

A Survey of RNA Editing in the Human Brain

Matthew James Blow

Darwin College

October 2004

This dissertation is submitted for the degree of Doctor of Philosophy

DECLARATION

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. The dissertation does not exceed the word limit set by the Biology Degree Committee.

ACKNOWLEDGEMENTS

I wish to sincerely thank Mike Stratton for his enthusiasm, patience and wisdom throughout this project. I have been fortunate to have him as a supervisor. Thanks also to the entire cancer genome project for their help in many ways in the work that led to this thesis, but mostly for their good company. I am particularly grateful to my supervisors Richard Wooster and Andy Futreal for their guidance and sound advice. In addition, I would like to thank many other members of the Sanger institute, Nav Navaratnam from the MRC Clinical Sciences Centre at Hammersmith Hospital and Nick Coleman from the Hutchison/MRC Research Centre at Addenbrooke's Hospital for their useful comments. Finally, I would like to thank Trish and my family, especially Mum, Dad and Sarah, for their enduring love and encouragement.

ABSTRACT

RNA editing is a post-transcriptional modification of RNA that occurs in prokaryotes, plants and animals. It occurs by a range of mechanisms including nucleotide insertions and deletions and base substitutions. The aim of these studies was to provide an extensive and systematic survey of the classes and distribution of editing in human mRNA. More than 3Mb of sequence from a human brain cDNA library were compared to genomic DNA sequences from the same individual and to the reference human genome sequence. Approximately 1 in 2,000 nucleotides in the RNA sample from which the library was constructed were shown to be edited. All edits were adenosine to inosine (A > I), predominantly in Alu repeats in intronic and non-coding RNAs. No edits were found in coding sequence. Analysis of the genome in the vicinity of edited sequences strongly supports the notion that formation of intramolecular double-stranded RNA (dsRNA) by inverted sequence copies underlies most A>I editing. The likelihood of editing is increased by presence of the two inverted copies within the same intron, proximity of the two copies and a high local density of inverted copies. A > I editing exhibits some sequence specificity, and is less likely at an adenosine 3' to a guanosine and more likely at an adenosine 5' to a guanosine. Simulation of the dsRNA molecules that underlie known edits indicates that there is a greater likelihood of A > I editing at A:C mismatches than at other mismatches or at A:U matches. However, because A:U matches in dsRNA are more common than all mismatches, overall the likely effect of editing is to increase the number of mismatches in dsRNA. The potential functions of A>I RNA editing have been considered in the light of this survey.

TABLE OF CONTENTS

DECLARATION	2
ACKNOWLEDGEMENTS	3
ABSTRACT	4
TABLE OF CONTENTS	5
LIST OF TABLES	11
LIST OF FIGURES	12
1 INTRODUCTION	14
1.1 GENERAL INTRODUCTION	14
1.2 THE HUMAN GENOME	15
1.2.1 Transposable elements in the human genome	17
1.3 INTRODUCTION TO RNA	20
1.3.1 Messenger RNA (mRNA)	22
1.3.2 Ribosomal RNA (rRNA)	23
1.3.3 Transfer RNA (tRNA)	24
1.3.4 Spliceosomal RNAs (snRNAs)	25
1.3.5 Small nucleolar RNAs (snoRNAs)	25
1.3.6 Miscellaneous non-coding RNAs	26
1.3.7 Double-stranded RNA (dsRNA)	Error! Bookmark not defined.
1.4 GENERAL INTRODUCTION TO RNA EDITING	31
1.4.1 RNA editing of tRNA in <i>Escherichia coli</i>	33
1.4.2 RNA editing of Paramyxovirus RNA by polymerase stuttering	

1.4.3	Guided uridylyate insertion and deletion RNA editing in Trypanosome kinetoplasts	34
1.4.4	Nucleotide insertion and nucleotide substitution RNA editing in <i>Physarum polycephalum</i> mitochondria	35
1.4.5	Nucleotide substitution RNA editing in yeast	36
1.4.6	Nucleotide substitution RNA editing in Plant organelles	36
1.4.7	Nucleotide substitution RNA editing of <i>Caenorhabditis elegans</i> RNAs	37
1.4.8	Nucleotide substitution RNA editing in <i>Drosophila melanogaster</i>	38
1.4.9	Nucleotide substitution RNA editing in squid	40
1.4.10	Nucleotide substitution RNA editing in <i>Xenopus laevis</i>	40
1.4.11	Nucleotide substitution RNA editing of mammalian RNAs	40
1.5	RNA EDITING IN HUMANS	41
1.5.1	Human A > I RNA editing enzymes	45
1.5.2	Human A > I editing substrates	49
1.5.3	The function of A > I editing	55
1.5.4	A > I editing and human disease	56
1.5.5	Human C > U RNA editing enzymes	57
1.5.6	Human C > U editing substrates	60
1.5.7	C > U editing and disease	62
1.5.8	Rare RNA edits of other classes	63
1.6	PROJECT INTRODUCTION	65
2	METHODS	66
2.1	LABORATORY METHODS	66
2.1.1	Construction of a human cerebral cortex cDNA library	66
2.1.2	Sequencing of cDNA clones	67
2.1.3	Sequencing of PCR and RT-PCR products	69
2.2	COMPUTATIONAL METHODS	72
2.2.1	Programs and databases	72

2.2.2	Custom Perl programs	73
2.2.3	Detection of high quality sequence variants	77
2.2.4	Analysis of edited Alu sequences	81
3	SEQUENCING AND EVALUATION OF A HUMAN BRAIN cDNA LIBRARY	83
3.1	INTRODUCTION	83
3.2	RESULTS	84
3.2.1	Construction of a human brain cDNA library	84
3.2.2	Evaluation of the cDNA library	86
3.2.3	Sequencing of 10,000 clones from a human brain cDNA library	88
3.2.4	Automated alignment of 9,341 cDNA clones to the human genome reference sequence	89
3.2.5	Evaluation of cDNA library composition by the genomic distribution of cDNA clones	96
3.2.6	Evaluation of cDNA library by annotation of known genes	101
3.3	DISCUSSION	107
3.3.1	Choice of experimental strategy for a survey of RNA editing	107
3.3.2	Choice of tissue for a survey of RNA editing	109
3.3.3	Extent to which the cDNA library is representative of the human brain transcriptome	110
3.3.4	Sequence class composition of the cDNA library	111
4	IDENTIFICATION OF NOVEL RNA EDITS IN HUMAN BRAIN	113
4.1	INTRODUCTION	113
4.2	RESULTS	113
4.2.1	Computational detection of high quality candidate RNA edits from human brain cDNA	113

4.2.2	Extensive A > I RNA edits but no other class of RNA edits are present in human brain cDNA	115
4.2.3	A > G / T > C variants are all likely A > I edits	122
4.2.4	RNA editing is absent from mitochondrial transcripts in human brain	123
4.2.5	The estimated frequency of RNA editing in the human brain	126
4.3	DISCUSSION	127
4.3.1	Classes of RNA editing in the human brain	127
4.3.2	Frequency of RNA editing in the human brain	128
5	THE CHARACTERISTICS OF A > I EDITED TRANSCRIPTS FROM HUMAN BRAIN	129
5.1	INTRODUCTION	129
5.2	RESULTS	130
5.2.1	A > I RNA editing targets a wide variety of human brain transcripts	130
5.2.2	A > I RNA editing is predominantly in non-coding RNA	131
5.2.3	RNA editing of translated exons is a rare event in human brain	133
5.2.4	A > I RNA editing is associated with Alu repeat sequences	135
5.2.5	The presence of an anti-sense repeat in the same transcript increases the likelihood of RNA editing of Alu sequences	137
5.2.6	The presence of an anti-sense Alu in the same intron increases the likelihood of RNA editing	139
5.2.7	The proximity of inverted Alu sequence influences the likelihood of RNA editing	141
5.2.8	The amount of inverted Alu sequence is associated with the likelihood of RNA editing	144
5.2.9	The orientation of Alus with respect to transcription has no impact on RNA editing	147

5.2.10	The orientation of Alus with respect to each other has no impact on RNA editing	147
5.2.11	Further analysis of Alus that have an inverted repeat in the same intron but are apparently unedited	150
5.2.12	The genome wide distribution of inverted Alus within 2kb in the same intron	151
5.2.13	The role of dsRNA formation in non-Alu edited sequences.	152
5.3	DISCUSSION	154
5.3.1	Sequence class composition of RNA editing substrates	154
5.3.2	Association of RNA editing with repeat sequences	155
5.3.3	The role of dsRNA formation in RNA editing	156
5.3.4	Edited Alus with no inverted copy in the same intron	158
5.3.5	Unedited Alus with an inverted copy in the same intron	158
5.3.6	RNA editing of non-Alu repeat sequences	159
6	THE ROLE OF LOCAL SEQUENCE EFFECTS IN RNA EDITING	161
6.1	INTRODUCTION	161
6.2	RESULTS	162
6.2.1	Local sequence preferences A > I RNA editing	162
6.2.2	BLAST alignment of inverted Alus indicates base-pairing preferences for A > I RNA editing	165
6.2.3	Alu multiple sequence alignments indicate base-pairing preferences for A > I RNA editing	167
6.2.4	A > I RNA editing results in a marginal decrease in base pairing in predicted dsRNA	169
6.2.5	Distribution of A > I editing sites in the Alu consensus sequence	170
6.3	DISCUSSION	173
6.3.1	Local sequence preferences of Alu A > I editing	173
6.3.2	Distribution of A > I edits in the Alu consensus sequence	174
6.3.3	Base-pairing preferences of Alu A > I editing	175

6.3.4 The overall effect of A > I editing on base-pairing in dsRNA177

7	GENERAL DISCUSSION	178
	7.1 FUTURE CHALLENGES	178
	7.2 THE FUNCTION OF A > I EDITING	180
8	REFERENCES	186

LIST OF TABLES

Table 1-1	Characteristics of human protein coding genes	16
Table 1-2	The repeat composition of the human genome	18
Table 1-3	The major families of non-coding RNA found in eukaryotic cells	21
Table 1-4	Overview of the dominant types and targets of RNA editing	32
Table 1-5	Known RNA edits in human transcripts	44
Table 3-1	Categorisation of cDNA clone sequences based on their alignment to the human genome using BLAT.	87
Table 3-2	Evaluation of the sequence composition of a human brain cDNA library.	88
Table 3-3	Genome-wide distribution of cDNA clones.	97
Table 3-4	Classification of cDNA clones according to overlap with gene annotation in the EnSEMBL genome database.	103
Table 3-5	The 20 most commonly sequenced genes in the cDNA library.	105
Table 4-1	List of evaluated sequence variants in mitochondrial cDNA clone sequences	125
Table 4-2	Estimation of the frequency of non A > I RNA editing in the human brain.	126
Table 5-1	Distribution of RNA edits by repeat class and subclass.	136
Table 6-1	A > I editing at different RNA base pairings	165

LIST OF FIGURES

Figure 1-1 Evolution of Alu sequences	20
Figure 1-2 The effect of RNA editing on base pairing in RNA	42
Figure 2-1 Automated detection and annotation of sequence variants	79
Figure 3-1 Analysis of Human Cerebral cortex nucleic acid preparations	86
Figure 3-2 Processing of cDNA clone sequence data	91
Figure 3-3 Discrimination of identical and non-identical overlapping clones.	95
Figure 3-4 The proportion of cDNA clones derived from each chromosome	99
Figure 3-5 Mitochondrial cDNA clones	101
Figure 3-6 Sequence class composition of the cDNA library.	106
Figure 4-1 Summary of the identification of 1,727 novel A>I RNA edits	116
Figure 4-2 Confirmation of RNA editing of heavily edited sequences	119
Figure 4-3 Variants identified incorrectly by automated detection	120
Figure 5-1 Breakdown of RNA edits by gene class	131
Figure 5-2 Distribution of A > I RNA edits by sequence class	132
Figure 5-3 Summary of the analysis of the subset of 286 variants from translated exon sequence.	134
Figure 5-4 Proportion of edited and unedited Alus with additional Alus in the same intron.	139
Figure 5-5 Proportion of edited and unedited Alus from introns of different sizes	141
Figure 5-6 Proportion of edited and unedited Alus with additional Alus within 0 to 1 kb in the same intron	142
Figure 5-7 Distance from edited and unedited Alus to the nearest Alu in the same intron.	144
Figure 5-8 Amount of flanking Alu sequence at different distances from edited and unedited Alus	146
Figure 5-9 Orientation of Alu sequences with respect to each other	148

Figure 5-10 Amount of anti-sense Alu sequence at different distances from edited and unedited Alus in 'Tails-Out' or 'Tails-in' orientation	149
Figure 6-1 Sequence context of adenosines in edited Alu sequences	163
Figure 6-2 Tri-nucleotide sequence context of adenosines in edited Alu sequences	164
Figure 6-3 Effect of sequence composition on the likelihood of RNA editing	168
Figure 6-4 Frequency of editing at adenosines in edited sense and anti- sense Alus	171