

An Analysis of Transcript Variation in Human Xp11.23

By
Tamsin Lee Eades

Thesis submitted for
the degree of Doctor of Philosophy

The Wellcome Trust Sanger Institute
and
Darwin College,
University of Cambridge
August 2005

This thesis 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 word limit set by the Biology Degree Committee.

Tamsin Eades

Abstract

Prior to its completion, it was conjectured that the human genome sequence contained in excess of 100,000 genes, whereas current estimates from the completed genome sequence have reduced this figure to between 20,000 and 25,000. This dramatic reduction could be partially attributed to the phenomenon of alternative splicing, where one gene has the ability to produce more than one functional mRNA transcript.

The aim of this thesis was to obtain a greater appreciation for the diversity of transcripts that can be generated from a single gene. Towards this end, the human genome sequence provided the definitive substrate for the study of transcript variation. A gene-rich region of the X chromosome was selected for analysis, where the frequency and the associated functional consequences of alternative splicing were assessed.

This thesis describes an analysis of the genome features and gene content in human Xp11.22-p11.3. Human Xp11.23 has the highest density of genes in the human X chromosome, and this analysis identified 77 known and 11 novel genes. The pseudogene content of this region was also high - it contained 59 processed and 7 non-processed pseudogenes. More detailed investigation of these revealed that the number and diversity of gene products generated from genes within Xp11.23 greatly exceeded the number of coding regions that it contained. In order to further study the impact of alternative splicing in Xp11.23 detailed analysis was completed on 18 genes using bioinformatic and comparative analysis together with a targeted RT-PCR sequencing strategy. This analysis identified more than 120 transcripts variants. Preliminary tissue profiling of these transcripts was completed using RT-PCR.

The functional consequences of alternative splicing were then investigated for one gene, polyglutamine binding protein 1, *PQBP1*. This ubiquitously expressed gene has been associated with various disease phenotypes including X-linked mental retardation, Renpenning syndrome and other neurodegenerative disorders. In concert with expression and evolutionary analysis, a cloned open reading frame

collection was generated, for 16 transcript variants. The relative abundance of minor transcript variants was determined in a panel of 20 human tissues where it was found that together the minor variants accounted for less than 10% of the all *PQBP1* transcripts.

Following *in silico* analysis of the predicted protein sequences, it was found that transcript variation was associated with variable inclusion of a nuclear localisation signal. Sub-cellular localisation analysis of the transcript variants in CHO-K1 and Cos7 mammalian cell lines showed that three isoforms were not redirected to the nucleus after translation. Further analysis revealed that 11 of the *PQBP1* transcripts harboured a premature termination codon. Preliminary mRNA kinetic assays, and nonsense mediated decay inhibition assays confirmed that at least five of these transcripts promoted rapid degradation via transcript surveillance pathways.

Acknowledgements

I would like to thank my supervisor Mark Ross, for his encouragement, support and advice throughout the entirety of this project. I feel incredibly fortunate to have had a supervisor who has always made time for me and has taught me the value of meticulous experimental planning, execution and analysis. Special thanks are also extended to Alison Coffey and Colette Johnston who not only proof-read this thesis but also gave valuable advice on *everything*. I have benefited so much from your wisdom, and friendship - thank you.

I am also thankful for the computational support that I have received from numerous teams at the Sanger Institute throughout this project. Carol Scott and Sarah Hunt from human genetics informatics have provided perl scripts and advice on routine computational analysis of large amounts of sequence. Thanks to Darren Grafham for prioritising some human and mouse clones for finishing. James Gilbert and Stephen Keenan from the Institute's informatics team carried out the computational analysis of the genome sequence, while some of the genome annotation described in this thesis was completed by Jennifer Harrow and the human and vertebrate analysis (HAVANA) team at the Sanger Institute.

Thanks are also extended to people who helped me with my experiments. Firstly, to the current and past members of the X chromosome group and DNA collections team, Christine Burrows, Frances Lovell, Ruth Bennett, Eleanor Howard and Hazel Arbury thank you to you all for helping "find my feet" at the Sanger, for your practical advice and for lending solutions/equipment in times of need. I shared a panel of human total RNA samples with Jackie Bye and am extremely grateful and thankful for her advice and patience with all RNA work. Many resources used for the sub-cellular localisation experiments were kindly provided by team 62 (molecular genomics and proteomics). Special thanks are extended to its members especially Ian Dunham, John Collins, Charmain Wright and James Grinham. Philippe Couttet, gave me much needed advice on NMD and mRNA stability experiments and provided me with the pTRE-TIGHT (tet-off) plasmid. Thanks also to Ian Barrett and Gareth Howell for giving me such a friendly introduction to the Sanger Institute.

Finally, thank you to all of my friends and family who have been so patient and tolerant over the past 3 ½ years. It's now time to repay many favours and perhaps I will now get to spend a Christmas at home!!! I am especially indebted to my parents and to Graeme, I know that I wouldn't have made it to this stage without you, thank you.

This project was funded by the Wellcome Trust and the Sanger Institute.

Table of Contents

Abstract	3
Acknowledgements	5
Table of Contents	7
List of Figures	14
List of Tables	17
List of Appendices.....	19
Abbreviations.....	20
Chapter 1: Introduction	25
1.1 The Human Genome Project.....	26
1.1.1 The human X chromosome.....	28
1.1.2 Applications of the human genome sequence	31
1.2 Describing the genomic landscape.....	33
1.2.1 G+C content.....	34
1.2.2 CpG islands.....	34
1.2.3 Repeat Elements	35
1.2.4 Segmental duplications	35
1.2.5 Pseudogenes	36
1.2.6 Gene content	36
1.2.7 Viewing genome sequence information	37
1.3 The human transcriptome.....	38
1.3.1 Non-coding genes	38
1.3.2 Protein Coding Genes.....	39
1.4 Gene identification.....	40
1.4.1 Using computational analysis to identify genes.....	43
1.4.2 Gene expression analysis aids gene identification.....	46
1.4.3 Completion of gene structures.....	47
1.5 Messenger RNA splicing	48
1.5.1 Mechanisms of mRNA splicing	49
1.5.2 Alternative mRNA splicing	51
1.5.3 Regulation of splicing	52
1.5.4 cDNA and EST sequences facilitate the identification and characterisation of alternative splicing events	55
1.6 Functional consequences of alternative splicing	58

1.6.1	Expression profiling of alternative splice forms.....	58
1.6.2	Tissue specific regulation of alternative splicing.	61
1.6.3	Evolutionary Conservation of Alternative splicing	62
1.6.4	Aberrant mRNA splicing and the role of transcript variation in disease.	65
1.6.5	Tools to study isoform function.....	66
1.6.6	Nonsense mediated decay	68
1.7	Aims of this thesis	71
Chapter 2: Materials and Methods		73
2.1	Chemical Reagents	74
2.2	Enzymes and commercially prepared kits	74
2.3	Hybridisation membranes, and X-ray and photographic film	75
2.4	Solutions and buffers.....	75
2.4.1	Buffers	75
2.4.2	Northern blotting solutions	76
2.4.3	Electrophoresis solutions and Western blotting solutions	77
2.4.4	Immunofluorescence solutions.....	77
2.4.5	Media.....	77
2.4.6	General DNA preparation solutions.....	78
2.5	Size Markers	78
2.6	<i>E. coli</i> strains.....	78
2.7	Mammalian Cell Lines	79
2.8	RNA samples	79
2.8.1	Sources of human total RNA.....	79
2.8.2	Additional sources of total RNA.....	80
2.9	cDNA libraries	80
2.10	Primer sequences	81
2.11	Key World Wide Web addresses.....	82
2.12	Mammalian cell culture	83
2.12.1	Growing and harvesting cells.....	83
2.12.2	Transfection	83
2.13	RNA manipulation	84
2.13.1	Preparation of RNA from cellular extracts.....	84
2.13.2	DNase treatment of total RNA	84
2.13.3	cDNA synthesis.....	84

2.13.4 Northern blotting.....	85
2.14 DNA manipulation	86
2.14.1 Purification of DNA	86
2.14.2 Alkaline phosphatase treatment of DNA.....	86
2.14.3 Restriction digests	87
2.14.4 Mini-preps of plasmid DNA.....	87
2.14.5 Midi and maxi preps of plasmid DNA	87
2.14.6 Quantification	87
2.15 Polymerase Chain Reaction.....	88
2.15.1 STS Pre-Screen.....	89
2.15.2 cDNA library screening	89
2.15.3 Vectorette PCR	89
2.15.4 Colony PCR	90
2.15.5 RT-PCR.....	90
2.15.6 Quantitative PCR.....	91
2.16 Electrophoretic analysis of DNA, RNA and proteins.....	92
2.16.1 Agarose electrophoresis.....	92
2.16.2 SDS-PAGE	93
2.17 Bacterial cloning	93
2.17.1 Preparation of chemically competent <i>E.coli</i>	93
2.17.2 Subcloning.....	93
2.17.3 Transformations	94
2.18 Identification of transcript variants.....	94
2.19 <i>PQBP1</i> open reading frame cloning.....	94
2.19.1 A-tailing of purified PCR products.....	95
2.19.2 Ligation into pGEM®T-Easy.....	95
2.20 Preparation of constructs for immunofluorescence	95
2.20.1 N-terminal T7 tag	95
2.20.2 C-terminal T7 tag	96
2.20.3 Non-directional cloning	96
2.21 Transcript stability assays	96
2.21.1 Preparation of inserts	96
2.21.2 Preparation of constructs.....	96
2.21.3 Ligation of Vector and DNA.....	97
2.21.4 Transfection of mammalian cells.....	97
2.21.5 Time course experiment	97

2.21.6	Real-time PCR analysis	97
2.21.7	Analysis of results	98
2.22	Translation inhibition time course analysis	98
2.22.1	Real-time PCR analysis	98
2.23	Detection of luciferase activity	99
2.23.1	Preparation of cell lysates	99
2.23.2	Luciferase activity assay	99
2.23.3	Bradford assay	99
2.24	Western blotting	99
2.24.1	Preparation of samples for Western blotting.	99
2.24.2	Electrophoresis of proteins using SDS-PAGE	100
2.25	Intracellular localisation	100
2.25.1	Transfection of mammalian cells for immunofluorescence. ...	100
2.25.2	Fixation of cells	100
2.25.3	Antibody staining and visualisation	101
2.26	Computational Analysis	101
2.26.1	Xace	101
2.26.2	Blixem.....	102
2.26.3	RepeatMasker	102
2.26.4	GAP4	103
2.26.5	Perl scripts	103
2.26.6	Emboss	103
2.26.7	Excel analysis	103
2.27	Sequence Analysis	104
2.27.1	Transcript annotation	104
2.27.2	Alignment of nucleic acid and protein sequences	105
2.27.3	Calculation of sequence identities and similarities	105
2.27.4	Phylogenetic analysis of protein sequences	105
2.28	Comparative sequence analysis	106
2.28.1	zPicture	106
Chapter 3: Gene annotation and analysis of human Xp11.22-p11.3		107
3.1	Introduction	108
3.1.1	Xp11.22-p11.3	108
3.2	Sequence analysis	112
3.2.1	Repeat analysis	113

3.2.2	G+C content.....	113
3.3	Annotation of transcripts mapped to human Xp11.22-p11.3.	114
3.3.1	Annotation of known genes.....	115
3.3.2	Annotation of novel transcripts in Xp11.22-p11.3	124
3.3.3	Annotation of pseudogenes.....	124
3.3.4	Distribution of genes	127
3.4	Experimental verification of novel and putative genes.	129
3.4.1	Antisense transcripts	136
3.4.2	Transcript features	137
3.5	Assessing the completeness of annotated genes	137
3.5.1	Polyadenylation sites	138
3.5.2	Transcription start sites.....	141
3.5.3	Experimental evidence confirms the transcript size of RP11- 339A18.4.....	144
3.5.4	Alternative splicing	146
3.6	Duplication events.....	147
3.6.1	Duplication of the <i>SSX</i> family.	147
3.6.2	Genes located in an inverted repeat.....	150
3.7	Discussion	152
Chapter 4:	Identification of alternative transcripts in human Xp11.23	157
4.1	Introduction	158
4.2	Using mouse transcript and genome information to identify additional alternative transcripts.....	163
4.2.1	Identification of mouse orthologues.....	163
4.2.2	Features of interest in the mouse genome.	166
4.2.3	Comparison of human and mouse transcript variants.	167
4.3	Identification of novel transcripts for human Xp11.23 by cDNA screening and sequencing	172
4.3.1	Primer Design	172
4.3.2	Optimisation of PCR conditions	172
4.3.3	cDNA screening	173
4.3.4	Cloning and sequencing.	173
4.3.5	Identification of novel <i>RBM3</i> transcripts.....	175
4.3.6	Identification of novel <i>PQBP1</i> transcripts	181

4.3.7 Tissue specific amplification profiles	188
4.3.8 Summary of results.....	190
4.4 Analysis of alternative splicing events	199
4.4.1 Describing the variation in transcript structures	199
4.4.2 Location of transcript variation.....	199
4.4.3 Analysis of exon junctions	200
4.4.4 The association of transcript diversity with other features.	205
4.5 Discussion	208
Chapter 5: <i>PQBP1</i> transcript diversity: comparative and expression analysis	214
5.1 Introduction	215
5.1.1 <i>PQBP1</i>	215
5.2 Identifying additional <i>PQBP1</i> transcripts by random open reading frame cloning	219
5.2.1 Amplification, cloning and identification of alternative variants.	219
5.2.2 Description of transcript variants	224
5.2.3 Analysis of splice site scores for <i>PQBP1</i> alternative transcripts	224
5.3 Comparative analysis of <i>PQBP1</i> locus.....	226
5.3.1 Comparative gene analysis	226
5.3.2 Phylogenetic analysis of <i>PQBP1</i> peptide sequences.	232
5.3.3 Comparative genome analysis of the <i>PQBP1</i> locus in eight vertebrate species.	233
5.3.4 Sequence variation around splice sites	236
5.3.5 Conservation of exon 2a	239
5.4 Expression profiling of <i>PQBP1</i>	240
5.4.1 Known data on tissue-expression of <i>PQBP1</i>	240
5.4.2 Analysis of <i>PQBP1</i> gene expression by quantitative PCR.....	242
5.4.3 Primer Design	243
5.4.4 Sensitivity, linearity and amplification efficiencies.....	248
5.4.5 Quantitation of reference <i>PQBP1</i>	249
5.4.6 Expression profiling and quantification of alternative variants	251
5.5 Discussion	254
5.5.1 Comparative sequence analysis highlights potential causes for <i>PQBP1</i> transcript variation.....	254

5.5.2	Expression studies of <i>PQBP1</i>	257
5.5.3	Conclusions.....	259
Chapter 6:	Functional analysis of <i>PQBP1</i> transcript variants	261
6.1	Introduction	262
6.2	Identification of open reading frames in <i>PQBP1</i> alternative transcripts.....	263
6.2.1	Predicted domains found in <i>PQBP1</i> variant proteins	265
6.3	Intracellular localisation of <i>PQBP1</i> transcripts.....	268
6.3.1	Preparation of constructs.....	268
6.3.2	Western blot analysis of <i>PQBP1</i> -T7 fusion proteins.....	271
6.3.3	Subcellular localisation of <i>PQBP1</i> transcripts in cos-7 and CHO cells.	273
6.4	Analysis of mRNA stability of <i>PQBP1</i> transcript variants.	276
6.4.1	Preparation of constructs.....	277
6.4.2	Optimisation of experimental protocol.....	279
6.4.3	Quantitative analysis of mRNA stability	280
6.5	Degradation of alternatively spliced <i>PQBP1</i> transcripts by nonsense mediated decay.	288
6.6	Discussion	294
6.6.1	Splicing patterns affect the sub-cellular location of <i>PQBP1</i> transcript variants.	294
6.6.2	Comparison of methods used to determine mRNA stability.	295
6.6.3	Not all <i>PQBP1</i> transcripts containing a PTC are targeted for rapid degradation.	297
6.6.4	Conclusions.....	300
Chapter 7:	Discussion.....	301
7.1	Summary	302
7.2	The human genome sequence and alternative splicing	304
7.3	Future directions	310
Bibliography		315
Appendices		354

List of Figures

Figure 1.1 mRNA splicing.....	51
Figure 1.2 Types of splicing variation.....	52
Figure 1.3 Overview of nonsense mediated decay.....	70
Figure 3.1 The human X chromosome and clones mapped to human Xp11.22-p11.3	111
Figure 3.2 Visualisation of sequence data in ACeDB.....	112
Figure 3.3 Visualisation of sequence alignments using BLIXEM.	115
Figure 3.4 Example of a known gene structure, <i>PCSK1N</i>	116
Figure 3.5 Screening of fosmid for 5' end of <i>GRIPAP1</i>	118
Figure 3.6 Diagram illustrating a “pseudogene” (pseudogene structure), for the locus bA198M15.1.....	125
Figure 3.7 Chromosomal origin of pseudogenes identified in human Xp11.22- p11.3.....	126
Figure 3.8 Gene annotation of human Xp11.22-p11.3.....	128
Figure 3.9 Success rates at various stages of analysis to attempt to confirm and extend novel and putative gene structures.	131
Figure 3.10 Confirmation and extension of novel cds RP11-54B20.4.	132
Figure 3.11 Annotation of <i>RBM3</i> and its antisense gene <i>AC1145L6.5</i> on the genome sequence, AC115618.	136
Figure 3.12 Analysis of polyadenylation signals in human Xp11.22-p11.3.....	140
Figure 3.13 Distribution of predicted transcript start sites (TSSs) for genes and pseudogenes in human Xp11.22-p11.3.	143
Figure 3.14 State of completion of annotated gene structures.....	144
Figure 3.15 Confirmation of gene expression, <i>RP11-339A18.4</i>	145
Figure 3.16 Transcript variants of the gene <i>UBE1</i>	146
Figure 3.17 Annotation of <i>SSX</i> gene family members.	149
Figure 3.18 Gene duplication in Xp11.23.....	151
Figure 4.1 Experimental strategy to identify novel transcripts in human Xp11.23..	159
Figure 4.2 Order and orientation of genes involved in this study	160
Figure 4.3 Annotation of mouse orthologues	165
Figure 4.4 Annotation of an antisense transcript to <i>Wdr13</i> in the mouse.	166
Figure 4.5 Annotation of EST sequences spanning <i>PIM2</i> and <i>DXIm46e</i>	167

Figure 4.6 Number of alternative transcripts that were annotated for each of the gene pairs.	168
Figure 4.7 Identification of novel transcribed regions using comparative analysis..	171
Figure 4.8 Transcript profiling of <i>RBM3</i> in 29 different tissues.....	176
Figure 4.9 cDNA screens of <i>RBM3</i>	177
Figure 4.10 Sequencing of <i>RBM3</i> fragments.....	178
Figure 4.11 Tissue specific expression patterns of <i>RBM3</i>	180
Figure 4.12 The location of primers used to screen for novel <i>PQBP1</i> transcripts.	182
Figure 4.13 Identification of novel transcripts for <i>PQBP1</i> - cDNA screens	184
Figure 4.14 The novel exon 2a lies within an <i>Alu</i> repeat	186
Figure 4.15 Identification of a 21 bp deletion that is exclusive to the adrenal gland sample	187
Figure 4.16 Example of cDNA screens that displayed tissue specific expression profiles	189
Figure 4.17 Summary of the number of novel transcripts identified by cDNA screening for 18 genes in Xp11.23	190
Figure 4.18 Transcript structures for genes in human Xp11.23.....	191
Figure 4.19 Types of alternative splicing events observed in 18 genes from human Xp11.23.....	199
Figure 4.20 Classification of splice site sequences	201
Figure 4.21 Splice site scores for 17 human genes in Xp11.23.	203
Figure 4.22 Correlation of gene features with transcript numbers	207
Figure 5.1 <i>PQBP1</i> and its neighbours.....	215
Figure 5.2 Overview of the cloning protocol.....	221
Figure 5.3 Number of <i>PQBP1</i> transcripts cloned for each variant.	222
Figure 5.4 Exon/intron structures of <i>PQBP1</i> alternative transcripts.....	223
Figure 5.5 Identification of a <i>PQBP1</i> like pseudogene in <i>C. familiaris</i>	229
Figure 5.6 Potential duplication of <i>PQBP1</i> locus in <i>F. rubripes</i>	231
Figure 5.7 Phylogenetic tree of <i>PQBP1</i> peptides.	232
Figure 5.8 Comparative Genome Analysis of the <i>PQBP1</i> locus.	235
Figure 5.9 Sequence variations around exon, intron junctions in different species.	237
Figure 5.10 Multiple sequence alignments of splice sites used in alternative transcripts.	238

Figure 5.11 Global alignment of genome sequences containing the gene <i>PQBP1</i> using MultiContigView at Ensembl.	239
Figure 5.12 EST expression profile for <i>PQBP1</i>	241
Figure 5.13 Expression profile of <i>PQBP1</i> as extracted from the Gene Expression Atlas.	242
Figure 5.14 Location of primers used to determine the abundance of <i>PQBP1</i> alternative transcripts.	245
Figure 5.15 Specificity of <i>PQBP1</i> alternative transcript primers.	246
Figure 5.16 Effect of yeast cDNA on real-time PCR amplification efficiency of <i>PQBP1</i> transcripts.	248
Figure 5.17 Quantification of <i>PQBP1</i> alternative transcripts by real-time PCR. .	249
Figure 5.18 Comparison of amplification efficiencies for the primer pairs <i>GAPDH</i> and <i>PQBP1.Q10</i>	250
Figure 5.19 Relative abundance of <i>PQBP1</i> expression.	251
Figure 5.20 Relative abundances of <i>PQBP1</i> transcript variants.	252
Figure 6.1 Identification of open reading frames in <i>PQBP1</i> transcripts.	264
Figure 6.2 Predicted motif patterns in <i>PQBP1</i> putative proteins.	266
Figure 6.3 Prediction of subcellular localisation of <i>PQBP1</i> isoforms.	266
Figure 6.4 Schematic of subcellular localisation protocol.	270
Figure 6.5 Western Blot analysis confirms <i>PQBP1</i> expression for some, but not all constructs.	272
Figure 6.6 Localisation of <i>PQBP1</i> alternative isoforms in Cos-7 cells.	274
Figure 6.7 Schematic of gene regulation on the BD™ Tet-Off System.	277
Figure 6.8 Preparation of constructs to study <i>PQBP1</i> mRNA decay rates.	278
Figure 6.9 Preparation and analysis of pTRE-TIGHT- <i>PQBP1</i> plasmids.	279
Figure 6.10 Dose response curve for the CHO-AA8-Luc Control cell line.	280
Figure 6.11 Schematic of experimental protocol used to assess mRNA decay rates.	282
Figure 6.12 Analysis mRNA decay by real-time PCR.	285
Figure 6.13 Effect on <i>PQBP1</i> transcript levels of inhibiting protein translation using anisomycin, cycloheximide and puromycin.	291

List of Tables

Table 1.1 Full-length cDNA sequencing projects.	42
Table 1.2 Type of BLAST (or BLAST like analysis) used in gene annotation.....	45
Table 1.3 A comparison of alternative splicing databases.....	57
Table 2.1 Strains of <i>E. coli</i> used in this study	79
Table 2.2 Mammalian cell lines used in this study.	79
Table 2.3 cDNA samples used in this study.....	80
Table 2.4 cDNA libraries used in this study.	81
Table 2.5 Key World Wide Web addresses used in this study	82
Table 2.6 Amounts of DNA and reagents used in mammalian cell transfections....	84
Table 2.7 Real time PCR control primers.....	92
Table 2.8 Emboss applications used in this study	104
Table 3.1 Repeat content of human Xp11.23	113
Table 3.2 “Orphan” or “short” fosmids that could harbour the 5’ end of GRIPAP1...	118
Table 3.3 Known genes annotated in human Xp11.22- p11.3	119
Table 3.4 Processed pseudogenes located in Xp11.22-Xp11.3	126
Table 3.5 Experimental verification of novel genes and transcripts and putative Genes	133
Table 3.6 Overview of transcript features annotated in human Xp11.22-Xp11.3..	137
Table 4.1 Number of alternative transcripts annotated for each gene between the markers between markers <i>DXS6941</i> and <i>DXS9784</i>	161
Table 4.2 Sequence identity of human and mouse orthologues.	164
Table 4.3 Analysis of the conservation of alternative exon junctions and first and last exons in human and mouse genes	169
Table 4.4 Mouse specific exons that are conserved in human	170
Table 4.5 Summary of cDNA screening, ligations and sequencing reactions for each gene analysed in this study	174
Table 4.6 Summary of <i>RBM3</i> transcript variants	179
Table 4.7 Summary of <i>PQBP1</i> transcript variants.	185
Table 4.8 Variation in splice site sequences for alternatively spliced exons.....	202
Table 5.1 <i>PQBP1</i> associated X-linked mental retardation phenotypes.	217
Table 5.2 Description of <i>PQBP1</i> transcript variants.....	222
Table 5.3 Species and sequence assembly versions used in comparative analysis.	226
Table 5.4 Identification of homologues of the <i>PQBP1</i> locus in other vertebrates.	227

Table 5.5 Exon/Intron Structure of <i>PQBP1</i> homologues.....	228
Table 5.6 Potential <i>PQBP1</i> retroposed pseudogenes	229
Table 5.7 Sequence assemblies and chromosome co-ordinates of genomic sequences used for comparative analysis.....	233
Table 5.8 Primer sequences and additional details of primers used in quantitative analysis of alternative <i>PQBP1</i> transcripts.....	245
Table 5.9 Tissue samples with statistically significant differences variations in transcript abundance.	253
Table 6.1 Predictions regarding the nuclear localisation of <i>PQBP1</i> isoforms.....	267
Table 6.2 Expected sizes of T7 tagged proteins.....	272
Table 6.3 Analysis of mRNA decay obtained from <i>PQBP1</i> alternative transcripts..	284
Table 6.4 Antibiotics used to inhibit translation in HEK293 cells.....	289
Table 6.5 Primer pairs used to quantify alternative <i>PQBP1</i> transcripts.	290
Table 6.6 Summary of results obtained in this chapter.	297

List of Appendices

APPENDIX I	Primers used to screen vectorette cDNA libraries for the expression of novel genes.....	355
APPENDIX II	Primers used to identify transcript variants	356
APPENDIX III	Primer pairs used to screen human cDNA samples for novel transcript variants.....	360
APPENDIX IV	Transcript variants identified for 18 genes in human Xp11.23.	364
APPENDIX V	Multiple sequence alignment of <i>PQBP1</i> transcripts	370
APPENDIX VI	Primers used to quantify <i>PQBP1</i> alternative transcripts	377
APPENDIX VII	Multiple sequence alignment of <i>PQBP1</i> peptides.....	378
APPENDIX VIII	Primer combinations and sequences used in the preparation of T7 epitope:: <i>PQBP1</i> (variant) pCDNA.3 constructs.....	380
APPENDIX IX	Control primers used in real-time PCR analysis	381

Abbreviations

1st EF	First exon finder
ABI	Applied Biosystems
ACeDB	<i>A. C. elegans</i> database
ACTB	beta actin
AEDB	alternative exon database
ANS	anisomycin
ASAP	the Alternative Splicing Annotation Project
ASD	alternative splicing database project
ASDB	alternative splicing database
ASG	alternative splicing gallery
AT	annealing temperature
ATP	adenosine 5'-triphosphate
BAC	bacterial artificial chromosome
bis-acrylamide	(N, N'-methylene-)bis-acrylamide
BLAST (-n -p)	basic local alignment search tool (-nucleotide -protein)
BLAT	basic local alignment tool
Blixem	BLast matches In an X-windows Embedded Multiple alignment
BMD	Becker muscular dystrophy
B-ME	B-mercaptoethanol
bp	base pair
BSA	bovine serum albumin
°C	degrees Celsius
cDNA	complementary DNA
CDS (c-)	coding sequence (consensus)
CEN	centromere
CHX	cycloheximide
CNS	central nervous system
CpG	cytidyl phosphoguanosine dinucleotide
Cps	counts per second
C _T	cycle threshold
CT- antigen	cancer testis antigen
CTP (d-)	cytidine 5'-triphosphate (deoxy-)
DAPI	4',6-Diamidino-2-phenylindole
dbEST	database of expressed sequence tags

DDBJ	DNA DataBase of Japan
DEPC	diethyl pyrocarbonate
DKFZ	Deutsches Krebs Forschung Zentrum (German cancer research centre)
DMD	Duchenne muscular dystrophy
DMSO	dimethyl sulphoxide
DNA	deoxyribonucleic acid
Dnase I	deoxyribonuclease A
dNTP	deoxyribonucleotide 5' triphosphate
dsDNA	double-stranded deoxyribonucleic acid
dsRNA	double-stranded ribonucleic acid
EBI	European Bioinformatic Institute
ECL	enhanced chemiluminescence
ECR	evolutionary conserved region
EDTA	ethylenediamine tetra-acetic acid
EJC	exon junction complex
EMBL	European Molecular Biology Laboratory
EMBOSS	European Molecular Biology Open Software Suite
ePCR	electronic polymerase chain reaction
ES cell	embryonic stem cell
ESE	exon splicing enhancer
ESS	exon splicing silencer
EST	expressed sequence tag
FITC	fluorescein isothiocyanate
g	gram
<i>g</i>	force of gravity (relative centrifugal force)
<i>GAPDH</i>	Glyceraldehyde-3-phosphate dehydrogenase
G banding	Geisma banding
GTP (d-)	guanosine 5'-triphosphate (deoxy-)
HASDB	human alternative splicing database
HAVANA	human and vertebrate genome annotation and analysis
HEPES	N-[2-hydroxyethyl]piperazin-N'-[2-ethansulphonic acid]
HGMP	Human Genome Mapping Project
HGNC	HUGO Gene Nomenclature Committee
HGP	Human Genome Project
hnRNP	heterogenous nuclear ribonucleoprotein
HRP	horseradish peroxidase

HMM	hidden Markov model
HS	hierarchical shotgun
HUGO	Human genome organisation
ICI	Imperial Chemical Industries
INDELs	insertion or deletion of deoxyribonucleic acid
IHGSC	International Human Genome Sequencing Consortium
IPTG	isopropyl β -D-thiogalactoside
kb	kilobase pairs
kDa	kilodalton
l	litre
LB	Luria-Bertani
LINE	long interspersed nuclear element
LTR	long terminal repeat
M	molar
Mb	megabase pairs
MER	medium reiterative repeat
MGC	Mammalian Gene Collection
μ g	microgram
μ l	microlitre
μ M	micromolar
min(s)	minute(s)
mg	milligram
ml	millilitre
mM	millimolar
mm	millimetre
mya	million years ago
NCBI	National Centre for Biotechnology Information
NEDO	New Energy and Industrial Technology Development Organisation
ng	nanogram
nm	nanometre
NMD	nonsense mediated decay
O/N	overnight
oligo-dT	deoxyribothymidyl oligonucleotide
OMIM	On-line Mendelian Inheritance in Man
ORF	open reading frame
PAC	P1-derived artificial chromosome

PAGE	polyacrylamide gel electrophoresis
PBS (-T)	phosphate buffered saline (-Tween 20)
PCR	polymerase chain reaction
RASL	RNA annealing selection and ligation
pg	picogram
PALSdb	putative alternative splicing database
PE	Perkin Elmer
poly(A)	polyadenylation
PTC	premature termination codon
PUR	puromycin
RNA	ribonucleic acid
RNAi	ribonucleic acid interference
mRNA	messenger ribonucleic acid
rRNA	ribosomal ribonucleic acid
snRNA	small nucleolar ribonucleic acid
hnRNA	heteronuclear ribonucleic acid
hnRNP	heteronuclear ribonucleoprotein
snRNP	small nucleolar ribonucleoprotein
dsRNA	double-stranded ribonucleic acid
siRNA	short-interfering ribonucleic acid
tRNA	transfer ribonucleic acid
miRNA	micro-ribonucleic acid
RISC	ribonucleic acid induced silencing complex
RNase A	ribonuclease A
rpm	revolutions per minute
RT	room temperature
RT-PCR	reverse transcription polymerase chain reaction
SAGE (Long-)	serial analysis of gene expression
SDS	sodium dodecyl sulphate
s	seconds
siRNA	short interfering RNA
SINE	short interspersed nuclear element
SNP	single nucleotide polymorphism
SSAHA	sequence Search and Alignment by Hashing Algorithm
ssDNA	single-stranded ribonucleic acid
SSH	suppressive subtractive hybridisation

SSX	synovial sarcoma X-linked
STS	sequence tagged site
TAP	transcript assembly programme
tBLAST -n, -x	translated blast local alignment sequence tool -vs nucleotide database -vs protein database
TLC	thin layer chromatography
TEL	telomere
TEMED	N,N,N',N'-tetramethylethylenediamine
TFB (-I, -II)	Transformation buffer -I, -II
TIGR	the Institute for Genomic Research
Tris	tris(hydroxymethyl)aminoethane
TSS	transcription start site
TTP	thymidine 5'-triphosphate
TTS	transcription termination site
U	unit
UCSC	University of California, Santa Cruz
UTR	untranslated region
uv	ultraviolet
V	volt
VEGA	vertebrate genome annotation
v/v	volume/volume
W	watt
w/v	weight/volume
WAS	Wiskott Aldrich syndrome
Wash U.	Washington University
WGS	whole genome shotgun
WTSI	Wellcome Trust Sanger Institute
Xace	X chromosome version of ACeDB
X-gal	5-bromo-4-chloro-3-indolyl-B-D-galactoside
XIC	X-inactivation centre
Xist	X inactive specific transcript
YAC	yeast artificial chromosome