

Gene identification and characterisation on the human X chromosome

Ian Barrett

Darwin College, University of Cambridge

January 2004



This dissertation is submitted for the degree of Doctor of Philosophy.

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except where specifically indicated in the text.

This dissertation does not exceed the size limit for the Biology Degree Committee.

Abstract

This thesis investigates the gene composition and evolution of regions of the human X chromosome, including data from comparative genome analysis of other organisms.

Chapter Three presents studies undertaken to annotate genes within the Xq22-q23 region of the human X chromosome. Selected features of the region are discussed, including investigation of alternative polyadenylation site usage, an insertion of the mitochondrial genome into the nuclear genome, and an inverted duplication and potential gene fusion event involving the NXF2 and TCP11-like genes.

As a result of the annotation described in Chapter Three, extensive paralogy within the Xq22 region was discovered, along with additional examples of paralogy between Xp and Xq22-q23. Work in subsequent Chapters attempted to characterise these aspects further and provide information on the evolution of the regions.

Chapter Four describes work undertaken to map and sequence the region of the mouse genome corresponding to human Xq22-q23, in order to investigate the evolution of the Xq22 paralogues. Annotation of genes within the region of the mouse X chromosome and the orthology of the human and mouse regions is described. Features of the mouse region, such as the presence of two large repeat families, are also discussed.

Chapter Five presents phylogenetic and expression profile analysis of the Xq22 paralogues, and examines orthology in the corresponding region of the mouse genome.

Chapter Six includes a discussion of Xp/Xq paralogy, and presents studies providing evidence for a segmental duplication leading to this paralogy. In addition, orthologues of the genes involved in the paralogy are identified in the marsupial mouse, *Sminthopsis macroura*, and their genomic localisations determined. Evidence suggesting a minimum age of the duplication is presented. Comparative analysis of human, mouse, *Fugu rubripes* and *Sminthopsis macroura* genomic sequence is described.

Acknowledgements

I don't really know how to begin thanking everyone. I've been extremely fortunate to have benefited from the kindness, help, advice and expertise of so many people during these studies, and have had great times along the way (as well as numerous genome T-shirts). The selfless help I've received from so many people really has been very touching and humbling.

Firstly, thanks to Tim, David and Andy for your support and understanding over the last months. Thank you also to the BGS and Education Committee for their support. Financial support from the MRC and The Wellcome Trust is gratefully acknowledged.

A huge thanks to all the Sanger Institute cloning, sequencing and finishing teams for all their work spanning sequence from human, mouse and that furry little fella, the marsupial mouse – without you none of this work would have been possible. Thanks especially to Kirsten McLay, Darren Grafham, Christine Bird, Siobhan Whitehead and Lucy Matthews for answering so many of my queries. Thanks as well to Sean Humphray and Glen Threadgold for all the help with the mouse mapping. A big thanks to Paul Hunt for putting up with me and for all your gridding and being generally very helpful, and to Carol Carder for showing me the ropes, and letting me upset the carefully balanced workings of your lab! Thanks as well to the now HAVANA group, for the sequence analysis, especially Stephen Keenan, whom I badgered on many occasions.

I'd also like to thank Jess Tyson, Karen Woodward and Willem Rens for our collaborations, which I enjoyed very much, and from which I learned a great deal. A particular thanks to Willem for the gifts of resources and advice, and your infectious enthusiasm!

Thank you Jackie Bye and Susan Rhodes for all the work you put into the cDNA resources, and especially to you Jackie for all the advice and encouragement, and our discussions! I'm watching out for that next invention... Thank you to John Collins and David Beare for all the advice on the transcript mapping, and for paving the way.

A very big thanks to all the FISH gang and Nigel Carter. I had a great time as an "imposter" in the lab. Thanks very much to you all for your help in times of need in getting things done. I won't forget my last day at the Sanger, checking out some FISH images (nor my special FISH-cake Louisa....!). Thanks particularly to Ruby Banerjee, Susan Gribble and especially to Deb Burford for really coming through when things were tight (hope I can repay you in some way in your PhD studies).

Thanks to Christina Hedberg-Delouka and Christine Rees for keeping me in check (as much as you could...) and also to Joan Green and Andrew King for help and advice on all things library-related (and sorry for the often late returns..!). Thanks to Richard Summers as well for all the reprographics support and photos for my leaving gifts.

Thanks to Kate Rice and David Bentley for the chats and advice, I hope HumGen continues to go from strength to strength. I feel very lucky to have been part of it. Thank you to David Vetrie, Stephan Beck and Nabeel Affara for helpful advice. A big thanks to Cords, Matt, Rob and Em for being good mates and also to Paw and Vicki for a lot of laughs and chats. Thanks to "Tammar" Tamsin as well for all the encouragement and chats, and congratulations on your rowing achievements. Can ya tell wot it is yit? "Frantastic" Fran Lovell and Christine Burrows, I especially owe you guys a great deal in all the help you gave me when times were tight (and even when they weren't). Thanks also to Adam Whittaker for all the hyb jokes, and Jen Conquer, Helen Steingruber and Barbara Gorick for all the office banter! Thanks very much

to “Pod’s Angels” Hazel Arbery and Ruth Bennett as well, for your real efforts in the last push. You’re real stars.

Carol Scott, I bet you’ll be glad I’ve finished! I can’t thank you enough for all the remote-control help you’ve given me, and thanks also to you and Sarah Hunt for all the help while I was actually a genuine member of the Sanger Institute.

Gareth “Gazza” Howell, thanks for all you’ve taught me and for all the clowning around, although I’m glad to be able to walk the corridors without apprehension.... On a more serious note, thanks for all the work you put into Xq22, which formed the basis for this work. Thanks as well to Simon Gregory for the help, advice and laughs, just a shame you weren’t around for the rugby this year...! Thanks both for all your advice and friendship, and for all the good times (plenty more to come).

Pod (a.k.a. Alison Coffey), you really have been a star. I’ve learned so much from you and you really have been an inspiration in more ways than one. Thanks for everything, and especially for all the help in the final push – definitely couldn’t have done it without you.

Mark Ross, what can I say? You’ve taught me a lot over the last years, more than you’ll probably realise and not just good science. You’re an excellent scientist to learn from and a great mentor. Thanks very much from both of us for all you’ve done.

People often say they couldn’t have done it without you etc. but in this case especially I really couldn’t. I’ve been helped by, and learned from, more people than I’ve been able to mention here. I’ll always be grateful to everyone (you and I know who you are) for all your help, and for all the fantastic memories I’ve got from my time at the Sanger, a very special place.

And now thanks to all our family and friends for all their support over the years and getting me this far. Thanks to Andy for the chats and to Mark for being a good mate.

Thank you especially to my long-suffering wife Gill (well, not long yet but pretty suffering) for all your support and love over good times and bad and for all your help, especially your help with the final printing and formatting. You’ve really held things together over the last year, which has truly tested us both. Can’t wait to get on with the rest of our married life, and see what it holds for us.

Thank you to Dad for being understanding when I didn’t feel like giving an update on progress and for all your love, support, interest and encouragement. You really are someone to look up to.

And finally, a very special thanks to Mum for all that you’ve given and taught me. There just isn’t enough space to put it all here. You’re an incredible inspiration and very much loved. I’d like to dedicate this work to you and Dad. Sleep well.

Table of Contents

Title.....	i
Abstract.....	ii
Acknowledgements.....	iii
Table of contents.....	v
List of figures.....	xi
List of tables.....	xix
Glossary of abbreviations.....	xxii
1 Chapter1 - Introduction.....	1
1 Introduction.....	2
1.1 Genome mapping and sequencing	2
1.1.1 Genomes	2
1.1.2 Genome mapping and sequencing	3
1.1.3 The human and mouse genome projects	5
1.1.4 Future directions and related studies	7
1.2 Gene identification	8
1.2.1 Genes	8
1.2.2 cDNA-based gene identification methods (direct selection) and exon trapping	11
1.2.3 Sequence-based gene and regulatory-sequence identification methods	12
1.2.4 Genomic sequence analysis	15
1.3 Gene duplications and evolution of genomes	17
1.3.1 General evolution framework	17
1.3.2 Whole genome duplication hypothesis	20
1.3.3 Segmental duplications and tandem duplications	21
1.4 Comparative genomic analysis	25
1.4.1 Identification of functionally important sequences	25
1.4.2 Evolutionary studies	27
1.4.3 Value of comparative genomic analysis for functional studies	28
1.4.4 Current and future prospects/projects	29
1.5 The X chromosome	30
1.5.1 Human X chromosome overview	30
1.5.2 Sex chromosome evolution	32

1.5.3	Human Xq22-q23	35
1.5.3	Human Xq22-q23	36
2	Chapter 2 - Materials and Methods	38
2	Materials and Methods	39
2.1	Chemical reagents	39
2.2	Enzymes and commercially prepared kits	39
2.3	Nucleotides	40
2.4	Solutions	40
2.4.1	Buffers	40
2.4.2	Electrophoresis solutions	40
2.4.3	Media	41
2.4.4	DNA labelling and hybridisation solutions	41
2.4.5	General DNA preparation solutions	42
2.4.6	FISH solutions	42
2.5	Size markers	43
2.6	Hybridisation membranes and X-ray and photographic film	43
2.7	Sources of genomic DNA	43
2.8	Sources of RNA	43
2.9	Sources of cells for <i>Sminthopsis macroura</i> and mouse FISH	44
2.10	Bacterial clone libraries	45
2.10.1	Cosmid libraries	45
2.10.2	PAC and BAC libraries	45
2.10.3	cDNA libraries	45
2.11	Primer sequences	47
2.12	Key World Wide Web addresses	47
2.13	Isolation of bacterial clone DNA	49
2.13.1	Miniprep of BAC DNA	49

2.13.2	Microprep of BAC DNA for restriction digest fingerprinting50
2.14	Bacterial clone Hind III/Sau 3A 1 fluorescent fingerprinting51
2.15	Agarose gel preparation and electrophoresis52
2.16	Applications using the polymerase chain reaction53
2.16.1	General primer design53
2.16.2	Oligonucleotide preparation53
2.16.3	Amplification of genomic DNA by PCR53
2.16.4	Colony PCR of STSs from bacterial clones54
2.17	Radiolabelling of DNA probes by direct incorporation54
2.18	Hybridisation of radiolabelled DNA probes55
2.18.1	Hybridisation of DNA probes derived from STSs to same-species BAC filters55
2.18.2	Hybridisation of DNA probes derived from human STSs to <i>Sminthopsis macroura</i> BAC filters55
2.18.3	Stripping radiolabelled probes from hybridisation filters56
2.19	Clone library screening56
2.19.1	cDNA library screening by PCR56
2.19.2	Single-sided specificity PCR (SSPCR) of cDNA57
2.20	RT-PCR expression profiling using total RNA samples59
2.20.1	Generation of cDNA from total RNA59
2.20.2	Screening of cDNA samples by PCR59
2.21	FISH of BAC-derived probes to <i>Sminthopsis macroura</i> metaphase preparations60
2.21.1	Metaphase slide preparation60
2.21.2	Nick-translation labelling of BAC clones60
2.21.3	Hybridisation of FISH probes to <i>Sminthopsis macroura</i> metaphase slides61
2.21.4	Washing of <i>Sminthopsis macroura</i> metaphase slides and detection of signal62
2.21.5	Two-colour detection using a Cadenza instrument63
2.22	Fibre-FISH of BAC-derived probes to <i>Mus musculus</i> DNA fibres63
2.22.1	<i>Mus musculus</i> DNA fibre slide preparation63
2.22.2	Hybridisation of FISH probes to <i>Mus musculus</i> DNA fibre slides64
2.23	Mapping and sequence analysis software and databases64

2.23.1	IMAGE	64
2.23.2	FPC	65
2.23.3	Xace and other custom mapping databases	65
2.23.4	BLIXEM	66
2.23.5	RepeatMasker	67
2.24	Alignment of nucleic acid and protein sequences and phylogenetic analysis	67
2.24.1	Alignment of nucleic acid and protein sequences	67
2.24.2	Calculation of sequence identities and similarities	68
2.24.3	Phylogenetic analysis of nucleic acid and protein sequences	68
2.25	Comparative sequence analysis	69
2.25.1	PIPmaker	69
2.25.2	VISTA	69
3	Chapter 3 - Gene annotation of the human Xq22-q23 region.....	70
3.1	Introduction	71
3.2	Generation of an annotated gene map of human Xq22-q23	73
3.3	Selected features of the region	99
3.3.1	Discovery of extensive paralogy within human Xq22 and between Xp and Xq22-q23.....	99
3.3.2	NXF2 inverted repeat and gene fusion	100
3.3.3	Alternative 3'-UTR usage	105
3.3.4	Mitochondrial insertion into the nuclear genome at Xq22	109
3.4	Discussion	111
4	Chapter 4 - Genomic landscape of the mouse genomic region equivalent to human Xq22-q23	113
4.1	Introduction	114
4.2	Assembly of a sequence-ready BAC contig for mouse X E3-F2	116
4.3	Identification of genes and their structures using sequence analysis	129

4.4	Comparative analysis of the human and mouse Xq22-q23/E3-F2 region	134
4.4.1	Orthologues of human Xq22 genes	134
4.4.2	PARL repeats	138
4.4.3	Pramel3L repeats	142
4.4.4	The mouse Nxf2 locus	146
4.4.5	A mouse gene supporting the presence of a novel gene in human Xq22	146
4.5	Discussion	148
5	Chapter 5 - Characterisation of extensive gene duplication discovered within human Xq22-q23	150
5.1	Introduction	151
5.1.1	Thymosin-beta genes	153
5.1.2	NADE family genes	153
5.1.3	NXF family and TCP11-like genes	154
5.1.4	ALEX family genes	154
5.1.5	GASP family genes	155
5.1.6	pp21/TCEAL1 family genes	155
5.1.7	Rab-like genes	155
5.1.8	Histones and cU46H11.CX.1/cU116E7.CX.1 genes	155
5.1.9	Tex genes and COL4A5/COL4A6	155
5.2	Duplicated genes within Xq22-q23: sequence analysis, comparative analysis, phylogenetic analysis and RT-PCR expression profiles	156
5.2.1	Thymosin-beta genes	161
5.2.2	NADE family genes	167
5.2.3	NXF family and TCP11-like genes	172
5.2.4	ALEX family genes	179
5.2.5	GASP family genes	185
5.2.6	pp21/TCEAL genes	191
5.2.7	RAB-like genes	197
5.2.8	Histones and cU46H11.CX.1/cU116E7.CX.1 genes	200
5.2.9	Duplicated pseudogenes, TEX genes and COL4A5/COL4A6	203
5.3	Discussion	206

6	Chapter 6 - Characterisation of a regional duplication represented on human Xq22-q23 and Xp	208
6.1	Introduction	208
6.2	Characterisation of the Xq22-q23/Xp regional duplication	211
6.2.1	Extent of the duplication and genes involved	211
6.3	Identification of orthologues of the duplicated genes in the marsupial mouse, <i>Sminthopsis macroura</i>	218
6.4	Genomic localisation of the <i>Sminthopsis macroura</i> orthologues by FISH	234
6.5	Dating the Xq22-q23/Xp regional duplication	250
6.5.1	Gene-based evidence from the scientific literature	250
6.5.2	Comparative analysis of the <i>Fugu rubripes</i> genome	252
6.6	Comparative analysis of <i>Sminthopsis macroura</i> genomic sequence ...	264
6.6.1	Comparative analysis of sequence composition for human, mouse and <i>Sminthopsis macroura</i>	266
6.6.2	Comparative sequence analysis of the CSTF2/NOX1/XK-L region in human, mouse, <i>Sminthopsis macroura</i> and <i>Fugu rubripes</i> using PIP and VISTA	269
6.7	Discussion	274
7	Chapter 7 - Discussion	277
7.1	Summary of thesis results and discussion of major themes	278
7.2	Future directions	285
	References.....	290
	Appendices.....	307

List of Figures

Chapter 1

Figure 1-1	Schematic cladogram of selected branches of chordate/vertebrate evolutionary relationships	19
Figure 1-2	Schematic representation of proposed genome duplication events during vertebrate evolution	20
Figure 1-3	Overview of the regions being mapped and sequenced by each of the major sequencing centres	31
Figure 1-4	Schematic diagram of the human X and Y chromosomes	33
Figure 1-5	Figure depicting selected aspects of mammalian phylogenies	36
Figure 1-6	Diagram illustrating conservation between human and marsupial X chromosomes	37

Chapter 2

Figure 2-1	Diagram illustrating the SSPCR procedure	58
------------	--	----

Chapter 3

Figure 3-1	G-banded ideogram of the human X chromosome (Francke), illustrating the region Xq22-q23 chosen for this study	73
Figure 3-2	The image above illustrates how results of sequence analyses were collated and viewed within Xace	75
Figure 3-3	Example of an mRNA match viewed using BLIXEM	76
Figure 3-4	Example of a “gene” (GD_mRNA) structure	77
Figure 3-5	Example of a “predicted gene” (GD_composite) structure	77
Figure 3-6	Diagram illustrating a “pseudogene” (pseudogene) structure	78

Figure 3-7	Prescreening, poolscreening and SSPCR for STS stSG84336	88
Figure 3-8	Diagram illustrating annotation of a gene structure using SCCD sequence	89
Figure 3-9	Transcript map of human Xq22-q23	97
Figure 3-10	Observations of Xp/Xq paralogues	99
Figure 3-11	Diagram showing results of Dotter analysis of the genomic sequence flanking the two NXF2 loci (against itself)	100
Figure 3-12	A schematic representation of the inverted repeat containing the TCP11-like and NXF2 loci	102
Figure 3-13	Expression profiling of NXF2 variants	103
Figure 3-14	A schematic representation of the primer positions within the 3' UTRs of (a) ALEX3, (b) CSTF2 and (c) TBG	104
Figure 3-15	Images of 2.5% agarose gels containing RT-PCR products for primers designed to 3'-UTR variants	105
Figure 3-16	A tabulated summary of the RT-PCR results	106
Figure 3-17	Xace representation of mitochondrial genome-encoded protein matches within clone bA522L3	108
Figure 3-18	Diagram illustrating BLASTN matches between human Xq22 sequence (from clone bA522L3) and the mitochondrial genome	109
Figure 3-19	Schematic diagram of a model for integration of the mitochondrial genome in sequence accession AL590407	110
Chapter 4		
Figure 4-1	Figure illustrating the mouse X chromosome showing blocks with conserved synteny to the human X chromosome	115

Figure 4-2	Diagram illustrating the STS-based hybridisation strategy used to isolate mouse RPCI-23 BAC clones	120
Figure 4-3	A first-generation BAC contig map of the mouse X E3-F2 region	121
Figure 4-4	Example of detection of contig overlaps utilising WGS assemblies	123
Figure 4-5	Finalised BAC contig map of the mouse X E3-F2 region	124
Figure 4-6	Results of fibre-FISH	125
Figure 4-7	Diagram illustrating a section of contig 24 illustrating the tiling path	126
Figure 4-8	Diagram illustrating overlapping clones from contig 24 based on fingerprint data	127
Figure 4-9	Diagram illustrating a “gene” structure for the Plp gene	130
Figure 4-10	Diagram illustrating a “predicted gene” structure for locus bM182N4.MX.3	131
Figure 4-11	Diagram illustrating a “pseudogene” structure for the pseudogene locus bM389M3.MX.4	131
Figure 4-12	Diagram illustrating transcript map of annotated loci in the <i>Mus musculus</i> X E3-F2 region	132
Figure 4-13	Schematic diagram of PARL-like loci within the <i>Mus musculus</i> X E3-F2 region	140
Figure 4-14	Diagram illustrating Dotter analysis of the <i>Mus musculus</i> X E3-F2 region containing PARL-like repeats	141
Figure 4-15	Schematic diagram of Pramel3L loci within the <i>Mus musculus</i> X E3-F2 region	144
Figure 4-16	Diagram illustrating Dotter analysis of the <i>Mus musculus</i> X E3-F2 region containing Pramel3L repeats	145

Figure 4-17	Alignment of human gene prediction cU101D3.GENSCAN.3 and part of mouse mRNA AK017555.1	147
-------------	--	-----

Chapter 5

Figure 5-1	Diagram illustrating positions of duplicated loci within Xq22	152
Figure 5-2	Figure showing a schematic representation of thymosin-beta paralogue gene order and orientation within human Xq22, and the corresponding orthologous region in <i>Mus musculus</i> (mouse)	162
Figure 5-3	Figure illustrating alignments of thymosin-beta paralogues	163
Figure 5-4	Phylogenetic analysis of human and mouse thymosin-beta genes	164
Figure 5-5	RT-PCR profiling of human thymosin-beta family genes	165
Figure 5-6	A schematic representation of NADE family gene order and orientation within human Xq22 and the corresponding orthologous region in mouse	168
Figure 5-7	Alignments of NADE family predicted peptides	169
Figure 5-8	Phylogenetic analysis of NADE Xq22 paralogues	170
Figure 5-9	RT-PCR expression profiling of NADE genes	171
Figure 5-10	Figure showing a schematic representation of NXF and TCP11-like paralogue gene order and orientation within human Xq22, and the corresponding orthologous region in mouse	174
Figure 5-11	Alignments of NXF family genes and TCP11-like genes	175
Figure 5-12	Figure illustrating phylogenetic analysis of human and mouse NXF genes	176
Figure 5-13	Figure illustrating phylogenetic analysis of human and mouse TCP11-like genes	177
Figure 5-14	RT-PCR expression profiling of NXF family genes	178

Figure 5-15	Figure showing a schematic representation of ALEX paralogue gene order and orientation within human Xq22, and the corresponding orthologous region in mouse	181
Figure 5-16	Figure illustrating alignment of ALEX-like predicted peptides, for human and mouse	182
Figure 5-17	Phylogenetic analysis of human and mouse ALEX-like genes	183
Figure 5-18	RT-PCR expression profiling of ALEX-like genes	184
Figure 5-19	Figure showing a schematic representation of GASP paralogue gene order and orientation within human Xq22, and the corresponding orthologous region in mouse	187
Figure 5-20	Figure illustrating alignments of GASP paralogue predicted peptides	188
Figure 5-21	Figure illustrating phylogenetic analysis of GASP-like genes	189
Figure 5-22	RT-PCR expression profiling of GASP-like genes	190
Figure 5-23	Alignment of GASP and ALEX family predicted peptides showing conservation between the sequences	191
Figure 5-24	Figure showing a schematic representation of pp21/TCEAL1 paralogue gene order and orientation within human Xq22, and the corresponding orthologous region in mouse	193
Figure 5-25	Figure illustrating alignments of pp21/TCEAL human paralogue predicted peptides	194
Figure 5-26	Figure illustrating phylogenetic analysis of pp21/TCEAL genes	195
Figure 5-27	RT-PCR expression profiling of pp21-like genes	196
Figure 5-28	Figure showing a schematic representation of RAB-like paralogue gene order and orientation within human Xq22, and the corresponding orthologous region in mouse	198

Figure 5-29	Figure illustrating alignments of RAB-like paralogues	199
Figure 5-30	Figure showing a schematic representation of histone and cU46H11.CX.1/cU116E7.CX.3 paralogue gene order and orientation within human Xq22, and the corresponding orthologous region in mouse	201
Figure 5-31	Figure illustrating alignments of cU46H11/cU116E7 genes	202
Figure 5-32	RT-PCR expression profiling of cU46H11.CX.1	203
Figure 5-33	Partial alignment of TEX mRNA sequences	204
Figure 5-34	Figure showing a schematic representation of pseudogene, TEX and COL4A5/COL4A6 paralogue gene order and orientation within human Xq22	205
 Chapter 6		
Figure 6-1	Observations of Xp/Xq paralogues	210
Figure 6- 2	Gene order and orientation of paralogous pairs on Xq and Xp (Block 1)	214
Figure 6-3	Schematic representation of paralogy between Xp21.3-p11.4 and Xq22.1 (Block 2)	215
Figure 6-4	Schematic representation of chromosomal locations of autosomal genes with paralogues on the X chromosome	217
Figure 6-5	Strategy for design of primers to amplify probes for use in a reduced-stringency hybridisation approach to identify <i>Sminthopsis macroura</i> BAC clones	221
Figure 6-6	Diagram showing the genes for which probes were designed to identify orthologues in <i>Sminthopsis macroura</i> , and their positions on the human X chromosome	222

Figure 6-7	An example of a hybridisation of a pool of five probes to filters of the <i>Sminthopsis macroura</i> library	223
Figure 6-8	An example of the second round of the reduced-stringency hybridisation procedure	224
Figure 6-9	FPC representation of contig 2	232
Figure 6-10	Diagram illustrating genes for which <i>Sminthopsis macroura</i> positive BACs were selected for FISH analysis and sequencing	236
Figure 6-11	Karyogram showing <i>Sminthopsis macroura</i> normal karyotype ideogram and representative DAPI-stained chromosomes from metaphase chromosome preparations from a male <i>Sminthopsis macroura</i> cell line (2n=18) used for FISH analyses	238
Figure 6-12	FISH of <i>Sminthopsis macroura</i> BAC clones against spreads of metaphase chromosomes	242
Figure 6-13	FISH of <i>Sminthopsis macroura</i> BAC clones against spreads of metaphase chromosomes	243
Figure 6-14	FISH of <i>Sminthopsis macroura</i> BAC clones against spreads of metaphase chromosomes	244
Figure 6-15	FISH of <i>Sminthopsis macroura</i> BAC clones against spreads of metaphase chromosomes	245
Figure 6-16	FISH of <i>Sminthopsis macroura</i> BAC clones against spreads of metaphase chromosomes	246
Figure 6-17	Example of results from co-localisation experiments using FISH of <i>Sminthopsis macroura</i> BAC clones against spreads of metaphase chromosomes	248
Figure 6-18	Schematic representation of selected <i>Fugu rubripes</i> WGS sequence scaffolds with information regarding putative <i>Fugu</i> orthologue gene	

	order, transcription direction and shared synteny with human Xp/Xq paralogue and non-Xp/Xq paralogue orthologues	255
Figure 6-19	Phylogenetic trees constructed for the MID genes	259
Figure 6-20	Phylogenetic trees constructed for the RAB genes	260
Figure 6-21	Phylogenetic trees constructed for the SYTL genes	261
Figure 6-22	Phylogenetic trees constructed for the Sushi-repeat genes	262
Figure 6-23	Phylogenetic trees constructed for the lipophilin genes.	263
Figure 6-24	PIP plot of the human Xq22 region encompassing genes CSTF2, NOX1 and XK-L	271
Figure 6-25	VISTA plot of the human Xq22 region encompassing genes CSTF2, NOX1 and XK-L	273
Figure 6-26	Diagram summarising analyses presented in this Chapter and providing a model for the evolution of the Xp/Xq paralogous regions	276

List of Tables

Chapter 1

Table 1-1	Table of major human and mouse cDNA sequencing projects in progress at time of writing	13
-----------	--	----

Chapter 2

Table 2-1	Clones and appropriate antibiotics	41
Table 2-2	Sources of total RNA used for RT-PCR experiments	44
Table 2-3	cDNA libraries used for SSPCR	46
Table 2-4	Vector-specific primer sequences for primers used in SSPCR	47
Table 2-5	Primer combinations used in SSPCR	57

Chapter 3

Table 3-1	Results of STS pre-screens and poolscreens	81
Table 3-2	Genes, predicted genes and pseudogenes annotated within Xq22-q23	90

Chapter 4

Table 4-1	Table of genes used to screen for mouse BACs	119
Table 4-2	Table of clones selected for sequencing and status of the region	128
Table 4-3	List of annotated loci within the region	133
Table 4-4	Human Xq22 genes (listed Cen to Tel) and their likely mouse orthologues	134
Table 4-5	Sequence identities (% identity) of PARL-like nucleotide sequences	135
Table 4-6	Sequence identities (% identity) of Pramel3L nucleotide sequences	142

Chapter 5

Table 5-1	Results of colony screens using expression profiling STS primers against genomic clones from the Xq22 tiling path	160
-----------	---	-----

Chapter 6

Table 6-1	Gene structure information obtained from Ensembl v15.33.1.	212
Table 6-2	Sequence and structural comparisons of paralogous gene pairs	213
Table 6-3	Table showing results from the second round of <i>Sminthopsis macroura</i> BAC library screening after increasing stringency washes	225
Table 6-4	Results from the <i>Sminthopsis macroura</i> BAC library screening	226
Table 6-5	Paralogous gene pairs for which their respective probes detected clones in common	230
Table 6-6	Non-paralogous genes for which their respective probes detected clones in common	231
Table 6-7	<i>Sminthopsis macroura</i> Hind III/Sau3A fingerprinting results	232
Table 6-8	<i>Sminthopsis macroura</i> BAC clones chosen for FISH analysis and sequencing	235
Table 6-9	Localisation data for FISH of <i>Sminthopsis macroura</i> BAC clones against spreads of metaphase chromosomes	241
Table 6-10	Co-localisation experiments by FISH of <i>Sminthopsis macroura</i> BAC clones against spreads of metaphase chromosomes	247
Table 6-11	<i>Fugu rubripes</i> orthologues (as determined by reciprocal BLAST analysis) collated from Ensembl (<i>Fugu</i>) release 15.2.1 and Ensembl (Human) release 15.33.1	254
Table 6-12	Human gene structure information obtained from Ensembl v15.33.1 and the Xq22-q23 transcript map described in Chapter 3, and <i>Fugu</i> gene structure information obtained from Ensembl (<i>Fugu</i>) v15.2.1	256

Table 6-13	Human gene structure information obtained from Ensembl v15.33.1 and the Xq22-q23 transcript map described in Chapter 3, and Fugu gene structure information obtained from Ensembl (Fugu) v15.2.1	257
Table 6-14	Marsupial clone sequences and genes annotated	266
Table 6-15	Sequence composition data from RepeatMasker analysis of marsupial, human and mouse orthologous regions	268

Abbreviations

µg	microgram
µl	microlitre
µM	micromolar
µm	micrometre
°C	degrees celsius
ACeDB	A <i>C. elegans</i> Database
AT	annealing temperature
BAC	bacterial artificial chromosome
BLAST	basic local alignment search tool
Bp	base pair
cDNA	complementary deoxyribonucleic acid
CEN	centromere
cm	centimetre
cM	centimorgan
CpG	cytidyl phosphoguanosine dinucleotide
Ctg	contig
DDC	duplication degeneration complementation
DDW	double distilled water
DMD	duchenne muscular dystrophy
DNA	deoxyribonucleic acid
EBI	European Bioinformatics Institute
EST	expressed sequence tag
FISH	fluorescence <i>in situ</i> hybridisation
g	gram or force of gravity
HAVANA	Human And Vertebrate ANalysis and Annotation
HTGS	high throughput genomic sequence
HUGO	Human Genome Organisation
kb	kilobase pairs
LCR	low copy repeat
LD	linkage disequilibrium
LINE	long interspersed nuclear element
LTR	long terminal repeat
M	molar
Mb	megabase pairs
MHC	major histocompatibility complex
MIR	medium interspersed repeat
mL	millilitre
mM	millimolar
mm	millimetre
mRNA	messenger ribonucleic acid
Mya	million years ago
NCBI	National Centre for Biotechnology Information
ng	nanogram
nm	nanometre
numts	nuclear mitochondrial insertions
OMIM	Online Mendelian Inheritance in Man
ORF	open reading frame
PAC	P1-derived artificial chromosome

PAR	pseudoautosomal region
PBS	phosphate-buffered saline
PC	personal computer
PCR	polymerase chain reaction
RACE	rapid amplification of cDNA ends
RFLP	restriction fragment length polymorphisms
RH	radiation hybrid
RNA	ribonucleic acid
rpm	revolutions per minute
RT	reverse transcriptase
RT-PCR	reverse transcriptase-polymerase chain reaction
SINE	short interspersed nuclear element
SNP	single nucleotide polymorphism
SSPCR	single-sided specificity polymerase chain reaction
STS	sequence tagged site
TEL	telomere
U	units
UTR	untranslated region
UV	ultra-violet
V	volts
VEGA	VErtebrate Genome Annotation
WGS	whole-genome shotgun
XAR	X added region
XCR	X conserved region
Xic	X inactivation centre
XIST	X inactivation specific transcript
YAC	yeast artificial chromosome