Chapter II Materials and Methods

All solutions used are listed at the end of this chapter.

2.1 DNA manipulation methods

2.1.1 Polymerase Chain Reaction (PCR)

PCR was performed in 0.5 ml microcentrifuge tubes in a DNA Thermal Cycler (Perkin Elmer) or in a 96 well micro titer plate (Costar Thermowell™ C- or M-type) in an Omnigene (Hybaid) (C-type) or a PTC-225 (MJ Research) (M-type). For most applications, 15 µl reactions were prepared.

1. A premix sufficient for the number of planned reactions was prepared, allowing for a 1X reaction mix once the DNA template (section 2.1.1.2) was added (usually 10 µl of mix and 5 µl of template).
2. The final reaction contained 1X buffer (Buffer 1 unless otherwise specified), 200 µM of each of the four nucleotides (Pharmacia), 40 ng of each primer, and 0.5 units /µl of DNA polymerase (*Taq* (Amplitaq) or *Pfu* (Promega)).
3. The amplifications were performed under the same cycling profile (except where specified): 94°C for 5 minutes, followed by 35 cycles at 94°C for 30 seconds, annealing temperature (specific to each primer) for 30 seconds and 72°C for 30 seconds, and finally followed by 1 cycle at 72°C for 5 minutes.
4. Reaction products were visualised by agarose gel electrophoresis and staining with ethidium bromide (section 2.1.2).

DNA templates

The templates used were

1. Bacterial colonies picked into 100 µl of sterile water and 5 µl used directly.
2. cDNA or bacterial pools.

3. DNA excised from an agarose gel into 100 µl sterile water, left overnight and 5 µl used directly.
4. Human (Sigma D-3035) or mouse genomic DNA at 12.5 ng/µl.

2.1.2 Gel electrophoresis

1. An agarose gel was prepared (2.5% for most PCR amplified products and 1% for fragments over 1 kb) in 1X TBE and ethidium bromide (250 ng/µl).
2. DNA was added to the appropriate amount of 6X loading buffer (e.g. 5 µl of PCR product and 1 µl of 6X loading buffer) and loaded. In the case of Buffer 2, the samples were loaded directly.
3. Size markers (1 kb ladder, Gibco-BRL) were also loaded.
4. Minigels were run at 80 Volts for 10-15 minutes and larger gels were run at 200 Volts for the time required to obtain satisfactory separation.
5. DNA was visualised under UV on a transilluminator and photographed with a Polaroid camera.

2.1.3 Restriction enzyme digests

2.1.3.1 Liquid DNA

1. Up to 10 µg of bacterial clone, or plasmid, DNA was used in a reaction containing the appropriate 1X buffer, 1mM spermidine, 100 µg/ml BSA and 20-50 units of the appropriate enzyme.
2. The DNA was digested for 2 hours or overnight at the appropriate temperature for the enzyme.
3. The DNA was subjected to agarose gel electrophoresis and visualised (section 2.1.2).

2.1.3.2 PCR products

1. After PCR amplification, the required amount (usually 5-10 µl) was transferred to a new 0.5 ml microcentrifuge, and 5 units of the restriction enzyme added.
2. DNA was digested for 1 hour at the recommended temperature and visualised by gel electrophoresis.

2.1.4 DNA purification

2.1.4.1 Ethanol precipitation

1. In a 1.5 ml microcentrifuge tube, 0.1 volumes of 3M sodium acetate and either 1 volume of isopropanol or two and half volumes of ethanol were added to the DNA.
2. The samples were mixed well by vortexing and incubated for 20 minutes at –20°C.
3. The DNA was then pelleted in a microcentrifuge at 13,000 g and washed once with 70% ethanol.
4. The pellet was left to dry and then resuspended in the appropriate amount of $T_{0.1}E$.
5. The recovery was tested by gel electrophoresis (section 2.1.2).

2.1.4.2 Gel purification

The DNA fragment was excised from the agarose gel with a clean scalpel.

1. The gel slice was weighed in a 1.0 ml eppendorf tube.
2. The gel slice was then purified using a Qiaquick Gel Extraction Kit™ (Qiagen) according to the manufacturers instructions.
3. The recovery was tested by gel electrophoresis (section 2.1.2).

2.1.4.3 ExoSAP purification of PCR products

1. A premix, sufficient for the number of planned reactions, was prepared, allowing for a 1X reaction mix once the PCR reaction was added (usually a 15 µl PCR reaction volume).

2. The final reaction contained 1X reaction buffer (SAP Buffer), 1X Dilution buffer, 1 unit/µl of Shrimp Alkaline Phosphatase (USB) and 1unit/µl exonuclease I (USB).

3. The mixture was incubated at 37°C for 30 minutes, followed by 80°C for 15 minutes.

2.2 Clone resources

2.2.1 Libraries used

Different types of clone resources have been used throughout this project. The Sanger Institute clone resource group, who also provided the arrayed filters and PCR pools for screening, maintains the clone resources.

2.2.1.1 Bacterial clone libraries

The RPCI-23 female (C57Bl/6J) mouse BAC library (Osoegawa *et al.*, 2000) was screened in this study. Library details are shown in table 2.1.

Table 2.1: Details of the mouse genomic library used

Library	Library type	Library code	Antibiotic	Vector	Cloning site	Genomic digest
RPCI-23	BAC	bM	Chloramphenicol 12.5 µg/ml	pBACe3.6	*Eco*RI	*Eco*RI

2.2.1.2 cDNA libraries

cDNA libraries used during the course of this project are described in table 2.2.

Table 2.2: cDNA resources used during the course of this project

cDNA library code	cDNA library description	Source/reference	Vector	Vectorette PCR
T	Adult testis	CLONTECH	pCDM8	+
FB	Fetal Brain	Invitrogen	pcDNAI	+
FL	Fetal Liver	Invitrogen	pcDNAI	+
FLu	Fetal Lung	Invitrogen	pcDNAI	+
HL60	Peripheral blood	Invitrogen	pcDNAI	+
AH	Adult Heart	Invitrogen	pcDNA3	+
ALu	Adult Lung	CLONTECH	pcDNAI	+
SK-N-MC	Neuroblastoma cells	Invitrogen	pcDNAI	-
PF	Adult brain	Pfizer	pcDNAI	-
U937+*	(Monocyte -NOT activated- from a patient with promonocytic leukaemia)	Simmons (1993)	pCDM8	-
U937AC*T	(Monocyte -PMA activated- from a patient with promonocytic leukaemia)	Simmons (1993)	pCDM8	-
H9*	Placental, full tem normal pregnancy	Simmons (1993)	pH3M	-
YT*	HTLV-1 +ve adult leukaemia T cell	Simmons (1993)	pH3M	-
NK*	Natural killer cell	Simmons (1993)	pH3M	-
Daudi*	B lymphoma	Simmons (1993)	pH3M	-
HPBall*	T cell from a patient with acute lymphocytic leukaemia	Simmons (1993)	pH3M	-
BM*	Bone marrow	Simmons (1993)	pH3M	-
DX3*	Melanoma	Simmons (1993)	pH3M	-

* Generously provided by David Simmons, Oxford (Simmons, 1993).
+ Screened by vectorette PCR. Remaining libraries are available to screen by single sided specificity PCR (Huang *et al*., 1993).

2.2.2 cDNA clone synthesis

2.2.2.1 Nested PCR

1. A premix sufficient for the number of planned reactions was prepared, allowing for a 1X reaction mix once the DNA template (1 µl cDNA from RT-PCR (section 2.5.5) or 1.5 µl of first round reaction mix) was added.
2. The final reaction volume of 25 µl contained 1X *Pfu* DNA polymerase 10X buffer with $MgSO_4$ (Promega) (PCR buffer 2), 200 µM each of the four nucleotides (Amersham Pharmacia Biotech), 40 ng of each primer and 3 units/µl of *Pfu* DNA polymerase (Promega).
3. First round amplification was performed with the external pair of nested primers, under the cycling profile: 95°C for 2 minutes, followed by 35 cycles at 95°C for 45 seconds, annealing temperature (specific to each primer) for 30 seconds and 72°C for 3 minutes, finally followed by 1 cycle at 72°C for 5 minutes.
4. Two volumes of deionised water were added to 1 volume of the reaction mix. 1.5 µl of the diluted reaction mix was used as a template for second round PCR.
5. Steps 1 to 3 were repeated with the internal set of nested primers.
6. Reaction products were visualised by agarose gel electrophoresis and staining with ethidium bromide (section 2.1.2).
7. Selected bands were cut out and the DNA extracted from the gel (section 2.1.4).

The PCR product was ligated into the pGEM®-T Easy vector (Promega) according to the manufacturers' instructions and used to transform competent cells (section 2.2.2.5).

2.2.2.2 Addition of T7 tag

A schematic of the overall strategy is shown in chapter V.

C-terminus

1. PCR (section 2.1.1) utilising the proofreading enzyme *Pfu* (Promega) was used to introduce suitable restriction sites flanking the open reading frame of the cDNA clone. The designed primers (section 2.4.1) also incorporated a Kozak consensus sequence (Kozak, 1987) prior to the start codon and removed the stop codon.
2. The PCR products were then digested (section 2.1.3) and subcloned (section 2.2.2.4) into pBlue-CT7 (a kind gift from Dr Begoña Aguado (HGMP Resource Centre)).
3. The plasmid was then digested using suitable restriction enzymes and the cDNA plus C-terminal T7.Tag was subcloned (section 2.2.2.4) into the expression vector pCDNA3 (Invitrogen) (a kind gift from B Aguado).
4. The plasmid was then miniprepped (section 2.2.2.6) and the insert sequenced (Elizabeth Huckle, Sanger Institute) using appropriate oligonucleotides.

N-terminus

1. PCR (section 2.1.1), utilising the proofreading enzyme *Pfu* (Promega), was used to introduce suitable restriction sites flanking the open reading frame of the cDNA clone. The designed primers (section 2.4.1) retained the stop codon.
2. The PCR products were digested (Section 2.1.3) and subcloned (section 2.2.2.4) into pCDNA3-NT7 (section 2.2.2.3).
5. The plasmid was then miniprepped (section 2.2.2.6) and the insert sequenced (E. Huckle) using appropriate oligonucleotides.

2.2.2.3 Generation of pCDNA-NT7

A new expression vector containing a *Not*I restriction site and preceded by the T7.Tag sequence was created for used in mammalian cell expression systems. A schematic of the vector is shown in chapter V.

1. The vector pBlue-NT7 (a kind gift from B Aguado) was digested with *Bam*HI.
2. The complementary oligonucleotide 5'-GAT CCA GCG GCC GCT G-3' was heated to 65°C for 10 minutes and then snap cooled on ice. The oligonucleotide was then subcloned into pBlue-NT7 (section 2.2.2.4).
3. PCR (section 2.1.1) utilising the proofreading enzyme *Pfu* (Promega) was used to introduce suitable restriction sites (*Hind*III and *Xba*I) flanking the open reading frame of the T7.Tag. The designed primers (section 2.8.4.2) also incorporated a Kozak consensus sequence (Kozak, 1987) prior to the start codon.
4. The PCR product was subcloned (section 2.2.2.4) into pCDNA3 (Invitrogen) (a kind gift from B. Aguado). As *Xba*I is a methyl-sensitive restriction enzyme, the plasmid was transformed into the *E. coli* strain INV110 (Invitrogen), which has disrupted *dam* and *dcm* genes. Methylation of the plasmid *Xba*I site is thus prevented.
5. The plasmid was then miniprepped (section 2.2.2.F) and the insert sequenced (E. Huckle) using appropriate oligonucleotides to confirm its integrity.

2.2.2.4 Subcloning

1. The vector and insert to be used were digested with the appropriate restriction enzymes (section 2.1.3).
2. The restriction products were gel purified (section 2.1.4). Concentration of the vector and insert was estimated from appearance on the gel against the size markers.
3. An approximate 3:1 molar ratio of the insert and vector (roughly 150 ng of insert: 50 ng of vector) was ligated together in a final reaction volume of 10 µl, including 1 µl of 10X Ligation buffer, and 1 unit of T4 DNA ligase (Roche).
4. The ligation reaction was incubated at 4°C overnight.

2.2.2.5 Transformation

1. 2 µl of the ligation reaction was added to a sterile 1.5 ml microcentrifuge tube on ice.
2. Tube(s) of frozen JM109 High Efficiency Competent Cells (Promega) were removed from –70°C storage and thawed on ice.
3. 50 µl of cells were carefully transferred into each tube and gently flicked to mix. The tubes were incubated on ice for 20 minutes.
4. The cells were heat-shocked for 45 seconds in a water bath at 42°C. The tubes were then immediately returned to ice for 2 minutes.
5. 950 µl of LB broth was added to the tubes, which were then incubated for 1.5 hours at 37°C, 150 rpm.
6. 100 µl of each transformation reaction was plated onto duplicate LB/ 100 µg/ml ampicillin/ 0.5 mM IPTG/ 80 µg/ml X-Gal plates.
7. The plates were incubated overnight at 37°C.
8. White colonies were picked into 100 µl sterile water and used as a PCR template in order confirm identity of the insert using appropriate primers (section 2.1.1).

2.2.2.6 Bacterial clone minipreps

1. A single colony was inoculated into 10 ml of LB broth containing the appropriate antibiotic and grown overnight at 37°C, shaking at 250 rpm.
2. The cells were pelleted at 1500 g and the media fully drained. It was important to ensure that all the media was removed at this stage to prevent inhibition of digestion.
3. The pellet was resuspended in 200 µl of GTE on ice.
4. On ice, 400 µl of fresh 0.2M NaOH and 1% SDS was added with gentle inversion of the tube. The tube was left on ice for 5 minutes.
5. 300 µl of 5M Acetate, 3M K^+ was added. The tube was gently inverted to mix and then incubated on ice for 10 minutes.

6. The precipitate was pelleted in a microcentrifuge at 13,000 g.
7. The supernatant was transferred to a new tube. If the supernatant was still cloudy, step 6 was repeated.
8. 600 µl of cold isopropanol was added to the cleared supernatant.
9. The DNA was pelleted in a microcentrifuge at 13,000g.
10. The DNA pellet was resuspended in 200 µl TE and extracted with 200 µl phenol/isoamyl-alcohol/chloroform (25:1:24).
11. The DNA was ethanol precipitated (section 2.1.4)
12. The DNA was pelleted in a microcentrifuge and washed with 70% ethanol.
13. The DNA was resuspended in 30 µl $T_{0.1}E$ and stored at –20°C.

2.2.2.6 Bacterial clone micropreps

1. A single colony was inoculated into 500 µl of LB broth containing the appropriate antibiotic and grown overnight at 37°C shaking at 300 rpm.
2. 250 µl of culture was aliquoted into a 96 well round-bottom plate (Costar).
3. The cells were pelleted at 2500g for 4 minutes. The plate was inverted to drain the supernatant.
4. The pellet was resuspended in 25 µl.
5. On ice, 25 µl of fresh 0.2M NaOH and 1% SDS was added and mixed by gently tapping. The plate was left on ice for 5 minutes.
6. 25 µl of 5M Acetate, 3M K^+ was added and mixed by tapping gently. The plate was then incubated on ice for 10 minutes.
7. The well contents were transferred to a 96 well filter-bottom plate.
8. The filter plate was taped on top of a 96 well round-bottom plate containing 100 µl of isopropanol.

9. The precipitate was pelleted at 2500 rpm for 2 minutes and the filter plate discarded. The round-bottom plate was incubated at room temperature for 30 minutes and then spun at 3200 rpm for 20 minutes at room temperature.
10. Supernatant was removed by inverting the plate.
11. 100 µl of 70% ethanol was added to each well and the plate tapped gently. The plate was then spun at 3200 rpm, for 20 minutes at room temperature.
12. The supernatant was removed by inversion and the pellet dried at room temperature.
13. The DNA was resuspended in 5 µl of $T_{0.1}E$ with RNase (10 µl of 1 mg/ml RNase per 1 ml of $T_{0.1}E$).

2.2.3 Vectorette Library Synthesis

2.2.3.1 Library titration

1. The cDNA library, consisting of bacteria stored in glycerol, was defrosted on ice. 2 µl of the library was diluted in 198 µl LB. Six tenfold serial dilutions were prepared and 100 µl of each plated onto LB agar plates containing the appropriate antibiotic.
2. The plates were left inverted at 37°C for 4 hours. They were then transferred to 30°C for 16 hours and then back to 37°C for an additional 4 hours to promote growth of separated colonies.
3. The colonies were counted and the library titre estimated.

2.2.3.2 High density liquid pools

1. Ten tubes of 20 ml LB, with the appropriate antibiotic, were prepared.
2. 250,000 clones, diluted in LB, were added to each tube (2,500,000 clones in total).
3. The clones were grown at 37°C for 20 hours at 240 rpm.

4. Meanwhile, dilutions of 1000, 100, 10 and 1 clone(s) were plated onto LB agar plates containing the appropriate antibiotic to check the titration. The plates were inverted and left to grow overnight at 37°C.

2.2.3.3 Low density plated pools

1. 25 X 20,000 clones were plates onto LB agar plates with Hybond N+ filters (500,000 clones in total).
2. The plates were then inverted and the clones left to grow for 4 hours at 37°C, followed by 16 hours at 30°C and a further 4 hours at 37°C.

2.2.3.4 DNA extraction

1. Filters were rolled up and put in a 50 ml falcon tube with 20 ml of SET. The cells were shaken off and the filters removed.
2. DNA was extracted using the bacterial clone miniprep protocol (section 2.2.2.6).
3. 1 µl of RNase (10 ng/ml) was added to the extracted DNA and incubated at 37°C for 1 hour.
4. 0.01 µl of the DNA was visualised by gel electrophoresis to check the extraction outcome.

2.2.3.5 Library preparation

1. 1 mg of DNA was digested with the appropriate enzyme in 30 µl (section 2.1.3).
2. 70 µl of water was added and the DNA extracted with 100 µl phenol/isoamyl-alcohol/chloroform (25:1:24).
3. The DNA was ethanol precipitated (section 2.1.4).
4. The DNA was pelleted in a microcentrifuge and washed with 70% ethanol.
5. The DNA pellet was resuspended with 100 µl of ligation buffer.

6. 10 µl of 1pmol/µl annealed vectorette vectorette bubbles appropriate to the enzyme used, 1.1 µl adenosine 5'-triphosphate and 2.5 units of T4 DNA Ligase (Amersham Pharmacia Biotech) were added.
7. The tubes were incubated at 16°C overnight.
8. The mixes were diluted to 500 µl to generate Stock Pools
9. Equal volumes of sets of five plates Stock Pools were mixed to generate stocks of Super Pools.
10. 1/100 dilutions of stock Super Pools were prepared using $T_{0.1}E$. These pools were screened in the first round of PCR (section 2.3.3).
11. 1/100 dilutions of the plated Stock Pools were prepared using $T_{0.1}E$. These pools were screened in the second round of PCR (section 2.3.3).
12. 1/10 dilutions of the plated and liquid Stock Pools were prepared using $T_{0.1}E$. These pools were used for both PCR pool screening and vectorette PCR (section 2.3.3).

2.3 Screening

2.3.1 Probe labelling

DNA probes were labelled, either by PCR, or by random hexamer labelling (Feinberg & Vogelstein, 1983; Hodgson & Fisk, 1987).

2.3.1.1 PCR labelling of STSs

1. The required fragment was amplified from either genomic DNA or cDNA as appropriate.
2. The fragments were separated on a 2.5% gel (section 2.1.2), cut out and transferred to a microcentrifuge tube containing 100 µl of sterile deionised water. The DNA was allowed to diffuse out of the gel slice (at least one hour).

3. Using PCR buffer 1, 9 μl reactions were set up containing the required primers, 2 μl of the liquid surrounding the gel slice, nucleotides (except dCTP) and DNA polymerase. A single drop of mineral oil was placed on top of the reaction mixture.
4. 1 μl of [α-^{32}P] dCTP (3000 Ci/mol, Amersham Pharmacia Biotech) was added.
5. PCR was performed in a DNA Thermal Cycler (Perkin Elmer) under the following cycling profile, 94°C for 5 minutes, 20 cycles of 93°C for 30 seconds, 55°C for 30 seconds, 72°C for 30 seconds and 1 cycle of 72°C for 5 minutes.
6. After completion, the probes were either denatured in the thermal cycler at 99°C for 5 minutes and then snap chilled, or were added to competitor DNA (section 2.3.1.3) and denatured in a boiling water bath for 5 minutes, before being snap chilled.
7. The probes were then added to the hybridisation mix.

2.3.1.2 Random hexamer labelling for probes

1. 20 ng of DNA was added to a microcentrifuge tube. The volume was made up to 17.5 μl with sterile water.
2. The DNA was denatured in a boiling waterbath for 5 minute, then snap chilled on ice.
3. 5 μl of OLB3, 1 μl of 10 mg/ml BSA, 0.5 units of Klenow DNA polymerase (GibcoBRL) and 0.5 μl of [α-^{32}P] dCTP (3000 Ci/mol, Amersham Pharmacia Biotech) were added to the tube.
4. The reactions were incubated at room temperature for 3 hours and then denatured in a boiling bath for 5 minutes.

2.3.1.3 Competitive reassociation of radio labelled probe

Many of the probes designed from mouse BAC end sequences were of relatively uncharacterised genomic content and were likely to contain repetitive DNA. This was

suppressed with mouse genomic DNA (Sealey *et al.*, 1985). Potential microsatellite sequences were also competed with poly CA/GT (Pharmacia 27-7940).

1. 125 µl of mouse DNA (10 mg/ml, a kind gift from George Stavrides, Sanger Institute), 125 µl of 20X SSC, 5 µl of poly CA/GT (1 mg/ml), was added to a screw cap microfuge tube and the total volume of the mixture made up to 0.5 ml of water minus the volume of the labelling reaction.
2. The labelled DNA was added to the competition mixture and the tube boiled in a water bath for 5 minutes.
3. The tube was snap chilled on ice and the probe was then added to the hybridisation buffer.

2.3.2 Library screening

2.3.2.1 Screening pools by PCR

Using PCR buffer 1, sufficient reaction mix was set up, containing the required primers, nucleotides and 5µl of DNA from the pools making a total of 15 µl, to screen all the pools for the library.

1. PCR amplification was performed under the cycling profile of 94°C for 5 minutes, 35 cycles of 93°C for 30 seconds, annealing temperature specific for each set of primers for 30 seconds, 72°C for 30 seconds followed by 1 cycle of 72°C for 5 minutes.
2. The reaction products were visualised by gel electrophoresis (section 2.1.2).

2.3.2.2 Screening of library filters by hybridisation of PCR-labelled probes

Gridded filters of both the mouse RPCI-23 BAC library and the region-specific subset of bacterial clones were generated by the Sanger Institute clone resource group. These were then screened.

1. STSs were labelled as described (section 2.3.1).

2. Up to 30 filters were sequentially placed in a 15x10x5 cm sandwich box with sufficient hybridisation buffer to cover the filters. A plastic sheet, cut to size, was placed on top to reduce evaporation. The filters were pre-hybridised at 65°C for 2 hours (Innova 4000, New Brunswick Scientific).
3. The filters were removed and the denatured probe was added to hybridisation solution in the box and mixed.
4. The filters were added one by one back into the box and each was carefully submerged under the hybridisation mix. The plastic sheet was replaced on top.
5. After hybridisation overnight at 65°C, the filters were washed by rinsing twice in 2X SSC at room temperature for 5 minutes. The filters were then washed twice in 0.5X SSC and 1% N-lauroyl-sarcosine at 65°C for 30 minutes, before rinsing twice in 0.2X SSC at room temperature for 5 minutes.
6. The washed filters were wrapped in Saran wrap (Dow Chemical Co.) and exposed to pre-flashed Fuji Medical X-ray film (036010) (or equivalent) overnight with two intensifying screens at –70°C in the first instance. This was repeated if longer or shorter exposures were needed.
7. Occasionally filters were re-washed to 0.2X SSC with 1% N-lauroyl-sarcosine at 65°C for 30 minutes if required (i.e. high background).
8. The autoradiographs were developed and labelled with the name of the filter and the data entered into 22ace (section 2.7.1).

2.3.3 Vectorette PCR

This method was adapted from the original (Riley *et al.*, 1990) to isolate cDNA fragments from a cDNA library. The method was developed by Dr. John Collins (Sanger Institute) and libraries were made by J. Collins, G. Stavrides and M. Goward (section 2.2.3).

2.3.3.1 Identification of positive pools

1. For each library, 5 super pools were screened with the appropriate STS.
2. For each positive super pool, the 5 individual constituent pools were screened with the appropriate STS.

2.3.3.2 First round PCR

1. Using PCR buffer 3, 15 µl reactions were set up using 5 µl of a 10X concentrated PCR pool.
2. The reaction was transferred to the Omnigene (Hybaid) and incubated for 1 cycle at 94°C for 5 minutes.
3. After 4 minutes, the program was paused and 1 µl of enzyme mix (0.12 µl *Taq* (Amplitaq), 0.12 µl *Taq*Extender (Promega), 0.12 µl PerfectMatch (Promega) and 0.64 µl sterile water) was added. The cycling was then allowed to continue, for 1 minute at 94°C, 17 cycles of 94°C for 5 minutes, 68°C for 30 seconds and 72°C for 3 minutes, followed by 18 cycles of 94°C for 5 cycles, 60°C for 30 seconds and 72°C for 5 minutes and finally 72°C for 5 minutes.
4. The fragments were then visualised by gel electrophoresis on a 1% agarose gel (section 2.1.2) and the fragments were cut out and stored overnight in 100 µl of sterile water at 4°C.

2.3.3.3 Second round PCR

1. To obtain DNA for sequencing the liquid from around the gel slice was reamplified using standard PCR conditions. Four 15 µl reactions were used to obtain sufficient DNA for sequencing.

2. The amplification was performed under the following cycling conditions, 1 cycle of 95°C for 5 seconds followed by 20 cycles of 94°C for 5 seconds, 60°C for 30 seconds, 72°C for 3 minutes and finally 1 cycle of 72°C for 5 minutes.
3. The pooled reactions were separated by gel electrophoresis on a 1% agarose gel (section 2.1.2). The fragments were cut out and purified using the Qiaquick Gel Extraction Kit™ (Qiagen) according to the manufacturers instructions.
4. The templates and oligonucleotides were used in cycle sequencing (E. Huckle).

2.4 Landmark production

2.4.1 Primer design

The primers were designed using the Primer3 program (Rozen and Skaletsky, 1998; http://www.genome.wi.mit.edu/genome_software/other/primer3.html) from http://www.sanger.ac.uk/cgi-bin/primer3.cgi. Additional primers for amplification and subsequent cloning of full-length cDNA (section 2.2.2) were designed manually.

2.4.2 Primer synthesis

1. Primers were synthesised at the Sanger Institute by David Frazer. A subset of the primers were synthesised by GenSet. Primer concentrations were supplied in both cases.
2. Primers were stored at –20°C and working dilutions for PCR prepared at 100ng/µl for each primer in pairs.
3. The primers were tested at three different annealing temperatures, 55°C, 65°C and 65°C, using the standard cycling on Thermal cyclers to establish optimal PCR conditions.

2.4.3 Fingerprinting

*Hind*III fingerprinting of bacterial clones was performed with the help of Owen McCann (Sanger Institute), using the standard protocol below (Marra *et al.*, 1997).

2.4.3.1. Digestion

1. Bacterial clones were microprepped (section 2.2.2.6) by Carol Carder (Sanger Institute) or M. Goward.
2. 2.6 µl of water, 0.9 µl of the appropriate buffer and 20 units of *Hind*III (Boerhinger) were added to each well, mixed by gentle tapping and then the plate centrifuged up to 1000 g to collect the contents.
3. The plate was incubated at 37°C for 2 hours.
4. The reaction was terminated by addition of 2 µl of 6X Dye Buffer II and the plate centrifuged up to 1000 g to collect the contents.

2.4.3.2 Gel preparation and loading

1. A 1% gel mix was prepared using 450 ml of 1X TAE and 4.5g agarose and poured at 4°C. A 121-well comb was placed in the gel and allowed to set for 45 minutes. The comb was then removed.
2. 3-4 l of 1X TAE was added to the gel tank.
3. 0.8 µl of the marker (Promega, DG1931) was loaded in the first well and then in every fifth well.
4. 1.0 µl of each sample was then loaded into the empty wells.
5. The gel was run at 90 V for 30 minutes at room temperature. Once the dye front had advanced beyond the wells, the gel tanks were transferred to a refrigerated room and run at 4°C for 15 hours at 90 V.

2.4.3.3 Gel staining

1. The gel was trimmed to ~19 cm and stained with vistra green stain for 45 minutes. The gel was covered whilst staining to prevent degradation of the vistra green.

2. The gel was then rinsed with 0.5 l of deionised water. The gel was visualised and the image recorded using a Molecular Dynamics scanner.

2.4.4 SNP verification

1. Candidate SNPs were identified by comparison of cDNA sequence to the genomic DNA. Primers were designed flanking approximately 400 bp of sequence surrounding the candidate variant.
2. Fragments containing the SNP of interest were amplified from the DNA of 24 individuals (Set M24PDR of 24 human DNAs from the Coriell cell repository).
3. Amplification was tested by electrophoresis (section 2.1.2) of 5 µl of the product.
4. The PCR products were purified using the ExoSAP protocol (section 2.1.4).
5. The recovery of the purification was tested by electrophoresis (section 2.1.2).
6. The fragments and correct primers were used in cycle sequencing (E. Huckle).
7. The resultant sequences were aligned in a Gap4 database (Dr. Kate Rice, Sanger Institute).

2.5 RNA manipulation

2.5.1 Steps taken to limit contamination with RNase

Autoclaved plasticware (tubes, pipette tips etc.) was used and bench surfaces, racks etc. were cleaned before use with RNase*Zap*® (Ambion).

All reagents for RNA work were made up with Diethylene Pyrocarbonate (DEPC) water.

1. 0.1% DEPC in deionised water was mixed, and left overnight in a fume hood.
2. The DEPC water was autoclaved before use.

2.5.2 RNA resources

RNA used in this project was obtained from a number of sources.

Table 2.3: RNA resources used during the course of this project

	Tissue	Source	Supplied as		Tissue	Source	Supplied as
A	Heart	Clontech	Human MTN™ Blot (7760-1)	11	Spleen	Stratagene	Total RNA
B	Brain (whole)	Clontech	Human MTN™ Blot (7760-1)	12	Stomach	Stratagene	Total RNA
C	Placenta	Clontech	Human MTN™ Blot (7760-1)	13	Colon I	Stratagene	Total RNA
D	Lung	Clontech	Human MTN™ Blot (7760-1)	14	Colon II	*	Tissue
E	Liver	Clontech	Human MTN™ Blot (7760-1)	15	Rectum	Stratagene	Total RNA
F	Skeletal muscle	Clontech	Human MTN™ Blot (7760-1)	16	Breast	Stratagene	Total RNA
G	Kidney	Clontech	Human MTN™ Blot (7760-1)	17	Ovary	*	Tissue
H	Pancreas	Clontech	Human MTN™ Blot (7760-1)	18	Uterus	Stratagene	Total RNA
I	Fetal brain	Clontech	Human MTN™ Blot (7756-1)	19	Cervix I	Stratagene	Total RNA
J	Fetal lung	Clontech	Human MTN™ Blot (7756-1)	20	Cervix II	*	Tissue
K	Fetal liver	Clontech	Human MTN™ Blot (7756-1)	21	Testis I	Clontech	Total RNA
L	Fetal kidney	Clontech	Human MTN™ Blot (7756-1)	22	Testis II	Invitrogen	Total RNA
1	Kidney I	Invitrogen	Total mRNA	23	Fetal brain I	Stratagene	Total RNA
2	Kidney II	Clontech	Total mRNA	24	Fetal brain II	Clontech	Total RNA
3	Liver I	Stratagene	Total mRNA	25	Fetal heart I	Stratagene	Total RNA
4	Liver II	*	Tissue	26	Fetal heart II	Stratagene	Total RNA
5	Cerebrum	*	Tissue	27	Fetal liver I	Stratagene	Total RNA
6	Skeletal muscle	*	Tissue	28	Fetal liver II	Stratagene	Total RNA
7	Skin	*	Tissue	29	Fetal lung I	Stratagene	Total RNA
8	Tonsil	*	Tissue	30	Fetal lung II	Stratagene	Total RNA
9	Lymphoblast cell line	#	Harvested cells	31	Fetal spleen	Stratagene	Total RNA
10	Thyroid	Stratagene	Total RNA	32	Fetal bladder	Stratagene	Total RNA

* Supplied as tissue, from Tissue Bank, Department of Histopathology, Addenbrookes Hospital, Cambridge.
# Cells supplied as a kind gift from Dr Nigel Carter, Sanger Institute.

2.5.3 RNA isolation

Total RNA was isolated from human tissue samples and cell lines after homogenisation in TRIzol reagent (Chomczynski & Sacchi, 1987).

1. The sample was homogenised in 1 ml of TRIzol reagent per 50-100 mg of tissue, or 10^7 cells.

2. The homogenised sample was incubated at room temperature for 5 minutes.
3. 0.2 ml of chloroform per 1 ml of TRIzol reagent was then added. The tube was shaken vigorously for 15 seconds and incubated at room temperature for 2-3 minutes.
4. The tube was centrifuged at no more than 12000 g for 15 minutes at 4°C.
5. The aqueous upper phase was transferred to a new tube and 0.5 ml of isopropanol per 1 ml of TRIzol reagent used was added.
6. The tube was incubated at room temperature for 10 minutes and then centrifuged at no more than 12000 g for 15 minutes at 4°C.
7. The supernatant was removed and the pellet washed once with 75% ethanol, adding at least 1 ml per 1ml of TRIzol reagent used.
8. The tube was centrifuged at 7500 g for 5 minutes at 4°C.
9. The pellet was dried at room temperature and resuspended in 100 µl of DEPC water. The sample was heated to 55°C for 10 minutes, then stored in 75% ethanol at –70°C.

2.5.4 Ethanol precipitation

1. 0.025 volumes of 3M sodium acetate were added to 1 volume of RNA in 75% ethanol.
2. The samples were mixed well by vortexing and incubated for 20 minutes at –70°C.
3. The RNA was then pelleted in a microcentrifuge at 4°C at 13,000g and washed once with 70% ethanol.
4. The pellet was left to dry at room temperature.

2.5.5 Reverse Transcription PCR (RT-PCR)

2.5.5.1 Preparation of RNA sample prior to RT-PCR

1. 10 µg of RNA was ethanol precipitated (section 2.5.3).
2. The RNA pellet was resuspended in 79 µl of DEPC water.

3. 10 µl of DNase I buffer (GibcoBRL), 1 µl DNase I (GibcoBRL) and 1 µl RNAguard (Amersham Pharmacia Biotech) was added to the tube and incubated at room temperature for 15 minutes.
4. 10 µl of 25 mM EDTA was added to stop the reaction. The tube was then incubated at 65°C for 10 minutes.
5. The tube was briefly chilled on ice and the RNA ethanol precipitated (section 2.5.3).

2.5.5.2 First strand cDNA synthesis

1. To a 10 µg RNA pellet, 2 µl of oligo (dT) (500 ML/ml) was added and the solution was made up to 24 µl with DEPC water.
2. The mixture was heated to 70°C for 10 minutes and then chilled briefly on ice. The contents of the tube were collected by brief centrifugation.
3. 8 µl of First Strand buffer (GibcoBRL), 4 µl of DTT (0.1M) (GibcoBRL), and 2 µl of dNTP mix (10 mM) were added. The tube contents were mixed gently and incubated at 42°C for 2 minutes.
4. 400 units of reverse transcriptase (SuperScript II, GibcoBRL) was added to the reaction and mixed by gentle pipetting.
5. The reaction was incubated at 42°C for 50 minutes.
6. The reaction was then inactivated by heating at 70°C for 15 minutes.
7. RNA complementary to the DNA was removed by addition of 2 units of E. coli RNase H (Promega) and incubating at 37°C for 20 minutes.
8. The cDNA was then used as a template in PCR (section 2.1.1).

2.5.6 Northern blotting

2.5.6.1 Probe generation

1. Probes were generated by PCR from cDNA templates (table 2.2), using primers designed to flank intronic sequence where possible.
2. The STS was radiolabelled by PCR (section 2.3.1).
3. β-actin control probes (Clontech) were radiolabelled by random hexamer labelling (section 2.3.1).

2.5.6.2 Hybridisation

The human Multiple Tissue Northern (MTN) blots (Nos. 7760-1 and 7756-1; Clontech) contain 2 µg of poly (A) mRNAs from different adult and fetal human tissues.

1. The blots were pre-hybridised for 1 hour and then hybridised for 18 hours at 65°C in hybridisation buffer.
2. The blots were washed twice in 2X SSC, 0.05% SDS for 10 minutes at room temperature, then twice in 0.1X SSC, 0.1% SDS for 10 minutes at 50°C.
3. The blots were then subjected to autoradiography at –70°C for an average of 3 days.

2.6 Cell Culture and Protein Manipulation

2.6.1 SDS-PAGE

SDS-PAGE was carried out using a Mini-PROTEAN®Electrophoresis cell (Biorad).

2.6.1.1 Gel preparation

1. The gel unit was assembled according to the manufacturer's instructions
2. A separating gel mix was prepared (12% separating gel was prepared for SDS treated proteins in the approximate molecular weight range of 10-100 k Daltons; lower or higher percentage gels were prepared as required) from a 30% acrylamide/bis stock

(Severn Biotech), containing 0.375M Tris-HCl (pH 8.8), 0.1% SDS, 0.05% ammonium persulfate, 0.05% TEMED and deionised water. The TEMED and ammonium persulfate were added last.

3. The separating gel was poured. A 2mm layer of distilled water was added to the top of the gel. The gel was then allowed to polymerise for 10 minutes.
4. The distilled water was poured off.
5. A 4% stacking gel mix was prepared from a 30% acrylamide/bis stock (Severn Biotech), containing 0.125M Tris-HCl (pH6.8), 0.1% SDS, 0.05% ammonium persulfate, 0.1% TEMED and deionised water. The TEMED and ammonium persulfate were added last. The gel comb(s) were inserted and the stacking gel poured on top of the separating gel.
6. The gel was allowed to polymerise for 30 minutes.

2.6.1.2 Running the gel

1. Cultured cells were harvested in 1X protein sample buffer and boiled at 95°C for 5 minutes, then loaded.
2. Size markers (Benchmark prestained protein ladder, GibcoBRL) were also loaded.
3. Gels were run at 200 Volts for approximately 45 minutes.

2.6.1.3 Electrophoretic transfer

Proteins were transferred to a nitrocellulose membrane using the Mini Trans-Blot® Electrophoretic Transfer cell (Biorad).

1. Nitrocellulose membrane (Hybond ECL, Amersham Pharmacia Biotech) and two pieces of Whatman 3MM were cut to size and soaked in transfer buffer.
2. The gel was equilibrated in transfer buffer.

3. The nitrocellulose membrane was placed on top of the gel. The two were then sandwiched between the Whatman papers. A glass tube was used to remove air bubbles. The sandwich was placed between fibre pads into the electrophoretic transfer cell (Biorad).
4. Electrophoretic transfer was run at 100V for 1 ½ hours.

2.6.2 Western blotting

1. Proteins were transferred to a nitrocellulose membrane (Hybond ECL, Amersham Pharmacia Biotech) (section 2.6.2.3) and blocked in 10% milk powder/0.1% Tween-20/phosphate buffered saline (PBS).
2. The blot was incubated with a mouse anti-T7 monoclonal antibody (stock at ~1 mg/ml) (Novagen #69522-4), at a dilution of 1/2500 in 10% milk powder/0.1% Tween-20/PBS.
3. The blot was washed three times for 10 minutes in 0.1% Tween/PBS at room temperature.
4. The secondary antibody, a sheep-anti-mouse-IgG HRP-conjugate (stock at ~0.32 mg/ml)(Sigma #A67782) was used at a dilution of 1/7500 in 10% milk powder/0.1% Tween-20/PBS.
5. The blot was washed three times for 10 minutes in 0.1% Tween/PBS at room temperature.
6. The signal was detected using ECL (NEN) according to the manufacturer's instructions and visualised by autoradiography.

2.6.3 Cell culture and transfection

1. COS-7 cells (SV40 transformed African Green monkey kidney) were grown in Dulbecco's modified Eagle's medium (DMEM) with 10% foetal bovine serum (FBS) and 100 µg/ml penicillin and 100 µg/ml streptomycin at 37°C in 5% CO_2.

2. Cells were seeded in a 24 well plate at ~40,000 cells/cm^2 in 1 ml of DMEM with FBS.
3. After 24 hours, the cells were then transfected with 0.6 μg DNA using the standard DEAE-dextran protocol (Seed & Aruffo, 1987).
4. Cells were incubated with the DNA for 3 hours, then shocked with 10% DMSO for 2 minutes.
5. The cells were then washed twice with PBS.
6. 600 μl of DMEM with 10% FBS was added and the cells were incubated at 37°C.
7. The cells were harvested after two and three days in 1X protein sample buffer.
8. 25 μl was loaded on a 12% polyacrylamide gel and the proteins were separated by SDS-PAGE (section 2.6.1), before Western blotting (section 2.6.2.4).

2.6.4 Immunofluorescence

1. COS-7 cells were seeded at ~20,000 cells/cm^2 onto coverslips in a 6 well plate.
2. Cells were transfected with 2 μg DNA using the DEAE-dextran method (Seed & Aruffo, 1987).
3. Cells were incubated with the DNA for 3 hours, then shocked with 10% DMSO for 2 minutes.
4. The cells were then washed twice with PBS.
5. 3 ml DMEM/10% FBS was added and the cells were incubated at 37°C for three days.
6. Cells were washed in 250mM Hepes (pH 7.4), and then fixed in 4% paraformaldehyde in 250 mM Hepes (pH 7.4). The reaction was quenched in 50 mM NH_4Cl.
7. The cells were permeabilised with 0.05% w/v saponin/0.2% gelatine in PBS.
8. The cells were then stained with mouse anti-T7. Tag monoclonal antibody (Novagen), at a dilution of 1/100 in 0.05% saponin/0.2% gelatine/PBS at room temperature.
9. The cells were rinsed twice in 0.05% saponin/PBS and then washed three times for 10 minutes in 0.05% saponin/PBS.

10. The secondary antibody, a goat anti-mouse-IgG FITC-conjugate (stock at ~1.1 mg/ml) (Sigma #F2012) was used at a dilution of 1/100 in 0.05% saponin/0.2% gelatine/PBS at room temperature.
11. The cells were rinsed twice in 0.05% saponin/PBS and then washed twice for 10 minutes in 0.05% saponin/PBS. The cells were finally washed twice for 10 minutes in PBS.
12. Coverslips were mounted onto slides with vectashield (Vector Laboratories) and visualised using a confocal microscope (Nikon 800, using the Microradiance confocal system (BIORAD) and Lasersharp image analysis software (BIORAD)).

2.7 Computational analysis

Details of most of the programs and scripts used can be found on the Sanger Institute WWW pages (http://www.sanger.ac.uk/Software). Some of the main programs and computational protocols are discussed below.

2.7.1 ACeDB

The data produced as part of this project was entered into the lab database 22ace (Dunham *et al.*, 1994), the chromosome 22 implementation of ACeDB (Durbin and Thierry-Mieg, 1991). The data entry can either be done by obtaining write access and editing the database in real time, or by importing files prepared previously in a format readable by the database. The database is used for a variety of data, such as sequence, gene annotations and library screen results. The navigation through the database is by clickable links similar to hypertext links, which bring up new windows. There are different graphical representations for sequence data, genetic map data and peptide data. Exampled are shown throughout this thesis (for more detail see http://www.acedb.org).

2.7.2 Sequence analysis

Finished clone sequences were subjected to the standard Sanger Institute computational analysis (Dunham *et al.*, 1999). In brief, the sequences were analysed for repeats and the repeats masked using RepeatMasker (Smit and Green, unpublished). The masked sequences were used in similarity searches against the public domain DNA and protein databases using the BLAST suite of programs. A variety of exon and gene prediction programs, including Genscan (Burge & Karlin, 1997) was used to predict possible gene structures. The unmasked sequence was used in GC content analysis and prediction of CpG islands, tandem repeats, tRNA genes (Fichant & Burks, 1991) and exons. The completed sequences were visualised in the DNA map display in 22ace (section 2.7.1). For more detail, see http://www.sanger.ac.uk/HGP/Humana/human_analysis.shtml.

2.7.3 Gene annotation

1. To align cDNA or EST sequences with genomic DNA, the obtained cDNA sequence or assembled ESTs together with the genomic region to which the gene localised were used with the est2genome program (Mott, 1997).
2. The output file from est2genome was converted to ACeDB format using estg2ace (Dunham, unpublished).
3. The resultant file was imported into the 22ace database.

2.7.4 BLAST

In addition to the above described similarity searches, which were performed as part of the automatic analysis, additional BLAST (Altschul *et al.*, 1997) searches were performed using available websites (section 2.8.5).

2.7.5 Perl scripts

It is often more efficient to analyse large amounts of data using scripts. Perl is a computer language widely used by the bioinformatics community for data management, data format conversion and cgi (common gateway interface) scripts for web forms (Stein & Thierry-Mieg, 1998). Perl scripts were kindly provided by Dave Beare (Sanger Institute) and Dr. Ewan Birney (European Bioinformatics Institute) to aid parts of the analysis described in this project. Additional scripts by Dr. Ian Dunham (Sanger Institute) and Dr. Luc Smink (CIMR, Cambridge) have also been utilised.

Table 2.4: Perl scripts used during the course of this project.

Script	Function	Author
MethComp	Uses GFF file to compare specificity/sensitivity of 'methods' for gene identification/annotation. Compares against the reference set of exons as defined in the keyset of genes (structures) and the GFF	D. Beare
gff2ps	Parses gff format to postscript format	E. Birney
estg2ace	Parses est2genome output to ace format	I. Dunham
e-profile	Classifies results of BLASTn searches of dbEST into tissue origin.	L. Smink and D. Beare
MatchReport	Submits BLAST jobs to multiple databases and processes the output into a variety of formats.	L. Smink, D. Beare and I. Dunham

2.7.6 Calculations of specificity and sensitivity of sequence data

2.7.6.1. Background

The correlation of different types of sequence evidence with the annotated set of genes from 22q13.31 was measured by comparing the alignments of sequence evidence (potential coding value) against the annotated gene features along the test sequence. Analysis of the alignment can take place at the nucleotide, exon and/or gene level as appropriate (see section 2.7.6.3). This has been one of the most widely used approaches in evaluating the accuracy of coding

region identification and gene structure prediction methods. A brief explanation is provided here (for further details see Burset & Guigo, 1996).

A 2x2 contingency table can be used to represent the relationship between the true and putative coding nucleotides on a test sequence (figure 2.1).

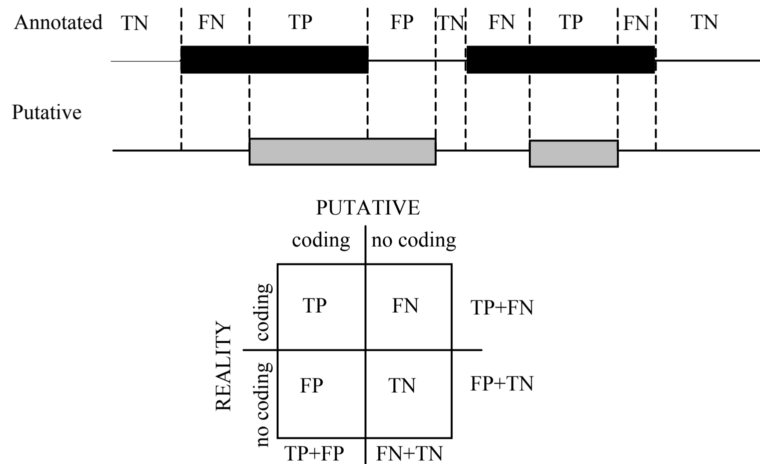

Figure 2.1: Measures of sequence correlation with annotated gene structures

The left upper cell of the table contains the number of coding nucleotides that also aligned with putative coding nucleotides (the true positives, TP), while the right lower cell contains the number of noncoding nucleotides that do not align with putative coding nucleotides (the true negatives, TN). The other two cells register the number of nucleotides in which the annotation and sequence evidence disagree: the number of coding nucleotides that do not align with putative coding nucleotides (the false negatives, FN) and the number of noncoding nucleotides which align with putatively coding sequence (the false positives, FP). Measures of sensitivity (Sn) and specificity (Sp) can be derived from this table and are usually defined as:

$$Sn = \frac{TP}{TP+FN} \qquad Sp = \frac{TP}{TP+FP}$$

Sn is the proportion of coding nucleotides that correctly align with putatively coding nucleotides and Sp the proportion of putative coding nucleotides that are actually coding.

2.7.6.2 MethComp

The correlation of different types of sequence evidence with the annotated sets of genes and pseudogenes from 22q13.31 was calculated using the perl script MethComp (Dave Beare, unpublished). This program analyses the nucleotide alignments of different sequence evidence against annotated exons from genome feature format (gff) files dumped from the chromosome 22 ACeDB database. The alignment of a putative coding feature within the test sequence is described as a 'hit'. In all cases, multiple overlapping hits are counted as one. The following calculations are performed:

Total coverage = No. of base hits/test region size.
Sn (Sensitivity) = No. of bases which hit reference exons/total no. of bases within reference exons = TP/(TP+FN)
Sp (Specificity) = No. of bases which hit reference exons/total coverage = TP/(TP+FP)
Exon hits = No. of reference exons hit/total no. of reference exons.
Gene hits = No. of reference genes hit/ total no. of reference genes.

2.7.6.3 Analysis of the accuracy of Genscan and Fgenesh.

Equivalent calculations of specificity and sensitivity are also calculated at the exonic and genic level in analysis of gene prediction programs. It is assumed that an exon (or gene) has been predicted correctly, only when both its boundaries (and internal exon structure) have been predicted correctly. Predicted exons or genes that only overlap true exons or genes are counted as false predictions. Sn is the proportion of coding exons or genes that correctly align with putatively coding exons or genes and Sp is the proportion of putative coding exons or genes that are actually coding. In this case, only protein coding sequence is taken into account.

Given the stringent criteria used to consider an exon or gene as correctly predicted, two additional measures of specificity and sensitivity are computed. These are the proportion of true exons or genes without overlap to predicted exons or genes – the Missing Exons (ME) or Missing Genes (MG) – and the proportion of predicted exons without overlap to actual exons – the Wrong Exons (WE) or Wrong Genes (WG).

$$ME \text{ (or G)} = \frac{\text{number of Missing Exons (or Genes)}}{\text{number of annotated exons (or genes)}}$$

$$WE \text{ (or G)} = \frac{\text{number of Wrong Exons (or Genes)}}{\text{number of predicted exons (or genes)}}$$

2.7.6.4 Promoter predictions

The results of the algorithms CPGFIND (Micklem, unpublished) PromoterInspector (Scherf *et al.*, 2000) and Eponine (Down, unpublished) were also correlated with the 38 annotated protein-coding genes within 22q13.31. In this case, correlation limits were set at 6 kb upstream to 0.5 kb downstream of the annotated transcription start site, Unlike CPGFIND and PromoterInspector, Eponine attempts to make strand-specific predictions. Only predictions on the same strand as the annotated gene were counted as a positive correlation. The specificity and sensitivity of each prediction type was calculated as before.

2.7.7 Phylogenetic analysis

1. Each of the 27 full-length protein sequences was used to search the NCBI nonredundant protein sequence database (http://www.ncbi.nlm.nih.gov), using the gapped BLASTP program (section 2.7.4).
2. The BLAST alignments were inspected by eye. Entries which were redundant, contained point mutations with respect to sequences already included in the analysis, or corresponded to sequences known to be previously submitted partial versions of the gene of interest, were excluded. Additionally, entries that demonstrated only partial matches were removed, as the sequences involved shared only some functionally similar parts (e.g. multidomain proteins.)
3. Sequence data were aligned using the default options of clustalw (Thompson *et al.*, 1994)(http://www.ebi.ac.uk/clustalw).

4. Neighbour-joining (NJ) analyses(Saitou & Nei, 1987) of the amino acid alignments were produced using Phylowin (Galtier *et al.*, 1996). Robustness of the NJ trees was tested by bootstrap analyses with 500 pseudo-replications per tree.
5. Potential orthologues, identified from the phylogenetic trees, were then compared against the NCBI nonredundant protein sequence database to ensure that orthologous pairs fulfilled the requirements of being the two most similar proteins between two different organisms (Huynen & Bork, 1998; Tatusov *et al.*, 1997; Tatusov *et al.*, 1996).
6. The chromosomal position of potential mouse orthologues was verified as far as possible by BLASTN comparison of the nucleotide sequence against the available mouse genomic sequence (http://mouse.ensembl.org).

2.8 Materials

2.8.1 Buffers

1X TE

- 10 mM Tris-HCl (pH 8.0)
- 1 mM EDTA

1X $T_{0.1}E$

- 10 mM Tris-HCl (pH 8.0)
- 0.1 mM EDTA

10X PCR buffer 1

- 670 mM Tris-HCl (pH 8.8)
- 166 mM $(NH_4)_2SO_4$ (enzyme grade)
- 67 mM MgCl
- (pH 8.8)

28% Sucrose solution

- 1X TE
- 28% w/v sucrose
- 0.008% w/v cresol red

***Pfu* 10X reaction buffer (PCR buffer 2)**

- 200 mM Tris-HCl (pH 8.8)
- 100 mM KCl
- 100mM $(NH_4)_2SO_4$
- 20 mM $MgSO_4$
- 1.0% Triton®X-100

DNase I reaction buffer

- 200 mM Tris-HCl (pH 8.4)
- 20 mM $MgCl_2$
- 500 mM KCl

First Strand buffer

- 250 mM Tris-HCl (pH 8.3)
- 375mM KCl
- 15 mM $MgCl_2$

10X Ligase buffer (Roche)

- 660 mM Tris-HCl
- 50 mM $MgCl_2$
- 10 mM dithioerythritol
- 10 mM ATP

(pH 7.5)

6X Glycerol loading dyes (I)

- 30% v/v glycerol
- 0.1% w/v bromophenol blue
- 0.1% w/v xylene cyanol
- 5 mM EDTA (pH 7.5)

6X Dye Buffer (II)

- 0.25% bromophenol blue
- 0.25% xylene cyanol
- 15% Ficoll (Type 400: Pharmacia)

Vistra Green stain

For 1 gel:

- 0.01M Tris HCl
- 0.0001M EDTA (pH 7.4)
- 50 µl Vistra green (Amersham RPN5786)

10X TAE

- 890mM Tris base
- 0.05M EDTA
- 5.71% glacial acetic acid (JTBaker)

10X TBE

- 890 mM Tris base
- 890 mM Borate
- 20mM EDTA (pH 8.0)

20X SSC

- 3 M NaCl
- 0.3 M Trisodium citrate

100X Denhardt's

- 20 mg/ml Ficoll 400-DL
- 20 mg/ml polyvinylpyrrolidine 40
- 20 mg/ml BSA (pentax fraction V)

Hybridisation buffer

- 6X SSC
- 2 mg/ml Ficoll 400-DL
- 2 mg/ml polyvinylpyrrolidone 40
- 2 mg/ml BSA (pentax fraction V)
- 1% N-lauroyl-sarcosine
- 50mM Tris-HCl (pH 7.4)
- 10% w/v dextran sulphate

SAP buffer

- 200 mM Tris HCl (pH 8.0)
- 100 mM $MgCl_2$

ExoSAP dilution buffer

- 50mM Tris HCl (pH 8.0)

OLB3

- 240 mM Tris-HCl
- 75 mM β-mercaptoethanol
- 0.1 mM dATP
- 0.1 mM dGTP
- 0.1 mM dTTP
- 1 M HEPES (pH 6.6)
- 0.1 mg/ml hexadeoxyribonucleotides (2.1 OD units/ml)

GTE

- 50 mM glucose
- 1 mM EDTA
- 25 mM Tris-HCl (pH 8.0)

3 M K^+/5 M Ac^-

- 60 ml 5 M potassium acetate
- 11.5 ml glacial acetic acid
- 28.5 ml H_20

Protein transfer buffer

- 10% 10X Protein running buffer
- 25% 100% Ethanol

Protein sample buffer

- 2% w/v SDS
- 10% v/v glycerol
- 60 mM Tris-HCl (pH 6.8)
- 0.01% w/v Bromophenol blue
- 5% v/v β-mercaptoethanol

10X Protein running buffer

- 30 g/l Tris base
- 144 g/l glycine
- 10 g/l SDS

2.8.2 Cell culture

2.8.2.1 Growth media

LB

- 10 mg/ml bacto-tryptone
- 5 mg/ml yeast extract
- 10 mg/ml NaCl
- (pH 7.4)

2.8.2.2 Antibiotic concentrations

Mouse RPCI-23 BAC clones: 12.5 μg/ml chloramphenicol.

Human cDNA clones: 100 μg/ml ampicillin.

Blue/white selection of cDNA clones: 100 μg/ml ampicillin/ 0.5 mM IPTG/ 80 μg/ml X-Gal.

2.8.3 Size markers

2.8.3.1 1 kb ladder (GibcoBRL)

This contains 1 to 12 repeats of a 1018 bp concatenated fragment and vector fragments from 75 to 1636 bp, thus producing the following sized fragments (bp):

Table 2.5: 1 kb ladder (GibcoBRL)

Band no.	Size (bp)	Band no.	Size (bp)
1	12216	12	1635
2	11198	13	1018
3	10180	14	516/506
4	9162	15	394
5	8144	16	344
6	7125	17	298
7	6108	18	220
8	5090	19	200
9	4072	20	154
10	3054	21	142
11	2036	22	75

2.8.3.2 Wide Range Analytical Marker DNA (Promega)

The Analytical Marker DNA, Wide Range, provides an evenly spaced distribution of 32 DNA fragments ranging from 702bp to 29,950bp in size and was used for band sizing in fingerprint experiments. This marker is composed of a mixture of restriction enzyme digests of Lambda DNA and ϕX174 DNA.

2.8.3.3 Benchmark™ Prestained Protein Ladder (GibcoBRL)

This ladder for SDS-PAGE consists of 10 proteins ranging in apparent molecular weight from approximately 10 to 200 kDa. The proteins are rendered blue by a proprietary method that covalently couples dyes to the proteins. The fourth protein band from the top is coupled with a pink dye for easy orientation.

Table 2.6: Benchmark™ Prestained Protein Ladder (GibcoBRL)

Band no.	Apparent molecular weight (kDa)
1	172.6
2	111.4
3	79.6
4	61.3
5	49.0
6	36.4
7	24.7
8	19.2
9	13.1
10	9.3

2.8.4 Primer sequences

2.8.4.1Vectorette primer

244 CGA ATC GTA ACC GTT CGT ACG AGA ATC GCT

T7 TAC GAC TCA CTA TAG GGA GA

SP6 CAT ACG ATT TAG GTG ACA C

2.8.4.2 Production of pCDNA3-NT7

pCDNA3-NT7 cassette	GAT CCA GCG GCC GCT G
stpCDNA3-NT7 S	GGC CAA GCT TGC CAC CAT GGC TAG CAT GAC
stpCDNA3-NT7 A	GGC CTC TAG ATC CAG CGG CCG CAG GAT CCC G

2.8.4.3 Other STSs

The primers of all other STSs used are listed in appendix 1.

2.8.5 URLs and ftp sites

Table 2.7 Useful URL and ftp sites

Title	URL
BLAST services at the NCBI	http://www.ncbi.nlm.nih.gov/BLAST
Clustalw	http://www.ebi.ac.uk/clustalw
Ensembl	http://www.ensembl.org/
Ensembl (Mouse)	http://mouse.ensembl.org
Entrez browser	http://www3.ncbi.nlm.nih.gov/Entrez
GeneCards: human genes, proteins and diseases (Weizmann)	http://bioinfo.weizmann.ac.il/cards
Genomatix PromoterInspector	http://www.genomatix.de/cgi-bin/promoterinspector/promoterinspector.pl
GFP project	http://www.dkfz-heidelberg.de/abt0840/GFP/
Humace home page	http://inteweb.sanger.ac.uk/LocalUsers/humace
Human Gene Nomenclature Database	http://www.gene.ucl.ac.uk/cgi-bin/nomenclature/searchgenes.pl
Human working draft genome browser	http://www.infobiogen.fr/services/GoldenPath/mirror/goldenPath/gbdDescriptions.html
InterPro	http://www.ebi.ac.uk/interpro/
Maps of Human and Mouse homology	http://www.ncbi.nlm.nih.gov/Homology/
MatInspector	http://transfac.gbf.de/cgi-bin/matSearch/matsearch2.pl
Mouse BAC ends	http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
Mouse genome database	http://www.informatics.jax.org
PipMaker	http://bio.cse.pse.edu/pipmaker
PIX at the HGMP	http://www.hgmp.mrc.ac.uk/Registered/Webapp/pix/
Primer3	http://www.sanger.ac.uk/cgi-bin/primer3.cgi
REBASE - Restriction Enzymes	http://rebase.neb.com/rebase/rebase.html
RepeatMasker web server	http://ftp.genome.washington.edu/cgi-bin/RepeatMasker
RPCI-23	http://www.chori.org/bacpac/23framefmouse.htm
Sanger Institute	http://www.sanger.ac.uk
Sanger Institute: Human analysis	http://www.sanger.ac.uk/HGP/Humana/human_analysis.shtml
Sanger Institute: SRSWWW	http://www.sanger.ac.uk/srs6
Search Evaluated MEDLINE	http://www.biomednet.com/db/medline
Sequence Logos	http://www.bio.cam.ac.uk/cgi-bin/seqlogo/logo.cgi
SignalP server	http://www.cbs.dtu.dk/services/SignalP
The HGMP Resource Centre	http://www.hgmp.mrc.ac.uk
The IMAGE Consortium	http://image.llnl.hov
The NCBI BLAST server.	http://www.ncbi.nlm.nih.gov/BLAST
The Sanger Institute: BLAST server	http://intweb.sanger.ac.uk/LocalUsers/humace/BLAST/Internal_blast_server.shtml
Web SequenceLogo main form	http://www.bio.cam.ac.uk/cgi-bin/seqlogo/logo.cgi
Unigene-Human	http://www.ncbi.nlm.gov/Schuler/UniGene/Hs.Home.html