 µM each dATP, dCTP, dGTP, dTTP<br>0.04 U/µl Taq DNA Polymerase |

## 2. 3 ISOLATION OF CLONED DNA AND PCR PRODUCTS

The cloned DNA was digested with an appropriate restriction endonuclease flanking the cloning site to separate insert DNA from vector DNA. Digested DNA and PCR products were separated in agarose gels using electrophoresis. High molecular weight YAC DNA was separated from yeast DNA using Pulsed Field Gel Electrophoresis (PFGE). Size-fractionated DNA was stained with ethidium bromide and visualized under a short-wave UV transilluminator. After size confirmation in the agarose gel, digested DNA,

PCR products, and YAC DNA were loaded on a low melting temperature gel and excised from the gel under a long-wave UV transilluminator, leaving behind its vector DNA and unwanted PCR products. The purified DNA was labeled and used as a probe.

### 2. 3. 1 Cloned DNA Digestion by Restriction Endonuclease

All endonuclease digestions were performed using temperature and buffer conditions as instructed by the supplier, New England Biolabs (NEB). The amount of restriction enzyme and the duration of digestion were determined under the assumption that one unit (1 U) of enzyme digests 1 μg of DNA in 1 h. The double enzyme digests were performed in a universal enzyme buffer (T4 polymerase buffer), omitting the precipitation procedures between the two enzymatic reactions. The reactions were stopped by enzyme inactivation at 68°C for 10 min.

### 2. 3. 2 DNA Fragment-Size Separation using Gel Electrophoresis (GE)

To examine DNA fragment-size by its mobility, a sample-portion of the digested DNA and PCR products were fractionated using agarose gel electrophoresis in 1 x TBE buffer. The concentration and the type of agarose gel were determined by the DNA fragment-size to be detected. An 0.8 cm thick slab of 0.3% (w/v) agarose gel (SeaKem, F. M. C.) was used for DNA molecules in the size range 10-50 kb, 0.8% (w/v) SeaKem agarose gel for the size range 1.0 - 10.0 kb, and 2% (w/v) NuSieve gel (F. M. C.) for PCR fragments in the range 0.1 - 1.0 kb in size. DNA fragments were separated with 30 V (constant voltage) overnight. The markers used were λ*Hind*III (in 5% glycerol, 0.05% bromophenol blue and 0.05% xylene cyanol) for fragment sizes

between 0.3-10.0 kb or ϕx174RF/*Hae*III (in 50% glycerol, 0.1 M EDTA, pH 8.0, and 0.1% orange G) for sizes between 0.1-1.0 kb.

When constructing restriction maps, a more accurate measurement of fragment size was required. However, the relationship between standard size and distance travelled in any given gel was non-linear. Linearity was improved by taking the log of both standard size and distance travelled (Table 2. 2), but some non-linearity remained, making a straight line calibration curve inaccurate. To improve accuracy, a cubic polynomial was fitted to the log of standard distance, for example,

$$\ln(standard\ \ length) = a + b\ \ln(distance) + c\ \ln(distance)^2 + d\ \ln(distance)^3$$

The parameters were estimated for each gel using polynomial least squares regression. This was done in Cricket Graph, which showed the fitted curve plotted on the standard sizes (Fig. 2. 1). The calibration curve was used to estimate the sizes of sample fragments for a gel using a spreadsheet, EXCEL. An example for the calibration curve in Table 2. 1 is shown in Table 2. 3.

| Std. Length | Distance | Ln (Std. Length) | Ln (Distance) |
|---|---|---|---|
| 23130 | 3.10 | 10.049 | 1.131 |
| 9416 | 5.30 | 9.150 | 1.668 |
| 6550 | 6.80 | 8.788 | 1.917 |
| 4361 | 9.60 | 8.380 | 2.262 |
| 2322 | 13.10 | 7.750 | 2.570 |
| 2027 | 14.00 | 7.614 | 2.637 |

Table 2. 2. Example of the relationship between standard lambda *Hind*III fragment length anddistance travelled on a gel, and of log of standard length and log of distance.

Estimates of fragment sizes are shown in the last column. Because the calibration curve estimates log size, size in base pairs is achieved by calculating $e^{\ln(size)}$. The calibration curve is most reliable for unknown fragment sizes falling within the range of the standard sizes (2027-23130 bp).

$$y = 15.6988 - 8.5524x + 3.9473x^2 - 0.708x^3$$

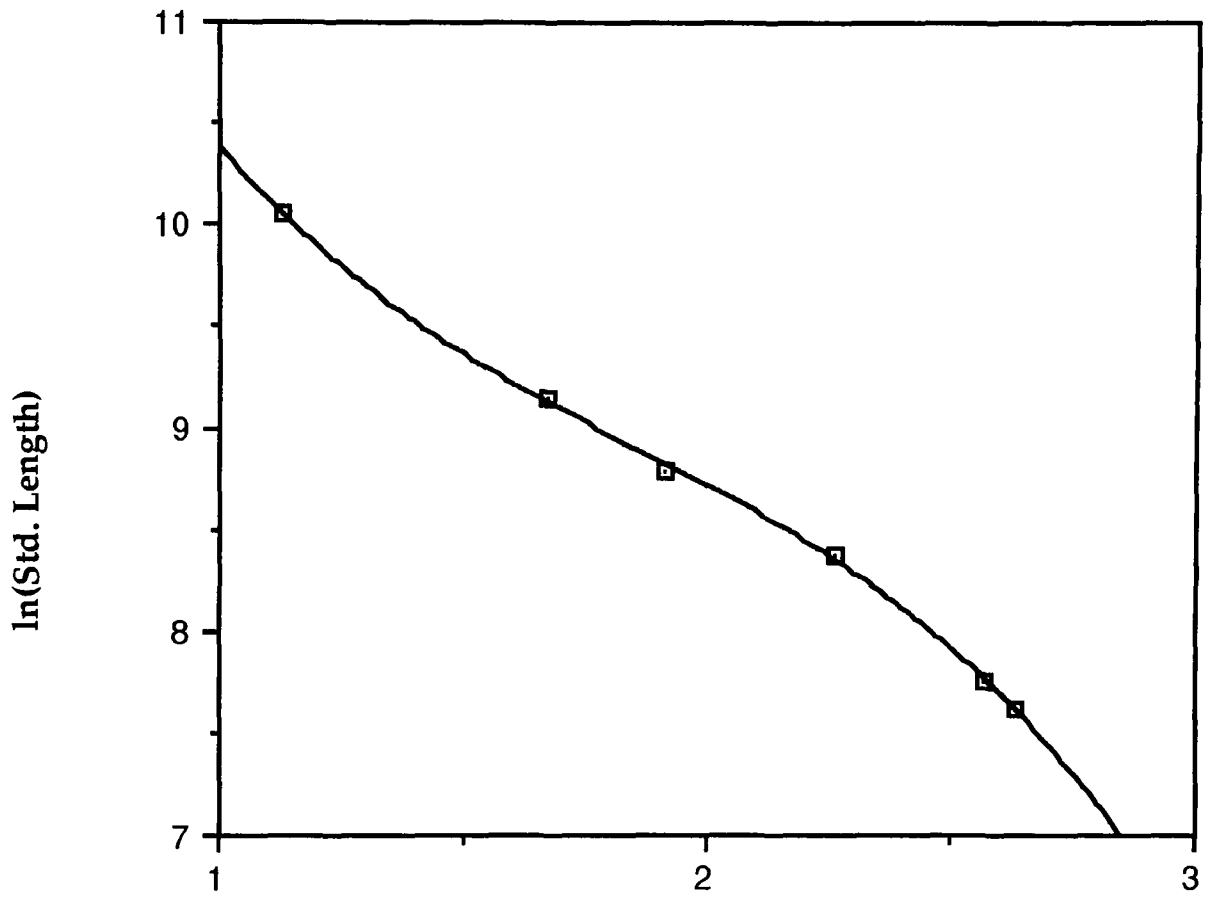



Fig. 2. 1. Example cubic polynomial regression of log standard length on log distance of lambda *Hind*III fragments. Values taken from Table 2. 2. In the equation, y is ln(std. length) and x is ln(Distance). The equation was used to estimate unknown fragment size from migration distances.

| Migration of unknowns | Log Distance | Est Log Frag size | Est Frag size of unknowns |
|---|---|---|---|
| 6.70 | 1.90 | 8.84 | 6907 |
| 7.15 | 1.97 | 8.76 | 6376 |
| 7.75 | 2.05 | 8.66 | 5758 |
| 8.37 | 2.12 | 8.56 | 5199 |
| 8.98 | 2.19 | 8.46 | 4708 |
| 9.48 | 2.25 | 8.38 | 4341 |
| 9.90 | 2.29 | 8.31 | 4054 |
| 10.00 | 2.30 | 8.29 | 3988 |
| 16.45 | 2.80 | 7.16 | 1282 |

Table 2. 3. Example use of the equation in Fig. 2. 1 to estimate sizes of unknown fragments from the log of migration distances on the gel. The equation estimates log of fragment size (column 3), and the antilog (*e*) of column 3 yields fragment size in base pairs, column 4.

### 2. 3. 3 YAC DNA Isolation using Pulsed Field Gel Electrophoresis (PFGE)

The YAC blocks stored in $TE_{50}$ were washed three times in 20 volumes of 1 x TE for 30 min each time and cooled on ice to make them firmer. Ten agarose blocks were loaded into a single long slot of the well prepared in a 1.3% (w/v) SeaPlaque agarose (F. M. C.) gel. Yeast chromosomes of *Saccharomyces cerevisiae* and MidRange II PFG ladder (NEB) were loaded as DNA size markers. Separation of YAC DNA from Yeast DNA was carried out in 0.5 x TBE using Pulsed Field Gel Electrophoresis (PFGE) in a contour-clamped homogeneous electric field gel (CHEF-DRII) system (Biorad) at 170 V (5.2 V/cm) with switch times of 60 s for 18 h, 90 s for 10 h and 120 s for 8 h.

### 2. 3. 4 Visualization of DNA using Ethidium Bromide Staining

The migration of DNA fragments in the agarose gel was visualized by staining with the intercalating dye, ethidum bromide (EtBr). A final

concentration of 0.5 μg/ml EtBr (Sigma) was mixed during gel formation. Agarose gel containing RNA and YAC DNA was stained after completion of the migration at room temperature for 20 min. The intercalated DNA with EtBr was visualized under short wave (302 nm) UV transillumination and photographed using the Eagle Eye II camera (Stratagene).

| | |
|---|---|
| 5 x TBE<br>(1 litre) | 54.0 g Tris base<br>27.5 g Boric acid<br>20.0 ml 0.5 M EDTA, pH 8.0 |

### 2. 3. 5 Purification of Insert DNA and PCR Products

To isolate the desired DNA fragments, the remainder of the digested DNA and PCR products were reloaded on a 1% (w/v) SeaPlaque agarose gel (F. M. C.), which has a feature of Low Melting Temperature (LMT). The desired DNA was separated in 1 x Tris-acetate-EDTA (TAE) buffer, pH 8.0 (Sambrook *et al.*, 1989) and excised using a scalpel from the gel under long-wave UV transillumination. The insert DNA was recovered from the excised agarose gel by digesting overnight at 37 °C with beta-agarase (New England, Biolabs) and purified using phenol-chloroform extraction and ethanol precipitation. For use in DNA sequencing, the excised gel was quickly purified using Qiagen columns (Qiagen Inc.).

## 2. 4 MEASUREMENT OF DNA CONCENTRATION

DNA concentration was measured using a fluorometer or a spectrophotometer according to the type of nucleic acid. The number of cells were counted using a Haemocytometer.

### 2. 4. 1 DyNA Quant 200 Fluorometer (Hoefer)

The DNA concentration was measured by fluorescence emission using a

DyNA Quant 200 Fluorometer (Hoefer) and calf thymus DNA as a standard. Hoechst 33258 (H 33258) dye has fluorescence excitation and emission spectrums of 365 nm and 458 nm respectively when bound to AT region of double stranded DNA, but 345 nm and 492 nm respectively when unbound. The Fluorometer excites a DNA sample at 365 nm and measures fluorescence in the 458 nm range, and is thus a measure of DNA concentration. Known concentrations of calf thymus DNA are used as a calibration. Calf thymus DNA is similar in AT concentration to human DNA (42%). Fluorescence measurement, which is free from RNA contamination, was used determine DNA quantities required in nick translation, ligation, and sequencing.

### 2. 4. 2 Spectrophotometer with Optical Density (OD)

Fluorescence measurement was not as suitable for plasmid DNA, which has different H 33258 binding characteristics arising from differences in DNA conformation (being supercoiled, relaxed and circular). Plasmid DNA and single stranded nucleic acids (which do not form a complex with dye) were therefore measured by optical density (OD) using an ultraviolet absorbance spectrophotometer (Beckman DU-62). The reading was recorded at 260 nm in comparison to a reading with no DNA (1 x TE). The conversion factors of 1.0 absorbance at 260 nm ($A_{260}$) corresponds to 50 μg/ml of double-stranded DNA and 33 μg/ml of oligonucleotides (Sambrook *et al.*, 1989). The DNA concentration was calculated by multiplying the $OD_{260}$ value of the sample by the dilution factor and also multiplied by its conversion factor. The purity of the DNA was checked against a purity indicator of absorbances at 260 nm and 280 nm ($A_{260}/A_{280}$), which has a ratio of 1.8. When a DNA preparation had a

ratio less than 1.8, protein contamination was indicated and a treatment of phenol-chloroform extraction was used for further DNA purification (2. 1. 1).

### 2. 4. 3 Cell Counts with Haemocytometer

Cells were rinsed, resuspended in 50 ml 1 x PBSA , and counted using a Haemocytometer (Weber England). The well of the chamber is divided into nine large squares, each $10^{-4}$ ml in volume. The wells were filled with the diluted cells and Trypan Blue (Weber England) and covered with a coverslip. To count the cells in the solution, as cells per ml, the number of cells in one square was multiplied by the dilution factor, the total volume in the tube, and $10^4$.

| | |
|---|---|
| 10 x PBSA<br>(1 litre) | 80 g NaCl<br>2.5 g KCl<br>14.3 g $Na_2HPO_4$<br>2.5 g $KH_2PO_4$ and autoclave |

## 2. 5 NUCLEIC ACID TRANSFER TO MEMBRANES

Southern blots embedded with DNA fragments were hybridized to probe through complementary association. Various kinds of DNA fragments, genomic (2. 5. 2), YAC, cDNA, cosmid, phage, plasmid, and PCR-synthesized DNA fragments (2. 5. 1) were used. Northern blots for expression studies were prepared from total cellular RNA extracted from an EBV transformed lymphoblastoid cell line. Northern blots embedded with size-fractionated RNA molecules from various tissues were hybridized with probes such as cDNA, phage, plasmid, and PCR-synthesized DNA fragments (2. 5. 3).

### 2. 5. 1 Southern Blots

To allow repeated use of the data, a replica of the DNA migrations in the

agarose gel was made on Southern blots (Southern, 1975). Prior to blotting, the gel was soaked in denaturing solution (0.5 M NaOH, 1.5 M NaCl) for 30 min to yield to single-stranded DNA (ssDNA) for the filter-hybridization experiments that were to follow. The negatively charged DNA was transferred to a positively charged nylon membrane (Hybond $N^+$, Amersham) in denaturant by capillary action overnight. The membranes were rinsed in 4 x SSC and air dried. The DNA was fixed by baking the membranes at 80°C for 10 min and UV cross-linked using a Stratalinker (Stratagene).

| | |
|---|---|
| 20 x SSC<br>(20 litres) | 175.3 g NaCl<br>88.2 g Tris-Na Citrate and autoclave |

### 2. 5. 2 Patient Genomic DNA Blots

To identify junction fragments, patient genomic Southern blots were prepared for hybridization with markers flanking the translocation breakpoint. Patient, normal male and normal female genomic DNA, the latter two for controls, were digested, purified, and precipitated. Five different enzyme (*Eco*RI, *Pst*I, *HindIII*, *Kpn*I and *Sna*BI) reactions took place separately in 250 µl of total reaction buffer following the NEB protocol (2. 3. 1). Each pellet was resuspended in 12 µl 1 x TE and re-digested in 20 µl of total reaction buffer for 6 h. All digests were loaded in 400 ml of 1% (w/v) SeaKem (F. M. C.) agarose gel and separated by electrophoresis in 1 x TBE overnight at 50 V. The gel was blotted to Hybond $N^+$ (Amersham) filters (2. 5. 1).

### 2. 5. 3 Patient Northern Blots

Total cellular RNA was extracted from an EBV transformed lymphoblastoid cell line using TRIzol (Gibco, BRL). 30 µg of RNA was separated on a 1%

(w/v) agarose gel containing 6% (v/v) formaldehyde and transferred to a Hybond N+ (Amersham) filter. The loading buffer was 1 x MOPS, 100% (v/v) deionized formamide, and 37% (v/v) formaldehyde. 23 µl of loading buffer were used for each 7 µl of RNA.

| | |
|---|---|
| 10 x MOPS | 200 mM MOPS (Sigma )<br>10 mM EDTA (Merck)<br>85 mM NaAc (Merck) |

## 2. 6 CLONING, LIBRARY CONSTRUCTION AND FILTER LIFTS

Physical mapping has been facilitated by the availability of existing library filters containing human YAC, cosmid, or cDNA clones (2. 6. 1) provided by the ICRF Reference Library system. A patient genomic library and sublibraries of YACs or cosmids were individually constructed as described in 2. 6. 2.

### 2. 6. 1 Existing Library Filters

The ICRF Human YAC, 48XXXX, Library filter contains 20,000 clones with an average DNA size of 700 kb (Larin *et al.*, 1991). This library enriched in X chromosomes increased the extraction-rate of X chromosome specific YACs. High density gridded colony filters of cosmid libraries were also available from the ICRF and used for hybridization with unique probes and whole YACs as probes. The libraries were constructed from digests of DNA from 'flow sorted' human X-chromosomes (ICRF FSCX) and ligated into Lawrist 4 vectors (Baxendale *et al.*, 1991; Nizetic *et al.*, 1991). These library filters were screened in duplicate. Human multiple tissue northern blots and human cDNA testis library cloned in λg10 vectors were available from Clontech (Sambrook *et al.*, 1989).

### 2. 6. 2 Constructed Cosmid Library Filters

Because of missing regions of interest in the ICRF FSCX library, a cosmid library and its replicas were constructed from the YACs spanning the translocation breakpoint. Constructing a cosmid library involved preparation of vector and insert DNA, ligation of the two, packaging and transfection. The transfected cosmids were screened with total human DNA as a probe and positives were picked in to microtitre plates and transferred on to membranes.

#### 2. 6. 2a Preparation of vector DNA

A 20 μg of SuperCos 1 (Stratagene) was digested with 9 U/μg of *Xba*I in a 200 μl volume at 37°C overnight. The DNA was extracted, precipitated (See Section 2. 1) and resuspended in 1 x TE to a concentration of 1 mg/ml. The complete digestion was examined electrophoretically with approximately 500 ng of the digest in a 0.8% (w/v) agarose gel. The *Xba*I digested vector was dephosphorylated with 0.65 U/μg calf intestinal alkaline phosphatase (CIAP) in T4 polymerase buffer at 37°C for 30 min to prevent vector-vector ligation. The reaction was stopped with trinitriloacetic acid (TNA) at the concentration of 0.015 M at 68°C for 20 min. The vector was again extracted and precipitated before being resuspended at a concentration of 1 mg/ml. The efficiency of ligation after dephosphorylation was tested by the comparison of three reactions with SuperCos/XbaI/CIAP DNA and ligase buffer (1) with no ligase, (2) with 0.5 μl T4 DNA ligase (NEB, 400 U/μl), and (3) with 0.5 μl T4 DNA ligase and 0.5 μl T4 polynucleotide kinase (NEB, 10 U/μl). Each reaction contained 300 ng, were incubated for 1 h at 37°C and loaded on a 0.8% (w/v)

agarose gel. The efficiency of dephosphorylation was observed in larger bands seen in the ligase only reaction and in the ligase plus kinase reaction. The vector preparation was completed with the digestion of the vector to create a cloning site with 5 U/µl *Bam*HI for 2 h at 37°C followed by extraction, precipitation, and resuspension at a concentration of 1 mg/ml.

2. 6. 2b Preparation of cosmid insert DNA from YAC DNA

SuperCos 1 accommodates DNA fragments ranging in size from 30 to 45 kb. YAC DNA was partially digested with a frequent cutting enzyme, *Mbo*I, generating a random population of overlapping fragments with compatible sticky ends to the *Bam*HI cloning site of the SuperCos 1 cosmid vector. Generating the appropriate insert DNA size was carried out by testing four different incubation times for digestion. One YAC block was melted in 20 µl 10 x T4 polymerase salts and 40 µl $dH_2O$ at 68°C for 10 min and cooled to 37°C to add the restriction enzyme of 0.5 U *Mbo*I with 20 µl BSA (1 mg/ml) and 20 µl DTT (5 mM). The block was divided into four and each incubated individually, one fourth for 30 sec, 1 min, 5 min, and 10 min. The enzyme was inactivated at 68°C and brought to 37°C for beta-agarase (3 U, NEB) treatment for 3 h at 37°C. 20 µl of each digest was fractionated by electrophoresis on a 0.3% (w/v) agarose gel against lambda *Hind*III marker. The digests with suitable size were dephosphorylated with 0.1 U CIAP in total volume of 220 µl 1 x T4 polymerase buffer for 30 min at 37°C. The phosphatase reaction was stopped with trinitriloacetic acid at a concentration of 0.015 M for 20 min at 68°C. 5 µl of Dextran T40 (10 mg/ml) was added as carrier. The DNA was extracted twice with phenol-chloroform, once with

chloroform, and then resuspended in 1 x TE. A further 5 µl dextran T40 and 0.1 M NaCl were added for the DNA precipitation in two volumes of 100% (v/v) ethanol. The pellet was resuspended in 10 µl of 1 x TE.

2. 6. 2c Ligation of insert DNA to SuperCos 1

The insert DNA (2.5 µg) was ligated to SuperCos 1 (1 µg) in a 20 µl of ligation buffer, 400 U NEB ligase, 10 mM ATP at 16°C overnight.

2. 6. 2d Packaging

The *in vitro* packaging of the ligated material was carried out using freeze thaw lysate and sonic extract (Gigapack II Gold, Stratagene) following the protocols provided by the supplier.

2. 6. 2e Preparation of host cells

A single colony of the XL1 Blue host bacteria was inoculated in 50 ml Terrific Broth (TB) with 0.2% maltose and 10 mM $MgSO_4$ overnight. The pellet produced by spinning the culture at 3,000 rpm at 4°C for 10 min was resuspended in 10 mM $MgSO_4$ and adjusted to obtain an $OD_{600}$ of 0.5.

2. 6. 2f Titering and plating the cosmid library

The packaged cosmids were transfected to the XL1 Blue host bacteria. To calculate the titre of the library, the packaged DNA was diluted in 1:10 and 1:50. 25 µl of each of the dilutions and 25 µl of fresh overnight culture with XL1 Blue cells ($OD_{600}$ =0.5) were incubated at room temperature for 30 min. The transformed cells were allowed to recover and express the ampicillin resistance in 200 µl fresh LB broth, spread on to LB plates containing 50 µg/ml ampicillin and incubated at 37°C overnight.

The yeast genome is about 15 Mb and average YAC size is about 500 kb,

representing 3.3% of the total DNA in the yeast cell. For one genome equivalent (GE), about 430 clones (yeast and YAC DNA of about 15,000 kb/average cosmid insert sizes of 35 kb ≈ 430 clones) were needed. The amount of packaged DNA needed for the coverage of seven genome equivalents, about 3,000 clones, was mixed with an equal volume of the $OD_{600}$ = 0.5 cells in 10 mM MgSO4 at room temperature for 30 min. After incubation for 1 h in 4 volumes of LB, the cells were plated at low density (1000 colonies per 20 x 20 cm plate), and incubated at 37°C overnight.

2. 6. 2g Replicating the cosmid library

To produce a replica of the cosmid library, Hybond-$N^+$ nylon membrane (Amersham) filter was pre-wetted on LB plates containing 50 μg/ml ampicillin, placed on the colonies of the master plate, and marked with a needle to aid in realignment after hybridization. Each filter was placed colony-side up on an LB plate containing 50 μg/ml ampicillin. Both master and replica plates were incubated for another 8 hours. The master plate was then sealed and stored at 4°C. The replica plates were prepared for hybridization by placing it colony side up on 3MM Whatman paper pre-wetted in three different solutions for 3 min each. The first solution was 0.5 M NaOH, the second 1 M Tris-HCl, pH 7.6, and the third 1 M Tris-HCl, pH 7.6, 1.5 M NaCl. The filters were immersed in 2 x SSC and rubbed to remove the bacterial debris. The filters were rinsed with fresh 2 x SSC, dried at 80°C for 2 h, and UV-crossed.

### 2. 6. 3 Phage Library from the Patient Genomic DNA

The lambda bacteriophage cloning system was used to generate a patient

genomic library cloning the junction fragments spanning the translocation breakpoint. DNA isolated from ML was digested and cloned in the lambda replacement vector, Lambda DASH II/*Bam*HI Vector provided from Stratagene Ltd.

2. 6. 3a Preparation of insert DNA

10.0 μg genomic DNA from the patient, ML, was digested with 10 U of *Bgl*II at 37°C for several hours. The reaction was terminated by adding a final concentration of 15 mM EDTA and incubation at 68°C for 15 min. The digests were extracted sequentially with phenol, chloroform, precipitated in ethanol, washed with 70% (v/v) ethanol, and resuspended in 1 x TE.

2. 6. 3b Ligation of insert DNA to Lambda DASH II

Two enzymes, *Bam*HI (G**GATC**C) used for the vector and *Bgl*II (A**GATC**T) used for the insert, possessed the compatible cohesive ends to form a hybrid site. At 16°C, 1.0 μg of *Bam*HI-digested vector arms and 0.4 μg of *Bgl*II-digested-insert DNA were ligated overnight in 10 μl reaction buffer.

2. 6. 3c Packaging

The ligated mixture was packaged using GigaPack II Gold packaging extracts (Stratagene), following the Stratagene protocols.

2. 6. 3d Preparation of host cells

The XL1 Blue MAR bacteria were used as host cells. Preparation was the same as 2. 6. 2e for the XL1 Blue host bacteria in constructing of the cosmid library, and adjusted to obtain an $OD_{600}$ of 0.5.

2. 6. 3e Titering the phage library

The packaged phages were transfected to the XL1 Blue MAR host bacteria. To

calculate the titre of the library, the packaged DNA was diluted to $10^{-1}$, $10^{-2}$, $10^{-3}$, and $10^{-4}$. Genomic DNA is about $3 \times 10^6$ kb and average phage DNA is 15 kb. For one genome equivalent (GE), therefore, required about 200,000 clones (genomic DNA/average phage insert sizes of 15 kb). The titre was calculated to obtain about 600,000 clones, the amount of packaged DNA needed to cover three genome equivalents.

2. 6. 3f Plating the phage library

Ten 1.5% NZCYM 22 x 22 cm plates (60,000 pfu per plate) and 300 ml 0.7% NZCYM Top Agar were prepared. The latter was kept warm in a water bath adjusted to 48°C. Packaged phage DNA in SM buffer were incubated with equal volume of the XL1 Blue MAR host bacteria ($OD_{600}$ =0.5) for 15 min at 37°C and then mixed with 25 ml of Top Agar, plated on the solidified 1.5% NZCYM plates, and incubated at 37°C overnight.

| | |
|---|---|
| SM buffer | 2.9 mg NaCl |
| (500 ml) | 2.0 g MgSO4 |
| | 25 ml 1 M Tris-Cl, pH 7.5 |
| | 2.5 ml 2% gelatin and autoclave |

2. 6. 3g Lifting filters to replicate the phage library plates

The plates were dried under a fume hood at room temperature for 30 min and cooled at 4°C in a cold room. Hybond-$N^+$ nylon membrane (Amersham) filters were placed on the phage plaque and marked with a needle to aid realignment after hybridization. The master plate was sealed and stored at 4°C. The lifted filters were placed on 3MM Whatman paper pre-wetted in denaturant (0.5 M NaOH) for 2 min and neutralized twice on 3MM Whatman paper pre-wetted in a neutralization solution (1 M Tris-HCl, pH 7.6) for 2 min. The filters were washed twice in 2 x SSC for 10 min, dried at 80°C for 2 h, and

UV-crossed.

### 2. 6. 4 Plasmid Subclone Filters

DNA fragments up to about 5 kb in length were cloned into *p*Bluescript and PCR products into *p*AMP or *p*GEM vectors. All three plasmids carry the ampicillin resistant ($amp^R$) and LacZ′ genes. The enzyme, beta-galactosidase becomes functional in the presence of the complementary LacZ′ genes from the host bacteria, *E. Coli*, and the plasmid. This enzyme breaks down lactose to glucose and galactose. When a lactose analogue, Xgal (5-bromo-4-chloro-3-indolyl-beta-D-galactopyranoside), is added to the agar with an inducer of the enzyme, IPTG (isopropyl-thiogalactoside), the cells synthesize beta-galactosidase and become blue. The cloning sites of these vectors are in the LacZ′ gene, and a DNA insert prevents beta-galactosidase production. Thus white colonies identified recombinants and blue colonies non-recombinants.

2. 6. 4a Ligation

Cosmids were digested with *Hind*III and purified using a QiaQuick column (Qiagen Inc.). *p*Bluescript vector was digested with *Hind*III and dephosphorylated with CIAP (2. 6. 2a). 500 ng of digested cosmid DNA was ligated to 25 ng of the vector using 0.5 µl T4 DNA ligase (NEB, 400 U/µl) at 16°C overnight.

PCR products amplified in a 100 µl reaction were purified by PCR-QiaQuick column. 500 ng of purified DNA were ligated to 50 ng *p*AMP or *p*GEM vectors using 0.5 µl T4 DNA ligase (NEB, 400 U/µl) at 16°C overnight.

2. 6. 4b Transformation

The ligation (10 ng DNA) was added to a tube containing 100 µl of DH5-alpha

competent cells (Gibco, BRL). The transformation took place on ice for 3 min and was stimulated by raising the temperature to 42°C for 45 sec (heat-shock). 1 ml of LB was added to the transformed cells and incubated for 1 h at 37°C on the shaker. Cells were spread onto LB- ampicillin plates (50 μg/ml) pre-treated previously with 25 μg/ml Xgal and 25 μl/ml IPTG. White colonies were picked for their recombinants and placed in a microtitre plate.

### 2. 6. 5 Cosmid End-Cloning

To construct a contig covering the whole YAC, the end of the cosmids were isolated using a cosmid-end cloning technique. 10 μl of cosmid DNA were digested with *Sac*I in 30 μl reaction buffer (1 x NEB buffer, 500 mM Spermidine and 100 μg/ml RNA) at 37°C overnight. The reaction was terminated by heat inactivation at 68°C for 10 min. Half of the digests were used to test for complete digestion. The remainder were purified and ligated as described in 2. 7. 3, transformed into DH5-alpha cells (Gibco, BRL) and plated onto LB plates containing 50 μg/ml ampicillin. Two of the colonies were picked and purified.

Cosmid end-cloning results in the ligation of vector plus flanking fragments of an insert. Cosmids with no *Sac*I sites in the vector were digested with *Sac*I. This resulted in *Sac*I-insert-*Not*I cloning site-vector-*Not*I cloning site-insert-*Sac*I fragments. These ligated at their *Sac*I cohesive sites to produce circular plasmids, thus cloning the two ends of the original insert.

The success of the 'end-cloning' was observed with the agarose gel results with *Sac*I, *Not*I , and *Sac*I/*Not*I digests, generating 1, 2 and 3 bands, respectively. The two end fragments of the insert were recovered from the

*Sac*I/*Not*I digests on 1% (w/v) SeaPlaque agarose gel (F. M. C.) as described previously (2. 3. 5).

End clones were used as probes to isolate more cosmids for walking in both directions. The isolated cosmid clones were ordered using the ICRF 'clone-to well' computing program.

## 2. 7 PROBE CONSTRUCTION USING ISOTOPES

Membrane-based methods of DNA and RNA analysis were once totally dependent on the use of isotopic probes. A non-isotopic protocol introduced by Hugues Roest Crollius (Hans Lehrach's Lab, ICRF) was used for hybridization to Southern blots or membrane lifts of cloned DNA from YACs, cosmids and plasmids. Isotopic probes, however, were preferable for hybridization to northern and human genomic blots and when screening primary library filters such as YAC, cosmid, phage, and cDNA libraries.

### 2. 7. 1 Isotopic Labeling of Double-Stranded DNA

Single copy and PCR synthesized DNA fragments were labeled by incorporating the radioactive isotope of phosphorus, $^{32}P$ using random primed DNA labeling (Baxendale *et al.*, 1991; Feinberg and Vogelstein, 1983). DNA was denatured at 100°C for 5 min and cooled in the presence of a mixture of all possible hexanucleotides. The annealed hexanucleotides were primed to synthesize new complementary DNA strands incorporating the dNTP, one of which has a labeled group. The synthesis is catalyzed by the Klenow subunit of DNA polymerase I. A T7 QuickPrimer kit (Pharmacia) provided a simple protocol to prepare the following mixture.

(1) 34.0 µl denatured DNA fragment (30 ng)
(2) 0.1 µl reagent mix containing unlabeled nucleotides (100 µM of each dATP, dGTP and dTTP)
(3) 5 µ [α–$^{32}$P]-dCTP (3000 Ci/m mol)
(4) 1 µl T7 polymerase

After an 1 h incubation at 37°C, the labeled probes were separated from unincorporated nucleotides by precipitating with 20 mM EDTA, 20 µg yeast tRNA, 2.5 M ammonium acetate and 2.5 volumes of absolute ethanol, resuspended with 100 µl 1 x TE, and then denatured at 100°C for 10 min prior to hybridization.

### 2. 7. 2 Isotopic Labeling of High Molecular weight dsDNA

Cosmid DNA and YAC DNA were labeled using the same method as described in 2. 7. 1 and prior to hybridization to target DNA, they were preassociated with the addition of competitor DNA to suppress false positives arising from repetitive sequences. For cosmid DNA, this non-specific background noise was reduced by the addition of 125-500 µg of sonicated human placental DNA (Sigma). For YAC DNA, 125-500 µg of sonicated human placental DNA (Sigma) and 50 µg of random cosmid DNA were added. Prior to hybridization, the competitor DNA was mixed with a labeled probe and denatured at 100°C for 10 min and incubated at 65°C for 3 h for pre-annealing in 5 x SSC.

### 2. 7. 3 Isotopic Labeling of Single-Stranded DNA

Labeling of single-stranded oligonucleotides was performed by Kinase-mediated 'end-labeling'. Oligonucleotides synthesized without a phosphate group at the 5'-termini were labeled by transferring the γ-$^{32}$P from [γ-$^{32}$P]dATP

by T4 polynucleotide kinase. The reaction mixture, 1 μg of oligonucleotide, 20 μCi [γ-$^{32}$P]dATP, and 3 U of T4 polynucleotide kinase (NEB) was incubated at 37°C for 2 h, and used directly for hybridization without ethanol precipitation.

## 2. 8 PROBE CONSTRUCTION WITH NON-ISOTOPES

Non-isotopic labeling involved the chemical coupling of a nucleotide precursor and a reporter molecule, Digoxigenin. Antibody bound to the reporter group and carrying fluorescent dye was detected by a fluorimetric assay.

### 2. 8. 1 Non-Isotopic Labeling of Double-Stranded DNA

150 ng of double-stranded DNA was labeled for filter hybridization using random-prime labeling with Digoxigenin-11-dUTP (DIG-11-dUTP) (Boehringer-Mannheim) following the supplier's instructions.

### 2. 8. 2 Non-Isotopic Labeling of High Molecular Weight dsDNA

High molecular weight dsDNA containing repetitive elements was pre-annealed with 125-500 μg of sonicated human placental DNA (Sigma) and 50 μg cosmid DNA as described in 2. 7. 2.

### 2. 8. 3 Non-Isotopic Labeling via PCR

DNA fragments with known primers were labeled by the incorporation of DIG-11-dUTP via PCR. Standard PCR was carried out with an addition of 15 mM DIG-11-dUTP followed the protocol by Hugues Roest Crollius (Hans Lehrach's Lab, ICRF).

### 2. 8. 4 Non-Isotopic Labeling of Single-Stranded DNA

Oligonucleotides were labeled with terminal deoxynucleotidyl transferase

added by tailing the probe with a mixture of nucleotides and DIG-ddUTP. (Boehringer-Mannheim).

## 2. 9 HYBRIDIZATION, POST-HYBRIDIZATION WASH & DETECTION

The hybridization, post-hybridization wash, and detection were performed under standard conditions with slight variations in buffer components, hybridization time, and temperature differences, to suit a particular probe, for example, whether it is labeled isotopically or non-isotopically, unique or microsatellite repeat sequence, or long or short.

Prior to hybridization, all filters were submerged for 4-5 h in hybridization buffer (pre-hybridization). Filters embedded with cosmid, cDNA, and YAC DNA were prehybridized overnight with 100 μg/ml denatured, sheared placenta total human DNA (Sigma) and then hybridized overnight to mask repeated DNA sequences.

### 2. 9. 1 Isotopic Probes

#### 2. 9. 1a Hybridization

Isotopically labeled oligonucleotides were hybridized to filters in Church buffer (Church and Gilbert, 1984) at 65°C. Probes with repeated DNA sequences were hybridized at 65°C overnight in a buffer containing 0.9 M NaCl, 1% (w/v) SDS and 40 μg/ml tRNA. Hybridization using DNA fragments as probes was performed in a formamide-based hybridization buffer at 42°C overnight. Formamide lowers the hybridization temperature and enhances probe hybridization by acting as a chemical denaturant. Two polar groups (amino and carbonyl molecules) of formamide competed for

hydrogen-bond formation with animo and carbonyl groups of bases, leaving the complementary strands separated.

The ratio of a labeled DNA and a buffer was adjusted to a final concentration of 1.0 x $10^6$ c.p.m./ml.

| | |
|---|---|
| Church buffer | 5 x SSC<br>5 x Denhardts solution<br>0.1% (w/v) SDS |
| Hybridization buffer | 50% (v/v) formamide<br>4 x SSC<br>50 mM Na phosphate buffer, pH 7.2<br>1 mM EDTA<br>10% (w/v) Pharmacia dextran sulphate<br>1% (w/v) SDS<br>200 μg/ml Sigma herring sperm DNA<br>25 μg/ml yeast tRNA<br>10 x Denhardt's solution |
| Denhardt's solution | 20% (w/v) SBSA Fraction V<br>2 % (w/v) Ficoll 400<br>2% (w/v) polyvinylpyrollidone |
| Sodium phosphate buffer<br>(1 M) | 134 g $Na_2HPO_4.7H_2O$<br>Adjust a pH to 7.2 with about 4 ml $H_3PO_4$ |
| 20 x SSC<br>(10 litre) | 1753 g NaCl<br>882 g sodium citrate pH to 7.0 with NaOH |

2. 9. 1b Post-Hybridization-Wash

Filters were washed to remove non-specific hybridization probes using a series of solutions with a gradual increase in stringency. The stringency was increased by either increasing the temperature or decreasing the salt (SSC) concentration. In general, filters were washed in 1 x SSC and 1% (w/v) SDS twice at 42°C and then twice at 65°C for 30 min in 0.1 x SSC and 0.1% (w/v) SDS. Filters with oligo-hybridization were gently washed once with 2 x SSC/0.1% (w/v) SDS at 50°C for 10 min. For hybridization with microsatellite repeat probes, filters were stringently washed twice in 0.5 x

SSC/0.1% (w/v) SDS at 65°C for 20 min.

2. 9. 1c Detection

Washed and air-dried filters were wrapped between sheets of Saran Wrap, placed into a cassette, followed by a sheet of autoradiographic X-ray film (Kodak, X-OMAT LS) and then an intensifying-screen. The film was exposed at -70°C for 1-7 days, and developed using an X-OGRAPH X150 (Imaging system). For microsatellite repeate DNA probes, the film was exposed at room temperature for not longer than 2 h.

2. 9. 1d Re-Use of Filters

Filters were 'stripped' for re-use after each hybridization with 0.1 M NaOH, 0.1% (w/v) SDS and 0.2 mM EDTA at 60°C for 20 min, washed with 0.1% (w/v) SDS and 0.2 mM EDTA at room temperature twice, rinsed in 4 x SSC at room temperature, dried, and re-probed.

**2. 9. 2 Non-Isotopic Probes**

2. 9. 2a Hybridization

Hybridization was performed in modified Church buffer ((5% (w/v) SDS, 0.5 M Sodium phosphate)) (Church *et al.*, 1984) at 65°C overnight.

2. 9. 2b Post-Hybridization Wash

The filters were washed twice in 2 x SSC and 0.1% (w/v) SDS at room temperature for 10 min and then once in 0.1 x SSC and 0.1% (w/v) SDS at 65°C for 20 min.

2. 9. 2 c Detection

Prior to detection, the filters were incubated in 5% fat-free milk, 1 x PBS for 5

min, which helped reduce background noise. To the same, but with a fresh milk solution, the 0.1 U/ ml anti-DIG antibodies were added and incubated with gentle agitation for conjugation to the DIG-dUTP. The filters were then washed in 1 x PBS four times for 10 min each, submerged in 0.1 M diethanolamine, pH 9.0, 1 mM $MgCl_2$ for 10 min, and drained on 3MM paper (Whatman). 0.6 mg/ml Attophos (Europa) in 2.4 diethanolamine was applied using a sponge pad for detection, visualized, and photographed under long wave UV (365 nm) using Eagle Eye II camera (Stratagene) 2 h after Attophos application.

2. 9. 2d 'Filter-Strip' and Re-Use

Filters were 'stripped' after each hybridization with 0.4 M NaOH, 0.1% (w/v) SDS at 50°C for 2-10 min and washed with 2 x SSC, 0.1% (w/v) SDS, 0.1 M Tris-HCl, pH 7.4 at room temperature twice and then incubated in Church buffer at 65°C overnight before re-use.

## 2. 10 EXON-TRAPPING/EXON-AMPLIFICATION

RNA splicing involves excision of introns and fusion of exons at the exon/intron boundaries, which *in vivo* are recognized by splicesomes. Using RNA splicing machinery, the Exon Amplification technique traps expressed sequences (exons), and recovers them from cDNA copies by PCR.

An expression plasmid vector, *p*SPL3, is a shuttle vector which can be propagated in E. coli and in monkey COS cells using an SV40 origin of replication (Buckler *et al.*, 1991; Church *et al.*, 1994). This vector contains a multiple cloning site between two minigenes that had been used to design

specific primers.

Five stages of exon trapping assay were followed: (1) subclone a cloned genomic DNA fragment into *p*SPL3 vector, (2) transfect the recombinant DNA into COS cells, (3) extract RNA transcribed from SV40 promoter using the COS cell RNA splicing mechanism, and (4) convert the RNA to cDNA using reverse transcriptase.

### 2. 10. 1 Transformation

Two cosmids, C9A and C7B, which span the site of the translocation breakpoint on chromosome 8 were digested with *Bgl*II and *Bam*HI, ligated into the *Bam*HI site of the *p*SPL3 exon-trapping vector, transformed into DH5-alpha).

### 2. 10. 2 Transfection

Recombinant DNA was Qiagen-prepared (Qiagen Inc.) and transfected by electroporation into COS7 cells using Gene Pulser apparatus (Biorad), following the manufacturer instructions.

### 2. 10. 3 RNA Extraction and cDNA PCR

After 72 h of growth the cells were harvested and RNA extracted using a TRIzol (Gibco, BRL) kit. The first strand of the cDNA was synthesized using the SA2 primer and reverse transcriptase (NEB). The first round of PCR amplification of the cDNA was performed with the SA2 and SD6 primers.

### 2. 10. 4 Isolation of Exons

PCR products larger than 110 bp vector product were separated in 1.3% (w/v) SeaPlaque (F. M. C.) low melting agarose gel, excised, and purified using the QiaQuick PCR Purification kit (Qiagen Inc.). The products were

PCR-amplified again using the nested primers, SA4 and SD2. The PCR products from the second round of amplification were sized on gels, excised, purified, and then ligated into *p*AMP1 vector (Clone Amp System, BRL) and sequenced.

## 2. 11 FLUORESCENCE-BASED CYCLE SEQUENCING

### 2. 11. 1 Automated DNA Cycle Sequencing

PCR fragments and dsDNA templates were purified for sequence analysis using PCR Preparation DNA Purification Resin (Promega) and the Magic miniprep kit (Promega), respectively. Sequence analysis was performed with Taq DyeDeoxy™ Terminator Cycle Sequencing kits on the Applied Biosystems model 373A DNA sequencing system. AmpliTaq DNA polymerase FS lacked the 5′ to 3′ exonuclease (5′ exo-) activity, but still possessed the 3′ to 5′ exonuclease (3′ exo+) activity for proof reading. Thus, only small amounts of dideoxynucleotides are required and removal of unincorporated dideoxynucleotides by phenol-chloroform extraction was not necessary. This produced much clearer results because phenol extraction can reduce the quality of the signal. The reaction mix further contained (1) A-dye (green), C-dye (blue), G-dye (black) and T-dye (red) terminators, (2) dATP, dCTP, dTTP, and the deoxyinosine triphosphate (dITP), the last used in place of dGTP so as to minimize band compressions, and (3) Tris-HCL, $MgCl_2$ and thermal stable pyrophosphatase.

The fluorescently labeled dye terminators were incorporated along the template, which was either 800 ng of dsDNA or 1μg of PCR products, using an

MJ PCR amplification system (MJ PTC-225 Thermal Cycler) with 25 cycles of denaturation (30 sec at 96°C), annealing (15 sec at 50°C) and extension (4 min at 60°C) proceeded by 3 min of denaturation at 96°C and followed by a final extension at 60°C for 10 min.

The small amounts of unincorporated dideoxynucleotides were removed by ethanol precipitation. The pellets were resuspended in a loading buffer (5 parts deionized formamide to 1 part 25 mM EDTA pH 8.0, and Blue dextran dye). Samples were denatured, loaded onto a 4.75% acrylamide gel, and subjected to electrophoresis on an ABI 373A DNA Sequencer (Applied Biosystems) following the ABI protocols. The dideoxynucleotides incorporated dye-terminators were monitored and recorded fluorescence signal as the DNA passed through a fixed point in the gel.

| | |
|---|---|
| 4.75% acrylamide gel | 9.5 ml Concentrate<br>35.5 Diluent<br>5 ml Buffer<br>400 µl 10% (w/v) ammonium persulfate<br>20 µl TEMED |

### 2. 11. 2 Sequence Analysis

The sequence data obtained were edited and aligned using Factura and Sequence Navigator software (Applied Biosystems), respectively. The edited sequences were analyzed by (1) the Gene Recognition and Analysis Internet Link (GRAIL) (Uberbacher and Mural, 1991), which uses a multi-sensory-neural network approach to find genes in DNA sequences and by (2) the Basic Local Alignment Search Tool-X (BLASTX) (Gish and States, 1993), which translates nucleotide sequences in all six reading frames and searches a nonredundant protein database. Nucleotide homologies were

searched using BLASTN. All BLAST searches were done by the NCBI server and GRAIL analyses by the ORNL server. Contigs were constructed by sequence-walking. Walking was achieved by generating primers near each end of existing sequence so as to extend the contig outwards by sequencing. Overlapping end sequences were identified using the Genome Assembly Program (GAP) (Bonfield *et al.*, 1995) or BESTFIT program of the GCG software package (Devereux *et al.*, 1984).

## 2. 12 PCR-BASED LIBRARY SCREENING

PCR screening was available for an ICI-YAC library and a Clontech cDNA library.

### 2. 12. 1 PCR YAC Library Screening

The ICI-YAC library was screened by PCR to isolate chromosome-8 specific YACs that spanned the translocation breakpoint. This library contained 35,000 YAC clones which were individually deposited in wells of 365 microtiter dishes. Each dish, with 12 columns and 8 rows of wells, contained 96 YAC clones. The dishes were arranged into 40 stacks of 9 each. The screening was performed by PCR assays in four stages. The primary, secondary, tertiary and quaternary stages determined respectively, which stack, plate, column and row contained the desired YAC. In the primary screening 9 plates containing 864 clones from each stack were pooled, thus requiring 40 PCR assays. Positive stacks in the primary screening entered into the secondary screening. Each plate of a stack was pooled, thus requiring 9 PCR assays per stack. Positive plates from the secondary screening entered

into the tertiary stage. Each column of a plate was pooled and 12 PCR assays per plate were performed. Positives columns in the tertiary screening entered into the quaternary screening. Each of the 8 rows of a column were assayed by PCR to locate the exact well of the library containing the desired YAC.

### 2. 12. 2 PCR cDNA Library Screening

Clontech cDNA libraries were available as concentrated cDNA cloned in phage from kidney, adult liver, placenta, fetal lung, testicle and temporal cortex. This was diluted 1/50 before the PCR screening to achieve a titre between $5 \times 10^8/\mu l$ to $1.4 \times 10^8/\mu l$.

## 2. 13 POLYMORPHISM ANALYSIS

PCR products were analyzed for heteroduplexes using conformation-sensitive gel electrophoresis (CSGE). Detected mismatches were then sequenced.

### 2. 13. 1 Conformation-Sensitive Gel Electrophoresis (CSGE)

When a single-base mismatch (heteroduplex) in a double-stranded PCR product is placed in a mild denaturant, its DNA conformation bends at the mismatched site (Bhattacharyya and Liley, 1989). Such heteroduplexes can be observed in gel-electrophoresis through differences in their migration in comparison with homoduplexes. Optimal separation is achieved using the CSGE protocol (Ganguly *et al.*, 1993). A 10% polyacrylamide gel was polymerized in 10% ethylene glycol, 15% (v/v) formamide and Tris-taurine buffer, and solidified between standard sequencing gel glass plates 1 mm thick. 12µl PCR product (250-650 bp) was denatured at 98°C for 5 min, incubated at 68°C for 1 h to allow formation of heteroduplexes, cooled to

room temperature, and separated electrophoretically at 500 V for 16 h. The gel was stained with 1 mg/ml ethidium bromide for 20 min, rinsed in $dH_2O$ for 10 min, visualized under short wave (302 nm) UV transillumination and photographed using the Eagle Eye II camera (Stratagene).

| | |
|---|---|
| Tris-taurine buffer | 89 mM Tris-HCl<br>7 mM Taurine<br>0.1 mM EDTA |

### 2. 13. 2 Sequence Confirmation

Positive results from the CSGE analysis were further investigated at the nucleotide level by sequencing (2. 11).

## 2. 14 PANELS FOR X-INACTIVATION STUDY

The X-inactivation assays were performed by Dr. A. P. Miller, according to following protocol.

Cell lines were grown in alpha-MEM media supplemented with 10% fetal calf serum, penicillin and streptomycin, and L-glutamine. Hybrids were maintained under the appropriate selection: AHA-11aB1 was grown in HAT at 37°C, t60-12 was grown in HAT at 39°C. Cells were harvested from a single t175 flask just prior to confluence for RNA isolation (Brown *et al.*, 1990). RNA was extracted from cell lines using RNAzol (Cinna/Biotecx) using two sequential extractions of the RNA to completely eliminate DNA contamination. Reverse transcription reactions were done with and without reverse transcriptase to control for DNA contamination within the RNA sample. Approximately 5 μg of total RNA was reversed transcribed in PCR buffer (Gibco, BRL) using random hexamers to prime the RNA strand. The

resulting cDNA samples were diluted 1:2 and 1 μl was subjected to PCR with gene specific primers. PCR conditions consisted of an initial heat denaturation at 95°C for 2 min followed by 35 cycles of 94°C for 15 s; 55°C for 15 s; 72°C for 40 s in a 50 μl reaction with 1.5 mM $MgCl_2$. PCR products were size fractionated on 2.0% (w/v) agarose gels, visualized with EtBr and photographed.

## 2. 15 CYTOGENETIC ANALYSIS

X inactivation and ML's Karyotype were examined by Dr. V. J. Buckle, MRC, IMM, Oxford using bromodeoxyuridine (BrdU) and DAPI analyses, respectively.

### 2. 15. 1 X-inactivation

BrdU, a thymidine analogue, is readily incorporated into chromosomes during DNA synthesis and can be detected by staining with the bisbenzimide dye, Hoescht 33258. This dye binds to the AT regions and gives off a bright fluorescence. When BrdU is incorporated in AT regions, however, the Hoescht fluorescence is quenched and the regions appear darkly stained. This allows detection of the difference between early and late replication under UV fluorescence.

To test for X inactivation, BrdU is added late in the cell cycle. Chromosomes that replicate early incorporate thymidine and stain brightly. Chromosomes that replicate late incorporate BrdU and stain darkly. In general, the inactivated X chromosome is less accessible to the replication machinery due to its condensed conformation. Its DNA therefore begins to

replicate late in S phase in comparison with the normal X and stains more darkly of the two in the BrdU analysis. The inactive X replicating late incorporates more BrdU and is more darkly stained than an active X chromosome that replicates early (Camargo and Cervenka, 1982).

### 2. 15. 2 Karyotype

The fluorescent dye, DAPI (4', 6-diamidiono-2-phenylindole), preferentially binds to AT-rich DNA and stains darkly chromosomal segments rich in AT-sequences. This creates banded chromosomes (a karyogram), which can be viewed by UV fluorescence. The DAPI staining located the translocation breakpoint at finer resolution.

## 2. 16 FLUORESCENCE *IN SITU* HYBRIDIZATION (FISH).

YACs and cosmid DNA probes with fluorescent dyes were hybridized to their complementary sequences (target DNA) on spreads of metaphase chromosomes prepared on slides. YACs and cosmid DNA were labeled with either biotin and digoxigenin as reporter molecules and indirectly detected by reporter-binding-molecules conjugated with fluorescent dye, avidin and antiDIG, respectively. Signal amplification was performed by repeating the primary detection, *e. g.* laying on top of the existing one, commonly, avidin FITC (fluorescein isothiocyanate), biotinylated anti avidin-avidin FITC, and again avidin FITC (Fig. 2. 2).

For dual detection, one probe was labeled with biotin and detected with avidin TR (texas red), biotinylated goat anti avidin, and avidin TR, with the other probe with digoxigenin conjugated with mouse anti DIG, anti mouse

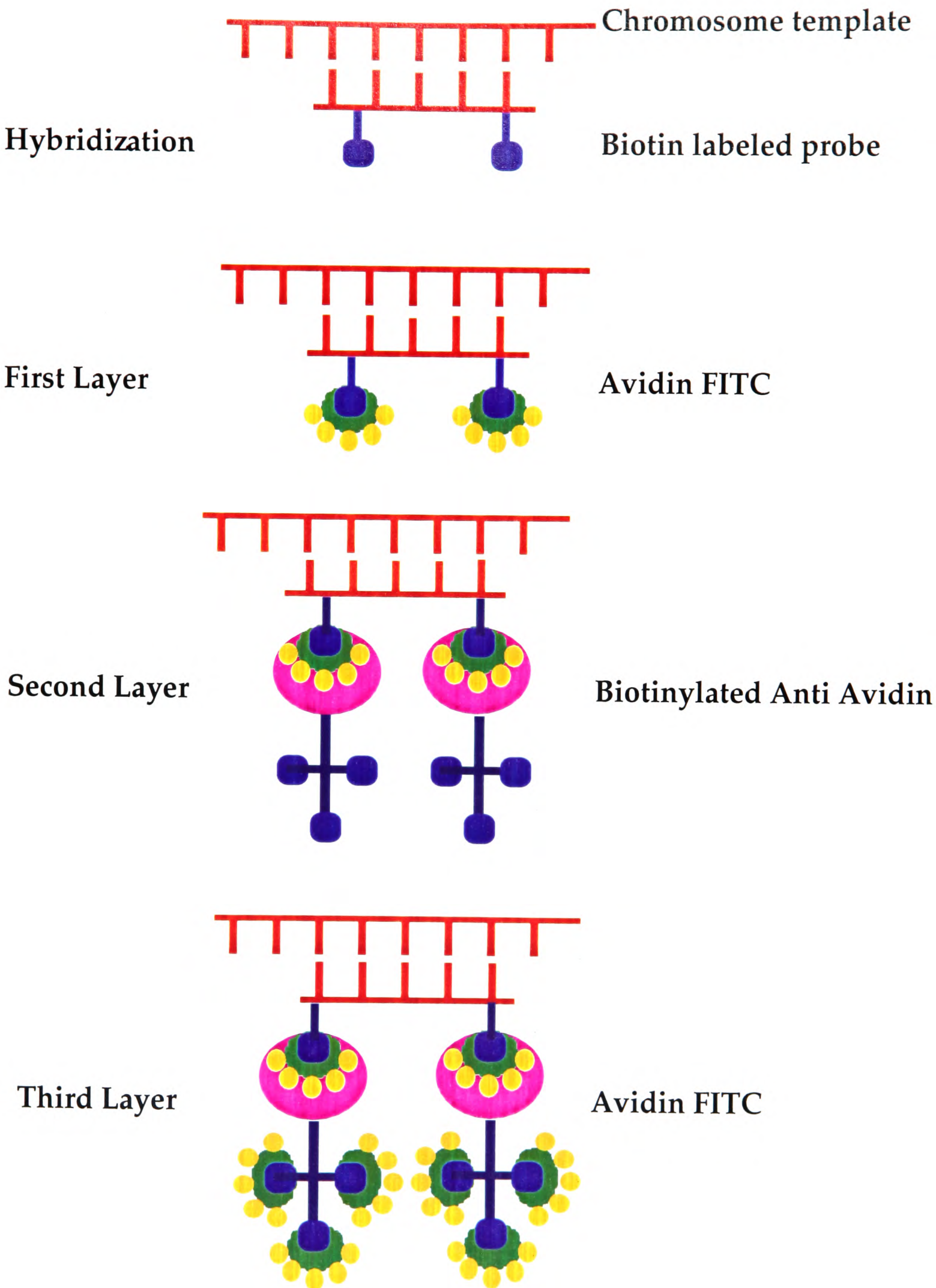



Fig. 2. 2. Signal amplification of probe-hybrids. Biotin labeled probes were conjugated with Avidin FITC, itself subsequently conjugated by biotinylated Anti-Avidin. The signals were amplified in the third layer of Avidin FITC.

(rabbit), and treated with tyramide signal amplification.

Probes and their target DNA were denatured both chemically using formamide and physically using heat, hybridized, and visualized by counter stains.

### 2. 16. 1 Probe Preparation by Nick Translation

YAC DNA prepared in solution using 4.5 M guanidium hydrochloride extraction (2. 1. 3a) was directly used as a probe. YAC DNA blocks were purified using a Double GeneClean Kit (Bioline 101 Inc.) prior to FISH.

YAC DNA and cosmid DNA were fragmented into an average length of about 300 bp and labeled with biotin-14-dATP using nick translation (BioNick™ Labeling System, Gibco, BRL). This length maximized the penetration of the probe and minimized the background signals on the slide. The reaction was at 16°C for 2 h and took place in a final volume of 50 µl containing 1 µg of the probe, 1 µl of 1 mM Biotin, 5 µl of 10 x nick translation buffer (0.5 M Tris-HCl, pH 7.5, 50 mM $MgCl_2$ and 0.5 mg/ml nuclease-free BSA), 5 µl 100 mM DTT, 5 µl dNTP mix (0.5 mM each of dATP, dCTP, dGTP and 0.1 mM dTTP), 3 µl 3.5 U/ml DNA polymerase I, 3 µl DNase (1 ng/µl). The reaction was terminated with 10 µl 0.25 M EDTA or by 'snap freezing' on dry-ice for 30 min. The fragment size was checked by running 100 ng of the labeled probe on 2% (w/v) gel with a ϕx174RF/*Hae*III marker. DNA which fragmented correctly between 100 bp and 500 bp was purified using a Select-B column (5'-3' Inc), precipitated with 0.1 volume of NaOAc, pH 5.2 and 2 volumes of ethanol, carriers of 50 µg of tRNA and 50 µg of salmon sperm DNA (500 bp), at 13,000 rpm for 20 min. The pellet was resuspended in $dH_2O$

to a final concentration of 50 ng/μl for YACs and 20 ng/μl for cosmid, and stored at -20°C. 300 ng/μl YAC probes and 80 ng/μl for cosmid probes, denatured at 90°C for 10 min, were used for each hybridization slide (20 x 20 mm) (Table 2. 4).

### 2. 16. 2 Slide Preparation

Metaphase and interphase spreads prepared from phytohaemagglutinin (PHA) stimulated lymphocytes, 0.5 ml of whole blood, 10 ml of RPMI culture medium, and PHA were incubated for 72 h at 37°C. Cells were harvested in

| | Probes | | |
|---|---|---|---|
| Conditions | Cosmid | YAC | Centromere |
| Probe DNA concentration per slide (10 μl hybridization mix) | 80 ng | 300 ng | 15 ng |
| Cot-1 competitor DNA concentration per slide | 2.5 μg | 7.5 μg | - |
| Pre-annealing time (min) | 15 min | 15 min | - |
| Hybridization time | O/N | 4 days | O/N |

Table 2. 4 Hybridization conditions for three kinds of probes. The concentrations for each probe were prepared for 10 μl of hybridization mix using a coverslip 22 mm x 22 mm in size, 5 μl for a coverslip of 18 mm x 12 mm, and 1 μl for a circular coverslip 6 mm in diameter.

metaphase by synchronizing the lymphocyte cell cultures at the S phase using BrdU (a final concentration 200 μg/ml of medium) (Buckle and Rack, 1993). BrdU was used since the more commonly used thymidine was toxic to these lymphocytes. The cells were spread on a slide in fixative of three parts analar methanol to one part glacial acetic acid at 4°C and stored at -20°C in a sealed container with desiccate. Prior to hybridization the slides were treated with

100 µl RNase (100 µg/ml) at 37°C for 1 h, washed in 2 x SSC, fixed in PBS/50 mM $MgCl_2$ for 5 min and PBS/50 mM $MgCl_2$/1% (v/v) formaldehyde for 10 min, rinsed in PBS and then dehydrated through an ethanol series (30 sec in each of 10, 50, 70, 90 and 100% (v/v) ethanol). The chromosomes were denatured for 5 min at 72°C in 70% (v/v) formamide/0.1 mM EDTA in 2 x SSC. The slides were washed twice at 4°C in 2 x SSC and then dehydrated through an ethanol series as described above.

Metaphase and interphase spreads were also prepared from an EBV transformed lymphoblastoid cell line of the patient. Lymphoblastoid cells were grown until a dividing population was obtained with around 5 x $10^6$ cells per 10 ml of RPMI culture medium. Thymidine (a final concentration of 300 µg/ml) was added to synchronize the lymphocyte cell culture. The rest of the procedure was the same as described in 2. 15. 1.

Interspersed repeat sequences of whole YAC and cosmid DNA were suppressed prior to hybridization by Cot-1 DNA (Gibco, BRL) (7.5 µg Cot-1 DNA for 300 ng YAC probe and 2.5 µg for 80 ng cosmid probe) (Table 2. 4).

| | |
|---|---|
| Complete RPMI | 100 ml RPMI 1640<br>1 ml Pen/Strep (Gibco stock = 5,000 U/ml Pen, 5,000 µg/ml Strep)<br>1 ml L-glutamine (100 x Gibco stock = 200 mM) |
| RPMI + 20 % FCS + PHA | 80 ml complete RPMI<br>20 ml fetal calf serum (FCS)<br>2 ml phytohaemagglutinin, M form (PHA) |

### 2. 16. 3 Hybridization and Post-Hybridisation Wash

The Cot-1 competitor and 300 ng of the biotin labeled YAC DNA or 80 ng of cosmid per slide (20 x 20 mm) were dried down together in a vacuum pump

and resuspended in a final concentration of hybridization mix of 50% (v/v) formamide/1 x SSC and 10% (w/v) dextran sulphate. Formamide present in the hybridization mixture aided in dissolving probe and maintained the morphology of the chromosomes by lowering the DNA/DNA association temperature, from 65°C to 42°C. The hybridization time was 4 nights for YAC probes and overnight for cosmid probes as summarized in the Table 2. 4.

### 2. 16. 4 Detection

The signals were detected by applying in layers avidin FITC (5 μg/ml), biotinylated goat anti-avidin (5 μg/ml) and avidin FITC (5 μg/ml). The slides were mounted in antifade (Vector Laboratories) containing propidium iodide (PI) and DAPI. The signals were captured as digital images with an MRC 600 scanning confocal microscope.

For the analysis of two probes with dual detection, one probe was labeled with biotin and other with digoxigenin, DIG-11-dUTP (Boehringer-Mannheim). Biotin labeled probes were detected with Avidin-Texas Red (5 μg/ml) (Vector Laboratories), biotinylated goat Anti-Avidin (5 μg/ml) and then Avidin-Texas Red (5 μg/ml) and counterstained with PI and DAPI. The DIG labeled probes were detected by layers of 0.2 μg/ml of anti DIG (mouse), 1 mg/ml) of anti-mouse FITC (goat), and counterstained with TO-PRO-3 (Molecular Probes) and DAPI. Weak Texas Red signals were further amplified using a TSA-Indirect kit (Tyramide Signal Amplification, Du Pont NEN). Two-colour signals were collected independently via an MRC 1024 scanning confocal microscope (Biorad) and collectively visualized using Lasersharp Software (Biorad).

## 2.17 CHEMICAL AND MATERIAL SUPPLIERS

| | |
|---|---|
| Amersham | Amersham International plc. Amersham Place, Little Chalfont, Amersham, Buckinghamshire, HP7 9NA. |
| Beckman | Beckman Instruments Inc., 1117 California Avenue, Palo Alto, CA 94304, USA (UK supplier: Beckman RIIC Ltd., Progress Road, Sands Ind. Estate, High Wycombe, Bucks. |
| Bio 101 | Bio 101 Inc., P. O. Box 2284, La Jolla, CA 29038-2284, USA |
| BDH | BDH Chemicals Ltd., Merck Ltd., Merck House, Broom Road, Poole, Dorset, BH15 1TD. |
| Bio-Rad | Bio-Rad Laboratories Ltd., Bio-Rad House, Maylands Avenue, Hemel Hempstead, Hertfordshire, HP2 7TD. |
| Boehringer | Boehringer Mannheim GmbH, Biochemica, D6800 Mannheim 31, Germany. (UK supplier: BCL Boehringer, Mannheim House, Bell Lane, Lewes, East Sussex.) |
| Clontech | Clontech Laboratories, 4030 Fabian Way, Palo Alto, CA 94303, USA. |
| Difco | Difco Laboratories Ltd., P. O. Box 14B, Central Avenue, East Molesey, Surrey, KT8 0SE. |
| Du Pont | Du Pont Instruments, Peck's Lane, Newton, CT 06470, USA. (UK supplier: Diagnostics and Biotechnology Systems, Wedgwood Way, Stevenage, Hertfordshire, SG1 4QN. |
| Eppendorf | Eppendorf Gerätebau, Nether & Hinz GmbH, Postf 65 06 70, 2000 Hamburg 65, Germany. (UK supplier: Anderman & Co. Ltd., 145 London Road, Kingston-upon-Thames, Surrey. |
| Falcon | Becton Dickinson, 490B, Lakeside Drive, Sunnyvale, CA 94086, USA. (UK supplier:Between Towns Road, Cowley, Oxford.) |
| Fluka | Fluka AG, CH-9470 Buchs, Switzerland. (UK supplier: The Old Brickyar, New Road, Gillingham, Dorset, SP8 4JL.) |
| FMC Corp | FMC Bio Products, 5 Maple Street, Rockland, ME 04841, USA. (UK supplier: flowgen Instruments Ltd., Broad Oak Enterprise Village, Broad Oak Road, Sittingbourne, Kent, ME9 8AQ.) |
| Fuji | Fuji Medical Systems, 90 Viaduct Road, Stanford, CT 06907, USA. Fuji Photofilm Co., 26-30 Nishizabu 2-chome, Minato-ku, Tokyo 106, Japan. |
| Gibco-BRL | Bethesda Research Laboratories Inc., P. O. Box 6009, |

| | |
|---|---|
| | Gaithersburg, MD 20877, USA. (UK supplier: Life Technologies Ltd., Unit 4, Cowley Mill Trading Estate, Longbridge Way, Uxbridge, UB8 2YG.) |
| IBI | International Biotechnologies Inc., P. O. Box 9558, 25 Science Park, New Haven, CT 06535, USA. (UK supplier: 36 Clifton Road, Cambridge, CB1 4ZR.) |
| ICN | ICN Biochemicals, division of ICN Biomedicals Inc., Cleveland, OH 44128, USA. |
| Kodak | Eastman Kodak Company, Rochester, NY 14650, USA. (UK supplier: Kodak Ltd., P. O. Box 33, Swallowdale Lane, Hemel Hempstead, Herts.) |
| Life Sciences | Life Sciences Inc., 2900 72nd Street North, St. Petersburg, FL 33710, USA. (UK supplier: Anglian Biotechnology Ltd. And Northumbria Biological Ltd.) |
| NBS | New Brunswick Scientific (UK) Ltd., Edison House, 163 Dixons Hill Road, North Mymms, Hatfield, AL9 7JE. |
| NEB | New England Biolabs In., 32 Tozer Road, Beverley, MA 01915, USA. (UK supplier: CP Laboratories, P. O. Box 22, Bishiop's Stortford, Herts, CM23 3 DX.) |
| Perkin-Elmer | Perkin-Elmer Cetus, 761 Main Avenue, Norwalk, CT 06859, USA. |
| Pharmcia | Pharmacia LKB, S-75182 Uppsala, Sweden. (UK supplier: Pharmacia Ltd., Pharmacia House, Midsummer Boulevard, Central Milton Keynes, Bucks.) |
| Promega | Promega Biotech, 2800 S. Fish Hatchery Road, Madison WI 53711, USA (UK supplier: P and S biochemicals, 38 Queensland Street Liverpool L7 3JG.) |
| Qiagen | Qiagen Inc. 9259 Eton Avenue, Chatsworth, CA 91311, USA. |
| Sigma | Sigma Chemical Company Ltd., Fancy Road, Poole, Dorset, BH17 7NH. |
| Stratagene | Stratagene Ltd., 11099 North Torrey Pines Road, La Jolla, CA 92037, USA. (UK supplier: 140 Cambridge Innovation Centre, Cambridge Science Park/Milton Road, Cambridge, CB 4 4GF.) |
| US Biochemical | United States Biochemical Corporation, P. O. Box 22400, Cleveland, OH 44122, USA. (UK supplier: Cambridge BioScience, 25 Signet Court, Newmarket Road, Cambridge, CB5 8LA.) |
| Whatman | Whatman Ltd., Springfield Mill, Maidstone, Kent ME14 2LE. |

# CHAPTER III

## IDENTIFICATION OF THE TRANSLOCATION BREAKPOINT

The results of this study are divided into four chapters. In this chapter, the isolation of a junction fragment containing the translocation breakpoint is presented. Chapter IV presents the identification of the Gastrin Releasing Peptide Receptor gene on chromosome X in relation to the translocation breakpoint. Chapter V covers the physical mapping of chromosome 8, initiated by sequence walking from the junction fragments, and in Chapter VI, the results are presented for fine mapping revealing the relation between the translocation breakpoint and the Syndecan-2 gene on chromosome 8.

### 3.1 CYTOGENETIC ANALYSIS

The project began by confirming the status of the X-inactivation and the location of the translocation breakpoint. Examination of patient metaphase cells, using the BrdU and DAPI analyses by Dr. V. J. Buckle, verified that the normal X chromosome is inactivated and the translocation occurred reciprocally between Xp22.13 and 8q22.1.

#### 3.1.1 X-Inactivation

The patient lymphoblastoid cells were grown in the presence of BrdU supplied late in the cell cycle and stained with dye (2. 14. 1). Her normal X chromosome was darkly stained (Fig. 3. 1), indicating late replication due to inactivation. The der X chromosome fluoresced brightly, indicating early

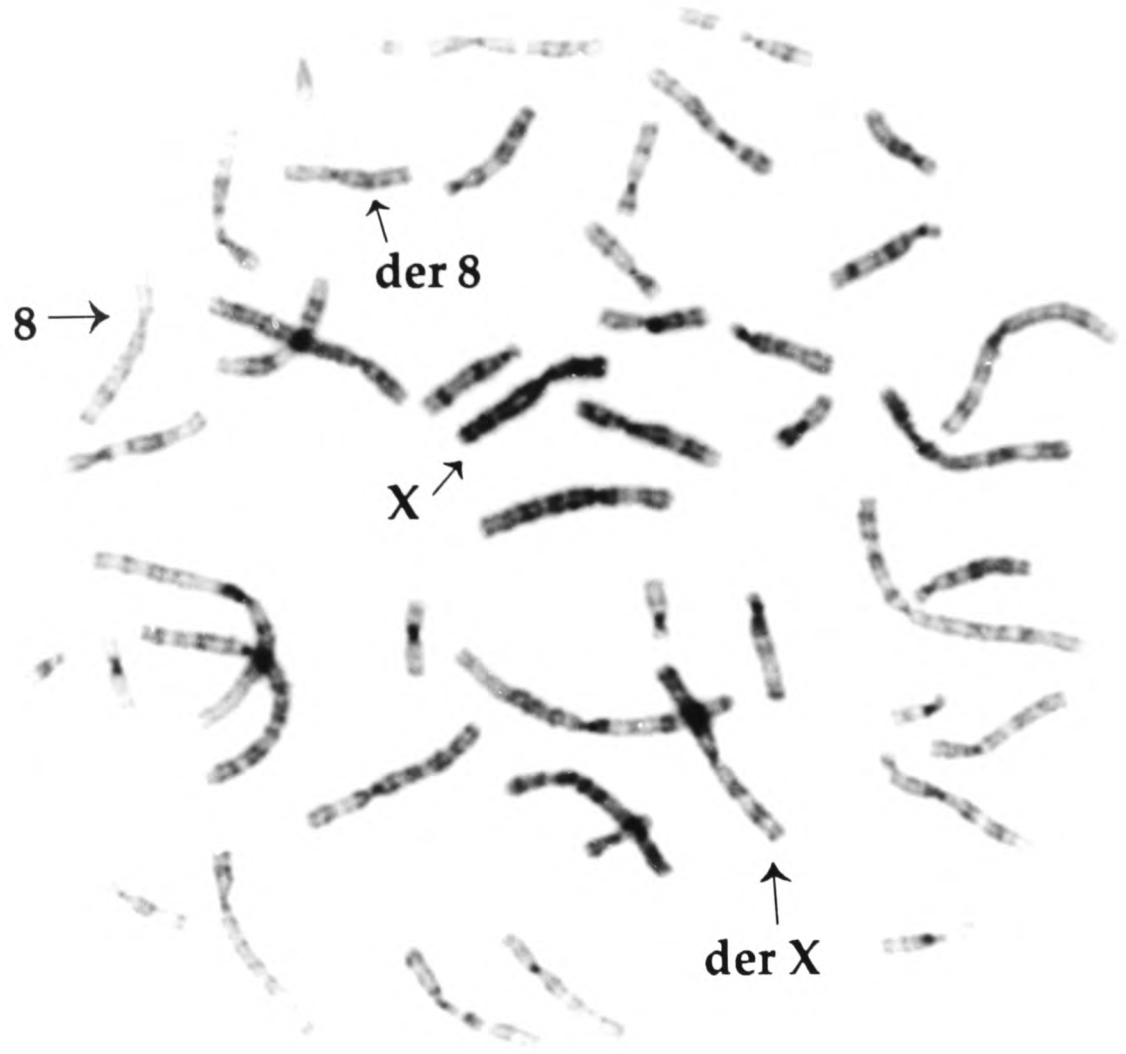



Fig. 3. 1. ML's karyotype is 46,X,t(X;8)(p22.13;q22.1). Inactivation of the patient's normal X chromosome was examined by incorporation of bromodeoxyuridine during culture. The darkly stained normal X chromosome confirmed late replication and indicated an inactive state.

replication and its status as the active X chromosome.

### 3. 1. 2 Karyotype

High-resolution banding using DAPI staining (2. 14. 2) resolved ML's translocation breakpoint to a sub-sub band of Xp22.13 and a sub-band of 8q22.1. The BrdU and DAPI techniques determined her karyotype as 46,X,t(X;8)(p22.13;q22.1), with an inactivated normal X chromosome, an activated X chromosome with the translocated 8q22-qter (derivative X or der(X)), an activated normal chromosome 8, an activated chromosome 8 with the translocated Xp22-pter (derivative 8 or der(8)) (Fig. 3. 2) and a normal replication pattern of the remaining genome.

## 3. 2 FLANKING MARKERS AND YAC ISOLATION

FISH was used to identify YACs containing the translocation breakpoint.

### 3. 2. 1 FISH Probes

Fragments containing the translocation breakpoint were first identified from the X chromosome. The analysis began with the X chromosome because 9 markers (DXS9, GLYR, DXS207, DXS197, DXS43, DXS257, DXS365, DXS274 and DXS41) (Fig. 3. 3) had already been used to isolate 33 YACs in the region of the translocation by Dr. F. Francis in Hans Lehrach's Lab, using a hybridization-based screening approach (Alitalo *et al.*, 1995; Francis *et al.*, 1994). Two more markers in the region, ZFX and POLA, were used to isolate further YACs from the ICRF human YAC library (2. 6. 1).

Nine positive YACs with ZFX and POLA were fractionated by PFGE, photographed, and Southern blotted. Hybridization of these YACs with ZFX

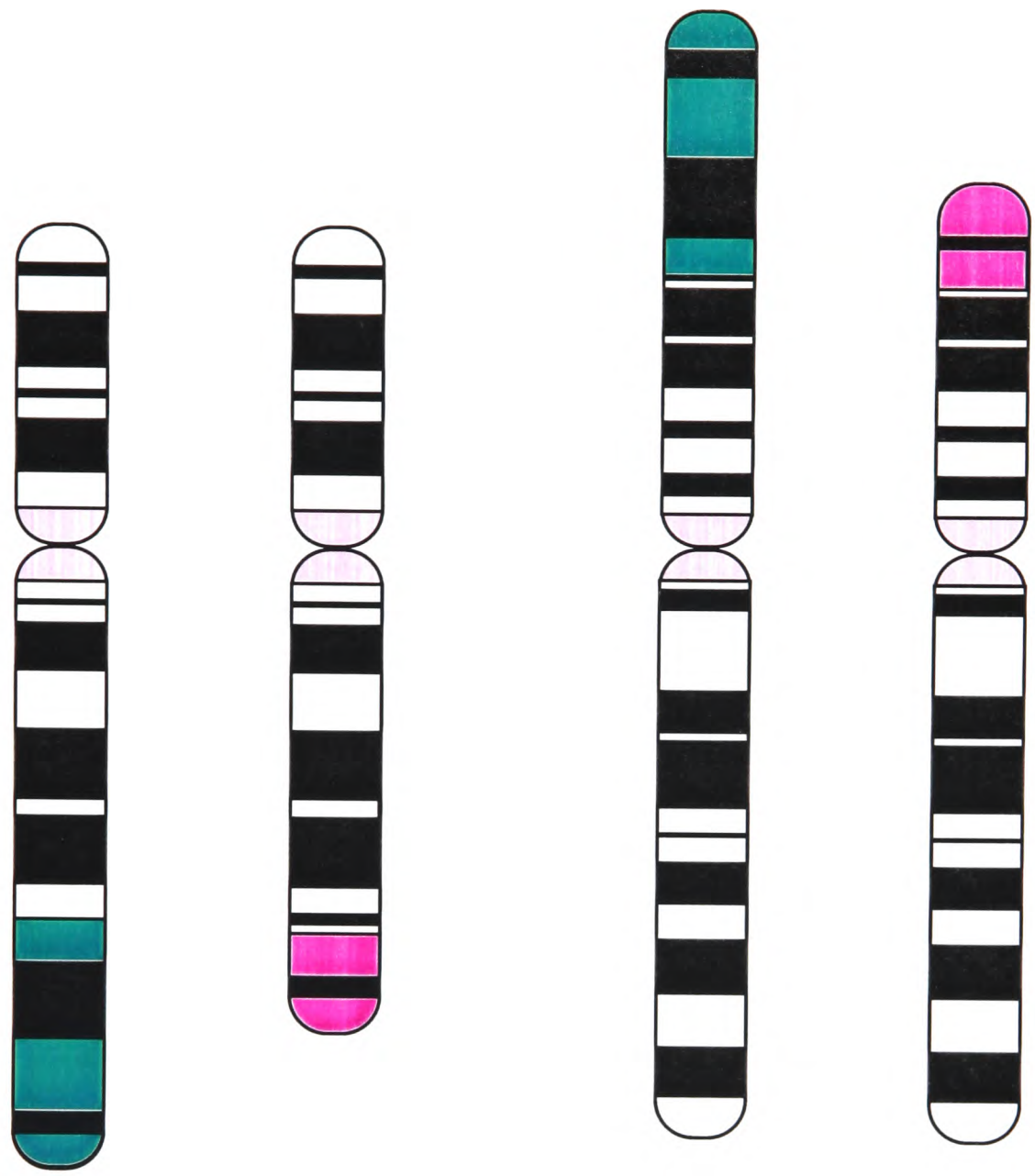



Fig. 3. 2. Schematic representation of ML's reciprocal X;8 translocation. Dark pink represents the translocated X region from Xp22.13-Xpter and green the translocated 8 region from 8q22.1-8qter. The chromosomes are G-banded.

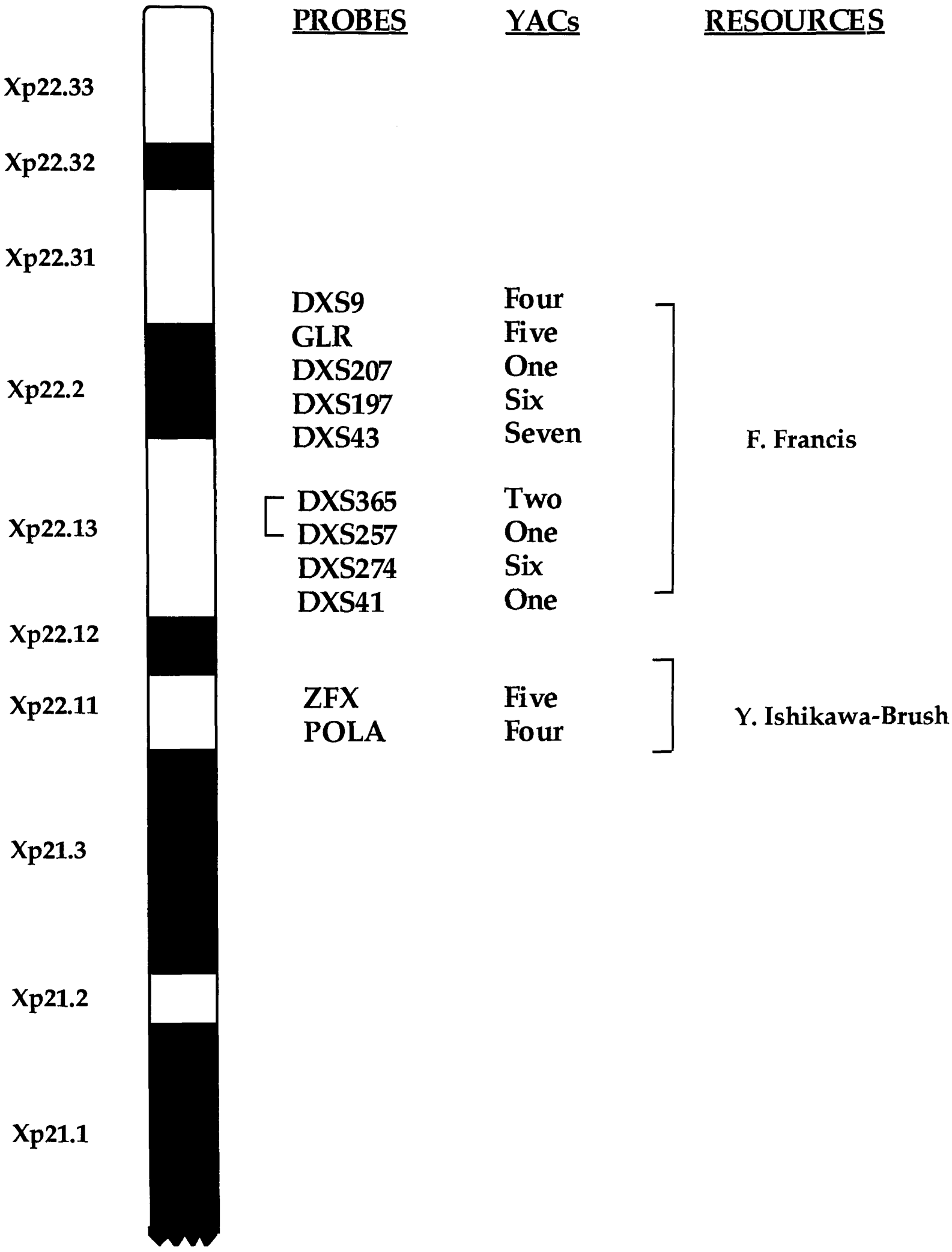



Fig. 3. 3. Origin and location of probes used in screening the ICRF YAC library, and the number of YACs isolated by each probe.

and POLA probes ensured that these YACs contained the region of ZFX and POLA. One positive YAC band was expected. If more than one YAC band was detected by *p*YAC vector or total human DNA hybridization, then a deletion or cotransformants (more than one YAC) may have occurred. The positive bands were sized in comparison with the known sizes from yeast (*S. cerevisiae*) chromosomes photographed on an ethidium bromide stained gel. The sizes are shown in Table 3. 1.

### 3. 2. 2 FISH Analysis of YACs

All 42 YACs were examined for chimerism by FISH (2.15) using normal male lymphoblastoid metaphase chromosome spreads. Non-chimeric YACs hybridize on the X chromosome only while chimeric YACs hybridize on two or three chromosomes. About 28% of the YACs from this region were found to be chimeric and 10% contained two YACs by PFGE analysis.

The FISH analysis of the translocation breakpoint used patient lymphoblastoid metaphase chromosome spreads. Non-chimeric YACs were examined in the proximal to distal direction. 23 were found to be proximal to the breakpoint, 14 distal, and five crossed the translocation breakpoint (Table 3. 1). The five YACs crossing the translocation breakpoint were detected by three positive signals on the normal X chromosome, der(X), and der(8), which were detected and distinguished from other chromosomes by DAPI staining (2. 14. 2). The X chromosome was tagged with an X centromere specific alpha satellite marker. They were YAC-57 (ICRFy900B0544) (Fig. 3. 4), YAC-58 (ICRFy900E1092), YAC-60 (ICRFy900D0758), YAC-61 (ICRFy900E1112), and YAC-63 (ICRFy900B1148). The map position of the translocation

| Locus | ICRF Ref.No. | YIB No. | Size | Status | Position |
|---|---|---|---|---|---|
| DXS9 | ICRFy900E11113 | YAC-43 | 1.0 Mb | Chimeric | Distal |
| DXS9 | ICRFy900D0367 | YAC-44 | 460 kb | Non-chimeric | Distal |
| DXS9 | ICRFy900F127 | YAC-45 | 370 kb | Non-chimeric | Distal |
| DXS9 | ICRFy900E02312 | YAC-46 | 320 kb | Non-chimeric | Distal |
| GLYR | ICRFy900F1020 | YAC-47 | 370 kb | Chimeric | Distal |
| GLYR | ICRFy900A0230 | YAC-48 | 1.0 Mb | Chimeric | Distal |
| GLYR | ICRFy901A084 | YAC-49 | 1.0 Mb | Co-transformant | Distal |
| GLYR | ICRFy905A1217 | YAC-50 | 600 kb | Non-chimeric | Distal |
| GLYR | ICRFy905D0287 | YAC-51 | 650 kb | Non-chimeric | Distal |
| 207/197 | ICRFy904G07244 | YAC-52 | 350 kb | Non-chimeric | Distal |
| DXS197 | ICRFy900B03154 | YAC-53 | 630 kb | Co-transformant | Distal |
| DXS197 | ICRFy900G1096 | YAC-54 | 400 kb | Non-chimeric | Distal |
| DXS197 | ICRFy900B12119 | YAC-55 | 350 kb | Co-transformant | Distal |
| DXS197 | ICRFy900E118 | YAC-56 | 650 kb | Chimeric | Distal |
| 197/43 | ICRFy900B0544 | YAC-57 | 700 kb | Non-chimeric | Crossing |
| 197/43 | ICRFy900E1092 | YAC-58 | 680 kb | Non-chimeric | Crossing |
| DXS43 | ICRFy900D11134 | YAC-59 | 520 kb | Chimeric | |
| DXS43 | ICRFy900D0758 | YAC-60 | 480 kb | Non-chimeric | Crossing |
| DXS43 | ICRFy900E1112 | YAC-61 | 520 kb | Non-chimeric | Crossing |
| DXS43 | ICRFy900B0544 | YAC-62 | 1.1 Mb | Chimeric | |
| DXS43 | ICRFy900B1148 | YAC-63 | 520 kb | Non-chimeric | Crossing |
| DXS43 | ICRFy900C1064 | YAC-64 | 650 kb | Chimeric | |
| DXS43 | ICRFy900C1064 | YAC-65 | 420 kb | Chimeric | |
| DXS257 | ICRFy900A0922 | YAC-66 | 370 kb | Non-chimeric | Proximal |
| DXS257 | ICRFy900E01138 | YAC-67 | 560 kb | Non-chimeric | Proximal |
| DXS365 | ICRFy900H0391 | YAC-68 | 370 kb | Chimeric | |
| DXS274 | ICRFy904G0480 | YAC-69 | 370 kb | Co-transformant | Proximal |
| DXS274 | ICRFy904D0787 | YAC-70 | 440 kb | Non-chimeric | Proximal |
| DXS274 | ICRFy904B0683 | YAC-71 | 660 kb | Non-chimeric | Proximal |
| DXS274 | ICRFy904B0583 | YAC-72 | 440 kb | Non-chimeric | Proximal |
| DXS274 | ICRFy900H0439 | YAC-73 | 250 kb | Co-transformant | |
| 274/41 | ICRFy904A01173 | YAC-74 | 380 kb | Non-chimeric | Proximal |
| DXS41 | ICRFy904F02311 | YAC-75 | 370 kb | Non-chimeric | Proximal |
| ZFX | ICRFy900A0601 | YAC-42 | 800 kb | Chimeric | |
| ZFX | ICRFy900A0501 | YAC-41 | 300 kb | Non-chimeric | Proximal |
| ZFX | ICRFy900A0401 | YAC-40 | 600 kb | Non-chimeric | Proximal |
| ZFX | ICRFy900A0301 | YAC-39 | 580 kb | Non-chimeric | Proximal |
| ZFX | ICRFy900A0201 | YAC-38 | 650 kb | Non-chimeric | Proximal |
| POLA | ICRFy902A06103 | YAC-15 | 300 kb | Non-chimeric | Proximal |
| POLA | ICRFy902F0612 | YAC-14 | 900 kb | Chimeric | |
| POLA | ICRFy902G0821 | YAC-13 | 500 kb | Non-chimeric | Proximal |
| POLA | ICRFy902G0921 | YAC-12 | 500 kb | Non-chimeric | Proximal |

Table 3. 1. Results of the FISH analyses of the X-chromosome ICRF YACs showing their chimeric status and position relative to the translocation breakpoint. Also shown is the size of the YACs, their isolating probes, ICRF reference numbers, and YIB numbers used in this text.

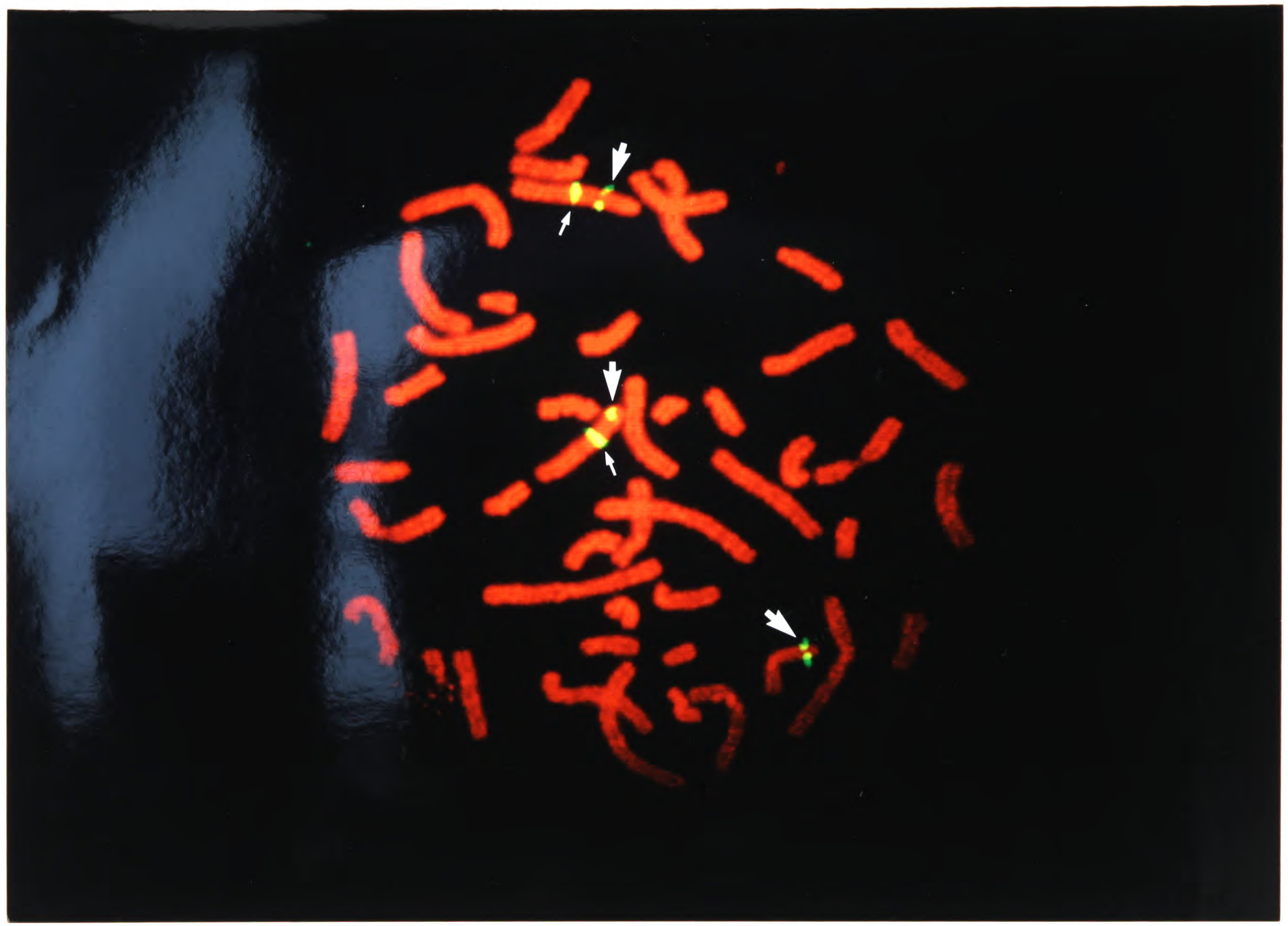


Fig. 3. 4. Hybridization of YAC-57 on normal X, der(X) and der(8) chromosomes (large arrow) indicates the YAC spans the translocation breakpoint. The X alpha satellite marker was used to identify the chromosome centromere (small arrow).

breakpoint was now narrowed to about a 500 kb fragment. The proximal and distal flanking markers were DXS43 and DXS197, respectively. The previously unknown map of two marker clusters, [GLR, DXS43, and DXS197] and [DXS207 and DXS9], was determined to be:

"Xpter-DXS9-GLR-DXS207-DXS197-DXS43-Xcen"

by this FISH experiment (Francis *et al.*, 1994).

## 3. 3 COSMID ISOLATION

Cosmids were used to isolate the translocation breakpoint in smaller fragments of 40-50 kb in size. YACs spanning the translocation breakpoint were used as probes to screen for cosmids in the ICRF FSCX cosmid library. None of the cosmids identified spanned the translocation breakpoint. It was therefore necessary to construct cosmid libraries from these YACs to generate better coverage of the critical region in cosmids.

### 3. 3. 1 Whole YAC Screening of the ICRF FSCX Cosmid Library

The two shortest non-chimeric YACs, YAC-60 (480 kb) and YAC-61 (520 kb), were used to isolate cosmids by whole YAC screening (Fig. 3. 5) of the ICRF FSCX cosmid library (2. 6. 1). 58 cosmids were isolated. From now on these will be referred to as the Lawrist cosmids after the name of the vector used to clone them.

To construct the contig, these cosmids were digested by *Hind*III. However, only 20% were digested completely under standard conditions. Complete digestion was achieved using a poorer nutrient medium and a reduced incubation time for cosmid growth. This appeared to prevent the



Fig. 3. 5. Hybridization of YAC-60 as a probe onto the ICRF flow-sorted X chromosome cosmid library.

formation of rearrangements. One portion of the digested cosmid DNA was run on gels and blotted to filters. The remaining portions were further digested with *Alu*I to fragment it into smaller pieces and purified to generate probes. The filters and probes were used for cosmid-cosmid hybridization in an attempt to construct a contig covering about 500 kb (the length of a YAC). The Southern blots were also hybridized with the publicly available probes, DXS197, 315, 4077 and DXS43, and 18 additional probes generated by Alu-PCR amplification (2. 2. 6) derived from the five YACs crossing the breakpoint. This produced a contig with two gaps (Fig 3. 6).

In parallel with the construction of the contig, FISH was used to identify cosmids spanning the translocation breakpoint. Eleven cosmids were found to be proximal and eight distal (Table 3. 2) to the translocation breakpoint until Cos-17 displayed the pattern of crossing the breakpoint (Fig 3. 7). This cosmid hybridized on the normal X chromosome, der(X) and der(8). The X chromosome centromeres were easily identified by the X alpha satellite marker. To confirm the FISH result, Cos-17 was hybridized to the *Eco*RI, *Hind*III, *Pst*I, *Sna*BI, and *Kpn*I digested genomic DNA of ML, AL (ML's mother), normal male, and normal female. Junction fragments contain altered restriction sites for these enzymes, producing different size of fragments from that of the normal. In the hybridization results, however, no difference in fragment size was observed between Cos-17 and ML's genomic DNA blots (Fig. 3. 8). This result did not confirm the FISH results that Cos-17 spanned the translocation breakpoint. However the result is ambiguous. Either (1) the FISH observation is a false positive because of peculiarities in

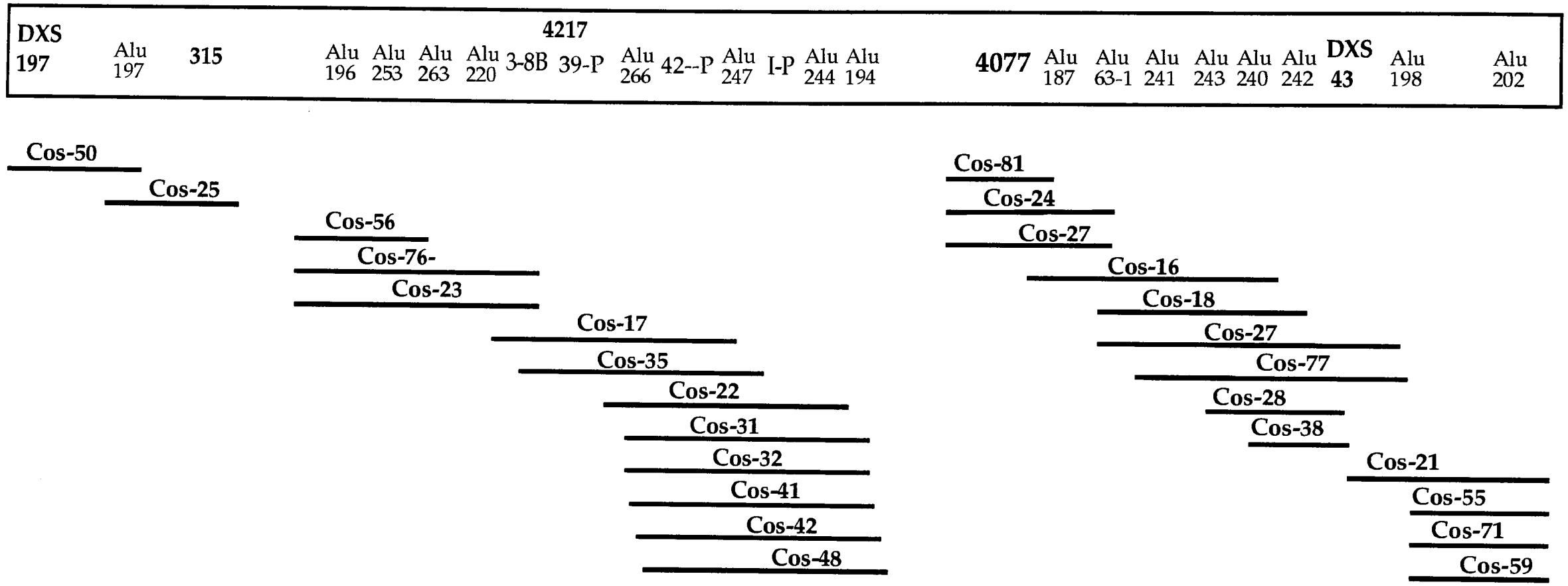



Fig. 3. 6. Contig of the Lawrist cosmids showing the probes above the cosmids they isolated. The contig contained two gaps.

| ICRF cosmids | Proximal | Distal |
|---|---|---|
| Cos-15 | + | |
| Cos-16 | + | |
| Cos-18 | + | |
| Cos-19 | + | |
| Cos-20 | + | |
| Cos-21 | + | |
| Cos-27 | + | |
| Cos-32 | + | |
| Cos-50 | + | |
| Cos-51 | + | |
| Cos-56 | + | |
| Cos-17 | + | + |
| Cos-25 | | + |
| Cos-26 | | + |
| Cos-37 | | + |
| Cos-40 | | + |
| Cos-47 | | + |
| Cos-48 | | + |
| Cos-49 | | + |
| Cos-57 | | + |
| Cos-58 | | + |

Table 3. 2 FISH results showing the relationship of 21 Lawrist cosmids with respect to the translocation breakpoint. There were 11 proximal and nine distal cosmids. Cos-17 was indicated to possibly span the translocation breakpoint.

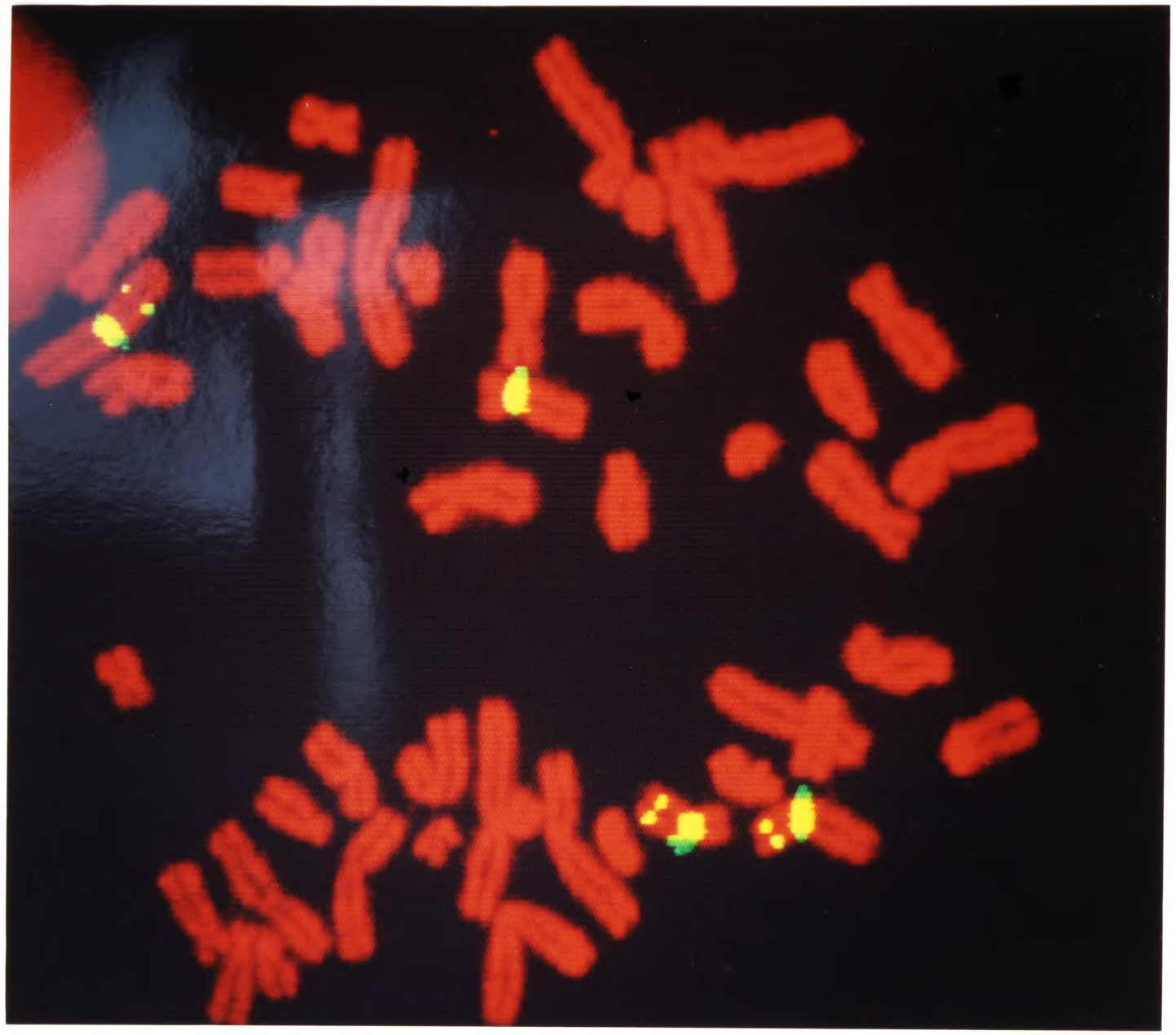


Fig. 3. 7. Fluorescence *in situ* hybridization of Cos-17 on normal X, der(X) and der(8) chromosomes indicating the cosmid spanning the translocation breakpoint. Chromosomes X and der(X) were tagged by an X-centromere specific alphoid repeat probe. Chromosomes 8 and der(8) were tagged by an 8-centromere specific alphoid repeat probe. Cos-17 is shown hybridizing to chromosome X, der(X) and der(8).

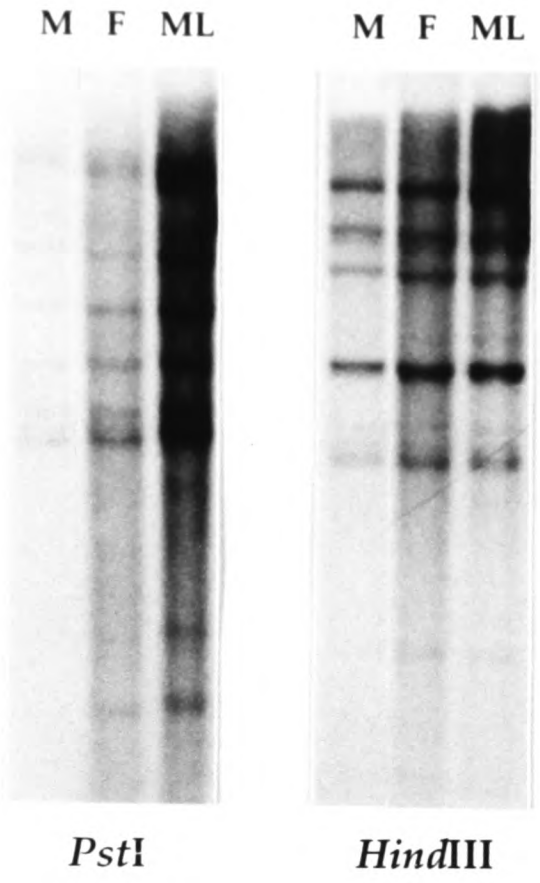



Fig. 3. 8. Cos-17 hybridization to normal female and patient (ML) genomic DNA digested either by *Hind*III or *Pst*I, indicating the absence of a junction fragment.

Cos-17, such as rearrangement, or repeat sequences common to Cos-17, chromosome X local repeat sequences that are located both proximally and distally to the translocation breakpoint (e. g. retroviral insertions), or (2) the rearranged restriction sites in the region of the translocation breakpoint were very similar to the restriction sites on the normal X. However the second possibility is very unlikely to have occurred for each of five different restriction enzymes.

In an attempt to resolve this ambiguity, the analysis was repeated using other cosmids overlapping with Cos-17. Three overlapping cosmids, Cos-22, Cos-31, and Cos-35, were identified through similarities in their separate *Hind*III and *Eco*Rl electrophoretic banding patterns and hybridization results

(Fig. 3. 9). The *Hind*III contig of these four cosmids is shown in Table. 3. 3. Fragment sizes were determined by polynomial regression (2. 3. 2). FISH

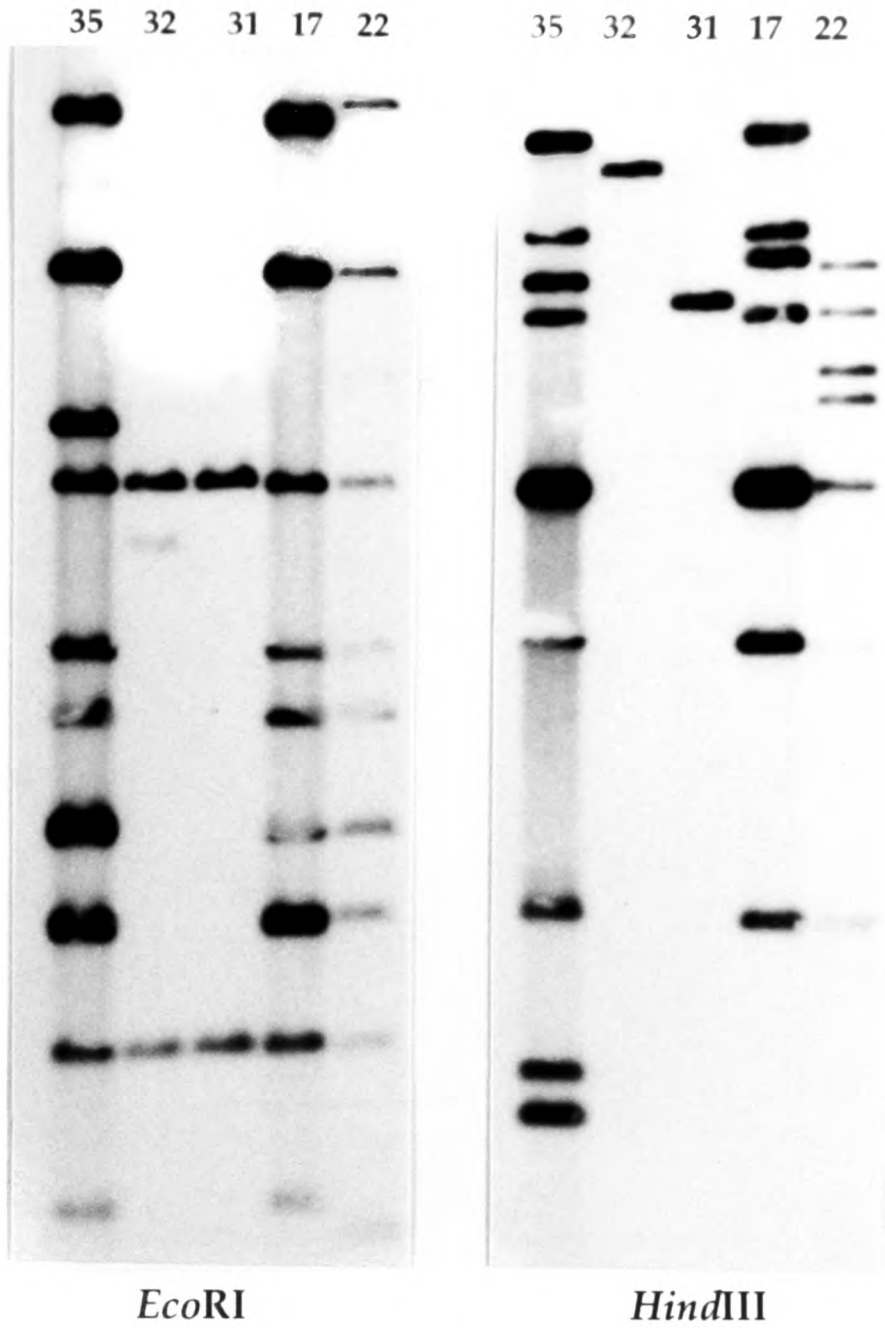



Fig. 3. 9. Identification of Lawrist cosmids overlapping Cos-17 through close similarities in banding patterns. Lawrist Cos-35 was hybridized onto southern blots of *Hind*III or *Eco*RI digests of Cos-22, Cos-17, Cos-31, Cos-32 and Cos-35.

| Cosmids | 6.6 kb | 1.6 kb | 8.7 kb | 3.9 kb | 4.1 kb | 5.8 kb | 7.0 kb | 4.3 kb | 1.3 kb | 6.3 kb |
|---|---|---|---|---|---|---|---|---|---|---|
| Cos-17 | + | + | + | + | + | + | + | + | + | + |
| Cos-22 | | | | + | + | + | + | + | + | + |
| Cos-35 | | | + | + | + | + | + | | | |
| Cos-31 | | | | | | | | + | + | + |

Table. 3. 3. A *Hind*III contig [6.5, 1.6]- 8.7-[3.9, 4.1, 5.8, 7.0]-[4.3, 1.3, 6.3] of four Lawrist cosmids: Cos-17, Cos-22, Cos-31, and Cos-35. The order of the fragments in square brackets and the orientation of the cosmids were unknown.

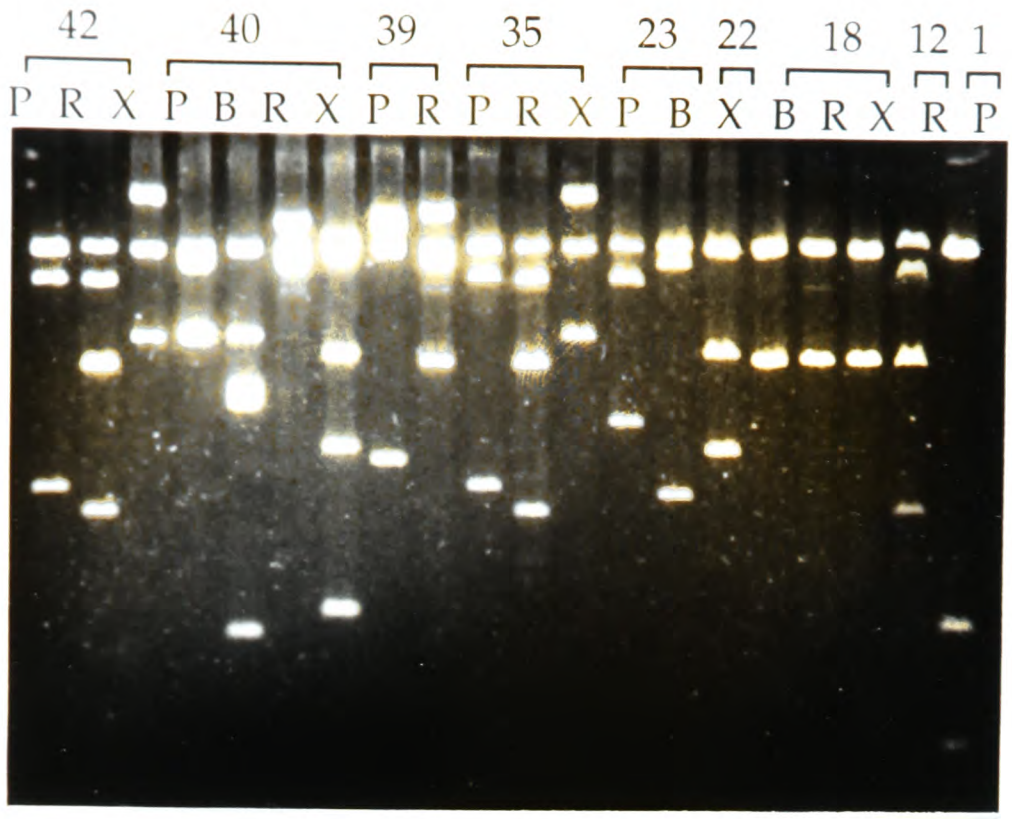
analysis and genomic DNA hybridization failed to determine the position of these cosmids with respect to the translocation breakpoint because of high background signal.

To obtain clearer hybridization results, shorter probes of average size of 4-5 kb were prepared from subcloning of *Hind*III fragments of Cos-17 in *p*Bluescript (2. 7. 5). These subclones were digested with *Hind*III to separate insert DNA from the Lawrist vectors and further double digested individually using one of either, *Xba*I, *Bgl*II, *Eco*RI, or *Pst*I to produce smaller probes. All digests were Southern blotted and hybridized with a sonicated human placental DNA probe (2. 9. 2). Fragments hybridizing with the probe contained repeat sequences such as LINE-1 and Alu. The remaining fragments contained potentially unique sequences. Of these, 18 (Table 3. 4) were excised from gels (Fig. 3. 10A and 10B), and purified for use as probes.

| ***Xba*I** | ***Bgl*II** | ***Eco*RI** | ***Pst*I** |
|---|---|---|---|
| 18-X | 23-B | 12-R | 1-P |
| 22-X | 40-B | 35-R | 23-P |
| 35-X | | 39-R | 35-P |
| 40-X | | 40-R | 39-P |
| 42-X | | 42-R | 42-P |

Table 3. 4. Eighteen fragments containing unique sequences from human genomic DNA, derived from double digests of either *Xba*I , *Bgl*II , *Eco*RI , or *Pst*I on *Hind*III subclones of Cos-17.

These 18 probes were hybridized to YACs spanning the translocation breakpoint, and to normal male and female genomic DNA to confirm they were derived from these YACs and they were unique in genomic DNA. 12 of the probes satisfied these conditions and they were hybridized to Southern



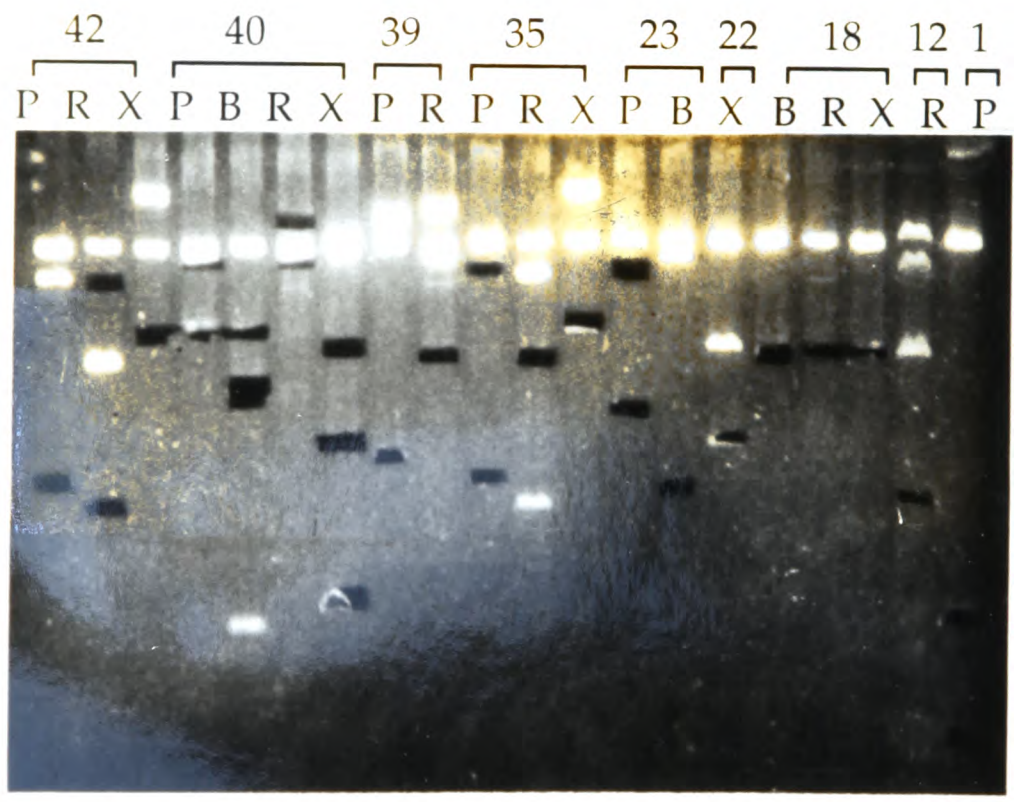




Fig. 3. 10. Isolation of 18 unique fragments from Cos-17-*Hind*III subclones. (A) *Hind*III subclones digested with *Xba*I, *Bgl*II, *Eco*RI or PstI, separated by electrophoresis and stained with EtBr. (B) Excision of unique fragments after identified through negative hybridization signals with human placenta DNA.

blots of ML's genomic DNA digested with *Eco*RI, *Bgl*II, *Eco*RV, *Pst*I, *Kpn*I and *Snb*I. The results of the hybridization are shown in Fig. 3. 11. None of the probes indicated the presence of a junction fragment.

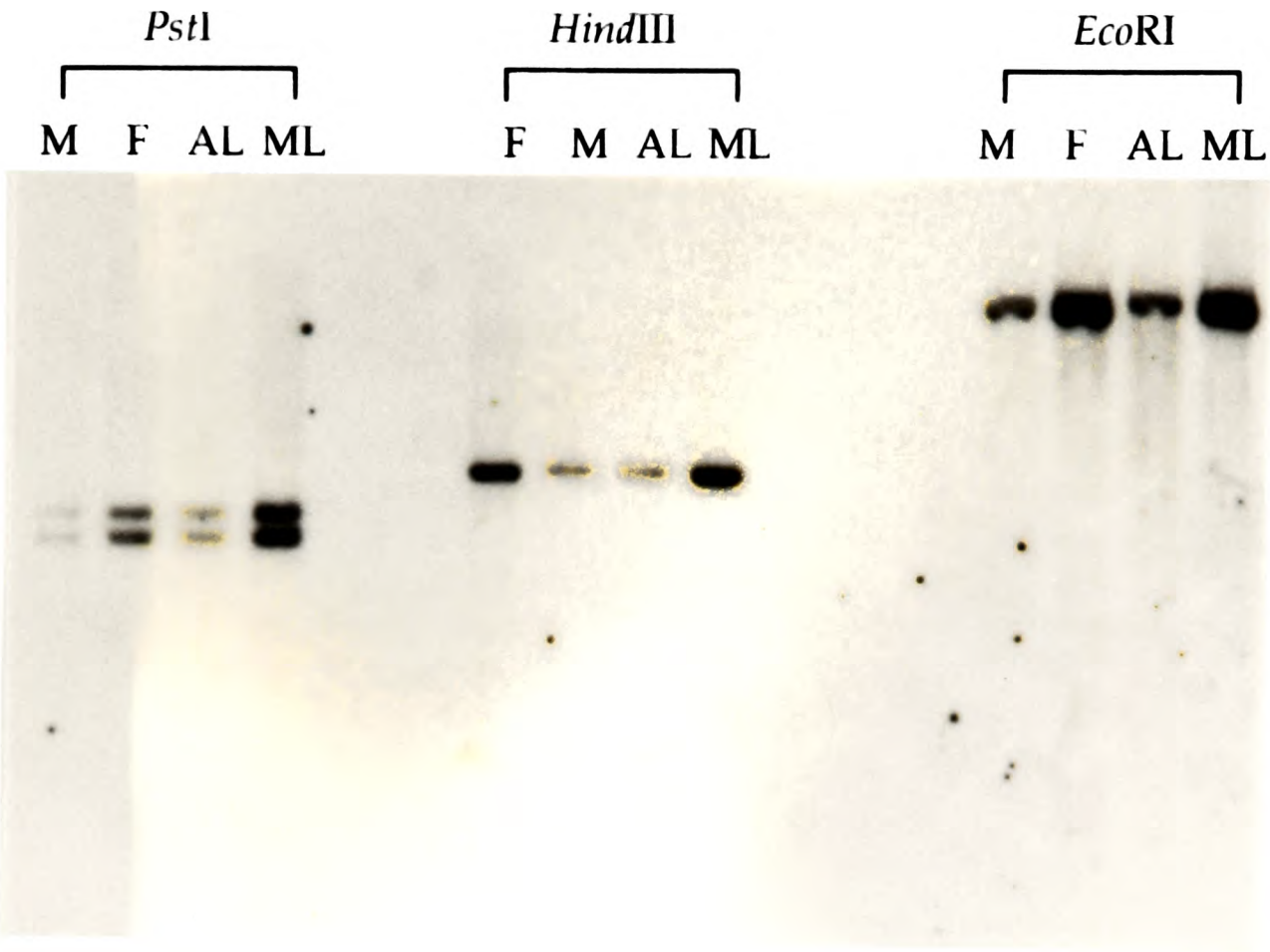
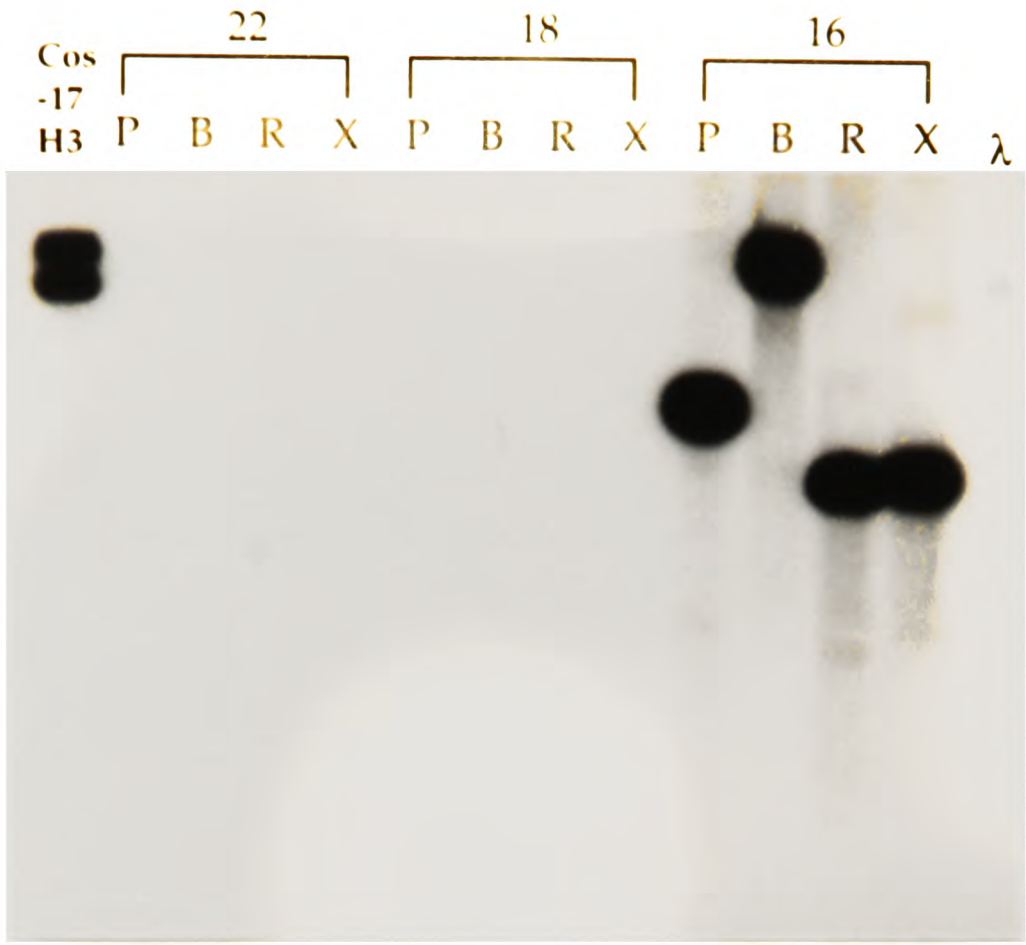



Fig. 3. 11. Example hybridization of a Cos-17 unique sequence subclone onto different digests of patient and normal genomic DNA. Legend: ML, patient DNA; AL, ML's normal mother's genomic DNA; F and M, normal female and male genomic DNA.

### 3. 3. 2 Search for Polymorphic Di- and Tri-repeats

One useful outcome of the Cos-17 analysis was the discovery of di- and tri-nucleotide repeat sequences in the Cos-17 *Hind*III subclones. Such microsatellite repeats are useful in family studies. Primers with a length of 24 nt, equivalent to two possible di-nucleotide repeats or 20 possible tri-nucleotide repeats (2. 2. 7), were labeled using the kinase-mediated method (2. 9. 3) and hybridized onto Southern blots of *Hind*III inserts. Several positive repeats were found using CA, CAC, CAG, and TCC probes (Fig. 3. 12) and sent



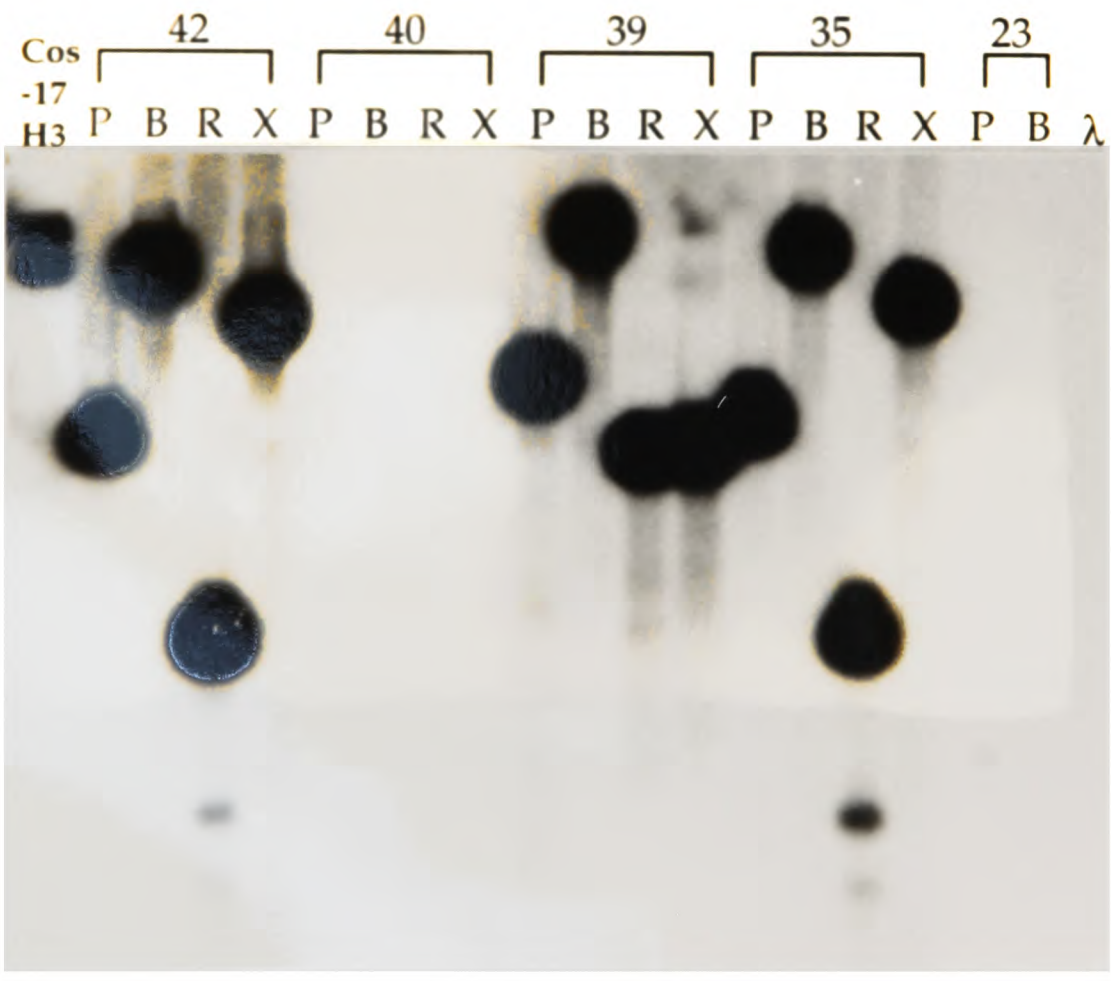




Fig. 3. 12. Hybridization of (CA)n probes to Cos-17 subclones (double digested with *Hind*III and other enzyme as indicated) indicating the presence of (CA)n repeats in the subclone. Legend: P, *Pst*I, B, *Bgl*II, R, *Eco*RI, and X, *Xba*I.

to Dr. J Powell, Institute of Psychiatry, London, for generation of polymorphic markers, who found that one CA repeat was polymorphic with four alleles and a heterozygosity of 0.38 (Drs. J. Powell and E. Maestrini, personal communication). The data from a total of 135 X chromosomes (45 females and 45 males) are shown in Table 3. 5.

| Allele | Base pair | Frequency |
|---|---|---|
| 1 | 103 | 0.00741 |
| 2 | 111 | 0.0815 |
| 3 | 113 | 0.77 |
| 4 | 115 | 0.141 |

Table 3. 5. Cos-17 CA repeat subclone alleles determined from 135 X chromosomes, 45 females and 45 males (E. Maestrini, personal communication).

### 3. 3. 3 Construction of the Cosmid Library

Because none of the Lawrist cosmids could be clearly demonstrated to span the translocation breakpoint it was necessary to construct new cosmid libraries from YACs spanning the translocation breakpoint.

Two cosmid libraries, subcloned into the SuperCos vector, were constructed (2. 6. 2) from the two smallest YACs spanning the breakpoint (YAC-60 (480 kb) and YAC-61 (520 kb)). The replicated filters of each library were hybridized with total human placenta DNA to isolate human recombinant clones. 225 positive cosmids, referred to as the SuperCos cosmids from now on to distinguish them from Lawrist cosmids, were picked onto three microtitre plates and replicated into filters using a 96-pin device (Fig. 3. 13).

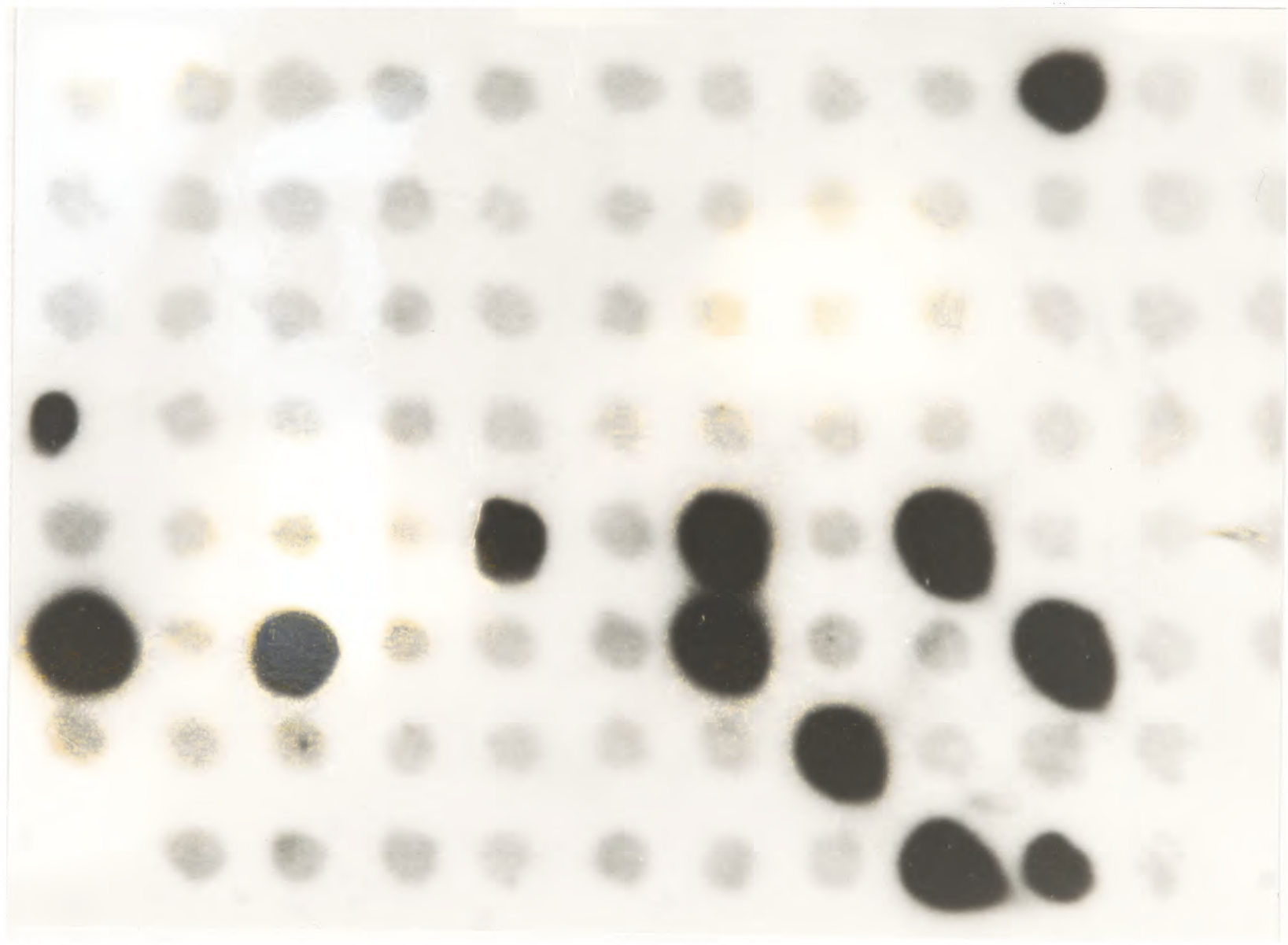


Fig. 3. 13. Example of a replica filter transferred cosmid clones using a 96-pin device from a cosmid library constructed from YAC-60. Hybridization of a GRPR exon-1 probe onto one of these replica is shown.

### 3. 3. 4 Construction of the Cosmid Contig

The 225 SuperCos cosmids were ordered into a contig to extend the Lawrist contig, using the same probes as before (3. 3. 1). Another 10 probes were generated from the gastrin releasing peptide receptor (GRPR) gene which had just been mapped to Xp22.1, the region of ML's translocation breakpoint (Alitalo *et al.*, 1995). The cDNA sequence of GRPR (database accession number M73481) is about 1.7 kb (1628 bp) and contains three exons, 800 bp, 350 bp and 550 bp in length, respectively (Fig. 3. 14). Primers were designed to

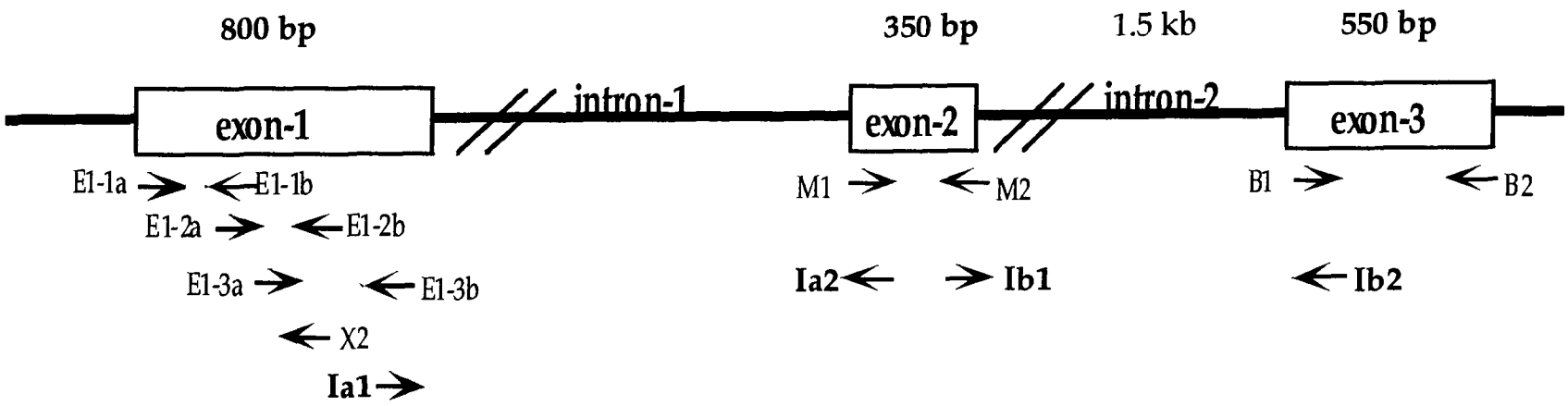



Fig. 3. 14. The structure of the GRPR gene, showing the location of primers.

amplify the three exons and two introns by PCR: E1-1a and E1-3b for exon-1, M1 and M2 for exon-2, B1 and B2 for exon-3, Ia1 and Ia2 for intron-1, and Ib1 and Ib2 for intron-2. All exons and intron-2 were amplified, but intron-1 appeared to be too long for PCR amplification. Intron-2 was found to be about 1.5 kb in length (Fig. 3. 14). A further six PCR products were amplified using various combinations of the GRPR primers and Alu primers such as

Collins-5, Collins-3 (Tagle and Collins, 1992), 2729 and 3144 (Monaco *et al.*, 1991) (Fig. 3. 15). All ten probes were hybridized onto the *Hind*III Southern blots of the Lawrist and SuperCos cosmids.

Four GRPR probes identified the relative location of the GRPR gene in the contig. Only Lawrist cosmids which had mapped as proximal to the breakpoint in the FISH experiments (Table 3. 2) were positive with the exon-2, intron-2 and exon-3 probes. No distal cosmids hybridized with the GRPR probes, nor did the exon-1 probe hybridize with any Lawrist cosmid. For the SuperCos contig, 16 cosmids hybridized with exon-1 only and eight were found to cover the whole GRPR gene, suggesting a GRPR gene length within 40-50 kb. The contig covered 500 kb with one gap (Fig. 3. 16).

## 3. 4 SUMMARY

The translocation breakpoint in the patient ML was mapped at high resolution to t(X;8)(p22.13;q22.1) by DAPI staining and karyotype analysis. The BrdU incorporation analysis confirmed that the translocated X chromosome is active and the normal X chromosome is inactive. FISH analysis, using 42 YACs from the Xp22 region, identified the translocation breakpoint in 5 YACs of about 500 kb each in Xp22.13. Two of these YACs were used to construct two cosmid libraries and a cosmid contig with one gap. Eight cosmids were identified as both spanning the translocation breakpoint and containing the whole of the gastrin releasing peptide receptor gene (GRPR), thus associating the GRPR gene with the translocation breakpoint.

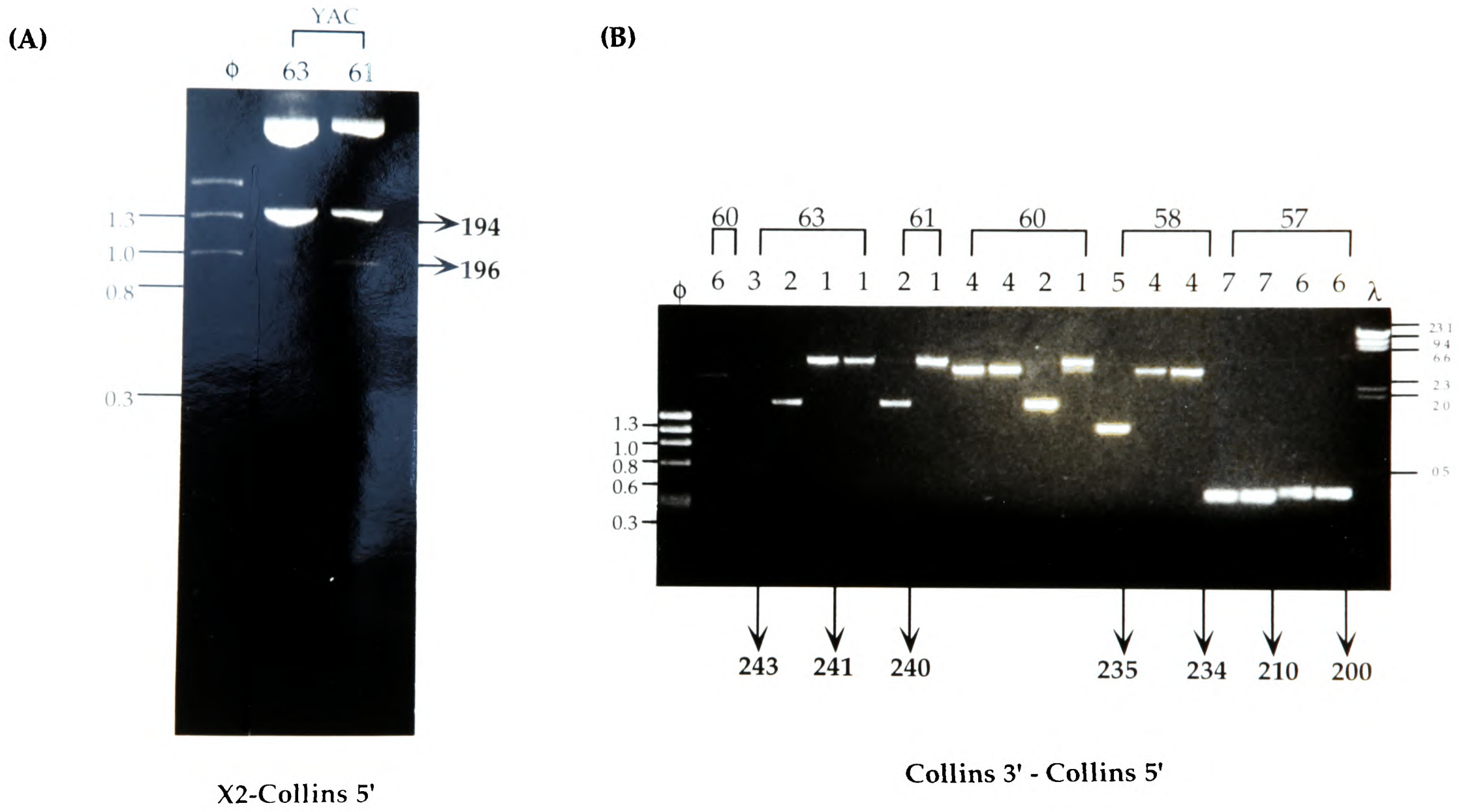



Fig. 3. 15. (A) The four PCR products amplified from the GRPR exon-1 (X2) 3′-end primer (Fig. 3. 14) and an Alu Collins 5′ primer (2. 2. 6). The fragments used are designated Alu194 and Alu196 in order of large to small. (B) Alu243, Alu241, Alu240, Alu235, Alu234, Alu210, and Alu200 were amplified from an Alu Collins 3′ and 5′ primers (2. 2. 6). The templates used were YAC-63, -61, -60, -58, and -57. The marker is a ϕx174 DNA-*Hae*III digest.

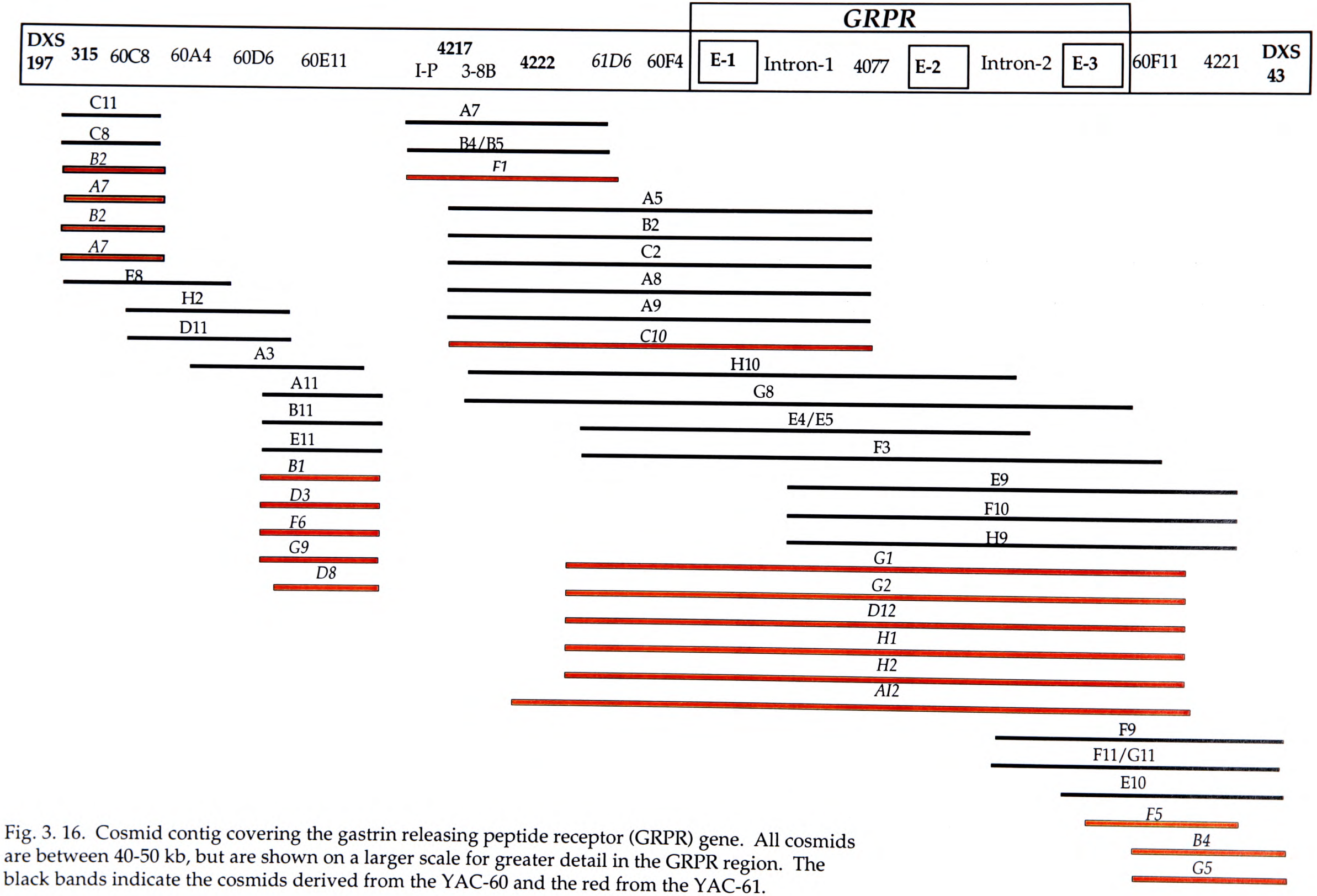



Fig. 3. 16. Cosmid contig covering the gastrin releasing peptide receptor (GRPR) gene. All cosmids are between 40-50 kb, but are shown on a larger scale for greater detail in the GRPR region. The black bands indicate the cosmids derived from the YAC-60 and the red from the YAC-61.

# CHAPTER IV

## TRANSLOCATION WITHIN THE GRPR GENE

Placing the GRPR gene in the cosmid contig established it as a candidate gene for the site of the translocation breakpoint. Construction of a restriction map (4. 3. 2) and sequencing (4. 3. 3) confirmed the gene had been disrupted by the translocation. GRPR expression was studied in relation to X inactivation (4. 2. 2) and tissue specificity (4. 2. 3). The translocation junction fragments were cloned in phage (4. 3. 1) and then subcloned for sequence analysis.

### 4. 1 IDENTIFICATION OF THE TRANSLOCATION BREAKPOINT

The location of the breakpoint in relation to the GRPR gene was examined using two approaches, hybridization of GRPR cDNA probes to Southern blots of genomic DNA isolated from ML and physical mapping in the GRPR gene region.

#### 4. 1. 1 Isolation of GRPR cDNA

cDNA of the GRPR gene was prepared by (1) PCR amplification of first strand cDNA from total normal female lymphoblastoid cell RNA (2. 2. 8a) and (2) PCR screening of a testis cDNA library (Clontech) (2. 12. 2).

The three exons and combinations of exons 2 and 3 were amplified from first strand cDNA by PCR using the GRPR primers described in section 3. 3. 4 and Fig. 3. 14. However, a full length, 1.7 kb, GRPR cDNA product

could not be generated. Although the 5′ exon-1 primer (E1-1a) and the 3′ exon-3 primer (B2) (Fig. 3. 14) amplified several fragments, none could be distinguished as the full length fragment.

Screening the testis cDNA library by PCR using these same primers amplified many non-specific fragments but again none could be identified as full length GRPR cDNA.

Identification of full length cDNA may have required making adjustments to the PCR conditions and changing the cDNA source to a tissue which expresses GRPR at a higher level. However, before this was attempted, the physical mapping strategy on cloned DNA had already identified the region of translocation breakpoint and work on the generation of a full-length cDNA construction was discontinued.

### 4. 1. 2 Physical mapping in the GRPR gene region

Patient genomic DNA was digested separately with *Hin*dIII, *Eco*RI, *Eco*RV, *Pst*I and *Bgl*II, Southern blotted and hybridized with three of the eight SuperCos cosmids covering the GRPR region: G2, H1, and H2, all of which were derived from the YAC61 (marked in red in Fig. 3. 16). All three cosmids clearly identified junction fragments in all *Pst*1 shown in (Fig. 4. 1) but the *Hin*dIII digests. Although puzzling at the time, the discrepancy of the *Hin*dIII result was later resolved (4. 3. 2).

Identification of the translocation breakpoint in the GRPR gene was approached through *Hin*dIII subcloning SuperCos cosmids carrying the whole GRPR region and the region of the translocation breakpoint. *Sau*3A subcloning of the same cosmids were used to identify some of the intron

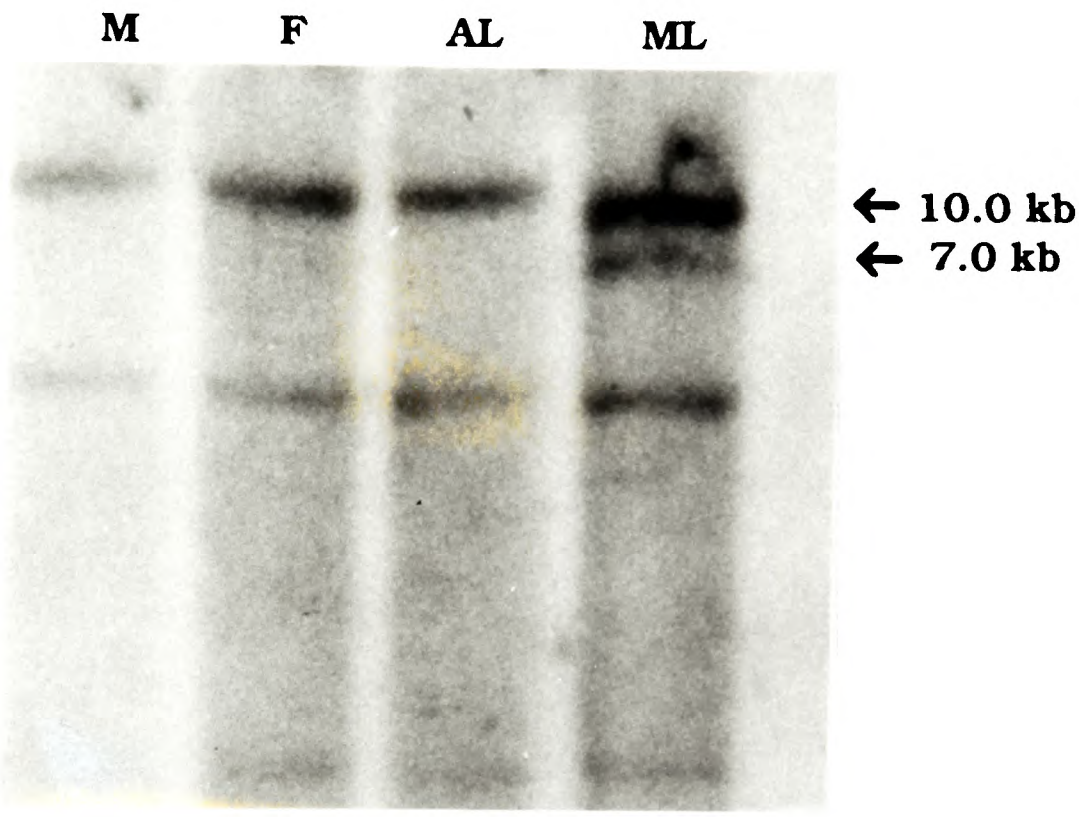



Fig. 4. 1. Whole SuperCos G2 cosmid hybridization to Southern blots of the patient genomic DNA digested with *Pst*I, identifying the junction fragment in the patient. Legend: ML, patient; AL, patient's normal mother; F, female; M, male.

sequence adjacent to some of the exons for generation of PCR primers flanking each exon.

For the latter experiment, cosmid G2 was digested by *Sau*3A, and subcloned into *Bam*HI-CIAP treated *p*Bluescript. Because *Sau*3A and *Bam*HI are isoschizomers, small G2 fragments of about 420 bp, convenient for sequencing could be easily ligated into vector. The recombinants were purified and about 10% were sequenced. SauB6 and SauA11 subclones contained part of GRPR exon-1, SauE8 and SauE9 subclones contained part of GRPR exon-2, and SauC11 contained small part of GRPR exon-3. The primers 243 for the GRPR exon-1, I2mm and I1mm for the exon-2, and 368 for the exon-3 (Fig. 4. 2) were generated and used to screen for GRPR mutations and analysis of splice sites.

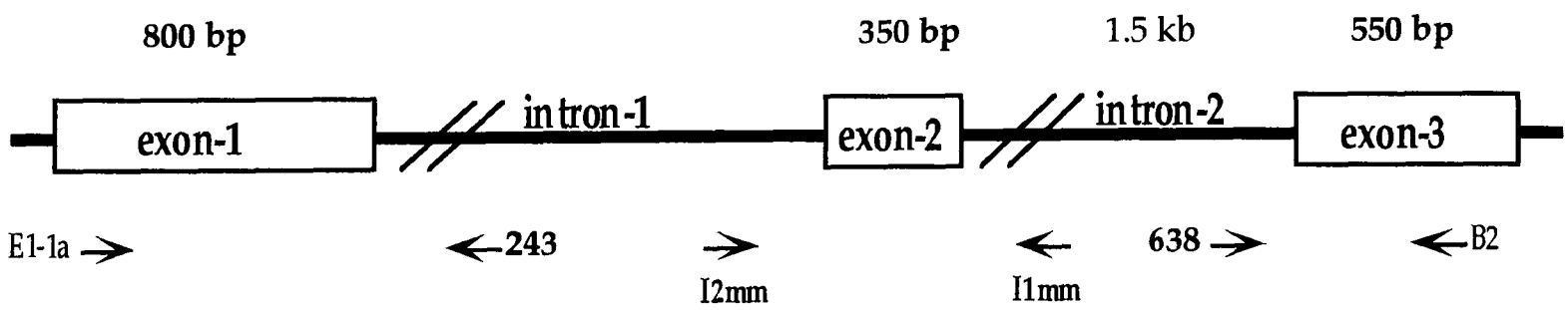



Fig. 4. 2. Location of the primers generated and used to screen for GRPR mutations and analysis of splice sites.

At the same time, the existence of microsatellite polymorphisms in the *Sau*3A subclones of the SuperCos cosmid G2 was examined because these may be useful in family studies. The microsatellite repeat oligonucleotide probes

previously used to detect microsatellites in the Lawrist cosmid Cos-17 (2. 2. 7) were hybridized with filters containing the *Sau*3A subclones. These same microsatellite repeat sequences, CA, CGG, CAG and TCC, were also found in the SuperCos cosmid. These await further analysis to determine if they are polymorphic.

In the hybridization method for the identification of the translocation breakpoint in the GRPR, G2 and H1, two of the eight SuperCos cosmids spanning the translocation breakpoint and covering the GRPR gene region were digested by *Hin*dIII and subcloned into *p*Bluescript. These fragments were larger than the *Sau*3A fragments, averaging about 4 kb in size, and reduced the number of hybridizations needed to isolate the flanking markers of the junction fragment. The subclones were separated from vector and purified for use as hybridization probes. They were radiolabeled and hybridized to Southern blots of (1) patient genomic DNA digested separately by *Hin*dIII, *Eco*RI, *Eco*RV, *Pst*I and *Bgl*II, and (2) *Hin*dIII digests of Lawrist and SuperCos cosmids. These hybridizations were used to map the probes into the cosmid contig and further refine cosmid overlaps until probes identifying the junction fragments in the patient genomic DNA blot were detected. This occurred with the 10th and 11th probe, G2-7a and G2-7b, two *Hin*dIII fragments identified in the same plasmid subclone. The resulting map is:

**(G2-5a)-(G210-1)-(exon-1)-(G2-7b)-(G2-7a)-(4077)-(G2-6)-(G2-1)-(G2-5b)**
**-(G2-1)-(H10-5a)-(exon-2)-(intron-2)-(exon-3)-(H10-10)**

G2-7a and G2-7b, from now on referred to as 7a and 7b, detected junction fragments in *Pst*I, *Eco*RV, *Bgl*II, and *Eco*RI digests of patient DNA (Fig. 4. 3).

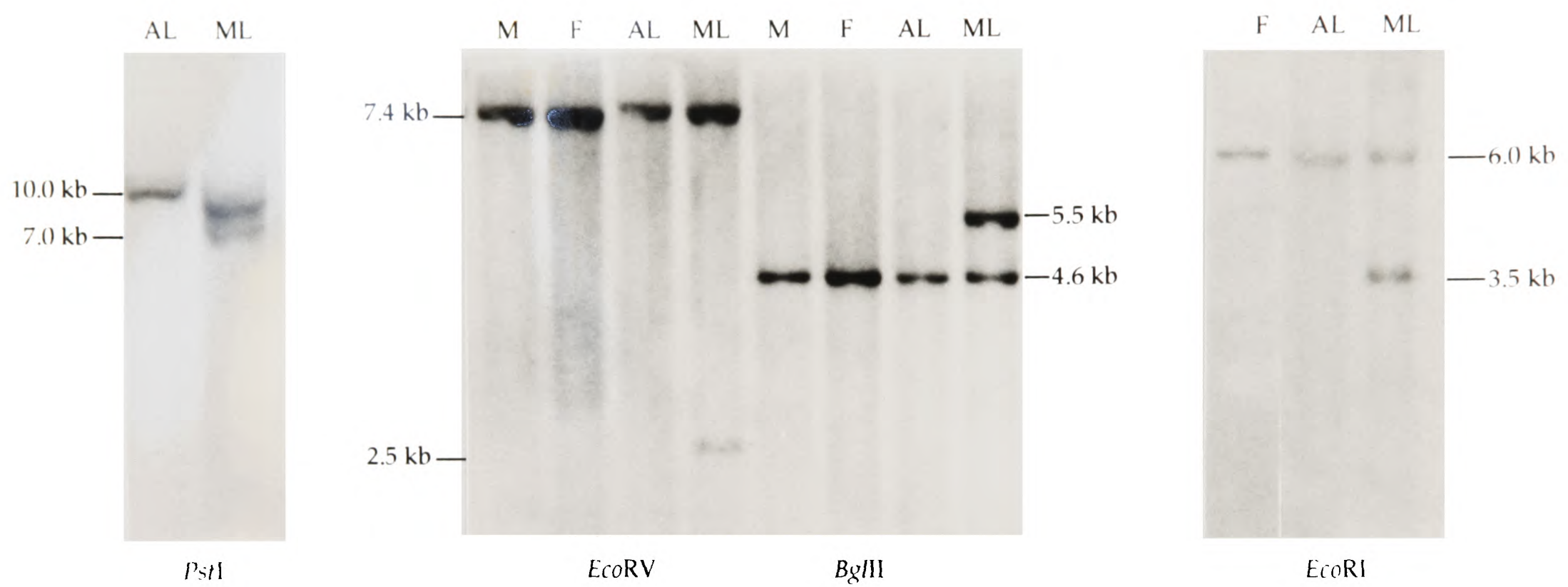



Fig. 4. 3. Detection of the translocation junction fragment in patient genomic DNA digested with *Pst*I, *Eco*RV, *Bgl*II, and *Eco*RI, using probe 7b. Legend: ML, patient; AL, patient's normal mother; F, female; M, male.

7a hybridized with small *Hind*III fragments (about 1-2 kb) in both patient and control DNA. This apparently contradictory result was later clarified (4. 3. 2). The hybridization pattern by 7a and 7b were not identical, indicating that they were unique and flanked the breakpoint on the X chromosome. These subclones displayed an interesting relationship corresponding to the two breakpoints. In the normal X chromosome both 7a and 7b probes hybridized with a 4.6 kb *Bgl*II fragment from ML's digested genomic DNA, while 7a hybridized only with a 5.0 kb *Bgl*II fragment on either der(X) or der(8), and 7b hybridized only with a 5.5 kb *Bgl*II fragment on either der(X) or der(8). This relationship confirmed the breakpoint was most likely between 7a and 7b. (Fig. 4. 4). Both subclones hybridized with cosmids E9, F10 and H9. These contain exon-2, intron-2 and exon-3, but not exon-1 of the GRPR gene. The subclones also hybridized with cosmids A5, B2, C2, A8, and A9, which are positive for exon-1 of the GRPR gene, but not exon-2, intron-2 nor exon-3. These results placed 7a and 7b within the first intron of the GRPR gene (Fig. 4. 5) and thus identified intron-1 to be the site of the translocation breakpoint.

## 4. 2 The GRPR Gene

The translocation breakpoint disrupted the GRPR gene. The isolation of GRPR polymorphic markers and the expression of the X-linked gene, GRPR, was examined.

### 4. 2. 1 Polymorphism in Exon-2 of the GRPR Gene

Disruption of the GRPR gene in association with the patient's phenotype stimulated a search for GRPR mutations in other autistic patients. Three pairs of primers (Fig. 4. 2), generated from sequencing the GRPR exon areas

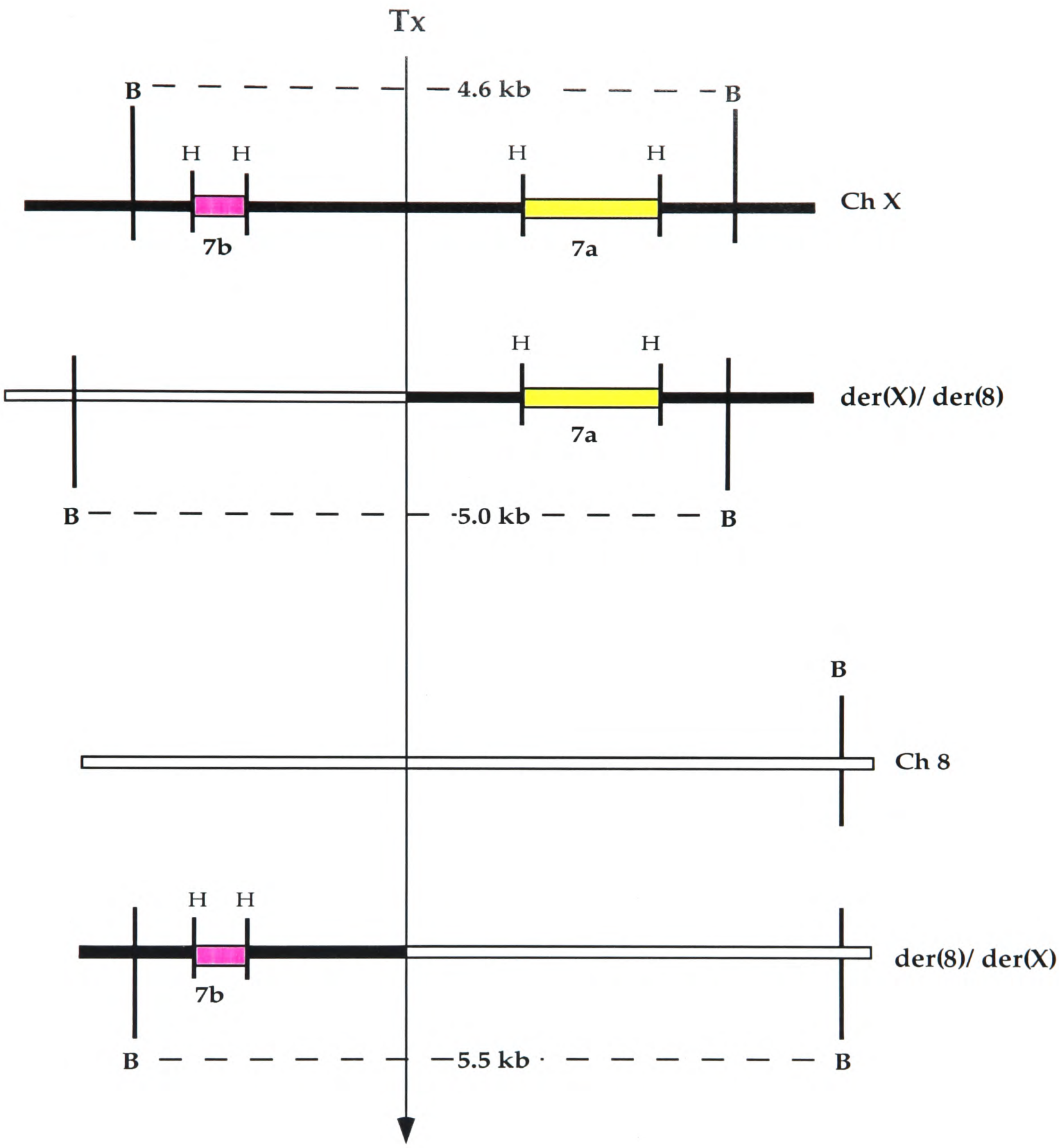



Fig. 4. 4. Preliminary analysis showing relationships between the translocation breakpoint and the SuperCos X chromosome cosmid subclones, 7a and 7b, resulting at the time from hybridizations with patient genomic DNA. Both subclones hybridized with normal X and each hybridized separately with a derivative chromosome, but none with chromosome 8. "Tx" indicates then the possible position of the translocation breakpoint.

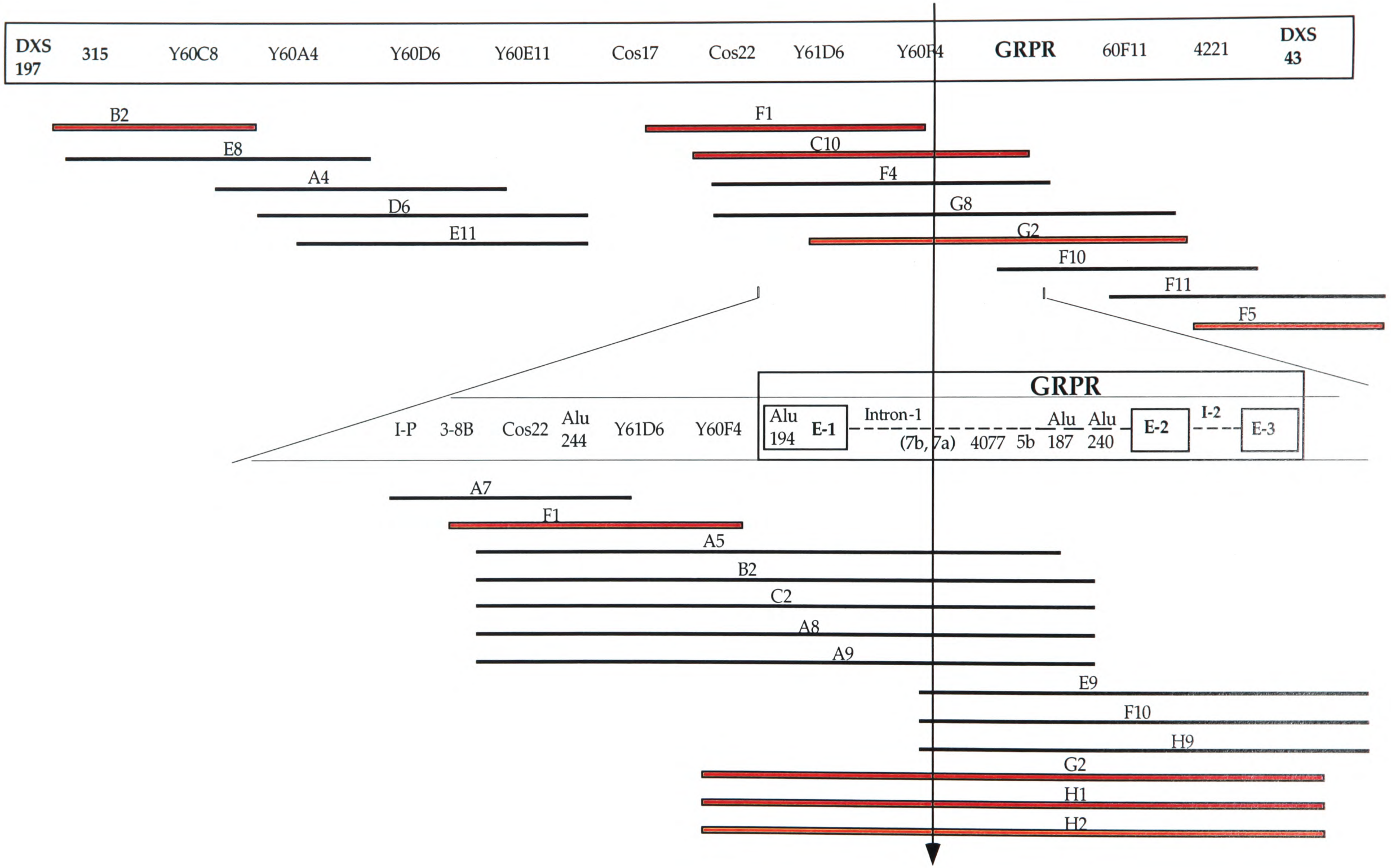



Fig. 4. 5. Position of the translocation in the SuperCos cosmid contig. Cosmids A5, B2, C2, A8, A9 contains exon-1 (E-1), 7b, and a part of 7a. Cosmids E9, F10, H9 contained E-2, E-3, and the majority part of 7a. This located the translocation breakpoint within the first intron of GRPR.

were sent to Dr. G. Brown (University of Oxford, UK) and Dr. A. Poustka (DKFZ, Germany), who examined unrelated autistic patients for GRPR splice site alterations and other mutations. Testing two autistic young males (AJ and MJ), R. Brown and G. Brown (personal communication) found a polymorphism in exon-2 of the GRPR gene using the Single-Stranded Conformation Polymorphism (SSCP). I confirmed this result using the Conformation-Sensitive Gel Electrophoresis (CSGE) technique (2. 13. 1) and later by sequencing. Two alleles were detected, a thymine for AL, a cytosine for MJ, both at position +851 in exon-2 (Fig. 4. 6). The translocation patient ML was heterozygous for the two alleles. Both TCT and TCC code for serine. S. Klanch and A. Poutska examined the exon-2 polymorphism and the other exons of GRPR in five patients and found no etiologic mutations by direct sequencing. Thus, seven unrelated autistic male patients so far studied did not have alterations in splice sites nor mutations in the GRPR coding region.

### 4. 2. 2 The GRPR gene and Escape from X-Inactivation

It is known that some areas of the X chromosome escape inactivation, such as the pseudoautosomal region and small regions nearby. Although the GRPR gene is far from this region, confirmation of its activation status is important in understanding ML's phenotype. The inactivation assay was performed by Dr. A. P. Miller at Huntington F. Willard's laboratory (Case Western Reserve University). The GRPR exon-2 end and exon-3 end primers (Ib1-B2) (Fig. 4. 7A) amplified a product of 574 bp from cDNA using RT-PCR of RNA isolated from a panel of mouse/human somatic cell hybrids. GRPR expression was examined from normal active and inactive X chromosomes. The panel

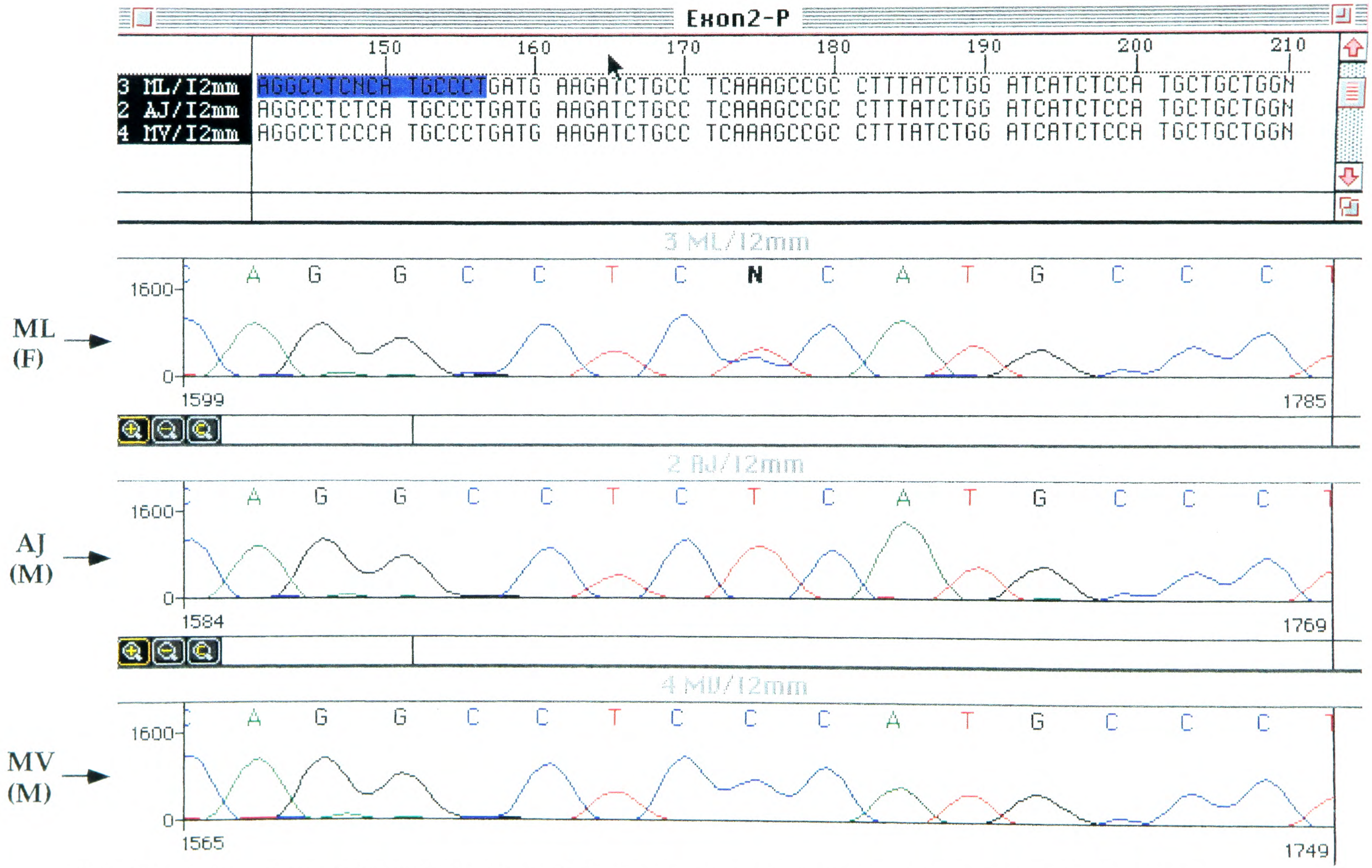



Fig. 4. 6. Sequence analyses of GRPR indicating a base substitution mutation in exon-2. Male AJ is hemizygous for thymidine, MV is hemizygous for cytosine, and female ML is heterozygous for both.

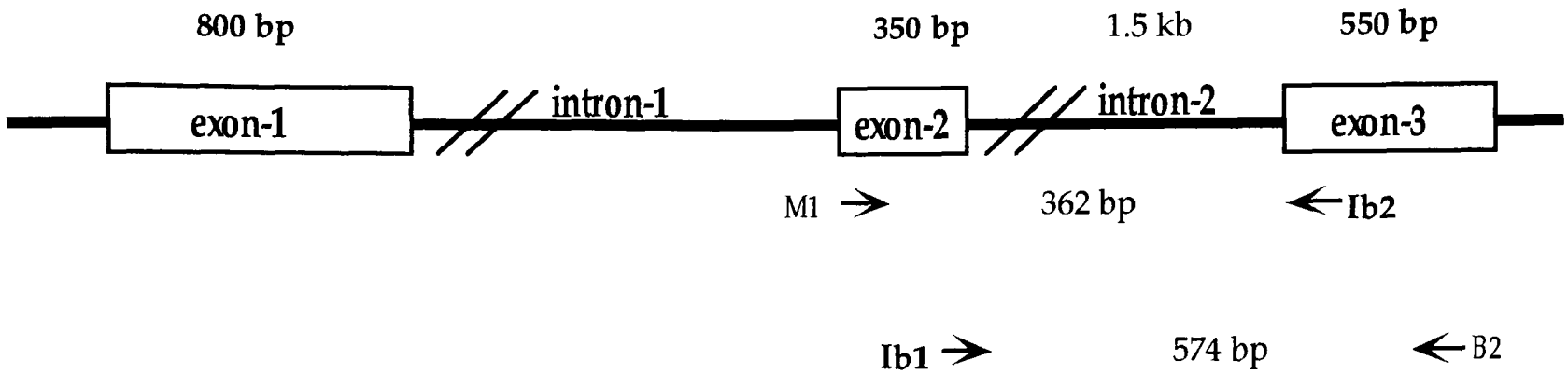



Fig. 4. 7A. Primers used for the X-inactivation analysis

consisted of two active X containing hybrids, Xa1 (AHA-11aB1) and Xa2 (t60-12), and four inactive X containing hybrids Xi1 (t11-4Aaz5), Xi2 (t48-1a-1Daz4a), Xi3 m(t75-2maz34-4a) and Xi4 (t86-B1maz1b-3a) (Brown *et al.*, 1990; Miller *et al.*, 1995; Willard *et al.*, 1993). Expression of GRPR was detected from both the active and inactive X chromosomes indicating that it escapes X-inactivation (Fig. 4. 7B) at least in this assay.

### 4. 2. 3 Tissue Specific Expression of GRPR

GRPR expression was detected in pancreas after a long exposure, using GRPR exon-2 probe derived from a PCR product between M1 and M2 (Fig. 3. 14), but not in brain, placenta, kidney, prostate, testis, small intestine, colon, heart, lung, liver, skeletal muscles and ovary on northern blot analysis of poly(A)+ RNA (Clontech) (Fig. 4. 8). Northern blots containing total RNA isolated from control and patient lymphoblastoid cells showed no hybridization signal using GRPR exon probes (data not shown).

## 4. 3 JUNCTION FRAGMENT

The junction fragment was cloned into phage using the *Bgl*II digest of patient DNA. The *Bgl*II digest produced a junction fragment of 5.5 kb which is

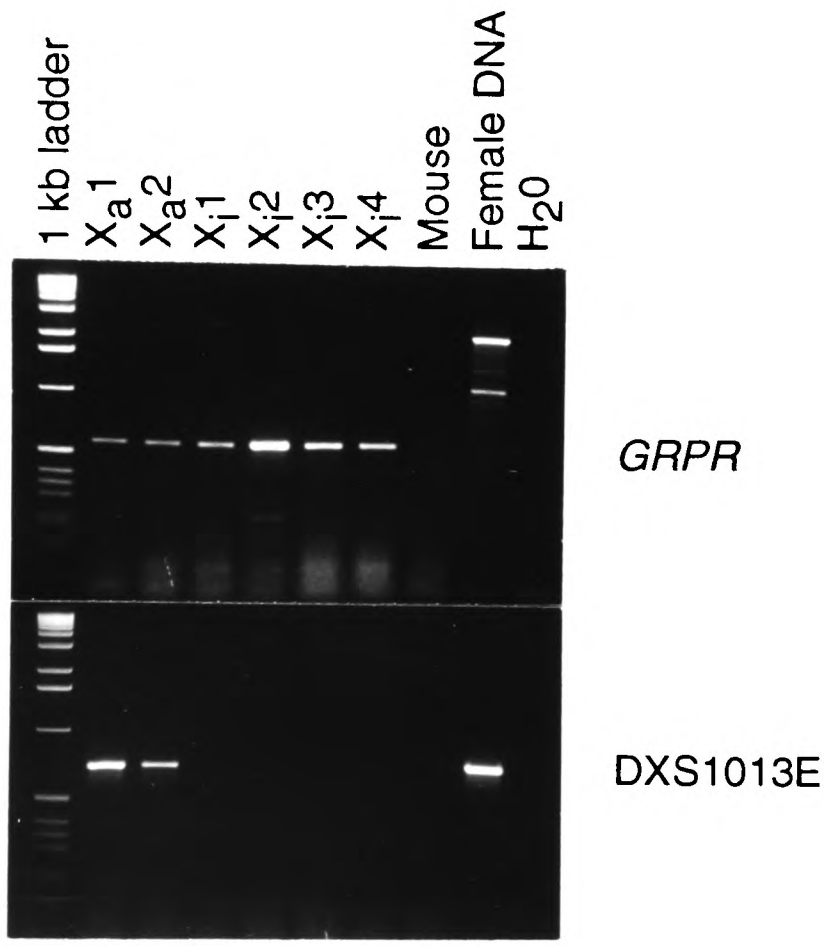



Fig. 4. 7B. Escape from X inactivation of GRPR. The top figure demonstrates that GRPR is expressed from both the active and inactive X chromosomes. Primers used to detect expression are Ib1/B2 which amplify a product of 574 bp from cDNA, part of exon-2 and exon-3 (Ib1-B2), and a larger product from DNA. DXS1013E expression is used as a control for a gene that is subject to inactivation. These primers amplify a product of about 700 bp. Lane 'mouse' refers to the parental mouse line (tsA1S9) and is a negative control for the amplification of murine products. The DNA lane corresponds to the female cell line GM7002. The lane on the far right is a H20 lane and molecular weight makers (1 kb ladder, Gibco BRL) are in the far left lane.

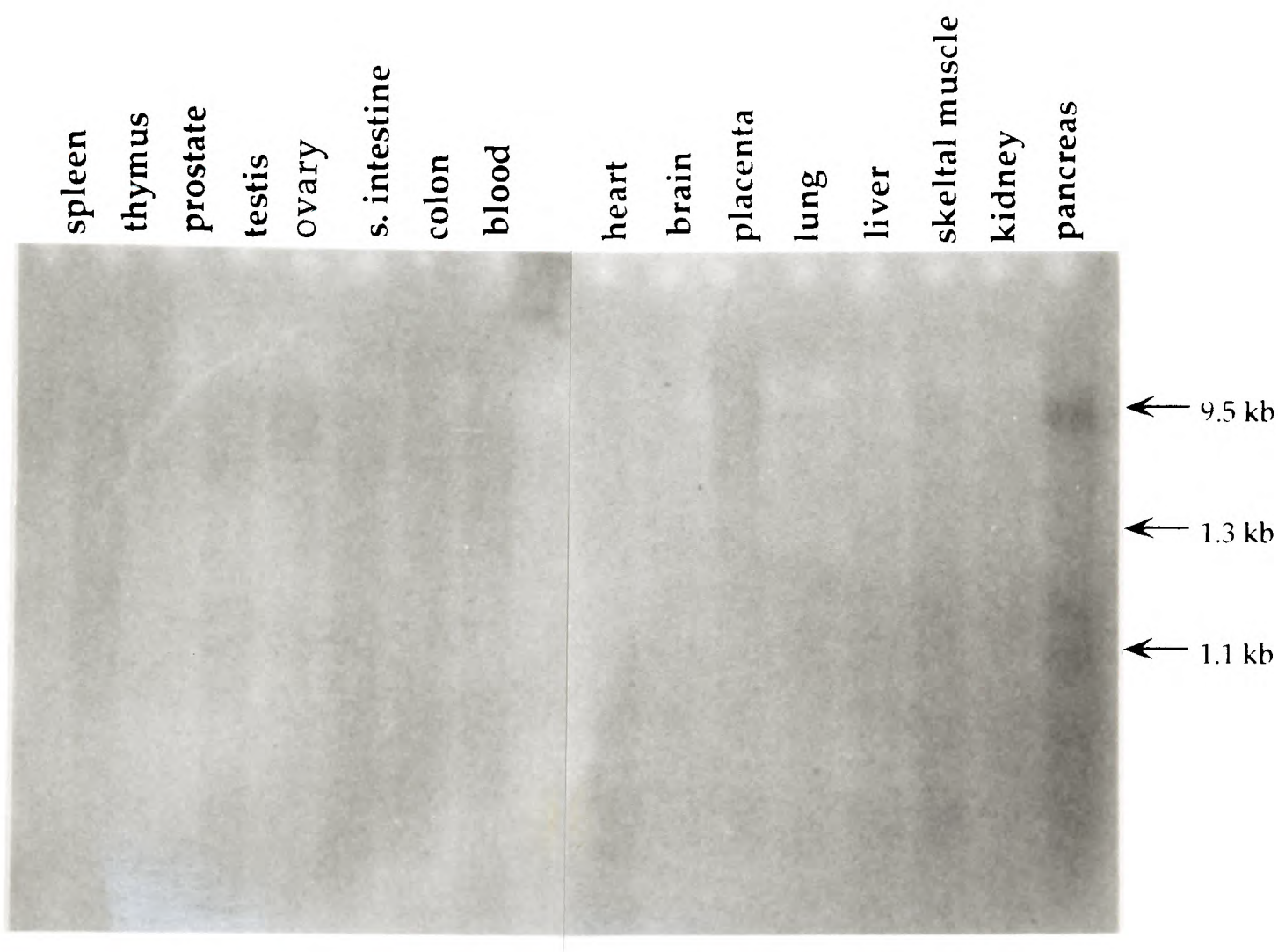



Fig. 4. 8. Northern blot hybridization analysis of GRPR expression. Expression was detected in poly$A^+$ RNA from pancreatic tissue.

different within 1 kb of the normal X *Bgl*II fragment of 4.6 . This similarity in size facilitated equal representation of both fragments for cloning into phage. The positive phages were subcloned into *Bam*HI-digested *p*Bluescript, digested with a combination of enzymes to construct a restriction map, and sequenced.

### 4. 3. 1 Cloning the Junction Fragment

Genomic DNA from ML was digested with *Bgl*II and a phage library of 600,000 pfu was constructed (2. 6. 3). Duplicate library filters were hybridized with probes 7a and 7b separately. As predicted, three unique types of phage clones were identified, one with a normal chromosome X fragment (positive 7a and 7b), one with a der(X) fragment and one with a der(8) fragment (positive with only 7a or 7b). Six clones of each type were identified from the 3-genome-fold coverage library (Fig. 4. 9).

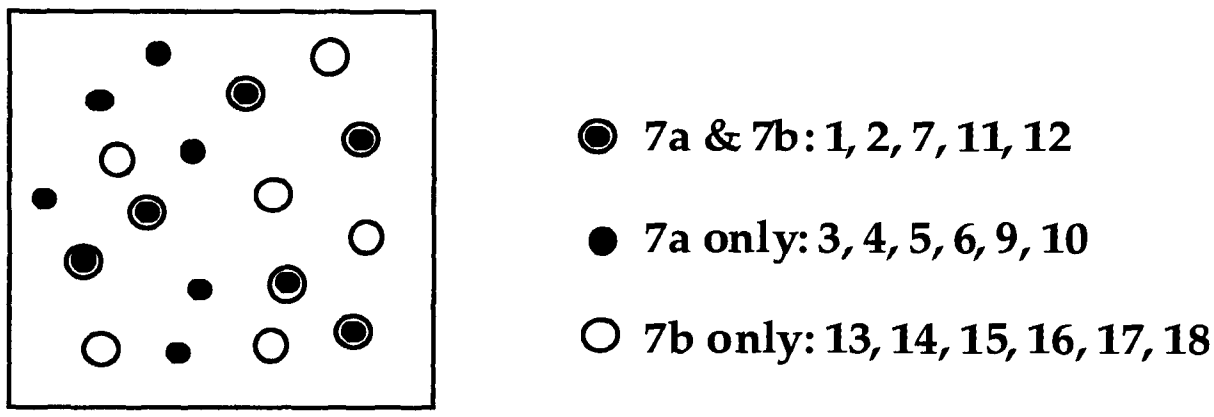



Fig. 4. 9. Schematic representation, summarizing 3 separate hybridization of (1) 7a and 7b, (2) 7a only, and (3) 7b only, onto a patient phage library. The circles represent separate and overlapping hybridizations.

DNA isolated from the positive phage clones was digested with *Bgl*II or *Bgl*II/*Sal*I, Southern blotted (Fig. 4. 10A), and hybridized with 7a or 7b (Fig. 4. 10B). These positives were ligated into the *Bam*HI-CIAP treated *p*Bluescript,

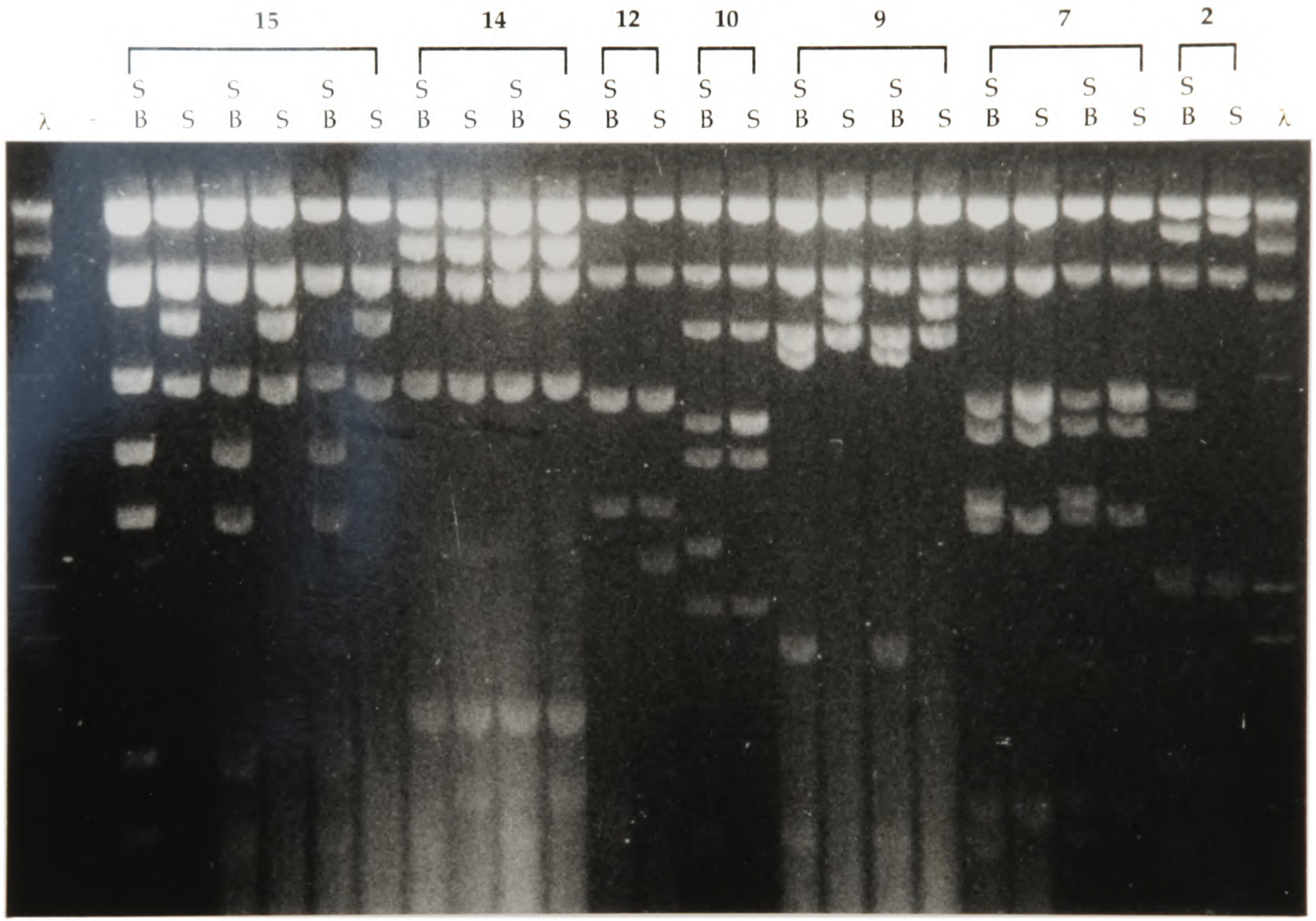



Fig. 4. 10A. EtBr stained gel photo containing the patient genomic DNA cloned into bacteriophage EMBL3 and digested with *Bgl*II (B) or *Bgl*II/*Sal*I (B/S). EtBr stained gel photo visualizing the patient genomic DNA digested with *Bgl*II (B) or *Bgl*II/*Sal*I (B/S).

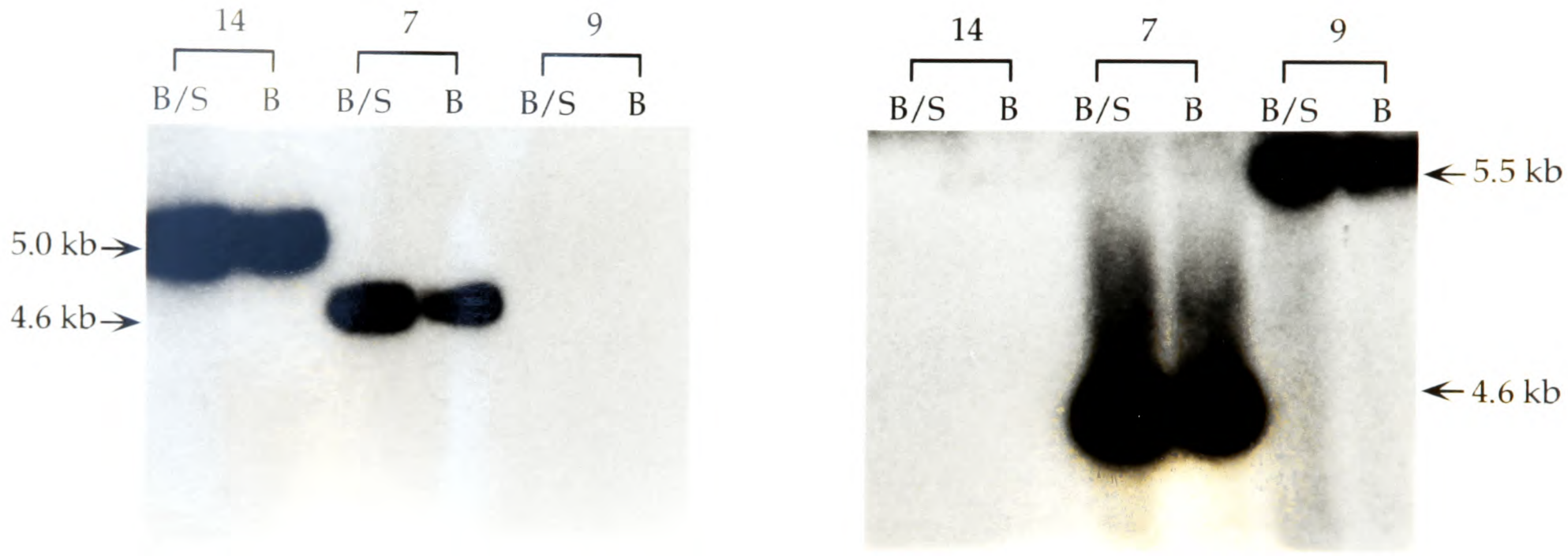



Fig. 4. 10B. Hybridization of SuperCos cosmid subclones 7a and 7b to patient genomic DNA cloned in phage and digested with *Bgl*II (B) or *Bgl*II/*Sal*I (B/S).

and transformed in competent cells. The subclones were digested with *Not*I and *Sma*I to separate them from the plasmid vector, Southern blotted, and hybridized with 7a and 7b probes. Three types of plasmid clones were again identified. The first, designated 7-6, positive with both 7a and 7b, contained a 4.6 kb *Bgl*II fragment derived from the normal X chromosome. The second, designated 14-16, was positive with 7a only and contained a 5.0 kb *Bgl*II fragment, and the third, designated 9-2, was positive with 7b only and contained a 5.5 kb *Bgl*II fragment. The second and third two *Bgl*II fragments are the two separate translocation junction fragments derived from either the der(X) or der(8) chromosomes.

### 4. 3. 2 Construction of a Restriction Map

A restriction map of these three plasmid subclones was constructed using single (*Not*I, *Hind*III, *Sma*I, *Xba*I, *Spe*I and *Acc*I), double (*Not*I/*Sma*I, *Xba*I/*Hind*III), and triple (*Sma*I/*Not*I/*Hind*III, *Xba*I/*Sma*I/*Not*I, *Spe*I/*Sma*I/*Not*I, *Acc*I/*Sma*I/*Not*I) enzyme digestions and hybridization with 7a and 7b. This hybridization result clarified the discrepancy seen in 4. 1. 2. All three fragments hybridized with both 7a and 7b probes, but the 9-2 fragment hybridized faintly with the 7a probe, suggesting the translocation occurred within the 7a probe (Fig. 4. 11A and 11B). Using the fragment sizes estimated using the regression technique of the section 2. 3. 2, the restriction map was constructed and placed 7a adjacent to 7b and confirmed the translocation breakpoint to be within 7a (Fig. 4. 12). The translocation breakpoint occurred only 250 bp away from the *Hind*III site. This small fragment was not detected on the Southern blots of patient *Hind*III genomic

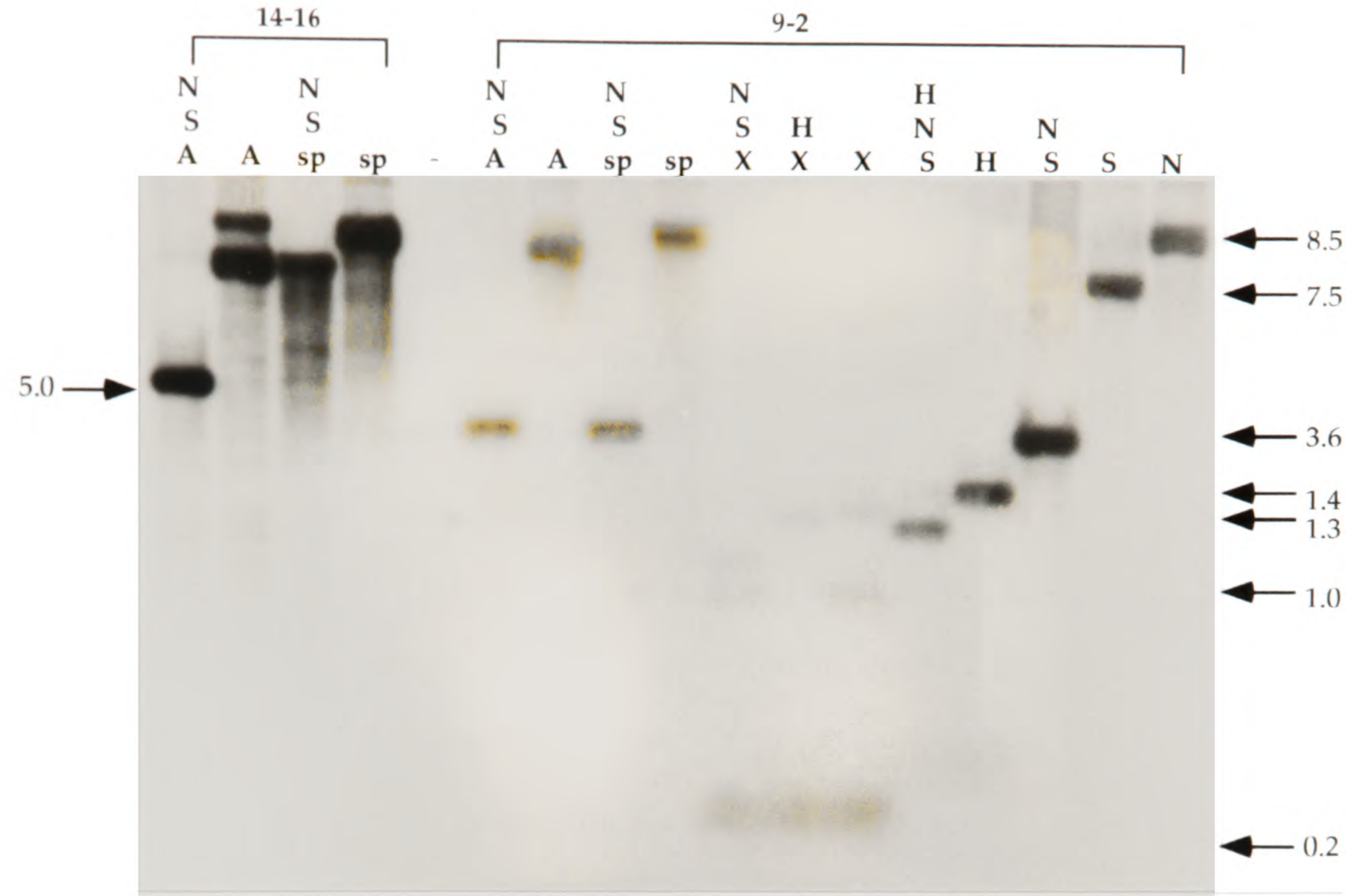



Fig. 4. 11A. Hybridization of 7a to multiple digests of patient genomic DNA, first cloned in phage and then subcloned into plasmids. Arrows indicate small fragments resulting from the translocation. Legend: N, *Not*I; H, *Hind*III; X, *Xal*I; S, *Sma*I; sp, *Spe*I, and A, *Acc*I. Fragment lengths shown in kb.

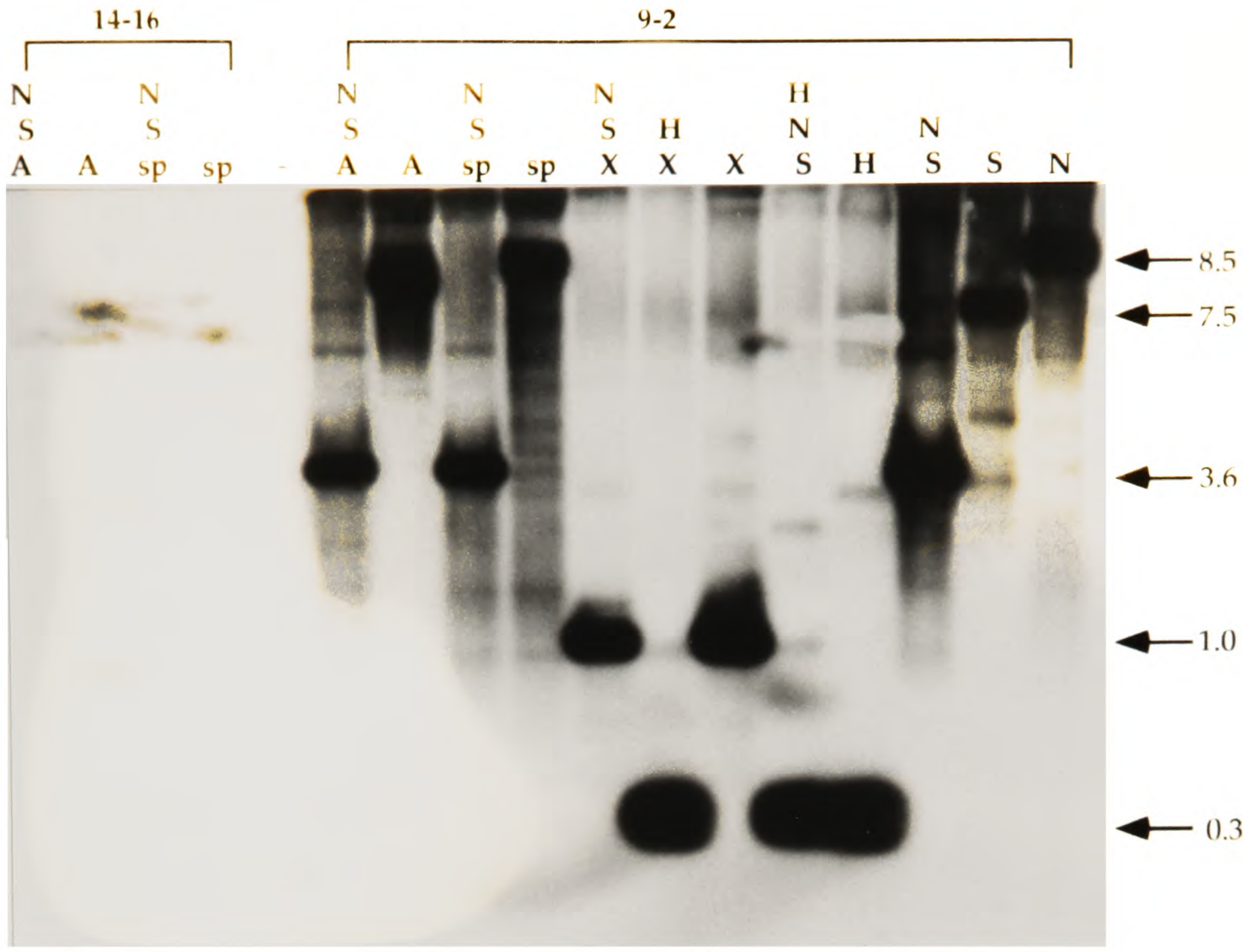
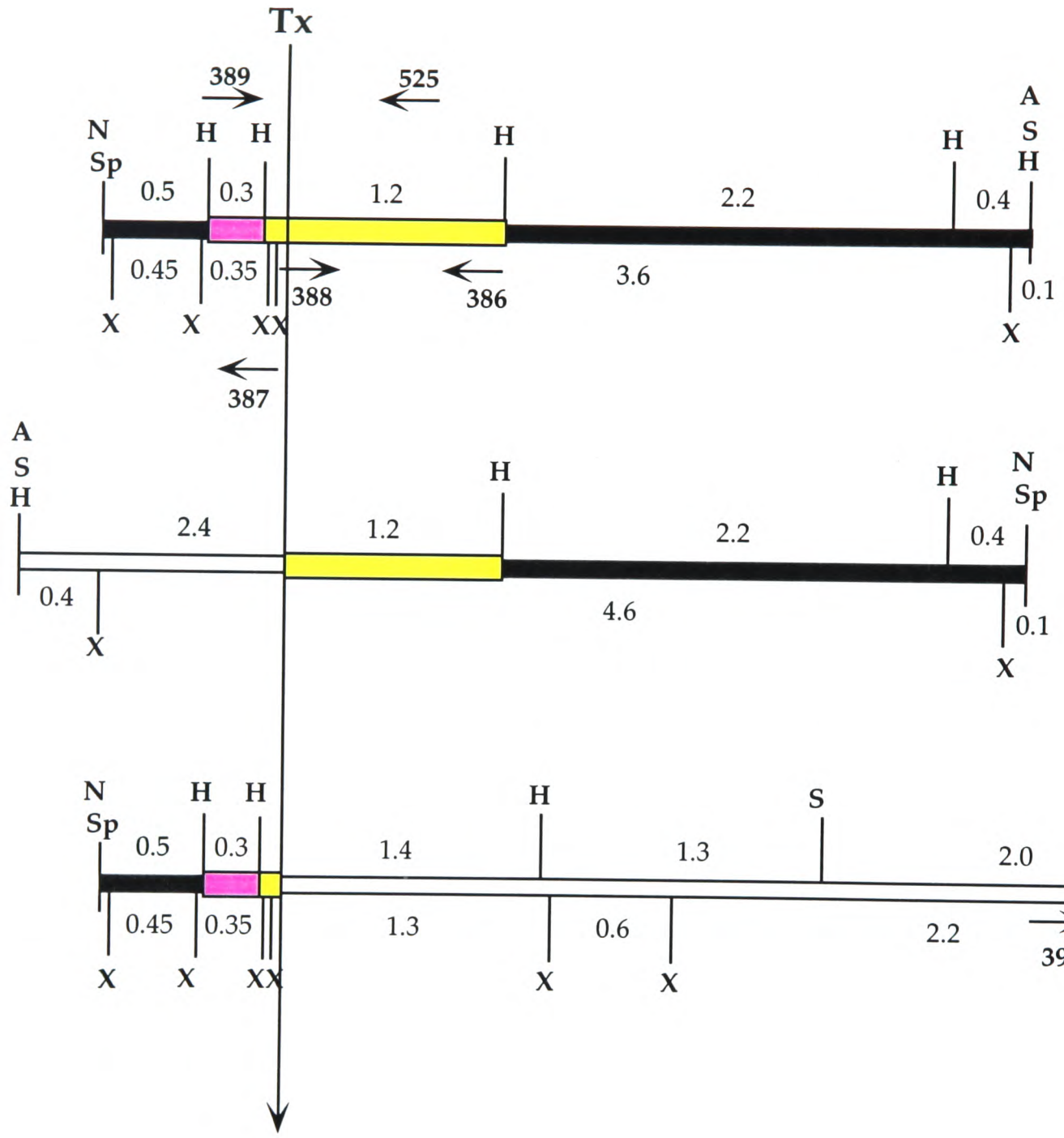



Fig. 4. 11B. Hybridization of 7b to multiple digests of patient genomic DNA, first isolated into phage clones and then subcloned into plasmid. Arrows indicate small fragments resulting from the translocation. Legend: N, *Not*I; H, *Hind*III; X, *Xal*I; S, *Sma*I; sp, *Spe*I, and A, *Acc*I. Fragment lengths shown in kb.




**7 - 6**
4.6 kb

**14 - 16**
5.0 kb

**9 - 2**
5.5 kb

Fig. 4. 12. Relationships of subclones 7a and 7b derived from SuperCos G2, and restriction maps of the junction fragment subclones of der(8) (9-2) and der(X) (14-16), and the normal X chromosome (7-6). The restriction enzymes used were *Bgl*II (B), *Hin*dIII (H), *Sma*I (S), *Xba*I (X), *Acc*I (A), and *Spe*I (Sp). Filled and open bars indicate chromosome X and 8, respectively. Location of primers, 389, 525, 387, 388, and 386, used in identifying the translocation breakpoint at the nucleotide level. Also shown for PCR amplification (Fig. 4. 13) is a pair of chromosome 8 specific primer, 392 and 391. "Tx" indicates the position of the translocation breakpoint. Fragment lengths shown in kb.

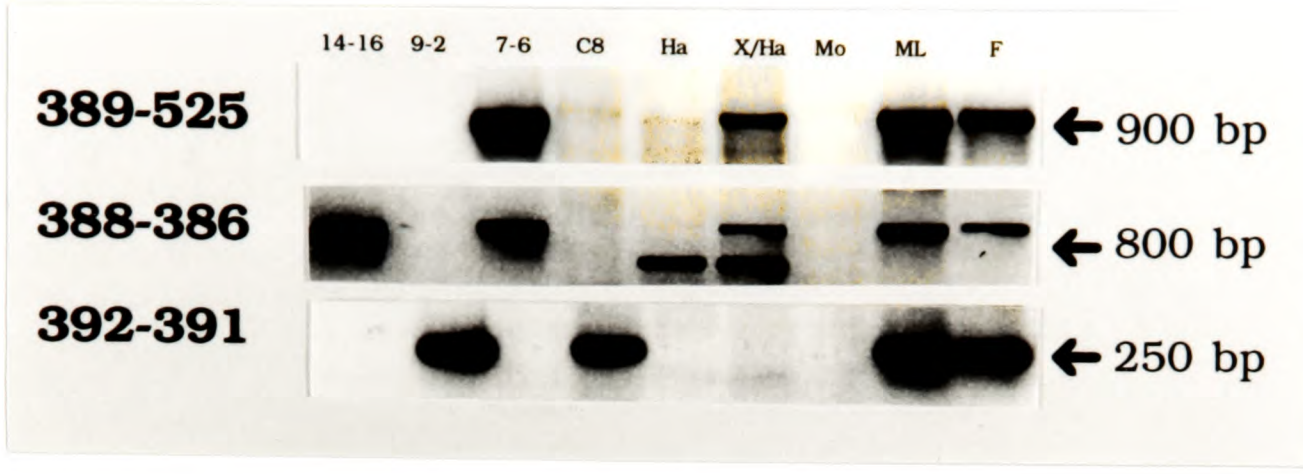
DNA isolated from ML. The small fragments were clearly shown on multiple digests of patient genomic DNA subcloned as high copy in plasmids (Fig. 4. 11A and 11B).

### 4. 3. 3 Sequencing the Junction Fragments

The patient *Bgl*II junction fragments, 7-6 from the normal X chromosome, 9-2 and 14-16 from either der(X) or der(8) were sequenced beginning from the ends, using the *p*Bluescript vector primers (Table 2. 1B). *Hind*III fragment 7a and 7b, identified in the same plasmid subclone from cosmid G2, were separately subcloned in *Hind*III digested *p*Bluescript and also sequenced using vector primers. Sequence walking was achieved by constructing new primers at each walking-step (Fig. 4. 12) and identified the translocation breakpoint at the nucleotide level. The translocation was identified at the point where the sequences of the junction fragments from the derivative chromosomes differed from the normal of sequence. The PCR products using these primers also confirmed the translocation breakpoint (Fig. 4. 13).

The size of intron-1 was determined by sequencing the long PCR products of 9 kb and 10 kb, between the 5′ end of the 9-2 junction fragment and the 3′ end of the GRPR exon-1 and between the 3' end of the 14-16 junction fragment and the 5' end of the GRPR exon-2, respectively (Fig. 4. 14 and 15). The intron-1 was about 24 kb and the whole GRPR gene is then about 27 kb. This determined the orientation of the junction fragments with respect to the centromere and telomere shown in the figure 4. 15. This also indicated the 9-2 junction fragment was located on the der(8) and the 14-16 junction fragment on the der(X) (Fig. 4. 16).

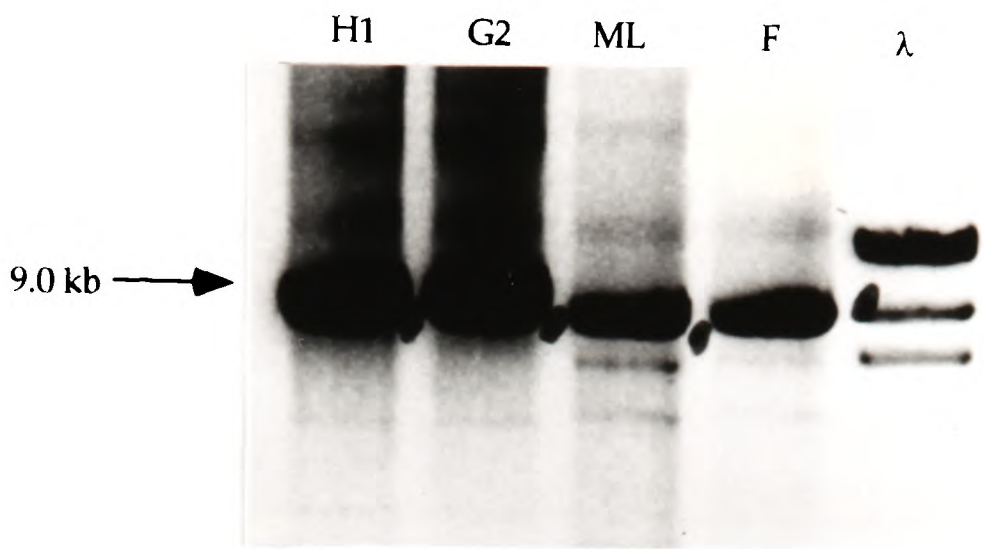
(A)



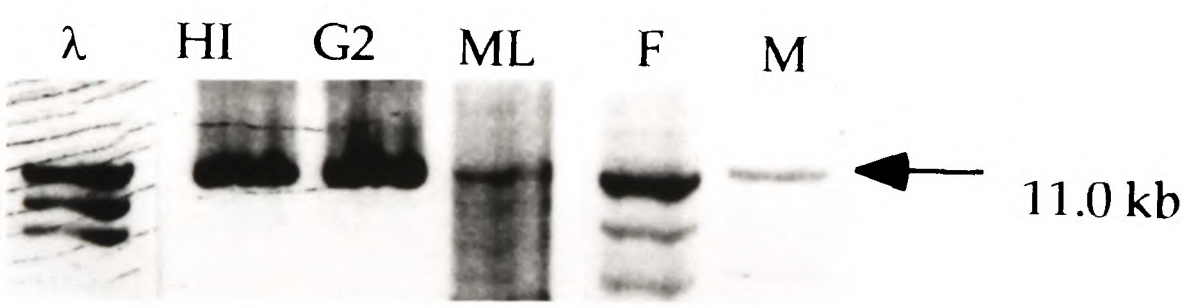

(B)

| | 389-525 | 388-386 | 389-387 | 392-391 |
|---|---|---|---|---|
| 7-6 | + | + | + | - |
| 14-16 | - | + | - | - |
| 9-2 | - | - | + | + |
| c8 | - | - | - | + |

Fig. 4. 13. (A) Confirmation of the translocation breakpoint by PCR amplification. Subclone 7-6 was amplified using PCR primers derived from both 7a and 7b. Subclone 9-2 was amplified using PCR primers from 7b only, and subclone 14-16 from 7a only. (B) All primers are indicated in Figure 4. 12, In the table, + and - indicate the presence and absence of a PCR product. Legend: F, normal female; Ml, patient; Ha, hamster; X/Ha, human X/hamster cell hybrid (HORL9X) (Heisterkamp, 1982); Mo, IR mouse (Nabholz *et al.*, 1969); c8, also known as C4a, human chromosome 8/mouse cell hybrid (Heisterkamp, 1982).




**Ia1 - 641**




**642 - Ia2**

Fig. 4. 14. Long PCR amplifications for orienting the junction fragments with respect to the GRPR gene. The primers used are indicated in Figure 4. 15.

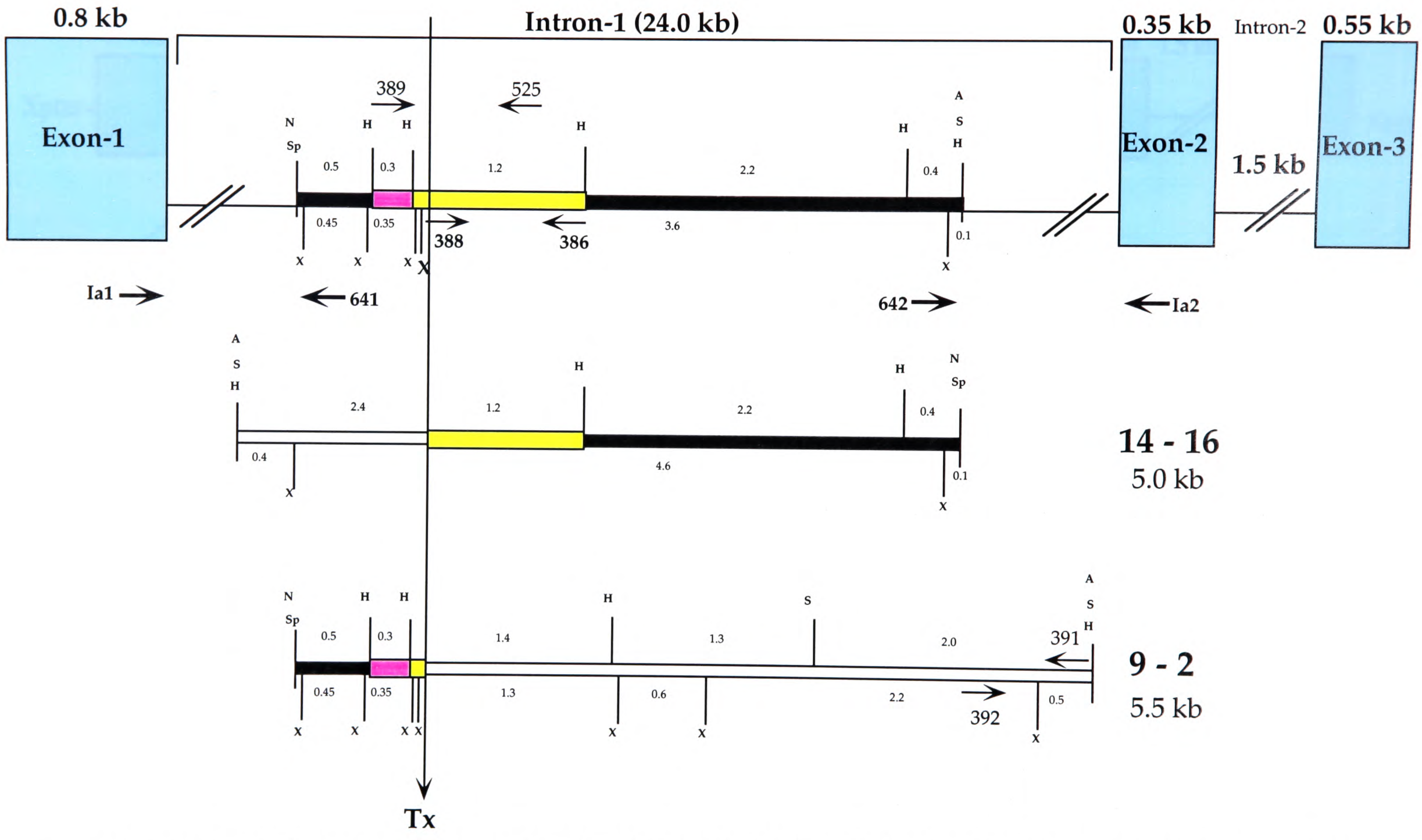



Fig. 4. 15. Relationships of subclones 7a and 7b derived from SuperCos G2, and restriction maps of the junction fragment subclones of der(8) (9-2) and der(X) (14-16), and the normal X chromosome (7-6). The restriction enzymes used were *Bgl*II (B), *Hin*dIII (H), *Sma*I (S), *Xba*I (X), *Acc*I (A), and *Spe*I (Sp). Filled and open bars indicate chromosome X and 8, respectively. Location of primers used in determining the orientation of the junction fragment.

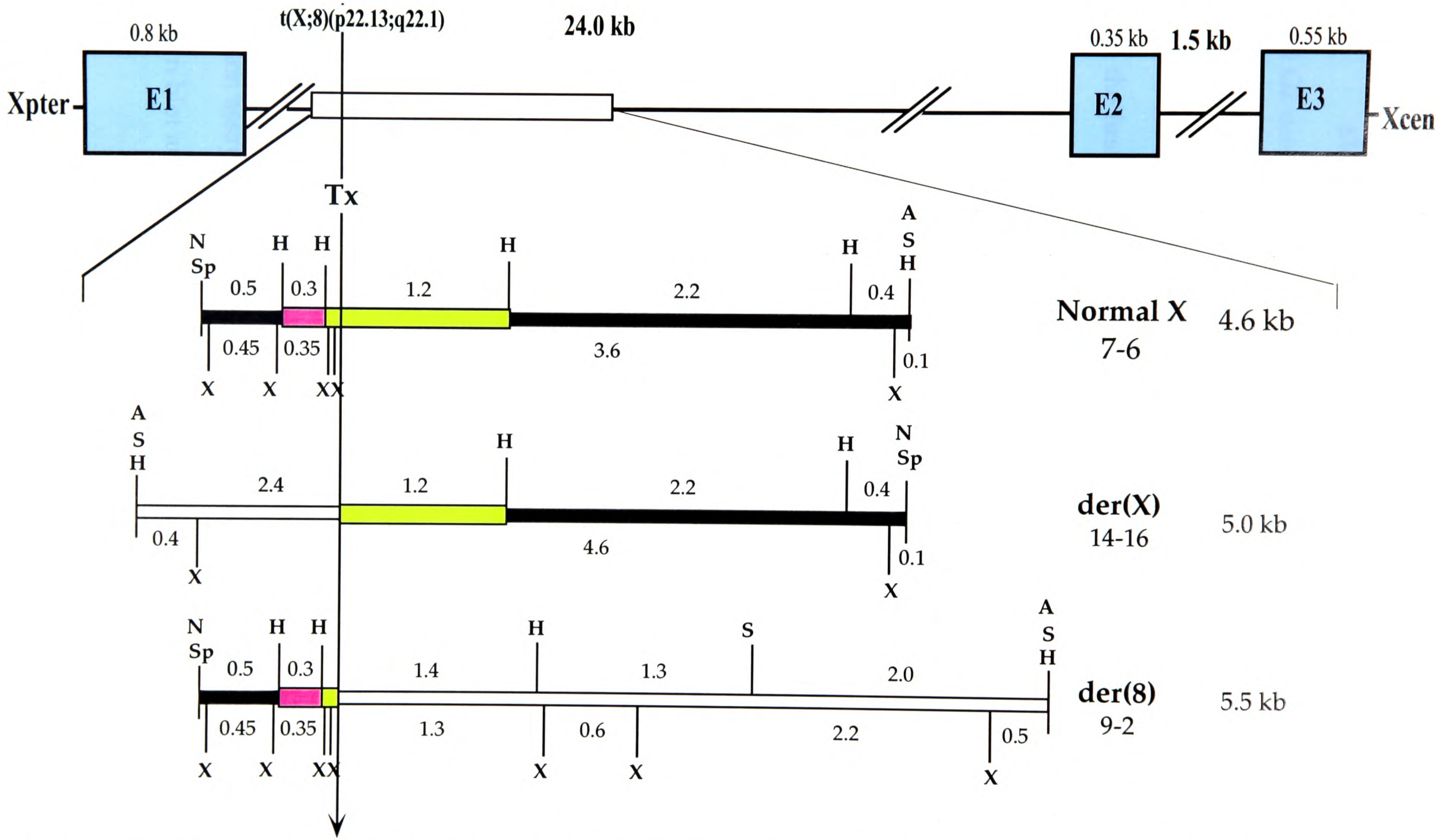



Fig. 4. 16. Identification and orientation of the two derivative chromosomes with respect to the *Bgl*II junction fragments, indicating 9-2 is on der(8), 14-16 on der(X) and 7 6 is on the normal X chromosome. Filled and open bars indicate chromosome X and 8, respectively.

## 4. 4 SUMMARY

Cosmid G2 containing the GRPR gene and the translocation breakpoint were digested with *Hind*III and subcloned into *p*Bluescript. One of the subclones contained two *Hind*III co-ligated inserts, 7a and 7b, both of which were found to lie in the first intron of the GRPR gene in the cosmid contig. They identified the translocation junction fragments in patient genomic DNA digested with *Pst*I, *Eco*RV, *Bgl*II, and *Eco*RI. These junction fragments were cloned in phage by hybridizing 7a and 7b to the patient phage library constructed with *Bgl*II digests of the patient genomic DNA and subsequently subcloned in plasmid. 7a identified a 5.0 kb junction fragment (14-16), and 7b a 5.5 kb junction fragment (9-2). Restriction mapping and sequencing of the junction fragments and fragments 7a and 7b, revealed the translocation breakpoint to be within fragment 7a. Long PCR identified the two derivative chromosomes with respect to the *Bgl*II junction fragments, indicating 14-16 is on der(X), 9-2 is on der(8), and for GRPR that exon-1 is on der(X) and exon-2 and -3 are on der(8).

The GRPR gene is about 27 kb in length and contains a substitution polymorphism in exon-2. The first intron of the gene is about 24 kb in length and is disrupted by the translocation breakpoint. Using a somatic cell hybrid expression assay, the GRPR gene was shown to escape X-inactivation. Northern blot analysis indicated expression in pancreas.

# CHAPTER V

## CHROMOSOME 8 AND THE SYNDECAN-2 (SDC2) GENE

Having identified the gene disrupted by the translocation breakpoint on the X chromosome, the goal was then to identify any gene in the vicinity of or disrupted by the translocation breakpoint on chromosome 8. Of the DNA fragments available for cloning, ranging from YACs at about 500 kb to plasmids at 5 kb, cosmids at 40-45 kb are best suited for cloning genes. The average gene density of genes in the nuclear genome is also about one in 40-45 kb. The specific goal was therefore to identify cosmids containing both the translocation breakpoint and a gene on chromosome 8.

Two strategies were followed. The first was to examine by FISH whether existing cosmids in the region spanned the translocation breakpoint (5. 1). The second strategy was to (a) generate chromosome 8 specific primers by sequencing through from the cloned chromosome X translocation junction fragment to chromosome 8 using der(X) and der(8) plasmid constructs as templates (5. 2. 1), (b) use these primers to isolate chromosome 8 specific YACs by PCR screening of the ICI YAC library (5. 2. 2), (c) subclone the YACs to construct a cosmid library and (d) generate a cosmid contig, (e) use Exon Trapping (5. 3. 2) and direct cDNA library screening (5. 3. 1) to isolate cosmids containing the translocation breakpoint and chromosome 8 gene(s), and (f) identify gene(s) through database searches using genomic DNA sequence (5. 4. 3).

## 5.1 EXAMINATION OF EXISTING CHROMOSOME 8 COSMIDS

At the time of the analysis the most important candidate region for multiple exostoses genes was 8q24 (later confirmed to contain the multiple exostoses gene, EXT1 (Ahn *et al.*, 1995)). However, Drs. K. Yoshiura and Y. Nakamura of the Cancer Institute, Tokyo, Japan, have a patient with the translocation breakpoint in the 8q23 region, t(8;13)(q23.2;q32), associated with multiple exostoses (Yoshiura *et al.*, 1994). Similar to the translocation breakpoint seen in ML, their patient's translocation breakpoint, is proximal to the 8q24. It was therefore useful to analyze their cosmids for map position relative to ML's translocation breakpoint.

Eight cosmids which had mapped by FISH to 8q23 region (Emi *et al.*, 1992) and seven cosmids adjacent to the 8q23 translocation breakpoint were sent for analysis. Among these, four were proximal and three were distal, but none of their cosmids spanned the translocation breakpoint in 8q23 in their analysis (Fig 5. 1). The translocation breakpoint occurred between the proximal marker, cCI8-553 and 600 kb from the distal marker, cCI8-1512, which displayed a translocation junction fragment in the patient's DNA digested with *Bss*HII (Yoshiura *et al.*, 1994).

To determine the relative position of the two translocations, two of their cosmids, flanking the breakpoint on each side, cC18-553 and cC18-1512, were analyzed by FISH on ML's lymphoblastoid metaphase chromosome spreads. The same FISH hybridization pattern in both, cC18-553 and cC18-1512, confirmed that the translocation breakpoint of patient ML occurred at the different site from that of the Japanese patient. Both markers which

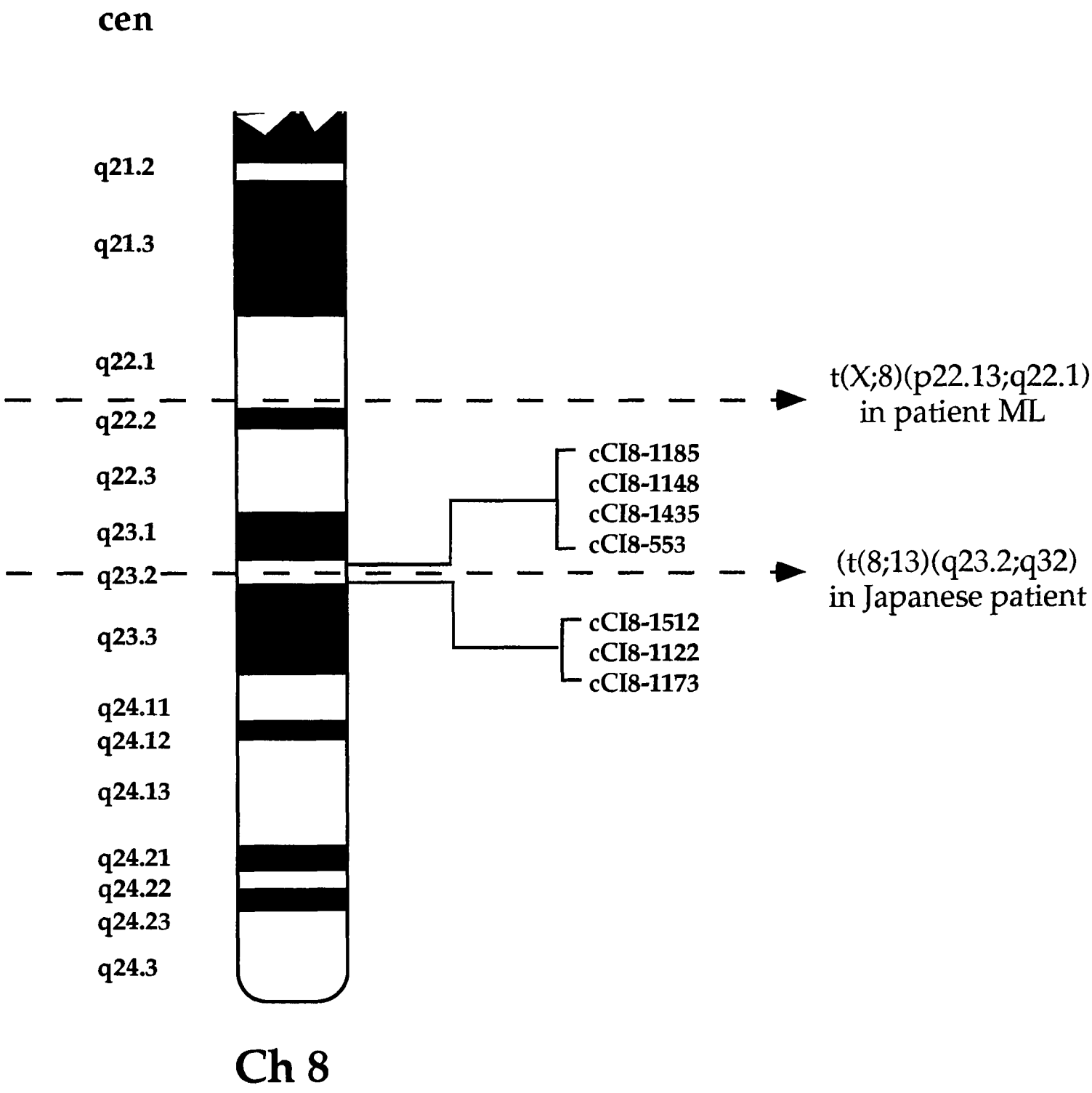



Fig. 5. 1. Relationship between the chromosome 8 markers and the translocation breakpoint of a Japanese patient with multiple exostoses (Yoshiura *et al.* 1994) and patient ML.

were hybridized on der(X) of patient ML positive with the X specific alphoid marker also confirmed that both markers are distal to the ML's translocation breakpoint (data not shown). All other markers were also identified to be distal to ML's translocation breakpoint, suggesting the order of "8cen - t(X;8q22) (ML) - t(8q23;13) (Japanese patient) - 8q24 (EXT1) -8qter" (Fig. 5. 1).

## 5. 2 PHYSICAL MAPPING OF CHROMOSOME 8q22.1

No other chromosome specific cosmids in the region were available for analysis and it was necessary to construct a cosmid contig. To achieve this, chromosome 8 specific probes in the region of ML's translocation breakpoint were generated to isolate YACs from the ICI YAC library (5. 2. 2). Sequencing through the translocation breakpoint allowed isolation of chromosome 8 specific primers (5. 2. 1). This sequencing also allowed characterization of the translocation at the DNA level (5. 3 and 5. 4).

### 5. 2. 1 Isolation of the Chromosome-8 Specific primers

To generate chromosome-8 specific primers, the region of the breakpoint was sequenced by primer walking from chromosome X to 8. Primer 33 was designed from the sequence generated by X-specific primer 65 using the 14-16 plasmid as a template, and primer 67 was designed from the sequence generated by X-specific primer 64 using the 9-2 plasmid template (Fig. 5. 2). These two primers (33 and 67) were confirmed to be chromosome 8 specific by PCR amplification of DNA from a somatic cell hybrid specific for chromosome 8 (Heisterkamp *et al,* 1982) but not for chromosome X (Fig 5. 3). This PCR product was 240 bp in length and designated as a sequence tagged

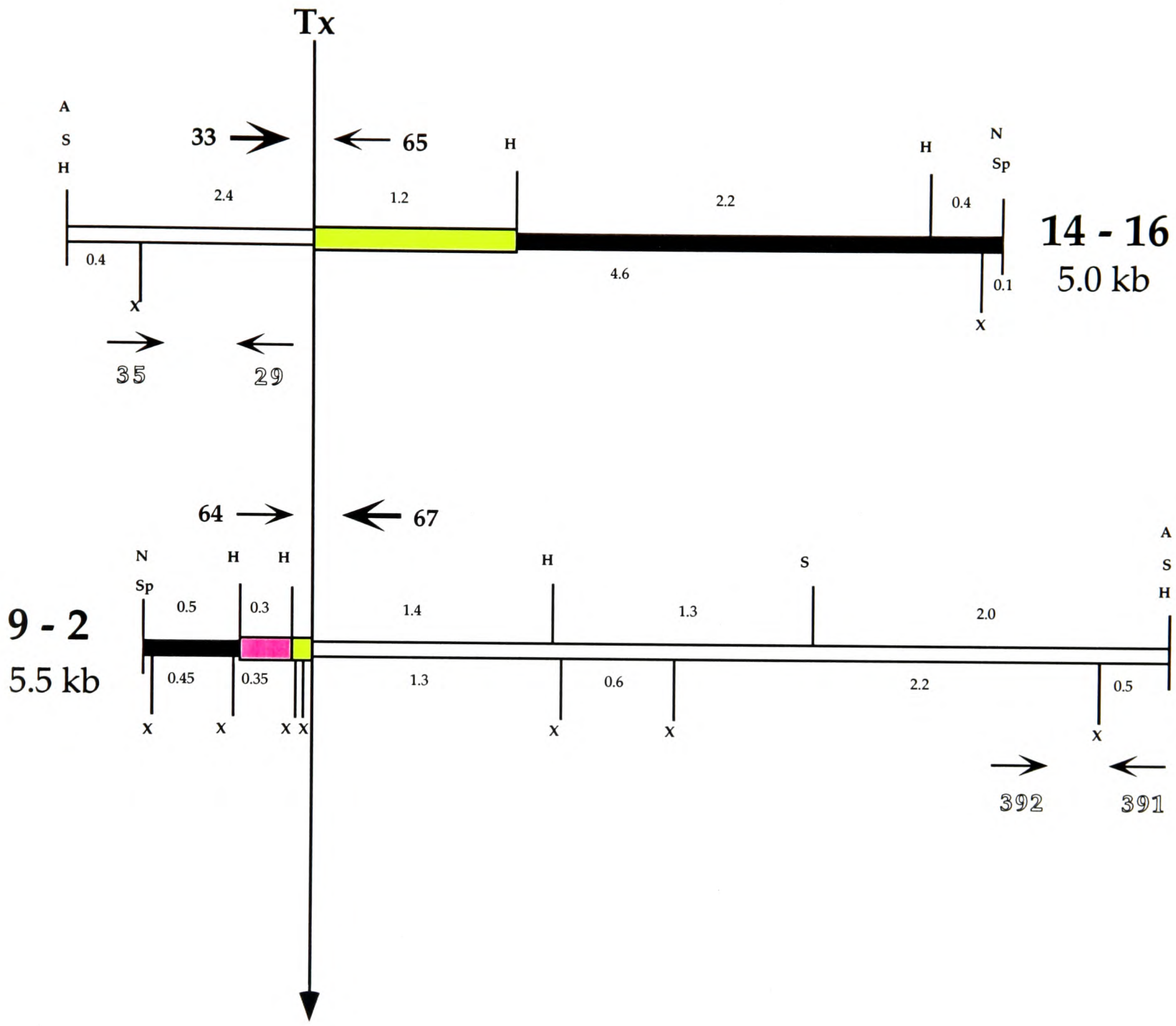



Fig. 5. 2. Schematic diagram of chromosome 8 specific primers 33 and 67. The primers were generated from sequencing across the translocation breakpoint from the two chromosome X junction fragments 14-16 and 9-2, using primers 65 and 64, respectively. Also shown for future reference are two pairs of primers, 35-29 and 392-391.

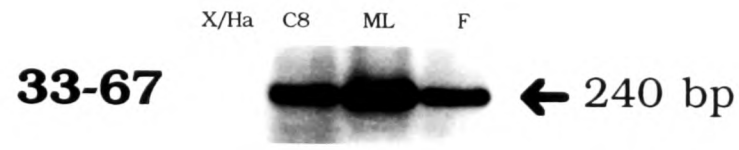



Fig. 5. 3. Confirmation that primers 33 and 67 are specific to chromosome 8. PCR amplification using X/Ha, human X/hamster cell hybrid (HORL9X) (Heisterkamp *et al.*, 1982), and c8, also known as C4a, human chromosome 8/mouse cell hybrid (Heisterkamp *et al.*, 1982) as templates. Amplification of primers 33-67.

site (STS)3367 or STS3367 STS3367 contained the chromosome 8q22.1 region of the translocation breakpoint and was used as a probe for hybridization to Southern blots of genomic DNA isolated from ML. The probe identified translocation junction fragments derived from several restriction enzymes, confirming that it crossed the breakpoint (Table 5. 1 and Fig. 5. 4).

| | Normal Female | Patient ML | |
|---|---|---|---|
| | Normal Frag. Size | Normal | JF size |
| *Hind*III | 2.0 kb | 2.0 kb | 1.4 kb |
| *Eco*RI | 7.5 kb | 7.5 kb | 3.2 kb |
| *Pst*I | 1.2 kb | 1.2 kb | 2.8 kb |
| *Bgl*II | 5.4 kb | 5.4 kb | 5.0 kb |

Table 5. 1. Translocation junction fragment sizes detected by hybridizing probe STS3367 to Southern blots of patient genomic DNA digested with *Hind*III, *Eco*RI, *Pst*I and *Bgl*II.

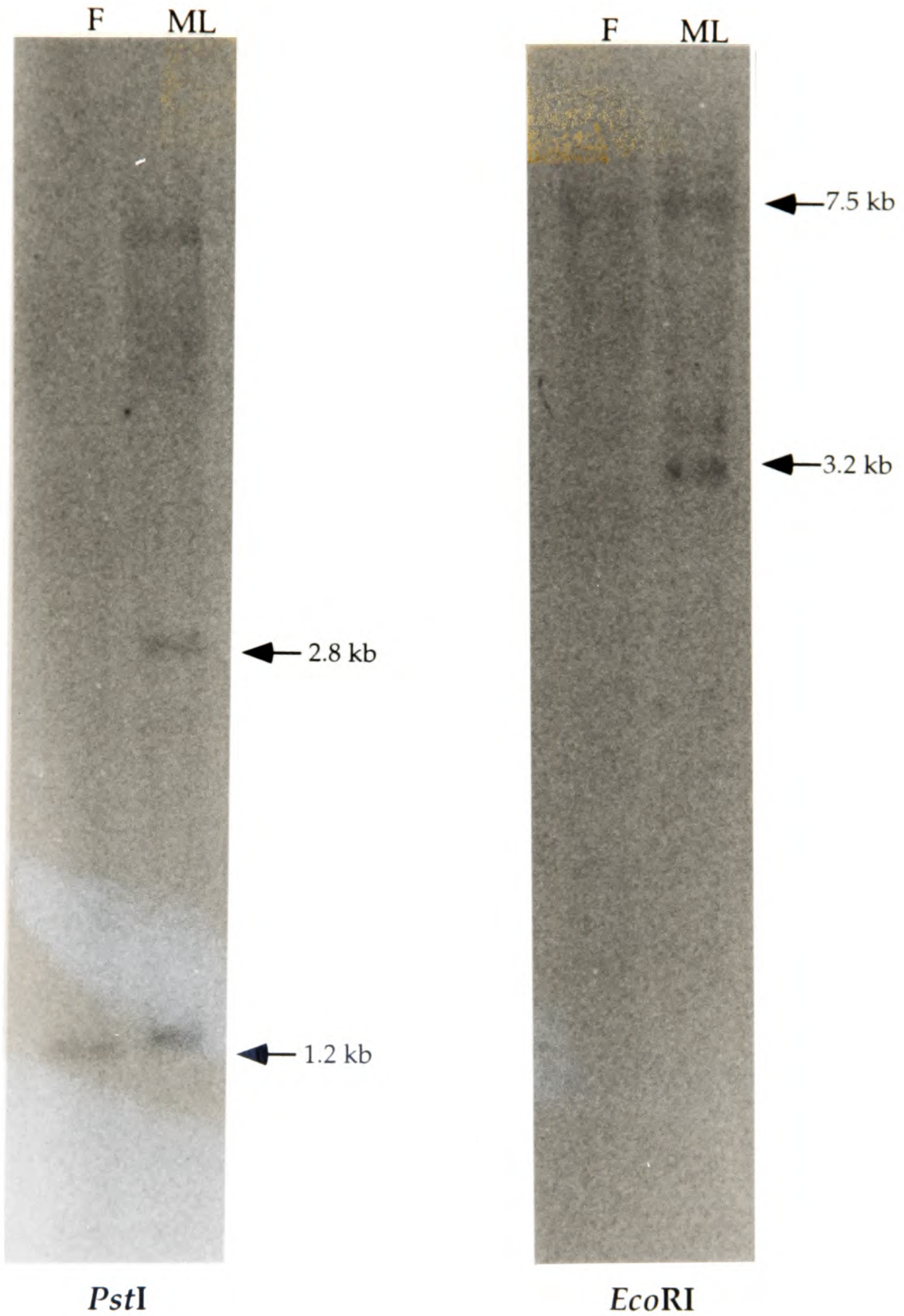



Fig. 5. 4. Detection of the translocation junction fragment in patient genomic DNA digested with *Eco*RI and *Pst*I and hybridized with probe STS3367. Legend: F, normal female; ML, patient.

### 5. 2. 2 Isolation of YACs

Primers, 33 and 67, were used to isolated three YACs, 2B, 13B, and 38B, by PCR screening of the ICI YAC library (Fig. 5. 5A) (2. 12. 1). Three YACs were fractionated by PFGE (2. 3. 3), visualized by EtBr staining, sized using yeast chromosome standards (Fig. 5. 5B), and Southern blotted. The blot was hybridized with STS3367 to confirmation chromosome 8 specificity (Fig. 5. 5C). YAC2B is about 450 kb, YAC13B about 500 kb, and YAC38B 100 kb.

### 5. 2. 3 Construction of Cosmid Libraries

Construction of the cosmid libraries from YACs was performed using the same procedure as with chromosome X (Chapter III, page 130). Three YACs were confirmed to be non-chimeric by FISH (Fig. 5. 6). Two YACs about 500 kb in length (2B and 13B) were partially digested with *Mbo*I, ligated to the SuperCos vector, *in vitro* packaged and a library was plated at low density. Each library was replicated on filters, which were hybridized with total human placenta DNA to identify human insert containing cosmids. 288 cosmids were positive and picked into microtitre plates and transferred to membrane filters using a 96-pin device. These filters were hybridized with the STS3367 probe which identified ten cosmids spanning the translocation breakpoint: 2E, 7B, 3H, 7G, 9G, II4A, 7C, 7C, III3F, and 9A.

## 5. 3 IDENTIFICATION OF THE SYNDECAN-2 (SDC2) GENE

To identify gene(s) in cosmid DNA, cosmids spanning the translocation breakpoint were used to screen a testis cDNA library by hybridization and as templates for exon trapping. All sequences specific to chromosome 8

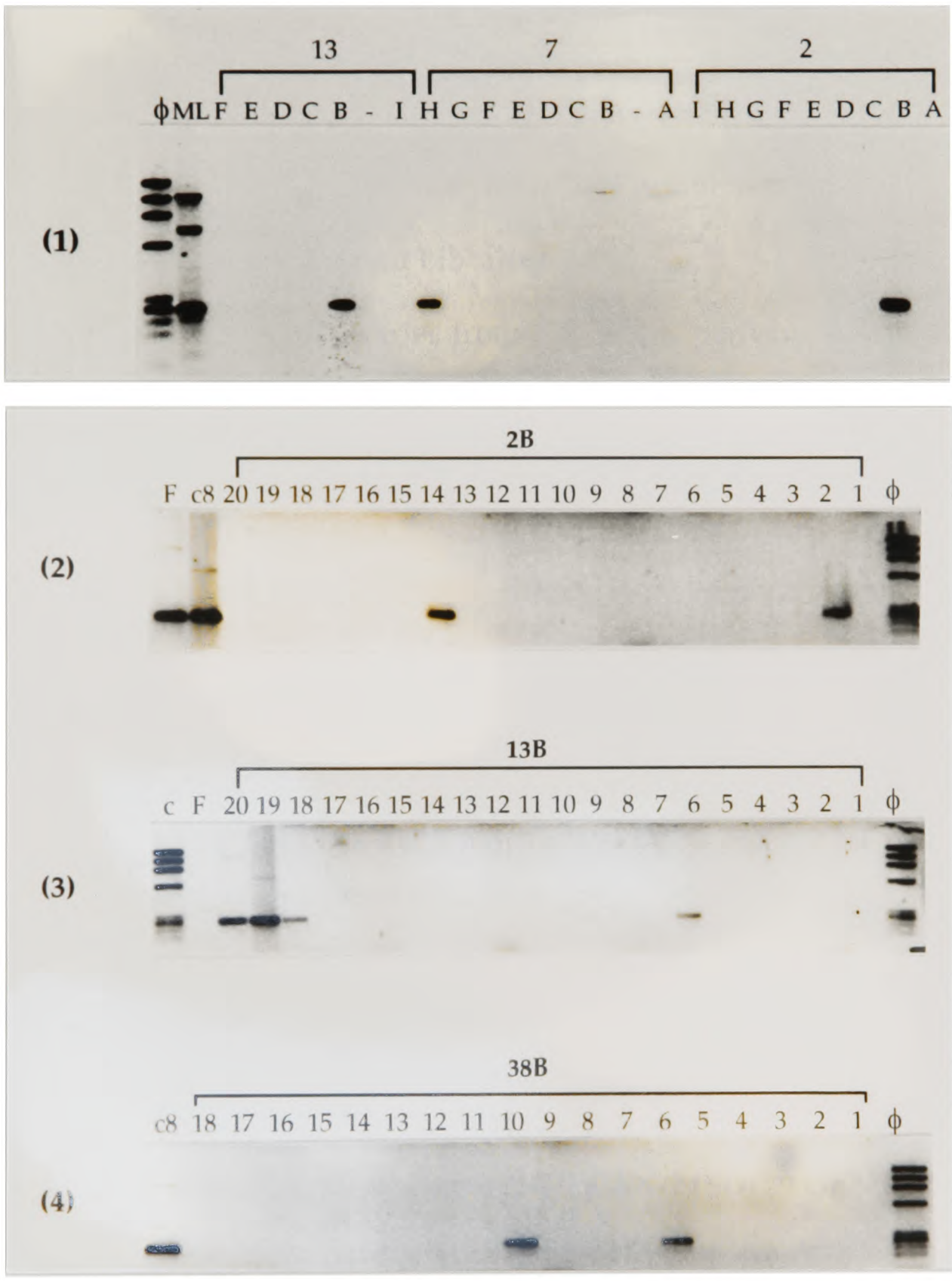



Fig. 5. 5A. Isolation of chromosome 8 YACs 2B, 13B and 38B by PCR screening of the ICI YAC library using primers 33 and 67. (1) Primary YAC pool screening, (2), (3) and (4) tertiary YAC screening. "c" indicates a control containing no DNA and c8, also known as C4a, human chromosome 8/mouse cell hybrid (Heisterkamp, 1982).

*et al.*

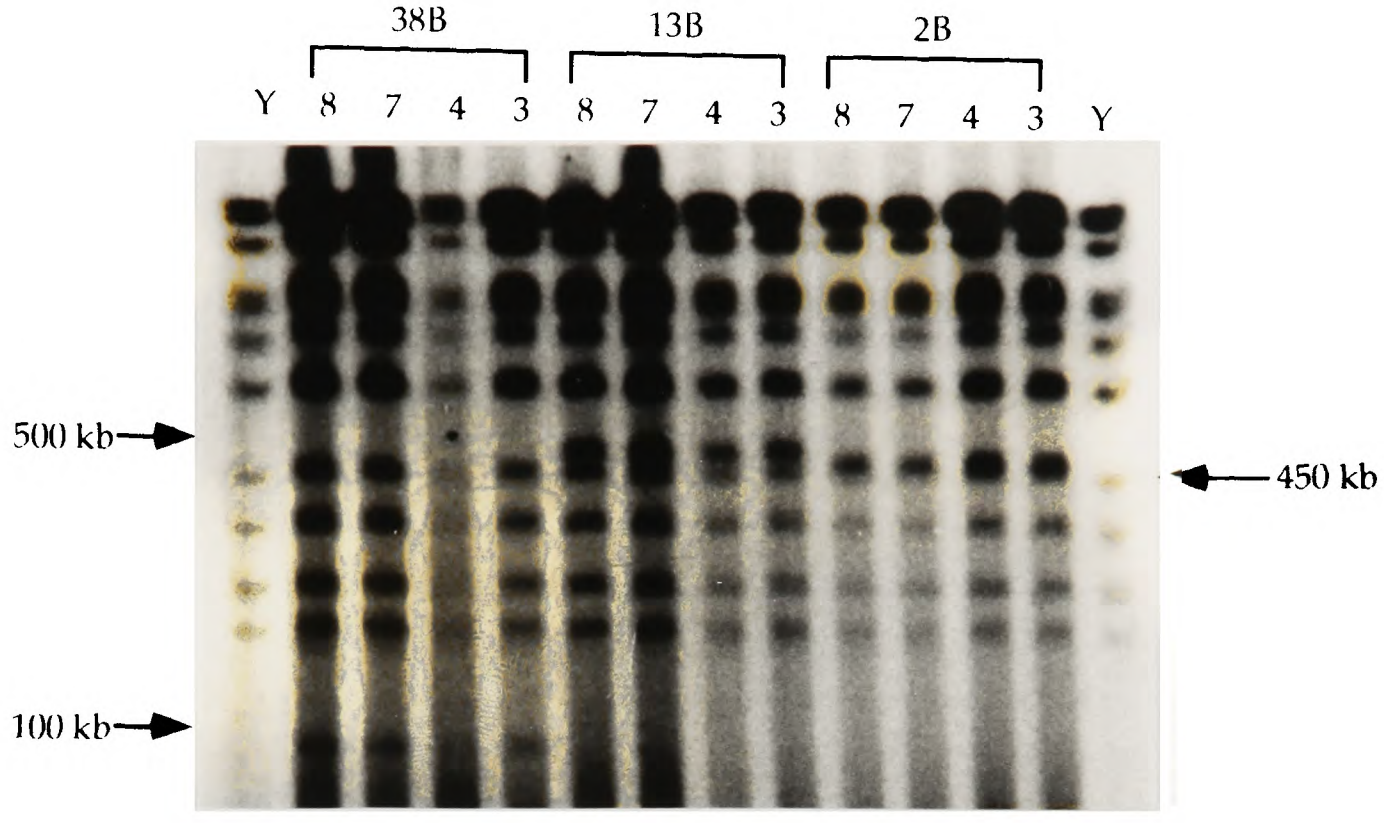



Fig. 5. 5B. Size determination of chromosome 8 YACs, 2B, 13B, and 38B by PFGE fractionation of the YACs and sizing against known yeast chromosomes.

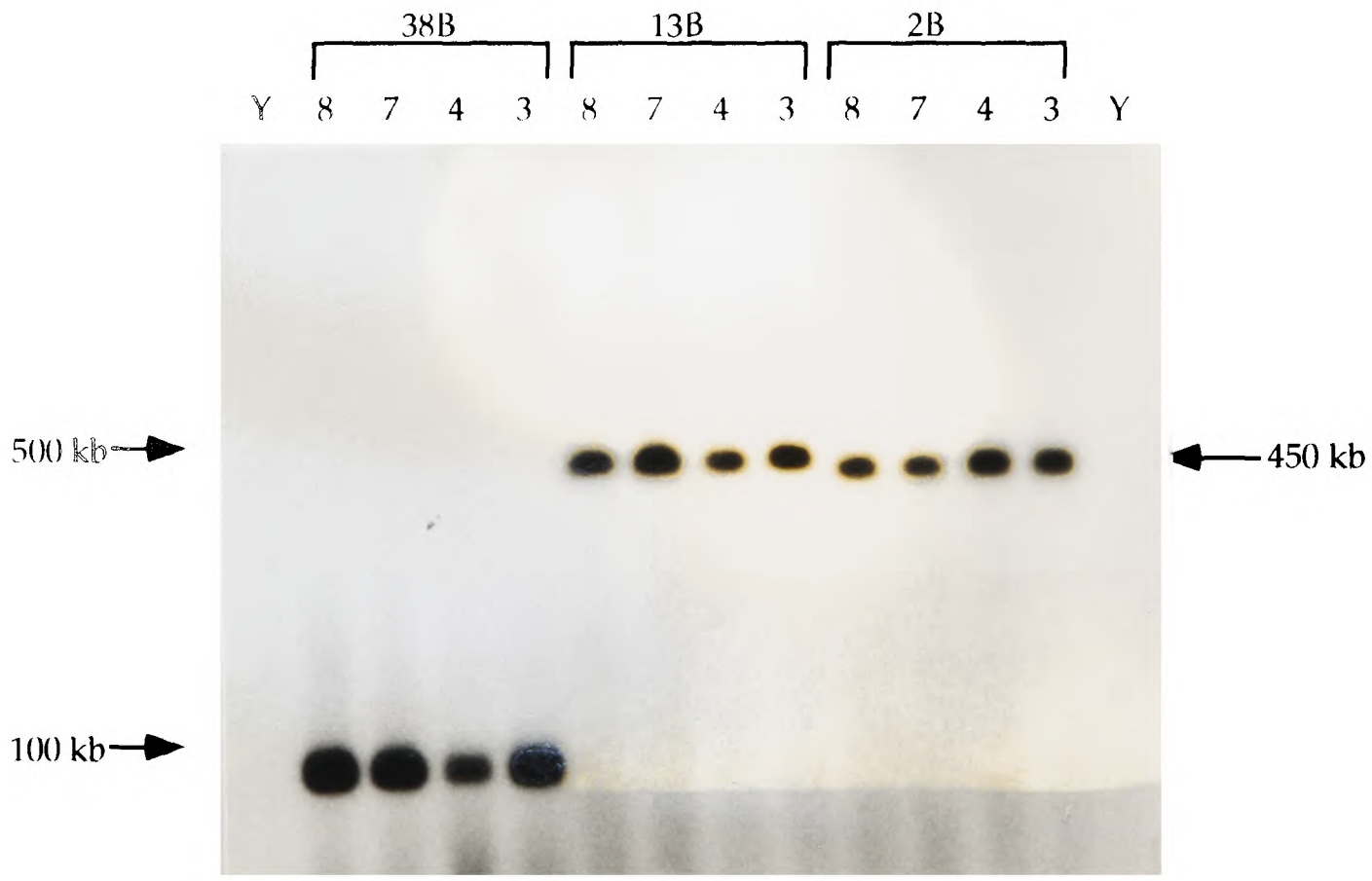



Fig. 5. 5C. Confirmation that YAC 2B, 13B and 38B are specific to chromosome 8. Hybridization by chromosome 8 specific probes STS3367 to the Southern blot of the PFGE fractionated gel of Fig. 5. 4B.

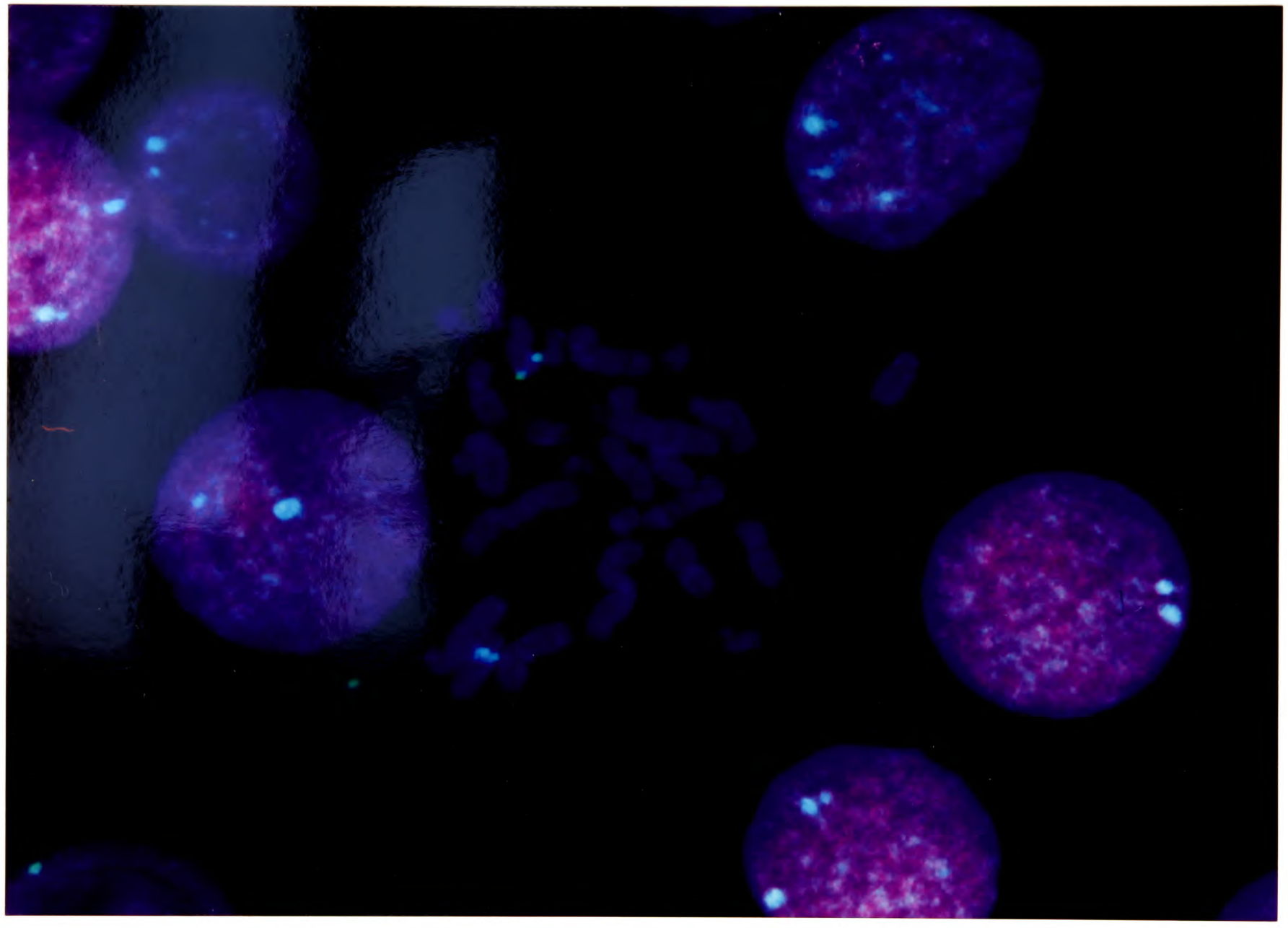


Fig. 5. 6. FISH confirmation that YAC, 2B, is non-chimeric by FISH on a normal female lymphoblastoid metaphase chromosome spread. Two hybridization signals indicate YAC, 2B, hybridized to normal chromosome 8 only.

generated from both der(X) and der(8) were examined for the presence of open reading frames using GRAIL, BLASTX, and BLASTN.

### 5. 3. 1 Testis cDNA Screening

Two cosmids, 2E and 7B (5. 2. 2), out of ten cosmids spanning the translocation breakpoint, were used as hybridization probes to screen a human testis cDNA library. Positive cDNAs (A1, A3, B1, B4, B5, B6, C1, C2, D1, and D4) were purified and subcloned. B1, B4, C1, D1, and D4 were sequenced and overlapped with each other (Fig. 5. 7). The BLASTN search identified the Syndecan-2 gene with 100% homology using sequence from each of these cDNA fragments (SDC2). This gene is also known as human heparan sulfate proteoglycan or fibroglycan [MIM: J04621] (McKusick, 1994). It had been previously localized by Marynen *et al.* by *in situ* hybridization and hybridization to DNA of somatic cell hybrids to 8q22-q24 (Marynen *et al.*, 1989), in the region of ML's translocation breakpoint.

### 5. 3. 2 Construction of a Cosmid Contig

The translocation breakpoint and SDC2, were both found within a 40 kb region (i. e. the size of a cosmid). To determine whether the SDC2 gene was disrupted by the translocation breakpoint, a high resolution (or density) cosmid contig was constructed.

PCR products were generated using primers from the exon sequences of SDC2 and the end clones (Fig. 5. 8) (2. 6. 5) derived from the end subcloning of cosmids which hybridized to either or both SDC2 or STS3367. These end clones were non-radioactively labeled and used as probes for hybridization to the Southern blots containing *Hind*III digested cosmid DNA. A cosmid

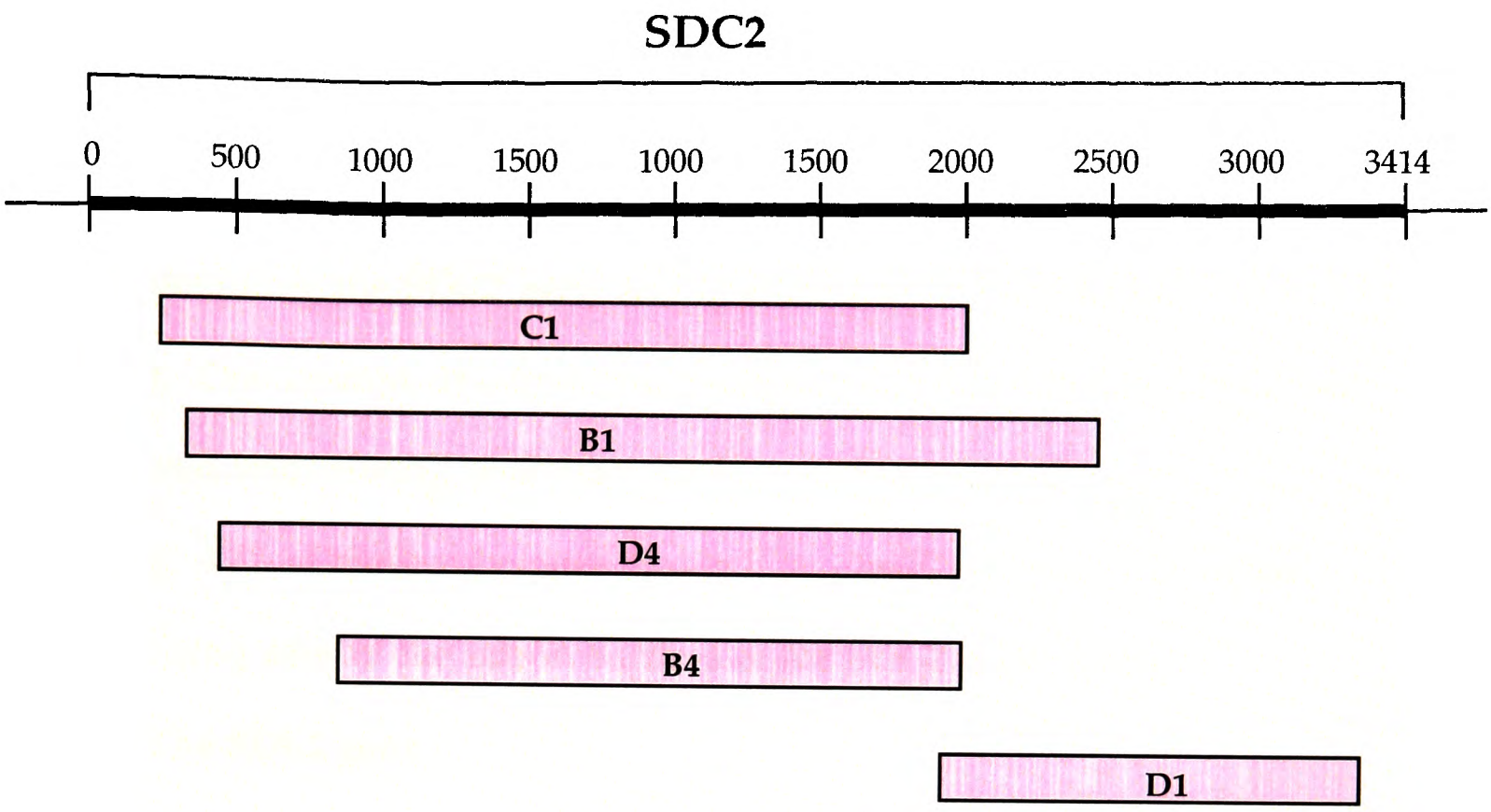



Fig. 5. 7. Relative positions of cDNA clones, C1, B1, D4, and D1 with respect to the SDC2 gene

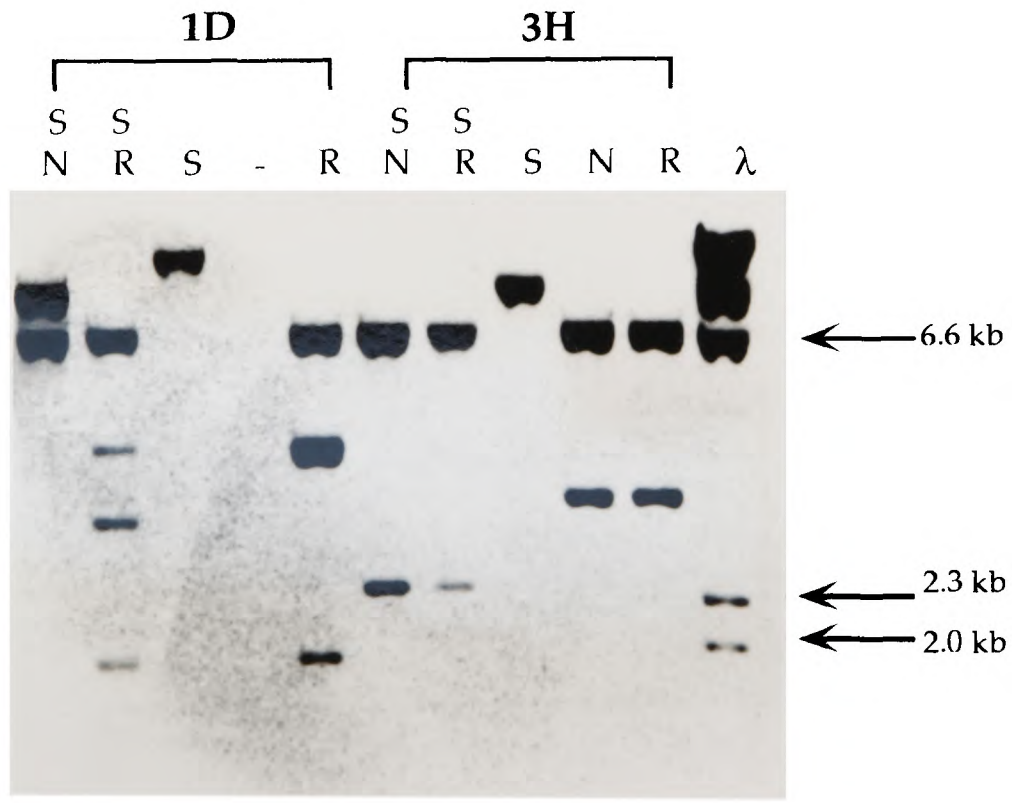



Fig. 5. 8. Examples of cosmid end clones isolated from cosmid 1D and 3H using the method of 2. 6. 5. The SuperCos vector about 7.0 kb was separated with *Eco*RI (R) or *Not*1 (N) digests. The insert-insert ligation was separated with *Sac*I digest.

contig covering this region was constructed using the program, "clone-to-well (2. 6. 5)" (Fig. 5. 9) with one gap. This gap left unresolved the relationship between the SDC2 gene and the translocation breakpoint. The cluster of the cosmids, 4B, 6G, 1D, 1G, 9H, 10D, 3-7c and 3-6c were separated from that of the cosmids containing the SDC2 gene and spanning the translocation breakpoint. One cosmid, 1G, from this cluster and 2E and 3H from the latter were digested with *Hind*III and subcloned into *p*Bluescript (2. 6. 4). Sequencing 40% of these subclones (Table 5. 2) identified four sites of the AG-GT splicing sites of the intron borders of the SDC2 gene (below).

### 5. 3. 3 The SDC2 gene

The SDC2 gene sequence in the database covers 3.4 kb of the 3' end. Its ATG-start and TAA-stop sites were at the position of 588 and 1190nt, respectively. The four introns identified were at 648 (>785 bp), 761 (>150 bp), 894 (>1.7 kb), and 1027 nt (915 bp) (Fig. 5. 7) by sequencing the *Hind*III subclones of cosmid 2E.

In the database the 5′ end of the SDC2 gene was not resolved, and the 3 YACs isolated with STS3367 contained 4 exons also from the 3' end. The isolation of the 5′ end of the gene was attempted using a 5′ RACE method (2. 2. 8). However, the second round of the nesting-PCR did not amplify any PCR product. Before resolving this, the contig confirmed that the translocation breakpoint did not disrupt the SDC2 gene. The cosmids, 10D, 6E, 1D, 1G, 10F, 11E, and 1C contain the 3′ end of the SDC2 gene. The cosmid, 2E, 3H, and 7G span both the SDC2 gene and the translocation breakpoint whereas 9A only the translocation breakpoint. The translocation occurred

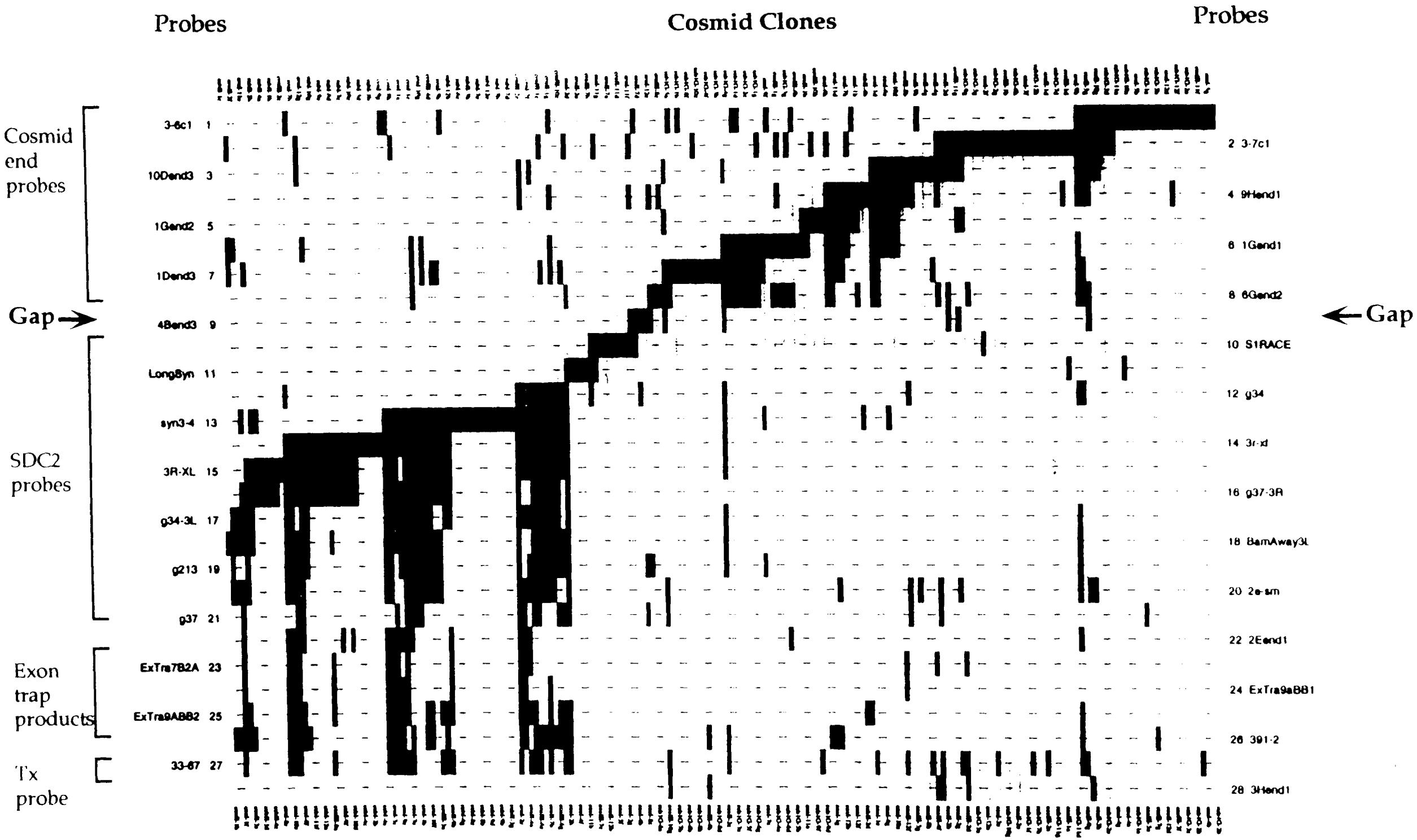



Fig. 5. 9. Cosmid contig with one gap generated by the 'clone-to-well' computer program. Tx probe: probe containing the translocation breakpoint.

| Name | Size | GRAIL | BLASTN | BLASTX |
|---|---|---|---|---|
| G266 | 0.25 kb | ( - ) | Cosmid D22S280, partial Alu repeat 57%/80 bp (0.00058) | Hypothetical Protein(yeast) 47%/20 aa (0.0003) |
| G23 | 0.7 kb | ( - ) | cDNA clone 5′ | Retinoid "X" receptor gamma |
| G216 | 0.82 kb | ( - ) | Human gene | Proline rich protein |
| G37 | 1.0 kb | HSPG | Chloroplast ATP | cDNA clone 5′ |
| G223 | 1.2 kb | ( - ) | C. elegans cosmid ZK1193 56%/116 bp (0.93) | US5432155 patent 87%/8 aa (0.83) |
| G34 | 1.3 kb | HSPG | Gene for PRPP synthetase | Signalling protein |
| G4 | 1.4 kb | HSPG | HSPG | HSPG 3′ end |
| G10 | 2.4 kb | "Marginal" | Human tissue plasminogen activator 88%/170 bp (4.1e-130) | Alu subfamily 84%/167 aa (3.9e-23) |
| G1 | 2.6 kb | "Good" | cDNA clone 3′ | HSPG 3' end |
| G11 | 3.7 kb | "Good" | LINE protein ORF2 87%/220 bp (4.6e-63) | Unknown protein (Homo sapiens) 79%/73 aa (8.3e-38) |
| G18 | 5.3 kb | ( - ) | Enhancer region | Alu repeat |
| G16 | 5.5 kb | ( - ) | H. repeat polymorphism gene | Basement membrane collagen |
| Sub29 | 754 bp | ( - ) | Cosmid CA repeat DXS1211 region 76%/132 bp (1.1e-44) | Testis-specific RNA 63%/11 aa (0.57) |
| Sub39 | 359 bp | ( - ) | Cytochrome oxidase 43%/90 bp (0.9) | Alpha apoprotein 32%/25 aa (0.5) |
| Sub44 | 885 bp | ( - ) | Cosmid 11p area 78%/192 bp (2.7e-60) | Alu subfamily 70%/61 aa (8.6e-23) |
| Sub36 | 664 bp | ( - ) | ATP synthetase alpha subunit 76%/98 bp (6.1e-13) | None |
| Sub66 | 420 bp | ( - ) | Myosin light chain-2 67%/68 bp (0.12) | Cytomegalo virus 62%/16 aa (0.99) |
| Sub48 | 480 bp | ( - ) | Haemophilus influenza 59%/102 bp (0.34) | Patent US4912038 62%/14 aa (0.99) |
| IG8FT3 | 527 bp | ( - ) | HSPG 100%/62 bp (1.5e-15) | HSPG 100%/20 aa (0.0001) |
| IG7CT7 | 341 bp | ( - ) | N. meningitides 1bpA gene 70%/67 bp (0.0084) | P-glycoprotein-1 42%/21 aa (0.91) |
| IG1H | 579 bp | ( - ) | Mycobacterium tuberculosis 61%/101 bp (0.0013) | Mycobacterium tuberculosis 41%/70 aa (4.3e-18) |
| IG7G | 581 bp | ( - ) | Cytoplasmic filament protein 72%/48 bp (1.0) | Haemophilus influenza 36%/92 aa (3.5e-17) |
| IG6CT7 | 409 bp | ( - ) | Acinetobacter (pobA)gene) 78%/46 bp (0.17) | None |

5. 2. Sequence analysis of *Hind*III subclones from cosmid 2E, IG, and 3H.

| Name | Size | GRAIL | BLASTN: N(P) | BLASTX: N(P |
|---|---|---|---|---|
| IG6AT7 | 574 bp | ( - ) | H. annuus SF3 gene 73%/56 bp (0.10) | US5508384 patent 30%/30 aa (0.99) |
| IG9DT7 | 560 bp | ( - ) | B. olevaceae PC2 gene 63%/85 bp (0.12) | T cell receptor alpha V3 55%/20 aa (0.93) |
| IG9DK | 356 bp | ( - ) | C. elegans cosmid B0416 64%/71 bp (0.79) | Tyrosine kinase receptor 40%/25 aa (0.99) |
| IG9DT3 | 438 bp | ( - ) | H. Wilms tumour gene 68%/228 bp (6.9e-47) | Protein-lysosine Kinase 36%/36 aa (0.59) |
| IG5AT3 | 490 bp | ( - ) | Cosmid 19q13 area 87%/41 bp (0.0051) | Proline rich C.elegans 40%/22 aa (0.99) |
| IG5AT7 | 559 bp | ( - ) | MHC alu element 87%/111 bp (4.4e-45) | H. Alu subfamily SX 78%/38 aa (1.3e-16) |
| IG6C | 445 bp | ( - ) | H. dystrophin gene 78%/60 bp (5.4e-0.5) | None |
| IG6BT3 | 701 bp | ( - ) | Transposon-like element 80%/47 bp (0.019) | Kallikrein 75%/12 aa (1.0) |
| IG6BT7 | 569 bp | ( - ) | Transposon-like element 80%/47 bp (0.019) | Kallikrein 71%/17 aa (0.99) |
| IG1BT3 | 525 bp | ( - ) | C. elegans cosmid B0361 | Kallikrein 75%/12 aa (0.54) |
| IG9AT7 | 380 bp | ( - ) | Cosmid DXS679 region 75%/48 bp (0.59) | Neurotoxin 75%/8 aa (0.94) |
| 3H2ET3 | 390 bp | "Good" | Xq28 region 70%/220 bp (1.2-80) | LINE-1 Reverse transcriptase 87%/50 aa (4.1e-42) |
| 3H2ET7 | 402 bp | "Good" | DXS6791 region 73%/270 bp (1.1e-90) | LINE protein 90%/52 aa (6.2e-27) |
| 3H3FT3 | 540 bp | ( - ) | C. elegans 72%/55 bp (0.17) | msg 2 protein 43%/23 aa (0.41) |
| 3H3FT7 | 273 bp | ( - ) | Fragile X 81%/40 bp (0.99) | Murine hepatitis virus 50%/15 aa (0.18) |
| 3H8HT3 | 270 bp | ( - ) | C. elegans cosmid C09G9 66%/72 bp (0.51) | ORF1 European hare syndrome 66%/9 aa (0.003) |
| 3H12D3 | 487 bp | ( - ) | H. Xq23 cosmid U21a12 77%/75 bp (1.5e-84) | Haemophilus influenza 53%/13 aa (0.83) |
| 3H12D7 | 461 bp | ( - ) | DXS6791 region 81%/130 bp (2.0e-36) | ORF4 Porcine respiratory coronavirus 87%/8 aa (0.051) |
| 3H9GT3 | 632 bp | "Good" | H. LINE1 repeat 93%/338 bp (5.7e-119) | Leu zipper protein p40 94%/52 aa (1.5e-27) |
| 3H9GT7 | 418 bp | ( - ) | H. L1 repetitive 93%/136 bp (3.0e-77) | Reverse transcriptase ORF2 87%/49 aa (1.2e-35) |
| Bamnow | 856 bp | ( - ) | H. neurokin a receptor gene 79%/105 bp (2.6e-27) | SSV7 protein (yeast) 25%/60 aa (0.998) |
| 310now | 702 bp | ( - ) | HSPG 94%/278 bp (1.1e-111) | K09e3.7 gene C. elegans 39%/23 aa (0.73) |

Table 5. 2. Continued

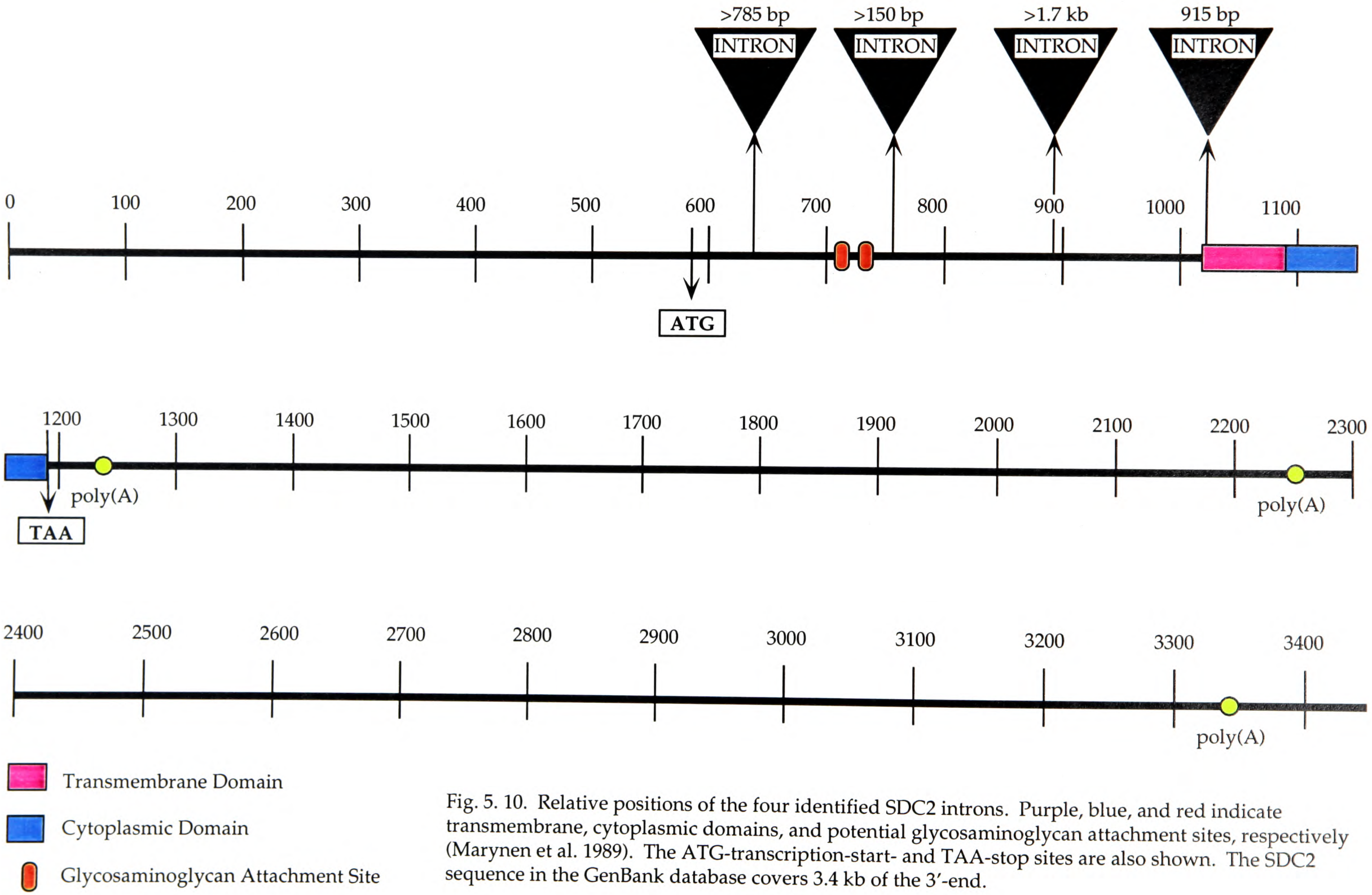



Fig. 5. 10. Relative positions of the four identified SDC2 introns. Purple, blue, and red indicate transmembrane, cytoplasmic domains, and potential glycosaminoglycan attachment sites, respectively (Marynen et al. 1989). The ATG-transcription-start- and TAA-stop sites are also shown. The SDC2 sequence in the GenBank database covers 3.4 kb of the 3′-end.

within a cosmid length, less than 30 kb distal to the 3' end of the SDC2 gene. (Fig. 5. 11). Dual-colour FISH analysis of cosmids in the 5′ end of the contig (10D, 6E, 1D, and 1G) showed them not to be translocated to der(X) (Fig. 5. 12). This result confirmed the orientation of the SDC2 gene with respect to the translocation breakpoint (Fig. 5. 11), and demonstrated that a fusion gene between SDC2 and GRPR was not possible from this translocation (Fig. 5. 13).

## 5. 4 SEARCH FOR GENES OTHER THAN SDC2

All STSs specific to chromosome 8, der(X) and der(8) were examined for the presence of open reading frames using GRAIL and BLASTX. To identify gene(s) in cosmid DNA, cosmids spanning the translocation breakpoint were used for exon trapping.

### 5. 4. 1 Homology Studies of Chromosome 8 STS fragments

Sequencing of STS3367 crossing the translocation breakpoint revealed no significant homology nor evidence for exons. Hybridization of the STS3367 to human multitissue northern blots (Clontech) revealed no obvious transcripts (data not shown). Sequences from a 180 bp fragment generated from the 29-35 primers on the chromosome 8 region of the 14-16 junction fragment and from a 220 bp fragment from the 391-392 primers (Fig. 5. 2) on the chromosome 8 region of the 9-2 junction fragment, identified significant similarities with O-family repeats and LINE sequence, respectively (Table 5. 3). Results of the computer similarity searches by FASTA, BLASTN, and BLASTX (2. 11. 2) of these chromosome 8 regions of the junction fragments

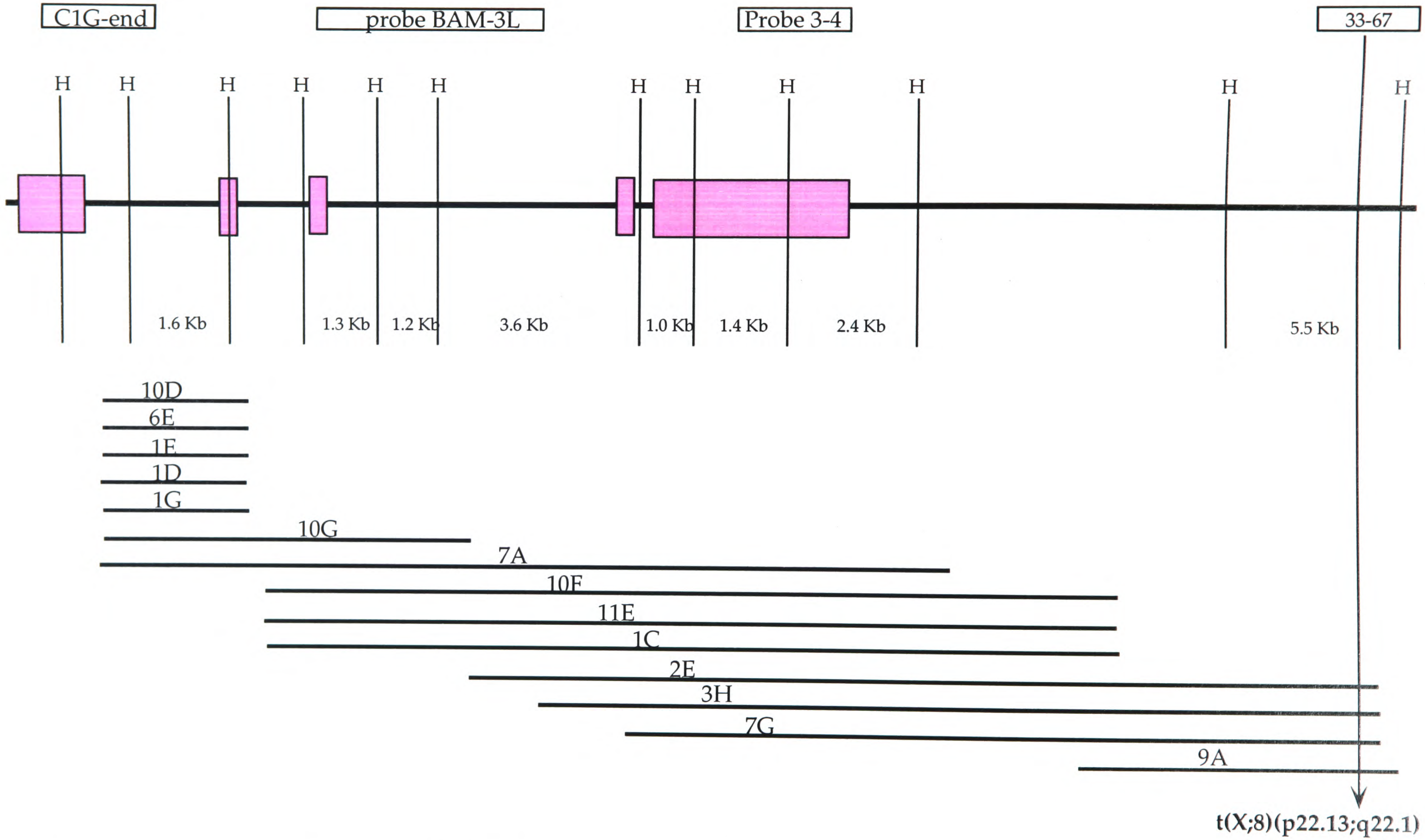



Fig. 5. 11. Location of the SDC2 gene and ML's translocation breakpoint within the chromosome 8 cosmid contig. The contig was constructed from the probes shown at the top of the figure.

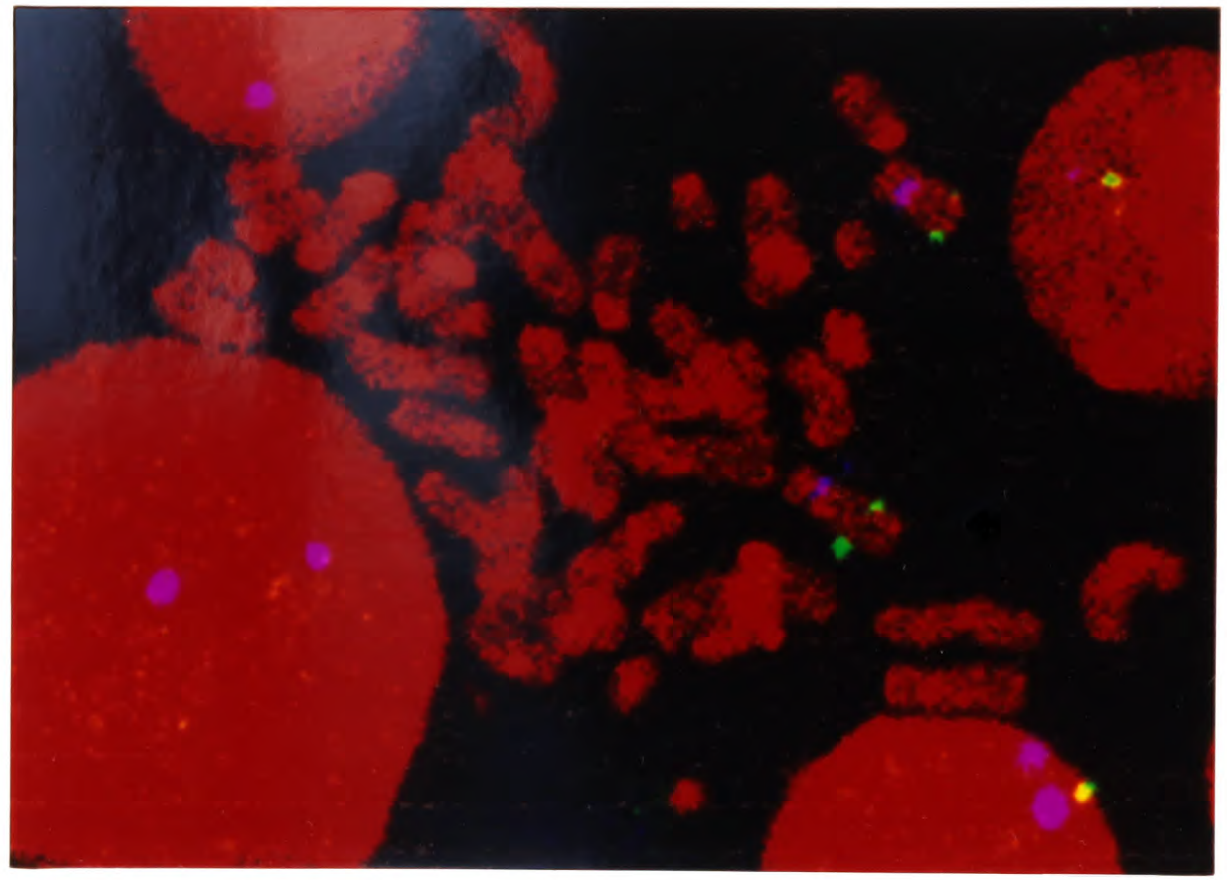


Fig. 5. 12. FISH confirmation of the orientation of SDC2 with respect to the translocation breakpoint. Hybridized to the der(8) with 8 -specific alphoid marker (purple) indicates cosmid 2E (yellow & green) contains the 3′ end of SDC2 and the translocation breakpoint

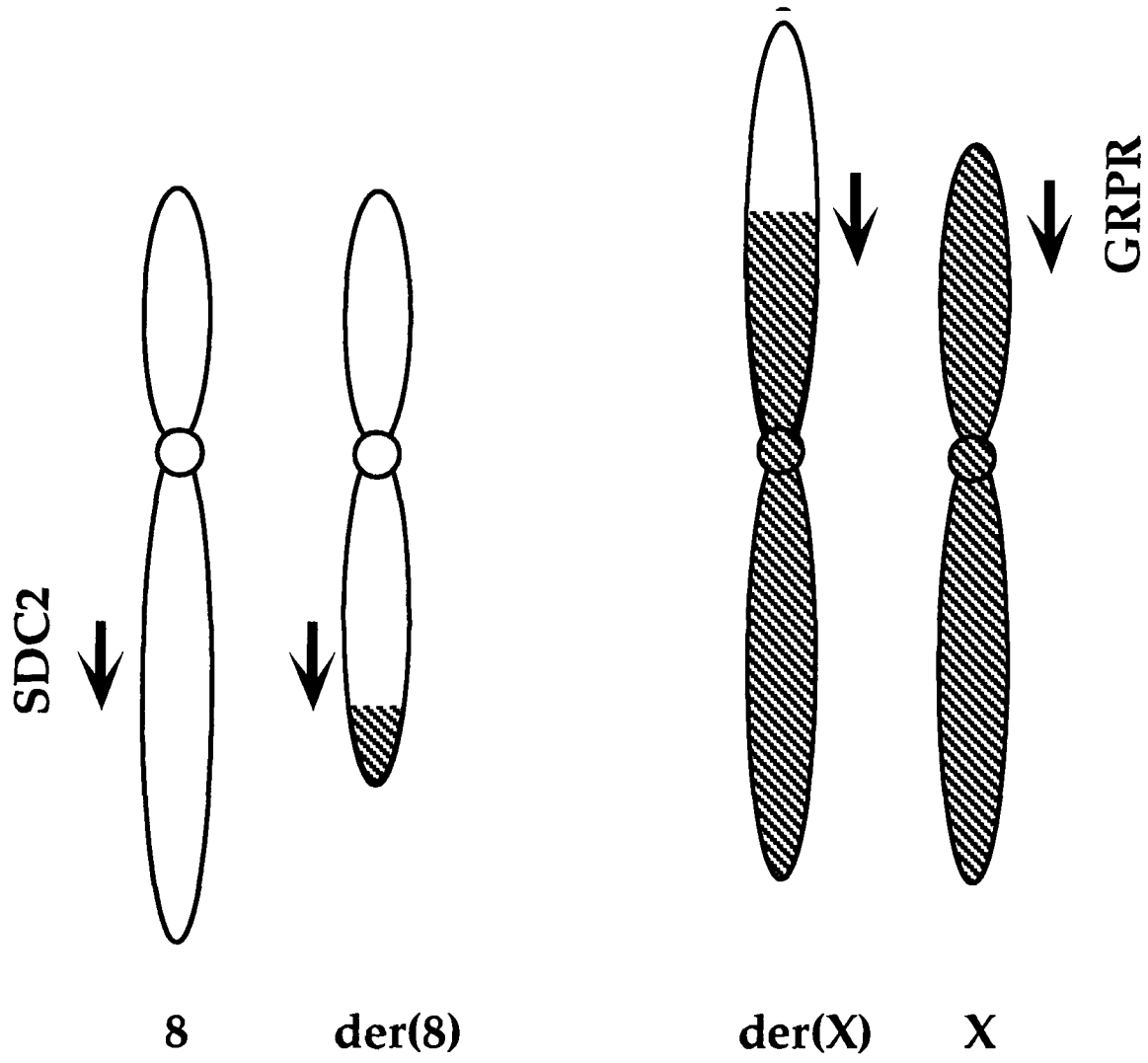



Fig. 5. 13. Schematic diagram showing the orientation of the GRPR and SDC2 genes with respect to the translocation breakpoint and the incompatibility for a fusion gene.

| Ch8 frag. | Size | GRAIL | BLASTN | BLASTX |
|---|---|---|---|---|
| STS3367 | 240 bp | ( - ) | Human oligodendrocyte myelin glycoprotein 72%/36 bp | None |
| 391-392 (9-2/T7) | 220 bp | "Marginal" | Iduronate 2-sulfatase gene 75%/206 bp ($1.7 \times 10^{-46}$) | L1H5′ region 36%/240 aa ($9 \times 10^{-10}$) |
| 29-35 (14-16/T3) | 200 bp | ( - ) | PII-like protein gene 74%/47 bp (0.14) | O-family |

Table 5. 3. Results of computer searches using the GRAIL, BLASTN and BLASTX, for homology with fragments generated from primers in chromosome 8 region junction fragments. GRAIL categorizes levels of homology as excellent, good, or marginal, and (-) for non-homology. For BLASTN and BLASTX, the length of the homologous region is given, the percent of homology and the probability ($p$) of random homology.

are summarized in Table 5. 3. GRAIL searches DNA and amino acid sequence for evidence of open reading frames. BLASTN compares DNA sequences with other sequences in a composite DNA database for similarities. BLASTX translates a submitted query nucleotide sequence in all six reading frames and compares using amino acid sequences with protein sequences in its nonredundant database for similarities.

### 5. 4. 2 Exon Trapping in the 8q22 region

Evidence of genes other than SDC2 in the region of the translocation was investigated by exon-trapping (Buckler *et al.*, 1991) (2. 10) and partial genomic sequence analysis. Two cosmids (2E and 9A) spanning the breakpoint were chosen for analysis; SDC2 was present in one (2E) but not in the other. Four exon-trap products were purified and sequenced, one of which was produced from both cosmids (Table 5. 4). This product was purified and used as a probe to screen a testis cDNA library and a human multitissue northern blot

| Name | Size | GRAIL | BLASTN: P(N) | BLASTX: P(N) |
|---|---|---|---|---|
| 9A-BB1<br>7B-1 | 128 bp | "Good" | M. fervidus RNA(0.16) | Zinc finger (0.053)<br>LINE protein ORF2 50%/16 aa(0.3) |
| 7B-2A | 97 bp | "Marginal" | Alu repetitive element (3e-17) | Ig gamma chain (0.27) |
| 7B-2B | 68 bp | ( - ) | H. Alu repeat(0.48) | C. elegans (0.99) |
| 9A-BB2 | 42 bp | ( - ) | IAP gene long terminal repeat (0.14) | PAH gene (0.98) |

Table 5. 4. Homology searches using the computer programs GRAIL, BLASTN and BLASTX of fragments generated from Exon Trapping from chromosome 8 cosmids (9A and 7B) spanning the translocation breakpoint. GRAIL categorizes levels of homology as excellent, good, or marginal, and (-) for non-homology. For BLASTN and BLASTX, the length of the homologous region is given, the percent of homology and the probability (p) of random homology.

(Clontech). Both results were negative. The consensus splice "AG-GT" sites are found in the genomic sequence (Fig. 5. 14). GRAIL, BLASTN and BLASTX analysis identified the three other exon-trapped products as Alu and LINE-related sequences.

### 5. 4. 3 Homology Studies of Chromosome 8 *Hind*III fragments

Exon trapping (5. 3. 2) did not identify any homologies other than with Alu and LINE-1 sequences. 25 *Hind*III subclones of cosmid 1G, 2E and 3H (Fig. 5. 11) were partially sequenced to generate about 41.7 kb of genomic sequence (Table 5. 2; 5. 3. 3). No evidence was obtained for genes other than SDC2 in this region by GRAIL, BLASTN and BLASTX searches of these sequences (2. 11. 2).

### 5. 4. 4 Expression of the SDC2 gene

These analyses have shown the SDC2 gene, although not directly disrupted, to be most closely associated with the translocation breakpoint. An analysis

CACTGGACTGGCTTACATCCAGTGCATTTAATTTAGCATTTGGAGAAATTGCTCT
GTGACCTGACCGAATGTAGGTCACGTAAATTAAATCGTAAACCTCTTTAACGAGA

GGCTGCCATCACTGAATAGTTTGGCACTGTGCTGTCACCC**AGGAAAGAGATGCT**
CCGACGGTAGTGACTTATCAAACCGTGACACGACAGTGGG**TCCTTTCTCTACGA**

**GGTGAACAGAACCAGGATTACGAGAAAGTACAAACATTTGTTAAAGGGAGC**
**CCACTTGTCTTGGTCCTAATGCTCTTTCATGTTTGTAAACAATTTCCCTCG**

**CTGATAGGCAAAAGAGCAGTTGGCAGTGAGGAGGACTGAGCCACAACCCAA**
**GACTATCCGTTTTCTCGTCAACCGTCACTCCTCCTGACTCGGTGTTGGGTT**

**GTTTCATGAG**GTAAGACCTCCGGCCCCAGGAAAGGACACCAATGCAAGGCATTG
**CAAAGTACTC**CATTCTGGAGGCCGGGGTCCTTTCCTGTGGTTACGTTCCGTAAC

CCTGCCCCTGG
GGACGGGGACC

Fig. 5. 14. Identification of the consensus splice "AG-GT" sites (underlined) in the genomic sequence containing the exon trap product.

using human multitissue northern blots (Clontech) indicated ubiquitous expression in prostate, testis, ovary, small intestine, colon, heart, lung, liver, skeletal muscle, kidney, and pancreas, except spleen, brain, and blood (Fig. 5. 15). To examine the expression level of SDC2, a northern blot of total RNA isolated from the patient's lymphoblastoid cells was screened. The results indicated no expression of SDC2 in either patient or control lymphoblastoid cells, but positive for the expression in control fibroblast cells (data not shown). Thus the expression level of SDC2 in the patient remains unresolved since the patient's fibroblast cells were unavailable.

## 5. 5 BALANCED RECIPROCAL TRANSLOCATION

To examine the mechanism of the translocation, the regions of the breakpoint from the two normal and two derived translocation chromosomes were sequenced and compared.

STS3367 was purified and sequenced from both ends using its PCR primers, 67 and 33. By comparing the four sequences of the normal chromosomes 8 and X and both translocation breakpoint junction fragments, four nucleotide sequence, GGCA, was found on all four chromosomes. The translocation occurred within the 5'-GGCA-3' sequence without the loss or gain of a single nucleotide (Fig. 5. 16).

## 5. 6 SUMMARY

Three chromosome 8-specific YACs were isolated by PCR screening of the ICI-YAC library. Primers were designed from derivative chromosomal sequences adjacent to the translocation breakpoint. Cosmid libraries were constructed

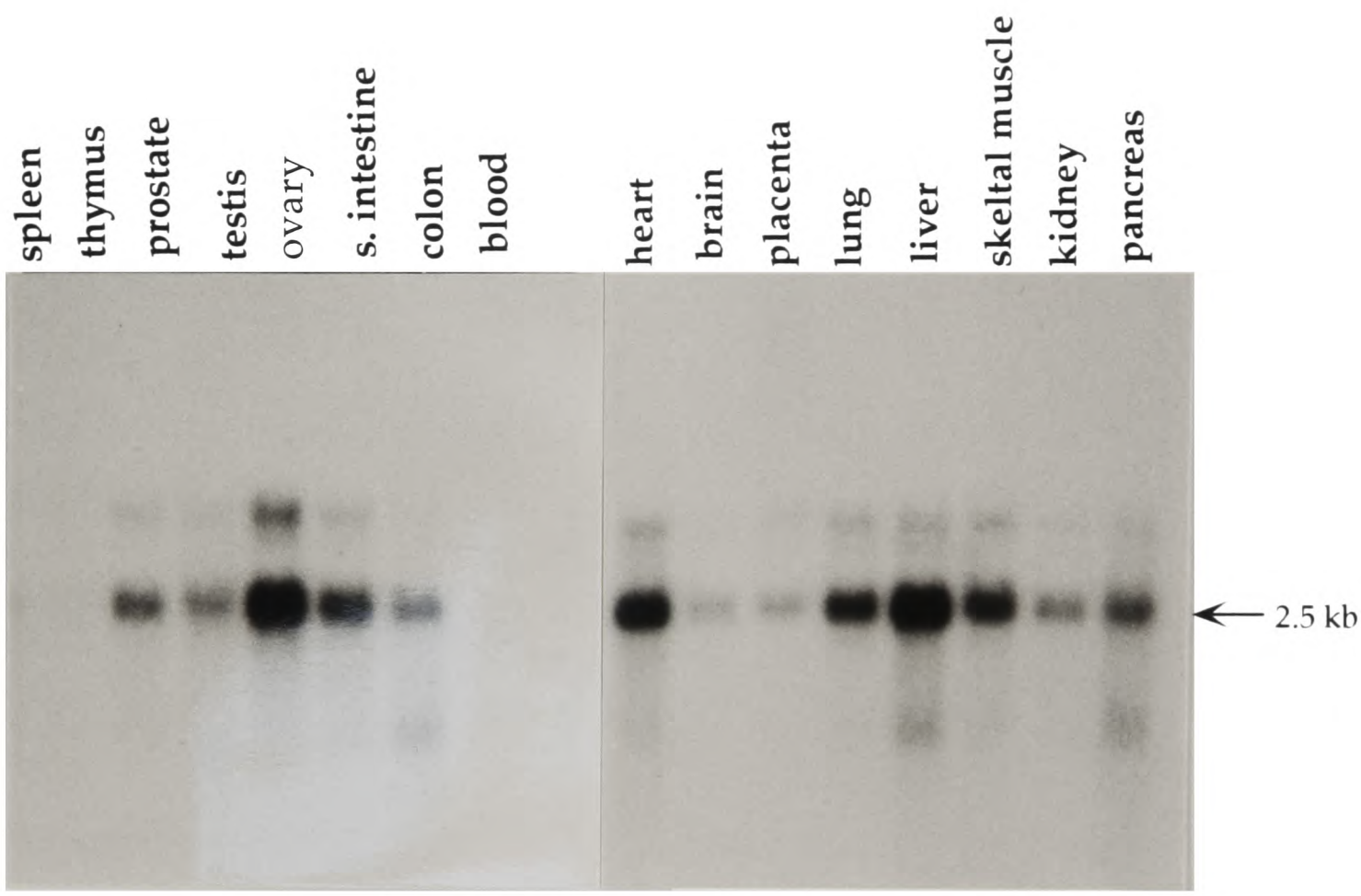



Fig. 5. 15. Northern blot hybridization analysis of SDC2 expression using A1 cDNA as a probe. Expression was detected in polyA+ RNA from most of the tissues.

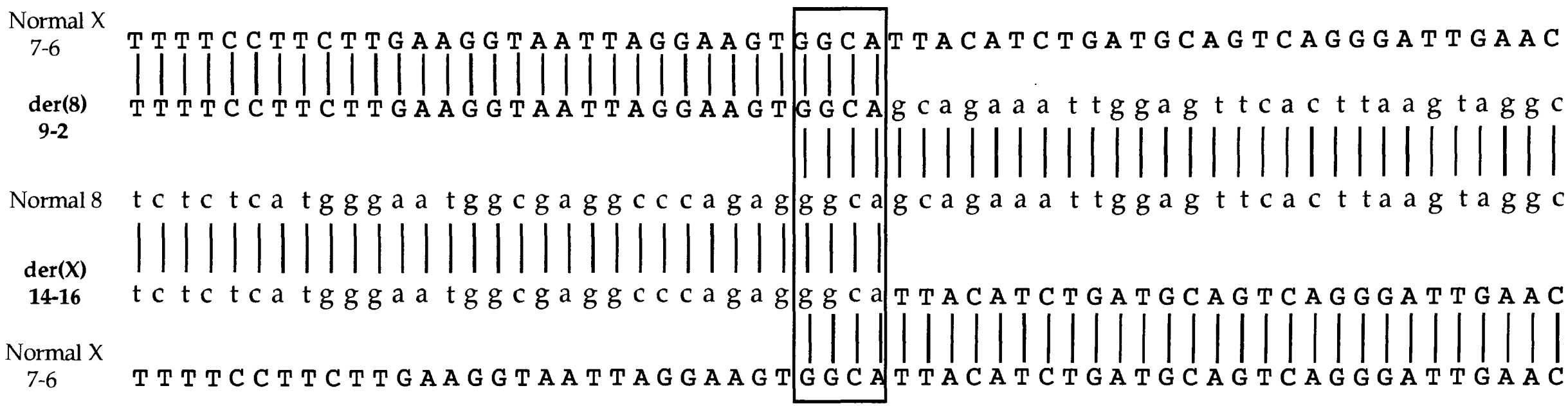



Fig. 5. 16. Nucleotide sequences of subclones from der(8) (9-2) and der(X) (14-16), normal chromosome X (Normal X: 7-6) and normal chromosome 8 (Normal 8) spanning the translocation breakpoint. Chromosomes X sequences are indicated in uppercase and chromosome 8 sequences in lowercase. The box indicates the 5′-GGCA-3′ sequence where the translocation occurred.

from two of the non-chimeric YACs with an average size of 500 kb and the cosmids were ordered in a cosmid contig using STSs generated by an end-cloning technique. A human testis cDNA library was screened using two of these cosmids which spanned the translocation breakpoint as probes. Ten positive cDNAs were isolated, all of which contained a part of the Syndecan-2 (SDC2) gene. The cosmid contig placed the translocation breakpoint approximately 30 kb distal to the SDC2 gene. SDC2 appears to be the only gene associated with the translocation breakpoint as confirmed by sequencing about 41.7 kb either side of the translocation breakpoint.

The GenBank (1. 4. 4) database covers 3,414 bp of the 3′ end of the SDC2 gene containing the start and stop codons. Four introns were identified, but the whole gene size has been unresolved. The SDC2 gene is expressed ubiquitously except in brain, spleen and blood. The translocation breakpoint was reciprocal at '5′-GGCA-3′ sequence and occurred with the transcriptional orientation opposite to the GRPR gene, giving rise to incompatibility for a fusion gene of the GRPR and SDC2 genes.

# CHAPTER VI

## DISCUSSION

A young adult female, ML, manifests mental retardation, epilepsy, impairment in reciprocal social interaction, short stature, mild brachycephaly, and multiple exostoses. Her karyotype exhibits a balanced translocation of 46,X,t(X;8)(p22.13;q22.1). The molecular genetic investigation in this patient identified the translocation breakpoint on both chromosomes X and 8 (6. 1 and 6. 2). The translocation breakpoint on the X chromosome occurred in the first intron of the gastrin releasing peptide receptor (GRPR) gene (6. 3) and on chromosome 8 occurred approximately 30 kb distal to the 3' end of the Syndecan-2 gene (SDC2) (6. 4).

### 6. 1 SUMMARY OF THE RESULTS

A cytological abnormality, an X;8 translocation in patient, ML, provided a candidate region, without the need for linkage analysis, and thus a quick route to identification of genes which when disrupted caused her autism and multiple exostoses. Using a positional cloning strategy, 44 YACs specific to the Xp22 candidate region were hybridized to metaphase DNA spreads from the patient by fluorescence *in situ* hybridization. Five YACs isolated by the DXS43 probe were identified to be spanning the translocation breakpoint. Two out of the five YACs were used to isolate 256 cosmids by screening the ICRF flow-sorted X chromosome cosmid library and by constructing cosmid libraries. The cosmids were radioactively labeled and

hybridized to Southern blots containing patient genomic DNA. The breakpoint was identified in eight cosmids by the presence of patient-specific restriction fragments (junction fragments).

The Gastrin Releasing Peptide Receptor (GRPR) gene was mapped by Alitalo *et al.* (1995) adjacent to DXS43 in the Xp22.1 region during the time that a cosmid contig was being constructed covering the YAC originally isolated with DXS43 (Alitalo *et al.*, 1995). The positional candidate strategy was then undertaken by studying the location of the translocation relative to GRPR and its feasibility as a candidate gene for the phenotype of the patient. Its sequence was extracted from the database and used to design PCR primers. The PCR products were non-radioactively labeled and hybridized to Southern blots containing *Hind*III-digested cosmids. This confirmed that eight cosmids spanning the translocation breakpoint also contained the GRPR gene. Two plasmid subclones, 7a and 7b, derived from one of the cosmids were identified to flank the translocation breakpoint by hybridizing to the same Southern blots from the patient genomic DNA as described above.

For the isolation of the junction fragments, patient genomic DNA was digested with *Bgl*II and cloned in phage to construct a patient genomic library. The filters replicated from this library were hybridized with subclones, 7a and 7b, which identified phages positive to both 7a and 7b, only to 7a, and only to 7b. These positives were subcloned into plasmids for sequencing. The plasmid designated as 7-6 hybridized positively to both 7a and 7b, indicating that this clone contained the fragment from the normal chromosome X. The der(X) and der(8) patient fragments were also cloned in 14-16 positive to 7a

and 9-2 positive to 7b, respectively. The restriction map of these three clones identified the location of the translocation breakpoint and their sequencing identified it at the nucleotide level.

The relationship between the GRPR gene and the translocation breakpoint was resolved by physical mapping of this region. The subclones, 7a and 7b hybridized to eight cosmids. Five out of eight hybridized with the exon-1 of the GRPR gene, but not exon-2, intron-2 nor exon-3. Of the remaining clones, three, hybridized with exon-2, intron-2 and exon-3, but not exon-1. The translocation breakpoint therefore occurred within the first intron of the GRPR gene, proximal to exon-1.

The analysis of chromosome 8 began with isolating a chromosome 8 specific STS, STS3367, by sequencing across the translocation breakpoint using primers from both the der(X) and der(8) plasmids. The hybridization of this STS to Southern blots containing patient genomic DNA identified translocation junction fragments with several endonucleases, confirming that it was adjacent to the breakpoint.

A common sequence, GGCA, was found in the regions of the translocation breakpoint from all four chromosomes, the der(X), der(8), the normal chromosome X and 8. The translocation had occurred within the 5'-GGCA-3' sequence without the loss or gain of a single nucleotide.

To identify coding regions on chromosome 8, three YACs were isolated by PCR screening of the ICI YAC library, using the STS3367 primers. Cosmid libraries were constructed from two of the YACs and screened by the STS3367. Two, out of seven cosmids spanning the translocation breakpoint,

were used as hybridization probes to screen a human testis cDNA library. The Syndecan-2 gene (SDC2) was identified through a BLASTN search of all the positive cDNA clones. Physical mapping placed the translocation breakpoint at less than 30 kb distal to the 3' end of the SDC2 gene on chromosome 8.

## 6. 2 CHROMOSOMAL TRANSLOCATION

Little is known of the variety of mechanisms causing constitutional chromosomal translocations. The X/autosome translocations associated with Duchenne muscular dystrophy (DMD) reviewed by van Bakel *et al.* (1995) have involved small deletions, duplications or insertions at the translocation breakpoints (van Bakel *et al.*, 1995). The present translocation resulted in no such loss or gain, not even of a single nucleotide. This may be related to a proposed mechanism of translocation, that of simultaneous chromosomal breaks and repair within the GGCA sequence common to both X and 8 chromosomes.

The chromosomal translocation of the patient ML was not seen in her mother nor her brother. They are phenotypically and clinically normal. There is no genetic information in her deceased father and the parental origin of the translocation has not been investigated. However, the X/autosome translocations investigated in eight DMD females were found to be all paternal in origin (Bodrug *et al.*, 1990; Robinson *et al.*, 1990). Four other X/autosome translocations which were not associated with DMD were also found to come from the father (reviewed by Robinson *et al.*, 1990). A preferential paternal origin was suggested to be a feature of the X/autosome

translocations because the probability of all 12 translocations having the same parental origin by chance is less than 1 in 4000 (Robinson *et al.*, 1990).

A chromosomal break can cause (1) a loss of function phenotype if it disrupts the coding sequence of a gene or a nearby regulatory region, (2) a gain of function with inappropriate expression by generating a fusion gene by splicing regulatory sequences from one gene to distal coding sequences of another gene, and (3) an alteration of gene expression of a gene located hundreds of kilobases away by affecting the chromatin structure (position effect). The der(X) chromosome from patient ML was previously shown to be active (Bolton *et al.*, 1995) and the disruption of the GRPR gene and a position effect of the SDC2 may be responsible for the phenotype of the patient.

## 6. 3 GASTRIN RELEASING PEPTIDE RECEPTOR (GRPR)

GRPR is a receptor for gastrin-releasing peptide (GRP). GRP is one of the only two known mammalian homologues of the amphibian peptide bombesin (Spindel *et al.*, 1984). The other counterpart is neuromedin B (NMB) (Krane *et al.*, 1988). Bombesin is a tetradeca-peptide originally purified from the skin of the European frog *Bombina bombina* (Anastasi *et al.*, 1971). Over 10 bombesin-like peptides have been isolated and classified into three subfamilies according to their COOH-terminal tripeptides. These are the bombesin, ranatensin, and litorin families (Table 6. 1). GRP belongs to the bombesin subfamily and NMB to that of ranatensin. No mammalian member of the litorin subfamily has been yet identified. Bombesin and GRP are identical in 90% of their C-terminal amino acids. The genes for NMBR,

| Species | Peptide | Human homolog | Sequence |
|---|---|---|---|
| *Bombina bombina* | Bombesin | GRP | Glp-Gln-Arg-Leu-Gly-Asn-Gln-Trp-Ala-Val-Gly-His-Leu-Met-$NH_2$ |
| *Rana pipiens* | Ranatensin | NMB | Glp-Val-Pro-Gln-Trp-Ala-Val-Gly-His-Leu-Met-$NH_2$ |
| *Litoria aurea* | Litorin | - | Glp-Gln-Trp-Ala-Val-Gly-His-Leu-Met-$NH_2$ |

Table 6. 1. Bombesin superfamily and GRP. Glp stands for pyro-glutamate, a form of glutamate in which a side chain on the amino acid binds to one end of the amino acid to form a small ring.

which is a receptor for NMB, and GRPR contain three exons each and encode 384 and 390 amino acid proteins, respectively, with 56% identity at the amino acid level (Corjay *et al.*, 1991). Both contain an extracellular domain, seven hydrophobic transmembrane domains, and a cytoplasmic domain, typical of trimeric guanine nucleotide binding protein-coupled (trimeric G protein) receptors (Dohlman *et al.*, 1987). These receptors are the largest family of cell surface receptors and their amino acid sequences have been conserved in particular in the seven hydrophobic transmembrane domains (Fig. 6. 1). Because of this structure, they are also known as serpentine or seven-pass transmembrane receptors.

These receptors mediate cellular responses to signalling molecules-ligands such as hormones, local mediators, and/or neurotransmitters. The binding of such a ligand alters the conformation of the cytoplasmic domain of the receptor, exposing binding sites for trimeric G proteins, which bind to the receptor. Initially, the G protein has GDP bound to its alpha subunit, but the complex of ligand-receptor with the trimeric G protein weakens the affinity of the G protein for GDP, allowing GTP to bind

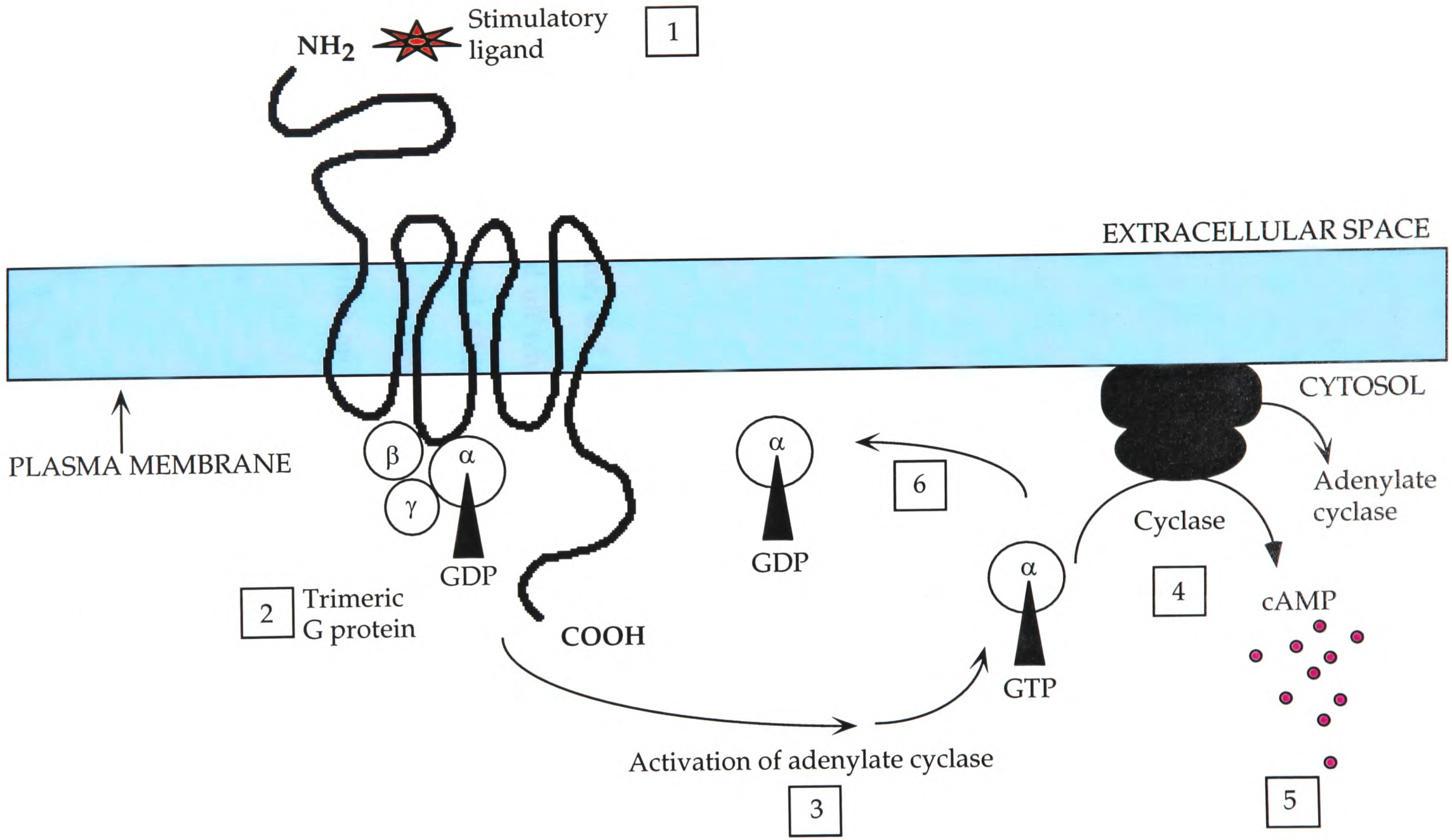



Fig. 6. 1. Schematic diagram of an intracellular signalling mechanism involving G-protein linked receptors. (1) activation of the receptor by a ligand binding, (2) phosphorylation of the trimeric G-protein in the intracellular domain of the receptor, (3) activation of adenylate cyclase, (4) production of cyclic AMP, and (5) relay of the signal by altering other specific proteins in the cell.

(Fig. 6. 1). By this GTP association, the alpha subunit of the G protein dissociates from the receptor complex and binds and activates a plasma membrane enzyme, adenylate cyclase to produce cyclic AMP (cAMP), as a second messenger. cAMP activates either cyclic-AMP-dependent protein kinase (A-kinase), $Ca^{2+}$/Calmodulin-dependent protein kinase (CaM-kinase), or phospholipase C-β. The latter forms two pathways either generating inositol phospholipid, which releases $Ca^{2+}$ from the endoplasmic reticulum, or diacylglycerol, which activates C-kinase. A rise in $Ca^{2+}$ levels also stimulates A-kinase and CaM-kinase. In response to an extracelluar signal, C-kinase, A-kinase, and CaM-kinase phosphorylate specific target proteins on serine or threonine residues and thus alters the activity of these proteins. The G protein alpha subunit loses a phosphate in stimulating adenylate cyclase and reverts to GDP bound peptide which then reassociates with the ligand-bound receptor, and the signalling process is renewed. These responses are turned off when the extracelluar signalling ligand is removed.

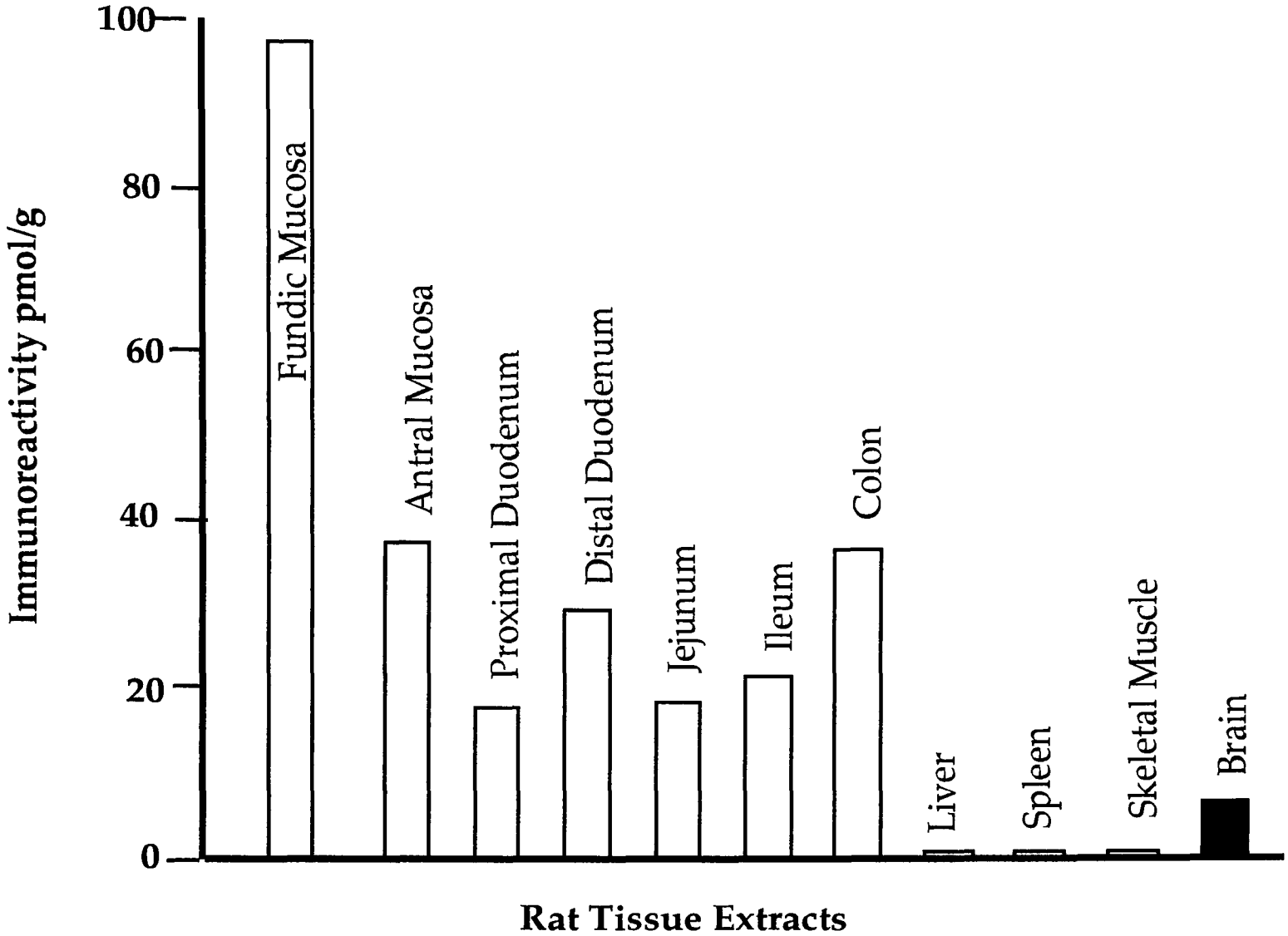
Both GRPR and NMBR receptor genes are thought to have evolved from duplication of a common ancestor (Corjay *et al.*, 1991), but are encoded by different genes which are not clustered in the genome (Table 6. 2). GRP

| Name | Locus | No. of Amino acids | Reference |
|---|---|---|---|
| GRPR | Xp22.13 | 384 amino acids | (Alitalo *et al.*, 1995) |
| GRP | 18q21 | 27 amino acids | (Lebacq-Verheyden *et al.*, 1987) |
| NMB | 16q21-qter | 390 amino acids | (Daly *et al.*, 1991) |
| NMB | 5q22-qter | 76 amino acids | (Richard *et al.*, 1994) |

Table 6. 2. Human bombesin-like receptors and ligands, indicating chromosome location and coded peptide length.

and NMB are present in the gastrointestinal tract, lungs and also in brain (Spindel *et al.*, 1986; Wang and Cutz, 1994). The tissue distribution of GRP extracted from the rat (Sprague-Dawley) showed the highest concentration in gastric mucosa (98 pmol/g tissue) and high concentrations throughout other parts of gut such as in the antral mucosa, the proximal duodenum, distal duodenum, jejunum, ileum and colon (average 27 pmol/g) (Fig. 6. 2A), and was suggested to have the function of a paracrine hormone (Spindel *et al.*, 1986; Wang *et al.*, 1994). GRP was also found in brain (6 pmol/g) (Walsh *et al.*, 1979) and was suggested to be involved in various neurobiological activities as a neurotransmitter (Spindel *et al.*, 1986; Wang *et al.*, 1994). Walsh *et al.* found pronounced regional differences in the brain, with highest concentrations in the hypothalamus (29 pmol/g), intermediate amounts in the cortex and midbrain (93 pmol/g) and virtually undetectable concentrations in the cerebellum and pons-medulla (1 pmol/g) (Fig. 6. 2B). A similar result was found in the dog (Walsh and Wong, 1978). Brown *et al.* (1977) found essentially the same result in the rat except for a high concentration in pons-medulla (Brown *et al.*, 1977). He also detected a small concentration of GRP in the lung as did Wharton *et al.* (Wharton *et al.*, 1978).

GRP functions as a growth factor, neurotransmitter, and paracrine hormone. GRP stimulates proliferation of 3T3 mouse fibroblasts, normal human bronchial epithelia cells, and induces cell hyperplasia *in vitro* in the rat (Rosengurt and Sinnett-Smith, 1983). Some human small-cell lung cancer (SCCL) cell lines secrete GRP and express a single class of high-affinity receptor for GRP (Moody *et al.*, 1981). Cuttitta *et al.* (1985) demonstrated



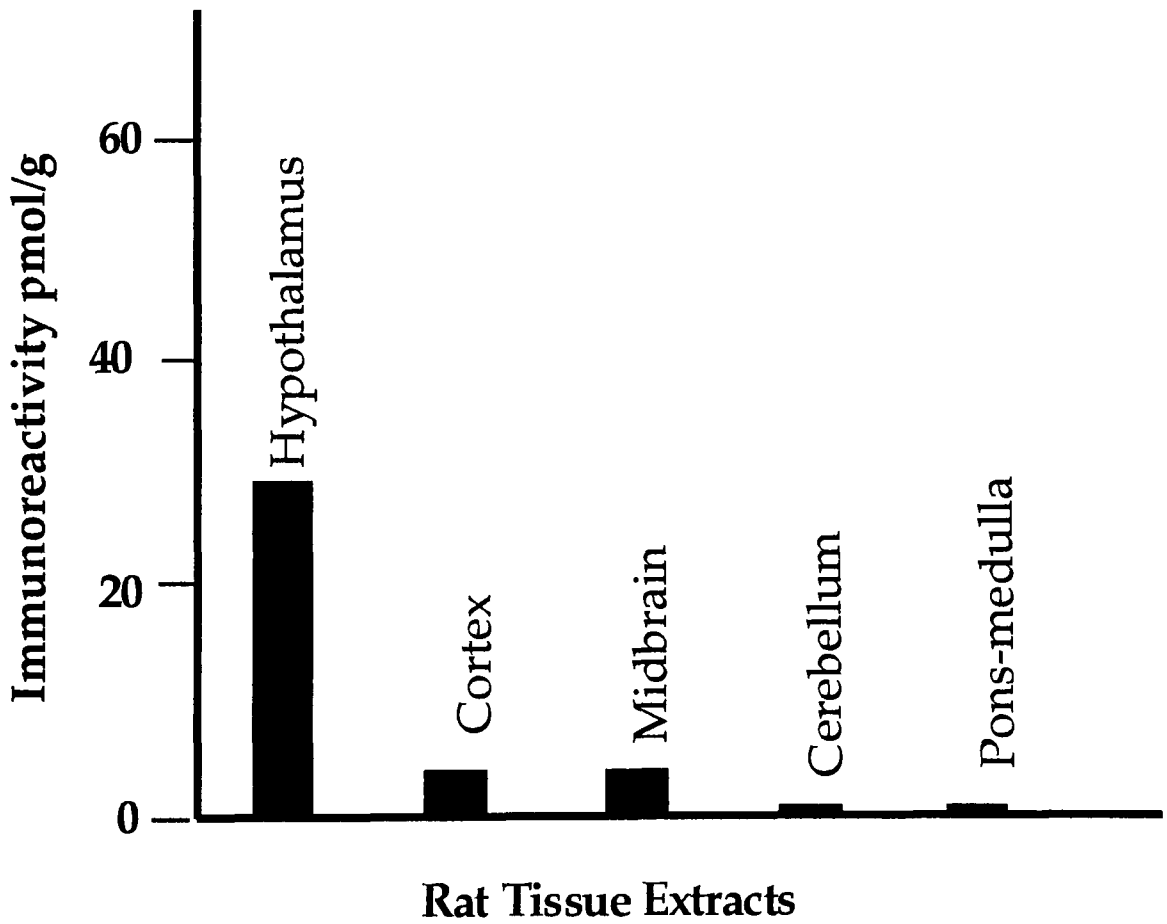




Fig. 6. 2A and 2B. Distribution of immunoreactive bombesin in acetic acid extracts of rat tissue and brain, respectively, from Walsh *et al.*, 1979. The mean of 6 animals.

small-cell lung tissue clonal growth by adding GRP exogenously and growth inhibition of SCCL cell lines *in vitro* and *in vivo* by blocking the receptor binding sites for the hormone using monoclonal antibodies to GRP. These suggested that GRP can be a potent mitogen in SCCL (Cuttitta *et al.*, 1985) and that GRP can function as an autocrine growth factor.

Wada *et al*, (1991) studied the expression of GRPR mRNA in the brain using *in situ* hybridization on rat brain sections. Rodent GRPR and NMBR coding regions show 90% and 89% amino acid identity with those of human, respectively (Corjay *et al.*, 1991). They reported distinct regional distributions of GRPR and NMBR mRNA and hypothesized independent roles for the two receptors in the central nervous systems (Wada *et al.*, 1991). Antisense $^{35}$S-labeled GPPR and NMBR cDNA probes were separately hybridized to coronal sections of the olfactory or hypothalamic regions of rat brain. GRPR mRNA expression was not detected in the olfactory regions of the brain, two regions where NMBR mRNA expression was highest. The GRPR mRNA was expressed throughout the limbic system (Wada *et al.*, 1991). The limbic system includes all of the interior parts of the cortex such as hypothalamus, hippocampus, amygdala, mamillary body, cingulate and parahippocampal gyri, anterior commissure, medial olfactory area, ventral striatum, orbitofrontal area, anterior insular cortices, etc., including the border region of the fornix encircling the upper cortex (Fig. 6. 3). The most intensive expression of GRPR mRNA was found in hypothalamic area, hippocampus, and amygdaloid nucleus and the mamillary nucleus (Wada *et al.*, 1991). The limbic system has long been considered to be involved in emotion, affection,

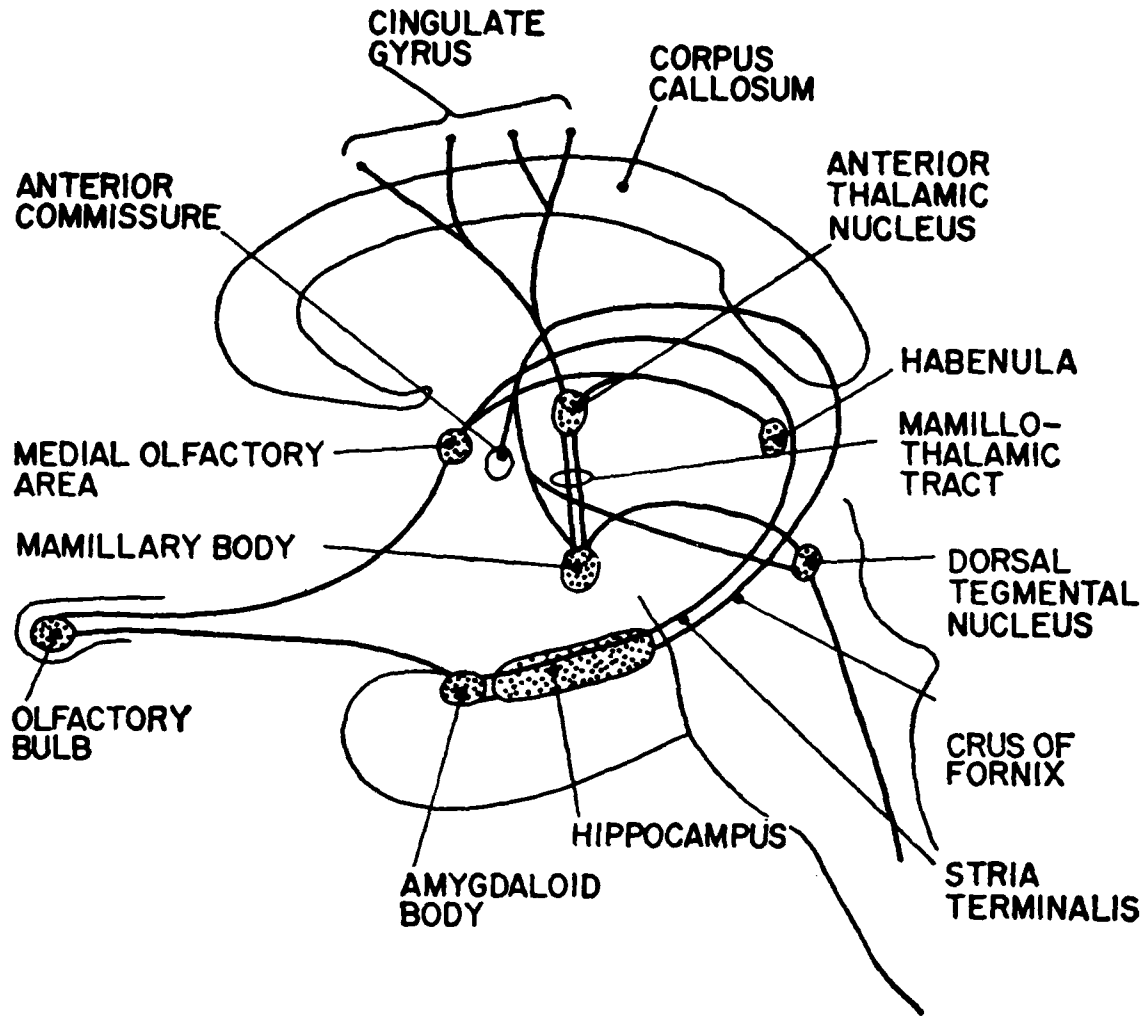



Fig. 6. 3. Schematic diagram of the limbic system (connections of the hippocampus and fornix) from Passmore and Robson, 1996.

attention, obsession, reward, punishment, pleasure and fear (Bauman and Kemper, 1985; Ciaranello *et al.*, 1982; Devinsky *et al.*, 1995; Gingell *et al.*, 1996; LeDoux, 1987; LeDoux, 1994; Peterson *et al.*, 1996).

The role of the amygdala in learning and remembering fear was elucidated in rat experiments and reviewed by LeDoux (1994). Visual dangers are first processed by the thalamus, which passes the information directly to the lateral nucleus of the amygdala. These stimuli are transmitted to the basolateral, the accessory basal and the central nuclei in the amygdala. The central nucleus is connected to the brain stem, bringing about a physiological response to the danger (Fig. 6. 4) (LeDoux, 1994). Simultaneously, the cortex receives the information from the thalamus. This information is then relayed to the amygdala, causing heart rate and blood pressure to increase and muscles to contract as the fear response (LeDoux, 1994).

GRP may be involved in these biological activities. Injection of GRP into the paraventricular nucleus of the rat hypothalamus, results in an increase in the blood pressure, and contraction of smooth muscles (Lebacq-Verheyden *et al.*, 1987; Spindel *et al.*, 1986). The gastric pH was also increased, gastric acid secretion decreased, and food intake reduced (reviewed by Tache (Tache and Brown, 1982). Other biological activities associated with GRPR in mammals are stimulation of pancreatic enzyme secretion, stimulation of intestinal motor activity, and contraction of the gall bladder (Lehy *et al.*, 1983).

GRP was expressed at different stages of rat development. GRP was detected only in the lower part of the gut in 1-2 day old rats, and in the pylorus, small intestine and large intestine at 8-16 days from birth, suggesting

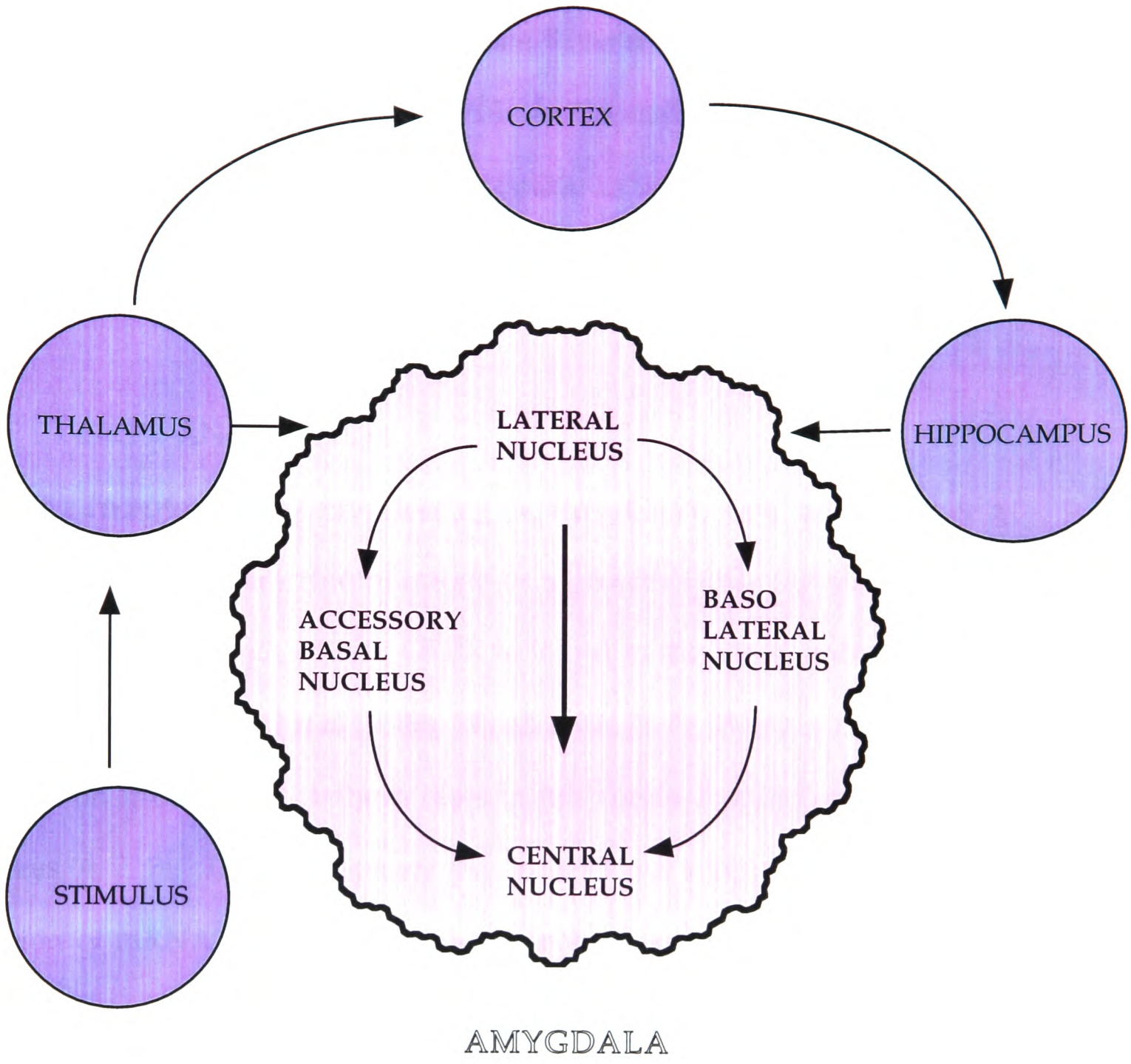



Fig. 6. 4. Schematic diagram of the amygdala's role in learning about and remembering fear. A visual stimulus, for example, the sight of a viva examiner, is first processed by the thalamus. The signal will be passed to the lateral nucleus and transmitted to the basolateral, the accessory basal, and the central nuclei in the amygdala. At the central nucleus which connects to the brain stem, bringing about physiological changes such as causing heart rate and blood pressure to increase and muscles to contract. At the same time, the stimulus will be also received by the visual cortex from thalamus and relayed to the hippocampus. When the cortex determines that the object is not a threatening object, the message to the amygdala will calm the fear response (LeDoux, 1994).

the function of a paracrine hormone. Brain concentrations increased slowly and reached adult levels by 26 days (Walsh *et al.*, 1979).

Other peptides, which are present both in gut endocrine cells and in central peripheral neurons, are substance P, neurotensin, cholecystokinin, enkephalin, vasoactive intestinal peptide and somatostatin. Substance P has the same carboxy-terminal Leu-Met dipeptide as bombesin, but is rarely found in SCCL and binds to a different set of receptors than bombesin (Moody *et al.*, 1981).

## 6. 4 GRPR AND AUTISM

GRPR is a receptor for gastrin-releasing peptide (GRP). GRP acts as either a neurotransmitter, a paracrine hormone or a growth factor (Lebacq-Verheyden *et al.*, 1987; Spindel *et al.*, 1986). GRPR is found in the brain, gastrointestinal tract and lungs (Spindel *et al.*, 1986; Walsh *et al.*, 1979; Wang *et al.*, 1994) although on multitissue northern blots in this thesis, it was detected only in pancreas (4. 2. 2). This discrepancy may result from a higher sensitivity for detection of mRNA signals in specific regions on brain sections in contrast to mRNA extracted from whole brain tissue. Involvement of multiple regions of the limbic system has long been implicated as a candidate neural substrate underlying the social deficits in autism (Bachevalier, 1996; Bauman *et al.*, 1985; Ciaranello *et al.*, 1982; Dawson, 1996; Demasio and Maurer, 1978; Devinsky *et al.*, 1995; Gingell *et al.*, 1996; LeDoux, 1987; LeDoux, 1994; Peterson *et al.*, 1996; Tager-Flusberg, 1996).

Dawson (1996), in his literature reviews on autistic neuropsychology,

classified autistic behavioural symptoms into four categories: orienting to social stimuli, motor imitation, joint attention, and empathy (perception of emotion) (Table 6. 3), and suggested a link between autism and impairment of a particular part of the limbic system, the medial temporal lobe, which

| Symptom | Possible Impairment | Substrate | Reference |
|---|---|---|---|
| Orienting to social stimuli | Recognition of affection | Amygdala | LeDoux, 1987 |
| | Social stimulus-Reward | Amygdala | Jones & Mishkin, 1972 |
| | Disengage/shift Attention | Cerebellum | Courchesne *et al.*, 1994 |
| Motor imitation | Perception of body movement | Amygdala | Brothers *et al.*, 1990 |
| | Long-term memory | Amygdala & Hippocampus | Mishkin & Appenzellar, 1987 |
| | Motor planning & execution | Frontal lobe & Basal ganglia | Damasio & Maurer, 1987 |
| Joint attention | Perception of gaze direction | Amygdala | Brothers *et al.*, 1990 |
| | Rapid shifts in attention | Cerebellum | Courchesne *et al.*, 1994 |
| Empathy | Perception of emotion | Amygdala | Adophs *et al.*, 1994 |

Table 6. 3. Hypothesized neural substrates for early symptoms of autism derived partially from "From mind to molecule: Researchers try to unravel the complexity of autism." Journal of NIH Research, 7, 1995 and (Dawson, 1996).

contains the amygdala, hippocampus, and entorhinal cortex (Dawson, 1996). In a study on macaques, a variety of neurons specialized in the processing of facial recognition and affective expressions, gestures, and gaze were found to be contained in the medial temporal lobe (Perrett *et al.*, 1992). A link between autism and the medial temporal lobe was also suggested by Bachevalier (1996) (1. 2. 3). He observed that monkeys with operated neonatal amygdala lesions

and humans with brain lesions in the medial temporal lobe developed stereotypic repetitive behaviours, withdrew in social interactions, and lost facial and body expressions (Bachevalier, 1996). Collectively, Tager-Flusberg (1996) speculated there are two different brain areas involved in autism. The dysfunction of the medial temporal lobe may link to the impairment in 'social affective relatedness' and cortex damage to cognitive impairments (Tager-Flusberg, 1996). Cognitive memory was subdivided on the basis of three patients with brain injuries, into two components, context-rich event memory (episodic) and universal factual memory (semantic). Episodic memory appears to be fully dependent on the hippocampal component, which is suited to promote flexible associations by recognizing relations among items and differentiating overlapping patterns whereas semantic memory, learning vocabulary and information, depend primarily on the entorhinal cortices underlying the medial temporal lobe (Devinsky *et al.*, 1995; Eichenbaum, 1997; Vargha-Khadem *et al.*, 1997). The findings that these two are partly dissociable and independent (Vargha-Khadem *et al.*, 1997) may explain the coexistence of abysmal and remarkable mental abilities in some autistic patients. Some autistic individuals exhibit excellent memory in the 'Block Design' test of the Wechsler Scales of Intelligence (1. 2. 2), and for poems, names, and complex patterns (Asperger, 1944; Kanner, 1943). On the other hand, they often lack in recognition to social stimuli (Table 6. 3), acquiring less and less social and communicative skills with age (Frith, 1993).

GRP functions as a neurotransmitter (Lebacq-Verheyden *et al.*, 1987; Spindel *et al.*, 1986). Comings *et al.* (1988) found an association between the

GRP gene and another neurological disorder, Tourette's syndrome [GTS;MIM 137580]. GTS is characterized with obsession, echolalia, grunting, coprolalia, tic-like movements, and self-mutilation (43%) (van Woert *et al.*, 1977). The frequency of GTS has not been systematically reported but is estimated to be 28.7 per 100,000 in school children of Monroe County in New York (Kurlan *et al.*, 1987). About 75% of the patients are male, with onset occurring between ages 2 and 14 years. Comings *et al.* (1988) injected GRP into the brains of mice and reported the appearance of GTS-like behaviours involving tics and facial punching (Comings and Coimings, 1988). GTS obsessive behaviour was shown to be linked to the limbic system (Devinsky *et al.*, 1995; Peterson *et al.*, 1996) and tic movement to the corpus callosum, thalamus, and midbrain (Peterson *et al.*, 1996). Studying a translocation patient, t(7;18) (q22;q22.1), Comings *et al.* (1986) localized GTS at 18q22.1 (Comings *et al.*, 1986), where GRP has mapped (Lebacq-Verheyden *et al.*, 1987). Donnai (1987) presented further evidence for the location of the GTS at 18q22.1 in the deletion study in the GTS female patient (Donnai, 1987). The mapping of GTS, however, has not been resolved. Brett (1990) mapped GTS to Chromosome 3 (Brett *et al.*, 1990), and Taylor (1991) reported an association of GTS with a deletion of 9pter (Taylor *et al.*, 1991). Interestingly, Sverd (1991) and Comings and Comings (1991) reported the co-existence of autism and GTS in two different family studies (1. 2. 7) (Comings and Comings, 1991; Sverd, 1991).

Over 3,000 expressed genes have been found to be active in the developing hippocampus, one part of the limbic system (Adams *et al.*, 1993). Besides these numerous candidate genes, in autism, there could be other

genes which are not normally expressed in the brain (Cook, 1996). Given the high sibling risk, lambda S (1. 2. 6), Pickles *et al.* (1995) estimated an involvement of 3-4 loci, and possibly up to 10 loci for autism (Pickles *et al.*, 1995). If this is correct, one of the susceptibility loci for autism may be X-linked, or the phenotype partially sex-limited, because theincidenceof autism is 3-4 times higher in males than in females (1. 2. 2). GRPR is on the X chromosome and in a different region from that recently excluded from autism by segregation and linkage analysis of multiplex autism families (Hallmayer *et al.*, 1996; Hallmayer *et al.*, 1994),

GRPR has been found to clearly escape X-inactivation in an assay using somatic cell hybrids (Ishikawa-Brush *et al.*, 1997). All genes in the major pseudoautosomal region (PAR1) and the minor pseudoautosomal region (PAR2) have been shown to be expressed on both active and inactive X chromosomes (Fig. 6. 5 and Table 6. 4). Several other X-linked genes mapped close to the pseudoautosomal region are also known to escape X-inactivation. The non-pseudoautosomal genes, ZFX, RPS4X and UBE1 (Fig. 6. 5 and Table 6. 6), have homologues on the Y chromosome and are also known to escape X-inactivation. Some active genes, outside the pseudoautosomal region and without a Y homolog, also escape X-inactivation. A cluster of such genes was found in the Xp11.21-p11.22 region (Miller *et al.*, 1995). GRPR is similar to these latter genes, but is located outside of the Xp cluster. GRPR is a coding gene, maps to Xp22.13 which is far proximal to the pseudoautosomal region and does not have a Y homolog.

The quantitative expression of GRPR loci on the active and inactive X

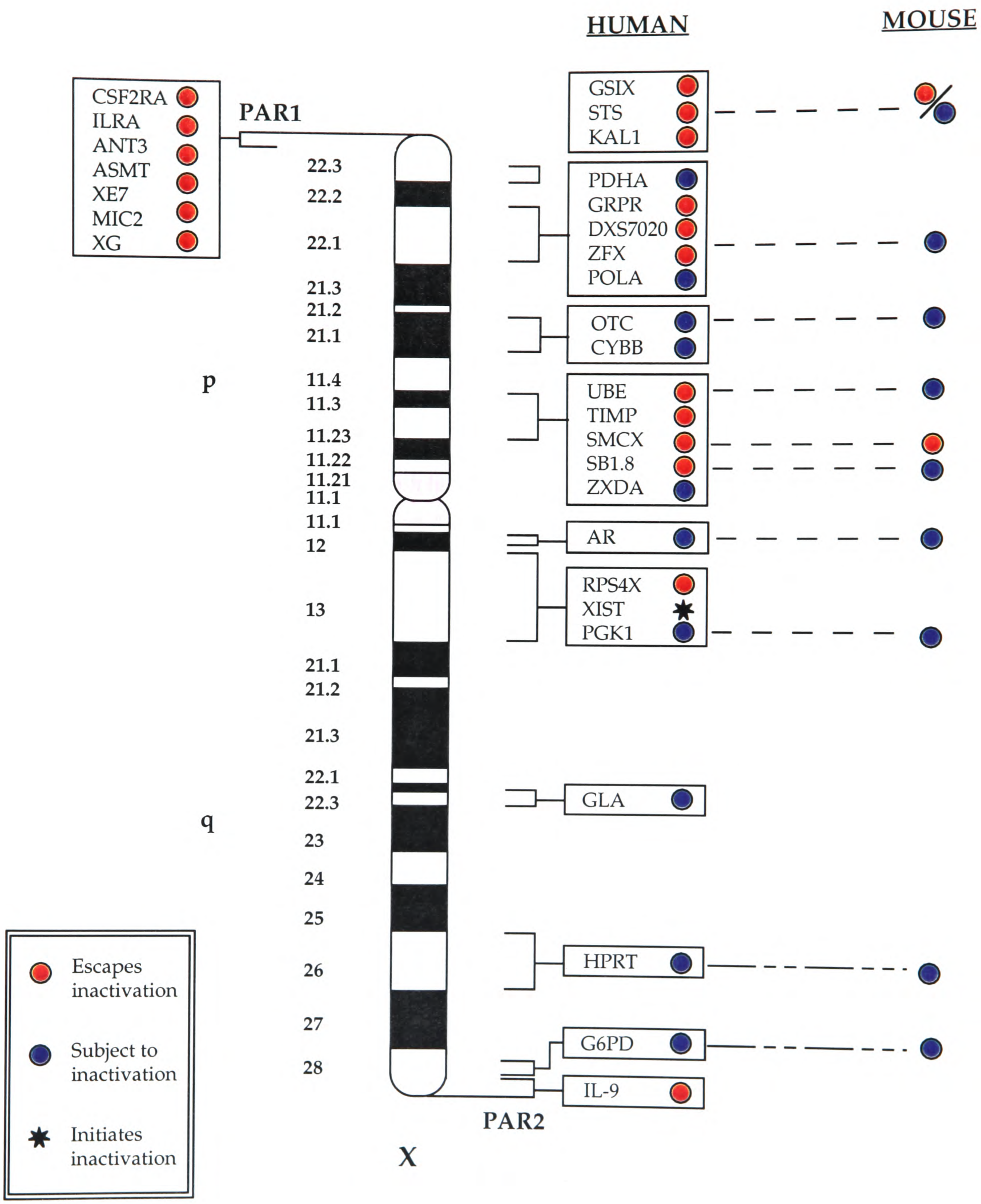



Fig. 6. 5. Distribution of the human (left) and mouse (right) genes and inactivation status on the human X chromosome. PAR1 and PAR2 indicate the major and minor pseudoautosomal regions. The XIST locus marked * initiates X inactivation and is expressed on the inactivated X, but not on an active X. Legend is contained in Table 6. 4.

| Gene symbol | Name | Location | Homolog Y |
|---|---|---|---|
| CSF2RA | granulocyte-macrophage colony-stimulating factor receptor | PAR1 | CSF2AY |
| ILRA | interleukin-3 receptor | PAR1 | ILRAY |
| ANT3 | adenine nucleotide translocator 3 | PAR1 | ANT3 |
| ASMT | encodes acetyl seronin N-methyl transferase | PAR1 | ASMTY |
| XE7 | encodes XE7 antigen: homology to CD99 | PAR1 | XE7Y |
| MIC2 | encodes cell surface antigen CD99 | PAR1 | MIC2Y |
| XG | encodes Xg blood group regulator | PAR1 | XGY |
| GSIX | encodes Xg blood group antigen | Xp22.3 | - |
| STS | steroid sulfatase gene | Xp22.3 | - |
| KAL1 | Kallman syndrome gene | Xp22.3 | - |
| PDHA1 | pyruvate dehydrogenase E1 alpha polypeptide | Xp22.1 | - |
| GRPR | Gastrin Releasing Peptide Receptor | Xp22.1 | - |
| DXS7020E | EST escaping X-inactivation | Xp22.2 | - |
| ZFX | zinc finger protein of unknown function | Xp22.1 | ZFY |
| POLA | polymerase (DNA directed) , alpha | Xp21.3 | - |
| OTC | ornithinine carbamoyltransferase | Xp21.1 | - |
| CYBB | cytochrome b-245, beta polypeptide | Xp21.1 | - |
| UBE1 | ubiquitin activating enzyme E1 | Xp11.23 | - |
| TIMP1 | tissue inhibitor of metalloproteinase | Xp11.23 | - |
| SMCX | encodes H-Y antigen | Xp11.23 | SMCY |
| XE169X | =SMCX | Xp11.23 | - |
| SD1.8 | =DXS423 | Xp11.23 | - |
| AR | androgen receptor | Xq12 | - |
| RSP4X | encodes ribosomal protein S4 | Xq13 | RSP4XY |
| XIST | X chromosome inavtivation center | Xq13 | - |
| PGK1 | phosphogycerate kinase 1 | Xq13 | - |
| GLA | galactosidase, alpha | Xp21.3 | - |
| HRPT | hydroxanthine guanosine phosphoribosyltransferase | Xq26 | - |
| G6PD | glucose-6-phosphate dehydrogenase | Xq28 | - |
| IL-9 | Interleukin-9 receptor gene | PAR2 | IL-9R |

Table 6. 4. Legend to Fig. 6. 5.

chromosome has not been determined. It is possible that the GRPR gene from the inactive X chromosome may be expressed at a fraction of the active allele. In a study of the STS gene (see the above and Fig. 6. 5) located nearby the pseudoautosomal region, expression of the allele escaping X-inactivation was found to be 30% of the active allele (Migeon *et al.*, 1982). The human ubiquitin activating enzyme E1 (UBE1) gene was also expressed from the inactive X chromosome at about 30% of active X levels (Carrel *et al.*, 1996). A reduction in GRPR expression such as with these genes may lead to low dosages, insufficient to prevent autism in this patient.

Dosage-sensitive phenotypes have been studied in the X-linked gene for hypoxanthine phosphoribosyl transferase (HPRT) (Sege-Peterson *et al.*, 1992). HPRT is subject to X-inactivation in the female and the effect of decreasing the enzyme dosage in males and the corresponding phenotype are shown in the Table 6. 5. Loss of function becomes evident at 60% of the normal level of the HPRT gene product. Hyperuricemia occurs with less than

| HPRT activity (% of normal) | Phenotype |
|---|---|
| > 60% | Normal |
| 8-60% | Neurologically normal, hyperuricemia (gout) |
| 1.6-8% | Neurological problem (choreoathetosis) |
| 1.4-1.6% | Lesch-Nyhan syndrome (self-mutilation, choreoathetosis) |
| <1.4% | Lesch-Nyhan syndrome and mental retardation |

Table 6. 5. Effects of decreasing levels of hypoxanthine adenine phosphoribosyl transferase on phenotype from Sege-Peterson *et al.*, 1993.

60%, choreoathetosis with less than 8%, Lesch-Nyhan syndrome with less than 1.6%, and Lesch-Nyhan syndrome plus mental retardation with less

than 1.4% expression. Pathogenic effects caused by gene dosage often show variable phenotypic expression, even within families (Fisher and Scambler, 1994). Genes like HPRT whose products function as a metabolic or developmental switch appear to be dosage-sensitive (Fisher *et al.*, 1994). The gene products, which are part of a signalling system, such as a receptor or DNA binding site, are also dosage-sensitive (Fisher *et al.*, 1994). GRPR is a G-protein receptor (6. 1). It is conceivable that a decrease in dosage may cause malfunctions between interacting products.

It is unknown whether inactivation patterns for the GRPR gene are consistent throughout every tissue and cell within the body. This does not appear to be true for imprinted genes. A maternally encoding gene for potassium channel KVLQT1 (or KCNA9) is expressed tissue-specifically and imprinted in most tissues with the exception of the heart. Mutations in expressed KVLQT1 biallelically in the heart are involved in cardiac arrhythmia syndrome. In contrast, Beckwith-Wiedemann syndrome is caused by the expression of the paternally imprinted KVLQT1 in tissues other than the heart due to maternal translocations, suggesting paternally expressed gene product of growth enhancers, or a deficiency of maternally expressed gene product as a growth suppressor (Lee *et al.*, 1997; Mannens and Wilde, 1997; Neyroud *et al.*, 1997; Reik and Maher, 1997). Although the human and mouse X inactivation patterns are different (Fig. 6. 5), the murine homolog, H-Y antigen, Smcx, was also shown to vary among different tissues from newborn and adult mice (Carrel *et al.*, 1996; Shardown *et al.*, 1996). It is, therefore, possible that GRPR could escape inactivation in test tissues such as

fibroblasts and lymphoblasts (and other tissues) described in the section 4. 2. 2, but could be subject to inactivation in the brain, resulting in no GRPR expression and the observed phenotype.

Significantly, Rao *et al.* (1994) reported one autistic patient had a duplication of Xp22.1-pter (1. 2. 8), where ML's translocation breakpoint was found and also where the GRPR gene was mapped (Rao *et al.*, 1994).

## 6. 5 SYNDECAN-2 (SDC2) GENE

The Syndecans and glypicans (Fig. 6. 6) are two major families of cell surface proteoglycans, which regulate the activities of secreted signalling molecules. These two families associate with the cell-membrane in different ways. The glypicans are anchored to the peripheral membrane through glycosyl-phosphatidyl-inositol linkage (Fig. 6. 6). Syndecans have transmembrane-spanning regions. All proteoglycans except glypicans contain a core protein with conserved hydrophobic transmembrane and cytoplasmic domains, but divergent extracelluar domain (Fig. 6. 7). The latter carries several glycosaminoglycan (GAG) chains that vary in number of chains and length (Table 6. 6). In particular, their pattern of sulfation in the GAG chains (e. g. chondroitin, keratan, dermatan or heparan sulfate) provides selective expression of sulfated chain sequences associated with specific function. For example, the aggrecan, carrying chondroitin and keratan sulfates, binds with link protein (dipeptide; see Fig. 6. 7) to hyaluronan to form multimolecular aggregates in cartilage (Hardingham and Fosang, 1992). Decorin and serglycin, carrying chondroitin and dermatan sulfates, bind to transforming growth

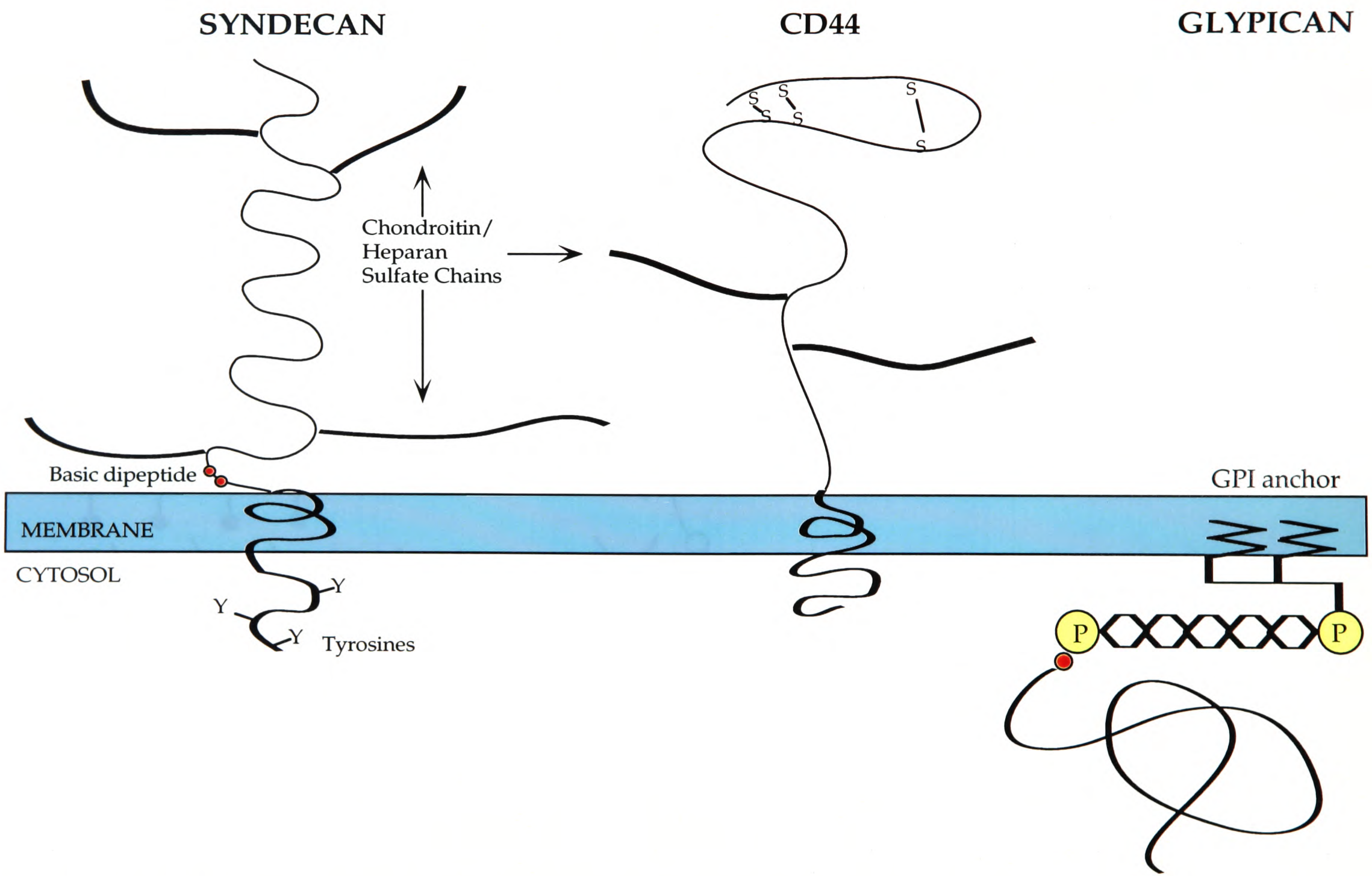



Fig. 6. 6. Schematic diagram of the Syndecan, CD44, and Glypican proteoglycans. Y indicates the attachment for a tyrosine internalization signal in the cytoplasmic domain, and 'GPI' the glycosyl phosphatidylinositol anchor (Hardingham *et al.*, 1992).

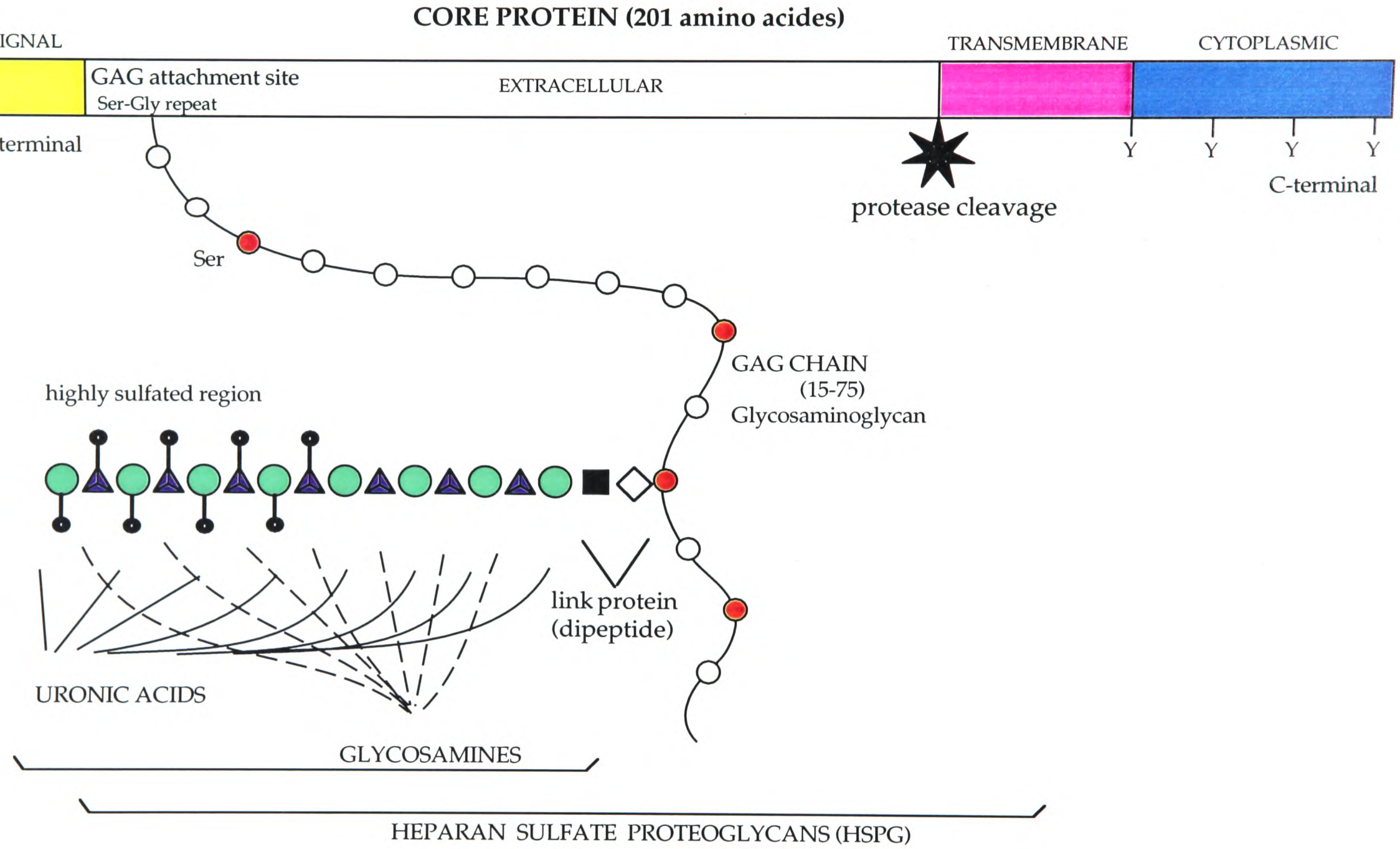



Fig. 6. 7. Schematic structure of the core protein of proteoglycans, showing the signal, extracellular, transmembrane, cytoplasmic domains, and the attachment of the GAG chain at the serine-glycine repeat sequence of the extracelluar domain, and disaccharide chains of uronic acids and glycosamines, attached to serines of the GAG chain. The protease cleavage site is shown at the boundary of the extracellular and transmembrane domains. Y indicates the attachment for a tyrosine internalization signal in the cytoplasmic domain.

| Proteoglycan | M Wof Core Protein | Type of GAG chain | No of GAG chain | Loction | Functions | Locus | myc gene | Original designation |
|---|---|---|---|---|---|---|---|---|
| Aggrecan | 210,000 | chondroitin sulfate + keratan sulfate | ~130 | cartilage | mechanical support | | | |
| Decorin | 40,000 | chondroitin sulfate/ dermatan sulfate | 1 | connective tissues | binds to type I collagenand TGF-β | 12q & Xq | | |
| Serglycin | 20,000 | chondroitin sulfate/ dermatan sulfate | 10-15 | secretory vesicles in white blood cells | packaging secretory molecules | | | |
| CD44 | 40,000 | chondroitin sulfate heparan sulfate | 0 | lymphocytes | hemopoietic epithelial cell | 11pter-p13 | | Pgp-1, H-CAM |
| Betaglycan | 36,000 | chondroitin sulfate/ heparan sulfate | 1 | cell surface and matrix | binds to TGF-β | | | TGF-β typeIII |
| Syndecan-1 | 32,000 | chondroitin sulfate heparan sulfate | 5 | fibroblast and epithelial cell surface | cell adhesion; bind FGF | 2p23 | Nmyc | FGF binding HSPG |
| Syndecan-2 | 20,200 | heparan sulfate | 4 | mesenchyme cells | precursors of hard & connective tissue | 8q22.1 | myc | fibroglycan |
| Syndecan-3 | 45,000 | chondroitin sulfate heparan sulfate | 8 | smooth muscle cells | adhesive cell-matrix interaction | 1p32 | Lmyc | N-syndecan |
| Syndecan-4 | 19,500 | heparan sulfate | 3 | smooth muscle cells adipocytes | focal adhesion component | 20q12-13 | Bmyc | ryudocan amphglycan |

Table 6. 6. Example and comparison of some human proteoglycans.

factor–β (TGF-β) at the surface of collagen fibrils and packages secretory molecules in white blood cells for storage, respectively (Hardingham *et al.*, 1992). Betaglycan, also known as type TGF-β III receptor, carries chondroitin and dermatan or heparan sulfates and binds to TGF-β on the cell surface (Lopez-Casillas *et al.*, 1991; Wang *et al.*, 1991). CD44 (Fig. 6. 6), carrying chondroitin and heparan sulfates, is the main cell surface receptor for hyaluronan with a postulated role in matrix adhesion, lymphocyte activation and lymph node homing (Brown *et al.*, 1991; Hardingham *et al.*, 1992) (Table. 6. 6). The switching of different chain types also modifies the organization of the pericellular and extracellular matrices for different biological needs.

Syndecans (syndecan meaning 'to bid together' in Greek) are the most ubiquitous cell surface heparan sulfate proteoglycans (HSPG) with a membrane-spanning core protein. Their GAG chains consist of disaccharide repeats of a uronic acid and one of the aminosugar of either N-acetylglucosamine or N-acetylgalactosamine, which is sulfated (Fig. 6. 7). One tyrosine at the end of the transmembrane domain and three tyrosines in the cytoplasmic domain (Fig. 6. 6 and 6. 7) are conserved among the Syndecans, but the mechanism of their function as signal transducters is unknown. The protease cleavage site at the beginning of the transmembrane domain (Fig. 6. 7) is also invariant, which may facilitate the release of bound ligands from the Syndecan extracellular domain (Bernfield *et al.*, 1992).

Syndecans evolved from a common ancestor and have maintained a high conservation of signal (95%), hydrophobic transmembrane (96%) and

cytoplasmic domains (100%) among mouse, rat, hamster and human. The extracellular domain is variable in length and amino acid sequence (70% sequence identity between human and mouse), but maintains the same location of the GAG attachment at the serine site (Doolittle, 1992) (Fig. 6. 7).

There are four Syndecans, numbered as based on the first member to be identified (Table 6. 6 and Fig. 6. 8). Syndecan-1 (SDC1) originated from the cDNA of mouse and human mammary epithelia, hamster kidney fibroblasts, and rat microvascular endothelia; Syndecan-2 (SDC2) from the cDNA of human fetal lung fibroblasts, mouse 14 day embryos, and rat liver; Syndecan-3 (SDC3) from the cDNA of chick embryo limb buds and new born rat Schwann cells; and Syndecan-4 (SDC4) from the cDNA of chick embryo limb buds and rat microvascular endothelia (Reviews of Bernfield *et al.*, 1992) (Bernfield *et al.*, 1992). SDC1 and SDC3 carry both chondroitin sulfate and heparan sulfate while SDC2 and SDC4 only heparan sulfate. These two sets show similarities within and differences between sets, both structurally and functionally. The sequence identity in conserved transmembrane and cytoplasmic domains (rat) between SDC2 and SDC4 are 77%, between SDC1 and SDC3 70% while SDC2/SDC4 and SDC1/SDC3 are 65%. SDC2 and SDC4 are similar in size, about 200 amino acids, and almost half the size of SDC1 and SDC3 (Fig. 6. 8). The level of expression of both SDC2 and SDC4 increases as oligodendrocytes differentiate from proliferative late progenitors to postmitotic mature oligodendrocytes. In contrast, the expression of SDC1 and SDC3 shows an inverse developmental pattern. This suggests at least two levels of regulation exist during the differentiation process of oligodendrocyte lineage progression

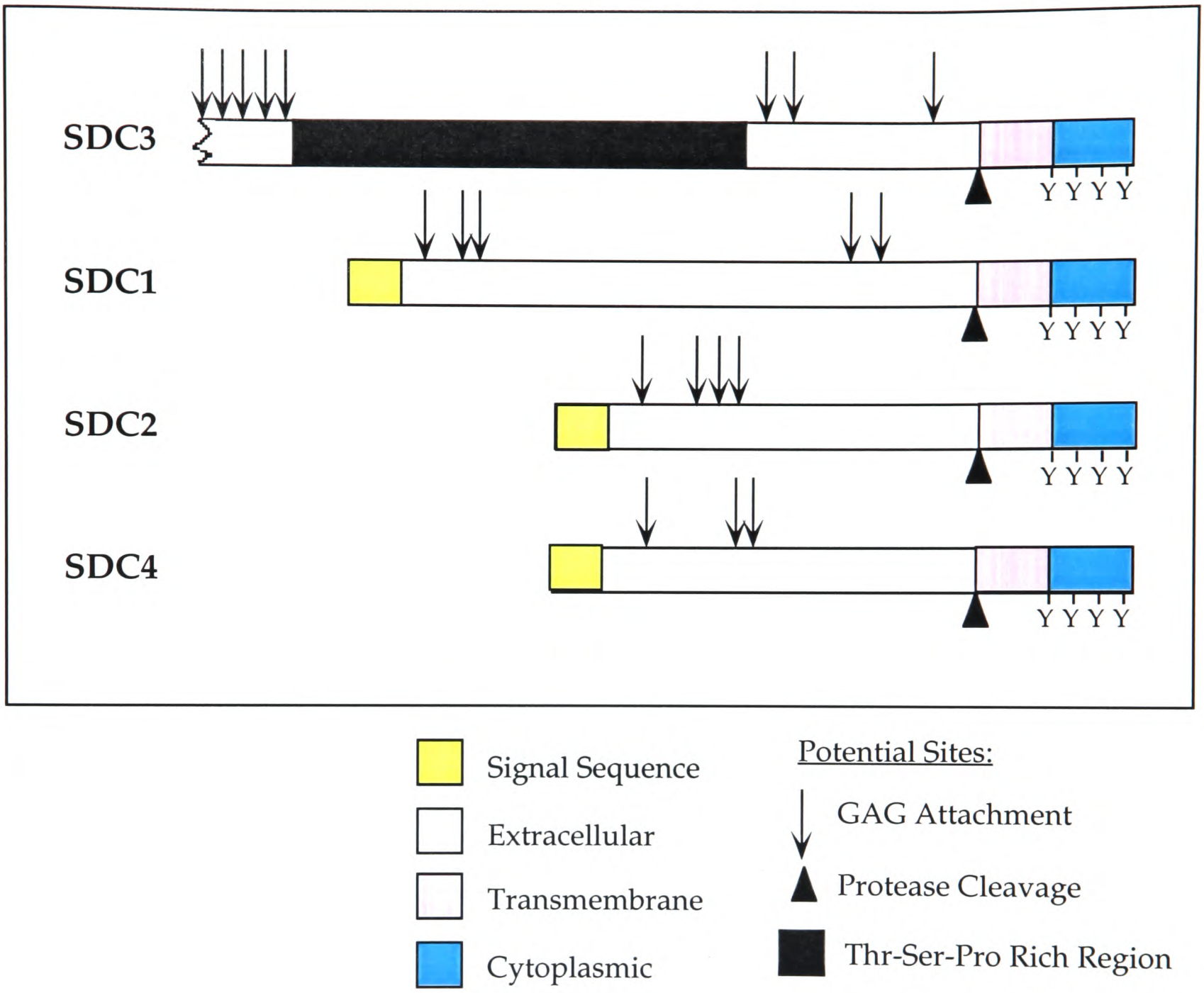



Fig. 6. 8. Core proteins of the four known human syndecans showing the number and locations of the GAG attachment sites. Y indicates the attachment for a tyrosine internalization signal in the cytoplasmic domain.

and myelinogenesis (Bansal *et al.*, 1996). Normal hepatic stellate cells synthesize all four Syndecans, and the expression of SDC1 and SDC3 was altered in activated hepatic stellate cells in liver fibrosis (chronic cholestatic liver disease). This suggests that SDC1 and SDC3 can activate growth factors, giving rise to fibrogenesis (Roskams *et al.*, 1996).

The individual Syndecans have distinct roles. SDC1 which localizes to basolateral surfaces on fibroblast and epithelial cells functions as a matrix and growth factor receptor. SDC1 binds diverse proteins found in the cellular microenvironment including extracellular matrix components (e. g. fibronectin), peptide growth factors (e. g. FGF family), protease inhibitors (e. g. antithrombin III), and circulating lipoprotein (Fig. 6. 9) (Reviewed by Bernfield *et al.*, 1992) (Bernfield *et al.*, 1992). SDC1 has been the most extensively studied of the four Syndecans. It is likely that the other Syndecans may also bind to a variety of molecules in the cellular microenvironment via their heparan sulfate chains, acting as receptors or as co-receptors for ligands as seen in SDC1. Some proteins like viral coat proteins may possibly bind to the cell surface of the negatively charged heparan sulfate before cellular infection (Shieh *et al.*, 1992). SDC3 expression has been detected in heart, vascular smooth muscle and endothelia, stratified epithelia, and neural cells, including Schwann cells and neonatal cortical neurons (Carey *et al.*, 1992). It is also expressed in high amounts during a critical phase of chicken limb chondrogenesis in which limb mesenchymal cells condense and interact with one another before depositing a cartilage matrix (Gould *et al.*, 1992). SDC4, the least studied Syndecan, has been found

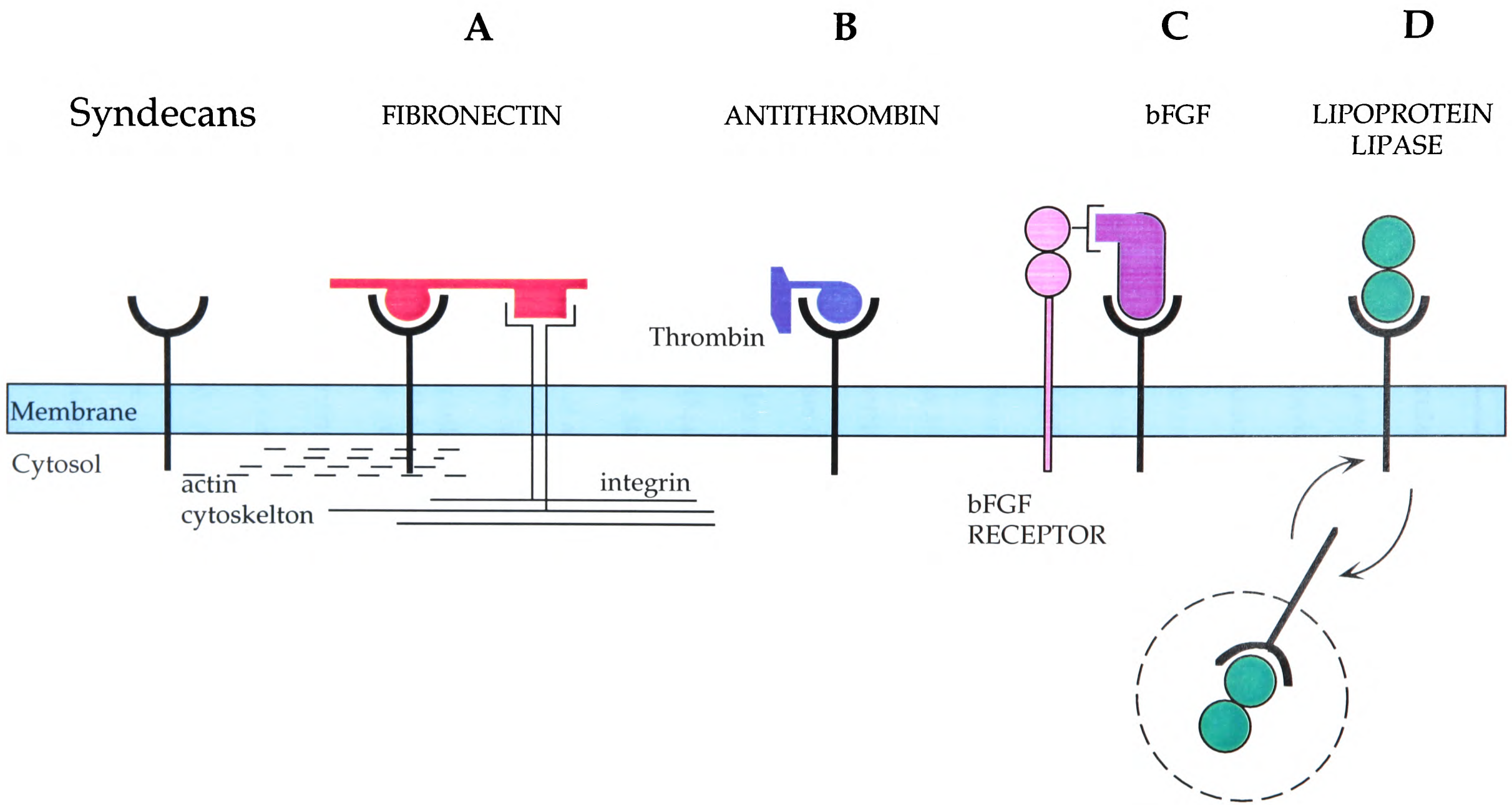



Fig. 6. 9. Possible functions of the syndecans at the cell surface. Syndecans bind a variety of molecules via their extracellular heparan sulfate chains and can act as the sole or co-receptors for ligands. (A) Cells bind to fibronectin via SDC1; (B) Antithrombin III binds to SDC1 and SDC4 which allows the formation of a complex with thrombin (Kojima *et al.*, 1992); (B) bFGF binds to SDC1 (Bernfield *et al.*, 1992); (D) Lipoprotein lipase is bound to the cell surface HSPG of a Syndecan followed by cycling to and from an intracellular compartment (Bernfield *et al.*, 1992).

in fibroblasts, adipocytes, vascular smooth muscle and endothelia (Kojima *et al.*, 1992b) and appeared to be associated with focal adhesions with fibronectin, laminin or type 1 collagen (Woods and Couchman, 1994).

SDC2, also known as human heparan sulfate proteoglycan or fibroglycan [MIM: J04621] (McKusick, 1994), is the Syndecan of relevance to the patient's phenotype of this thesis. Molecular genetic studies have shown that 3.4 kb of the 3′ end of this gene was isolated by screening a human lung fibroblast expression cDNA library with monoclonal antibody 6G12 (Marynen *et al.*, 1989). The epitope of this antibody was shared by two core proteins which were derived from heparitinase treatment of cell surface-associated HSPG of human lung fibroblasts (Lories *et al.*, 1987). The SDC2 gene was mapped to chromosome 8q23 by *in situ* hybridization and hybridization to DNA of somatic cell hybrids (Marynen *et al.*, 1989). Sequencing revealed four potential GAG attachment sites near the N-terminus, and a hydrophobic membrane domain (Marynen *et al.*, 1989).

Like the other Syndecans, SDC2 interacts as a receptor or a co-receptor for diverse proteins, both intracellularly, with the actin containing cytoskeleton, and on the outer surface of the cell membrane, with growth factor and matrix components that support the shaping, maintenance and repair of an organism (David *et al.*, 1993). The latter binding is predominantly between the highly anionic sulfate groups on the GAG and clusters of basic amino acid arranged in a three-dimensional array on the protein. SDC2, whose cytoplasmic domain is closely homologous to SDC4, does not have a focal adhesion component (Woods *et al.*, 1994).

SDC2 occurs in mesenchyme derived from the paraxial and lateral plate mesoderm and from the neural crest, near and on cells during middle and late stages of development in embryonic tissues. It is exclusively expressed on mesenchymal cells that form the precursors of the connective and hard tissues (David *et al.*, 1993). SDC2 antibodies have revealed large amounts of SDC2 on and around certain mesenchymal cells, but no detection in epithelial cells at all stages of development (David *et al.*, 1993). In contrast, SDC1 is mainly detected in epithelia but was also expressed in some mesenchymal cells, only transiently (Bernfield *et al.*, 1992; David *et al.*, 1993). During tooth, kidney and limb development, SDC1 was transiently expressed on the tip epithelium before condensation (Solursh *et al.*, 1990). The expression of SDC1 is lost before condensation (Bernfield *et al.*, 1992). SDC2 was weakly expressed in the early embryo, accumulated during the morphogenetic phase and at the moment of cell lineage differentiation, and persisted in the perichondrium, periosteum and connective tissue cells (David *et al.*, 1993). In situ hybridization analyses confirmed that the expression of SDC2 was limited to mesenchymal tissues. During development, this partial overlap of the expression of SDC1 and SDC2 may undergo crucial changes by complementing the expression (Bernfield *et al.*, 1992).

SDC3 may also associate with mesenchymal cell condensation. SDC3 was expressed in high amounts at the onset of chicken limb chondrogenesis where limb mesenchymal cells condensed and interacted with one another before depositing a cartilage matrix (Gould *et al.*, 1992). SDC3 contains a 182

amino acid domain with threonine, serine, and proline repeats (Fig. 6. 7). This domain is often identified in mucin-like proteins and suggested that SDC3 mediates adhesive cell-matrix interactions (Bernfield *et al.*, 1992; Gould *et al.*, 1992).

The four Syndecan genes are dispersed on different chromosomes and each Syndecan gene is located near a member of the myc gene family. SDC1 is next to Nmyc, SDC2 close to myc, SDC3 near Lmyc, and SDC4 on the same chromosome as Bmyc (Spring *et al.*, 1994) (Table 6. 7).

| Syndecan | Locus | Myc | Locus |
|---|---|---|---|
| 1 | 2p23 | Nmyc | 2p24.1 |
| 2 | 8q22.1 | Myc | 8q24.12-q24.13 |
| 3 | 1p32 | Lmyc | 1p34.3 |
| 4 | 20q12-13 | Bmyc | 20q |

Table 6. 7. The Myc family of cellular oncogene.

## 6. 6 SDC2 AND MULTIPLE EXOSTOSES

The translocation breakpoint in patient ML occurred in what appears to be a non-coding region on chromosome 8 (5. 3. 3), about 30 kb distal to the Syndecan-2 (SDC2) gene. Although the SDC2 gene was not disrupted, its gene expression might be altered by a change of chromatin structure (position effect) resulting from the translocation (6. 2).

Placing a gene in another region of the genome can silence or enhance its transcription because of an alteration of chromatin domain or structure. Alterations such as these are called 'position effect' mutations and may result

from unequal crossing over or from chromosomal aberrations such as translocations, retroviral integrations, inversions, deletions, or duplications (Bedell *et al.*, 1996). Position effects are of two types: variegated (V) and stable (S). V-type position effects can follow placement of a gene from a decondensed chromatin (euchromatin) region to a condensed (heterochromatin) region of the genome (Table 6. 8). These two cytologically distinct regions can be recognized by incorporation of bromodeoxyuridine

| Eukaryotic Genome | | | | |
|---|---|---|---|---|
| Chromatins | Interphase | Replication | BrdU stain | Characteristics |
| Euchromatin | Decondensed | Early | Light | Single-copy sequences, genes |
| Heterochromatin | Condensed | Late | Dark | middle-repetitive sequences<br>highly repetitive sequences |

Table 6. 8. Two cytological states, euchromatin and heterochromatin, of the eukaryotic genome.

(BrdU) (2. 15. 1) during the cell cycle. The euchromatin region, which mostly contains genes, is decondensed in interphase and replicates early. When BrdU is added late in the cell cycle (2. 15. 1), this region can be seen as a light band by incorporating less BrdU. In contrast, the AT rich heterochromatin region is highly condensed, replicates late, and is visible as a dark band in BrdU staining. Normally these states remain unchanged, but when a gene is integrated into a heterochromatic region, the activity of the translocated gene can be repressed because the chromatin remains condensed and transcriptional inactivation of the heterochromatin spreads to the

euchromatin (Bedell *et al.*, 1996; Milot *et al.*, 1996). This heterochromatin spreading can extend as far as 2 Mb of DNA into the euchromatic region. The extent of spreading often varies among cells, giving rise to a gradient effect (Bedell *et al.*, 1996). Consequently the final phenotype may be variegated with patches of normal and mutant tissues.

One possible example of a V-type position effect is described in a benign, slowly progressive myopathy, Facioscapulohumeral muscular dystrophy (FSHD: MIM 158900). FSHD maps to nearby heterochromatin, in the telomeric region of chromosome 4q. When Southern blots of *Eco*RI-digested DNA were hybridized to a subtelomeric probe, a very large hybridizing fragment was seen in normal individuals, but DNA from FSHD patients showed much smaller bands. These patients were considered to have a reduced copy number of a 3.2 kb repetitive sequence that is recognized by the probe. The 3.2 kb sequence was not transcribed or expressed in the patients (Hewitt *et al.*, 1994). The FSHD gene may be located proximally to the tandem 3.2 kb repeat and silenced by a position effect by the deletion, which moved the gene closer to the 4q telomere (Bengtsson *et al.*, 1994). V-type position effects can also be caused by spreading of X-inactivation from the X-inactivation centre into the autosomal segment of the X chromosome in an X:autosomal translocations (Bedell *et al.*, 1996; Cattanach, 1994).

S-type position effects involve only euchromatin and are also known as *cis-trans* position effects. Some genes cluster as a gene family (e. g. alpha- and beta- globin gene, Class I HLA heavy chain gene, HOX gene, and histone

gene families) and occupy an extended domain at a specific subchromosomal region. If such multigenes in a family coordinate for a multimeric protein consisting of subunits coded by two different alleles, the intragenic/interallelic recombination by unequal crossing-over can delete one or more genes and change the properties of the multimeric protein. For example, the *cis* configuration ($m^1$ $m^2$/+ +) is phenotypically normal. The *trans* configuration ($m^1$ +/+ $m^2$) is mutant. This observation suggests that the mRNA transcribed from a (+ +) chromatid functions normally, whereas the mRNAs transcribed from ($m^1$ $m^2$), ($m^1$ +), or (+ $m^2$) chromatids do not. One of the classic examples of this type is described below in the alpha- and beta-globin gene clusters.

A haemoglobin molecule, an oxygen-carrying pigment, is a tetromeric protein composed of four subunits, two polypeptide chains of alpha-globin, two of beta–globin, and four iron-containing ring compounds. The alpha-globin gene family clustered in 30 kb on chromosome 16 encodes a zeta gene ($\zeta$) and two active alpha genes ($\alpha_1$ and $\alpha_2$). The beta-globin gene family clustered in 50 kb on chromosome 11 consists of episilon ($\varepsilon$), two different gamma genes ($G\gamma$ and $A\gamma$), a delta gene ($\delta$), and a beta gene ($\beta$). Each gene is expressed sequentially at a different stage of development (Fig. 6. 10A). The $\zeta$ and $\varepsilon$ genes are only active in the embryo, the $\gamma$ genes normally in only the fetus, and $\beta$ and $\delta$ genes only after birth (Fig. 6. 10A) (Cao *et al.*, 1996; Emery, 1984). The function of haemoglobin is reduced or prevented by mutation in

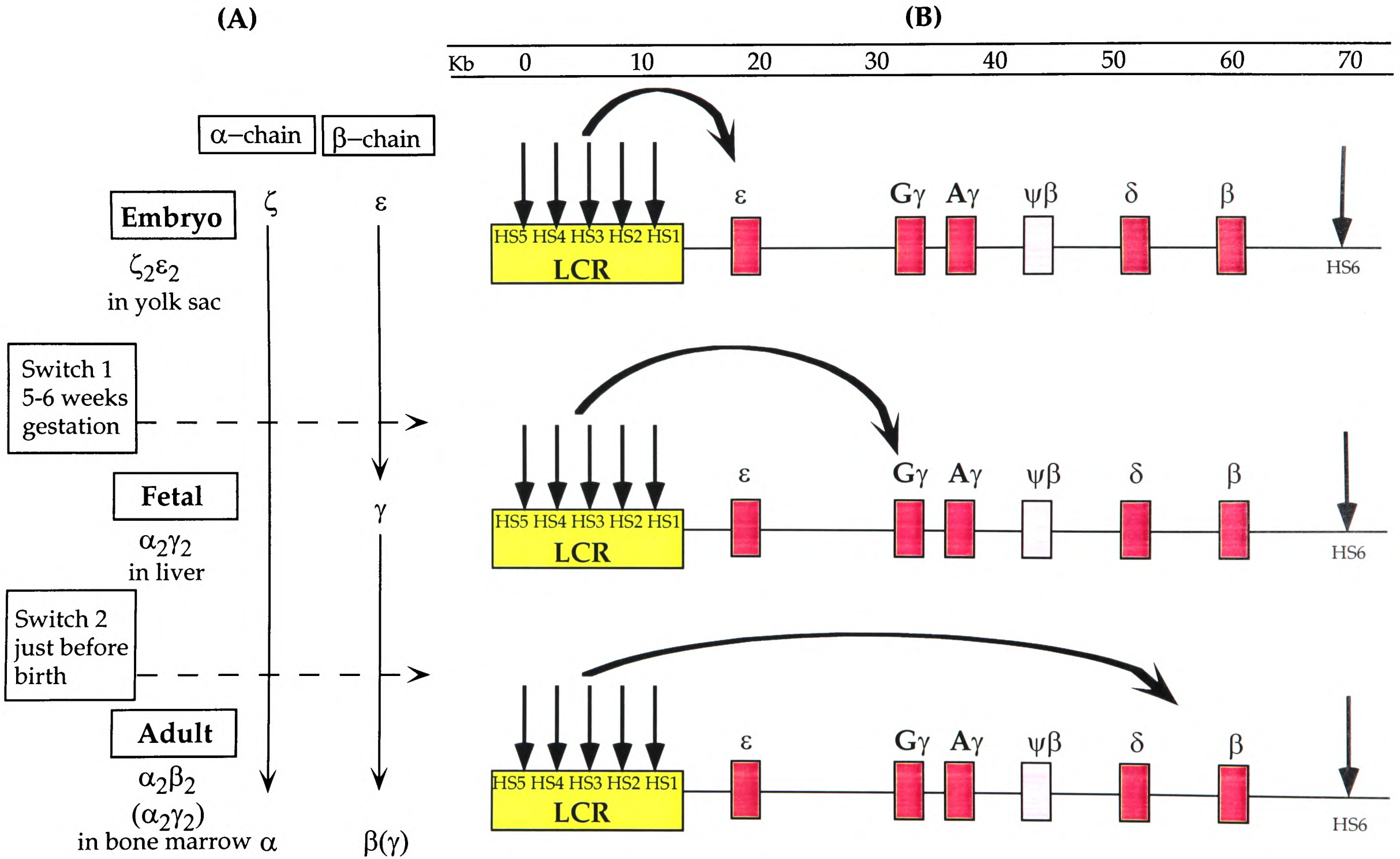



Fig. 6. 10 (A). Human haemoglobin switching occurs at two distinct developmental stages. (B) Gene expression in the beta-globin gene cluster is controlled by common locus control regions (LCR). The LCRs consists of five erythroid-specific DNase 1-hypersensitive sites (HS1-5). The 'HS6' region at the 3' end may also be involved in the switching. Horizontal arrows mark the direction of transcription of expressed genes. The vertical arrows indicate an enhancer effected by the LCR on the indicated genes. The ψβ gene is a pseudogene.

the gene coding for either the α or β polypeptides and also from intragenic recombination produced by unequal crossing over in the globin gene clusters, giving rise to anemias known as 'thalassaemia' (Cao *et al.*, 1994; Emery, 1984). Many of the most severe thalassaemias result from deletions of part of a cluster by unequal crossing-over. The α-thalassaemia are usually caused by gene dosage effects. Normal individuals have four alpha-globin genes (αα/αα). Individuals with two alpha-globin genes (α-/α- or αα/-- phase) suffer mild symptoms. With only one alpha gene (α-/--), the excess beta chains form an unusual tetramer $\beta_4$, which causes HbH disease and the complete absence of alpha genes (--/--) results in a lethal thalassaemia, hydrops fetalis.

The unequal crossing-over in the alpha-globin cluster appears more common, probably because the introns in the alpha genes are much shorter and present less impediment to mispairing between nonhomologous genes. However, the unequal crossing-over between linked genes in the beta-globin cluster occurs in Hb Lepore and Hb Kenya types of thalassaemia, producing fused genes, a δβ gene and a GγAγ gene, respectively.

Deletion of the regulatory domain of a gene(s) can alter its expression. The globin genes in the cluster coordinate their expression using a *cis*-acting regulatory region, the locus control region (LCR) (Cao *et al.*, 1994). The LCR is located about 40 kb and 10 kb upstream of the alpha- and beta-globin clusters, respectively. The LCR deletion caused the Hispanic, English

and Dutch γζβ thalassaemias although their β-globin gene is intact (Townes and Behringer, 1990). Five DNase I hypersensitive (HS) sites were found in a 21 kb in length of the β-globin LCR and one downstream (Fig. 6. 10B). In cells where the globin genes are not expressed, the DNA in these genes are relatively resistant to DNase I and they are tightly packaged into chromatin. In erythroid cells, the LCR is an HS site and exposes binding sites for *trans*-acting factors, which alter the chromatin configuration required for subsequent globin gene expression (Smith and Stillman, 1989). The two HS sequences, 5 and 6, at opposite ends of the locus (Fig. 6. 10B) were found to have a high similarity to those of the scaffold-associated regions (Grosveld *et al.*, 1987), suggesting chromatin structure associates with gene expression.

A *cis*- acting sequence similar to the globin LCR has been identified which governs parent-specific methylation and gene expression of a megabase-size chromosomal region at 15q11. This region contains two adjacent but oppositely imprinted domains. Lack of expression of the paternally imprinted domains causes Prader-Willi syndrome (PWS: MIM 176270) characterized by gross obesity and moderate mental retardation. Lack of expression of the maternally imprinted domains causes Angelman syndrome (AS: MIM 105830) characterized by growth retardation and severe mental retardation. Deletion of the paternal copy of PWS or the maternal copy of AS has no effect, but deletion of either of the imprinting control centers, some distance away from these genes can cause either syndrome.

In the example of campomelic dysplasia (CMPD1; MIM 211970) and

aniridia (AN1;MIM 106210), affected patients were known to have a causative translocation breakpoint, but the breakpoints occurred 80-100 kb outside the genes and did not physically disrupt them. CMPD1 is a dwarfism with skeletal defects such as congenital bowing and angulation of the long bones, and is also associated with sex-reversal. Foster *et al.* (1994) cloned a translocation chromosome and identified an SRY-related gene, SOX9, located 88 kb distal to the chromosome 17q translocation on the 5′ end. This gene is involved in both bone formation and control of testis growth. Nullisomy for SOX9 is lethal. SOX9, like DSS (dosage-sensitive sex-reversal gene) on chromosome Xp21, causes male-to-female sex reversal with varying degrees of external genitalia development. Three patients were sex-reversed (46, XY females) and their 17q breakpoints were mapped 50 kb or more from 5′ end of the SOX9 gene.

The translocation breakpoint also identified to have occurred on the 3′ end of a gene in two aniridia pedigrees. AN1 is characterized by a visible absence of the iris and is caused by loss-of-function mutations of the PAX6 gene on 11p13. The translocation breakpoint in these two pedigrees did not disrupt the PAX6 gene, but in both cases the chromosomal breakpoints occurred 85-125 kb distal to the 3′ end of PAX6 (Fantes *et al.*, 1995). Crolla *et al.* (1996) described another case of a position effect in one sporadic aniridia patient whose translocation breakpoint occurred 100 kb distal to the 3′ end of PAX6 (Crolla *et al.*, 1996).

The translocation reported here nearby the SDC2, or possibly in relation to another gene in 8q22.1 may be another example of a position effect.

Other genes assigned to this region on 8q22.1 are shown in Fig. 6. 11 and Table 6. 10.

Cohen Syndrome (COH1; MIM 216550), characterized by hypotonia, short stature, obesity and prominent incisors, has been assigned to the 8q22-q23 (Wood *et al.*, 1993). The patient ML shows a few features of COH1 syndrome conditions such as hypotonia, short stature, and small hands, but she does not exhibit the abnormal facies, opthalmological abnormalities, nor joint hypermobility that characterize Cohen's syndrome (Bolton *et al.*, 1995).

The genes for erythrocyte carbonic isoenzymes, CA1, CA2, and CA3 (MIM 114800) have been assigned to 8q22. Four other evolutionary related CA genes (CA4, CA5, CA6 and CA7) have been identified on other chromosomal regions (Davis *et al.*, 1987). CA1 and CA2 are found to be deficient in the syndrome of osteopetrosis with renal tubular acidosis (MIM 259730) (Kendall and Tashian, 1977) and CA3 is found in high concentration in muscle (Wade *et al.*, 1986).

Cleidocranial dysplasia (CCD; MIM 119600), featured by persistently open skull sutures, dental anomalies, and often vertebral malformation, appears to be heterogeneous. Three patients (mother and daughter, and unrelated infant) had rearrangement of chromosome 8q22 (Ramesar *et al.*, 1996). In linkage studies in CCD 5 families, the CCD gene was assigned to 6p21 (Feldman *et al.*, 1995) and the gene for bone morphogenetic protein-6 (BMP6; MIM 112266) located on chromosome 6 has been suggested to be a candidate gene, but the locus on chromosome 8q22 has not been completely excluded.

Klippel-Feil syndrome (KFS; MIM 148900), characterized vertebral

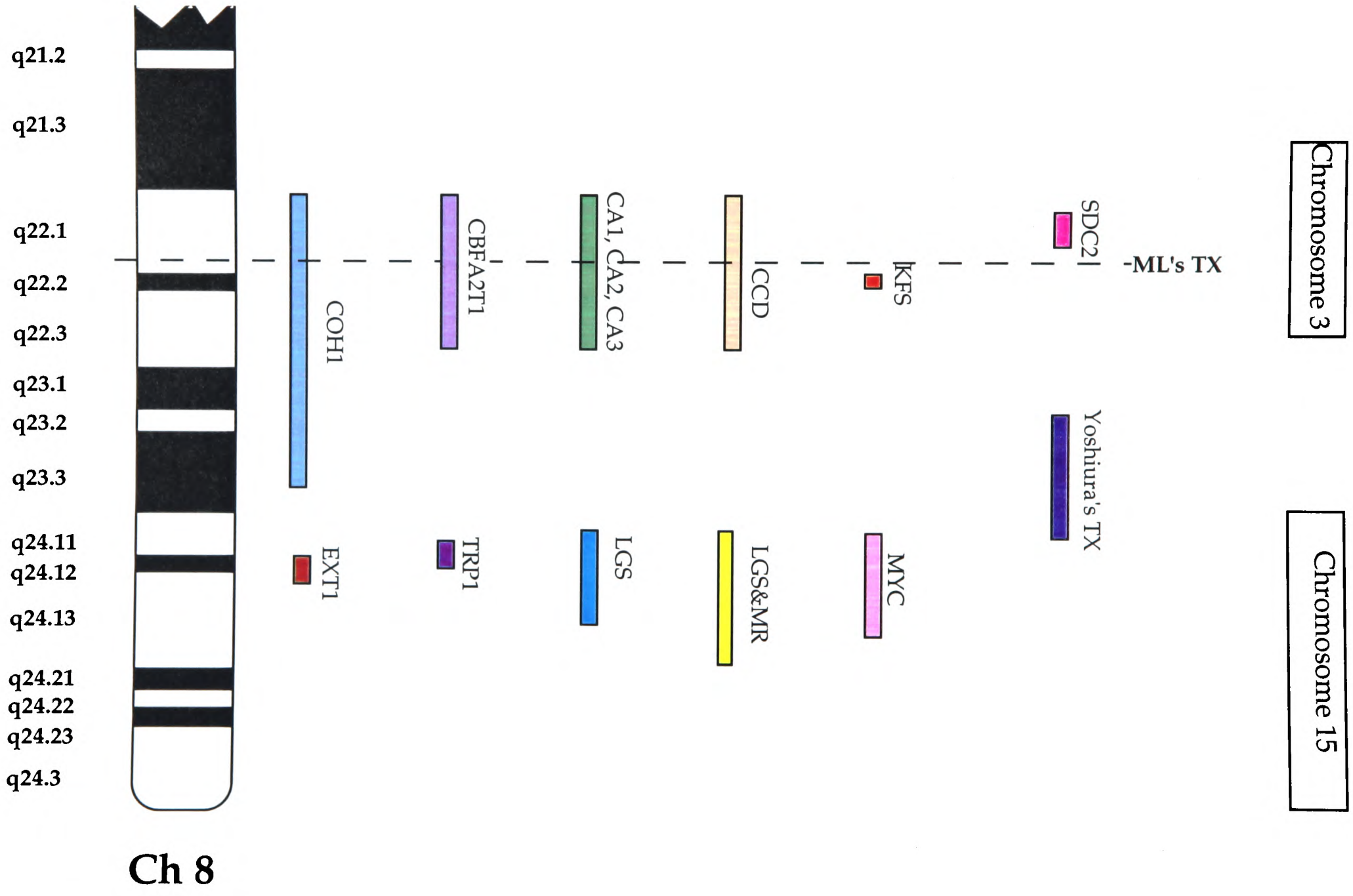



Fig. 6. 11. Genes in the 8q22-24 region and murine chromosomal homologs.

| Gene | Name | Location | Comments | MIM # |
|---|---|---|---|---|
| SDC2 | Syndecan-2 | 8q22.1 | associated with bone formation | 142460 |
| MYC | Myc family of cellular oncogenes | 8q24.12-q24.13 | associated with Burkitt lymphoma | 133700 |
| KFS | Klippel-Feil syndrome | 8q22.2 | characterized cervical vertebral fusion | 148900 |
| CCD | Cleidocranial dysplasia | 8q22.1 | associated with skeletal disorders | 119600 |
| CA 1, 2, and 3 | carbonic isoenzymes | 8q22 | found in erythrocytes | 114800 |
| MYBL1 | avian myeloblastosis viral oncogene homolog-like 1 | 8q22-q23 | expressed mainly in testis | 159405 |
| CBFAT1 | core-binding factor alpha subunit 2 | 8q22 | associated with acute myeloid leukemia | 133435 |
| COH1 | Cohen Syndrome | 8q22-q23 | characterized with hypotonia, abnormal facies, and prominent incisors | 216550 |
| ODF1 | outer dense fiber of sperm tails 1 | 8q22 | transcribed in primary spermatocytes | 182878 |
| EXT1 | Multiple exostoses Type 1 | 8q12.13 | presence of multiple exostoses | 133700 |
| TRPS1 | Trichorhinophalangeal syndrome | 8q24.12 | characterized with abnormal facies | 190350 |
| LGS | Langer-Giedion syndrome | 8q24.11-q24.13 | characterized with abnormal facies and multiple exostoses | 150230 |
| LGS & MR | LGS and mental retardation | 8q24.11-q24.13 | characterized with abnormal facies and multiple exostoses | 150230 |
| Yoshiura's translocation | Multiple exostoses | 8q22-q23 | translocation associated with multiple exostoses | 133700 |

Table 6. 9. Genes in the 8q22-24 region.

fusion. Cervical and upper thoracic vertebra fusion is classified for type I, hemivertebrae and occipitoatlantal fusion for type II, and cervical, lower thoracic or lumbar fusion for type III (Gunderson *et al.*, 1967), and probably further existence of a fourth type associated with sacral agenesis (Raas-Rothschild *et al.*, 1988). Clark *et al.* (1995) identified cosegregation with a paracentric inversion inv(8)(q22.2q23.3) (Clarke *et al.*, 1995) among a 4-generation KFS family with cervical fusion with malformation of largyngeal cartilages and mild to severe vocal impairment.

The ODF1 (outer dense fibre of sperm tails 1 ), MYBL1 (V-MYB avian myeloblastosis viral oncogene homolog-like 1 ), and CBFA2T1 (core-binding factor alpha subunit 2 translocation to 1) have been also assigned to 8q22 region (Fig. 6. 11 and Table 6. 10).

Alteration of SDC2 gene expression in patient ML has not yet been determined, but the translocation may have altered nearby chromatin structure and modified gene expression. The translocation breakpoint in patient ML occurs within euchromatic regions of Xp22.13 and 8q22.1, and would seem not to represent a V-type position effect. On the der(8), the translocated portion of the X chromosome to chromosome 8 does not contain the X-inactivation centre (which maps to Xq13 of the long arm of the X chromosome) and this excludes X-inactivation spreading into the chromosome 8 region of the translocation. On the der(X), the X-inactivation centre on Xq13 is too distant from the translocation breakpoint to influence expression of the chromosome 8 material. It is unknown whether a locus control region for the SDC2 gene exists, but if it does, it is more likely to be at

the 5′ end of the gene. The translocation occurred 30 kb outside of the 3′ end of SDC2 and disruption of an LCR region, although possible, is less likely than if it had occurred outside the 5′ end. The translocations occurring outside the 3′ end of the PAX6 gene have been suggested to be disruptive of the gene by virtue of the position of the translocation. The translocation observed in patient ML with respect to SDC2 may be another example of this phenomenon.

The expression of SDC2 has been shown to be confined in mesenchymal cells that form the precursors of connective and skeletal tissues during middle and late states of development in embryonic tissues (David *et al.*, 1993). The biological functions of SDC2 are consistent with the pathological nature of multiple exostoses (1. 3).

## 6.7 FUTURE WORK

The biological functions of GRPR and SDC2 are consistent with a role in the pathophysiology of autism and multiple exostoses respectively in patient ML. Confirmation of whether a defective GRPR or SDC2 causes autism or multiple exostoses should come from observing rearrangements in the GRPR or SDC2 gene in other unrelated patients, and through analysis of the effect of these rearrangements on gene function.

For hereditary multiple exostoses, 70% of mutations have been detected in EXT1 or EXT2 [Philippe, 1997 #174]. Therefore the remaining 30% should be examined for mutations in SDC2 to determine its involvement in multiple exostoses.

Point mutations of both GRPR and SDC2 genes can be screened by RT-PCR assay from mRNA of large numbers of unrelated patients at the cDNA level (2. 2. 8) using primers designed to amplify coding DNA. Deviant heteroduplexes can be detected by conformation-sensitive gel electrophoresis (CSGE 2. 14) and polymorphisms or mutations can be confirmed by sequencing (2. 13. 2). Mutations of the GRPR or SDC2 exon/intron boundaries can also be rapidly scanned by PCR analysis at the genomic cDNA level.

The cDNA of the GRPR and SDC2 genes are not large, about 3.0 kb. With 3-5 exons screening for mutations should be relatively easy. However, Autism and multiple exostoses appear to be heterogeneous and cosegregation of a single locus with disease may not be easy to detect. Individuals carrying a susceptibility allele may have a higher risk for these diseases, but may still be unaffected. In addition, it will be difficult to distinguish between neutral and pathogenic variants and there is a possibility that some neutral mutations detected in patient samples, particularly if they are rare in the general population, will not also be detected in the control group. Thus a mutation may look statistically related to the phenotype but may in fact be neutral. A mutation cannot be classified definitely as pathogenic until confirmation by functional analysis.

Alternatively, or concurrently, the association between the disease and known GRPR or SDC2 polymorphisms should be compared by genotypes of unrelated affected and unaffected individuals from the population (1. 4. 2. 1d). The transmission disequilibrium test (TDT) should be undertaken to avoid

biased ascertainment and genotype the parents of the patients (1. 4. 2. 1d; page 145). For autism, Cos-17, DXS43, and GRPR exon-2 contain polymorphic sites (3. 3. 2; page 130), which could be used in these studies. Polymorphism in SDC2 are currently unknown, but the same method (3. 3. 2; page 130) can be applied for a search for polymorphic di-and tri-repeat sites for in SDC2 gene.

A candidate gene should have an expression pattern consistent with the disease phenotype or at least a subset of expressing tissues. The hypothesis that the mutation (a leucine substitution) caused Gerstmann-Straussler-Scheiner (GSS) syndrome was confirmed by the transgenic approach. The mouse prion protein gene modified with a leucine substitution was injected into fertilized oocytes to produce transgenic mice which later developed neurodegeneration characterized in GSS syndrome (Hsiao *et al.*, 1990). Similarly, transgenic mice with a mutant ataxin gene containing the expanded CAG repeat developed ataxia and Purkinje cell degeneration, confirming the cause of spinocerebellar ataxia type 1. To model these disorders caused by 'gain-of-function' mutations, a mutant gene was transferred into fertilized oocytes and integrated randomly in the mouse genome.

For some disorders, the phenotype is reversible, in particular, those where mutations cause loss of function. The cultured disease cells from a patient cell line that displays the mutant phenotype because of loss of function may be complemented by the transfection of a cloned normal allele (Strathdee *et al.*, 1992). Diseases caused by loss of gene function can be also confirmed by producing a human disease-like phenotype in a 'knock-out'

mouse in which the gene has been experimentally inactivated. This can be achieved by introducing a germline mutation to the mouse ortholog (gene targeting). Cloned DNA can be transferred into fertilized oocytes by pronuclear microinjection, or a retroviral vehicle. Because of technical difficulties, limited-scale production, and random insertion, mouse embryonic stem (ES) cell lines are commonly used as a vehicle for transferring exogenous DNA into the germline. Mouse ES cells, derived from the inner cell mass of the blastocyst of 3.5-day post-coitum embryos, can be cultured *in vitro*. They retain the capacity to differentiate into the different tissue types of the adult mouse. While in culture, ES cells can be genetically modified by insertion of exogenous DNA or by the introduction of subtle mutations. The modified ES cells can then be injected into isolated blastocysts of another mouse strain and then implanted into a pseudopregnant foster mother of the same strain as the blastocyst. The developing embryo is a chimera, containing two populations of cells derived from different zygotes, those of the blastocyst and the implanted ES cells. Fully transgenic mice can be accomplished by backcrossing of chimeras and subsequent interbreeding.

Gene expression and function of GRPR (Wada *et al.*, 1991) and SDC2 (David *et al.*, 1993; Marynen *et al.*, 1989) have been investigated by manipulations *in vitro*, in cultured cells, or cell extracts. Gene targeting allows the study of gene function in whole organisms in a variety of different cellular environments. Autism and multiple exostoses may have a complex etiology with multiple genetic and environmental components. However, once candidate genes matched to a pathogenesis are identified, breeding

experiments can be used to bring different combinations of disease genes together, and the effect of different genetic backgrounds in different strains of mice and of different environmental factors can be investigated. Animal models of complex genetic diseases such as Alzheimer's, atherosclerosis, and hypertension, are being established.

Dosage imbalance of GRPR and long-range gene irregulation of SDC2 as might have occurred in patient ML (6. 3 and 6. 5) may be examined by YAC transgenics (Rosenberg *et al,* 1993). YACs containing specific copy-numbers of a gene can be separated from yeast cells by PFGE (2. 3. 3). The purified YAC can be transferred into ES cells by using an artificial lipid vesicle (liposome). The lipid bilayer of the liposome fuses to phospholipids of the plasma membrane and the cloned genes can be entered into the cytoplasm of a cell and migrate to the nucleus through the pores of the nuclear envelope. This system allows the transfer of not only large genes but also *cis*-acting regulatory elements. A 660-kb hypoxanthine phosphoribosyltransferase (Jakobovits *et al.,* 1993) and 400 kb-amyloid protein precursor gene (Lamb *et al.,* 1993) were early examples of larger genes transferred into mouse ES cells. This methods was also used to study dosage imbalances in Down syndrome, trisomies 13 and 18, and Charcot-Marie-Tooth syndrome type 1A (Patel and Lupski, 1994).

In some cases, animal models of disease have phenotypes that are markedly different from the homologous human phenotype (Erickson *et al,* 1990). The Lesch-Nyhan syndrome phenotype did not manifest in the HPRT- mice produced by gene targeting because mice have an alternative purine salvage pathway when HPRT is deficient (Wu and Melton, 1993).

When the salvage pathway was also disrupted, the HPRT- mice manifested the disease. This study presented complexities which when analyzed led to new understanding of biochemical pathways and the relationship between gene and phenotype. Animal models of disease allow detailed examination of the physiology of disease and provide an opening for novel therapy and treatment.

In summary, future work in the genetics of Autism and multiple exostoses as a result of this thesis will be the identification and examination of polymorphisms and mutations in these two genes and their functional importance to the disease phenotypes. Whether or not these are genes of major effect, they may be significant in uncovering currently unknown biochemical pathways important to normal and variant phenotypes.

# REFERENCES


Adams, M., Soares, M., Kerlavage, A., Fields, C. and Venter, J. (1993) 3400 expressed sequence tags identify diversity of transcripts in human brain. *Nature Genet.*, **2**, 173-179.

Ahn, J., Ludecke, H.-J., Lindow, S., Horton, W.A., Lee, B., Wagner, M.J., Horsthemke, B. and Wells, D.E. (1995) Cloning of the putative tumour suppressor gene for hereditary multiple exostoses (EXT1). *Nature Genet.*, **11**, 137-143.

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

Anand, R. and Southern, E.M. (1990) In *Pulsed field gel electrophoresis--A Practical Approach*, (Rickwood, D. & Hames, B.D. eds.), Gel Electrophoresis of Nucleic acids. IRL Press, Oxford,

Anand, R., Villasante, A. and Tyler-Smith, C. (1989) Construction of yeast artificial chromosome libraries with large inserts using fractionation by pulsed field gel electrophoresis. *Nucleic Acids Res.*, **17**, 3425-3433.

Anastasi, A., Erspamer, V. and Bucci, M. (1971) Isolation and structure of bombesin and alytensin, two analogous active peptides from the skin of the European amphibians Bombina and Alytes. *Experientia*, **27**, 166-167.

Anderson, G.M. and Hoshino, Y. (1987) In *Handbook of Autism and Pervasive Developmental Disorders.*, (Cohen, D.J., Donnellan, A. & Paul, R. eds.), Neurochemical studies of autism. Wiley, New York,

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

Asperger, H. (1944) Die autistischen psycopathen im kindesalter. *Archiv fur Psychiatrie und Nervenkrankheiten.*, **119**, 76-136.

Bachevalier, J. (1996) Brief Report: Medial temporal lobe and autism: a putative animal model in primates. *J. Autism Dev. Disord.*, **26**, 217-220.

Bailey, A., Le Couteur, A., Gottesman, I., Bolton, P., Simonoff, E., Yuzda, E. and Rutter, M. (1995) Autism as a strongly genetic disorder: evidence from a British twin study. *Psychol. Med.*, **25**, 63-77.

Bailey, A., Le Couteur, A., Rutter, M., Pickles, A., Yuzda, E., Schmidt, D. and Gottesman, I. (1991) *Obstetric and neurodevelopmental data from the British twin study of autism.* London.

Baker, P., Piven, J., Schwartx, S. and Patil, S. (1994) Brief report: duplication of chromosome 15q11-13 in two individuals with autistic disorder. *J. Autism Dev. Disord.*, **24**, 529-535.

Ballabio, A. (1991) Contiguous deletion syndromes. *Curr. Opin. Genet. Dev.*, **1**, 25-29.

Bansal, R., Kumar, M., Myrratm, J. and Ofeufferm, S.E. (1996) Developmental and FGF-2-mediated regulation of Syndecans (1-4) and Glypican in oligodendrocytes. *Mol. Cellular Neuroscience*, **7 (4)**, 276-288.

Baron-Cohen, S., Wheelwright, S., Stott, C., Bolton, P. and Goodyer, I. (1997) *Is there a link between engineering and autism?* SAGE, London.

Bartack, L., Rutter, M. and A, C. (1975) A comparative study of infantile autism and specific developmental receptive language disorder. I. The children. *Br. J. Psychiat.*, **126**, 127-145.

Bauman, M. (1991) Microscopic neuroanatomic abnormalities in autism. *Pediatrics*, **31**, 791-796.

Bauman, M.L. and Kemper, T.L. (1985) Histonantomic observations of the brain in early infantile autism. *Neurology*, **35**, 866-874.

Baxendale, S., Bates, G.P., MacDonald, M.E., Gusella, J.F. and Lehrach, H. (1991) The direct screening of cosmid libraries with YAC clones. *Nucleic Acids Res.*, **19**, 6651.

Bedell, M.A., Jenkins, N.A. and Copeland, N.G. (1996) Good genes in bad neighbourhoods. *Nature Genet.*, **12**, 229-232.

Bengtsson, U., Altherr, M.R., Wasmuth, J.J. and Winokur, S.T. (1994) High resolution fluorescence *in situ* hybridization to linearly extended DNA visually maps a tandem repeat associated with facioscapulohumeral muscular dystrophy immediately adjacent to the telomere of 4q. *Hum. Mol. Genet.*, **3**, 1801-1805.

Bernfield, M., Kokenyesi, R., Kato, M., Hinkes, M.T., Spring, J., Gallo, R.L. and Lose, E.J. (1992) Biology of the Syndecans: a family of transmembrane heparan sulfate proteoglycans. *Annu. Rev. Cell Biol.*, **8**, 365-393.

Bhattacharyya, A. and Liley, D.M.J. (1989) The contrasting structures of mismatched DNA sequences containing looped-out bases (bulges) and multiple mismatches (bubbles). *Nucleic Acids Res.*, **17**, 6821-6840.

Bhojraj, S.Y. and Panjwani, J.S. (1993) A new management approach to decompression, posterior stabilization, and fusion for cervical laminar exostosis with cord compression in a case of diaphyseal aclasis. Case report and review of the literature. *Spine*, **18 (10)**, 1376-1379.

Blanton, S.H., Hougue, D., Wagner, M., Wells, D., Yound, I.D. and Hecht, J.T. (1996) Hereditary multiple exostoses: confirmation of linkage to chromosome 8 and 11. *Am. J. Med. Genet.*, **62**, 150-159.

Blomquist, H.K., Bohman, M., Edvinsson, S.O., Gillberg, C., Gustavson, K.-H., Holmgren, G. and Wahlstrom, J. (1985) Frequency of the fragile X syndrome in infantile autism: a Swedish multicenter study. *Clin. Genet.*, **27**, 113-117.

Bodrug, S.E., Roberson, J.R., Weiss, L., Ray, P.N., Worton, R.G. and van Dyke, D.L. (1990) Prenatal identification of a girl with a t(X;4)(p21;q35) translocation: molecular characterization, paternal origin and association with muscular dystrophy. *J. Med. Genet.*, **27**, 426-432.

Bolton, P. and Griffiths, P.D. (1997) Association of tuberous sclerosis of temporal lobes with autism and atypical autism. *Lancet*, **349**, 392-395.

Bolton, P., Macdonald, H., Murphy, M., Scott, S., Yuzda, E., Whitlock, B., Pickles, A. and Rutter, M. (1991) Genetic findings and heterogeneity in autism. *Psychia. Genet.*, **2**, 49.

Bolton, P., Macdonald, H., Pickles, A., Rios, P., Goode, S., Crowson, M., Bailey, A. and Rutter, M. (1994) A case-control family history study of autism. *J. Child Psychol. Psychiat.*, **35**, 877-900.

Bolton, P., Powell, J., Rutter, M., Buckle, V., Yates, J.R.W., Ishikawa-Brush, Y. and Monaco, A.P. (1995) Autism, mental retardation, multiple exostoses and short stature in a female with 46,X,t(X;8)(p22.13;q22.1). *Psychiat. Genet.*, **5**, 51-55.

Bolton, P. and Rutter, M. (1990) Genetic influences in autism. *Internatl. Rev. Psychiat.*, **2**, 67-80.

Bonfield, J.K., Smith, K.F. and Staden, R. (1995) A new DNA sequence assembly program. *Nucleic Acids Res.*, **24**, 4992-4999.

Brask, B.H. (1967) The need for hospital beds for psychiatric children: an analysis based on a prevalence investigation in the County of Arhus. *Ugeskr. Laeg.*, **129**, 1559-1570.

Brett, P., Curtis, D., Gourdie, A., Schnieden, V., Jackson, G., Holmes, D., Robertson, M. and Gurling, H. (1990) Possible linkage of Tourette syndrome to markers on short arm of chromosome 3 (C3p21-14) (Letter). *Lancet*, **336**, 1076.

Brown, C.J., Flenniken, A.M., Williams, B.R.G. and Willard, H.F. (1990) X chromosome inactivation of the human TIMP gene. *Nucleic Acids Res.*, **18**, 4191-4195.

Brown, M.R., Rivier, J. and Vale, W. (1977) Bombesin affects the central nervous system to produce hyperglycemia in rats. *life Sci.*, **21**, 1729-1977.

Brown, T.A., Bouchard, T., St. John, T., Wayner, E. and Carter, W.G. (1991) Human keratinocytes express a new CD44 core protein (CD44E) as a heparan sulfate intrinsic membrane proteoglycan with additional exons. *J. Cell Biol.*, **113 (1)**, 207-221.

Bryson, S.E., Clark, B.S. and Smith, I.M. (1988) First report of a Canadian epidemiological study of autistic syndromes. *J. Child Psychol. Psychiat.*, **29**, 433-445.

Buckle, V.J. and Rack, K.A. (1993) In *Human Genetic Disease Analysis--A Practical Approach*, (Davies, K.E. ed.), Fluorescence *in situ* hybridization. IRL Press, Oxford, pp. 59-82.

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.

Buhler, E.M. and Malik, N.J. (1984) The tricho-rhino-phalangeal syndrome(s): chromosome 8 long arm deletion: is there a shortest region of overlap between reported cases? TRP I and TRP II syndromes: are they separate entities? *Am. J. Med. Genet.*, **19**, 113-119.

Burd, L., Martsolf, J.T., Kerbeshian, J. and Jalal, S.M. (1988) Partial 6p trisomy associated with infantile autism. *Clinic. Genet.*, **33**, 356-359.

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*, **238**, 806-812.

Camargo, M. and Cervenka, J. (1982) Patterns of DNA replication of human chromosomes. II. Replication map and replication model. *Am. J. Hum. Genet.*, **34**, 757-780.

Cao, A., Rosatelli, M.C. and Galanello, R. (1996) Control of beta-thalassaemia by carrier screening, genetic counselling and prenatal diagnosis: the Sardinian experience. *Ciba-Found-Symp.*, **197**, 137-51.

Carey, D.J., Evans, D.M., Stahl, R.C., Asundi, V.K. and Conner, K.J. (1992) Molecular cloning and characterization of N-syndecan, a novel transmembrane heparan sulfate proteoglycan. *j. Cell Biol.*, **117 (1)**, 191-201.

Carrel, L., Clemson, C.M., Dunn, J.M., Miller, A.P., Hunt, P.A., Lawrence, J.B. and Willard, H.F. (1996) X inactivation analysis and DNA methylation studies of the ubiquitin activating enzyme E1 and PCTAIRE-1 genes in human and mouse. *Hum. Mol. Genet.*, **5 (3)**, 391-401.

Carrel, L., Hunt, P.A. and Willard, H.F. (1996) Tissue and lineage-specific variation in inactive X chromosome expression of the murine Smcx gene. *Hum. Mol. Genet.*, **5 (9)**, 1361-1366.

Cattanach, B.M. and Jones, J. (1994) Genetic imprinting in the mouse: Implications for gene regulation. *J. Inherited Metabolic Dis.*, **17**, 403-420.

Cawthon, R., Weiss, R., Xu, G., Viskochil, D., Culver, M., Stevens, J., Robertson, M., Dunn, D., Gesteland, R., O'Connell, P. and White, R. (1990) A major segment of the neurofibromatosis type 1 gene: cDNA sequence, genomic structure, and point mutations. *Cell*, **62**, 193-201.

Chelly, J., Tumer, Z., Tonnesen, T., Petterson, A., Ishikawa-Brush, Y., Tommerup, N., Horn, N. and Monaco, A.P. (1993) Isolation of a candidate gene for Menkes disease that encodes a potential heavy metal binding protein. *Nature Genet.*, **3**, 14-19.

Chess, S. (1971) Autism in children with congenital rubella. *J. Autism Child. Schizophrenia*, **1**, 33-47.

Chugani, H.T., Da Silvea, E. and Chugani, D.C. (1996) Infantile spasms: III prognostic implications of bitemporal hypometabolism in positron emission tomography. *Ann. Neurol.*, **39**, 643-649.

Chumakov, I.M., Le Gall, I., Billault, A., Ougen, P., Soularue, P., Guillou, S., Rigault, P., Bui, H., DeTand, M.F., Barillot, E., Abderrahim, H., Cherif, D., Berger, R., Le Paslier, D. and Cohen, D. (1992) Isolation of chromosome 21-specific yeast artificial chromosomes from a total human genome library. *Nature Genet.*, **1**, 222-225.

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 Genetics*, **6**, 98-105.

Church, G.M. and Gilbert, W. (1984) Genomic sequencing. *Proc. Natl. Acad. Sci. USA*, **81**, 1991-1995.

Ciaranello, R.D., van den Berg, S.R. and Anders, T.F. (1982) Intrinsic and Extrinsic Determinants of Neuronal Development: Relation to Infantile Autism. *J. Autism Dev. Disord.*, **12**, 115-144.

Clarke, R.A., Singh, S., McKenzie, H., Kearsley, J.H. and Yip, M.-Y. (1995) Familial Klippel-Feil syndrome and paracentric inversion inv(8)(q22.2q23.3). *Am. J. Hum. Genet.*, **57**, 1364-1370.

Coleman, M. and Gillberg, C. (1985) *The biology of the Autistic Syndromes.* Praeger, New York.

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**, 347-350.

Comings, D.E. and Comings, B.G. (1988) In *Genetics of Neuropsychiatric Diseases.*, (Wetterberg, L. ed.), The genetics of Tourette syndrome and its relationship to other psychiatric disorders. Macmillan, London.

Comings, D.E. and Comings, B.G. (1991) Clinical and genetic relationships between autism-pervasive developmental disorder and Tourette syndrome: a study of 19 cases. *Am. J. Med. Genet.*, **39**, 180-191.

Comings, D.E., Comings, B.G., Dietz, G., Muhleman, D., Okada, T.A., Sarinana, F., Simmer, R. and Stock, D. (1986) Evidence the Tourette syndrome gene is at 18q22.1 (Abstract). Berlin, 1-620.

Cook, A., Raskind, W., Blanton, S., H, Pauli, R.M., Gregg, R.G., Francomano, C.A., Puffenberger, E., Conrad, E.U., Schmale, G., Schellenberg, G., Wijsman, E., Hecht, J.T., Wells, D. and Wagner, M.J. (1993) Genetic heterogeneity in families with hereditary multiple exostoses. *Am. J. Hum. Genet.*, **53**, 71-79.

Cook, E.H. (1996) Brief Report: Pathophysiology of autism: Neurochemistry. *J. Autism Dev. Disord.*, **26**, 221-225.

Cook, E.H., Courchesne, R., Lord, C., Cox, L.N., Yan, S., Lincoln, A., Haas, R., Courchesne, E. and Leventhal, B.L. (1997) Preliminary evidence of linkage between the serotonin transporter and autistic disorder. *Mol. Psychiat.*, **2**, 247-250.

Cook, E.H.J., Lindgren, V., Leventhal, B.L., Courchesne, R., Linconln, A., Shulman, C., Lord, C. and Courchesne, E. (1997) *Am. J. Hum. Genet.*, **60**, 928-934.

Corjay, M.H., Dobrzanski, D.J., Way, J.M., Viallet, J., Shapira, H., Worland, P., Sausville, E.A. and Battey, J.F. (1991) Two distinct bombesin receptor subtypes are expressed and functional in human lung carcinoma cells. *J. Biol. Chem.*, **266**, 18771-18779.

Courchesne, E., Saitoh, O., Yeung-Courchesne, R., Press, G.A. and Lincoln, A.J. (1994) Abnormality of cerebellar vermian lobules VI and VII in patents with infantile autism: Identification of hypoplastic and hyperplastic subgroups with MR imaging. *Am. J. Roentgenology*, **162**, 123-130.

Courchesne, E., Townsend, J. and Saitoh, O. (1994) The brain in infantile autism: Posterior fossa structures are abnormal. *Neurology*, **44**, 214-223.

Courchesne, E., Yeung-Courchesne, R., Press, G.A., R, H.J. and Jernigan, T.L. (1988) Hypoplasia of cerebellar lobules VI and VII in autism. *New Engl. J. Med.*, **318**, 1349-1354.

Cox, D.R., Burmeister, M., R, P.E., 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.

Cremers, F.P., van de Pol, D.J., van Kerkhoff, L.P., Wieringa, B. and Ropers, H.H. (1990) Cloning of a gene that is rearranged in patients with choroideraemia. *Nature*, **347**, 674-677.

Crolla, J.A., Cross, I., Atkey, N., Wright, M. and Oley, C.A. (1996) FISH studies in a patient with sporadic aniridia and t(7;11)(q31.2;p13). *J. Med. Genet.*, **33**, 66-68.

Cross, S.H. and Bird, A.P. (1995) CpG islands and genes. *Curr. Opin. Genet. Dev.*, **5**, 309-314.

Curatolo, P., Cusmai, R., Cortesi, F., Chiron, C. and Jambaque, I., Dulac, O. (1991) Neuropsychiatric aspects of tuberous sclerosis. *Ann. N. Y. Adad. Sci.*, **615**, 125-127.

Cuttitta, F., Carney, D.N., Mulshine, J., Moody, T.W., Fedorko, J., Fischler, A. and Minna, J.D. (1985) Bombesin-like peptides can function as autocrine growth factors in human small-cell lung cancer. *Nature*, **316**, 823-826.

Daly, M.C., Spindel, E.R., Giladi, E., Grzeschik, K.-H. and Naylor, S.L. (1991) PCR analysis of somatic cell hybrids localizes the neuromedin ß receptor (hNMBr) gene to chromosome 6q21-qter. *Cytogenet. Cell Genet.*, **58**, 1909.

Damasio, H., Maurer, R.G., Damasio, A.R. and Chui, H.C. (1980) Computerized Tomographic Scan findings in patients with autistic behavior. *Arch. Neurol*, **37**, 504-510.

David, G., Bai, X.M., van der Schueren, B., Marynen, P., Cassiman, J.-J. and van den Berghe, H. (1993) Spatial and temporal changes in the expression of fibroglycan (Syndecan-2) during mouse embryonic development. *Development*, **119**, 841-854.

Davies, J.L., Kawaguchi, Y., Bennett, S.T., Copeman, J.B., Cordell, H.J., Pritchard, L.E., Red, P.W., Gough, S.C.L., Jenkins, S.C., Palmer, S.M., Balfour, K.M., Rowe, B.R., Farrall, M., Barnett, A.H., Bain, S.C. and A, T.J. (1994) A genome-wide search for human type 1 diabetes susceptibility genes. *Nature*, **371**, 130-136.

Davis, M.B., West, L.F., Barlow, J.H., Butterworth, P.H.W., Lloyd, J.C. and Edwards, Y.H. (1987) Regional localization of carbonic anhydrase genes CA1 and CA3 on human chromosome 8. *Somat. Cell Molec. Genet.*, **13**, 173-178.

Dawson, G. (1996) Brief report: Neuropsychology of autism: a report on the state of the science. *J. Autism Dev. Disord.*, **26**, 179-184.

Delong, G.R. (1978) In *Autism: A Reappraisal of Concepts and Treatment.*, (Rutter, M. & Schopler, E. eds.), A neuropsychologic interpretation of infantile autism. Plenum Press, New York,

DeLong, G.R., Bean, C. and Brown, F.R. (1981) Acquired reversible autistic syndrome in acute encephalopathic illness in children. *Arch. Neurology*, **38**, 191-194.

Demasio, A.R. and Maurer, R.G. (1978) A neurological model for childhood autism. *Arch. Neurol.*, **35**, 777-786.

Devereux, J., Haeberli, P. and Smithies, O. (1984) A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Res.*, **12**, 387-395.

Devinsky, O., Morrell, M.J. and Vogt, B.A. (1995) Contributions of anterior cingulate cortex to behaviour. *Brain*, **118 (1)**, 279-306.

Dohlman, H.G., Caron, M.G. and Lefkowitz, R.J. (1987) A family of receptors coupled to guanine-nucleotide regulatory proteins. *Biochem.*, **26**, 2657-2663.

Donnai, D. (1987) Gene location in Tourette syndrome. *Lancet*, **I**, 627.

Doolittle, R.F. (1992) Reconstructing history with amino acid sequences. *Protein Sci.*, **1**, 191-200.

Dracopoli, N.C. (1995) *Current Protocols in Human Genetics.* John Wiley & Sons, Chichester.

Eden, M.S., Thelin, J.W., Michalski, K. and Mitchel, J.A. (1985) Partial trisomy 6p and partial monosomy 9p from a *de novo* translocation 46,XY,-9,+der(9)t(6:9)(p21:p24). *Clinic. Genet.*, **28**, 375-384.

Eder, E.G., Oberbauer, R.W. and Ranner, G. (1993) Cervical cord compression in hereditary multiple exostoses. *J. Neurosurg. Sci.*, **37 (1)**, 53-56.

Eiberg, H. and Mohr, J. (1985) Linkage relationships of paraoxonase (PON) with other markers: indication of PON-cystic fibrosis synteny. *Clin. Genet.*, **28**, 265-271.

Eichenbaum, H. (1997) How does the brain organize memories? *Science*, **177**, 330-332.

Erickson, R.P., Bevilacqua, A., Ventra, P.J., Karolyi, J. and Tashian, R.E. (1990) Ectopic expression of chloramphenicol acetyltransferase (CAT) in the cereblIum in mice transgenic for a carbonic anhydrase II promoter-CAT construct that is without apparent phenotypic effect. *Mol. Reproduction Dev.*, **27 (2)**, 102-109.

Emery, A.E.H. (1984) *An introduction to recombinant DNA.* John Wiley & Sons, Chilchester.

Emi, M., Takahashi, E., Koyama, K., Okui, K., Oshimura, M. and Nakamura, Y. (1992) Isolation and mapping of 88 new RFLP markers on human chromosome 8. *Genomics*, **13**, 1261-1266.

Faham, S., Hileman, R.E., Fromm, J.R., Linhardt, R.J. and Rees, D.C. (1996) Heparin structure and interactions with basic fibroblast growth factor. *Science*, **271**, 1116-1120.

Fantes, J., Redeker, B., Breen, M., Boyle, S., Brown, J., Fletcher, J., Jones, S., Bickmore, W., Fukushima, Y., Mannens, M., Danes, S., van Heyningen, V. and Hanson, I. (1995) Aniridia-associated cytogenetic rearrangments suggest that a position effect may cause the mutant phenotype. *Hum. Mol. Genet.*, **4**, 415-422.

Fein, D., Humes, M., Kaplan, E., Lucci, D. and Waterhouse, L. (1984) The question of left hemisphere dysfunction in infantile autism. *Psycholocial Bulletin*, **95**, 258-281.

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

Feldman, G.J., Robin, N.H., Brueton, L.A., Robertson, E., Thompson, E.M., Siegel-Bartelt, J., Gasser, D.L., Bailey, L.C., Zackai, E.H. and Muenke, M. (1995) A gene for cleidocranial dysplasia maps to the short arm of chromosome 6. *Am. J. Hum. Genet.*, **56**, 938-943.

Fish, G.S., Cohen, I.L., Wolf, E.G., Brown, W.T., Jenkins, E.C. and Gross, A. (1986) Autism and the fragile-X syndrome. *Am. J. Psychiat.*, **143**, 71-73.

Fisher, E. and Scambler, P. (1994) Human haploinsufficiency-One for sorrow, two for joy. *Nature Genet.*, **7**, 5-7.

Folstein, S. and Rutter, M. (1977) Infantile autism: a genetic study of 21 twin pairs. *J. Child Psychol. Psychiat.*, **18**, 297-321.

Folstein, S.E. and Rutter, M. (1988) Autism: Familial aggregation and genetic implications. *J. Autism Dev. Disord.*, **18**, 3-30.

Foster, J.W., Dominguesz-Streglich, M.A., Guioli, S., Kwok, C., Weller, P.A., Stevanovic, M., Weissenback, J., Mansour, S., Young, I.D., Goodfellow, P.N., Brook, J.D. and Schafer, A.J. (1994) Campomelic dysplasia and autosomal sex reversal caused by mutations in an SRY-related gene. *Nature*, **372**, 525-530.

Fountain, J., Wallace, M. and Bruce, M. (1989) physical mapping of a translocation breakpoint in neurofibromatosis. *Science*, **244**, 1085-1087.

Francis, F., Benham, F., See, C.G., Fox, M., Ishikawa-Brush, Y., Monaco, A.P., Weiss, B., Rappold, G., Hamvas, R.M.J. and Lehrach, H. (1994) Identification of YAC and cosmid clones encompassing the ZEX-POLA region using irradiation hybrid cell lines. *Genomics*, **20**, 75-85.

Friedman, E. (1969) The 'autistic syndrome' and phenylketonuria. *Schizophrenia*, **1**, 249-261.

Friend, S.H., Bernards, R., Rogelj, S., Weinberg, R.A., Rapaport, J.M., Albert, D.M. and Dryja, T.P. (1986) A human DNA segment with properties of the gene that predisposes to retinoblastoma and osteosarcoma. *Nature*, **323**, 643-646.

Frischauf, A.M. (1987) In *Guide to molecular cloning techniques.*, (Berger, S.L. & Kimmel, A.R. eds.), Construction and characterization of a genomic library in lambda. Academic press,

Frith, U. (1993) Autism. *Scientific American*, 78-83.

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.

Fu, Y.H., Kuhnl, D.P.A. and Pizzuti, A. (1991) Variation of CGG repeat at the fragile X site results in genetic instability: resolution of the Sherman paradox. *Cell*, **67**, 1047-1058.

Ganguly, A., Rock, M. and Prockop, D. (1993) Conformation-sensitive gel electrophoresis for rapid detection of single base differences in double-stranded PCR products and DNA fragments: Evidence for solvent-induced bends in DNA heteroduplexes. *Proc. Natl. Acad. Sci. USA*, **90**, 10325-10329.

Ghaziuddin, M., Sheldon, S., Tsai, L.Y. and Alessi, N. (1993) Abnormalities of chromosome 18 in a girl with mental retardation and autistic disorder. *J. Intellectual Disability Res.*, **37**, 313-317.

Gillberg. (1984) Infantile autism and other childhood psychoses in a Swedish region: epidemiological aspects. *J. Child Psychol. Psychiat.*, **25**, 35-43.

Gillberg, C., Anderson, L., S, S. and Borjesson. (1987) Infantile autism in children of immigrant parents. *B. J. Psyciat.*, **150**, 865-858.

Gillberg, C. and Coleman, M. (1992) *Clinics in Developmental Medicine*. (2nd edn.), The Biology of the Autistic Syndrome. Mac Keith Press, London.

Gillberg, C. and Wahlstrom, J. (1985) Chromosome abnormalities in infantile autism and other childhood psychoses: a population study of 66 cases. *Dev. Med. Child Neurol.*, **27**, 293-304.

Gillberg, I.C., Gillberg, C. and Ahlsen, G. (1994) Autistic behaviour and attention deficits in tuberous sclerosis: a population-based study. *Dev. Med. Child Neurol.*, **36**, 50-56.

Gingell, K., Parmar, R. and Sungum-Paliwal, S. (1996) Autism and multiple pituitary deficiency. *Dev. Med. Child Neurol.*, **38 (6)**, 545-553.

Gish, W. and States, D. (1993) Identification of protein coding regions by database similarity search. *Nature Genet.*, **3**, 266-272.

Gitschier, J., Wood, W.I., Goralka, T.M., Wion, K.L., Chen, E.Y., Eaton, D.H., Vehar, G.A., Capon, D.J. and M, L.R. (1984) Characterization of the human factor VIII gene. *Nature*, **312**, 326-330.

Goldstein, G.I. and Lancy, D.F. (1985) In *Cognitive Development in Atypical Children: Progress in cognitive Development Research*, (Siegel, L.S. & Morrison, F.J. eds.), Cognitive development in autistic children. Springer, New York,

Goodman, R. (1990) Technical note: are perinatal complications causes or consequences of autism? *J. Child Psychol. Psychiat.*, **31**, 809-812.

Gould, S.E., Upholt, W.B. and Kosher, R.A. (1992) Syndecan 3: a member of the syndecan family of membrane-intercalated proteoglycans that is expressed in high amounts at the onset of chicken limb cartilage differentiation. *Proc. Natl. Acad. Sci. USA*, **89 (8)**, 3271-3275.

Grosveld, F., van Assendelft, G.-B., Greaves, D.R. and Kollias, G. (1987) Position-independent, high-level expression of the human beta-globin gene in transgenic mice. *Cell*, **51 (6)**, 975-986.

Gunderson, C.H., Greenspan, R.H., Glaser, G.H. and Lubs, H.A. (1967) The Klippel-Feil syndrome: genetic and clinical reevaluation of cervical fusion. *Medicine*, **46**, 491-512.

Hallmayer, J., Hebert, J.M., Spiker, D., Lotspeich, L., McMahon, W.M., Petersen, P.B., Nicholas, P., Pingree, C., Lin, A.A., CavalliSforza, L.L., Risch, N. and Ciaranello, R.D. (1996) Autism and the X chromosome: Multipoint sib-pair analysis. *Arch. Gen. Psychiatry*, **53**, 985-989.

Hallmayer, J., Pintado, E., Lotspeich, L., Spiker, D., McMahon, W., Petersen, P.B., Nicholas, P., Pingree, C., Kraemer, H.C., Wong, D.L., Ritvo, E., Lin, A., Hebert, J., Cavalli-Sforza, L.L. and Ciaranello, R.D. (1994) Molecular analysis and test of linkage between the FMR-1 gene and infantile autism in multiplex families. *Am. J. Hum. Genet.*, **55**, 951-959.

Hallmayer, J., Spiker, D., Lotspeich, L., McMahon, W.M., Petersen, P.B., Nicholas, P., Pingree, C. and Ciaranello, R.D. (1994) Male-to Male transmission in extended pedigrees with multiple cases of autism. *Am. J. Med. Genet.*, **67**, 13-18.

Hardingham, T.E. and Fosang, A.J. (1992) Proteoglycans: many forms and many functions. *FASEB*, **6**, 861870.

Hashimoto, T., Tayam, M., Miyazaki, M., Murakawa, K. and Kuroda, Y. (1994) MRI measurements of the brain stem and cerebellum in high functioning autistic children. *No to Hattatsu*, **26**, 3-8.

Hecht, J., Hougue, D., Wang, Y., Blanton, S., Wagner, M., Strong, L. and Raskind, W. (1997) Hereditary multiple exostoses (EXT): Mutational studies of familial EXT1 cases and EXT-associated malignancies. *Am. J. Hum. Genet.*, **60**, 80-86.

Hecht, J.T., Hogue, D., Strong, L.C., Hansen, M.F., Blanton, S.H. and Wagner, M. (1995) Hereditary multiple exostosis and chondrosarcoma: linkage to chromosome 11 and loss of heterozygosity for EXT-linked markers on chromosomes 11 and 8. *Am. J. Hum. Genet.*, **56**, 1125-1131.

Heisterkamp, N. (1982) Chromosomal location of human cellular homologs of 2 viral oncogenes. **299**, 747-749.

Hennekam, R.C.M. (1991) Hereditary multiple exostoses. *J. Med. Genet.*, **28**, 262-266.

Herault, J., Petit, E., Marineau, J., Cherpi, C., Perrot, A., Barthelemy, C., Lelord, G. and Muh, J.P. (1996) Serotonin and autism: Biochemical and molecular biology features. *Psychiat. Res.*, **65**, 33-43.

Hewitt, J.E., Lyle, R., Clark, L.N., Valleley, E.M., Wright, T.J., Wijmenga, C., **van** Deutekom, J.C.T., Francis, F., Sharpe, P.T., Hofker, M., Frants, R.R. and Williamson, R. (1994) Analysis of the tandem repeat locus D4Z4 associated with facioscapulohumeral muscular dystrophy. *Hum. Mol. Genet.*, **3**, 1287-1295.

Hoshino, Y., Tachibana, J.R., Watanabe, M., Murata, S., Sokoyama, F., Kaneko M, Yashima, Y. and Kumoshiro, H. (1984) Serotonin metabolism and hypothalamic-pituitary function in children with infantile autism and minimal brain dysfunction. *Jpn J. Psychiatry Neuronl*, **26**, 937-945.

Hou, J., Parrish, J., Ludecke, H.-J., Sapru, M., Wang, Y., Chen, W., Hill, A., Siegel-Bartelt, J., Northrup, H., B, E.F.F., Chinault, C., Horsthemke, B., Wagner, M.J. and Wells, D.E. (1995) A 4-megabase YAC contig that spans the Langer-Giedion syndrome region on human chromosome 8q24.1: use in refining the location of the trichorhinophalangeal syndrome and multiple exostoses genes (TRPS1 and EXT1). *Genomics*, **29**, 87-97.

Hsiao, K.K., Scott, M., Foster, D., Groth, D.F., Dearmond, S.J. and Prusiner, S.B. (1990) Spontaneous neurodegeneration in transgenic mice with mutant prion protein. *Science*, **250**, 1587-1590.

Hunt, A. and Dennis, J. (1987) Psychiatric disorder among children with tuberous sclerosis. *Dev. Med. Child Neurol.*, **29**, 190-198.

Ishikawa-Brush, Y., Powell, F.P., Bolton, P., Miller, A.P., Francis, F., Willard, H.F., Lehrach, H. and Monaco, A.P. (1997) Autism and multiple exostoses associated with an X;8 translocation occurring within the GRPR gene and 3' to the SDC2 gene. *Hum. Mol. Genet.*,

Jakobovits, A., Moore, A.L., L, G.L., Vergara, G.J., Maynard-Currie, C.E., Austin, H.A. and Klapholz, S. (1993) Germ-line transmission and expression of a human-derived yeast artificial chromosome. *Nature*, **362**, 255-258. James, A.L. and Barry, R.J. (1980) A review of psychophysiology in early onset psycosis. *Schizophrenia Bulletin*, **6**, 506-525.

Jeffreys, A.J., Wilson, V. and Thein, S.L. (1987) Hypervariable 'minisatellite' regions in human DNA. *Nature*, **314**, 67-73.

Kanner, L. (1943) Autistic disturbances of affective contact. *Nervous Child*, **2**, 217-250.

Kendall, A.G. and Tashian, R.E. (1977) Erythrocyte carbonic anhydrase I: inherited deficiency in humans. *Science*, **197**, 171-172.

Khoury, M.J. (1994) Case-parental control method in the search for disease-susceptibility genes. *Am. J. Hum. Genet.*, **55**, 414-415.

Knowlton, R.G. and Cohen-Haguenauer, O. (1985) A polymorphic DNA marker linked to cystic fibrosis is located on chromosome 7. *Nature*, **318**, 380-382.

Kojima, T., Shworak, N.W. and Rosenberg, R.D. (1992b) Molecular cloning and expression of two distinct cDNA-encoding heparan sulfate proteoglycan core proteins from a rat endothelial cell line. *J. Biol. Chem.*, **267 (7)**, 4870-4877.

Kolvin, I., Ounsted, C. and Roth, M. (1971) Studies in the childhood psychoses. V. Cerebral dysfunction and childhood psychoses. *B. J. Psyciat.*, **118**, 407-414.

Kotsopoulous, S. and Kutty, K.M. (1979) Histidinemia and infantile autism. *J. Autism Dev. Disord.*, **9**, 55-60.

Krane, I.M., Naylor, S.L., Helin-Davis, D., Chin, W.W. and Spindel, E.R. (1988) Molecular cloning of cDNAs encoding the human bombesin-like peptide neuromedin B: chromosomal localization and comparison to cDNAs encoding its amphibian homolog ranatensin. *J. Biol. Chem.*, **263**, 13317-13323.

Krooth, R.S., Macklin, M.T. and Hilbish, T.F. (1961) Diaphysial aclasis (multiple exostoses) on Guam. *Am. J. Hum. Genet.*, **13**, 340-347.

Kunkel, L.M., Monaco, A.P., Middlesworth, W., D, O.H. and Latt, S.A. (1985) Specific cloning of DNA fragments absent from the DNA of a male patient with an X chromosome deletion. *Proc. Natl. Acad. Sci. USA*, **82**, 4778-4782.

Kurlan, R., Behr, J., Medved, L., Shoulson, I., Pauls, D. and Kidd, K.K. (1987) Severity of Tourette's syndrome in one large kindred: implication for determination of disease prevalence rate. *Neurol.*, **44**, 268-269.

Lamb, B.T., Sisodia, S.S., Lawler, A.M., Slunt, H.H., Kitt, c.A., Kearns, W.G., Pearson, P.L., Price, D.L. and Gearhart, J.D. (1993) Introduction and expression of the 400 kilobase Amyloid precursor proten gene in transgenic mice. *Nature Genet.*, **5 (1)**, 22-30.

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

Le Couteur, A., Rutter, M., Lord, C., Rios, P., Robertson, S., Holgrafer, M. and McLennan, J. (1989) Autism diagnostic interview: a standardized investigator-based instrument. *J. Autism Dev. Disord.*, **19**, 363-387.

Le Merrer, M., Legeal-Mallet, L., Jeannin, P.M., Horsthemke, B., Schinzel, A., Plauchu, H., Toutain, A., Achard, F., Munnich, A. and Maroteaux, P. (1994) A gene for hereditary multiple exostoses maps to chromosome 19p. *Hum. Mol. Genet.*, **3**, 717-722.

Le Merrer, M., Othmane, K.B., Stanescu, V., Lyonnet, S., **van** Maldergem, L., Royer, G., Munnich, A. and Maroteux, P. (1992) The gene for hereditary multiple exostoses does not map to the Langer-Giedion region (8q23-24). *J. Med. Genet.*, **29**, 713-715.

Lebacq-Verheyden, A.-M., Bertness, V., Kirsch, I., Hollis, G.F., McBride, O.W. and Battey, J. (1987) Human gastrin-releasing peptide gene maps to chromosome band 18q21. *Somat. Cell Molec. Genet.*, **13**, 81-86.

LeDoux, J.E. (1987) In *Handbook of physiology: Section 1. The nervous system: Vo. 5. Higher functions of the brain, part I.*, (Blum, F. edn.), Emotion. American Physiological Society., Bethesda,

LeDoux, J.E. (1994) Emotion, memory and the brain. *Scientifc American*, **6**, 32-39.

Lee, M.P., Hu, R.-J., Johnson, L.A. and Feinberg, A.P. (1997) Human *KVLQT1* gene shows tissue-specific imprinting and encompasses Beckwith-Wiedemann syndrome chromosomal rearrangements. *Nature Genet.*, **15**, 181-185.

Legeai-Mallet, L., Jeannin, P., Maroteaux, P., Munnich, A. and Le Merrer, M. (1995) Natural clinic and genetic history of hereditary multiple exostoses. *Am. J. Hum. Genet.*, **57**, A325.

Lehy, T., Accary, J.P., Labeille, D. and Dubrasquet, M. (1983) Chronic administration of bombesin stimulates antral gastrin cell proliferation in the rat. *Gastroenterology*, **84**, 914-919.

Lewis, M.H., Bodfish, J.W., Powell, S.B. and Golden, R.N. (1995) Clomipramine treatment for repetitive movement disorders in mental retardation: A double-blind comparison with placebo. *Am. J. Mental Retardation*, **100**, 299-312.

Lockyer, L. and Rutter, M. (1969) A five- to fifteen-year follow-up study of infantile psychosis: III. Psychological aspects, *BJP*, **115**, 865-882.

Lopez-Casillas, F., Cheifetz, S., Doody, J., Andres, J.L., Lane, W.S. and Massague, J. (1991) Structure and expression of the membrane proteoglycan betaglycan, a component of the TGF-ß receptor system. *Cell(4)*, **785-795**,

Lord, C., Rutter, M., Goode, S., Heemsbergen, J., Jordan, H., Mawhood, L. and Schopler, E. (1989) Autism diagnostic observation schedule: a standardized observation of communicative and social behaviour. *J. Autism Dev. Disord.*, **19**, 185-212.

Lord, C., Rutter, M. and Le Couteur, A. (1994) Autism Diagnosis Interview-revised: A revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *J. Autism Dev. Dis*, **24/5**, 659-685.

Lord, C., Schopler, E. and Revicki, D. (1982) Sex differences in autism. *J. Autism Dev. Disord.*, **12**, 317-330.

Lories, V., de Boeck, H., David, G., Cassiman, J.J. and van den Berghe, H. (1987) Heparan sulfate proteoglycans of human lung fibroblasts. Structural heterogeneity of the core proteins of the hydrophobic cell-associated forms. *J. Biol. Chem.*, **262**, 854-859.

Lotter, V. (1966) Epidemiology of autistic conditions in young children: I. Prevalence. *Social Psychiat.*, **1**, 124-137.

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

Ludecke, H.-J., Wagner, M.J., Nardmann, J., La Pillo, B., Parrish, J.E., Willems, P.J., Haan, E.A., Frydman, M., Hamers, G.J.H., Wells, D.E. and Horsthemke, B. (1995) Molecular dissection of a contiguous gene syndrome: localization of the genes involved in Langer-Giedion syndrome. *Hum. Mol. Genet.*, **4**, 31-36.

Mannens, M. and Wilde, A. (1997) KVLQT1, the rhythm of imprinting. *Nature Genet.*, **15**, 113-114.

Mariner, R., Jackson III, A.W., Levitas, A., Hagerman, R.J., Braden, P.M., McBogg, P.M., Smith, A.C.M. and Berry, R. (1986) Autism, mental retardation, and chromosomal abnormalities. *J. Autism Dev. Disord.*, **16**, 425-440.

Marynen, P., Zhang, J., Cassiman, J.-J., **van** den Berghe, H. and David, G. (1989) Partial primary structure of the 48- and 90-kilodalton core proteins of cell surface-associated Heparan Sulfate Proteoglycans of lung fibroblasts. *J. Biol. Chem.*, **264**, 7017-7024.

Matsuno, T., Ichioka, Y., Yagi, T. and Ishii, S. (1988) Spindle-cell sarcoma in patients who have osteochondromatosis: a report of two cases. *J. Bone Joint Surg.*, **70A**, 137-141.

McBride, P.A., M, A.G., Hertzig, M.E., Sweeney, J.A., Kream, J., Cohen, D.J. and Mann, J.J. (1989) Serotonergic responsivity in male young adults with autistic disorder: results of a pilot study. *Arch. Gen. Psychiatry*, **46**, 312-221.

McDougle, C., J, Price, L.H. and Volkmar, F.R. (1994) Recent advances in the pharmacotherapy of autism and related conditions. *Child Adol. Psychiat. Clinics N. Am.*, **3**, 71-89.

McDougle, C.J., Naylor, s.T., Cohen, D.J., Aghajanian, G.K., Heninger, G.R. and Price, L.H. (1996) Effects of tryptophan depletion in drug-free adults with autistic disorder. *Arch. Gen. Psychiat.*, **53**, 993-1000.

McGaughran, J.M., Ward, H.B. and Evans, D.G.R. (1995) WAGR syndrome and multiple exostoses in a patient with del(11)(p11.2p14.2). *J. Med. Genet*, **32**, 823-824.

McKusick, V.A. (1994) *Mendelian Inheritance in man.* The Johns Hopkins University Press, Baltimore.

Mercer, J.F.B., Livingston, J., Hall, B., Paynter, J.A., Begy, C., Chandrasekharappa, S., Lockhart, P., Grimes, A., Bhave, M., Siemieniak, D. and Glover, T.W. (1993) Isolation of a partial candidate gene for Menkes disease by positional cloning. *Nature Genet.*, **3**, 15-19.

Migeon, B.R., Shapiro, L.J., Norum, R.A., Mohandas, T., Axelman, J. and Dabora, R.L. (1982) Differential expression of steroid sulphatase locus on active and inactive human X chromosome. *Nature*, **299**, 838-840.

Miller, A.P., Gustashaw, K., Wolff, D.J., Rider, S.H., Monaco, A.P., Eble, B., Schlessinger, D., Gorski, J.L., van Ommen, G.J., Weissenbach, J. and Willard, H.F. (1995) Three genes that escape X chromosome inactivation are clustered with a 6 Mb YAC contig and STS map in Xp11.21-p11.22. *Hum. Mol. Genet.*, **4**, 731-739.

Milot, E., Fraser, P. and Grosveld, F. (1996) Position effects and genetic disease. *TIG*, **12 (4)**, 123-126.

Monaco, A.P., Bertleson, C.J., Middlesworth, W., Colletti, C.A., Fischbeck, K.H., Bartlett, R., Pericack-Vance, M.A., Roses, A.D., Aldridge, J. and Kunkel, L.M. (1985) Detection of deletions spanning the Duchenne muscular dystrophy locus with a DNA segment tightly linked to the disease. *Nature*, **316**, 842-845.

Monaco, A.P., Lam, V.M.S., Zehetner, G., Lennon, G.G., Douglas, C., Nizetic, D., Goodfellow, P.N. and Lehrach, H. (1991) Mapping irradiation hybrids to cosmid and yeast artificial chromosome libraries by direct hybridization of Alu-PCR products. *Nucleic Acids Res.*, **19**, 3315-3318.

Monaco, A.P. and Larin, Z. (1994) YACs, BACs, PACs, and MACs: artificial chromosomes as research tools. *TIBTECH*, **12**, 280-286.

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

Moody, T.W., Pert, C.B., Gazdar, A.F., Carney, d.N. and Minna, J.D. (1981) High levels of intracellular bombesin characterize human small-cell lung carcinoma. *Science*, **214**, 1246-1248.

Morton, N.E. (1955) Sequential tests for the detection of linkage. *Am. J. Hum. Genet.*, **7**, 277-318.

Murray, A.W. and Szostak, J.W. (1983) Construction of artificial chromosomes in yeast. *Nature*, **305**, 189-193.

Nabholz, M., Miggiano, V. and Bodmer, W. (1969) Genetic analysis of human-mouse somatic cell hybrids. *Nature*, **223**, 358-363.

Newson, E., Dawson, M. and Everard, P. (1984) The natural history of able autistic people: their management in social context. Summary of the report to the DHSS in four parts. *Communication*, **18**, 1-4.

Neyroud, N., Tesson, F., Denjoy, I., Leibovici, M., Donger, C., Barhanin, J., Faure, S., Gary, F., Coumel, P., Petit, C., Schwartz, K. and Guicheney, P. (1997) A novel mutation in the potassium channel gene KVLQT1 causes the Jervell and Lange-Nielsen cardioauditory syndrome. *Nature Genet.*, **15**, 186-198.

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 cosmid libraries of the human chromosomes X and 21. *Proc. Natl. Acad. Sci. USA*, **88**, 3233-3237.

Ogle, R.F., Dalzell, P., Turner, G., Wass, D. and Yip, M.-Y. (1991) Multiple exostoses in a patient with t(8;11)(q24.11;p15.5). *J. Med. Genet*, **28**, 881-883.

Ornitz, E.M. (1978) In *Autism: A Reappraisal of Concepts.*, (Rutter, M. & Schopler, E. edn.), Biological homogeneity or heterogeneity. Plenum Press, New York,

Ornitz, E.M., Guthrie, D. and Farley, A.H. (1977) The early development of autistic children. *J. Aut. Child Schizophrenia*, **7**, 207-230.

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., Wagner, M.J., Hecht, J.T., Scott Jr, C.I. and Wells, D.E. (1991) Molecular analysis of overlapping chromosomal deletions in patients with Langer-Giedion syndrome. *Genomics*, **11**, 54-61.

Passmore, R. and Robson, J.S. (1996) In *A Companion to Medical Studies.*: Anatomy, Biochemistry and Physiology. Blackwell, Oxford,

Perrett, D.I., Hietanen, M., Oram, M.W. and Benson, P.J. (1992) In *Processing the facial image*, (Bruce, V., Cowey, A., Ellis, A.W. & Perrett, D.I. eds.), Organization and functions of cells responsive to faces in the temporal cortex. Clarendon, Oxford,

Patel, P.I. and Lupski, J.R. (1994) Charcot-Maie-Tooth disease: A new paradigm for the mechanism of inherited disease. *TIG*, **10 (4)**, 128-133.

Peterson, B.S., Bronen, R.A. and Duncan, C.C. (1996) Three cases of symptom change in Tourette's syndrome and obsessive-compulsive disorder associated with paediatric cerebral malignancies. *J. Neurol. Neurosurg. Psychiat.*, **61 (5)**, 497-505.

Philippe, C., Porter, D.E., Emerton, M.E., Wells, D.E., Simpson, A.H.R.W. and Monaco, A.P. (1997) Mutation screening of EXT1 and EXT2 genes in patients with hereditary multiple exostoses. *Am. J. Hum. Genet.*, **in press**,

Pickles, A., Bolton, P., Macdonald, H., Bailey, A., Le Couteur, A., Sim, C.-H. and Rutter, M. (1995) Latent-class analysis of recurrence risk for complex phenotypes with selection and measurement error: a twin and family history study of autism. *Am. J. Hum. Genet.*, **57**, 717-726.

Piven, J., Arndt, S., Bailey, J. and Andreasen, N. (1996) Regional brain enlargement in autism: a magnetic resonance imaging study. *J. Am. Acad. Child Adol. Psyciatry*, **4**, 530-536.

Poncet, M. and Habib, M. (1994) Isolated defect in motivated behaviour following basal ganglia damage. *Revue Neurologique (Paris)*, **150 (8-9)**, 588-593.

Raas-Rothschild, A., Goodman, R.M., Grunbaum, M., Berger, I. and Mimouni, M. (1988) Klippel-Feil anomaly with sacral agensis: an additional subtype, type VI. *Craniofac. Genet. Dev. Biol.*, **8**, 297-301.

Ramesar, R.S., Greenberg, J., Martin, R., Goliath, R., Bardien, S., Mundlos, S. and Beighton, P. (1996) Mapping of the gene for cleidocranial dysplasia in the historical Cape Town (Arnold) kindred and evidence for locus homogeneity. *J. Med. Genet.*, **33**, 511-514.

Rao, P.N., Klinepeter, K., Stewart, W., Hayworth, R., Grubs, R. and Pettenati, M.J. (1994) Molecular cytogenetic analysis of a duplication Xp in a male: further delineation of a possible sex influencing region on the X chromosome. *Hum. Genet.*, **94**, 149-153.

Ray, P.N., Belfall, B. and Duff, C. (1985) Cloning of the breakpoint of an X;21 translocation associated with Duchenne muscular dystrophy. *Nature*, **318**, 672-675.

Reik, W. and Maher, E.R. (1997) Imprinting in clusters: lessons from Beckwith-Wiedemann syndrome. *TIG*, **13 (8)**, 330-334.

Richard, I., Broux, O., Chiannilkulchai, N., Fougerousse, F., Allamand, V., Bourg, N., Brenguier, L., Devaud, C., Pasturaud, P., Roudaut, C., Lorenzo, F., Sebastiani-Kabtchis, C., Schultz, R.A., Polymeropoulos, M.H., Gyapay, G., Auffray, C. and Beckmann, J.S. (1994) Regional localization of human chromosome 15 loci. *Genomics*, **23**, 619-627.

Ridley, R. and Baker, H.F. (1983) In *Ethopharmacology: Primate models of neuropsychiatric disorders.*, (Miczek, K.A. ed.), 'Is there a relationship between social isolation, cognitive inflexibility, and behavioural stereotypy? An analysis of the effects of amphetamine in the marmoset'. Liss, New York.

Riikonen, R. and Amnell, G. (1981) Psychiatric disorders in children with earlier infantile spasms. *Dev. Med. Child Neurol.*, **23**, 747-760.

Riley, J., Butler, R., Ogilvie, D., Finniear, R., Jenner, D., Powell, S., Anand, R., Smith, J.C. and Markham, A.F. (1990) A novel, rapid method for the isolation of terminal sequences from yeast artificial chromosome (YAC) clones. *Nucleic Acids Res.*, **18 (10)**, 1887-2890.

Risch, N. (1990) Genetic linkage and complex diseases, with special reference to psychiatric disorders. *Genet. Epidemiology*, **7**, 3-16.

Risch, N. (1990) Linkage strategies for genetically complex traits I. Multilocus models. *Am. J. Hum. Genet.*, **46**, 222-228.

Robinson, D., Boyd, Y., Cockburn, D., Collinson, M., Craig, I. and Jacobs, P.A. (1990) The parental origin of *de novo* X-autosome translocations in females with Duchenne muscular dystrophy revealed by M27ß methylation analysis. *Genet. Res., Camb.*, **56**, 135-140.

Robson, K.J.J., Chandra, T., MacGillivray, R.T.A. and Woo, S.L.C. (1982) Polysome immunoprecipitation of phenylalanine hydroxylase mRNA from rat liver and cloning of its cDNA. *Proc. Natl. Acad. Sci. USA*, **79**, 4701-4705.

Rolf, L., Haarmann, F., Grotemeyer, K. and Kehrer, H. (1993) Serotonin and amino acid content in platelets of autistic children. *acta Psychiatr Scand*, **87**, 312-316.

Rosengerg, C., Voltz, A.K., Lawler, A.M., Lamb, B.T., Stetten, G. and Gearhart, J.D. (1996) Alterations of yeast artificial chromosome transgenic sequences in stretched embryonic stem-cell chromatin visualized by fluorescence in situ hybridization. *Cytogenet. Cell Genet.*, **75 (1)**, 67-70.

Rosengurt, E. and Sinnett-Smith, J. (1983) Bombesin stimulation of DNA synthesis and cell division in cultures of Swiss 3T3 cells. *Proc. Natl. Acad. Sci. USA*, **80**, 2936-2940.

Roskams, T., Rosenbaum, J., de Vos, R., David, G. and Desmet, V. (1996) Heparan sulfate proteoglycan expression in chronic cholestatic human liver diseases. *Hepatology*, **24 (3)**, 524-532.

Rutter, M. (1970) Autistic children: infancy to adulthood. *Seminars in Psychiatry*, **2**, 435-450.

Rutter, M. (1991) *The New Genetics of Mental Illness.* (McGuffin, P. & Murray, R. eds.), Autism as a genetic disorder. Butterworth-Heinemann, Oxford.

Rutter, M., Bolton, P., Harrington, R., Le Couteur, A., Macdonald, H. and Simonoff, E. (1990) Genetic factors in child psychiatric disorders--I. a review of research strategies. *J. Child Psychol. Psychiat.*, **31**, 3-37.

Rutter, M. and Lockyer, L. (1967) A five to fifteen year follow-up study of infantile psychosis: I Description of sample. *B. J. Psychiat.*, **113**, 1169-1182.

Rutter, M., Macdonald, H., Le Couteur, A., Harrington, R., Bolton, P. and Bailey, A. (1990) Genetic factors in child psychiatric disorders - II. Empirical findings. *J. Child Psychol. Psychiat.*, **31**, 39-83.

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

Schain, R.J. and Freedman, D.X. (1961) Studies on 5-hydroxylindole metabolism in autistic and other mentally retarded children. *J. Pediat.*, **58**, 315-320.

Schmale, G.A., Conrad, E.U. and Raskid, W.H. (1994) The natural history of hereditary multiple exostoses. *J. Bone Joint Surg. Am.*, **76**, 986-992.

Schmickel, R.D. (1986) Contiguous gene syndrome. *J. Pediat.*, **109**, 231-240.

Schopler, E., Andrews, E.E. and Strupp, K. (1979) Do autistic children come from middle class parents? *J. Autism Dev. Disord.*, **9**, 139-152.

Schuler, G.D., Boguski, M.S., Stewart, E.A., Stein, L.D., Gyapay, G., Rice, K., White, R.E., Rodrigues-Tome, P., Aggarwal, A., Bajorek, E., Bentolila, S., Birren, B.B., Butler, A., Castle, A.B., Chiannikulchai, N., Chu, A., Clee, C., Cowles, S., Day, J.R., Dibling, T., Drouot, N., Dunham, I., Duprat, S., East, C., Edwards, C., *et al.* (1996) A gene map of the human genome. *Science*, **274**, 540-546.

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

Sege-Peterson, K., Chamber, J., Page, T., Jones, O.W. and Nyhan, W.L. (1992) Characterization of mutations in phenotypic variants of hypoxanthine phosphoribosyltransferase. *Hum. Mol. Genet.*, **1 (6)**, 427-432.

Shardown, S., Norris, D., Fisher, A. and Brockdorff, N. (1996) The mouse Smcx gene exhibits developmental and tissue specific variation in degree of escape from X inactivation. *Hum. Mol. Genet.*, **5 (9)**, 1355-1360.

Shiang, R., Thompson, L.M., Zhu, Y.-Z., Church, D.M., Fielder, T., Bocian, M., Winokur, S.T. and Wasmuth, J.J. (1994) Mutations in the transmembrane domain of FGFR3 cause the most common genetic form of Dwarfism, achondroplasia. *Cell*, **78**, 335-342.

Shieh, M.T., WuDunn, D., Montgomery, R.I., Esko, J.D. and Spear, P.G. (1992) Cell surface receptors for herpes simplex virus are heparan sulfate proteoglycans. *Dev. Biol.*, **140(1)**, 83-92.

Smalley, S.L. (1991) Genetic Influences in autism. *Psychiatr. Clin. North Am.*, **14**, 125-139.

Smalley, S.L., Asarnow, R.F. and Spence, A. (1988) Autism and Genetics: a decade of research. *Arch. Gen. Psychiatry*, **45**, 953-961.

Smith, S. and Stillman, B. (1989) Purification and characterization of CAF-1, a human cell factor required for chromatin assembly during DNA replication in vitro. *Cell*, **58**, 15-25.

Solomon, L. (1964) Hereditary multiple exostosis. *am. J. Hum. Genet.*, **16**, 351-363.

Solursh, M., Reiter, R.S., Jensen, K.L., Kato, M. and Bernfield, M. (1990) Transient expression of a cell surface Heparan Sulfate Proteoglycan (Syndecan) during limb development. *Dev. Biol.*, **140**, 83-92.

Southern, E.N. (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol*, **98**, 503-518.

Southern, E.N., Anand, R., Brown, W.R.A. and Fletcher, D.S. (1987) A model for the separation of large DNA molecules by crossed field gel electrophoresis. *Nucleic Acids Res.*, **15**, 5925-5943.

Spielman, R.S., McGinnis, R.E. and Ewens, W.J. (1993) Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). *Am. J. Hum. Genet.*, **52**, 506-516.

Spindel, E.R., Chin, W.W., Price, J., Rees, L.H., Besser, G.M. and Habener, J.F. (1984) Cloning and characterization of cDNAs encoding human gastrin-releasing peptide. *Proc. Natl. Acad. Sci. USA*, **81**, 5699-5703.

Spindel, E.R., Zilberberg, M.D., Habener, J.F. and Chin, W.W. (1986) Two prohormones for gastrin-releasing peptide are encoded by two mRNAs differing by 19 nucleotides. *Proc. Natl. Acad. Sci. USA*, **83**, 19-23.

Spring, J., Goldberger, O.A., Jenkins, N.A., Gilbert, D.J., Copeland, N.G. and Bernfield, M. (1994) Mapping of the syndecan genes in the mouse: Linkage with member of the Myc gene family. *Genomics*, **21(3)**, 597-601.

Steffenburg, S., Gillberg, C., Hellgren, L., Andersson, L., Gillberg, I., Jakobsson, G. and Bohman, M. (1989) A twin study of autism in Denmark, Finland, Iceland, Norway and Sweden. *J. Child Psychol. Psychiat.*, **30**, 405-416.

Stickens, D., Burbee, D., Ramos, P. and Evans, G. (1996) Identification of a novel gene family responsible for hereditary multiple exostoses (EXT). *Am. J. Hum. Genet.*, **56**, A6.

Stickens, D., Clines, G., Burbee, D., Ramos, P., Thomas, S., Hogue, D., Hecht, J.T., Lovett, M. and Evans, G.A. (1996) The EXT2 multiple exostoses gene defines a family of putative tumour

suppressor genes. *Nature Genet.*, **14**, 25-32.

Strathdee, C.A., Gavish, H., Shannon, W.R. and Buchwald, M. (1992) Cloning of cDNA for Fanconi's anaemia by functional complementation. *Nature*, **356**, 763-767.

Stutsman, R. (1948) *Merril-Palmer Scale of Mental tests. Reprint of Part III , Mental Measurement of Pre-School Children*. Stoelting Co., Chicago.

Sverd, J. (1991) Tourette syndrome and autistic disorder: a significant relationship. *Am. J. Med. Genet.*, **39**, 173-179.

Tache, Y. and Brown, M. (1982) On the role of bombesin in homeostasis. *Trends Neurosci.*, **5**, 245-248.

Taft, L.T. and Cohen, H.J. (1971) Hypsarrhythmia and infantile autism: A clinical report. *J. Aut. Child Schizophrenia*, **1**, 327-336.

Tager-Flusberg, H. (1996) Brief Report: current theory and research on language and communication in autism. *J. Autism Dev. Disord.*, **26**, 169-172.

Tagle, D. and Collins, S. (1992) An optimized Alu-PCR primer pair for human specific amplification of YACs somatic cell hybrid. *Hum. Mol. Genet.*, **1**, 121-122.

Taylor, L.D., Krizman, D.B., Jankovic, J., Hayani, A., Steuber, P.C., Greenberg, F., Fenwick, R.G. and Caskey, C.T. (1991) 9p monosomy in a patient with Gilles de la Tourette's syndrome. *Neurology*, **41**, 1513-1515.

Thein, S.L. and Hinton, J. (1993) A simple and rapid method of direct sequencing using dynabeads. *British J. Heaematology*, **79**, 113-115.

Ton, C.T., Hirvonen, H., Miwa, H., Weil, M.M., Monaghan, P., Jordan, T., van Heyningen, V., Hastie, N.D., Meijers-Heijboer, H., Drechsler, M., Royer-Pokora, B., Collins, F., Saroop, A., Strong, L.C. and Saunders, G.F. (1991) Positional cloning and characterization of a paired box- and homeobox-containing gene from the aniridia region. *Cell*, **67**, 1059-1074.

Townes, T.M. and Behringer, R.R. (1990) Human globin locus activation region (LAR): role in temporal control. *TIG*, **6 (7)**, 219-223.

Tsui, L.C. and Buchwald, M. (1985) Cystic fibrosis locus defined by a genetically linked polymorphic DNA marker. *Science*, **230**, 1054-1057.

Tully, E.M., Mellon, C.D. and McMahon, W.M. (1993) Autism associated with 9p-deletion syndrome. *Biol. Psychiatry*, **33**, 37A-162A.

Tumer, Z., Chelly, J., Tommerup, N., Ishikawa-Brush, Y., Tonnessen, T., Monaco, A. and Horn, N. (1992) Characterization of a 1.0 Mb YAC contig spanning two chromosome breakpoints related to Menkes disease. *Hum. Mol. Genet.*, **7**, 483-489.

Uberbacher, E.E. 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.

van Bakel, I., Holt, S., Craig, I. and Boyd, Y. (1995) Sequence analysis of the breakpoint regions of an X;5 translocation in a female with Duchenne muscular dystrophy. *Am. J. Hum. Genet.*, **57**, 329-336.

van Slegtenhorst, M., de Hoogt, R., Hermans, C., Nellist, M., Janssen, B., Verhoef, S., Lindhout, D., van den Ouweland, A., Halley, D., Young, J., Burley, M., Jeremiah, S., Woodward, K., Nahmias, J., Fox, M., Ekong, R., Osborne, J., Wolfe, J., Povey, S., Snell, R.G., Cheadle, J.P., Jones, A.C., Tachataki, M., Ravine, D., Sampson, J.R., *et al.* (1997) Identification of the Tuberous Sclerosis gene TSC1 on chromosome 9q34. *Science*, **277**, 805-808.

van Woert, M.H., Yip, L.C. and Balis, M.E. (1977) Purine phosphoribosyltransferase in Gilles de la Tourette syndrome. *New Eng. J. Med.*, **296**, 210-212.

Vargha-Khadem, F., Gardian, D.G., Watkins, K.E., Connelly, A., van Paesschen, W. and Mishkin, M. (1997) Differential effects of early hippocampal pathology on episodic and semantic memory. *Science*, **277**,

Vostanis, P., Harrington, R., Prendergast, M. and Farndon, P. (1994) Case reports of autism with interstitial deletion of chromosome 17 (p11.2 p11.2) and monosomy of chromosome 5 (5pter->5p15.3). *Psychiat. Genet.*, **4**, 109-111.

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.

Wada, E., Way, J., Shipira, H., Kusano, K., Lebacq-Verheyden, A., Coy, D., Jensen, R. and Battery, J. (1991) cDNA cloning, characterization and brain region-specific expression of a neuromedin-ß-preferring bombesin receptor. *Neuron*, **6**, 421-430.

Wade, R., Gunning, P., Eddy, R., Shows, T. and Kedes, L. (1986) Nucleotide sequence, tissue-specific expression, and chromosome location of human carbonic anhydrase III: the human CAIII gene is located on the same chromosome as the closely linked CAI and CAII genes. *Proc. Natl. Acad. Sci. USA*, **83**, 9571-9575.

Wainwright, B.J. and Scambler, P.J. (1985) Localisation of cystic fibrosis locus to human chromosome 7cen-q22. *Nature*, **318**, 384-385.

Walker, H.A. (1978) A dermatoglyphic study of autistic patients. *J. Aut. Child Schizophrenia*, **7**, 11-21.

Wallace, M.R., Marchuk, d.A., Andersen, L.B., Letcher, R., Odeh, H.M., Saulino, A.M., Fountain, J.W., Brereton, A. and Nicholson, J. (1990) Type 1 neurofibromatosis gene: Identification of a large transcript disrupted in three NF1 patients. *Science*, **249**, 181-186.

Walsh, J., Wong, H.C. and Dockray, G.J. (1979) Bombesin-like peptides in mammals. *FASEB*, **38**, 2315-2319.

Walsh, J.H. and Wong, H. (1978) In *Methods of Hormone Radioimmunoassay*, (Jaffe, B.M. & Behrman, H.R. eds.), Bombesin-like peptides. Academic, New York,

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., Weiseenback, J. and Goodfellow, P.N. (1994) A method for constructing radiation hybrid maps of whole genomes. *Nature Genet.*, **7**, 22-28.

Wang, D. and Cutz, E. (1994) Simultaneous detection of messenger ribonucleic acids for bombesin/gastrin-releasing peptide and its receptor in rat brain by nonradiolabeled double *in situ* hybridization. *Lab. Investig.*, **70**, 775-780.

Wang, X., Lin, H.Y., Ng-Eaton, E., Downward, J., Lodish, H.F. and Weinberg, R.A. (1991) Expression cloning and characterization of the TGF-ß type III receptor. *Cell*, **67 (4)**, 797-805.

Warren, S.T., Knight, s.J.L. and Peters, J.F. (1990) Isolation of the human chromosomal band Xq28 within somatic cell hybrids by fragile X site cleavage. *Proc. Natl. Acad. Sci. USA*, **87**, 3856-3860.

Wharton, J., Polak, J.M. and Bloom, S.R. (1978) Bombesin-like immunoreactivity in the lung. *Nature*, **273**, 769.

White, R. and Woodward, S. (1985) A closely linked marker for cystic fibrosis. *Nature*, **318**, 382-384.

Wicklund, C.L., Pauli, R.M., Johnston, D. and Hecht, J.T. (1995) Natural history study of hereditary multiple exostoses. *Am. J. Med. Genet.*, **55**, 43-46.

Willard, H.F., Brown, C.J., Carrel, L., Hendrich, B. and Miller, A.P. (1993) Epigenetic and chromosomal control of gene expression: Molecular and genetic analysis of X chromosome inactivation. *Cold Spring Harbor Quant. Symp. Biol.*, **58**, 315-322.

Williamson, D.A. and Bolton, P. (1995) Brief report: atypical autism and tuberous sclerosis in a sibling pair. *J. Autism Dev. Disord.*, **25**, 435-442.

Wing, L. (1979) In *Estimating Needs for Mental Health Care*, (Hafner, H. ed.), Mentally retarded children in Camberwell. Springer, Berlin,

Wing, L. (1981) Sex ratios in early childhood autism and related conditions. *Psychiat. Res.*, **5**, 129-137.

Wood, S., Ben Othmane, K., Bergerheim, U.S.R., Blanton, S.H., Bookstein, R., Clarke, R.A., Daiger, S.P., Donis-Keller, H., Drayna, D., Kumar, S., Leach, R.J., Ludecke, H.-J., Oshima, J., Sadler, L.A., Spurr, N.K., Steinbrueck, T., Trapman, J., Wagner, M., Wang, Z., Wells, D. and Westbrook, C.A. (1993) Report of the first international workshop on human chromosome 8 mapping. *Cytogenet. Cell Genet.*, **64**, 134-141.

Woods, A. and Couchman, J.R. (1994) Syndecan 4 heparan sulfate proteoglycan is a selectively enriched and widespread focal adhesion component. *Mo. Biol. Cell.*, **5 (2)**, 183-192.

Wu, Y., -Q, Heutink, P., de Vries, B.B.A., Sandkuijl, L.A., van den Ouweland, A.M.W., Niermeijer, M.F., Galjaard, H., Reyniers, E., Willems, P.J. and Halley, D.J.J. (1994) assignment of a second locus for multiple exostoses to the pericentric region of chromosome 11. *Hum. Mol. Genet.*, **3**, 167-171.

Wu, C.-L. and Melton, D.W. (1993) Production of a model for Lesch-Nyhan syndrome in hypoxanthine phosphoribosyltransferase-deficient mice. *Nature Genetics*, **3**, 235-240.

Wuyts, W., van Hul, W., Wauters, J., Nemtsova, M., Reyniers, E., van Hul, E., de Boulle, K., de Vries, B.B.A., Hendrickx, J., Herrygers, I., Bossuyt, P., Balemans, W., Fransen, E., Vits, L., Coucke, P., Nowak, N.J., Shows, T.B., Mallet, L., van den Ouweland, A.M.W., McGaughran, J., Halley, D.J.J. and Willems, P.J. (1996) Positional cloning of a gene involved in hereditary multiple exostoses. *Hum. Mol. Genet.*, **5**, 1547-1557.

Yoshiura, K., Inazawa, J., Koyama, K., Nakamura, Y. and Niikawa, N. (1994) Mapping of the 8q23 translocation breakpoint of t(8;13) observed in a patient with multiple exostoses. *Genes, Chromosomes & Cancer*, **9**, 57-61.

The extracted article "Autism and Multiple Exostoses Associated with an X;8 Translocation Occurring Within the *GRPR* Gene and 3′ to the *SDC2* Gene" can be found at http://dx.doi.org/10.1093/hmg/6.8.1241