Characterisation of the TSC1 Candidate Region

on Human Chromosome 9q34

by

Karen Jane Woodward

A thesis submitted for the degree of

Doctor of Philosophy

at the

University of London

June 1995

The Galton Laboratory

Department of Genetics and Biometry

University College London
Abstract

Tuberous sclerosis (TSC) is an autosomal dominant disorder with clinical symptoms ranging from minor skin lesions to more serious manifestations such as mental retardation and seizures. Linkage analysis established heterogeneity and assigned genes to 9q34 (TSC1) and 16p13.3 (TSC2). Detection of TSC-associated deletions on 16p13.3 refined the localisation of the gene and TSC2 has now been identified. Meiotic recombination events in TSC families have defined a consensus candidate region for TSC1 between D9S149 and D9S114, but further refinement gives conflicting positions for the gene.

A positional cloning strategy has been applied to characterise the TSC1 candidate region and identify the gene. Radiation hybrids retaining markers from the target interval were characterised and used as a resource for region specific DNA. The clones isolated plus ones from other sources were confirmed to map to 9q34 by FISH. Localisations were refined by using three lymphoblastoid cell lines with translocations breakpoints within 9q34, thereby subdividing the band into 4 intervals. A total of 32 loci from 28 contigs, and 43 anonymous contigs were mapped enabling a FISH map to be constructed.

Cosmids mapping to the TSC1 candidate region were subsequently screened for transcribed sequences by exon amplification. The initial study led to the identification VAV2, a good candidate for TSC1. It was followed by more extensive analysis of 6 contigs mapping to the same FISH interval as D9S149 and D9S114. A total of 171 putative exons were cloned, 48 potentially different products were sequenced and 34 were inferred to be authentic exons. A subset from 4 contigs were confirmed to be
conserved sequences by cross-species hybridisation and cDNA clones were isolated and partially characterised. The exons and cDNAs isolated are important resources for construction of a 9q34 transcription map and may facilitate identification of the TSC1 gene.
<table>
<thead>
<tr>
<th>Table of contents</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Title page</td>
<td>1</td>
</tr>
<tr>
<td>Abstract</td>
<td>2</td>
</tr>
<tr>
<td>Contents</td>
<td>4</td>
</tr>
<tr>
<td>List of figures</td>
<td>11</td>
</tr>
<tr>
<td>List of tables</td>
<td>15</td>
</tr>
<tr>
<td>Abbreviations</td>
<td>16</td>
</tr>
<tr>
<td>Contributors</td>
<td>19</td>
</tr>
<tr>
<td>Acknowledgements</td>
<td>20</td>
</tr>
</tbody>
</table>

### Chapter 1

1. **Introduction**

1.1. **Clinical aspects of Tuberous Sclerosis (TSC)**

1.2. **Positional Cloning**

1.2.1. **Disease gene localisation**

- 1.2.1.1. Structural abnormalities
- 1.2.1.2. Linkage analysis
- 1.2.1.3. Linkage disequilibrium
- 1.2.1.4. Comparative mapping

1.2.2. **Physical mapping**

- 1.2.2.1. Fluorescence in situ hybridisation
- 1.2.2.2. Radiation hybrid mapping
- 1.2.2.3. Pulsed field gel electrophoresis
- 1.2.2.4. Contig assembly
- 1.2.2.5. Sequence tagged sites

1.2.3. **Isolation of transcribed sequences**

- 1.2.3.1. Classical approaches
- 1.2.3.2. Exon trapping
- 1.2.3.3. cDNA selection
- 1.2.3.4. cDNA sequencing
- 1.2.3.5. Genomic DNA sequencing & computer analysis
- 1.2.3.6. Alternative strategies

1.2.4. **Analysis of candidate genes**
1.3. Positional cloning of the TSC gene(s) 52

1.3.1. Linkage analysis 52

1.3.2. Identification of TSC2 55

1.3.3. Characterisation of TSC2 55

1.3.4. Mapping human chromosome 9q34 57
   1.3.4.1. Genetic mapping 58
   1.3.4.2. Comparative mapping 58
   1.3.4.3. Physical mapping 61

1.3.5. Isolation of transcribed sequences from 9q34 62

1.4. Aims of the project 62

Chapter 2 65

2. Materials and Methods 66

2.1. Materials 66
   2.1.1. cDNA libraries 66

   2.1.2. Cell culture media 66
      2.1.2.1. Bacterial cell culture 66
      2.1.2.2. Mammalian cell culture 67

   2.1.3. Cell lines 67
      2.1.3.1. Lymphoblastoid cell lines 67
      2.1.3.2. COS-7 cell line 67

   2.1.4. Chemicals 68

   2.1.5. Cosmids 68

   2.1.6. Enzymes 70

   2.1.7. Radiation reduced hybrids 70

   2.1.8. Standard solutions and buffers 70

   2.1.9. Zoo blots 71
2.2. Methods

2.2.1. Standard DNA techniques

2.2.1.1. Restriction endonuclease digestion

2.2.1.2. Agarose gel electrophoresis

2.2.1.3. Southern blotting

2.2.1.4. Radiolabelling DNA probes

2.2.1.4.a. Random oligonucleotide priming

2.2.1.4.b. End labelling

2.2.1.5. Hybridisation of DNA probes

2.2.1.6. Post-hybridisation washes and signal detection

2.2.1.7. Mini preparation of cosmid DNA

2.2.1.8. Maxi preparation of cosmid/plasmid DNA

2.2.1.9. DNA extraction from cultured cells

2.2.1.10. Polymerase chain reaction (PCR)

2.2.2. Mammalian cell culture

2.2.2.1. Lymphoblastoid cell lines

2.2.2.2. COS-7 cell line

2.2.3. Cosmid library construction

2.2.3.1. DNA preparation

2.2.3.2. Packaging reaction

2.2.3.3. Replica plating

2.2.3.4. Screening for cosmids containing human DNA

2.2.4. Fluorescence in situ hybridisation

2.2.4.1. Metaphase preparations

2.2.4.1.a. Lymphocytes

2.2.4.1.b. Lymphoblastoid cell lines

2.2.4.1.c. Hybrid cell lines

2.2.4.2. Slide preparation & prehybridisation treatment

2.2.4.3. Biotin-labelling the probe by nick translation

2.2.4.4. Preparation of labelled probe

2.2.4.4.a. Cosmid probes

2.2.4.4.b. Total human DNA probe

2.2.4.5. Hybridisation

2.2.4.5.a. Cosmid probes

2.2.4.5.b. Total human DNA probe

2.2.4.6. Post-hybridisation washes

2.2.4.6.a. Cosmid probes

2.2.4.6.b. Total human DNA probe

2.2.4.7. Detection of hybridisation

2.2.4.8. Microscopy

2.2.5. Exon Amplification

2.2.5.1. Subcloning genomic DNA
2.2.5.1. a. Dephosphorylation of vector DNA
2.2.5.1. b. Digestion of genomic DNA
2.2.5.1. c. Ligation reaction
2.2.5.2. Preparation of competent cells
2.2.5.3. Transformation
2.2.5.4. Isolation of DNA
2.2.5.5. Transfection into COS-7 cells
2.2.5.6. Isolation of RNA
2.2.5.7. Reverse transcriptase PCR
2.2.5.7. a. Using pSPL1
2.2.5.7. a. i. cDNA synthesis
2.2.5.7. a. ii. PCR
2.2.5.7. b. Using pSPL3
2.2.5.7. b. i. cDNA synthesis
2.2.5.7. b. ii. Primary PCR
2.2.5.7. b. iii. BstXI enzyme digestion
2.2.5.7. b. iv. Secondary PCR
2.2.5.8. Cloning and analysis of PCR products
2.2.5.8. a. Using pSPL1
2.2.5.8. a. i. Identifying and purifying exons
2.2.5.8. a. ii. Secondary PCR
2.2.5.8. a. iii. Verification of a true exon
2.2.5.8. a. iv. Cloning
2.2.5.8. b. Using pSPL3
2.2.5.8. b. i. Identifying and purifying exons
2.2.5.8. b. ii. Cloning
2.2.5.8. b. iii. Verification of a true exon

2.2.6. Sequencing
2.2.6.1. Sequencing reactions
2.2.6.1. a. Using pSPL1
2.2.6.1. b. Using pSPL3
2.2.6.2. Polyacrylamide gel electrophoresis
2.2.6.3. Signal detection

2.2.7. Screening cDNA libraries
2.2.7.1. Lambda ZAP®II cDNA library (Stratagene)
2.2.7.2. Gridded cDNA filters (ICRF)

Chapter 3

3. Results

3.1. Physical mapping human chromosome 9q34
3.1.1. Characterisation of radiation hybrids
3.1.2. Generation of radiation hybrid cosmid libraries 109

3.1.3. Isolation of cosmids using hybrid Alu-PCR 114

3.2. Generation of a FISH map for distal chromosome 9q 117

3.3. Isolation of transcribed sequences from 9q34 130

3.3.1. Exon amplification using pSPL1 130
   3.3.1.1. Hybridisation analysis 135
   3.3.1.2. Chimerism 138
   3.3.1.3. Sequence analysis 140

3.3.2. Exon amplification using pSPL3 149
   3.3.2.1. Sequence analysis 171
      3.3.2.1.a. Contig SURF 175
      3.3.2.1.b. Contig D9S10 181
      3.3.2.1.c. Contig D9S114 183
      3.3.2.1.d. Contig 37 187
      3.3.2.1.e. Contig 166 188
      3.3.2.1.f. Contig L-4DD1 188
   3.3.2.2. Hybridisation analysis 189
      3.3.2.2.a. Contig SURF 191
      3.3.2.2.b. Contig D9S114 194
      3.3.2.2.c. Contig 37 198

3.3.3. Isolation of cDNAs from contig SURF, D9S114 & 37 200
   3.3.3.1. Screening a gridded cDNA library 200
      3.3.3.1.a. Hybridisation and sequence analysis 200
   3.3.3.2. Screening lambda ZAP®II cDNA library 202
      3.3.3.2.a. Hybridisation analysis 205
      3.3.3.2.b. Sequence analysis 217
         3.3.3.2.b.i. Clone 10 217
         3.3.3.2.b.ii. Clone 5 218
         3.3.3.2.b.iii. Clone 7 219
         3.3.3.2.b.iv. Clone 20 219
      3.3.3.3.c. FISH analysis 219
         3.3.3.3.c.i. Clone 5 219
         3.3.3.3.c.ii Clone 20 220

Chapter 4 222

4. Discussion 223

4.1. Positional cloning the TSC1 gene 223
4.1.1. Localisation of the TSC1 candidate region

4.1.1.1. Haplotype analysis
4.1.1.2. Loss of heterozygosity
4.1.1.3. DNA rearrangement

4.1.2. Physical mapping human chromosome 9q34

4.1.2.1. Isolation and mapping of clones from radiation hybrids
4.1.2.2. Mapping clones encoding genes
4.1.2.3. Mapping the TSC1 candidate region

4.1.3. Isolation of transcribed sequences from 9q34

4.1.3.1. Exon amplification using pSPL1
4.1.3.2. Exon amplification using pSPL3

4.1.4. Characterisation of transcribed sequences and their candidates for TSC1

4.1.4.1. Contig SURF
4.1.4.1.a. Surfeit genes
4.1.4.2. Contig D9S10
4.1.4.2.a. VAV2
4.1.4.3. Contig D9S114
4.1.4.3.a. D9s114 Bam A, D9S114 Bam C & D9S114 Pst B
4.1.4.3.a.i. Characterisation of cDNA clones
4.1.4.3.b. D9S114 Bam B
4.1.4.3.b.i. Characterisation of a cDNA clone
4.1.4.4. Contig 37
4.1.4.4.a. Ctg37 Bam A
4.1.4.4.a.i. Characterisation of a cDNA clone

4.1.5. Exon amplification and alternative approaches for isolating transcribed sequences

4.1.6. Recognised genes on chromosome 9q34 which may be candidates for TSC1

4.2. Characterisation of TSC2 and possible predictions for TSC1

4.3. Other phakomatoses

4.4. TSC3?

4.5. Animal models for tuberous sclerosis
4.6. Positional cloning and alternative approaches to gene identification

4.7. Summary and future work

Bibliography

Publications arising from this thesis

Appendix

1. Materials
   1.1. Standard solutions and buffers
   1.2. Bacterial cell culture
   1.3. Mammalian cell culture
   1.4. Solutions for FISH

2. Primer sequences

3. Exon sequences
   3.1. Exon amplification using pSPL1
   3.2. Exon amplification using pSPL3

4. cDNA sequences
List of figures

1.1 Schematic outline of the multistep process of positional cloning 27
1.2 Overview of the procedure of exon amplification 42
1.3 Schematic representation of exon amplification 43
1.4 Schematic representation of exon amplification using pSPL3 44
1.5 The composite genetic map of human chromosome 9q31-qter 59
1.6 Human chromosome 9 aligned with the homologous regions of mouse chromosomes 2, 4, 13 and 19 60
1.7 Physical map of human chromosome 9q34 64
3.1 FISH of biotinylated total human DNA onto radiation hybrid metaphase chromosomes 107
3.2 Deletion map for panel of radiation hybrids 110
3.3 FISH mapping cosmids generated from 17B and 20A radiation hybrid libraries, onto normal metaphase chromosomes 112
3.4 FISH mapping cosmids isolated from genomic libraries using Alu-PCR probes, onto normal human metaphase chromosomes 114
3.5 FISH of pooled Alu-PCR derived cosmids onto normal human metaphase chromosomes 116
3.6 FISH mapping cosmids encoding genes onto normal human metaphase chromosomes 118
3.7 Schematic representation of the four intervals used to position cosmids by FISH on human chromosome 9q34 119
3.8 Cosmid clones mapping to interval 1, proximal to SD-1 cell line translocation breakpoint 121
3.9 Cosmid clone mapping to interval 2, distal to SD-1 and proximal to 9T12 cell line translocation breakpoint 122
3.10 Cosmid clone mapping to interval 3, distal to 9T12 and proximal to 9T01 cell line translocation breakpoint 123
3.11 Cosmid clones mapping to interval 4, distal to 9T01 cell line translocation breakpoint 124
3.12 FISH map of cosmids localised to distal chromosome 9q 127
3.13 FISH of pooled Alu-PCR derived cosmids onto metaphase chromosomes from 9T12 translocation cell line  
3.14 RT-PCR analysis using vector pSPL1  
3.15 Vector to vector splicing  
3.16 Vector to genomic splicing event that gives rise to an amplified exon  
3.17 Hybridisation of putative exons to their corresponding genomic clones  
3.18 Hybridisation of putative exons to zooblots  
3.19 Sequence analysis of chimeric cKW1B PCR product  
3.20 Cryptic splicing event  
3.21 Sequence similarity between a product from cosmid 17B2 and the human Alu-Sx subfamily consensus sequence  
3.21 Sequence analysis of 17BE PCR product (exon e)  
3.22 Sequence similarity between the product from cosmid 17BE (exon e) and VAV  
3.23 Hybridisation of exon e to cDNA clones  
3.24 Contig number 54 which contains the Surfeit gene cluster  
3.25A Contig number 41 which contains marker D9S10  
3.25B Region of contig number 41 that was used for exon amplification  
3.26 Contig number 80 which contains marker D9S114  
3.27 Contig number 37  
3.28 Contig number 166  
3.29 Contig number 10 which contains the left hand end of YAC 4DD1 (L-4DD1)  
3.30 Shot-gun cloned cosmids into pSPL3 vector  
3.31 Primers used for exon amplification with vector pSPL3  
3.32 The effect of BstXI digestion on RT-PCR products  
3.33 RT-PCR analysis and cloned products from contig SURF
3.34 RT-PCR analysis and cloned products from contig D9S10
3.35 RT-PCR analysis and cloned products from contig D9S114
3.36 RT-PCR analysis and cloned products from contig 37
3.37 RT-PCR analysis and cloned products from contig 166
3.38 RT-PCR analysis and cloned products from contig L-4DD1
3.39 Organisation of overlapping exon sequences within SURF Bam A
3.40 Sequence similarity between SURF Bam B and ribosomal protein L7a
3.41 Organisation of the human and mouse Surfeit gene cluster
3.42 Possible organisation of SURF Bam A, SURF Bam B, SURF Pst B and SURF Pst D within the human Surfeit gene cluster
3.43 Artefact amplification product composed of entirely HIV-1 sequence
3.44 Sequence similarity between D9S114 Bam A and a yeast hypothetical guanine nucleotide binding protein
3.45 Sequence similarity between D9S114 Bam A and the human set gene
3.46 Hybridisation of putative exons from the SURF contig to their corresponding genomic clones
3.47 Organisation of SURF Bam A, SURF Bam B and SURF Bam C within their corresponding genomic clones and orientation of the Surfeit gene cluster
3.48 Hybridisation of putative exons from the SURF contig to zooblots
3.49 Hybridisation of putative exons from the D9S114 contig to their corresponding genomic clones
3.50 Hybridisation of putative exons from the D9S114 contig to zooblots
3.51 Hybridisation of putative exon from contig 37 (Ctg Bam A) to the corresponding genomic clone and a zooblot
3.52 Hybridisation of pooled exons from contigs SURF, D9S114 and 37 to a gridded human foetal brain cDNA library
3.53 Hybridisation of exons from the SURF contig to cDNA clones
3.54 Hybridisation of pooled exons from contigs D9S114 and 37 to a
human foetal brain cDNA library

3.55 cDNA clones digested with restriction enzymes EcoRI and BgII

3.56 Hybridisation of exons D9S114 Bam A and D9S114 Pst B onto Southern blot of cDNA clones

3.57 Hybridisation of exons D9S114 Bam B and Ctg37 Bam A onto Southern blot of cDNA clones

3.58 Hybridisation of polyT onto Southern blot of cDNA clones

3.59 BamHI restriction enzyme digest of cosmids that take a minimum tiling path across contigs D9S10 and D9S114

3.60 Hybridisation of cDNA clone 5 to Southern blot of cosmids that take a minimum tiling path across contigs D9S10 and D9S114

3.61 Hybridisation of cDNA clone 7 to Southern blot of cosmids that take a minimum tiling path across contigs D9S10 and D9S114

3.62 Hybridisation of cDNA clone 10 to Southern blot of cosmids that take a minimum tiling path across contigs D9S10 and D9S114

3.63 FISH mapping of cDNA clones 5 and 20 onto normal human metaphase chromosomes

4.1 Critical recombination events within TSC pedigrees that have excluded linkage to TSC2 and positioned the TSC1 gene

4.2 Deletion map of 17B, 19B and 20A radiation hybrids

4.3 Summary of transcribed sequences isolated using pSPL1

4.4 Summary of transcribed sequences isolated using pSPL3

4.5 Genetic mapping of exon sequences on mouse chromosome 2 compared to human chromosome 9q34

4.6 Structural domains within VAV proto-oncogene
<table>
<thead>
<tr>
<th>Table</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Diagnostic features of tuberous sclerosis</td>
<td>5</td>
</tr>
<tr>
<td>2.1</td>
<td>Karyotypes of translocation cell lines used for FISH</td>
<td>67</td>
</tr>
<tr>
<td>2.2</td>
<td>Cosmids isolated from radiation hybrids 17B and 20A</td>
<td>69</td>
</tr>
<tr>
<td>3.1</td>
<td>Marker retention analysis of radiation hybrids</td>
<td>104</td>
</tr>
<tr>
<td>3.2</td>
<td>FISH results using biotinylated total human DNA as probe onto radiation hybrid metaphase chromosomes</td>
<td>105</td>
</tr>
<tr>
<td>3.3</td>
<td>Summary of FISH and marker analysis for radiation hybrids</td>
<td>108</td>
</tr>
<tr>
<td>3.4</td>
<td>Eighteen genes mapped by FISH with their clone name and previously reported map position</td>
<td>125</td>
</tr>
<tr>
<td>3.5</td>
<td>Markers mapped by FISH with their clone names and previously reported map position</td>
<td>126</td>
</tr>
<tr>
<td>3.6</td>
<td>Summary results of exons isolated using pSPL1</td>
<td>142</td>
</tr>
<tr>
<td>3.7</td>
<td>Cosmid contigs used for exon amplification using pSPL3</td>
<td>149</td>
</tr>
<tr>
<td>3.8</td>
<td>Numbers of exons cloned from the six contigs screened</td>
<td>170</td>
</tr>
<tr>
<td>3.9</td>
<td>Forty-eight exons isolated from 9q34 with their approximate length, sequenced length and database similarities</td>
<td>172</td>
</tr>
<tr>
<td>3.10</td>
<td>Characterisation of 20 cDNA clones isolated from the Stratagene library</td>
<td>212</td>
</tr>
<tr>
<td>4.1</td>
<td>Summary of some of the advantages and disadvantages of exon amplification</td>
<td>262</td>
</tr>
<tr>
<td>4.2</td>
<td>Summary of some of the advantages and disadvantages of cDNA selection</td>
<td>263</td>
</tr>
</tbody>
</table>
Abbreviations

A adenine or adenosine
BamHI restriction enzyme cutting at GGATCC
BglII restriction enzyme cutting at AGATCT
bp base-pairs
BstXI restriction enzyme cutting at CCANNNNNTGG
C cytosine or cytidine
cDNA complementary DNA
cM centimorgan
CT cranial tomography
DAPI 4, 6-diamidino-2-phenylindole
dATP deoxy-adenosine triphosphate
dCTP deoxy-cytosine triphosphate
ddH_{2}O deionised and distilled water
DEPC diethyl pyrocarbonate
dGTP deoxy-guanosine triphosphate
DNA deoxyribonucleic acid
dNTP (unspecified) deoxy-nucleoside triphosphate
DMEM Dulbecco’s modified Eagles medium
DMSO dimethylsulphoxide
dTTP deoxythymidine triphosphate
dTTP deoxythymidine triphosphate
DTT dithiothreitol
dUMP deoxy-uridine monophosphate
EcoRI restriction enzyme cutting at GAATTC
EDTA ethylenediaminetetraacetic acid
EST expressed sequence tag
FCS foetal calf serum
FISH fluorescence in situ hybridisation
g grams
G guanine or guanosine
HAT hypoxanthine aminopterin thymidine
**HindIII** restriction enzyme cutting at AAGCTT

HPRT hypoxanthine phosphoribosyl transferase

kb kilobase-pairs

*KpnI* restriction enzyme cutting at GGTACC

L litre

LacZ β-galactosidase gene

M moles

*MboI* restriction enzyme cutting at GATC

Mb megabase-pairs

MRI magnetic resonance imaging

*NdeI* restriction enzyme cutting at CATATG

PBS phosphate buffered saline

PCR polymerase chain reaction

PFGE pulsed field gel electrophoresis

pfu plaque forming units

pM picomolar

pmol picomole

pSPL1 First splicing vector used for exon amplification

pSPL3 Second splicing vector used for exon amplification

*PstI* restriction enzyme cutting at CTGCAG

RFLP restriction fragment length polymorphism

RNA ribonucleic acid

RNase ribonuclease

*Sall* restriction enzyme cutting at GTCGAC

*Sau3AI* restriction enzyme cutting at GATC

SDS sodiumdodecyl sulphate

SSC saline sodium citrate

STS sequence tagged site

SV40 Simian virus 40

T thymidine or thymine

Tris tris(hydroxymethyl)aminomethane

TSC Tuberous Sclerosis Complex

u units
U uracil or uridine
UDG uracil DNA glycosylase
μg microgram
μl microlitre
μM micromolar
v/v volume for volume
w/v weight for volume
YAC yeast artificial chromosome
Xgal 5-bromo-4-chloro-3-indolyl-β-D-galactoside
Contributors

Mari-Wyn Burley
Helped with screening the Stratagene cDNA library.

Jude Fitzgibbon
Generated the radiation hybrids, characterised them by marker analysis and produced a deletion map. Constructed the 20A hybrid library, isolated clones by using hybrid Alu-PCR products to screen a total human genomic library and isolated RXR cosmids.

Margaret Fox
Mapped cDNA clones by FISH.

Nick Hornigold
Assembled cosmid contigs and helped with the selection of clones for FISH.

Darren Griffin
Characterised a large number of the radiation hybrids by FISH.

Joseph Nahmias
Isolated the majority of the cosmids mentioned in this thesis and helped to assemble them into contigs. Also helped to characterise the radiation hybrids by marker analysis and construct the deletion map.

Alison Pilz
Genetically mapped exon sequences in the mouse genome and placed them on a comparative map of human chromosome 9q34.

Lynne West
Helped to map some of the cosmid contigs by FISH.

Janet Young
Helped to sequence some of the exons and sequenced all of the cDNA clones.
Acknowledgements

I would like to thank my supervisors Prof. Sue Povey and Dr. Jonathan Wolfe and all members of the chromosome 9 group especially Jude, Joseph, Ali, Janet, Mari-Wyn and Nick. In addition to contributing to my work, they have provided encouragement and inspiration throughout.

I am also grateful to Dr. Joy Delhanty for letting me work in her lab (208) and of course all of my friends at The Galton Laboratory. Special thanks go to Simon, Dagan, Darren and Richa whose pranks have been most entertaining and to Margaret whose technical expertise has been invaluable.

Finally, I would like to thank my Mum and Dad and John as it has been their patience, understanding and encouragement that has kept me going.
Chapter 1

Introduction
Chapter 1

1. Introduction

1.1. Clinical aspects of tuberous sclerosis (TSC)

Tuberous sclerosis is a disease best recognised for its combination of neurological and dermatological features. It is characterised by hamartomas which are benign overgrowths predominantly of a cell type or tissue type that normally occurs in the organ. Although these lesions mainly develop in the skin, brain, heart, eye and kidney, they can arise in virtually any organ in the body and hence the term tuberous sclerosis complex (TSC) is given.

The major neurological features of tuberous sclerosis include mental retardation and seizures which affect approximately 38% and 62% of patients respectively (Webb et al., 1991). Webb et al. (1991) found that mental retardation always occurred in association with seizures and if seizures did not develop in the first few years of life subsequent mental retardation was rare. An early age of seizure onset and more severe mental retardation was found to correlate with larger and more numerous cortical tubers (Gomez, 1991). Additional neurological features detected by cranial tomography (CT) or magnetic resonance imaging (MRI) include subependymal glial nodules and giant cell astrocytomas. These are hamartomas with identical histological features (Gomez, 1991) and they are present in approximately 90% and 5-10% of TSC individuals respectively (Fryer et al., 1990).
Dermatological features of the disease include angiofibromas, ungual fibromas, forehead fibrous plaques, shagreen patches and hypomelanotic macules. Facial angiofibromas develop in a characteristic “butterfly” shape across the nose and cheeks, usually between the third and fifth year of life and rarely after puberty (Gomez, 1991). These lesions are hamartomas and are present in between 40-90% of TSC patients (Fryer et al., 1990). Ungual fibromas are hamartomas which develop from the nails of hands or feet in between 15-50% of patients (Fryer et al., 1990) and they do not usually appear until the second decade of life (Gomez, 1991). Shagreen patches are thickened areas of discoloured skin which may be present at birth or develop throughout life in approximately 25% of patients (Fryer et al., 1990). Forehead fibrous plaques are also present in approximately 25% of patients (Fryer et al., 1990) but they are usually seen at birth. Finally hypomelanotic macules are white spots which are clearly visible in ultraviolet light and occur in between 80-90% of TSC patients (Fryer et al., 1990).

Other clinical symptoms include retinal phakomas which are hamartomas found in approximately 50% of patients (Robertson, 1991); cardiac rhabdomyomas which are hamartomas present in approximately 55% of patients (Fryer et al., 1990) and can be detected in foetal life as early as 22 weeks’ gestation (Webb et al., 1993) and renal angiomyolipomas or cysts which are present in 60% and 20% of TSC patients respectively (Fryer et al., 1990). These two types of renal abnormalities can occur separately or together and they are frequently multiple and bilateral (Bernstein et al., 1991). Renal cysts are often small and scattered and may be present without causing any symptoms (Bernstein et al., 1991). However they may also be numerous and large at birth and associated with severe early onset hypertension and eventual renal failure.
Renal carcinoma has also been reported in patients with TSC (Stapleton et al., 1980). Renal carcinoma has also been reported in patients with TSC (Weinblatt et al., 1987) but it is a rare occurrence (Fryer et al., 1991).

The variability in expression of tuberous sclerosis is enormous both within and between families and this causes particular difficulties in diagnosis. Diagnosis is based on the clinical criteria defined by Gomez, (1988 and 1991) and the features of the disease have been summarised in table 1.1. A single definite criterion is sufficient for diagnosis but more than one of the presumptive or suspect criteria are required. The probability is greatly increased if the person under investigation has an affected first degree relative. As the phenotype is so varied rigorous investigation is necessary for exclusion of the disease. However accuracy and early detection has been dramatically improved by the advancement of modern imaging techniques such as CT, MRI and ultrasound.

Tuberous sclerosis has an autosomal dominant pattern of inheritance (Gunther and Penrose, 1935) and an estimated incidence at birth of 1 in 5,800 (Osborne et al., 1991). A number of reports have found higher estimates of prevalence in children compared to adults (Hunt and Lindenbaum, 1984; Sampson et al., 1989b). These may reflect a less accurate diagnosis in the older groups but also an increased early mortality associated with the disease. It is possible that disease gene carriers with only minor skin manifestations do not come to medical attention and therefore true prevalence could be higher. The mutation rate in TSC is high as approximately 60% of affected individuals are sporadic cases which represent new mutations (Sampson et al., 1989b). Penetrance of the gene is high but the great variability in expression causes problems in determining if it is complete (Webb and Osborne, 1991). True non-penetrance can only be considered if the obligate gene carrier is examined using the modern methods of
Table 1.1

Diagnostic features of Tuberous Sclerosis

<table>
<thead>
<tr>
<th>Organ</th>
<th>Definite Criteria</th>
<th>Presumptive Criteria</th>
<th>Suspect Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>CNS</td>
<td>Cortical tuber</td>
<td></td>
<td>Infantile spasms</td>
</tr>
<tr>
<td></td>
<td>Subependymal nodules</td>
<td></td>
<td>Other seizures</td>
</tr>
<tr>
<td></td>
<td>Giant cell astrocytoma</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Retina</td>
<td>Multiple phakomas</td>
<td>Single phakoma</td>
<td></td>
</tr>
<tr>
<td>Skin</td>
<td>Facial angiofibromas</td>
<td>Confetti-like spots</td>
<td>Hypomelanotic macules</td>
</tr>
<tr>
<td></td>
<td>Ungual fibroma</td>
<td>Shagreen patch</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Fibrous forehead plaque</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kidneys</td>
<td>Multiple angiomyolipomas</td>
<td>Single angiomyolipoma</td>
<td>Cysts</td>
</tr>
<tr>
<td>Heart</td>
<td>Multiple rhabdomyomas</td>
<td>Single rhabdomyoma</td>
<td></td>
</tr>
<tr>
<td>Lungs</td>
<td>Lymphangiomyomatosis</td>
<td>Honeycomb image</td>
<td></td>
</tr>
<tr>
<td>Teeth</td>
<td>Enamel pits</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gingiva</td>
<td></td>
<td>Fibromas</td>
<td></td>
</tr>
<tr>
<td>Rectum</td>
<td>Polyps</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thyroid</td>
<td></td>
<td>Adenoma</td>
<td></td>
</tr>
<tr>
<td>Adrenal</td>
<td></td>
<td>Angiomyolipoma</td>
<td></td>
</tr>
<tr>
<td>Gonads</td>
<td></td>
<td>Angiomyolipoma</td>
<td></td>
</tr>
<tr>
<td>Liver</td>
<td></td>
<td>Angiomyolipoma</td>
<td></td>
</tr>
<tr>
<td>Bones</td>
<td></td>
<td>Cysts Osteom thickening</td>
<td></td>
</tr>
</tbody>
</table>

Table taken from Gomez, 1991 and Fryer, 1991
imaging and therefore previously reported cases (Baraister and Patton, 1985; Gunther and Penrose, 1935) can only be supported if the individuals are re-examined. Although the reports of Wilson and Carter (1978) and Lowry et al. (1979) were prior to these rigorous examinations, they do represent a rather different problem. Both cases are of non-consanguineous parents giving birth to a second affected child despite being clinically normal themselves. This may be due to non-penetrance or variable disease expression. However it could also be due to germ-line mosaicism in one of the parents or even non paternity. To explain these events, greater knowledge of the molecular basis of tuberous sclerosis is required. The etiological mechanism leading to the disease is poorly understood and therefore a positional cloning approach was taken to identify the TSC gene(s).

1.2. Positional Cloning

Positional cloning is a multistep process (see figure 1.1) which is initiated by mapping a disease gene to a particular region of a chromosome. This may be possible by the identification of structural abnormalities, genetic linkage analysis or comparative mapping. Gene localisation is then followed by molecular analysis of the target region and may include finer genetic mapping, linkage disequilibrium studies, physical mapping, identification of transcribed sequences and ultimately searching for mutations in candidate genes. The techniques which had been used or considered in identifying the cause of TSC are outlined below.
Figure 1.1

Schematic outline of the multistep process of positional cloning

1. Disease or mutant phenotype
   - Structural abnormality
   - Genetic linkage analysis
   - Comparative mapping

2. Chromosome localisation

3. Finer genetic mapping
   - linkage disequilibrium studies
   - physical mapping

4. Contig construction
   - (cosmids, YACs, P1 clones)

Gene isolation

Traditional approaches
1. Conservation on zooblots
2. CpG islands
3. cDNA library screening with YAC, cosmids

More recent approaches
1. exon amplification
2. cDNA selection
3. genomic sequencing and computer analysis
4. regionally mapped candidate genes and ESTs

5. Mutation analysis
6. Analysis of function

Taken from Monaco, 1994 (see text for details)
1.2.1. Disease gene localisation

1.2.1.1. Structural abnormalities

Many of the genes identified by positional cloning have been localised by structural abnormalities that are visible cytogenetically. Examples of such diseases include chronic granulomatous disease (Royer-Pokora et al., 1986); Duchenne muscular dystrophy (Monaco et al., 1986); retinoblastoma (Friend et al., 1986) and neurofibromatosis 1 (Wallace et al., 1990). In some cases structural aberrations have also been detected by Southern blotting. The diseases caused by tri-nucleotide repeat expansions such as fragile X (Verkerk et al., 1991); X-linked spinal and bulbar muscular atrophy (La Spada et al., 1991) and myotonic dystrophy (Harley et al., 1992) are good examples.

1.2.1.2. Linkage analysis

Disease genes not so readily marked by a structural abnormality can be localised by meiotic linkage mapping, provided that pedigrees are available where the disease gene is segregating. The technique uses the principle that if two loci are closely linked on a chromosome it is less likely that they will become separated by meiotic recombination thus they will be inherited together. The likelihood of linkage between a particular marker and disease gene is calculated using mathematical models. It is represented as a lod score, a logarithmic function of the odds of linkage versus non-linkage at a set recombination fraction (θ). The recombination fraction is defined as the probability with which a recombination event occurs. If it is equal to zero then adjacent loci appear completely linked, whereas if it is equal to one half they segregate independently. It can be used to give an estimate of the genetic distance between two loci which is measured
in so-called crossover units or centimorgans (cM), with 1cM equivalent to the distance of 1% meiotic recombination.

High resolution genetic maps greatly accelerate the localisation of monogenic disease genes by linkage analysis. They also facilitate the localisation of polygenic disease by allowing an exclusion mapping strategy. However their construction is largely dependant on the isolation of DNA polymorphisms for use as markers. A large number of markers are restriction fragment length polymorphisms (RFLPs) detected by Southern blotting (Southern, 1975). Unfortunately these are laborious to detect and are often uninformative for linkage analysis. The discovery of simple tandem repeat or microsatellite marker loci which are highly polymorphic and can be detected by PCR (Weber and May, 1989) has led to the construction of dense genetic maps for the human genome (Weissenbach et al., 1992; Gyapay et al., 1994) and even for other mammals such as mouse (Dietrich et al., 1994), rat (Serikawa et al., 1992) and pig (Johansson et al., 1995). In humans, the Généthon genetic linkage map now consists of 2,066 (AC)_n short tandem repeats across the genome with an average spacing of 2.9 centiMorgans (cM) (Gyapay et al., 1994). Therefore once a gene has been localised to a particular chromosome, finer genetic mapping and positional cloning is possible. By analysing recombination events in affected individuals it is often possible to map a disease gene to a genetic interval of 1cM. The markers flanking the target region are then very useful reference points for building and verifying physical maps.

1.2.1.3. Linkage disequilibrium

Linkage disequilibrium has been used in positional cloning approaches for a number of disease genes including cystic fibrosis (Estivill et al., 1987), myotonic dystrophy
(Harley *et al.*, 1992), torsion dystonia (Ozelius *et al.*, 1992b; Risch *et al.*, 1995), Neurofibromatosis 1 (Jorde *et al.*, 1993) and diastrophic dysplasia (Hästbacka *et al.*, 1992). It may be considered when the disease gene in question has not been associated with a structural abnormality or when the localisation by genetic linkage analysis has a low resolution of within several million base pairs. The technique is most useful for locating disease genes that are caused by one or a few mutations and in populations such as the Ashkenazi Jews that have descended from a small founder population (Risch *et al.*, 1995). By assuming that genetic markers around the disease gene will recombine least often, a common haplotype within the vicinity of the gene on the affected chromosome may be detected which reflects the ancestral haplotype and therefore the founder mutation. Strong linkage disequilibrium with a particular extended haplotype and the disease gene may then facilitate localisation. It has recently been successful in the isolation of the diastrophic dysplasia gene by analysing the isolated Finnish population (Hästbacka *et al.*, 1994).

### 1.2.1.4. Comparative mapping

Comparative mapping is an approach for identifying conserved synteny groups between distantly related species. The technique requires the construction of high resolution genetic maps and these now exist for both the human and mouse genome (Gyapay *et al.*, 1994; Encyclopaedia of the mouse genome III, 1993). The rapid advancement in the mouse map followed the discovery that wild mice species such as *Mus spretus* can interbreed with laboratory strains (Avner *et al.*, 1988). When the two strains are crossed fertile F1 females are produced which can then be backcrossed with either *Mus spretus* or laboratory mouse parent strain.

By analysing the segregation pattern of polymorphic markers in the
progeny, genetic order can be determined. In addition, the large numbers of progeny produced in a backcross (The European backcross collaborative group, 1994; Rowe et al., 1994) allows high resolution mapping of all regions of the genome. A large number of chromosomal regions conserved by gene content and marker order have been identified between the mouse and human genome and the comparative maps constructed can then be used to position loci within these homologous regions. In addition, the conserved linkage groups can be used to map mouse mutations which may correspond to genetically determined human homologues. These localisation may facilitate the identification of the human disease gene loci and thereby provide an animal model for the disease process. However although these maps are powerful tools some caution must be taken when interpreting the results as discrepancies in loci order make occur as a result of evolutionary changes (Pilz., 1995).

1.2.2. Physical mapping

Physical maps show the position of various DNA landmarks relative to one another along the chromosome. There are many different mapping techniques available and each one is characterised by the use of different landmarks. The techniques to be covered here include mapping by fluorescence *in situ* hybridisation, somatic cell hybrids, pulsed field gel electrophoresis, contig construction and sequence tagged sites.

1.2.2.1. Fluorescence *in situ* hybridisation

Fluorescence *in situ* hybridisation (FISH) is one of the most direct approaches for localising DNA sequences on human metaphase chromosomes and enables rapid mapping of genes (Wiegant et al., 1991; Jones et al., 1993). Unique DNA cloned in plasmids (Lichter et al., 1988), cosmids (Lichter et al., 1990a) or yeast artificial
chromosomes (YACs) (Dreisen et al., 1991) can be mapped using a strategy of chromosomal in situ suppression hybridisation (CISS) (Lichter et al., 1990a). The technique uses an excess of unlabelled total human DNA or DNA enriched for repeat sequences (Cot-1 DNA) to prevent non-specific hybridisation between probe and genomic repeat sequences. Lichter et al. (1990a) used this technique to map cosmid clones onto extended (prometaphase) chromosomes and constructed a high resolution map of chromosome 11. The clones were localised by fractional length measurements and therefore banding of the chromosome was unnecessary. Cosmids have also been mapped by direct R-banding FISH and high resolution cytogenetic maps of various chromosomes constructed (Takahashi et al., 1992; Takahashi et al., 1993; Takahashi et al., 1994). As the cloned material mapped can be associated with DNA markers, FISH is a powerful tool for integrating cytogenetic, physical and genetic maps of the genome (Gingrich et al., 1993; Green et al., 1994c).

Metaphase mapping can resolve DNA probes separated by more than 1Mb (Lawrence et al., 1990). For resolution in the order of 50-1000kb, FISH has been applied to chromatin fibres of interphase cell nuclei (Lawrence et al., 1988; Lawrence et al., 1990; Trask et al., 1989; Trask et al., 1991). Nuclei are harvested in the G0/G1 stage of the cell cycle to avoid confusing doublet signals. Sequence order can be determined by comparison of the average distance between different pairs of probes (Trask et al., 1991). The advent of multiple colour FISH (Nederlof et al., 1990; Ried et al., 1992; Dauwerse et al., 1992) has increased the potential for physical DNA mapping and ordering. However the three-dimensional structure of the chromatin fibres within the nucleus can interfere with reliable orders of hybridisation signals especially when they are above 1Mb (Den Dunnen et al., 1992). Recent techniques have been developed using linear decondensed
nuclear DNA which alleviate this problem and also give greater resolution (Heng et al., 1992; Wiegant et al., 1992; Parra and Windle, 1993). Heng et al. (1992) used chromatin fibres released from nuclei to resolve sequences between 21-350kb apart. An alternative protocol used highly extended DNA loops surrounding the nuclear matrix in a halo-like structure to resolve sequences separated by 10-200kb (Wiegant et al., 1992). Resolution in the range of less than 5kb and greater than 700kb has been reported by Parra and Windle, (1993). DNA was forcibly stretched in a linear fashion to create a direct visual hybridisation (DIRVISH) DNA map. These methods of high resolution mapping allow visualisation of overlapping cosmid and YAC clones and measurement of gaps between adjacent contigs.

The technique of FISH has also been successfully applied to the evaluation of hybrid cell lines use in mapping (Pinkel et al., 1986). Biotinylated total human DNA is used as a probe against hybrid metaphase chromosomes and human material present is specifically stained. Changes in chromosome structure such as interspecies translocations are easily detected and by repeating the experiment over a period of time the stability of the human chromosomal material can also be determined.

Rapid characterisation of human DNA within interspecies hybrid cell DNA is possible using a technique known as reverse in situ hybridisation (Boyle et al., 1990). There are two approaches to this technique, one is to use total hybrid cell DNA as a probe against normal human metaphase chromosome spreads, using conditions that suppress signal from ubiquitous repetitive DNA (Pinkel et al., 1988). The second is to amplify human specific interspersed repetitive sequences (Alu and L1) within the hybrid cell DNA by
PCR and then use this as a probe on human metaphase chromosomes (Lichter et al., 1990b).

### 1.2.2.2. Radiation hybrid mapping

The ability to separate individual human chromosomes from the rest of the genome either by rodent-human somatic cell hybrids or by physical cell sorting is important in map construction as the chromosome of interest can be focused on and specific probes generated. The technique our group has employed is that of radiation hybrids.

Radiation hybrid mapping is based on the principle that the further apart two DNA markers are on a chromosome the more likely a given dose of radiation will break the DNA between them and they appear on separate chromosomal fragments. It was first described by Goss and Harris (1975), to determine the position of four genes on human chromosome Xq. Human diploid male lymphocytes were irradiated to fragment the chromosomes and then immediately fused to hamster cells. The recipient hamster cell line was deficient for the enzyme hypoxanthine phosphoribosyl transferase (HPRT) and growing the hybrid cells in HAT medium selected for human-hamster hybrid cells containing the wildtype human HPRT gene. Each hybrid clone contained a unique set of chromosomal fragments and was typed for the presence or absence of the other three human genes. By measuring the frequency with which the three genes were co-transferred with the selectable marker, their order and the relative distances between them could be estimated.

Cox et al. (1989) modified this procedure to isolate specific regions of the human genome that do not contain selectable markers. A rodent-human somatic cell hybrid
containing the human chromosome of interest was lethally irradiated with X-rays and the chromosomal fragments were rescued by fusion to a hamster HPRT deficient cell line. Growth in HAT medium selected for hybrid clones containing the wildtype hamster HPRT from the irradiated cell line. There was no positive selection for human chromosomal fragments but sixty percent of those characterised contained human material.

Panels of radiation hybrids have been generated for a number of human chromosomes using this approach and they have great potential in physical mapping (Benham et al., 1989; Goodfellow et al., 1990; Burmeister et al., 1991; Florian et al., 1991). Direct physical mapping from radiation hybrids is most reliable when a large number of hybrids are analysed. The statistical analysis of co-segregation of markers in a hybrid panel can produce lod scores similar to those used in meiotic mapping and allowed the construction of a physical map in close agreement with that obtained by pulsed field gel electrophoresis (Burmeister et al., 1991). A single hybrid is not a reliable tool for mapping but can be a valuable resource for isolation of region specific probes (Gillett et al., 1993). Specific regions of human DNA can be isolated by generation of lambda or cosmid libraries from hybrid DNA and then selection for human clones (Gusella et al., 1980; Florian et al., 1991). Alternatively, interspersed repetitive sequence PCR (IRS-PCR) (Ledbetter et al., 1990) such as Alu-PCR (Nelson et al., 1989) can be used for selective amplification of only human DNA. The products can be used as region specific probes for screening genomic libraries (Nahmias et al., 1995) or as a source for sequence tagged sites (STS) (Cole et al., 1991).
More recently, panels of radiation hybrids for the whole human genome have been constructed by using a diploid human fibroblast as the chromosome donor (Walter et al., 1994), as originally described by Goss and Harris (1975). The level of radiation used for construction of the hybrids will affect the number of breaks within the human DNA and therefore the resolution of the mapping panel, by increasing the dose of radiation the resolution will be increased. The approach will have great potential in genetic mapping as a high resolution map of a single panel of 100-200 hybrids should allow accurate localisation of an anonymous marker or gene by a single set of PCR reactions (Walter et al., 1994). However these hybrids will generally be less useful for the generation of clones resources.

1.2.2.3. **Pulsed field gel electrophoresis**

Pulsed field gel electrophoresis (PFGE) can separate DNA fragments several million base pairs in size by agarose gel electrophoresis. When used in conjunction with rare restriction enzymes to digest the genomic DNA it is a good method for constructing long-range physical maps of human chromosomes. DNA fragments between 50 and 2000kb in size are detected by hybridisation of DNA probes to Southern blots. Analysis of the hybridisation pattern allows the order of non-methylated restriction sites and the distances between them in the genomic DNA to be determined. This approach of long range restriction mapping within a target region has proved successful in gene identification (Rommens et al., 1989) because the rare endonucleases used, identify clusters of non-methylated CpG dinucleotides often present at the 5' end of genes.
1.2.2.4. Contig assembly

The generation of comprehensive ordered sets of overlapping clones (i.e. contigs) for genomic DNA provides the basis for more detailed mapping and aids the identification of new landmarks. Pioneering physical mapping projects on *Escherichia coli* (Smith *et al.*, 1987a; Daniels and Blattner, 1987); yeast (Olson *et al.*, 1986) and *Caenorhabditis elegans* (Coulson *et al.*, 1986) DNA have demonstrated the feasibility of using this approach on genomes of small to moderate complexity. Now the approach is being applied to the human genome and various strategies have been employed for map construction. Examples include fingerprint hybridisation (Zuo *et al.*, 1993; Nizetic *et al.*, 1994; Wapenaar *et al.*, 1994), restriction fingerprinting with digital analysis (Nahmias *et al.*, 1995) and sequence tagged site (STS) mapping (Green *et al.*, 1994c) using Généthon markers.

The major advantage to clone based physical maps is the immediate access to the DNA itself and the availability of chromosome specific libraries has greatly accelerated map construction (McCormick *et al.*, 1993; Graw *et al.*, 1992). Yeast artificial chromosomes (YACs) have an important role to play in mapping large genomic regions as they have the ability to isolate very large DNA segments (5x10^4 to over 1x10^6) (Burke *et al.*, 1987). Green *et al.* (1994c) generated a YAC-based physical map of human chromosome 7 by using Généthon genetic markers as a framework. However chimerism, rearrangements and deletions within YACs have proved to be a major disadvantage (Zuo *et al.*, 1993). Alternatively smaller recombinant clones such as cosmids (Nahmias *et al.*, 1995) or P1 clones (Albertsen *et al.*, 1994; Neuhausen *et al.*, 1994) have been used which are less likely to contain the same rearrangements and allow more detailed physical mapping. The integration of both YAC and cosmid
resources proved very successful in the mapping of the *C. elegans* genome (Coulson *et al.*, 1988). YACs were used to link and orientate cosmid contigs relative to one another by hybridisation of YAC DNA to gridded cosmid filters. Integration has now been successfully applied to regions of the human genome. For example, by using a combination of YACs and cosmids, greater than 90% of chromosome 21 has been cloned (Nizetic *et al.*, 1994). YAC contigs have been constructed for long-range clone continuity and have then been converted into cosmids for detailed molecular analysis (Zuo *et al.*, 1993; Jones *et al.*, 1994; Nizetic *et al.*, 1994; Wapenaar *et al.*, 1994).

1.2.2.5. *Sequence tagged sites*

Sequence tagged sites (STSs) are short regions of unique DNA which can be detected by PCR (Saiki *et al.*, 1988) and which can be used as landmarks on a physical map (Olson *et al.*, 1989). The aim of the Human Genome Mapping Project is to create an STS map for the entire genome with a mapped site every 100kb. Landmarks used in other types of physical maps will ideally be translated into STSs by sequencing a short fragment of DNA at that particular site. By recording each mapped sequence in a database available to everybody, an integrated map of the human genome will be created and it would alleviate the problem of exchanging cloned DNA material. STS-content mapping strategies (Green *et al.*, 1991) have now been applied to many regions of the human genome (Cole *et al.*, 1991; Coffey *et al.*, 1992; Green *et al.*, 1994c).

1.2.3. *Isolation of transcribed sequences*

The isolation of transcribed sequences from the human genome will significantly aid the identification of unknown genes and our understanding of the structural organisation of
the chromosomes. Coding sequences represents only a small proportion (3-5%) of the human genome and therefore it will be an enormous task to identify the 60,000-70,000 estimated genes (Fields et al., 1994). A number of strategies have been employed which have either focused on small regions of the genome to isolate specific disease genes or alternatively have applied a global approach to transcription map a whole chromosome or the whole genome.

1.2.3.1. Classical approaches

Classical approaches to identifying coding sequences include interspecies cross-hybridisation to search for evolutionary conserved genomic sequences which are then used as probes to screen cDNA libraries (Call et al., 1990; Monaco et al., 1986). However not all unique or low copy sequences are conserved and allow detection by Southern hybridisation techniques. CpG island mapping studies of genomic clones has also been employed. Undermethylated CpG dinucleotides are often associated with the 5' end of genes, particularly those which are ubiquitously expressed (Bird et al., 1986). This approach led to the identification of the cystic fibrosis gene (Rommens et al., 1989). However there are certain limitations as a considerable distance may lie between the CpG island and the transcribed sequence and also not all genes are marked by the presence of an island. Both of these techniques are labour intensive and can only be applied to small regions of the genome.

Direct methods of identifying expressed sequences include the use of complete genomic clones as probes to screen cDNA libraries under conditions that suppress hybridisation due to repeat sequences (Elvin et al., 1990; Snell et al., 1993). This technique proved successful in the identification of the type 1 neurofibromatosis gene (Wallace et al.,
Alternatively cDNA clones specifically depleted of repetitive sequences have been used to screen libraries of genomic clones (Hochgeschwender et al., 1989). It is essential that the depletion of repeats is complete for both approaches to prevent background false positive signals.

### 1.2.3.2. Exon trapping

A number of recent strategies have recovered transcribed sequences from cloned mammalian genomic DNA on the basis of selection for functional splice sites (Buckler et al., 1991; Auch and Reth, 1990; Duyk et al., 1990; Hamaguchi et al., 1992; Ozawa et al., 1993; Nehls et al., 1994) or the recovery of the last exons of genes (Nisson and Watkins, 1993; Nisson and Krizman, 1994). The main advantages to the approach is that large genomic regions of DNA can be screened and the identification of genes is independent to their pattern of expression.

The technique of exon amplification was described by Buckler et al., (1991) and has proved successful in the identification of a number of genes. These include human Xp21 glycerol kinase gene (Walker et al., 1993); neurofibromatosis type 2 tumour suppressor gene (Trofatter et al., 1993); Huntington’s disease gene (Huntington’s disease collaboration research group., 1993); other genes within the same region of chromosome 4p16 (Ambrose et al., 1993; Duyao et al., 1993) and a copper transporter gene defective in individuals with Menkes disease (Vulpe et al., 1993). The method has also been used to isolate exons from human chromosome 9q (Church et al., 1993; Church et al., 1994), chromosome 17q21 (Abel et al., 1994) and the human major histocompatibility class II region on chromosome 6 (North et al., 1993).
An overview of exon amplification is shown in figure 1.2 and schematic representation of the technique using the initial mammalian expression vector, pSPL1 is shown in figure 1.3. Genomic DNA of interest is shot-gun cloned into the BamHI cloning site within the HIV-1 tat intron, which is flanked by 5' donor and 3' acceptor splice sites of the HIV-1 tat gene. After propagation in E.coli cells, the recombinant clones are transfected into COS-7 cells. Transcription is driven from the SV40 early region promoter and transcripts are processed in vivo to remove the tat intron sequences. If the cloned genomic DNA contains an exon in the correct orientation, vector and insert splice sites pair and the exon is retained in the mature RNA message. Cytoplasmic RNA is isolated from the COS-7 cells two to three days after transfection and subjected to RNA-based PCR (RT-PCR) using beta-globin specific oligonucleotides as primers in the reaction. A constant sized amplification product is produced when the vector 5' donor and 3' acceptor splice sites pair. Any product larger than this vector only product suggests that an exon has been trapped. Further analysis is required to confirm that the PCR product is a novel exon. The product can be subcloned, sequenced or used as a hybridisation probe to screen zoolblots or cDNA libraries.

A number of artefacts were produced when using pSPL1 (Church et al., 1993; North et al., 1993) and as a result the vector was modified to give vector pSPL3 (Church et al., 1994). The schematic representation of the procedure using this vector is illustrated in figure 1.4. Genomic fragments are cloned into a multiple cloning site within the HIV-tat intron and then electroporated into COS-7 cells as before. After RNA isolation and synthesis of double stranded cDNA, the DNA is digested with the restriction enzyme BstX1. The vector contains a BstX1 site at each end of the multiple cloning site, within
Overview of the procedure of exon amplification (Buckler et al., 1991)

1. Digest and dephosphorylate pSPL1
2. Digest genomic DNA with appropriate enzyme
3. Subclone genomic DNA into pSPL1
4. Transform into E. coli
5. Isolate miniprep DNA
6. Transfect into COS-7 cells
7. Isolate RNA 48-72 hours later
8. cDNA synthesis
9. PCR
10. Characterisation of putative exons
Figure 1.3

Schematic representation of exon amplification (Buckler et al., 1991)

1) Cloning genomic DNA fragments into pSPL1

2) Transfection into COS-7 cells and isolation of cytoplasmic RNA 48-72 hours later

3) RT-PCR analysis using primers from the flanking B-globin sequence to amplify the cloned exon
Figure 1.4

Schematic representation of exon amplification using pSPL3

1. Genomic DNA cloned into multiple cloning site

2. RNA

3. Primary PCR

4. BstXI digestion step

5. Secondary PCR

SDv - vector 5' splice donor site
SAv - vector 3' splice acceptor site
MCS - multiple cloning site containing BstXI site at each end

<table>
<thead>
<tr>
<th>Represents</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>exon sequence derived from pSPL3</td>
<td></td>
</tr>
<tr>
<td>HIV-1 tat intron sequence</td>
<td></td>
</tr>
</tbody>
</table>
the HIV-1 tat intron. Therefore false positives which arise by cryptic splicing and retain this part of the vector sequence, are cleaved and removed as templates in the following PCR reaction. In addition, half BstX1 sites have been added adjacent to the vector 5' and 3' splice sites. When an exon is not trapped, these sites come together to create a full BstX1 site. Therefore after digestion these molecules are also eliminated as templates for PCR. When an exon is trapped in the mature RNA transcript the BstX1 site is not formed and therefore the molecules are amplified by PCR.

Duyk et al. (1990) shot-gun cloned genomic DNA into a retroviral shuttle vector pETV-SD, downstream of an exon-trap cassette. The cassette consists of a functional 5' splice site from the human beta-globin gene and an intervening sequence containing the E.coli beta-galactosidase gene which acts as a marker. During a retroviral lifecycle the DNA is transcribed in vivo. Transcripts derived from recombinant clones containing a functional 3' splice site are spliced to result in loss of the beta-galactosidase gene. After a second round of viral replication, COS cells are infected. The viral RNA genome is reverse transcribed and then amplified as a circular DNA episome due to the presence of the SV40 origin of replication. The episomal DNA is recovered from the COS cells, linearised and then transformed into bacterial cells. Clones are grown on plates containing antibiotic and X-gal for blue/white colour selection. Hydrolysis of X-gal by beta-galactosidase produces the characteristic blue colour indicative of Lac+ phenotype, whereas clones that do not contain a functional beta-galactosidase gene (LacZ) are white. White colonies are picked and then analysed by sequencing to confirm the presence of a true exon.
The technique was modified to use vector pETV-SD2 (Datson et al., 1994). The exon trap cassette also contains the 5' splice site of the human beta-globin gene and an internal LacZ gene, but the retroviral lifecycle has been eliminated. Genomic clones are shot-gun cloned into the exon trapping vector, propagated in E.coli and then electroporated into COS cells. Transcription is driven from the SV40 promoter. If the cloned DNA contains a functional 3' splice site in the correct orientation, processing of the mature message results in loss of the beta-galactosidase gene. Following RNA isolation, RT-PCR is performed with vector specific primers containing NotI sites. RT-PCR products are digested with NotI and circularised to give rise to a functional plasmid. Clones containing a trapped exon are selected as white colonies and have a size advantage during PCR.

The techniques described by Auch and Reth, (1990), Hamaguchi et al. (1992) and Ozawa et al. (1993) follow broadly similar principles to exon amplification (Buckler et al., 1991) but use different exon trap vectors. Auch and Reth, (1990) described vector pL531n which contained the LTR of Rous Sarcoma Virus as a strong promoter, the origin of SV40 for plasmid amplification in COS cells, a unique Kpnl cloning site and 5' splice donor site, 3' splice acceptor site and polyadenylation site of the rat preproinsulin gene. Hamaguchi et al. (1992) constructed vector pMHC2 and genomic DNA was partially digested and cloned into the BgIII site within intron 10 of the p53 gene. Vector pEXT2 was constructed by Ozawa et al. (1993), it contained the rat α2 macroglobulin gene as a trapping cassette and had the advantage of a multiple cloning site.
The technique described by Nehls et al. (1994) is unique because it uses a lambda phage vector called lambda GET. Genomic DNA between 6.5-19kb in size is cloned into the phage vector between 5’ donor and 3’ acceptor splice sites and automatically subcloned into multi-copy plasmids. RNA transcripts derived from these clones are isolated and amplified by RT-PCR. Trapped exons are detected by the size of the RT-PCR product.

1.2.3.3. cDNA selection

The strategy of cDNA selection uses genomic DNA clones to construct enriched region specific cDNA libraries. It proved to be successful in the isolation of a novel gene defective in individuals with X-linked agammaglobulinaemia (XLA) (Vetrie et al., 1993). Lovett et al. (1991) and Parimoo et al. (1991) described techniques of immobilising YAC or cosmid clones on filters and using them as a target to hybridised homologous cDNA clones from conventional amplified libraries. After extensive washing the cDNAs were eluted and amplified by PCR using flanking vector primers. The procedure was repeated to give two rounds of selection and the result was an enrichment of two thousand fold for the cDNAs of interest. Both techniques were successful and the only significant difference between them was that Parimoo et al. (1991) used a strategy of quenching repetitive sequences within the genomic DNA whereas Lovett et al. (1991) used a strategy of blocking the repeats within the cDNA.

The technique progressed to using biotin-streptavidin capture systems (Korn et al., 1992; Morgan et al., 1992; Tagle et al., 1993). Korn et al. (1992) used biotinylated cosmids from chromosome Xq28 which were repeat depleted, to hybridise to amplified cDNA sequences. The resulting hybrids were captured on streptavidin coated magnetic beads and non-specific cDNAs removed by stringent washing. The cDNAs were eluted
and then enriched by a second round of selection using fresh biotinylated cosmid DNA. Tagle et al. (1993) successfully applied the approach to many sources of cDNA and genomic clone pools in parallel.

Transcription maps for specific regions of the human genome have now been constructed using the technique of cDNA selection. Sedlacek et al. (1993b) focused on a 300kb region around the G6PD gene on chromosome Xq28 and identified 11 different genes. Gecz et al. (1993) constructed a 1Mb transcription map on chromosome Xq13.3 around the PGK1 locus and identified three new transcripts. In addition, a similar selection method has been used to generate sublibraries of conserved sequences from mouse and pig using human cosmids from Xq28 (Sedlacek et al., 1993a).

1.2.3.4. cDNA sequencing

A rapidly growing approach to isolating transcribed sequences is to partially sequence randomly selected cDNA clones and generate expressed sequence tags (ESTs) (Adams et al., 1991). The development of automated sequencing has made the technique feasible and the sequences produced are becoming useful markers in mapping the human genome (Olson et al., 1989). The advantage to these markers compared to the sequence tagged sites (STSs) more commonly used, is that they point directly to an expressed gene.

Adams et al. (1991) partially sequenced 600 randomly selected human brain cDNA clones and examined them for similarities in the GenBank nucleic acid database. A total of 377 new genes were identified, of which 48 showed significant similarities to genes from other organisms. If the gene from the other organism has already been mapped,
then the chromosomal localisation can be used to confirm or expand comparative maps. In addition, 46 of the ESTs were localised to individual human chromosomes by using PCR and somatic cell hybrid mapping panels. The aim of the project is to tag most of the human genes with an EST and create a database that will aid disease gene identification, isolation of coding sequences and physical mapping.

1.2.3.5. Genomic DNA sequencing and computer analysis

Several groups have employed sequencing strategies to analyse human genomic DNA ranging in size from 58-106kb (McCombie et al., 1992; Martin-Gallardo et al., 1992). 'Shot-gun' sequencing methods have been used which require the construction of libraries of overlapping DNA fragments. However random sequencing leads to high levels of redundancy and therefore alternative techniques of ordered shot-gun sequencing (Chen et al., 1993) and genome sample sequencing (Smith et al., 1994c) have also been described.

The development of automated sequencing has led to the production of large amounts of sequencing information. Computational methods of analysing the data are now being used to identify coding regions within the genomic DNA. These software programmes include BLASTN, BLASTX, GRAIL and FASTA. BLASTN is used to compare a nucleotide query sequence against a nucleotide sequence database and is based on the Basic Local Alignment Search Tool algorithm (BLAST) (Altschul et al., 1990). The same algorithm is used for BLASTX (Gish and States, 1993) but this translates the nucleotide query sequence into each of the three reading frames on both DNA strands and then searches for similarities to members of a protein sequence database. Statistically significant local alignments are made irrespective of their length and for
BLASTX a continuous stretch of 30-40 amino-acids is normally sufficient. BLASTX is more sensitive than BLASTN because similar peptides can be encoded by highly divergent nucleotide sequences. However an up-to-date sequence database is critical to both searches (review Altshul et al., 1994).

GRAIL uses a completely different set of principles compared to BLAST. A multiple sensor-neural network approach is used for locating protein-coding regions in human DNA sequences (Uberbacher and Mural, 1991). Coding nucleotide sequences are constrained by the relative abundance of aminoacids as well as the differential usage of synonymous codons. Various codon bias and hexamer frequency methods are combined in GRAIL and evaluated over 100 nucleotide windows of the genomic sequences.

1.2.3.6. Alternative strategies

Human transcribed sequences have been isolated from specific regions of the genome using human-rodent somatic cell hybrids (Liu et al., 1989; Corbo et al., 1990; Jones et al., 1991a). Human cDNA synthesis was primed from heterogeneous nuclear RNA by oligonucleotides derived from human specific Alu sequences (Corbo et al., 1990) or consensus 5' splice sequences (Liu et al., 1989). Jones et al. (1991a) used a technique of subtractive cDNA hybridisation to enrich for sequences expressed in one cell line and not in another. A tissue-specific extinguisher (TSE1) which encodes a regulatory subunit of protein kinase A was cloned as a result. However these approaches are limited by the low level of expression of many tissue specific genes in hybrid cell lines.

Coincidence sequence cloning is a novel method which enables the sequences that are shared between two complex and partially coincident DNA resources to be recovered.
(Brookes and Porteous, 1991). The approach combined human DNA with DNA from a mouse-human somatic cell hybrid and was able to exclusively recover the human DNA products present in the hybrid genome. The technique has recently been revised to an end ligation and coincident DNA selection (EL-CSC) procedure which could potentially be used to identify transcribed sequences in genomic DNA by mixing with cloned or uncloned cDNA resources (Brookes et al., 1994).

Another new approach to identifying transcribed sequences from specific chromosomal regions is preparative in situ hybridisation (Hozier et al., 1994). The procedure involves hybridisation of a PCR amplified cDNA library to denatured human metaphase chromosome spreads that are fixed on a slide. After hybridisation the chromosomes are stained to produce a characteristic banding pattern and then the region of interest is microdissected from the rest of the genome. This material is then amplified using PCR and cloned to produce a specific cDNA sub-library which can then be analysed further.

1.2.4. Analysis of candidate genes

Once coding sequences from the region of interest have been isolated further investigation is required to determine if they are candidate genes for the disease in question. The expression pattern of the transcript can be investigated by Northern blot analysis of RNA from different tissues. Then any quantitative or qualitative differences in expression are analysed by a comparison of RNA from patients to that of normal controls. The analysis for disease specific mutations is vital and the sporadic patients are helpful since comparison of their DNA with that of their parents allows the distinction between causative mutations and polymorphism to be made more easily. Any physical rearrangements in patient genomic DNA detected by Southern blot analysis or PFGE
using the candidate gene as a probe would suggest its involvement in the disease. An alternative approach is to search for sequence differences between patient and normal individuals. Direct sequencing the entire candidate gene would be extremely time consuming, fortunately other techniques have now been developed to scan for mutations. These include Single Strand Conformational Polymorphism (SSCP) (Orita et al., 1989) which enable molecules differing by only a single nucleotide to be detected by a change in electrophoretic mobility; Denaturing Gradient Gel Electrophoresis (DGGE) (Myers et al., 1987) which enables different alleles to be identified by comparing denaturing profiles between DNA duplexes and chemical cleavage mismatch (Cotton et al., 1988) which identifies mutations by a susceptibility to chemical cleavage at that site.

1.3. Positional cloning of the TSC gene(s)

1.3.1. Linkage analysis

Fryer et al. (1987) studied the linkage analysis of 26 polymorphic markers in 19 TSC families and found that the disease gene was linked to the ABO blood group locus on the distal long arm of chromosome 9. Confirmation of the assignment was obtained by a report of linkage between TSC and the restriction fragment length polymorphism at the Abelson oncogene locus (ABL) (Connor et al., 1987). However subsequent reports failed to substantiate linkage between TSC and ABO (Northrup et al., 1987; Renwick, 1987; Smith et al., 1987b; Kandt et al., 1989). In addition, Povey et al. (1988) reported two families with recombination between TSC and ABL. Therefore heterogeneity became a possibility, with more than one TSC locus whose mutation gives rise to the disease phenotype.
The first case of chromosomal abnormality associated with TSC was reported by Clark et al. (1988). A liveborn infant with trisomy for 11q23.3-qter was diagnosed as having TSC on the basis of cardiac rhabdomyomas, multifocal cortical tubers and subependymal giant cell astrocytomas. A careful examination of the parents showed no evidence of TSC. Therefore a TSC locus in the region of chromosome 11q23.3-qter was postulated. Linkage analysis supported the localisation of a TSC locus on chromosome 11q14-q23, as positive lod scores were generated for 5 markers from this region (Smith et al., 1990). The TSC locus was postulated in the vicinity of the tyrosinase probe (TYR) but there was an absence of linkage to a locus on chromosome 9q (Smith et al., 1990).

The second case of a chromosome rearrangement identified a TSC patient with a de-novo reciprocal translocation t(3;12)(p26.3;q23.3) and a third TSC locus was suggested on chromosome 12q22-q24 (Fahsold et al., 1991). However this was never independently confirmed (Povey et al., 1992).

Genetic heterogeneity was suggested by several authors (Sampson et al., 1989a; Haines et al., 1989; Smith et al., 1989; Janssen et al., 1990; Kandt et al., 1991) and led to an international TSC collaborative study (Haines et al., 1991a). It was concluded that approximately 40% of families were linked to a locus on chromosome 9q between ORM and MCT136. Analysis of the families which did not show linkage to 9q suggested a TSC locus on chromosome 11q but it was not well localised. A few families remained unlinked to either 9q or 11q and therefore there was a possibility of a third locus. Further analysis by Haines et al., (1991b) supported heterogeneity and a TSC locus in
the 9q32-34 region for approximately 33% of families. There was no evidence of linkage to markers on chromosome 11q22 and therefore another TSC locus in the genome was suggested. Haines et al. later showed a higher proportion of chromosome 9 linked families (50%) when unaffected individuals were removed from the linkage analysis (Povey et al., 1992). It indicated the importance of penetrance by suggesting that it is possible to carry the disease gene and yet remain unaffected and stressed that critical recombination events used to localise the TSC gene(s) should be confined to affected individuals.

Genetic heterogeneity was confirmed with one TSC locus definitely assigned to chromosome 9q but the majority of TSC families remaining unlinked (Povey et al., 1992). Therefore the linkage analysis continued with a genome search (Kandt et al., 1992). Five families were chosen that were not linked to chromosome 9q and they were investigated with DNA markers from the appropriate regions of chromosome 9,11,12 as well as other markers in the genome. One of these markers was D16S283 in the region of the gene for autosomal dominant polycystic kidney disease type 1 (PKD1). It was chosen because it is an informative marker and also because polycystic kidneys are associated with both diseases. All families gave positive lod scores with this marker and provided evidence for a second TSC locus on chromosome 16p13.3. The linkage was confirmed by a number of groups (Smith et al., 1992; Pericak-Vance et al., 1992; Povey et al., 1994a).

Following discovery of the locus on chromosome 16 the linkage data was re-evaluated and approximately 50% of TSC families were linked to a locus on chromosome 9q and approximately 50% were linked to a locus on chromosome 16p (Kwiatkowski et al.,
Although there is not any linkage data for a third locus, there are a few unexplained families that do not appear to segregate for TSC1 or TSC2 and many families where there are insufficient data to confirm linkage to either loci (Povey et al., 1994a). Therefore the possibility of another disease gene cannot be excluded.

1.3.2. Identification of TSC2

As the mutation rate for tuberous sclerosis is high, it was assumed that a proportion of sporadic patients may have large deletions which might facilitate identification of the disease gene (The European chromosome 16 tuberous sclerosis consortium, 1993). A total of 255 unrelated patients were analysed by pulsed field gel electrophoresis using a single-copy probe from the candidate region on chromosome 16. In 5 cases, smaller aberrant bands were detected which suggested that they might be caused by constitutional interstitial deletions. A long-range restriction map and contig map of the area was constructed which allowed the deletions to be precisely mapped. Probes corresponding to the deletions were used to screen human foetal brain and human kidney cDNA libraries. One of the transcripts isolated was found to be disrupted in all five deletions. Further evidence of its involvement in TSC was provided when a subclone of the cDNA was used as a hybridisation probe to screen 260 unrelated TSC patients and a further 5 mutations were detected.

1.3.3. Characterisation of TSC2

Northern blot analysis showed the cDNA had a 5.5kb transcript which was widely expressed and sequencing confirmed a coding sequence of 5474bp. The protein was
predicted to be 1784 amino acids and has been called Tuberin. Computational searches for sequence similarities at the protein level identified a small region of homology with the rap1-GTPase-activating protein (GAP) (GAP3) (Rubinfeld et al., 1991). The functional significance of this homology has not yet been elucidated. Cellular localisation and tissue distribution of normal and aberrant tuberin is currently being investigated by generating antibodies to fusion proteins and synthetic peptides (Halley et al., 1994). An excess of 500 TSC patients have now been studied by Southern blot analysis, to detect gross rearrangements such as deletions in TSC2. (Pericak-Vance et al., 1995). Unfortunately only approximately 3% of these have shown a mutation (Pericak-Vance et al., 1995).

There have been several reports of loss of heterozygosity on chromosome 16p13.3 in hamartomas from tuberous sclerosis patients (Smith et al., 1993; Green et al., 1994a) which suggest that the TSC2 gene product may function as a tumour suppressor gene. If this is the case tuberous sclerosis would follow Knudson’s two hit mutation hypothesis (Knudson, 1971). A germ-line mutation is inherited from the affected parent and a second somatic mutation in the allele inherited from the unaffected parent is required to produce the disease phenotype at the cellular level. Other examples of tumour suppressor genes include retinoblastoma; neurofibromatosis and p53 (reviewed Cowell et al., 1992). There have been suggestions that TSC1 may also act as a tumour suppressor gene because loss of heterozygosity of microsatellite markers from 9q34 has been detected in hamartomatous lesions (Carbonara et al., 1994; Green et al., 1994b).
1.3.4. Mapping human chromosome 9q34

There is significant interest in mapping human chromosome 9q because many disease genes have been assigned to the region (reviewed Povey et al., 1994b). Friedrich's ataxia has been mapped to 9q12-q21 (Chamberlain et al., 1988); Gorlin syndrome and multiple self-healing epitheliomata to 9q22.3-q31 (Farndon et al. 1992; Goudie et al., 1993); Fukuyama congenital muscular dystrophy to 9q31-q33 (Toda et al., 1993); dysautonomia to chromosome 9q31 (Blumenfeld et al., 1993) and tuberous sclerosis, nail-patella syndrome and idiopathic torsion dystonia to chromosome 9q34 (Fryer et al., 1987; Renwick and Lawler, 1955; Ozelius et al., 1989). There is also evidence that genes involved in bladder cancer and ovarian cancer are localised on chromosome 9 as loss of heterozygosity has been reported for both (Keen and Knowles, 1994; Osborne et al., 1994). Two regions of deletion have been found for bladder cancer between 9p21-p22 and 9q13-q34.1 (Keen and Knowles, 1994). A common region of deletion for ovarian tumours was between 9q31-q33 (Osborne et al., 1994).

The gene for Hereditary haemorrhagic telangiectasia type 1 was mapped by linkage to 9q33-q34 (McDonald et al., 1994) and has been recently identified (McAllister et al., 1994). It is hoped that comprehensive genetic, physical, transcription, comparative and cytogenetic maps of human chromosome 9q will accelerate the identification of the other disease genes.
1.3.4.1. Genetic mapping

A number of genetic maps for human chromosome 9 have been constructed (Lathrop et al., 1988; Ozelius et al., 1992a; Kwiatkowski et al., 1992; Wilkie et al., 1992; Henske et al., 1993; Attwood et al., 1994). A CEPH consortium linkage map comprised of 42 ordered loci was constructed as a framework for the whole of chromosome 9 (Attwood et al., 1994). A higher resolution map specific for chromosome 9q34 was constructed by mapping 16 highly informative markers in 17 Venezuelan reference families (Henske et al., 1993). Figure 1.5 shows a composite sex-averaged genetic linkage map for human chromosome 9q31-qter.

Genetic mapping in tuberous sclerosis families has produced a number of conflicting results but the consensus localisation for the TSC1 gene is between markers D9S149 and D9S114 (Kwiatkowski et al., 1993; Povey et al., 1994b). The genetic distance for this interval is approximately 3.6cM (Povey et al., 1994b).

1.3.4.2. Comparative mapping

A comprehensive comparative map between human chromosome 9 and the mouse has been constructed by interspecies backcross linkage analysis. (Pilz et al., 1994; Povey et al., 1994b; Pilz et al., 1995a). Mouse homologues of human loci known to be localised on chromosome 9 have been mapped and regions of homology with mouse chromosome 2, 4, 13 and 19 identified (figure 1.6). The order of loci in the conserved regions can be used to predict the probable order of loci in the human genome and may facilitate the identification of mouse mutants for disease genes mapping to human chromosome 9.
Figure 1.5

The composite genetic linkage map of human chromosome 9q31-qter (sex averaged)

Taken from The Third International Workshop on Chromosome 9 (Povey et al., 1994b)
Figure 1.6

Human chromosome 9 aligned with the homologous regions of mouse chromosomes 2, 4, 13 and 19

Taken from The Fourth International Workshop on Chromosome 9 (Pericak-Vance et al., 1995)
1.3.4.3. Physical mapping

Somatic cell hybrids containing specific regions of chromosome 9q are very useful resources for mapping (Povey et al., 1992) and include hybrids CJ9q (640-63a12) which contains the entire distal arm of human chromosome 9 (Jones and Kao, 1984) and DORA-24-4 which contains human 9q34 distal to the ABL oncogene (Carritt., unpublished). Radiation reduced hybrids have been generated for use as tools in mapping and also as a source of region specific DNA which can be cloned or used as a probe to screen libraries (Henske et al., 1992; Florian et al., 1991; Nahmias et al., 1995).

A flow-sorted chromosome 9 gridded cosmid library (LL09NC01"P") has been constructed by Pieter de Jong and co-workers and has been used by a number of groups (Povey et al., 1994b). Additional chromosome 9 specific libraries were prepared by Graw et al. (1992), Henske et al. (1992) and Takahashi et al. (1994) using cosmid clones and McCormick et al. (1993) using YAC clones. Many groups are using the cloned material to assemble contigs of overlapping clones across chromosome 9q (Barroso et al., 1994; Nahmias et al., 1995; Murrell et al., 1995; Zhou et al., 1995) which can be used to estimate physical distances between markers and also to isolate expressed genes.

FISH interphase mapping by Leversha et al. (1994) has provided order and distance measurements for three separate regions on 9q which include D9S151-D9S22, D9S58-D9S60 and AK1-D9S67 (Povey et al., 1994b). Physical maps have also been constructed by pulsed field gel electrophoresis (Henske et al., 1994; Harris et al., 1993).
and PCR analysis of flow-sorted chromosomes containing specific translocation breakpoints within chromosome 9 (Zhou et al., 1992).

Figure 1.7 shows the physical distances calculated between markers in the region of chromosome 9q34 using these techniques. The candidate region for TSC1 is between markers D9S149 and D9S114 and is approximately 2.5Mb. The cloned DNA material from this interval is an important resource for isolating expressed genes which would be candidates for tuberous sclerosis.

1.3.5. Isolation of transcribed sequences from 9q34
The most extensive isolation of expressed sequences from human chromosome 9 has been carried out by Church et al. (1993 and 1994) using the technique of exon amplification (Buckler et al., 1991). The aim is to create a high resolution transcription map which can be integrated with the physical map and will aid identification of the disease genes. It has been achieved for part of the consensus TSC candidate region between DBH and D9S114 on chromosome 9q34 (Murrell et al., 1994; Murrell et al., 1995). The map consisted of ordered exons, STSs and simple sequence repeats with a marker every 5-10kb. The exons were used to isolate cDNA clones which were analysed as candidates for tuberous sclerosis. Other potential candidate genes have isolated in the TSC target region by Smith et al. (1994b), Kumar et al. (1994) and Henske et al. (1995) but as yet the gene remains unidentified.

1.4. Aims of the project
The project aimed to use a positional cloning approach to characterise the TSC1 candidate region on human chromosome 9q34 and identify the TSC1 gene. Radiation
hybrids that had been previously characterised for markers retained from 9q34 were thought to be an important source of region specific DNA. My work was to continue their characterisation by FISH and then construct cosmid libraries from hybrid DNA to improve the physical map. Once cloned DNA had been isolated and mapped by FISH to 9q34, the aim was to screen for expressed sequences by exon amplification.
Figure 1.7

Physical map of human chromosome 9q34

Table:

<table>
<thead>
<tr>
<th>Markers</th>
<th>PFGE</th>
<th>FISH</th>
<th>Contig</th>
</tr>
</thead>
<tbody>
<tr>
<td>D9S125</td>
<td>650kb</td>
<td></td>
<td></td>
</tr>
<tr>
<td>D9S149</td>
<td></td>
<td>2.5Mb</td>
<td></td>
</tr>
<tr>
<td>ABO</td>
<td>450kb</td>
<td>650kb</td>
<td></td>
</tr>
<tr>
<td>DBH</td>
<td>1.8Mb</td>
<td>75kb</td>
<td></td>
</tr>
<tr>
<td>D9S122</td>
<td>610kb</td>
<td>300kb</td>
<td>110kb</td>
</tr>
<tr>
<td>D9S10</td>
<td></td>
<td>20kb</td>
<td></td>
</tr>
<tr>
<td>D9S66</td>
<td></td>
<td>305kb</td>
<td>1025kb</td>
</tr>
<tr>
<td>D9S114</td>
<td></td>
<td>1.4Mb</td>
<td>230kb</td>
</tr>
<tr>
<td>D9S23</td>
<td></td>
<td>225kb</td>
<td></td>
</tr>
<tr>
<td>D9S14</td>
<td></td>
<td>60kb</td>
<td></td>
</tr>
<tr>
<td>D9S67</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Chapter 2

Materials and Methods
Chapter 2

2. Materials and Methods

2.1. Materials

2.1.1. cDNA libraries

A human foetal brain cDNA library in the Lambda ZAP®II vector was purchased from Stratagene. Oligo (dT) and random primers were used for construction and the estimated titre of the library was $2.5 \times 10^{10}$ pfu/ml. The average insert size was 1kb and the cloning site was EcoRI.

Gridded human foetal brain cDNA filters were obtained from Imperial Cancer Research Fund (Lehrach et al., 1990). The library (number 507) was constructed by Sebastian Meier-Ewert from 17 week embryo poly(A)+ RNA and the vector used was pSPORT (BRL). Five filters with different set numbers were provided, giving a total of 100,000 clones.

2.1.2. Cell culture Media

2.1.2.1. Bacterial cell culture

Bacterial culture media was purchased from Difco laboratories, prepared using deionised distilled water and sterilised by autoclaving at 15lbs psi, 121°C for 20-30 minutes. The components for media are given in appendix 1.2.
2.1.2.2. Mammalian cell culture

Mammalian cell culture media including RPMI-1640 medium, Eagles minimal essential medium (MEM), non essential amino acids and GPS were obtained from Gibco BRL. The components for mammalian cell culture are given in appendix 1.3.

2.1.3. Cell lines

2.1.3.1. Lymphoblastoid cell lines

Three lymphoblastoid cell lines were used in this study and their karyotypes are shown in table 2.1. SD-1 has the characteristic Philadelphia translocation (Dhut et al., 1991) and was obtained from Dr. B. D. Young. Cell lines 9T01 and 9T12 were collected by Prof. M. Ferguson-Smith (Povey et al., 1992) and purchased from the European Collection of Animal Cell Cultures at Porton Down.

Table 2.1
Karyotypes of translocation cell lines used for FISH

<table>
<thead>
<tr>
<th>Cell Line</th>
<th>Karyotype</th>
</tr>
</thead>
<tbody>
<tr>
<td>SD-1</td>
<td>46 XX t (9; 22) (q34.1; q11)</td>
</tr>
<tr>
<td>9T12</td>
<td>46 XY t (9; 20) (q34.3; q11.2)</td>
</tr>
<tr>
<td>9T01</td>
<td>46 XX t (9; 15) (q34.3; q24)</td>
</tr>
</tbody>
</table>

2.1.3.1. COS-7 cell line

The COS-7 cell line is derived from CV-1, an African Green Monkey kidney cell line, that has been transformed by an origin-defective mutant of SV40 (Gluzman, 1981). The cells can therefore support the growth of recombinant SV40 viruses and allow a high
level of replication and transcription from the exon amplification vectors pSPL1 and pSPL3.

### 2.1.4. Chemicals

All chemicals were of AnalaR quality and purchased from British Drug House (BDH), Poole, Dorset unless otherwise stated. Radiochemical included redivue [α-32P] dCTP and [γ-32P] dATP were purchased from Amersham and [α-35S] dATP from NEN/Dupont.

### 2.1.5. Cosmids

The eighteen clones encoding genes are shown in table 3.4. Cosmids encoding RXRA were provided by Dr. J. Fitzgibbon, clones encoding ABL and ASS were provided by Dr. R. M. Harris and the cosmid encoding ABO and DBH was provided by Mr. N. Hornigold. The remainder were isolated by Dr. J. Nahmias from the Lawrence Livermore chromosome 9 specific cosmid library LL09NC01. This was provided by Pieter de Jong and co-workers and co-ordinate numbers are given.

The 14 cosmids containing markers are shown in table 3.5. Dr. D. Kwiatkowski provided most of these, but those containing D9S207, D9S740, D9S10, D9S114 and D9S298 were provided by Dr. G. Vergnaud, Dr. J. Armour, Dr. J. Wolfe, Dr. J. Fitzgibbon and Dr. J. Nahmias respectively.

A small number of cosmids were isolated from radiation hybrids by either generation of a library from hybrid DNA (section 3.1.2.) or by using hybrid Alu-PCR products as
probes to screen a total human genomic library (section 3.1.3.). These clones are listed in table 2.2.

Table 2.2.
Cosmids isolated from radiation hybrids 17B and 20A

<table>
<thead>
<tr>
<th>Cosmid</th>
<th>Source</th>
<th>Isolated By</th>
</tr>
</thead>
<tbody>
<tr>
<td>cKW1B*</td>
<td>17B hybrid genomic library</td>
<td>K. Woodward</td>
</tr>
<tr>
<td>cKW1*</td>
<td>17B hybrid genomic library</td>
<td>K. Woodward</td>
</tr>
<tr>
<td>cKW3</td>
<td>17B hybrid genomic library</td>
<td>K. Woodward</td>
</tr>
<tr>
<td>cKW10*</td>
<td>17B hybrid genomic library</td>
<td>K. Woodward</td>
</tr>
<tr>
<td>cJF20A1</td>
<td>20A hybrid genomic library</td>
<td>Dr. J. Fitzgibbon</td>
</tr>
<tr>
<td>cJF20A3</td>
<td>20A hybrid genomic library</td>
<td>Dr. J. Fitzgibbon</td>
</tr>
<tr>
<td>cJF20A5</td>
<td>20A hybrid genomic library</td>
<td>Dr. J. Fitzgibbon</td>
</tr>
<tr>
<td>cJF20A7</td>
<td>20A hybrid genomic library</td>
<td>Dr. J. Fitzgibbon</td>
</tr>
<tr>
<td>cJF17BA</td>
<td>17B Alu PCR to screen genomic library</td>
<td>Dr. J. Fitzgibbon</td>
</tr>
<tr>
<td>cJF17B2 (17B2)*</td>
<td>17B Alu PCR to screen genomic library</td>
<td>Dr. J. Fitzgibbon</td>
</tr>
<tr>
<td>cJF17BE (17BE)*</td>
<td>17B Alu PCR to screen genomic library</td>
<td>Dr. J. Fitzgibbon</td>
</tr>
<tr>
<td>cJF17BF</td>
<td>17B Alu PCR to screen genomic library</td>
<td>Dr. J. Fitzgibbon</td>
</tr>
</tbody>
</table>

* These cosmids were screened for transcribed sequences by exon amplification (section 3.3). Cosmids cJF17BA and cJF17B2 both contained the D9S114 polymorphism and cJF17F contained marker D9S298.

A large number of clones were isolated by using Alu-PCR products from radiation hybrids as to screen the Lawrence Livermore chromosome 9 specific cosmid library. These cosmids plus those from other sources (section 3.1.3.) were fingerprinted and assembled into contigs (Nahmias et al., 1995).
2.1.6. Enzymes

All restriction endonucleases were purchased from Gibco BRL except for BstXI which was obtained from New England Biolabs. T4 DNA ligase and Moloney Murine Leukemia Virus reverse transcriptase (M-MLV RT) were obtained from Gibco BRL, T4 polynucleotide kinase from Amersham, calf intestinal phosphatase from Pharmacia, Taq DNA polymerase from HT Biotechnology Ltd, Proteinase K from Boehringer Mannheim and pancreatic ribonuclease A (RNase A) from Sigma Chemical Company.

2.1.7. Radiation reduced hybrids

The radiation reduced hybrid cell lines used in this study were generated at the Galton Laboratory, UCL by Dr. J. Fitzgibbon, using the irradiation fusion technique of Goss and Harris (1975) (Fitzgibbon, 1993). They were derived from the HPRT+ Chinese hamster-human somatic cell hybrid line, 64063a12 (CJ9q) that contains chromosome 9q as its sole human component. This donor cell line was irradiated with a 45,000 rad dose of X-rays and then fused to an unirradiated HPRT- hamster recipient cell line, WG3H. There was no direct selection for the retention of human DNA but cells were grown in HAT media to select for the rescue of the hamster HPRT+ gene.

2.1.8. Standard solutions and buffers

All solutions and buffers were prepared using distilled water and then sterilised by autoclaving at 15lbs psi, 121°C for 20-30 minutes. The components for standard solutions and buffers are given in appendix 1.1.
2.1.9. Zoo blots

Genomic DNA from orang-utan, dog, rat, hamster and *Xenopus* were obtained from Prof. S. Povey*, DNA from zebra and vole were obtained from Dr. B. Carritt* and mouse DNA was a gift from Dr. A. Pilz*. Human DNA was isolated from a normal individual by Luiza Boules at University College London and African green monkey DNA was isolated by myself from COS-7 cells. Genomic DNA from guanaco and trout were gratefully received from the Animal and Tissue blood bank at the Institute of Zoology, London Zoo.

* MRC Human Biochemical Genetics Unit, The Galton Laboratory, UCL, London.

2.2. Methods

2.2.1. Standard DNA techniques

2.2.1.1. Restriction endonuclease digestion

Between 5-10μg genomic DNA and 0.5-1μg cosmid DNA was digested to completion with 20 units of the required enzyme, using the conditions recommended by the enzyme manufacturer.

2.2.1.2. Agarose gel electrophoresis

DNA fragments were size separated on 1-2% w/v agarose gels (Sigma) prepared in 1 x TBE (appendix 1.1). Ethidium bromide (Sigma) was incorporated at a concentration of 0.5μg/ml and 1 x TBE was used as the electrophoretic running buffer. The samples were mixed with 0.1 x volume of loading buffer (appendix 1.1) and then loaded into the wells of the gel. The voltage and duration of electrophoresis was dependent on the size
of the DNA fragment to be analysed. On completion the DNA was visualised by ultra
violet (UV) transillumination. DNA fragment length was estimated by comparison to a
1kb DNA size ladder (Gibco BRL) that was electrophoresed beside the sample.

2.2.1.3. Southern blotting

DNA was transferred from agarose gels to nylon membrane by the technique of
Southern blotting (Southern, 1975). The gel was agitated in 0-25N hydrochloric acid
(HCL) for 10 minutes to partially nick and depurinate the DNA. After a brief rinse in
distilled water the gel was submerged in denaturing solution (appendix 1.1) for 30
minutes followed by neutralising solution (appendix 1.1) for 30 minutes. GeneScreen
Plus membrane (NEN Research Products) was cut to the size of the gel, pre-wet in
distilled water and then equilibrated in the 10 x SSC (appendix 1.1) for 10 minutes. The
membrane was placed on top of the gel and DNA was transferred by capillary blotting
using 10 x SSC as the transfer buffer. The next day the DNA was fixed to the membrane
by baking at 80°C for 2 hours.

2.2.1.4. Radiolabelling DNA probes

2.2.1.4.a. Random oligonucleotide priming

The method of using random hexanucleotides to prime DNA synthesis from a single
stranded template was introduced by Feinberg and Volgelstein, (1983). The reaction
was performed using a multiprime DNA labelling kit (Amersham) according to the
manufacturers instructions using [$\alpha^{32}$P] dCTP (3000 Ci mmol$^{-1}$) as the isotope. 25ng
DNA probe was denatured at 99°C for 5 minutes in a volume of 30µl and then quenched on ice. 10µl buffer 1 (containing dATP, dTTP, dGTP and reaction buffer) and
5μl buffer 2 (containing random hexanucleotide primers) were added with 30Ci radiolabelled nucleotide dCTP (10Ci/μl) and 2 units DNA polymerase 1 'Klenow' fragment. Labelling proceeded at room temperature overnight or 37°C for 2 hours.

More recently a rediprime DNA labelling system (Amersham) was used with redivue [α³²P] dCTP (Amersham) according to the manufacturers recommendations. This radiochemical has the advantage of being stable at 4°C and has a red dye incorporated to improve its visibility. Nonamer primers were used instead of hexamers and therefore the labelling reaction was complete in 10 minutes. Once completed, 50μl 1 x TNE/0.1% SDS was added to the reaction and it was spun through a 1ml sephadex G50 column at 2000rpm for 5 minutes to remove unincorporated nucleotides.

2.2.1.4.b. End labelling

Short oligonucleotides such as poly(T)_{13-20} were radiolabelled by attaching γ-phosphate of ATP to the 5’ terminus of the molecule using T4 polynucleotide kinase. The reaction was as follows; 10pmol DNA, 1μl reaction buffer (supplied with the enzyme), 2μl [γ-³²P] dATP (10Ci/μl), 2μl enzyme and sterile distilled water upto a volume of 10μl. The reaction proceeded at 37°C for 50 minutes and was terminated by incubation at 65°C for 5 minutes. Unincorporated nucleotides were removed by centrifugation through a 1ml sephadex G25 column at 2000rpm for 5 minutes.

2.2.1.5. Hybridisation of DNA probes

GeneScreen Plus membrane was pre-wet in 2 x SSC and then sealed in a plastic bag or bottle (Hybaid) with 10-20ml hybridisation buffer (appendix 1.1). The plastic bag was
submerged in a 65°C shaking waterbath or the bottle rotated in a 65°C hybridisation oven (Hybaid) for approximately 2 hours.

Sonicated salmon sperm DNA (Sigma) was added to the radiolabelled DNA to give a final concentration in the hybridisation buffer of 100μg/ml. The mixture was denatured by boiling for 5 minutes and then quenched on ice. Hybridisation followed addition of the probe mixture to the hybridisation buffer and proceeded at 65°C for at least 16 hours.

2.2.1.6. Post-hybridisation washes and signal detection

The membrane was initially washed twice in 2 x SSC for 5 minutes at room temperature. More stringent washes followed and after each one the level of radioactivity was monitored with a Geiger counter. Further washes were only performed if a significant level of radioactivity could still be detected. These washes were as follows in increasing order of stringency; 30 minutes in 2 x SSC/1% SDS 60°C, 30 minutes in 0.5 x SSC/1% SDS 60°C and 30 minutes in 0.1 x SSC/1% SDS 60°C. After the final wash excess moisture was removed by blotting with 3MM paper (Whatman) and the damp membrane was wrapped in plastic. Care was taken not to let the membrane dry completely as it then becomes impossible to remove non-specifically bound probe. The membrane was exposed to X-ray film (Kodak) in a light-proof cassette with intensifying screens at -70°C for 1-7 days. The probes were removed from the GeneScreen plus membrane by boiling in 0.1 x SSC/1% SDS for 10-30 minutes and stored at -20°C until required.
2.2.1.7. Mini preparation of cosmid DNA

A 10ml bacterial culture was grown overnight in LB medium (appendix 1.2) with appropriate antibiotic selection (appendix 1.1). The culture was centrifuged at 3000 rpm for 10 minutes, the pellet resuspended in 200µl lysis buffer (50mM glucose, 10mM EDTA, 25mM Tris pH 8) and then the solution transferred to a 1.5ml eppendorf tube. After 10 minutes at room temperature 400µl fresh 0.2M NaOH/1% SDS was added and the tube was placed on ice for 5 minutes. Following the addition of 300µl 3M sodium acetate pH 5.2, the incubation proceeded for a further 10 minutes on ice. Centrifugation at top speed in a microfuge for 5 minutes pelleted the cell debris and the supernatant was transferred to a clean tube. 600µl isopropanol was added and the tube was incubated at -70°C for 10 minutes followed by centrifugation for 5 minutes. The pellet was allowed to air dry and then resuspended in 200µl 0.3M sodium acetate pH 6. Following two phenol extractions and one chloroform extraction, the DNA was precipitated with 200µl isopropanol at -70°C for 10 minutes. The tube was centrifuged in a microfuge for 5 minutes, the pellet washed with 70% ethanol and then resuspended in 50µl TE (appendix 1.1). Finally 1µl of 10mg/ml RNase A (appendix 1.1) was added and the tube incubated at 37°C for 15 minutes. The DNA miniprep was stored at -20°C until required.

In addition to the method above, Wizard™ miniprep DNA purification systems (Promega) were used according to the manufacturers instructions.

2.2.1.8. Maxi preparation of cosmid/plasmid DNA
Wizard™ maxiprep DNA purification systems (Promega) were used to isolate cosmid and plasmid DNA according to the manufacturers recommendations.

2.2.1.9. DNA extraction from cultured cells

Cells from at least three large flasks (160cm²) were washed thoroughly in Hanks’ balanced salt solution (appendix 1.3). Then 7ml lysing solution (appendix 1.1) was added and incubation at 37°C proceeded for at least 5 hours. Three extractions with phenol:chloroform:isoamylalcohol (25:25:1) were followed by two extractions with chloroform. The DNA was precipitated by the addition of 0.1 x volume of 3M sodium acetate and two volumes of ice-cold ethanol. Centrifugation pelleted the DNA and then it was washed in 70% ethanol, air-dried and resuspended in an appropriate volume of TE.

2.2.1.10. Polymerase chain reaction (PCR)

During preparation of the PCR reactions certain precautions were taken to minimise the risk of DNA contamination. The reactions were set up in a hood using Gilson pipettes specific for PCR, the tips and eppendorfs were autoclaved and gloves were worn throughout the procedure.

A reaction mix was prepared using sterile distilled water to give a final concentration of 1x reaction buffer (HT Biotechnology Ltd), 200μM of each dNTP (Pharmacia), 500pM/ml of each oligonucleotide primer (Oswell DNA service) and 0.01 U/ml DNA Taq polymerase (HT Biotechnology Ltd). 100μl of the mix was aliquotted into each 0.5ml eppendorf tube, DNA was added to all but one tube and each were overlaid with
50μl paraffin oil (BDH). The DNA was denatured at 94°C for 5 minutes and then 30-35 cycles of denaturation, primer annealing and extension followed. A Hybaid Omnigene PCR machine was used for all the reactions.

The PCR conditions are specific for each reaction as they depend upon primer composition and DNA sequence to be amplified (sequences given in appendix 3). The annealing temperature of the primer was calculated by using the following equation (Maniatis et al., 1982); \[69.3 + (0.41 \times G + C\%) - 650/\text{bp of primer}\] .

2.2.2. Mammalian cell culture

Cell lines were cultured from stocks frozen in liquid nitrogen. The cells were thawed rapidly at 37°C and then transferred to a 25ml flask containing 15ml pre-warmed medium (appendix 1.3).

Stocks removed from liquid nitrogen were replenished as soon as possible. Cells were pelleted by centrifugation at 1000rpm for 5 minutes and then resuspended in 1-3ml glycerol medium (appendix 2A.3) depending on the size of the pellet and stored in 1ml aliquots.

2.2.2.1. Lymphoblastoid cell lines

Lymphoblastoid cell lines were cultured in RPMI-1640 medium supplemented with 10% foetal calf serum and 5% CO₂ (appendix 1.3). The cells were grown as a suspension with approximately 1 million cells per ml. As the cells divided the medium became depleted and the colour changed from cherry red to yellow. This was a good
indication of when the cells needed to be fed and it was usually every 2-3 days. The cells were allowed to settle at the bottom of the flask, the old medium was removed and double the amount of pre-warmed fresh medium was added. The cells were sub-cultured by removing the old medium and transferring the cells into 2-3 flasks of the same size with fresh medium.

2.2.2.2. COS-7 cell line

COS-7 African green monkey cells were propagated in Eagle’s minimal essential medium (MEM) supplemented with 10% inactivated foetal calf serum (appendix 1.3). The cells were incubated horizontally and adhered to the bottom of the 160cm$^2$ large flasks. Once the cells became confluent the culture was trypsinised and then divided between 3-6 flasks of equal size. To trypsinise the cells it was first necessary to remove the medium and wash with 5ml Hanks’ balanced salt solution (appendix 1.3). After removal of the Hanks’, 5ml trypsin/versene solution (appendix 1.3) was added. The flasks were incubated at 37°C until the cells became detached and then 10ml warm medium was added to deactive the trypsin. The solution was divided into new flasks and warm medium added to give a total volume of approximately 40ml. The cells were then returned to the incubator and allowed to grow at 37°C.

2.2.3. Cosmid library construction

2.2.3.1. DNA preparation

High molecular weight DNA was extracted from hybrid cell lines grown by Dr. J. Fitzgibbon. The DNA was partially digested with restriction endonuclease $MboI$ at different time points to produce fragments approximately 40kb in length. Once this was
achieved DNA was size fractionated by sodium chloride density gradient. Each fraction
was analysed by gel electrophoresis to identify those containing fragments 35-45kb.
Those of correct size were ligated into cosmid vector Lorist B (Cross and Little., 1986)
whose arms had previously been prepared by a method according to Little, (1987). The
ligation reaction consisted of approximately 0.5μg vector DNA, 1μg hybrid DNA, 2μl
5x T4 DNA ligase buffer and 1μl T4 DNA ligase in a volume of 10μl. The reaction
proceeded at 15°C overnight.

2.2.3.2. Packaging reaction

The linear cosmid molecules produced in the ligation were packaged into bacteriophage
lambda head particles. The reaction consisted of 1μl ligation mix, 7μl buffer A (20mM
Tris.Cl pH8.0, 1mM EDTA, 5mM MgCl₂, ½v 0.05% beta-mercaptoethanol), 1μl buffer
Q (6mM Tris.Cl pH7.5, 18mM MgCl₂, 15mM ATP pH7.6, 60mM spermidine pH7.5,
½v 0.2% beta-mercaptoethanol), 3.5μl sonicated extract and 5μl freeze thaw lysate. The
sonicated extract and freeze thaw lysate packaging extracts were prepared according to
Maniatis et al.(1982) by Steven Hall and stored at -70°C. They were placed on dry ice
prior to use and added to the reaction whilst thawing. Incubation proceeded at 30°C for
1 hour and then the reaction volume was increased to 200μl with SM buffer (appendix
1.2).

The RecA phenotype of the recipient bacterial strain, ED8767 was verified prior to use
by testing for UV sensitivity. If the bacterial strain is unable to repair UV damage, there
is a reduced incidence of recombination events between direct repeats in the cosmid and
therefore internal deletion events are less common.
A single colony of UV sensitive ED8767 was grown overnight in LB broth (appendix 1.2) with 0.2% maltose, to promote production of maltose receptor molecules to which the phage particles can adhere. The culture was pelleted at 4000 rpm for 5 minutes and resuspended in half the volume of 10mM MgSO₄.

An equal volume of cells and diluted phage were incubated at 30°C for 15 minutes after which a 4 x volume of LB medium was added. Incubation proceeded at 37°C for 40 minutes to allow the infected bacteria to express the antibiotic resistance genes. They were plated onto Hybond N filters on agar plates with kanamycin selection (appendix 1.1) and allowed to grow at 37°C overnight.

2.2.3.3. Replica plating

Two replicas of each plate were taken with orientation achieved by an asymmetrical pattern of pin-holes. The replicas and masters were allowed to regenerate and then the master stored at 4°C. The DNA from the replica plates was attached to the filter by denaturation in alkali (1.5M NaCl, 0.5M NaOH), followed by neutralisation (1.5mM NaCl, 0.5M Tris.HCl pH8.0) and washing in 2 x SSC to remove bacterial debris. Filters were air dried and then baked for 2 hours at 80°C.

2.2.3.4. Screening for cosmids containing human DNA

Filters were hybridised to radiolabelled total human DNA and if on development of the X-ray film duplicate signals were found on the replica plates, the corresponding colony was picked from the master. It was used to inoculate 10ml LB medium with kanamycin.
selection and was grown overnight. A glycerol stock was prepared by mixing 0.5ml culture with 0.5ml 30% glycerol and it was stored -70°C. The remainder of the culture was used for a cosmid DNA miniprep to be mapped onto human metaphase chromosomes by FISH.

2.2.4. **Fluorescence in situ hybridisation**

2.2.4.1. **Metaphase preparations**

2.2.4.1.a. **Lymphocytes**

1ml of heparinised whole blood from a male donor was added to a 25ml culture flask containing 17ml Iscoves modified DMEM medium (Imperial), 1% GPS (glutamine, penicillin, streptamycin) (appendix 1.3), 2ml foetal calf serum (FCS) and 200μl phytohaemagglutinin (PHA) (ICN labs). The culture was incubated at 37°C for 72 hours with gentle shaking daily. Thymidine (Sigma) was then added to give a final concentration of 0.3mg/ml and the culture was incubated at 37°C for a further 18 hours. Following addition of 2-deoxycytidine (Sigma) (final concentration of 2.3μg/ml) the incubation proceeded at 37°C for 3 hours 55 minutes. Colcemid (Gibco) (final concentration 0.1μg/ml) was added and a final incubation at 37°C followed for 20 minutes. The culture was then transferred to two sterile 10ml tubes and centrifuged at 1000 rpm for 5 minutes. The supernatant was removed and the pellet resuspended dropwise in 5ml pre-warmed 0.075M KCL hypotonic solution. The tubes were left at room temperature for 20 minutes and then centrifuged at 1000 rpm for 5 minutes. The supernatant was removed and the pellet resuspended dropwise in 10ml 3:1 absolute methyl alcohol:glacial acetic acid fixative. The process of centrifugation and
resuspension in fixative was repeated at least twice until the pellet was a clean white
colour. The final suspension was stored at -20°C until required.

2.2.4.1.b. Lymphoblastoid cell lines

Lymphoblastoid cells were transferred to a 10ml tube and colcemid was added to give a
final concentration of 0.05μg/ml. Following incubation at 37°C for one hour the tubes
were centrifuged at 1000rpm for 5 minutes. The cell pellet was gently resuspended in
10ml pre-warmed 0.075M KCL hypotonic solution and incubated at 37°C for 10
minutes. The cells were pelleted, fixed and stored as described above.

2.2.4.1.c. Hybrid cell lines

Hybrid cultures were grown by Dr. J. Fitzgibbon and flasks containing a reasonable
number of dividing cells were selected for harvesting. Colcemid was added to each flask
to give a final concentration of 0.025μg/ml and incubation proceeded at 37°C for 1
hour. The medium was removed to a tube and centrifuged at 1000 rpm for 5 minutes.
Cells were washed from the flask with 5ml of KCL/EDTA hypotonic solution (appendix
1.4) and this solution was added to the cells in the tube. A further 5ml of hypotonic
solution was added to the flask. Incubation of the flask and tube proceeded for 20
minutes at 37°C. The contents of the flask was added to the tube and it was centrifuged
at 1000 rpm for 5 minutes. The cells were fixed as described earlier and suspensions
stored at -20°C until required.

2.2.4.2. Slide preparation and pre-hybridisation treatment

Slides were cleaned and cooled by soaking in methanol containing 0.5% v/v
concentrated hydrochloric acid (HCL). Immediately prior to use the slide was wiped dry
with a lint free cloth, breathed upon and 1-2 drops of chromosome suspension from a 1ml plastic pipette were added. The slide was allowed to air dry slightly before the addition of two drops of fresh fixative. The slide was shaken so it was almost dry and then flooded with 70% glacial acetic acid. It was left for approximately 1 minute and then completely air dried by shaking. Slides were examined under phase-contrast microscopy and only those with a high number of metaphase spreads not surrounded by cytoplasm were used.

During the following procedures the slides were treated in 50ml or 100ml glass Coplin jars unless otherwise stated. Incubations under coverslips (22x50mm) were in a volume of 100μl unless specified otherwise.

The slides were washed for 10 minutes in PBS, dehydrated through an alcohol series (70%, 95% and 100% ethanol) and then air dried. Treatment with 100μg/ml RNase A in 2 x SSC pH7, followed for 1 hour at 37°C in a moist chamber. After two 5 minute washes in 2 x SSC at room temperature and one 5 minutes wash in proteinase K buffer (appendix 1.4) at 37°C, the slides were treated with proteinase K (50ng/ml in proteinase K buffer) for 7 minutes at 37°C. A rinse in formaldehyde buffer (appendix 1.4) was followed by a 10 minute post-fixation step in 1% formaldehyde in formaldehyde buffer. The slides were washed in PBS for 5 minutes, dehydrated through an alcohol series and air dried.

2.2.4.3. Biotin-labelling the probe by nick translation

The concentration of DNA from cosmid minipreps was measured by DNA fluorimetry using a TKO-100 minifluorimeter (Hoefer). 0.5-1μg DNA was labelled with biotin-14-
dTTP by nick translation using a BioNick™ labelling system (Gibco BRL). The reaction was performed according to the manufacturers instructions and is designed to generate small biotin labelled DNA probes between 50-500bp in size. Following incubation at 16°C for 1 hour the reaction was stopped with 5μl 300mM EDTA pH8 and unincorporated nucleotides were removed using a NICK™ column (Pharmacia). The column was developed with 1 x TNE buffer (appendix 1.1), the probe was added and then eluted with the second 400μl volume of 1 x TNE. The biotinylated probes were stored at -20°C until required.

2.2.4.4. Preparation of labelled probe

2.2.4.4.a. Cosmid probes

Cosmid probes were prepared with an excess of unlabelled Cαf-1 DNA to suppress the repetitive DNA sequences (Lichter et al., 1990a). For each hybridisation reaction the following components were added to a 1.5ml eppendorf tube; 200ng purified biotinylated probe, 10μg sonicated herring sperm DNA, 10μg yeast tRNA, 10μg Cαf-1 DNA (Gibco BRL), 0.1 x volume 3M sodium acetate and 2.5 x volume ice-cold absolute ethanol (99.7-100%). Following precipitation at -70°C for 1 hour the tubes were centrifuged in a cold-room microfuge for 15 minutes. The supernatant was removed and the pellet freeze-dried for 10 minutes. The pellet was resuspended in 10μl hybridisation mix (appendix 1.4) and incubated at 37°C for 10 minutes to ensure the DNA was properly dissolved. The probe mixture was denatured at 70°C for 5 minutes and the repeat sequences were allowed to pre-anneal to the Cαf-1 DNA at 37°C for approximately 90 minutes prior to hybridisation.

2.2.4.4.b. Total human DNA probe
Probe preparation was the same as above but repeat sequences were not suppressed. Therefore unlabelled \( C_{o-t-l} \) DNA was not added and there was no pre-annealing step.

2.2.4.5. Hybridisation

2.2.4.5.a. Cosmid probes

Chromosomal DNA on the slide was denatured under a coverslip with 100\( \mu \)l 70% formamide/2 x SSC at 80°C for 3 minutes. Immediate quenching in ice-cold 70% ethanol was followed by dehydration through an alcohol series and then the slide was air dried. The denatured and pre-annealed probe was applied under a 22mm round coverslip and the edges sealed with cow gum. Hybridisation proceeded at 37°C in a moist chamber over-night.

2.2.4.5.b. Total human DNA probe

The 10\( \mu \)l probe mixture was applied under a 22mm round coverslip and sealed with cow-gum. The probe and chromosomal DNA were denatured simultaneously in a 80°C oven for 5 minutes. Hybridisation proceeded at 37°C in a moist chamber over-night.

2.2.4.6. Post-hybridisation washes

2.2.4.6.a. Cosmid probes

Coverslips were removed and the slides were washed three times for 5 minutes in 50% formamide/2 x SSC at 45°C. This was followed by three 5 minute washes in 0.1 x SSC at 65°C and a single wash of 5 minutes in 4 x SSC/0.05% Tween20 at room temperature.
2.2.4.6.b. Total human DNA probe

The slides were initially washed three times for 5 minutes in 70% formamide/2 x SSC at 37°C and then three times for 5 minutes in 2 x SSC at room temperature. The third and final wash was the same as above.

2.2.4.7. Detection of hybridisation

Biotinylated DNA was detected using procedures described by Pinkel et al. (1986). The slides were first equilibrated in 4 x SSC/5% non-fat milk (Marvel) (4xSSC/NFM) for 20 minutes. All following incubations proceeded at room temperature for 20 minutes and were carried out in a volume of 100μl under a 22x50mm coverslip. The initial incubation was with avidin-FITC (Vector Laboratories) at a concentration of 5μg/ml in 4 x SSC/NFM and from this step onwards the slides were protected from exposure to the light. The slides were then washed three times for 5 minutes in 4 x SSC/0.05% Tween 20 and the hybridisation signal was amplified by incubation with 5μg/ml biotinylated goat anti-avidin (Vector Laboratories) in 4 x SSC/NFD, followed by washes as before and a second incubation with 5μg/ml avidin-FITC. After a 5 minute wash in 4 x SSC/0.05% Tween 20 and two 5 minute washes in PBS, the slides were drained and mounted in 20μl Vectorshield medium (Vector Laboratories) containing 10μg/ml 4, 6-diamidinophenylindole (DAPI) and 2μg/ml propidium iodide (PI) as counterstains for both Q-banding and R-banding.

2.2.4.8. Microscopy

The preparations were analysed by confocal microscopy using a laser scanning confocal microscope (BIO-RAD MRC 600). The laser beam was used to excite the
fluorochromes which then emit light of a specific wavelength. The emissions were collected separately by two photomultiplier tubes using the appropriate filter sets and then the images merged. Polaroid photographs were produced using a Mitsubishi colour video copy processor.

2.2.5. Exon Amplification

2.2.5.1. Subcloning genomic DNA

2.2.5.1.a. Dephosphorylation of vector DNA

The terminal 5' phosphate groups of linearised vector DNA were removed by treatment with calf intestinal alkaline phosphatase (Pharmacia) to prevent recircularization of the vector during the ligation reaction. Vector pSPL1 was linearised with restriction enzyme BamHI and vector pSPL3 was linearised with single enzyme digests of both BamHI and PstI. Digests were performed in One-Phor-All Plus buffer (Pharmacia) using 2-5μg DNA. On completion, 0.1 unit calf intestinal alkaline phosphatase was added and incubation at 37°C proceeded for 30 minutes. The reaction was stopped by inactivating the enzyme at 85°C for 15 minutes. Two phenol extractions and one chloroform extraction followed. The DNA was precipitated by addition of 0.1 x volume 3M sodium acetate, 2 x volume ethanol and incubation at -70°C for 15 minutes. Following centrifugation the pellet was washed in 70% ethanol, air-dried and resuspended in an appropriate volume TE (100-500ng/μl).

2.2.5.1.b. Digestion of genomic DNA

Approximately 1μg cosmid DNA was digested with the appropriate restriction endonuclease(s). Cosmids to be sub-cloned into vector pSPL1 were digested with
BamHI, BglII and double digests with both enzymes present. Single enzyme digests using BamHI and PstI were performed for cosmids sub-cloned into vector pSPL3. After complete digestion the DNA was ethanol precipitated and resuspended in an appropriate volume TE (50-250ng/μl).

2.2.5.1.c. Ligation reaction

The ligation reactions were usually performed with a vector : insert ratio of 1: 0.5. However in some instances a number of ligations were needed with different ratios in order to obtain maximum shot-gun cloning efficiencies.

Each reaction consisted of : 100ng dephosphorylated vector DNA, 50ng digested insert DNA, 2μl 5 x T4 ligase buffer (Gibco BRL) and 1U T4 ligase (1U/μl) in a final volume of 10μl. Incubation proceeded at 15°C overnight and then the reactions were stored at -20°C until required.

2.2.5.2. Preparation of competent cells

Cells were prepared as described by Hanahan, (1983). A single colony of JM101 from a LB agar plate was used to inoculate 10ml LB medium (appendix 1.2). After shaking at 37°C overnight the culture was used to inoculate 200ml pre-warmed LB medium. Once the absorbance reached 0.45 at 550nm the culture was ready for the preparation of competent cells and all the following steps were performed in a cold room at 4°C using cold pipettes and tubes. The culture was divided into four 50ml tubes (Falcon 2070) and placed on ice for 15 minutes. The cells were pelleted by centrifugation at 2500rpm for 15 minutes at 4°C and gently resuspended in 2ml TFB1 (appendix 1.2). A further 10ml TFB1 was added and the cells were left on ice for 15 minutes. After centrifugation at
2500rpm for 15 minutes at 4°C the cells were resuspended in 1.6ml TFB2 (appendix 1.2) and incubated on ice for a further 15 minutes. The cells could be used immediately, stored at 4°C for use in the next few days or stored at -70°C in 200μl aliquots until required.

2.2.5.3. Transformation

10-50ng DNA was added to 200μl of competent cells and they were mixed gently. Incubation on ice proceeded for 60 minutes, the cells were heatshocked in a waterbath at 42°C for 90 seconds and then they were incubated on ice for a further 2 minutes. 800μl prewarmed LB medium (42°C) was added and the cells were incubated at 37°C for 60 minutes shaking at 225rpm. One-tenth of the volume of the transformation was plated onto LB agar plates with ampicillin (100μg/ml) antibiotic selection. The plates were allowed to dry at room temperature for 10 minutes and then incubated at 37°C overnight. The remaining nine-tenths of the transformation was used to inoculate 10ml LB medium containing ampicillin (100μg/ml) and incubated with gentle shaking at 37°C overnight.

Analysis of the plates the following day determined the efficiency of the transformation. Control reactions consisted of; ligation and transformation of dephosphorylated vector only; transformation containing no DNA and transformation of a known amount of super-coiled DNA. If the transformation was a success the efficiency was approximately 10⁸ transformants/μg of super-coiled plasmid DNA and the frequency of non-recombinant colonies was less than 10%. If these requirements were satisfied plasmid DNA was isolated from the overnight cultures.
2.2.5.4. Isolation of DNA

A small scale purification of plasmid DNA from the overnight culture was performed using the Wizard™ Miniprep DNA purification system (Promega) according to the manufacturer’s recommendations.

The success of shot-gun cloning using a single restriction enzyme was examined by digesting the DNA with the appropriate enzyme to release the insert DNA. The products were visualised on an ethidium bromide stained 0.8% agarose gel and compared to a digest of the original cosmid used. A very intense band should be visible corresponding to the vector DNA and all the cosmid DNA fragments should also be present.

2.2.5.5. Transfection into COS-7 cells

The cells were grown in large 160cm\(^2\) flasks until they reached 75-80% confluence (approximately 12 x 10\(^6\) cells). Then they were trypsinised and pelleted by centrifugation at 1200 rpm, 4\(^\circ\)C, 7-8 minutes. The supernatant was removed and the cells washed in 30-40ml ice-cold PBS. The cells were pelleted again and resuspended in 0.7ml cold PBS per transfection. One large flask was sufficient for three transfections (approximately 4 x 10\(^6\) cells each).

For the transfection, 0.7ml cells were added to a sterile pre-chilled 0.4cm electroporation cuvette (Bio-rad) with 1-20\(\mu\)g miniprep DNA made up to a total volume of 100\(\mu\)l with cold PBS. The reaction was allowed to stand on ice for 10 minutes, the cells were then gently resuspended, placed in the electroporation apparatus (Bio-rad)
and pulsed at a setting of 1.2KV, 25uf. The cuvette was immediately returned to ice and
the cells allowed to recover for approximately 10 minutes. They were then transferred
to a small 25ml tissue culture flask containing 10ml prewarmed medium. Incubation
proceeded at 37°C for 2-3 days and then the RNA was isolated.

2.2.5.6. Isolation of RNA

During the isolation of the cytoplasmic RNA the cells and solutions were kept ice-cold,
care was taken to ensure that the conditions were free of RNase by pre-treatment of the
solutions and plasticware with DEPC (Maniatis et al., 1982).

The medium was removed from the flasks and the cells washed three times with cold
PBS. The flasks were placed on a bed of ice and a final volume of 10ml PBS was
added. The cells were removed from the flask using a sterile cell scraper and the
solution was transferred to a 10ml cold sterile centrifuge tube. After centrifugation at
1200 rpm, 4°C for 7-8 minutes the supernatant was removed and the pellet resuspended
in 300μl TKM 10:10:1 (appendix 1.1). Following a 5 minute incubation on ice, 5μl
10% Triton X-100 was added. The tube remained on ice for a further 5 minutes. The
nuclei were pelleted by centrifugation at 1500 rpm, 4°C for 5 minutes and the
supernatant removed to a cold microfuge tube. Then 300μl phenol:chloroform (1:1) was
added, the tube was vortexed, centrifuged at room temperature for 3 minutes and the
aqueous layer removed to a clean, cold tube. 12μl 5M NaCl and 750μl ethanol were
added and the tube was placed in a -70°C freezer to precipitate the RNA. Centrifugation
in a coldroom microfuge for 30 minutes pelleted the RNA. Finally it was washed with
70% ethanol, dried and resuspended in 20μl DEPC-treated water. The RNA was stored
at -70°C and thawed on ice when required.
2.2.5.7. Reverse transcriptase PCR amplification

2.2.5.7.a. Using pSPL1

2.2.5.7.a.i. cDNA synthesis

1-3 μg RNA was added to a tube containing: 1 x PCR buffer (HT Biotechnology Ltd), 200μM each dNTP, 1μM 3' oligonucleotide SA2 and 4mM DTT (Gibco BRL). This was heated to 65°C for 5 minutes and then 4U RNasin ribonuclease inhibitor (Promega) and 200U M-MLV reverse transcriptase (200U/μl) were added to give a final volume of 25ml. The reverse transcription reaction (RT) proceeded for 90 minutes at 42°C.

2.2.5.7.a.ii. PCR

The entire 25μl RT reaction was used for PCR and the reaction conditions were 1 x PCR buffer (HT Biotechnology Ltd); 200μM each dNTP; 1μM 3' oligonucleotide SA2; 1μM 5' oligonucleotide SD2 and sterile distilled water to a final volume of 100μl. Denaturation occurred at 94°C for 5 minutes, 0.5U Taq polymerase (5U/μl) (HT Biotechnology Ltd) was added and 35 cycles of denaturation at 94°C for 1 minute, primer annealing at 58°C for 2 minutes and extension at 72°C for 3 minutes took place. The RNA/PCR products were visualised on 2% agarose gels with ethidium bromide staining.

2.2.5.7.b. Using pSPL3

2.2.5.7.b.i. cDNA synthesis
The reverse transcription reaction was the same as that for pSPlJ 1 except 1 x first strand buffer (Gibco BRL) was used instead of PCR buffer, the concentration of dNTP was increased from 200μM to 500μM and the reaction volume was 20μl.

2.2.5.7.b.ii. Primary PCR

The RT reaction was divided in half and used for two 100μl PCR reactions each with the following conditions; 1 x PCR buffer (HT Biotechnology Ltd), 200μM of each dNTP, 1μM 3' oligonucleotide primer dUSA4 and 1μM 5' oligonucleotide dUSD2. After denaturation at 94°C for 5 minutes, 0.5U Taq polymerase (5U/μl) (HT Biotechnology Ltd) was added. One of the tubes was amplified using 30 cycles of 94°C for 40 seconds and 72°C for 1 minute, whereas the second was amplified for only 5 cycles and the reaction was held at 55°C for BstX1 enzyme digestion.

2.2.5.7.b.iii. BstX1 enzyme digestion

The primary PCR with only 5 cycles had 20U BstX1 restriction endonuclease added and the digestion proceeded for 2 hours at 55°C.

2.2.5.7.b.iv. Secondary PCR

Following enzyme digestion the PCR amplification continued for a further 25 cycles using the same conditions. The addition of further Taq polymerase enzyme was unnecessary.

2.2.5.8. Cloning and analysis of PCR products

2.2.5.8.a. Using pSPL1

2.2.5.8.a.i. Identifying and purifying exons
PCR products were analysed on 2% agarose gels and visualised using ethidium bromide. Any bands larger than the size expected to be produced by vector only splicing suggested that a recombinant has been amplified. These products were separated on low melting point agarose gels (Nusieve) using 1 x TAE (appendix 1.1) as the electrophoretic running buffer. The gel was visualised using low wavelength UV light and the band corresponding to an amplified exon was excised using a clean scalpel. The DNA was purified using a Wizard™ PCR prep DNA purification system (Promega) according to the manufacturer’s recommendations.

2.2.5.8.ii. Secondary PCR

The purified DNA fragment was used as a template in a second PCR amplification reaction using a pair of internal oligonucleotides SA1 and SD1 that flank the vector slice junctions.

2.2.5.8.iii. Verification of a true exon

The secondary PCR products were sequenced and used as probes to screen Southern blots of cosmids, zooblots and cDNA libraries.

2.2.5.8.iv. Cloning

The trapped exons that were shown to consist of unique sequences were cloned using a TA cloning kit (Invitrogen). The secondary PCR product was cloned into the plasmid vector pCR™II according to the manufacturer’s instructions.

2.2.5.8.b. Using pSPL3

2.2.5.8.b.i. Identifying and purifying exons
Same as for pPSL1.

2.2.5.8.b.ii. Cloning

The dUSA2 and dUSD4 primers used in the PCR reaction have deoxy-dUMP residues at the 5' end and facilitate rapid cloning of the products using the CLONEAMP™ system (Gibco BRL). The procedure was performed as the manufacturers’ instructed. Between 10-50ng purified PCR product was added to a 0.5ml eppendorf tube with 50ng pAMP1 vector DNA, 1x annealing buffer (20mM Tris-HCL (pH8.4), 50mM KCL, 1.5mM MgCl2), 1 U Uracil DNA glycosylase and sterile distilled water to make a final volume of 20μl. The tube was incubated at 37°C for 30 minutes to allow simultaneous deglycosylation of dUMP residues and annealing of vector and insert DNA. Following incubation the tube was placed on ice and 5μl of the reaction was used for transformation. The remainder of the reaction was stored at -20°C.

The transformation was performed with Epicurian Coli XL1-Blue supercompetent cells (Stratagene) using the protocol supplied. The cells were thawed on ice and 50μl were aliquotted into pre-chilled 15ml Falcon 2059 polypropylene tubes. To improve transformation efficiency, B mercaptoethanol (25mM final concentration) was added to the cells and they were incubated on ice for 10 minutes with gentle agitation every 2 minutes. Following addition of 5ng DNA the cells were swirled gently and allowed to sit on ice for 30 minutes. A heat pulse in a 42°C waterbath for 45 seconds was proceeded by 2 minutes on ice. Then 0.9ml pre-warmed (42°C ) SOC medium (appendix 1.2) was added and the tubes were incubated for 1 hour at 37°C shaking at 225rpm. 200μl of the transformation was plated on LB agar plates containing ampicillin antibiotic (100μg/ml) and Xgal (30μg/ml) and IPTG (30μg/ml) (appendix 1.2) for
blue/white colour selection. The plates were incubated overnight at 37°C and inspected the following day. Recombinant colonies could be easily detected by their white colour corresponding to disruption of the \textit{lac Z} gene. A maximum of 30 white colonies were picked from each agar plate and they were used to inoculate 100μl LB medium in wells of a microtitre plate. The plates were incubated at 37°C overnight shaking at 255rpm. Each well culture was confirmed to contain an insert in the pAMP1 vector by PCR amplification using a pair of internal PCR primers that flank the splice sites (SA5 and SD5 or KSFl and KSR1). The reaction mix was the same as usually only 1μl culture was used as the template DNA and the initial denaturation was increased from 5 minutes to 10 minutes.

2.2.5.8.b.iii. Verification of a true exon

Same as for pPSL1 except that cloned exons were sequenced instead of secondary PCR products.

2.2.6. Sequencing

2.2.6.1. Sequencing reactions

The chain termination method of sequencing (Sanger \textit{et al.}, 1977) was carried out by using the Sequenase Version 2 sequencing system (USB).

2.2.6.1.a. Using pSPL1

The secondary PCR products were used as templates for sequencing and the forward and reverse PCR primers SA1 and SD1 were used to prime the reaction. The annealing reaction consisted of: 200-500ng PCR product in a volume of 6μl with distilled water,
1μl of the appropriate primer (10pM/μl), 2μl reaction buffer and 1μl dimethyl sulfoxide. It was denatured by heating to 99°C for 3 minutes and then immediately frozen on a mix of dry-ice and ethanol. The labelling reaction consisted of 1μl DTT (0.1M), 2μl labelling mix (Sequenase labelling mix diluted 1 in 5 with distilled water), 0.5μl α35S dATP (1415 Ci mmol⁻¹) and 2μl Sequenase Version 2 enzyme (diluted 1 in 4 with Sequenase enzyme dilution buffer). The labelling reaction was added to the annealing reaction, centrifuged and allowed to thaw at room temperature. After 5 minutes, 3.5μl reaction was added to each of the four termination mixes ddG, ddA, ddT and ddC. Following 5 minutes at 42°C the reaction was stopped with 4μl formamide stop solution provided.

2.2.6.1.b. Using pSPL3

Plasmid miniprep DNA containing the cloned exon(s) was made single-stranded by alkaline-denaturation using the protocol provided with the DNA sequencing kit (Sequenase). Initially, 3-5μg plasmid DNA and 0.1 x volume 2M NaOH/2mM EDTA were combined in an eppendorf tube. After an incubation at 37°C for 30 minutes, 0.1 x volume 3M sodium acetate pH5.3 was added to neutralised the mixture. The DNA was precipitated following the addition of 2 x volume ethanol and placement at -70°C for 15 minutes. The DNA was pelleted by centrifugation, washed in 70% ethanol and resuspended in 7μl distilled water.

The annealing reaction consisted of : 7μl single-stranded plasmid DNA, 2μl Sequenase reaction buffer and 1μl of the appropriate primer (10pM/μl). After incubation at 65°C for 2 minutes the tube was placed in a beaker of water at 65°C and allowed to cool slowly at room temperature. The annealing was complete once the temperature was...
below 35°C and it was then placed on ice. The labelling reaction consisted of 1µl DTT (0.1M), 2µl labelling mix (Sequenase labelling mix diluted 1 in 5 with distilled water), 0.5µl α35S dATP (1415 Ci mmol⁻¹) and 2µl Sequenase Version 2 enzyme (diluted 1 in 4 with Sequenase enzyme dilution buffer). The two reactions were combined and left at room temperature for 5 minutes. Meanwhile 2.5µl of each termination mix was aliquotted in 0.5µl eppendorf tubes and pre-warmed to 37°C. 3.5µl reaction mix was added to each termination mix and incubated at 37°C for 5 minutes. The reaction was stopped by the addition of 4µl stop solution.

2.2.6.2. Polyacrylamide gel electrophoresis

The sequencing apparatus used was 21 x 50cm or 21 x 40cm and supplied by Bio-rad. It was cleaned with distilled water and ethanol prior to use and the top plate was treated with approximately 3mls of Gel-Slick™ to prevent the gel from sticking. The gel was prepared with 60mls acrylamide mix (7M urea, 6% acrylamide (Severn Biotech Ltd) and 1 x TBE), 74µl TEMED (Bio-rad) and 74µl freshly prepared 25% M/V ammonium persulphate (Bio-rad). The thickness of the gel was 0.4mm and a straight line was created at the top by inserting the comb upside down. Once the gel had been allowed to set for one hour this was removed and inserted in the opposite orientation to produce the wells into which the sample was loaded.

The gel was pre-run at 45 watts for 20 minutes in pre-warmed 1 x TBE to achieve an optimum running temperature of 50°C. Once this was obtained the samples were denatured at 75°C for 3 minutes and 3µl was loaded into a 24 well sharks tooth comb or 2µl into a 48 well comb. The length of time the samples were allowed to run depended
on the relative positions of the bromophenol blue and xylene cyanol dyes which were present in the formamide stop solution.

### 2.2.6.3. Signal detection

When the sequencing apparatus was disassembled the gel remained on the bottom glass plate, it was transferred to 3mm filter paper covered with cling film and then dried for 2 hours at 80°C using a Bio-rad gel drier. Signals were detected by autoradiographic exposure with B-max hyperfilm (Amersham) for 12-72 hours at room temperature.

### 2.2.7. Screening cDNA libraries

#### 2.2.7.1. Lambda ZAP®II cDNA library (Stratagene)

A human foetal brain cDNA library (Stratagene) was screened according to the manufacturers recommendations, using XL1-Blue as the host strain. The concentration of the library was calculated by following the titering procedure and found to be $1.1 \times 10^{10}$ pfu/ml. An appropriate volume of the library was then plated onto large (22 x 22cm) NZY plates (appendix 1.2) to give 500,000 pfu/plate. These were grown overnight at 37°C and transferred onto nylon membrane (Hybond N) in duplicate using pin holes for orientation. After 7 minutes in denaturing solution (appendix 1.1) and 6 minutes in neutralising solution (appendix 1.1) the filters were rinsed in 2 x SSC and blotted briefly on Whatmann 3MM paper. The DNA was crosslinked to the filters by baking at 80°C for 2 hours.

Two filters each containing approximately 500,000 plaques were prehybridised at 65°C for a minimum of two hours, in 20 ml hybridisation buffer (appendix 1.1). 25ng PCR product was radio-labelled and a pool of 4-5 different exons was added to the filters.
After hybridisation overnight, the filters were washed and the signals detected by autoradiography. Plaques corresponding to duplicate signals were picked into 1ml SM buffer (appendix 1.2) containing 20μl chloroform and were used for a secondary screen. The phage were diluted and titered with host cells on small (9cm) plates and lifts were taken from those plates with approximately 200 plaques. The procedure was repeated and if positive signals corresponded to single plaques a tertiary screen was unnecessary. The plaque was picked into 1ml SM buffer with 20μl chloroform and was stable at 4°C for up to one year.

The EXASSIST™/SOLR™ system (Stratagene) was used for \textit{in vivo} excision of the pBluescript® phagemid from the Lambda ZAP™ II vector. Freshly prepared XL1-Blue and SOLR cells were used as hosts and the procedure was taken from the instruction manual. SOLR cells were used with the Exassist helper phage because they contain an amber mutation which prevents replication of the phage genome, therefore only the replication of the excised phagemid is permitted. Following a successful rescue the excised phagemids grew as colonies on LB agar plates with ampicillin selection (appendix 1.2). Overnight cultures were grown for each phagemid, a glycerol stock was prepared and DNA was isolated for further analysis.

\textbf{2.2.7.2. Gridded cDNA filters (ICRF)}

Gridded human foetal brain cDNA library filters (ICRF library no. 507 see materials) were prehybridised in Church hybridisation buffer (appendix 1.1) and were screened with pools of trapped exons as described above. The co-ordinates of positive signals were deduced and the clones requested. The positive plasmid clones were received as
bacterial stabs and they were streaked onto 2 x YT agar plates (appendix 1.2) with ampicillin selection. A single colony was picked and grown overnight in 2 x YT broth (appendix 1.2) with ampicillin selection. After preparation of a glycerol stock, miniprep plasmid DNA was isolated and analysed by restriction enzyme digestion and Southern blotting.
Chapter 3

Results
Chapter 3

3. Results

3.1. Physical mapping human chromosome 9q34

The physical mapping strategy used by our research group was designed to identify genomic DNA clones from the TSC1 candidate region and to produce an overlapping clone map (Nahmias et al., 1995). Radiation hybrids prepared by Dr. J. Fitzgibbon, from the somatic cell hybrid CJ9q (containing chromosome 9q as its sole human component) provided a useful basic resource of region specific DNA (see section 2.1.7).

3.1.1. Characterisation of radiation hybrids

A total of 39 hybrids were generated and they were characterised for the retention of a series of well mapped markers on chromosome 9q from D9S48 to D9S11 (Fitzgibbon, 1993) (see table 3.1). 25 hybrids were positive for at least one of the markers tested and 16 of these retained loci distal to AK1 on chromosome 9q34.

Further characterisation by fluorescence in situ hybridisation (FISH) was performed by myself and Dr. D. Griffin (Griffin, 1992). Biotinylated human DNA was used as a probe onto hybrid metaphase chromosome spreads to detect the number of discrete human fragments present in each hybrid cell line (see table 3.2). 37 hybrids were analysed and 26 of these showed at least one hybridisation signal. Human DNA was present in the hybrid cell lines as multiple or single fragments which were fused to
Table 3.1

Marker retention analysis of radiation hybrids

<table>
<thead>
<tr>
<th>Hybrid</th>
<th>Markers used</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>D 9 S</td>
</tr>
<tr>
<td></td>
<td>9</td>
</tr>
<tr>
<td>1A</td>
<td>NT</td>
</tr>
<tr>
<td>2A</td>
<td>NT</td>
</tr>
<tr>
<td>3A</td>
<td>NT</td>
</tr>
<tr>
<td>5A</td>
<td>NT</td>
</tr>
<tr>
<td>6A</td>
<td>NT</td>
</tr>
<tr>
<td>6B</td>
<td>NT</td>
</tr>
<tr>
<td>6C</td>
<td>NT</td>
</tr>
<tr>
<td>7A</td>
<td>NT</td>
</tr>
<tr>
<td>7B</td>
<td>NT</td>
</tr>
<tr>
<td>8C</td>
<td>NT</td>
</tr>
<tr>
<td>9B</td>
<td>NT</td>
</tr>
<tr>
<td>12C</td>
<td>NT</td>
</tr>
<tr>
<td>13A</td>
<td>NT</td>
</tr>
<tr>
<td>13B</td>
<td>NT</td>
</tr>
<tr>
<td>13C</td>
<td>NT</td>
</tr>
<tr>
<td>14A</td>
<td>NT</td>
</tr>
<tr>
<td>17A</td>
<td>NT</td>
</tr>
<tr>
<td>17B</td>
<td>NT</td>
</tr>
<tr>
<td>18A</td>
<td>NT</td>
</tr>
<tr>
<td>18B</td>
<td>NT</td>
</tr>
<tr>
<td>19B</td>
<td>NT</td>
</tr>
<tr>
<td>20A</td>
<td>NT</td>
</tr>
<tr>
<td>20B</td>
<td>NT</td>
</tr>
<tr>
<td>21A</td>
<td>NT</td>
</tr>
</tbody>
</table>

- Hybrid testing positive for a particular marker
- Hybrid testing negative for a particular marker
- Hybrid not tested for a particular marker
- Hybrid testing positive for markers distal to AK1 on 9q34

Table taken from Fitzgibbon, 1993.
For details of the markers and assays tested see Fitzgibbon, 1993.
25 hybrids were positive for at least one of the markers tested, the remaining 14
(3B, 3C, 4B, 8A, 8B, 9A, 9C, 11A, 11B, 12A, 15B, 16B, 16C and 22B) were
negative for all.
Table 3.2

FISH results using biotinylated total human DNA as probe onto radiation hybrid metaphase chromosomes

<table>
<thead>
<tr>
<th>Hybrid</th>
<th>Number of human fragments</th>
<th>Stability of human fragment(s)</th>
<th>Type of human fragment</th>
</tr>
</thead>
<tbody>
<tr>
<td>1A</td>
<td>1</td>
<td>100%</td>
<td>Small, solitary</td>
</tr>
<tr>
<td>2A</td>
<td>1</td>
<td>80%</td>
<td>Large, fused to rodent chromosome</td>
</tr>
<tr>
<td>2B</td>
<td>Multiple (&gt;3)</td>
<td>60%</td>
<td>Different content in each cell</td>
</tr>
<tr>
<td>3A</td>
<td>5</td>
<td>80%</td>
<td>4 small, 1 large</td>
</tr>
<tr>
<td>3C</td>
<td>1</td>
<td>60%</td>
<td>Tiny and integrated</td>
</tr>
<tr>
<td>5A</td>
<td>Multiple (&gt;3)</td>
<td>Different</td>
<td>Different in each cell</td>
</tr>
<tr>
<td>6A</td>
<td>1</td>
<td>95%</td>
<td>Fragments of equal size fused</td>
</tr>
<tr>
<td>6B</td>
<td>Multiple (&gt;3)</td>
<td>Different</td>
<td>Different in each cell</td>
</tr>
<tr>
<td>6C</td>
<td>Multiple (&gt;3)</td>
<td>Different</td>
<td>Different in each cell</td>
</tr>
<tr>
<td>7A</td>
<td>Multiple (&gt;3)</td>
<td>Different</td>
<td>Different in each cell</td>
</tr>
<tr>
<td>7B</td>
<td>Multiple (&gt;3)</td>
<td>Different</td>
<td>Different in each cell</td>
</tr>
<tr>
<td>8B</td>
<td>1</td>
<td>50%</td>
<td>Small, solitary</td>
</tr>
<tr>
<td>9A</td>
<td>Multiple (&gt;3)</td>
<td>Different</td>
<td>Different in each cell</td>
</tr>
<tr>
<td>9B</td>
<td>Multiple (&gt;3)</td>
<td>Different</td>
<td>Different in each cell</td>
</tr>
<tr>
<td>12C</td>
<td>2</td>
<td>100%</td>
<td>Small, both very close and integrated</td>
</tr>
<tr>
<td>13B</td>
<td>1</td>
<td>30%</td>
<td>Tiny, solitary</td>
</tr>
<tr>
<td>14A</td>
<td>3</td>
<td>Different</td>
<td>Different in each cell</td>
</tr>
<tr>
<td>15B</td>
<td>1</td>
<td>80%</td>
<td>Tiny, integrated</td>
</tr>
<tr>
<td>17A</td>
<td>1</td>
<td>80%</td>
<td>Small, solitary</td>
</tr>
<tr>
<td>17B</td>
<td>1</td>
<td>99%</td>
<td>Small, integrated</td>
</tr>
<tr>
<td>18A</td>
<td>Multiple (&gt;3)</td>
<td>Different</td>
<td>Different in each cell</td>
</tr>
<tr>
<td>18B</td>
<td>Multiple (&gt;3)</td>
<td>Different</td>
<td>Different in each cell</td>
</tr>
<tr>
<td>19B</td>
<td>Multiple (&gt;3)</td>
<td>Different</td>
<td>Different in each cell</td>
</tr>
<tr>
<td>20A</td>
<td>1</td>
<td>50%</td>
<td>Small, solitary</td>
</tr>
<tr>
<td>20B</td>
<td>1</td>
<td>40%</td>
<td>Small, solitary</td>
</tr>
<tr>
<td>21A</td>
<td>1</td>
<td>25%</td>
<td>Small, integrated</td>
</tr>
</tbody>
</table>

Stability was determined by analysing the retention of human fragments in 20 hybrid metaphase chromosome spreads. When stability was less than 100% the remaining cells showed either variable or no human content.

Shaded hybrids were analysed by myself and the other were analysed by Dr. D. Griffin. An addition nine hybrids (3B, 4B, 8A, 8C, 9C, 11A, 11B, 12A and 13A) analysed by myself and two analysed by Dr. D. Griffin (16C and 16B) did not show a hybridisation signal. 13C and 22B were not tested.
hamster chromosomes, integrated into hamster chromosomes or present as solitary fragments with no visible hamster element (figure 3.1). The stability of each human fragment was determined by analysing its retention in twenty metaphase chromosome spreads (table 3.2). A number of hybrids contained multiple human fragments (for example 5A, 9A and 14 A) and these showed different stability frequencies for each fragment.

Hybrids showed a correlation between FISH and marker analysis results (see table 3.3). For example hybrid 14A showed three human fragments by FISH and three non-contiguous markers in the retention study. However inconsistencies were also found. Two hybrids (8C and 16B) were positive for marker analysis but no human material was detected by FISH. It suggested that when the cell line was resuscitated from liquid nitrogen for expansion for the FISH analysis, the human DNA might have been lost. Alternatively the human DNA fragments may be very small and therefore below the limit of resolution for FISH, or retained in only a small percentage of the cells so therefore the likelihood of detection was much lower. Hybrids such as 20A, which showed a single fragment by FISH, but more than one by marker analysis, may also be explained by the presence of small human fragments that cannot be detected by FISH. Alternatively, it may be that a number of human DNA fragments fused to each other after irradiation.

Other contradictory results included four hybrids which showed human fragment(s) by FISH but were negative for all the markers tested (3C, 8B, 9A and 15B) and hybrids such as 6C which showed a single fragment in the retention study but multiple fragments by FISH. It is thought that these differences probably reflect the paucity of
Figure 3.1

FISH of biotinylated total human DNA onto radiation hybrid metaphase chromosomes

A) Hybrid 6A
   - Single human fragment fused to a rodent chromosome

B) Hybrid 17B
   - Single human fragment integrated into a rodent chromosome

C) Hybrid 20A
   - Single solitary human fragment

D) Hybrid 9A
   - Multiple human fragments
Table 3.3

Summary of FISH and marker analysis for radiation hybrids

<table>
<thead>
<tr>
<th>Hybrid</th>
<th>No. of human fragments by FISH</th>
<th>No. of markers positive</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>1A</td>
<td>1</td>
<td>1</td>
<td>Small human fragment, results correlate</td>
</tr>
<tr>
<td>2A</td>
<td>1</td>
<td>2</td>
<td>Large human fragment, reasonable correlation</td>
</tr>
<tr>
<td>2B</td>
<td>Multiple (&gt;3)</td>
<td>5</td>
<td>Markers contiguous, no correlation</td>
</tr>
<tr>
<td>3A</td>
<td>5</td>
<td>1</td>
<td>No correlation of results</td>
</tr>
<tr>
<td>3B</td>
<td>0</td>
<td>0</td>
<td>Correlation of results</td>
</tr>
<tr>
<td>3C</td>
<td>1</td>
<td>0</td>
<td>Small human fragment, reasonable correlation</td>
</tr>
<tr>
<td>4B</td>
<td>0</td>
<td>0</td>
<td>Correlation of results</td>
</tr>
<tr>
<td>5A</td>
<td>Multiple (&gt;3)</td>
<td>23</td>
<td>Large region of 9q, reasonable correlation</td>
</tr>
<tr>
<td>6A</td>
<td>1</td>
<td>2</td>
<td>Contiguous markers, correlation</td>
</tr>
<tr>
<td>6B</td>
<td>Multiple (&gt;3)</td>
<td>2</td>
<td>Contiguous markers, reasonable correlation</td>
</tr>
<tr>
<td>6C</td>
<td>Multiple (&gt;3)</td>
<td>4</td>
<td>Contiguous markers, no correlation</td>
</tr>
<tr>
<td>7A</td>
<td>Multiple (&gt;3)</td>
<td>1</td>
<td>No correlation</td>
</tr>
<tr>
<td>7B</td>
<td>Multiple (&gt;3)</td>
<td>2</td>
<td>Non contiguous markers, reasonable correlation</td>
</tr>
<tr>
<td>8A</td>
<td>0</td>
<td>0</td>
<td>Correlation of results</td>
</tr>
<tr>
<td>8B</td>
<td>1</td>
<td>0</td>
<td>Reasonable correlation</td>
</tr>
<tr>
<td>8C</td>
<td>0</td>
<td>1</td>
<td>Reasonable correlation</td>
</tr>
<tr>
<td>9A</td>
<td>Multiple (&gt;3)</td>
<td>0</td>
<td>No correlation of results</td>
</tr>
<tr>
<td>9B</td>
<td>Multiple (&gt;3)</td>
<td>2</td>
<td>Non contiguous markers, reasonable correlation</td>
</tr>
<tr>
<td>9C</td>
<td>0</td>
<td>0</td>
<td>Correlation of results</td>
</tr>
<tr>
<td>11A</td>
<td>0</td>
<td>0</td>
<td>Correlation of results</td>
</tr>
<tr>
<td>11B</td>
<td>0</td>
<td>0</td>
<td>Correlation of results</td>
</tr>
<tr>
<td>12A</td>
<td>0</td>
<td>0</td>
<td>Correlation of results</td>
</tr>
<tr>
<td>12C</td>
<td>2</td>
<td>12</td>
<td>Most markers grouped, reasonable correlation</td>
</tr>
<tr>
<td>13A</td>
<td>0</td>
<td>2</td>
<td>No correlation</td>
</tr>
<tr>
<td>13B</td>
<td>1</td>
<td>2</td>
<td>Non contiguous markers, reasonable correlation</td>
</tr>
<tr>
<td>13C</td>
<td>NT</td>
<td>1</td>
<td>No comment as FISH not tested (NT)</td>
</tr>
<tr>
<td>14A</td>
<td>3</td>
<td>5</td>
<td>Non contiguous markers good correlation</td>
</tr>
<tr>
<td>15B</td>
<td>1</td>
<td>0</td>
<td>Reasonable correlation</td>
</tr>
<tr>
<td>16B</td>
<td>0</td>
<td>0</td>
<td>Correlation of results</td>
</tr>
<tr>
<td>16C</td>
<td>0</td>
<td>0</td>
<td>Correlation of results</td>
</tr>
<tr>
<td>17A</td>
<td>1</td>
<td>4</td>
<td>Markers non contiguous, no correlation</td>
</tr>
<tr>
<td>17B</td>
<td>1</td>
<td>6</td>
<td>5 markers contiguous, good correlation</td>
</tr>
<tr>
<td>18A</td>
<td>Multiple (&gt;3)</td>
<td>4</td>
<td>Non contiguous markers, reasonable correlation</td>
</tr>
<tr>
<td>18B</td>
<td>Multiple (&gt;3)</td>
<td>1</td>
<td>No correlation</td>
</tr>
<tr>
<td>19B</td>
<td>Multiple (&gt;3)</td>
<td>5</td>
<td>Non contiguous markers good correlation</td>
</tr>
<tr>
<td>20A</td>
<td>1</td>
<td>5</td>
<td>4 markers contiguous, good correlation</td>
</tr>
<tr>
<td>20B</td>
<td>1</td>
<td>5</td>
<td>Same as 20A</td>
</tr>
<tr>
<td>21A</td>
<td>1</td>
<td>2</td>
<td>Non contiguous markers, reasonable correlation</td>
</tr>
<tr>
<td>22B</td>
<td>NT</td>
<td>0</td>
<td>No comment as FISH not tested (NT)</td>
</tr>
</tbody>
</table>

The results were defined as having reasonable correlation if there was one fragment by FISH but zero or two by markers analysis or if there was none by FISH but one by marker analysis.
markers tested from proximal chromosome 9q.

At the time of the analysis, the conservative definition of the position of the TSC1 gene was distal to D9S64 and proximal to D9S14 (Povey et al., 1992). Hybrids with markers retained in this region included 5A, 6B, 12C, 17B, 19B, 20A and 20B. Hybrid 5A was positive for 23 of the 24 markers tested and was not analysed further because the aim was for an enrichment of only the TSC1 candidate region. In addition, as hybrid cell lines 20A and 20B retained identical markers and they were both shown to contain a small, solitary human fragment by FISH they were considered to be the same and it was only 20A that was studied further. The remaining hybrids were characterised with an additional 8 markers from chromosome 9q34 within the TSC1 candidate region and a deletion map was constructed (see figure 3.2). This deletion map in combination with the results of the FISH analysis, was then used to select hybrids which could be useful resources in the identification of genomic clones.

3.1.2. Generation of radiation hybrid cosmids libraries

Hybrids 6C, 17B, 12C, 19B and 20B were all positive for markers within the TSC1 candidate interval (see figure 3.2), but hybrids 6C and 19B contained multiple human fragments and 12C contained two human fragments by FISH. These results suggested that the human chromosomal material within the hybrid cell line was fragmented. In contrast, hybrids 17B and 20A both showed a single human DNA fragment by FISH which was stably maintained in each respective cell line (see table 3.2 and figure 3.1). Together these two hybrids retained a large proportion of the candidate region and therefore genomic cosmids libraries were generated from each.
Figure 3.2

Deletion map for panel of radiation hybrids (taken from Nahmias et al., 1995)

<table>
<thead>
<tr>
<th>LOCUS</th>
<th>CENTROMERE</th>
<th>HYBRIDS</th>
</tr>
</thead>
<tbody>
<tr>
<td>D9S48</td>
<td>17A</td>
<td>19B</td>
</tr>
<tr>
<td>D9S15</td>
<td>17B</td>
<td></td>
</tr>
<tr>
<td>GSN</td>
<td></td>
<td>12C</td>
</tr>
<tr>
<td>D9S49</td>
<td></td>
<td>12C</td>
</tr>
<tr>
<td>AK1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>D9S65</td>
<td>17A</td>
<td>19B</td>
</tr>
<tr>
<td>ASS</td>
<td></td>
<td>12C</td>
</tr>
<tr>
<td>ABL</td>
<td></td>
<td></td>
</tr>
<tr>
<td>D9S113</td>
<td></td>
<td></td>
</tr>
<tr>
<td>D9S64</td>
<td></td>
<td></td>
</tr>
<tr>
<td>D9S125</td>
<td></td>
<td>19B</td>
</tr>
<tr>
<td>D9S149</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ABO</td>
<td></td>
<td></td>
</tr>
<tr>
<td>D9S150</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DBH</td>
<td>6C</td>
<td>12C</td>
</tr>
<tr>
<td>D9S122</td>
<td></td>
<td></td>
</tr>
<tr>
<td>D9S10</td>
<td>17B</td>
<td></td>
</tr>
<tr>
<td>D9S114</td>
<td></td>
<td></td>
</tr>
<tr>
<td>D9S298</td>
<td></td>
<td>19B</td>
</tr>
<tr>
<td>D9S14</td>
<td></td>
<td></td>
</tr>
<tr>
<td>D9S67</td>
<td></td>
<td></td>
</tr>
<tr>
<td>D9S158</td>
<td></td>
<td></td>
</tr>
<tr>
<td>D9S17</td>
<td>12C</td>
<td></td>
</tr>
<tr>
<td>D9S7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>D9S11</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

TELOMERE
The libraries were prepared as described in section 2.2.3. Approximately 20,000 clones from each library were screened for the retention of human DNA inserts. Clones that hybridised to the radioactively labelled human DNA probe were picked and purified. A total of 4 cosmids were identified from each hybrid cell line and these were then mapped by FISH onto normal human metaphase spreads (examples are shown in figure 3.3). The four positive clones from hybrid 17B (cKW1B, cKW10, cKW3 and cKW1) and two from 20A (cJF20A1 and cJF20A3) were confirmed to be localised to human chromosome 9q34. An additional two clones from hybrid 20A (cJF20A5 and cJF20A7) were found to map to the centromere of chromosome 9. This result suggested that hybrid 20A contained human centromeric sequences and correlated with the human DNA being present as a solitary fragment which was stably maintained.

As such a small number of human cosmids were identified from each library, it was decided that instead of rescreening greater numbers of clones, a different approach should be used. At the same time Dr. J. Fitzgibbon was using Alu-PCR products from the radiation hybrids to screen human genomic libraries and as this technique had been successful, the cosmid libraries were not pursued any further.
Figure 3.3

FISH mapping cosmids generated from 17B and 20A radiation hybrid libraries, onto normal human metaphase chromosomes

A) cKW1
Isolated from hybrid 17B and mapping to 9q34

B) cKW1B
Isolated from hybrid 17B and mapping to 9q34

C) cJF20A1
Isolated from hybrid 20A and mapping to 9q34

D) cJF20A5
Isolated from hybrid 20A and mapping to the centromere of chromosome 9
3.1.3. Isolation of cosmids using hybrid *Alu*-PCR

*Alu*-PCR is a method of amplifying human specific sequences between *Alu* repeats, a PCR fingerprint pattern is generated which represents the human DNA content within a particular DNA mixture. Two oligonucleotides called *Alu* IV (Cotter *et al.*, 1991) and 5'R' were used which prime from the 3' and 5' end of *Alu* sequences respectively.

Initially *Alu*-PCR products were generated from hybrid 17B using primer *Alu* IV and these were then used to screen a human cosmid library (Fitzgibbon, 1993). Four cosmids were isolated which were mapped by FISH to human chromosome 9q34 (figure 3.4). In addition, the *Alu*-PCR products were biotin labelled and used as a reverse paint onto human metaphase chromosomes by FISH (L. West, personnel communication). The signal was only present on the distal arm of chromosome 9q and was consistent with the marker retention studies (Fitzgibbon, 1993).

As the *Alu*-PCR fingerprint of each hybrid showed preferential amplification of some *Alu* sequences (data not shown), further analysis used hybrids 17B, 20A and 19B, in an effort to obtain products from the entire TSC1 candidate region (see figure 3.2). In addition, both 5' and 3' primers were used to give a greater yield of products. These products were used to screen the Lawrence Livermore flow-sorted chromosome 9 gridded cosmid library and 1,431 different clones were identified (Nahmias *et al.*, 1995). As expected, because of the overlap in human DNA material between each hybrid cell line, many clones were identified by two or more different hybrid *Alu*-PCR products.
FISH mapping cosmids isolated from genomic libraries using *Alu*-PCR probes, onto normal human metaphase chromosomes

A) **cJF17BE**
Isolated by using *Alu*-PCR product from hybrid 17B and mapping to chromosome 9q34

B) **cJF17B2**
Isolated by using *Alu*-PCR product from hybrid 17B and mapping to chromosome 9q34

C) **cJF17BA**
Isolated by using *Alu*-PCR product from hybrid 17B and mapping to chromosome 9q34

D) **cJF17BF**
Isolated by using *Alu*-PCR product from hybrid 17B and mapping to chromosome 9q34
The efficacy of the selection strategy was analysed by pooling 96 individually grown clones from each hybrid Alu-PCR screen and then using DNA made from the pool as a probe onto normal human metaphase chromosomes. The results are shown in figure 3.5. The pool of randomly selected cosmid clones from the chromosome 9 library was a control in the experiment and this showed a hybridisation signal along the entire length of human chromosome 9. In contrast, all the hybrid Alu-PCR cosmid pools showed a main signal on human chromosome 9q34.

These clones plus ones from other sources have been fingerprinted and assembled in contigs (Nahmias et al., 1995). The other clones were derived from cosmid libraries generated from hybrid cell lines 17B and 20A (section 3.1.2), a female genomic library (Cachon-Gonzalez, 1991), a library constructed from somatic cell hybrid CJ9q (Florian, unpublished), a library constructed from somatic cell hybrid E6B (Henske et al., 1992) and additional cosmids that had previously been mapped by FISH to chromosome 9q34 (Graw et al., 1992; Takahashi et al., 1994). To date, 1894 clones have been fingerprinted giving rise to 172 contigs which contain a total of 1,116 cosmids. The next task was to localise contigs by FISH mapping at least one clone from each onto human metaphase chromosomes.
Figure 3.5

FISH of pooled Alu-PCR derived cosmids onto normal human metaphase chromosomes

A) Hybrid 17B
5' Alu-PCR primer used.
Signal on 9q34

B) Hybrid 19B
5' Alu-PCR primer used. Signal on 9q34

C) Hybrid 20A
5' Alu-PCR primer used.
Signal on 9q33-q34, 9q21-22 and 9p12

D) Control
Random cosmids picked from the same library
3.2. Generation of a FISH map for distal chromosome 9q

In addition to Alu-PCR products many probes corresponding to genes or mapped loci were used to screen the cosmid libraries (Nahmias et al., 1995). The clones isolated were mapped onto human metaphase chromosomes and examples are shown in figure 3.6. If a gene was mapped by this technique to human chromosome 9q34 it could, by localisation alone, be a candidate for the tuberous sclerosis gene. Therefore it became important to map clones more precisely than simply a cytogenetic R-band.

Metaphase chromosomes from cell lines carrying a translocation enable clones to be positioned above or below the breakpoint by FISH. Three lymphoblastoid cell lines with balanced reciprocal translocations involving human chromosome 9q34 (see table 2.1 for karyotypes) were used to localise clones to one of four intervals on the chromosome (figure 3.7). The cell line SD-1 has the characteristic Philadelphia translocation with a breakpoint in the Abelson proto-oncogene on chromosome 9q34.1 (Dhut et al., 1991). Cell lines 9T12 and 9T01 were partially characterised by PCR marker analysis on flow-sorted chromosomes to have breakpoints between DBH and RXRA and distal to RXRA respectively (Zhou et al., 1992). By testing more loci the localisations of these breakpoints have been refined to between D9S114 and D9S298 and to between D9S740 and D9S207 respectively (Woodward et al., 1995).
Figure 3.6

FISH mapping cosmids encoding genes onto normal human metaphase chromosomes.

A) PBX3

B) DBH

C) RXRA

D) NOTCH1
Figure 3.7

Schematic representation of the four intervals used to position cosmids by FISH on human chromosome 9q34

NOTE: Intervals are not to scale
Examples of clones mapping to each of the four intervals on human chromosome 9q34 are shown in figures 3.8-3.11. A total of 18 genes (table 3.2) and 14 DNA markers (table 3.3) have been regionally localised using this approach and a FISH map of distal chromosome 9q was constructed (figure 3.12). The order of loci was taken from a recent consensus map of chromosome 9q34 (Povey et al., 1994b) but it is approximate as some loci have a range of map positions. All the cosmids encoding genes apart from Abelson proto-oncogene (ABL) and collagen type V alpha 1 (COL5A1) and all the markers except D9S740 and D9S207 form part of contigs (Nahmias et al., 1995) and a different contig has been shown on each line. An additional 43 cosmid contigs have been mapped but they are as yet unassigned to marker loci. These are represented in figure 3.12 by the column headed anonymous contigs. One clone from each contig has been shown for simplicity, but in most cases more than one was actually mapped. A total of 64 contigs were mapped, 26 were localised to interval 1, 20 to interval 2, 9 to interval 3 and 9 to interval 4.

The majority of clones mapped were from the Lawrence Livermore library and coordinate numbers are given. Clones with a prefix cKW were derived from a hybrid 17B genomic library (see section 3.1.2. and table 2.2) and clones with a cFF prefix were derived from a CJ9q cosmid library (Florian, unpublished).
Figure 3.8

Cosmid clones mapping to interval 1, proximal to SD-1 cell line translocation breakpoint

A) FPGS cosmid
SD-1 cell line metaphase chromosomes. Small arrow points to the normal chromosome 9 and large arrow points to the derivative chromosome 9.

B) ASS cosmid
SD-1 cell line metaphase chromosomes. Small arrow points to the normal chromosome 9 and large arrow points to the derivative chromosome 9.
Figure 3.9

Cosmid clone mapping to interval 2, distal to SD-1 and proximal to 9T12 cell line translocation breakpoint

A) SURF1 cosmid

SD-1 cell line metaphase chromosomes. Small arrow points to the normal chromosome 9 and large arrow points to the derivative chromosome 22.

B) SURF1 cosmid

9T12 cell line metaphase chromosomes. Small arrow points to the normal chromosome 9 and large arrow points to the derivative chromosome 9.
Figure 3.10

Cosmid clone mapping to interval 3, distal to 9T12 and proximal to 9T01 cell line translocation breakpoint

A) LL09 242 C4 cosmid
9T12 cell line metaphase chromosomes. Small arrow points to the normal chromosome 9 and large arrow points to the derivative chromosome 20.

B) LL09 242 C4 cosmid
9T01 cell line metaphase chromosomes. Small arrow points to the normal chromosome 9 and large arrow points to the derivative chromosome 9.
Figure 3.11

Cosmid clones mapping to interval 4, distal to 9T01 cell line translocation breakpoint

A) NOTCH1 cosmid
9T01 cell line metaphase chromosomes. Small arrow points to the normal chromosome 9 and large arrow points to the derivative chromosome 15.

B) LL09 9 F10 cosmid
9T01 cell line metaphase chromosomes. Small arrow points to the normal chromosome 9 and large arrow points to the derivative chromosome 15.
Table 3.4

Eighteen genes mapped by FISH with their clone name and previously reported map location.

<table>
<thead>
<tr>
<th>NAME</th>
<th>LOCUS</th>
<th>PREVIOUS LOC. N</th>
<th>MAP REFERENCE</th>
<th>PROBE SOURCE</th>
<th>COS-MID</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abelson proto-oncogene</td>
<td>ABL</td>
<td>9q34.1</td>
<td>Heisterkamp et al., 1982</td>
<td>Dr. Harris*</td>
<td>ABL</td>
</tr>
<tr>
<td>ABO blood group</td>
<td>ABO</td>
<td>9q34.1-q34.2</td>
<td>Cook et al., 1978</td>
<td>Dr. Yamamoto</td>
<td>ABO17</td>
</tr>
<tr>
<td>Argininosuccinate synthetase 1</td>
<td>ASS</td>
<td>9q34.1</td>
<td>Carritt et al., 1977</td>
<td>Dr. Harris*</td>
<td>ASS</td>
</tr>
<tr>
<td>CAN D9S46E</td>
<td>CAN</td>
<td>9q34</td>
<td>von Lindern et al., 1992</td>
<td>RT-PCR</td>
<td>31 G2</td>
</tr>
<tr>
<td>Complement component 8 gamma peptide</td>
<td>C8G</td>
<td>9q22.3-q32</td>
<td>Yuille et al., 1992</td>
<td>RT-PCR</td>
<td>10 E7</td>
</tr>
<tr>
<td>Carboxyl ester lipase</td>
<td>CEL</td>
<td>9q34.3</td>
<td>Taylor et al., 1992</td>
<td>RT-PCR</td>
<td>196 G6</td>
</tr>
<tr>
<td>Collagen type V alpha 1</td>
<td>COL5A1</td>
<td>9q34.2-q34.3</td>
<td>Greenspan et al., 1992</td>
<td>RT-PCR</td>
<td>87 F11</td>
</tr>
<tr>
<td>Dopamine beta-hydroxylase</td>
<td>DBH</td>
<td>9q34</td>
<td>Craig et al., 1988</td>
<td>Dr. Mallet</td>
<td>DBH15</td>
</tr>
<tr>
<td>Endoglin</td>
<td>ENG</td>
<td>9q34-qter</td>
<td>Fernandez-Ruiz et al., 1993</td>
<td>RT-PCR</td>
<td>53 A11</td>
</tr>
<tr>
<td>Foly polyglutamate synthetase</td>
<td>FPGS</td>
<td>9cen-9q34</td>
<td>Jones and Kao, 1984</td>
<td>Dr. Shane</td>
<td>220 F4</td>
</tr>
<tr>
<td>N-methyl D-aspartate receptor</td>
<td>GRIN1</td>
<td>9q34.3</td>
<td>Brett et al., 1994</td>
<td>Dr. Brett</td>
<td>144 E7</td>
</tr>
<tr>
<td>Pre-B-cell leukemia transcription factor 3</td>
<td>PBX3</td>
<td>9q33-9q34</td>
<td>Monica et al., 1991</td>
<td>Dr. Bech-Hansen</td>
<td>25 G7</td>
</tr>
<tr>
<td>Pregnancy-associated plasma protein-A</td>
<td>PAPPA</td>
<td>9q33.1</td>
<td>Sillahtaroglu et al., 1993</td>
<td>RT-PCR</td>
<td>88 H1</td>
</tr>
<tr>
<td>Progestagen-associated endometrial protein</td>
<td>PAEP</td>
<td>9q34</td>
<td>Van Cong et al., 1991</td>
<td>RT-PCR</td>
<td>57 G1</td>
</tr>
<tr>
<td>Retinoid X receptor alpha</td>
<td>RXRA</td>
<td>9q34</td>
<td>Jones et al., 1993</td>
<td>PCR</td>
<td>RXR4</td>
</tr>
<tr>
<td>Surfeit 1</td>
<td>SURF1</td>
<td>9q34.1</td>
<td>Yon et al., 1993</td>
<td>RT-PCR</td>
<td>39 C8</td>
</tr>
<tr>
<td>Homologue of Drosophila Notch</td>
<td>NOTCH</td>
<td>9q34.3</td>
<td>Ellisen et al., 1991</td>
<td>A.T.C.C.</td>
<td>285 A4</td>
</tr>
<tr>
<td>VAV like gene</td>
<td>&quot;VAV2&quot;</td>
<td>9q34.3</td>
<td>Woodward et al., 1994</td>
<td>Trapped exon</td>
<td>17BE</td>
</tr>
</tbody>
</table>

* Probes supplied directly as cosmids, other probes were used to screen libraries.
<table>
<thead>
<tr>
<th>LOCUS</th>
<th>PREVIOUS LOCATION</th>
<th>MAP REFERENCE</th>
<th>COSMID NAME</th>
</tr>
</thead>
<tbody>
<tr>
<td>D9S61</td>
<td>9q</td>
<td>Kwiatkowski et al., 1992</td>
<td>dk 1AD7</td>
</tr>
<tr>
<td>D9S62</td>
<td>9q</td>
<td>Kwiatkowski et al., 1992</td>
<td>dk 1AG5</td>
</tr>
<tr>
<td>D9S113</td>
<td>9q34.1-qter</td>
<td>Henske et al., 1993</td>
<td>5B1</td>
</tr>
<tr>
<td>D9S148</td>
<td>9q34.1-qter</td>
<td>Henske et al., 1993</td>
<td>C10</td>
</tr>
<tr>
<td>D9S64</td>
<td>9q34-qter</td>
<td>Kwiatkowski et al., 1992</td>
<td>dk 1AC6</td>
</tr>
<tr>
<td>D9S149</td>
<td>9q34.1-qter</td>
<td>Henske et al., 1993</td>
<td>D3</td>
</tr>
<tr>
<td>D9S150</td>
<td>9q34.1-qter</td>
<td>Henske et al., 1993</td>
<td>dk 1A1</td>
</tr>
<tr>
<td>D9S66</td>
<td>9q34-qter</td>
<td>Kwiatkowski et al., 1992</td>
<td>dk 4AF10</td>
</tr>
<tr>
<td>D9S10</td>
<td>9q34.3</td>
<td>Lathrop et al., 1988</td>
<td>MCT136-11</td>
</tr>
<tr>
<td>D9S114</td>
<td>9q34.1-qter</td>
<td>Henske et al., 1993</td>
<td>17BA</td>
</tr>
<tr>
<td>D9S298</td>
<td>9q34</td>
<td>Attwood et al., 1993</td>
<td>17BF</td>
</tr>
<tr>
<td></td>
<td>(F2-5)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>D9S67</td>
<td>9q34-qter</td>
<td>Kwiatkowski et al., 1992</td>
<td>dk 3AE11</td>
</tr>
<tr>
<td>D9S740</td>
<td>9q34</td>
<td>Armour et al., 1994</td>
<td>G2</td>
</tr>
<tr>
<td>D9S207</td>
<td>9q34.3</td>
<td>Woodward et al., 1994</td>
<td>C207</td>
</tr>
</tbody>
</table>
Figure 3.12

FISH map of cosmids localised to distal chromosome 9q

<table>
<thead>
<tr>
<th>CENTROMERE</th>
<th>GENES</th>
<th>MARKERS</th>
<th>ANONYMOUS CONTIGS</th>
</tr>
</thead>
<tbody>
<tr>
<td>9q33.1</td>
<td>PAPPA</td>
<td>D9S61</td>
<td></td>
</tr>
<tr>
<td>9q33-34.1</td>
<td>FPGS/ENG</td>
<td>D9S62</td>
<td></td>
</tr>
<tr>
<td></td>
<td>PBX3</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>ASS</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SD-1</td>
<td></td>
<td>207 H11</td>
<td></td>
</tr>
<tr>
<td>BREAKPOINT</td>
<td>CAN</td>
<td>CFP1/9A</td>
<td></td>
</tr>
<tr>
<td>9q34.1</td>
<td></td>
<td>23 D10</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1 D6</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>273 A1</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>220 F5</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>23 A4</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>119 C5</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>201 H3</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>33 D5</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>283 E12</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>281 H2</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>237 B10</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>52 A12</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>60 C11</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>102 E9</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>175 F6</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>267 A2</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>287 F3</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>213 E7</td>
<td></td>
</tr>
<tr>
<td>9T12</td>
<td></td>
<td>235 D4</td>
<td></td>
</tr>
<tr>
<td>BREAKPOINT</td>
<td>ABO/SURF1</td>
<td>D9S113/D9S148</td>
<td></td>
</tr>
<tr>
<td>9q34.3</td>
<td></td>
<td>180 F1</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>18 C2</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>CKW1B</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>73 H8</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>223 C7</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>36 B8</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>115 H6</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>84 E3</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>256 C3</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>192 G7</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>160 G6</td>
<td></td>
</tr>
<tr>
<td>9T12</td>
<td>RXRA</td>
<td>D9S298</td>
<td></td>
</tr>
<tr>
<td>BREAKPOINT</td>
<td>COL5A1</td>
<td>D9S67</td>
<td></td>
</tr>
<tr>
<td>9q34.3</td>
<td></td>
<td>D9S740</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>22 B1</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>227 B4</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>269 C10</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>242 C4</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>CKW3</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>CKW10</td>
<td></td>
</tr>
<tr>
<td>9T12</td>
<td>PAEP</td>
<td>D9S207</td>
<td></td>
</tr>
<tr>
<td>BREAKPOINT</td>
<td>NOTCH1</td>
<td>CFP2/2H</td>
<td></td>
</tr>
<tr>
<td>9q34.3</td>
<td>GRIN1</td>
<td>140 F1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CIG</td>
<td>267 D1</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>117 H2</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>226 A5</td>
<td></td>
</tr>
</tbody>
</table>

TELOMERE
Clones mapping proximal to SD-1 were further positioned cytogenetically according to R-bands. Clones encoding pregnancy associated plasma protein-A (PAPPA) mapped to band 9q33.1 and clones encoding folylpolyglutamate synthetase (FPGS) (see figure 3.8.), endoglin (ENG) and pre-B-cell leukaemia transcription factor 3 (PBX3) mapped to band 9q34.1.

The most recent flanking markers for tuberous sclerosis have been defined as D9S149 and D9S114 (Povey et al., 1994b) and both of these markers were localised to interval 2, between the SD-1 and 9T12 translocation breakpoints. This enabled contigs to be rapidly mapped and those within the same interval as the TSC1 candidate region were easily identified for further analysis. A total of 20 contigs were mapped to interval 2 and six of these were subsequently used for the isolation of transcribed sequences by exon amplification (see section 3.3.2).

The translocation cell lines were also used to map pools of cosmids isolated using hybrid Alu-PCR products as probe. Cosmids in common between hybrid cell lines (17B, 19B and 20A) were grown individually, pooled and the DNA prepared was mapped by FISH onto metaphase chromosomes from the 9T12 translocation cell line. Results are shown in figure 3.13. Cosmids isolated with hybrid 17B and 20A Alu-PCR were mainly positioned above the 9T12 breakpoint but there was a small hybridisation signal beneath. In contrast, cosmids in common between hybrids 19B and 20A were found to be equally distributed above and below the 9T12 breakpoint on chromosome 9q34. A similar result was found for those selected with hybrid 17B and 19B. Cosmids selected by all three hybrids 17B, 19B and 20A were found to map only above the 9T12 breakpoint.
FISH of pooled Alu-PCR derived cosmids onto metaphase chromosomes from 9T12 translocation cell line

A) Hybrids 17B & 20A
Cosmids identified with Alu-PCR products from both hybrids. Signal on 9 & both derivative chromosomes, mainly above the breakpoint

B) Hybrids 17B & 19B
Cosmids identified with Alu-PCR products from both hybrids. Signal on 9 and both derivative chromosomes

C) Hybrids 19B & 20A
Cosmids identified with Alu-PCR products from both hybrids. Signal on 9 & both derivative chromosomes

D) Hybrids 17B, 19B & 20A
Cosmids identified with Alu-PCR products from all hybrids. Signal on 9 and derivative chromosome 9
3.3. Isolation of transcribed sequences from chromosome 9q34

Genomic clones from human chromosome 9q34 were used to isolated transcribed sequences by exon amplification (Buckler et al., 1991). The immediate goal was to select clones that map within the TSC1 candidate region and thereby identify candidates for the tuberous sclerosis gene. However, all coding sequences identified would facilitate the construction of a transcription map for chromosome 9q34 that may provide insight into the organisation and function of this region of the genome. Initially mammalian expression vector pSPL1 was used for a small scale analysis using eight cosmids. Modification of the vector by Church et al. (1994) gave pSPL3 which was used for a more thorough investigation of 6 contigs from the region of interest.

3.3.1. Exon amplification using pSPL1

The eight clones analysed included cosmid DBH14 as a positive control as it was known to contain expressed sequences from the dopamine beta hydroxylase gene. The other clones included cKW1B, cKW1 and cKW10 which were generated from a 17B radiation hybrid genomic library (section 3.1.2) and were confirmed to map to chromosome 9q34 by FISH (see figure 3.3), 17B2 and 17BE which were isolated by screening a genomic library with Alu-PCR products from hybrid 17B (Fitzgibbon, 1993) and were also mapped to 9q34 by FISH (see figure 3.4) and finally clones MCT136-7 and MCT136-11 which were isolated by screening a genomic library with the MCT136 probe that contains marker D9S10 (Wolfe, unpublished). This marker was known to map to chromosome 9q34 within the TSC1 candidate region (Povey et al., 1992). At the
time these experiments were initiated the translocation cell lines were not available for
FISH mapping.

Three restriction enzyme digests were performed for each cosmid (BamHI, BglII and
BamHI plus BglIII) and the DNA fragments generated from each were shot-gun cloned
into the prepared vector. Each enzyme digest was treated separately and transfected into
COS-7 cells. Cytoplasmic RNA was extracted 48-72 hours later and exons were isolated
from the vector derived RNA sequences by RT-PCR using β-globin specific primers
(schematic representation shown in figure 1.3). RT-PCR analysis of pSPL1 vector
transfectants gave the expected product size of 429bp when primers SA2 and SD2 were
used. RT-PCR analysis of transfectants with shot-gun cloned cosmids yielded the same
429bp fragment but also larger products in at least one of the different enzyme digests
used. Examples are shown in figure 3.14 and lane 4 in each case shows additional larger
bands which are stained more weakly but suggest that an exon has been amplified from
the BamHI digested shot-gun clone. The 429bp product is always the most intense
because the predominant vector derived mature RNA message is generated by a vector
to vector splicing event (figure 3.15). Only when a cloned genomic fragment contains a
complete exon, in the correct orientation and both the splice sites present are functional,
will an exon be trapped. When all these criteria are satisfied the vector 5’ donor splice
site pairs with the genomic 3’ acceptor splice site and the genomic 5’ donor splice site
pairs with the vector 3’ acceptor splice site giving rise to an amplified exon (figure
3.16). These products are larger than the vector product by the size of the exon(s)
trapped. PCR bias in favour of the smaller and more abundant substrates is an additional
reason why the larger amplification products appear fainter on an agarose gel and this
might also explain the prominent bands smaller than 429bp, which are shown in a
Figure 3.14

RT-PCR analysis using vector pSPL1

A. Cosmid 17BE

B. Cosmid 17B2

C. Cosmid DBH14

D. Cosmid cKW1B

Lanes: 1. 1kb ladder, sizes shown in base pairs; 2. mock transfection, no DNA control; 3. pSPL1 vector only control transfection; 4. transfection of BamHI shot-gun clones; 5. transfection of BglII shot-gun clones; 6. transfection of BamHI/BglII shot-gun clones. Product migrating at 429bp is derived from vector only splicing. Products larger than 429bp in lanes 4-6 suggest an exon has been amplified.
Figure 3.15

Vector to vector splicing event

This occurs when there is no exon sequence cloned, when an exon is cloned but it is in the anti-sense orientation or when splice sites present are non-functional.

5' SSv - vector splice donor site
3' SSv - vector splice acceptor site

Shaded box at the end represents beta-globin sequence. Primers for RT-PCR are within this sequence. Striped box represents HIV-1 tat exon sequence.
Figure 3.16

Vector to genomic splicing event that gives rise to an amplified exon

This occurs when an entire exon with functional flanking splice sites is inserted into the vector in the 5' to 3' orientation.

- 5' SSv - vector splice donor site
- 3' SSv - vector splice acceptor site
- 5' SSg - genomic splice donor site
- 3' SSg - genomic splice acceptor site

Shaded box at the end represents beta-globin sequence. Primers for RT-PCR are within this sequence. Striped box represents HIV-1 tat exon sequence.
number of lanes in figure 3.14. They are particularly obvious in figure 3.14.C and even occur in lane 2 which is the mock transfection with no DNA. Therefore they are thought to be artefacts arising by amplification from the COS-7 cell background.

3.3.1.1. Hybridisation analysis

Initial analysis concentrated on the putative exons isolated from cosmids DBH14, cKW1B and 17BE. Individual products from the BamHI digested shot-gun clones which were larger 429bp, were excised from the agarose gel, purified and radio-labelleed to use as a probe on Southern blots containing the corresponding BamHI digested cosmid clone. Figure 3.17 shows that as expected the putative exons only hybridised to the cosmid from which they were derived. However probes from cKW1B and 17BE gave more than one hybridisation signal which suggested that the product was mixed. To try and purify the product it was separated by electrophoresis on an agarose gel, the band was stabbed with a plastic tip and then this was used as the source of DNA for a second PCR reaction using a pair of internal primers (SA1 and SD1) which flank the splice sites. The individual bands were excised, purified and the experiment repeated. Unfortunately the same result was obtained, a main hybridisation signal but additional fainter ones which could be detected on the autoradiograph after a longer exposure.

Putative exons from 17BE and cKW1B were also used as hybridisation probes on zooblots to detect cross-species conservation and confirm that they were a true exon. The largest product from cKW1B was only conserved in primates (figure 3.18A) and gave a smear in hamster, rat and mouse. The human BamHI fragment which hybridised was 6kb in size and this corresponded to the cosmid hybridisation result (figure 3.17B),
Figure 3.17

Hybridisation of putative exons to their corresponding genomic clones

A. DBH14 probe

B. cKW1B probe

C. 17BE probe

Southern blots of BamHI digested cosmid clones. Lanes: 1. DBH14; 2. MCT136-7; 3. cKW1B; 4. cKW1; 5. cKW10; 6. 17BE. Sizes of hybridisation signals given in kb.
Hybridisation of putative exons to zooblots

Genomic DNA digested with BamHI restriction enzyme and Southern blotted.
Lanes: 1 human; 2 orang-utan; 3 African green monkey; 4 zebra; 5 dog; 6 vole; 7 rat; 8 hamster; 9 mouse. Sizes for sites of hybridisation are given in kb.
although then it was the 11kb fragment which showed the most intense signal. As this larger fragment was not detected on the zooblot it was thought to represent an artefact. In contrast, the 17BE product was definitely conserved in primates, zebra, dog, vole, rat, hamster and mouse (figure 3.18B). The size of the human BamHI fragment was 1.2kb which correlated with the most intense signal on the cosmid Southern blot (figure 3.17C). The results suggested that the product isolated from cosmid 17BE was a true exon.

3.3.1.2. Chimerism

As a result of the problems in trying to purify individual PCR products, efforts were made to clone them. The internal oligonucleotide primers SA1 and SD1 which flank the splice sites, contain restriction sites SalI and BamHI respectively at their 5’ end for directional cloning. However when the PCR products from both DBH14 and cKW1B were digested with BamHI for cloning into a plasmid vector, they were cleaved into two fragments. This was very unexpected and suggested that maybe a cryptic splicing event had taken place. The cKW1B PCR product was sequenced directly and then compared to databases using the programmes BLASTN, BLASTX and FASTA. The results showed that the product was chimeric with HIV-1 tat vector sequence at the 3’ end corresponding to residues 6343-6413 of the HIV-1 envelope protein. The 5’ end displayed no significant sequence similarity. A similarity was only considered to be significant to a nucleotide database entry if is was ≥55bp in length, ≥ 60% similarity, p value ≤0.01 or to a protein database entry if it was ≥24 amino-acids in length, ≥55% similarity, p value ≤0.01 (Khan et al., 1992). The sequence of the cKW1B product is shown in figure 3.19.
Figure 3.19

Sequence analysis of the chimeric cKW1B PCR product

SD1

GGATCCGCGACGAAGACCTCCTCAAGGCGACTCATAAAGTTTCTCT

HIV-1 tat exon cKW1B

ATCAAAGCAAAAACGTAGGGCTCGGGGAGAAGGAAGACAGTACGGGCCCAAG

GGCCACCCCCGCAACGACATGACATCCACCCCACAGCCCATGGGTAGAG

HIV-1 to/intron

TCCTGCCCAGCTCCGCTCAGAGAATTCGGGAGCTGCACGGTGAG

HIV-1 tat intron

ACCTAGAGTCAAAGTGGCTCCCCGTGTGGGAGAAGGAAGCACCACCACCCT

Ndel

ATTTTGTCATCAGATGCTAAAGCATATGATACAGAGGCTGACGAGAG

HIV-1 tat exon

CCCAATCCCCAGGGGACCCCGACGACG

SA1

Sequence is shown 5’ to 3’. HIV-1 and cKW1B sequences are indicated by arrows.

Oligonucleotide primers SA1 and SD1 were used for the PCR reaction and for sequencing.
Chimeric products are produced when cryptic splice sites are used for RNA processing. A cryptic HIV-1 \textit{tat} splice donor site is found 66bp downstream of the \textit{Bam}HI cloning site and can be activated if the cloned genomic fragment contains a cryptic 3’ splice acceptor site or if a partial exon is cloned which contains a real 3’ splice acceptor site (Figure 3.20). Both situations can be detected by digestion with the restriction enzyme \textit{Ndel}. A restriction site is present downstream of the \textit{Bam}HI cloning site but upstream of the cryptic HIV-1 \textit{tat} splice donor site and therefore gives a constant sized 3’ DNA fragment. As the recognition sequence is infrequent in genomic DNA it is unlikely that an amplified exon will contain a site. However if it does, the probability of the site being positioned the correct distance from the PCR product termini to give the expected size fragment is very low.

3.3.1.3. Sequence analysis

A total of 20 products were identified in the analysis with average length of 119bp and a size range from 64-230bp. Nine of the products (45%) contained a recognition site for \textit{Ndel} that suggested they were chimeric and therefore they were eliminated from further analysis. This included the product from cosmid DBH14 that was supposed to be the positive control. However this high level of chimerism is consistent with results from other groups (Church \textit{et al}., 1993; North \textit{et al}., 1993). The remaining 11 PCR products were directly sequenced using the di-deoxy chain termination method (Sanger \textit{et al}., 1977) and were then used in BLASTN, BLASTX and FASTA database searches. The sequences are shown in appendix 3.1 and the summary of the results are shown in table 3.6.
This occurs when the cloned genomic DNA contains a cryptic splice acceptor site or when an exon is interrupted by the restriction enzyme digest used for cloning. The absence of a genomic splice donor site is compensated by activation of the vector cryptic splice site. It can be recognised by the presence of an *NdeI* restriction site.

5' SSv - vector splice donor site
3' SSv - vector splice acceptor site
5' SSc - cryptic splice donor site
3' SSg - genomic splice acceptor site

Shaded box at the end represents beta-globin sequence. Primers for RT-PCR are within this sequence. Striped box represents HIV-1 *tat* exon sequence.
Table 3.6

Summary results of exons isolated using pSPL1

<table>
<thead>
<tr>
<th>COSMID</th>
<th>ENY DIG</th>
<th>APPR LTH</th>
<th>Ndel SITE</th>
<th>HYB.N INFO.</th>
<th>SEQUENCE ANALYSIS</th>
<th>COMMENTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>cKW1B</td>
<td>BamHI</td>
<td>230</td>
<td>YES</td>
<td>S,SCP</td>
<td>5'NH, 3'HIV</td>
<td>CHIMERIC</td>
</tr>
<tr>
<td>cKW1B</td>
<td>BamHI</td>
<td>180</td>
<td>YES</td>
<td>S</td>
<td></td>
<td>CHIMERIC</td>
</tr>
<tr>
<td>cKW1B</td>
<td>BamHI</td>
<td>100</td>
<td>YES</td>
<td>NS</td>
<td></td>
<td>CHIMERIC</td>
</tr>
<tr>
<td>DBH14</td>
<td>BamHI</td>
<td>100</td>
<td>YES</td>
<td>S</td>
<td></td>
<td>CHIMERIC</td>
</tr>
<tr>
<td>17BE</td>
<td>BamHI</td>
<td>101</td>
<td>NO</td>
<td>S, C</td>
<td></td>
<td></td>
</tr>
<tr>
<td>17B2</td>
<td>BamHI</td>
<td>117</td>
<td>NO</td>
<td>NS</td>
<td></td>
<td>ARTEFACT</td>
</tr>
<tr>
<td>17B2</td>
<td>BglII</td>
<td>100</td>
<td>YES</td>
<td></td>
<td></td>
<td>CHIMERIC</td>
</tr>
<tr>
<td>17B2</td>
<td>BglII</td>
<td>74</td>
<td>NO</td>
<td>NH</td>
<td>NOVEL BY SEQ.</td>
<td></td>
</tr>
<tr>
<td>17B2</td>
<td>B/B</td>
<td>120</td>
<td>NO</td>
<td>NS</td>
<td></td>
<td>ARTEFACT</td>
</tr>
<tr>
<td>cKW10</td>
<td>BamHI</td>
<td>150</td>
<td>YES</td>
<td></td>
<td></td>
<td>CHIMERIC</td>
</tr>
<tr>
<td>cKW10</td>
<td>BamHI</td>
<td>85</td>
<td>NO</td>
<td>NH</td>
<td>NOVEL BY SEQ.</td>
<td></td>
</tr>
<tr>
<td>cKW10</td>
<td>B/B</td>
<td>120</td>
<td>YES</td>
<td></td>
<td></td>
<td>CHIMERIC</td>
</tr>
<tr>
<td>cKW1</td>
<td>BamHI</td>
<td>129</td>
<td>NO</td>
<td>S, C</td>
<td>NH</td>
<td>NOVEL PRODUCT</td>
</tr>
<tr>
<td>MCT136-7</td>
<td>BamHI</td>
<td>200</td>
<td>NO</td>
<td>NH</td>
<td>NOVEL BY SEQ.</td>
<td></td>
</tr>
<tr>
<td>MCT136-7</td>
<td>BamHI</td>
<td>80</td>
<td>YES</td>
<td></td>
<td></td>
<td>CHIMERIC</td>
</tr>
<tr>
<td>MCT136-7</td>
<td>BamHI</td>
<td>64</td>
<td>NO</td>
<td>NH</td>
<td>NOVEL BY SEQ.</td>
<td></td>
</tr>
<tr>
<td>MCT136-7</td>
<td>B/B</td>
<td>87</td>
<td>NO</td>
<td>2S</td>
<td>NH</td>
<td>NOVEL BY SEQ.</td>
</tr>
<tr>
<td>MCT136-11</td>
<td>BamHI</td>
<td>200</td>
<td>NO</td>
<td>NH</td>
<td>NOVEL BY SEQ.</td>
<td></td>
</tr>
<tr>
<td>MCT136-11</td>
<td>BamHI</td>
<td>80</td>
<td>YES</td>
<td></td>
<td></td>
<td>CHIMERIC</td>
</tr>
<tr>
<td>MCT136-11</td>
<td>BamHI</td>
<td>64</td>
<td>NO</td>
<td>NH</td>
<td>NOVEL BY SEQ.</td>
<td></td>
</tr>
</tbody>
</table>

ENY DIG - enzyme digest. B/B - BamHI and BglII
APPR LTH - approximate product length
HYB.N INFO.- Hybridisation information
S - exon hybridises specifically to cosmid analysed, 2S - to two cosmids or NS - nonspecifically to many cosmids
SCP - exon is single copy in primates or C - conserved by cross species hybridisation
NH - sequenced exon displayed no significant similarity to databases
NOVEL BY SEQ. - novel by sequence searches
Empty boxes indicate that no analysis was performed
One of the products from cosmid 17B2 (17B2Bam.seq) showed homology to the Alu-Sx subfamily consensus sequence (Accession Number U14574). The sequence similarity is shown in figure 3.21. The product also hybridised to many bands on cosmid Southern blots and was deduced to be an artefact. Amplification of Alu sequences have also been reported by other groups (Church et al., 1993; North et al., 1993; Church et al., 1994), but they never represented more than 13% of the total number of products. Therefore although some Alu sequences contain regions with high homology to acceptor and donor splice junctions, the majority are not amplified.

**Figure 3.21**

**Sequence similarity between a product from cosmid 17B2 and the human Alu-Sx subfamily consensus sequence**

92% identity in a 97 nucleotide overlap

17B2Bam 39 AGGTTGCAGTGAGCTGA-GATCGT TCCACTGCACTCAAC 1
Alu-Sx 214 AGGTTGCAGTGAGCCGAAGATCGCGCCACTGCACTCCAGC 255

17B2Bam 77 GGAGGCTGAGTCAGGGGAATTGCTTGAACCCGGGAGGCG 40
Alu-Sx 178 GGAGGCTGAGGCAGGAGAATCGCTTGAACCCGGGAGGCG 213

17B2Bam 97 CTGTAATCCCAGTTACTCG 78
Alu-Sx 158 CTGTAATCCCAGCTACTCG 177
Nine of the products showed no significant sequence similarities. One of these, a product from cKWl hybridised specifically to the correct cosmid on a Southern blot and was also conserved by cross-species hybridisation. Therefore it was assumed to be part of a novel gene. Although these products may be authentic exons they were not pursued because analysis continued using the modified vector pSPL3.

The most interesting product was derived from cosmid 17BE and called exon e (figure 3.21). Sequence searches found it to have 73% identity in a 74 nucleotide overlap with mouse vav oncogene mRNA (accession number X64361) and 72% identity in a 65 nucleotide overlap with the human VAV oncogene (accession number X16316) (figure 3.22).

The product was cloned using a TA cloning kit (Invitrogen) and mapped genetically in mouse by interspecific backcross analysis (Pilz et al., 1995). Exon e was localised to proximal mouse chromosome 2 which is a region homologous to human chromosome 9q33-q34. It was localised to the same position as Dbh, Notch1 and Surf which are mouse homologues of human genes DBH, NOTCH and SURF-1 respectively. These genes were mapped to human chromosome 9q34 in section 3.2. The result confirmed the localisation of a VAV-like gene (VAV2) on human chromosome 9q34 (Woodward et al., 1994), a finding which was simultaneously reported by Smith et al. (1994b).
Sequence analysis of 17BE PCR product (exon e)

SD1

GGATCCGCGACGAAGACCTCCTCAAGGCAGTCAGACTCATAAAGTTTCTCT

HIV-1 tat exon  exon e

ATCAAAGCAACGAGCATGACCTGGGGAGGACATCTACGACTGCGTCCCTGTG

AGGATGGAGGAGGACCATCTACGAGAACATCATCAAGGTGGAGGTGCAG

HIV-1 tat exon

CAGCCCATGACCCACCACCTCCAATCCCGAGGGGACCGACGTGCAG

SA1

Sequence is shown 5' to 3' and 17BE and HIV-1 sequences are indicated by arrows.
Oligonucleotide primers SA1 and SD1 were used for the PCR reaction and for sequencing.
Figure 3.22

Sequence similarity between the product from cosmid 17BE (exon e) and VAV

A. 73% identity in a 74 nucleotide overlap between exon e and M. musculus vav mRNA

| exon e 16 | GAGGACATCTACGACTCGT- CCCTGTGAGGATGA - - | 50 |
| vav 483 | GAGGACCTTTATGACTGCTGGGAAATGAGGAGGCGAGA | 520 |

| exon e 51 | GGGGACGACATCTACGAGGACATCAAGGTGGGA | 86 |
| vav 522 | GGGGACGAGATCTACGAGGACCATACATCAAGGTGGGA | 556 |

B. 72% identity in a 65 nucleotide overlap between exon e and Human VAV mRNA

| exon e 16 | GAGGACATCTACGACTCGT - - CCCTG - TGAGGATGG | 53 |
| vav 436 | GAGGACCTGTATGACTGCTGGAGAATGAGGAGGCGAGA | 477 |

| exon e 50 | AGGGGACGACATCTACGAGGACATCAT | 76 |
| vav 474 | AGGCGACGACATCTACGAGGACATCAT | 500 |
Screening of a human foetal brain cDNA library (Stratagene) was initiated using exon e as a probe. However a collaborating group (Kwiatkowski et al.) had also isolated exon sequences from VAV2 and had already isolated a cDNA clone. Therefore screening became unnecessary and the work was halted. Exon e was confirmed to be homologous to the VAV2 cDNA clone provided, by hybridisation (figure 3.23). Sequencing data for VAV2, provided prior to publication (Henske et al., 1995) indicated that the exon was placed at the 5' end of the cDNA from base pair 455 to base pair 557. The gene was suggested as a candidate for TSC1 by Smith et al. (1994b). However subsequent analysis by Henske et al. (1995) eliminated VAV2 as a candidate for TSC1 and this investigation will be described in detail later (section 4.1.4.2.a).
Figure 3.23

Hybridisation of exon e to cDNA clones

A. Southern blot

1. 1kb ladder; 2. HGMP-RC S-24-D06; 3. HGMP-RC S-28-D06; 4. HGMP-RC S-37-F06; 5. VAV2 cDNA from Stratagene foetal brain library (Henske et al., 1995). The HGMP clones were used as negative controls in the experiment.

cDNA clones digested with restriction enzymes HindIII and BamHI and Southern blotted.
3.3.2. Exon amplification using pSPL3

Six contigs which mapped to the interval between the SD-1 and 9T12 translocation breakpoint by FISH were chosen for exon amplification. Cosmids were selected to give a minimum tiling path across the contig and they were initially mapped by FISH to confirm their localisation within the target interval. The cosmids are shown in table 3.7. and the contigs which they are from are shown in figures 3.24 - 3.29.

Table 3.7

Cosmid contigs used for exon amplification using pSPL3

<table>
<thead>
<tr>
<th>CONTIG NO.</th>
<th>CONTIG NAME</th>
<th>COSMID LL09 NO.</th>
<th>COSMID VAX NO.</th>
<th>COSMID NAME</th>
</tr>
</thead>
<tbody>
<tr>
<td>54</td>
<td>SURF</td>
<td>255 A6</td>
<td>94286</td>
<td></td>
</tr>
<tr>
<td>54</td>
<td>SURF</td>
<td>211 A7</td>
<td>94076</td>
<td></td>
</tr>
<tr>
<td>41</td>
<td>D9S10</td>
<td>255 H4</td>
<td>94125</td>
<td></td>
</tr>
<tr>
<td>41</td>
<td>D9S10</td>
<td>124 E8</td>
<td>92632</td>
<td></td>
</tr>
<tr>
<td>41</td>
<td>D9S10</td>
<td></td>
<td>17BE</td>
<td></td>
</tr>
<tr>
<td>80</td>
<td>D9S114</td>
<td>157 B7</td>
<td>92764</td>
<td></td>
</tr>
<tr>
<td>80</td>
<td>D9S114</td>
<td>222 A10</td>
<td>92868</td>
<td></td>
</tr>
<tr>
<td>80</td>
<td>D9S114</td>
<td>189 A10*</td>
<td>93659*</td>
<td></td>
</tr>
<tr>
<td>37</td>
<td>37</td>
<td>288 E8</td>
<td>92730</td>
<td></td>
</tr>
<tr>
<td>37</td>
<td>37</td>
<td>180 F1</td>
<td>92791</td>
<td></td>
</tr>
<tr>
<td>166</td>
<td>166</td>
<td>84 B3</td>
<td>92650</td>
<td></td>
</tr>
<tr>
<td>166</td>
<td>166</td>
<td>194 E12</td>
<td>93701</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>L-4DD1</td>
<td>203 H12</td>
<td>94546</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>L-4DD1</td>
<td>256 C3</td>
<td>94543</td>
<td></td>
</tr>
</tbody>
</table>

* No longer in this contig, now in contig 193.
VAX NO.- Number used for contig assembly on the VAX computer.
LL09 NO.- Lawrence Livermore library co-ordinate number.
17BE was derived from a human genomic library (Cachon-Gonzalez, 1991)
Figure 3.24
Contig number 54 which contains the Surfeit gene cluster

Centromere ← 291E1 ← 211A7 ← 219C9 ← 174E4 ← 60F9 ← ABO.17

181A9 ← 212E10 ← 219C9 ← 174E4 ← 60F9

255A6 * ← 226D9 * ← 72D11.x ← 272H1 ← 126D9 ← 39C8 *

211A7 ← 212E10 ← 219C9 ← 174E4 ← 60F9 ← 3H6 ← 109C8

226D9 * ← 72D11.x ← 272H1 ← 126D9 ← 39C8 *

255A6 * ← 226D9 * ← 72D11.x ← 272H1 ← 126D9 ← 39C8 *

211A7 ← 212E10 ← 219C9 ← 174E4 ← 60F9 ← 3H6 ← 109C8

177E4 ← 226D9 ← 72D11.x ← 272H1 ← 126D9 ← 39C8 *

211A7 ← 212E10 ← 219C9 ← 174E4 ← 60F9 ← 3H6 ← 109C8

255A6 * ← 226D9 * ← 72D11.x ← 272H1 ← 126D9 ← 39C8 *

Cosmids used for exon amplification were 211 A7 and 255 A6
Figure 3.25A

Contig number 41 which contains the marker D9S10

Centromere  

<table>
<thead>
<tr>
<th>DBH.2</th>
<th>245G2 *</th>
</tr>
</thead>
<tbody>
<tr>
<td>DBH.2</td>
<td>245G2 *</td>
</tr>
<tr>
<td>211H</td>
<td>254C11</td>
</tr>
<tr>
<td>139E4</td>
<td>DBH.5</td>
</tr>
<tr>
<td>254D11</td>
<td>275C9 *</td>
</tr>
<tr>
<td>47F1</td>
<td>296D11</td>
</tr>
<tr>
<td>38A4</td>
<td>260D5</td>
</tr>
<tr>
<td>38E11</td>
<td>182E3.x</td>
</tr>
<tr>
<td>DK1A1</td>
<td>DBH.4</td>
</tr>
<tr>
<td>157B4</td>
<td>DBH.1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>MCT136-10</th>
<th>35A9</th>
<th>132G4</th>
<th>275A5</th>
<th>152F5</th>
</tr>
</thead>
<tbody>
<tr>
<td>65G5 *</td>
<td></td>
<td>124E8</td>
<td>78H5 *</td>
<td>dk.n</td>
</tr>
<tr>
<td>37C7 *</td>
<td></td>
<td>252D7</td>
<td>242B6</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>17B.e</td>
<td></td>
<td></td>
</tr>
<tr>
<td>95230</td>
<td></td>
<td>255H4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>124F</td>
<td>7A10</td>
<td>245G1 *</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MCT136-10</td>
<td>79B4</td>
<td>104B4 *</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>59H6</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Telomere

<table>
<thead>
<tr>
<th>218C2.x</th>
<th>270B12</th>
<th>157D3 *</th>
</tr>
</thead>
<tbody>
<tr>
<td>127G4</td>
<td>230F5</td>
<td></td>
</tr>
<tr>
<td>234G12</td>
<td>157D3.x</td>
<td></td>
</tr>
<tr>
<td>53E7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>73B9</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Pick here to go to the genetic map

Pick here to go to the genetic map

Pick here to go to the genetic map

D9S150 DBH D9S122 D9S10 D9S66
D9S66 exon E

Mu.bin.4 DBH DBH Mu.bin.12 Mu.bin.15 FISH dis SD1 FISH prox B00041 FISH dis SD1 W3.8.15 W3.7.2
Mu.bin.4 DBH Mu.bin.11 Mu.bin.14 Mu.bin.15 MCT136 FISH dis SD1 W3.8.14 FISH prox B00041
DBH DBH FISH prox B00041 FISH prox B00041 MCT136 MCT136 W3.8.13 FISH prox B00041 W3.7.25
DBH DBH Mu.bin.10 Mu.bin.14 Mu.bin.15 FF9a.6 MCT136 FISH prox B00041 franz.1 W3.7.27
Mu.bin.7 Mu.bin.14 FISH dis SD1 alu.17B+20A FISH dis SD1 alu.17B+20A W3.7.8 FISH dis SD1
FISH dis SD1 Mu.bin.14 FISH prox B00041 MCT136 FISH prox B00041 W3.8.7 W3.7.21
Mu.bin.14 Mu.bin.15 FISH prox B00041 FISH dis SD1 FISH dis SD1 W3.7.10
Figure 3.25B  
Region of contig number 41 that was used for exon amplification

<table>
<thead>
<tr>
<th></th>
<th>Centromere</th>
<th>MCT136-10</th>
<th>108A10*</th>
<th>291A9*</th>
<th>Telomere</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>35A9</td>
<td>65G5*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>132G4</td>
<td>37C7*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>275A5</td>
<td>78H5*</td>
<td>252D7</td>
<td>242B6</td>
<td></td>
</tr>
<tr>
<td></td>
<td>152F5</td>
<td>272C1</td>
<td>dk.n</td>
<td>17B.e</td>
<td></td>
</tr>
<tr>
<td></td>
<td>267E11*</td>
<td>95230</td>
<td>255H4</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>272C1</td>
<td>124E8.x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>95230</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>78H5*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>267E11*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>sg52</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>267E11*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>137B8</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>216C9</td>
<td>79B4*</td>
<td>104B4*</td>
<td>59H6</td>
<td></td>
</tr>
<tr>
<td></td>
<td>79B4*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>137B8</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>216C9</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Pick here to go to the genetic map

- D9S10
- D9S66
- exon E
- D9S66

<table>
<thead>
<tr>
<th></th>
<th>Mu.bin.14</th>
<th>Mu.bin.15</th>
<th>FISH prox B00041</th>
<th>FISH prox B00041</th>
<th>W3.8.13</th>
<th>W3.8.14</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mu.bin.14</td>
<td>FISH prox B00041</td>
<td>MCT136</td>
<td>FISH dis SD1</td>
<td>FISH prox B00041</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mu.bin.14</td>
<td>FISH dis SD1</td>
<td>FISH prox B00041</td>
<td>MCT136</td>
<td>FISH dis SD1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mu.bin.14</td>
<td>Mu.bin.15</td>
<td>alu.17B+20A</td>
<td>FISH prox B00041</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mu.bin.14</td>
<td>Mu.bin.15</td>
<td>FF9a.6</td>
<td>MCT136</td>
<td>FISH prox B00041</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Cosmids used were 255 H4, 124 E8 and 17BE
Figure 3.26
Contig number 80 which contains marker D9S114

Centromere →

<table>
<thead>
<tr>
<th>188E3 *</th>
<th>294C11</th>
</tr>
</thead>
<tbody>
<tr>
<td>188E4 *</td>
<td>251B1</td>
</tr>
<tr>
<td>172E2</td>
<td>157B7</td>
</tr>
<tr>
<td>138A10</td>
<td>70G3</td>
</tr>
<tr>
<td>58H6.x *</td>
<td>193F7</td>
</tr>
<tr>
<td>222A10</td>
<td>196A5 *</td>
</tr>
<tr>
<td>17B.2</td>
<td>260H8</td>
</tr>
<tr>
<td>17B.a</td>
<td>66A9 *</td>
</tr>
<tr>
<td>210D12</td>
<td>93171</td>
</tr>
</tbody>
</table>

Telomere

Pick here to go to the genetic map
D9S114

Mu.bin.26 FISH prox B00041
FISH dis SD1
FISH prox B00041
FISH prox B00041
FISH dis SD1

FISH dis SD1
FISH prox B00041
tom to 93149

Cosmids used for exon amplification were 157B7, 222A10 and 17B2
(This contig has only very recently been orientated)
Cosmids used for exon amplification were 288 E8 and 180 F1
This contig has not been oriented on the chromosome
**Figure 3.28**

**Contig number 166**

<table>
<thead>
<tr>
<th>Gene 1</th>
<th>Gene 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>38A7</td>
<td>272D1 *</td>
</tr>
<tr>
<td>217G4</td>
<td>249F8</td>
</tr>
<tr>
<td>44C3</td>
<td>15B4</td>
</tr>
<tr>
<td>52D4</td>
<td>128D2</td>
</tr>
<tr>
<td>44G5 *</td>
<td>84B3</td>
</tr>
<tr>
<td>194E12</td>
<td></td>
</tr>
</tbody>
</table>

**Cosmids used for exon amplification were 84 B3 and 194 E12**

This contig has not been orientated on the chromosome.
Figure 3.29  Contig number 10 that contains the left hand end of YAC 4DD1 (L-4DD1)

Cosmids used for exon amplification were 203 H12 and 256 C3
It must be stressed that the contigs shown here are up to date and that the isolation of new clones and contig assembly has progressed a great deal since this analysis began. Therefore although only two or three clones were chosen from each contig they did cover the entire contig at the time. In addition, some contigs have been rearranged since this work commenced. Clone 189 A10 was in contig number 80 but it has now moved to contig number 193, which is a small contig within interval 2 that has not been further characterised. Unfortunately there are not any diagrams of the initial contigs available, to demonstrate how they have evolved over time.

Four of the contigs used in the analysis include at least one marker. Contig 54 contains the Surfeit gene cluster, contig 41 contains D9S10, contig 80 contains D9S114 and contig 10 contains the left-hand end of YAC clone 4DD1 (i.e. L-4DD1). These markers were used as names in this thesis because the computer software used for contig assembly can result in contig numbers changing. As contigs number 37 and 66 did not include a marker these numbers were used as names throughout but care was taken to correlate exons isolated with cosmid co-ordinates rather than contig number.

Each cosmid was digested with BamHI and PstI and shot-gun cloned into the prepared pSPL3 vector. The resultant miniprep DNA was confirmed to have all the possible DNA fragments cloned, by digestion with the appropriate restriction enzyme to release the insert. The digest was electrophoresed on an agarose gel beside an appropriate digest of the original cosmid and this allowed a direct comparison of the bands in common. Examples are shown in figure 3.30. The intense DNA fragment 6kb in size which is present in lane a. for each cosmid is the pSPL3 vector. The insert sizes were found to range from 1 - 10kb, the optimum sizes are 1 - 4kb (Buckler et al., 1991).
Figure 3.30

Shot-gun cloned cosmids into pSPL3 vector

**A. BamHI digested cosmids**

<table>
<thead>
<tr>
<th>Lanes</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>a</td>
<td>b</td>
<td>a</td>
<td>b</td>
<td>a</td>
<td>b</td>
<td>a</td>
<td>b</td>
<td>a</td>
</tr>
<tr>
<td></td>
<td>a</td>
<td>b</td>
<td>a</td>
<td>b</td>
<td>a</td>
<td>b</td>
<td>a</td>
<td>b</td>
<td>a</td>
</tr>
</tbody>
</table>

B. *PstI* digested cosmids

<table>
<thead>
<tr>
<th>Lanes</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>a</td>
<td>b</td>
<td>a</td>
<td>b</td>
<td>a</td>
<td>b</td>
<td>a</td>
<td>b</td>
<td>a</td>
</tr>
<tr>
<td></td>
<td>a</td>
<td>b</td>
<td>a</td>
<td>b</td>
<td>a</td>
<td>b</td>
<td>a</td>
<td>b</td>
<td>a</td>
</tr>
</tbody>
</table>

Lanes: a. shotgun cloned cosmid; b. restriction enzyme digested cosmid. 1. 17BE; 2. 255 H4; 3. 124 E8; 4. 180 F1; 5. 288 E8; 6. 194 E12; 7. 84 B3; 8 122 D6; 9. 1kb ladder.
After verifying the efficiency of the shot-gun cloning experiment, the miniprep DNA was transfected into COS-7 cells as for pSPL1. Each cosmid digest was transfected individually and then a pooled transfection was performed for each contig with both of the enzyme digests.

The schematic representation of the technique using pSPL3 is shown in figure 1.3 and the primers which can be used for RT-PCR, PCR and sequencing are shown in figure 3.31 (sequences are given in appendix 2). Reverse-transcription and 5 cycles of PCR were followed by digestion with restriction enzyme BstXI. This step should cleave vector only amplification products and cryptic splicing events which retain either side of the multiple cloning site and thereby eliminate them as substrates for the following 25 cycles of PCR. The differences between RT-PCR products with and without a digestion step are shown in figure 3.32. The 177bp vector only RT-PCR product was reduced following addition of BstXI but it was definitely not eliminated. However the enzyme digestion did result in the RT-PCR products greater than 177bp in size becoming more intense and more clearly defined. It suggested that the reduction in amount of vector only template allowed more efficient amplification of the larger substrates and that the complexity of the these products was reduced by elimination of the false positives.

The results of RT-PCR analysis with BstXI digestion, for each of the 6 contigs analysed are shown in figures 3.33A-3.38A. A large number of bands greater than the 177bp vector only RT-PCR product were produced for each contig with both of the enzyme digests used. All of the 12 cosmid yielded a product although clone 211 A7 from the SURF contig did not show any for the BamHI transfection. Most cosmids generated
### Figure 3.31

**Primers used for exon amplification with vector pSPL3**

<table>
<thead>
<tr>
<th>Primer Name</th>
<th>Position on Vector (bp)</th>
<th>Vector Position</th>
</tr>
</thead>
<tbody>
<tr>
<td>SD6</td>
<td>607-626</td>
<td>5' SSv</td>
</tr>
<tr>
<td>SD2</td>
<td>651-671</td>
<td></td>
</tr>
<tr>
<td>SD5</td>
<td>687-697</td>
<td></td>
</tr>
<tr>
<td>SA5</td>
<td>698/3095</td>
<td></td>
</tr>
<tr>
<td>SA4</td>
<td>3096-3105</td>
<td></td>
</tr>
<tr>
<td>SA2</td>
<td>3178-3199</td>
<td></td>
</tr>
<tr>
<td>dUSD2</td>
<td>3245-3265</td>
<td></td>
</tr>
<tr>
<td>KSFl</td>
<td>651-671</td>
<td></td>
</tr>
<tr>
<td>KSRl</td>
<td>671-690</td>
<td></td>
</tr>
<tr>
<td>dUSA4</td>
<td>698/3095-3113</td>
<td></td>
</tr>
</tbody>
</table>

5' SSv - pSPL3 splice donor site  
3' SSv - pSPL3 splice acceptor site
Figure 3.32

The effect of *Bst*XI digestion on RT-PCR products

<table>
<thead>
<tr>
<th>BstXI</th>
<th>Non BstXI</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>5</td>
<td>7</td>
</tr>
<tr>
<td>6</td>
<td>8</td>
</tr>
<tr>
<td>7</td>
<td>9</td>
</tr>
</tbody>
</table>

Lanes: 1. 1kb ladder; 2. mock transfection, no DNA control; 3. pSPL3 control; 4. 211 A7 PstI; 5. 255 A6 PstI; 6. pool 211 A7 and 255 A6 PstI; 7. 211 A7 BamHI; 8. 255 A6 BamHI; 9. pool 211 A7 and 255 A6 BamHI. Note that 177bp fragment is pSPL3 vector only.
multiple bands which suggested that a number of exons had been trapped. A larger number were detected for the PstI digested shot-gun clones in contigs SURF, D9S114, Ctg37 and Ctg166 compared to the BamHI digest. However there was no obvious difference in the yield between digests for contigs D9S10 and L-4DD1.

Overall the pooled transfections did not produce a significantly larger number of RT-PCR products than the single cosmids. There were more in the SURF BamHI pool, Ctg166 PstI pool and the L-4DD1 PstI pool compared to the individual cosmids. However there were fewer in D9S114 PstI pool and the rest were approximately the same.

The RT-PCR products were electrophoresed on a low melting point gel and those greater than 177bp were excised, purified and shot-gun cloned into the pAMP1 vector (BRL). The white colonies were picked, grown overnight and then confirmed to contain an insert by PCR using internal primers (KSFl and KSR1) which flank the splice sites. RT-PCR products were cloned for all the single cosmid transfections and for all the pools except Ctg166. A great deal of redundancy within each contig was expected and a rapid method of reducing the numbers of products to be analysed was required. Therefore the PCR products were electrophoresed on agarose gels and those of different sizes for each digest selected. These are shown in figures 3.33-3.38B. This selection method cannot discriminate between different exons of the same length. Glycerol stocks were therefore prepared for all of the clones so that an additional screen would be possible at a later date.
Figure 3.33

RT-PCR analysis and cloned products from contig SURF

A. RT-PCR analysis

<table>
<thead>
<tr>
<th>Lane</th>
<th>BamHI</th>
<th>PstI</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>3 4</td>
<td>3 4</td>
</tr>
</tbody>
</table>

1000 - 177bp vector

A. Lanes: 1. 1kb ladder; 2. cosmid 211 A7; 3. cosmid 255 A6; 4. pooled cosmids 211 A7 and 255 A6. B. 1. 1kb ladder; products amplified using PCR primers KSFl and KSR1. Sizes shown in base pairs.

B. Cloned products

<table>
<thead>
<tr>
<th>Lane</th>
<th>BamHI</th>
<th>PstI</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A B C</td>
<td>A B C D</td>
</tr>
</tbody>
</table>

1000 -

A. Lanes: 1. 1kb ladder; 2. cosmid 211 A7; 3. cosmid 255 A6; 4. pooled cosmids 211 A7 and 255 A6. B. 1. 1kb ladder; products amplified using PCR primers KSFl and KSR1. Sizes shown in base pairs.
Figure 3.34

RT-PCR analysis and cloned products from contig D9S10

A. RT-PCR analysis

B. Cloned products

A. Lanes: 1. 1kb ladder; 2. mock transfection, no DNA control; 3. pSPL3 vector only control; 4. cosmid 124 E8; 5. cosmid 255 H4; 6. pooled cosmids 124 E8, 255 H4 and 17BE. B. 1. 1kb ladder; products amplified using PCR primers KSF1 and KSR1. Sizes shown in base pairs.
Figure 3.35

RT-PCR analysis and cloned products from contig D9S114

A. RT-PCR analysis

B. Cloned products

A. Lanes: 1. 1kb ladder; 2. cosmid 157 B7; 3. cosmid 222 A10; 4. pooled cosmids 157 B7, 222 A10 and 17B2. B. 1. 1kb ladder; products amplified using PCR primers KSF1 and KSR1. Sizes shown in base pairs.
Figure 3.36

RT-PCR analysis and cloned products from contig 37

A. RT-PCR analysis

B. Cloned products

A. Lanes: 1. 1kb ladder; 2. mock transfection, no DNA control; 3. pSPL3 vector only control; 4. cosmid 288 E8; 5. cosmid 180 F1; 6. pooled cosmids 288 E8 and 180 F1. B.1. 1kb ladder; products amplified using PCR primers KSF1 and KSR1. Sizes shown in base pairs.
Figure 3.37

RT-PCR analysis and cloned products from contig 166

A. RT-PCR analysis

BamHI

<table>
<thead>
<tr>
<th>Lane</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Size</td>
<td>1000</td>
<td>500</td>
<td>400</td>
<td>320</td>
</tr>
</tbody>
</table>

PstI

<table>
<thead>
<tr>
<th>Lane</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Size</td>
<td>1000</td>
<td>500</td>
<td>400</td>
</tr>
</tbody>
</table>

B. Cloned products

BamHI

<table>
<thead>
<tr>
<th>Lane</th>
<th>A</th>
<th>B</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Size</td>
<td>1000</td>
<td>500</td>
<td>400</td>
</tr>
</tbody>
</table>

PstI

<table>
<thead>
<tr>
<th>Lane</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Size</td>
<td>1000</td>
<td>500</td>
<td>400</td>
<td>320</td>
<td>300</td>
<td>200</td>
</tr>
</tbody>
</table>

- 177bp vector

A. Lanes: 1. 1kb ladder; 2. cosmid 84 B3; 3. cosmid 194 E12; 4. pooled cosmids 84 B3, and 194 E12. B. 1. 1kb ladder; products amplified using PCR primers KSF1 and KSR1. Sizes shown in base pairs.
Figure 3.38

RT-PCR analysis and cloned products from contig L-4DD1

A. RT-PCR analysis

- BamHI
- PstI

B. Cloned products

A. Lanes: 1. 1kb ladder; 2. mock transfection, no DNA control; 3. pSPL3 vector only control; 4. cosmid 203 H12; 5. cosmid 256 C3; 6. pooled cosmids 203 H12 and 256 C3.

B. 1. 1kb ladder; products amplified using PCR primers KSFl and KSR1. Sizes shown in base pairs.
A total of 171 RT-PCR products were cloned, 27 from contig SURF, 29 from contig D9S10, 32 from contig D9S114, 29 from contig 37, 28 from contig 166 and 26 from contig L-4DD1. Table 3.8 shows how the numbers of different sized products corresponded to the numbers of colonies picked. Clones of different sizes from the same contig and the same enzyme digest were named from A-F with decreasing order of size. Often many clones of the same size were obtained from one or more cosmids from the same contig and provided that they were derived using the same enzyme digest, they were given the same letter code. For example in contig D9S10, the BamHI digest of cosmid 124 E8 yielded 5 clones identical in size to two of those from 17BE (BamHI) and a further two from 255 H4 (BamHI). These product were assumed to be the same as they were derived from overlapping clones. They were designated C as they were the third smallest clone isolated from the BamHI digest of contig D9S10.

The pooled transfections for L-4DD1 BamHI, D9S114 PstI and D9S10 PstI generated all the different sized products for their corresponding digests. In the majority of cases it was a combination of products from each individual cosmid and the pool which gave the total number. This result was encouraging as it made the rather labour intensive analysis seem more worthwhile. However as the recombinant clones were only plated once, if the procedure had been repeated with a increased number of clones picked and characterised, a greater representation of products may have been obtained. This may also explain why in some cases the numbers of RT-PCR products cloned, did not reflect
## Table 3.8 Numbers of exons cloned from the six contigs screened

<table>
<thead>
<tr>
<th>Contig</th>
<th>Cosmid</th>
<th>Digest</th>
<th>No. of clones picked</th>
<th>No. of difft sizes</th>
<th>Clones typed by size for each enz.</th>
</tr>
</thead>
<tbody>
<tr>
<td>SURF</td>
<td>211 A7</td>
<td>BamHl</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>SURF</td>
<td>255 A6</td>
<td>BamHl</td>
<td>9</td>
<td>3</td>
<td>2A, 1B, 5C</td>
</tr>
<tr>
<td>SURF</td>
<td>POOL</td>
<td>BamHl</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>SURF</td>
<td>211 A7</td>
<td>Pstl</td>
<td>3</td>
<td>1</td>
<td>3D</td>
</tr>
<tr>
<td>SURF</td>
<td>255 A6</td>
<td>Pstl</td>
<td>15</td>
<td>3</td>
<td>1B, 2C, 12D</td>
</tr>
<tr>
<td>SURF</td>
<td>POOL</td>
<td>Pstl</td>
<td>1</td>
<td>1</td>
<td>1A</td>
</tr>
<tr>
<td>D9S10</td>
<td>124 E8</td>
<td>BamHl</td>
<td>5</td>
<td>1</td>
<td>5C</td>
</tr>
<tr>
<td>D9S10</td>
<td>255 H4</td>
<td>BamHl</td>
<td>3</td>
<td>2</td>
<td>2C, 1D</td>
</tr>
<tr>
<td>D9S10</td>
<td>17BE</td>
<td>BamHl</td>
<td>10</td>
<td>2</td>
<td>8B, 2C</td>
</tr>
<tr>
<td>D9S10</td>
<td>POOL</td>
<td>BamHl</td>
<td>3</td>
<td>2</td>
<td>1A, 2B</td>
</tr>
<tr>
<td>D9S10</td>
<td>124 E8</td>
<td>Pstl</td>
<td>1</td>
<td>1</td>
<td>1B</td>
</tr>
<tr>
<td>D9S10</td>
<td>255 H4</td>
<td>Pstl</td>
<td>3</td>
<td>1</td>
<td>3A</td>
</tr>
<tr>
<td>D9S10</td>
<td>POOL</td>
<td>Pstl</td>
<td>4</td>
<td>2</td>
<td>2A, 2B</td>
</tr>
<tr>
<td>D9S114</td>
<td>157 B7</td>
<td>BamHl</td>
<td>10</td>
<td>5</td>
<td>2A, 3B, 1C, 1D, 3E</td>
</tr>
<tr>
<td>D9S114</td>
<td>222 A10</td>
<td>BamHl</td>
<td>9</td>
<td>5</td>
<td>2A, 1B, 1C, 1D, 4E</td>
</tr>
<tr>
<td>D9S114</td>
<td>POOL</td>
<td>BamHl</td>
<td>4</td>
<td>3</td>
<td>1B, 2D, 1E</td>
</tr>
<tr>
<td>D9S114</td>
<td>157 B7</td>
<td>Pstl</td>
<td>1</td>
<td>1</td>
<td>1B</td>
</tr>
<tr>
<td>D9S114</td>
<td>222 A10</td>
<td>Pstl</td>
<td>4</td>
<td>3</td>
<td>1A, 1B, 2C</td>
</tr>
<tr>
<td>D9S114</td>
<td>POOL</td>
<td>Pstl</td>
<td>4</td>
<td>3</td>
<td>1A, 2B, 1C</td>
</tr>
<tr>
<td>37</td>
<td>288 E8</td>
<td>BamHl</td>
<td>5</td>
<td>3</td>
<td>1A, 3C, 1D</td>
</tr>
<tr>
<td>37</td>
<td>180 F1</td>
<td>BamHl</td>
<td>9</td>
<td>4</td>
<td>3A, 2B, 3C, 1D</td>
</tr>
<tr>
<td>37</td>
<td>POOL</td>
<td>BamHl</td>
<td>4</td>
<td>2</td>
<td>1B, 2D, 1E</td>
</tr>
<tr>
<td>37</td>
<td>288 E8</td>
<td>Pstl</td>
<td>5</td>
<td>3</td>
<td>2A, 2B, 1C</td>
</tr>
<tr>
<td>37</td>
<td>180 F1</td>
<td>Pstl</td>
<td>6</td>
<td>4</td>
<td>1A, 2B, 2C, 1D</td>
</tr>
<tr>
<td>37</td>
<td>POOL</td>
<td>Pstl</td>
<td>1</td>
<td>1</td>
<td>1D</td>
</tr>
<tr>
<td>166</td>
<td>84 B3</td>
<td>BamHl</td>
<td>5</td>
<td>2</td>
<td>3A, 2B, 1C</td>
</tr>
<tr>
<td>166</td>
<td>194 E12</td>
<td>BamHl</td>
<td>5</td>
<td>2</td>
<td>1B, 4C</td>
</tr>
<tr>
<td>166</td>
<td>84 B3</td>
<td>Pstl</td>
<td>7</td>
<td>4</td>
<td>1B, 1D, 3E, 2F</td>
</tr>
<tr>
<td>166</td>
<td>194 E12</td>
<td>Pstl</td>
<td>11</td>
<td>5</td>
<td>2A, 2B, 2C, 3D, 2E</td>
</tr>
<tr>
<td>L-4DD1</td>
<td>203 H12</td>
<td>BamHl</td>
<td>3</td>
<td>2</td>
<td>2C, 1E</td>
</tr>
<tr>
<td>L-4DD1</td>
<td>POOL</td>
<td>BamHl</td>
<td>7</td>
<td>5</td>
<td>1A, 1B, 3C, 1D, 1E</td>
</tr>
<tr>
<td>L-4DD1</td>
<td>203 H12</td>
<td>Pstl</td>
<td>2</td>
<td>1</td>
<td>2D</td>
</tr>
<tr>
<td>L-4DD1</td>
<td>256 C3</td>
<td>Pstl</td>
<td>8</td>
<td>4</td>
<td>1A, 2B, 2C, 3D</td>
</tr>
<tr>
<td>L-4DD1</td>
<td>POOL</td>
<td>Pstl</td>
<td>6</td>
<td>3</td>
<td>3A, 2B, 1D</td>
</tr>
</tbody>
</table>

Products range from A-F with decreasing order of size for each of the 6 contigs and for each of the 2 digest used. Numbers reflect the proportions of different sized products cloned from each of the 12 categories.
the actual numbers of RT-PCR products present in the transfection (see figures 3.33-3.38). It was usually the largest products which were not cloned and a bias in favour of smaller fragments during the shot-gun cloning could be an explanation. Therefore it may be necessary to clone these larger fragments by excising them individually from the agarose gel.

3.3.2.1. **Sequence analysis**

One example of the different sized products from each contig digest was selected for further analysis. A number of products had identical sizes to different cosmids within the same contig digest and as they were considered to be the same, one was selected at random. 48 products were chosen from the total number of 171 products cloned and they were all characterised by sequence analysis. The numbers selected from each digest ranged from 2 to 6, giving a total per contig ranging from 6 to 9. The average putative exon size was 189bp with a range from 69-520bp. They were named according to the contig they were from (D9S10, Ctg37 etc.), the digest used and a letter from A to F with decreasing order of size as shown in table 3.8. The sequences obtained (appendix 3.2) were then compared to nucleotide and protein databases using BLASTN, BLASTX and FASTA computer programme searches. The results are shown in table 3.9. In some instances the full nucleotide sequence was not deduced but approximate length and sequenced length are shown. A sequence was considered as having a significant match to a nucleotide database entry if the similarity was ≥55bp in length, ≥60% similarity, p value ≤0.01 and to a protein database entry if the similarity was ≥24 amino-acids in length, ≥55% similarity, p value ≤0.01 (Khan et al., 1992). Similarities were defined as weak if they were just within these limits and strong if the similarity was ≥70%, p value ≤e-5. Only sequences with strongly significant matches were analysed further.
Table 3.9

Forty-eight exons isolated from 9q34, with their approximate length, sequenced length and database similarities

<table>
<thead>
<tr>
<th>EXON NAME</th>
<th>CONTIG</th>
<th>ENZ</th>
<th>LET</th>
<th>APPR. LTH</th>
<th>SEQ. LTH</th>
<th>SEQUENCE SIMILARITY</th>
</tr>
</thead>
<tbody>
<tr>
<td>SURF</td>
<td>Bam HI</td>
<td>A</td>
<td>290</td>
<td>252</td>
<td>FASTA chicken L7a (72% in 65) (D14522)</td>
<td></td>
</tr>
<tr>
<td>SURF</td>
<td>Bam HI</td>
<td>B</td>
<td>159</td>
<td>159</td>
<td>FASTA human L7a (100% in 81) (X52138)</td>
<td></td>
</tr>
<tr>
<td>SURF</td>
<td>Bam HI</td>
<td>C</td>
<td>101</td>
<td>101</td>
<td>FASTA human L21 (96% in 101) (U14967)</td>
<td></td>
</tr>
<tr>
<td>SURF</td>
<td>Pst I</td>
<td>A</td>
<td>280</td>
<td>177</td>
<td>BLASTN ampicillinase (e-41) (M14017)</td>
<td></td>
</tr>
<tr>
<td>SURF</td>
<td>Pst I</td>
<td>B</td>
<td>240</td>
<td>151</td>
<td>5' chicken L7a (D14522) 3' BLASTN HIV (e-8)</td>
<td></td>
</tr>
<tr>
<td>SURF</td>
<td>Pst I</td>
<td>C</td>
<td>120</td>
<td>67</td>
<td>BLASTN HIV (e-21) SAc/SDc</td>
<td></td>
</tr>
<tr>
<td>SURF</td>
<td>Pst I</td>
<td>D</td>
<td>81</td>
<td>81</td>
<td>FASTA chicken L7a (73% in 81) (D14522)</td>
<td></td>
</tr>
<tr>
<td>D9S10</td>
<td>Bam HI</td>
<td>A</td>
<td>198</td>
<td>198</td>
<td>BLASTN mouse otx1 homeobox (e-4) (P80205)</td>
<td></td>
</tr>
<tr>
<td>D9S10</td>
<td>Bam HI</td>
<td>B</td>
<td>116</td>
<td>116</td>
<td>BLASTN HIV (e-42) SAc/SDc</td>
<td></td>
</tr>
<tr>
<td>D9S10</td>
<td>Bam HI</td>
<td>C</td>
<td>103</td>
<td>103</td>
<td>FASTA human VAV (73% in 69) (X64361)</td>
<td></td>
</tr>
<tr>
<td>D9S10</td>
<td>Bam HI</td>
<td>D</td>
<td>69</td>
<td>69</td>
<td>FASTA human VAV (64% in 69) (X64361)</td>
<td></td>
</tr>
<tr>
<td>D9S10</td>
<td>Pst I</td>
<td>A</td>
<td>134</td>
<td>134</td>
<td>FASTA human cDNA (60% in 136) (X57348)</td>
<td></td>
</tr>
<tr>
<td>D9S10</td>
<td>Pst I</td>
<td>B</td>
<td>106</td>
<td>106</td>
<td>FASTA human integrin (63% in 52) (M14785)</td>
<td></td>
</tr>
<tr>
<td>D9S114</td>
<td>Bam HI</td>
<td>A</td>
<td>273</td>
<td>273</td>
<td>BLASTX yeast hypothetical guanine nucleotide binding protein (e-15) (P25382)</td>
<td></td>
</tr>
<tr>
<td>D9S114</td>
<td>Bam HI</td>
<td>B</td>
<td>230</td>
<td>230</td>
<td>BLASTN human set gene (e-26) (M93651)</td>
<td></td>
</tr>
<tr>
<td>D9S114</td>
<td>Bam HI</td>
<td>C</td>
<td>200</td>
<td>200</td>
<td>BLASTX yeast hypothetical guanine nucleotide binding protein (e-8) (P25382)</td>
<td></td>
</tr>
</tbody>
</table>

Continued on next page
<table>
<thead>
<tr>
<th>EXON NAME</th>
<th>CONTIG</th>
<th>APPR.</th>
<th>SEQ.</th>
<th>SEQUENCE SIMILARITY</th>
</tr>
</thead>
<tbody>
<tr>
<td>D9S114</td>
<td>Bam HI</td>
<td>D</td>
<td>146</td>
<td>146</td>
</tr>
<tr>
<td>D9S114</td>
<td>Bam HI</td>
<td>E</td>
<td>116</td>
<td>116</td>
</tr>
<tr>
<td>D9S114</td>
<td>Pst I</td>
<td>A</td>
<td>210</td>
<td>164</td>
</tr>
<tr>
<td>D9S114</td>
<td>Pst I</td>
<td>B</td>
<td>198</td>
<td>198</td>
</tr>
<tr>
<td>D9S114</td>
<td>Pst I</td>
<td>C</td>
<td>74</td>
<td>74</td>
</tr>
<tr>
<td>Ctg37</td>
<td>Bam HI</td>
<td>A</td>
<td>260</td>
<td>246</td>
</tr>
<tr>
<td>Ctg37</td>
<td>Bam HI</td>
<td>B</td>
<td>230</td>
<td>162</td>
</tr>
<tr>
<td>Ctg37</td>
<td>Bam HI</td>
<td>C</td>
<td>160</td>
<td>82</td>
</tr>
<tr>
<td>Ctg37</td>
<td>Bam HI</td>
<td>D</td>
<td>140</td>
<td>96</td>
</tr>
<tr>
<td>Ctg37</td>
<td>Bam HI</td>
<td>E</td>
<td>88</td>
<td>88</td>
</tr>
<tr>
<td>Ctg37</td>
<td>Pst I</td>
<td>A</td>
<td>190</td>
<td>104</td>
</tr>
<tr>
<td>Ctg37</td>
<td>Pst I</td>
<td>B</td>
<td>170</td>
<td>124</td>
</tr>
<tr>
<td>Ctg37</td>
<td>Pst I</td>
<td>C</td>
<td>116</td>
<td>116</td>
</tr>
<tr>
<td>Ctg37</td>
<td>Pst I</td>
<td>D</td>
<td>149</td>
<td>149</td>
</tr>
<tr>
<td>Ctg166</td>
<td>Bam HI</td>
<td>A</td>
<td>210</td>
<td>126</td>
</tr>
<tr>
<td>Ctg166</td>
<td>Bam HI</td>
<td>B</td>
<td>120</td>
<td>103</td>
</tr>
<tr>
<td>Ctg166</td>
<td>Bam HI</td>
<td>C</td>
<td>110</td>
<td>34</td>
</tr>
<tr>
<td>Ctg166</td>
<td>Pst I</td>
<td>A</td>
<td>260</td>
<td>165</td>
</tr>
</tbody>
</table>

Continued on next page
Table 3.9 Continued

<table>
<thead>
<tr>
<th>EXON NAME APPR. LTH</th>
<th>SEQUENCED LTH</th>
<th>SEQUENCE SIMILARITY</th>
</tr>
</thead>
<tbody>
<tr>
<td>EXON NAME CONTIG</td>
<td>ENZ</td>
<td>LET</td>
</tr>
<tr>
<td>Ctg166</td>
<td>PstI</td>
<td>B</td>
</tr>
<tr>
<td>Ctg166</td>
<td>PstI</td>
<td>C</td>
</tr>
<tr>
<td>Ctg166</td>
<td>PstI</td>
<td>D</td>
</tr>
<tr>
<td>Ctg166</td>
<td>PstI</td>
<td>E</td>
</tr>
<tr>
<td>Ctg166</td>
<td>PstI</td>
<td>F</td>
</tr>
<tr>
<td>L 4DD1</td>
<td>BamHI</td>
<td>A</td>
</tr>
<tr>
<td>L 4DD1</td>
<td>BamHI</td>
<td>B</td>
</tr>
<tr>
<td>L 4DD1</td>
<td>BamHI</td>
<td>C</td>
</tr>
<tr>
<td>L 4DD1</td>
<td>BamHI</td>
<td>D</td>
</tr>
<tr>
<td>L 4DD1</td>
<td>BamHI</td>
<td>E</td>
</tr>
<tr>
<td>L 4DD1</td>
<td>PstI</td>
<td>A</td>
</tr>
<tr>
<td>L 4DD1</td>
<td>PstI</td>
<td>B</td>
</tr>
<tr>
<td>L 4DD1</td>
<td>PstI</td>
<td>C</td>
</tr>
<tr>
<td>L 4DD1</td>
<td>PstI</td>
<td>D</td>
</tr>
</tbody>
</table>

ENZ - enzyme digest used for shot-gun cloning
LET - letter code given to product
APPR. LTH - approximate product length
SEQ LTH - sequenced product length
SA/SDc - vector cryptic splice acceptor and donor sites used

174
A putative exon was only considered to be identical to a known sequence if there was ≥97% identity (Khan et al., 1992).

Sequence similarities for each of the six contigs are described separately in the following sections (3.3.2.1.a - 3.3.2.1.f). The results obtained were used to select a subset of the putative exons for further analysis. These were then hybridisation to their corresponding cosmids and to zooblots to detect evolutionary conservation. Once their authenticity had been confirmed a number exons were then used as probes to screen cDNA libraries and identify transcripts.

3.3.2.1.a. Contig SURF

Eight different sized products were cloned from the SURF contig and four of these showed strong similarity to ribosomal protein L7a (SURF Bam A, SURF Bam B and SURF Pst B and SURF Pst D) (see figure 3.39). The largest product SURF Bam A was approximately 270bp in length and 252bp were sequenced. It was found to overlap both SURF Pst D and SURF Pst B. SURF Pst D was sequenced completely and its entire length overlapped SURF Bam A from base-pairs 7 to 75. It suggested that SURF Bam A could start at the same position as SURF Pst D but that it must be composed of more than one exon. The overlap between SURF Pst B and SURF Bam A extended for 59 base pairs, from position 143 to 200bp on SURF Bam A and position 3 to 60bp on SURF Pst B. The end of the overlap coincided with a PstI site within SURF Bam A. From position 61bp onwards, SURF Pst B was composed of HIV-1 tat intron sequence. The presence of the PstI restriction enzyme site within the exon meant that an incomplete exon was cloned for the PstI digested shot-gun clones. SURF Pst B was amplified by the pairing of vector splice donor site with cloned splice acceptor site, the
absence of the cloned splice donor site was compensated by activation of the cryptic vector splice donor site at position 1134bp on the pSPL3 vector and it paired with the vector splice acceptor site. The result was a chimeric product. A schematic representation of cryptic splicing is shown in figure 3.20 and the HIV-1 sequence amplified in each case is shown in figure 3.19.

**Figure 3.39**

**Organisation of overlapping exon sequences within SURF Bam A**

The gene encoding ribosomal protein L7a is known as SURF-3 (Giallongo *et al.*, 1989). It is localised within the Surfeit gene cluster on human chromosome 9q34 (Yon *et al.*, 1989) and was known to be within the contig studied. The cluster contains at least six closely spaced genes which have been designated SURF-1 to SURF-6. The similarities between SURF Bam B and the chicken and human L7a ribosomal protein are shown in figure 3.40. The similarity between SURF Bam A and L7a ribosomal protein was found
further upstream of that for SURF Bam B, but both were within the 5' flanking region of
the gene. The similarities were all found between the minus DNA strand of the
sequenced product and the plus DNA strand of ribosomal protein L7a. Exon
amplification requires exons to be cloned in the 5' to 3' orientation for correct splicing
and cloning into the pAMP1 vector (BRL) using Uracil DNA Glycosylase is also in the
5' to 3' direction. Therefore the only explanation was that the products isolated were
from another SURF gene further upstream of SURF-3 that had coding sequence on the
opposite DNA strand. A representation of the Surfeit gene cluster is shown in figure
3.41 and the organisation suggests that the products could be derived from SURF-5. No
sequence information for human SURF-5 is in the Embl or GenBase databases and the
direction of transcription is not known. However analysis of the mouse gene cluster has
shown that the direction of transcription for 5 of the Surf genes, Surf-1 to Surf-5,
alternates with respect to that of its neighbour (Yon et al., 1993). The close clustering of
the Surfeit genes and their associated CpG islands has been conserved even in the
chicken, which is separated from mammals by 600 million years of divergent evolution
(Colombo et al., 1992). Therefore it seems likely that the direction of transcription for
SURF-5, could be the same in the human genome as that in the mouse, i.e. opposite to
SURF-3.

If SURF Bam A, SURF Bam B, SURF Pst B and SURF Pst D are derived from SURF-
5, they gives us some indication of the number of exons that must be present, assuming
that the correct splicing events have taken place. The products have been positioned
within SURF-5 in figure 3.42 but they only represent the minimum number. Exon
amplification relies on the presence of both functional 5' and 3' splice sites and thereby
eliminates the possibility of the most terminal exons being amplified at both the 5' and
Figure 3.40

Sequence similarity between SURF Bam B and ribosomal protein L7a

A. 84% similarity in a 117 nucleotide overlap with chicken L7a (accession no. D14522)

<table>
<thead>
<tr>
<th>SURF Bam B 120</th>
<th>CTTGGCGGTCTTGATGATCTCGGTGAAGTTGTCCATGA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chicken L7a 468</td>
<td>CTTAGCGGTCTTTAATGATCTCGGTAAAGTTGTCCATAA</td>
</tr>
<tr>
<td>SURF Bam B 82</td>
<td>TGGACTTAATGTCGTCCTTCAGCCGCTTGTTGTAGGAC</td>
</tr>
<tr>
<td>Chicken L7a 506</td>
<td>TGGATTTGACGTACATCTTAAGGCCTTGTTGTAGGAC</td>
</tr>
<tr>
<td>SURF Bam B 43</td>
<td>TGCAGCAGCGTCTCTTTACTTGCTGTGGGAGCTCCTCTCTG</td>
</tr>
<tr>
<td>Chicken L7a 545</td>
<td>TGCAGCAGGGTCTGTTACTCTGTCGACAGAACCAGCTG</td>
</tr>
<tr>
<td>SURF Bam B 2</td>
<td>CTG 4</td>
</tr>
<tr>
<td>Chicken L7a 582</td>
<td>CTG 584</td>
</tr>
</tbody>
</table>

B. 100% identity in a 81 nucleotide overlap with human L7a (accession no. X52138)

<table>
<thead>
<tr>
<th>SURF Bam B 120</th>
<th>CTTGGCGGTCTTGATGATCTCGGTGAAGTTGTCCATGA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human L7a 398</td>
<td>CTTGGCGGTCTTGATGATCTCGGTGAAGTTGTCCATGA</td>
</tr>
<tr>
<td>SURF Bam B 82</td>
<td>TGGACTTAATGTCGTCCTTCAGCCGCTTGTTGTAGGAC</td>
</tr>
<tr>
<td>Human L7a 436</td>
<td>TGGACTTAATGTCGTCCTTCAGCCGCTTGTTGTAGGAC</td>
</tr>
<tr>
<td>SURF Bam B 43</td>
<td>TGCAG 39</td>
</tr>
<tr>
<td>Human L7a 475</td>
<td>TGCAG 479</td>
</tr>
</tbody>
</table>
Figure 3.41

Organisation of the human and mouse Surfeit gene cluster

A. Human Surfeit locus

![Diagram of the human Surfeit locus with CpG islands and gene positions]

B. Mouse Surfeit locus

![Diagram of the mouse Surfeit locus with CpG islands and gene positions]

Taken from Yon et al. (1993). Arrow heads show the direction of transcription, dotted lines indicate the uncertainty of extent and position of the SURF-1, SURF-2 and SURF-4 genes. Putative CpG islands are shown as black boxes with the distances between in kilobase pairs.
3’ end. As SURF Bam B was the product which was positioned the most 3’ on SURF-3, it would be the most 5’ on SURF-5. SURF Bam A and SURF Pst D begin at approximately the same position but the former is 189bp larger in size. Therefore it is suggested that the next complete exon amplified is SURF Pst D. Between the end of SURF Pst B and the start of SURF Pst D within SURF Bam A, there is 66bp sequence which may represent the third exon. The region from the beginning of SURF Pst B to the end of SURF Bam A could represent the fourth exon amplified. These results suggest the minimum number of exons within SURF-5 is six.

**Figure 3.42**

Possible organisation of SURF Bam A, SURF Bam B, SURF Pst B and SURF Pst D within the human Surfeit gene cluster

The direction of transcription for SURF-3 is shown by an arrow, the presumed direction for SURF-5 is shown by a dotted arrow. The boxes represent products cloned and the numbers inside indicate the different possible exons which have been amplified. The shaded box represents HIV-1 *tat* intron sequence.
Although SURF Bam C was derived from the same cosmid (255 A6) as the products with strong similarity to ribosomal protein L7a, it showed strong similarity to human L21 (96% on 101bp, p value e-30). This is also a ribosomal gene but it has not been localised to the Surfeit gene cluster. SURF Bam C also had strong similarity to clone P94_70 (p value e-36) which was an amplified exon isolated by Church et al. (1994) and was most probably identical. In addition, there was strong similarity to a cDNA clone (p value e-33) from the Human Genome Mapping Project Resource Centre (HGMP-RC) (M-02-H01, human foetal organ mixture tissue 3' end). These results suggest that it could be a true exon. However it could also be a processed pseudogene which is localised between SURF-5 and SURF-6 as the distance between these genes has not been published. Other possibilities include it being derived from SURF-6 as no sequence information is available for this gene. Alternatively it could be derived from the untranslated leader sequence of SURF-5 or SURF-6 or from an aberrant splicing event within the intronic sequences.

The remaining two products from the SURF contig were deduced to be artefacts as they contained either entirely HIV-1 sequence (SURF Pst C) or pSPL3 vector sequence (SURF Pst A).

3.3.2.1.b. Contig D9S10

Six different sized products were cloned from contig D9S10. Two products (D9S10 Bam C and D9S10 Bam D) showed strong similarity to human VAV mRNA (accession number X16316) and provided further confirmation of a VAV-like gene (VAV2) on human chromosome 9q34. D9S10 Bam C had identical sequence to exon e which was
amplified from cosmid 17BE using vector pSPL1. Figure 3.22 showed the similarity between exon e and the mouse and human VAV sequence. D9S10 Bam D was found to overlap the human VAV mRNA immediately upstream of this sequence. Therefore it was assumed that the intronic sequence between these two exons contained a BamHI restriction enzyme site which resulted in them being cloned into pSPL3 on separate genomic fragments. Sequence data for a VAV2 cDNA was provided by Dr. D. Kwiatkowski prior to publication (Henske et al., 1995) and D9S10 Bam C was localised to base pairs 455-557, the same position as exon e. D9S10 Bam C was localised immediately upstream at basepairs 386-454.

D9S10 Pst A and D9S10 Pst B showed similarity to human sequences (see table 3.9) and D9S10 Bam A showed similarity to the mouse otx1 homeobox gene. D9S10 Bam B was deduced to be an artefact as it was composed of entirely HIV-1 sequence. The 116bp product was produced by activation of vector cryptic splice donor and acceptor sites. Figure 3.43 shows this sequence with the flanking vector sequence which acts as splice junctions.
**Figure 3.43**

**Artefact amplification product composed of entirely HIV-1 sequence**

position 1324bp on pSPL3 vector

```
gtgttag TTTAAAGTGCACTGATTTGAAGAATGATACTAATACCAATAGTAGT
AGCGGGAGAATGATAATGGAGGAAGGAGAGATAAAAAACTGCTCTTTTAAT
ATCAGCACAAGCATAAGAG gtaagg
```

position 1439bp on pSPL3 vector

Lower case letters show the vector sequence which acts as splice sites. Arrows indicate the beginning and end of the amplified product.

---

**3.3.2.1.c. Contig D9S114**

Eight amplification products were cloned and sequenced from contig D9S114. The largest product was D9S114 Bam A and it showed strong similarity to the beta subunit of a number of guanine nucleotide-binding proteins (G proteins), regulatory proteins and transcription factors. All similarities were between plus strands of each sequence and in frame 3 for D9S114 Bam A. The strongest was with a yeast hypothetical G protein (p value e-15) which belongs to the beta transducin family. Figure 3.44 shows the similarities between the two sequences. The second strongest similarity was with the human LIS-1 protein (accession number S36113) (p value e-12) which is deleted in
patients with Miller-Dieker syndrome (Reiner et al., 1993). The LIS-1 (lissencephaly-1) gene has been localised to human chromosome 17p13.3 and the amino-acid sequence shows significant similarity to beta subunits of heterotrimeric G proteins. Other proteins with significant similarity to D9S114 Bam A included the yeast regulatory protein TUP1 (accession number JN0133), yeast probable transcription factor associated protein (accession number S34023), Drosophila transcription factor IID-associated protein (accession number S33263) and African clawed frog beta-transducin repeat containing protein (accession number B48088).

Figure 3.44

Sequence similarity between D9S114 Bam A and a yeast hypothetical guanine nucleotide binding protein

Frame +3, identities 28/57 (49%), similarities 35/57 (61%), p value 3.9e-15

<table>
<thead>
<tr>
<th>D9S114 Bam A</th>
<th>Yeast protein</th>
</tr>
</thead>
<tbody>
<tr>
<td>24</td>
<td>403</td>
</tr>
<tr>
<td>KAVSSVKFSPNGGWLASSSADKLIKIWGAYDGKFE</td>
<td>KLVNHVAFSPDGRYIVSASPDNSIKLWDGRDGKFI</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>D9S114 Bam A</th>
<th>Yeast Protein</th>
</tr>
</thead>
<tbody>
<tr>
<td>128</td>
<td>438</td>
</tr>
<tr>
<td>KTMSGHKLGISDVAWSSDSNLL</td>
<td>STFRGHIASVYQVAWSSDCRLL</td>
</tr>
<tr>
<td>194</td>
<td>459</td>
</tr>
<tr>
<td>T GH + VAWSS DL</td>
<td></td>
</tr>
</tbody>
</table>

Taken from a BLASTX search, the D9S114 Bam A sequence was translated in all three frames on both DNA strands and then compared to the protein database
D9S114 Bam C gave the same sequence similarities as D9S114 Bam A and they were found to overlap. Both sequences started at the same 5' position and were almost identical for 100bp. Then D9S114 Bam A had an additional sequence of 74bp which was not present in D9S114 Bam C. The sequence similarity continued at the end of this 74bp sequence until the 3' end. It suggested that the extra sequence in D9S114 Bam A was an exon not present in D9S114 Bam C. Exon skipping or alternative splicing does occur during splicing of normal cellular genes (Andreadis et al., 1987), these results suggest that it also occurs during splicing of the pSPL3 derived transcripts.

The complete sequence of D9S114 Bam B was 229bp in length and it showed two separate regions of strong similarity with the human set gene as illustrated in figure 3.45. The similarity between D9S114 Bam B and set was within the 3' coding region and the 3' untranslated region of the gene. The set gene had been previously localised to human chromosome 9q34 proximal to the ABL oncogene at 9q34.1 (von Lindern et al., 1992). However the cosmid that D9S114 Bam B was isolated from, had been mapped by FISH distal to the SD-1 translocation breakpoint which is itself within the ABL oncogene on human chromosome 9q34.1. This result and the position of the cosmid within contig D9S114 did not correspond with the previous localisation for set. However the strongest region of similarity from position 138-229bp on D9S114 Bam B was within the 3' end of the set gene and it had been reported that this region was not single copy in the human genome. Therefore it suggests that this product represents an additional copy of the gene within human chromosome 9q34.
D9S114 Bam D overlapped D9S114 Bam B for its entire length but showed no sequence similarity to the set gene. Both products started at the same position but D9S114 Bam D ended at position 138 on D9S114 Bam B which is where the strong similarity with set began. The minus strand of D9S114 Bam D showed strong similarity to the plus strand of the Miller-Dieker lissencephaly protein (LIS-1) (p value 5e-17). This protein also showed strong similarity to D9S114 Bam A and D9S114 Bam C but in both cases it was between plus strands. It suggested that the gene that D9S114 Bam B and D are derived from overlaps the gene that D9S114 Bam A and C are derived from, but they are on opposite DNA strands. The plus strand of D9S114 Bam D shows strong similarity to the human Ig associated MstII repeat of an epsilon heavy chain constant region.

D9S114 Pst B showed 99% similarity in 75 nucleotides with human clone P94_100 which was isolated from a human chromosome 9 cosmid by Church et al. (1994) and therefore they were assumed to have identity. It also had similarity to a human neuronal growth protein (GAP-43) (accession number M25667) (66% similarity in 73bp).

D9S114 Pst C showed similarity to a human sequence in the database and D9S114 Bam E and D9S114 Pst A showed similarity to silkworm and hamster sequences respectively (see table 3.9). No artefacts were detected from this contig.
Figure 3.45

Sequence similarity between D9S114 Bam A and the human set gene

A. Region with 92% similarity in a 92 nucleotide overlap

|            | D9S114 Bam B | 138 | ATATCCTTGCA GTTTAAGATGATACTTTAAAAAT |
|            | Human set    | 1454| ACATCCTTGCA GTTTAAGATGACACTTTAAAAAT |
|            | D9S114 Bam B | 172 | GAATTATCTCCTAATGATTACTTGA GCCCTGCCA |
|            | Human set    | 1488| AAATTCTCTCCTAATGACCTTGA GCCCTGCCA |
|            | D9S114 Bam B | 206 | ATCAATGGGA AATCAGCAGAACC |
|            | Human set    | 1522| CTCAATGGGA AATCAGCAGAACC |

B. Region with 80% similarity in a 25 nucleotide overlap

|            | D9S114 Bam B | 42  | GTGGCTGAAGGGA GGAAGAAGGAGCAG |
|            | Human set    | 763 | GAGGATGAAGGGA GGAAGGATGAAG |

3.3.2.1.d. Contig 37

Nine different sized products were cloned and sequenced from contig 37. Ctg37 Bam A showed strong similarity to a rat protein tyrosine kinase. Ctg37 Bam C, Ctg37 Bam D, Ctg37 Bam E, Ctg37 Pst A and Ctg37 Pst B all showed similarity to non-human sequences in the databases searched (see table 3.9). Ctg37 Pst D was found to be a chimeric product with HIV-1 sequence at the 3’ end. Artefacts included Ctg37 Bam B
which contained Lorist B sequence and Ctg37 Pst C which contained only HIV-1 sequence. The HIV-1 sequence was identical to D9S10 Bam B and SURF Pst C and therefore the same cryptic vector splice junctions had been used (see figure 3.43).

3.3.2.1.e. Contig 166

Nine different products were cloned and sequenced from contig 166. Of these, five were deduced to be artefacts. Three were composed entirely of HIV-1 sequence as shown in figure 3.43 (Ctg166 Bam B, Ctg166 Pst E and Ctg166 Pst F) and two contained Lorist B vector sequence (Ctg166 Pst B and Ctg166 Bam C). Ctg166 Pst A and Ctg166 Pst C showed strong similarity to human sequences, Ctg166 Pst D had strong similarity to a human procollagen sequence and Ctg166 Bam A showed similarity to chick metallothionein (see table 3.9).

3.3.2.1.f. Contig L-4DD1

A total of 9 different sized products were cloned and sequenced from contig L-4DD1 and four of these were found to be chimeric with HIV-1 sequence at the 3’ end. The total number of chimeric products for the 48 amplification products sequenced was only 6 and therefore it was deduced that the BstXI digestion was incomplete for this contig. The RT-PCR analysis shown in figure 3.38 appears to confirm this as many bands were visible and they were rather smeary compared to RT-PCR analysis of the other contigs (figures 3.33-3.37).
The remaining products showed strong similarity to either human (L-4DD1 Bam C, L-4DD1 Pst A and L-4DD1 Pst B) or non-human sequences in the database (L-4DD1 Bam B, L-4DD1 Bam E) (see table 3.9).

3.3.2.2. Hybridisation analysis

The aim of the hybridisation analysis was to confirm that amplification products were true exons. Initially Southern blot analysis was used to confirm that they hybridised to their corresponding genomic clones. Evolutionary conservation of a DNA sequence has been previously used to detect genes in genomic DNA (Monaco et al., 1986; Rommens et al., 1989; Call et al., 1990) and therefore the products were also hybridised to zooblots with DNA from a number of different species. Cross-species hybridisation indicated that the product was part of a gene and it was then used to screen a cDNA library and thereby isolate homologous transcripts. A foetal brain cDNA library was used because of its complexity and because of the expectation that TSC1 is expressed in brain.

Analysis was concentrated on a comparatively small number of amplification products and the criteria for selection was based on positional information for the contig they were from and sequence information. Three putative exons from the Surfeit contig (SURF Bam A, SURF Bam B and SURF Bam C) were analysed further because they were known to lie within the conservative flanking markers for TSC1 (Nahmias et al., 1995) and very little is known about function of the Surfeit genes.
Two products from contig D9S114 (D9S114 Bam A and D9S114 Pst B) and one product from contig 37 (Ctg37 Bam A) were selected primarily by their sequence similarities, which might suggest a possible role as a TSC1 candidate gene. It has been proposed that TSC1 and TSC2 function as tumour suppressor genes (Smith et al., 1993; Green et al., 1994a; Green et al., 1994b; Carbonara et al., 1994) and this might suggest an involvement in signalling pathways related to cellular growth control. The small region of similarity between the protein sequence of Tuberin (1593-1631) and a human GTPase-activating protein (GAP-3) also suggests its function in signal transduction. D9S114 Bam A and D9S114 Bam C both showed strong similarity to the beta subunit of G proteins and D9S114 Pst B showed similarity to a human GTPase-activating protein (GAP-43). Although the significance of these similarities was uncertain, their potential for isolating a candidate gene seemed promising. Protein kinases are also involved in the regulation of information transfer within and between cells, activation of receptor linked and cytoplasmic protein tyrosine kinases is crucial in the control of cell growth and differentiation (Ullrich and Schlessinger, 1990). Therefore Ctg37 Bam A was also of interest as it had strong similarity with a rat protein tyrosine kinase. Similarly, L-4DD1 Bam B may also be of value, but it was not pursued in this study because a collaborating group was involved in a comprehensive search for transcribed sequences in this particular region of the genome (Kwiatkowski, personal communication). Although contig positional information could not confirm that the three putative exons selected were localised in the candidate region, they definitely deserved investigation. D9S114 is the flanking marker for the consensus target region and therefore depending on the position of the corresponding genes from this contig with respect to D9S114, they may be in the critical region. As for contig 37, the only positional information known is that it lies within the SD-1 and 9T12 translocation.
breakpoints. A complete clone map has not yet been attained for this region, but it is hoped that as contig assembly progresses, this contig will be mapped more precisely.

D9S114 Bam B was also analysed further, as sequence searches showed similarity to set, a gene that had been previously localised proximal to ABL (von Lindern et al., 1992). The identification of a transcript for this particular putative exon would clarify whether this was a homologous gene, a multiple copy of the set gene or maybe that the initial positioning had been incorrect.

3.3.2.2.a. Contig SURF

Three products from the SURF contig (SURF Bam A, SURF Bam B and SURF Bam C) are shown to hybridise to genomic clones in figure 3.46. SURF Bam A and SURF Bam C hybridised to both cosmids 255 A6 and 211 A7 but the 2 PCR products hybridised to different sized BamHI fragments. In contrast, SURF Bam B only hybridised to 255 A6. SURF Bam A and SURF Bam B showed strong similarity to the 5' flanking region of L7a ribosomal protein (SURF-3) and it was suggested that they were derived from SURF-5. The organisation of the amplification products within SURF-5 that is shown in figure 3.42 and the hybridisation patterns in figure 3.46 suggest that the 5' end of the gene is within cosmid 255 A6 and that the 3' end is within 211 A7. This is shown in figure 3.47. As the contig has been orientated on the chromosome and 211 A7 is known to be more proximal that 255 A6, the Surfeit gene cluster has thereby been orientated with SURF-5 positioned proximal to SURF-3.
Figure 3.46

Hybridisation of putative exons from the SURF contig to their corresponding genomic clone

A. SURF Bam A

B. SURF Bam B

C. SURF Bam C

Figure 3.47

Organisation of SURF Bam A, SURF Bam B and SURF Bam C within their corresponding genomic clones and orientation of the Surfeit gene cluster

Centromere —— Telomere

SURF Bam C —— SURF Bam A —— SURF Bam B

211 A7

B B B

9kb 6kb 4kb

255 A6

B B B

SURF-5 ———— SURF-3

B - BamHI restriction site with sizes in kilobase pairs indicated.
Arrows for the SURF genes represent direction of transcription.
Cosmid clones are represented as black lines. Products are shown as boxes. SURF Bam B hybridises to only 255 A6, SURF Bam A and SURF Bam C hybridise to both. Note that the diagram has not been drawn to scale.

SURF Bam C was also positioned within the overlap between cosmids 211 A7 and 255 A6 but as it hybridised to different sized BamHI fragments compared to SURF Bam A and different results were obtained on a sequence search, it suggests that this product was derived from a different gene.
Products SURF Bam A and SURF Bam B were conserved between human, orang-utan, African green monkey and dog genomic DNA as shown in figure 3.48. SURF Bam A was also conserved in mouse genomic DNA and therefore it was used as a hybridisation probe on the European Interspecific backcrosses (EUCIB; The European backcross collaborative group, 1994). Large numbers of progeny were produced by interspecific backcross between C57BL/6 (laboratory) and *Mus spretus* (wildtype) mouse strains that allow high resolution genetic mapping of markers. The analysis was performed by Dr. A. Pilz (1995) and SURF Bam A was positioned on proximal mouse chromosome 2, a region homologous to distal human chromosome 9 (9q33-q34). The localisation was the same as mouse homologues for DBH and SURF-1 and was consistent with SURF Bam A being an expressed sequence from the Surfeit gene cluster.

3.3.2.2.b. Contig D9S114

Figure 3.49 shows three products from D9S114 hybridised to their corresponding genomic clone. Both D9S114 Bam A and D9S114 Pst B mapped to cosmid 157 B7. As different restriction enzyme digests were used for the Southern blots, it was not determined if the same genomic fragment was homologous. D9S114 Bam B mapped to cosmid 222 A10. These results were all consistent with the cosmids from which the amplification products were derived.

Figure 3.50 shows D9S114 Bam A and D9S114 Pst B used as hybridisation probes on zooblots. Both products were conserved in human, orang-utan, African green monkey and dog genomic DNA. D9S114 Bam A was also conserved in guanaco, mouse, rat, trout and *Xenopus*. It cannot be assumed that D9S114 Pst B was not also conserved in these species because the zooblot used only had four different types of genomic DNA.
Hybridisation of putative exons from the D9S114 contig to their corresponding genomic clones

A. D9S114 Bam A

B. D9S114 Bam B

C. D9S114 Pst B

Figure 3.50

Hybridisation of putative exons from the D9S114 contig to zooblots

EUCIB interspecific backcross analysis was used to genetically map D9S114 Bam A in mouse. The amplification product was localised to proximal mouse chromosome 2, distal to mouse homologues of VAV2, DBH and SURF-1 and proximal to RXRA and COL5A1 (Pilz, 1995).

3.3.2.2.c. Contig 37

Ctg37 Bam A mapped to a 3kb genomic fragment in both cosmids 180 F1 and 288 E8. (figure 3.51A). Hybridisation of the product to a zooblot showed it to be conserved in primates, dog and mouse (figure 3.51B), however these were the only species tested.
Figure 3.51

Hybridisation of a putative exon from contig 37 (Ctg37 Bam A) to the corresponding genomic clones and a zooblot

A. Cosmid Southern blot

B. Zooblot

A. Cosmid DNA digested with *BamHI* restriction enzyme. Lanes: 1. 194 E12; 2. 84 B3; 3. 180 F1; 4. 288 E8; 5. 26 10D; 6. 222 A10; 7. 157 B7; 8. 124 E8; 9. 255 H4.

B. Genomic DNA digested with *BamHI* restriction enzyme. Lanes: 1. human; 2. orangutan; 3. African green monkey; 4. dog; 5. mouse. Sizes of hybridisation signals given in kilobase pairs (kb).
3.3.3. Isolation of cDNAs from contig SURF, D9S114 & 37

The amplification products that were conserved on zooblots and therefore deduced to be transcribed sequences, were used as hybridisation probes to screen cDNA libraries. These included SURF Bam A, SURF Bam B, D9S114 Bam A, D9S114 Pst B and Ctg37 Bam A. Products SURF Bam C and D9S114 Bam B were also used, although they had not been hybridised to a zooblot.

3.3.3.1. Screening a gridded cDNA library

A human foetal brain gridded cDNA reference library was used initially which consisted of 100,000 clones (Lehrach \textit{et al.}, 1990). A pool of five probes including SURF Bam A, SURF Bam B, SURF Bam C, D9S114 Bam A and Ctg37 Bam A were used in a single screen and three positive clones were obtained (figure 3.52). The co-ordinates 70X,54Y; 70X,72Y and 118X,118y correspond to clones ICRFp507001115; ICRFp507101115 and ICRFp507109147 respectively.

3.3.3.1.a. Hybridisation and sequence analysis

Clones ICRFp507001115 and ICRFp507101115 were received with further information regarding their identity. They had both been previously recognised by a SURF-5 probe (T. Duhig, ICRF, London). This correlated with the suggestion that SURF Bam A and SURF Bam B were both transcribed sequences from SURF-5 (see section 3.3.2.1.a).

DNA was isolated from the clones, it was digested and then Southern blotted. A cDNA clone from the HGMP-RC (M-02-H01) was also included in the analysis because it showed strong similarity to SURF Bam C by sequence search (see section 3.3.2.1.a).
Hybridisation of pooled exons from contigs SURF, D9S114 and 37 to a gridded human foetal brain cDNA library

Pooled radio-labelled probes included SURF Bam A, SURF Bam B, SURF Bam C, D9S114 Bam A and Ctg37 Bam A. Arrows show the sites of hybridisation and the corresponding co-ordinate numbers.
SURF Bam A, SURF Bam B and SURF Bam C were each hybridised to the Southern blot and the results are shown in figure 3.53. SURF Bam A and SURF Bam B both hybridised to ICRFp507001115 and not ICRFp507011115. This corresponded to the cDNA clone that was recognised as SURF-5 and showed the most intense hybridisation signal on the autoradiograph (see figure 3.52, co-ordinate 70X,54Y). The clone was partially sequenced by Janet Young, but no significant similarity was found on searching the protein or nucleotide database.

SURF Bam C did not hybridise to this clone (ICRFp507001115) but as expected it did hybridise to the one it showed sequence similarity to (HGMP M-02-H01). Clone ICRFp507109147 was assumed to be homologous to either D9S114 Bam A or Ctg37 Bam A. However at the time of this analysis, positive cDNA clones for both of these amplification products had already been identified using a different library (Stratagene) and therefore they were not pursued.

### 3.3.3.2. Screening lambda ZAP® II cDNA library

A human foetal brain lambda ZAP®II library (Stratagene) was screened with a pool of amplification products that included D9S114 Bam A, D9S114 Bam B, D9S114 Pst B and Ctg37 Bam A. A million plaques were screened and twenty cDNA clones gave hybridisation signals that were present in duplicate (examples shown in figure 3.54). These plaques were picked and rescreened until they were present as single clones. An in vivo excision procedure was used to obtain the pBluescript® phagemid from the lambda ZAP® II vector and DNA was isolated from each.
Figure 3.53

Hybridisation of exons from the SURF contig to cDNA clones

A. cDNA Southern blot

1kb 1 2 3 4

2.0 -

1.0 -

0.4 -

B. SURF Bam A

1 2 3 4

C. SURF Bam B

1 2 3 4

D. SURF Bam C

1 2 3 4

A. cDNA clones digested with restriction enzymes HindIII and PstI and Southern blotted, 1kb ladder shown. B. Autoradiograph of probe SURF Bam A.
C. Autoradiograph of probe SURF Bam B. D. Autoradiograph of probe SURF Bam C.
Lanes: 1. cDNA 70X, 54Y; 2. cDNA 70X, 72Y; 3. cDNA 118X, 118Y; 4. cDNA HGMP-RC M-02-H01. Sizes of hybridisation signals given in kilobase pairs (kb).
Hybridisation of pooled exons from contigs D9S114 and 37 to a human foetal brain cDNA library

A. replica 1

B. replica 2

Pooled radio-labelled exons included D9S114 Bam A, D9S114 Bam B, D9S114 Pst B and Ctg37 Bam A.
3.3.3.2.a. Hybridisation analysis

The cDNA clones were numbered from 1 to 20, digested with restriction endonucleases and Southern blotted. Enzyme EcoRI was used to release the insert and BglII was used to cleave the insert but not the vector (figure 3.55). Previous restriction analysis of clones 16 and 18 had found them to be identical and therefore clone 18 was not present on the gel. Southern blots were individually hybridised with each of the four exons. Hybridisation with D9S114 Bam A and D9S114 Pst B is shown in figure 3.56 and twelve of the clones were homologous to both sequences (As clone 18 was identical to number 16 gave a total of 13.) Strength in hybridisation signal did vary between probes for some of the clones. For example clones 1, 4, 5 and 16 were much weaker with D9S114 Pst B than D9S114 Bam A. As approximately the same amount of DNA was loaded onto the gel, the significance of this finding was unclear, but it suggested that some of the clones only contained partial exon sequences. The only major difference between hybridisation patterns for each exon was that clone 17 was positive with D9S114 Pst B but not D9S114 Bam A. Therefore it was deduced that both D9S114 Bam A and D9S114 Pst B were transcribed sequences from the same gene.
Note that twenty cDNA clones were isolated. However number 18 was not present on this gel as it had already been confirmed to be identical to number 16.
Figure 3.56

Hybridisation of exons D9S114 Bam A and D9S114 Pst B onto Southern blot of cDNA clones.

A. D9S114 Bam A

B. D9S114 Pst B

Nineteen cDNA clones were present on the Southern blot (see figure 3.55). Clones that hybridised to the exons have been numbered.
**Figure 3.57 A** shows hybridisation of D9S114 Bam B to the cDNA clones. Positive signals were obtained for clones number 3 and 7. These clones were not homologous to any of the other probes used.

**Figure 3.57B** shows hybridisation of Ctg37 Bam A to the cDNA clones and positive signals were obtained for both numbers 15 and 20. These clones were not homologous to any of the other probes used. Therefore cDNA clones had been identified for all four exons used in the library screen. Clones number 13 and 14 were not positive with any of the probes and were therefore assumed to be due to errors in picking.

A restriction enzyme digest of the cDNA clones was also hybridised with radio-labelled polyT (Pharmacia) to determine clones that could be complete at the 3' end. Figure 3.58 shows that the positive clones were numbers 10, 13 and 20.
Figure 3.57

Hybridisation of exons D9S114 Bam B and Ctg37 Bam A onto Southern blot of cDNA clones

A. D9S114 Bam B

B. Ctg37 Bam A

Nineteen cDNA clones were present on the Southern blot (see figure 3.55). Clones that hybridised to the exons have been numbered.
Hybridisation of polyT onto Southern blot of cDNA clones.

A. Twenty cDNAs digested with EcoRI and BglII

B. Autoradiograph

cDNA clones that hybridised to polyT have been numbered.
The largest clones homologous with each exon were identified for further analysis. Clone number 5 had an insert size of approximately 2.5kb, it was strongly homologous to D9S114 Bam A but weakly homologous to D9S114 Pst B. Clone number 10 had an insert size of approximately 3.0kb and gave an equal hybridisation signal for both D9S114 Bam A and D9S114 Pst B. Clone number 7 had an insert size of approximately 2.0kb and was homologous to D9S114 Bam B. Finally clone number 20 had an insert size of approximately 3.0kb and was homologous to Ctg37 Bam A. The selected cDNA clones were radiolabelled and hybridised to cosmid blots to confirm their localisation. The results are shown in a summary table (table 3.10). As expected they all mapped to the cosmids from which the exon sequences they were homologous to, were derived.

The majority of cDNA clones identified were from contig D9S114 and as this marker is the distal flanking marker for TSC1, clones number 5, 7 and 10 were mapped further. Contig assembly had progressed significantly since the analysis began and the joining of contig D9S114 and contig D9S10 was thought to be imminent (N. Hornigold, personal communication). Therefore cosmids were selected to take a minimum tiling path from one end of contig D9S10 (see figure 3.25A) and across contig D9S114 (see figure 3.26). Cosmids were digested with BamHI restriction enzyme (figure 3.59), Southern blotted and individually hybridised with clones 5, 7 and 10. The results are shown in figures 3.60-3.62. Clone 5 was homologous to cosmid 157 B7, clone 7 was homologous to cosmids 66 A9 and 17B2 and clone 10 was mainly homologous to cosmids 206 C11, 210 D12 and 157 B7. However clone 10 also had weak hybridisation signals for most of the other cosmids on the Southern blot suggesting that it contained some repetitive sequence in addition to the unique sequence. This was confirmed by hybridisation to a
Table 3.10
Characterisation of 20 cDNA clones isolated from the Stratagene library

<table>
<thead>
<tr>
<th>cDNA no.</th>
<th>Exon hybridisation (hybn)</th>
<th>cDNA size</th>
<th>PolyT hybn</th>
<th>Cosmid hybridisation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>D9S114 Bam A / Pst B</td>
<td>2.5kb</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>D9S114 Bam A / Pst B</td>
<td>3.6</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>D9S114 Bam B</td>
<td>0.5kb</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>D9S114 Bam A / Pst B</td>
<td>1.1kb</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>D9S114 Bam A / Pst B</td>
<td>2.8kb</td>
<td>NO</td>
<td>2.5kb BamHI 157 B7</td>
</tr>
<tr>
<td>6</td>
<td>D9S114 Bam A / Pst B</td>
<td>3.6kb</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>D9S114 Bam B</td>
<td>2.6kb</td>
<td>NO</td>
<td>222 A10</td>
</tr>
<tr>
<td>8</td>
<td>D9S114 Bam A / Pst B</td>
<td>3.7kb</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>D9S114 Bam A / Pst B</td>
<td>3.0kb</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>D9S114 Bam A / Pst B</td>
<td>3.5kb</td>
<td>YES</td>
<td>2.5kb BamHI 157 B7</td>
</tr>
<tr>
<td>11</td>
<td>D9S114 Bam A / Pst B</td>
<td>3.5kb</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>D9S114 Bam A / Pst B</td>
<td>1.4kb</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>NONE</td>
<td>1.2kb</td>
<td>YES</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>NONE</td>
<td>1.8kb</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>Ctg37 Bam A</td>
<td>3.0kb</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>D9S114 Bam A / Pst B</td>
<td>2.5kb</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>D9S114 Pst B</td>
<td>0.8kb</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>D9S114 Bam A / Pst B</td>
<td>2.5kb</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>D9S114 Bam A / Pst B</td>
<td>1.8kb</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>Ctg37 Bam A</td>
<td>3.0kb</td>
<td>YES</td>
<td>180F1/288E8</td>
</tr>
</tbody>
</table>

Shaded rows indicate cDNAs analysed further.
Figure 3.59

*Bam*HI restriction enzyme digest of cosmids that take a minimum tiling path across contigs D9S10 and D9S114

Figure 3.60

Hybridisation of cDNA clone 5 to Southern blot of cosmids that take a minimum tiling path across contigs D9S10 and D9S114

Figure 3.61

Hybridisation of cDNA clone 7 to Southern blot of cosmids that take a minimum tiling path across contigs D9S10 and D9S114

Figure 3.62

Hybridisation of cDNA clone 10 to Southern blot of cosmids that take a minimum tiling path across contigs D9S10 and D9S114

Southern blot of genomic DNA as smears were visible in each lane rather than discrete bands (data not shown).

Clones 5 and 10 were assumed to be derived from the same gene as they were homologous to the same exons, but the genomic extent of clone 10 was greater. Both clones hybridised to the cosmid from which the exon sequences were derived (157 B7) but clone 10 also hybridised to a cosmid from the opposite end of contig D9S114 (210 D12) (see figure 3.26) and a cosmid from the D9S10 contig (206 C11) (see figure 3.25A). As this clone was assumed to have a polyA tail it was suggested that the 3’ end of the gene was within contig D9S10.

3.3.2.2.b. Sequence analysis

Clones 5, 7, 10 and 20 were partially sequenced by Janet Young using M13 vector primers (appendix 4) and then compared to the exons which had been isolated. This was possible using a BLAST search, as with the help of Dr. J. White all the sequences had been entered into a separate database. Nucleotide and protein databases were also searched using BLAST, BLASTX and FASTA computer programmes.

3.3.2.2.b.i. Clone 10

The sequence obtained from clone 10 showed partial D9S114 Pst B, D9S114 Bam A and D9S114 Bam C exon sequences adjacent to one another at one end of the clone. It was consistent with the three exons hybridising to the same cosmid. However only the 3’ end of D9S114 Pst B was present because of an interruption during cDNA cloning. The sequence identity between D9S114 Bam A, D9S114 Bam C and cDNA 10 halted at the point where the two exons differed and then continued into unrecognisable
sequence. This particular position was thought to be the start of an alternatively spliced exon which was present in D9S114 Bam A but missing in D9S114 Bam C. Therefore maybe as more sequence information becomes available for clone 10, the 3’ ends of D9S114 Bam A and D9S114 Bam C may become apparent.

As the orientation of the exons was known, the end of the cDNA with these sequences present, was confirmed to be the 5’ end. However the clone could not be complete at because exon amplification does not allow for selection of the first or last exons of genes. Surprisingly, sequence analysis at the opposite end of the clone 10 did not reveal a polyA tail. This might suggest that the clone is chimeric and that the sequence homologous to polyT is within the middle. Alternatively, the clone may contain a high proportion of A residues which have resulted in a false positive signal. The 3’ end of this clone was assumed to hybridise to contig D9S10 and if it is chimeric, the probability of having two different cDNAs in the same clone which both mapped to chromosome 9 must be very low. Sequence database searching using the 5’ end did show strong similarity to the human fragile X gene (FMR1 accession number L29074) and thereby generated further doubt on the authenticity of this clone. Therefore although further sequencing would elucidate the problem, analysis turned to clone number 5 which was homologous to the same exon sequences.

3.3.2.2.b.ii. Clone 5

Clone 5 was sequenced at each end and a BLASTX search revealed strong similarity with G proteins and proteins containing WD40 repeats. This was encouraging as the similarities were consistent with those obtained from searches with the exon sequences.

A strategy of sequence walking was applied by Janet Young and to date the clone is
nearly completely sequenced (approximate size 2.5kb, 2.0kb sequenced). The exons are present within the middle of the cDNA, they are complete and they have enabled the clone to be orientated. The alternatively spliced exon thought to be represented within D9S114 Bam A and not D9S114 Bam C is present within this clone. The work is ongoing and the aim is to identify a full length transcript for this gene. As a large number of cDNA clones were isolated using these exon probes but not analysed, the initial work will involve their characterisation.

3.3.2.2.b.iii. Clone 7

Each end of the clone has been sequenced and compared to protein and nucleotide databases, however no similarity with the set gene or any other sequence was found. It would seem to indicate that this is a novel gene. Further characterisation of this cDNA clone is currently in progress (Janet Young, personal communication).

3.3.2.2.b. iv. Clone 20

Again, each end of the clone has been sequenced but computer searches did not reveal and significant similarities. Further characterisation of this cDNA clone is currently in progress (Janet Young, personal communication).

3.3.2.2.c. FISH analysis

3.3.2.2.c.i. Clone 5

Clone number 5 has been confirmed to map to human chromosome 9q34 by FISH performed by Margaret Fox (see figure 3.63A). However additional sites of hybridisation were also detected including chromosome 1p. These extra signals may reflect pseudogenes or maybe indicate that the gene is a member of a gene family. The
sequence similarities would certainly suggest that the latter is very likely. The gene does
seem an attractive candidate for TSC1 but more extensive characterisation will be
required. Primarily, the genomic extent of the full length transcript must be determined.
If this indicates that the gene is proximal to D9S114 then it may have even greater
importance.

3.3.2.2.c.ii. Clone 20

The clone has been localised to human chromosome 9q34 by FISH (performed by Dr.
M. Fox) and is shown in figure 3.63B. The significance of this clone with respect to
identification of the TSC1 gene, will become clearer once the contig has been more
firmer localised.
Figure 3.63

Fish mapping of cDNA clones 5 and 20 onto normal human metaphase chromosomes

A. Clone number 5

Signal on chromosome 9q34 (large arrows) but additional sites including chromosome 1p (small arrows)

B. Clone number 20

Signal on chromosome 9q34
Chapter 4

Discussion
Chapter 4

4. Discussion

4.1. Positional cloning of the TSC1 gene

4.1.1. Localisation of the TSC1 candidate region

4.1.1.1. Haplotype analysis

Definition of the TSC1 candidate region on chromosome 9 by linkage analysis, is complicated by locus heterogeneity, as it is difficult to distinguish recombination events from non-linkage. The problem is magnified by predominant small family size, as only a large family can contain a critical recombinant and still maintain good evidence for linkage to chromosome 9. Consequently, only one reported tuberous sclerosis family shows absolutely convincing segregation for TSC1 and this is a large four generation family (lod score 3.86, $\theta = 0$) (Nellist et al., 1993).

Povey et al. (1994a) reported four families with lod scores of between 1.0 and 2.0 and several families with lod scores about 1.0 for markers on chromosome 9q34. It is reasonable to deduce that these families are linked to TSC1, especially if there are clear recombination events in affected individuals that exclude the TSC2 gene, however it is not completely foolproof. In addition, a few families do not appear to segregate for either TSC1 or TSC2 and may suggest the possibility of a third locus (Povey et al., 1994a). No significant positive lod scores have been obtained for any other regions of the genome, but the possible existence of a TSC3 locus cannot be excluded and will be discussed further in section 4.4.
When this study commenced the candidate region for TSC1 was considered to be between D9S64 and D9S14 and this interval had a genetic distance of 15cM (Povey et al., 1992). Subsequent identification of a second TSC locus on chromosome 16p13.3 (TSC2) (Kandt et al., 1992) and re-evaluation of the linkage data indicated that the vast majority of TSC families could be accounted for by a mutation in either the TSC1 or TSC2 gene. Therefore haplotype analysis of TSC pedigrees that have excluded linkage to TSC2 have been used to refine the position of the TSC1 gene.

The current consensus definition of the TSC1 candidate region is between D9S149 and D9S114 (Povey et al., 1994b; Pericak-Vance et al., 1995) (figure 4.1). This interval has a genetic distance of 3.6cM (Povey et al., 1994b) and a physical distance of between 1.5 and 3.0Mb (Povey et al., 1994b; Nahmias et al., 1995). Although these flanking markers have not been disputed, they have only been verbally reported (Third International Workshop on chromosome 9) and defined by a small number of critical recombination events. The reports were made by Dr. M. Smith and Dr. H. Northrup respectively and were both defined by recombination events in affected individuals. Although this interval has been referred to as the consensus candidate region in this thesis, it may be wise to remain a little sceptical until details of pedigrees including haplotype analysis and clinical investigation have been properly documented. There are many more families which have recombinant events that are either more proximal or more distal to these flanking markers (Povey et al., 1992; Kwiatkowski et al., 1993). But again most of these have only been reported verbally and therefore are not entirely convincing. In contrast, a comprehensive study within 32 families by Povey et al. (1994a) suggested that TSC1 would lie between D9S113 and D9S298. It may be more prudent to consider
these markers as the flanking markers of the disease gene (figure 4.1) as the results have been reported in detail and are well substantiated.

Additional critical recombination events in TSC1 families have been reported which reduce the candidate region even further. Unfortunately they suggest conflicting positions for the TSC1 gene. The events are summarised in figure 4.1. The different localisations may be due to incorrect diagnosis of the individuals concerned, incorrect assessment of linkage to a locus on chromosome 9, a new mutation arising in an existing family, one very large gene, two genes or haplotyping errors.

Nellist et al. (1993) reported a large chromosome 9 linked family with two key recombination events which placed the TSC1 gene proximal to DBH and distal to Gelsolin (GSN). The family was later analysed with marker D9S150 and the candidate region was refined as proximal to D9S150 (Sampson and Harris, 1994). These critical recombination events occurred in unaffected individuals that had been through an extensive clinical investigation (Nellist et al., 1993). Although penetrance of the disease is high it may not be complete and therefore if these individuals are examples of non-penetrance it would place TSC1 proximal to GSN or distal to D9S150.

Further evidence for a proximal localisation of the TSC1 gene has been provided by Janssen et al. (1994) after linkage analysis in 14 Dutch families. However it was not until very recently that a pedigree and details of the family were available for analysis (Janssen, 1995). A critical recombination event in an affected individual placed the gene proximal to ABO and thereby reduced the TSC1 candidate region to an interval between D9S149 and ABO. The genetic distance for this interval has been defined as 1cM.
Figure 4.1

Critical recombination events within TSC pedigrees that have excluded linkage to TSC2 and positioned the TSC1 gene

<table>
<thead>
<tr>
<th>Centromere</th>
<th>Distal marker</th>
<th>Proximal marker</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ASS</td>
<td>Povey et al., 1994b</td>
</tr>
<tr>
<td></td>
<td>D9S113</td>
<td></td>
</tr>
<tr>
<td></td>
<td>D9S64</td>
<td>Povey et al., 1992*</td>
</tr>
<tr>
<td></td>
<td>D9S125</td>
<td></td>
</tr>
<tr>
<td></td>
<td>D9S149</td>
<td>Povey et al., 1994b*</td>
</tr>
<tr>
<td></td>
<td>ABO</td>
<td>Janssen et al., 1994*</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Janssen, 1995</td>
</tr>
<tr>
<td></td>
<td>D9S150</td>
<td>Nellist et al., 1993^</td>
</tr>
<tr>
<td></td>
<td>DBH</td>
<td>Pericak-Vance et al., 1993*</td>
</tr>
<tr>
<td></td>
<td>D9S122</td>
<td></td>
</tr>
<tr>
<td></td>
<td>D9S10/D9S66</td>
<td>Pitiot et al., 1994*</td>
</tr>
<tr>
<td></td>
<td>D9S114</td>
<td>Povey et al., 1994b*</td>
</tr>
<tr>
<td></td>
<td>D9S298</td>
<td>Povey et al., 1994a</td>
</tr>
<tr>
<td></td>
<td>D9S14</td>
<td>Povey et al., 1992*</td>
</tr>
<tr>
<td></td>
<td>D9S67</td>
<td></td>
</tr>
<tr>
<td></td>
<td>D9S158</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Telomere</td>
<td></td>
</tr>
</tbody>
</table>

Note the diagram is not to scale.
* Pedigree details not published
^ Critical recombinant unaffected
■ Consensus candidate region for TSC1 (Povey et al., 1994b; Pericak-Vance et al., 1995)
(Povey et al., 1994b) and the physical distance is less than 2Mb (see PFGE data in figure 1.7).

In contrast, recombination events in affected individuals from two different TSC families have localised TSC1 distal to DBH (Pericak-Vance et al., 1993) and distal to D9S10 (Pitiot et al., 1994). This would reduce the candidate region to an interval between D9S10 and D9S114, giving a genetic distance of 2cM (Povey et al., 1994b) and a physical distance of approximately 325kb (Murrell et al., 1995). However neither of these key recombination events have been published with any pedigree details and therefore it is hard to be entirely convinced of the results.

With such controversy over gene localisation, it is difficult to know which reports are correct. Murrell et al. (1995) constructed a complete clone map for a 1.7Mb region from DBH-D9S67 after concluding that the most probable localisation for TSC1 was between DBH and D9S114. Zhou et al. (1995) reported that the most likely candidate region for TSC1 was between D9S149 and D9S66, therefore a physical map was constructed between these two markers. Our research group has taken a more cautious approach and applied a mapping strategy which covers the entire candidate region (Nahmias et al., 1995; Woodward et al., 1995). Furthermore, it seemed prudent to select contigs from various positions within the candidate region for the subsequent isolation of transcribed sequences, as there was not overwhelming evidence for either a more proximal or more distal localisation of the TSC1 gene.
4.1.1.2. Loss of heterozygosity

On reviewing a book entitled Tuberous Sclerosis (Gomez, 1979) which described clinical details of the disease, Comings (1980) came to the conclusion that TSC would be analogous to retinoblastoma. He proposed that Knudson's two hit hypothesis (Knudson, 1971) would also apply to tuberous sclerosis and therefore two steps would be necessary for tumour (hamartoma) development. The first step would be a mutation in the gamete which resulted in an inherited predisposition to cancer and the second would be a somatic mutation which resulted in hamartoma development in any tissue. With these hereditary forms of cancer one normal allele of the gene can prevent tumour development and therefore the term tumour suppressor gene is given.

Unfortunately it seems that Comings' (1980) prediction was not taken very seriously as it was not until 1993 that evidence of TSC2 functioning as a tumour suppressor gene was reported (Smith et al., 1993). By using a panel of informative genetic markers the loss of one allele in tumour material can be detected by loss of heterozygosity. No allele loss was found in the vicinity of TSC1 on 9q34, but a renal angiolipoma from one patient and hamartomous lesions from the kidneys of another, showed allele loss for markers on 16p13.

A number of reports followed which demonstrated allele loss for markers in the region of TSC2 (Green et al, 1994a) and TSC1 (Green et al., 1994b; Carbonara et al., 1994). The loss of heterozygosity on chromosome 16p13.3 was consistent with the localisation of the cloned TSC2 gene (The European chromosome 16 tuberous sclerosis consortium, 1993). It is hoped that studies of allele loss on chromosome 9q34 may facilitate a refinement in the TSC1 candidate region and thereby aid positional cloning of the gene.
Carbonara et al. (1994) analysed a giant cell astrocytoma from an individual that was a member of a chromosome 9 linked TSC family. Allele loss was found for markers D9S149, D9S150, D9S122, D9S66 and D9S158 but markers D9S125, D9S114 and D9S67 were uninformative (see order of markers in figure 4.1). These results support the localisation of the TSC1 gene on chromosome 9q34 but could not reduce the candidate region. In this case the paternal haplotype had been lost which contained the putative normal TSC1 locus. This would result in complete loss of function of the TSC1 protein product and it is consistent with Knudson’s two hit hypothesis (Knudson, 1971).

No loss of heterozygosity was detected for markers in the vicinity of TSC2 or a further six recognised tumour suppressor genes (neurofibromatosis type 1, NF1; p53, TP53; familial breast carcinoma, BRCA1; von Hippel-Lindau, VHL; adenomatous polyposis coli, APC and familial malignant melanoma, MLM/CDKN2). However a small region of allele loss was detected on chromosome 9p21 which suggested the localisation of a new putative tumour suppressor gene. Only marker D9S104 had been lost and the development of more markers in this region will help to define the critical region.

Green et al. (1994b) studied hamartomatas from four sporadic and two familial cases of TSC. A total of eight markers were analysed which included ASS, D9S64, D9S149, D9S150, DBH, D9S66 and D9S67 (see order of markers in figure 4.1). One renal angiolipoma had lost markers ABO, DBH and D9S66 but not markers D9S149 or D9S67 and it was uninformative for D9S150. These results position the TSC1 gene between D9S149 and D9S67 and agree with the localisation obtained by haplotype analysis.
These results seem encouraging but the use of loss of heterozygosity studies in localising a disease gene should not be over-estimated. For example, the BRCA1 gene is thought to act as a tumour suppressor gene, loss of heterozygosity was found for markers from 17q12-q21 and the gene was subsequently cloned (Miki et al., 1994). However loss of heterozygosity was also found for a discrete region of 17q distal to the BRCA1 gene in 73% of breast and ovarian tumours analysed (Godwin et al., 1994) and therefore suggested the involvement of a second tumour suppressor gene.

In addition, the general consensus of opinion at the Fourth International Chromosome 9 Workshop (Pericak-Vance et al., 1995) was that allele loss on chromosome 9 is much less frequent than the allele loss on chromosome 16. This observation raises some questions as to whether TSC1 actually does function as a tumour suppressor gene, it is possible that the mechanism of TSC1 is different from that of TSC2.

4.1.1.3. DNA rearrangement

A sporadic TSC1 patient has been reported that has two maternal alleles for the D9S66 dinucleotide repeat polymorphism in 3 different tissue types tested (Smith et al., 1994a). Two colour interphase FISH has also shown a duplication for D9S66 that does not extend as far as ABL and in addition, PFGE has revealed a novel MluI fragment in this patient that is not in 15 other individuals including the parents. Karyotyping lymphoblastoid cells from the patient has not detected any cytogenetic abnormality but as a number of techniques have been used, the evidence seems quite convincing that a rearrangement has occurred.
The family is informative for 2 DBH polymorphisms, D9S122 and D9S149 and no indication of dosage differences have been detected with these markers in the affected individual (Smith et al., 1994a). In addition, no duplication was detected by two colour FISH for cosmid 17B2 which contains the D9S114 polymorphism (Smith, personal communication). It would therefore suggest that TSC1 is localised to an interval between D9S122 and D9S114, which has a physical distance of approximately 435bp (Murrell et al., 1995). The result correlates with the more distal localisation of TSC1 and supports the haplotype analysis reported by Pitiot et al. (1994).

If the tumour suppressor model for TSC1 is correct it suggests that the breakpoints of the DNA duplication may be within the disease gene and thereby result in loss of function. However a duplication may also result in over expression of one or more genes in the region (i.e. dosage effect) or altered gene expression by a position effect within the duplicated region. An example of an autosomal dominant disorder that is associated with a DNA duplication is Charcot-Marie-Tooth disease Type 1A (CMT1A) (Lupski et al., 1991). Although genetic heterogeneity for CMT1 had been suggested, the duplication was found to be transmitted to all affected individuals in seven multigeneration families without recombination and also present in several sporadic patients (Lupski et al., 1991). However in this case, the genetic heterogeneity for tuberous sclerosis and the detection of the duplication in an only a single sporadic patient does give rise to some reservation in interpreting the result, as there is always the chance, although maybe rather slight, that this DNA duplication is a red herring and that the actual disease gene is TSC2 or even TSC3.
4.1.2. Physical mapping human chromosome 9q34

An important part of positional cloning is physical mapping the region of the genome where the gene is thought to lie. In this study it involves the region on human chromosome 9q34 between markers D9S149 and D9S114. A combination of techniques were used to identify cosmid clones from the target region and radiation hybrids were a useful source of DNA. The cosmids isolated were then fingerprinted, organised into contigs (Nahmias et al., 1995) and mapped by FISH (Woodward et al., 1995).

Fluorescence in situ hybridisation has taken a major role in the physical mapping strategy. It has been used to analyse radiation hybrids for their human content and the results were consistent with the pattern of markers retained from human chromosome 9q (see section 3.1.1.)

FISH was also used to map cosmid clones onto normal human metaphase chromosomes (figure 3.3-3.5). However the resolution of FISH mapping was improved by using metaphase chromosomes from cell lines SD-1, 9T12 and 9T01. Both 9T12 and 9T01 had been previously used to map nine markers by PCR analysis on flow-sorted chromosomes (Zhou et al., 1992). However the approach used in this study, eliminates the need for flow-sorting and a variety of genomic clones (cosmid, YAC, phage) can be easily localised.

Other groups have mapped clones to chromosome 9 by fractional length measurements (Graw et al., 1992) or direct R-banding FISH (Takahashi et al., 1994). The method used in this study was more rapid than the former and more accurate than the latter as both cytogenetic R-bands and position with respect to the translocation breakpoint can be
used for localisation, therefore enabling band 9q34 to be sub-divided into 4 intervals (figure 3.7). The only limit to the resolution of mapping using this approach is the number and availability of cell lines with appropriate breakpoints.

Interphase mapping has also been used to order cosmids on chromosome 9q34 (own data not shown) but the analysis was labour intensive and therefore superseded by the more efficient mapping using translocation breakpoints. Unfortunately this technique does not enable the clones to be ordered within each interval though. Once the clones are assembled into contigs with known loci, this is no longer a disadvantage because there are good physical and genetic maps available for chromosome 9q34 (Povey et al., 1994b; Pericak-Vance et al., 1995). However contigs that are as yet anonymous in each interval remain unordered. Further interphase mapping and high resolution mapping using extended DNA fibres (Heiskanen et al., 1994) is currently being pursued (Ekong, personal communication). The technique should allow contigs to be ordered and also give an estimation of distances between them.

4.1.2.1. Isolation and mapping of clones from radiation hybrids

Two different approaches were used to isolate human clones from radiation hybrids. The direct approach in which cosmid libraries were generated from the hybrids and screened with human DNA, led to the isolation of only 4 human clones from each of the hybrids 17B and 20A. These low yields were disappointing and demonstrated the difficulty in constructing a representative library when the human DNA component is so much smaller than the total DNA used as the starting material. Within each hybrid cell the hamster chromosomes were sometimes greater than diploid, whereas the human DNA was only present as a single fragment. Furthermore, the human DNA fragment
was not always present in every cell (see table 3.2), the stability in hybrid 17B was 99% but this compares to only 60% in hybrid 20A. An additional problem with this approach was that five cosmids were identified as positive with human DNA, but they did not give a hybridisation signal on human metaphase chromosome by FISH. These were deduced to be hamster clones.

The second approach used Alu-PCR products from hybrid 17B to screen a total human genomic cosmid library (Fitzgibbon, 1993). Ten clones were identified but only four were found to map to human chromosome 9q34 (figure 3.4), the remainder mapped to other regions of the genome. These background hybridisation signals were obviously a disadvantage but the problem was resolved by using a gridded flow-sorted chromosome 9 library (LL09NC01P). The efficiency using this library is greater because chromosome 9 is present in 6 copies and by screening the same clones every time there is consistency with each experiment. However this library does have at least one gap in the vicinity of ABO that has been detected by our group and others (Povey et al., 1994b). The cosmid encoding this gene was isolated from a total genomic library (Cachon-Gonzalez, 1991). It emphasises the importance of every library, irrespective of size, as they can be used to complement one another in achieving a complete clone map for a particular target region.

All the cosmids isolated from hybrid 17B were mapped either to the interval between the SD-1 and 9T12 translocation breakpoints or to the interval between the 9T12 and 9T01 breakpoints (see table 3.4 and figure 3.12). These localisations were consistent with the retention marker analysis (figure 3.1) because the 9T12 breakpoint was mapped within the region of DNA from D9S150 to D9S67 (figure 3.12) that is retained by
hybrid 17B. Cosmid fingerprinting positioned seven of the eight clones within contigs and 17BE and cKW1B were both localised to contig number 41 which contained D9S10 and D9S66 (see figure 3.25A). These two clones were isolated using different techniques but were positioned next to each other within the TSC1 candidate region. This demonstrates the value of using both approaches, as they complement one another and thereby reduce the effects of any particular bias, such as organisation of Alu sequences in the genome.

The efficiency of isolating cosmids using Alu-PCR products from hybrid DNA was analysed by FISH. As expected, pools of cosmids from each hybrid mapped to chromosome 9q34 and a random pool of cosmids mapped along the entire length of the chromosome (figure 3.5). However the cosmids isolated using hybrid 20A gave additional hybridisation signals on chromosome 9p12 and 9q21-22, around the centromere. As two of the four clones isolated from the hybrid 20A cosmid library mapped to the centromere of chromosome 9 this result was not surprising.

Pools of cosmids that were identified by hybrids 17B, 19B and 20A were analysed by FISH to metaphase chromosomes from the 9T12 cell line (figure 3.13) whose translocation breakpoint had been mapped between markers D9S114 and D9S298 (see figure 3.12). The main signal for cosmids identified using hybrids 17B and 20A was proximal to the breakpoint and according to retention studies this would represent the region from D9S150 to D9S114 (see figure 4.2). However the precise localisation of the breakpoint between markers D9S114 and D9S298 and the extent of human DNA from hybrid 20A distal to D9S114 is not known. Therefore the cosmids could also be derived from the region between D9S114 and the breakpoint where DNA from the two hybrids
Figure 4.2
Deletion map of 17B, 19B and 20A radiation hybrids

Note: Dotted lines extending from hybrids represent possible extent of hybrid DNA.
Figure not drawn to scale and adapted from Nahmias et al. (1995).
may also overlap. A smaller hybridisation signal was also present distal to the 9T12 breakpoint and this could be due to an overlap in human DNA between hybrid 17B and 20A between D9S67 and D9S158 or between the breakpoint and D9S298.

The pools of cosmids identified using hybrids 17B and 19B were evenly distributed proximal and distal to the breakpoint. A group of loci from D9S298 to D9S67 were retained in both these hybrids which could account for the signal distal to the breakpoint. The proximal signal suggests that DNA from hybrid 19B overlaps DNA from hybrid 17B between D9S114 and the breakpoint.

Cosmids identified by both hybrids 19B and 20A were also positioned on either side of the breakpoint. The proximal signal may be from the interval between D9S149 and ABO or between D9S114 and the breakpoint. The distal signal may be from the interval between the breakpoint and D9S298, or from D9S67-D9S158.

Finally, the cosmids identified by all three hybrids were only localised proximal to the breakpoint. As the TSC1 candidate region has been positioned proximal to D9S114 (Povey et al., 1994b) it suggested that this category of cosmids were of major interest. The result also provided further information on the localisation of human DNA that was in common between the three hybrids. According to marker analysis the only region where the DNA content of hybrid 17B and 19B could overlap to give a proximal hybridisation signal is between D9S114 and the breakpoint. Therefore the proximal fragment of hybrid 20A could not extend beyond the breakpoint or it would also overlap with hybrids 17B and 19A. Consequently, the signal distal to the breakpoint for cosmids identified by both 20A and 17B must be between D9S67 and D9S158 and distal to the
19B fragment. However this would not allow a region distal to the breakpoint where human DNA from hybrids 20A and 19B overlapped.

It was concluded that the general consensus of FISH results supported the marker retention analysis and therefore they could be used infer cosmid position. However as these localisations relied on the assumption that there were no additional fragments in each hybrid other than those detected by marker analysis, they were not entirely consistent. Therefore although in principal this method of positioning seemed effective, in practice one clone from each contig was mapped by FISH onto the translocation breakpoint cell lines as it was much more accurate.

4.1.2.2. Mapping clones encoding genes

By mapping the eighteen genes with respect to the three translocation breakpoints their physical localisations were refined. With one exception (C8G) the data was consistent with earlier studies. Complement component 8 gamma peptide (C8G) had been previously reported to map to 9q22.3-q32 by PCR on flow-sorted translocation chromosomes (Yuille et al., 1992). A genomic clone encoding C8G was positioned distal to the 9T01 breakpoint. This localisation on human chromosome 9q34.3 was supported by comparative mouse mapping data as interspecific backcross linkage analysis mapped the mouse homologue of C8G to the same position on mouse chromosome 2 as GRIN 1 (Pilz et al., 1995a). However the C8G probe derived by RT-PCR (Pilz et al., 1995a) and mapped in mouse, was the same as that used to isolate a human genomic clone (Nahmias et al., 1995) and localise the gene in this study.
Therefore it conceivable, although unlikely, that the wrong gene was mapped in both experiments.

The Surfeit gene cluster (SURF) had been previously reported to map distal to the Philadelphia translocation on chromosome 9q34 (Yon et al., 1993). This localisation was confirmed but also refined as it mapped proximal to the 9T12 breakpoint. The carboxyl ester lipase gene (CEL) had been previously mapped to chromosome 9q34.3 (Taylor et al., 1991). In this study, CEL was mapped to the interval between the SD-1 and 9T12 breakpoints. The localisation of the collagen type V alpha 1 gene (COL5A1) was refined from 9q34.2-9q34.3 (Greenspan et al., 1992) to the interval between the 9T12 and 9T01 breakpoint. Similarly the localisation of the progestagen-associated endometrial protein (PAEP) was refined from chromosome 9q34 (Van Cong et al., 1991) to the region distal to the 9T01 breakpoint at 9q34.3.

4.1.2.3. Mapping the TSC1 candidate region

Both D9S149 and D9S114 mapped to the interval between the translocation breakpoint for cell line SD-1 and 9T12 (see figure 3.12). Therefore any gene mapping outside this region such as RXRA, COL5A1 and NOTCH1 can be confidently excluded as a candidate for TSC1. Conversely genes and contigs mapping within this interval can be easily identified for further analysis. The physical distance between these two breakpoints is approximately 3.7Mb (Povey et al., 1994b) and the TSC1 candidate region is between 1.5 and 3.0Mb in size (Nahmias et al., 1995; Povey et al., 1994b). Although determining contigs mapping within the TSC1 candidate region would require further work because clones within each interval have not been ordered, mapping contigs on the translocation breakpoint cell lines is an efficient method of rapidly
reducing the number of contigs to be analysed. In this study, 20 of the 64 contigs mapped were localised to the interval between the SD-1 and 9T12 breakpoints. These contigs facilitate a detailed analysis of the region of the genome where the flanking markers for TSC1 are known to map and six of them have now been searched for transcribed sequences.

4.1.3. Isolation of transcribed sequences from 9q34

Two separate exon amplification (Buckler et al., 1991) experiments were performed in this study. The preliminary experiment used vector pSPL1 and 8 cosmids were screened (see summary in figure 4.3) which were considered to be in the TSC1 candidate region by their isolation from hybrid 17B or position on the 9q34 physical map (see section 3.3.1). The second was more extensive as 6 cosmid contigs were analysed and the modified vector pSPL3 (Church et al., 1994) was used (see summary in figure 4.4). Contigs were selected by their localisation within the consensus flanking markers for TSC1 or within the interval between the SD-1 and 9T12 translocation breakpoints (section 3.3.2). A number of different regions were analysed because there was no strong evidence for a more refined candidate interval at the time the contigs were chosen. Both experiments were successful in identifying transcribed sequences from genes and led to the isolation and characterisation of cDNA clones. In addition, both yielded exon sequences that had also been isolated by other groups (Kwiatkowski, personal communication; Church et al., 1994). This suggested the preferential use of particular splice sites in the processing of transcripts generated from the exon amplification vector and illustrated the reproducibility of the technique.
Figure 4.3

Summary of transcribed sequences isolated using pSPL1

Centromere

<table>
<thead>
<tr>
<th>Loci</th>
<th>Cosmids Screened</th>
<th>No. of putative exons isolated</th>
<th>Novel genes identified</th>
</tr>
</thead>
</table>
| SD-1 breakpoint
| D9S113 | DBH-14 | 1 |
| D9S64 | MCT136-7 | 4 |
| D9S149 | MCT136-11 | 3 |
| CEL | cKW1B | 3 |
| ABO | 17BE | 1 |
| SURF | | |

9T12 breakpoint

| D9S114 | 17B2 | 4 |
| D9S298 | cKW1 | 1 |
| | cKW10 | 3 |

Telomere

Note that cell lines SD-1 and 9T12 were only available after the analysis commenced

Consensus candidate region for TSC1 (Povey et al., 1994b; Pericak-Vance et al., 1995)

Diagram not to scale
**Figure 4.4**

**Summary of transcribed sequences isolated using pSPL3**

<table>
<thead>
<tr>
<th>Loci</th>
<th>Contig Name</th>
<th>Cosmids Screened</th>
<th>Putative exons</th>
<th>cDNAs isolated</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>SD-1 breakpoint</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D9S113</td>
<td>CEL</td>
<td>L-4DD1 (10)</td>
<td>Bam A-E</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>203 H12</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>256 C3</td>
<td>Pst A-D</td>
<td></td>
</tr>
<tr>
<td>D9S149</td>
<td>CEL</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D9S64</td>
<td>CEL</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>9T12 breakpoint</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D9S114</td>
<td>CEL</td>
<td>17B2 (80)</td>
<td>Bam A-C</td>
<td>16 clones (1-12, 16-19)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>157 B7</td>
<td>Pst A-C</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>222 A10</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>CEL</td>
<td>37*</td>
<td>Bam A-E</td>
<td>2 clones (15, 20)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>288 E8</td>
<td>Pst A-D</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>180 F1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>CEL</td>
<td>166*</td>
<td>Bam A-C</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>84 B3</td>
<td>Pst A-F</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>194 E12</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Telomere</strong></td>
<td>CEL</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Contigs are localised between SD-1 and 9T12 translocation breakpoints but no further positional information is currently known.

- Consensus candidate region for TSC1 (Povey et al., 1994b; Pericak-Vance et al., 1995) Diagram is not to scale. # cDNA from SURF contig see section 3.3.2.1.a.
4.1.3.1. Exon amplification using pSPL1

In summary, the technique was successful in yielding 18 different products from 8 cosmids. Six were unique by sequence search, one had strong similarity to VAV and eleven were either chimeric or artefacts which were eliminated from the analysis.

Although the cosmids in this early study had been mapped by FISH to normal metaphase chromosomes, they were only mapped with respect to the translocation breakpoints after the experiment had commenced. Cosmids cKW1 and cKW10 were localised to interval 3 between the 9T12 and 9T01 translocation breakpoint, whereas the other clones were localised to interval 2. Therefore the two novel products from cKW1 and cKW10 were excluded from further analysis because they mapped outside the TSC1 candidate region.

Recent fingerprint analysis has positioned cosmids cKW1B, 17BE, MCT136-7, MCT136-11 and DBH within contig 41 (Nahmias et al., 1995). This contig contains markers D9S10 and D9S66 and it is shown in figure 3.25A. Physical and genetic mapping has ordered these markers and DBH on the chromosome (Povey et al., 1992) and thereby localised the contig within the flanking markers for TSC1 (see figure 4.4). All the products isolated from cKW1B were chimeric, one of these was single copy in primates and although it was not pursued it may still be of use as a probe for isolating cDNA clones. The product isolated from cosmid 17BE had strong homology to VAV and will be discussed in the section titled characterisation of transcribed sequences and their
candidates for TSC1 (section 4.1.4.2.b.). An additional three unique products derived from the MCT 136 cosmid were also isolated from this contig. They were novel by sequence search and could have potential in identifying candidate genes. However as the contig was rescreened using the modified vector pSPL3 these products were not analysed further.

Cosmid 17B2 was positioned within contig 80 (see figure 3.26) and proved to contain the D9S114 polymorphism (Nahmias et al., 1995) which is the distal flanking marker for TSC1 (Povey et al., 1994b). One of the products isolated was novel by sequence search and could have potential for isolating a candidate gene. However it was not studied further because analysis of the contig and even this particular cosmid continued using pSPL3.

4.1.3.2. Exon amplification using pSPL3

The analysis using pSPL3 was more thorough than that using pSPL1. Six cosmid contigs were studied which at the start of the experiment consisted of 2-3 clones each. The clones were analysed individually and as a pool and the patterns of products generated from each were compared. The products were immediately cloned and this facilitated their analysis by sequencing and eliminated any mixed products which might confuse the results.

As with pSPL1, there was considerable differences in the pattern of amplification products depending on the restriction enzyme used for cloning. Variations are seen because an exon will not normally be amplified if it contains a recognition sequence for the enzyme used for cloning. In addition, restriction fragment lengths can affect the
cloning efficiency and therefore an exon present on a large genomic fragment could also be missed. A partial digest with a frequent cutting enzyme such as Sau3AI may reduce this problem. However combining different restriction enzymes also provides a greater opportunity for an exhaustive search of exon sequences. As the modified vector pSPL3 contains a multiple cloning site, it therefore has greater flexibility than pSPL1 and other previously reported exon trapping vectors which only contain one (Auch and Reth, 1991; Hamaguchi et al., 1992).

The pattern of products generated for pooled cosmids using pSPL3 was not consistently more complex than that of individual cosmids (figures 3.33-3.38). This suggested a decrease in efficiency of the technique with increasing complexity of genomic DNA. However Church et al. (1993) used pools of 5-10 cosmids and although the average number of products isolated per cosmid was almost halved, the level of efficiency with respect to the total amount of genomic DNA present was not significantly compromised.

4.1.3.2.a. Sequence analysis

In summary, 48 of the cloned products were characterised by sequence analysis, 14 were found to be artefacts and the remaining 34 were inferred to contain transcribed sequences. Seven products (14.5%) had strong similarity to ribosomal protein L7a (Yon et al., 1993), VAV2 (Smith et al., 1994b; Woodward et al., 1994; Henske et al., 1995) or set (von Lindern et al., 1992) which were known genes on human chromosome 9q. In addition, two products (4%) had very strong homology (94%) to exon sequences isolated from chromosome 9 by Church et al. (1994) and two products (4%) had identity to sequences isolated by Dr. D. Kwiatkowski from the same region. These results were encouraging because they demonstrated that the technique had been successful. As
mentioned earlier the isolation of exon sequences that were identical to those from other
groups demonstrated that preferential splice sites were probably used in the procedure.

The level of chimerism was significantly lower using pSPL3 as only six (12.5\%) of the
products cloned contained HIV sequence at the 3' end, whereas 45\% of the products
isolated using pSPL1 were chimeric. Therefore although the \textit{BstXI} restriction digest
step did not eliminate vector amplification (see figure 3.32), it was nevertheless
successful in reducing chimerism. If the \textit{BstXI} digestion had been complete, chimerism
would have been eliminated and SURF Pst B would not have been amplified. This
product contained exon sequence with strong similarity to ribosomal protein L7a but it
was interrupted by a \textit{PstI} site. It illustrates how chimerism can recover exons not
normally amplified and provided that the products are large enough they may still have
use as hybridisation probes. However four of the chimeric products did not show any
significant sequence similarity at the 5' end. This may be because they are novel exon
sequences or because they have been produced by activation of a cryptic genomic splice
acceptor site. It is possible that cryptic splice sites within genes are activated when they
are taken out of their genomic context on a DNA fragment which does not contain any
other splice signals. It may be that the cryptic sites remain silent \textit{in vivo} because they are
followed by a correct splice site or that they are present on a non-coding strand.
Ten of the products (21%) cloned were artefacts, compared to two products (10%) isolated using pSPL1. These artefacts fell into two categories and included Lorist vector and HIV-1 tat intron sequence. In contrast to using pSPL1, Alu sequences were not amplified with pSPL3. Although these artefacts are a disadvantage they are easily detected by sequence searches or by hybridisation to cosmid Southern blots. In addition, as the technique requires two functional splice sites in the selection procedure, it should provide greater stringency than other exon trapping systems which only require one (Duyk et al., 1990; Krizman and Berget, 1993).

4.1.4. Characterisation of transcribed sequences and their candidates for TSC1

The limiting factor to exon amplification in this study was not the isolation of putative exons but the confirmation that they do actually represent authentic coding region. Sequence analysis was a rapid method of screening the products and selecting those for further analysis. It was also very efficient as all those investigated further were confirmed to be genuine transcribed sequences.

4.1.4.1. Contig SURF

4.1.4.1.a. Surfeit genes

Sequence searches showed that four of the products from the SURF contig had strong similarity to the nucleotide sequence of ribosomal protein L7a. This was initially misleading as the similarity was within the 5' flanking sequence of the SURF-3 gene and on the minus DNA strand. It was deduced that the products were derived from the Surfeit gene upstream of SURF-3 which was SURF-5 and the finding was consistent
with the alternating direction of transcription for adjacent Surfeit genes in the mouse
(Yon et al., 1993) (see figure 3.41B). The amplified exons were positioned within
SURF-5 according to their localisation within the 5' flanking region of SURF-3 (figure
3.41). Although it seems unlikely that they are cloning artefacts, the possibility cannot
be eliminated.

Furthermore, SURF Bam A was mapped genetically in the mouse genome by Dr. A.
Pilz to the same position as mouse homologues of DBH and SURF-1 and therefore it
was consistent with SURF Bam A being a transcribed sequence from the Surfeit gene
cluster. The mouse homologue of SURF Bam A (Surf Bam A) mapped proximal to
homologues of RXRA, COL5A1, and D9S114 Bam A and distal to the mouse
homologue of VAV2 (see figure 4.5).

The Surfeit genes are known to be very tightly clustered, Yon et al. (1993) found
SURF-3, SURF-1 and regions of SURF-5 and SURF-2 within a 16kb genomic DNA
fragment. One of the SURF-5 sequences isolated, hybridised to both of the cosmids
analysed (see figure 3.46) and therefore it was assumed that SURF-6 would be within
211 A7 and SURF-3, SURF-1 and SURF-2 and SURF-4 would be within 255 A6. As
the contig has been orientated with respect to the chromosome (see figure 3.24), the
Surfeit cluster of genes can also now be orientated (see figure 3.46) SURF-6 being the
most proximal and SURF-4 being the most distal.

However transcribed sequences from SURF-3 were not identified and although it is
possible that one of the products with identity to human L21 (SURF Bam B) may be
derived from SURF-6, it will not be confirmed until the nucleotide or protein sequence
is entered into the databases. Losses are inevitable but the cloning efficiency of the
Figure 4.5

Genetic mapping of exon sequences on mouse chromosome 2 compared to human chromosome 9q34

Mouse

<table>
<thead>
<tr>
<th>centromere</th>
<th>Human</th>
</tr>
</thead>
<tbody>
<tr>
<td>Notch1</td>
<td>NOTCH1</td>
</tr>
<tr>
<td>Vav2 (exon e)</td>
<td>RXRA,COL5A1</td>
</tr>
<tr>
<td>Dbh, Surf1, Surf Bam A</td>
<td>D9S114 Bam A</td>
</tr>
<tr>
<td>D9S114 Bam A</td>
<td>VAV2 (exon e)</td>
</tr>
<tr>
<td>Rxra, Col5a1</td>
<td>DBH</td>
</tr>
<tr>
<td>Cel</td>
<td>SURF1, Surf Bam A</td>
</tr>
<tr>
<td></td>
<td>CEL</td>
</tr>
<tr>
<td>telomere</td>
<td>centromere</td>
</tr>
</tbody>
</table>

The order of the human chromosome has been inverted to aid comparison.

Taken from Pilz, 1995.
experiment was analysed before transfection and all the possible genomic fragments had been cloned. A restriction map of the human Surfeit gene cluster does indicate two BamHI and four PstI sites (Yon et al., 1993) which may account for loss of some exons. One of these BamHI sites was shown within SURF-5 and therefore corresponds with the result from this study (see figure 3.47). An alternative explanation is that SURF-1, SURF-6 and SURF-4 are intronless genes or contain only one intron and therefore the selection strategy means they are likely to be missed. However this does not apply to SURF-3 as it is known to contain at least 8 exons and 7 introns (Giallongo et al., 1989). Bias in pAMP1 cloning of amplification products seems the most plausible explanation and if a greater number of recombinant clones had been picked and characterised these genes may have been represented.

The SURF contig also contained the ABO blood group locus and was positioned distal to D9S149 and proximal to D9S150 (Nahmias et al., 1995). Therefore it is within the consensus flanking markers for TSC1. However the region encompassed by these genes is excluded from the candidate regions defined by critical recombination events described by Pericak-Vance et al. (1993), Pitiot et al. (1994) and Janssen et al. (1994) (see figure 4.1). As these localisations are controversial and little information is available regarding the Surfeit housekeeping genes, their possibility as candidates for TSC1 cannot be excluded. One of the cDNA clones isolated in this experiment (ICRFp50700115) was homologous to two of the SURF-5 exon sequences and had been previously identified by another research group as SURF-5. This clone would be a useful resource in assessing the potential of SURF-5 as a candidate for TSC1. As yet, sequence analysis has been unable to detect any significant similarities between this clone and any other sequence in the nucleotide and protein databases.
4.1.4.2. Contig D9S10

4.1.4.2.a. VAV2

One product isolated using pSPL1 (exon e) and two products isolated using vector pSPL3 (D9S10 Bam C and D9S10 Bam D) showed strong similarity to the mouse and human VAV proto-oncogene. Exon e was conserved in primates (human, orang-utan, monkey), zebra, dog and rodents (vole, rat, hamster and mouse) and according to cosmid contig information mapped slightly distal of D9S10 and D9S66 (see figure 3.25B). The localisation contributed to the mapping of VAV2 on human chromosome 9q34 (Smith et al., 1994b; Woodward et al., 1994) and positioned the gene within the consensus flanking markers for TSC1 and the more distal candidate region reported by Pitiot et al. (1994) and Pericak-Vance et al. (1993).

Comparative mapping of exon e in the mouse genome localised VAV2 to a region of known synteny between mouse chromosome 2 and human chromosome 9q34. Initially, the map position could not be genetically separated from mouse homologues of human genes NOTCH1, SURF and DBH (Pilz et al., 1995a). However mapping using a larger backcross (European backcross collaborative group, 1994) allowed the mouse homologue of VAV2 to be positioned on the mouse chromosome proximal to homologues of DBH, SURF-1 and SURF Bam A and distal to the homologue of NOTCH1 (Pilz, 1995). As mouse chromosome 2 is inverted with respect to human chromosome 9 this localisation is completely consistent with the order in the human genome (see figure 3.12 and figure 4.5).
VAV proto-oncogene was localised to human chromosome 19p12-13.2 (Martinerie et al., 1990) and found to be specifically expressed in cells of hematopoietic origin including erythroid, lymphoid and myeloid lineages (Katzav et al., 1989). Molecular analysis of the predicted amino-acid sequence indicated a novel structure with a number of different domains. These domains are illustrated in figure 4.6 and include helix-loop-helix, leucine zipper, a highly acidic domain, guanine nucleotide exchange factor, two putative nuclear localisation signals, zinc finger, SH2 and two SH3 domains (Katzav et al., 1991). It is loss of the amino-terminal helix-loop-helix domain which results in activation of VAV oncogenic potential (Katzav et al., 1991).

Cytoplasmic signalling proteins contain Src homology 2 (SH2) domains that mediate their interaction with receptor tyrosine protein kinases (Koch et al., 1991). There is also evidence that Src homology 3 (SH3) domains mediate protein-protein interaction in signal transduction activated by protein tyrosine kinases (Cicchetti et al., 1992). The identification of these domains within VAV implied that it was a substrate for tyrosine protein kinases and indicated that it may have a role in transduction of tyrosine phosphorylation signalling (Bustelo et al., 1992; Margolis et al., 1992).

VAV also shared homology with guanine nucleotide releasing factor (GRF) domains in the yeast (CDC25; Jones et al., 1991), rodent (Ras-GRF; Shou et al., 1992) and human (Dbl; Hart et al., 1991) genome (Gulbins et al., 1993). These proteins have been found to activate Ras-related small guanosine triphosphates (GTP)-binding proteins (G proteins). Ras proteins cycle between an active GTP bound state and an inactive GDP bound state. GRFs mediate exchange of bound GDP to GTP and therefore these proteins
result in Ras stimulation. The GRF domain within VAV suggested that it may have a role in signal transduction through Ras activation.

**Figure 4.6**

**Structural domains within VAV proto-oncogene**

VAV2 was characterised by Henske *et al.* (1995) and found to contain the same major domains as VAV. In contrast to VAV, a wide pattern of tissue expression was detected for VAV2 (Henske *et al.*, 1995). With these results and the knowledge that oncogenes can act as tumour suppressor genes (Mulligan *et al.*, 1993), the gene seemed an attractive candidate for TSC1. Henske *et al.* (1995) performed Southern blot studies in 89 TSC patients (7 from families linked to a locus on chromosome 9) but only two different RFLPs and one rare variant were found. There were no changes detected which were inherited with the tuberous sclerosis phenotype. Therefore an intensive sequencing
analysis of the complete VAV2 coding region was performed on RT-PCR products from 6 patients. Prior detection of a common polymorphism at position 1785bp was used to select analysis on heterozygotes and 4 of these were from TSC families linked to a locus on chromosome 9. Three further sequence variants were reported but they were silent changes with respect to the amino-acid sequence. No deletions or significant mutations were found and therefore the authors concluded that VAV2 was excluded as a candidate (Henske et al., 1995).

In a parallel study performed by myself, the 5’ end of VAV2 was analysed using the technique of single strand conformational polymorphism (SSCP) (Orita et al., 1989) to search for possible disease related variants in 16 TSC patients. This region was chosen because loss of the 5’ end of VAV activates its oncogenic potential (Katzav et al., 1991) and it suggested that a similar finding might be detected in TSC patients if this is a candidate for TSC1. The complete 2.7kb cDNA sequence of VAV2 was available from Henske et al. (1995) prior to publication and four pairs of oligonucleotide primers were designed to amplify from position 1-1002bp of the transcript by RT-PCR, in four overlapping fragments between 200-350bp. RNA was extracted from lymphoid lines derived from 8 TSC patients that were members of 6 different TSC families. Four families were thought to be segregating for TSC1, the remaining two did not appear to segregate for either TSC1 or TSC2 (Povey et al., 1994a). As the first pair of primers were localised within exon sequence and could be amplified by PCR from genomic DNA, a further 8 TSC patients from different families considered to be segregating for TSC1 were also studied just for this fragment. Unfortunately, analysis of the RT-PCR and PCR fragments by SSCP did not detect any electrophoretic variants in the TSC patients when compared to normal controls (own data not shown). With these negative
findings and the knowledge that Henske et al. (1995) had carried out a more extensive analysis but not found any indication of a mutation, the analysis was discontinued. However these investigations by SSCP alone, would not be sufficient to eliminate VAV2 as a candidate gene.

4.1.4.3. Contig D9S114

4.1.4.3.a. D9S114 Bam A, D9S114 Bam C & D9S114 Pst B

Sequence analysis of D9S114 Bam A and D9S114 Bam C showed that both products had strong similarity to the beta-transducin or WD-40 repeat found within the beta subunit of heterotrimeric guanine nucleotide binding proteins (G proteins). There are two different types of G proteins, the "small G proteins" which exist as a single polypeptide such as Ras, or the heterotrimeric G proteins which are composed of alpha, beta and gamma subunits. The small G proteins are involved with regulation of cell growth, protein secretion and intracellular vesicle interaction (Hall, 1990). Both types of G proteins are associated with signal transduction from cell surface receptors, they are thought to act as switches which regulate information transfer depending on the nucleotide bound (review Simon et al., 1991). Signal transduction is initiated by ligand binding (hormones, growth factors etc.) to a receptor protein. As a result the receptor is stabilised in an active form and transmits a signal across the cell membrane, thereby activating the heterotrimeric G protein. The GDP bound to the alpha subunit is exchanged for GTP and the alpha-GTP subunit dissociates from the beta-gamma subunit. Both of these proteins may then interact with effector molecules to amplify the signal. Termination of the signal occurs when GTP bound by the alpha subunit is hydrolysed to GDP and then the alpha subunit reassociates.
The beta subunit is characterised by eight segments of amino acid sequence, that each contain a 40 amino acid motif with certain amino acids, including a tryptophan-aspartic acid pair. The letter codes for these amino acids are W and D respectively and therefore this motif is commonly known as a WD-40 repeat. However heterotrimeric G-proteins are not the only proteins which contain this WD-40 repeat as the sequence database search with D9S114 Bam A showed. The other proteins have diverse roles and examples include the yeast CDC4 protein which is involved in cell cycle regulation (Peterson et al., 1984), yeast PRP4 which is involved in RNA splicing (Dairymple et al., 1989) and Drosophila Enhancer of split which is a neurogenesis protein (Hartley et al., 1988). The specific function of the conserved amino acids are not yet apparent, however they are thought to have a role in protein-protein interactions (Dairymple et al., 1989).

D9S114 Pst B showed strong similarity to a human GTPase-activating protein (GAP-43). Hydrolysis of GTP to GDP is accelerated by GTP activating proteins (GAPs) and therefore these proteins inactive GTP binding proteins such as Ras. It suggested that this exon in addition to D9S114 Bam A and D9S114 Bam C was involved in signal transduction pathways. Isolation of cDNA clones homologous to these exon sequences confirmed that they were in fact derived from the same gene.

Cross-species hybridisation of a transcribed sequence from this gene (D9S114 Bam A) found it to be conserved in primates (human, orang-utan, monkey), dog, guanaco, rodent (mouse, rat), fish (trout) and amphibian (Xenopus) genomic DNA (figure 3.50).
D9S114 Bam A was genetically mapped in the mouse genome by Dr. A. Pilz to proximal mouse chromosome 2, the region homologous to human chromosome 9q34. It was localised proximal to mouse homologues of RXRA and COL5A1 and distal to mouse homologues of DBH, SURF-1 and SURF Bam A (see figure 4.5). As mouse chromosome 2 is inverted with respect to human chromosome 9q34 this order would not agree with the human order of genes. The order of genes from centromere to telomere in this region, as deduced by human mapping is CEL, SURF, DBH, VAV2, D9S114 Bam A, (RXRA, COL5A1) NOTCH1 (Povey et al., 1994b). The order inferred by comparative mapping in the mouse genome was CEL, (RXRA, COL5A1), D9S114 Bam A, (SURF1, DBH), VAV2, NOTCH1 (Pilz, 1995) and therefore the only inconsistency was with COL5A1, RXRA and D9S114 Bam A. These results suggested that evolutionary changes have taken place between the human and mouse genomes at this particular point, however the possibility of errors in mapping cannot be eliminated. If the differences are real, it indicates that an inversion involving these three genes has occurred between these two genomes (Pilz, 1995).

4.1.4.3.a.i. Characterisation of cDNA clones

Thirteen cDNA clones homologous to both D9S114 Bam A and D9S114 Pst B were isolated and two of the largest ones were selected for further analysis (clone 5 and 10). Initially it was assumed that the 3' end of clone 10 was complete (figure 3.58) and that it extended into contig D9S10 (figure 3.62). The results suggested that the gene was transcribed in a telomere to centromere direction proximal to D9S114 and would therefore be within the TSC1 candidate region. As the physical distance between D9S10 and D9S114 is approximately 325kb (Murrell et al., 1995) it also suggested that the
gene contained large intronic sequences. However sequencing at either end of the cDNA did not detect a polyA tail and therefore it was possible that this clone was chimeric.

Analysis turned to clone 5. It was confirmed to map to human chromosome 9q34 by FISH but additional sites of hybridisation were also detected on chromosome 1q (see figure 3.63). As database searches with the cDNA sequence confirmed it to be strongly similar to proteins containing a WD-40 repeat, it was thought this other hybridisation signal might represent another member of the large family of genes with WD-40 or β transducin like repeats. (However the possibility of it being a chimeric clone has not yet been eliminated.) The similarities also suggested that the protein product of this gene might have a role in signal transduction and complete loss of function as proposed in the tumour suppressor model, would affect the signalling pathways and may lead to uncontrollable cell growth and tumour development. Therefore this gene seemed a very good candidate for TSC1 and the analysis continued.

Hybridisation of the cDNA to cosmids from contig D9S114 found only 157 B7 to be homologous (figure 3.60). Comparison of the clones within the contig to a complete contig map published by Murrell et al. (1995) have very recently orientated this contig and as a result 157 B7 has been positioned distal to marker D9S114. Therefore if the critical recombinant verbally reported by Dr. H. Northrup (Povey et al., 199b) is correct, it means that cDNA 5 is outside the TSC1 candidate region. The DNA duplication reported by Smith et al. (1994) which extends between D9S122 and D9S114 would also suggest that this is the case. However there is a possibility that the full length cDNA might extend proximal to D9S114 and that this critical recombinant is intragenic and that the junctions for the duplication are within the candidate gene.
Although it now seems less likely that this is the TSC1 gene, cDNA 5 is currently being analysed on genomic DNA of TSC patients in the hope that a large rearrangement will be detected. As with any positional cloning effort the final proof that the gene identified is actually the disease gene, depends on discovering a mutation that distinguishes normal from affected individuals. In addition, work is currently underway to isolate and sequence the full length transcript as whatever the outcome it is still a very interesting novel gene.

4.1.4.3.b. D9S114 Bam B

The similarity between D9S114 Bam B and set was within the 3' coding region and the 3' untranslated region of the gene. The gene was discovered in leukemic cells from a patient with acute undifferentiated leukemia (AUL) and fused to CAN (von Lindern et al., 1992). Subsequent mapping of the corresponding cDNA localised the set gene centromeric to ABL on chromosome 9q34. The 3' end was known to have a high proportion of acidic residues which suggested it may be a nuclear protein but at present nothing is known about its function (von Lindern et al., 1992).

4.1.4.3.b.i. Characterisation of a cDNA clone

Partial sequencing of the largest cDNA clone homologous to D9S114 Bam B (number 7) did not show any similarity to set. Therefore it was thought that this gene may contain certain domains with similarity to set but there was no reason to suspect that the initial localisation of set was incorrect (von Lindern et al., 1992).
The cDNA clone mapped to two adjacent cosmids within the D9S114 contig (figure 3.61) and the most distal one (17B2) contained the distal flanking marker for TSC1. Therefore the recent orientation of this contig would position the gene within the consensus candidate gene for TSC1, but the D9S114 critical recombination event would have to be intragenic. As a result it must be considered a candidate for the TSC1 gene and is currently being investigated further.

4.1.4.4. Contig 37

4.1.4.4.a. Ctg37 Bam A

This exon showed strong similarity to rat protein tyrosine kinase and when used as a probe to screen a cDNA library two clones were isolated.

4.1.4.4.a.i. Characterisation of a cDNA clone

The largest cDNA was confirmed to map to human chromosome 9q34 by FISH but partial sequencing did not show any significant similarity to any other sequence in the database. As the localisation of contig 37 with respect to the flanking markers for TSC1 is not currently known, present work involving interphase and extended DNA FISH mapping is being pursued to order the contig on 9q34 (Ekong, personal communication). Once the result of this study is known, the importance of this gene with respect to TSC1 will be much clearer.
4.1.5. Exon amplification and alternative approaches for isolating transcribed sequences

Exon amplification was found to be an efficient method of isolating transcribed sequences and the advantages and disadvantages of the technique are shown in table 4.1. In this study, it was considered successful if a minimum of one exon sequence was amplified from each gene present. However the requirement for 5' and 3' functional splice sites means that intronless genes and genes with only one intron are likely to be missed. In addition, the first and last exons of genes are also not usually amplified unless they are incorporated in products arising from cryptic splicing. It follows that if an amplification product is not isolated from a genomic DNA fragment it cannot be assumed that the region is devoid of exons. Therefore a thorough search for exon sequences in the genome may require a combined approach which utilises a number of different techniques for isolating transcribed sequences (see introduction section 1.3.5.).

The method of exon trapping described by Krizman and Berget (1993) is specifically designed to select for the 3'-terminal exons and would therefore be a useful technique for complementing exon amplification. Alternatively the strategy of cDNA selection (Lovett et al., 1991; Parimoo et al., 1991) could also be used and the advantages and disadvantages of this technique are shown in table 4.2. One of the major disadvantages is that it is dependent on the presence of the relevant gene sequences in the cDNA library tested. Conventional amplified cDNA libraries were used initially (Lovett et al., 1991; Parimoo et al., 1991) but the sequences they contained were not always representative and therefore rare transcripts were unlikely to be present in those derived from complex tissues. More recently complex sets of uncloned cDNA sources were
### Table 4.1

**Summary of some advantages and disadvantages of exon amplification**

<table>
<thead>
<tr>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>No information about source or developmental stage of expression required</td>
<td>Intronless gene or genes with a single exon are likely to be missed</td>
</tr>
<tr>
<td>Equal representation of genes with low and high levels of RNA expression</td>
<td>Terminal exons are likely to be missed</td>
</tr>
<tr>
<td>Requirement for 2 functional splice sites improves stringency of the technique</td>
<td>PCR bias for sequence and length i.e. larger exons less likely to be amplified</td>
</tr>
<tr>
<td>Efficient, ~10% artefacts with pSPL3 which fall into a small number of categories that can be easily detected</td>
<td>Artefacts (HIV-1, Alu or vector sequences)</td>
</tr>
<tr>
<td>Multiple cloning site within pSPL3, allows a combined restriction enzyme approach to give a greater opportunity for isolating exons</td>
<td>Exons missed if they are cleaved by the restriction enzyme, if they are on a large genomic fragment that has a reduced likelihood of being cloned or if they are cloned in the wrong orientation</td>
</tr>
<tr>
<td>Cryptic splicing can recover exons not normally amplified (i.e. if restriction site is present within the exon sequence)</td>
<td>Cryptic splicing events amplify processed intron sequences (levels reduced with modified vector pSPL3)</td>
</tr>
<tr>
<td>Short average length of exons means they can be sequenced in a single gel run and can then be either pursued or ignored</td>
<td>Short average length of exons may be a problem with hybridisation</td>
</tr>
<tr>
<td>Can be used to characterise differentially spliced transcripts</td>
<td>Preferential use of certain splice sites, some exons may be skipped</td>
</tr>
<tr>
<td>Analysis of small defined chromosomal regions or large scale screening is possible</td>
<td>Reduced efficiency with larger scale screening</td>
</tr>
<tr>
<td>Can be used to characterise the intron/exon structure of genes</td>
<td></td>
</tr>
</tbody>
</table>
Table 4.2

Summary of some advantages and disadvantages of cDNA selection

<table>
<thead>
<tr>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Large target sizes (1 Mb) can be easily used in the analysis</td>
<td>Must have an appropriate source of cDNA that contains the target cDNA</td>
</tr>
<tr>
<td>Very sensitive, a cDNA can be enriched 100,000 fold</td>
<td>PCR-based technique, therefore sequence and length bias</td>
</tr>
<tr>
<td>Can use complex, uncloned primary cDNAs to compensate for low level expression</td>
<td>Gene families and pseudogenes can complicate analysis</td>
</tr>
<tr>
<td>Can sample many tissues in a single experiment</td>
<td>Low level repeats, mitochondrial cDNAs and ribosomal sequences can complicate analysis</td>
</tr>
<tr>
<td>cDNA average length between 0.5-1.0kb</td>
<td>cDNA sensitive to exon length, may not be a good hybridisation probe</td>
</tr>
</tbody>
</table>

Adapted from Lovett, 1994
used (Morgan et al., 1992) which should allow the corresponding gene to be detected even if it is expressed at low levels. However problems can arise if the target gene is transiently expressed or if there is significant hybridisation to non-unique sequences, such as related members of a gene family and pseudogenes. The selected material may also be contaminated with ribosomal cDNAs, low-level repeats and mitochondrial sequences which can complicate the analysis. As with exon amplification, there is also a bias for small fragments to be amplified by PCR and sequencing errors which may be introduced with each round of amplification. However two of the major advantages to cDNA selection are that large genomic regions greater than 1Mb can be easily screened and that the procedure is very sensitive with possibly 100,000 fold enrichment of a specific cDNA sequence (Lovett, 1994).

It does seems that the universal limiting factor to these techniques is not the isolation of sequences but confirmation that they do actually represent authentic coding regions. Difficulties can arise if the products are from genes that are expressed at low levels or in limited cell types or developmental stages. The short average length of products can also result in problems during Northern and Southern analysis and the screening of cDNA libraries. For example the average length of products isolated by exon amplification have been reported as 116bp (Koyama et al., 1995) or 135bp (Church et al., 1994). This is comparable with the average lengths from this study which were 119bp using pSPL1 and 189bp using pSPL3. In contrast the average lengths of sequences isolated by cDNA selection are between 0.5-1.0kb (Lovett, 1994). Although these products have a longer length and appear to have a hybridisation advantage over exon sequences, this might not be the case for genomic DNA, if the corresponding gene is composed of small exons (Lovett, 1994).
4.1.6. Recognised genes on chromosome 9q34 which may be candidates for TSC1

Genes that have been localised between the conservative flanking markers for TSC1 include ABO, the Surfeit gene cluster, DBH, VAV2 (Smith et al., 1994b; Woodward et al., 1994; Henske et al., 1995) and CEL (Nahmias et al., 1995). Cosmid contigs have been constructed around all these loci (Nahmias et al., 1995) and their order on the chromosome with respect to the other genes, markers and cosmids is shown in figure 3.12 and figure 4.4. As transcribed sequences were identified from the Surfeit gene cluster and VAV2, their possibilities of being candidates for TSC1 were discussed in section 4.1.4.1.a. and 4.1.4.2.a. respectively. This section therefore only considers the likelihood of ABO, DBH and CEL.

The ABO blood group locus was the original marker used to confirm linkage of TSC1 to chromosome 9 and it encodes the A and B blood group glycosyltransferases. Glycosyltransferases are a family of enzymes which catalyze transglycosylation reactions between sugar-nucleotide donor and acceptor substrates. The role and pattern of expression for the ABO blood group locus does not make it a likely candidate for TSC1. In addition, the critical recombination events described by Janssen et al. (1994), Pitiot et al. (1994) and Pericak-Vance et al. (1993) which refine the TSC1 candidate region, would also eliminate ABO.

DBH is an enzyme which is in the catecholamine neurotransmitter pathway and involved in the conversion of dopamine to noradrenaline. The neurological manifestations of TSC suggested that DBH would be a good candidate for TSC1
(Fahsold et al., 1991). However the tumour suppressor gene hypothesis for TSC1 has reduced the potential of DBH as a candidate. In addition, neither of the controversial recombination events which minimise the TSC1 candidate region actually include DBH (Nellist et al., 1993; Pitiot et al., 1994).

Carboxyl ester lipase (CEL) is a major component of pancreatic juice and it hydrolyses cholesterol esters in addition to a number of other dietary esters (Taylor et al., 1991). Although CEL is not a likely candidate gene for TSC1, it is a highly polymorphic locus and therefore it may be a useful marker in genetic linkage analysis. CEL is within contig number 10 which also contains the left hand end of YAC 4DD-1 (see figure 3.19 and figure 4.5). The contig has been positioned between the flanking markers for D9S149 and D9S114 (Nahmias et al., 1995) and therefore the CEL polymorphism could play a role in refining the localisation of TSC1. Exon sequences have been isolated from this contig and they may be useful in identifying TSC1. However their value would be significantly reduced if the critical recombination events described by Pitiot et al. (1994) and Pericak-Vance et al (1993) were correct.

The human C3G gene was recently mapped to chromosome 9q34.3 by FISH (Takai et al., 1994). C3G contains a proline rich SH-3 binding domain (Tanaka et al., 1994) which indicates a role in signal transduction mediated by tyrosine kinases (Pawson and Gish, 1992). It also has a sequence at the carboxyl terminus which is homologous to guanine nucleotide releasing proteins for Ras including CDC25, Ste6 and Sos (Tanaka et al., 1994). The protein is ubiquitously expressed in human foetal and adult tissues and must be considered a candidate. A cosmid contig exists around the gene (N. Hornigold, personal communication) and has been localised between the SD-1 and 9T12
translocation breakpoints (Ekong, personal communication), but it has not yet been positioned with respect to the flanking markers for TSC1.

4.2. Characterisation of TSC2 and possible predictions for TSC1

The cloning of the TSC2 gene and the characterisation of the protein product may now facilitate identification of candidates for TSC1. As the numbers of families linked to a TSC1 or TSC2 locus are approximately even and there is no obvious difference in family size, assuming that the pattern is equivalent for sporadic patients, it suggests that the mutation rate is the same for both genes. The large number of sporadic patients indicates that the mutation rate is high and therefore suggests that the genes are susceptible to mutation. The minimum mutation rate was estimated at $2.5 \times 10^{-5}$ mutations per gene per generation, assuming that there is complete penetrance, a birth frequency of 1 in 12,000 and the proportion of new mutations was at least 60% (Sampson et al., 1989b). As the birth prevalence estimated by Osborne et al. (1991) was 1 in 5,800, this would effectively double the estimated mutation rate. This high mutation rate suggested that the gene would be very large however this does not correlate with the finding for TSC2, as the 5.5kb transcript only covers a genomic region of approximately 45kb (Sampson and Harris, 1994).

No major phenotypic differences have been detected between patients linked to TSC1 or TSC2 and this may suggest that both gene products act in the same biochemical pathway or protein complex. In addition, the pattern of expression detected for TSC2 is
likely to be the same as TSC1 and therefore a ubiquitously expressed transcript is expected.

The evidence that both TSC1 and TSC2 function as tumour suppressor genes (Smith et al., 1993; Green et al., 1994a; Carbonara et al., 1994; Green et al, 1994b) in accordance with Knudson's two hit hypothesis, suggests that a cell must be reduced to the homozygous null state before cellular growth and differentiation becomes disordered. It also suggests that cells within each tumour would be clonal in origin and this has now been demonstrated by Green et al. and was reported at the Fourth International Chromosome 9 Workshop (Pericak-Vance et al., 1995). TSC like tumours also occur sporadically in patients not affected by tuberous sclerosis and therefore it will be interesting to investigate whether these lesions also show loss of heterozygosity for markers in the region of the TSC1 and TSC2 loci.

The detection of large deletions which disrupted the TSC2 gene and reduced levels of expression for the TSC2 transcript in affected individuals, both indicated that constitutional mutations within the gene were likely to result in loss of function. These findings were consistent with the gene being a tumour suppressor and may now be expected for the TSC1 gene.

The functional motifs within the predicted protein tuberin, included four hydrophobic domains which may be involved in membrane anchorage and four potential glycosylation sites downstream of the last putative transmembrane domain. A periodic array of leucine residues was also detected which suggested the presence of a leucine zipper which is involved in protein-protein interactions. The functional significance of
these domains will hopefully be elucidated once the cellular localisation of the protein is known.

4.3. Other phakomatoses

The phakomas found in tuberous sclerosis are also characteristic of other diseases including Neurofibromatosis 1 (NF1), Neurofibromatosis 2 (NF2) and von Hippel Lindau disease (VHL). Evidence suggests that the genes responsible for these diseases, NF1 (Legius et al., 1993), NF2 (Rouleau et al., 1993) and VHL (Latif et al., 1993) also function as tumour suppressors because a similar combination of inactivating constitutional and somatic mutations have been detected. Neurofibromin, is the protein product of the type 1 neurofibromatosis gene, it showed significant sequence similarity with ras p21 GTPase activating protein (GAP) (Xu et al., 1990) and was reported to regulate ras activity (Martin et al., 1990). Tuberin, the protein product of TSC2 has a small region of homology with rap 1 GAP (GAP3) (European chromosome 16 tuberous sclerosis collaborative consortium, 1993) which specifically stimulates the GTPase activity of the rap 1 protein (p21 rap 1). Rap 1 is another member of the group of homologous GTPases involved in regulation of cell proliferation and differentiation and although tuberin shows no homology with ras GAP or neurofibromin, it suggests that these proteins may have roles in the same cellular process. However until the proteins from these phakomatoses are identified and their functions become known, the relationships between the gene products can only be considered as speculation.
4.4. TSC3 ?

There is no strong evidence for a third tuberous sclerosis locus, but as a small number of families do not appear to segregate for either TSC1 or TSC2, it cannot be ignored (Povey et al., 1994a). However if TSC3 does exist in the genome, it must be a minor disease gene locus as the majority of families are linked to TSC1 or TSC2 (Kwiatkowski et al., 1993; Povey et al., 1994a; 1994b). The existence of TSC3 may explain the conflicting positions for the TSC1 gene that have been described in section 4.1.1.1.

If genetic counselling for tuberous sclerosis patients is to be offered with complete certainty, it is imperative that all the disease genes are identified. However as only a few families remain unlinked to a locus on chromosome 9 or 16 and these ones tend to have less family members, the likelihood of localising the gene by linkage analysis is very small. Therefore attention must be paid to reports of chromosome abnormalities or patients with multiple congenital abnormalities include TSC and may facilitate localisation.

4.5. Animal models for tuberous sclerosis

The identification of an animal mutation that is homologous to a human mutation which causes a specific genetic disease provides the investigator with an animal model. This can become a powerful tool in understanding the aetiology of the disease and for attempting to correct the phenotype. The construction of comparative maps facilitates the identification of homologous mutations as conserved linkage groups between species are deduced. The candidate region for TSC1 on human chromosome 9q34

270
shows homology to mouse chromosome 2. Mouse mutations have been reported in this region but as yet, none have correlated with the disease (Povey et al., 1994b). The mouse homologue of TSC2 was mapped to mouse chromosome 17 (Olsson et al., 1994; Olsson et al., 1995; Pilz et al., 1995b), however a mouse mutation in this region which gives rise to an equivalent disease phenotype has not been reported.

The Eker rat is an example of an inherited predisposition to renal carcinoma which has an autosomal dominant pattern. The mutation has recently been mapped to rat chromosome 10 which is a region homologous to human chromosome 16p13.3 and a germline insertion was detected in the rat homologue of the TSC2 gene (rTsc2) (Yeung et al., 1994; Kobayashi et al., 1995). The normal rTsc2 transcript was a similar size to that of TSC2 (5.5kb) and was expressed in a wide range of tissues including kidney, brain, heart, lung, muscle, liver and spleen. The Eker mutation resulted in a rearrangement within the open-reading frame which produced a larger transcript (7.5kb) but the normal 3' end which had homology to rap 1 GAP was lost (Yeung et al., 1994; Kobayashi et al., 1995). Normal kidneys of the heterozygote rat expressed both the normal and abnormal transcripts, whereas primary tumour and cell lines only expressed the abnormal transcript (Yeung et al., 1994). These findings were consistent with Knudson's two hit hypothesis and suggested that rTsc2 functioned as a tumour suppressor (Yeung et al., 1994; Kobayashi et al., 1995). It supported the evidence that human TSC2 also functions as a tumour suppressor and may have an important role in studying TSC2 function. However there are phenotypic differences between the Eker rat and TSC in humans which are not been fully explained. These differences include renal carcinomas in the rat compared to angiomyolipomas in humans and no clinical epilepsy
or other CNS involvement in the rat. Therefore tuberin may play a more general role in renal oncogeneisis.

4.6. Positional cloning and alternative approaches to gene identification

In the last nine years the positional cloning approach has been used to identify at least 40 genes (Collins, 1995). The majority of these were localised by a structural aberration that could be detected cytogenetically and examples include tuberous sclerosis (TSC2) (The European chromosome 16 tuberous sclerosis consortium, 1993) and polycystic kidney disease (ADPKD) (The European polycystic kidney disease consortium, 1994). Another type of DNA rearrangement which can be detected by Southern blot analysis or PCR amplification is the expanded tri-nucleotide repeat which proved to be the mutational basis of disease for fragile X (Verkerk et al., 1991) and Huntington’s disease (The Huntington’s disease collaborative research group, 1993).

The localisation of disease genes not so readily marked have relied on genetic linkage mapping. The technique is most informative if the diseases are common monogenic traits with large numbers of families available for analysis. However if a specific mutation can be identified in affected individuals that is not present within normal chromosomes, the method of allelic association can be used to narrow down the candidate region for disease genes. It has been very important in the identification of myotonic dystrophy (Harley et al., 1992), cystic fibrosis (Estivill et al., 1987) and diastrophic dysplasia (Hästbacka et al., 1994).
Over the past few years there has been a rapid increase in the resources and techniques available for isolating transcribed sequences from genomic DNA (see section 1.2.4.) and consequently transcription maps (Gardiner et al., 1994; Rommens et al., 1994) and large numbers of ESTs have been generated (Adams et al., 1993). The ESTs are being developed by partial sequencing of randomly selected cDNA clones directly from a cDNA library (Adams et al., 1991; Adams et al., 1993). They have been deposited into public databases at a phenomenal rate and will significantly aid the identification of disease genes using a candidate gene approach. It is hoped that eventually these sequences will be associated with mapping information (Polymeropoulos et al., 1993) and then their potential in gene identification will be greatly increased.

The candidate gene approach depends on partial functional information about the disease in question and examples include Li - Frameni syndrome in which the tumour suppressor gene, p53 was studied (Malkin et al., 1990) and hereditary haemorrhagic telangiectasia type 1 in which endoglin was studied (McAllister et al., 1994). This approach does not require as extensive detail as the functional cloning approach which facilitates disease gene identification solely on knowledge of the biochemical defect (e.g. phenylketonuria and sickle cell anaemia). Furthermore it does not require the isolation of new genes but relies on information available from those already identified.

A number of genes have recently been identified by a combination of different approaches. For example, Vetrie et al. (1993) isolated the X-linked agammaglobulinemia (XLA) by positional cloning using the technique of cDNA selection. But the same gene was simultaneously identified by Tsukada et al. (1993)
using a functional approach by searching for genes regulating B cell development. In addition, Walker et al. (1993) identified the glycerol kinase gene by positional cloning using the technique of exon amplification (Buckler et al., 1991) but Sargent et al. (1993) mapped the same gene to human Xp21 by using a randomly expressed sequence tag (EST) with homology to bacterial glycerol kinase and therefore a candidate gene approach. The latter has been referred to as a positional-candidate approach (Ballabio, 1993; Collins, 1995) and it is thought that as the density of mapped transcripts rapidly increases, this will soon be the predominant approach for disease gene identification.

4.7. Summary and future Work

The physical map of human chromosome 9q34 constructed by our group is progressing rapidly and it is hoped that a complete cosmid contig between the consensus flanking markers for TSC1 will soon be achieved. The radiation hybrids that retained specific regions of chromosome 9q34 were a powerful tool for isolating DNA clones and therefore their characterisation by FISH and marker analysis was vital to the mapping strategy used (section 3.1.1.). Although only a few cosmids were isolated by myself in this study, they were confirmed to map to chromosome 9q34 and therefore contributed to the map (section 3.1.2.). These and many more additional clones have been mapped by FISH using the translocation breakpoint cell lines and the physical map generated (section 3.2, figure 3.12) has been extremely useful for the selection of cosmids for the isolation of transcribed sequences. However future work may involve the identification of cell lines with translocation breakpoints between SD-1 and 9T12, as they would be extremely valuable in refining the TSC1 candidate interval detected by FISH.
Although the ideal single method for isolating transcribed sequences has not yet been invented, the technique of exon amplification was found to be very efficient. Many putative exon sequences from the TSC1 candidate region have been generated and these are important resources for physical mapping chromosome 9q34, for construction of a transcription map and for isolation and characterisation of candidate genes for TSC1 (section 3.3).

The putative exons selected for further analysis were chosen by their physical localisation and/or their similarities to other sequences in the database. Although the recombination events that position the TSC1 gene are conflicting, the distal localisation proposed by Pitiot et al. (1994) was further strengthened by the DNA duplication in a sporadic patient reported by Smith et al. (1994a). Therefore putative exons from contig D9S10 and D9S114 were of major interest as they mapped within this more distal region. A significant amount of effort was put into isolating and characterising cDNA clones but unfortunately VAV2 (D9S10) no longer seems to be a candidate (see section 4.1.4.2.a.) and recent orientation of the D9S114 contig would suggest that the cDNAs homologous to exons D9S114 Bam A and D9S114 Pst B may no longer be in the candidate region (see section 4.1.4.3.a.). However a full length transcript has not been identified for this particular gene and the pattern of expression is as yet unknown. The sequence similarity to heterotrimeric G proteins might suggest a role in signal transduction and as this is a very interesting novel gene that has not been definitely eliminated as a candidate, the investigation continues. In addition, cDNA clones homologous to exon D9S114 Bam B are potential candidates for TSC1 and are also currently being analysed further.
The SURF contig was studied because very little is known about this cluster of housekeeping genes that have been very highly conserved through evolution (Yon et al., 1993). cDNA clones for SURF-5 and another gene in the region have been identified and may be useful for further analysis. Although their positions are more proximal within the consensus candidate region it is impossible to eliminate them as candidates on this basis alone, as there is no single critical recombination event which refines the target interval and is entirely convincing (section 4.1.1).

Finally, transcribed sequences were also isolated from contig 37 and cDNA clones identified. However as the only mapping information available for this contig is a localisation within the SD-1 and 9T12 translocation breakpoints, the cDNAs have not been investigated beyond FISH mapping to chromosome 9q34 and partial sequencing.

Although the cloning and characterisation of the TSC1 candidate region may have brought us very close to the TSC1 gene in molecular terms, the gene remains unidentified. A refinement in the candidate region on chromosome 9q34 would significantly aid gene identification but as described earlier, genetic heterogeneity seriously hampers the linkage analysis (section 4.1.1.1). Allelic association has been useful for a number of diseases (section 1.2.1.3) but as the mutation rate for TSC is high, it is thought that the approach is unlikely to be successful. The serendipitous discovery of a chromosomal abnormality on chromosome 9q34 associated with the disease would greatly facilitate the positional cloning approach. However, as the physical and transcription maps are rapidly improving, it may be the candidate gene approach which ultimately proves to be successful.
Once the TSC1 gene is identified the difficult task of screening for mutations will commence. Over 500 TSC patients have been screened for gross rearrangements within TSC2 by Southern blot analysis, however less than 3% have shown deletions (Pericak-Vance et al., 1995). If the same follows for the TSC1 gene, a more sensitive and efficient method will certainly be required. Assuming that both TSC1 and TSC2 are tumour suppressor genes and that the mutation results in a loss of function, the direct protein truncation test may be the most applicable method of mutation detection (Roest et al., 1993). The technique has been very successful in the identification of translation terminating mutations in the adenomatous polyposis coli gene (APC) (Van der Luijt et al., 1994) and Duchenne muscular dystrophy gene (Roest et al., 1993). The advantage to the method is that a wide spectrum of mutations can be rapidly detected within a large gene, however they must all result in a termination of translation.

The characterisation of mutations in TSC1 and TSC2 will be of interest and of practical value in genetic counselling and risk assessment. It is possible that some correlation between the nature of the germline mutation and the severity of the disease may be found. However, the variation within families suggests that occurrence and timing of the ‘second hit’ may be critical, or that variation in other genes may be important from this respect. It is hoped that the discovery or production of an animal model may provide some information on the developmental stage at which an effective second hit occurs and on the normal role of the genes involved. The elucidation of the function of the TSC1 and TSC2 genes is not only necessary for the understanding of the pathogenesis of tuberous sclerosis but may also reveal a previously unknown control pathway for cell proliferation and differentiation.
Bibliography
Bibliography


Pilz, A., Glenister, P., Povey, S., Lyon, M and Abbott, C. (1995b). The mouse homologue of the human tuberous sclerosis 2 (TSC2) gene maps to chromosome 17, but does not fall within the \textit{t}^{\text{w18}} or \textit{t}^{\text{h20}} deletions. \textit{Genomics} \textbf{26}: 420-421.


moesin-, ezrin-, radixin-like gene is a candidate for the neurofibromatosis 2 tumor suppressor. *Cell* 72: 791-800.


Publications arising from this thesis
Publications arising from this thesis


Appendix
Appendix

1. Materials

1.1. Standard solutions and buffers

**Church hybridisation buffer**
0.5M NaPi pH 7.2; 7% SDS; 1mM EDTA

**Denaturing solution**
1.5M NaCl; 0.5M NaOH

**Formamide stop solution**
98% deionised formamide; 10mM EDTA, pH8.0; 0.025% w/v bromophenol blue; 0.025% xylene sulphate.

**Hybridisation buffer**
1M NaCl; 10% w/v dextran sulphate; 1% w/v SDS.

**Loading buffer (agarose gels)**
0.25% w/v bromophenol blue; 0.25% w/v xylene cyanol; 25% w/v Ficoll in distilled H2O. The Ficoll was dissolved at in water at 65°c prior to the addition of the other dyes.

**Lysing solution**
0.5% SDS, 100g/ml proteinase K in STE

**Neutralizing solution**
1.5M NaCl; 0.5M Tris-HCL pH7.5

**10x PCR buffer (HT Biotechnology Ltd)**
500mM Tris-HCL, pH 9.0; 500mM KCl; 70mM MgCl2; 160mM (NH4)2SO4

**PBS**
150mM NaCl; 10mM Na2HPO4/NaH2PO4, pH 7.4
For 200ml 1x PBS : added one tablet (Sigma) to 200ml sterile distilled water.

**RNase A**
Purchased in a dessicated form and made to a stock concentration of 10mg/ml using sterile water. The solution was boiled for 10 minutes to inhibit DNase activity and then stored at -20°c.

**1M sodium phosphate solution pH7.2 (NaPi)**
Prepared by titration of 1 molar solutions of Na2HPO4.12H2O and NaH2PO4.2H2O

**SSC**
150mM NaCl; 15mM sodium citrate, pH7

**STE**
0.1M NaCl, 10mM Tris-HCl pH 8.0, 1mM EDTA pH 8.0

**TAE**
40mM Tris-acetate, pH8.0; 40mM sodium acetate; 1mM EDTA.

**TBE**
90mM Tris-HCL, pH8.0; 90mM boric acid; 2mM EDTA.

**TE**
10mM Tris-HCL pH8; 1mM EDTA pH7.6

**TNE**
100mM Tris-HCLpH8; 100mMNaCl; 1mM EDTA pH8.0

**TKM 10:10:1**
10mM Tris pH 7.5, 10mM KCL, 1mM MgCl₂

### 1.2. Bacterial cell culture

**Ampicillin**
Prepared to a stock concentration of 50mg/ml using sterile water, filter sterilised (pore size 0.22uM) and stored at -20°c. Working concentration 50-100µg/ml.

**IPTG**
Stock solution made to 200mg/ml. 4µl of the stock was added to a 9mm pre-made agar plate with appropriate antibiotic selection 30 minutes prior to use to give a working concentration of 30µg/ml.

**Kanamycin**
Prepared to a stock concentration of 25mg/ml using sterile water, filter sterilised (pore size 0.22uM) and stored at -20°c. Working concentration 12.5-25µg/ml.

**LB medium**
1% "/v tryptone; 0.5% "/v yeast extract; 0.5% "/v NaCl.

**LB agar**
LB medium with 1.5% "/v agar

**LB glycerol medium**
LB medium with 5% "/v glycerol.

**TFB1**
10mM MES; 100mM RbCl₂; 10mM CaCl₂, 2H₂O; 50mM MnCl₂, 4H₂O
All salts should be added as solids and the pH adjusted to 5.8 with acetic acid. The solution is then filter sterilised (0.22μM pore size) and stored at 4°C for upto one year.

**TFB2**

10mM MOPS; 75mM CaCl₂, 2H₂O; 10mM RbCl₂, 15% glycerol
The solution should be adjusted to pH 6.5, filter sterilised and stored at 4°C.

**NZY agar**
NZY broth with 1.5% w/v agar.

**NZY broth**
1% w/v NZ amine (casein hydrolysate); 0.5% w/v yeast extract; 0.5 %w/v NaCl; 0.2% w/v MgSO₄·7H₂O. Adjust to pH 7.5 with NaOH.

**SM buffer**
per 1L : 5.8g NaCl; 2g MgSO₄·7H₂O; 50ml 1M Tris-HCL (pH 7.5); 5ml 2% w/v gelatin.

**2 x YT medium**
% w/v tryptone; w/v yeast extract; w/v NaCl.

**XGAL**
Stock solution made to 20mg/ml in dimethylformamide. 40μl of the stock was added to a 9mm pre-made agar plate with appropriate antibiotic selection 30 minutes prior to use to give a working concentration of 30μg/ml.

### 1.3. Mammalian cell culture

**Eagles minimal essential medium (MEM)**
For 500ml : 390ml ddH₂O; 45ml Eagles MEM (10x); 50ml foetal calf serum, 5ml GPS; 5ml 100 x sodium pyruvate; 5ml non-essential aminoacids and 5ml Hpes (1M). Correct to pH 7.2 with 1-2ml NaHCO₃/NaOH.

**Glycerol medium for freezing**
For 100ml : Mix together 72ml ddH₂O and 10ml glycerol and sterilise by autoclaving. Add 8ml Eagles MEM; 1ml Hpes (1M) and 20ml foetal calf serum to the sterile bottle and bring to pH 7.2 with 0.25-2.5ml NaHCO₃/NaOH.

**GPS**
For 100ml : 200mM L-Glutamine (ICN Labs); 600mg Penicillin (U.C.H. pharmacy); 1g Streptomycin (Sigma). Freeze in 1ml aliquots and dilute to a working concentration of 1% v/v in medium.

**HANKS’ balanced salt solution**
For 300ml stock solution : 32g NaCl; 1.6g KCL; 0.36g Na₂HPO₄·12H₂O; 0.24g KH₂PO₄; 4g glucose; 0.08g phenol red. Made upto correct volume with ddH₂O, autoclave and stored at 4°C.
Working concentration: dilute 1 in 12 with ddH₂O, autoclave and store at 4°C. Before use bring to pH 7.0-7.2 with NaHCO₃/NaOH.

NaHCO₃/NaOH
For 100ml: 5.6g NaHCO₃; 6ml 10M NaOH. Mix thoroughly and sterilise by membrane filtration.

RPMI - 1640 medium
For 100ml: 76ml H₂O; 9ml RPMI (x10); 10ml fetal calf serum; 1ml GPS; 1ml non-essential amino acids; 3ml 5.3% w/v NaHCO₃; 1-2ml 1M NaOH.

Trypsin/Versene solution
For 100ml: 0.02g EDTA in 100ml sterile Hanks’ balanced salt solution (working concentration). Before use add 5-10ml 1% trypsin solution (ICN labs) and correct pH with NaHCO₃/NaOH.

1.4. Solutions for FISH

Deionised Formamide
For 100ml: 5g ion-exchanger Amberlite monobed MB-1 mixed resin was added to 100ml formamide. It was sealed in a Duran bottle, covered from light and agitated gently for 2 hours. The solution was filtered twice through Whatman 1MM paper and used immediately or stored -70°C.

Formaldehyde buffer
50mM MgCl₂ in 1 x PBS

Hybridisation mix
50% v/v deionised formamide; 2 x SSC, 10% w/v dextran sulphate. Incubation at 65°C proceeded for 2 hours to ensure that the dextran sulphate was completely dissolved and then the mix was stored at -20°C.

KCL/EDTA hypotonic solution
For 100ml: 0.3g KCL (3%); 0.02g (0.2%) EDTA; 0.48g (4.8%) Hepes; pH 7.4. Sterilised by autoclaving and store 4°C.

10x Phosphate buffer
0.5M NaH₂PO₄; Na₂HPO₄ pH7

Proteinase K
Purchased in a dessicated form and made to a stock concentration of 50μg/ml using sterile water.

Proteinase K buffer
20mM Tris-HCL; 2mM CaCl₂ pH7.4
2. Primer Sequences

SD6     TCT GAG TCA CCT GGA CAA CC
SD2*    GTG AAC TGC ACT GTG ACA AGC
SA2*#   ATC TCA GTG GTA TTT GTG AGC

SD1*    CCC GGA TCC GCG ACG AAG ACC TCC TCA AGG C
SA1*    CCC GTC GAC GTC GGG TCC CCT CGG GAT TGG

dUSD2#  CUA CUA CUA CUA GTG AAC TGC ACT GTG ACA AGC TGC

dUSA4#  CUA CUA CUA CUA CAC CTG AGG AGT GAA TTG GTG G

SD5     CCC TCG AGG TCG ACC CAG C
SA5     CTA GAA CTA GTG GAT CTC CAG C

KSF1#    CTG CAC GTG CTC TAG AGT CG
KSR1#    CCT CGG GAG ATC TCC AGG T

* primers used with pSPL1

# primers used most successfully with pSPL3
3. Exon sequences

3.1 Exon sequences using pSPL1

1bv2.seq (without HIV)
Derived from cosmid CKW1B. Length: 164

```
1  aaactgaggc tcggggagaa ggaagacagt acgggcccaa gggccacccc
gccaagcagc atgacatcca ccccacagcc catggtgtagg aatctgcgccc
101 actccgcct cactgccgct acaaattggg agctgcacgg tgagacctag
151 agtcactgct gatc
```

17BEBam1.seq
Derived from cosmid 17BE, contig with D9S10. Length: 101

```
1  cgagcatgac ctggggagga catctacgac tgcgtccctg tgaggatgga
ggggacgaca tctacgagga catcatcaag gtggaggtgc agcagcccat
51 gggacgaca ttgctgacgca tctacgagga catcatcaag gtggaggtgc agcagcccat
101 g
```

17B2 Bam.seq
Derived from cosmid 17B2, contig with D9S114. Length: 112

```
1  ttctctatca aagcagttgg agtgtcgcct cccccctctg ctggtgcagc
cacctgcct cccgggctca ggaattcccc tgtcactgcct tccgagtaga
tctgggattacag
```

17B2 BB.seq
Derived from cosmid 17B2, contig with D9S114. Length: 128

```
1  accaggtggt ccacagggac gagctttgct gctctcctctc atcgcacagt
cgtgcaagac ccacagctgc cttctcctctg cctccccctc atcgcacagt
51 cgtgcaagac ccacagctgc cttctcctctg cctccccctc atcgcacagt
tccgccgggc ccgccggcct cacacgac
```

17B2Bgl.seq
Derived from cosmid 17B2, contig with D9S114. Length: 74

```
1  cttctctgtg tagtgcacag gctctcctctc atcgcacagt
cgtgcaagac ccacagctgc cttctcctctg cctccccctc atcgcacagt
tccgccgggc ccgccggcct cacacgac
```

kw10bam2.seq
Derived from cosmid CKW10. Length: 85

```
1  ggtcaccag ggaaggccc gcagccgagc cagcccgagc agccctgtgc
51 agggactcct gtcttgagag ggggaaggg gcagccgagc cagcccgagc
```

kw1bam1v2.seq
Derived from cosmid CKW1. Length: 144

```
1  acctaggtgc tcacagctgc gcacaggtgc gctctcctctc atcgcacagt
cgtgcaagac ccacagctgc cttctcctctg cctccccctc atcgcacagt
tccgccgggc ccgccggcct cacacgac
```
mct7bam1.seq
Derived from cosmid MCT136-7. Length: 200

1 gcceggcctac tcgctgcgag atacccagca cgcggggcac ttcggagcct
51 gtcgagcgtg ggagcaccag cactcacagg gcagcctgcc acgcttccag
101 gatcgtgagg agttcggtgc accaaagccc tcacccgttg gatccgtgagg
151 aaggaagca cccaccacct attaacatca gatgctaatg atatatacga

mct7bam3.seq
Derived from cosmid MCT136-7. Length: 64

1 cgtctctct tcgcaagggg atctgctgc cagcggcgct ccggatgccg
51 gaccggaccct ccat

mct7bb1.seq
Derived from cosmid MCT136-7. Length: 87

1 gggacaccttc taccagggat acatgtgtac caagtcgtgg tcggggcaca
51 caaggagtgc ctgggaatac tagctcctcga gatgtc

mct11bam1.seq
Derived from cosmid MCT136-7. Length: 200

1 gcceggcctac tcgctgcgag atacccagca cgcggggcac ttcggagcct
51 gtcgagcgtg ggagcaccag cactcacagg gcagcctgcc acgcttccag
101 gatcgtgagg agttcggtgc accaaagccc tcaccccgttg gatccgtggg
151 aaggaagca cccaccacct attaacatca gatgctaatg atatatacga

mct11bam3.seq
Derived from cosmid MCT136 7. Length: 64

1 cgtctctct tcgcaagggg atctgctgc cagcggcgct ccggatgccg
51 gaccggaccct ccat
3.1 Exon sequences using pSPL3

SURFBamA.seq
cosmid LL09 255A6, length: 252
1
gacgagacgc aggtgtcacg ggccactcag ggtgaacagg acaattacga
51
gatgcatgtg cgagccgcca acatcgtccg agccggcgag tccctgtatga
101
gctggtgtc cgaccccaag cagttctcag tctctaatga ctteccctcc
151
201
251

SURFBamB.seq
cosmid LL09 255A6, complete, length: 159
1
gtcccaggtc tencnctgcg ctcgccatgg cccagcagag
gagccctgccc cagagcaagg agacgtgctgc agtcctacca aacacgcgcc
101
tgaaggacga catfagttgc atcatggaca acctacca gatcatcaag
151
acreccag

SURFBamC.seq
cosmid LL09 255A6, complete, length: 101
1
gaattcatag ggaagaggtt ccagcagctc aggctccttc ccattgggtc
51
tacagtctg tgccttcggg tggagcaggc tggcgcttca gttgaatcc
101

SURFPstA.seq
cosmid pool (LL09 255 A6; LL09 211 A7), length: 177
1
gcgaccgagt tgctcttgcc cggcgtcaac acgggataat accgcgccac
51
atacgtgatc atctctcgggg tggagggagca ttcgcctagc acgcgccctc
101
aatcactgaa ctcgtctgat gacgttgtgt ggtggtctgctg gttgctacg
151
gtgcacccaa ctgatcttca gcatctt

SURFPstB.seq
cosmid LL09 255 A6, length: 151
1
gattccctcc gtgaacgagg ccattgacca gcgcaaccag cacgtgacga
51
cacgtcagga cccagatatat ctgcctcgttg tggaggggaag
101
cacaccaacc aactctatgt tgtctctgac gatcctactg atatacga
151

SURFPstC.seq
cosmid LL09 255 A6, length: 67
1
gggagaatga taatggagaa aggagagata aaaaactgct cttcctaatat
51
cacacaagc aataagag
**SURFPstD.seq**
cosmid LL09 255 A6, complete, length: 81
1
  atggaggacg agacgcaggt gtctcgggct actcagggat gaacaggaca
51
  attacagatg caigtgccag ccgccaacat c

**D9S10BamA.seq**
cosmid pool(LL09 255H4, LL09 124 E8 and 17BE), complete, length: 198
1
  ctggccgatc tctggcctct tcaactgcat caccatccac cctctggaca
d9
  tcgcggccgg cgtgtggatg atcatgaatg ccttcatctt gttgctgtgt
101
  gaggcgccct tcgctgtgcc ctctatcgag cgttgccgga
151
  gaagggtgac cggctgtgcgt cctggcagaa ggctgctcttc tactgcgg

**D9S10BamB.seq**
cosmid pool(LL09 255H4, LL09 124 E8 and 17BE), complete, length: 116
1
  ttttaaagtgc actgatttga agaatgatac taataccaat agtagtagcg
51
  ggagaatgat aatgganaaa ggagagataa aaactgtcct ttctaatatc
101
  agcacaagca taagag

**D9S10BamC.seq**
cosmid LL09 255 H4, complete, length: 103
1
  cgagcatgac ctgggggagg acatctacga ctgcgtcccg tgtgaggatg
51
  gaggggacga catctacgag gacatctca aggtggaggt gcagcagccc
101
  atg

**D9S10BamD.seq**
cosmid LL09 255H4, complete, length: 69
1
  gccctttccc tcagaggaga ccacagagaa tgacgatgac gtctaccgca
51
  gctgaggaga gctggecga

**D9S10PstA.seq**
cosmid LL09 124 E8, complete, length: 134
1
  caaggtcata gtgaagtgacc attactgagg caccacagtc ccagcagcat
51
  caagggcga aaacttgacct gcaacagaag ggtggtcagag ccagaagagg
101
  agcagctgac accacaggag gcacccaggg ctga

**D9S10PstB.seq**
pool of cosmids (LL09 255 H4; LL09 124 E8; 17BE), complete, length: 106
1
  gtctcagcag agatgctctc tcctcacaag acgcctctctg gacccacagg
51
  ctaagctcc caaaaagccc tgtgtgcttc ccattgtage ctgattatc
101
  tgtagg

**D9S114BamA.seq**
cosmid LL09 157B7, complete, length: 273
D9S114BamB.seq
cosmid LL09 222 A10, complete, length: 230
1 ccctacacctg tgaagccaaa ctatgctcta aagttcaccc ttgctggcca
51 caccaaagca gtgtcctcgc tgaattcag cccgaatgga gggtggctgg
101 caagttcata tgtctgataaa ctatattaaa ttgsggcgc gtatagtgag
151 aaatgtgaga aaaccatgtc tgtgacaag ccctttatgt tgtatgtgac
201 ctgtgctgca gattctacc ctcttctgc tgcctcagat gaccaacacct
251 tgaagatag ggacgtgagc tcg

D9S114BamC.seq
cosmid LL09 157 B7, 3' complete, length: 196
1 ccctacacctg tgaagaaact atgctctaaa gttttctctt gctggccaca
51 ccaaagcagt gtgtcctcgc aatttcttca cctcatgtga gggataaagc
101 gtgtcctcgc gatgtgagtc cagttttctc aggctgtaga ggtgtgtgc
151 tgtctgccca gatgtgagtc cagttttctc aggctgtaga ggtgtgtgc
201 tgaagatag ggacgtgagc tcg

D9S114BamD.seq
cosmid LL09 222 A10, complete, length: 146
1 gctgtacagg cttctgttct gggaatacct caggaaactt acaatcgtgg
51 ctgaaggtga aggagaagca gtcacgttta ccgggccagc gcagcaggaag
101 agagtgaagg gcgaagctc actacacttc aaacaaccga gacgatctt

D9S114BamE.seq
cosmid LL09 222 A10, complete, length: 116
1 tgtcaggagg ggagagcata aaactccgt caatacctct ggagtgaatg
51 aacgaatgaa tgttctagct acgagagagt attcctacct catttatcct
101 cacgccaatc ctatga
g

D9S114PstA.seq
cosmid pool (LL09 157 B7, LL09 222 A10, 17B2 and 26.10D-
this cosmid is no longer in the contig), length: 164
1 gcggggcgcct tgtgacctccg cgacagcag gaattctgct ggagtgaatg
51 cgccctcaac ggtgctgctc acctttcctc aggagcagc agacctggtttt
101 cgccctcaac ggtgctgctc acctttcctc aggagcagc agacctggtttt
151 ggagtgaagc cc

D9S114PstB.seq
cosmid pool (LL09 157 B7, LL09 222 A10, 17B2 and 26.10D-

333
this cosmid is no longer in the contig) complete, length: 198
1  gatcgggat atatatgtga ctgcactgtg acttgctgca cgctctagag
51  tcgagggagc aactgcctcg tcaccgggtc cctccaccct tgtctcctgt
101 gcggccagcg tcagagccat ggcgacggag gagaagaagc ccgagaccga
151 ggccgccaga gcacagccaa cccttccgca tcggccacgc agagcaag

D9S114PstC.seq
cosmid pool (LL09 157 B7, LL09 222 A10, 17B2 and 26.10D-
this cosmid is no longer in the contig) complete, length: 74
1  attctgcttc ctagagactt ctccagcctc gggaaagttg gctccatgga
51  gccatccggg ccactcatgt gtgg

Ctg37BamA.seq
cosmid LL09 180 F1, length: 246
1  tgacagttac ccccaacctg accagtgtct acattgccaa ccacagggct
51  ccctcaccac acttgcgcgtt gaggatcgag ggctccagtc tcgctgcttg
101 gagaatatgc aclccctca caaccaagta cccctcctct ccggagatatg
151 gctcggctga gtacagagtg gccctcaccg aggaccggct gccccgcctg
201 gaggagatcc gcctcgcgttt ccctccagg aagatgtgaa catcag

Ctg37BamB.seq
pool of cosmids (LL09 288 E8; LL09 180 F1), length: 162
1  ttaaacgcga aaaggggcat aaatagcgaa aacccgcgag gtcgccgccc
51  cgtaacctgt cggatcaccg gaaaggaccc gtaaagtgat aatgattatc
101 atctacatat cacaacgtgc gtggaggcca tcaaaccagc tcaaatatca
151 attatagacgc ag

Ctg37BamC.seq
cosmid LL09 180 F1, length: 82
1  cgctgacaca ggctacaaca caggcatgca gcagcacact gtattcctag
51  tggggaggtt ctccctgcag ttctgggatc ag

Ctg37BamD.seq
cosmid LL09 180F1, length: 96
1  ccctcagggt gcctggcgcg gaagagctgt gtggcctggc aagcaagcct
51  gggagccggc gggaggtctg ccnggcctgc cccntnccct gacccg

Ctg37BamE.seq
pool of cosmids(LL09 288 E8; LL09 180 F1), complete, length:88
1  ccctcagat gcgtgcatag ctatgaacc atatcgcgtc tgaagggcnng
51  gttctgggag agacttaatg aagagcttgg ccacggag

Ctg37PstA.seq
cosmid LL09 288 E8, length: 104
1  tccttttagca ggttggagcc ccagggcctg ggacagcctg ccgctgccag
51  caacctceca ctgctgcecta ggttgcagcg cccactgtca tgcttctgaa
101  gaag

Ctg37PstB.seq
cosmid LL09 288 E8, 5' complete, length: 124
1  tcatctccca aggccgcctt cccagcggcc gcctgcctcc ctgctgactc
51  ctgcccgtgt ctcttgtttc aacgttgcgc gatctctggc ctcttcaact
101  gcatcaccat ccaccectgt aaca

Ctg37PstC.seq
cosmid LL09 288E8, complete, length:116
1  tttaaagtgc actgatttga agaatgatac taataccaat agtagtagcg
51  ggaagaatgat aatggagnaa ggagagataa aaaaactgctc tttcaatatc
101  agecacaagna taagag

Ctg37PstD.seq
cosmid pool (LL09 288 E8 and LL09 180 F1), complete, length: 149
1  gacctgatca aggtgcatca cagcttcctg agggccatcc gtgtccggat
51  ggtggggggc agcacgctgg ccaaggtctt cctcgatttc aaggaaaggc
101  ctggagatct cccgagggac ccgacagccg aagaatagaa gaagaaggt

Ctg166BamA.seq
cosmid LL09 84 B3, 5' complete 3' not complete, length: 126
1  gtgcaagtgc aacctgcatg ccacctgtgc tccatgcgcg agggcagcct
51  gcagtgcgag tgcgagcaca acaccaccgg ccccgactgc ggcaagtgca
101  agaagaattt ccgtcacccg gtctgg

Ctg166BamB.seq
cosmid LL09 84 B3, HIV SDC2, 3' complete, length: 103
1  ctgatagtga aatgtcatt aataccactgt tcatgcggag agggcagcct
51  gcagttgcag tgagcgcaca acaccaccgg ccccgactgc ggcaagtgca
101  aagaagaagtt ccgtcacccg gtctgg

Ctg166BamC.seq
cosmid LL09 84 B3, length: 34
1  ggggttcatt cctgtcaact taatccttga aggg

Ctg166PstA.seq
cosmid LL09 194 E12, length: 165
1  tctggagctc gacgcccgtc gcaagatgga ttacacaaaa caaaccagtg
51  gatatgaa agaagtctgc aatcagcagc gacagcagt gacgcagcagc
101  agggcaggag ggggtctctc ggaatggatgt tctaccctcc ctgtcctttc
151  agcataggga tttaa

Ctg166PstB.seq
cosmid LL09 84 B3, length: 142
   1  taaaaacatt taagttatga cgaagaacgg aaacctttaa acgggaaat
   51  tttcataaat agcgaaaaacc ccgcgagtgc ccgccccgttaa ctctgcggat
  101  caccggaaag gacccgtagt gtatgttttct ctcataatca ca

Ctg166PstC.seq
cosmid LL09 194E12, complete, length: 169
   1  gatccactgt accagcagca gctgctgatc tttcaagaga ctgatggacg
   51  ttaggttctt tccaaacttg tgtcatttta atttcctcaa caacccgca
  101  aagacggttg gtatcagag gaactgagat gaggaggt gaatgacct
  151  acccaaggtc acatggtca

Ctg166PstD.seq
cosmid LL09 84B3, length: 137
   1  acagggagac agcgtgtaag gcagatnctc agggagcgga cagccgcaag
   51  caaggcaga gactggagtg acgcagctgc aagcccagga acnccagagn
  101  tgcacggcgc caccaggagc gaggaggg cgagagg

Ctg166PstE.seq
cosmid LL09 194E12, length: 109
   1  gcactgattt naagaatgat actaatacca atagtagtag cgggagaatg
   51  ataatngaga aaggagagca taaattnctg ctctttcaat atcagcacaa
  101  gcataanag

Ctg166PstF.seq
cosmid LL09 84B3, complete, length: 74
   1  ggctgctatt aacaagagat ggtggtaata acaacaatgg gtccgagatc
   51  ttcaagacctg gaggaggaga tatg

4DD1BamA1.seq
cosmid pool(LL09 203 H12; LL09 256 C3), length: 141, 3' end overlaps exon DK7,
does not hit 4DD1BamA2.seq
   1  ctcagagagg gaacgggtca tgactaaagt ctctggctca gaggacggtg
   51  cagctgtggt tctgcccag ctcgggctgc ctcggcacag gcctgaagc
  101  gaatgaccag gacgctttgct acaacgatcct attttaggat

4DD1BamA2.seq
cosmid pool (LL09 203 H12; LL09 256 C3), length: 198, SDc 3' end HIV has been
removed, overlaps DK1 does not hit 4DD1BamA1.seq
1  aagaagaatc tgcaaaaggc ttgcggaggc tgcattttca tctttccacc
51  atttctgaag cgtggcccct tctgctctta ggggagcctg ctgaacatgt
101  gtctcttg ggctatctgg ggtcagcgtc taggtatggc ttccagcaca
151  gccaagaagc tccttcceca ggtgcgiggg gcccaggaaac cagccctg

4DD1BamB1.seq
cosmid pool (LL09 203 H12; LL09 256 C3), does not hit 4DD1BamB2.seq, length: 241
1  ggaatttctc ttcaaatgac tgcgtgacct gtctctgaca cccttctgtg
51  tgaaccacgg cctttacttt ctgaatttgg acactttgac atctggagcc
101  ttgctaaccc tggaggagtc ttctcctgga gttggctaat ttctagaaat
151  agcagatact gaatttcac ccagcgcctc tcceccacca cctaaagcaa
201  acctgggaac ctaaacacct gcgcagagtc ttcagagggc g

4DD1BamB2.seq
cosmid pool (LL09 203 H12; LL09 256 C3), does not hit 4DD1BamB1.seq, length: 242
1  cttgatgcctttctgataag tagcagatgta cttgccttca ttcctatggc
tggtctccct caccagccta aagcacaacc tgggaaccct tcaaacacct
tgcgtgacata ccctggaggaa gttgcgiggg gcccggcggc tcceccacca cctaaagcaa
201  acctgggaac ctaaacacct gcgcagagtc ttcagagggc g

4DD1BamC.seq
cosmid pool (LL09 203 H12; LL09 256 C3), length: 203
1  acacccttct gttgcgccca cggcctttac tttctgaatt tggacacttt
gacatctggg gctttattg cttggttcct cttggttcct
tgcgtgacata ccctggaggaa gttgcgiggg gcccggcggc tcceccacca cctaaagcaa
201  acctgggaac ctaaacacct gcgcagagtc ttcagagggc g

4DD1BamD.seq
cosmid pool (LL09 203 H12; LL09 256 C3), length: 147
1  gccctttct cttctggggc gcctttacn ttctgaattt nnacactttg
acacccctct gttggaaccn cggcctttac tttctgaatt tggacacttt
gacatctggg gctttattg cttggttcct cttggttcct
tgcgtgacata ccctggaggaa gttgcgiggg gcccggcggc tcceccacca cctaaagcaa
201  acctgggaac ctaaacacct gcgcagagtc ttcagagggc g

4DD1BamE.seq
cosmid LL09 203H12, about 200bp, 3' complete, length: 130
1  gccctttct cttctggggc gcctttacn ttctgaattt nnacactttg
acacccctct gttggaaccn cggcctttac tttctgaatt tggacacttt
gacatctggg gctttattg cttggttcct cttggttcct
tgcgtgacata ccctggaggaa gttgcgiggg gcccggcggc tcceccacca cctaaagcaa
201  acctgggaac ctaaacacct gcgcagagtc ttcagagggc g

4DD1PstA.seq
cosmid LL09 256 C3, length: 168..
1  gcacccgatcc tcctctctgg aaaaagggcac acacccgggtg cggcgcggttt
51  aggacaccca tggccccatcc tcagtcactg gtgcatggcc cttccccccc
101  aacctgaaaa ttccactggt cccccacccc accgagcctt aggatgecca
151  ggctgtttcc cteccagca

4DD1PstB.seq
cosmid pool (LL09 203 H12; LL09 256 C3), length: 140
1  atgatccca gacactgagt gagacatggg aaggaagccg gtgctcagag
51  agggaacgg tcatgactaa agtctctcgg ccagagacg gtgcagctgt
101  ggttctggtcc cagctccggc gtcctcggca cagagcttga

4DD1PstC.seq
cosmid LL09 256C3, about 180bp, 3’ complete, length: 116
1  gggacngcgg cncngccctg cagggatccc agatatctcg tgntcccgta
51  cctgtgtgga aggaagcgac caccactcta ttttgctggt cagatgcgta
101  agcatatgat acagag

4DD1PstD.seq
cosmid LL09 203 H12, length: 111
1  ctctagagtc gacccagcag atcccagata tctggtgatc ccgtacctgt
51  gtggaaggaa gcaaccacca cttcattttg tgtcatcagat gcxaacgat
101  atgatcagag
4. cDNA Sequences

cDNA clone number 5
length: 1195

   1  ccctgtaata aaaaatattaa aacttttgtt gactctgttt gaacacaata
   51  gactaggaac tcctccttag acaagttgtag aattggcaag aatgtcaacc
  101  aacaactttct aagacacttct cgggtgcttc cagcctcttc ccggcaaat
  151  ccatttcga gaggaggctc aatcagcagc cagctctgac cccagcgc
  201  cacttgccgg gaccaccttcc agacacttcc cacagccagt attccgggtc
  251  cacttcggga caggcagag ggaaaggacct tagcagctc
  301  tcctccag ttaatttttg tcgctatttt ctacgctgava gaggcgatga
  351  ggtttttgg tggcgagcaca gctttgatgt gttttagcct
  401  ggtttttgg tggcgagcaca gctttgatgt gttttagcct
ggtttttgg tggcgagcaca gctttgatgt gttttagcct
ggtttttgg tggcgagcaca gctttgatgt gttttagcct
ggtttttgg tggcgagcaca gctttgatgt gttttagcct
ggtttttgg tggcgagcaca gctttgatgt gttttagcct
ggtttttgg tggcgagcaca gctttgatgt gttttagcct
ggtttttgg tggcgagcaca gctttgatgt gttttagcct
ggtttttgg tggcgagcaca gctttgatgt gttttagcct
ggtttttgg tggcgagcaca gctttgatgt gttttagcct
ggtttttgg tggcgagcaca gctttgatgt gttttagcct
  451  ggtttttgg tggcgagcaca gctttgatgt gttttagcct
  501  ggtttttgg tggcgagcaca gctttgatgt gttttagcct
  551  ggtttttgg tggcgagcaca gctttgatgt gttttagcct
  601  ggtttttgg tggcgagcaca gctttgatgt gttttagcct
  651  ggtttttgg tggcgagcaca gctttgatgt gttttagcct
  701  ggtttttgg tggcgagcaca gctttgatgt gttttagcct
  751  ggtttttgg tggcgagcaca gctttgatgt gttttagcct
  801  ggtttttgg tggcgagcaca gctttgatgt gttttagcct
  851  ggtttttgg tggcgagcaca gctttgatgt gttttagcct
  901  ggtttttgg tggcgagcaca gctttgatgt gttttagcct
  951  ggtttttgg tggcgagcaca gctttgatgt gttttagcct
1001  ggtttttgg tggcgagcaca gctttgatgt gttttagcct
1051  ggtttttgg tggcgagcaca gctttgatgt gttttagcct
1101  ggtttttgg tggcgagcaca gctttgatgt gttttagcct
1151  ggtttttgg tggcgagcaca gctttgatgt gttttagcct

M13 universal primer length: 383

   1  gcggagcaca tagcaagaat taatgcaac aacagattc aagcagattg
   51  ggtttttgg tggcgagcaca gctttgatgt gttttagcct
ggtttttgg tggcgagcaca gctttgatgt gttttagcct
ggtttttgg tggcgagcaca gctttgatgt gttttagcct
ggtttttgg tggcgagcaca gctttgatgt gttttagcct
ggtttttgg tggcgagcaca gctttgatgt gttttagcct
ggtttttgg tggcgagcaca gctttgatgt gttttagcct
  101  ggtttttgg tggcgagcaca gctttgatgt gttttagcct
ggtttttgg tggcgagcaca gctttgatgt gttttagcct
ggtttttgg tggcgagcaca gctttgatgt gttttagcct
ggtttttgg tggcgagcaca gctttgatgt gttttagcct
ggtttttgg tggcgagcaca gctttgatgt gttttagcct
  151  ggtttttgg tggcgagcaca gctttgatgt gttttagcct
  201  ggtttttgg tggcgagcaca gctttgatgt gttttagcct
  251  ggtttttgg tggcgagcaca gctttgatgt gttttagcct
  301  ggtttttgg tggcgagcaca gctttgatgt gttttagcct
  351  ggtttttgg tggcgagcaca gctttgatgt gttttagcct
  401  ggtttttgg tggcgagcaca gctttgatgt gttttagcct
  451  ggtttttgg tggcgagcaca gctttgatgt gttttagcct
  501  ggtttttgg tggcgagcaca gctttgatgt gttttagcct
  551  ggtttttgg tggcgagcaca gctttgatgt gttttagcct
  601  ggtttttgg tggcgagcaca gctttgatgt gttttagcct
  651  ggtttttgg tggcgagcaca gctttgatgt gttttagcct
  701  ggtttttgg tggcgagcaca gctttgatgt gttttagcct
  751  ggtttttgg tggcgagcaca gctttgatgt gttttagcct
  801  ggtttttgg tggcgagcaca gctttgatgt gttttagcct
  851  ggtttttgg tggcgagcaca gctttgatgt gttttagcct
  901  ggtttttgg tggcgagcaca gctttgatgt gttttagcct
  951  ggtttttgg tggcgagcaca gctttgatgt gttttagcct
1001  ggtttttgg tggcgagcaca gctttgatgt gttttagcct
1051  ggtttttgg tggcgagcaca gctttgatgt gttttagcct
1101  ggtttttgg tggcgagcaca gctttgatgt gttttagcct
1151  ggtttttgg tggcgagcaca gctttgatgt gttttagcct

M13 reverse primer length 253

   1  tggtagtattta taaacctgccgct ctaaagtttct ttcttctggt tcataaatttt
   51  tggtagtattta taaacctgccgct ctaaagtttct ttcttctggt tcataaatttt
  101  tggtagtattta taaacctgccgct ctaaagtttct ttcttctggt tcataaatttt
  151  tggtagtattta taaacctgccgct ctaaagtttct ttcttctggt tcataaatttt
  201  tggtagtattta taaacctgccgct ctaaagtttct ttcttctggt tcataaatttt
  251  tggtagtattta taaacctgccgct ctaaagtttct ttcttctggt tcataaatttt
  301  tggtagtattta taaacctgccgct ctaaagtttct ttcttctggt tcataaatttt
  351  tggtagtattta taaacctgccgct ctaaagtttct ttcttctggt tcataaatttt
cDNA clone number 10
M13 universal primer, length: 580

1  ccattgttcc ctttttgca aatggtgagc
dna clone number 10
M13 reverse primer, length: 456

1  ccccccccttc gctacacca ctctacgccg ccgcacggcc

cDNA clone number 20
M13 universal primer, length: 268

1  cctgggcac aacctggtc tggagacgca cggccagcgc
cDNA clone number 20
M13 reverse primer, length: 389

1  cccttgttgg atgagggcag aatcactgga atcactgtgt
251 catcctggct aacatgttaa cccgtctct actaaaatat caaaaattagc
tgggcgtgtt gggtggcttg taattcca gcatacctg agactgagggc
301 atgagaactg cctgaamcc aagaaattgga ggttgcart
351