

**Bangor University**

## **DOCTOR OF PHILOSOPHY**

### **How could clock gene variants contribute to the causes of autism?**

Nicholas, Bradley

*Award date:*  
2011

*Awarding institution:*  
Bangor University

[Link to publication](#)

#### **General rights**

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

#### **Take down policy**

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.



PRIFYSGOL  
**BANGOR**  
UNIVERSITY

**How could clock gene variants contribute to  
the causes of autism?**

Thesis submitted in accordance with the requirements of  
Bangor University for the degree of Doctor in Philosophy

**B. D. Nicholas**

Bangor University, UK

School of Biological Sciences

The North West Cancer Research Fund Institute

November 2011



## Declaration and Consent

### Details of the Work

I hereby agree to deposit the following item in the digital repository maintained by Bangor University and/or in any other repository authorized for use by Bangor University.

**Author Name:**.....

**Title:**.....

**Supervisor/Department:**.....

**Funding body (if any):**.....

**Qualification/Degree obtained:**.....

This item is a product of my own research endeavours and is covered by the agreement below in which the item is referred to as “the Work”. It is identical in content to that deposited in the Library, subject to point 4 below.

### Non-exclusive Rights

Rights granted to the digital repository through this agreement are entirely non-exclusive. I am free to publish the Work in its present version or future versions elsewhere.

I agree that Bangor University may electronically store, copy or translate the Work to any approved medium or format for the purpose of future preservation and accessibility. Bangor University is not under any obligation to reproduce or display the Work in the same formats or resolutions in which it was originally deposited.

### Bangor University Digital Repository

I understand that work deposited in the digital repository will be accessible to a wide variety of people and institutions, including automated agents and search engines via the World Wide Web.

I understand that once the Work is deposited, the item and its metadata may be incorporated into public access catalogues or services, national databases of electronic theses and dissertations such as the British Library’s EThOS or any service provided by the National Library of Wales.

I understand that the Work may be made available via the National Library of Wales Online Electronic Theses Service under the declared terms and conditions of use (<http://www.llgc.org.uk/index.php?id=4676>). I agree that as part of this service the National Library of Wales may electronically store, copy or convert the Work to any approved medium or format for the purpose of future preservation and accessibility. The National Library of Wales is not under any obligation to reproduce or display the Work in the same formats or resolutions in which it was originally deposited.

**Statement 1:**

This work has not previously been accepted in substance for any degree and is not being concurrently submitted in candidature for any degree unless as agreed by the University for approved dual awards.

Signed ..... (candidate)

Date .....

**Statement 2:**

This thesis is the result of my own investigations, except where otherwise stated. Where correction services have been used, the extent and nature of the correction is clearly marked in a footnote(s). All other sources are acknowledged by footnotes and/or a bibliography.

Signed ..... (candidate)

Date .....

**Statement 3:**

I hereby give consent for my thesis, if accepted, to be available for photocopying, for inter-library loan and for electronic repositories, and for the title and summary to be made available to outside organizations.

Signed ..... (candidate)

Date .....

**Statement 4:**

I agree to submit my thesis (the Work) electronically via Bangor University's e-submission system, however I **opt-out** of the electronic deposit to the Bangor University (BU) Institutional Digital Repository, the British Library ETHOS system, and/or in any other repository authorized for use by Bangor University, due to lack of permissions for use of third party material.

Signed ..... (candidate)

Date .....

**In addition to the above I also agree to the following:**

1. That I am the author or have the authority of the author(s) to make this agreement and do hereby give Bangor University the right to make available the Work in the way described above.
2. That the electronic copy of the Work deposited in the digital repository and covered by this agreement, is identical in content

to the paper copy of the Work deposited in the Bangor University Library, subject to point 4 below.

3. That I have exercised reasonable care to ensure that the Work is original and, to the best of my knowledge, does not breach any laws – including those relating to defamation, libel and copyright.
4. That I have, in instances where the intellectual property of other authors or copyright holders is included in the Work, and where appropriate, gained explicit permission for the inclusion of that material in the Work, and in the electronic form of the Work as accessed through the open access digital repository, *or* that I have identified and removed that material for which adequate and appropriate permission has not been obtained and which will be inaccessible via the digital repository.
5. That Bangor University does not hold any obligation to take legal action on behalf of the Depositor, or other rights holders, in the event of a breach of intellectual property rights, or any other right, in the material deposited.
6. That I will indemnify and keep indemnified Bangor University and the National Library of Wales from and against any loss, liability, claim or damage, including without limitation any related legal fees and court costs (on a full indemnity bases), related to any breach by myself of any term of this agreement.

Signature: .....

Date : .....

## **ACKNOWLEDGEMENTS**

My thanks must primarily go to my supervisor Dr. Thomas Caspari who has generously and enthusiastically supported my work on this thesis and to Autism Cymru, who along with the European Social Fund provided the financial backing for the work. I also thank Professor D. E. J. Linden for his advice regarding establishing the studentship and I am especially grateful to Professor Michael J. Owen, Dr. George Kirov, Dr. Varuni Rudrashingam, Dr. Dawn C. Wimpory and Dr. Susan Nash for their support.

I gratefully acknowledge the resources provided by the Autism Genetic Resource Exchange (AGRE) Consortium and the participating AGRE families. The Autism Genetic Resource Exchange is a program of Cure Autism Now and is supported, in part, by Grant MH64547 from the National Institute of Mental Health to Daniel H Geschwind (PI). Without such resources genetic studies of autism would not be possible in the manner currently conceived.

The personal story concluded by this thesis is a long one, with many contributing characters and it would be remiss of me to omit to mention those who have played an indirect but important part. So, in order of appearance, I thank: Gaynor and Hywel Edwards; Rachael Williams and Rodney Stinton; John Bryant; Nick Battey, Lilla Cooper and Nick Pipe. Special thanks are due to my wife Dr. Dawn Wimpory for drawing my attention towards autism in the first place, and later for her clinical insight and tenacity of purpose in maintaining clearly defined

How could clock gene variants contribute to the causes of autism?

standards of sample selection in the clock gene screening. As well as thanking Dawn for her clinical work and moral support I also thank my children Elen and Moi for tolerating (and occasionally not tolerating) my out-of-hours work on this thesis.

## ABSTRACT

Autism is a strongly genetic disorder where risk-conferring variation in a number of genes contributes to the phenotype. Twin studies show incomplete concordance for autism in monozygotic sibling pairs, suggesting that environmental and/or epigenetic effects also contribute to the disorder.

This thesis investigates association of clock gene variants with autistic disorder. Significant indirect positive genetic association for autistic disorder was found for two single nucleotide polymorphisms in the clock gene *PER1* and for two single nucleotide polymorphisms in the clock gene *NPAS2*.

Bioinformatics analysis of these single nucleotide polymorphisms showed: SNP rs885747 disrupts a splicing enhancer/suppressor element, SNP rs34705978 is within a differentially methylated control element, SNP rs6416892 is four nucleotides from the tissue specific binding site of sterol regulatory element binding transcription factor 2 and SNP rs1811399 alters the structure of a candidate microRNA. Investigation within the most significant haplotype in *NPAS2* highlighted a conserved circadian regulatory element (RRE) whilst that of *PER1* contained alternative and essential splice site SNPs. Analysis of genes containing conserved circadian regulatory elements, the E-box, D-box and RRE, showed that some of the strongest candidate genes for

How could clock gene variants contribute to the causes of autism?

autism, schizophrenia and bipolar disorder are likely to be circadian clock controlled genes.

Synchronization of high frequency oscillations between different brain regions is a correlate of normal brain function that is altered in autism; a role for clock genes in regulating the dendritic architecture of high frequency neural oscillators is discussed. Interplay between the molecular processes of the circadian clock, the sex determination pathway and alternative splicing is highlighted as the basis of a hypothesis suggesting how clock gene mutations might also determine short period oscillator phenotypes.

## Table of Contents

<b>DECLARATION</b>	I
<b>ACKNOWLEDGMENTS</b>	IV
<b>ABSTRACT</b>	V
<b>Table of Contents</b>	VIII
<b>List of Figures</b>	XIII
<b>List of Tables</b>	XV

### 1 Introduction.

1.1 What is autistic disorder?	Page 1
1.2 Are there biological markers for autism?	Page 2
1.3 Sleep and circadian rhythm disturbance is associated with autism.	Page 5
1.4 Autism genetics.	Page 8
1.4.1 MicroRNA genes.	Page 11
1.5 Epigenetics and autism.	Page 18
1.5.1 Epigenetic vs. conventional genetic models of autism heritability.	Page 19
1.5.2 Gene silencing by DNA methylation.	Page 25
1.5.3 MECP2: an epigenetic regulator protein.	Page 28
1.5.4 DNA methylation of clock genes in autism.	Page 30
1.6 Clock genes: The mammalian circadian molecular clock.	Page 31
1.7 The circadian clock and genome stability: autism and cancer.	Page 37
1.7.1 Clock genes and the DNA damage response.	Page 39
1.7.2 DNA damage checkpoints.	Page 41
1.7.3 Circadian control of apoptosis.	Page 42
1.7.4 The circadian clock and cancer development.	Page 44
1.7.5 Cancer and autism.	Page 48
1.8 Autism and social timing: Timing in social communication is a cross species phenomenon with genetically predetermined aspects.	Page 50
1.8.1 Clock genes and sex-determining genes are implicated in forms of social communication in <i>Drosophila</i> .	Page 53
1.9 Aims and structure of the thesis.	Page 58

### 2 Publication. Association of *PER1* and *NPAS2* with autistic disorder: support for the clock genes/social timing hypothesis.

2.1 General introduction and contributions to the publication.	Page 61
2.1.1 Experimental contributions.	Page 61
2.1.2 Authorship of the paper.	Page 62



2.2 Abstract.	Page 62
2.3 Introduction.	Page 63
2.3.1 Whole-genome screens and candidate gene studies for autism	Page 64
2.3.2 Timing difficulties and autism	Page 65
2.3.3 The hypothesis.	Page 67
2.3.4 Molecular genetic background: clock genes.	Page 68
2.3.5 The candidate genes and their relevance to the hypothesis	Page 70
2.4 Materials and methods.	Page 72
2.4.1 Subjects.	Page 72
2.4.2 Screening the candidate genes.	Page 75
2.4.3 Statistics.	Page 77
2.5 Results.	Page 77
2.6 Discussion.	Page 81

### **3 Are autism-associated genes clock-controlled genes and what possible effects might the autism-associated SNPs in *PER1* and *NPAS2* have?**

3.1 Introduction.	Page 86
3.2 Which autism candidate genes are clock-controlled genes?	Page 89
3.2.1 Introduction.	Page 89
3.2.2 Methods.	Page 90
3.2.3 Results.	Page 91
3.2.4 Conclusion.	Page 93
3.3 Searching for regulatory features coincident with the autism-associated clock gene SNPs and haplotypes.	Page 94
3.3.1 Introduction.	Page 94
3.3.2 Methods.	Page 95
3.3.3 Results.	Page 99
3.3.3i <i>NPAS2</i> , SNPs individually.	Page 99
3.3.3ii <i>NPAS2</i> , most significant haplotype.	Page 100
3.3.3iii <i>PER1</i> , SNPs individually	Page 109
3.3.3iv <i>PER1</i> , most significant haplotype.	Page 112
3.3.4 Conclusion	Page 113
3.4 Do the autism-associated SNPs in <i>NPAS2</i> and <i>PER1</i> coincide with transcription factor binding site motifs? TESS analysis of rs1811399, rs34705978 rs885747 and rs6416892.	Page 116
3.4.1 Introduction.	Page 116
3.4.2 Method	Page 117
3.4.3 Results	Page 117
3.4.4 Conclusion	Page 119

### **4 Investigating possible novel functionality of the autism-associated clock gene SNPs with web based bioinformatics modeling tools.**

4.1 Introduction	Page 120
4.2 Could the autism-associated SNP rs885747 affect splicing of <i>PER1</i> ?	Page 121

4.2.1 Introduction	Page 121
4.2.2 Method	Page 122
4.2.3 Results	Page 122
4.2.4 Conclusion	Page 123
4.3 Structural analysis of RNA transcripts containing autism-associated SNPs in <i>NPAS2</i> , <i>PER1</i> , <i>EN2</i> and the psychosis susceptibility variant rs1344706 in <i>ZNF804A</i>	Page 124
4.4 Web publication: Autism-associated SNPs in the clock genes <i>NPAS2</i> , <i>PER1</i> and the homeobox gene <i>EN2</i> alter DNA sequences that show characteristics of microRNA genes.	Page 127
4.4.1 Contributions to this publication	Page 127
4.4.2 Abstract	Page 127
4.4.3 Introduction	Page 128
4.4.4 Materials and Methods	Page 132
4.4.5 Results	Page 136
4.4.5i Analysis of EN2 intron 1 sequence fragment containing the autism-associated SNPs rs1861972 (A/G) and rs1861973 (T/C)	Page 136
4.4.5ii Analysis of the intron region of NPAS2 containing the autism-associated SNP rs1811399 (C/A).	Page 140
4.4.6 Discussion	Page 146
4.5 Further candidate MIR analysis using MirBase, The Vienna Web Server and The RNA Self Containment Web Server.	Page 151
4.5.1 Introduction	Page 151
4.5.2 Method	Page 151
4.5.3 Results	Page 153
4.5.3i. <i>NPAS2</i> rs1811399 hairpin	Page 153
4.5.3ii. <i>EN2</i> rs1861973 hairpin	Page 163
4.5.3iii. <i>PER1</i> rs885747 hairpin	Page 167
4.5.4 Conclusion.	Page 169
4.5.5. Analyzing Self Containment with the RNA Self Containment Web Server.	Page 170
4.5.5i Introduction.	Page 170
4.5.5ii Method.	Page 170
4.5.5iii Results.	Page 170
4.5.5iv Conclusion.	Page 171
4.5.6. Searching MirBase for sequence homology between the <i>NPAS2</i> , <i>PER1</i> and <i>EN2</i> hairpins using BLAST and SSEARCH tools.	Page 171
4.5.6i. Introduction.	Page 172
4.5.6ii. Method.	Page 172
4.5.6iii. Results.	Page 172
4.5.6iv. Conclusion.	Page 173
4.5.7 General conclusion from the RNA secondary structure analysis.	Page 173

4.6 Analysis of the results of three microRNA expression studies in autism in relation to clock gene targets.	Page 178
4.6.1 Introduction	Page 178
4.6.2 Method	Page 179
4.6.3 Results	Page 179
4.6.4 Conclusion	Page 182

## 5 Discussion

5.1 Summary of the main findings	Page 184
5.1.1 Clock controlled genes include autism-associated genes.	Page 185
5.1.2 The autism associated clock gene SNPs rs885747, rs6416892 and rs34705978 have predicted function.	Page 185
5.1.3 Could point mutation generate de novo silencing RNAs?	Page 186
5.1.4 Autism and prostate cancer share disease-associated clock gene SNPs: autism as a male sex hormone/clock gene related disorder.	Page 189
5.1.5 How could clock genes affect the frequency of neural oscillators? Tuning of oscillators by clock genes: a hypothesis.	Page 191
5.1.5i Hypothesis.	Page 194
5.1.6 Convergence between the circadian molecular clock and microRNA biogenesis pathways: a genetic focus for neuropsychiatric phenotypes.	Page 196
5.1.7 There are many ways to break a clock.	Page 199
5.1.8 Final conclusion and indications for further research.	Page 200

## APPENDICES

### Appendix A

#### **Methylation status of a CpG rich element in *PER1*: a preliminary investigation.**

A.1 Introduction.	Page 201
A.2 Materials and methods.	Page 203
A.2.1 Materials	Page 203
A.2.1i Buffers and solutions	Page 203
A.2.1ii Primer Sequences.	Page 204
A.2.2 Methods	Page 206
A.2.2i Agarose gel electrophoresis.	Page 206
A.2.2ii Polymerase Chain Reaction (PCR): Standard procedure.	Page 207
A.2.2iii PCR: methylation-specific procedure.	Page 208
A.3 Results.	Page 210
A.3.1 Standard PCR reactions.	Page 210
A.3.2 Methylation specific PCR reactions.	Page 211
A.4 Discussion.	Page 213

## Appendix B

### Does NPAS2 form a complex with HAT1?

B.1 Introduction.	Page 216
B.2 Materials and methods.	Page 218
B.2.1 Media, buffers, solutions, gels. and antibodies.	Page 218
B.2.2 Tissue culture with HeLa cells.	Page 221
B.2.3 Protein extraction from HeLa cells.	Page 222
B.2.4 Protein fractionation by size exclusion chromatography.	Page 222
B.2.5 SDS Polyacrylamide gel electrophoresis (SDS PAGE) and western blot.	Page 223
B.3 Using protein fractionation, SDS-PAGE and western blot to investigate NPAS2 and HAT1 protein complex formation.	Page 225
B.3.1 Methods.	Page 225
B.3.2 Results.	Page 226
B.3.3 Discussion	Page 228

## Appendix C

### ZNF804A and SNP rs1344706: A psychosis-susceptibility variant and its possible relevance to autism.

C.1 Introduction.	Page 230
C.2 Method.	Page 231
C.3 Results:	Page 231
C.3.1 Autism Chromosomal Rearrangement Database.	Page 231
C.3.2 Candidate MIR scans using MirBASE and The Vienna Web Server	Page 232
C.4 Conclusion	Page 239

## Appendix D

### Human PER1: a comparison with *Drosophila* PERIOD in relation to the K&H cycle determining region and an autism/cancer associated haplotype.

D.1 Introduction	Page 240
D.2 Method	Page 241
D.3 Results	Page 242
D.4 Conclusion	Page 246

<b>References.</b>	Page 251
--------------------	----------

## List of Figures

Figure 1-1.	Chromosomal anomalies linked to autism.	Page 13
Figure 1-2.	MicroRNA biogenesis.	Page 16
Figure 1-3.	The circadian molecular clock in relation to psychiatric disorders.	Page 36
Figure 1-4.	Circadian clock proteins PER, CRY and TIMELESS in the ATM and ATR checkpoint pathways.	Page 43
Figure 1-5.	Cell cycle of a typical somatic cell.	Page 47
Figure 2-1.	Linkage disequilibrium within <i>NPAS2</i> .	Page 79
Figure 3-1.	Banding pattern of human chromosome 2.	Page 101
Figure 3-2.	Alternative transcripts of <i>NPAS2</i> .	Page 104
Figure 3-3.	Differentially methylated regions in <i>NPAS2</i> .	Page 105
Figure 3-4.	Differentially methylated region C in <i>NPAS2</i> .	Page 106
Figure 3-5.	Banding pattern of human chromosome 17.	Page 110
Figure 3-6.	Alternative transcripts of <i>PER1</i> .	Page 111
Figure 3-7.	The autism-associated SNP rs6416892 is located in a SREBF2 binding-peak.	Page 112
Figure 4-1.	Splice Scan2 alternative splicing: <i>PER1</i> .	Page 123
Figure 4-2.	The <i>EN2</i> candidate microRNA containing the autism-associated SNP rs1861973.	Page 138
Figure 4-3.	Effects of SNPs rs1861972 and rs1861973 on secondary structure of mRNA transcribed from an <i>EN2</i> intron fragment.	Page 139
Figure 4-4.	A or C Alleles of rs1811399 determine two different mRNA structures.	Page 140
Figure 4-5.	Candidate microRNA from intron 1 of <i>NPAS2</i> .	Page 143
Figure 4-6.	<i>PER1</i> intron 12 hairpins.	Page 145
Figure 4-7.	The <i>NPAS2</i> hairpin: A allele, RNAfold, Vienna Web Server.	Page 155
Figure 4-8.	The <i>NPAS2</i> rs1811399 hairpin: A allele; RNAfold, Vienna RNA Web Server, MFE and centroid structures.	Page 156
Figure 4-9.	The <i>NPAS2</i> hairpin: C allele; RNAfold, Vienna RNA Web Server.	Page 157
Figure 4-10.	The <i>NPAS2</i> hairpin: C allele; RNAfold, Vienna RNA Web Server MFE and centroid structures.	Page 158
Figure 4-11.	Sequence alignment of rs1811399 hairpin	Page 159
Figure 4-12.	Conservation of structure of the rs1811399 hairpin in primates	Page 160
Figure 4-13.	Consensus structure of the rs1811399 hairpin in primates.	Page 161
Figure 4-14.	Dot plot for the rs1811399 hairpin.	Page 162
Figure 4-15.	The <i>EN2</i> hairpin, T allele, RNAfold, Vienna RNA Web Server.	Page 164
Figure 4-16.	The <i>EN2</i> hairpin, C allele, RNAfold, Vienna RNA Web Server.	Page 165
Figure 4-17.	Sequence conservation of the <i>EN2</i> rs1861973 region in dot and bracket notation.	Page 166
Figure 4-18.	Sequence conservation in the <i>EN2</i> rs1861973 region.	Page 166
Figure 4-19.	Best matches from MirBASE searches using the <i>EN2</i> , <i>NAPS2</i> and <i>PER1</i> hairpins.	Page 172

Figure 4-20.	NPAS2 candidate microRNA and hsa-let-7f-2 compared.	Page 174
Figure 5-1	Possible effects of candidate microRNAs in NPAS2, that targets CLOCK and in PER1 that targets PER1.	Page 189
Figure 5-2.	Possible co-reliance of the circadian molecular clock and MicroRNA biosynthetic pathways in neurons.	Page 198
<b>Appendix</b>		
Figure A.1.	Bisulphite reaction and the conversion of cytosine to uracil via a sulphonated intermediate.	Page 202
Figure A-2.	<i>PER1</i> sequence fragments.	Page 205
Figure A-3.	PCR amplifications of fragments of human <i>PER1</i> in autistic male child AUO 64004.	Page 212
Figure A-4.	Sample A: AUO 64001 and B: AUO 66401 amplified with primers PER1-1 and PER1-2;	Page 213
Figure B-1.	Autoradiograph of western blots of nuclear and cytoplasmic protein extracts from HeLa cells.	Page 226
Figure B-2.	Western blot of size-exclusion column chromatography fractions from a nuclear extract of HeLa cells.	Page 228
Figure C-1.	ZNF804A hairpin with rs1344706 C allele.	Page 234
Figure C-2.	ZNF804A hairpin with rs1344706 A allele.	Page 235
Figure C-3.	Alignment of protein sequence fragment of ZNF804A and GPATCH8 showing conservation of the amino acid sequence of the zinc finger DNA binding domain.	Page 236
Figure C-4.	Structure of a candidate pre-MIR in the ZNF804A hairpin.	Page 236
Figure C-5.	P-values for genotyped SNPs in <i>ZNF804A</i> reproduced from Williams, Norton et al. (2010).	Page 237
Figure C-6.	The sense-antisense complementarity between <i>ZNF804A</i> and the 3'UTR of UHMK1.	Page 238
Figure D-1.	Protein domain structure of PER1.	Page 243
Figure D-2.	Protein sequence alignment: <i>Drosophila</i> PER and human PER1.	Page 244
Figure D-3.	Alignment of <i>Drosophila</i> PER sequence and human PER1.	Page 245
Figure D-4.	Re-alignment of <i>Drosophila</i> PER sequence and human PER1 centered on the Thr-Gly repeat region.	Page 246

## List of Tables

Table 2-1.	Candidate genes and their chromosomal locations.	Page 67
Table 2-2.	Diagnostic criteria and demographics of the sample.	Page 73
Table 2-3.	Distribution of percentile ranking (PR) from ability /behavioral measures and corresponding intelligence levels.	Page 73
Table 2-4.	Results of the transmission disequilibrium tests.	Page 80
Table 3-1.	Bioinformatics tools.	Page 88
Table 3-2.	Autism genes and circadian control element genes.	Page 92
Table 3-3.	Overlap between autism genes, circadian element genes and cortex genes.	Page 93
Table 3-4.	Synonyms of rs1811399.	Page 96
Table 3-5.	Summary of the autism associated SNPs and haplotypes in <i>NPAS2</i> and <i>PER1</i> .	Page 98
Table 3-6.	The conserved RREs of <i>NPAS2</i> .	Page 102
Table 3-7.	The conserved RRE in the first intron of <i>NPAS2</i> .	Page 103
Table 3-8.	Cell-type specific transcription factor binding sites in the <i>NPAS2</i> haplotype.	Page 107
Table 3-9.	Results from GLIDERS database search showing SNPs in linkage disequilibrium to rs1811399.	Page 108
Table 3-10.	Synonyms of rs885747.	Page 109
Table 3-11.	Five common variants within the region of <i>PER1</i> delineated by the best autism-associated haplotype.	Page 113
Table 3-12.	Summary of the features coincident with, and delineated by, the <i>NPAS2</i> and <i>PER1</i> autism -associated SNPs and haplotypes respectively.	Page 113
Table 3-13.	TESS transcription factor binding site predictions.	Page 118
Table 4-1.	The predicted targets of the candidate microRNAs in <i>EN2</i> , <i>NPAS2</i> and <i>PER1</i> .	Page 146
Table 4-2.	Reference sequences and annotations for the <i>NPAS2</i> rs1811399 hairpin.	Page 153
Table 4-3.	Vienna RNA Web server RNAfold output statistics for the <i>NPAS2</i> hairpin containing the SNP rs1811399.	Page 154
Table 4-4.	Vienna RNA Web server RNAfold output statistics for the <i>NPAS2</i> rs1811399 hairpin for five primate sequences.	Page 154
Table 4-5.	<i>EN2</i> sequence fragment.	Page 163
Table 4-6.	Statistics for the <i>EN2</i> hairpin containing the SNP rs1861973.	Page 163
Table 4-7.	Statistics for the conservation of RNA structure around SNP rs1861973 analysed with the Vienna Web Server.	Page 167
Table 4-8.	Sequence and SNP details of intron 12 <i>PER1</i> .	Page 167

Table 4-9.	Vienna RNA Web Server RNAfold output statistics for the PER1 hairpin containing the autism-associated SNP rs885747 and rs885953 C.	Page 168
Table 4-10.	Vienna RNA Web Server RNAfold output statistics for the PER1 hairpin containing the autism-associated SNP rs885747 and rs885953 G.	Page 168
Table 4-11.	Structural conservation for the PER1 hairpin	Page 169
Table 4-12.	Self Containment Index data for the EN2, NAPS2 and PER1 hairpins.	Page 171
Table 4-13.	Summary of the attributes of the hairpin structures.	Page 173
Table 4-14.	Genotypes for the autism associated SNP rs1861973 in neurotypical populations.	Page 177
Table 4-15.	EI-Neel. MicroRNAs up-regulated.	Page 179
Table 4-16.	EI-Neel. MicroRNAs down-regulated.	Page 180
Table 4-17.	Talebizadeh. MicroRNAs up-regulated.	Page 180
Table 4-18.	Talebizadeh. MicroRNA down-regulated.	Page 180
Table 4-19.	Sarachana. MicroRNAs up-regulated.	Page 181
Table 4-20.	Sarachana. MicroRNAs down-regulated.	Page 181
Table 4-21.	Overlap between the results of EI-Neel, Talebizadeh and Sarachana.	Page 182.

## Appendix

Table A-1.	Native primers for <i>PER1</i> .	Page 204
Table A-2.	Fragment size calculations from Ensembl database sequences.	Page 204
Table A-3.	Methylation Specific Primers for <i>PER1</i> .	Page 204
Table A-4.	Thermal cycler settings for standard PCR reactions.	Page 207
Table A-5.	Thermal cycler settings for the methylation -specific PCR reactions.	Page 209
Table A-6.	Primer pairs: <i>PER1</i> and the observed fragment sizes.	Page 211
Table B-1a.	SDS Polyacrylamide resolving gel mix.	Page 219
Table B-1b.	SDS Polyacrylamide stacking gel mix.	Page 220
Table C-1.	<i>ZNF804A/B</i> in autism cases from the ACRD.	Page 232
Table C-2.	Stability of the ZNF804A hairpin containing the SNP rs1344706.	Page 233
Table C-3.	Predicted DROSHA scores for the ZNF804A hairpin.	Page 237
Table C-4.	Targets of the ZNF804A candidate mir 5'arm.	Page 237
Table C-5.	Targets of the ZNF804A candidate mir 3'arm.	Page 237



# 1 Introduction

## 1.1 What is autistic disorder?

The severe neurodevelopmental disorder known as autistic disorder is characterized by impaired verbal and non-verbal communication, impaired reciprocal social interaction and a markedly restricted repertoire of activities and interests. These features co-occur in an individual with autistic disorder and are detectable by three years of age (APA 1994). The prevalence of autistic disorder is 0.1–0.2% with a recurrence risk ratio giving a twenty five fold increase of developing the disorder over population base-rates for a first degree relative of an individual with autistic disorder, a measure known as the relative risk (Freitag 2007). 90% of the phenotypic variation in autism is attributable to genetic variation and twin studies give a 60% concordance rate in monozygotic twins but no concordance in dizygotic twins (Bailey et al. 1995; Hyman 2008). Males are also at higher risk of autistic disorder than females in the approximate ratio of 4:1 (Fombonne 2005). Such data suggest that genetic factors dominate the occurrence of autistic disorder within the population but environmental factors such as adverse intra-uterine conditions or potentially protective effects may also be involved (Santangelo and Tsatsanis 2005). Work on differences between monozygotic embryos in mouse suggests it is also possible that epigenetic effects could contribute to the discordance observed in twin studies (Piotrowska et al. 2001; Piotrowska and Zernicka-Goetz 2001; Torres-Padilla et al. 2007).

Outside of autistic disorder but excluding other syndromes such as Rett syndrome and fragile X syndrome that have autistic behaviour as part of

the phenotype (APA 1994), individuals may show milder forms of impairment that are similar to autistic disorder. This has led to the idea of a spectrum of disorders with autistic features and the term autism spectrum disorders (ASD) is now in common use. ASD is thus used to refer to individuals who show deficit in one or perhaps two of the domains of the triad of impairments that define autistic disorder or individuals who show some measure of weakness in each domain but to a degree that would not elicit a diagnosis of autistic disorder according to the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV) (APA, 1994). There is currently no clinically defined cut-off point on the autism spectrum that distinguishes individuals with an ASD from neuro-typical individuals. This thesis primarily concerns the well-defined category of autistic disorder.

## **1.2 Are there biological markers for autism?**

Despite six decades of research on autism there are still no biological, genetic, or physiological markers, unequivocally linked with the disorder. There is however, some consensus that: certain brain regions such as the cerebellum, (Bauman and Kemper 1985; Bauman and Kemper 1988; Amaral et al. 2008; Scott et al. 2009), circadian rhythm disturbance including altered serotonin and melatonin levels (Nir et al. 1995; Cook and Leventhal 1996; Kulman et al. 2000; Anderson 2002; Anderson et al. 2002; Tordjman et al. 2005; Corbett et al. 2006), abnormalities in cholesterol metabolism (Tierney et al. 2001; Tierney et al. 2006; Bukelis et al. 2007) and increased oxidative stress and impaired methylation capacity

(James et al. 2004; James et al. 2006; James et al. 2008) reflect different aspects of damage to a pathway that leads to autism.

The brain region most consistently linked to autism is the cerebellum. From 1986 (Bauman and Kemper 1986) to current studies (Scott et al. 2009) differences have been noted in the cerebella of autistic individuals that suggest changes in this brain region are linked to autism. However, with the exception of a small number of studies, such as the Imbalanced Spectrally Timed Adaptive Resonance Theory (Grossberg and Seidman 2006) there has been little attempt to understand how defects in a brain region generally thought of as mainly concerned with motor coordination, could contribute to affective deficits such as lack of empathy requiring a “theory of mind” that is diagnostically central to the presentation of autism (Baron-Cohen et al. 1985). It is noteworthy that variants of the gene for ENGRAILED2 (*EN2*, which has a pivotal function in cerebellum and hind-brain development (Sgaier et al. 2007)), has been repeatedly shown to be associated with autism (Petit et al. 1995; Benayed et al. 2005; Brune et al. 2008; Wang et al. 2008).

The results of investigations into autism-associated changes in the morphology of other brain regions are somewhat less consistent. Overall, the results of studies aimed at detecting structural brain differences linked to autism are difficult to interpret. This is mostly due to inconsistency in subject age, intelligence, degree of autism and allowed levels of co-morbidities amongst the samples and across the different studies.

Some regions have been targeted for investigation on the basis of involvement in distinct behaviours that are disturbed in autism. The Amygdala Theory of Autism (Baron-Cohen et al. 2000) focuses on this

brain region's involvement in emotional responses (Cahill and McGaugh 1998). The cerebrum, the seat of cognitive function in humans, has been examined as a candidate region in autism because of altered levels of cognitive functioning in individuals with the disorder.

Measurements of brain volume and growth trajectories show altered patterns of brain growth for autistic children that can be summarized as an early period of excess growth followed by a period of repressed growth, as compared with typically developing infants and children (Courchesne et al. 2003). More recently, nuclear magnetic resonance imaging highlights regions of the temporal lobe that are affected in autism (Ecker et al. 2010; Salmond et al. 2005).

Neural activity of the brain generates oscillating electro-magnetic fields that can be detected by skin-surface electrodes placed on the scalp (Buzsaki and Draguhn 2004). Investigation of autism using this principle together with functional nuclear magnetic resonance imaging (fNMI) indicates that the patterns of intra-brain communication between specialized neural networks are altered in autism (Minshew and Williams 2007). This may reflect reports of anatomical differences in the inter-hemispheric portal, the corpus callosum in autism (Piven et al. 1997). Abnormal functional connectivity is proposed to be a characteristic of the autistic brain where an exaggeration of local connectivity takes precedence over neurotypical integrative connectivity (Just et al. 2007). Coherent neuronal oscillations (< 0.1 Hz) of electrical activity of the resting brain (default mode network) show characteristic connectivity signatures of neuropsychiatric disorders including autism (Buckner and Vincent 2007; Broyd et al. 2009).

Genes involved with synapse formation and structure are affected in certain autism families (Durand et al. 2007; Yamakawa et al. 2007) although recent population based studies do not support a role for common variants of the synaptic scaffold protein, SH3 and multiple ankyrin repeat domains 3 (SHANK3) in autism (Qin et al. 2009; Sykes et al. 2009). Nevertheless, multiple lines of evidence support the idea that abnormal synapse formation is the basis of autism pathology (Bourgeron 2009) and the evidence for this idea is strongest when the broadest autism phenotype is considered (Alarcon et al. 2008). In the wider population, genetic variation in any of these synaptic genes does not seem to confer increased autism-specific risk however, that suggests a generalized vulnerability of brain synapses in ASDs and schizophrenia (Friedman et al. 2008; Burbach and van der Zwaag 2009; Kirov et al. 2009).

### **1.3 Sleep and circadian rhythm disturbance is associated with autism.**

Everyday experience shows that human behavior is ordered into repeating sleep-wake cycles. It is less evident to casual observation however, that even in the absence of physical cues such as dawn and dusk (and/or social and metabolic cues e.g. set meal-times) humans, in common with other mammals, display an approximately 24h periodicity in activity and physiological measures (Aschoff 1965; Aschoff et al. 1971). These so called circadian rhythms are also detectable in the workings of the individual organs of the body (Brown et al. 2002; Dardente and Cermakian 2007; Kowalska and Brown 2007) and even persist in cell lines (Balsalobre et al. 1998). The circadian system of mammals may thus be

envisaged as a series of slave oscillators, the organ systems, under the endocrine and neurological control of a central coordinating pacemaker.

The central organismal pacemaker in mammals is located in the brain, the suprachiasmatic nuclei of the anterior hypothalamus (SCN). The phase of the self sustaining 24h cycles of neural activity in the SCN is set by neural cues from the retina to which the SCN is linked via the bundle of nerve fibers known as the retinohypothalamic tract (RHT). Via this tract, the SCN receives photic cues generated by the action of light on the rod, cone and melanopsin photoreceptors of the retina (Berson et al. 2002; Hattar et al. 2002). Subsequently, the SCN emits neuronal and hormonal signals, including glucocorticoid hormone that entrains the circadian rhythms of the peripheral organs *via* glucocorticoid hormone receptors (Balsalobre et al. 2000; Hastings et al. 2003).

Other factors in addition to the action of light are capable of resetting the phase of circadian rhythms in mammals. For example feeding regimes (Stokkan et al. 2001), temperature rhythms (Brown et al. 2002), social cues (Aschoff et al. 1971; Levine et al. 2002; Mistlberger and Skene 2004) and hormonal signals (Balsalobre et al. 2000), can each entrain or modulate circadian rhythms.

Within the mammalian brain it is apparent that some regions show circadian rhythms whilst others do not and that the phase of one region with respect to another varies. Thus the brain contains multiple circadian oscillators, which may have bearing on the phenomenon of neural network oscillations that appear to be a requirement for temporal ordering and long term storage of information (Abe et al. 2002; Buzsaki and Draguhn 2004; Feillet et al. 2008). Further, it would appear that the circadian rhythm of

the sleep-wake cycle is required for appropriate cognitive development through the processes of memory consolidation (Stickgold 2005).

Sleep disturbance is a common problem of people with autistic disorder (Elia et al. 2000). Parents of autistic children report difficulty settling their child for sleep and nights with frequent wakening (Richdale and Prior 1995). Melatonin is the mammalian endocrine signal for darkness. Its precursor, serotonin, as well as the level of urinary cortisol, all vary in a circadian manner and each of these measures are significantly altered in autism (Richdale and Prior 1992; Nir et al. 1995; Cook and Leventhal 1996; Corbett et al. 2006). Melatonin is currently used in the pharmacological treatment of poor sleep in autism (Andersen et al. 2008).

Further evidence of dysregulation of the melatonin system in autism is seen in the association of coding sequence variants and a promoter variant in the acetylserotonin O-methyltransferase (ASMT) gene with ASDs. ASMT also called hydroxyindole O-methyltransferase (HIOMT), catalyses the final step in the serotonin to melatonin biosynthetic pathway (Melke et al. 2008). However, what may have been a neat picture is blurred by the concurrent finding that the abnormal melatonin levels and promoter variants, though statistically associated with ASDs, are also found in the non-autistic parents of the affected children (Melke et al. 2008). This is interpreted as altered melatonin levels are contributory but not in themselves causative of ASDs. Taken together these factors represent a disruption of circadian regulation in autism that may extend down to the level of the circadian molecular clock (Nicholas et al. 2007; Melke et al. 2008; Hu et al. 2009).

Alteration in clock genes that drive the circadian cycle could derive altered melatonin levels, since the circadian rhythm in melatonin level is under the control of the circadian molecular clock. Disruption of the core circadian clock mechanism could thus disrupt melatonin levels but presumably give additional phenotypes because of pleiotropic effects of the clock genes themselves.

Boucher (2001) proposed that an integrated series of oscillators in the brain spanning a frequency spectrum from milliseconds to hours is damaged in autism. However, it is clear that circadian disruption accompanies many neuropsychiatric disorders and it is currently unclear to what degree, and in what way, circadian rhythm collapse contributes to each disorder specifically, (Barnard and Nolan 2008).

Strong indication that the circadian disruption in autism is rooted at the molecular level comes from the finding that altered clock gene expression profiles can distinguish severe autism from neurotypical controls and even from mildly affected autism spectrum cases (Hu et al. 2009a; Hu et al. 2009b). These gene-expression experiments were carried out on autism lymphoblastoid cell lines and thus it would appear very unlikely, that the altered clock gene expression is due to global neural or humoral effects manifesting at a systems level by an altered clock in the suprachiasmatic nucleus. Rather, this suggests that the circadian rhythm anomalies in autism result from factors that affect the core molecular clock mechanism.

#### **1.4 Autism Genetics.**

The heritability of autistic disorder has been described as involving a number of genes at unlinked and some epistatic loci that together



contribute to the phenotype (Pickles et al. 1995). There is considerable comorbidity of mental retardation in autism (at least 75%) and this is a major problem for researchers investigating the genetics of this disorder (Fombonne 1999). The comorbidity of mental retardation may be seen to dilute the autism specific phenotype from the point of view of selecting a homogeneous sample for genetic analysis. For example, whole genome screens where large sample size is required have generally had to sacrifice sample homogeneity, to ensure a sufficiently large number of samples to prioritize power requirements.

Severely autistic individuals with normal or superior IQ measures indicate however, that autism is not a result of mental retardation or must always co-occur with mental retardation. Thus certain studies have addressed the problem of sample heterogeneity by stricter control of IQ, selecting for more intellectually able subjects that met strict diagnostic criteria for autistic disorder, (Nicholas et al. 2007) or by dividing the sample into a number of sub categories based on analysis of the behavioural phenotype prior to screening (Hu and Steinberg 2009).

An alternative explanation of autism heritability is gleaned from analysis of autism risk in multiplex families (>1 affected member) where strong evidence for dominant transmission to male offspring is seen. This multiplex family data gives a good fit to a simple genetic model in which most autism families fell into two types. Type 1 families (a small minority of autism families) represent a risk of autism in male offspring of ~50% and type 2 (the vast majority of autism families) where male offspring have a lower risk. Sporadic autism with high penetrance in males and relatively poor penetrance in females is proposed to account for the risk in Type 2

families (the low-risk families) and high-risk families are from those offspring, most often females, who carry a new causative mutation but are unaffected and in turn transmit the mutation in dominant fashion to their offspring (Zhao et al. 2007). However, validating this proposed pattern of autism heritability would require an explanation that incorporated the function of any genes found to be associated with the disorder in terms of a penetrance difference in males and females. It must also be noted that this model is somewhat at odds with the most recent genome wide scans that indicate numerous common variants contribute to autism risk (Ma et al. 2009).

There are linkage findings for autism in at least two independent studies in regions 2q, 3q25–27, 3p25, 6q14–21, 7q31–36 and 17q11–21. (Bailey et al. 1998; Buxbaum et al. 2001; Cantor et al. 2005) and a genome wide meta-analysis found significance for autistic disorder on 7q21-35 (Trikalinos et al. 2006). However, the genome screen results have been inconsistent, possibly because these studies have until recently, been generally under-powered to find genes of small effect set in a context of sample heterogeneity and diagnostic differences.

The most recent autism Genome Wide Association Studies employed a much larger sample, highlighting non-coding SNPs in an intergenic region on chromosome 5p as significantly associated with autism. This region is flanked by the cadherin genes *CDH9* and *CDH10*; reinforcing the implication of impaired neuronal connectivity in autism (Ma et al. 2009; Wang et al. 2009).

Candidate gene studies in autism have tested numerous genes, on the basis of proximity to the peak scores of whole-genome linkage studies or

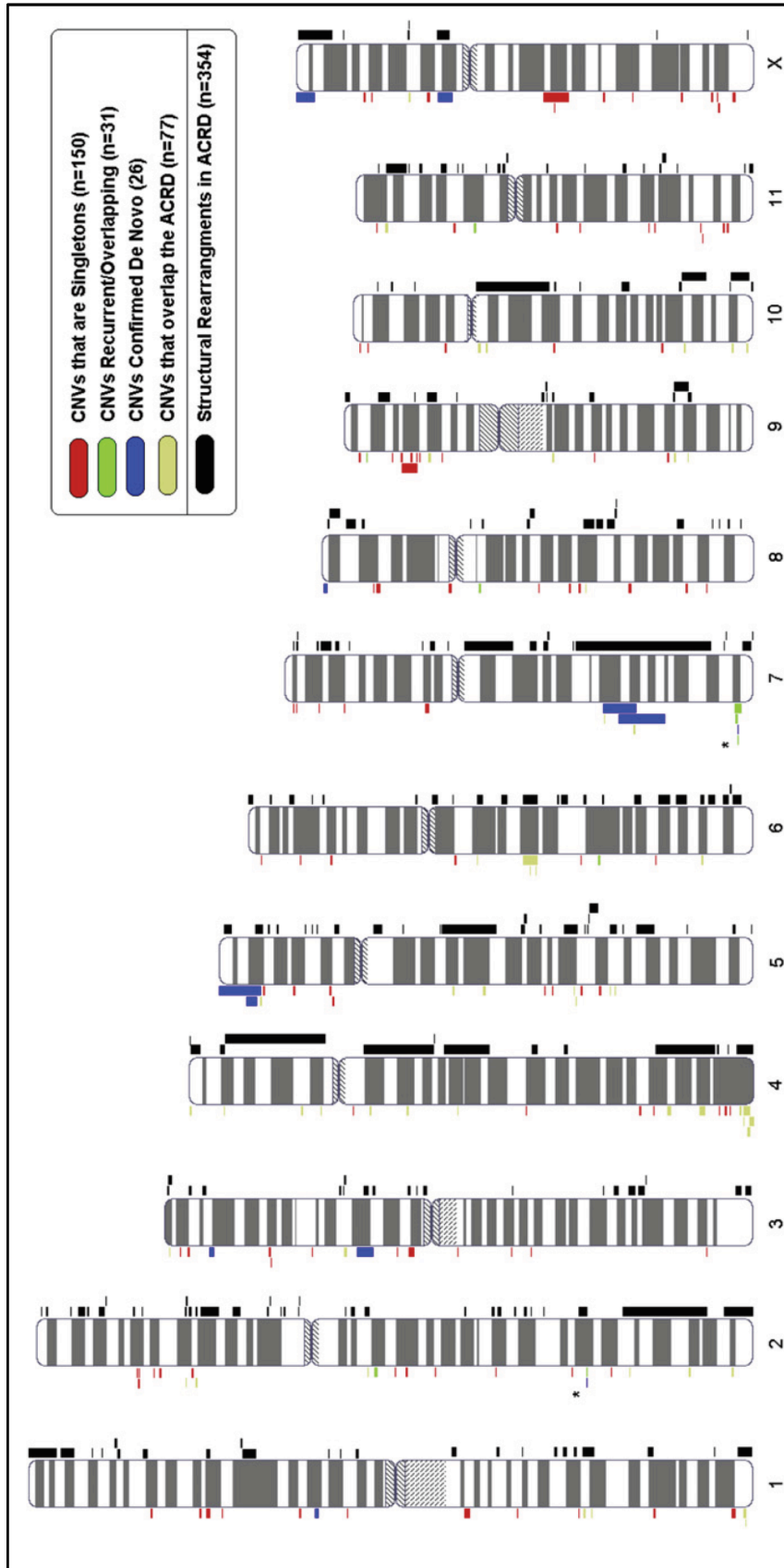
by their implication in neurological systems e.g. the serotonergic system. With the exceptions of the genes for the neurotransmitter receptor protein gamma-aminobutyric acid (GABA) A receptor, beta 3 (GABRB3); the homeobox protein ENGRAILED2, that is developmentally active in the formation of the cerebellum; and the serotonin transporter solute carrier family 6 (neurotransmitter transporter, serotonin), member 4 protein (SLC6A4 also known as 5-HTT), these studies remain largely unreplicated (Cook et al. 2001; Buxbaum et al. 2002; Benayed et al. 2005). A further complication is that the scope of candidate gene studies in autism is sometimes limited by the lack of credible, integrated supporting hypotheses that link the function of the candidate genes to what is known of the pathophysiology of autism.

#### 1.4.1 MicroRNA genes

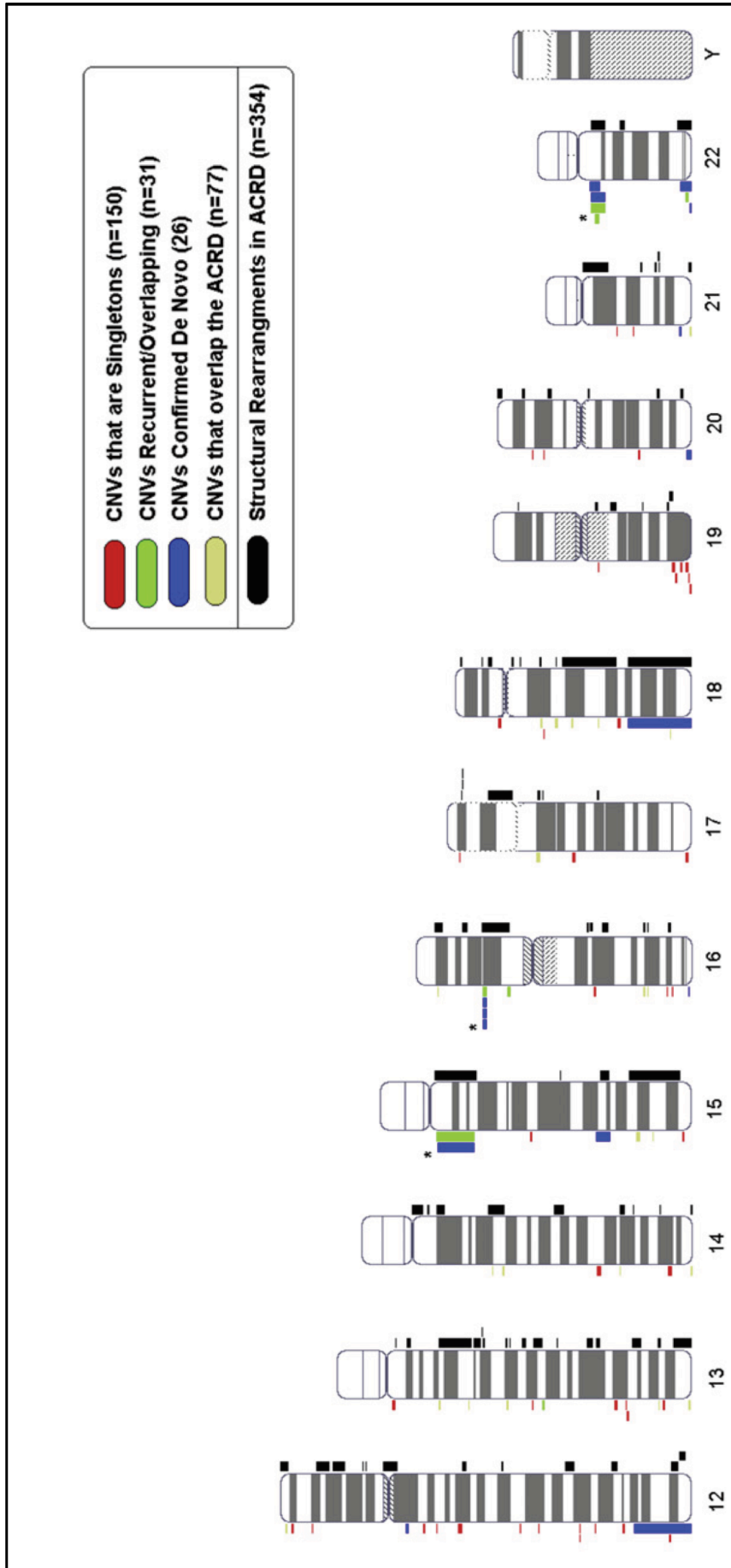
The exons of protein coding genes represent a small proportion (~1.2%) of the human genome (Lunter et al. 2006). The initial idea that the remaining non-coding regions represented genomic “junk” has been gradually eroded by findings that non-coding regions may contain control elements essential to the regulation of gene expression and certain non-coding transcripts are functional molecules in their own right. It now appears that most of the human genome is transcribed (Hayashizaki and Carninci 2006) and apart from transfer RNA (tRNA) and ribosomal RNA (rRNA), a growing collection of non-coding RNA (ncRNA) species are identified. Amongst these ncRNAs, microRNAs are of particular relevance to autism because of their important role in neural patterning (Kosik 2006),

developmental timing (Lee et al. 1993; Reinhart et al. 2002) and survival of purkinje neurons (Schaefer et al. 2007).

The precursors of microRNAs (pre-microRNA) may be derived from intergenic regions (~50%), introns of protein coding gene regions (~50%) or occasionally overlap protein-coding regions e.g. *MIR155* (Rodriguez et al. 2004; Saini et al. 2007). All microRNA gene transcripts contain sequences capable of internal, complementary base pairing and thus helical hairpin structures (pri-microRNA hairpins) form within these transcripts. These pri-microRNAs may be up to tens of kilobases in length and may contain a number of hairpin regions. The hairpins are recognized by drosha ribonuclease type III (DROSHA) and cleaved from the rest of the transcript. The ~60-100nt long hairpins (called pre-microRNAs) are transported to the cytoplasm by EXPORTIN-5 where the loops and tails are removed by the dicer1 ribonuclease type III (DICER) and the remaining ~22nt double stranded microRNA (the mature microRNA or miR) is subsequently incorporated into the RNA-induced silencing complex (RISC) that contains argonaught protein (EIF2C) and the neuropsychiatrically relevant fragile x mental retardation protein (FMRP) and DiGeorge critical region eight protein (DGCR8) (Verkerk et al. 1991; Yu et al. 1991; Wang et al. 2007). The gene specific targeting of this silencing complex is determined by the ~22nt nucleotide sequence of the mature microRNA loaded RISC and particularly the seed (the first 2 to 8 nucleotides, 5' to 3') of the mature microRNA (Lewis et al. 2005). Binding of mature microRNA loaded RISC to the UTRs of target genes causes translational repression of the target or rapid degradation of the transcript of the target gene (Wightman et al. 1993; Giraldez et al. 2006).



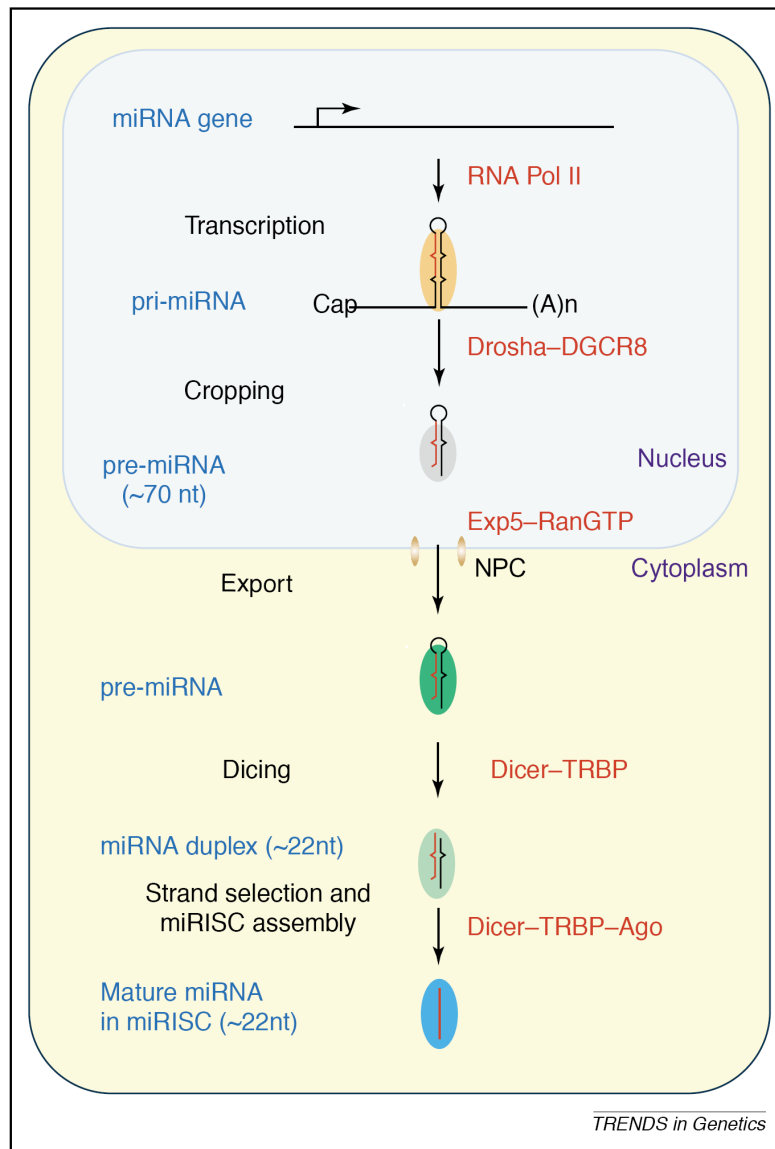
**Figure 1-1. Chromosomal anomalies linked to autism.** Reproduced from Marshall et al. (2008) shows the location of numerous copy number variations (CNVs) found in autism cases. CNVs are small duplications or deletions in chromosomes covering a limited number of genes. The green bars on chromosome 2q, 6q, 7q, 8q, 11q, 13q, 15q, 16p and 22q show regions that have been repeatedly observed in autistic individuals. The asterisk shows overlap between the recurrent and de novo CNVs.



**Figure 1-1 (continued) Chromosomal anomalies linked to autism.** Some of these regions however, particularly the 16p\* locus but also the 22q11 region may be better described as associated with neuropsychiatric disorder as CNVs in these chromosome regions also confer risk of schizophrenia and bipolar disorder (McCarthy et al. 2009). The acronym ACRD refers to the Autism Chromosome Rearrangement Database (Marshall et al. 2008).

Unlike protein coding genes, microRNA genes are difficult to predict bioinformatically and certain *bona fide* microRNAs are physically illusive due to expression patterns that are extremely restricted: spatially, to certain cell types; temporally, or both. For example miRNA *Isy-6* in *Caenorhabditis elegans* is expressed in but a few neurons at a specific developmental time. Nevertheless, this restricted pattern of expression of *Isy-6* is a critical determinant of the bilateral asymmetry of chemoreceptors in the worm that are required for typical behaviour in response to certain chemical stimuli (Johnston and Hobert 2003).

Although there is an absolute requirement for hairpin structure in pri-microRNA and pre-microRNA transcripts the calculated mean free energy (MFE) of such structures cannot be used predictively with complete confidence (Rivas and Eddy 2000). Nevertheless, a statistical approach shows that the MFE of microRNA hairpins is significantly lower than that for random sequences. Additionally, sequence alignments of multiple related species shows many microRNAs are evolutionarily conserved. Sequence conservation has been previously used to identify evolutionarily conserved microRNA genes in the human genome and it is likely that the majority of this type of microRNA gene have now been catalogued in human. Other microRNAs are however primate or species-specific (Mor et al. 2011) and it is likely that numerous further examples of this type of molecule will emerge with future research.



**Figure 1-2. MicroRNA biogenesis.** Reproduced from Kim and Nam (2006): a model for microRNA biogenesis where the blue shaded region is the nucleus and the yellow shaded region the cytoplasm. MicroRNA biogenesis is envisaged as a 5-stage process of: 1) transcription with Pol II; 2) cropping of the pri-microRNA by the microprocessor complex Drosha-DGCR8; 3) Export of the pre-microRNA from the nucleus by exportin-5, 4) removal of the loop by DICER and finally incorporation of one strand of the mature miR duplex into the rna induced silencing complex.

The most parsimonious verification of microRNA genes relies on a number of stages: firstly, biochemical confirmation with directional cloning may be employed (Ambros and Lee 2004). In this technique, the 19-25nt



RNA fraction of a total RNA sample is collected by polyacrylamide gel electrophoresis, eluted, ligated to 5' and 3' adaptor molecules, and amplified by RT-PCR to construct a cDNA library. Sequencing of individual clones allows subsequent checking for genomic position with BLAST searches. If the BLAST searches show the clones are not derived from the degradation of known small RNAs (rRNA and tRNA), then: bioinformatics confirmation of a ~80nt hairpin structure; evolutionary conservation of the hairpin sequence (and particularly that of the mature microRNA (miR) and seed region of the hairpin) and detection by Northern Blot analysis is required for confirmation. It should be noted however, that the requirement for sequence conservation in the above verification process would preclude a positive result for *bona fide* processed hairpins that are species-specific or which were well conserved on a structural level but only weakly conserved in terms of sequence.

Anomalous microRNA expression has been detected in autism (Abu-Elneel et al. 2008; Talebizadeh et al. 2008; Sarachana et al. 2010) and the notion that autism-associated SNPs in brain expressed introns could disrupt novel microRNAs was proposed by this author in [Precedings.nature.com](http://Precedings.nature.com) (hdl:10101/npre.2008.2366.1 2008). SNPs in X-linked microRNAs are shown to alter processing and targeting of microRNA (Sun et al. 2009) and these may have neurological significance. To date however, no autism-associated SNPs are co-located in fully confirmed microRNA genes. However, SNPs showing association with autism and that are not located in protein coding regions or control elements could be considered for co-location with novel microRNAs. It is not inconceivable that such orphan SNP associations might indirectly

affect neuropsychiatrically relevant proteins, via an effect on novel microRNAs that have synaptic genes, for example, among their targets.

The recent discovery that the microRNAs MIR132 and MIR219 play an important role in the circadian molecular clock and neural functioning indicates a complex interaction between circadian biology and microRNA biogenesis (Cheng et al. 2007). Anomalies in the expression of any given microRNA will be functionally manifest through dysregulation of its target genes. The targets of some brain expressed microRNA that show circadian patterns of expression may include genes relevant to synaptic function and circadian disruption may thus affect brain synapses through dysregulation of microRNAs that target the mRNA of synaptic proteins.

### **1.5 Epigenetics and autism.**

Discordance for autism in monozygotic twin sibling pairs suggests environmental and/or epigenetic factors might contribute to the aetiology of the disorder and an epigenetic contribution to autism aetiology is also supported by gene expression analysis of monozygotic twins discordant for autism (Bailey et al. 1995; Hu et al. 2006b). Thus any full explanation of the aetiology and prevalence of autism should account for this discordance and the complex patterns of heritability in autism. It is therefore pertinent to consider how: epigenetic mechanisms regulate gene expression, how certain epi-mutations are inherited and what examples of epigenetic psychiatric disorders there are. In this molecular-genetic context, epigenetic, means heritable traits that do not involve changes to the sequence of the DNA nucleotide code and the molecular mechanisms

supporting this form of heritability (Waddington 1953; Holliday and Pugh 1975).

#### 1.5.1 Epigenetic vs. conventional genetic models of autism heritability.

Although a plausible model of inheritance has been proposed (for the broader autism phenotype) that relies upon conventional genetics, this model also relies on evoking a large yet unexplained difference of penetrance for genes conferring risk of autism in males as compared with females (Zhao et al. 2007). Epigenetic factors could contribute to this differential penetrance and the discordance in twin studies (Bailey et al. 1995; Skuse et al. 1997; Skuse 1999; Kaminsky et al. 2006). Furthermore, the existence of a plausible conventional model does not rule-out the possibility of epigenetic mechanisms playing a major role in the disorder.

Given the phenotypic overlap between autism, Rett syndrome and Fragile-X syndrome, it is noteworthy that Fragile-X syndrome and Rett syndrome (DSM-IV), though primarily monogenic disorders, manifest disease through aberrant epigenetic effects. In Prader-Willi syndrome and Angelman syndrome, disorders that also present with aspects of the autism phenotype, pathological effects on imprinting (silencing of chromosome regions in a parent-of-origin manner) are causative. At least two studies have considered autism as an imprinting disorder when viewed from an evolutionary standpoint (Isles et al. 2006; Crespi 2008; Ubeda and Gardner 2010).

A large proportion of all the known imprinted genes (~80 coding genes and ~37 non-coding RNAs) are expressed in the human brain and it has been proposed that this expression pattern supports the hypothesis that

imprinted genes are critical elements in the development and maintenance of social behaviours in humans (Davies et al. 2005; Isles et al. 2006).

In *Apis mellifera* (honeybee), epigenetic effects appear to be required to maintain insect social organisation, as the colony social structure and division of labour is reliant on *de novo* DNA methylation (Kucharski et al. 2008). Queen bees or workers are derived from the same clonal larvae, larval fate being determined by whether or not a given larva is fed royal jelly (Colhoun and Smith 1960; Schmitzova et al. 1998; Barchuk et al. 2007). The gross difference in phenotype and fecundity between workers and queen bees appears to be due to *de novo* DNA methylation as knockdown of the *de novo* DNA methyltransferase Dnmt3 phenocopies the effect of royal jelly feeding (Kucharski et al. 2008). This system presents a good example of the complex interactions between: innate behaviours (feeding regimes), environmental effects (nutrients received), endo-phenotype (worker etc.) and social organisation that are maintained by epigenetic regulatory molecules.

Population genetics modelling suggests that genomic imprinting (the silencing of certain chromosomal regions in a sex-dependent manner) is driven by natural selection in a population where there is asymmetric relatedness between parents and offspring. In this model the matriarchal genome benefits from equal resourcing to all of the mother's offspring over time (as all her offspring will certainly be her own). In contrast and where it is assumed that all the mother's offspring will not (with certainty) share the same father, the paternal genome benefits from the transmission of genes that prioritise taking resource from the mother to enhance the foetus. For example, populations where males are typically more transient than

females would derive such a selective pressure. This conflict or kinship theory is an extension of the parent-offspring conflict theory that may be extended to encompass later stages of development (beyond resource allocation to the foetus) to cover aspects of differential sociability between males and females (Trivers 1974; Haig and Westoby 1989; Moore and Haig 1991; Isles et al. 2006).

In mammals, DNA methylation and histone modification associated with chromatin remodelling are the primary areas of understanding of the epigenetic process. The extra genetic information endowed by DNA methylation and histone modification can be envisaged as an overlay on the DNA nucleotide sequence code, imparting extra information to certain code regions without changing the underlying nucleotide sequence. In vertebrates, DNA methylation involves covalent modification (methylation) of certain cytosine residues of CG dinucleotides within a DNA sequence. When these methylated CG dinucleotides occur in regulatory regions of genes they provide binding sites for molecules such as Methyl CpG Binding Protein 2 (MECP2) that can act as gene expression silencers. The CG dinucleotides are often referred to as CpG dinucleotides, the p representing the phosphodiester bond between the nucleotide residues. In mammals 60-90% of all the CpGs are methylated however CpG rich regions that are coincident with the promoter regions of genes, so-called CpG islands, are predominantly unmethylated (Bird 1986). Within genes, differentially methylated regions occur, often in introns or overlapping the start sites of alternative transcripts. These differentially methylated regions are involved in the regulation of tissue specific gene expression (Miyazaki et al. 2009).

Chromosomal DNA molecules are closely associated with globular nuclear proteins called histones (Kornberg 1974). Octomers of two each of histones H2A, H2B, H3 and H4 make up the nucleosome core around which ~150 nucleotides of the DNA double helix is coiled at approximately 1.6 left handed turns per histone (Luger et al. 1997). Each nucleosome core is separated by a stretch of linker DNA that may be 10 to 100nt in length dependent on species and tissue type and histone H1 binds to the point of entry and exit of the DNA filament on the nucleosome core. Phenotypic effects that depend on covalent modification of the histone proteins of nucleosomes may be transmitted down a lineage of dividing cells by virtue of the incomplete disassociation of DNA and nucleosomes during mitosis (Felsenfeld and Groudine 2003). The modification of histones can involve acetylation, methylation, phosphorylation sumoylation and ubiquitination of particular amino acid residues particularly within the tail of the histones. Each modification alters the strength of the Histone-DNA binding (Crosio et al. 2003; Levenson and Sweatt 2006; Tsankova et al. 2007). Most notably, acetylation of the K14 and K9 lysines of the tail of histone H3 by histone acetyltransferase enzymes (HATs) generally corresponds with a transcriptionally active chromatin structure while methylation of the K9 of H3 is associated with a silent chromatin state.

The clarity of this generalisation is however confounded when multiple modifications occur on the same Histone molecule as when tri-methylation of K9 is associated with transcriptionally active promoters. Tight binding of DNA to histones occurs where genes are transcriptionally repressed while an active chromatin state correlates with an open chromatin structure where the DNA and histones are not strongly associated (Mizzen et al.

1996; Taunton et al. 1996). Modulation between these two forms of chromatin in different tissues and at different times is a fundamental mechanism that underlies the process of differential gene expression required for differentiation, development and circadian gene expression (Doi et al. 2006; Berger 2007).

A further level of complexity involves the protein CCCTC-binding factor, (zinc finger protein) (CTCF). Together with other elements of the epigenetic regulatory process, CTCF plays a role in the formation of higher-order chromatin topology, arranging chromatin into complex loop structures, thought to be pertinent to patterns of gene expression. In this way, long-range chromatin interactions mediated by CTCF can organise distant groups of genes into transcriptionally active units (Cai et al. 2006; Zhao et al. 2006).

Long range changes in the structure of chromatin as well as local changes in the physical and functional accessibility of the DNA to transcription factors are sometimes termed chromatin remodelling, a process that may occur in response to environmental or developmental stimuli. In a qualitatively similar manner differentially methylated regions within genes are recognised by chromatin remodelling complexes that can alter the proximities of regulatory and coding regions *within* the gene. By way of the formation of loop structures, the differentially methylated regions come to lie at the neck of the loop, thus bringing the region immediately upstream of a differentially methylated region into close proximity to the sequence immediately downstream of the next downstream differentially methylated region. Regulation of this form of structural reorganisation within the gene can be correlated with differential

tissue expression patterns and parent-of-origin effects (Murrell et al. 2004).

Histone modifications and subsequent chromatin remodelling can thus be directed by the patterns of DNA methylation at control elements within genes. MECP2 for example recruits histone deacetylase to certain methylated promoters that leads to histone modification (Jones et al. 1998; Kokura et al. 2001). Aberration of the DNA methylation system is causative in Rett syndrome, Prada-Willi syndrome and Angelman syndrome and altered patterns of DNA methylation may also play a role in autism (Lintas and Persico; Nagarajan et al. 2008; Nguyen et al. 2010).

In mammals DNA methylation patterns undergo dynamic changes during development with two periods of global methylation remodelling that occur a) during gametogenesis and b) in the zygote prior to implantation (Morgan et al. 2005). Methylation remodelling during gametogenesis involves a wave of erasure of all methylation marks followed by a wave of remethylation that includes parent (maternal or paternal) specific imprints. This remethylation is mainly accomplished by the de novo methyltransferases DNMT3A and DNMT3B (Jeltsch 2006). The erasure of methylation followed by remethylation in mammalian gametes results in the sperm genome being predominantly methylated compared to the oocyte genome, which is less globally methylated (Morgan et al. 2005). Immediately after fertilization, the sperm genome is rapidly globally demethylated (apart from imprinted regions) while that of the oocyte is gradually demethylated (apart from imprinted regions) (Mayer et al. 2000; Oswald et al. 2000). Upon implantation the methylation



levels in the zygote begin to increase as development proceeds (Meehan 2003).

### 1.5.2 Gene silencing by DNA methylation.

DNA methylation in mammals is generally correlated with the suppression of transcriptional activity. This may be local, as when methylation of CpG dinucleotides in the promoters of genes correlate with suppression of activity of those genes, or, widespread, for example when methylation of many genes on one of the pair of X chromosomes in the human female are inactivated (X-chromosome inactivation). In the case of X-chromosome inactivation, this process ensures that the effective gene dosage of critical X-linked genes in human females is not double that of males.

In humans, chromosomes 7 and 15 also show imprinted regions. Within these regions, in a similar manner to the global hypermethylation of the inactivated X chromosome in females, hypermethylation of CpG dinucleotides leads to gene silencing. However, human somatic cells are diploid and thus imprinting typically occurs on only one of a pair of given chromosomes. Whether the paternal or maternal, copy of a given pair of imprintable chromosomes has its imprintable region hypermethylated, derives the concept of a parent-of-origin effect for the imprint. Thus some imprinted regions may be sex specific, where, for a given gene region, it is typically the male or female copy that is silenced. Serious diseases linked to defective imprinting include the congenital cancer-predisposition syndrome Beckwith-Wiedemann syndrome (DeBaun et al. 2002; Weksberg et al. 2010) and the neurodevelopmental disorders of Prader-

Willi Syndrome (Bittel and Butler 2005) and Angelman Syndrome (Lalande and Calciano 2007).

Prader-Willi Syndrome is characterised by symptoms that include obesity through compulsive eating, muscular hypotonia and mental retardation. The disorder is most commonly caused by lack of a functional paternally derived copy of the chromosomal region 15q11-q13, or maternal uniparental disomy of chromosome region 15q11-q13. In neurotypical individuals certain genes within this region are differentially expressed depending on whether the genes are imprinted on the parentally inherited copy of chromosome 15 or the maternally inherited copy of chromosome 15. Imprinting of chromosome region 15q11-q13, is controlled by the Prader-Willi Syndrome/Angelman Syndrome Imprinting Centre (PWS/AS-IC), within chromosome region 15q11-q13. Thus, in addition to gross deletion and rearrangements of this region, aberrant methylation of genes within chromosome region 15q11-q13 (caused by mutation of the PWS/AS-IC) can cause Prader-Willi Syndrome. About 70% of cases are caused by a *de novo* paternally derived chromosome 15q11-q13 deletion while maternal disomy of chromosome region 15q11-q13 accounts for about 25% of cases. Microdeletions or epimutations of the PWS/AS-IC, along with chromosome 15 translocations, complete the remainder.

The chromosome region 15q11-q13 is also affected in Angelman syndrome and cytogenetic analysis of autism cases shows the chromosome 15q11-q13 imprinted region is also the most common chromosomal region affected (Marshall et al. 2008). Angelman syndrome is characterized by: mental retardation, impaired movement or balance,

characteristic abnormal behaviors, and severe speech and language difficulties. Angelman Syndrome is caused by a lack of a functional (maternal) copy of ubiquitin protein ligase E3A (*UBE3A*) that also lies within the chromosome region 15q11-q13 (Kishino et al. 1997; Matsuura et al. 1997). In a mouse model of Angelman Syndrome *Ube3a* regulates the degradation of guanine nucleotide exchange factor ephexin-5 (*ARHGEF15*) that in turn, regulates brain synapse formation. *UBE3A* is expressed from both paternal and maternal alleles in most tissues but preferential expression of the maternal allele occurs in human brain (Margolis et al. 2010).

For Prader-Willi Syndrome and Angelman Syndrome, perhaps the best “catch all” explanation of the genetic mechanism underlying their pathology involves the functional units of the PWS/AS-ICs: the PWS-SRO and the AS-SRO (SRO = Smallest Region-of-deletion Overlap) (Horsthemke and Wagstaff 2008). In terms of these critical regulatory regions that determine imprinting status in 15q11-q13, deletion of PWS-SRO endows the gene activity and epigenetic modification of 15q11-q13 typical of the maternal state irrespective of its maternal or paternal derivation. While on the other hand AS-SRO deletion in a chromosome with an intact PWS-SRO determines the paternal state regardless of maternal or paternal derivation. Typically, methylation of the PWS-SRO determines the maternal state for 15q11-q13, while lack of methylation of the PWS-SRO determines the paternal state for 15q11-q13. With reference to the periods of demethylation and remethylation during embryogenesis described above, it is clear that imprinting defects in the PWS/AS region could arise from failure to demethylate the PWS-SRO

during spermatogenesis, from failure to methylate the PWS-SRO during oogenesis, or from failure to maintain PWS-SRO methylation after fertilization.

### 1.5.3 MECP2: an epigenetic regulator protein.

The binding of some transcription factors such as E2F Transcription Factor 1 (E2F) or cAMP Responsive Element Binding Protein 1 (CREB) is directly inhibited by the methylation of their respective DNA motif targets and this inhibition correlates with the loss of transcriptional activation normally associated with E2F or CREB binding (Iguchiariga and Schaffner 1989; Campanero et al. 2000). More generally, methyl-CpG binding proteins, e.g. MECP2, MBD1, MBD2 and MBD4 have a higher affinity for methylated as opposed to unmethylated DNA and on association with CpG dinucleotides within promoters and regulatory regions, recruit chromatin remodeling complexes to form a repressive chromatin environment (Jones et al. 1998; Nan et al. 1998; Wade et al. 1999; Fuks et al. 2000). These methyl-CpG binding proteins thus link the phenomena of DNA methylation and chromatin remodeling.

The prototypical methyl CpG binding protein Methyl CpG Binding Protein 2, (MECP2), the gene affected in the autism related disorder of Rett syndrome, plays a role in the process of epigenetic regulation of gene expression (Amir et al. 1999). The high affinity of this DNA binding protein for methylated CpG dinucleotides in gene promoters initially suggests a role as a global transcriptional silencer (Jones et al. 1998; Nan et al. 1998) however later work with *Mecp2* null mice indicates a more limited and selective repressive scope (Tudor et al. 2002; Jordan et al. 2007).

Transcriptional profiling of *Mecp2* knockout cell lines indicates MECP2 binding may also activate certain promoters (Chahrour et al. 2008). Nevertheless, MECP2 mediated regulation of brain-derived neurotrophic factor (BDNF) via the promoter of brain-specific BDNF alternative transcript III is established. MECP2 occupies this promoter in polarised postmitotic neurons. On depolarisation and calcium influx MECP2 leaves the BDNF III promoter, permitting expression of the neural growth stimulant BDNF (Chen et al. 2003; Martinowich et al. 2003) suggesting a mechanism for use dependent synaptic connectivity that might support the processes of learning and memory (An et al. 2008; Kuczewski et al. 2009).

Research into the function of MECP2 is also particularly relevant to the field of neuropsychiatry as mutation of *MECP2* causes the neuropsychiatric disorder of Rett Syndrome (with autism as part of the phenotype). Intriguingly, over-expression of *MECP2* also results in neuropsychiatric phenotypes and autism and it has been proposed that normal neuronal function and development requires a homeostatic mechanism whereby the levels of MECP2 are maintained within strict limits in neurons (Ramocki and Zoghbi 2008). Too much MECP2 and neurons develop hyper-arborisation while too little leads to neurons that are depleted in dendritic spine number and complexity. Males with duplications at the MECP2 locus show an autism phenotype that overlaps with that of typical (MECP2 deleted) Rett syndrome (Meins et al. 2005; Van Esch et al. 2005; del Gaudio et al. 2006; Friez et al. 2006; Smyk et al. 2008). Thus Rett Syndrome, one of the most common causes of mental retardation in females (Moretti and Zoghbi 2006), is now considered to be

phenocopied by *MECP2* deletions and duplications that could both positively and negatively affect *MECP2* levels.

#### 1.5.4 DNA methylation of clock genes in autism.

The social timing hypothesis posits brain oscillators, that are reliant on clock genes that drive circadian rhythms, malfunction in autism (Wimporly et al. 2002). In the hypothesis, this author proposed that epigenetic and/or genetic mutations leads to pathological clock gene expression patterns in autism. Analysis of gene expression patterns in twins discordant for autism show epigenetic effects are likely (Hu et al. 2006b) and a methylation screen of monozygotic twins with discordant autism diagnoses (and their non autistic siblings) highlighted the clock regulated gene B-cell CLL/lymphoma 2 (*BCL2*) and the clock gene RAR-related orphan receptor A (*RORA*) as differentially methylated in the autistic twin compared with the co-twin and unaffected sib. Immunohistochemical analysis of autism brain tissue showed altered expression of *BCL2* and *RORA* compared to controls and genome-wide significant, epigenetic, parent-of-origin effects for autism at the clock homologue (mouse) (*CLOCK*) locus tentatively implicate clock gene methylation in the disorder (Fradin et al. 2010; Nguyen et al. 2010). A study of clock gene promoter methylation in mouse, found major changes in clock gene E-Box region methylation around the perinatal period of development, suggesting a role for developmentally programmed methylation of clock gene regulatory elements (Ji et al. 2010).

## 1.6 Clock genes: The mammalian circadian molecular clock.

In mammals, the core molecular clock, described for example by Looby and Loudon (2005) comprises a suite of epistatic genes that operates as a system of transcription/translation auto-regulatory feedback loops. In Figure 1-3 a core element, aryl hydrocarbon receptor nuclear translocator-like (*ARNTL*) is shown under the synonym *BMAL1*.

A heterodimer consisting of CLOCK/*ARNTL* or neuronal PAS domain protein 2 (*NPAS2*)/*ARNTL* is the positive element of the circadian molecular clock that induces the transcription of the negative elements of the feed back loop (Gekakis et al. 1998; Bunger et al. 2000). The negative elements: period homolog 1 (*Drosophila*) and homologs 2 and 3 (*PER1*, *PER2*, *PER3*); cryptochrome 1 (photolyase-like) (*CRY1*) and cryptochrome 2 (photolyase-like) (*CRY2*); and nuclear receptor subfamily 1, group D, member 1 (*NR1D1* also known as *REV-ERB $\alpha/\beta$* ) repress the action of the CLOCK/*ARNTL* activator and thus derive oscillating expression of the CLOCK/*ARNTL* activator via protein turnover of the PER/*CRY* complex (Shearman et al. 2000). The microprotein, inhibitor of DNA binding 2, dominant negative helix-loop-helix protein (*ID2*) has recently been shown to also negatively regulate the CLOCK /*ARNTL* activator (Duffield et al. 2009). The established roles for this protein in development and immunity suggest a mechanism for clock regulated developmental processes and interaction between the circadian clock and the immune response (Hacker et al. 2003; Rankin and Belz 2011).

A role for the *Drosophila* TIM paralogue TIMEOUT in the mammalian circadian circuit awaits clarification, nevertheless mammalian TIMELESS shows ~24h oscillation and the protein physically associates with PER and

CRY while knockdown of *Tim* in SCN (rat) disrupts neuronal activity rhythms. This role of *TIMELESS* in SCN neuronal activity rhythms may impact circadian behaviour (Barnes et al. 2003; Jin et al. 1999; Kume et al. 1999). Additionally, association of *TIMELESS* with depression and sleep disturbance may reflect clock function (Utge et al. 2010).

The CLOCK/ARNTL heterodimer drives rhythmic expression of clock-controlled genes e.g. basic helix-loop-helix family, member e40 (*BHLHE40* also known as *DEC1*), basic helix-loop-helix family, member e41 (*BHLHE41* also known as *DEC2*), WEE1 homolog (*S. pombe*) (*WEE1*), v-myc myelocytomatosis viral oncogene homolog (avian) (*MYC* also known as *C-MYC*) and MAGE-like 2 (*MAGEL2*) by binding to E-box activator elements (*CACGTG* or *CACNNGT*) in their promoter regions (Hogenesch et al. 1998; Honma et al. 2002; Fu et al. 2002; Matsuo et al. 2003; Kozlov et al. 2007; Mercer et al. 2009).

In addition to the E-box, the D site of albumin promoter (albumin D-box) binding protein (DBP) is under clock-control and binds and activates genes containing D-Box elements (*TTATG[T/C]AA*) (Ueda et al. 2005). The core clock elements, NR1D1 and the RAR-related Orphan receptor A, (*RORA*) regulate in a clock-controlled manner via RRE elements (*[A/T]A[A/T]NT[A/G]GGTCA*) that are not necessarily within the promoter regions of genes (Harding and Lazar 1993). Thus the circadian regulation of genes containing E-Box D-Box and RRE circadian regulatory elements, represent the clock-controlled output from the central oscillator. Some of these clock-controlled genes encode transcription factors, so that directly or indirectly, it is estimated that some 15% of the whole mammalian transcriptome shows circadian regulation per tissue (though a different set



of genes representing this fraction of the transcriptome may be cycling in each of the different tissues (Yang et al. 2007)).

Chromatin modification plays a central role in the mechanism of transcriptional activation/repression seen in the circadian molecular clock. CLOCK functions as a histone acetyltransferase (Doi et al. 2006) and the cyclic expression of CLOCK/ARNTL is also accompanied by a synchronous rhythm in H3 histone acetylation (H3K9) on the promoters of *PER* and *CRY*, catalysed by CLOCK and the histone acetyltransferase E1A binding protein p300 (EP300) (Etchegaray et al. 2003).

To balance the system, experiments with mouse embryo fibroblasts show sirtuin 1 (*SIRT1*), coding a histone de-acetylase that acts upon H3 K9/K14 of ARNTL and other circadian promoters, is under circadian control (Nakahata et al. 2008; Asher et al. 2008). Cyclic methylation of H3K27 on *PER* and possibly *CRY* promoters by the polychrome protein, enhancer of zeste homolog 2 (*Drosophila*) (EZH2) is also shown to be required for the maintenance of circadian rhythm, which, taken altogether, shows chromatin modification as an important facet of the circadian clock mechanism (Etchegaray et al. 2006).

The negative elements of the system (PERs and CRYs) operate via a PER/CRY heterodimer that inhibits the activating effects of the CLOCK/ARNTL heterodimer. The binding of CLOCK to ARNTL causes acetylation of ARNTL at K537 which in turn facilitates binding to and repression of the CLOCK/ARNTL heterodimer by the PER/CRY complex (Hirayama et al. 2007b). This complex also contains splicing factor proline/glutamine-rich (SFPQ/PSF) that recruits SIN3 homolog A, transcription regulator (SIN3A) and HDACs to the *PER1* promoter, where

deacetylation of H3K9 occurs, likely reversing the activating effect of the histone acetylase CLOCK (Duong et al. 2011).

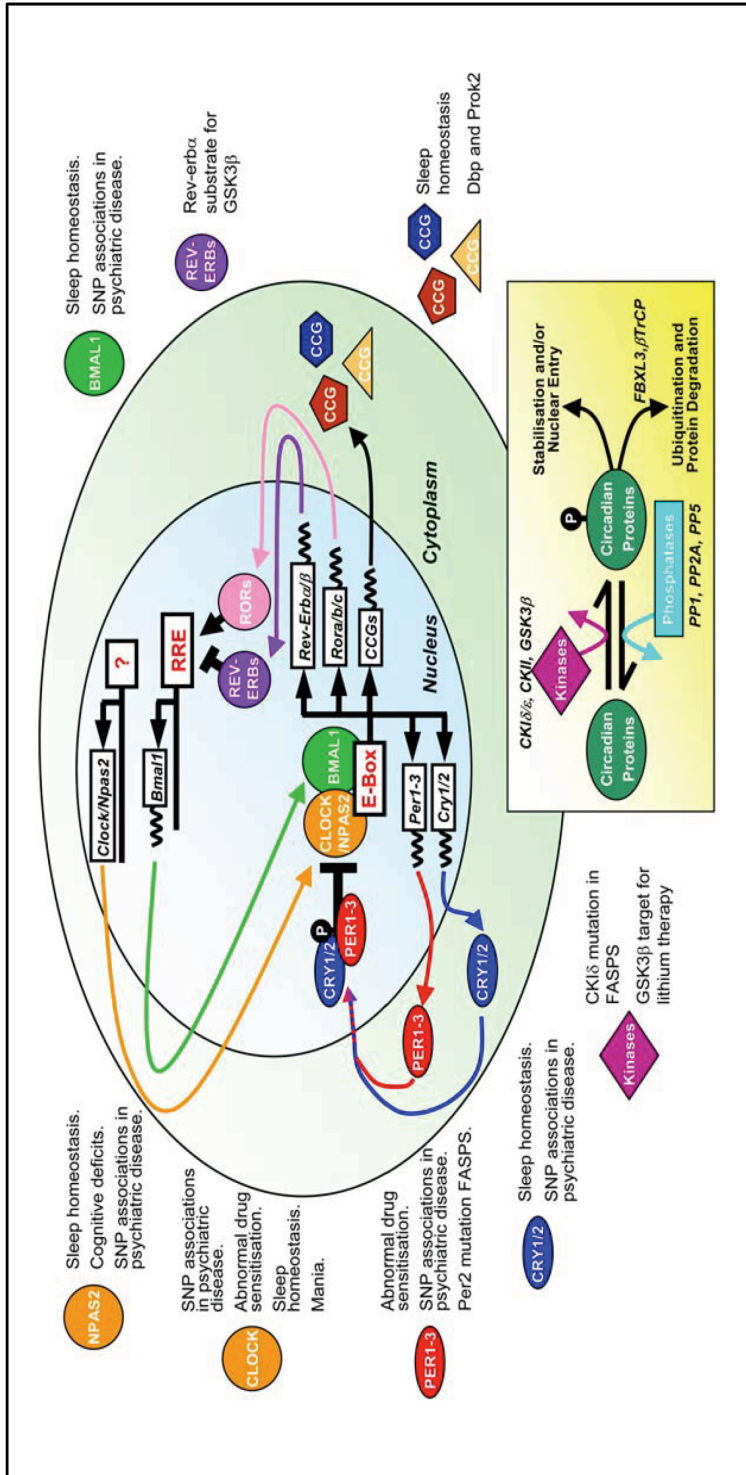
The CLOCK/ARNTL heterodimer also induces *NR1D1* (*REV-ERB $\alpha/\beta$* , a lithium sensitive gene) that represses expression of ARNTL (Preitner et al. 2002; Ueda et al. 2002; Yin and Lazar 2005; Yin et al. 2006). Protein turnover eventually degrades the negative elements, and restoration of the activator complex allows the cycle to start over. The inductive effect of protein turnover of the negative elements is also reinforced by the function of the retinoic acid orphan receptor alpha, beta and gamma proteins (RORs). Whilst being clock-controlled, also induce expression of *ARNTL*, *NPAS2* and *CLOCK* in a feed-forward circuit of the circadian molecular clock (Sato et al. 2004; Akashi and Takumi 2005; Crumbley et al. 2010; Crumbley and Burris 2011). The duration of the molecular migration of proteins and mRNA to and from the nucleus respectively, together with the protein turnover time, broadly defines the ~24h period of this oscillatory system.

To integrate this core clock mechanism into the life of the organism as a whole, genes associated with resetting of the clock (e.g. *CRY1*, *CRY2*) serve in matching environmental time with biological time by resetting the clock in response to environmental cues such as day length (Berson et al. 2002). Casein kinase 1, epsilon (CSNK1E) regulates the clock by affecting the stability of the PER/CRY complex through phosphorylation and other post translational processes e.g. sumoylation and ubiquitination of clock proteins also play important roles in the circadian clock mechanism (Lee et al. 2001; Cardone et al. 2005; Busino et al. 2007). Output pathways transduce the clock's time-of-day signals into multiple physiological

responses that endow organisms with the ability to maintain a state of appropriate physiological readiness that anticipates the environmental demands associated with a particular time of day or night.

Recently it has become apparent that small molecules also play a role in the central circadian oscillator. The discovery that normal clock function in mouse requires the microRNAs miR219 and miR132, shows that the molecular clock in mammals is more complex than previously thought, involving regulation of gene transcripts by RNA interference.

MIR219 is a clock-controlled gene, while MIR132 is light regulated. Target validation in vivo confirms: regulatory factor X, 4 (*RFX4*); PH domain and leucine rich repeat protein phosphatase 1, (*PHLPP1 (SCOP)*); and, Rho GTPase activating protein 32 (*ARHGAP32*) as targets for these microRNAs. These proteins are predicted to exert their confirmed influence on clock function through roles in the processes of signal integration, chromatin remodeling and cellular excitability; that are thus modulated by oscillating levels of miR132 and miR219 (Cheng et al. 2007). This microRNA requirement is additional to transcription/translation auto-regulatory feedback and post-transcriptional regulation by kinases and the deadenylase CCR4 carbon catabolite repression 4-like (*S. cerevisiae*) (*CCRN4L* also known as nocturnin) (Garbarino-Pico and Green 2007).



**Figure 1-3. The circadian molecular clock and psychiatric disorders.**

Reproduced from Barnard and Nolan (2008). Interactions of the clock genes in the central molecular clock mechanism of mammals. The main activator, the CLOCK/ARNTL or NPAS2/ARNTL heterodimer (ARNTL = BMAL1 in the diagram) has affinity for promoter located E Box elements. CLOCK/ARNTL binding at E Boxes (CACGTG or CACMGT) induces clock-controlled genes (CCGs e.g. *WEE1*) that include the transcription factors CRY and PER (inhibitors of the activator complex) such that one activation-repression cycle defines the circadian period of ~24h. CSNK1E (shown as CK1e in the diagram) phosphorylates the PER/CRY complex leading to its degradation and concomitant reactivation of the CLOCK/ARNTL complex. The diagram also indicates that the majority of these genes have been linked to neuropsychiatric disorders. GSK3β is highlighted as a target for lithium treatment in bipolar disorder. The feed-forward circuit indicated with a question mark is now confirmed and RORA/NR1D1 promotes the transcription of NPAS2/CLOCK in the core activator (Akashi M. and T. Takumi 2005; Crumbley et al. 2010; Crumbley and Burris, 2011).

The introduction of miRNAs to the core clock-mechanism also adds complexity in terms of our understanding of the pleiotropic effects of the clock, as microRNAs are powerful agents of multi-targeted gene regulation. The inclusion of genes of particular relevance to autism such as *BDNF* and *MECP2* (see page 29) amongst the verified targets of the circadian mirR219 (Klein et al. 2007), suggests a possible role for circadian microRNAs in neuropsychiatric disorders. Links between the circadian molecular clock and the microRNA biogenesis pathway thus suggests how an altered circadian clock could manifest currently unexplained effects on memory formation (Stickgold 2005) and the co-occurrence of circadian and neurobehavioral phenotypes (Figure 1-3).

In addition to their role as clock elements, certain clock genes e.g., *PER1*, *PER2* and *NPAS2* appear to play roles in signaling pathways, in DNA repair and in cancer (Chilov et al. 2001; Rutter et al. 2001; Dioum et al. 2002; Fu et al. 2002; Gery et al. 2006). *RORA* and *PER2* function in cerebellum development and stem cell proliferation in the adult hippocampus respectively (Gold et al. 2007; Borgs et al. 2009) and such studies point toward a broader functionality for clock genes beyond a specific role in circadian rhythm generation.

### **1.7 The circadian clock and genome stability: autism and cancer.**

After many years of research on mutations that disrupt the cell division cycle in both higher and lower organisms, it has only recently become clear that the circadian molecular clock plays a crucial role in the temporal structuring of the cell division cycle (Matsuo et al. 2003; Chen et al. 2007). Crosstalk between the circadian clock and the cell division cycle is

purported to give at least two fundamental advantages. Firstly, the temporal separation of DNA replication from incompatible metabolic processes protects DNA from, for example, oxidative damage (Chen and McKnight 2007). Secondly, processes that are vulnerable to adverse cyclic environmental factors, e.g. ultra violet radiation in sunlight, may be temporally-protected by being gated into safe time-windows, in this example, night time (Danilova et al. 2004).

Similarities between the signalling pathways that co-ordinate the DNA damage response and the circadian clock, are presented as suggesting a common origin for these two phenomena (Uchida et al. 2010). This idea, developed on a zebra fish model, is in keeping with wider evidence that the circadian molecular clock is involved in both temporal regulation (Reppert and Weaver 2002) and genome repair and replication (Gery et al. 2006; Kondratov and Antoch 2007).

In zebrafish, a common light induced signalling pathway regulates both the circadian clock and the cell cycle. Reactive oxygen species (ROS) produced by the action of sunlight on the fish derives photo-oxidative stress that activates a redox signalling cascade involving the cell cycle control regulator zWEE1 (Hirayama et al. 2005), the clock gene product zCRY1a and the DNA repair protein zPHR (Hirayama et al. 2007a; Hirayama et al. 2009). In this system, sunlight is an environmental hazard, a zeitgeber to the circadian clock and a promoter of photolytic DNA damage repair.

When yeast is grown under naturalistic, low-nutrient conditions, the cells show a 4 to 5 hour rhythm of alternating glycolytic (reductive) and respiratory (oxidative) phases. DNA replication is confined to the reductive

phase of this cycle when the potential for oxidative damage to the replicating DNA is less (as compared to during the oxidative phase). Mutants that force the cells to undergo DNA replication in the oxidative phase derive higher levels of spontaneous mutation while deletion of the checkpoint kinase, suppressor of Mec1 Lethality (*sm11*) that un-couples the circadian rhythm from these metabolic rhythms also de-synchronises the metabolic and cell cycles and increases mutation rates (Chen et al. 2007).

#### 1.7.1 Clock genes and the DNA damage response.

The DNA damage response can be envisaged as comprising not only DNA repair and the activation of DNA damage checkpoints but also apoptosis, and transcriptional reprogramming, and that each of these is gated by the circadian-clock. The DNA repair process includes mechanisms that involve direct repair, nucleotide excision repair, base excision repair, and double-strand break repair. The circadian clock may play a role in a number of these processes but so far, only nucleotide excision repair is shown to be under the strict control of the circadian molecular clock (Kang and Sancar 2009).

The Nucleotide excision repair mechanism is responsible for correcting a wide range of DNA damage that includes pyrimidine dimers and the di-adducts induced by *cis*-platin treatments. In this process, firstly, replication protein A1, 70kDa (RPA), xeroderma pigmentosum, complementation group A (XPA) and xeroderma pigmentosum, complementation group C (XPC) bind to the DNA damage site and recruit the general transcription factor IIH complex (TFIIH complex) that allows unwinding of the DNA

strands into the pre-incision complex 1. XPC bound to RAD23 homolog B (*S. cerevisiae*) (XPC-RAD23B) leaves the complex and is replaced by excision repair cross-complementing rodent repair deficiency, complementation group 5 protein (ERCC5) to derive the pre-incision complex 2 where XPA recruits excision repair cross-complementing rodent repair deficiency, complementation group 4 protein (ERCC4) to form the pre-incision complex 3. ERCC4 (within the complex) cuts the DNA strands ~23 nucleotides 5' of the damage and ERCC5 cuts the DNA strands between ~4 nucleotides 3' of the lesion (Huang et al. 1992). Thus a ~28 nucleotide oligomer is released together with the repair factors that have served their role. RPA remains bridging the excision gap and recruits replication factor C (activator 1) 1, 145kDa protein (RFC1) and proliferating cell nuclear antigen protein (PCNA). DNA polymerases delta and or epsilon extend the DNA strands across the gap that is finally ligated by ligase I, DNA, ATP-dependent (LIG1). Disruption of the excision repair pathway by mutation causes the cancer prone and photosensitivity related syndrome of xeroderma pigmentosum. Individuals with this condition also show higher rates of neurological abnormalities (Sancar 1996).

XPA is a clock-controlled gene. It shows circadian patterns of expression in mouse liver and brain (Kang et al. 2009; Kang et al. 2010) and has two E-boxes in the XPA 5' promoter region. XPA is ubiquitinated by hect domain and RLD 2 (HERC2), which is itself clock regulated in phase with XPA possibly contributing to entraining the half-life turnover time for XPA to ~3h (Sancar et al. 2010; Kang et al. 2010). Thus a critical component of the nucleotide excision repair mechanism is under circadian control and therefore, perturbations in the normal running of the circadian



molecular clock or the organismal clock driven by the SCN could compromise the process of nucleotide excision repair.

### 1.7.2 DNA damage checkpoints.

To maintain genome integrity, DNA damage repair must be carried out before the process of DNA replication and cell division is complete. The mammalian cell cycle has two time points, called check points, where progress through the cycle is slowed or halted in response to the presence of damage to the dividing DNA. Firstly the ATR-CHK1 pathway is activated in response to DNA damage induced by ultra violet light (UV), chemicals that produce similar modification to that of UV, or chemicals that cause stalled replication forks. Secondly the ATM-Chk2 pathway is the primary DNA damage response to double-strand breaks that result from ionizing radiation (IR) or chemicals that mimic the effects of IR.

The structure of these two damage response pathways can be considered in terms of damage sensors, signal transducers and effector proteins. The primary sensor kinases, ataxia telangiectasia mutated (ATM) and ataxia telangiectasia and Rad3 related (ATR) relay the damage response signal through the signal transducing kinases of CHK1 and CHK2 and on via effectors such as tumor protein p53 (TP53 also known as p53), cell division cycle 25 homolog C (*S. pombe*) (CDC25) and cell division cycle 45 homolog (*S. cerevisiae*) (CDC45) to cyclin-dependent kinase 1 (CDK1 also known as CDC2) and cyclin-dependent kinase 2 (CDK2) where phosphorylation of CDK1 and CDK2 results in cell cycle arrest at the G2/M and G1/S boundaries respectively. It is now clear that

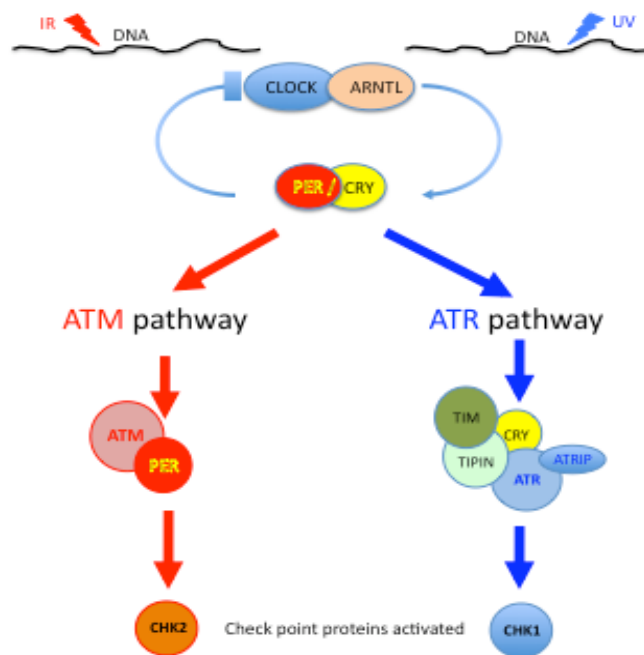
certain clock genes also take part directly in this signal transduction mechanism.

In the ATM pathway; CRY, timeless homologue (*Drosophila*) (TIMELESS) and timeless interacting protein (TIPIN) join ATR and ATR interacting protein (ATRIP) in a complex that activates the CHK1 protein (Unsal-Kacmaz et al. 2005). In the ATM pathway PER1 interacts with ATR in the activation of CHK2 (Gery et al. 2006). The circadian clock thus interfaces with the cell cycle in two ways; firstly, by producing circadian oscillations in the availability of clock-controlled cell cycle proteins such as; WEE1, C-MYC, cyclin-dependent kinase inhibitor 1A (p21, Cip1) (CDKN1A) and XPA and secondly, by elements of the circadian molecular clock itself e.g. PER, TIMELESS and CRY proteins being part of the damage response signal transduction cascade (Fu et al. 2002; Matsuo et al. 2003; Sancar et al. 2004; Unsal-Kacmaz et al. 2005; Gery et al. 2006; Grechez-Cassiau et al. 2008). The terms serial connection and direct connection respectively have been applied to these two forms of clockgene involvement in the workings of the cell cycle.

### 1.7.3 Circadian control of apoptosis.

During the cell division cycle and in response to the presence of DNA damage, commitment to activate DNA damage response pathways or apoptotic pathways will be determined largely by the degree of any DNA damage carried by the cell. Apoptosis of cells carrying irreparable mutations thus protect the organism from harbouring a growing number of defective mutant cells. Apoptosis also plays a benign role in embryogenesis and development where programmed cell death is an

intrinsic part of the expansion and re-modelling of tissue structures and systems.



**Figure 1-4. Circadian clock proteins PER, CRY and TIMELESS in the ATM and ATR checkpoint pathways.**

IR indicates DNA damage induced by ionising radiation and UV indicates DNA damage induced by ultra violet radiation. Mutations in circadian clock genes may thus make neurons more vulnerable to the effects of DNA damage. The diagram also suggests that DNA damage might affect the clock by limiting the availability of the PER/CRY complex when either or both of these proteins are recruited to the checkpoint pathways.

The two canonical apoptotic pathways, the death receptor pathway and the mitochondrial dependent apoptotic pathway, are under circadian control by virtue of the central role of clock-controlled genes in both these pathways. In the death receptor pathway the product of the clock-controlled tumor necrosis factor (TNF) gene, on binding to membrane bound death receptors, initiates a cascade of caspase enzymes that leads to expression of DNA fragmentation factor, 40kDa, beta polypeptide

(caspase-activated DNase) DFFB/CAD and subsequent chromosomal breakdown (Hotchkiss et al. 2009). In the mitochondrial apoptotic pathway TP53 acts as a regulator of the pro-apoptotic proteins BCL2-associated X protein (BAX) and BCL2-antagonist/killer 1 (BAK1). Though not strictly shown to be rhythmic, perhaps because of the complexity of its interactions, TP53 expression is upregulated by ARNTL and down regulated by CRY proteins and thus is potentially under the influence of oscillating levels of these canonical clock proteins (Mullenders et al. 2009).

#### 1.7.4 The circadian clock and cancer development.

The onset of cancer is thought to involve a breakdown of the cell's normal response to DNA damage and as clock genes operate in DNA damage response pathways, it is not surprising that the disruption of circadian rhythms is associated with increased rates of cancer and epidemiology implicates disruption of circadian rhythms per se as an oncogenic factor (Davis et al. 2001). However, experiments with clock gene mutant mice also suggest a complex clock gene specific link to cancer development.

For example  $Cry1^{-/-} Cry2^{-/-}$  mice are arrhythmic but do not show increased rates of cancer as compared with wild type mice (Gauger and Sancar 2005). Further,  $p53^{-/-}$  knockout in mice normally increases cancer rates but in  $Cry1^{-/-} Cry2^{-/-}$  mice,  $p53$  knockout reduces the rate of early onset cancer and extends their median lifespan. Investigations of this paradox in cell lines showed  $p53^{-/-}$  cells compared with  $Cry1^{-/-} Cry2^{-/-} p53^{-/-}$  cells were similar with respect to UV-induced DNA damage repair and UV-induced ATM and ATR checkpoint activation but differed with respect to

sensitivity to genotoxin induced apoptosis (including UV-induced apoptosis) which was raised in the *Cry1<sup>-/-</sup> Cry2<sup>-/-</sup> p53<sup>-/-</sup>* mice (Ozturk et al. 2009). The absence of an expected incremental increased rate of cancer due to the *p53<sup>-/-</sup>* mutation added to a *Cry1<sup>-/-</sup> Cry2<sup>-/-</sup>* background may be explained by a compensating effect of *Cry1<sup>-/-</sup> Cry2<sup>-/-</sup>* dependent increased apoptosis in the *Cry1<sup>-/-</sup> Cry2<sup>-/-</sup> p53<sup>-/-</sup>* triple mutant.

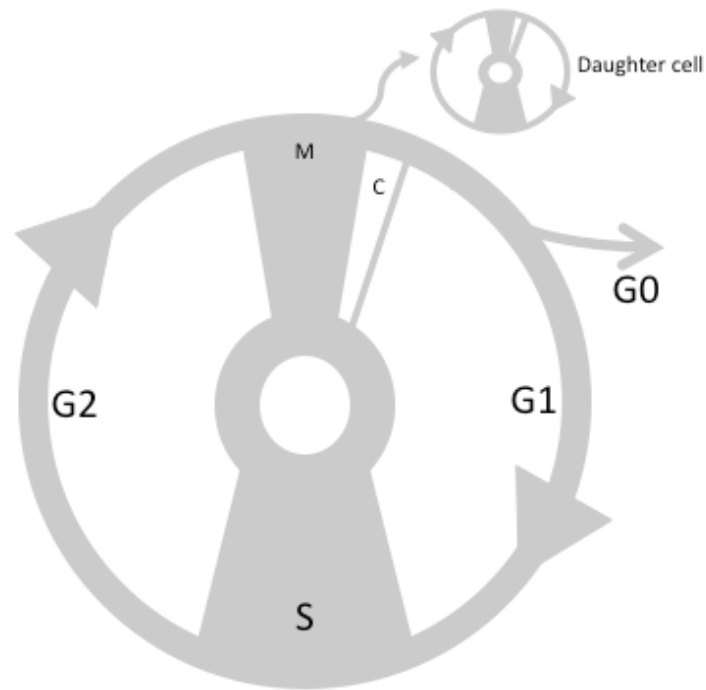
Studies with *Clock* and *Bmal1* mutant mice further indicate that rather than circadian disruption per se, it is pathological changes in the levels of certain clock proteins that are the causative link between circadian clock disruption and cancer (Gorbacheva et al. 2005; Kondratov et al. 2006; Antoch et al. 2008). In this respect it is noteworthy that altered expression of PER proteins are linked to breast tumour development (Fu et al. 2002) which may be driven by anomalous promoter methylation of PER proteins (Chen et al. 2005). Even though the evidence reviewed by Sancar et al., (2010) argues in favour of a focus on specific clock genes in cancer development i.e. PER, CRY and TIM, other studies not cited by their review highlight other clock and clock related genes in cancer development e.g. *NPAS2* (Hoffman et al. 2008), *EZH2* (Etchegaray et al. 2006; Puppe et al. 2009) and *RACK1* (Hu et al. 2006a; Al-Reefy et al. 2010; Robles et al. 2010). The growing number of reports of other clock and clock-controlled genes in cancer development suggests that attempts to define the role of the circadian clock in cancer development as centred on specific clock genes or on the circadian rhythm itself is somewhat tautological.

Although the exact mechanism by which clock genes impact upon DNA damage repair mechanisms is currently not fully understood

nevertheless, certain circadian clock proteins, CRY, PER1 and TIMELESS are shown to take part in the molecular mechanisms that underlie the DNA damage response and DNA repair mechanisms in mammalian cells (Unsal-Kacmaz et al. 2005; Gery et al. 2006).

An extra facet is added by considering the need for chromatin remodelling during the DNA damage repair process. The canonical circadian protein CLOCK is a histone acetyltransferase and together with K (lysine) acetyltransferase 5 (KAT5 also known as Tip60) is involved with DNA repair that depends on chromatin remodeling (Miyamoto et al. 2008). The histone acetyltransferase activity of CLOCK is also required for its role in circadian rhythm generation (Doi et al. 2006). The histone acetyltransferase EP300 appears to modulate the tissue specificity of clock gene expression by forming either a co-activator or co-repressor complex with PCAF/CLOCK/ARNTL or HDAC3/CLOCK/ARNTL respectively that targets the CLOCK/ARNTL heterodimer, with concomitant activation or repression respectively (Hung et al. 2007; Hosoda et al. 2009). EP300 acetylates H3-K56 and co-localizes with DNA repair sites, while SIRT1 deacetylates H3-K56 (Das et al. 2009).

Intriguingly SIRT1 is a critical component of the feedback loop that links circadian rhythms with metabolism. SIRT1 regulates the core circadian clock by its requirement for high amplitude circadian expression of the core clock genes. Also, cellular oxidative capacity (NAD<sup>+</sup> levels) shows a circadian rhythm driven by CLOCK/ARNTL and SIRT1 (Nakahata et al. 2009).



**Figure 1-5. Cell cycle of a typical somatic cell.**

M = mitosis and C = Cytokinesis. During G1, Cellular contents excluding the chromosomes are duplicated, S: chromosomes are duplicated, G2: most DNA damage repair occurs.

SIRT1 regulates the circadian expression of nicotinamide phosphoribosyltransferase (NAMPT, the rate-limiting enzyme in the NAD<sup>+</sup> synthesis pathway) and thus determines a circadian rhythm in the availability of its own cofactor, NAD<sup>+</sup>. This is reminiscent of the situation in yeast where the circadian and metabolic cycles are coupled (Chen et al. 2007). On this theme, recent findings implicate SIRT1 in the maintenance of genome stability via a mechanism involving the deacetylation of H3-K56 (Yuan et al. 2009) and in NAD<sup>+</sup> related excitotoxicity of energetically compromised neurons (Liu et al. 2009).

In the SCN, circadian rhythms driven by the CLOCK/ARNTL heterodimer can be phenocopied by the NPAS2/ARNTL heterodimer

(Kennaway et al. 2006; DeBruyne et al. 2007). However, NPAS2 does not possess HAT activity. Therefore, the reliance of CLOCK/ARNTL driven circadian rhythms on the HAT activity of the CLOCK protein (Doi et al. 2006) suggests that the NPAS2/ARNTL heterodimer generates circadian rhythms by virtue of other HAT containing molecules that associate with NPAS2/ARNTL on circadian promoters. Asher and Schibler (2006) propose EP300 and ATR as possible cofactors lending extra HAT capacity to the CLOCK/ARNTL or NPAS2/ARNTL complex.

#### 1.7.5 Cancer and autism.

If disrupted circadian molecular clocks contribute to autism, it follows that altered cancer rates could be more likely in this subpopulation. This increased risk may be additional to an increased risk of cancer in autism that is due to overlap between genes linked to autism development that are also linked to cancer development.

Certain cancer related genes also appear to be involved in autism. For example, mutations in Tuberous Sclerosis Complex (TSC) genes encoding hamartin, tuberous sclerosis 1 (*TSC1*) and tuberin, tuberous sclerosis 2 (*TSC2*) the phosphatase and tensin homolog (*PTEN*) and the met proto-oncogene (hepatocyte growth factor receptor) (*MET*) are each linked to forms of cancer and increased risk of autism (Wiznitzer 2004; Gentile et al. 2008; Orlova and Crino 2010). However, whether the neuropsychological phenotype is primarily due to secondary effects caused by the physical disruption of brain tissues by tumors (*TSCs*) or altered brain growth patterns (*PTEN* and *MET*) remains to be seen. Nevertheless, where these factors are taken into account, there remains a



highly significant association between breast and uterine cancer and autism (Kao et al. 2010). This finding might have relevance to altered clock gene expression in autism (Hu et al. 2009a) as a number of studies implicate altered clock gene expression in breast cancer development (Chen et al. 2005; Zhu et al. 2008).

Increased rates of cancer in autism (Rzhetsky et al. 2007; Kao et al. 2010) Increased genome instability (Pinto et al. 2010) in the disorder; anomalous expression of clock genes in severe autism (Hu et al. 2009a) and abnormal patterns of promoter methylation of the clock gene *RORA* (Nguyen et al. 2010) together with genetic association of *PER1* and *NPAS2* (Nicholas et al. 2007) tentatively suggests that circadian clock anomalies might contribute to both the neuropsychiatric phenotype and increased cancer susceptibility in the severe autism subpopulation.

### **1.8 Autism and social timing: timing in social communication is a cross species phenomenon with genetically predetermined aspects.**

Interpersonal timing difficulty in autism has been highlighted as a critical element of this predominantly male-affecting disorder and hypotheses and evidence of a role for circadian, communicative and/or neurological aspects of timing in autism are documented (Segawa et al. 1981; Newson 1984; Richdale and Prior 1992; Courchesne et al. 1994; Boucher 2001; Brock et al. 2002; Wimpory et al. 2002; Welsh et al. 2005; Hesling et al. 2010).

Boucher (2000, 2001) and Brock et al. (2002) suggest a core timing-deficit presenting different manifestations by its effect on the elements of

an integrated system of neural and physiological oscillators in autism. Wimpory et al. (2002) proposed a causative, concurrent and developmental role for timing deficit in autistic disorder and this author contributed the idea that this deficit may be derived from pathological genetic/epigenetic variations in clock genes (Wimpory et al. 2002).

In terms of the above hypotheses, timing deficit may manifest in autism as both temporally measured anomalies in social interaction (social timing) and apparently disparate symptoms such as anomalous performance in tasks involving perceptual and cognitive coherence that require synchronization of neural oscillators in distinct brain regions. (Feldstein et al. 1982; Tantam et al. 1993; Sears et al. 1994; Grice et al. 2001; Townsend et al. 2001; Inui and Asama 2003; Schmitz et al. 2003; Szelag et al. 2004; Brown et al. 2005; Gowen and Miall 2005; Haist et al. 2005; Trevarthen and Daniel 2005; Bebko et al. 2006).

In typically developing infants, rhythmic turn taking in proto-conversations precedes the emergence of verbal language, and this coordinated interpersonal timing facilitates social interaction (Jaffe et al. 2001). In a study of typically developing and Down syndrome infants, it was shown that the ability to coordinate interpersonal timing in the Down syndrome children was similar to that of normal children (though Down syndrome children were delayed in the development of this faculty) (Jasnow et al. 1988). This, together with the observations that very young infants engage synchronously in proto-conversations with their caregiver, suggests that this social timing ability is not strongly dependent on intellect and that it precedes (and is therefore not derived from) the cognitive-symbolic functioning associated with word language. This point is re-

focused by the Social-Timing hypothesis that posits a gene-derived deficit in social timing ability in autistic infants disrupts early infant-mother interaction and consequently contributes to the development of abnormal sociability and language (Wimpory et al. 2002).

Jasnow et al., (1988) proposed that the ability of infants to perform synchronized turn-taking interactions from an early age reflects a functionally adaptive social predisposition common to many species including frogs, birds, and insects that use sonic communication in a social context (Jasnow et al. 1988). For example, a neural network in the brain of the female cricket *Gryllus bimaculatus* acts as a temporal filter such that only certain temporal patterns within the male cricket's song will elicit phonotactic response in the females (Schildberger 1984). A similar (in principle) neural auditory filter has been described in toads.

The closely related species *Bufo americanus* and *Bufo woodhousii fowleri* often inhabit the same ponds and are capable of inter-breeding. The identity of these two species is however maintained primarily by differences in the temporal structure of the mating call for each species. Neurons in the mid brain of the toads perform time domain analyses of inputs from the auditory pathway and act as a neural filter, only transducing signals corresponding to mating calls possessing the appropriate species-specific temporal pattern (Rose and Capranica 1984). Neurons of the inferior colliculus (*Rattus*) process the temporal patterning of sounds to determine social relevance (Moller 1983) indicating that a neural mechanism for the detection of temporally patterned socially relevant auditory output from conspecifics is likely to be widespread amongst animal species (Langner 1992). Genetically determined neural

circuitry is also required for the production of temporally patterned socially relevant auditory output from conspecifics. This concept is reinforced by genetic studies in the fruit fly *Drosophila*.

In *Drosophila*, the *period* gene (*per*) influences female receptivity to song and temporal elements of song production in males (Kyriacou and Hall 1980; Kyriacou et al. 1992; Greenacre et al. 1993). Song production in *Drosophila* is innate (Billeter et al. 2006) and a number of areas of the male *Drosophila* brain are implicated in song production with particular focus on the “3 + 1” neuron bundles and Giant Descending Neurons that are required to be male typical (i.e. not feminized) for song production (Moran and Kyriacou 2009).

There is evidence to support the notion that behavioural rhythms: seasonal, circadian, ultradian and sonic communicative rhythms, may share genetic substrates. For example: Seasonality in humans (Kovanen et al. 2010), time point of mating in the melon fly *Bactrocera cucurbitae* (Miyatake et al. 2002), duration of copulation in *Drosophila* (Beaver and Giebultowicz 2004) and interpulse interval in pulsatile secretion of leutenising hormone and cortisol in Syrian hamster (Loudon and Wayne 1994) are all influenced by clock gene mutations. Thus the characteristic impairments of temporal aspects of perceptual and productive prosody and circadian anomalies in autism (Hesling et al. 2010; Peppe et al. 2007; Glickman 2010), fosters the notion that clock genes may be involved in the aetiology of the disorder.

### 1.8.1 Clock genes and sex-determining genes are implicated in forms of social communication in *Drosophila*.

Female *Drosophila* show a circadian rhythm of mating activity that is dependent on a functional circadian molecular clock in the fly's lateral neurons of the brain and it is this cyclic receptivity of the female that determines mating (Sakai and Ishida 2001). Prior to mating, male flies court females with gesture, grooming and a courtship song that is comprised of two parts: a hum (sine song) and a series of pulses (pulse song). In pulse song, the rests between the pulses (interpulse intervals, *ipi*) show a cyclic modulation in length, over the duration of delivery of the song. Studies with *Drosophila per* mutants show this cycle of ~55 seconds (K&H cycle) is regulated by the clock gene *per* (Kyriacou and Hall 1980; Alt et al. 1998) and is altered in a qualitatively similar way to which the same *per* mutations affect the circadian cycle of locomotion. In *Bactrocera cucurbitae* (melon fly), circadian mutants with short circadian locomotor rhythms also produce short pulse train intervals in the male courtship song (Miyatake and Kanmiya 2004).

These observations show *per* operating not only in a circadian oscillator but also in modulating high-frequency oscillators controlling motor function associated with a communicative behavior (Konopka and Benzer 1971; Konopka et al. 1996; Ritchie et al. 1999). Further, where male flies interrupt their song, they restart singing in phase with the initial portion of the song and each fly's song may start at a different phase of the K&H cycle. This suggests that the *per*-determined ~55s K&H cycle is integral with a concurrent, short-interval timing process.

Germ line transformation experiments in *D. melanogaster* and *D. simulans* show that the region of *per* that determines the species-specific variation in pulse song lies within the central and largest exon of the gene (Wheeler et al. 1991). This K&H cycle determining region (KDR) consists of a Thr-Gly repeat motif flanked upstream by 60 amino acids and downstream by 122 amino acids. The Thr-Gly repeat motif is polymorphic within each of these two species, suggesting that it is not the number of repeats that regulates the K&H cycle. More likely, species-specific amino acid substitutions in the 122 amino acids on the 3' flank of the Thr-Gly repeat motif are the K&H cycle determinants (Wheeler et al. 1991).

The Thr-Gly-repeat region forms a conserved type-2 beta turn secondary structure (Guantieri et al. 1999) even though the amino acid repeat sequence varies between *Drosophila* species. This structure/region appears to function in some way that compensates for the effects of temperature variation on the circadian clock (Sawyer et al. 1997). Possibly, the length of the Thr-Gly region may effect temperature regulation indirectly, and at the level of transcript processing, by influencing alternative splicing of the *per* transcript.

Initially three transcripts were identified for the *per* locus in *Drosophila* however, only two transcripts were subsequently confirmed *in vivo* (Citri et al. 1987; Cheng et al 1998). These two transcripts differ only in that an 89bp intron is excised from transcript B compared to transcript A, where the latter retains this intron (located in the 3'UTR of the gene). Relative levels of transcript A and B determine the length of the cycle of locomotor activity. High temperature and long days enhance the level of transcript A by suppressing the splicing of the alternative 3'UTR intron and this leads

to a delay in the onset of evening activity and a longer midday nap. This is in keeping with a summer solar regimen and indicates an adaptive mechanism for seasonal compensation of the clock (Cheng et al. 1998; Collins et al. 2004).

This circadian phenotype may be manifest *via* the regulatory effect of microRNA/s acting on the 3'UTR of the *per* mRNA. Removal of the UTR intron from transcript A (to form transcript B) would deny transcript B any microRNA target sites present in the excised intron. Thus regulation of relative levels of *per* by microRNA may be responsible for the phase shifts in response to temperature. This possibility is an analogy of microRNA regulation of *Drosophila Ubx* where the length of the *Ubx* 3' UTR is determined in a cell type and developmental stage specific manner (not by temperature), the shorter transcript denying the possibility of regulation by *mir-iab-4* (Miura et al. 2011).

Alternative splicing determines relative levels of the long and short isoforms of the fungal clock protein FRQ, endowing the *Neurospora* circadian clock with temperature compensatory capacity. In this case temperature determines the formation of an alternative intron in the 5' region of the gene and splicing of this intron removes the translation initiation site for the long transcript (Diernfellner et al. 2005). These examples suggest alternative splicing is possibly a conserved mechanism for buffering the clock against diurnal temperature fluctuation. The mechanism for this process appears to rely on the physical effect of temperature on the binding of the spliceosome to weak splice signals within the pre-mRNA, where cool conditions promote binding while higher

temperatures inhibit binding and therefore splicing (Ladd and Cooper 2002; Low et al. 2008).

Splicing of the 3' terminal intron in *per* is regulated by protein arginine methyl transferase 5, (*PMRT5 csul/dart5*) that catalyses methylation of arginine residues in Sm spliceosomal proteins and histones (Bedford and Richard 2005; Gonsalves et al. 2006; Anne et al. 2007; Sanchez et al. 2010). The requirement for alternative splicing of transcripts of genes that function in clock output pathways (e.g. the *Drosophila* genes *takeout* and *slowpoke (KCNMA1)*) indicates a co-requirement of alternative splicing and the circadian clock.

Phospholipase-C appears to operate as a signaling molecule in the temperature sensitive splicing of the 3'UTR intron in *per* (Collins et al. 2004). The association of circadian regulatory elements with the mammalian orthologue of *norpA* (phospholipase C, beta 4 (*PLCB4*)) (Kumaki et al. 2008) further suggests a role for the circadian clock in alternative splicing, by regulating signaling in this process. The involvement of the circadian clock in alternative splicing is more explicit in *Arabidopsis* where activation of the spliceosome by PMRT5 dependent arginine methylation is under circadian control via clock regulation of PMRT5 (Sanchez et al. 2010).

The non-POU domain containing, octamer-binding protein NONO is also implicated in splicing, RNA export and transcriptional repression (Shav-Tal and Zipori 2002). NONO forms complexes with PER and the histone methyl transferase WDR5 and their location within the cell varies as a correlate of *PER* expression (Brown et al 2005; Gori et al. 2001; Wysocka et al. 2003). Thus PER might regulate the functionality of these



proteins (in splicing and histone methylation) through compartmentation. *NONO/nonA* is essential for normal circadian rhythm in *Drosophila* and mammals as it is required for cAMP-dependent activation of CREB target genes. Additionally the *Drosophila* orthologue *nonA* regulates aspects of male courtship behaviour that includes song production (Kyriacou and Hall 1980; Kulkarni et al. 1988; Campesan et al. 2001; Amelio et al. 2007).

The circadian clock and sex-determination pathways thus share certain genes required for the manifestation of male courtship behaviour. The sexual identity of the nervous system in *Drosophila* is determined by, the gene *fruitless (fru)*. The protein encoded by the gene *transformer* determines male-specific and female-specific splice variants of mRNA transcribed from *fru* and this process is tissue-specific, in accord with the somatic sex of the fly (Ryner et al. 1996; Demir 2005; Vilella et al. 2005; Manoli et al. 2005; Salz 2011). Intriguingly, *fru* and another gene critical to the sex determination pathway, *Sexlethal (Sxl)* are clock regulated in *Drosophila* brain (Kadener et al. 2006).

The genes *takeout (to)* and *Neuropeptide F (npf)* are downstream of *fru* in the sex determination pathway and each are also output genes for the circadian clock. Transcription of *to* in brain fat bodies is induced in response to the male-specific forms of the transcripts of genes *doublesex* and *fru* leading to a significant enrichment of *to* product in male *Drosophila* brain. *To* plays a role in determining male courtship behaviour and circadian regulation of *to* appears to be primarily involved with the timing of feeding behaviour and a sexually dimorphic, locomotor phenotype, is also determined by this gene (So et al. 2000; Dauwalder et al. 2002; Meunier et al. 2007).

Sexual dimorphism of a temporal phenotype is also seen in the crepuscular habit of *Drosophila*. Flies show two daily peaks in activity (dawn and dusk) and the timing of these peaks is slightly different for males as compared with females, such that anticipatory behaviour occurs ~1h earlier in males (Helfrich-Förster 2000). Neuropeptide F is shown to modulate the timing of the evening peak in males by clock-controlled cell type specific expression in dorsolateral neurons (LN<sub>d</sub>) (Lee et al. 2006).

Such convergence of the circadian clock and sex determining pathways appear to represent a mechanism underlying the expression of temporal phenotypes that are sexually dimorphic. Epigenetic processes (that determine cell type specific patterns of gene expression) the circadian molecular clock and the sex determination pathways are thus implicated in aspects of developmental timing and in particular, sexually dimorphic temporal behaviours. Thus the circadian and behavioural timing anomalies, and, the strong sex bias towards males in autism, may reflect a disturbance in the neurotypical interplay between the circadian clock, and sex-and-cell-type specific splicing.

## **1.9 Aims and structure of the thesis.**

The first aim of this work was to test the hypothesis, proposed by this author (Wimpory et al. 2002), that genes encoding the central oscillator of the circadian molecular clock are implicated in autistic disorder. If the association test proved positive, the possible effects of the associated variants would be examined and a hypothesis developed to explain how such clock gene variants could affect autism relevant neurology and/or development. The work aimed to be integrative and draw together genetic,

epigenetic, physiological and psychological aspects of timing and circadian malfunction in autism.

The structure of the thesis is outlined as follows. Chapter 2 tests genetic association of clock genes with autism by using the transmission disequilibrium test (TDT) on a cohort of DNA samples from autistic children and their parents, this test gave positive results.

In chapter 3 the autism-associated SNPs were cross-checked against up-dated versions of the human genome map. Particular attention was given to checking whether any clock gene autism-associated SNPs co-locate with known regulatory features and that might be disrupted by such SNPs. Also, a database representing genes known to contain conserved circadian regulatory elements (E-box, D-Box and RRE) was scanned for overlap with a set of autism candidate genes. The set of putative autism genes (that now appears to contain a number of first order clock controlled genes) likely represents clock-controlled elements of a molecular pathway that is dysfunctional in autism.

Chapter 4 investigates possible novel functionality of the autism-associated SNPs by using bioinformatics tools that employ algorithms to calculate bespoke solutions for the effect of SNPs on alternative splicing and RNA secondary structures. This chapter includes the web publication “Autism-associated SNPs in the clock genes NPAS2, PER1 and the homeobox gene EN2 alter DNA sequences that show characteristics of microRNA genes”. This author carried out all the bioinformatics work for this publication and wrote the manuscript.

A corollary of the idea that clock genes may harbor microRNAs is considered in the last section of chapter 4. The results of three published

studies that found dysregulation of known microRNAs in autism were re-investigated to determine whether these “autism microRNAs” have clock genes amongst their predicted targets.

Chapter 5 discusses aspects of how variants in clock genes might alter neural oscillators. A hypothesis is presented that might account for the effect of clock gene mutations on behaviours regulated by high frequency biological oscillators and the observed comorbidity of neuropsychiatric disorder and circadian rhythm disturbance: The thesis closes with a summary of the main findings, conclusions and indications for future research.

Appended to the thesis are descriptions of two laboratory-based experiments to investigate possible DNA methylation effects in *PER1* and a possible role for HAT1 in NPAS2 protein complex formation. Constraints on time, the unavailability of clinical samples and a promising line of bioinformatics enquiry developing out of RNA transcript analysis determined that the informatics and bioinformatic analysis of the results of the association test, would become the main focus of the rest of the thesis. Two pieces of bioinformatic work; on the structure of *PER1* and *Drosophila* *PERIOD*, and transcript analysis of a schizophrenia associated gene (*ZNF804A*), are also appended.

## **2 Publication: Association of *PER1* and *NPAS2* with Autistic disorder: support for the clock genes/social timing hypothesis**

### **2.1 General introduction and contributions to the publication.**

The following association study (Nicholas et al. 2007) (Published in *Molecular Psychiatry* 2007 12, 581–592; doi:10.1038/sj.mp.4001953) tests the hypothesis that there is association between strictly diagnosed autistic disorder and clock gene variants. The genes *PER1*, *PER2*, *PER3*, *CLOCK*, *NPAS2*, *ARNTL*, *TIM*, *CRY1*, *CRY2*, *DBP* and *CSNK1*, were each tested in the association study.

#### 2.1.1 Experimental contributions.

This author was first to propose that clock genes are implicated in autism and that aberrant DNA methylation of clock genes and/or clock gene variations contribute to the disorder. This author contributed the argument for investigating clock genes, in terms of multi-level timing dysfunction in autism. In the funding proposal for this study, this author used cytogenetic findings and evidence from autism sleep studies to show that clock gene anomalies could not be ruled out. This author designed the experimental outline for the association study, led the selection of the candidate genes and contributed to the informatics work surrounding sample selection. This was the first genetic study of clock genes in autism and since publication three further investigations have given results in support of a role for clock genes in autism (Hu et al. 2009a; Nguyen et al. 2010; Fradin et al. 2010).

The psychological element of the hypothesis, that timing deficit may be innate in children with autism and that this timing deficit has a developmental effect on the emergent social-symbolic functioning of autistic infants, was contributed by D. Wimpory. D. Wimpory also selected the autism DNA samples for the transmission disequilibrium test on the basis of the clinical diagnosis of the donors, a strictly autistic disorder phenotype.

G. Kirov made the detailed plan of the experimental procedure and V. Rudrashingham performed the practical work. G. Kirov oversaw the laboratory practice and the specifics of the sequencing, genotyping and genetics. M.J. Owen oversaw the interpretation of the genetic analysis.

#### 2.1.2 Authorship of the paper.

This author wrote the “hypothesis” and “The candidate genes and their relevance to the hypothesis” sections of the paper; G. Kirov wrote the methods and results section; D. Wimpory and this author wrote the introduction and the details of sample selection. This author, D. Wimpory and M. J. Owen wrote the discussion section.

## **2.2 Abstract.**

Clock gene anomalies have been suggested as causative factors in autism. We screened eleven clock/clock-related genes in a predominantly high-functioning Autism Genetic Resource Exchange sample of strictly diagnosed autistic disorder progeny and their parents (110 trios) for association of clock gene variants with autistic disorder. We found significant association ( $P < 0.05$ ) for two single-nucleotide polymorphisms

in *PER1* and two in *NPAS2*.

Analysis of all possible combinations of two-marker haplotypes for each gene showed that in *NPAS2*, 40 out of the 136 possible two-marker combinations were significant at the  $P < 0.05$  level, with the best result between markers rs1811399 and rs2117714,  $P = 0.001$ . Haplotype analysis within *PER1* gave a single significant result: a global  $P = 0.027$  for the markers rs2253820–rs885747. No two-marker haplotype was significant in any of the other genes, despite the large number of tests performed. Our findings support the hypothesis that these epistatic clock genes may be involved in the etiology of autistic disorder. Problems in sleep, memory and timing are all characteristics of autistic disorder and aspects of sleep, memory and timing are each clock-gene-regulated in other species. We identify how our findings may be relevant to theories of autism that focus on the amygdala, cerebellum, memory and temporal deficits. We outline possible implications of these findings for developmental models of autism involving temporal synchrony/social timing.

### **2.3 Introduction.**

This study aims to evaluate the hypothesis that clock genes are implicated in autistic disorder. This severe neuro-developmental disorder is characterized by three areas of abnormality: impairment in communication (verbal and non-verbal) and reciprocal social interaction together with a markedly restricted repertoire of activities and interests, all in evidence before 3 years of age (APA. 1994). The primary focus of this paper is autistic disorder, as opposed to the more heterogeneous autistic

spectrum disorders (ASDs). The prevalence of autistic disorder is 0.1–0.2%; (Chakrabarti and Fombonne 2001) autistic disorder has a 60% concordance rate in monozygotic twins but no concordance is shown for autistic disorder in dizygotic twins (Bailey et al. 1995). Such data suggest that there is a high degree of, but not complete, genetic control over the occurrence of autistic disorder; adverse inter-uterine or potentially protective effects may be involved in the manifestation of this disorder (Santangelo and Tsatsanis 2005). The heritability of autistic disorder is best explained by a model involving a number of genes at unlinked and some epistatic loci that together contribute to the phenotype (Pickles et al. 1995). There is considerable (at least 75%) comorbidity of mental retardation in autism (Fombonne 1999). Our study addressed sample heterogeneity by selecting for more intellectually able subjects that met strict diagnostic criteria for autistic disorder, thereby increasing experimental power.

### 2.3.1 Whole-genome screens and candidate gene studies for autism

There are significant whole-genome linkage findings for autism on 2q, (Palferman et al. 2001) 3q (Auranen et al. 2002) and 17q (Alarcon et al. 2005). 2q is highlighted as most likely harboring a locus for autistic disorder (Philippe et al. 1999; Buxbaum et al. 2001; Palferman et al. 2001; Shao et al. 2002). However, the overall genome screen results are inconsistent, with a possible explanation being that most/all of these studies are underpowered to find genes of small effect set in the context of sample heterogeneity and diagnostic differences. Candidate gene studies in autism have tested numerous genes, often on the basis of their



proximity to LOD score peaks from whole-genome linkage studies. Such studies remain largely un-replicated, with the possible exceptions of *GABRB3* (Cook et al. 1998; Buxbaum et al. 2002) *ENGRAILED2* (Gharani et al. 2004; Benayed et al. 2005) and the serotonin transporter 5-HTT (Cook et al. 1997; Tordjman et al. 2001; Yirmiya et al. 2001; Kim et al. 2002).

The scope of candidate gene studies in autism is generally limited by the lack of credible supporting hypotheses. Our candidate gene study was prompted by the hypothesis (Wimpory et al. 2002) that autism reflects a disturbance of clock gene function, from a molecular level to the manifestation of autistic disorder as a psychological phenomenon.

### 2.3.2 Timing difficulties and autism

timing difficulties have been proposed as being central to autism e.g. (Newson 1984), with hypotheses encompassing circadian, communicative and/or neurological aspects of timing (Segawa et al. 1981; Courchesne et al. 1994; Richdale and Prior 1995; Boucher 2000; Boucher 2001; Brock et al. 2002; Wimpory et al. 2002; Welsh et al. 2005). For example, Boucher (2001) suggests a core timing deficit presenting different manifestations by its effect on the elements of an integrated system of neural and physiological oscillators. Wimpory et al., (2002) hypothesize a causative, concurrent and developmental role (Wimpory and Nash 1999) for timing deficit in autistic disorder and that this deficit is derived from pathological variations in the structure/function of clock/clock-related genes (Wimpory et al. 2002).

Drawing on the above hypotheses, we conceive that timing deficit may

manifest in autism as both temporally measured anomalies and apparently disparate symptoms (such as anomalous performance in tasks involving perceptual and cognitive coherence (see review, (Happé and Frith 2006), relational memory (episodic (Boucher and Lewis 1989; Bowler et al. 2000; Millward et al. 2000; Salmond et al. 2005) and diachronic thinking (Boucher et al. 2007)); each with some temporal and/or clock gene dependency. Temporally measured anomalies in autism (including high-functioning autism and Asperger syndrome) range from circadian/sleep architecture (outlined below) to brain oscillations involved in neural binding; (Grice et al. 2001; Brown et al. 2005) information processing; (Szelag et al. 2004) attention (Townsend et al. 2001; Haist et al. 2005; Zwaigenbaum et al. 2005) including rapid attention-switching (Courchesne et al. 1994a; Townsend et al. 1999); and, motor coordination (from posture to eye-blink) (Sears et al. 1994; Inui and Asama 2003; Schmitz et al. 2003; Gowen and Miall 2005). They also extend to the reciprocity/temporal synchrony skills required for conversation (Feldstein et al. 1982; Tantam et al. 1993; Trevarthen and Daniel 2005; Bebko et al. 2006).

Objectively recorded atypical sleep architecture in young adults with high-functioning autism and Asperger syndrome (Limoges et al. 2005) and in children with autistic disorder (Elia et al. 2000) shows association between sleep profiles and autistic symptomatology (Elia et al. 2000; Limoges et al. 2005). Circadian hormone (melatonin) anomalies are found in autistic disorder (Nir et al. 1995; Kulman et al. 2000; Tordjman et al. 2005) as well as altered serotonin levels in autism (Cook and Leventhal 1996; Anderson 2002). An altered circadian clock mechanism affecting normal sleep wake cycles could have an effect on the levels of these

hormones that is additional to any other specific transport and reuptake effects.

### 2.3.3 The hypothesis.

Our study tests the hypothesis that there is association between strictly diagnosed autistic disorder and alterations in clock genes (specifically *PER1*, *PER2*, *PER3*, *CLOCK*, *NPAS2*, *ARNTL* (*BMAL1*), *TIM*, *CRY1*, *CRY2*, *DBP* and *CSNK1E*. Table 2-1).

**Table 2-1. Candidate genes and their chromosomal locations.**

<i>TIM</i>	12 q13.3
<i>ARNTL</i>	11 p15.2
<i>CLOCK</i>	4 q12
<i>CSNK1E</i>	22 q31.1
<i>CRY1</i>	12 q23.3
<i>CRY2</i>	11 p11.2
<i>DBP</i>	19 q13.33
<i>NPAS2</i>	2 q11.2
<i>PER1</i>	17 p13.1
<i>PER2</i>	2 q37.3
<i>PER3</i>	1 p36.23

There is a high degree of functional and sequence similarity of canonical clock genes across widely different organisms (Allada et al. 1998; Darlington et al. 1998; Gekakis et al. 1998; Rutila et al. 1998; Hogenesch et al. 2000) and, in addition to affecting the circadian cycle, clock gene anomalies are specifically linked to sleep disorders (in humans (Ebisawa et al. 2001; Toh et al. 2001), altered sleep architecture and contextual memory (in mice (Garcia et al. 2000; Franken et al. 2006) and communicative timing and memory formation (in *Drosophila* (Kyriacou and Hall 1980; Alt et al. 1998; Ritchie et al. 1999; Sakai et al. 2004) all are

relevant to aspects of autism: sleep, (Elia et al. 2000; Limoges et al. 2005) memory (Boucher and Lewis 1989; Bowler et al. 2000; Millward et al. 2000; Toichi and Kamio 2002; Toichi and Kamio 2003; Salmond et al. 2005) and communicative timing (Feldstein et al. 1982; Trevarthen and Daniel 2005).

#### 2.3.4 Molecular genetic background: clock genes.

In mammals, the core molecular clock model (see for example, Looby and Loudon (2005)) comprises a suite of epistatic genes that operate as an integrated system of transcription/translation autoregulatory feedback loops. CLOCK and ARNTL, the protein products of the genes *CLOCK* and *ARNTL*, positive elements of the system, heterodimerize and activate the transcription of the genes *PER1*, *PER2*, *PER3*, *CRY1*, *CRY2*, *REV-ERB $\alpha$*  and other clock-controlled genes. The gene products of *PER*, *CRY* and *REV-ERB $\alpha$* , the negative elements of the system, operate via a PER/CRY heterodimer that inhibits the activating effects of the CLOCK/BMAL1 heterodimer while REV-ERB $\alpha$  represses expression of *ARNTL*. Protein turnover eventually releases the genes from repression and the cycle starts over. The time taken for the molecular migrations of proteins and mRNA, respectively, to and from the nucleus, together with the protein turnover time, broadly defines the ~24h period of this oscillatory system. To integrate this core clock mechanism into the life of the organism as a whole, genes associated with resetting the clock (*CRY1*, *CRY2*) serve in matching environmental time with biological time by resetting the clock in response to environmental cues such as day length (Berson et al. 2002). *CSNK1E* is involved with regulating the clock by affecting the stability of

the PER/CRY complex through phosphorylation (Lee et al. 2001) while output pathways transduce the clock's time signals into physiological response. It is this molecular clock that endows living organisms with the ability to maintain a state of appropriate physiological readiness that anticipates the environmental demands associated with a particular time of day or night. In addition to their role as clock elements, certain clock genes (e.g., *PER1*, *PER2* and *NPAS2*) appear to play roles in signaling pathways and in DNA repair (Chilov et al. 2001; Rutter et al. 2001; Dioum et al. 2002; Fu et al. 2002; Gery et al. 2006). Genetic studies also point toward a broader functionality for clock genes beyond their role in circadian rhythms. For example, *per* in *Drosophila melanogaster* is shown to regulate a short-period oscillator involved with the fly's courtship song, (Kyriacou and Hall 1980) a primitive form of sonic communication. The male fly, as part of the mating ritual, produces a song by rhythmic beating of its wings. This song has several acoustic components, including pulse song where a series of rhythmic pulses are separated by inter-pulse intervals (ipi) of ~34-ms. Kyriacou and Hall (Kyriacou and Hall 1980) studied song structure in *Drosophila* circadian *per* mutants (Konopka and Benzer 1971) and showed a cyclic modulation of ipi-duration in the pulse song of *D. melanogaster*, the Kyriacou and Hall (K&H) cycle, of ~55 s is also under the control of *per* (Kyriacou and Hall 1980; Alt et al. 1998). The three circadian *per* mutants studied *per<sup>s</sup>* (19 h: short period) *per<sup>l</sup>* (28-h: long period) and *per<sup>0</sup>* (arrhythmic) showed K&H cycles that were also short, long and abolished respectively. Thus, *per*'s influence on the circadian cycle (~24h) and on a short-period (~55s) oscillatory function associated with the song appear to be qualitatively similar (Kyriacou and

Hall 1980). These observations show *per* operating not only in a circadian oscillator but also in modulating high-frequency oscillators controlling motor function associated with a communicative behavior (Konopka and Benzer 1971; Konopka et al. 1996; Ritchie et al. 1999). Kyriacou and Hall (Kyriacou and Hall 1980) also noted that where male flies interrupted their song, they restarted in phase with the initial portion of the song and that each fly's song may start at a different phase of the K&H cycle. This suggests that the *per*-determined K&H cycle is integral with a concurrent, short-interval timing process.

The clock genes *PER1* and *NPAS2* also play key roles in memory formation. Experiments with *Npas2* knockout mice show that the knockout mice performed statistically similarly to wild-type mice in a battery of behavioral tests apart from the test for cued and contextual fear. The results of this experiment suggest that *Npas2-LacZ(-/-)* mice are deficient in complex emotional memory specifically, but not in non-emotional memory (Garcia et al. 2000). In *Drosophila*, *per* plays a role in long-term memory formation, which is independent of the core circadian oscillator (Sakai et al. 2004).

### 2.3.5 The candidate genes and their relevance to the hypothesis

Because of the epistatic nature of the core clock genes, we decided to screen all the canonical clock elements and also *NPAS2*, *DBP*, *TIM* and *CSNK1E*. Apart from including the core elements (*PER1*, *PER2*, *PER3*, *ARNTL* (*BMAL1*), *CLOCK*, *CRY1* and *CRY2* (see Table 2-1) for their role in the central circadian oscillator, we also selected candidate genes on the basis of their noncircadian functions and/or association with syndromes

implying possible links with autism.

Thus, we included: *PER1*, for its association with long-term memory formation (Sakai et al. 2004) and high-frequency oscillators involved with communicative timing in *Drosophila* (Kyriacou and Hall 1980; Alt et al. 1998; Ritchie et al. 1999); *PER2*, for its implication in familial advanced sleep phase syndrome (Toh et al. 2001); and, both *CSNK1E* and *PER3*, for their effect on sleep disorders (Takano et al. 2004) and association with delayed sleep phase syndrome (Ebisawa et al. 2001; Toh et al. 2001). We included *NPAS2* because it is a paralogue of clock, expressed in the brain (Reick et al. 2001) and associated with complex (cued and contextual) memory in mice (Garcia et al. 2000). Subjects with autism also show impairment in complex memory (Williams et al. 2006); contextualized episodic memory is specifically impaired even in high-functioning autism and Asperger syndrome (Boucher and Lewis 1989; Bowler et al. 2000; Millward et al. 2000; Toichi and Kamio 2003; Salmond et al. 2005). *NPAS2* is also a transcriptional regulator of non-rapid eye movement sleep in mice (Franken et al. 2006); this is relevant, given the altered sleep architecture in high-functioning autism, autistic disorder and Asperger syndrome (Elia et al. 2000; Limoges et al. 2005). Finally, we included *DBP* for its role in the regulation of clock outputs such as sleep and locomotor activity in mice (Franken et al. 2000; Ripperger and Schibler 2006) and *TIMELESS (tim)* for its role in *Drosophila* circadian systems (Barnes et al. 2003).

## **2.4 Materials and methods.**

### 2.4.1 Subjects.

In the interests of sample homogeneity, we selected a predominantly high functioning sample where all subjects were strictly diagnosed with autistic disorder (detailed below). We obtained DNA first from 90 probands and all their parents (parent–offspring trios), of whom there were 65 male and 25 female subjects, from the Autism Genetic Resource Exchange (AGRE) (<http://www.agre.org/>). This is a publicly available database and central repository founded by the Cure Autism Now Foundation (Geschwind et al. 2001). In the second stage of the study, we obtained an additional 20 AGRE probands (14 males and six females) to form, together with first wave subjects, a larger sample in which we might attempt to replicate any positive results that reached a significance level of  $P < 0.05$  in the first wave.

The subjects were 91% Caucasian, 4% other and 5% unknown (see Table 2). All probands analyzed in this study met strict diagnostic criteria for autistic disorder using both the Autism Diagnostic Interview-Revised (ADI-R (Lord et al. 1994)) and the Autism Diagnostic Observation Schedule-Generic (ADOS-G (Lord et al. 2000)) compatible with the DSM-IV definition of autistic disorder that includes high-functioning individuals. (1) We excluded those with borderline or ‘not quite autism’ (NQA) sometimes accepted as a research diagnosis.



**Table 2-2. Diagnostic criteria and demographics of the sample.**

<b>Diagnostic measures</b>	<b>Total</b>	<b>1st stage</b>	<b>2nd stage</b>
Autistic disorder (ADI-R) <sup>a</sup>	100% (n = 110)	100% (n = 90)	100% (n = 20)
Autistic disorder (ADOS-G) <sup>b</sup>	100% (n = 110)	100% (n = 90)	100%(n = 20)
Male	72% (n = 79)	72% (n = 65)	70% (n = 14)
Female	28% (n = 31)	28% (n = 25)	30% (n = 6)
Caucasian	91%	91%	90%
Other race <sup>c</sup>	4%	3%	10%
Unknown race <sup>d</sup>	5%	6%	0%

**Abbreviations:**

ADI-R, Autism Diagnostic Interview Revised;

ADOS-G, Autism Diagnostic Observation Schedule-Generic.

<sup>a</sup> Met strict ADI-R criteria for autistic disorder.

<sup>b</sup> Met strict ADOS-G criteria for autistic disorder (only subjects with consistent diagnoses were accepted; we rejected those for whom ADI-R or ADOS-G re-assessment had changed their diagnostic status to autistic disorder from another diagnosis that failed to meet our criteria (e.g., ASD; 'not quite autism' (NQA) / borderline autistic disorder etc.).

<sup>c</sup> Mixed race and Pacific Islander.

<sup>d</sup> Mixed race.

<sup>e</sup> 5% mixed race and 5% Pacific Islander (n=1+1).

**Table 2-3. Distribution of percentile ranking (PR) from ability/behavioral measures and corresponding intelligence levels.**

	<b>Sample division</b>	<b>n</b>	<b>%Total sample</b>	<b>Men PR (IQ)<sup>a</sup></b>	<b>s.d</b>	<b>Range PR (IQ)<sup>a</sup></b>
<b>R</b>	1st wave	72	66	42 (IQ 97.3)	17.4	0.3–99 (IQ 58–136)
	2nd wave	19	17	21 (IQ 88.3)	17.5	1–90 (IQ 63–119)
	1st and 2nd wave	91	83	37 (IQ 95.4)	17.7	0.3–99 (IQ 58-136)
<b>P</b>	1st wave	6	6	5.5 (IQ 77)	9.7	0.1–25(IQ 51–90)
	2nd wave	1	1	0.1 (IQ 51)	N/A	0.1 (IQ 51)
	1st and 2nd wave	7	6	4.8 (IQ 75)	9.1	0.1–25 (IQ 51–90)
<b>V</b>	1st wave only	12	11	5.7	14.1	0.1–50

**Abbreviations:****R**, Assessment measure: Raven's Progressive Matrices**P**, Assessment measure: Peabody Picture Vocabulary Scales**V**, Assessment measure: Vineland Adaptive Behavior Scales

N/A, not applicable.

**Abbreviations (continued):**

<sup>a</sup>RPM non-verbal IQ estimations, together with PPVT-III, and VABS-composite percentile rankings, were obtained from AGRE. For comparison purposes in this table, Wechsler Percentile-IQ Correspondence Scale was employed to give percentiles and IQ approximations (for RPM and PPVT-III, respectively).

We also used assessment measures available from AGRE, prioritizing inclusion of subjects on the basis of intelligence (83%) over language (6%), and language over behavioral (11%) measures. The assessment measures for these three fields were Raven Progressive Matrices (RPM (Raven 1996)), Peabody Picture Vocabulary Test (PPVT-III (Dunn 1997)) and the Vineland Adaptive Behavior Scales (VABS (Sparrow 1984)) respectively. Our criterion for accepting subjects tested on the RPM was a non-verbal intelligence quotient (IQ) of at least 51. In practice, they all had IQs of 58 or more (mean 95.4, s.d. 17.7, range 58-136, n = 91). We accepted additional subjects (n = 7) with IQ approximations, equivalent to our RPM selection criteria, derived from PPVT-III percentiles using Wechsler Percentile-IQ Correspondence Scale (Wechsler 1997). Remaining subjects (n = 12) had VABS percentile rankings (0.1 and above) equivalent to those of our PPVT-III selection criterion (see Table 3). On the Wechsler Percentile-IQ Correspondence Scale a percentile rank of 2 corresponds to an IQ range of 68–70 inclusive and we therefore used an IQ cut-off of 68 (or percentile ranking of 2) to define 'high functioning' autism (cf. Mottron (Mottron 2004)). Our total sample was 85%

high functioning, with remaining subjects ensuring appropriate sample size. We thereby derived the most able autistic disorder sample available from AGRE at the time of subject selection.

Possible non-idiopathic autistic disorder cases were excluded, as were probands who had any known additional major medical or neurological condition that might have been perceived to cause secondary autism. Similarly, cases with a previous or concurrent alternative diagnosis (e.g. ASD on ADI-R and/or schizophrenia) were excluded, together with those with identified chromosomal aberrations.

#### 2.4.2 Screening the candidate genes.

Candidate clock genes were systematically screened for indirect association with autism by genotyping single-nucleotide polymorphisms (SNPs) covering the genes. We aimed at an average spacing of one common SNP at every 3–5kb, but this was not always possible, or practical. For example, some of the genes were too long to allow such a dense coverage within our budget (for example, for *NPAS2* we tried to cover some 160 kb), and in some cases there were long gaps with no validated common SNPs at the time we performed the work. In addition, for some SNPs our genotyping assay could not be designed or failed. These reasons prevented us from providing a uniform coverage of each gene. SNPs were chosen from the publicly available databases HAPMAP ([www.hapmap.org/](http://www.hapmap.org/)), NCBI ([www.ncbi.nlm.nih.gov/](http://www.ncbi.nlm.nih.gov/)), CHIP Bioinformatics tools ([snpper.chip.org/](http://snpper.chip.org/)) (Riva and Kohane 2002). Collaboration with the team of Professor Vishwajit Nimgaonkar, who screened these genes in a study of bipolar affective disorder (Mansour et al. 2005) provided details

on additional SNPs that they had identified. During the course of our study a number of these SNPs became available in the public databases. Genes where only a few or no SNPs were available were screened for mutations in the promoter, all exons and 3' UTR using denaturing high performance liquid chromatography (dHPLC) analysis on a WAVE dHPLC system (Transgenomics Inc., Cambridge, MA, USA). In addition, we screened the genes that showed evidence for association after individual genotyping of the first stage of our sample. For mutation screening, we used 16 autistic probands from our sample. Temperature and solution gradients for individual fragments for dHPLC were calculated using the Stanford Genome Technology Center DHPLC Melt Program freely available on the web (<http://insertion.stanford.edu/melt.html>). In addition to screening the promoter regions of these genes, we also screened predicted regulatory regions of the gene using the Cister program (<http://zlab.bu.edu/~mfrith/cister.shtml>). This freely available automated resource explores clusters of cis-elements in DNA sequences and such regions were also included for screening. On average, we covered genes for screening with B20–25 DNA fragments of 400–500 bp length each. The genes that we screened for mutations were *CLOCK*, *PER1* and *NPAS2*. Fragments showing possible mutations via dHPLC were sequenced using the ABI PRISM BigDye Terminator Cycle Sequencing Ready Reaction Kit, version 3.1 on an ABI PRISM 3100 Genetic Analyzer (Applied Biosystems, CA, USA). Individual genotyping was carried out using the Amplifluor SNPs Genotyping Systems (Serologicals Corporation, Norcross, GA, USA) (Myakishev et al. 2001), which is a one-step PCR-based reaction using allele-specific primers. All forward and reverse

primers for the Amplifluor reaction were designed using the automated primer design software, Amplifluor AssayArchitect, freely available through the Serologicals Corporation website ([www.assayarchitect.com](http://www.assayarchitect.com)). Amplifluor reactions were performed in a 5ml volume in 96-well black propylene plates (ABgene, Epsom, UK) with 15 ng of dried genomic DNA. Data were analyzed using an Analyst AD fluorescence multiwell plate reader (LJL Biosystems, CA, USA). Every assay designed was first tested on one 96-well plate of DNA samples and only SNPs that gave satisfactory clusters and no non-mendelizations were genotyped in the full sample of trios. On average, some 30% of all SNPs we attempted to genotype failed and were therefore discarded; we present results only on the SNPs that produced good traces.

### 2.4.3 Statistics.

For statistical analysis of genotyping results we used the transmission disequilibrium test (TDT103) that examines the transmission of alleles from heterozygous parents and is thus a robust test against population stratification. Haplotype analysis was performed with the program TDTPHASE (Dudbridge 2003).

## **2.5 Results.**

All results from individual genotyping are presented in Table2-4. The table also shows the genomic positions of SNPs according to the Golden Path database (<http://www.genome.ucsc.edu/>) (Build 35, May 2004). Initially, for the sample of 90 trios, we obtained significant evidence for association for two of the genes studied. In *PER1* two intronic SNPs gave significant results: rs885747 with a  $P = 0.014$  and rs6416892 with  $P = 0.035$ . Re-

sequencing the gene did not identify any novel polymorphisms in the 16 individuals we genotyped, although we found a number of SNPs already available in public databases. All six SNPs genotyped in the first wave were genotyped in the additional 20 trios. The two significant SNPs remained significant:  $P = 0.047$  and  $0.042$ , respectively. In *NPAS2* we found one significantly associated SNP in the first wave of genotyping: rs1811399 at  $P = 0.009$  level. *NPAS2* was re-sequenced in an attempt to identify markers that were more strongly associated, or have a putative functional significance. We identified several new SNPs and genotyped them individually: a C/T polymorphism in intron 3 (called NPAS2\_X3\_C\_T in Table 4), a G/A polymorphism in intron 7 (NPAS2\_IN7\_G\_A), a G/A polymorphism in exon 8 (NPAS2\_X8\_G\_A), an A/G polymorphism in intron 11 (NPAS2\_X11\_A\_G), an A/T polymorphism in intron 12 (NPAS2\_X12\_A\_T) and a C/T polymorphism in exon 15 of the gene (NPAS2\_X15\_C\_T). These, as well as all the markers typed in the first wave, were genotyped in the complete set of 110 trios (a total of 72 SNPs).

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)
rs3849381 (1)		0.90	0.89	0.18	0.72	0.31	0.72	0.21	0.36	0.18	0.62	0.14	0.63	0.07	0.03	0.16	0.37
rs1811399 (2)	0.08		0.64	0.30	0.79	0.22	0.48	0.41	0.01	0.01	0.14	0.13	0.13	0.08	0.05	0.02	0.04
rs2117714 (3)	0.11	0.04		0.86	1.00	0.75	0.17	0.05	0.19	0.34	0.25	0.11	0.22	0.26	0.06	0.24	0.20
rs356651 (4)	0.02	0.05	0.05		1.00	0.78	0.23	0.45	0.50	0.37	0.40	0.63	0.42	0.22	0.60	0.37	0.50
NPAS2_X3_C_T (5)	0.04	0.04	0.09	0.03		1.00	0.27	0.11	0.47	0.02	0.26	0.44	0.26	0.08	0.50	0.01	0.00
rs895521 (6)	0.05	0.03	0.04	0.52	0.04		0.74	0.29	0.17	0.30	0.23	0.03	0.26	0.21	0.11	0.30	0.38
rs13017718 (7)	0.01	0.00	0.01	0.00	0.03	0.01		0.88	1.00	1.00	1.00	1.00	1.00	0.76	1.00	0.80	0.56
rs895520(8)	0.02	0.03	0.00	0.05	0.00	0.02	0.09		1.00	1.00	1.00	0.32	1.00	0.98	0.64	0.97	0.03
NPAS2_INT7_G_A (9)	0.00	0.00	0.01	0.00	0.00	0.00	0.01	0.07		1.00	0.96	1.00	0.96	0.88	0.95	1.00	0.10
NPAS2_X8_G_A (10)	0.02	0.00	0.01	0.12	0.00	0.09	0.01	0.29	0.02		1.00	1.00	1.00	1.00	1.00	1.00	0.00
NPAS2_X11_A_G (11)	0.02	0.01	0.02	0.00	0.00	0.00	0.01	0.12	0.53	0.03		0.53	1.00	1.00	0.97	1.00	0.03
rs1562313 (12)	0.01	0.00	0.00	0.01	0.01	0.00	0.02	0.01	0.02	0.04	0.01		0.54	1.00	1.00	1.00	0.20
NPAS2_X12_A_T (13)	0.02	0.01	0.02	0.00	0.00	0.00	0.01	0.13	0.50	0.03	0.96	0.01		1.00	0.97	1.00	0.02
rs2305160 (14)	0.00	0.00	0.01	0.00	0.00	0.00	0.02	0.29	0.03	0.09	0.06	0.09	0.07		1.00	1.00	0.02
rs2305159 (15)	0.00	0.00	0.00	0.02	0.02	0.00	0.03	0.13	0.20	0.08	0.36	0.48	0.38	0.20		1.00	0.12
NPAS2_X15_C_T (16)	0.02	0.00	0.00	0.11	0.00	0.08	0.01	0.29	0.02	0.96	0.03	0.04	0.03	0.09	0.09		0.01
rs2305158 (17)	0.02	0.00	0.03	0.01	0.00	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	

**Figure 2-1. Linkage disequilibrium within *NPAS2*.**

Above the diagonal:  $D'$ , below the diagonal:  $r^2$ . Values above 0.5 are highlighted.

As shown in Table 2-4, only one of the new markers reached statistical significance in the full set of trios: NPAS2\_X3\_C\_T with a P-value of 0.028, whereas the originally significant marker, rs1811399, remained significant at  $P = 0.018$ .

Haplotype analysis: We performed analysis of all possible combinations of two-marker haplotypes for each gene. In *NPAS2*, 40 out of the 136 possible two-marker haplotype combinations were significant at the  $P < 0.05$  level, with the best result between markers rs1811399 and rs2117714,  $P = 0.001$ . If we perform a Bonferroni correction for 136 tests, this result would lose its significance; however, this correction is over

conservative due to the high number of SNPs that are in linkage disequilibrium in *NPAS2* (Figure 2-1). Within *PER1* there was a single significant result: a global  $P = 0.027$  for the haplotype analysis of markers rs2253820–rs885747. No two-marker haplotype was significant in any of the other genes, despite the high number of tests performed.

**Table 2-4. Results of the transmission disequilibrium tests.**

Gene/marker	Coordinate (bp)	Distance from previous marker (bp)	Number of trios typed	Allele	Frequency in parents	Frequency in children	Transmitted	Non-transmitted	P-value
ARNTL_rs2279287	13255061		87	A>G	0.305	0.328	35	27	0.31
ARNTL_rs2279285	13255263	202	32	T>G	0.703	0.688	8	10	0.63
ARNTL_rs1982350	13306707	51444	88	A>G	0.401	0.403	38	37	0.91
ARNTL_rs3789327	13341892	35185	85	A>G	0.556	0.559	37	36	0.91
ARNTL_rs3816358	13348048	6156	86	A>C	0.111	0.116	15	13	0.71
ARNTL_rs2278749	13354454	6406	86	T>C	0.201	0.174	20	29	0.2
ARNTL_rs2290035	13364347	9893	87	A>T	0.489	0.454	33	45	0.17
CLOCK_SNP4_G_A	56134430		88	G>A	0.949	0.949	9	9	1
CLOCK_rs1801260	56142297	7867	88	T>C	0.770	0.778	34	31	0.71
CLOCK_rs6811520	56156106	13809	85	A>G	0.309	0.282	27	36	0.26
CLOCK_rs2272073	56217272	61166	85	G>A	0.609	0.565	33	48	0.1
Cry1_3741890	105889492		31	G>T	0.436	0.436	14	14	1
Cry1_rs1921126	105914072	24580	84	A>G	0.545	0.560	45	40	0.59
Cry1_rs1861591	105919394	5322	83	G>A	0.542	0.566	45	37	0.38
Cry1_rs1420399	105919649	255	62	C>T	0.419	0.403	24	32	0.28
Cry1_rs2078074	105939273	19624	85	A>G	0.562	0.582	48	41	0.46
CRY2_rs7121775	45820899		31	C>T	0.2823	0.2742	12	13	0.84
CRY2_3_A_G	45823734	2835	87	A>G	0.287	0.293	35	33	0.81
CRY2_rs2292912	45834264	10530	87	C>G	0.2011	0.1897	26	30	0.59
CRY2_rs1401417	45836686	2422	31	C>G	0.7016	0.6935	11	12	0.84
CRY2_rs12364060	45846351	9665	29	C>T	0.724	0.724	11	11	1
CRY2_rs7933420	45853373	7022	87	T>A	0.5374	0.5517	45	40	0.59
CRY2_rs877412	45854227	854	84	A>C	0.5863	0.5893	31	30	0.9
CRY2_rs3824872	45862181	7954	32	C>A	0.7891	0.7656	9	12	0.5
CKI_rs2075983	37015198		90	T>C	0.2611	0.2889	39	29	0.22
CKI_rs2075984	37015389	191	106	G>T	0.5212	0.5283	49	42	0.46
CKI_rs1005473	37018873	3484	108	T>G	0.9444	0.9352	10	14	0.41
CKI_rs1534891	37019599	726	108	G>A	0.8727	0.8519	19	28	0.19
DBP_rs3745734	53826200		90	C>G	0.9333	0.9222	12	12	1
DBP_rs550455	53828327	2127	87	A>G	0.7098	0.7184	35	34	0.9
DBP_rs386551	53830098	1771	88	G>C	0.7045	0.7159	35	33	0.81
DBP_rs3848543	53833344	3246	88	G>A	0.8693	0.8523	17	21	0.51
NPAS2_rs3849381	100910337		107	C>T	0.2734	0.2944	48	39	0.34
NPAS2_rs1811399	100937532	27195	109	C>A	0.2294	0.1789	32	54	<b>0.018</b>
NPAS2_rs2117714	100980295	42763	106	A>G	0.7005	0.7358	48	33	0.1
NPAS2_rs356651	101000316	20021	106	C>T	0.1439	0.1321	25	30	0.5
NPAS2_X3_C_T	101000390	74	110	T>C	0.1886	0.2318	47	28	<b>0.028</b>
NPAS2_rs895521	101010613	10223	108	C>T	0.8333	0.8519	37	29	0.32
NPAS2_rs13017718	101012554	1941	108	T>C	0.0718	0.0694	14	15	0.85
NPAS2_rs895520	101036455	23901	105	A>G	0.4167	0.3952	37	46	0.32
NPAS2_INT7_G_A	101038851	2396	108	A>G	0.0856	0.088	18	17	0.87
NPAS2_X8_G_A	101039093	242	108	G>A	0.831	0.8565	32	21	0.13
NPAS2_X11_A_G	101043469	4376	108	A>G	0.8588	0.875	30	23	0.34
NPAS2_rs1562313	101045973	2504	109	C>T	0.8211	0.8028	27	35	0.31
NPAS2_X12_A_T	101046160	187	108	A>T	0.8542	0.875	32	23	0.22
NPAS2_rs2305160	101049822	3662	108	G>A	0.6829	0.6435	38	55	0.08
NPAS2_rs2305159	101049961	139	108	A>C	0.6898	0.6898	45	45	1
NPAS2_X15_C_T	101052709	2748	108	C>T	0.8264	0.8519	33	22	0.14
NPAS2_rs2305158	101070569	17860	110	A>G	0.2341	0.2318	40	41	0.91



**Table 2-4.** Results of the transmission disequilibrium tests (continued)

<i>Gene/marker</i>	<i>Coordinate (bp)</i>	<i>Distance from previous marker (bp)</i>	<i>Number of trios typed</i>	<i>Allele</i>	<i>Frequency in parents</i>	<i>Frequency in children</i>	<i>Transmitted</i>	<i>Non-transmitted</i>	<i>P-value</i>
Per1_rs2289591	7988735		106	T>G	0.2264	0.1887	29	45	0.063
Per1_rs2253820	7988894	159	108	A>G	0.2269	0.2315	30	24	0.41
Per1_rs885747	7991462	2568	107	C>G	0.6262	0.5841	32	50	<b>0.047</b>
Per1_rs3027178	7993810	2348	110	A>C	0.6864	0.7273	55	41	0.15
Per1_rs6416892	7983585	75	107	C>A	0.3364	0.3879	48	30	<b>0.042</b>
Per1_rs2585398	7995585	1700	104	G>T	0.5048	0.4663	39	51	0.21
Per2_In1_C_T	238926115		87	C>T	0.954	0.954	7	7	1
Per2_rs2304669	238947663	21548	32	G>A	0.1719	0.125	6	12	0.16
Per2_In17_G_T	238947792	129	90	G>T	0.9806	0.9889	5	2	0.26
Per2_Ex19F_L	238951288	3496	90	C>T	0.025	0.0111	2	7	0.1
Per2_Ex23C_T	238959378	8090	87	T>C	0.1408	0.1322	17	20	0.62
Per3_rs228729	7779961		87	C>T	0.6839	0.6379	28	42	0.094
Per3_rs228688	7813396	33435	24	T>G	0.3854	0.375	12	13	0.84
Per3-21-T/G	7814949	1553	86	G>T	0.1512	0.1395	18	22	0.53
Per3_rs2859387	7821514	6565	85	A>G	0.45	0.4059	34	47	0.15
Per3_rs228665	7825222	3708	84	C>G	0.3065	0.3333	37	30	0.39
Per3-27-C/T	7831399	6177	24	C>T	0.1667	0.1875	8	6	0.59
Per3_rs4908699	7839540	8141	85	A>G	0.1412	0.1176	18	26	0.23
TIME_rs2291739	55100920		88	C>T	0.4489	0.4659	50	44	0.54
TIME_rs2291738	55101548	628	86	G>A	0.5174	0.5	43	49	0.53
TIME_rs774047	55102189	641	85	G>A	0.5471	0.5412	44	46	0.83
TIME_rs774033	55111578	9389	88	G>A	0.5426	0.5284	44	49	0.6
TIME_rs2279665	55113961	2383	84	G>C	0.5387	0.5119	40	49	0.34

## 2.6 Discussion.

We hypothesized that clock genes are implicated in autistic disorder (Wimporly et al. 2002). In this candidate gene study, we found significant associations in two of the genes studied: *PER1* and *NPAS2*, with two SNPs in each gene reaching conventional levels of statistical significance. In addition, a high proportion of all possible haplotypes in *NPAS2* was also significant. It is difficult to interpret the real significance of these results. On the one hand, if we perform a correction for the multiple testing of all SNPs analyzed in this study, none of the results will remain significant. On the other hand, our sample size is too small to allow us to identify markers that would withstand correction for multiple testing, unless they conferred huge effects. This is unlikely to be the case in a disorder of complex inheritance such as autism where current models predict several interacting genes of small effect, each contributing to the phenotype of autistic disorder

(Pickles et al. 1995). In this context, we note that the NPAS2/ ARNTL (BMAL1) heterodimer is an activator of *PER1* (Reick et al. 2001); It follows that genetic variation affecting the function of NPAS2 could be additive to genetic variation affecting the function of PER1.

We used all available cases of autistic disorder in the AGRE database that satisfied our strict selection criteria and we were therefore unable to consider replicating our initial results in a large sample. We placed considerable emphasis on subject diagnosis and intelligence in order to select a very homogeneous sample; restricting the phenotype in this way also increased the experimental power over a limited sample size. We satisfied the strictest diagnostic criteria for autistic disorder using ADOS-G and ADIR for each case. By selecting for more able subjects, we derived a sample that was predominantly (85%) high functioning and unusually homogeneous in this field of research to date. To confirm or refute our results, future studies would need to collect other samples that have a similar composition.

The expression patterns of *NPAS2* (Zhou et al. 1997; Garcia et al. 2000) and *PER1*, (Sun et al. 1997) when considered together, cover brain areas found to be altered in individuals with autism: the cerebellum, forebrain and limbic system including the hippocampus and amygdala (Bauman and Kemper 1985; Bauman and Kemper 1986; Bauman and Kemper 1988; Raymond et al. 1996). Non-circadian expression of *PER1* occurs in the cerebrum, hypothalamus and cerebellum (granular layer), while circadian *PER1* expression is found in the SCN hypothalamus and Purkinje cells (Sun et al. 1997). The implication of *NPAS2* in complex emotional memory (Garcia et al. 2000) and our results associating *NPAS2*

with autistic disorder, lend tentative support to findings that the thalamo-cortico-amygdalo pathway (associated with complex emotional memory (Hamann 2001; Zald 2003; Phelps 2004) is dysfunctional in autism (Bauman and Kemper 1988; Kemper and Bauman 2002; Bauman and Kemper 2005). This may be relevant for related neurological and psychological theories that focus on episodic/contextual memory dysfunction and the hippocampus/limbic system (Ben Shalom 2003; Salmond et al. 2005) and/or the amygdala (Baron-Cohen et al. 2000; Howard et al. 2000; Bachevalier and Loveland 2006).

Evidence of altered sleep architecture in autism is also supported by our positive result for *NPAS2*, as *NPAS2* is a transcriptional regulator of non-rapid eye movement sleep in mice (Franken et al. 2006). Several of the sleep features attributable to *Npas2* knockout in mice are recognized in high-functioning autism and Asperger syndrome: longer sleep latency, prolonged waking after sleep onset and reduced non-rapid-eye-movement sleep (Limoges et al. 2005). Franken et al. (2006) also show that the effect of *Npas2* knock-out on sleep is modulated by the sex of the animal (with males being more adversely affected than females), an observation that may have relevance to autism, given the male to female ratio of 3.8:1 (Fombonne 1999).

The role of *per* in modulating high-frequency oscillators concerned with communicative timing (Kyriacou and Hall 1980; Alt et al. 1998; Ritchie et al. 1999) together with our findings of association of *PER1* with autistic disorder, strengthens the notion that temporal deficits are quintessential to autistic disorder (Segawa et al. 1981; Newson 1984; Courchesne et al. 1994; Richdale and Prior 1995; Boucher 2000; Boucher 2001; Wimpory et

al. 2002; Szelag et al. 2004; Welsh et al. 2005).

Purkinje neurons are important for learning appropriate timing (Koekkoek et al. 2003; Kotani et al. 2003) and their abnormally low number in the cerebella of autistic subjects (Ritvo et al. 1986; Kemper and Bauman 2002) is considered a keystone biological observation implicating cerebella dysfunction (Courchesne et al. 1994b; Akshoomoff et al. 2002). In this context, reports of a *PER1* interacting protein (PIPS) in rat that co-translocates with *PER1* into the nucleus (Matsuki et al. 2001) and which is further shown to be required for neuronal growth factor-mediated neuronal survival in P12 cells (Kiyama et al. 2006), tentatively suggests a role for *PER1* in the lack of Purkinje neurons of the cerebellum in autism. We are currently engaged in an investigation of PIPS as well as further analysis of *PER1* and *PER2*.

The gene *ENGRAILED2* (*EN2*) functions in the development of the cerebellum (Song and Joyner 2000) and has been shown to be associated with autistic spectrum disorders (Gharani et al. 2004; Benayed et al. 2005). A risk allele that interacts with *EN2* variants to perturb the normal spatial/temporal expression of *EN2* could alter normal brain development, a point that has resonance with *PER1*'s role in the cell division cycle (Matsuo et al. 2003) and dopamine D2 receptor-mediated signaling (Yujnovsky et al. 2006).

The timing, contextual/memory deficits and other implications suggested by the association of *PER1* and *NPAS2* with autistic disorder may manifest concurrently and/or developmentally, for example, through infant–adult interaction that is timing dependent (Murray and Trevarthen 1985; Nadel et al. 1999; Trevarthen and Aitken 2001; Crown et al. 2002).

Neonatal physiological oscillations (sleep–wake cyclicality and cardiac vagal tone) predict mother–infant synchrony at 3 months (Feldman 2006). In turn, the temporal organization of 3-month-old infants' social attention predicts mother–infant synchrony at 9 months (Feldman et al. 1996). Mother–infant synchrony itself predicts wide-ranging later developments (Feldman et al. 1996; Feldman et al. 1999; Jaffe et al. 2001). Temporal synchrony/reciprocity difficulties of autism (Feldstein et al. 1982; Tantam et al. 1993; Bebko et al. 2006) have also been recorded in autistic infancy (Kubicek 1980; Wimpory et al. 2000; Trevarthen and Daniel 2005) where they may serve as a functional deficit contributing to the developmental cascade of autistic disorder (Newson 1984; Courchesne et al. 1994; Wimpory et al. 1995; Trevarthen 2000; Wimpory et al. 2002).

Problems in timing, memory and sleep are all characteristics of autistic disorder and aspects of timing, memory and sleep are each clock-gene regulated in other species. The association of clock genes with autistic disorder suggests a role for these genes in autism. It also encourages replication of our study and the collection of further data, particularly on timing, sleep and emotional complex/contextual memory, in high-functioning autistic disorder.

### **3 Are autism-associated genes clock-controlled genes and what possible effects might the autism-associated SNPs in *PER1* and *NPAS2* have?**

#### **3.1 Introduction.**

To investigate the possible functionality of the autism-associated SNPs in *PER1* and *NPAS2* a detailed examination of the SNPs and the regions embedding the SNPs was carried out using internet-based bioinformatics software and databases. The aim of this analysis was to make some biological sense for why these non-protein-coding SNPs should show association with autistic disorder. Any positive findings pertaining to the functionality of these SNPs could then be the starting point of further investigations of the effect of these polymorphisms in living cells. Direct biochemical experimentation in autism is clearly limited, as it is not feasible to obtain live brain tissue samples from autistic individuals and the use of postmortem material would be appropriate only for final-and-conclusive, rather than preliminary investigations. Thus the use of bioinformatics to analyze possible implications of the association of these polymorphisms with autistic disorder is presented here as a practicable solution and as a next-step in investigating whether these SNPs have any functional effect.

The bioinformatic analysis began with the conventional and then developed to cover the possibility of novel, previously unrecognized, features at these locations. After re-checking and redefining the genome coordinates of the SNPs with the most up-to-date genetic maps of the

human genome publicly available, the direct effect of the SNPs on protein structure could be discounted as three of the SNPs were intronic while rs6416892 is down stream of the 3'UTR of *PER1*.

The new coordinates were different from those published (Nicholas et al. 2007) due to update of the human genome maps. However, the relative positions of the SNPs in the *NPAS2* and *PER1* sequences and to other features within the genes stayed the same between the old and new versions of the gene maps. Subsequently, the regions in the immediate vicinity of each SNP were scanned for control elements or regulatory regions. The haplotypes that were found to be significantly associated with autistic disorder were similarly scanned. Finally, transcript analysis of the mRNA strand that is predicted to be transcribed from these non-coding regions was used to look for mRNA secondary structures that could have regulatory potential.

This bioinformatic analysis took an integrative, broad approach and consideration of possible data sources indicated that the Ensembl Genome Browser would be the primary software tool, for its facility to bring together several data feeds from individual authoritative sources. Ensembl allows for example, multiple genomic features to be considered simultaneously and has the facility for comparing genomic regions between related species using the alignment tool. Additionally, the Ensembl BLAST search tool may be used to search genomes for similar DNA or protein sequences or to highlight the location of features of interest. Ensembl produces and maintains automatic annotation on the human genome and other selected genomes. This web based bioinformatics tool is a joint project between EMBL-EBI (The European

Bioinformatics institute, which is part of the European Molecular Biology Laboratory) and the Wellcome Trust Sanger Institute. Ensembl is an open access site publicly available on line at <http://www.ensembl.org>. In addition to Ensembl, a number of other online resources were used in this analysis. On-line tools used are listed in Table 3.1 and data was cross-referenced between databases where possible.

An analysis of autism genes and clock-controlled genes was also made by merging an autism candidate gene dataset with a dataset of genes showing circadian patterns of expression in mammalian brain and a dataset of genes containing circadian regulatory modules. This analysis aimed to illuminate which genes linked to autism might be impacted by functional genetic variation that affects the circadian rhythm and/or clock gene expression.

**Table 3-1. Bioinformatics tools.**

<b>Bioinformatics resource</b>	<b>Internet addresses and summary of functions</b>
Autism Chromosome Rearrangement Database	<a href="http://projects.tcag.ca/autism/">http://projects.tcag.ca/autism/</a> (Chromosomal anomalies found in autistic individuals)
ClustalW2	<a href="http://www.ebi.ac.uk/Tools/clustalw2/index.html">http://www.ebi.ac.uk/Tools/clustalw2/index.html</a> (Sequence alignment tool)
DBASS3	<a href="http://www.som.soton.ac.uk/research/geneticsdiv/dbass3/">http://www.som.soton.ac.uk/research/geneticsdiv/dbass3/</a> (Aberrant 3' splice sites causing disease in humans)
DBASS5	<a href="http://www.som.soton.ac.uk/research/geneticsdiv/dbass5/">http://www.som.soton.ac.uk/research/geneticsdiv/dbass5/</a> (Aberrant 5' splice sites causing disease in humans)
Ensembl	<a href="http://www.ensembl.org/index.html">http://www.ensembl.org/index.html</a> (Multi species comparative genomics site with BLAST etc)
GeneCards	<a href="http://www.genecards.org/">http://www.genecards.org/</a> (Summary of information linked to gene names)
Genotator	<a href="http://genotator.hms.harvard.edu/geno/disorder/autistic_disorder/">http://genotator.hms.harvard.edu/geno/disorder/autistic_disorder/</a> (A meta-query engine for disease associated genes)
GLIDERS	<a href="http://mather.well.ox.ac.uk/GLIDERS/">http://mather.well.ox.ac.uk/GLIDERS/</a> (Genome-wide Linkage DisEquilibrium Repository and Search engine)
HUGO Gene Nomenclature Committee	<a href="http://www.genenames.org/">http://www.genenames.org/</a> (Human Genome Organization consortium for gene nomenclature)



**Table 3-1. (Continued). Bioinformatics tools.**

KEGG	<a href="http://www.genome.jp/kegg/">http://www.genome.jp/kegg/</a> (Kyoto Encyclopedia of Genes and Genomes Integrated gene atlas with pathways analysis)
MirBASE	<a href="http://www.mirbase.org/">http://www.mirbase.org/</a> (MicroRNA sequence repository and sequence search facility)
MyHits	<a href="http://myhits.isb-sib.ch/cgi-bin/index">http://myhits.isb-sib.ch/cgi-bin/index</a> (Protein domains and motifs etc.)
NHGRI (GWAS)	<a href="http://www.genome.gov/26525384#1">http://www.genome.gov/26525384#1</a> (National Human Genome Research Institute. Database of Genome Wide Association Study results)
PEDB	<a href="http://promoter.cdb.riken.jp/circadian.html">http://promoter.cdb.riken.jp/circadian.html</a> (Mammalian Promoter Enhancer Database: Circadian regulatory elements)
RNAanalyzer	<a href="http://rnaanalyzer.bioapps.biozentrum.uni-wuerzburg.de/server.html">http://rnaanalyzer.bioapps.biozentrum.uni-wuerzburg.de/server.html</a> (Predicts the 2D folding of RNA Molecules)
RNA Self Containment Server	<a href="http://w1.genomics.upenn.edu/cgi-bin/sc/sc.cgi">http://w1.genomics.upenn.edu/cgi-bin/sc/sc.cgi</a> (Predicts a RNA hairpin structure's resistance to mutation)
Sequence Editor	<a href="http://www.fr33.net/seqedit.php">http://www.fr33.net/seqedit.php</a> (Sequence clean up and conversion tool)
SpliceScan2	<a href="http://splicescan2.lumc.edu/">http://splicescan2.lumc.edu/</a> (Predicts splice sites)
TargetScan	<a href="http://www.targetscan.org">http://www.targetscan.org</a> (microRNA target predictor)
TESS	<a href="http://www.cbil.upenn.edu/cgi-bin/tess/tess?RQ=WELCOME">http://www.cbil.upenn.edu/cgi-bin/tess/tess?RQ=WELCOME</a> (Transcription Element Search System)
Vienna RNA Web Servers	<a href="http://rna.tbi.univie.ac.at/">http://rna.tbi.univie.ac.at/</a> (Predicts the 2D folding of RNA molecules, structural conservation etc.)

## 3.2 Which autism candidate genes are clock-controlled genes?

### 3.2.1 Introduction.

CLOCK/ARNTL-binding elements (E-Box, = CACGTG), D site of albumin promoter (albumin D-box) binding protein (DBP)/Nuclear factor, interleukin 3 regulated (NFIL3) binding elements (D-Box = TTATG[T/C]AA), and nuclear receptor subfamily 1, group D, member 1 (NR1D1)/RORA binding elements (RRE = [A/T]A[A/T]NT[A/G]GGTCA) are regulatory motifs associated with circadian patterns of gene expression (Hayes et al. 2005;

Ueda et al. 2005). In terms of genome wide distributions of these elements, E-boxes are typically found associated with transcription start sites (promoter regions), while D-boxes and RREs are not more likely to be near TSS (Kumaki et al. 2008). If the hypothesis, that the circadian clock plays a role in autism is true, and if the genes found in autism association and linkage studies are *bona fide* autism genes, then it follows that clock-controlled genes should be represented in the set of autism candidate genes.

This experiment tests this argument by looking for overlap between genes showing circadian patterns of expression in mouse prefrontal cortex (Yang et al. 2007) (called here PF cortex genes), genes that contain conserved E-box, D-Box and RRE control elements (called here E-box, D-Box and RRE genes (Kumaki et al. 2008)) and genes that show positive association with autism here called autism genes, listed and in Genotator (Table 3-1) (Wall et al. 2010).

### 3.2.2 Method

The computer applications Microsoft Excel and Microsoft ACCESS were used to interrogate supplementary information associated with the publications: “Analysis and synthesis of high-amplitude Cis-elements” (Kumaki et al. 2008) and “Genome-wide expression profiling and bioinformatic analysis of diurnally regulated genes in the mouse prefrontal cortex” (Yang et al. 2007). Searches using The National Centre for Biotechnology Information (NCBI) PubMed database and data from genome wide association study (GWAS) database of The National Human Genome Research Institute (NHGRI) Bethesda, Maryland USA

([www.genome.gov](http://www.genome.gov)) were also used to compile a list of autism-associated genes.

The number of elements in each dataset was: E-box genes (n = 1,108), D-box genes (n = 2,314), RRE genes (n = 3,288), PF cortex genes (n = 2,645) and autism genes (n = 110). Using the standard tools of the Microsoft Excel and Microsoft ACCESS applications, Autism genes were screened for overlap with with E-box genes, D-Box genes, RRE genes and genes that showed circadian patterns of expression in mammalian prefrontal cortex in order to determine overlap between autism genes, genes in the cortex that show circadian regulation, and genes with circadian regulatory elements. This overlap should indicate autism genes that are likely to show circadian regulation in the human PF cortex. The strongest autism candidate genes (i.e. genes supported by a substantial body of research with compelling over-all positive results for autism) were investigated further by using Ensembl to check the position of the E-boxes, D-boxes and RREs in the candidate gene.

### 3.2.3 Results.

A number of the autism genes were found in the mouse prefrontal cortex diurnally regulated gene data set and some of these were also identified in the clock gene *cis*-acting element data sets (E-box genes, D-Box genes and RRE genes). Table 3-2 shows the results of the screen for overlap between the data sets of autism genes and each of PF cortex genes, E-box genes, D-Box genes and RRE genes. Table 3-3 shows the autism genes that are present in at least one of the clock element data sets and are represented amongst the PF cortex genes.

**Table 3-2. Autism genes and circadian control element genes.**

<b>Autism vs. D-BOX</b>	<b>Autism vs. E-BOX</b>	<b>Autism vs. RRE</b>	<b>Autism vs. PF CORTEX</b>
<i>ANK3</i>	<i>CACNA1C</i>	<i>ANK3</i>	<i>ADIPOR2</i>
<i>ARX</i>	<i>CENTG2</i>	<i>ARX</i>	<i>AHI1</i>
<i>CACNA1C</i>	<i>DYRK1A</i>	<i>ATP1A2</i>	<i>ANK3</i>
<i>CDH10</i>	<i>FBXL19</i>	<i>CACNA1C</i>	<i>ARX</i>
<i>CDH9</i>	<i>LARGE</i>	<i>CDH10</i>	<i>BCL2</i>
<i>CNTN4</i>	<i>RAI1</i>	<i>CENTG2</i>	<i>BDNF</i>
<i>CNTNAP2</i>	<i>RARA</i>	<i>CNTN4</i>	<i>BZRAP1</i>
<i>CSMD1</i>	<i>RELN</i>	<i>DMD</i>	<i>DMPK</i>
<i>DMD</i>	<i>RIMS3</i>	<i>GABRB2</i>	<i>DYRK1A</i>
<i>EN2</i>	<i>SHANK3</i>	<i>GRIK2</i>	<i>FBXL19</i>
<i>GAS1</i>	<i>TBX1</i>	<i>GRIN2A</i>	<i>FOXP2</i>
<i>GRIK2</i>		<i>GRM8</i>	<i>GABRB2</i>
<i>GRM8</i>		<i>ITGA4</i>	<i>GAS1</i>
<i>ITGA4</i>		<i>KCNMA1</i>	<i>GTF2IRD1</i>
<i>KCNMA1</i>		<i>LARGE</i>	<i>ITGA4</i>
<i>LARGE</i>		<i>MET</i>	<i>KRAS</i>
<i>MET</i>		<i>PARK2</i>	<i>LIMK1</i>
<i>PARK2</i>		<i>PRKCB1</i>	<i>LZTS2</i>
<i>PITX1</i>		<i>RARA</i>	<i>MET</i>
<i>PRKCB1</i>		<i>RELN</i>	<i>MSRA</i>
<i>RORA</i>		<i>RIMS3</i>	<i>NIPBL</i>
<i>SEMA5A</i>		<i>RORA</i>	<i>OMG</i>
		<i>SEMA5A</i>	<i>RFWD2</i>
			<i>SEMA5A</i>
			<i>SHANK3</i>
			<i>TSC1</i>

Overlapping phenotypes associated with circadian regulated genes

Blue = bipolar and schizophrenia.

Red = schizophrenia and autism

Green = bipolar, autism and schizophrenia

These fields were merged and filtered to find which autism genes contained clock regulatory elements and showed circadian expression in the prefrontal cortex.

**Table 3-3. Overlap between autism genes, circadian element genes and cortex genes.**

<b>Gene</b>	<b>Associated disorders etc.</b>
<i>ANK3</i>	Bipolar Disorder (Ferreira et al. 2008)
<i>ARX</i>	Mental retardation (Shoubridge et al. 2010 )
<i>DYRK1A</i>	Phosphorylates <i>SIRT1</i> (Guo et al. 2010) Regulates <i>CRY2</i> (Kurabayashi et al. 2010) Down Syndrome (Altafaj et al. 2001)
<i>FBXL19</i>	Gene of unknown function (located 16p11.2, an autism susceptibility locus)
<i>GABRB2</i>	Autism (Ma et al. 2005). Schizophrenia (Shi et al. 2008)
<i>GAS1</i>	Autism. (Weiss et al. 2009)
<i>ITGA4</i>	Autism (Correia et al. 2009). Immune response and cancer
<i>MET</i>	Autism & cancer development (Campbell et al. 2008).
<i>SEMA5A</i>	Autism (Weiss et al. 2009). Parkinsons disease (Anantharam et al. 2007)
<i>SHANK3</i>	Autism (Durand et al. 2007)

#### 3.2.4 Conclusion.

Of particular note is the finding that the pre-synaptic protein *SHANK3* is listed in the E-Box, autism genes and PF cortex gene tables. There is a strong case for involvement of *SHANK3* in autism aetiology (Moessner et al. 2007; Bourgeron 2009) but as far as this author is aware this is the first time that *SHANK3* has been highlighted as: containing a conserved clock-control element; being a synaptic gene involved in the autism development pathway and showing circadian oscillation of expression in mammalian prefrontal cortex. It is also very interesting to note that out of only a few genes that have reached genome wide significance in GWAS for autism (Yang et al. 2009; Weiss et al. 2009) two of these genes *SEMA5A* and *CDH10* are likely clock-regulated genes by virtue of conserved circadian control elements (Kumaki et al. 2008).

GWAS have recently begun to reinforce previous assertions that there is considerable genetic overlap between autism, bipolar disorder and

schizophrenia. For example SNPs in the schizophrenia susceptibility gene *ZNF804A* are also significantly associated with bipolar disorder and when the datasets are merged this association becomes more significant ( $P = 4.1 \times 10^{-13}$ ) (Williams et al. 2010). In common with some other neuropsychiatric disorders, bipolar disorder and certain types of schizophrenia present with circadian disruption and modest genetic association has been found between schizophrenia and clock genes (Mansour et al. 2009) while strong association between bipolar disorder and the circadian clock regulating kinase *GSK3 $\beta$*  is founded on genetic and pharmacological evidence (Lachman et al. 2007; Beaulieu et al. 2009). It appears plausible therefore, that the sets of genes conferring vulnerability to schizophrenia, bipolar disorder and autism overlap, and, this overlapping set of genes includes clock and clock-regulated genes. These findings encourage further genetic studies of clock genes in large data sets of pooled autism, schizophrenia and bipolar disorder samples and other studies focused on investigating circadian regulation of genes of major psychiatric relevance i.e. *SHANK3*, *GABRB2*, *CNTNAP2*, *CACNA1C*, *ANK3* and *EN2*.

### **3.3 Searching for regulatory features coincident with the autism-associated clock gene SNPs and haplotypes.**

#### 3.3.1 Introduction.

As new issues of the human genome map supersede older versions, refinement and corrections can adjust the co-ordinates of genes and new found features are added. Therefore, the most up to date publicly available

genome coordinates for the autism-associated SNPs were obtained and these locations scrutinized with the Ensembl genome browser. The genome sequence data feed for Ensembl is derived from The *Homo sapiens* High Coverage Assembly from the Genome Reference Consortium, which is an alliance of The Wellcome Trust Sanger Institute, The Genome Center at Washington University, The European Bioinformatics Institute and The National Center for Biotechnology Information. The data set consists of gene models built from the gene-wise alignments of the human proteome as well as from alignments of human cDNAs using the cDNA2genome model (Slater and Birney 2005).

### 3.3.2 Method

Using Ensembl the SNP locations were updated and searched for inclusion in, or proximity to, known regulatory features, functional elements and zones of sequence conservation. Areas of sequence conservation that fell outside exons and known regulatory features were given special consideration because such constraint might indicate sequences of as yet undetermined functionality.

From the Ensembl home page at <http://www.ensembl.org/index.html>, the species genome was selected, in this case “Human”, *Homo sapiens* and within the *Homo sapiens* section, the code for the single nucleotide polymorphism under investigation, e.g. rs1811399, was entered into the search box and submitted. For rs1811399, the server confirmed the query matched 1 unique database entry with 7 synonyms (Table 3-4).

**Table 3-4. Synonyms of rs1811399.**

-HGvbase:SNP001650070
Illumina_Human660W-quad:rs1811399
Illumina_CytoSNP12v1:rs1811399
ENSEMBL:Venter:ENSSNP1565019
ENSEMBL:Watson:ENSSNP6889409
ENSEMBL:celera:ENSSNP1565019
Illumina_Human1M-duoV3:rs1811399.

Following the hyperlink for rs1811399, the variation tab gave information on the alleles for rs1811399 (C/A) and the genome location for this SNP, 2:101479014 (forward strand). The figure 2 before the colon refers to the chromosome number and after the colon, the genome coordinates. Hyperlinks to the validation status, Linkage disequilibrium data and flanking sequence for rs1811399 were also to be found on this page.

By selecting the “region in detail” link, the genome browser displayed the chromosomal location and gene context of *NPAS2* and indicated the location of the SNP in *NPAS2* with a vertical red bar. By using the zoom feature in the “location” toolbar it was possible to zoom-in to determine the precise coordinates of features of interest such as regulatory features or conserved sequence elements that appeared close or coincident with the SNPs. After gaining a graphical fix on the SNP in relation to local regulatory features etc. the co-incidence or proximity of the features to the SNP was confirmed arithmetically by comparing the numerical genome coordinates of the SNP with the numerical co-ordinates of the feature.

Each graphical feature on the browser display has a unique Ensembl reference code that includes the genome coordinates. This information is accessible by following the hyperlink associated with the feature. Thus each SNP was screened for co-occurrence with regulatory elements, small RNAs etc. The locations of the SNPs and features were subjected to



crosschecks. For example, the location of each SNP as given in the text on the Ensembl variation page was cross referenced with the location of the SNP as shown on the scale bar at the bottom of the gene map on the Ensembl “region in detail” page. Each view was “customized” by switching on/off selectable tracks so that all relevant data feeds were searched.

Where sequence data external to Ensembl was used, for example from a database of high-amplitude *Cis*-elements in the mammalian circadian clock (Kumaki et al. 2008), the query sequences were used in Ensembl BLAST searches to find exact unique matches and thus locate and confirm the position of these regulatory elements in terms of the current genome coordinates. This process was repeated for each of the SNPs analyzed.

After the SNPs were investigated singly, the most significant haplotype pairs were analysed. In this context, the term haplotype describes pairs of SNPs on a single chromatid that are analysed together in statistical tests, such as is carried out with the programme TDTPHASE (Dudbridge 2003). Significant haplotype pairs define sequence blocks that may contain additional SNPs associated with the phenotype other than the SNPs making up the boundaries of the block. Haplotype analysis is valuable where the SNP density of an association study is low or for checking for functional associations by focussing on sections of the gene likely to harbour a functional variant of interest. Significant haplotype pairs may thus delineate a DNA region containing SNPs that are more strongly associated with the phenotype under investigation.

In order to increase the resolution of haplotypes, The GLIDERS (Genome-wide LInkage DisEquilibrium Repository and Search engine) tool

was used to attempt to impute linkage between the autism associated SNPs and other SNPs within the region that were not assayed in the association study. The GLIDERS search engine searches for SNPs that are in linkage disequilibrium with other SNPs genotyped in The International HapMap Consortium database. The search engine attempts to match the query SNP with one of a pair of SNPs that have been previously calculated to be in linkage disequilibrium. The parameters of the search can be varied such that any pair-wise combination of HapMap SNPs can be considered. The GLIDERS linkage disequilibrium statistics stored in the database were obtained by calculating the linkage disequilibrium between every possible HapMap phase 2 and 3 SNP pairing where both SNPs had minor allele frequency (MAF)  $\geq 5\%$ . The SNPs and haplotypes are listed in Table 3-5.

**Table 3-5. Summary of the autism associated SNPs and haplotypes in *NPAS2* and *PER1*.**

Gene name	SNP code	Genome coordinate*	Gene region	Alleles
<i>NPAS2</i>	<b><u>rs1811399</u></b>	2:101479014	Intron 1	A/C
	<u>rs2117714</u>	2:101521777	Intron 2	G/A
	<b>rs34705978</b>	2:101541872	Intron 3	T/C
<i>PER1</i>	<b><u>rs885747</u></b>	17:8050737	Intron 12	C/G
	<u>rs2253820</u>	17:8048169	Exon 17	T/C
	<b>rs6416892</b>	17:8042860	downstream	C/A

The SNPs underlined are the autism-associated haplotypes with the most significant p values. SNPs in bold have individually significant p values ( $< 0.05$ ). Alleles in red are autism-enriched alleles.

### 3.3.3 Results.

#### 3.3.3i NPAS2, SNPs individually (rs1811399 and rs34705978).

The NPAS2 gene is located on the long arm of chromosome 2 close to the centromere region in band q11.2 (forward strand). The closest fully annotated genes are *PDCL3*, *RPL31* and *TBC1D8*. *PDCL3* is approximately 0.25mb upstream of *NPAS2* while *RPL31* and *TBC1D8* are immediately downstream of *NPAS2*. A novel microRNA gene, ENSG00000238574, is located within *TBC1D8* towards the 3' end of the gene (Figure 3-1).

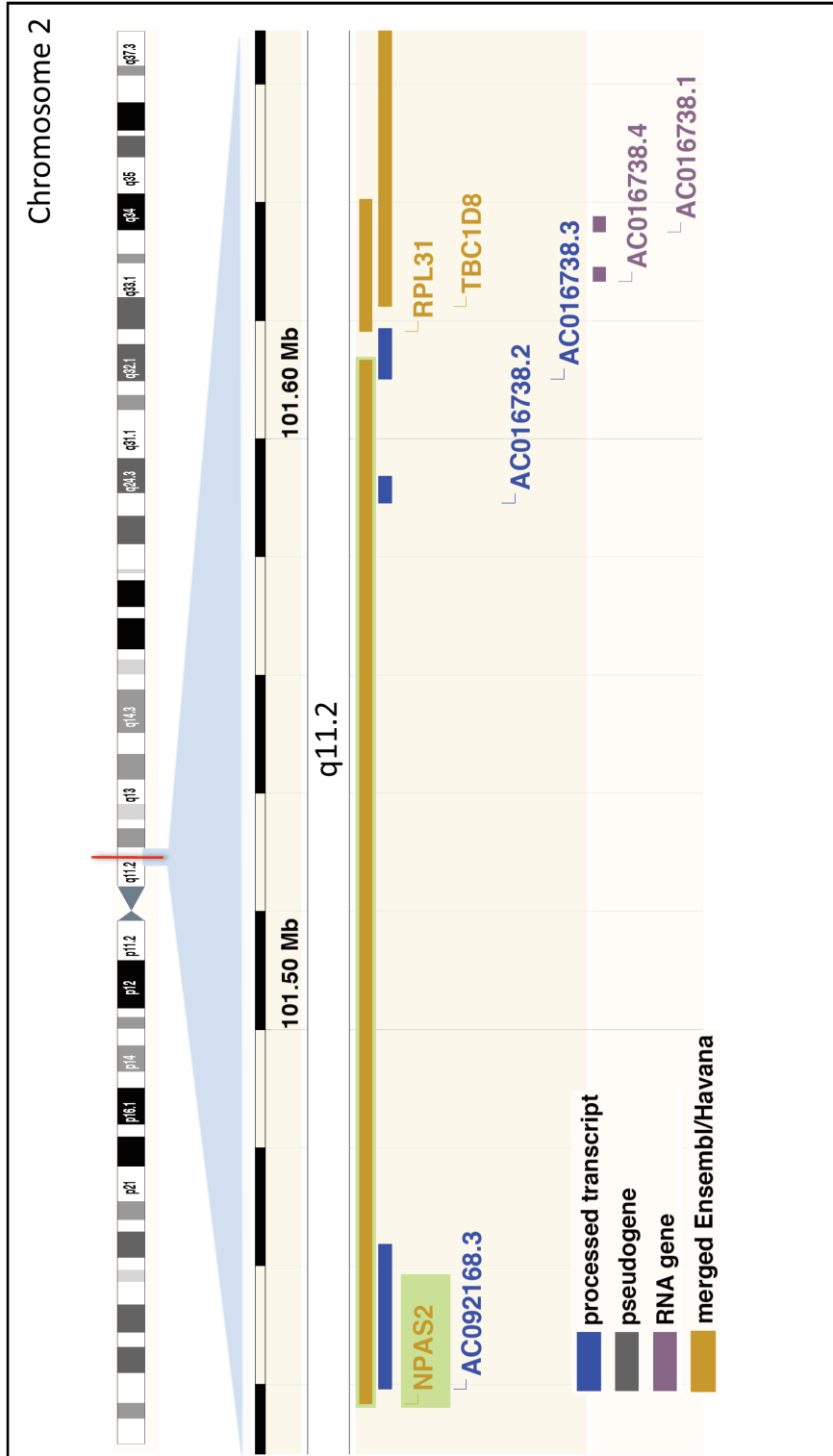
The autism-associated SNP rs1811399 is located in the first intron of *NPAS2* at genome coordinate 2:101479014. Using all possible filter tracks of Ensembl no features of interest coincided with the location of this SNP. The SNP rs2117714 (that together with rs1811399 makes up the best haplotype) is located down stream of rs1311399 just beyond the first exon. The SNP rs2117714, of itself, did not show significant association with autism nor was it found to coincide with any regulatory elements or other features of interest (Figure 3-2).

SNP rs34705978 showed significant association with autistic disorder from the association test (chapter 2) and is located in the third intron of *NPAS2* at 2:101541872, immediately down-stream of exon 3. There are currently no synonyms in the database for this SNP. This SNP is located within a regulatory feature ENSR00000123315 and its core maps to coordinates 2:101540977-101542157. This regulatory feature contains a differentially methylated region where CpG dinucleotides are methylated (or not) in a tissue dependent manner (Figure 3-3 and Figure 3-4).

### 3.3.3ii *NPAS2*, most significant haplotype.

Haplotype analysis (Dudbridge 2003) of *NPAS2* showed 40 out of the 136 possible two-marker combinations were significant at the  $P < 0.05$  level. The best result was between SNPs rs1811399 and rs2117714,  $P = 0.001$ . Using Ensembl this haplotype block of 42,763bp was searched for regulatory features. This region of genome coordinates 2:101479014 to 2:101521777 covers the second half of the first intron of *NPAS2* and just includes the very small second exon of the gene. This haplotype contains one of the three conserved RREs of *NPAS2* (Table 3-6) along with a number of cell-line specific protein binding sites (Table 3-8).

From the analysis of conserved circadian control elements (chapter 3.2) a RORA/NR1D1 binding element (RRE) was found to be located in the central region of the autism haplotype (Figure 3-2, Table 3-6 and Table 3-7). This binding site was checked with a whole genome (human) BLAST search to see if it matched the RRE listed in the database of conserved clock gene regulatory elements (Kumaki et al. 2008). Notably this is one of only four such RREs listed for *NPAS2* and it has the highest score. The score relates to the false discovery rate (FDR) where the FDR is inversely proportional to the score. The scoring formula was developed using Hidden Markov Model analysis trained on known circadian elements in the mouse genome (Kumaki et al. 2008).



**Figure 3-1. Banding pattern of human chromosome 2.** The location of *NPAS2* is indicated by the red vertical bar. The magnified region (indicated by the pale blue segment) shows *NPAS2* in the context of genes *RPL31* and a portion of *TBC1D8* immediately adjacent to the 3' end of *NPAS2*. *PDCL3* (not shown) is ~ 0.24Mb upstream of *NPAS2*. The scale bar indicates genome coordinates in mega bases.

BLAST search for the location of *NPAS2* RRE 1 (with the current Ensembl human genome sequence gave 100% unique match (23 bp) at location: 2 101504350 to 101504372 (+), within the first intron of *NPAS2* (RRE 1, Table 3-6). The conservation and the manner in which the database of RREs was compiled indicate this element is likely to be a *bona fide* RRE (Table 3-7) (Kumaki et al. 2008).

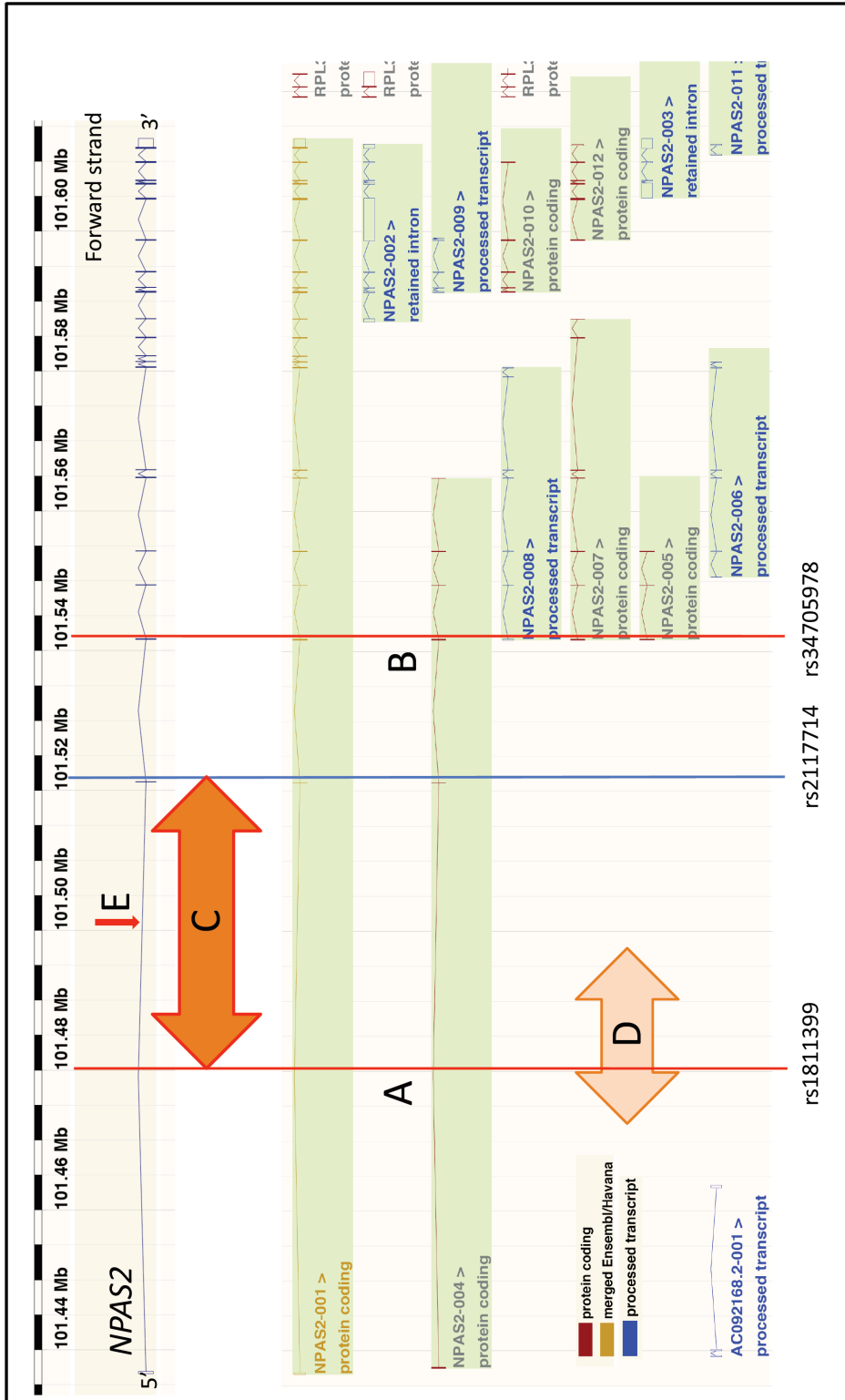
**Table 3-6.** The conserved RREs of *NPAS2*

<b>RORA / NR1D1 binding sites in <i>NPAS2</i></b>	
RRE 1	
Mouse	TTATGC TGACCCACTTT TGCTGT 
Human	TCAAGC TGACCTACTTT TCTTGT score 19.41 INTRON,
RRE 2	
Mouse	AGAGAA TGACCTACTTT ACTGGG 
Human	AGAGAA TGACCTACTTT ACAGGG score 16.8 PROMOTER,
RRE 3	
Mouse	GAAAAA TATGTAGGTCA GTGGAA 
Human	GAAAAA TATGTAGGTCA GTGGAA score 15.08 PROMOTER,
RRE 4	
Mouse	GATCCT TGACCCATTTT CCTGAC 
Human	CATCCT TGACCCATTTT CCGGAC score 13.78 PROMOTER,
Human BLAST with RRE 1 as query	
query	1 TCAAGCTGACCTACTTTTCTTGT 23 
Sbjct:	101504350 TCAAGCTGACCTACTTTTCTTGT 101504372 <i>NPAS2</i>

**Table 3-7.** The conserved RRE in the first intron of NPAS2.

<b>Species</b>	<b>Conserved RRE</b>	<b>Common name</b>
<i>Homo sapiens</i>	GCTGACCTACTTTTCTTGTTTGCCACAGCAGG	Human
<i>Pan troglodytes</i>	GCTGACCTACTTTTCTTGTTTGCCACAGCAGG	Chimpanzee
<i>Gorilla gorilla</i>	GCTGACCTACTTTTCTTGTTTGCCACAGCAGG	Gorilla
<i>Pongo pygmaeus</i>	GCTGACCTACTTTTCTTGTTTGCCACAGCAGG	Orang-utan
<i>Macaca mulatta</i>	GCTGACCTACTTTTCTTGTTTGCCACAGCAGG	Macac
<i>Callithrix jacchus</i>	GCTGACCTACTTTTCTTGTTTGCCACAGCAAG	Marmoset
<i>Tarsius syrichta</i>	GCTGACCTACTTTTCTTGTTTGCCACAGCAGG	Tarsier
<i>Microcebus murinus</i>	GCTGACCTACTTTTCTTGTTTGCCACAGCAGG	Mouse Lemur
<i>Otolemur garnettii</i>	GCTGACCTACTTTTCTTGTTTGCCACAGCAGG	Greater Galago
<i>Tupaia belangeri</i>	-CTGACCTACTTTTCTTGTTTGCCACAGCAGG	Treeshrew
<i>Cavia porcellus</i>	GCTGACCTACTTTT-----	Guinea pig
<i>Dipodomys ordii</i>	GCTGACCTACTTTGCT-GTTTGCCACAGCACA	Kangaroo Rat
<i>Mus musculus</i>	GCTGACCCACTTTTGCTGTGTGCCACAGCAGG	Mouse
<i>Rattus norvegicus</i>	GCTGCCCACTTTTGCTGTGTTCCACAGCAGG	Rat
<i>Ochotona princeps</i>	GCTGACCTACTTTTCTTGTTTGCCCTCAGCAGG	American Pika
<i>Oryctolagus cuniculus</i>	TCTGACCTACTTTCCCTGCTTGCCCTCAGCAGG	European Rabbit
<i>Bos Taurus</i>	GCTGACCTACTTTTCTTGTTTGCCCTGCAGAAGG	Cow
<i>Sus scrofa</i>	GCTGACCTACTTTTGCTGTGTGCCACAGCAAG	Pig
<i>Canis lupus familiaris</i>	GCTGACCTACTTTTCTTGTTTGCCACAGCAGG	Dog
<i>Felis catus</i>	GCTGACCTACTTTTCTTGTTTGCCACAGCAGG	Cat
<i>Equus caballus</i>	GCTGACCTACTT-TCCTGTGTGCCACAGCAGG	Horse
<i>Pteropus vampyrus</i>	GCTGACCTACTTTTCTTGTTTGCCCGCAGCAGG	Fruit Bat
<i>Erinaceus europaeus</i>	GCTGACCTACTTTTCTTGTTTGCTCA-AGCAGG	Hedgehog
<i>Dasypus novemcinctus</i>	ACTGACCTGCTTCCCC-GCCCGCC-ACGTCAGG	Armadillo
<i>Echinops telfairi</i>	GCTGACCTACTTTTCCCTGTTTGC-----AGG	Tenrec hedgehog
<i>Loxodonta Africana</i>	-CTGACCTACTTTCTTGTTTGCCACAGCAGG	African elephant
<i>Procavia capensis</i>	-CTGACCTACTTTGCTGTGTTTGCCACAGCAGG	Rock Hyrax

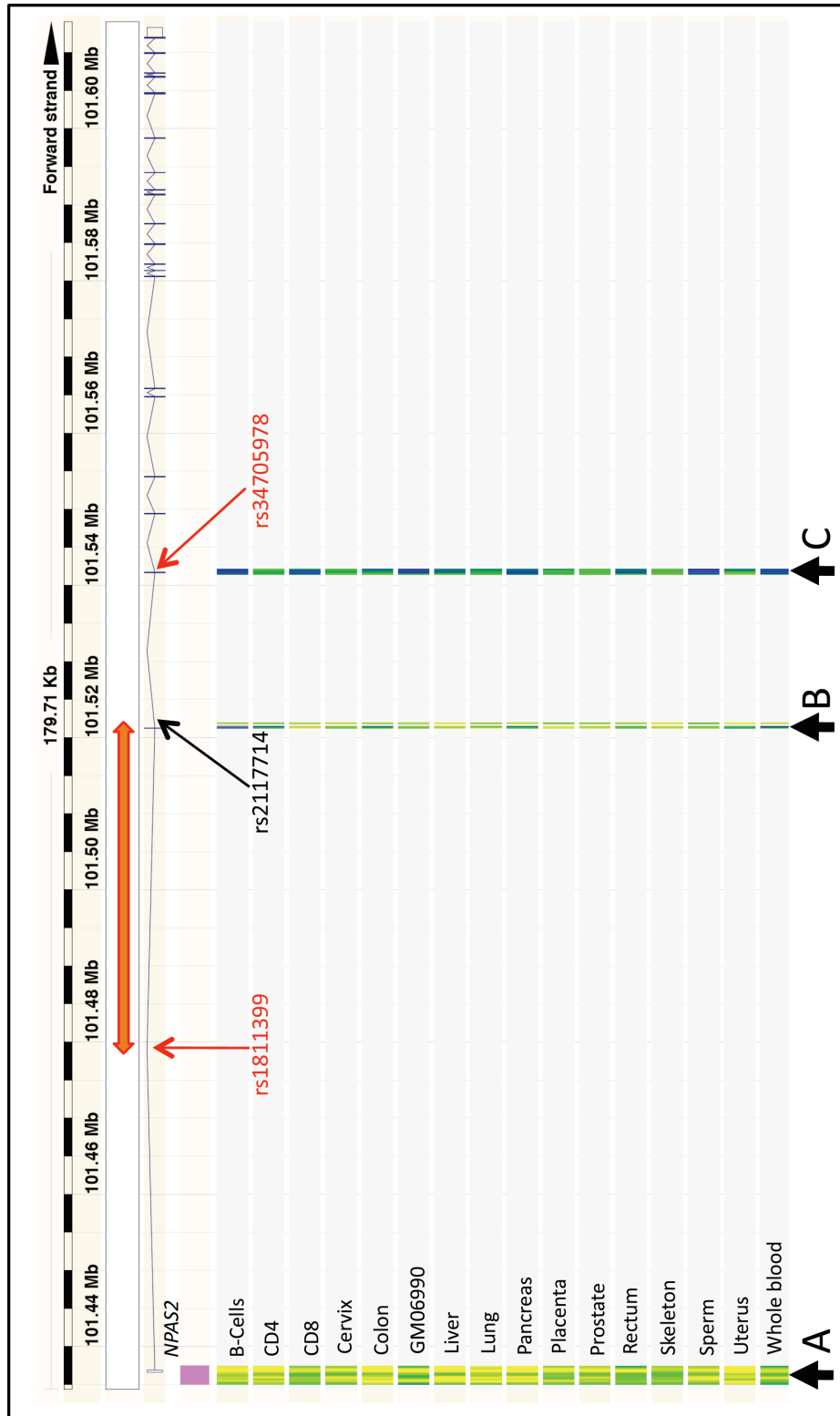
The grey shaded portion shows the RRE of consensus  
[A/T]A[A/T]NT[A/G]GGTCA



**Figure 3-2. Alternative transcripts of NPAS2.**

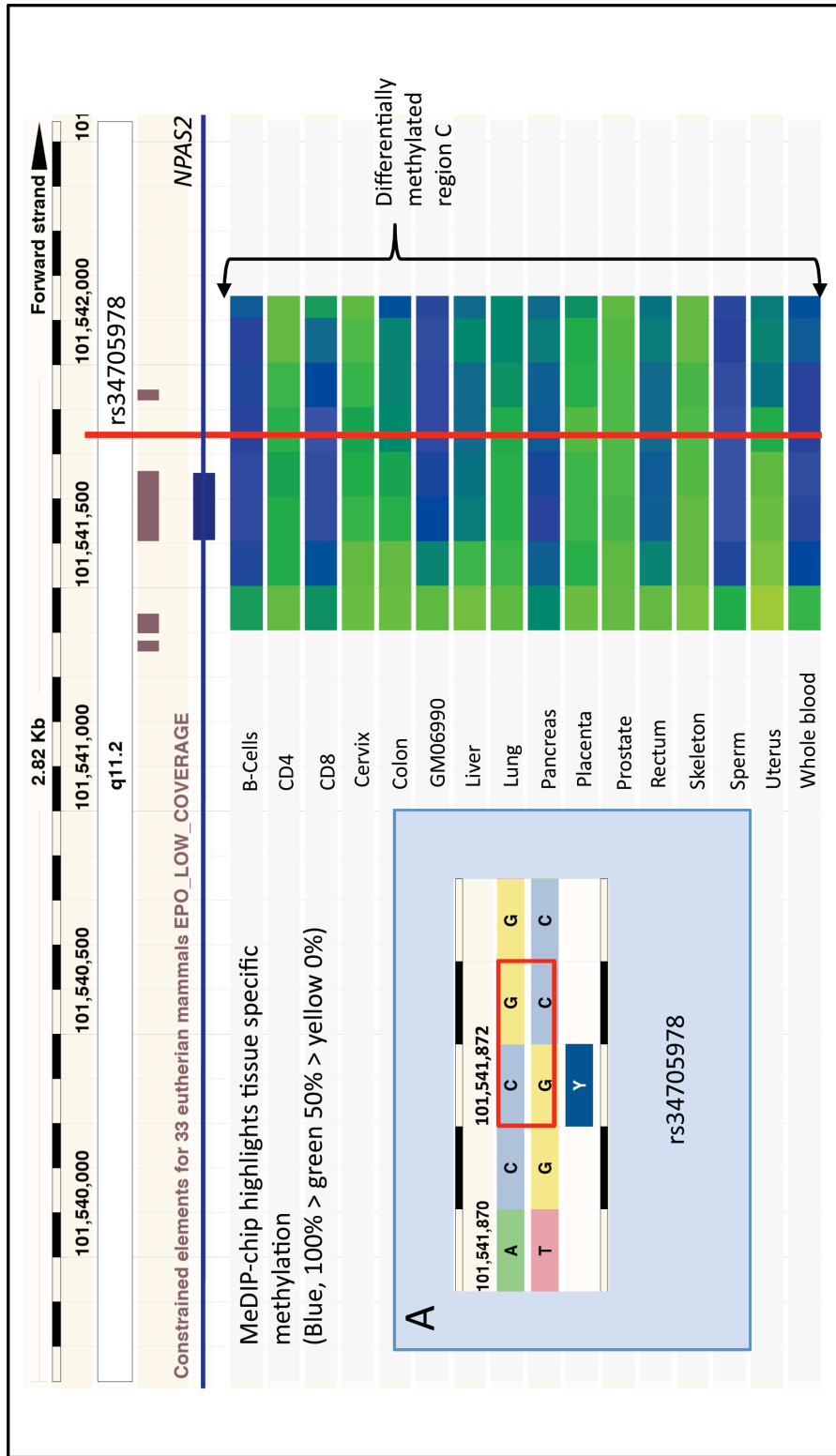
The red vertical lines A and B indicate the positions of autism-associated SNPs rs1811399 and rs34705978 relative to the introns and exons of NPAS2 transcripts. Region C shows the best haplotype and region D contains 8 SNPs in full linkage disequilibrium with rs1811399. The vertical red arrow E shows the position of the conserved RRE in the centre of the best haplotype.





**Figure 3-3. Differentially methylated regions in NPAS2.**

The three differentially methylated regions in NPAS2 are indicated by the bold black arrows A, B and C. The autism associated SNP rs34705978 lies within differentially methylated region C. The best haplotype from the association tests spans a region of intron 1 (shown by the horizontal red arrow) that contains the differentially methylated region B. Coloured shading shows the degree of methylation: deep blue, 100%, green 50% and yellow 0%.



**Figure 3-4. Differentially methylated region C in NPAS2.**

Methylation of CpG dinucleotides within region C is indicated : Deep blue, 100%, green 50% and bright yellow 0%. The different strata show tissue specific methylation patterns. The red vertical bar shows the position of the autism associated SNP rs34705978. Insert **A** highlights the sequence (top line + strand) containing the SNP which is indicated with the ambiguity code Y = (C/T). The red box show that the SNP is a CG dinucleotide and is thus potentially capable of differential methylation. This potential imprint is knocked out by the presence of the autism enriched T allele of this SNP.

**Table 3-8. Cell-type specific transcription factor binding sites in the *NPAS2* haplotype.**

Feature	location	Peak summit
Cjun - K562	2:101508611-101509013	101508776
Cfos - HeLa	2:101510247-101510767	101510499
Cmyc - HeLa Enriched	2:101510231-101510738	101510469
Max - HeLa Enriched	2:101510152-101510805	101510487
Cfos - K562	2:101510312-101510638	101510494
Cjun - K562	2:101510293-101510648	101510484
Cfos - HeLa	2:101519216-101519736	101519503
Max - HeLa	2:101519349-101519780	101519509
Cjun - K562	2:101519355-101519717	101519496

The most significant SNP (rs1811399) from the association study did not coincide with any known regulatory element, or other genomic feature listed in Ensembl (Table 3-8). It was considered possible that this SNP is in linkage with other near-by SNPs that are themselves functional in some way. For example perhaps this SNP is in linkage with common variants in or near RRE1. Understandably the RRE was not found to contain SNPs but common variants exist that flank this conserved element e.g. rs72627425 at 2:101504425.

In order to explore possible linkage of rs1811399 with other functional SNPs in the region, GLIDERS was used to search for SNPs in linkage disequilibrium with rs1811399. The results for this search show that a number of SNPs in the vicinity of rs1811399 are in strong linkage disequilibrium with this SNP (Table 3-9). However, the SNP in strongest linkage with rs1811399 was not found to coincide with any features likely to have functional consequence nor did the linkage disequilibrium region cover the RRE. The SNPs rs12472319 and rs12472321 (that are separated by 11 nucleotides), both are in linkage disequilibrium with rs1811399 (Table 3-9) and both lie in a DNase1 enriched site (HepG2 cells) and a CTCF binding site

**Table 3-9.** Results from GLIDERS database search showing SNPs in linkage disequilibrium to rs1811399.

SNP	Position	Feature	M.A.F.	Dist. to next marker	r <sup>2</sup>	D'	ChiSq	p-val	P value (BF.Corrected)
rs1561002	101465271	none	0.250	13kb	0.58	0.83	100.35	1.28e-23	8.98e-12
rs10206927	101467887	transposon	0.248	11kb	0.58	0.83	98.79	2.81e-23	1.97e-11
rs7582455	101471707	none	0.192	7kb	0.85	1.00	165.60	6.76e-38	4.74e-26
rs983287	101480401	none	0.188	1kb	0.82	1.00	152.91	4.01e-35	2.81e-23
rs2043534	101480885	none	0.259	1kb	0.80	1.00	168.58	1.51e-38	1.06e-26
rs930309	101484534	none	0.225	5kb	0.97	1.00	217.02	4.04e-49	2.83e-37
rs6542996	101485827	none	0.188	6kb	0.82	1.00	152.91	4.01e-35	2.81e-23
rs12472319	101487457	CTCF	0.259	8kb	0.80	1.00	168.58	1.51e-38	1.06e-26
rs12472321	101487468	CTCF	0.261	8kb	0.80	1.00	166.95	3.43e-38	2.41e-26
rs2164319	101496024	transposon	0.085	17kb	0.33	1.00	44.22	2.93e-11	1
rs13012765	101498174	DNase1 enriched	0.246	19kb	0.72	0.92	136.95	1.24e-31	8.70e-20
rs1369481	101511959	none	0.277	32kb	0.42	0.76	66.43	3.63e-16	0.00025

SNPs in green are located 5' of rs1811399. SNPs in black lettering are 3' of rs1811399

### 3.3.3iii PER1 SNPs individually (rs885747 and rs6416892)

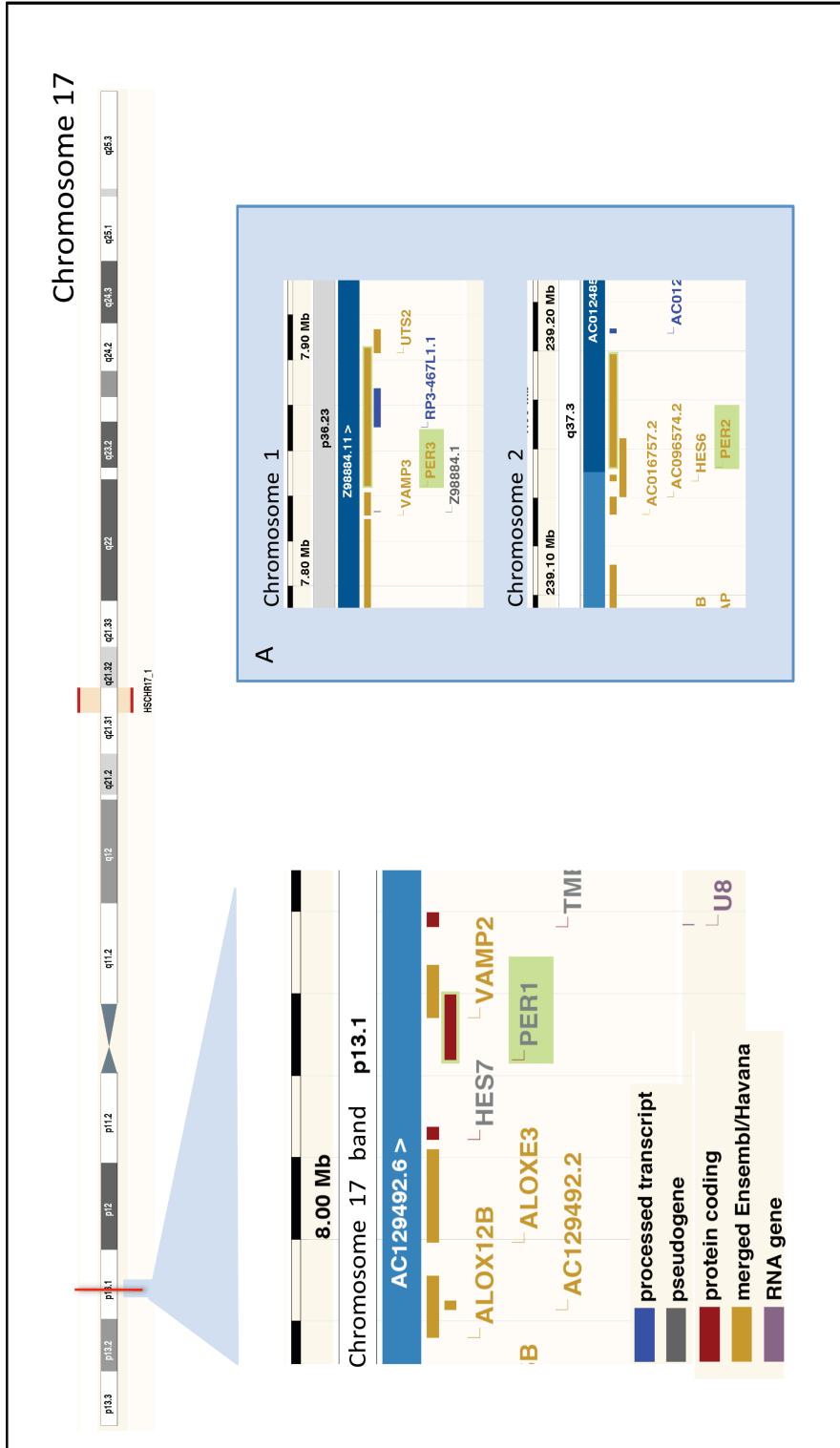
Using Ensembl as described, *PER1* is shown to be located on the short arm of chromosome 17 towards the telomere region (p13.1). The gene is flanked immediately upstream by the vesicle-associated membrane protein 2 gene, *VAMP2* and on the downstream side by *HES7* an element of the NOTCH signalling pathway and component gene of the somite segmentation clock (Figure 3-5). It is perhaps noteworthy that *HES7* is an essential gene in the ultradian segmentation clock pathway that regulates the ~2h waves of cell division that define the developing somites during embryogenesis (Kageyama et al. 2007).

The autism-associated SNP rs885747 lies approximately equi-distant from the 3' and 5' ends of *PER1* in Intron 12 at genome coordinate 17:8050737. This SNP has four synonyms (Table 3-10).

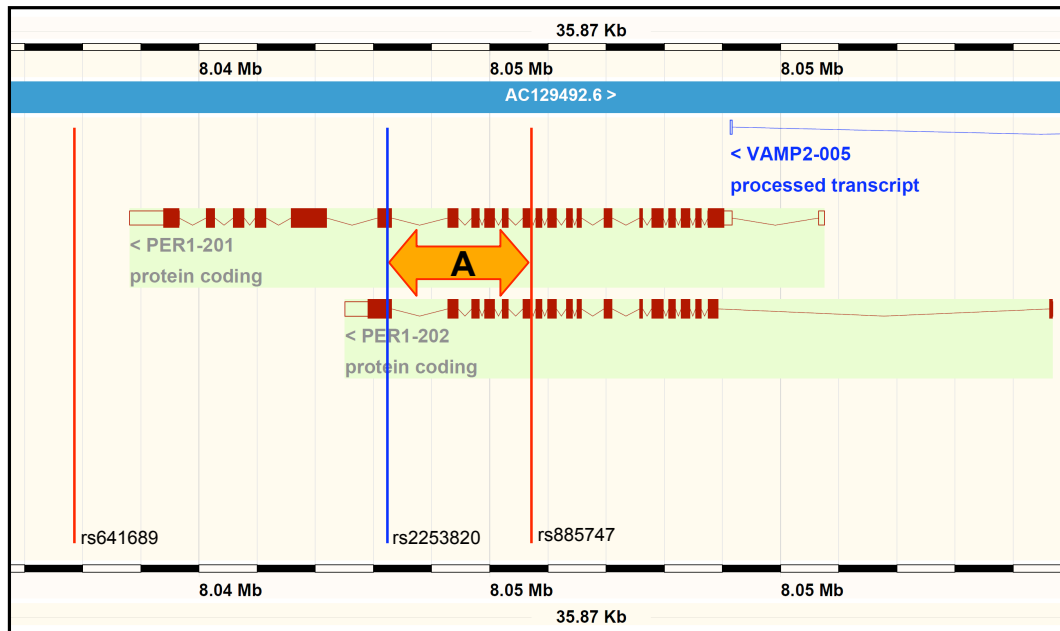
**Table 3-10. Synonyms of rs885747.**

TSC: TSC0109140
ENSEMBL Celera: ENSSNP1385387
ENSEMBL Venter: ENSSNP1385387
HGVbase: SNP000957527

As is the case for *NPAS2*, alternative transcripts are derived from the *PER1* locus. *PER1* is transcribed from the reverse strand with 2 alternative transcripts *PER1-201*, *PER1-202*. The SNP rs885747 is indicated in an intron of each of these transcripts. The SNP is not located within a regulatory region but its position close to an intron/exon boundary suggests it may have an influence on splicing of intron 12.



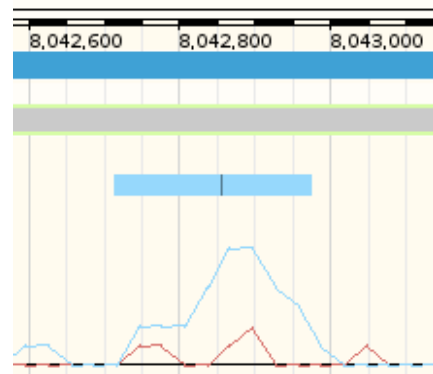
**Figure 3-5. Banding pattern of human chromosome 17.** The banding pattern of chromosome 17 with the location of *PER1* indicated by the red vertical bar towards the telomere on the short arm of the chromosome. The magnified region indicated by the pale blue triangular segment shows *PER1* in the context of immediately adjacent genes *HES7* and *VAMP2*. Insert A, the pale blue box, shows the locations of the paralogues *PER2* and *PER3*.



**Figure 3-6. Alternative transcripts of *PER1*.**

The red vertical lines show the positions of the autism-associated SNPs. The horizontal orange arrow marks the extent of the best haplotype. The blue vertical line is the SNP rs2253820 that together with rs885747 makes up the best haplotype.

The other autism associated SNP in *PER1*, rs6416892, has one synonym, ENSEMBL: Venter ENSSNP1385372. The SNP is located down-stream of *PER1* at 17:8042860 within the regulatory feature ENSR00000573321, core location 17:8041371-8043207. Detailed analysis of this regulatory region shows that it contains the following cell type specific binding sites validated in a number of different cell lines: FOS, MAX, CTCF, SREBF2, MYC, JUN, DNASE1. The SNP rs6416892 is located 5 nucleotides from the binding site peak of sterol regulatory element binding transcription factor 2 (SREBF2) in HepG2 cells (Figure 3-7). This SNP is also close to a CpG island at 8041980-8042415 and very close to a tRNA (Thr) gene that is located at 8042770-8042843.



**Figure 3-7. The autism-associated SNP rs6416892 is located in a SREBF2 binding-peak.**

The grey bar with green highlight shows a portion of regulatory region ENSR00000573321 containing FOS, MAX, CTCF, SREBF2, MYC, JUN, DNASE1 binding sites. The light blue bar shows the extent of the SREBF2 binding region. The SREBF2 binding-peak is shown by the vertical grey line within the light blue bar. The peak is 5nt 3' of the SNP.

#### 3.3.3iv *PER1*, most significant haplotype.

Haplotype analysis (Dudbridge 2003) within *PER1* gave a single significant result: a global  $P = 0.027$  for the markers rs885747-rs2253820 located at 17:8050773-17:8048169. Using Ensembl as described above, this region was shown to be devoid of any notable control elements but it does contain a number of essential splice site SNPs and other variations that could possibly cause functional changes in the *PER1* protein. It is possible these sSNPs are in linkage disequilibrium with rs885747. The SNPs that are essential splice sites and frame shift SNPs within the best haplotype of *PER1* (Table 3-11) may have a functional role in autism by



influencing which transcript variants for this gene reside in a given genotype.

**Table 3-11. Five common variants within the region of *PER1* delineated by the best autism-associated haplotype.**

SNP code	Location	Alleles	Type of SNP
GA032060	8048313	T/C	Essential splice site
GA013944	8049274	A/G	Essential splice site
rs35505654	8049703 & 8049704	-/C	Frame shift Coding
rs3027184	8050017	G/A	Non-synonymous coding
rs3027183	8050325	G/A	Splice Site

**Table 3-12. Summary of the features coincident with and delineated by the *NPAS2* and *PER1* autism-associated SNPs and haplotypes respectively.**

Gene	SNPs	Coincident features
<i>NPAS2</i>	rs1811399	None
	Haplotype rs1811399 - rs2117714	Conserved RRE (NR1D1/RORA) binding site. Cell-line specific TFBS and differentially methylated region.
	rs34705978	Polymorphic CpG within differentially methylated region
<i>PER1</i>	rs885747	Full linkage disequilibrium with rs6416892 in Ensembl populations. Possible splicing variant
	Haplotype rs885747- rs2253820	Haplotype contains three splice sites (one essential splice site) and two coding variants
	rs6416892	SNP within regulatory region at binding peak for SREBP2

### 3.3.4 Conclusion.

This analysis of the autism-associated SNPs in the clock genes *NPAS2* and *PER1*, made subsequent to the publication of the association study (Nicholas et al. 2007), updates the genomic coordinates for these SNPs. In doing so, the SNPs are now seen in the context of regulatory features

etc. (Table 3-12) that were unreported in the public databases at the time of publication.

This new analysis locates two of the significant SNPs, rs34705978 (*NPAS2*) and rs6416892 (*PER1*) within regulatory features. For rs34705978 in *NPAS2* the SNP (C/T with T enriched in the autism) lies in a differentially methylated control region. This SNP, when the C allele is present, forms a CpG pair (Figure 3-4) and thus the autism allele represents a knock out of a potential methylation site in this control element. This CpG pair is one of two pairs in the region that shows differential methylation in different tissues. Association of this SNP with autism suggests that this SNP may be a functional variant that could affect the tissue specific regulation of *NPAS2* expression. The importance of this regulatory feature may be reflected in the rarity of the T|T haplotype in human populations (Ensembl, Table 3-1).

For rs6416892 in *PER1*, Ensembl curated evidence (ENSR00000573321) from cell lines shows the SNP to be within a control region spanning bp 17:8041371-8043207. The SNP rs6416892 is located 5 nucleotides from the binding site peak of sterol regulatory element binding transcription factor 2 (SREBF2).

SREBF2 plays a major role in lipogenesis and cholesterol metabolism (Shimano et al. 1999) and is functionally implicated in a mouse model of Alzheimer's disease where high mitochondrial cholesterol levels (produced by over expressing SREBF2) derive amyloid-beta dependent neurotoxicity (Fernandez et al. 2009). This anti-psychotic-medication activated regulator of cholesterol biosynthesis is also implicated in the etiology of schizophrenia (Le Hellard et al.) and may play a role in autism via

SREBF2 dependent regulation of *PER1*. Abnormalities in cholesterol metabolism are reported in autism (Tierney et al. 2006) which is highly comorbid with the cholesterol metabolism disorder Smith-Lemli-Opitz syndrome where ~50% of cases meet ADI-R criteria for autistic disorder (Tierney et al. 2001; Bukelis et al. 2007).

SREBF2 functions in other pathways that might have specific relevance to autism. For example, this transcription factor regulates cell-type-specific genes in male germ cells where, like *PER1* it is highly expressed in a stage specific manner (Wang et al. 2002; Alvarez et al. 2003; Morse et al. 2003; Wang et al. 2004).

The most significant *PER1* haplotype in this study covers a relatively small section located centrally in *PER1* that contains splice site SNPs. (Two essential splice sites that were not genotyped in the association study). Experiments in *Drosophila* indicate that alternative splicing of *per* derives different circadian periods in the fly (Cheng et al. 1998). It is therefore noteworthy that these splice site SNPs in the most significant *PER1* haplotype suggest alternative splicing that could possibly affect circadian rhythms in humans.

Even though the SNP rs1811399 had the lowest p value of the four significant SNPs of the association study this SNP did not co-locate with any regulatory feature. The results of the GLIDERS analysis show strong linkage disequilibrium between this SNP and SNPs immediately downstream but this strong linkage disequilibrium region ended before the RRE and its peak did not coincide with any other regulatory elements.

Overall, this analysis of the clock gene autism-associated SNPs strengthens the notion that these are functional SNPs that may impact

tissue specific expression of *NPAS2* (rs34705978), cholesterol metabolism (rs6416892) and alternative splicing (rs885747). However, reasons for the association of rs1811399 with autistic disorder remain unclear at this point, the SNP appearing to have significance of itself but without being implicated in any regulatory region or feature.

### **3.4 Do the autism-associated SNPs in *NPAS2* and *PER1* coincide with transcription factor binding site motifs? TESS analysis of rs1811399, rs34705978 rs885747 and rs6416892.**

#### 3.4.1 Introduction.

Certain intronic motifs can act as binding sites for transcription factors that regulate the expression levels of the host gene (Tokuhiro et al. 2003). Thus it is possible that the autism-associated SNPs in the clock genes *NPAS2* and *PER1* affect the binding of transcription factors that regulate the temporal, tissue specificity or levels of clock gene expression. To begin testing whether the autism-associated SNPs in the clock genes *PER1* and *NPAS2* could function as transcription factor binding sites, a bioinformatics screen of the SNP regions was carried out using the internet-based Transcription Element Search System (TESS, Table 3-1 and described in Current Protocols in Bioinformatics ISBN: 978-0-471-25093-7). TESS identifies binding sites by using site or consensus strings and positional weight matrices from the TRANSFAC, JASPAR, IMD, and University of Pennsylvania's CBIL-GibbsMat database.

### 3.4.2 Method

TESS was interrogated with query sequences of 10 nucleotides flanking each side of the autism-associated SNPs. To test the system, the *EN2* SNPs rs1861972 and rs1861973 were entered into the TESS search. As expected, the search gave similar results to those published (Benayed et al. 2009). Using the autism-associated SNPs in *NPAS2* and *PER1* as query sequences a number of possible binding sites were found to overlap with the location of some of these SNPs. The blue or red bars indicated a transcription factor binding site on the same DNA strand as the gene or on the complimentary DNA strand respectively, described as forward or reverse with respect to the query sequence. The binding sites on the same strand as the gene were recorded (Table 3-13).

### 3.4.3 Results.

One transcription factor (TF), binding motif is indicated to overlap the SNP location rs1811399 GT-IIBa. This motif was first recognized as an enhancer element for SV40 viral replication (Xiao et al. 1987). However, this site is indicated for the reverse strand. The *NPAS2* gene is encoded on the forward strand and thus no score is given for the A allele of rs1811399 for overlap with a known transcription site according to TESS. One TF binding motif is indicated to overlap rs1811399 when the autism-enriched C allele is present. This NF-1 site is indicated for the forward strand, on which *NPAS2* is encoded. This motif is shown to play a role in the cell specific expression pattern of the neural cell adhesion molecule (NCAM) that is targeted by Nuclear Factor-1 type enhancer proteins (Hirsch et al. 1990). However the low score, as the full consensus is

TTGGC(N)<sub>5</sub>GCCAA (Gronostajski 2000), suggests that this is not a strong signal. For the *NPAS2* SNP rs34705978, TESS does not predict TF binding site overlap with the SNP's location for either the C or the autism-enriched T allele.

**Table 3-13. TESS transcription factor binding site predictions.**

Predicted sites overlapping the four autism-associated clock gene SNPS

rs1811399, rs34705978, rs885747 and rs6416892

Gene	SNPs	TFBs
NPAS2	<i>NPAS2</i> SNP rs1811399 = <b>A</b> CATGGTGATA <b>A</b> CAGCGGCTGC None	None
	<i>NPAS2</i> SNP rs1811399 = <b>C</b> CATGGTGAT <b>C</b> CAGCGGCTGC == == (8.00) NF-1	NF-1 (weak)
	<i>NPAS2</i> SNP rs34705978 = <b>C</b> CAGCCTTACC <b>C</b> GGTCGAGGGC None	None
	<i>NPAS2</i> SNP rs34705978 = <b>T</b> CAGCCTTAC <b>T</b> GGTCGAGGGC None	None
PER1	<i>PER1</i> SNP rs885747 = <b>G</b> GTGGGAAGC <b>G</b> GGGTCAAGCC = === (8.00) GAL4 = ===== (14.00) RXR- $\alpha$ / $\beta$	GAL4  RXR- $\alpha$ , RXR- $\beta$ (half site)
	<i>PER1</i> SNP rs885747 = <b>C</b> GTGGGAAGC <b>C</b> GGGTCAAGCC None	none
	<i>PER1</i> SNP rs6416892 = <b>A</b> ATCTTGAGGA <b>A</b> ACACCTGTTT ===== (10.00) GAL4	GAL4
	<i>PER1</i> SNP rs6416892 = <b>C</b> ATCTTGAGG <b>C</b> ACACCTGTTT ===== (8.7167) T-Ag	T-Ag

In *PER1* The SNP rs885747 (G allele) gave a GAL4 site and a RXR-alpha, RXR-beta site however these sites were knocked out by the presence of the C allele. For SNP rs6416892 the A allele gave a GAL4 site that was replaced in the same location by a T-Ag site by the autism enriched C allele. These results together with details of the sequences and positions of the binding motifs are shown in Table 3-13.

#### 3.4.4 Conclusion

The TESS analysis shows that for rs1811399 there is a weak match with an NF-1 binding site when the autism-enriched C allele is present. The poor score for this site suggests this is not a major feature. Nevertheless, an intronic Nuclear Factor-1 (NF-1) binding sites is involved in governing the cell-type-specific expression of the neural cell adhesion molecule gene *NCAM* (Schneegans et al. 2009) and NF-1 sites in the first intron of the cystic fibrosis gene regulate its cell-type-specific expression (Ott et al. 2009).

For rs885747 in *PER1*, with the G allele TFBS for GAL4 and RXR- $\alpha$ /RXR- $\beta$  are found. However, when the autism enriched C allele is present, these are lost. Retinoid X receptors are required for normal embryonic development (Mark et al. 2009) and circadian rhythms (McNamara et al. 2001). These ligand-inducible transcription factors (RXRs) bind to the hexameric DNA sequence AGGTCA. When this sequence occurs in promoter regions as a direct repeat with a one-base pair spacer, the motif facilitates binding of retinoid X receptor homodimers and the sequence GGGGTCA is the highest affinity retinoid X receptor binding site known. However, the sequence GGGGTCA does not occur as a repeat at this point in the *PER1* sequence and a half-site does not support 9-*cis*-RA induced RXR mediated gene expression (Subauste et al. 1994). RXR alpha is understood to undertake a hormone-dependent interaction with CLOCK and NPAS2 (McNamara et al. 2001) so an interaction of RXRs with another element of the circadian clock (*PER1*)

would at least be in keeping with these receptors' known function in circadian rhythms (McNamara et al. 2001).

For rs6416892, an overlapping TFBS is predicted for both alleles but with the GAL4 site predicted for the A allele being replaced by a T-Ag site when the autism enriched C allele is present. The Ensembl analysis given previously for this SNP reinforces the indication that this SNP might have functional implications in the regulation of *PER1*. It is however, not possible to draw any firm conclusions from the above analysis without further recourse to experiments such as foot-printing, gel mobility-shift assays and site directed mutagenesis to test the validity of these predicted binding sites.

## **4 Investigating possible novel functionality of the autism-associated clock gene SNPs with web based bioinformatics modeling tools**

### **4.1 Introduction**

Web based bioinformatics tools complement the large amount of data stored in web-server linked databases. These bioinformatics tools calculate bespoke solutions to specific queries such as alignments between sequences, secondary structural predictions etc. and can use algorithms that balance a number of conflicting variables e.g. those associated with complex dynamic processes such as intron splicing and the folding of a linear RNA molecule into a double-stranded helix. The following chapter describes the use of such tools to investigate possible



novel functional effects of the autism-associated clock gene SNPs and related questions. Particular focus is given to the SNPs rs885747 in *PER1* and rs1811399 in *NPAS2* as these SNPs did not show co-location with any regulatory feature or functional element highlighted by the database searches described above.

## **4.2 Could the autism-associated SNP rs885747 affect splicing of *PER1*?**

### 4.2.1 Introduction.

Serious human diseases e.g. Phenylketoneurea can be caused by intronic mutation that affect pre-mRNA splicing (Dworniczak et al. 1991). Intron splicing requires three elements within the intron: a 5' donor site signal, the branch point and the 3' polypyrimidine acceptor sequence signal. These features of themselves are insufficient to support accurate and efficient splicing of all exons. Exonic splicing enhancers, intronic splicing enhancers, exonic splicing silencers, and intronic splicing silencers (that are located close to the exon/intron boundaries) fine tune splice site signal recognition by the spliceosome (Mount 2000; Lim and Burge 2001).

These enhancer and silencer elements are particularly important when short exons (such as those in the central region of *PER1*) have the effect of weakening the splicing signals, or when the splicing signal motifs themselves are weak. The proximity of rs885747 to the intron-12 exon-13 boundary in *PER1* suggested that this SNP might influence splicing and this idea was tested using the bioinformatics tool SpliceScan2 (Churbanov et al. 2010).

The internet based bioinformatics tool SpliceScan2 (Table 3-1) scores splice sites based on an exon definition model. SpliceScan2 simultaneously scores exon length, acceptor and donor signals and the contribution of exonic/intronic enhancer/silencer elements, as it is the balance of all these factors that determines the pattern of splicing. The scores are in terms of the LOD (Log of the odds) (logarithm base 2) of a true splice site versus a decoy splice site. Selectable outputs from the SpliceScan2 server gives tables of LOD scores as well as a graphical display showing the position of introns exons and splice site enhancers and suppressors.

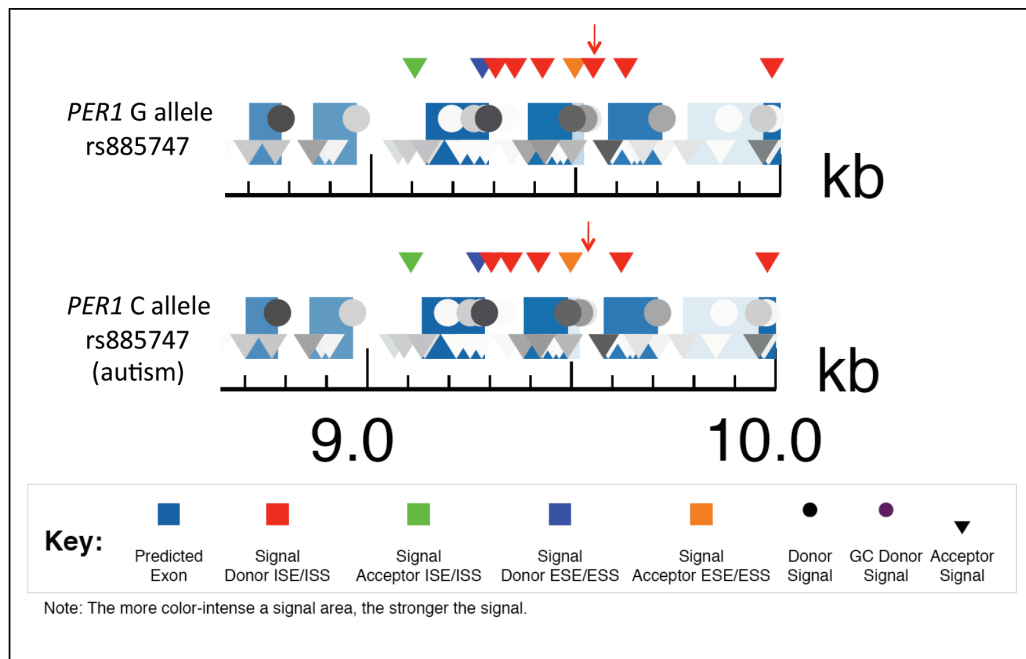
#### 4.2.2 Method.

Two copies of the nucleotide sequence of *PER1* were entered into SpliceScan2 in turn. The files were identical except that at the genome coordinate corresponding to rs885747 one sequence file contained rs885747 C allele while the second sequence file contained rs885747 G allele. The output, in the form of LOD scores was also noted. The graphical output was given as a portable document format file.

#### 4.2.3 Results.

The portable document format file was scrutinised for differences overall and in the region where the programme predicted the 12<sup>th</sup> intron exon boundary. For the input file containing the autism-associated C allele no 3' Intronic enhancer/silencer element was noted. However, the presence of an intronic enhancer/silencer element was noted for the file containing the G allele (Figure 4-1). The scores for this exon were also

recorded and scrutinised. The output files were identical apart from the differences associated with rs885747.



**Figure 4-1. Splice Scan2 alternative splicing: *PER1*.**

Splice Scan2 was used to test for effects of the SNP rs885747 on splicing of intron 12 in *PER1*. The software predicted the loss of an enhancer/suppressor signal for the autism-associated C allele of rs885747 as indicated by the red triangle missing in the C allele test (thin red arrow).

#### 4.2.4 Conclusion.

The autism-associated SNP is indicated to knockout an intronic splicing silencer/enhancer element and similar mutations are known which cause serious diseases e.g. Phenylketoneurea (Dworniczak et al. 1991). This intronic splicing silencer/enhancer element may have a phenotypic effect as two other reports find association of rs885747 with prostate cancer risk; the degree of disease aggressiveness in prostate cancer (Zhu

et al. 2009) and depression and early morning awakening, specifically in men (Utge et al. 2010). Loss of the splice site enhancer element by SNP rs885747 lowers the definition of the intron 12 boundaries (Figure 4-1) possibly introducing splicing errors into the mRNA in this region. Concomittant alterations in the protein sequence might alter the protein's reported interactivity with the AR (androgen receptor) compromising PER function in male typical developmental pathways (Cao et al. 2009). SNP rs885747 is therefore tentatively suggested to have some phenotypic impact through an influence on splicing and that this effect is sexually dimorphic.

#### **4.3 Structural analysis of RNA transcripts containing autism-associated SNPs in *NPAS2*, *PER1*, *EN2* and the psychosis susceptibility variant rs1344706 in *ZNF804A***

The loss of function of single genes can lead to specific disorders e.g. mutation of the gene for phenylalanine hydroxylase, *PAH*, causes Phenylketoneurea (Dilella et al. 1987) and loss of function of *FMR1* the fragile X mental retardation 1 gene derives Fragile-X Syndrome (Yu et al. 1991). However, the strongly genetic neuropsychiatric disorders of autism and schizophrenia have not been found to be the result of single gene mutation despite a great deal of genetic and molecular genetic investigation of these disorders. Moreover, current evidence suggests that, to a large degree, the heritable element of these disorders is contributed by a constellation of variants of small effect with perhaps epistatic effects magnifying the importance of some of these "small effect" genes. This may

be the case for the autism-associated variants in the epistatic clock genes *PER1* and *NPAS2* for example.

Where replication of results has confirmed that a particular risk variant is reliably linked to psychiatric disorder, (for example: the rs1861972-rs1861973 haplotype of *EN2* in autism; the *ZNF804A* SNP rs1344706 in schizophrenia; and, SNP rs1006737 of *CACNA1C* in bipolar disorder) paradoxically, these variants appeared at first to be non-functional polymorphisms in the introns of these genes (Benayed et al. 2005; O'Donovan et al. 2008; Ferreira et al. 2008). Despite detailed re-examination of *EN2* and *ZNF804A* no other SNPs were found to be more strongly associated with the phenotypes under investigation than the intronic SNPs and thus the risk alleles remain as intronic SNPs of currently unknown function (O'Donovan et al. 2008; Benayed et al. 2009).

This author suggests that certain autism-associated SNPs, appearing as non-functional polymorphisms in introns, may alter the secondary structure and function of the mRNA transcripts expressed from these introns (Nature Preceedings hdl:10101/npre.2008.2366.1.). For example, the intronic transcripts embedding rs18111399 in *NPAS2* and rs1861973 in *EN2* are predicted to take up allele dependent hairpin formations that resemble microRNA precursors and which may be processed by the microRNA biogenesis pathway into gene-silencing small RNAs.

However, Benayed et al. (2009) provide evidence for an effect of the *EN2* SNPs rs1861972 and rs1861973 on the levels of expression of *EN2* that is due the binding of transcription factors to the *EN2* intron at these SNP locations. These two possibilities are not necessarily mutually exclusive, as multiple predicted TFBS can coincide with the DNA

sequences corresponding to verified microRNA hairpins, as screening *bona fide* microRNA hairpins with TESS/Ensembl/MyHits tools can demonstrate.

Introns in numerous protein coding genes harbor microRNA genes that are transcribed along with the host gene and microRNAs are implicated in the development of Fragile-X syndrome Tourette's syndrome schizophrenia and autism (Abelson et al. 2005; Lin et al. 2006a; Kim and Kim 2007; Perkins et al. 2007; Abu-Elneel et al. 2008; Sarachana et al. 2010). This investigation considered whether the transcripts derived from: intron regions containing the autism-associated SNPs rs1811399 and rs34705978 in *NPAS2*; rs885747 rs6416892 in *PER1* and rs1861972 and rs1861973 in *EN2*, show characteristics typical of microRNA genes. The SNPs in *EN2* were included as they have perhaps the strongest case of any SNP for association with autism, the function of these SNPs are unresolved and *EN2* is listed amongst genes with conserved-D-Box elements suggesting *EN2* might be regulated by the circadian clock (Kumaki et al. 2008).

To justify searching for microRNA genes that, it might be argued, should have come to light in previous gene discovery studies, it is necessary to consider the primary bioinformatics methods use to discover microRNAs. Bioinformatics discovery of microRNA has followed the principles of sequence conservation of functional elements within related genomes, thus novel microRNAs have been discovered in a given genome, human say, by comparison with homologous sequences from other well-defined genomes e.g. mouse. However, non-conserved microRNAs are known that appear to be species specific or at least occur

only in closely related species hsa-mir-1273 for example in *Pan troglodytes* (chimpanzee) and *Homo sapiens*.

Laboratory based microRNA discovery has until very recently incurred a strong bias in favour of the discovery of constitutively and/or highly expressed microRNAs. Thus microRNAs that are neither highly expressed, expressed only at specific developmental stages, expressed only in particular tissues, or intermittently and are not evolutionarily conserved, could evade detection by routine laboratory based microRNA screens.

#### **4.4 Internet publication: Autism-associated SNPs in the clock genes *NPAS2*, *PER1* and the homeobox gene *EN2* alter DNA sequences that show characteristics of microRNA genes.**

(Nature Preceedings: <http://precedings.nature.com/documents/2366/>)

B. Nicholas, M. J. Owen, D. C. Wimpory and T. Caspari.

##### 4.4.1 Contributions to this publication.

This author (B. Nicholas) performed all the bioinformatic analysis and wrote the manuscript. M. J. Owen, D. C. Wimpory and T. Caspari contributed critical intellectual input in the preparation of the manuscript.

##### 4.4.2 Abstract.

Intronic single nucleotide polymorphisms (SNPs) in the clock genes *NPAS2* and *PER1* and the homeobox gene *EN2* are reported to be

associated with autism. This bioinformatic analysis of the intronic regions that contain the autism-associated SNPs rs1861972 and rs1861973 in *EN2*, rs1811399 in *NPAS2*, and rs885747 in *PER1*, shows that these regions encode RNA transcripts with predicted structural characteristics of microRNAs. These microRNA-like structures are disrupted *in silico* by the presence of the autism-enriched alleles of rs1861972, rs1861973, rs1811399 and rs885747 specifically, as compared with the minor alleles of these SNPs. The predicted gene targets of these microRNA-like structures include genes reported to be implicated in autism (*GABRB3*, *SHANK3*) and genes causative of diseases co-morbid with autism (*MECP2* and *RAI1*). The inheritance of the AC haplotype of rs1861972 - rs1861973 in *EN2*, the C allele of rs1811399 in *NPAS2*, and the C allele of rs1234747 in *PER1* may contribute to the causes of autism by affecting microRNA genes that are co-expressed along with the homeobox gene *EN2* and the circadian genes *NPAS2* and *PER1*.

#### 4.4.3 Introduction.

A number of genes are considered to contribute to the heritability of the neurodevelopmental disorder autism (APA 1994; Pickles et al. 1995). Numerous candidate gene studies have been deployed (reviewed by Yang and Gill, (2007) for example) to bring further detail to epidemiological studies of autism.

The choice of candidate genes has been guided by: the results of whole genome screens for autism (for example, Palferman et al. (2001)); cytogenetic studies of affected individuals (for example, Wolff et al.



(2002)); and, genetic studies of disorders that show co-morbidity with autism (reviewed, for example, by Zafeiriou, (2007)). Overall, these candidate gene studies have not yet found any protein coding sequence changes that are significantly associated with autistic disorder. Single nucleotide polymorphisms (SNPs) in the introns of genes have however, been reported to show significant association with autistic disorder and some of these results have withstood replication. In *EN2*, a gene involved in cerebellum development (Sgaier et al. 2007), the association of the intronic SNPs rs1861972 and rs1861973 with autistic disorder has been re-investigated a number of times with overall positive but complex results that support evidence for abnormal cerebellar development in autism (Kemper and Bauman 2002; Benayed et al. 2005; Brune et al. 2008; Wang et al. 2008).

The reported association of *PER1* and *NPAS2* with autistic disorder (Nicholas et al. 2007) is noteworthy given that specific sleep anomalies (Limoges et al. 2005) altered circadian rhythm (Corbett et al. 2006) and increased measures of oxidative stress are reported to be associated with autism (James et al. 2004).

*PER1* and *NPAS2* regulate sleep in mammals (Sun et al. 1997; Franken et al. 2006) and *NPAS2* effects redox signaling (Rutter et al. 2001; Dioum et al. 2002). The protein products of these genes interact (Reick et al. 2001) perhaps magnifying the effect of minor variations in these genes in autism. Altered expression levels of circadian genes have also been linked to bipolar disorder (Yang et al. 2009) and schizophrenia (Aston et al. 2004), both indicated to share genetic overlap with autism (Rzhetsky et al. 2007).

The autism-associated SNPs in *PER1*, *NPAS2* and *EN2* are however, all intronic. Parsimoniously they are considered to indicate that other functional changes may be in linkage disequilibrium with these autism-associated SNPs. But despite further investigation by the teams reporting association of rs1861972 (*EN2*), rs1861973 (*EN2*), rs1811399 (*NPAS2*), and rs885747 (*PER1*) with autism, no other autism linked functional changes in these genes were established (Benayed et al. 2005; Nicholas et al. 2007).

Recent findings in the field of small RNAs show that the introns of protein coding genes may harbor microRNA genes that encode small RNAs capable of regulating the function of many other genes at chromosomal locations remote from that of the micro-RNA gene itself (Lin et al. 2006b). MicroRNA genes within the introns of protein coding genes are transcribed concurrently along with the host gene and the genes *EN2*, *NPAS2* and *PER1* are all expressed in brain regions relevant to autism (Millen et al. 1994; Sun et al. 1997; Reick et al. 2001).

In this study we sought to investigate whether the intron regions containing the autism-associated SNPs: rs1861972 (*EN2*); rs1861973 (*EN2*); rs1811399 (*NPAS2*) and, rs885747 (*PER1*) show characteristics of human microRNA genes that might be affected by the autism-associated SNPs. MicroRNAs regulate development (Lee et al. 1993). They are abundant in the mammalian central nervous system (Bak et al. 2008) and are reported to play a role in Fragile-X syndrome (Lin et al. 2006a), Tourette's syndrome (Abelson et al. 2005) and perhaps schizophrenia (Perkins et al. 2007; Hansen et al. 2007) and autism (Abu-Elneel et al. 2008).

Mammalian mature microRNAs are short (21 to 23 nucleotide) RNA molecules that together with proteins of the RISC complex effect gene silencing (Lewis et al. 2005). The precursors of microRNAs are derived from intergenic microRNA genes (Saini et al. 2007) or microRNA genes within introns of protein coding genes. Indeed some 50% of human microRNA genes appear to be intronic (Rodriguez et al. 2004). MicroRNA gene transcripts contain complementary base paired hairpin regions (pri-microRNA hairpins) that are processed into pre-microRNA hairpins by the action of the enzyme DROSHA. The length, stability, architecture and sequence of the pri-microRNA hairpin are determinants of DROSHA-processing and distinguish microRNA hairpins from random hairpins within the genome (Stark et al. 2007a; Stark et al. 2007b). The ~60-100nt long pre-microRNA hairpins are transported to the cytoplasm by EXPORTIN-5 where the loops and tails are removed by the RNase III enzyme, DICER and the remaining ~22nt double stranded mature microRNA loaded into the protein complex RISC. The gene specific targeting of this silencing complex is determined by the nucleotide sequence of the mature microRNA loaded RISC and particularly the seed (the first 2 to 8 nucleotides, 5' to 3') of the mature 22-nucleotide microRNA. Binding of mature microRNA loaded RISC to the UTRs of target genes causes translational repression of the target (Wightman et al. 1993) or rapid degradation of the transcript of the target gene (Giraldez et al. 2006).

MicroRNAs regulate cellular processes that are relevant to the study of autism e.g.: developmental timing (Lee et al. 1993; Reinhart et al. 2002); cell death (Jovanovic and Hengartner 2006); the patterning of the nervous system (Kosik 2006) and, the survival of Purkinje neurons (Schaefer et al.

2007), a cell type affected in cases of autism (Kemper and Bauman 2002). SNPs in microRNA genes can profoundly affect the target specificity and gene silencing power of microRNAs (Duan et al. 2007; Saunders et al. 2007). We wondered, therefore, whether intronic SNPs that are associated with autistic disorder might represent allelic variation in microRNA genes transcribed from the introns of *EN2*, *NPAS2* and *PER1*. In this report we analyse the intronic autism-associated SNPs rs1861972 (*EN2*); rs1861973 (*EN2*); rs1811399 (*NPAS2*) and rs885747 (*PER1*) to consider whether the RNA transcripts of the intron regions containing these autism-associated SNPs have features typical of microRNAs and, if so, whether their predicted target genes are relevant to autism.

#### 4.4.4 Materials and Methods.

We carried out a bioinformatic analysis of the RNA transcripts encoded by introns 1, 2, and 12 of human *EN2*, *NPAS2* and *PER1*, respectively. Initially, we scanned for pri-microRNA-like structures using the web-based bioinformatics tool RNAanalyser (Table 3-1) (Bengert and Dandekar 2003). We subsequently determined whether the autism-associated SNPs rs1861972, rs1861973, rs1811399 and rs885747 are co-located within any computer predicted pri-microRNA-like structures. Where hairpins were found, special attention was given to: the overall energy of the hairpin, the number of nucleotides in the hairpin, the number and distribution of symmetric vs. asymmetric bulges, the number of nucleotides in each arm of the hairpin and the size of the terminal loop. The results were analysed with reference to the principles of microRNA discovery described by Berezikov et al. (2006) and Stark et al. (2007a).

Any pri-microRNA-like structures found to contain an autism-associated SNP were re-analysed to find the effect of the autism-associated SNPs on the structure of any pri-microRNA-like hairpins. This was an important step because even if the computer predicted promising pri-microRNA-like structures, unless one or other allele made some substantial difference to the hairpin, by altering its structure or affecting a potential seed sequence, the overall argument would fail. SNPs lying within candidate seed regions of any pre-microRNA-like structures were deemed of greater impact over SNPs lying outside the seed, but, SNPs that disrupted the configuration of the hairpin would *a priori* have greatest impact overall.

Using the bioinformatics tool Microprocessor SMV *in silico* DROSHA (Helvik et al. 2007) Pri-microRNA-like structures were screened for DROSHA processing sites that are required to generate a pre-microRNA from a pri-microRNA. To determine what would be the predicted target genes of a mature microRNA derived from any such hairpins found harboring one of the autism-associated SNPs, each arm of the pre-microRNA-like structures was analysed to indicate the likely 5' start of the mature microRNA region. Using the principles of mature microRNA recognition described by Stark et al. (2007a) and Berezikov et al. (2006) both the 5' and 3' arms of candidate pre-microRNAs were analysed since experimentally confirmed microRNAs have been reported to be produced from both 5' and 3' microRNA arms (hsa-mir-10b and hsa-mir-10b\*, for example (Michael et al. 2003).

Primary consideration was given to the number of complimentary matches between each candidate seed and the nucleotide sequences of

the 3' UTRs (microRNA targets) of all known protein coding genes in the human genome. For this analysis the TargetScan tool (Table 3-1) was used (Griffiths-Jones et al. 2006). It has been shown that verified microRNAs generally have numerous (~150-300) targets and the start point of verified mature microRNAs often coincides with a trend peak in a seed vs. number-of-seed-matches plot, for a given microRNA (Stark et al. 2007a). Thus if candidate seeds are considered (by a heptamer window moved one nucleotide at a time along a hairpin arm) and the number of predicted targets recorded for each seed, then a trend in the number of targets (increasing to a peak then decreasing over a section of the span of the microRNA arm) can indicate the start point of the mature microRNA contained within that arm (Stark et al. 2007a). However, when looking for target number trends it is important to consider the position of the candidate seeds in relation to the overall structure of the candidate hairpins. For example, seeds in the 3' arm which represent start points of candidate microRNAs shorter than 22 nucleotides were disregarded, as were seeds in 5' arms that would define candidate microRNAs where the 3' end of the mature microRNA coincided with the loop region of the candidate hairpin. We noted candidate seed sequences containing mismatches (likely start points) and uracil nucleotides at the start of a candidate microRNA because nucleotide-one of verified mature microRNAs is biased towards being uracil (Stark et al. 2007a).

BLASTN and SSEARCH tools were also used to consider sequence conservation. Searches were performed using the web-based tools in MirBASE (Table 3-1) (Griffiths-Jones et al. 2006). Thus a number of arguments were deployed against the notion that hairpins containing

autism-associated SNPs are *bona fide* microRNA genes.

The two SNPs (rs1861972 and rs1861973) in *EN2*, reported by Benayed et al. (2005) to be associated with autistic disorder, are physically close together in the single intron of *EN2* and were therefore analysed together. The analysis took account of all four possible combinations of allele; A-C, A-T, G-C and G-T. The results were compared with the haplotype analysis of Benayed et al. (2005) that showed the A-C haplotype to be significantly associated with autism ( $p = 0.0000067$  narrow phenotype). We also considered the effect of any other common SNPs reported in this region.

For rs1811399 in *NPAS2*, an intronic code block of some 300 nucleotides surrounding the SNP was analysed. This block contained no variation according to <http://www.ensembl.org> (release 50) apart from the autism-associated SNP. Thus only 2 structures were computed for the region containing *NPAS2* rs1811399, one for the A allele and one for the autism enriched C allele.

Two SNPs occur in intron 12 of *PER1* (Ensembl release 50), rs885747 (C/G), that is reported to be associated with autism (Nicholas et al. 2007) and rs885953 (G/C) that has not been investigated in autism. The four possible combinations of allele for the SNPs in intron 12 of *PER1* were used in the computation of transcript structures.

Finally using TargetScan (Table 3-1) (Griffiths-Jones et al. 2006) we analysed the best candidate mature microRNAs of any pre-microRNA-like structures found, to determine what would be the predicted targets for these microRNA-like structures. Sets of target genes thus derived were scanned for overlap and for genes with particular relevance to autism.

In summary, we set out to answer four main questions: Firstly, could the genomic DNA sequence fragment containing the autism-associated SNP generate a pri-microRNA-like hairpin? Secondly, do the autism-enriched alleles specifically disrupt pri-microRNA-like structures or change the seed sequence of a candidate mature microRNA-like region? Thirdly, if candidate microRNAs are detected, what are the predicted target genes of the candidate microRNAs? And finally, is there overlap between the targets of each of the candidate microRNAs and if so, are the common targets relevant to autism?

#### 4.4.5 Results.

##### 4.4.5i Analysis of *EN2* intron 1 sequence fragment containing the autism-associated SNPs rs1861972 (A/G) and rs1861973 (T/C).

The autism-associated haplotype rs1861972 (A)-rs1861973 (C) is estimated to contribute to the risk of disease in 40% of cases in the general population (Benayed et al. 2005). These SNPs are situated in the 3' half of the single large intron that divides the two exons of *EN2* (7q 36.3). The two SNPs are 151 nucleotides apart and were analysed together in a 300-nucleotide genomic DNA sequence fragment (Ensembl 2008 release 50).

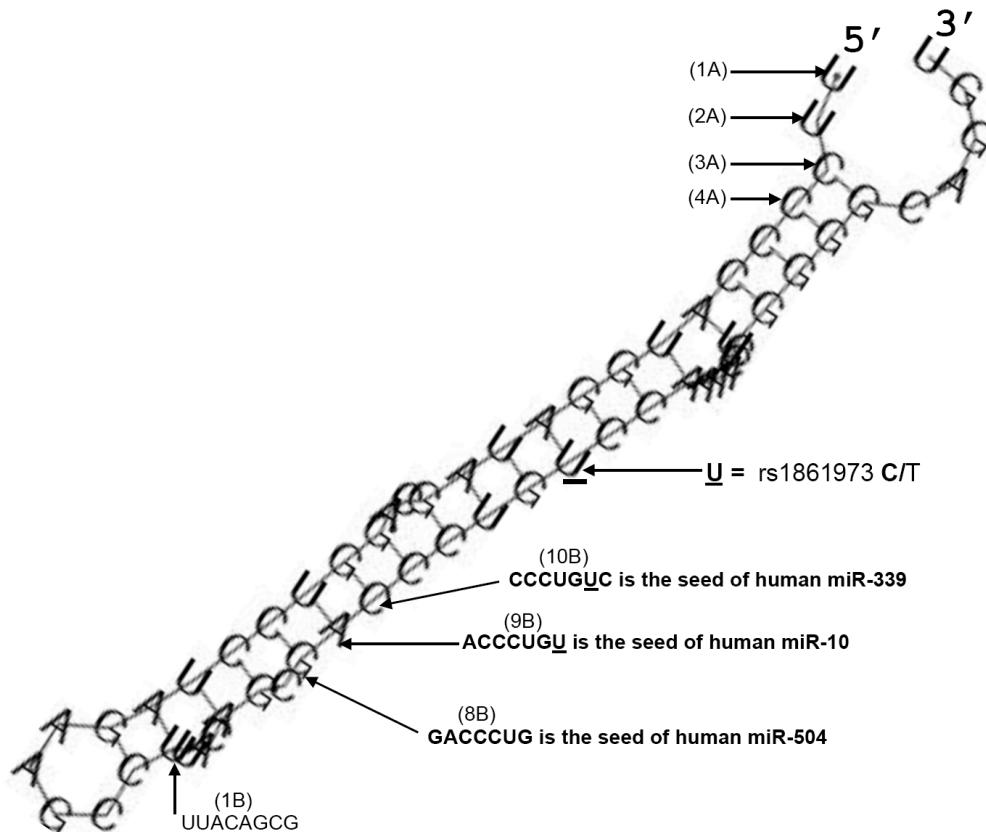
Two other SNPs are present in this sequence fragment; rs35529773 (C/-) for which linkage disequilibrium data was not publicly available and rs3824067 (T/A) that is well characterised and has T as the ancestral allele (Ensembl release 50). We took account of these SNPs in our analysis, computing transcript structures for all the possible combinations



of alleles (four-marker haplotypes) of the two autism-associated SNPs together with rs35529773 and rs3824067. Eight out of the sixteen possible allele combinations (four-marker-haplotypes) permitted the formation of a long 103 nucleotide pri-microRNA-like hairpin (Figure 4-3) with a 4-nucleotide loop. The 3' arm was 58 nucleotides long and the 5' arm 43 nucleotides long. (The total number of nucleotides in 90% of validated microRNAs ranges from 73 to 102 with arm length between 31 and 47 and loops between 4 and 26). This hairpin was always completely disrupted by the presence of the autism-associated rs1861972 (A)-rs1861973(C) haplotype but never disrupted by the presence of the rs1861972(G)-rs1861973(T) haplotype, regardless of which alleles of SNPs rs35529773 and rs3824067 accompanied rs1861972 A/G-rs1861973C/T (Figure 8-2). The rs1861972 (G)-rs1861973(C) haplotype induced a small central bulge in the main stem and a change in a candidate seed sequence of the mature mirRNA-like region of the hairpin. Further analysis using Microprocessor SMV *in silico* DROSHA (Helvik et al. 2007) gave a positive predictive value (PPV: the proportion of positive test results that are correct predictions) >0.47 for a DROSHA processing site for this hairpin with the T allele present in the hairpin and a PPV <0.3 for the hairpin with the C allele.

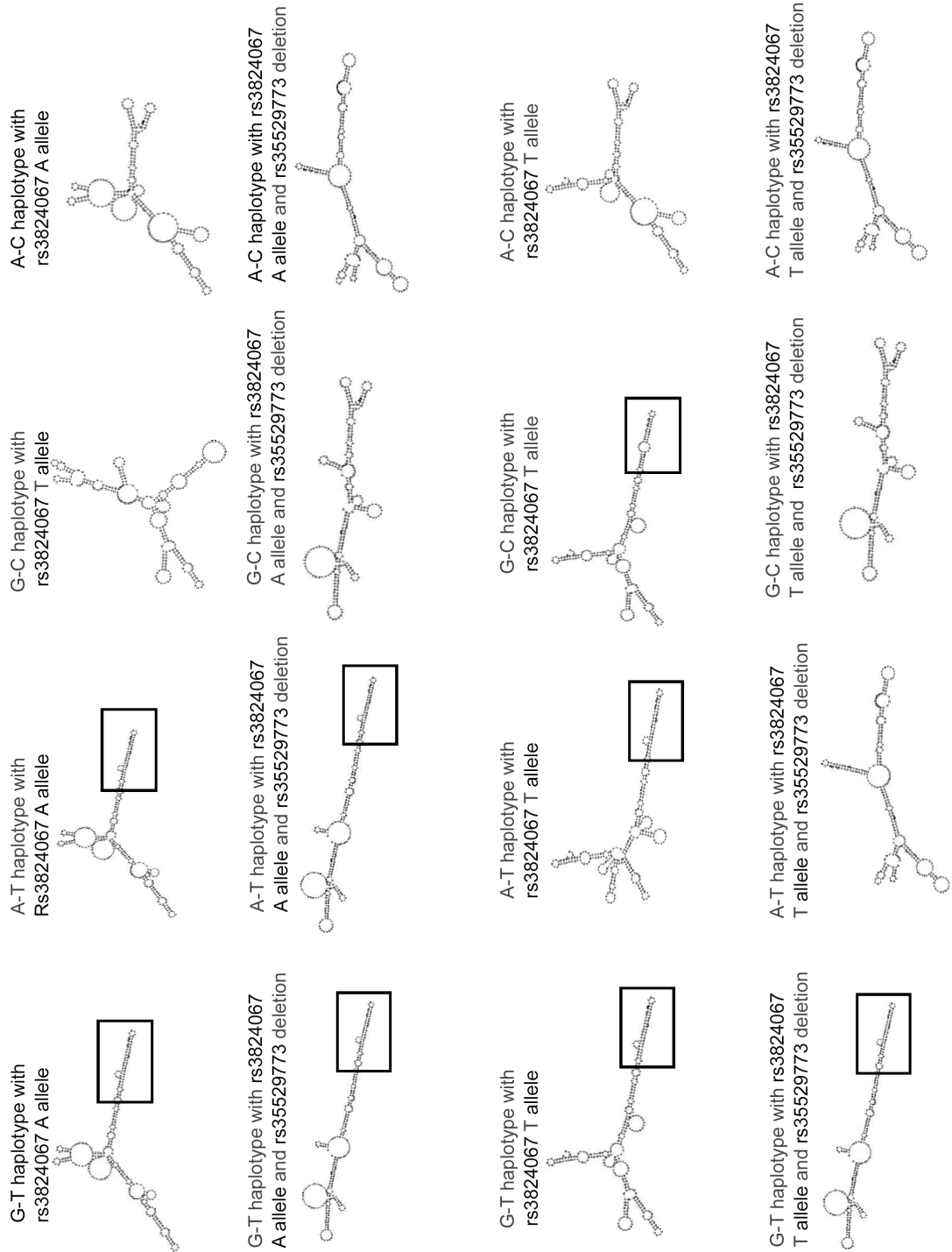
No trend peak was found in the number of seed matches along the 5' arm. The 3' arm however, showed two regions of target number trend with the central region of the 3' arm of this pre-microRNA-like structure encoding the seed sequences of three different established microRNAs. (More than one mature microRNA can be expressed from a given arm of a micro RNA (Yu et al. 2008)). The overlapping seed sequences of miR-10,

miR-339 and miR-504 each contain the autism-associated SNP (Figure 4-2). Intriguingly, miR-10 is reported to play a role in hind-brain development that is in keeping with the function of the *EN2* gene in cerebellum development and with reports of cerebellum anomalies in autism (Kemper and Bauman 2002; Sgaier et al. 2007; Woltering and Durston 2008). For the 3' arm of this *EN2* hairpin we took the two best candidate seeds for investigating predicted targets of this hairpin; UACAGCG and ACCCUGU, the seed of human miR-10.



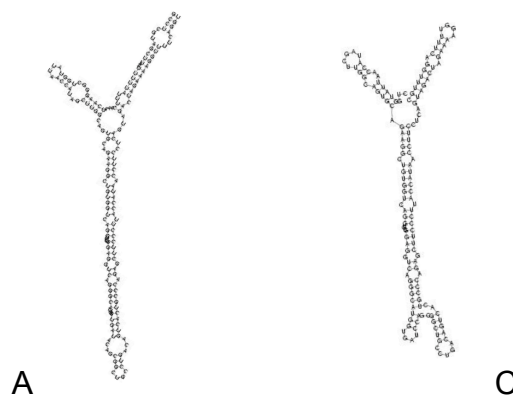
**Figure 4-2. The *EN2* candidate microRNA containing the autism-associated SNP rs1861973.** Note the seed sequences of three known human microRNAs are represented in the candidate mature microRNA region of the 3' arm.

**Figure 4-3. Effects of SNPs rs1861972 and rs1861973 on secondary structure of mRNA transcribed from an *EN2* intron fragment.**



4.4.5ii Analysis of the intron region of *NPAS2* containing the autism-associated SNP rs1811399 (C/A).

A 300 nucleotide DNA sequence fragment (Ensembl release 50) centred upon the SNP rs1811399 in *NPAS2* was analysed to determine the predicted structural characteristics of the RNA transcript of this intron region. Two DNA sequence fragments were analysed for this SNP, identical except that one contained the A allele and the other the autism enriched C allele of rs1811399. The results of the analysis gave two different RNA structures as shown in Figure 4-4. The 5' arm of the hairpin computed with the A allele of rs1811399 contains 50nt. The loop 4nt, the 3' arm, 44nt and the hairpin in total 101nt. These measures are in keeping with the majority of verified microRNAs (Stark et al. 2007a). The rs1811399 SNP is predicted to be located within the 5' arm of the long RNA hairpin containing the A allele of rs1811399, towards the loop of this pri-microRNA-like structure. Figure 4-4 shows the folding of the fragment containing the rs1811399 SNP. Disruption of the hairpin containing the A allele of rs1811399 is predicted when the A allele is replaced by the autism enriched C allele.



**Figure 4-4. A or C Alleles of rs1811399 determine two different mRNA structures.**

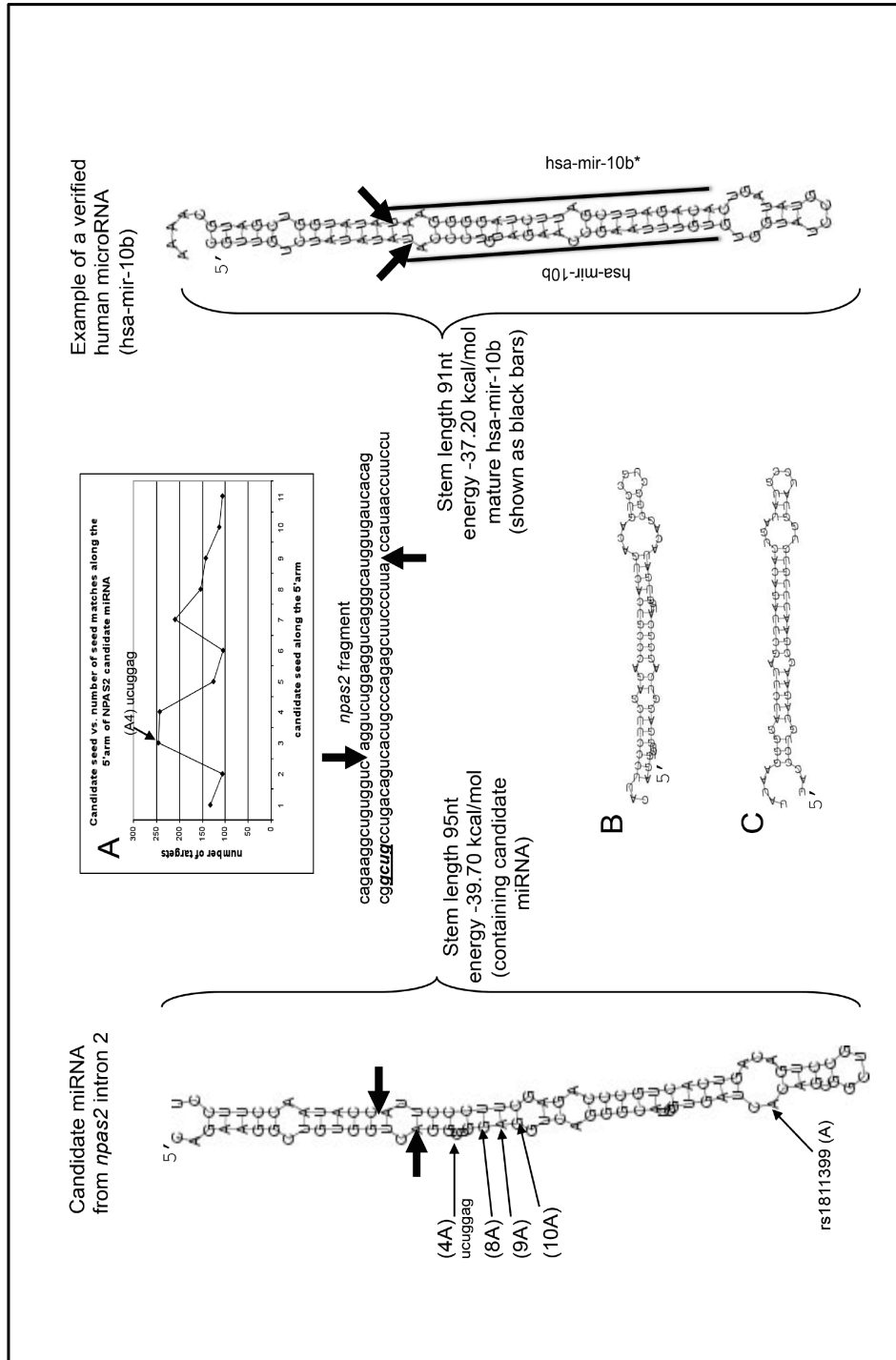
To further challenge the plausibility of this structure being a precursor for a functional microRNA, we searched for DROSHA binding sites within the pri-microRNA-like hairpin. Using the web based bioinformatics tool, Microprocessor SVM *in silico* DROSHA (Helvik et al. 2007) we were able to detect a DROSHA processing site that would derive a pre-microRNA-like structure shown in Figure 4-5. This was used to investigate the predicted targets of this hairpin using the web based tool TargetScan (Griffiths-Jones et al. 2006).

#### 4.4.5iii Analysis of intron 12 of *PER1* containing the autism-associated SNP rs885747 (C|G).

Intron 12 of *PER1* contains two SNPs: rs885747 that Nicholas et al (2007) found to be associated with autistic disorder (C allele) and rs885953 that was not analysed in their experiment. The PPV for the DROSHA sites in this hairpin were >0.38 for the C allele and <0.36 for the fragment containing the A allele. Seed-match target number trends were observed along the 5' and 3' arms of this hairpin. In the 5' and 3' arms, seed-match trend peaks coincided with base pairing mismatches and were taken as likely mature microRNA start points. There is remarkable conservation amongst primates of the predicted hairpin structure of the RNA transcripts of this intron. In *H. sapiens* this hairpin structure is strongly affected by the allele combination at these two loci (Figure 4-6). Due to the small size of the intron, the conservation of predicted hairpin structure amongst primates and the high GC content of this intron, 61.36% compared with 55.6 +/- 1.56 for short introns in general (Duan et al. 2007) we considered

that this intron may best be analysed in terms of the structural features of mirtrons (Berezikov et al. 2007; Ruby et al. 2007).

Mirtrons are microRNAs transcribed from small introns that by-pass the DROSHA cleavage step in the micro RNA pathway by way of splicing at the intron/exon boundaries and where the pri-microRNA hairpins lack an extended tail. The mature microRNA within such mirtrons is located in the closely base paired region distal to the central loop as compared with canonical pri-microRNAs where the mature microRNA is generally located proximal to the loop of the pri-microRNA. *In silico* analysis shows that the G-C and C-G haplotypes of these two SNPs allow folding into mirtron-like hairpins while the G-G and C-C haplotypes preclude hairpin formation. To predict start points of candidate mature microRNAs produced from these hairpins, trends in the number of seed-matches along the 5' and 3' arms were considered. For the hairpins containing the C-G and G-C haplotypes, two peaks indicated plausible start points, one in the 5' arm of the closely paired region and one in the 3' arm of the closely paired region. Seed-match trend peaks that fell outside of the closely paired region of the hairpin, in the more loosely base-paired region towards the loop, were disregarded as being less likely to contribute to a candidate mirtron (Figure 4-6). The G-C haplotype containing the autism enriched C allele of rs885747 presents a conformation that shifts the location of the 5' arm seed-match trend peak (relative to its position in the 5' end of the closely paired region in the C-G haplotype) to a region of the G-C hairpin where it is less likely to represent a start point for a mature microRNA; a start point corresponding to a mature microRNA with a 3' end extending beyond the closely base paired region of the hairpin.



**Figure 4-5.** Candidate microRNA from intron 1 of *NPAS2* and verified human microRNA *hsa-mir-10b* for comparison. Bold black arrows indicate the position of predicted DROSHA cleavage sites while fine black arrows indicate candidate seeds in the 5' arm of the candidate microRNA in *NPAS2*. Graph A shows a plot of the candidate seeds (1A, 2A, 3A etc.) vs. the number of predicted seed matches. B is the *NPAS2* candidate pre-microRNA and C the *hsa-mir-10b* pre-microRNA.

Subtle structural changes between G-C and C-G haplotypes affecting the region at the start of the closely-base-paired region of the 3' arm may have an effect, but no gross change in the 3' arm's closely-base-paired region is associated with the difference between the two alleles of rs885747 (according to our analysis). Therefore we focused on further investigation of the effect of rs885747 on the 5' arm's candidate mature microRNA and specifically the 22 nucleotide closely paired region at the 5' end of the hairpin as the region of the candidate mirtron affected by the autism-associated SNP rs885747.

#### 4.4.5iv Overlap between the sets of predicted targets for the *EN2*, *NPAS2* and *PER1* candidate microRNAs:

The predicted targets of the candidate microRNAs from *EN2*, *NPAS2* and *PER1* included a number of autism candidate genes: *RAI1*, *GABRB3*, *GABRB2*, *SHANK3*, *NRXN3*, *RELN*, *PITX1*, *SHANK3-INTERACTING-PROTEIN-1*, *A2BP1*, *STK39* and *DLX1* (Yang and Gill 2007; Zafeiriou et al. 2007). We further screened the total target data for overlap and found six target genes were common to the data sets corresponding to each of the candidate microRNA's predicted targets. The common targets were *ACVR1B*, *DAB2IP*, *MAP2K4*, *MTMR4* and *RAI1*. The relationship between autism relevant targets of the candidate microRNAs and the candidate host genes are summarized in Table 4-1.



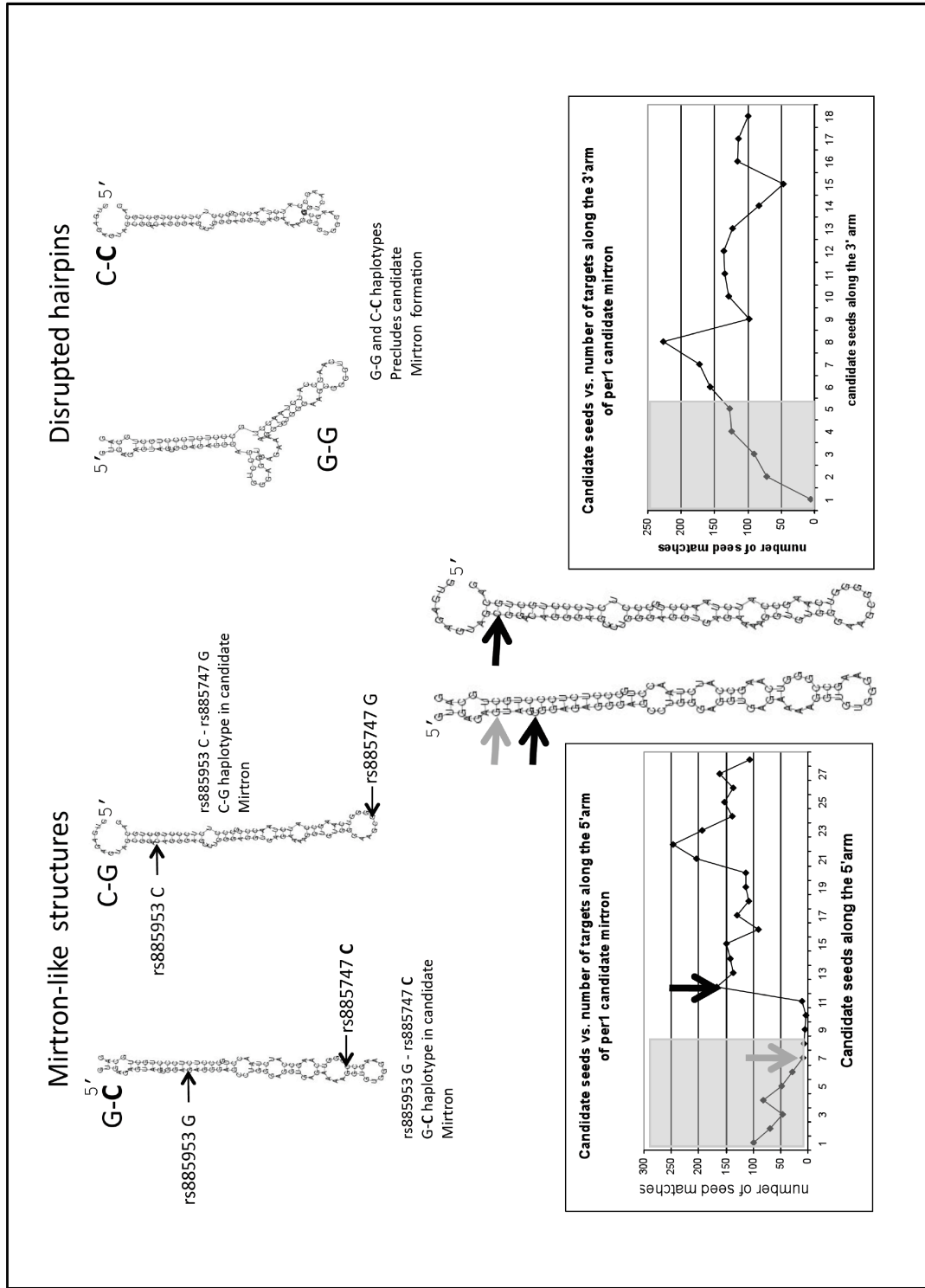


Figure 4-6. PER1 intron 12 hairpins

**Table 4-1. The predicted targets of the candidate microRNAs in *EN2*, *NPAS2* and *PER1*.**

<b>Host gene</b>	<b>NPAS2</b>		<b>EN2</b>		<b>PER1</b>	
Arm of candidate microRNA	5'	3'	5'	3'	5'	3'
Seed of candidate mature microRNA	UCUGGAG ACAGUCA		ACCCTGT TTACAGC hsa-mir-10		GGACAGG	
Number of predicted targets	246		235		178	
Autism relevant targets of the candidate microRNAs	<b>GABRB2</b> <i>MECP2</i> <b>NLGN2</b> ARID1A <b>GABRB3</b> <b>PITX1</b> <b>NRXN3</b> <i>RAI1</i> <b>RELN</b> <b>STK39</b> TLK1- MITF+ ProSAPiP1		<b>BDNF</b> <b>SHANK3<sup>^</sup></b> <i>NF1</i> FLT1+ <b>A2BP1<sup>^</sup></b> <b>AUTS2</b> <b>DLX1</b> <b>GABRB3</b> <b>HTR2A</b> <i>RAI1</i> ARID1A SCHIP1+ TLK1-		<i>DYRK1A</i> <i>RAI1</i> FLOT2- KIF1A	
<b>Targets common to candidate microRNAs from each gene</b> ACVR1B, DAB2IP, MAP2K4, MTMR4, RAI1						

**Table 4-1 (footnote).** The symbols + or - after a gene name indicate genes that are reported to show altered expression levels in autism. Similarly, ^ indicates genes that are found in micro-deletions associated with autism. Bold indicates genes that are reported to show positive association with autism and genes shown in italics are considered to play a causative role in disorders co-morbid with autism. The synaptic protein ProSAPiP1 is included as a SHANK3 interacting protein (Yang and Gill 2007).

#### 4.4.6 Discussion

The above findings suggest that introns 1, 1 and 12 of the genes *EN2*, *NPAS2* and *PER1* respectively, may harbor microRNA genes that are affected by the autism-associated SNPs rs1861972, rs1861973, rs1811399 and rs885747. We have shown that all of these SNPs alter microRNA-like structures, predicted for the mRNA transcripts of the genomic DNA sequence fragments containing these SNPs. The presence

of one (the autism enriched) but not the other, allele of each SNP disrupts hairpin structure or changes a candidate seed sequence and thereby the predicted target specificity of the microRNA-like structure.

For *EN2*, the RNA structural analysis was extended by the presence of four SNPs (rs3824067, rs1861972, rs35529773 and rs1861973) in the region of interest. We found that eight of the sixteen possible combinations of allele gave a long hairpin that always formed in the presence of the G-T rs1861972-rs1861973 haplotype. Notably, this hairpin structure never formed when the autism-associated A-C rs1861972-rs1861973 haplotype was present and which may relate to the highly significant association for the A-C haplotype observed by Benayed et al. (2005). Further comparing our results we see that in all of the samples presented by Benayed et al. (2005) (AGRE 1, AGRE 2, NIMH and the DSP siblings), the A-C haplotype was always over-transmitted from parents to affected individuals while the G-T haplotype was always under-transmitted to affected individuals and occurred less in the set of unaffected sibs in the DSP (discordant sib pair) test.

Our structural findings precisely mirror the data for Benayed et al. (2005), showing that the autism-associated A-C haplotype always accompanies hairpin disruption while the G-T haplotype always supports a predicted hairpin structure. Recently, Brune et al. (2008) confirmed association of rs1861972 with autism (rs1861973 was not analysed in their study). However, according to Brune et al. (2008) the A and G alleles of rs1861972 both proved positive but in different samples within their study. Our findings offer an explanation supporting and additional to that presented by Brune et al. (2008). Our structural analysis shows that the A

and C alleles of the rs1861972-2861973 haplotype are required for disruption of the hairpin regardless of the genetic background (i.e. which alleles of the other two SNPs in this fragment accompany the A-C rs1861972-rs1861973 haplotype). However, the G-C rs1861972-rs1861973 haplotype causes complete disruption when accompanied by the A allele of rs3824067 but only partial disruption with the T allele of rs3824067. If, in the populations studied by Brune et al. (2008) and Benayed et al. (2005), disruption of this *EN2* hairpin contributes to the cause of autism then the presence of either the A or G allele of rs1861972 could effect this disruption and thus be linked with autism. We therefore tentatively suggest that variation at rs1861972 and rs1861973 affects microRNA mediated regulation of the levels of targets such as *A2BP1*, *AUTS2*, *BDNF*, *GABRB3*, *HTR2A*, *NF1*, *SHANK3* and *RAI1* in cells expressing the homeobox gene *EN2* in autism (Table 4-1).

In *NPAS2*, the disruption of the hairpin by the presence of the autism enriched C allele of rs1811399 is also in keeping with the notion of a loss of microRNA mediated regulation in *NPAS2*-expressing cells in autism. This is consistent with the route to disequilibrium in autism reported for rs1811399 (Nicholas et al. 2007) as the enrichment of the C allele (disrupted hairpin) in autism occurred through an under-transmission of the A allele (intact hairpin). The gene targets of the candidate microRNA (Figure 4-5) containing the A allele of rs1811399 includes genes of particular relevance to autism: *RAI1*, *GABRB3*, *GABRB2*, *NLGN2*, *PITX1*, *SHANK3-INTERACTING-PROTEIN-1*, *STK39*, *NRXN3*, *DLX1* AND *RELN* (see Table1). We therefore tentatively suggest that the enrichment of the C allele of rs1811399 in autism compromises the potential for microRNA

dependent gene regulation in *NPAS2* expressing cells. The results for *PER1* suggest that intron 12 of *PER1* may encode a mirtron (microRNAs expressed from small introns). They also indicate how the autism-associated SNP rs885747 in intron 12 may cause disruption of the more likely functional structure represented by the C-G haplotype of rs885953-rs885747 when the rs885747 C allele is accompanied by the rs885953-G allele. Loss of the structure represented by the C-G haplotype of rs885953-rs885747 may equate to a loss of microRNA mediated regulation of targets that include the autism linked *RAI1* (Bi et al. 2007; Potocki et al. 2007) and the Down syndrome related kinase *DYRK1A* (Altafaj et al. 2001; Fotaki et al. 2002; Dowjat et al. 2007; Zafeiriou et al. 2007). If this microRNA exists and is active in autism then our model would also implicate rs885953, a SNP that is not currently fully characterized ([www.hapmap.org](http://www.hapmap.org)). Thus, for the autism-associated SNP rs1811399 in *NPAS2* and the rs1861972-rs2861973 A-C haplotype in *EN2* the autism-associated allele of rs885747 in *PER1* appears to force a structural change in a predicted hairpin that may result in the loss of microRNA mediated regulation.

It is likely that a number of genes contribute to the heritability of autism (Pickles et al. 1995). We therefore considered whether the effect of hairpin disruption by the autism-associated SNPs in the genes *EN2*, *NPAS2* and *PER1* might be additive, in terms of lost targeting, and we proposed that this could be represented by overlap between the data sets of targets of these candidate microRNAs and that the genes common to each data set should have relevance to autism. We pooled the target data for each of the candidate microRNAs in *EN2*, *NPAS2* and *PER1* and found five genes

that appeared in each of the target gene sets of the candidate microRNAs. These genes were *ACVR1B*, *DAB2IP*, *MAP2K4*, *MTMR4* AND *RAI1*. Intriguingly *RAI1* and *MAP2K4* are both located at the autism susceptibility locus 17p11.2, a chromosomal region where deletion or duplication is linked to Smith Magenis Syndrome or Potocki-Lupski Syndrome, respectively, and which convey an autism phenotype. Evidence suggests that genes in this region, especially *RAI1*, affect neural development in a dose-dependent manner (Zhang et al. 2004). *MAP2K4* (17p11.2) is linked to the cellular response to oxidative stress (Wang et al. 1998) and along with *NPAS2*, may be implicated in oxidative stress induced apoptosis of dopaminergic neurons (Anantharam et al. 2007). *ACVR1B* (12q 13.13) is the activinA receptor, type IB gene. Activin is found to modulate anxiety-related behaviour and adult neurogenesis in mouse and to play a role in recovery from ischemic brain injury (Ageta et al. 2008). *DAB2IP* (9q33.2) transduces TRAF2-induced ASK1-JNK activation (Zhang et al. 2004) and thus plays a central role in the oxidative stress response pathway (Shen et al. 2004) that is reported affected in autism (James et al. 2004). *MTMR4* (17q22) is a lipid phosphatase for phosphatidylinositol-3-phosphate (PTDINS3P) (Lorenzo et al. 2006) and disrupted phosphatidylinositol signaling is also reported in autism (Serajee et al. 2003).

It is not currently possible to verify the existence of microRNA genes on the basis of bioinformatic analysis alone and substantiation of the above findings would require additional experiments beyond the scope of this study. However, our results suggest a mechanism whereby certain intronic autism-associated SNPs may have functional significance and moreover how common SNPs may act in combination to alter phenotype.

## **4.5 Further candidate microRNA analysis using MirBASE, The Vienna RNA Web Server and the RNA Self Containment Web Server.**

### 4.5.1 Introduction.

Since the publication presented in chapter 4.4 additional bioinformatics tools have become publicly available and existing facilities expanded. For example, the Vienna RNA Web Server (Table 3-1) allows detailed analysis of hairpin structures as individual sequences and multi-species assemblies. This facility enabled the structures to be tested for evolutionary conservation independently of sequence conservation. Additions to the MirBASE collection were also taken into consideration. Thus The Vienna RNA Web Server was used to search for candidate microRNA genes in the 300bp intronic sequence fragments containing the autism-associated SNP: rs1811399, rs885747 and rs1861973; that showed positive results in scans with the RNA Analyser described in chapter 4.4.

### 4.5.2 Method

In addition to calculating the minimum free energy for an individual hairpin, the Vienna RNA Web Server allows multi species alignments to be analysed. For this analysis, sequences from five primates and containing the candidate human hairpins were used. These sequences were obtained with Ensembl sequence database and the Ensembl alignment tool or clustalW alignment tool. The Vienna RNA Web Server was used to calculate base pair probabilities, mean free energy and entropies for each element, as well as giving a “best fit” structure for the assembly overall, a

consensus alignment. The output shows whether the overall structure of the hairpin is withstanding base changes between species. If so, this is taken to indicate conservation of structure (although sequence variation is occurring that is not disrupting hairpin structure overall).

Dot plots of a hairpin assembly allowed exploration of the probability of the occurrence of a particular secondary structure contained in the whole Boltzmann ensemble (all possible structures a hairpin might take up, between a single strand and the structure with the lowest mean free energy). The output for this application is a base pairing probability matrix, the so-called dot plot, which takes the form of an upper, and a lower triangle of a diagonally divided square matrix. Each letter of the primary structure (linear input sequence) is assigned to a matrix index, i.e. arranged sequentially along the two dimensions of the square. The small squares that make up the matrix grid are shaded in to show base pairings. Shaded box (dot) size is proportional to the probability of base pairing. The lower triangle shows the secondary structure with minimal free energy and where the color of the dots represents sequence variation when an assembly of hairpins is analysed. Dot plots can show how some microRNAs that have disorganized MFE structures are still capable of making hairpins when the whole Boltzmann ensemble is considered. In these cases, a structure within the ensemble and which is not the MFE structure, is nevertheless sufficiently stable and common, to allow processing by the microprocessor complex.

Using the Vienna RNA Web Server, a 300 nucleotide DNA sequence fragment (centered upon the autism SNPs in *NPAS2*, *PER1* and *EN2*) was thus analyzed to determine the predicted structural characteristics of the



RNA transcript containing these SNPs. The effect of the SNPs on structures was noted. The graphical results were recorded as portable document files output from the Vienna RNA Web Server.

### 4.5.3 Results.

#### 4.5.3i NPAS2 rs1811399 hairpin.

The results for the *NPAS2* rs1811399 hairpin are shown below in the figures and tables. (The analysis also corrects a typographic error in the *NPAS2* sequence presented in Figure 4-5, the Nature Precedings submission (page 183) where a nucleotide (C) is gained at the 5' side of the first bulge distal to the central loop of the hairpin. The revised structures in this chapter show that no overall change of structure is made to the hairpin by this copying error).

**Table 4-2. Reference sequences for the *NPAS2* rs1811399 hairpin.**

<b>NPAS2 sequence fragment</b>	<b>Notes</b>
<p>300bp fragment starting 2:101478854</p> <p>TTCATCGGGCTGTATTTAACTTCTGCTTTTGTCTGA  AAAGCAAAGTTGGAGGGTTTCATTGCCCTCGTAGCTT  AATTTTATTTCAAATCAAGGGCTGGTATTAACCAT  AGCTTGGCAGTGCA<b>C</b>AAGGCTGTGGTCAGGTCTGGA  GGTCAGGGCATGGTGAT<b>M</b>CAGCG<b>GCTG</b>CCTGACAGT  CACTGCCAGAGCTTCCCTTACCATAACCTT<b>C</b>CTCA  GTAGACTAGAAAAGGTTTTTCAGGTTTGCCAGGTTA  TCCACACGTACCATGGCATAGTTTCTCCAGCAAGTC  AGCACACTGCCG</p> <p>C allele DROSHA fragment. Score -0.257</p> <p>GTCAGGTCTGGAGGTCAGGGCATGG<b>TGATC</b>CAGCGG  CTG<b>CCTGA</b>CAGTCACTGCCAGAGCTTCCCTTACCA  T (double loop =forked end)</p> <p>A allele DROSHA fragment. Score 0.068</p> <p>AGGTCTGGAGGTCAGGGCATGGTGAT<b>A</b>CAGCG<b>GCTG</b>  CCTG<b>ACAGTCA</b>CTGCCAGAGCTTCCCTTAC</p>	<p>M = rs1811399 A/C</p> <p><b>Grey</b> = fragment for <i>in silico</i> DROSHA.</p> <p><b>Pink</b> = the limit of the long hairpin structure</p> <p><b>Grey</b> = candidate pre-miRNA as predicted by Interagon (A allele)</p> <p><b>GCTG</b> = position of central loop</p> <p><u>CTGGAGG</u> top seed match 5'</p> <p><u>ACAGTCA</u> top seed match 3'</p>

**Table 4-2. Reference sequences for the *NPAS2* rs1811399 hairpin (foot note).**

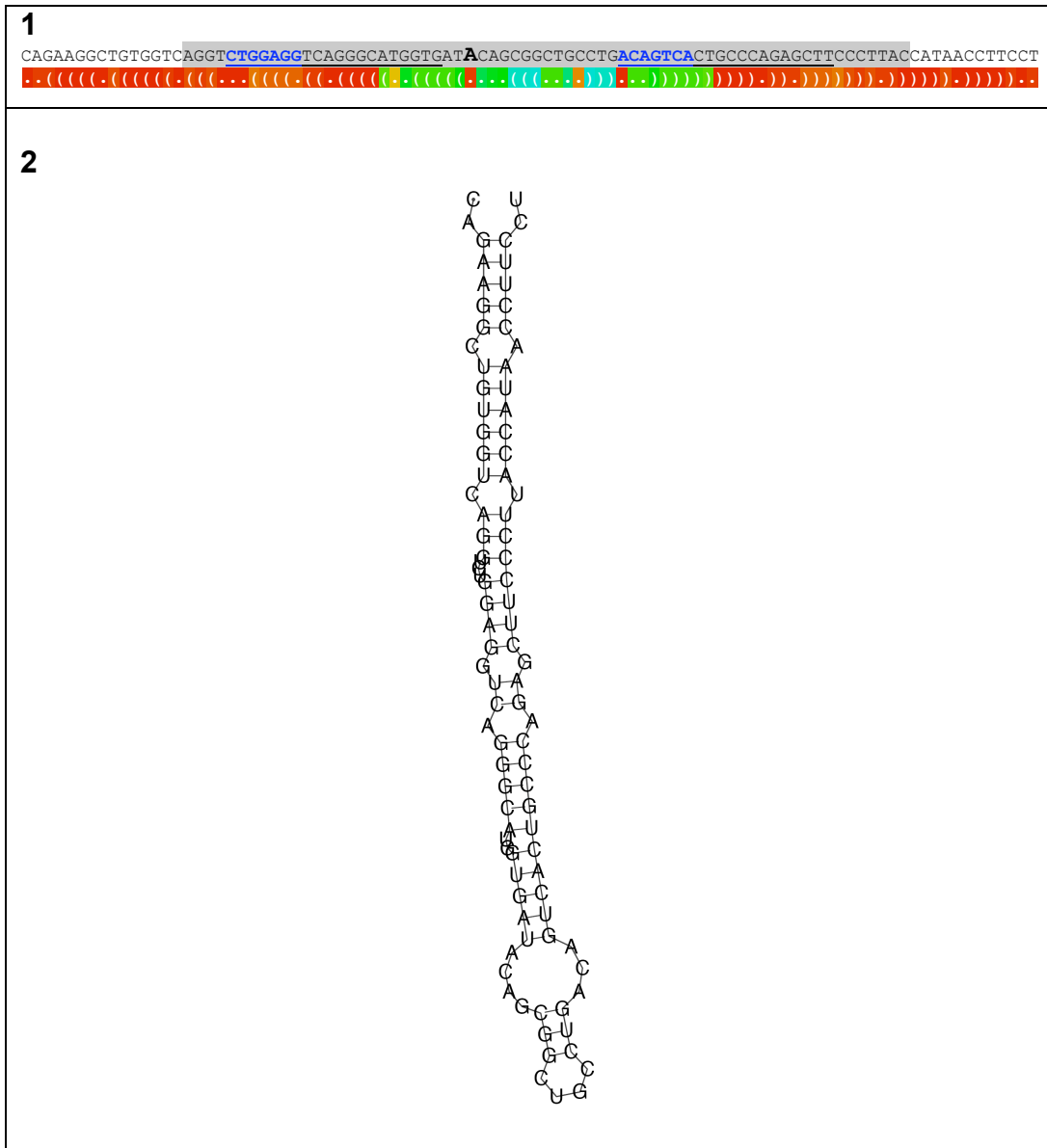
The predicted *in silico* DROSHA scores and predicted excision sites are as for the results given by the Interagon *in silico* DROSHA tool described in Chapter 4.4

**Table 4-3. Vienna RNA Web server RNAfold output statistics for the *NPAS2* hairpin containing the SNP rs1811399.**

	<i>NPAS2</i> hairpin		
	C Allele rs1811399	A Allele rs1811399	hsa-let-7f-2 (Forked MIR)
Optimal MFE structure	-41.60 kcal/mol	-39.50 kcal/mol	-42.1 kcal/mol
FE ensemble	-42.57 kcal/mol	-41.14 kcal/mol	-43.85 kcal/mol
Frequency of the MFE structure in the ensemble	20.65 %	6.94%	5.8%
Ensemble diversity	5.70	12.57	6.64
Centroid MFE	-41.60 kcal/mol	-35.21 kcal/mol	-41.90 kcal/mol

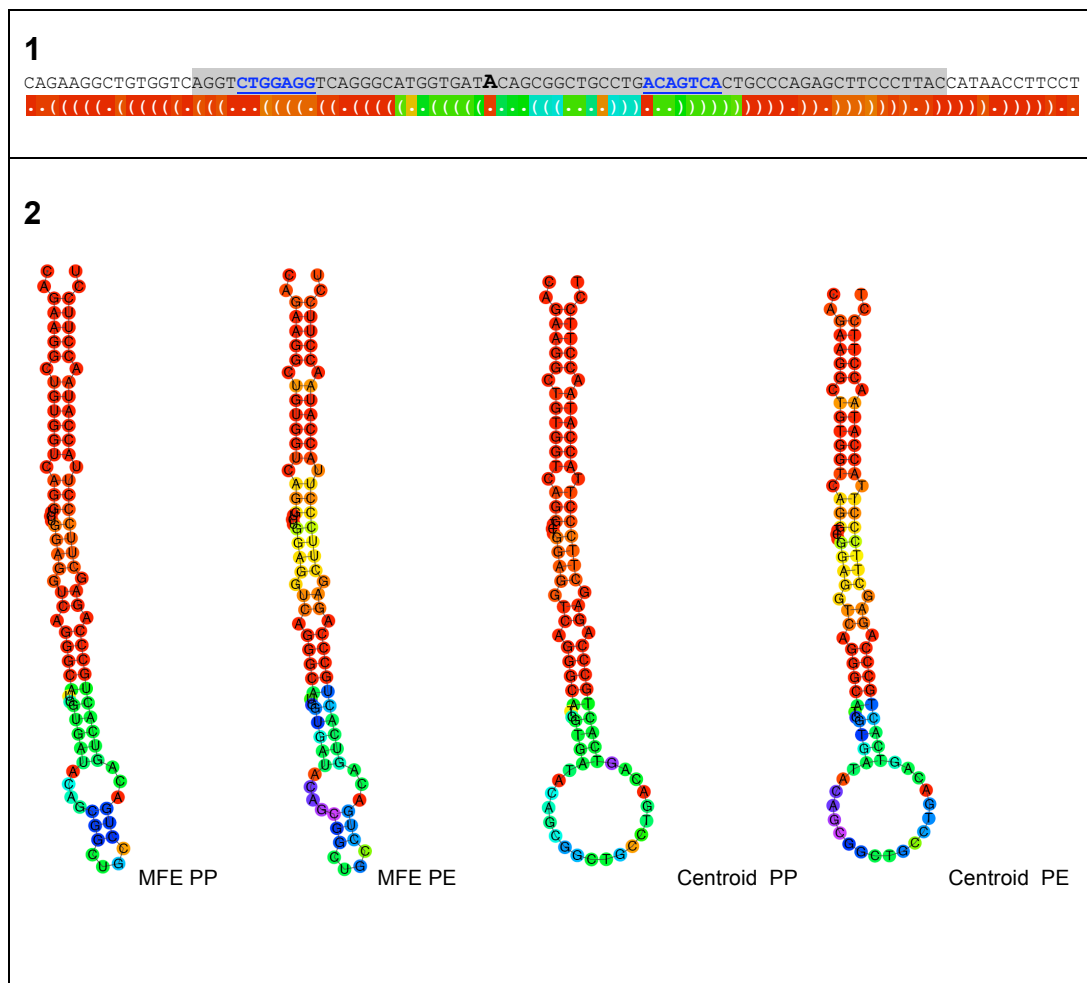
**Table 4-4. Vienna RNA Web Server RNAz output statistics for the *NPAS2* rs1811399 hairpin for five primate sequences.**

Length	92nt
strand	forward
Mean pair wise identity	95.92
Mean single sequence MFE	-41.22
Consensus MFE	-36.97
Energy contribution	-36.53
Covariance contribution	-0.44
Combinations/Pair	1.16
Mean z-score	-3.18
Structure conservation index	0.90
SVM decision value	3.77
SVM RNA-class probability	0.999597
Prediction	RNA



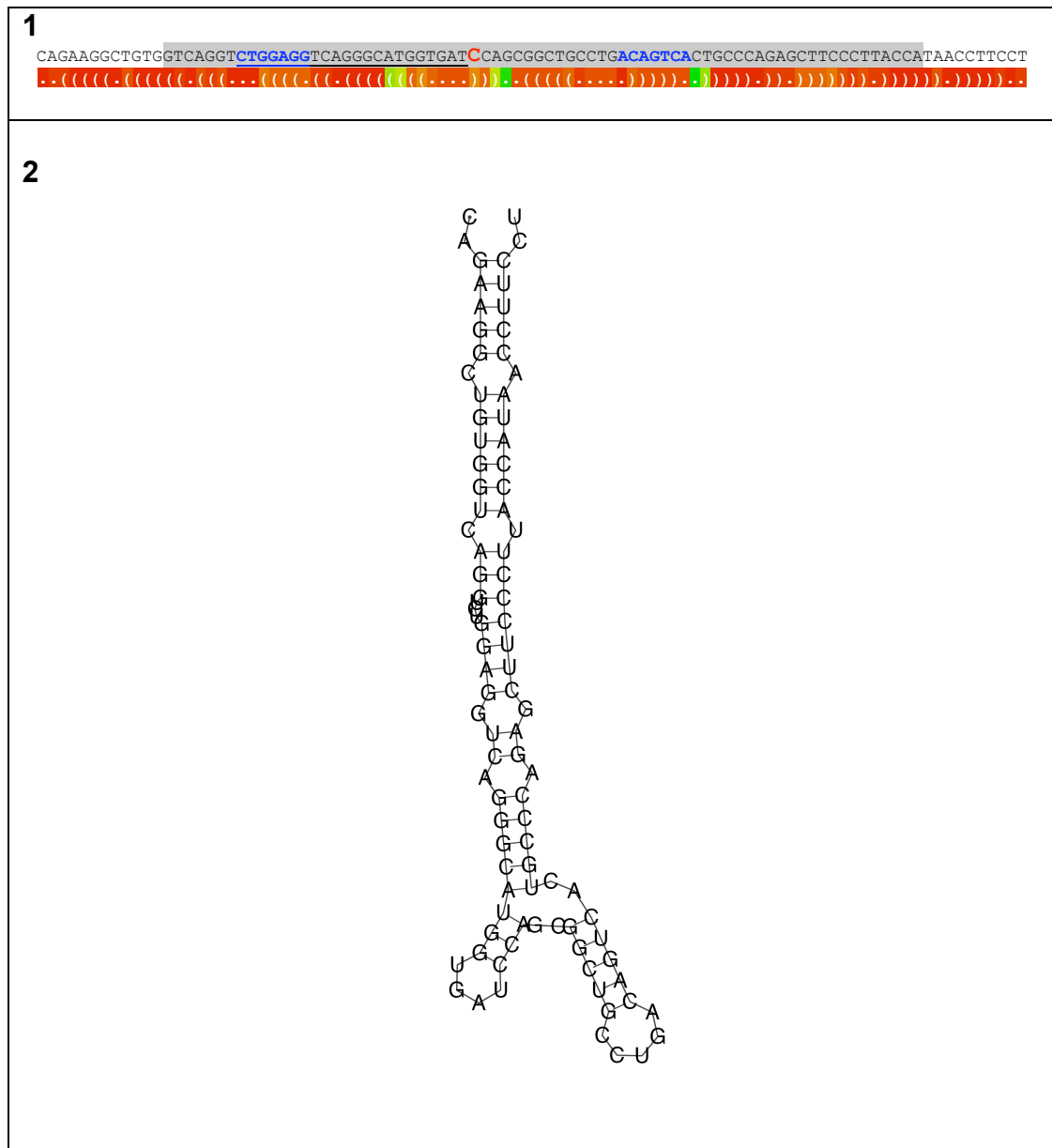
**Figure 4-7. The *NPAS2* rs1811399 hairpin: A allele; RNAfold, Vienna RNA Web Server.**

The *NPAS2* sequence shown in panel 1 and containing the SNP rs1811399 (A) is predicted to form a long RNA hairpin with the Vienna RNA Web Server RNAfold. The sequence is under laid by the secondary structure for the hairpin given in dot and bracket notation. The coloured strip indicates the base pairing probability where red =1 and violet = 0. Predicted seeds are shown in blue. The grey shading is the predicted pre-microRNA with the 5' and 3' candidate miRs underlined. In panel 2, the hairpin structure is shown graphically.



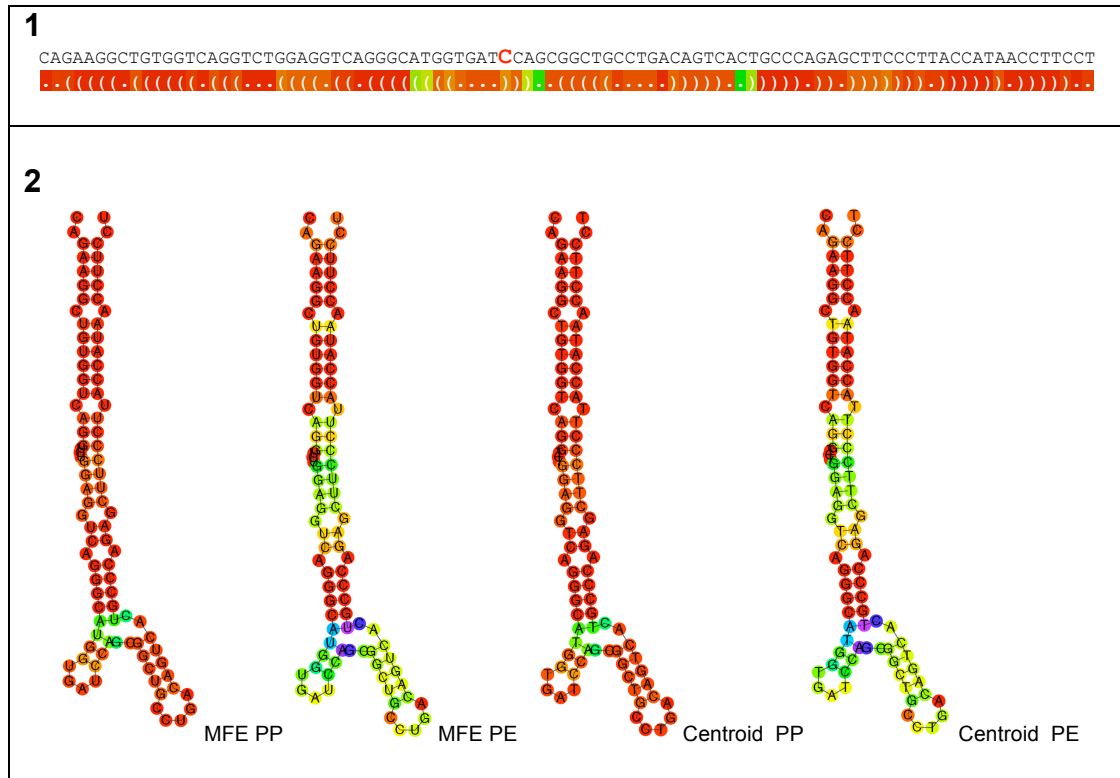
**Figure 4-8. The *NPAS2* rs1811399 hairpin: A allele; RNAfold, Vienna RNA Web Server MFE and centroid structures.**

Panel 1 is as described for figure 4-7 above. In panel 2, the hairpin structure is shown graphically. The mean free energy (MFE) structures are shaded to show positional probability of the bases (MFE PP) where red = 1 and violet = 0 and positional entropy (MFE PE) where red = 0 and violet = 1.4. The centroid structures, indicating a “common structure” for the whole assembly of possible structures have the same colour code.



**Figure 4-9. The *NPAS2* hairpin: C allele; RNAfold, Vienna RNA Web Server.**

Panel1: sequence detail in the hairpin structure and base pair probabilities are shown for the dot and bracket structure. The predicted seeds are shown in blue and the grey shading is the predicted pre-microRNA with the 5' candidate miR underlined. Note the 3' miR for the A allele hairpin now places across the central loop. Panel 2 MFE structure. Other conventions are as for figure 4-7.



**Figure 4-10. The *NPAS2* hairpin: C allele; RNAfold, Vienna RNA Web Server MFE and centroid structures.**

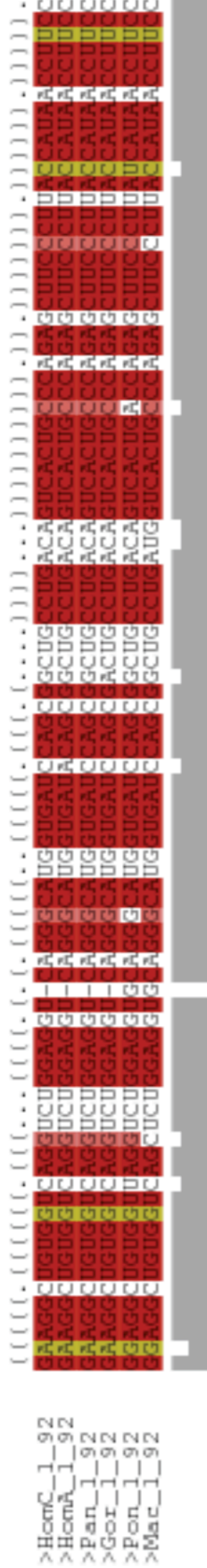
The *NPAS2* sequence shown in panel 1 and containing the SNP rs1811399 (C) is predicted to form a forked RNA hairpin. The sequence is under laid by the secondary structure for the hairpin given in dot and bracket notation. The shading of the sequence strip indicates the base pairing probabilities where red = 1 and purple = 0. In panel 2, the hairpin structure is shown graphically. The mean free energy (MFE) structures are shaded to show positional probability of the bases (MFE PP) where red = 1 and violet = 0 and positional entropy (MFE PE) where red = 0 and purple = 1.4). The centroid structures (that indicate an “average” structure for the whole ensemble of possible structures) follow the same convention.

H C	GAAGGCUCUGGUUAGGU <b>UCUGGAGGU</b> - <b>CAGGGCAUGGUGA</b> <u>U</u> <b>CAGCGGCUGCCUGACAGUCA</b> <b>CUGCCCAGAGCUUC</b> CCUUACCAUAACCUUCC (((((((((.....)))))))))	(-41.40)
H A	GAAGGCUCUGGUUAGGU <b>UCUGGAGGU</b> - <b>CAGGGCAUGGUGA</b> <u>U</u> <b>CAGCGGCUGCCUGACAGUCA</b> <b>CUGCCCAGAGCUUC</b> CCUUACCAUAACCUUCC (((((((((.....)))))))))	(-39.30)
Pan	GAAGGCUCUGGUUAGGU <b>UCUGGAGGU</b> - <b>CAGGGCAUGGUGA</b> <u>U</u> <b>CAGCGGCUGCCUGACAGUCA</b> <b>CUGCCCAGAGCUUC</b> CCUUACCAUAACCUUCC (((((((((.....)))))))))	(-41.40)
Gor	GAAGGCUCUGGUUAGGU <b>UCUGGAGGU</b> - <b>CAGGGCAUGGUGA</b> <u>U</u> <b>CAGCGGCUGCCUGACAGUCA</b> <b>CUGCCCAGAGCUUC</b> CCUUACCAUAACCUUCC (((((((((.....)))))))))	(-42.00)
Pon	GGAGGCUCUGGUUAGGU <b>UCUGGAGGU</b> <b>CAGGGCAUGGUGA</b> <u>U</u> <b>CAGCGGCUGCCUGACAGUCA</b> <b>CUGACCAGAGCUUC</b> CCUUACCAUAACCUUCC (((((((((.....)))))))))	(-42.40)
Mac	GGAGGCUCUGGUUAGGU <b>UCUGGAGGU</b> <b>CAGGGCAUGGUGA</b> <u>U</u> <b>CAGCGGCUGCCUGAUGGUCA</b> <b>CUGCCCAGAGCUUC</b> CCUUACCAUAACCUUCC (((((((((.....)))))))))	(-40.80)
Cns	GAAGGCUCUGGUUAGGU <b>UCUGGAGGU</b> <b>CAGGGCAUGGUGA</b> <u>U</u> <b>CAGCGGCUGCCUGACAGUCA</b> <b>CUGCCCAGAGCUUC</b> CCUUACCAUAACCUUCC (((((((((.....)))))))))	(-36.97)

**Figure 4-11. Sequence alignment of the rs1811399 hairpin**

Abbreviations: H C, human (C allele); H A, human (A allele); Pan, chimpanzee; Gor, gorilla; Pon, orangutan; Mac, macaque and Cns is the consensus sequence. Red text indicates candidate mature microRNA and blue text indicates the candidate seed sequence. Bold and underlined nucleotides are the alleles of rs1811399.

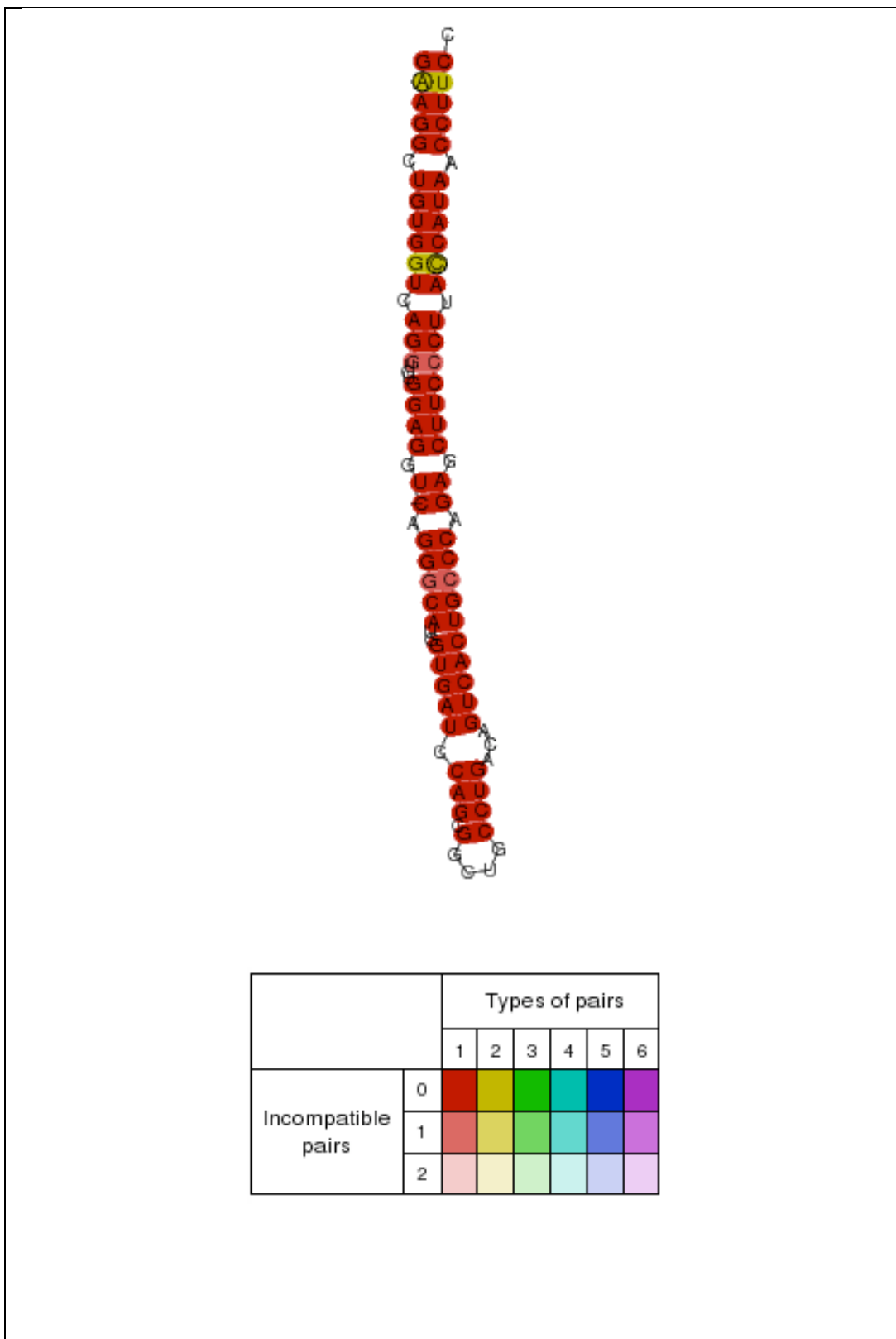
		Types of pairs					
		1	2	3	4	5	6
Incompatible pairs	0	Red	Yellow	Green	Teal	Blue	Purple
	1	Light Red	Light Yellow	Light Green	Light Teal	Light Blue	Light Purple
	2	Very Light Red	Very Light Yellow	Very Light Green	Very Light Teal	Very Light Blue	Very Light Purple



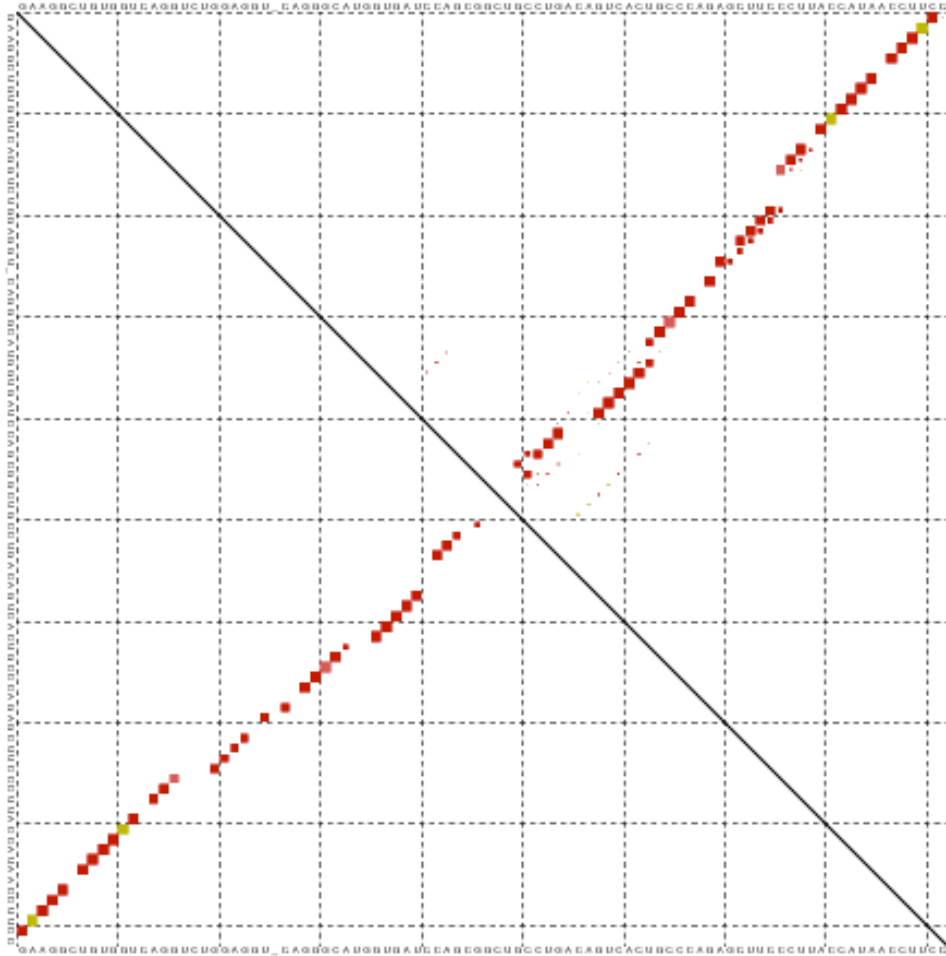
**Figure 4-12. Conservation of structure of the rs1811399 hairpin in primates.**

The upper dot and bracket track shows the conserved consensus hairpin structure in 5 primates. The colour (brown-purple) shows the number of structure-maintaining nucleotide pair changes for the consensus hairpin (0 for brown 6 for purple). The strongest colours mark pairs that maintain the structure in all species. Pairs that are incompatible (don't form hydrogen bonds) but nevertheless do not disrupt the overall hairpin structure are shown with their colour faded. Base changes that fall within bulges etc are indicated by the notched grey bar below the sequences.





**Figure 4-13. Consensus structure of the rs1811399 hairpin in primates.**  
 Conventions are as for figure 4-12 above.



**Figure 4-14. Dot plot for the rs1811399 hairpin.**

Conventions are as for Figure 4-14 above. The diagonal arrangement of dots indicates low diversity for the ensemble of possible structures with the hairpin formation dominant.

In Figure 4-9 and Figure 4-11, it is noteworthy that the formation of the forked hairpin (C allele) has the candidate 3' seed (as for the for the A allele ACAGUCA) now located in an unlikely region of the hairpin, i.e. encroaching into the end loop. it is possible that fork formation due to the presence of the C allele in the rs1811399 hairpin would thus alter the seed of the 3' candidate miR or disfavour the formation of the 3' candidate miR. Altogether.

#### 4.5.3ii *EN2* rs1861973 hairpin

For the *EN2* sequence there is conservation of an irregular structure amongst primates. The results for this SNP are shown below in the tables and figures.

**Table 4-5. *EN2* sequence fragment.**

<i>EN2</i> sequence fragment	Notes
300bp fragment starting 101478854 AAACCCAGAGGCGAGGTCACCACTCCCTGCCA <b>g</b> TGG CCTTGCCCCCTTCTTCCCCCACAGGGAACGCCAGGG GGTTGAGCCTCTTATCACCAAAAAGAAACTGATGAC ACTTCCCTCCTTCTG <b>C</b> TCTCCTCCCTCTGCCCTTTC CCCATGGATAGCAGGTCCTAG <b>AAGC</b> CTTACAGCGAC CCTG <b>C</b> CCAAAACCTGGGGCAGGTCACAGGGAGAAG GCCAGGTCAGGTTCATAAAGTCTGAATCCCAGTTGGG AGGCACAGTGGGGAGGGTCAGAAGTGGACCTGGACA AGGTCAGCTGGG PPV >0.47 for the predicted DROSHA processing site with the T allele  PPV <0.3 for the predicted DROSHA processing site with the C allele  N.B. the predicted pre-mir containing the C allele does not support hairpin formation.	<b>g</b> = rs1861972 G/A <b>C</b> = rs1861973 C/T  Grey = fragment for <i>in silico</i> DROSHA.  Pink = the limit of the long hairpin structure  Grey = candidate pre-miRNA as predicted by Interagon SVM  <b>GAAGCC</b> = position of central loop  <u>ACCCTGT</u> = seed of MIR10b

**Table 4-6. Statistics for the *EN2* hairpin containing the SNP rs1861973.**

	<i>EN2</i> hairpin rs1861973		
	C Allele rs1861973 (irregular structure)	T Allele rs1861973 (hairpin)	hsa-let-7f-2
Optimal MFE structure	-39.5 kcal/mol	-42.50 kcal/mol	-42.1 kcal/mol
MFE ensemble	-42.42 kcal/mol	-44.34 kcal/mol	-43.85 kcal/mol
Frequency of the MFE structure in the ensemble	0.87 %	5.07%	5.8%
Ensemble diversity	26.92	9.05	6.64
Centroid MFE	-30.74 kcal/mol	-42.4 kcal/mol	-41.90 kcal/mol

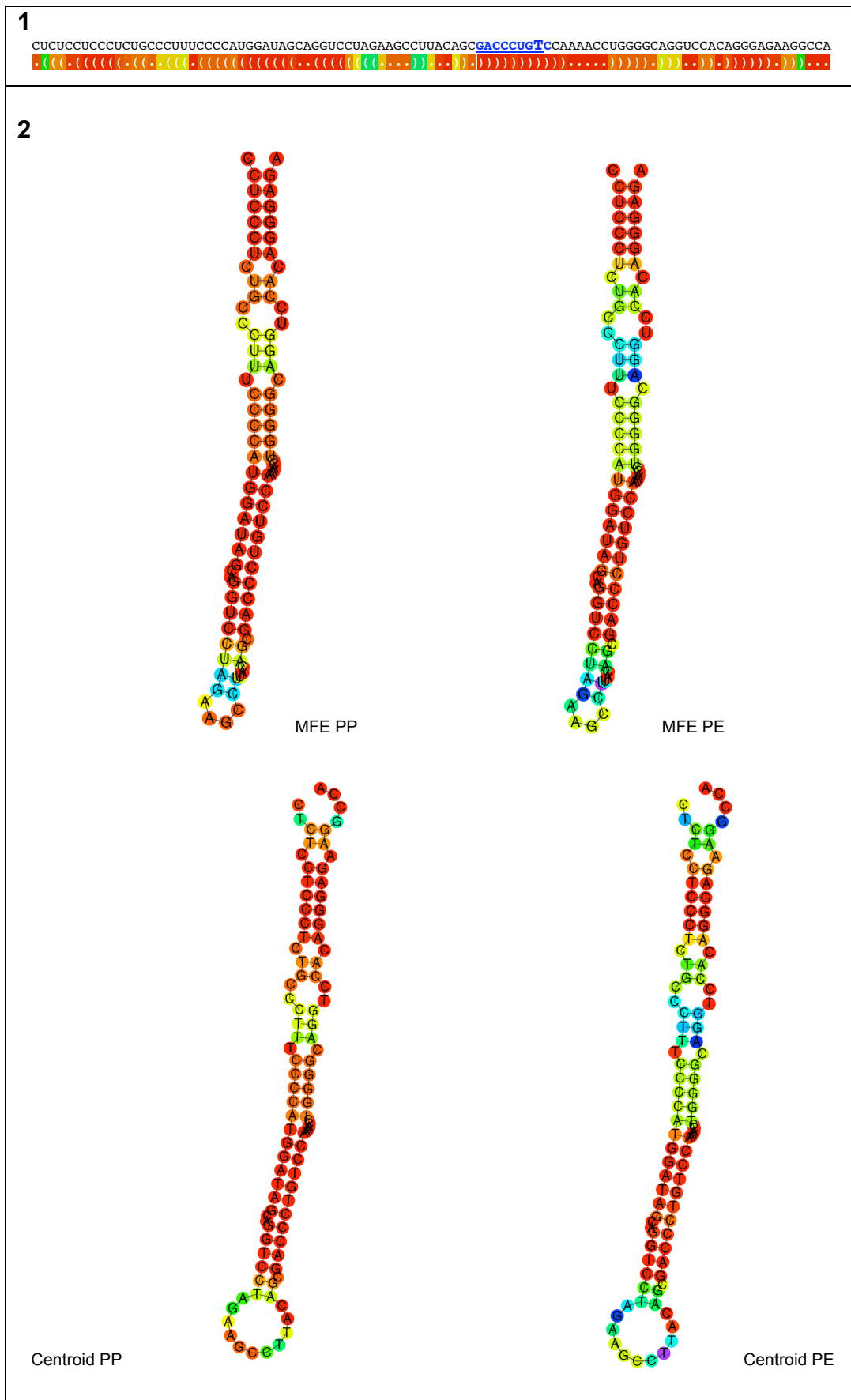
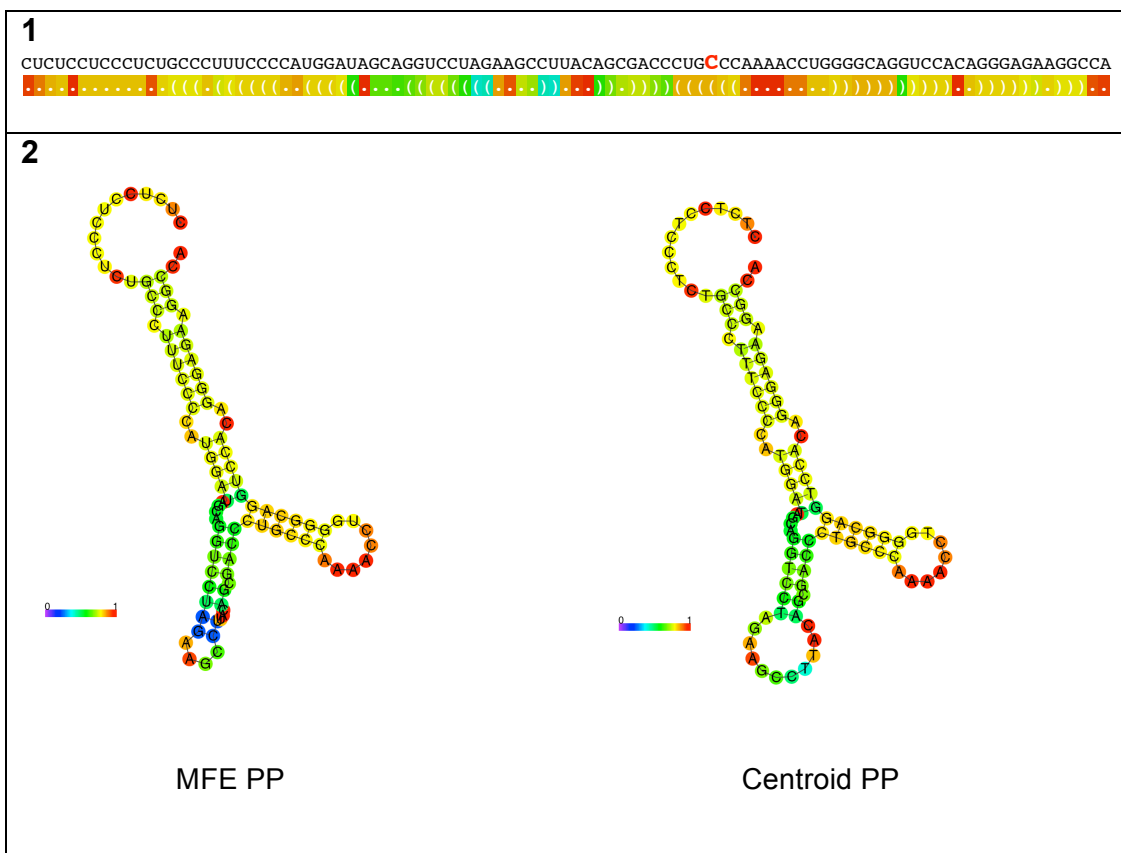


Figure 4-15. The *EN2* hairpin, T allele (RNAfold, Vienna RNA Web Server).

**Figure 4-15 (continued). The *EN2* hairpin, T allele (RNAfold, Vienna RNA Web Server).**

The *EN2* sequence shown in panel 1 and containing the SNP rs1861973 (indicated by **T**) is predicted to form a long stable RNA hairpin. The seeds of miR504, miR10 and miR339 are shown in blue. The sequence is under laid by the secondary structure for the hairpin given in dot and bracket notation. The shading convention and terms are as for *NPAS2* above in Figure 4-7.



**Figure 4-16. The *EN2* hairpin, C allele RNAfold, Vienna RNA Web Server.**

The *EN2* sequence shown in panel 1 and containing the SNP rs1861973 (**C**) is predicted to form an irregular RNA structure of high MFE. The sequence is under laid by the secondary structure given in dot and bracket notation. The shading convention and terms are as for *NPAS2* in Figure 4-7 above.



**Table 4-7. Statistics for the conservation of RNA structure around SNP rs1861973 analysed with the Vienna Web Server.**

Length	113
RNAz 1.0 Sequences	6
Columns	113
Reading direction	forward
Mean pairwise identity	75.99
Mean single sequence MFE	-42.72
Consensus MFE	-28.33
Energy contribution	-29.15
Covariance contribution	0.82
Combinations/Pair	1.40
Mean z-score	-3.79
Structure conservation index	0.66
SVM decision value	2.47
SVM RNA-class probability	0.994367
Prediction	RNA

#### 4.5.3iii *PER1* rs885747 hairpin

The sequence fragment containing intron 12 of *PER1* was analysed with the Vienna RNA analysis tools. The results are shown in the annotated tables and Figures.

**Table 4-8. Sequence and SNP details of intron 12 *PER1*.**

<i>PER1</i> sequence fragment	Notes
300bp fragment starting 101478854	
CCAGCTCCCTCCY TGGACACTGATATCCAGGAGCTGTCA	Y = rs3027181 C/T
GAGCAGATCCACCGGCTGCTGCTGCAGGTGAGAGTAGCG	G = rs885953 G/C
GA S AGGGAGCCTGGGAGGTGAGAAAAGGTGTGGGAAGC S	S = rs885747 C/G
GGGTCAAGCCATCTAACCTGCCCTCTCCCTGCTGCAGCC	
CGTCCACAGCCCCAGCCCCAC R GGACTCTGTGGAGTCGG	R = rs35859165 G/A
CGCCGTGACATCCCCAGGCCCTCTCCACAGCCCTGGGTC	
CTCCAGTGATAGCAACGGGGGTGATGCAGAGGGGCTGG	
GCCTCCTGCGCCAGTGAGTGACCTGCT	
	Grey = intron 12.

**Table 4-9. Vienna RNA Web Server RNAfold output statistics for the *PER1* hairpin containing the autism-associated SNP rs885747 and rs885953 C.**

	<b><i>PER1</i> hairpin</b> (Values when rs885953 is C)		
	<b>C Allele rs885747</b>	<b>G Allele rs885747</b>	<b>hsa-let-7f-2 (Forked MIR)</b>
Optimal MFE structure	-41.2kcal/mol	-39.20 kcal/mol	-42.1 kcal/mol
FE ensemble	-41.62 kcal/mol	-40.44 kcal/mol	-43.85 kcal/mol
Frequency of the MFE structure in the ensemble	9.93 %	13.44%	5.8%
Ensemble diversity	7.77	8.13	6.64
Centroid MFE	-40.90 kcal/mol	-42.4 kcal/mol	-41.90 kcal/mol

**Table 4-10. Vienna RNA Web Server RNAfold output statistics for the *PER1* hairpin containing the autism-associated SNP rs885747 and rs885953 G.**

	<b><i>PER1</i> hairpin</b> (Values when rs885953 is G)		
	<b>C Allele rs885747</b>	<b>G Allele rs885747</b>	<b>hsa-let-7f-2 (Forked MIR)</b>
Optimal MFE structure	-38.10kcal/mol	-35.9kcal/mol	-42.1 kcal/mol
FE ensemble	-39.87kcal/mol	-37.5 kcal/mol	-43.85 kcal/mol
Frequency of the MFE structure in the ensemble	5.69%	6.92 %	5.8%
Ensemble diversity	31.93	33.14	6.64
Centroid MFE	-24.78kcal/mol	-21.93 kcal/mol	-41.90 kcal/mol

Further examination with the structural conservation tools of the Vienna RNA Web Server indicated no hairpin structural conservation for this sequence in primates.



**Table 4-11 Structural conservation for the PER1 hairpin**

Location	0 – 105
Length	105
RNAz 1.0 Sequences	5
Columns	105
Reading direction	forward
Mean pairwise identity	89.33
Mean single sequence MFE	-43.28
Consensus MFE	-34.74
Energy contribution	-35.38
Covariance contribution	0.64
Combinations/Pair	1.10
Mean z-score	-1.38
Structure conservation index	0.80
SVM decision value	-0.01
SVM RNA-class probability	0.526804
Prediction	RNA

#### 4.5.4 Conclusion

This analysis of the autism-associated SNPs with the Vienna RNA Web Server has strengthened the possibility that the hairpin containing the SNP rs1811399 in *NPAS2* is a candidate microRNA. For *EN2 rs1861973* only the T allele supports hairpin formation. Structural conservation analysis suggests an irregular structure is conserved at this locus. The results for *PER1* do not further support a candidate microRNA in the 12<sup>th</sup> intron.

Confirmation of the positive results would have to come from experimental work in cell lines or tissue samples but these findings are encouraging especially for rs1811399 given the lack of any other clear indication of possible functionality for this SNP. The GLIDERS results (Table 3-9) showing a group of SNPs in linkage disequilibrium with rs1811399 further suggests this SNP is implicated in the association signal specifically, perhaps lending further support in favour of the functionality of the rs1811399 hairpin.

#### 4.5.5 Analyzing Self Containment with the RNA Self Containment Web Server.

##### 4.5.5i Introduction.

The RNA Self Containment Web Server (Table 3-1) calculates a Self-Containment Index (SC) (Lee and Kim 2008) that is a measure of the robustness of RNA hairpins to changes in their surrounding sequence context, which is proposed as a hallmark of structural modularity. SC values range from 0.0 (no self containment) to 1.0 (completely self contained). Based on empirical results, a typical mRNA sequence has an SC value of approximately 0.54 whereas microRNA hairpins have an average SC of 0.9 (Lee and Kim 2008).

##### 4.5.5ii Method.

Using the sequence fragments encoding the long hairpins containing SNPs rs1811399, rs885747 and rs1861973 as above. Each sequence was submitted in turn to the server and the data output for each hairpin noted. This was repeated for both alleles of the hairpins containing the SNPs.

##### 4.5.5iii Results.

The results are shown in Table 4-11. For the *NPAS2* hairpin the SC index is keeping with the values of microRNAs (~0.9). For the *PER1* and *EN2* hairpins the SC index does not support the notion that they are microRNA related hairpins.

**Table 4-12. Self Containment Index data for the *EN2*, *NPAS2* and *PER1* hairpins.**

Gene	Hairpin containing SNP	SC index
<i>NPAS2</i>	rs1811399 = C	0.91
	rs1811399 = A	0.87
<i>PER1</i>	rs885747 = C	0.75
	rs885747 = G	0.23
<i>EN2</i>	rs1861973 = T	0.48
	rs1861973 = C	0.50

#### 4.5.5iv Conclusion.

The results for the *NPAS2* hairpin are in keeping with the values expected for MicroRNAs. This is also reflected in the results of the Vienna RNA Web Server. For the *PER1* and *EN2* hairpins this test together with the results of the Vienna RNA Web Server indicates that these hairpins are not microRNA candidates.

#### 4.5.6 Searching MirBASE for sequence homology between the *NPAS2*, *PER1* and *EN2* hairpins using BLAST and SSEARCH tools.

##### 4.5.6i Introduction.

The MirBASE database is a searchable database of miRNA sequences that have been published that are curated together along with details of annotation of MicroRNAs. The search tools available in MirBASE include BLAST and SSEARCH functions that in this case were used to compare the hairpins containing rs1811388, rs885747 and rs1861973 with the sequences of all known microRNAs stored in the database.

4.5.6ii Method.

The hairpins, each containing the major and minor alleles of the autism-associated SNPs were used in the searches. The sequences were entered into the MirBASE search window and the appropriate search tool selected dependent on the string length of the query. For long hairpins, BLASTN and stem loop sequence were selected, while SSEARCH and “mature sequence” were used when searching with a ~22nt candidate mature sequence taken from each arm of the candidate microRNA under investigation.

4.5.6iii Results.

The results of the searches are shown below in Figure 4-19.

<b><i>NPAS2</i></b>	
score: 91 evalue: 0.23	
UserSeq	
35	uggugauacagcggcugccugacagucacugcccagag 72
40	uggggau <u>guugcagcugccugggagugacuuc</u> acacag 77
	hsa-mir-1301
<b><i>PER1</i></b>	
score: 99 evalue: 0.053	
UserSeq	
83	ggugugggaagccgggucaagccaucuaaccugcccucuccugc 39
32	ggugcggaugcagggugaggccuugu <u>accuuggcccagcccagc</u> 76
	hsa-mir-3620

**Figure 4-19. Best matches from MirBASE searches with the *EN2*, *NPAS2* and *PER1* hairpins.**

#### 4.5.6iv Conclusion

Although alignments could be made between the *NPAS2* query hairpin and known microRNAs the match is weak and the position of the mature MIR in relation to the homology suggests that this alignment is probably not biologically significant. For example in Figure 4-19 the highest scoring alignment matched the mature microRNA region of hsa-mir-1301 with the loop region of the *NPAS2* candidate microRNA. Also, the hsa-miR-1301 seed matches the region immediately upstream of the loop in the candidate MIR thus discounting these seed as functional in the *NPAS2* candidate MIR.

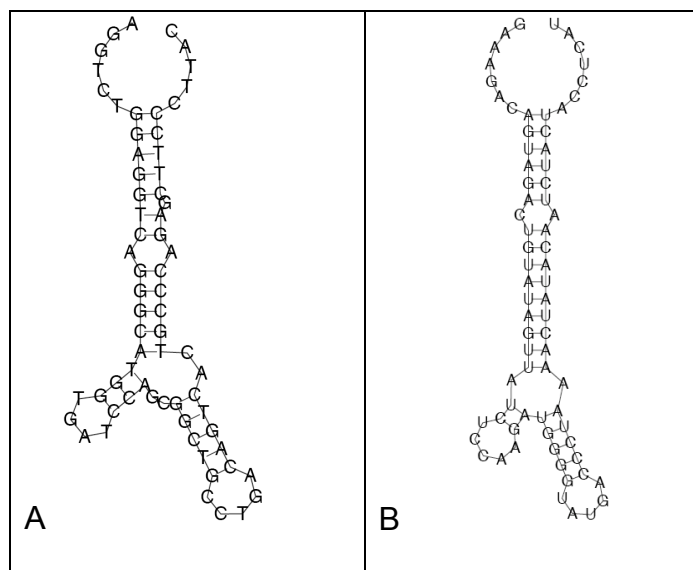
**Table 4-13. Summary of the attributes of the hairpin structures.**

	rs1811399 <i>NPAS2</i> hairpin	rs885747 <i>PER1</i> hairpin	rs1861973 <i>EN2</i> hairpin
Hairpin length >60nt	yes	yes	yes
Energy = or < 35kJ/mole	yes	yes	yes
Positive DROSHA score	yes	no	yes
Self Containment > 0.9	yes	no	no
Vienna RNA Web Servers Predicts hairpin conservation	yes	no	no

#### 4.5.7 General conclusion from the RNA secondary structure analysis.

The RNA secondary structure analysis, summarized in Table 4-13, indicates there are grounds to consider further investigation of the *NPAS2* hairpin in terms of a candidate microRNA. The SNP rs1811399 clearly affects the predicted structure of the *NPAS2* hairpin but the data does not allow determination of whether the A, C or both forms of the rs1811399

hairpin are the more likely to be processed into a silencing molecule. This is due to the following factors: bona fide microRNAs may have short forked ends (e.g. hsa-let-7f-2); the *in silico* DROSHA score is better for the A hairpin than the C hairpin (Figure 4-20); the MFE score is better for C form than the A hairpin (Table 4-3) thus taken together, the data suggest there is no outright winner between the two possible structures for the *NPAS2* candidate microRNA.



**Figure 4-20. *NPAS2* candidate microRNA and hsa-let-7f-2 compared.**

Similar structures for the forked microRNA hsa-let-7f-2 and the candidate hairpin containing the SNP rs1811399.

However consideration of the hairpin fine structure (Figure 4-8 and Figure 4-10) suggests that the position, and thus, the target complementarity of the seed sequence, may differ between the two structures.

Ensembl genotyped populations show that there is a strong bias against the homozygous CC genotype in all the Ensembl populations, even though the ancestral allele is C. The association study results (Table 2-4) also indicate that the significant association of the C allele of

rs1811399 with autism derives from an under-transmission of the A allele rather than the over-transmission of the C allele from parents to their children with autism. Tentatively these results suggest that the A allele of rs1811399 (long hairpin) confers some advantageous or protective effect in contemporary neurotypical populations that is not the case for the ancestral, autism-associated C allele.

It is interesting that the revised top targets of the 3' candidate seed (ACAGUCA) contain the dosage dependent autism genes, *RAI1* and *DGCR8*. Possibly the C allele of rs1811399 affects the availability of the 3' candidate miR only, by moving this sequence into the unfavourable position in the end loop (Figure 4-14). Loss of both copies of the 3' mirR (corresponding to the CC genotype) might thus disturb the levels of *RAI1* and *DGCR8* in cells expressing *NPAS2* and with a CC genotype. Also, the most likely 5' seed (and the next 3 seeds down stream) present *CLOCK* amongst their top conserved predicted targets, suggesting a possible mechanism for subduing *CLOCK* levels in cells expressing *NPAS2*. Lack of this possible functionality could thus cause undesirable *CLOCK* and *NPAS2* co-expression.

The evidence for the *PER1* hairpin being a candidate microRNA is weaker by not satisfying the DROSHA prediction, self-containment, or the Vienna Web server structural conservation tests. Nevertheless the intron is predicted to form a symmetrical hairpin in the presence of certain combinations of allele of the SNPs rs885953 and rs885747. These results, together with those of the *EN2* hairpin, are against the *PER1* and *EN2* hairpins being candidate microRNAs in the conventional sense.

Nevertheless, and especially in the case of the *EN2* hairpin, the structural analysis shows how SNPs might create *de novo* hairpins out of irregular RNA structures. Such hairpins, if processed by DROSHA etc. might have advantageous or deleterious effect. It is interesting to note that the *EN2* hairpin contains a candidate miR with a seed sequence identical to the seed of *MIR10*. The involvement of *MIR10* in the same developmental pathway as *EN2*, supports the notion that the formation of a novel hairpin, a novomir say, in the transcript derived from this *EN2* intron may interfere with brain development that is reliant on the expression of *EN2*. Notably the candidate seeds surrounding the SNP rs1861973 in *EN2* are those of the *bona fide* microRNAs *MIR10* (*BDNF*), *MIR339* (*CACNA1C SHANK3*), and *MIR504* (*NRXN1*). The genes in the brackets are amongst the top ten predicted targets for each of these microRNAs and each gene is strongly implicated in neuronal function/development and autism.

The frequency of genotypes of the SNP rs1861973 in neurotypical populations, support the hypothesis that the TT homozygote may be linked to a pathological phenotype. Ensembl data on genotypes shows that the TT genotype is the least frequent of the four possible genotypes in all the populations assessed and is absent in a number of populations (Table 4-14).

Additionally, *EN2* is a target of *MIR504* (Figure 4-3) thus expression of the rs1861973 hairpin (T allele dependent) could lead to the down regulation of *EN2* levels. This is in keeping with the finding of up regulation of *EN2* expression with the C allele of rs1861973 (Benayed et al. 2009).



**Table 4-14. Genotypes for the autism-associated SNP rs1861973 in neurotypical populations.**

<b>Population</b>	<b>C C</b>	<b>T C</b>	<b>T T</b>
Japanese, (Tokyo)	79	6	0
Han Chinese in Beijing	36	5	0
Chinese in Metropolitan Denver	71	14	0
Han Chinese in Beijing	32	11	0
Yoruba Ibadan, (Nigeria)	69	37	6
Mexican ancestry in Los Angeles	28	19	2
Utah, N. and W. European ancestry	61	47	5
Luhya in Webuye, Kenya	53	27	10
Toscans in Italy	43	39	6
Gujarati Indians in Houston	41	40	7
African ancestry in Southwest USA	18	26	5

This hypothesis is not refuted by recent results that find that the autism-associated C allele of rs1861973 is nevertheless, associated due its protective effect in the disorder (Yang et al. 2010). The presence in the gene pool of a dosage dependent pathological allele (T rs1861973) that affects molecular pathways already compromised in autism, could thus satisfy the otherwise complicated results for this gene in autism (Petit et al. 1995; Benayed et al. 2005; Brune et al. 2008; Wang et al. 2008). This analysis also suggests that the conservation of the irregular structure linked to the C allele of rs1861973 holds in abeyance the possibility of pathogenic hairpin formation with the T allele. In autism, gene variants conferring risk may compromise a molecular pathway containing *EN2* that becomes critically overburdened by the presence of the T allele of rs1861973.

## **4.6 Analysis of the results of three, autism microRNA expression studies in relation to clock gene targets.**

### 4.6.1 Introduction

If clock genes play a role in autism aetiology, then any microRNAs that are dysregulated in autism might feature clock genes prominently amongst their targets. Three studies to date have looked at microRNA expression in autism: Abu-Elneel et al. (2008), Talebizadeh et al. (2008) and Sarachana et al. (2010). The latter, performed pathway analysis that showed the circadian rhythm pathway was significantly associated with their data set of microRNAs dysregulated in autism. However, Abu-Elneel et al. (2008) did not undertake molecular pathway analysis, nor were specific circadian gene targets identified by Sarachana et al. (2010).

Target prediction is based on empirical sequence matching between microRNA seeds and matching sequences in the 3' untranslated region of genes. The prediction will of course generate false positive results, as the biological validity of the microRNA/target interaction will depend on the co-expression of the micro RNA and the target. In general this aspect is not currently possible to model bioinformatically. However, the target predictions are resistant to false negative results and thus inclusion of clock genes among the targets of the autism microRNAs could indicate that clock gene dysregulation in autism could (at least in part) be caused by the dysregulation of clock gene targeting microRNAs.

#### 4.6.2 Method

The microRNAs from each of the three studies were listed under the headings of El Neel, Talebizadeh and Saracharna. For each microRNA the current location was found with Ensembl and the location cross referenced with the Autism Chromosome Re-arrangement Database (ACRD, Table 3-1) to see if any of these microRNAs were to be found in insertions or deletions found in autistic individuals.

Using MirBASE the sequence of each microRNA was copied and the seed sequence entered into the TargetScan web server. The output file containing the predicted targets for each microRNA was then searched with a list of clock/clock related genes: *ARNTL*, *CLOCK*, *CRY1*, *CRY2*, *CSNK1E*, *NPAS2*, *NR1D1*, *PER1*, *PER2*, *PER3*, *RORA* and *SIRT1*.

#### 4.6.3 Results

The results are given in the following tables.

**Table 4-15. El-Neel. MicroRNAs up-regulated.**

MIR Up-regulated	Chr.	Clock gene target	ACRD
hsa-mir-484	16 p13.11	<i>PER1</i>	no
hsa-mir-21	17 q23.1	no	no
hsa-mir-212	17 p13.3	no	no
hsa-mir-23a	19 p13.13	no	no
hsa-mir-598	8 p23.1	no	no
hsa-mir-129	7 q32.1	no	yes
hsa-mir-431	14 q32.2	no	no
hsa-mir-7	9 q21.32	<i>CRY2</i>	no
hsa-mir-15a	13 q14.2	<i>CLOCK</i>	no
hsa-mir-27a	19 p13.13	no	no
hsa-mir-15b	3 q25.33	<i>CLOCK</i>	no
hsa-mir-148b	12 q13.13	<i>CLOCK</i>	no
<b>hsa-mir-132</b>	17 p13.3	<b><u>Circadian regulator MIR</u></b>	no
hsa-mir-128	2 q21.3	<i>SIRT1</i>	no
hsa-mir-93	7 q22.1	<i>CLOCK, NPAS2, CRY2</i>	no

**Table 4-16. El-Neel. MicroRNAs down-regulated.**

<b>MIR Down-regulated</b>	<b>Chr.</b>	<b>Clock gene target</b>	<b>ACRD</b>
hsa-mir-93	7 q22.1	<i>CLOCK, NPAS2, CRY2</i>	no
hsa-mir-106a	X q26.2	<i>CRY2, CLOCK, NPAS2</i>	no
hsa-mir-539	14 q32.31	no	no
hsa-mir-652	X q23	no	no
hsa-mir-550	7 p14.3	no	no
hsa-mir-432	14 q32.2	no	no
hsa-mir-193b	16 p13.12	no	no
hsa-mir-181d	19 p13.13	<i>SIRT1</i>	no
hsa-mir-146b	10 q24.32	no	no
hsa-mir-140	16 q22.1	no	no
hsa-mir-381	14 q32.31	no	no
hsa-mir-320a	8 p21.3	<i>NPAS2, PER2 CSNK1E</i>	no
hsa-mir-106b	7 q22.1	<i>CRY2, CLOCK, NPAS2</i>	no

**Table 4-17. Talebizadeh. MicroRNAs up-regulated.**

<b>MIR Up-regulated</b>	<b>location</b>	<b>Clock gene target</b>	<b>ACRD</b>
hsa-mir-23a	19 p13.13	no	no
hsa-mir-23b	9 q22.32	no	no
<b>hsa-mir-132</b>	17 p13.3	<b><u>Circadian regulator MIR</u></b>	no
hsa-mir-146a	5 q34	no	no
hsa-mir-146b	10 q24.32	no	no
hsa-mir-663	20 p11	no	no

**Table 4-18. Talebizadeh. MicroRNA down-regulated.**

<b>MIR Down-regulated</b>	<b>location</b>	<b>Clock gene target</b>	<b>ACRD</b>
hsa-mir-92	13 q31.3	<i>PER2</i>	no
hsa-mir-320	8 p21.3	<i>PER2, NPAS2</i>	no
hsa-mir-363	X q26.2	<i>PER2</i>	<u>no</u>

**Table 4-19. Saranchana. MicroRNAs up-regulated.**

<b>MIR UP-regulated</b>	<b>location</b>	<b>Clock gene target</b>	<b>ACRD</b>
hsa-mir-185	22 q11.21	no	no
hsa-mir-103	5 q34	<i>CLOCK, NPAS2, PER3</i>	yes
hsa-mir-107	10 q23.31	<i>CLOCK, NPAS2, PER3</i>	no
hsa-mir-29b	7 q32.3	<i>PER1, PER3,</i>	no
hsa-mir-194	1 q41	<i>CLOCK</i>	no
hsa-mir-524	19 q13.42	<i>NPAS2</i>	<b>yes</b>
hsa-mir-191	3 p21.31	no	no
hsa-mir-376a	14 q32.31	no	no
hsa-mir-451	17 q11.2	no	no
hsa-mir-23b	9 q22.32	no	no
hsa-mir-195	17 p13.1	<i>CLOCK</i>	no
hsa-mir-342	14 q32.2	no	<b>yes</b>
hsa-mir-23a	19 p13.13	no	no
hsa-mir-186	1 p31.1	no	no
hsa-mir-25	7 q22.1	<i>PER2</i>	no
hsa-mir-519c	19 q13.42	no	<b>yes</b>
hsa-mir-346	10 q23.2	no	no
hsa-mir-205	1 q32.2	no	no
hsa-mir-30c	1 p34.2	<i>CLOCK, PER2, PER3</i>	no
hsa-mir-93	7 q22.1	<i>CRY2, NPAS2</i>	no
hsa-mir-186	1 p31.1	no	no
hsa-mir-106b	7 q22.1	<i>CRY2, CLOCK, NPAS2</i>	no

**Table 4-20. Saranchana. MicroRNAs down-regulated.**

<b>MIR Down-regulated</b>	<b>location</b>	<b>Clock gene target</b>	<b>ACRD</b>
hsa-mir-182	7q32.2	<i>CLOCK</i>	no
hsa-mir-136	14q32.2	no	no
hsa-mir-518a	19q13.42	no	<b>yes</b>
hsa-mir-153	2q35	no	no
hsa-mir-520b	19 q13.42	<i>RORA</i>	<b>yes</b>
hsa-mir-455	9q32	no	no
hsa-mir-326	11q13.4	<i>CRY2</i>	no
hsa-mir-199b	9q34.11	<i>NPAS2</i>	no
hsa-mir-211	15q13.3	<i>SIRT1</i>	<b>yes, multiple</b>
<b>hsa-mir-132</b>	17p13.3	<b><u>Circadian regulator MIR</u></b>	no
hsa-mir-495	14q32.31	<i>CLOCK PER2</i>	no
hsa-mir-16	13q14.2	<i>CLOCK</i>	no
hsa-mir-190	15q22.2	<i>CLOCK</i>	no
<b>hsa-mir-219</b>	6p21.32	<b><u>Circadian regulator MIR</u></b>	no
hsa-mir-148b	12q13.13	no	no
hsa-mir-189/24-1	9q 22.32	<i>PER1 PER2 CRY2</i>	no
hsa-mir-133b	6p12.2	<i>SIRT1</i>	no
hsa-mir-106b	7q22.1	<i>CRY2, CLOCK, NPAS2</i>	no
hsa-mir-367	4q25	<i>PER2</i>	no
hsa-mir-139	11q13.4	no	no

**Table 4-21. Overlap between the results of El-Neel, Talebizadeh and Saranchana.**

<b>Common to the 3 studies</b>	<b>Clock gene targets</b>	<b>Common to 2 studies</b>	<b>Clock gene targets</b>
hsa-mir-132	<b>Circadian regulator</b>	hsa-mir-106b	<i>CRY2, CLOCK, NPAS2</i>
hsa-mir-23a	no	hsa-mir-146b	no
		hsa-mir-148b	no
		hsa-mir-23b	no
		hsa-mir-93	<i>CLOCK, NPAS2, CRY2</i>

#### 4.6.4 Conclusion

Most notably, all three studies find altered expression of the circadian microRNA *MIR132* and *hsa-mir-23a* that targets *CLOCK* at a nonconserved site in the 3'UTR. Secondly, clock genes contain 3'UTR target sites for almost all of these microRNAs and conserved sites as shown in the tables. All the microRNAs that are common to two studies also target core clock genes at unconserved sites and conserved sites as shown. *MIR132* and *MIR106b* are also reported to show altered expression patterns in schizophrenia (Kim et al. 2010).

The results of this meta-analysis show that circadian microRNAs are repeatedly found to be dysregulated in autism cases, which is compatible with a role for a disturbance of circadian clock function in autism. The involvement of *MIR132* in the circadian molecular clock (Cheng et al. 2007) implies that dysregulation of this microRNA could cause dysregulation of the core clock components. Also *MIR 182* is shown to be clock regulated and *CLOCK* is a potential target of this microRNA (Xu et al. 2007; Yang et al. 2008). It is noteworthy that different microRNAs were dysregulated in different autistic individuals and some of these microRNAs

are located in microdeletions found in individuals with autism, according to the ACRD. Each of these ACRD microRNAs target core clock genes.

Supposing the dysregulated microRNAs do contribute to autism, this could be seen as an example of how a phenotype that is linked to the disruption of a biological pathway, might be difficult to capture experimentally in terms of genotype. In this case, different genetic variations, in different microRNAs, in different individuals, might contribute to a similar outcome (circadian disruption) if the microRNAs have clock gene targets in common. Such a mechanism might be additive to variation in clock genes specifically.

## 5 Discussion

Disruption of circadian rhythms accompanies several neuropsychiatric disorders including autism (Barnard and Nolan 2008). Whether and to what degree however, altered circadian rhythms contribute to the causes or progress of these disorders is currently unclear. Circadian rhythms in mammals are generated by cyclic activity of transcription factors; the products of so called clock genes and the work described in this thesis begins testing the hypothesis that clock genes are implicated in autism.

### 5.1 Summary of the main findings.

This thesis reports that SNPs in the clock genes *NPAS2* and *PER1* are associated with autistic disorder. Further analysis of the autism-associated SNPs shows that:-

- The SNP rs34705978 (*NPAS2*) is a polymorphic nucleotide in a differentially methylated control element. The autism-enriched allele denies the possibility of methylation-directed regulation at this CpG pair.
- The most significant autism-associated haplotype in *NPAS2* contains a conserved RORA/NR1D1 binding site, a likely *NPAS2* regulatory motif.
- The SNP rs1811399 in *NPAS2* is located within a region of the gene that gives rise to an RNA transcript that is predicted to form a hairpin structure that is reminiscent of a pre-microRNA.
- The autism enriched allele of SNP rs885747 (*PER1*) knocks-out a predicted splice site enhancer /suppressor element.
- Autism and prostate cancer share disease associated loci in *PER1* and *NPAS2*.



- Cardinal autism genes have circadian regulatory elements and show diurnal patterns of expression in mammalian prefrontal cortex.

#### 5.1.1 Clock controlled genes include autism-associated genes.

Abnormal patterns of melatonin secretion and altered sleep profiles in autism indicate circadian disruption. Lymphoblastoid cell lines derived from autistic individuals have altered clock gene expression profiles indicating this disturbance is at the level of the circadian molecular clock (Hu et al. 2009a). The results presented in this thesis, support the idea that the circadian disruption seen in autism may stem from variation in the circadian molecular clock.

Evidence is also presented that shows several genes implicated in the disease process leading to autism have conserved clock control elements (E-box, D-Box and RRE) and/or show diurnal levels of expression in mammalian prefrontal cortex. These genes are therefore likely to be regulated by the circadian molecular clock and their typical expression altered by malfunction of the circadian molecular clock. Notably, the major regulator of synaptic plasticity, *MAPK1* (Samuels et al. 2008), is under circadian control along with *SHANK3* and *MET* that are strongly implicated in autism. Similarly, the five genes with genome wide levels of significance for association with autism: *SEMA5A*, *CNTNAP2*, *TAS2R1*, *CDH9* and *CDH10* are to be found in the set of genes with circadian regulatory elements (Kumaki et al. 2008).

#### 5.1.2 The autism-associated clock gene SNPs rs885747, rs6416892 and rs34705978 have predicted function.

The bioinformatic analysis of the SNPs rs885747 (splice site enhancer/suppressor), rs34705978 (SREBF2 binding site) and rs34705978 (allelic methylation target, CpG dinucleotide) indicates functionality leading to possible adverse phenotypic effects. This is in keeping with the low allele frequencies for the autism-associated alleles observed in the general population. The processes linked to each of these predicted functions are each capable of altering gene expression in a tissue-specific manner.

### 5.1.3 Could point mutation generate *de novo* silencing RNAs?

The autism-associated SNPs rs1811399 in *NPAS2*, rs1861973 in *EN2* and the psychosis variant rs1344706 in *ZNF804A* are located in intron regions that encode RNA transcripts predicted to form long RNA hairpin structures that are reminiscent of primary microRNAs. Analysis of these candidate microRNA hairpins tentatively supports the notion that the *NPAS2*, *EN2* and *ZNF804A* hairpins could be processed into silencing RNA molecules. If these candidate hairpins do generate microRNAs, why have they not been found in microRNA screens and previous bioinformatics discovery? This may be because: Firstly, bioinformatics discovery of microRNAs has primarily centered on looking for conserved sequences, an approach that would miss novel and evolutionarily recent hairpins. Secondly, microRNAs that are expressed at low levels in a tissue specific manner or at a particular developmental time point could evade tissue and developmental stage specific laboratory-based screens. Thirdly, the SNPs in these hairpins (that disrupt the hairpin structure) would militate against them being found consistently since individual genotype of the sample would

determine the presence of the hairpin. An interesting perspective gleaned from these results is that point mutation can generate hairpins from otherwise irregular RNA structures. This suggests that the potential for evolutionary effects from point mutation in non-coding regions may be underestimated if certain SNPs induce *de novo* hairpins that could be processed into functional gene silencing molecules.

The structure of the *NPAS2* rs1811399 hairpin is conserved between primates, it passed all of the bioinformatics quality controls and therefore received the most scrutiny given the available time and resources. Analysis of the predicted targets of the 5'arm of the *NPAS2* hairpin appears to match the biological function of *NPAS2* (its host gene). Two circadian genes *CRY2* and *CLOCK* are prominent predicted targets of this candidate microRNA, suggesting a role for this candidate microRNA in circadian rhythm regulation. A mechanism (a microRNA in *NPAS2* targeting *CLOCK*) that diminished *CLOCK* levels at the same time that expression levels of *NPAS2* are high would make sense as the levels of *CRY2* and *NPAS2* are normally out of phase, as the *CRY2* complex is an inhibitor of the *CLOCK/NPAS2/ARNTL* complex.

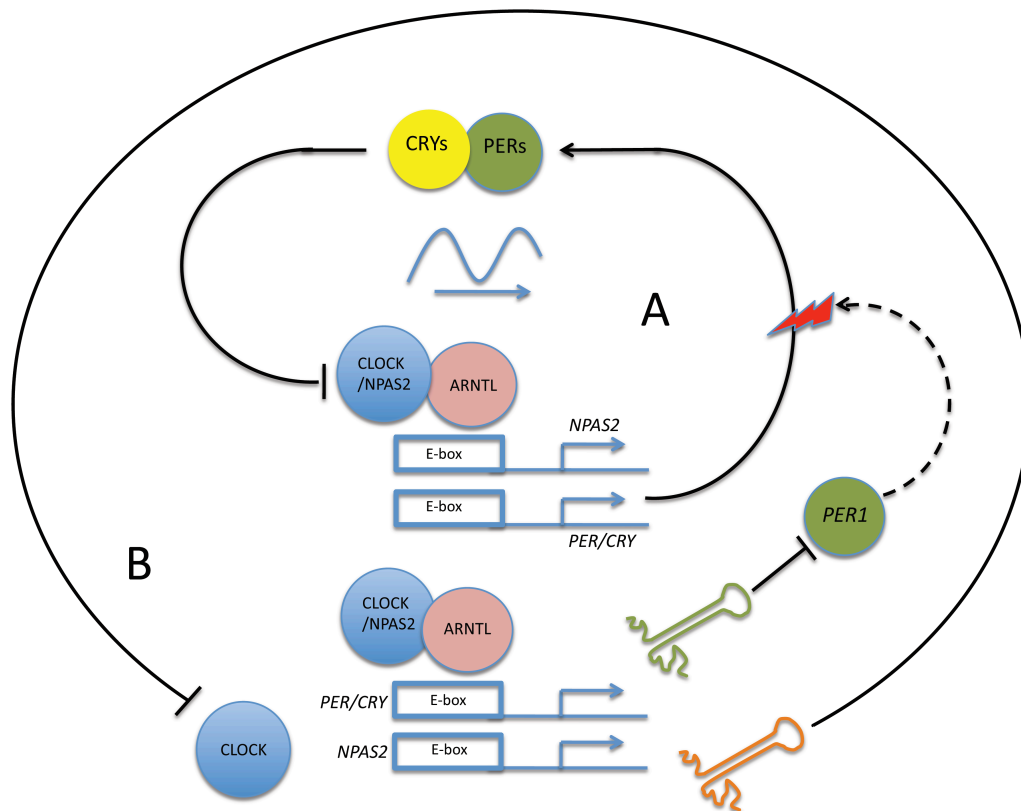
In relation to cell type, *NPAS2* and *CLOCK* are able to serve the same function in the circadian pacemaker (Reick et al. 2001) but in some tissues may be required to be exclusive with regards to their expression patterns. In tissues reliant on *NPAS2/CLOCK* functional exclusivity, disruption of this candidate microRNA could force co-expression of these proteins. Added to this, the seed GGAGAGG of the autism-associated C allele hairpin (rs885747 in *PER1*) targets *PER1*, with the possible implication of shortening the circadian cycle and denying *PER1* dependent pleiotropic

effects (Cermakian et al. 2001). It is currently unclear why clock gene paralogues should show characteristic expression patterns in the neurotypical brain although altered clock gene expression patterns appear to underpin circadian disruption in several neuropsychiatric disorders (Barnard and Nolan 2008). Thus forced co-expression of *NPAS2* and *CLOCK* (due to the lack of a clock-targeting microRNA in *NPAS2* and a shortened cycle due to the candidate microRNA in *PER1*) might thus affect memory formation and plasticity; known circadian rhythm correlates (Figure 5-1).

Up-regulation of *NPAS2* with a concomitant down regulation of the targets of the *NPAS2* candidate miRNA may be reflected in altered gene expression in a Parkinson's disease cell line model where up-regulation of *NPAS2* co-occurs with down regulation of neuronal pentraxin receptor (*NPTXR*) and semaphorin 5A (*SEMA5A*) (Anantharam et al. 2007). *NPTXR* is a predicted target of this *NPAS2* candidate microRNA as is autism-associated *SEMA5A* (Weiss et al. 2009). Genetic overlap is indicated for Parkinson's disease and autism (Rzhetsky et al. 2007).

#### 5.1.4 Autism and prostate cancer share disease-associated clock gene SNPs: autism as a male sex hormone/clock gene related disorder.

An outstanding feature of autism is the unexplained predominance of males with the disorder while the reason for male exclusivity in prostate cancer is obvious. The promoting effect of androgens on disease progression in prostate cancer reflects the implication of androgen sensitive molecular pathways in prostate cancer development. The extreme male brain theory of autism proposes that neurotypical brain



**Figure 5-1. Possible effects of candidate microRNAs in *NPAS2*, that targets *CLOCK* and in *PER1* that targets *PER1*.**

The candidate microRNA co-transcribed with *NPAS2* maintains tissue specific exclusivity of *NPAS2* in relation to *CLOCK*, Expression of the candidate microRNA in *PER1* leads to diminished *PER1* levels, shortening the circadian cycle (Cermakien et al. 2001) and possibly impacting on circadian plasticity (Mehnert et al. 2007)

tissue has sex-specific sensitivities to androgens and that these are altered in autism (Baron-Cohen 2002).

The *NPAS2* SNP rs2305160 (Ala394thr) is a biomarker for raised serum testosterone (Chu et al. 2008). It is possible therefore that the near-by autism-associated SNP rs34705978 (Nicholas et al. 2007) gains its significant association with autism through linkage disequilibrium with other functional common variants in *NPAS2* e.g. rs2305160 that are

associated with testosterone regulation. However, this study tested this androgen biomarker SNP (rs2305160) in autism (see Table 2-3) and although it showed a trend for association and was the 6<sup>th</sup> best p value of all the SNPs surveyed ( $p < 0.08$ ) it did not reach the  $p < 0.05$  level of significance. It is noteworthy that sex hormone levels are under circadian control (Vitzthum et al. 2009) and NPAS2 and PER1 interact with the androgen receptor (AR) (Mukhopadhyay et al. 2006; Cao et al. 2009). A possible explanation could be that although variants of NPAS2 linked to hormone levels may have an effect in autism, intronic SNPs (e.g. rs34705978) in regulatory regions of NPAS2 that directly affect the expression levels of NPAS2 via epigenetic regulation for example, have a stronger effect. Common mutations in AR are not associated with autism (Yan et al. 2004; Henningson et al. 2009).

Intriguingly the SNP rs885747 shown here to be associated with autism and that is predicted to be a splice site signal modulator, is shown to be associated with prostate cancer (Zhu et al. 2009). Together with rs2289591, rs885747 delineates a sequence block that covers the same region of PER1 as does the most significant autism haplotype. PER1 physically interacts with the androgen receptor (AR), inhibits transactivation of the AR in cells over expressing AR and interaction between PER1 and AR reduces the expression of androgen-sensitive genes following stimulation with dihydrotestosterone in LNCaP cells (Cao et al. 2009). The autism/prostate cancer haplotype contains a number of SNPs that enforce alternative splicing events. Also, the 3' end of the autism haplotype roughly coincides with the end point of the shorter of the two isoforms of PER1. This tentatively suggests that androgen sensitive

alternative splicing of *PER1* may be implicated in autism and prostate cancer.

Notably, there is also overlap between *NPAS2* SNPs associated with prostate cancer and autism. The *NPAS2* SNPs most strongly associated with prostate cancer overlap the best *NPAS2* haplotype for autism (Zhu et al. 2009) and the SNP rs17024926 (most strongly associated with prostate cancer) is at the centre of the autism haplotype close to the RRE identified in chapter 3 (1.7kb). Tentatively the SNPs adjacent to this RRE may thus be risk variants for autism and prostate cancer.

The extreme male brain theory of autism (Baron-Cohen 2002) is however difficult to reconcile alongside conditions such as adrenal hyperplasia that manifest high levels of fetal testosterone without increased risk of autism (Barbeau et al. 2009). Nevertheless, fetal testosterone levels are correlated with dimensions of personality that have been described as autistic traits (Chura et al. 2010) and thus it appears that androgen levels in the foetus might determine neural circuitry necessary for forms of cognition described as systematizing that are pathologically exaggerated in autism (Baron-Cohen 2002). As there is no genetic evidence in strong support of implicating the androgen receptor gene in autism (Yan et al. 2004; Henningson et al. 2009) a possible resolution could evoke variations in elements of the androgen receptor interactome e.g. *NPAS2* and *PER1* as risk factors for autism that interact with and could exaggerate effects in androgen related physiology.

#### 5.1.5 How could clock genes affect the frequency of neural oscillators?

##### Tuning of oscillators by clock genes: a hypothesis

A central tenet of the social timing hypothesis (Wimpory et al. 2002) is that clock genes are dysregulated in the autism brain through processes that include epigenetic (methylation) effects that determine tissue specific clock gene expression patterns. Altered clock gene expression patterns in autism are suggested to adversely affect neural oscillators linked to communication and cognition.

In *Drosophila*, circadian clock regulated neuropeptide (*pdf*) secretion causes diurnal remodeling of neuronal architecture in brain clock circuits. In motor neurons, the levels of expression of clock genes *per* and *tim* can determine the degree of neuronal branching (Mehnert et al. 2007; Fernández et al. 2008). Clock genes can regulate developmental timing (Kyriacou et al. 1990) and indications that the molecular circadian clock and the sex determining pathways work together to determine the sex-specificity of certain cell types supports the notion that the circadian clock may play a role in differentiation. Particularly, in ensuring the correct layout of male-typical and female-typical cell patterning within neural tissue (Kadener et al. 2006; Bur et al. 2009; Robinett et al. 2010).

The strong sex-bias towards males in the autism population increases very significantly for the subset of individuals who show no comorbid mental retardation and certain male-typical behaviours seem to be exaggerated to a pathological degree in the disorder (Baron-Cohen 2002). Even though the effects of higher levels of foetal androgens correlate with increased levels of autistic traits in individuals, levels of foetal androgens do not show significant association with a diagnosis of autistic disorder (Knickmeyer and Baron-Cohen 2006). This finding might be interpreted as suggesting foetal androgen levels may be contributory to autism but



require interaction with other pathways to elicit an autism phenotype. The possibility that genetic effects in the androgen system interact with genetic effects in the circadian clock in autism has not been explored.

Nevertheless, evidence against circadian and gender effects in autism notwithstanding, if clock gene variation contributes to autism, an explanation is required of how altered clock genes might affect neural oscillators (operating in milliseconds, seconds and minutes) that are involved in the behavioural phenotypes of autism. The genetic link between the circadian oscillator and other short duration timing phenomena such as the K&H cycle and duration of copulation in *Drosophila* (Kyriacou and Hall 1980; Beaver and Giebultowicz 2004) leaves unexplained the mechanism of how the molecules of the circadian clock might function to regulate such short duration phenomena.

Five publications to date implicate clock genes in autism (Wimpory et al. 2002; Bourgeron 2007; Nicholas et al. 2007; Hu et al. 2009a; Nguyen et al. 2010). This author proposes a hypothesis in terms of clock genes acting as morphogens, where this action is partly dependent on interplay with the sex-determination pathway in a cell-type-specific fashion. If clock gene determined neuronal branching patterns are a characteristic of certain neuronal oscillatory networks, it follows that altered expression patterns of clock genes in these tissues could alter the oscillatory properties of the network (if these depended on the degree of inter-branching). The basis for this tuning of oscillators by clock genes hypothesis is developed from the suggestion that clock genes regulate proteins involved in synaptic plasticity (Panda et al. 2002) the effect of *per* and *tim* on *Drosophila* neuronal structures (Mehnert et al. 2007), *per*

modulated oscillators regulating timing phenomena in *Drosophila* courtship behaviour (Kyriacou and Hall 1980) and on the mechanism of oscillatory circuits in insect brain (Lagier et al. 2007). In principle, it could be applicable to oscillators involved with social communication and cognitive functions in the human brain and builds on the tenet that oscillatory synchrony between brain regions is critical to normal cognition and memory function (Wang 2010).

#### 5.1.5i Hypothesis.

The clock genes *per* and *tim* regulate the size of *Drosophila* motor terminal boutons in a circadian manner and alter the number of neuronal projections in motor neurons in a developmental manner. Flies with mutated *tim* showed hyper-branching while mutants for *per* show hypo-branching, indicating *tim* suppresses branching while *per* promotes branching in these *Drosophila* motor neurons. The double mutant (*per/tim*) however has a normal phenotype with regards to this branching (Mehnert et al. 2007; Mehnert and Cantera 2008).

It is conceivable that differential clock gene expression in different cell types of the brain could developmentally regulate the degree of inter-arborization of adjacent neurons of different cell types. Further, the degree of interconnectivity between the cellular elements of a neuronal oscillator circuit could affect the overall frequency of the oscillator, measured in the order of milliseconds. For example, the frequency of  $\gamma$ 20-80Hz oscillations in mouse olfactory bulb that occurs in response to certain odors appears to be determined by a dendrodendritic inhibitory feed back circuit in mitral and granule cells of the bulb (two adjacent cell types). Mutation of

GABA<sub>A</sub>R- $\alpha$ 1 subunit that alters the signal strength of the inhibitory limb of the circuit also alters the frequency of the oscillator. Computer modeling based on these experimental findings shows that the effect of GABA<sub>A</sub>R- $\alpha$ 1 loss can be mimicked by a reduction in the number of inhibitory synapses (Lagier et al. 2007).

In principle, a neural oscillator can be envisaged where clock genes regulate the degree of arborization between adjacent neuron types (carrier and inhibitor) and therefore the number of inhibitory synapses. In this oscillator, epigenetic processes could regulate the tissue-specific expression pattern of clock proteins in a sex-appropriate manner. Confluence of the sex-determination pathway and circadian clock is envisaged to be a requirement for appropriate sex-specific splicing of transcripts that support sex-specific architecture of the oscillator circuit.

Promoter methylation of *PER* might determine cell-type specificity of expression and sex-specific alternative splicing (that requires clock regulated elements of the spliceosome) might determine sex-specific functionality. Thus mutation that affects the dosage, tissue distribution or pleiotropic function of *PER* (but not necessarily the circadian function) might also influence high frequency oscillator networks reliant on *PER*. *PER* forms complexes with the polypyrimidine tract-binding protein-associated splicing factor (PSF) and NONO (also implicated in splicing). Thus it is possible that *PER* may play a direct role in a protein complex concerned with the production of alternative transcripts (Shav-Tal and Zipori 2002; Duong et al. 2011).

### 5.1.6 Convergence between the circadian molecular clock and microRNA biogenesis pathways: A genetic focus for neuropsychiatric phenotypes.

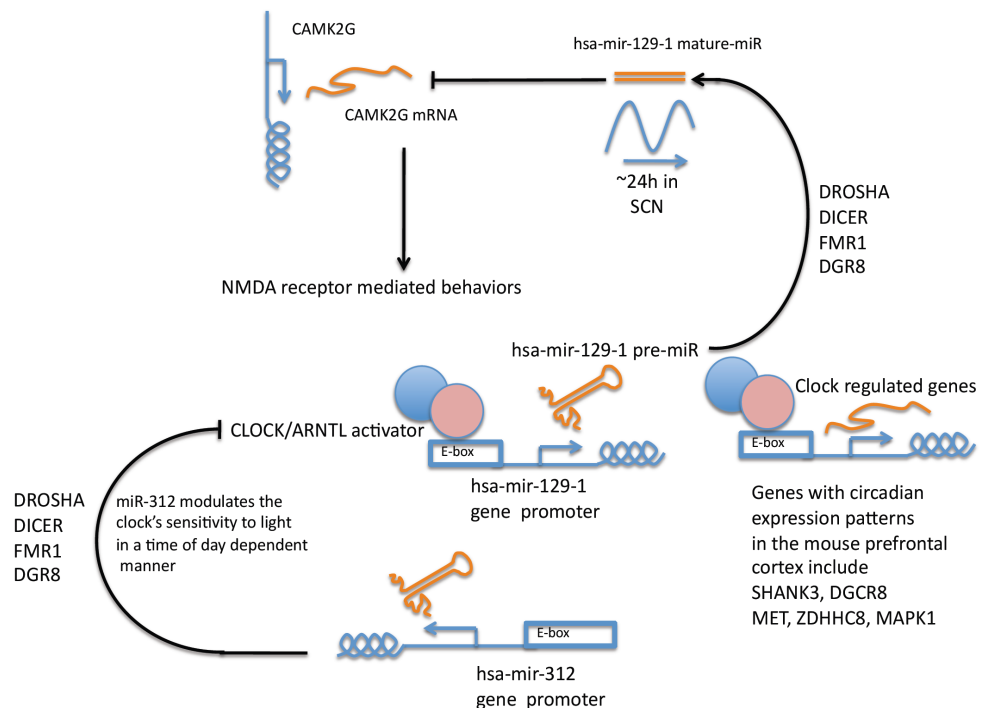
MicroRNAs have been recently discovered as components of the mammalian molecular clock. Circadian expression of *MIR129* regulates brain synaptic plasticity by *CAMK2* dependent regulation of NMDA receptor signaling, that is a target of dizocilpine, used in treating neuropsychiatric disorder (Cheng et al. 2007; Kocerha, et al. 2009). *MIR132* is also clock regulated and modulates the sensitivity of the SCN clock to resetting by light (Cheng et al. 2007). These recent discoveries may contribute to an explanation of why circadian dysfunction is common in neuropsychiatric disorder (Impey et al. 2010). Loss of function of the microRNA pathway may thus create neuropsychiatric phenotypes in part, by dysregulation of the circadian clock with concomitant effects on neuropsychiatrically relevant genes that are clock controlled e.g. monoamine oxidase A (*MAOA*) (Hampp et al. 2008).

A key element of the microRNA biogenesis pathway, The Fragile-X Mental Retardation Protein (FMRP) causes neuropsychiatric phenotypes that include profound circadian dysfunction and frequently autism (Verkerk et al. 1991; Yu et al. 1991). FMRP is an RNA binding protein important in RNA trafficking in axons (Lugli et al. 2005). FMRP regulates the degree of dendritic spine formation in neurons (Comery et al. 1997) and FMRP loss or over expression causes abnormal dendritic arbors and neural connectivity (Weiler et al. 1997). These effects are likely manifest through the proteins roles in RNA transport as well as being critical component of the RISC protein complex in the microRNA biosynthesis pathway (Li et al. 2008). MicroRNAs are shown to be critical to synaptic plasticity (Impey et

al. 2009). Thus loss of function of *FMR1* leading to altered dendritic arbors, may produce circadian disruption through interfering with the neuronal connectivity in the SCN and/or affect circadian clock function by knockout of *MIR312* and *MIR219-1* (upon which normal mammalian circadian clock function appears to depend (Cheng et al. 2007)). Additionally the E-box containing, circadian regulated and light sensitivity controlling *MIR132*, directly regulates synaptogenesis by activity-induced control of dendritic spine formation via regulation of RAC1-PAK signaling (Impey et al. 2009) that is the probable mechanism of this microRNA's influence on neuronal plasticity (Wayman et al. 2008).

Deletion of *DGCR8* on 22q11.21 contributes to the numbers of autism and schizophrenia cases (Vorstman et al. 2006; Kobrynski and Sullivan 2007). The *DGCR8* protein (also known as *pasha*) is, together with FMRP, a component of the microprocessor complex (Wang et al. 2007). Thus deletion or over expression of *DGCR8* by deletion or duplication of the 22q11.21 region should affect microRNA biogenesis which should, concomitantly, disrupt circadian rhythms that are dependent on the microRNA biogenesis pathway (Cheng et al. 2007). Recent findings show dysregulation of *DGCR8* with concomitant generalized unregulated microRNA levels in Schizophrenia (Beveridge et al. 2010). *DGCR8* appears to be a clock-controlled gene in the prefrontal cortex (of mouse) (Yang et al. 2007), thus making a potentially circular argument out of the question: does circadian disruption cause microRNA pathway dependent neuropsychiatric dysfunction, or is the converse true? In this context it is also noteworthy that *MECP2* is under homeostatic regulation by the circadian microRNA *MIR219* (Klein et al. 2007). Loss of function of

*MECP2* causes Rett syndrome a neuropsychiatric disorder with an autistic phenotype and a regression phase in early infancy that occurs at roughly the same developmental stage that autistic regression is observed (APA 1994).



**Figure 5-2. Possible co-reliance of the circadian molecular clock and MicroRNA biosynthetic pathways in neurons.**

Circadian regulation of an essential element of the microRNA pathway (*DGCR8*) creates co-reliance between these two pathways. Knockout of the two circadian microRNAs via knockout of *FMR1/DGCR8* derives circadian phenotypes. At least in mouse prefrontal cortex (PC), *DGCR8*, a critical component of the microRNA biosynthesis pathway, is under circadian clock control (Yang et al. 2007). Thus clock gene variation that affects the normal circadian expression of *DGCR8* in the PC could affect microRNA biosynthesis and thus microRNA dependent neuronal plasticity in this brain region. Synaptic proteins and autism candidate genes e.g. *SHANK3* and *MET* that are under circadian control in the PC are also highlighted. The model is proposed as a reason for the common co-morbidity between neuropsychiatric disorders and circadian rhythm dysfunction.

### 5.1.7 There are many ways to break a clock.

The number and variety of genes comprising the circadian molecular clock reflects the complexity of this circadian oscillator. Thus a global phenotype, circadian rhythm disruption, may be caused by many possible variations within each of the genetic elements that together comprise the circadian molecular clock. If, as proposed here, circadian rhythm dysfunction due to clock gene variation might contribute to autism, then the division of labour between the clock genes suggests that the impact of genetic variability within this gene network might well be understated in association tests, unless the analyses takes account of epistatic effects within the network as a whole. It is also clear that the one phenotype of circadian disruption may be accompanied by other pleiotropic effects that depend on which of the interacting clock genes is affected and even perhaps, different types of mutation within one particular clock gene.

Mutation that affects the tissue specific expression patterns of clock genes for example might be as powerful in producing neuropsychiatric phenotypes as gene knockouts, as it is clear that brain regions differ with regards circadian patterns of gene expression. However the impact of alterations of regionality of circadian expression has not been determined.

This concept of altered patterns of clock gene expression in the brains of individuals with autism is considered in the social timing hypothesis. Genetic and epigenetic mutation including altered clock gene methylation are proposed as determining factors for aberrant clock gene expression patterns in autism (Wimpory et al. 2002). Support for this notion has recently emerged in terms of: epigenetic effects at the *CLOCK* locus in a large autism population based study (Fradin et al. 2010); the

association of a polymorphic CpG site in *NPAS2* with autism (Chapter 2 and Chapter 7.3.3), and, altered promoter methylation of *RORA* in autism (Nguyen et al. 2010).

#### 5.1.8 Final conclusion and indications for further research.

Further genetic tests in larger samples are needed to confirm and measure the risk associated with the *NPAS2* and *PER1* variants showing positive association with autism in this study. The considerable overlap between autism candidate genes and clock-controlled genes suggests the role of the circadian clock in regulating these autism candidate genes should be explored further. Future genetic studies of clock genes and autism genes such as *SHANK3* may reveal epistatic effects.

The finding that the autism-associated SNP rs34705978 is within a differentially methylated region, together with recent evidence of altered clock gene methylation in autism, encourages further exploration of this SNP. If altered methylation of rs34705978 is associated with autism, the effect of the epi-allele could be additive to the effects of nucleotide substitution at this SNP.

The amenability of the circadian clock to non-invasive monitoring and to mild behavioral/pharmacological manipulation, makes the circadian molecular clock a potential therapeutic target in autism. Addressing pathological methylation might also be more tractable than compensation for DNA sequence changes.



## APPENDICES

### Appendix A

#### **Methylation status of a CpG rich element in *PER1*: a preliminary investigation.**

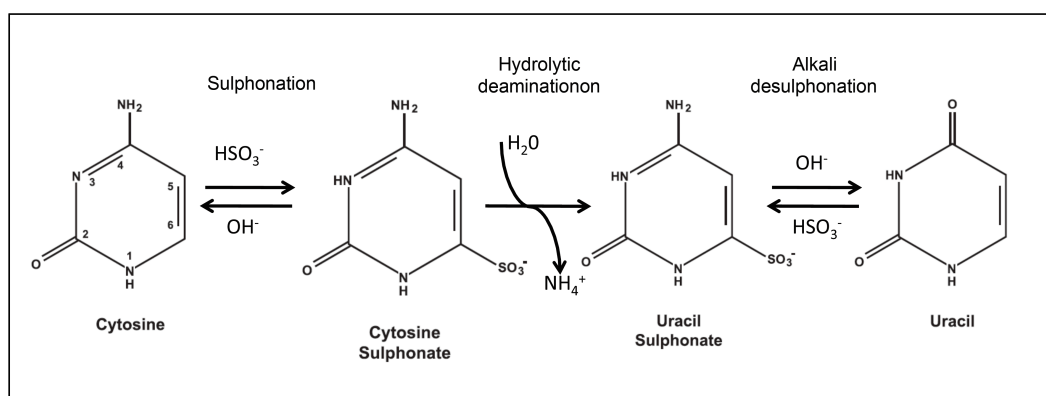
##### **A.1 Introduction.**

To investigate whether clock genes might be anomalously methylated in autism a preliminary investigation was carried out to test the feasibility of a small-scale study of clock gene methylation in autism DNA samples. This work was undertaken prior to the publication of the results showing abnormal methylation of the clock gene *RORA* and parent-of-origin effects on SNPs near *CLOCK* in autism (Nguyen et al. 2010; Fradin et al. 2010).

The association test described in chapter 2 identified two SNPs in *PER1* (rs885747 and rs6416892) that showed positive association with autistic disorder. A CpG repeat element (CGGCGCCGTG) was found close to both SNPs suggesting that the SNPs may be in linkage disequilibrium with methylation of this motif. The feasibility of analysing this sequence in autism samples was investigated in experiments with Polymerase Chain Reaction (PCR) and methylation-specific PCR and autism DNA samples (Table A-1, A-2, and A-3).

The technique of methylation-specific PCR (Herman et al. 1996) can be divided into two main steps (Figure A-1). Firstly, a modification step, where

all un-methylated cytosine residues in the DNA sample are converted to uracil (but methylated cytosine residues are protected and not converted). Secondly, the modified sample is split into two aliquots and amplified by standard PCR procedures. For example, a sequence TTAACGTTAA will be converted to TTAAUGTTAA after sodium bisulphite treatment if the central CpG dinucleotide was not methylated. If the central CpG dinucleotide is methylated, it will be left unconverted. One of the aliquots is amplified by PCR with primers designed for unmodified (methylated) DNA. For the other aliquot, a set of primers specific for the (unmethylated) DNA is used and therefore the modified sequence, if present, is amplified. Using paired-primer-controls the presence or absence of methylation of CpG dinucleotides within the primer sequences is thus determined by whether or not amplification occurs. PCR amplification is determined by agarose gel electrophoresis followed by ethidium bromide staining, visualisation by ultra violet (UV) fluorescence and image capture.



**Figure A-1.** Bisulphite reaction and the conversion of cytosine to uracil via a sulphonated intermediate.

## **A.2 Materials and Methods**

### A.2.1 Materials

- DNA Samples: Human genomic DNA supplied by the AGRE consortium (from autistic children and their parents) (Geschwind et al. 2001). The DNA samples were supplied dissolved in water in 96-well polypropylene plates and stored at -30°C. Stock solutions were at a concentration of 1µg DNA /100µl distilled water.
- Bisulphite conversion kit: Qiagen EpiTect Kit (catalogue No 59104/2006) for bisulphite conversion and clean up of DNA for methylation analysis.
- Fermentas, Gene Ruler™: Standard DNA markers.

#### A.2.1i Buffers and solutions

- Tris Borate EDTA Buffer (TBE): for agarose gel electrophoresis
  - 10 x Stock Solution
  - 0.89M Trisma base
  - 0.02M EDTA
  - 0.89M Boric acid
- DNA Sample Loading Buffer
  - 10x stock solution
  - 0.005% bromophenol blue,
  - 50% glycerol.
- Agarose gel mix
  - 1 x TBE buffer
  - 1% Agarose
  - 0.5µg/ml ethidium bromide

A2.1ii Primer Sequences.

Native Primers		
Code	Primer sequence	<i>PER1</i> genomic sequence
747-1	5'-CTCAGCTCTCCCTCCC-3'	5'-CTCAGCTCTCCCTCCC-3'
747-2	5'-GATGTCACGCGCCCG-3'	5'-CGGCGCCGTGACATC-3'
892-1	5'-CTCTGTCACCCAGCTGG-3'	5'-CTCTGTCACCY <sup>1</sup> AGCTGG-3'
892-2	5'-GCCACGGCGCCGGCG-3'	5'-CGCCGGCGCCGTGGC-3'
Aut-1	5'-CTCCTCCTTTCCCACTTCC-3'	5'-CTCCTCCTTTCCCACTTCC-3'
Aut-3	5'-CACACATCATCAACTCAC-3'	5'-GTGAGTTGATGATGATGTRTG-3'
PER1-1	5'-CAGGTAAGTGGCTGTGATCG-3'	5'-CAGGTAAGTGGCTGTGATC-3'
PER1-2	5'-CTCTTGCTGCTCTCAGTGGTC-3'	5'-GACCACTGAGAGCAGCAAGAG-3'
PER1-3	5'-CCTGGACACTGATATCCAGG-3'	5'-CCTGGACACTGATATCCAYG-3'
PER1-4	5'-GATCTGCTGGAAAGTCACCTG-3'	5'-CAGGTGACTTTCCASCCAGATC-3'
PER1-5	5'-GGAAGAATGGAGAAGCAAAGC-3'	5'-GGAAGAATGGAGAAGCAAAGC-3'
PER1-6	5'-TCCACCCGCGCCGCCTAGAG-3'	5'-CTCTAGGCGGCGGGGTGGA-3'

**Table A-1. Native primers for *PER1***

Y<sup>1</sup> = SNP rs117343376 C/T, R= SNP rs rs55655060 G/A, S = SNP rs760218 G/A, Y = SNP rs3027181 C/T. These SNPs were not publicly reported at the time of the experiment.

**Table A-2. Fragment size calculations from Ensembl database sequences.**

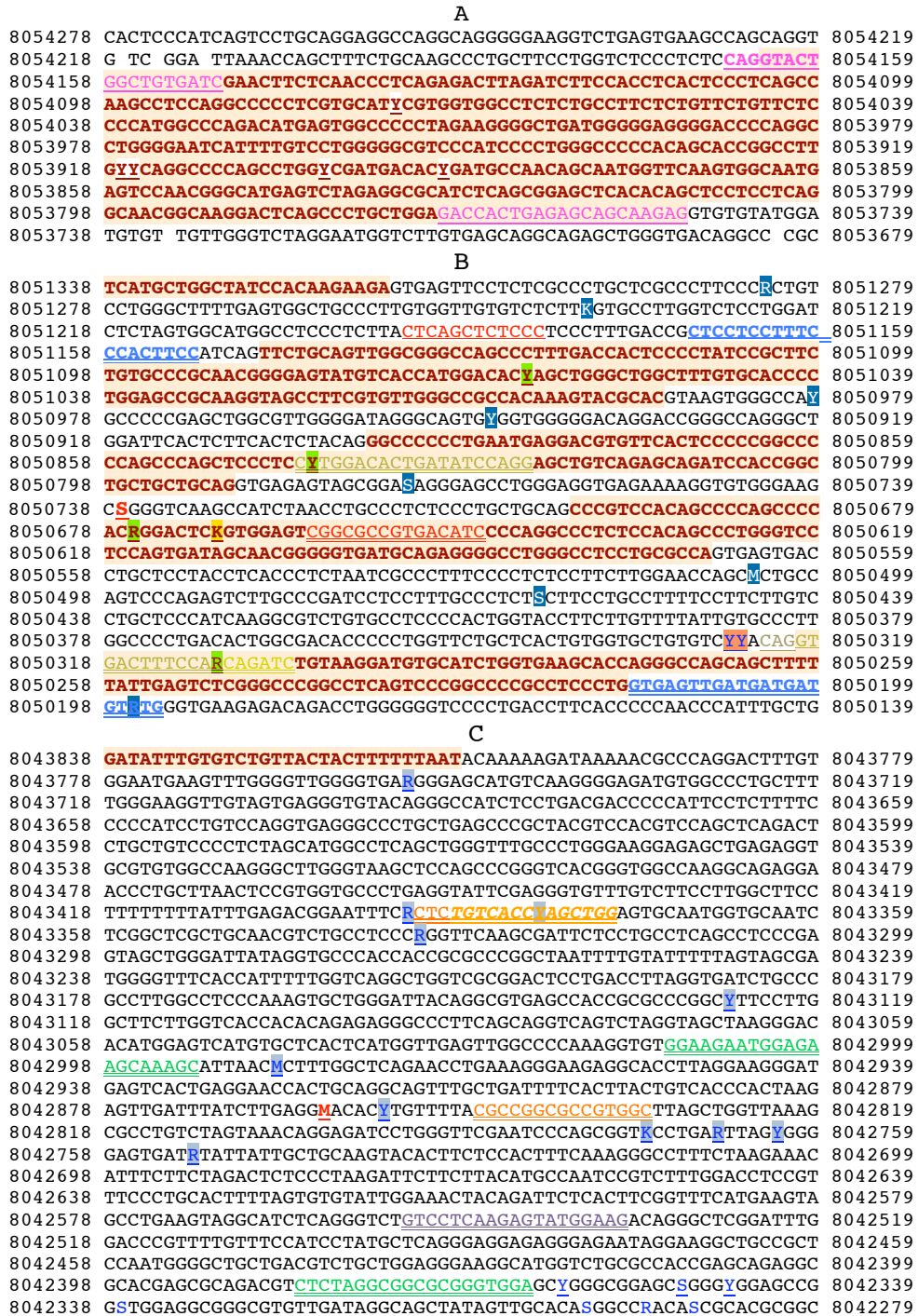
Primer pair code	Separation distance in bp
PER1-747-1 & PER1-747-2	547
PER1-892-1 & PER1-892-2	560
Aut-1 & Aut-3	976
PER1-1 & PER1-2	417
PER1-5 & PER1-6	649
PER1-747-1 & PER1-3	371
PER1-3 & PER1-4	540

**Table A-3. Methylation Specific Primers for *PER1*.**

Methylation Specific Primers		
Code	Primer sequence	Corresponding <i>PER1</i> genomic sequence etc.
Bis-892-1A*	5'-TATCACCCAACTAAAATAC-3'	5'-TGTCACCY <sup>1</sup> AGCTGGAGTGC-3' genome seq. 5'-GCACTCCAGCTGGZTGACA-3' antisense seq. 5'-GUAUTUUAGUTGGZTGAUA-3' mod antisense seq.
Bis-PER1-7	5'-CTTCCATACTCTTGAGGAC-3'	5'-GTCCTCAAGAGTATGGAAG-3' genome seq. 5'-GTUUTUAAGAGTATGGAAG-3' mod genome seq.

\*In primer Bis-892-1A Y<sup>1</sup> = SNPrs117343376 C/T

The oligonucleotides used as the primers were synthesized and supplied by MWG BIOTECH. Anzinger str. 7, 85560 Ebersberg, Germany.



**Figure A-2. PER1 sequence fragments.**

Sequence A shows the first large exon of *PER1* in brown shading and the primer pair PER1-1/PER1-2 is shown in pink letters covering the 5' and 3' ends of the exon.

Sequence B is the region of *PER1* centred on intron 12 where **S** = SNP rs885747. The

primer pairs are shown colour coded as in Table A-2 e.g. PER1 747-1/PER1 747-2 is in red letters. SNPs are shown highlighted with single letter ambiguity codes. In sequence C, brown lettering shows the end of the *PER1* 3'UTR. **M** = autism SNP rs6416892 and genome co-ordinates are given at the beginning and end of each line of the sequence.

### A.2.2 Methods.

#### A.2.2i Agarose gel electrophoresis.

The gel plate from an electrophoresis apparatus (e.g. Bio-Rad Mini-Sub) was prepared by sealing the ends with adhesive paper tape. The gel plate was placed on a level surface and the comb from the apparatus adjusted so that the teeth of the comb were level and about 1.5mm above the floor of the gel plate. Subsequently, agarose powder was dissolved in 1 x TBE buffer (1% weight to volume) by heating the mixture in a microwave oven. On cooling the molten gel solution to 55°C, ethidium bromide solution was added, to a final concentration of 0.5µg/ml and the solution was mixed carefully, to avoid the formation of bubbles. Typically 50ml of the molten gel was made at a time, enough to fill a mini gel plate to a depth of 5mm. The molten gel with ethidium bromide was poured into the gel plate on a level surface and allowed to set at room temperature (~20 minutes). After removing the comb followed by the sealing tapes from the set gel, the tray was placed in the electrophoresis chamber and covered with 1 x TBE buffer. 1 to 5µl of DNA sample was mixed with 5µl of 6 x loading buffer and the samples loaded into the wells of the gel with a micropipette. At least one lane on each gel was reserved for DNA fragment size markers e.g. Gene Ruler™. After connecting the electrophoresis chamber to the transformer, the gel was run at 120 V and the progress of the fragments observed by visually checking the movement of the marker dye-front in the

gel. When the marker dye had migrated to the end of the gel or to an appropriate prior point, the gel was transferred to a ultra-violet trans-illuminator for visualization and digital photography.

#### A.2.2ii Polymerase Chain Reaction (PCR): standard procedure.

Samples of human genomic DNA (1µg/100µl) from autism families, dissolved in nuclease free water and stored at -30°C, were thawed and a 20µl working aliquot taken from selected wells of the coded 96-well storage plates such that each DNA sample selected represented a single test individual (parent or child) with all codes recorded.

To make the reaction mixture, 4µl of human genomic DNA was added to a mixture consisting of: 5µl of a 5µM solution of primer 1, 5µl of a 5µM solution of primer 2, 5µl of 10x Taq buffer (Fermentas (NH<sub>4</sub>)<sub>2</sub> SO<sub>4</sub> solution free of magnesium chloride), 5µl of dNTPs 2mM, 3µl of MgCl (25mM, giving 1.5mM final concentration), 23µl purified water and 1µl (1.25 units) of Fermentas Taq polymerase enzyme. This mixture was made in a 500µl plastic PCR tube, mixed with a vortex mixer and the tube and contents transferred to a thermal cycler set to generate the conditions listed in Table A-4.

**Table A-4. Thermal cycler settings for standard PCR reactions.**

Step	Time	Temperature
1	2min	94°C
*2	30min	94°C
*3	30min	55°C
*4	60min	72°C
5	10min	72°C
6	hold	10°C

\* Steps 2, 3 and 4 of Table A-4 were repeated 30x before moving on to steps 5 and 6. When finished, the PCR tube contents were mixed with 5 $\mu$ l of 6 x gel loading buffer and 20 $\mu$ l of this final mixture run at 7V/cm for 45 minutes on a 1% agarose 8cm gel in TBE buffer.

#### A.2.2iii PCR: methylation-specific procedure.

Genomic human DNA samples were selected as for the standard PCR reaction described above. Each 20 $\mu$ l sample (0.2 $\mu$ g DNA) aliquot was placed in a sterile PCR tube to which was added 85 $\mu$ l of dissolved Bisulphite Mix from the EpiTect kit. The Bisulphite Mix was prepared by adding 800 $\mu$ l of RNase free water to each aliquot of Bisulphite Mix supplied in the kit and, so prepared, each aliquot of bisulphite mix was enough for 8 conversion reactions. 35 $\mu$ l of DNA Protect Buffer (supplied in the EpiTect kit) was then added to the PCR tube containing the 20 $\mu$ l sample aliquot and the 85 $\mu$ l of dissolved Bisulphite Mix. The contents were mixed thoroughly with a vortex mixer and the mixture checked to ensure the Protect Buffer turned from green to blue indicating sufficient mixing and the correct pH for the conversion reaction. The samples were kept at room temperature before being placed in the thermal cycler under the conditions shown in Table A-5. After bisulphite conversion in the thermal cycler the PCR tubes were spun in a micro-centrifuge at 3000rpm for 30 seconds after which the contents were transferred to 1.5ml micro centrifuge tubes.



**Table A-5. Thermal cycler settings for the methylation-specific PCR reactions.**

<b>Step</b>	<b>Time</b>	<b>Temperature</b>
Denaturation	5min	99°C
Incubation	25min	60°C
Denaturation	5min	99°C
Incubation	1h 25min	60°C
Denaturation	5min	99°C
Incubation	2h 55min	60°C
Hold	overnight	20°C

560µl of freshly prepared Buffer BL was then added to the mixture in each microcentrifuge tube. No carrier RNA is required if the sample contains >100ng DNA and this was the case for this experiment. The whole contents of each microcentrifuge tube was then added to an EpiTect spin column and spun in a microcentrifuge at maximum speed for 1min after which the flow through was discarded, the DNA having bound to the column. The column was then washed by adding 500µl of wash buffer (Buffer BW) followed by microcentrifugation of the spin column at maximum speed for 1 min. 500µl of desulphonation buffer (Buffer BD) was then added and the spin column incubated for 15 min at room temperature. The column was then spun in a microcentrifuge at maximum speed for 1 minute. After discarding the flow-through 500µl of Buffer BW was added to the spin column and spun at maximum speed in a microcentrifuge for 1 minute. This step was repeated once over. The spin column was then transferred to a new 1.5ml collection tube and spun at max in a microcentrifuge for 1 minute to remove any residual liquid from the column. Subsequently the spin column was placed into a new 1.5ml microcentrifuge tube and 20µl of Buffer EB was carefully dripped onto the centre of the membrane at the top of the spin column using a micropipette.

The purified DNA was collected by spinning the column for 1 minute at maximum speed in a microcentrifuge (~15,000 x g 12,000rpm). The concentration of the recovered DNA was measured using a quartz cuvette and a Sanyo spectrophotometer set to measure the absorbance of the solution at 260nm and where 1 absorbance unit was equivalent to 50µg/DNA per ml of solution.

### **A.3 Results.**

#### A.3.1 Standard PCR reactions.

Amplification of human DNA samples from neurotypical parents and autistic children using the primers described in Table A-1 was demonstrated. Table A-6 shows which samples were amplified: with which primer sets, the predicted fragment size of the amplified fragment, and the observed fragment size. Figures A-3 and A-4 show typical agarose gels from the experiments from which the data in Table A-6 was derived.

For most primer-pairs, amplification gave bands corresponding to fragment sizes in keeping with the predictions from the genomic *PER1* sequence. However, some tests gave multiple bands and some samples were recalcitrant to amplification. For example, the primer pair PER1-747-1 / PER1-747-2 gave unexpected results in family AU064. With sample AU064004 (autistic child) a band of ~1350 bp was seen, with a fainter band at ~1500bp (547bp expected) and multiple bands ~1200 & 500 & 250bp in the DNA sample from the mother. Amplifying the same region of sample AU064004 with different primer set PER1-747-1/ PER-3 (371 bp expected) also gave multiple bands of ~1350 & 900 & 500 & 400 bp.

A.3.2 Methylation specific PCR reactions.

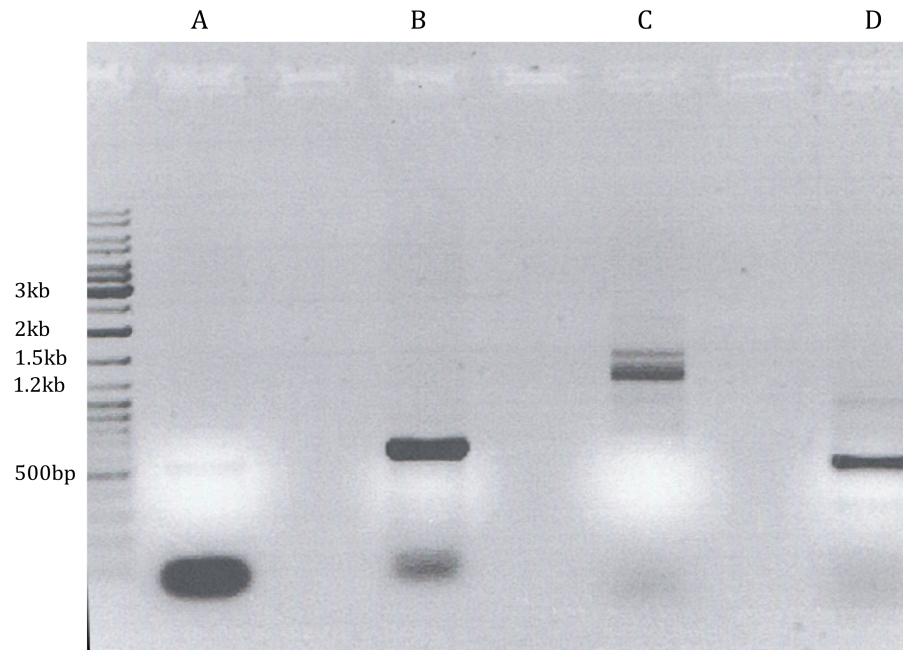
The PCR reactions set up with the methylation specific primers consistently showed no amplification product in contrast to the unmodified autism DNA. Firstly the yield of modified sample DNA was considered. Measurement with the Sanyo spectrophotometer and quartz cuvette at 260nm showed yields of nucleic acid of up to 70% of the starting sample

**Table A-6. Primer pairs *PER1* and the observed fragment sizes.**

Sample code	Primer set	Predicted fragment size (bp)	Observed fragment size (bp)
AU067404 (Aut)	Aut-1 Aut-3	976	~1000
AU067402 (NM)	Aut-1 Aut-3	976	~1000
AU004102 (NM)	PER1-1 PER1-2	417	~417
AU01001 (Aut)	PER1-1 PER1-2	417	~417
AU064001 (NF)	PER1-1 PER1-2	417	~417
AU064004 (Aut)	PER1-1 PER1-2	417	~417
AU066401 (NF)	PER1-1 PER1-2	417	>400 to <500
AU064004 (Aut)	PER1-5 PER1-6	649	~650
AU064001 (NF)	PER1-747-1 PER1-747-2	547	~1200 & 500 & 250
AU064004 (Aut)	PER1-747-1 PER1-747-2	547	~1350
AU064004 (Aut)	PER1-892-1 PER1-892-2	560	~560
AU064004 (Aut)	PER1-892-1 PER1-5	1030	~1030
AU064004 (Aut)	PER1-747-1 PER1-3	371	~1350 & 900 & 500 & 400
AU067208 (Aut)	PER1-892-1 PER1-5	1030	~700 & 800 & 300

based on OD of 1 = 50 µg/ml for double-stranded DNA. However, when this modified DNA was examined on an agarose mini gel, as a direct sample and after precipitating the DNA and re-dissolving in TE buffer, no

trace of DNA was apparent suggesting that the DNA may be highly fragmented.

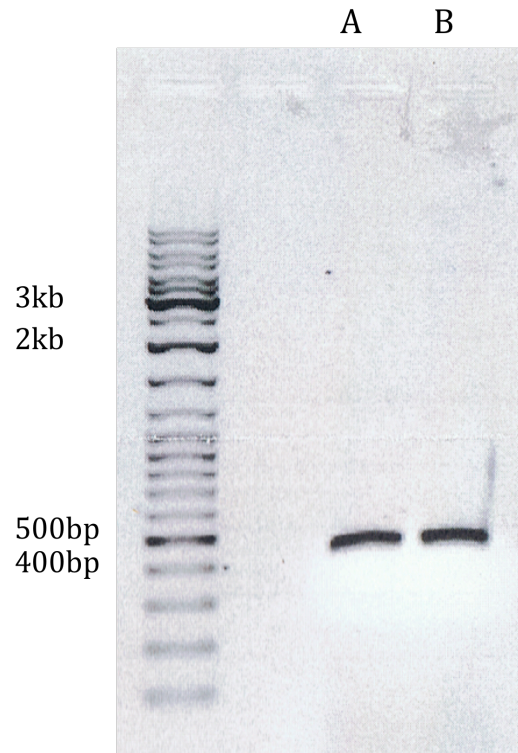


**Figure A-3. PCR amplifications of fragments of *PER1* in autistic male child AUO 64004.**

Tracks show: A = *PER1*-3 & *PER1*-4 (540), B = *PER1*-5 & *PER1*-6 (694), C = *PER1* 747-1 & *PER1* 747-2 (547) and D = *PER1* 892-1 & *PER1* 892-2 (560). Fragment sizes predicted from the genomic sequence are shown in the brackets. The weak amplification in lane A may be due to the presence of alleles of SNP rs76021 and/or SNP rs3027181 in the sample DNA sequence that were incompatible with primers *PER1*-3 & *PER1*-4. These SNPs were unreported and not accounted for in the primer design at the time of the experiment.

The modification procedure was repeated several times with different samples and with varying the parameters suggested in the manual supplied with the methylation specific PCR kit (temperature and time for

modification) however no signal was obtained from any of the amplifications using modified DNA as a substrate, with any primer pair.



**Figure A-4. Sample A: AUO 64001 and B :AUO 66401 amplified with primers PER1-1 and PER1-2.**

The gel shows a single band for both samples of between 400-500bp The fragment size expected from the sequence shown in Figure A-2 was 417bp.

#### **A.4 Discussion.**

These preliminary trials showed that generally, the human DNA samples were amenable to standard PCR amplification. However, some of the samples were recalcitrant to amplification (e.g. sample AUO64004 with primers PER1-3 & PER1-4 (Figure A-3)) or gave unexpected results as with primer-pair PER1-747-1/ PER1-747-2 (Figure A-3). Amplification of sample AUO64004 with primers PER1-3 & PER1-4 may have failed because of the presence of SNP rs760218 and/or SNP rs3027181 in the sample

DNA. These SNPs were not known at the time of experimentation and so the primer design did not take account of this possibility.

With primers PER1-747-1/ PER1-747-2 inappropriate size and multi banding was seen when the region was amplified in AUO64004 (autistic male child) and AUO64001 (his mother). Neither did varying the concentration of magnesium ions in the PCR reaction mix, nor using a different primer pair (PER1-747-1 / PER1-3) to amplify this region (11<sup>th</sup> and 12<sup>th</sup> Exons) help to achieve a single band. However amplification of a nearby region of sample AUO64004 with two other primer-pairs gave a single band of expected size. This suggests that there is not a general problem with the DNA of sample AUO64004. Given that two different primer pairs that covered the region PER1-747-1 to PER1-747-2 in these two related samples (mother and son) gave anomalous banding and that these samples gave expected results with primers amplifying a nearby region of *PER1*, it is possible that there is a sequence anomaly in these samples in the region delineated by the primer-pair.

The intended focus of this experiment was on finding good working matches between sample and primers such that unambiguous amplification of fragments containing the CpG rich element could be used as the basis for subsequent investigation within the sample population. Thus the possible reasons for the fragment size differences were not fully explored.

Technical problems precluded successful testing of the CGGCGCCGTG sequence with methylation specific PCR even though the trial methylation specific primers (Table A-3) were designed to contain no protected (methylated) CpGs (Ambiguities might arise if more than one

CpG occurred in the primer sequence and selective methylation of CpGs occurred). Nevertheless, if the results of this methylation specific PCR had shown positive results, such differences would require considerable further investigations in a sample size similar to that of the association test described in chapter 2, if these differences were to be shown to be associated with autism specifically. At the outset it was considered this might have been feasible with the samples to hand. However the results of this trial, apart from the technical difficulty with getting the methylation specific PCR working, indicated that time constraints/costs would not allow enough data to be collected to give meaningful results when for example, resequencing to allow for new SNPs is taken into account. For this reason and because the CpG motifs were not directly linked to the association signals described in chapter 2, in deciding which of the preliminary experiments to follow up, experiments that did not specifically involve the autism-associated SNPs were considered less important than experiments that could show results that were allele specific for these SNPs. Experiments that built directly on the positive results of the association study SNPs specifically were given priority.

## Appendix B

### Does NPAS2 form a complex with HAT1?

#### B.1 Introduction.

It is clear that the circadian clock plays an important role in genome maintenance but the details of exactly how this is achieved are still to be clarified. The brain is a highly metabolically active organ and terminally differentiated neurons in the brain are especially sensitive to damage by the bi-products of oxidative metabolism (Fishel et al. 2007). DNA repair in brain neurons is thus a vital process to maintain normal neuronal function and avoid apoptosis.

The large number of *NPAS2* haplotypes showing association with autism (see chapter 2), the protein's role as a redox signaling molecule (Rutter et al. 2001) and the findings that cerebellar purkinje cells (that are particularly sensitive to oxidative damage) are reported to undergo premature apoptosis in autism (Bauman and Kemper 1985) informed the notion that possibly currently undefined histone acetyltransferases required for DNA repair may partner *NPAS2* in a DNA damage response mechanism. *CLOCK*, a closely related protein to *NPAS2* (Reick et al. 2001), has inherent histone acetyltransferase activity and together with, K (lysine) acetyltransferase 5 (*KAT5* also known as *Tip60*), is involved with DNA repair that depends on chromatin remodeling (Miyamoto et al. 2008). Thus *NPAS2* might partner with a molecule that offers HAT activity and facilitates DNA repair.



Histone acetyltransferase 1 (HAT1) was selected as a candidate because: it is a histone acetyl transferase involved with the repair of DNA double stranded breaks (Qin and Parthun 2006) that are common in active neurons (Fishel et al. 2007), it is located in a region considered as an autism susceptibility locus (Marshall et al. 2008), it is linked to breast cancer development (Sorbello et al. 2003) which is significantly increased in autism (Kao et al. 2010) and its domain structure is somewhat similar to that of the CLOCK interacting protein HAT KAT5 (Tip60). Also, it appeared from review of available literature that HAT1 had received only modest investigation and the full spectrum of HAT1 function and interaction was yet to be defined.

The histone acetyltransferase HAT1 (Kleff et al. 1995) has two roles that are in keeping with its localization as a cytoplasmic protein and a nuclear protein (Ai and Parthun 2004) In the cytoplasm HAT1 acetylates newly synthesized H4 histone molecules. In the nucleus HAT1 takes part in DNA damage repair (Benson et al. 2007). In the cytoplasm HAT1 acetylates newly synthesized H4 histone molecules in a conserved pattern at lysines 5 and 12 but within one hour of acetylation, de-acetylation of H4 facilitates nucleosome formation (Jackson et al. 1976; Annunziato and Seale 1983). Experiments with yeast show that Hat1 (*s.cereviciae*) associates with Hat2 (RBBP7 in *H. sapiens* (Eberharter et al. 1996) in a complex known as the HAT-B complex (Verreault et al. 1998). The interaction of these two proteins increases the affinity of the HAT-B complex for H4 ten fold. Further, in the nucleus, Hat1p acetylates histone H4 at position 12 in the process of telomeric silencing (Kelly et al. 2000). In *S.pombe* *hat1* deletion has been shown to increase the sensitivity of

cells to methyl methanesulfonate, indicating a role for Hat1 in DNA double strand break repair (Qin and Parthun 2002; Barman et al. 2006; Qin and Parthun 2006) that can *in vivo* be substituted for by alternative means of acetylation of Lys-5 and Lys-12 (Benson et al. 2007).

## **B.2 Materials and Methods.**

### B2.1 Media, buffers, solutions, gels and antibodies.

Cell Culture: Reagents:

- Dulbecco's Modified Eagle Medium (D-MEM GlutaMAX™)
- Phosphate-Buffered Saline GIBCO® (PBS pH 7.4)
- Fetal Bovine Serum GIBCO®
- Trypsin-EDTA (Phosphate-Buffered Saline, pH 7.4 and containing 0.025% trypsin and 0.01% EDTA pH7.4)
- Penicillin-Streptomycin, liquid. (Invitrogen™ penicillin G (sodium salt) and streptomycin sulfate in 0.85% saline).

Protease inhibitors for soluble protein extract:

- 50mM NaF
- 1mM Na<sub>3</sub>VO<sub>4</sub>
- 5mM N-ethylmethimid
- 1mM PMSF
- 1 Protease Inhibitor Cocktail Tablet (Roche, cat no 11836 153 001) per 10ml HEPES buffer.

HEPES Buffer:

- 50mM HEPES, (pH 8.0)
- 150mM NaCl
- 0.1% NP40
- 10% glycerol
- 5mM EDTA

SDS PAGE Loading Buffer x2:

- 4% SDS

- 10% 2-mercaptoethanol
- 20% glycerol
- 0.004% bromophenol blue
- 0.125M TrisHCl pH 6.8
- SDS PAGE Running Buffer (pH ~8.3):
- 25mM Tris base
- 190mM glycine
- 0.1% SDS
- Western Blot Transfer Buffer (pH 8.3):
- 25mM Tris base
- 190mM glycine
- 10% methanol
- 0.1% SDS

Tris Buffer Saline x10, (TBS x 10):

- 24.23g Trizma HCl (pH 7.6).
- 80.06g NaCl
- 800ml Distilled water

Tris Buffer Saline Tween20, (TBST):

- 100ml of TBS x10
- 900ml Distilled water
- 1ml Tween20

Membrane Blocking Buffer:

- 3 - 5% powder milk or BSA in TBST buffer.

- Polyacrylamide Gel Mixes:

**Table B-1a. SDS Polyacrylamide resolving gel mix (10 ml).**

Reagent	Gel percentage		
	7.5%	10%	12.5%
10% SDS	100 $\mu$ l	100 $\mu$ l	100 $\mu$ l
1.5M Tris (pH 8.8)	2.5 ml	2.5 ml	2.5 ml
30% Acrylamide	2.5 ml	3.34 ml	4.2 ml
Distilled water	4.84 ml	4 ml	3.14 ml
10% APS	50 $\mu$ l	50 $\mu$ l	50 $\mu$ l
10 $\mu$ l Temed	10 $\mu$ l	10 $\mu$ l	10 $\mu$ l

**Table B-1b. SDS Polyacrylamide stacking gel mix.**

Reagent	volume
10% SDS	50 $\mu$ l
0.5M Tris (pH 6.8)	1.25 ml
30% Acrylamide	0.65 ml
Distilled water	3.05 ml
10% APS	50 $\mu$ l
10 $\mu$ l Temed	10 $\mu$ l

### Antibodies.

- NPAS2: Mouse antibody against a partial recombinant NPAS2 fragment. Storage buffer: 1 x PBS, pH 7.2. Working solution; 1 $\mu$ l antibody in 500  $\mu$ l Membrane Blocking Buffer. The antibody was supplied by The Abnova Corporation: 9F No. 108, Jhouzih St., Niehu, Taipei, Taiwan.
- HAT1: Rabbit polyclonal to HAT1 where the immunogen peptide is EKFLVEYKSAVEKK corresponding to amino acids 8-21 of human HAT1. Working solution: 1 $\mu$ l antibody in 500  $\mu$ l of Membrane Blocking Buffer. Antibody supplied by ABCam, 330 Cambridge Science Park, Cambridge, CB4 0FL UK.
- PCAF: Rabbit antibody supplied by Sigma Aldrich, Fancy Road, Poole, Dorset. BH12 4QH. Developed from synthetic peptide corresponding to amino acids 817-832 of human PCAF. Working solution: 1 $\mu$ l antibody in 500  $\mu$ l of Membrane Blocking Buffer.

### B.2.2 Tissue culture with HeLa cells.

HeLa cells were grown in plastic Petri dishes containing D-MEM plus fetal

calf serum (10% v/v), penicillin G (5000 U/ml) streptomycin 5000 $\mu$ g/ml, 2mM L-glutamine, and 1mM sodium pyruvate and incubated at 37°C in water saturated air and 5% CO<sub>2</sub>. The 60mm diameter plates of area 2,827mm<sup>2</sup> and containing pre-warmed culture medium were seeded at 0.8 x 10<sup>6</sup> cells per plate such that after ~2-3 days in the incubator the cells at confluency had a density of 3.2 x 10<sup>6</sup> cells per plate.

Cells were maintained in continuous culture by inoculating fresh medium in Petri dishes with a suspension of cells, made by removing the medium of a stock plate with an aspirator, washing the cells adhering to the plate with 1.5ml PBS, removing the supernatant with the aspirator and treating the cell layer with 1.5ml Trypsin-EDTA solution and incubating for about 1-3 minutes. Freeing the cells from the surface of the dish was judged to have occurred when the cell layer was seen to slide slowly down the dish when it was inclined and tapped (i.e. < 3 minutes). The freed cells were taken up by pipette and mixed with 5ml of pre-warmed culture medium (containing fetal bovine serum which neutralizes the trypsin) and precipitated by centrifugation at 1000rpm for 2 minutes. The supernatant was removed with the aspirator and the cells resuspended in 10ml of medium in a sterile plastic tube. The density of the cell suspension was checked with a haemocytometer and the appropriate volume of this preparation used to inoculate 60mm plates that contained 8ml of pre-warmed growth medium.

A single passage through this cycle took 2-3 days and this cycle of regenerating the cultures was maintained to ensure the fitness of the cells. The cells were checked by observation with a light microscope to estimate how close to confluence the spreading plaque was. The medium changed

from pink to yellow with age (decreasing pH), and was changed before it reached the yellow stage. At all stages where the cells were exposed (inoculation, trypsinisation etc.), the procedure was carried out using sterile technique and in a sterilized, laminar-flow, positive air pressure, filter hood.

### B.2.3 Protein extraction from HeLa cells.

90% confluent HeLa cells, adhering to a 60mm diameter Petri dish were washed three times with PBS. 700µl of IGEPAL CA-630 (Octylphenyl-polyethylene glycol) Lysis Buffer was then added to the washed cells. The protease inhibitors aprotinin (A) and Leupeptin (L), were added to the Lysis Buffer to give a final concentration of 1µg/ml and phenylmethanesulfonylfluoride (PMSF), to give a final concentration of 1mM. The plates containing the lysing cells were placed on an automatic rocker in a cold room at 4°C for 30 minutes. The orientation of each plate was changed by rotating each plate on the rocker by 90° at 15 minutes into the lysing process. The lysate so produced was transferred by micropipette into microcentrifuge tubes and spun at 13,000rpm in a bench top microcentrifuge for 15 minutes. The supernatant and pellet were both collected, and after removing the supernatant the pellet was mixed with enough fresh lysis buffer to resuspend it.

### B.2.4 Protein fractionation by size exclusion chromatography.

Size exclusion chromatography separates molecules in solution according to their size. Smaller molecules are able to enter the pores of the gel matrix of the column and take longer to elute as compared with the

solution surrounding the gel where larger molecules too big to enter into the gel matrix elute faster since they are excluded from the pores. A Superdex 200 HR column (Amersham Pharmacia Biotech) was placed at 4°C in a cold room and connected to an Aekta Basic 10 chromatography System. The column was equilibrated with HEPES Buffer and calibrated as described in the manual. 200µl to 500µl of soluble protein extract was loaded on the column and eluted at a rate of 0.25 ml/min. Fractions were monitored for protein with the UV sensor and computer interface of the system and the zone of fractions containing protein were saved and stored at -80 °C.

#### B.2.5 SDS Polyacrylamide gel electrophoresis (SDS PAGE) and western blot.

Casting and running the gel: Aurogene™ PAGE Sub Vertical Electrophoresis System VS10D, VS10DSYS, VS10DCAST and OmniPAGE Molecular Electroblothing Units SB10, VS10BI were used for running the SDS PAGE and western blots. The gel plate apparatus was set up as described in the manufacturer's recommendations and resolving gel mix introduced by pipette into the space between the two glass plates. A few drops of isopropanol were placed onto the surface of the gel to ensure a flat surface and to exclude air. After 15 minutes the isopropanol was poured off and the gel surface washed with a jet of distilled water from a wash bottle. The comb was then inserted into the gap between the plates but only to about 5mm depth to allow bubbles to escape when the liquid gel was pipetted in. The stacker gel mix was pipetted into the space between the plates making sure no bubbles formed at the bottom of the

teeth of the comb and subsequently, the comb was carefully pushed down to the correct level such that wells of approximately 1cm depth would be formed by the teeth displacing the stacker gel mix. The stacker gel was left to set for 10 minutes. After removing the comb the gel in its gel plate apparatus was set up in the electrophoresis tank and 1 x SDS PAGE running buffer was poured in. Bubbles that formed at the bottom of the gel plate and in the wells of the gel were removed with a jet of running buffer ejected from a syringe and needle. The protein samples were mixed with SDS PAGE Loading Buffer X2 (1:1 vol/vol) and heated at 95°C in a heating block for 5 minutes. An aliquot of pre-stained protein marker stock was also heated at 95°C for 5 minutes in the heating block.

Western blotting and visualizing the proteins with antibodies and chemiluminescence: After running, the gel was removed from the apparatus and set up in the Aurogene™ OmniPAGE Molecular Electro-blotting Units as described in the manufacturer's recommendations, with pre-wet (Western Blot Transfer Buffer) sponges and filter papers. Transfer took approximately 1 hr at 100 volts. Subsequently, and after disassembly of the blotting sandwich, the membrane was blocked with Membrane Blocking Buffer overnight at 4°C or for one hour at room temperature. To visualize the proteins on the western blots with antibodies and chemiluminescence, the blot was sealed in a plastic bag containing Membrane Blocking Buffer with all air bubbles removed by squeezing the liquid covering the blot in the bag to the open end of the bag before sealing. The bag and contents (Membrane Blocking Buffer plus primary antibody) was incubated on a rocking platform for 60 min at room temp, or overnight at 4°C, after which the blot was washed 3 x for 5 min with TBST.



Subsequently and in a similar manner to the first incubation, the blot was incubated for one hour at room temp with the secondary antibody diluted in Membrane Blocking Buffer and then washed 3 times for 10 min with TBST. Finally, the blot was immersed in the freshly mixed chemiluminescence reagents and immediately loaded into a loaded autoradiography cassette under darkroom conditions and procedures and exposed overnight.

### B.3 Using protein fractionation, SDS-PAGE and western blot to investigate NPAS2 and HAT1 protein complex formation.

#### B.3.1 Methods.

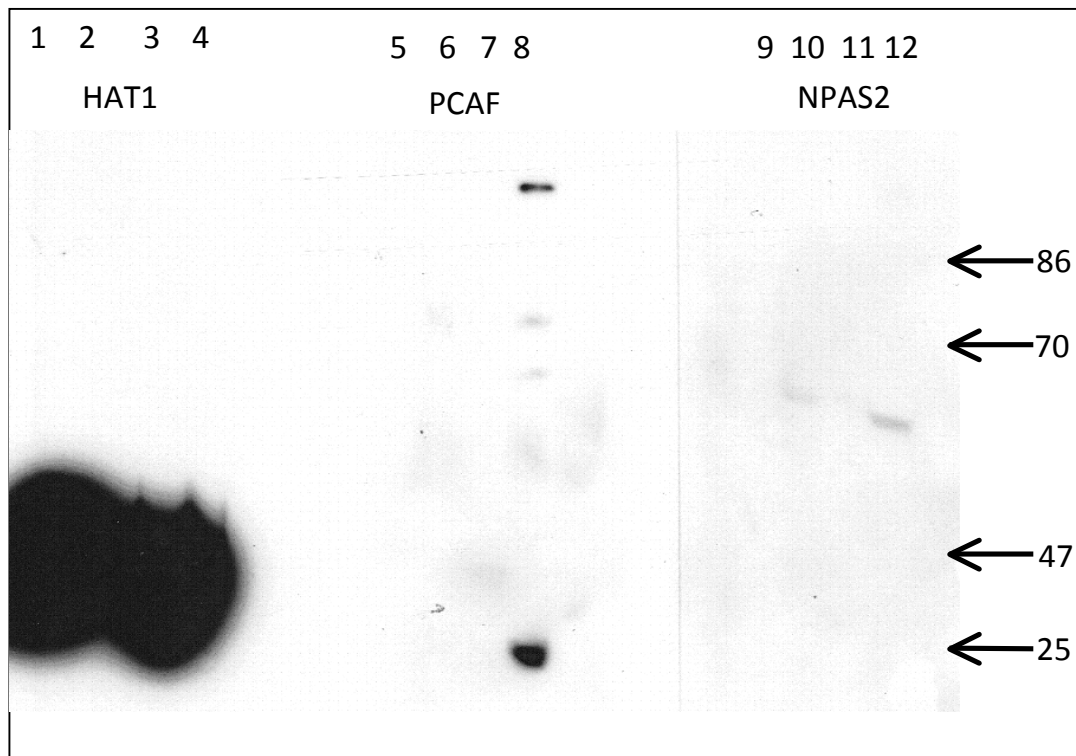
Nuclear and cytoplasmic protein extracts were prepared from HeLa cells and analysed with SDS-PAGE followed by western blot as described above. Protein fractionation (by size exclusion column chromatography) was also performed on protein extracts from HeLa cells and the fractions analysed with SDS-PAGE and western blot. The resulting filter bound proteins were challenged with antibodies against K(lysine) acetyltransferase 2B (KAT2B also known as p300/CBP-associated factor, PCAF) as a control, NPAS2 and HAT1.

The fraction number of eluate from the column correlated with the molecular weight range of native proteins and complexes in each fraction. Analyzing each fraction by SDS-PAGE segregated the individual proteins present in each fraction. Pre-stained protein size markers were used to determine the approximate molecular weight of the bands on the western blot membrane that showed reactivity with the antibodies against HAT1,

KAT2B and NPAS2.

### B.3.2 Results.

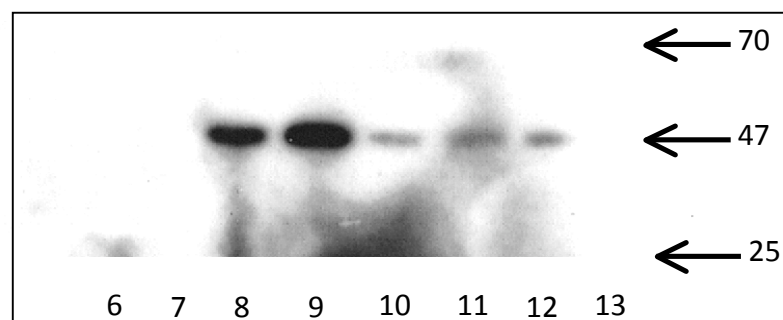
The experiment aimed to test whether cytoplasmic and/or nuclear extracts of HeLa cells contained HAT1/NPAS2 protein complexes. To test the competence of the system to detect protein bands likely to be NPAS2, HAT1 and PCAF, SDS-PAGE of nuclear and cytoplasmic protein extracts from HeLa cells, followed by western blot and probing with antibodies was carried out. A strong and clear band with the anti-HAT1 sera was seen and the pre-stained molecular weight markers indicated a molecular



**Figure B-1. Autoradiograph of western blots of nuclear and cytoplasmic protein extracts from HeLa cells.** Track1 and 2 represent cytoplasmic and nuclear extract respectively, both without protease inhibitors in the extraction mix. Tracks 3 and 4 represent cytoplasmic and nuclear protein extract with protease inhibitors. The same convention is kept for tracks 5-12. Tracks1-4 were probed with HAT1 antibody, 5-8 KAT2B (PCAF) antibody and tracks 9-12 with NPAS2 antibody.

weight for this band that was in keeping with the known molecular weight for HAT1 (~46 KDa). Only a faint band could be detected with the NPAS2 antibody. Increasing the sample loading of the gel and/or decreasing the stringency of the washes did not help to visualize a distinct band with the NPAS2 antibody. Moreover, the size of the faint band (~60KDa) (Figure B-1) was not in keeping with the molecular weight of the full length NPAS2 isoform, nor any of the other substantially smaller isoforms. The KAT2B (PCAF); (a control for the exclusion column experiment, a ~93KDa histoneacetyltransferase known to interact with a number of proteins (particularly, EP300 (264KD); TP53 (43.6KD); and CLOCK (Curtis, Seo et al. 2004)) showed 2 bands on the gel. The indicated molecular weight of the largest band was an approximate match with the molecular weight of this protein.

The sequence of fractions eluted from the non-denaturing exclusion column represented proteins and protein complexes of decreasing molecular size. Subsequent SDS-PAGE and western blot of this sequence of fractions showed cross-reactivity for HAT1 antibody in fractions 8-12 (Figure B-2). Fractions 8 and 9 appeared to contain most of the signal, with equal (reduced) reactivity in fractions 10-12.



**Figure B-2. Western blot of size-exclusion column chromatography fractions from a nuclear extract of HeLa cells.** The blot was probed with anti-HAT1 antibody. Lanes 6-13 represent fraction number and the figures with arrows indicate molecular weights in KDa. The signal covers an indicated molecular weight range of approximately 400 (fraction 8) to 200 (fraction 12) KDa.

### B.3.3 Discussion.

Following SDS-PAGE and western blot of HeLa protein extracts, HAT1 antibody in conjunction with chemiluminescent detection and autoradiography revealed a clear single band on blots of nuclear and cytoplasmic protein extracts (Figure B-1). This is in keeping with the occurrence of HAT1 as a nuclear and cytoplasmic protein. The signal for HAT1 was strong and similar results were obtained with or without protease inhibitors in the extraction buffer. For PCAF (KAT2B) that was to be used as a control in the column experiment, clear consistent bands were seen but protease inhibitors were required in the extraction medium. The PCAF antibody in addition to detecting a single clear band of appropriate size (~94 KDa) also detected some subsidiary bands. The product information suggested this was likely and the size of the lower molecular weight band was in keeping with an isoform of PCAF. Thus it appears that the extraction and detection system worked efficiently for HAT1 and PCAF.

For NPAS2 and despite repeated trials, no strong, clear bands ~90KDa could be seen with the NPAS2 antibody. The faint band seen at ~60KDa (Figure B-1) is present only in the tracks representing nuclear extract which is in keeping with NPAS2 as a nuclear protein. The NPAS2

antibody is raised against a partial recombinant NPAS2 corresponding to a fragment of the 18<sup>th</sup> exon of this 20-exon, 92KDa protein. NPAS2 has seven isoforms in humans: NPAS2-001, NPAS2-004, NPAS2-005, NPAS2-007, NPAS2-010, NPAS2-012 and NPAS2-201. Isoforms 004, 007, 010 and 012 are all approximately a quarter of the size of the full-length isoform in terms of amino acid number whereas isoform 005 is half this size. It appears unlikely therefore that this faint band ~60 KDa relates to one of these isoforms but rather suggest it is showing cross reactivity to some other component in the nuclear extract.

The results of the column chromatography suggest that HAT1 does form complexes with other proteins in the cell extract. The observation that HAT1 is found in fractions corresponding to proteins over 150 kDa does not rule out the possibility that HAT1 might complex with NPAS2. To further ascertain which proteins might form complexes with HAT1, stripping and re-probing the blots of the fractions showing HAT1 cross reactivity was planned. Subsequently reprobing with NPAS2 antibody might show cross reactivity with protein within the same fraction. A positive result, suitably controlled, could suggest that NPAS2 in the protein extract is complexed with another protein, possibly HAT1. Further tests would then be required, such as co-immunoprecipitation and yeast two-hybrid tests to validate this possibility. Costs, limited time and the imperative that investigations showing promise and that focused on the direct implications of the noncoding SNPs already showing positive association with autism would be favoured determined that this preliminary investigation though interesting was not pursued further.

## Appendix C

### ***ZNF804A* and the SNP rs1344706: A psychosis-susceptibility variant and its possible relevance to autism**

#### **C.1 Introduction.**

*ZNF804A* encodes a zinc finger protein of currently unknown function. Genetic analysis of *ZNF804A* shows positive association for SNPs in this gene with schizophrenia and bipolar disorder individually and more so when the data sets for these disorders are merged (Williams et al. 2010). The primary association of *ZNF804A* with schizophrenia susceptibility has thus been superseded by the idea that variations in this gene confer susceptibility to psychosis.

A large-scale retrospective analysis of autism comorbidities indicated genetic overlap between autism, schizophrenia and bipolar disorder (Rzhetsky et al. 2007) and this concept is reinforced by genetic studies (Carroll and Owen 2010). In order to investigate this probable genetic overlap between schizophrenia autism and bipolar disorder, *ZNF804A* was screened for inclusion in autism databases and with bioinformatics RNA structure analysis tools.

## **C.2 Method.**

Literature searches for *ZNF804A* were made followed by searches using The Autism Chromosomal Rearrangement Database (ACRD Table 3-1). Ensembl was also used to define the chromosomal location of *ZNF804A* and to look for paralogues. The ACRD was also searched for inclusion of any paralogues of *ZNF804A*. Bioinformatics transcript analysis was performed as described above (chapter 8) for the clock genes.

## **C.3 Results.**

### C.3.1 Autism Chromosomal Rearrangement Database.

Using *ZNF804A* as the query in searches with the ACRD, positive results were found for *ZNF804A* on chromosome 2q 32.1 where three reports of autism-associated micro-indels that contained the *ZNF804A* gene were listed (Marshall et al. 2008). Using Ensembl, searches were made for paralogues. *ZNF804B* was the only human paralogue of *ZNF804A*. This paralogue was also used as the query in searches with the ACRD. Intriguingly *ZNF804B* was also to be found amongst the autism-associated micro-indels (Table C-1) and further the nature of the anomalies (insertions) were in keeping with the effects of the pathological A allele of rs1344706 that increases the mRNA expression level of *ZNF804A* (Riley et al. 2010).

There is no mention of *ZNF804A* in the PubMed or Thompson Web of Science database in relation to autism, though *ZNF804A* has possible implication in cerebellum function and interacts with *ATXN1* (Lim et al. 2006).

**Table C-1. *ZNF804A/B* in autism cases from the ACRD.**

<b>Locus</b>	<b>Gain/Loss</b>	<b>Cytoband</b>	<b>Genes in each locus</b>
chr2:184152739..186310739	gain	2q32.1	<i>ZNF804A</i>
chr2:184505739..186310739	gain	2q32.1	<i>FLJ44048</i>
chr2:185184739..186774739	gain	2q32.1	
chr7:87960048-89325785	gain	7q21.13	<i>C7orf62</i>
	gain	7q21.13,	<i>ZNF804B</i>

With reference to Figure 1-1, it can be seen that the genes *ZNF804A* is located in one of the 12 recurrent autism indels recorded on the human karyotype diagram and *ZNF804B* is located in one of 26 de novo indels (Marshall et al. 2008).

### C.3.2 Candidate MIR scans using MirBASE and The Vienna RNA Web Server.

Analysing the sequence block containing the SNP rs1344706 gave a long, 145nt hairpin formation where rs1344706 alters the stability of the structure (Figure C-1, Figure C-2 and Table C-2). This candidate microRNA is predicted with Targetscan (Table 3-1) to have: p21 protein (Cdc42/Rac)-activated kinase 2 (*PAK2*); G patch domain containing 8 (*GPATCH8*); and, *ZNF804A* amongst its primary targets. BLAST searches show *GPATCH8* is the most similar protein to *ZNF804A* (Figure C-3) although the homology between *ZNF804A* and *GPATCH8* is restricted to the N-terminal end of the two proteins, encoding the zinc finger domains Znf\_C2H2 (IPR007087) and Znf\_U1 (IPR003604).

MirBASE was searched for homologies with known microRNAs however no homologies could be found that matched the stem regions of this hairpin. Some weak matches were found across the loop region or across regions of the hairpin that would define sequences too short to be



considered realistic mature microRNA candidates. This hairpin did not show sequence conservation amongst primates when analysed with the Vienna RNA Web Server.

**Table C-2. Stability of the ZNF804A hairpin containing the SNP rs1344706**

<b>ZNF804A rs1344706 hairpin</b>	<b>C Allele</b>	<b>A Allele</b>
MFE	<b>-42.90</b> kcal/mol	<b>-42.90</b> kcal/mol
MFE: thermodynamic ensemble	<b>-46.49</b> kcal/mol.	<b>-46.27</b> kcal/mol.
Frequency: MFE structure	<b>0.30</b> %.	<b>0.42</b> %.
Ensemble diversity	<b>54.15</b> .	<b>48.49</b>
Centroid secondary structure	<b>-23.21</b> kcal/mol	<b>-33.80</b> kcal/mol

Further hairpin scans were made throughout the intron of *ZNF804A* containing rs1344706 with particular attention given to the region with the high scoring SNP cluster as shown in Figure C-5. Intriguingly a ~349nt sequence was found within the region bounded by the cluster of high scoring SNPs that is identical to a sequence in *UHMK1* (Figure C-6). It is exceedingly unlikely that this match has occurred by chance ( $2.8e^{-262}$  according to Ensembl BLAST calculations). *UHMK1*, like *ZNF804A* is associated with schizophrenia (Puri et al. 2008), tentatively indicating possible relevance of this feature to schizophrenia. In *UHMK1* the *ZNF804A* homologous sequence is to be found within the 3'UTR .

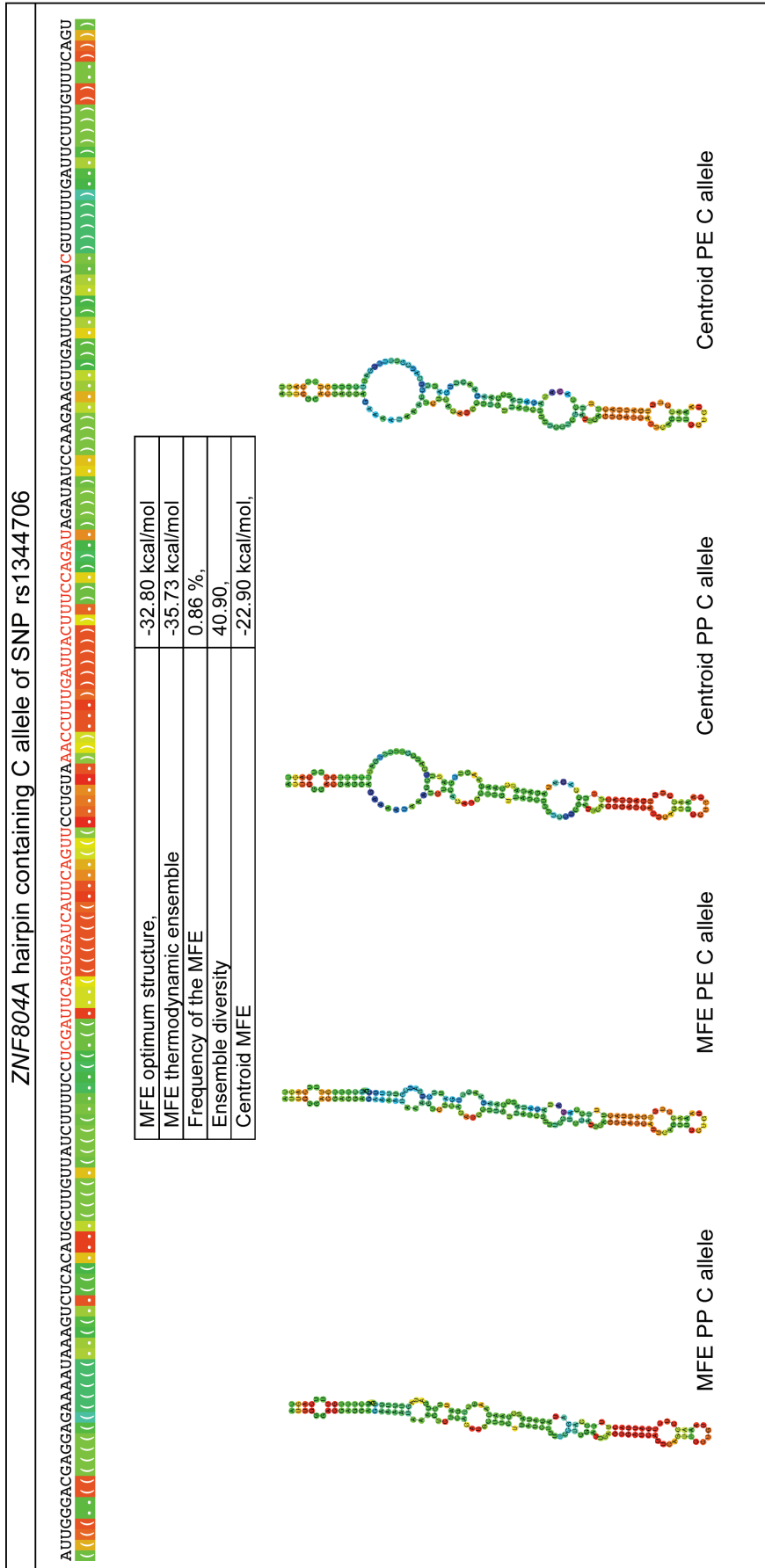


Figure C1. The sequence containing the SNP rs1344706 is predicted to form a long hairpin with the most stable region close to the loop. The SNP is marked in the sequence by the red C. The mean free energy (MFE) structures are shaded to show positional probability (MFE PP where red = 1 and purple = 0) and positional entropy (MFE PE where red = 0 and purple = 3) The centroid structures follow the same convention.

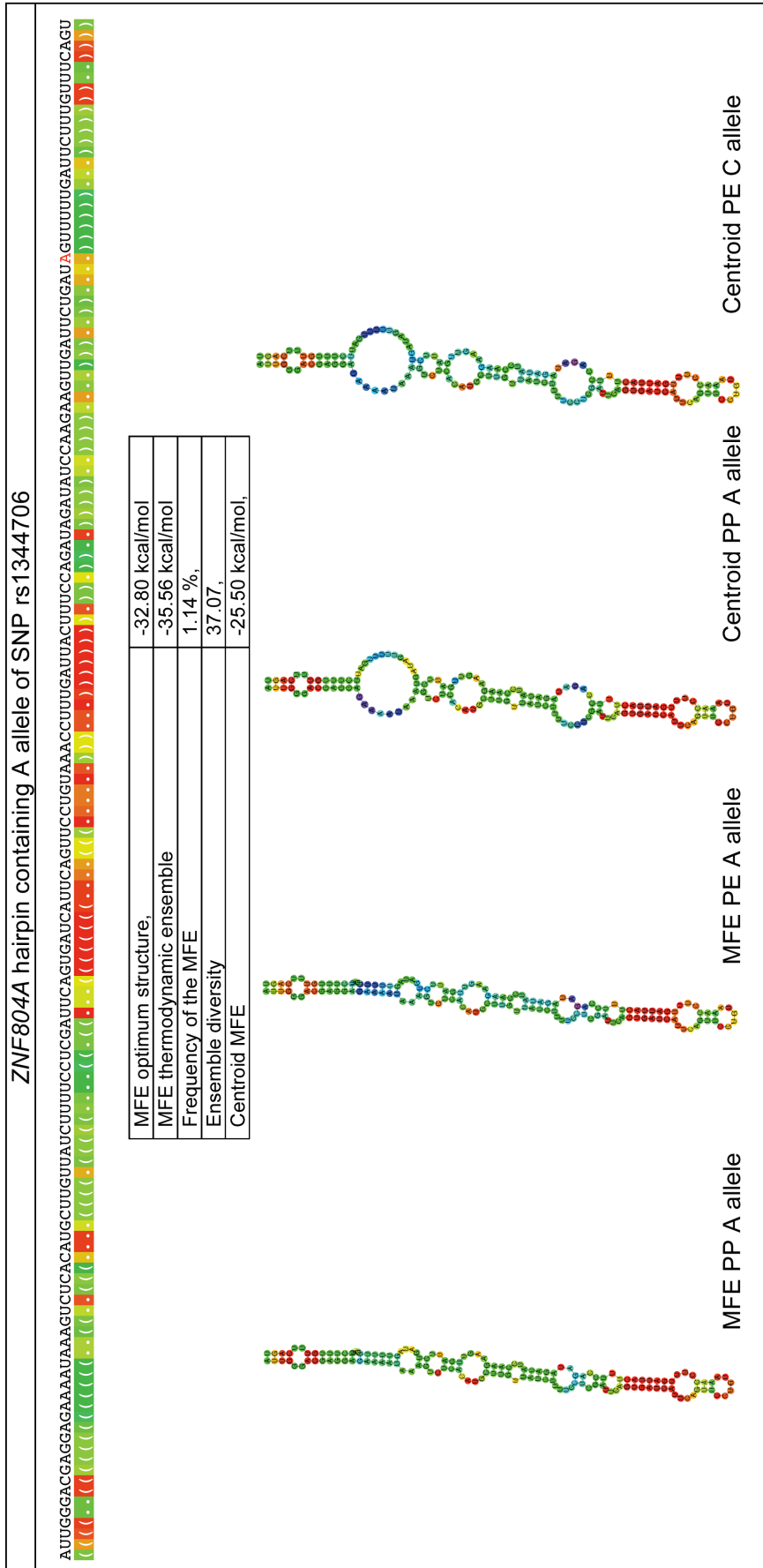
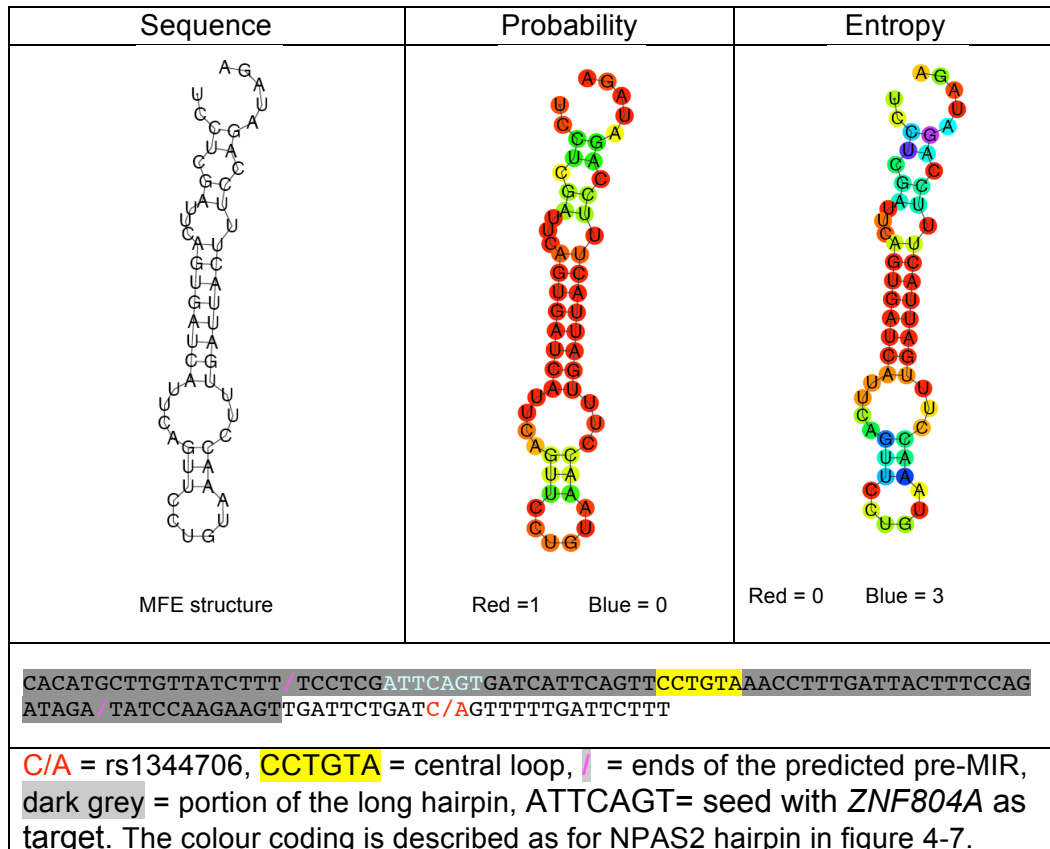


Figure C2. The ZNF804A sequence containing the SNP rs1344706 is predicted to form a long hairpin with the most stable region close to the loop, the SNP is marked in by the red A in the sequence. The mean free energy (MFE) structures are shaded to show positional probability (PP) where red = 1 and purple = 0 and positional entropy (PE) where red = 0 and purple = 2.6) The centroid structures follow the same convention.

GPATCH	KSLQGRDPIPIVVKYDVMGMGRMEMELDYAEDATERRRVLEVEKEDTEELRQKYKDYVD	120
ZNF	-----MECYIIVISSHLSNG-----HFRNIKGVFRGPLSKNGKNTLDYAE	41
	: **:. :. * . *: * : .: :* **:	
GPATCH	KEKAIKALEDLRANFYCELCDKQYQKHQEFDNHINSYDHAHKQRLKDLKQREFARNVSS	180
ZNF	KENTIKALEDLKANFYCELCDKQYYKHQEFDNHINSYDHAHKQRLKELKQREFARNVAS	101
	**.:*****:***** *****:*****:*****:	
GPATCH	RSRKDEKKQEKALRRLHELAEQRKQAECPGSGPMFKPTTVAVDEEGEDDKDESATNSG	240
ZNF	KSRKDERKQEKALQRLHKLAE LRKETVCAPGSGPMFKSTTVTVRENCNEISQRVVDSVN	161
	:*****:*****:***:*** **.: *****:***:* *:. * ..	
GPATCH	TGATASCGLGSEFSTDKGGPFTAVQITNTTGLAQAPGLASQGISFGIKNNLGTPLQKLG	300
ZNF	NQQDFKYTLIHSEENTKDATTVAEDPESANNYTAKNNQVGD-----QAQGIHRHKIGF	214
	. . * . . * . . * : . . . : . . . : : * :*:*	

**Figure C-3. Alignment of protein sequence fragment of ZNF804A and GPATCH8 showing conservation of the amino acid sequence of the zinc finger DNA binding domain.**



**Figure C-4. Structure of a candidate pre-MIR in the ZNF804A hairpin.**

**Table C-3. Predicted DROSHA scores for the ZNF804A hairpin**

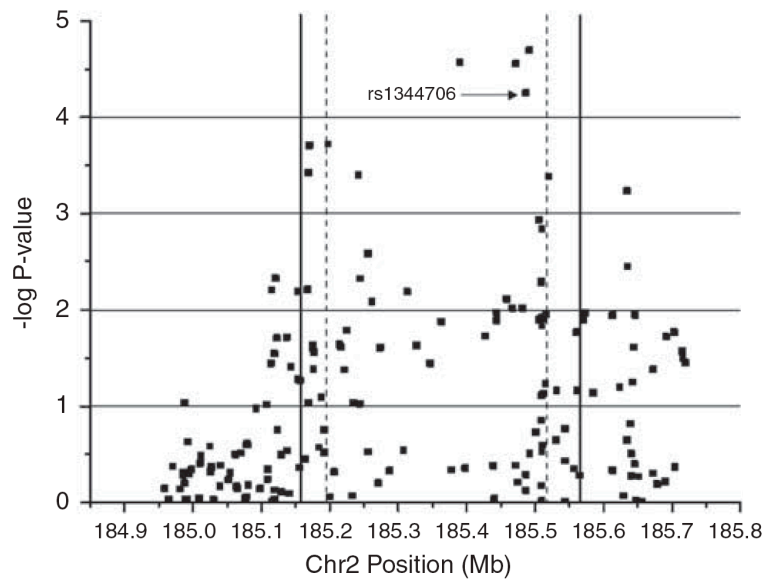
Predicted DROSHA site score	SNP
0.228	rs1344706 A allele
0.173	rs1344706 C allele

**Table C-4. Targets of the ZNF804A candidate mir 5'arm.**

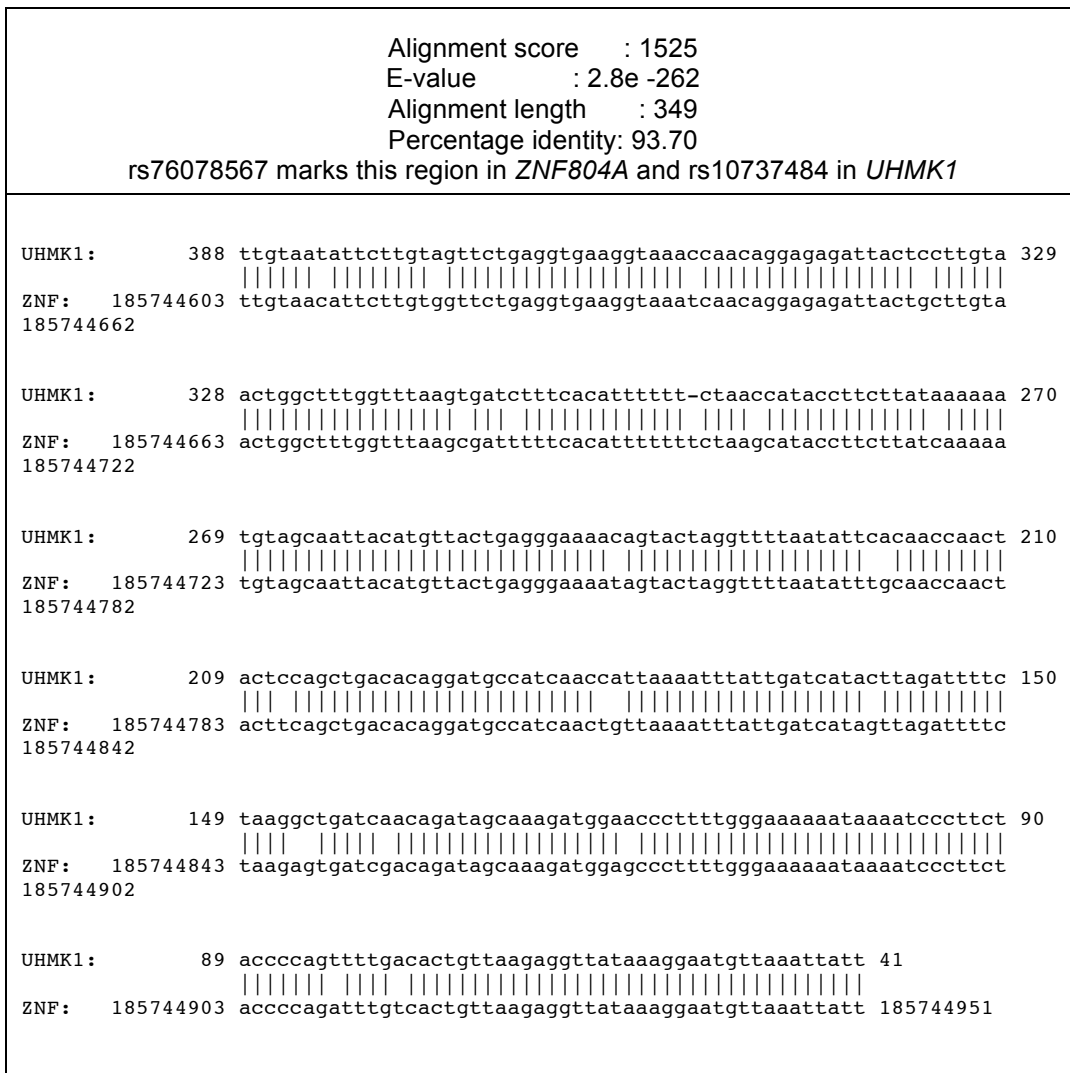
Candidate seed 5'	Number of matches	Schizophrenia gene targets
TCGATTC	5	none
CGATTCA	6	none
GATTCAG	207	<i>PAK2</i>
ATTCAGU	341	<i>ZNF804A</i>

**Table C-5. Targets of the ZNF804A candidate mir 3'arm.**

Candidate seed 3'	Number of matches	Schizophrenia gene targets
AACCUUU =242	5	none
ACCUUUG =251	6	none
CCUUUGA =341	207	none



**Figure C-5. P-values for genotyped SNPs in ZNF804A.** Reproduced from Williams et al. (2010).



**Figure C-6. Sense-antisense complementarity between ZNF804A and the 3'UTR of UHMK1.**

The position of this antisense complementary region in relation to the schizophrenia associated SNPs in ZNF804A is between rs1344706 (indicated with the arrow) and the SNP immediately to its left (Figure C-5). The UHMK1 sequence and the ZNF804A sequence are anti parallel in relation to their host genes, forward on UHMK1 and reverse on ZNF804A. Nevertheless the presence of an expressed pseudogene and ESTs for the reverse strand of ZNF804A at this location and ESTs from the forward and reverse strands of UHMK1 at this position indicates that RNA transcripts

with matching sequence are possible from either strand. No common variation is indicated within the *ZNF804A/UHMK1* complementary sequence. The SNP rs76078567 is immediately adjacent to this region but Williams et al. (2010) did not investigate this SNP.

#### C.4 Conclusion.

The schizophrenia risk allele (A) of rs1344706 is also associated with increased mRNA levels of *ZNF804A* but it appears that rs1344706 is not the eQTL and it is more likely to be in linkage disequilibrium with the true eQTL (Williams et al. 2010). The results from the Autism Chromosomal Rearrangement Database also shows that the autism-associated indels for *ZNF804A* and *ZNF804B* are all duplications indicating that over expression of *ZNF804A* may be relevant in autism as well as in schizophrenia.

The candidate microRNA indicated above is predicted (Targetscan) to target the schizophrenia associate genes *PAK2* (Mulle et al. 2010) and *ZNF804A* itself, as well as *GPATCH8* (that shows homology with the N terminal half of *ZNF804A* (Figure C-3)). The candidate microRNA is indicated to be more stable and have better Interagon SVM DROSHA score when the A allele of rs1344706 is present in the hairpin (Table C-3, Figure C-1 and Figure C-2). This is at odds with the A allele being associated with increased expression of *ZNF804A* in schizophrenia (Riley et al. 2010).

The likelihood of the *ZNF804A / UHMK1* antisense complementarity being due to chance is extremely remote (E-value  $2.8e^{-262}$ ) and the position of this feature in the 3'-UTR of *UHNK1* suggests that it could be a

regulatory natural antisense transcript potentially allowing the expression of *ZNF804A* to regulate the function of *UHNK1* or vice versa (Sun, Hurst et al. 2005; Faghihi and Wahlestedt 2009). A possible mechanism of action of *ZNF804A* in schizophrenia might be via *UHMK1* (functionally linked with RNA transport and axon development in neurons). SNP rs1344706 regulates *ZNF804A* expression directly and concurrently regulates *UHMK1* indirectly, in trans, through sense-antisense silencing (Cambray et al. 2009). This hypothesis implicates *ZNF804A* and *UHMK1* in the same genetic/biological system with regards to schizophrenia risk, although no genetic interaction for *ZNF804A* and *UHMK1* has been reported to date.

## Appendix D

### **Human *PER1*: a comparison with *Drosophila* *PERIOD* in relation to the K&H cycle determining region and an autism/cancer associated haplotype.**

#### **D.1 Introduction.**

The region of the *Drosophila* *PERIOD* gene (*per*) that determines the K&H cycle phenotype, is centrally located in the gene and immediately downstream of the region coding for the Thr-Gly repeats (Wheeler et al. 1991). The autism-associated haplotype also covers a relatively small central region of human *PER1*. With the autism Social Timing hypothesis in mind (Wimpory et al. 2002), the question arises, of whether the autism-linked



haplotype in *PER1* covers the region corresponding to the Kyriacou and Hall cycle determining region of *Drosophila per* (here called the KDR).

The human *PER1* locus is known to derive two protein isoforms, long and short. The short form superficially appears to be the long isoform terminated at a point corresponding to the mid region of the long form. The indication that the autism-associated SNP rs885747 and other SNPs within the most significant autism haplotype for *PER1* may influence splicing (Chapter 4.2) prompts investigation of the difference in biological potential of the two isoforms and, if recognized, the relevance of the KDR in *PER1*.

## **D.2 Method.**

Ensembl was used to scrutinize the overall protein domain structure of the two isoforms of *PER1*. Ensembl protein sequences of *PER1* and *Drosophila* *PERIOD* were also downloaded for analysis. Alignment of the protein sequences with ClustalW2 (Table 3-1) was followed by identification of the Thr-Gly repeat in the *Drosophila* sequence and corresponding regions in the human *PER1* sequence noted. The alignments took account of the two isoforms of human *PER1*.

The region immediately downstream of the Thr-Gly repeat was searched for possible protein modification sites and other functional motifs with the MyHits server and Ensembl tools (Pagni et al. 2004; Pagni et al. 2007). Alignments of the two isoforms of *PER1* (long, and short (Figure D-2)) were also used to determine whether any KDR-like region was missing from the short isoform. The autism-associated haplotype was also mapped onto the protein sequences to enable comment on its location, relative to

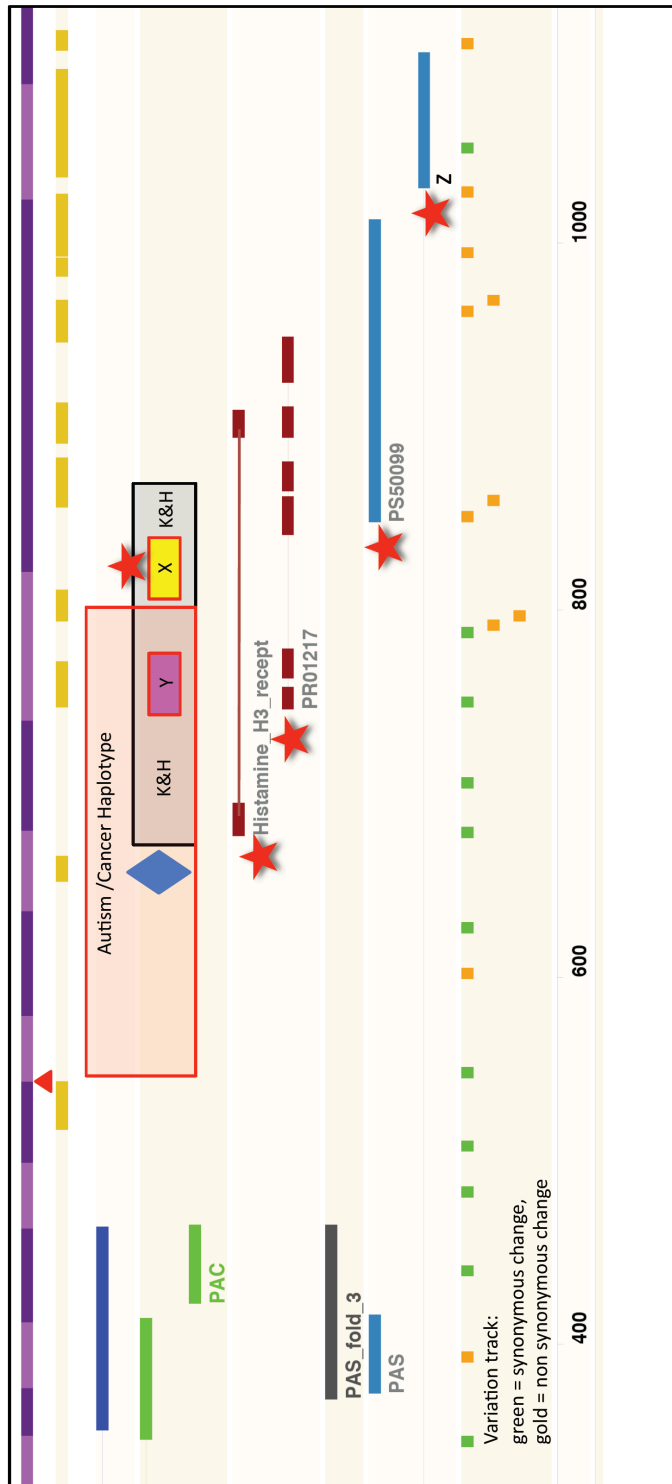
the short and long isoform. Any different functional capacity (suggested by differing domains and signatures) between the isoforms was noted and related to how this could have biological relevance.

### **D.3 Results.**

The KDR-like fragment was found to be present in the long isoform of human PER1 (Figure D-1) but the sequence of 56 amino acids towards the C-terminal end of the KDR-like region is altered in the short isoform (Figure D-4). The protein transduction domain, the histamine receptor domain, the proline rich region and the serine rich region are also absent in the short isoform (Figure D-1).

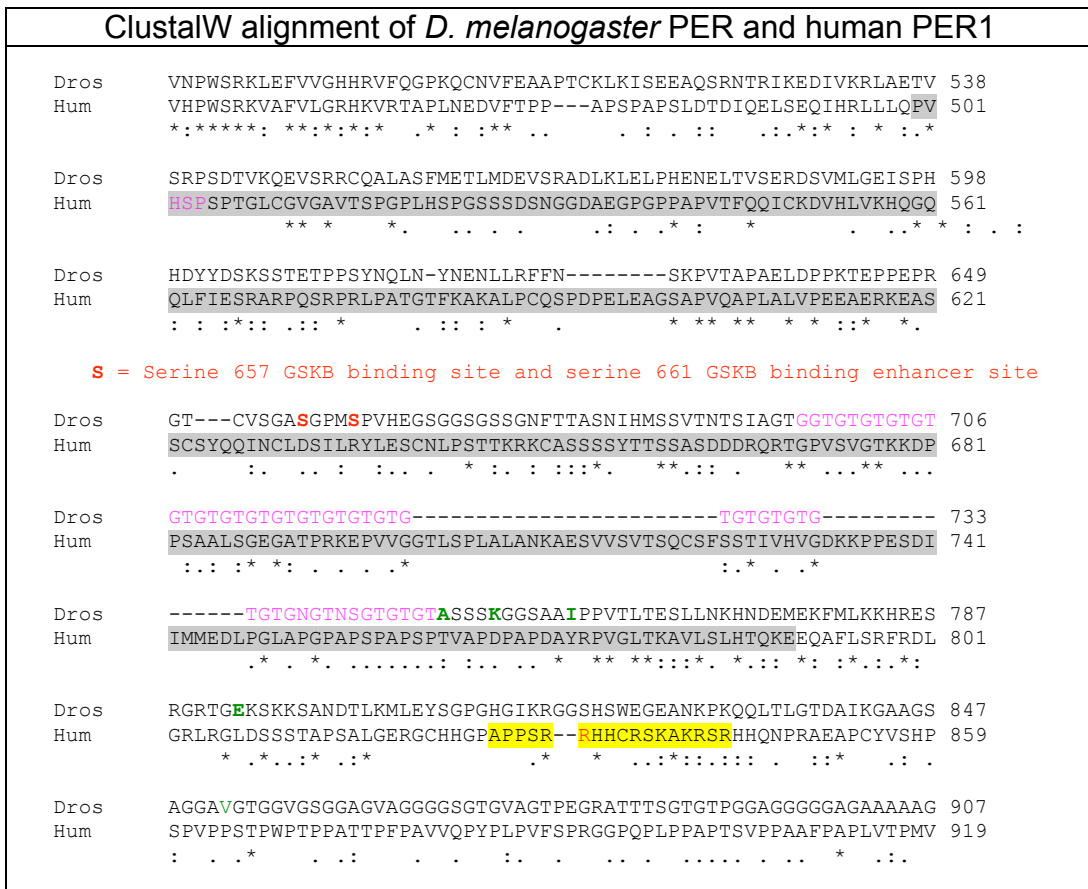
The *Drosophila* PER and human PER1 protein sequence alignments, showed that a region of PER1 that was similar to the *Drosophila* Thr-Gly repeats was identified within the autism haplotype (Figure D-2). Local re-alignment of the *Drosophila* Thr-Gly repeat and the corresponding human sequence showed that the downstream ends of these repeat regions show an identical pattern with respect to threonine residues (*Drosophila*) and proline residues (human). A phosphorylation site, at 15 or 16 amino residues from the ends of the *Drosophila* and human repeat regions respectively was also noted (Figure D-4).

The KDR in the *Drosophila* sequence was found to contain several predicted phosphorylation sites, as did the corresponding region of the human sequence (Figure D-3). This region of the human PER1 sequence was also found to contain a protein transduction zone (Yang et al. 2005) and to match the DOUBLETIME (CSNK1E) binding domain in *Drosophila* PER (Kim et al. 2007) (Figure D-3).



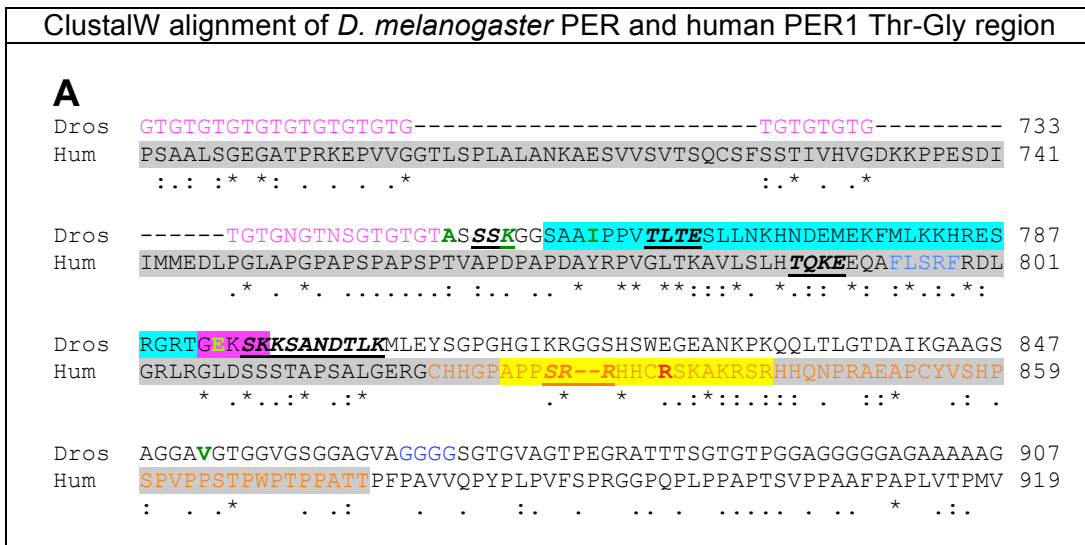
**Figure D1. Protein domain structure of PER1.**

Protein domain structure in the long isoform of PER1. The purple and pink bar along the top of the diagram shows a fragment of the 23 exon PER1 protein sequence. The red triangle indicates the intron 12 excision point. The red box shows the extent of the autism haplotype and the grey box shows the K&H cycle determining region-like sequence. The bright yellow bar (X) shows the location of the protein transduction domain and the bright pink bar (Y) the location of the Thr-Gly like repeats. Note that the short isoform loses the domains indicated with red stars. (The protein transduction domain; the extensin and proline rich domains (PRO1217 and PS50099) and the terminal serine rich region (Z)). The scale bar at the bottom gives the amino acid number of the protein sequence. The blue bar indicates the approximate position of the GSK3 phosphorylation site on the *Drosophila* alignment.



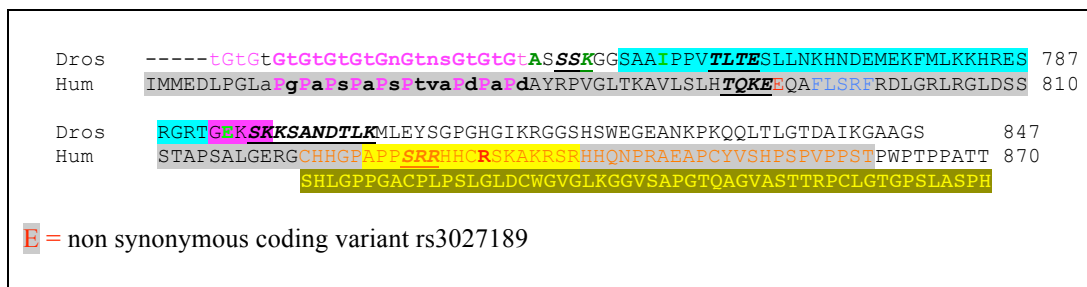
**Figure D-2. Protein sequence alignment: *Drosophila* PER and human PER1.**

The sequence segment shaded grey indicates the protein region that is covered by the most significant autism-associated haplotype. The protein transduction domain is shown in yellow and the K&H cycle mutations are shown in green letters (Wheeler et al. 1991). The pink lettering is the *Drosophila* Thr-Gly repeat. Red letters describe the GSK3B phosphorylation sites in *Drosophila* (Ko et al. 2010). Amino acid number is given at the right of the sequence.



**Figure D-3. Alignment of *Drosophila* PER and human PER1; KDR region detail.**

Sequence **A** shows ClustalW alignment of human PER1 (long isoform) and *Drosophila melanogaster* protein sequences. The bold black underlined letters are MyHits phosphorylation site predictions (CSNK1E/D, PKC). The start of the first phosphorylation site (human sequence) is coincident with the end of the autism haplotype. Residues that show species specific variation in *Drosophila* are shown in green (Wheeler et al. 1991). Turquoise highlight shows *Drosophila* PER DOUBLETIME (CSNK1E/D) binding domain (Kim et al. 2007) Grey shading with black lettering in the human sequence shows where PER1 (long) and PER1 (short) are identical. Grey shading with orange lettering shows the limit of PER1 (short). This terminal region of the short isoform, in orange, differs from that of PER1 long (long displayed). The C-terminal end of the KDR is indicated by the 3 G residues, shaded blue. The yellow highlight shows the protein transduction domain. The FLSRF motif (blue letters) is suggested to be a mammalian CSNK1E/D binding domain (Kim et al. 2007). The GEKSK motif in the *Drosophila* sequence overlapping a species-specific amino acid variation is predicted to be a nuclear localization signal (MyHits, Table 7-1).



**Figure D-4. Local alignment of *Drosophila* PER sequence and human PER1 centered on the Thr-Gly repeat region.**

Conservation of a repeating pattern of amino acid residues is seen for the end of the Thr-Gly repeat region (pink letters). Other colour codes are as for Figure D-3 above. The sequence shown with yellow letters and olive highlight is the C terminal region of the short PER1 isoform, that differs from that of the long as shown.

#### D.4 Conclusion

Sequence conservation allows the tentative identification of a region of the human PER1 protein that corresponds to the KDR of *Drosophila* PER. This KDR-like sequence in PER1 contains a functional membrane transduction domain (Yang et al. 2005). The corresponding sequence in *Drosophila* is different but as it is not feasible to predict with certainty whether such a domain is present on sequence alone, it would be interesting to discover whether this region of *Drosophila* PER also endows membrane transduction capability (Yang et al. 2005). The region also contains what appears to be a conserved repeat pattern identical to the terminal region of the *Drosophila* Thr-Gly repeat region (although the sequence of amino acids is not conserved) suggesting possible structural/functional conservation.

The autism haplotype in *PER1* spans exon 11 to 16 in the (20 exon) long isoform and corresponds to the last 6 exons of the short isoform (Figure D-1). An almost identical region is also delineated by SNPs associated with increased prostate cancer risk, especially the aggressive form of prostate cancer (Zhu et al. 2009). This autism haplotype contains a number of common variants that determine essential and alternative splice sites that may alter the protein structure in this region or may affect which isoform of *PER1* is expressed (Ladd and Cooper 2002).

Notably, the 5' limit of the autism and cancer haplotypes is delineated by the SNP rs885747, that this study indicates is a polymorphic splice site enhancer/suppressor (Chapter 4). The association of the same SNP and region in cancer and autism tentatively suggests that rs885747 might alter the strengths of the splice sites within this central region and possibly influence the balance of alternatively spliced isoforms.

A GSK3B phosphorylation site is present in the *Drosophila* sequence that is indicated to correspond to the centre of the autism/cancer haplotype. In *Drosophila* this phosphorylation site is important in *PER* transport from cytoplasm to nucleus (Ko et al. 2010). It is tempting to suggest that the genetic association of these SNPs is reflecting the effect of these and/or other SNPs (in linkage disequilibrium within the haplotype region) on a GSK3B phosphorylation site in mammalian *PER1*.

The *Drosophila* species-specific song-variants are however downstream of the GSK3B phosphorylation site and the Thr-Gly repeat. These amino acid changes are adjacent to a region of *Drosophila* *PER* known to bind the clock regulating kinase, DOUBLETIME (CSNK1E/D) (Kim et al. 2007). The boundaries of this binding region are seen here to

have further phosphorylation sites and a nuclear transportation signal motif that spans one of the *Drosophila* species-specific amino acid variants. This is in keeping with the understanding that multiple phosphorylation events, at multiple sites on PER, determine the rate of degradation and the balance of levels of nuclear and cytoplasmic PER (Leloup and Goldbeter 2011). Perhaps the circadian, ultradian and temperature compensating effects of PER may be resolved by a mechanism involving: the balance of cytoplasmic and nuclear PER; a role for PER in alternative splicing (including PER mRNA); variation that affects phosphorylation/location, and, variation that affects splicing.

With regards to the domain structure of the two isoforms of human PER1, the short and long isoform share PAS PAC domain-containing regions. The long isoform however, is distinguished as being the only one of the isoforms that has a histamine H3 receptor domain and it is likely that the different domain composition of each isoform is reflected in different functionality of the isoforms. The H3 receptor domain may indicate a neuronal signaling role for the long isoform since this domain is found in H3 histamine receptors in the presynaptic regions of histamine-containing neurons (Lovenberg et al. 1999). The *Drosophila* protein contains a Kv1.6 voltage-gated calcium channel signature in this region, tentatively suggesting that that some neuronal function for this region of the protein has been conserved.

The proline rich signature is implicated in mechanochemical functions linked to cell adhesion and vesicle trafficking (Williamson 1994), two processes that have relevance to cancer development and synaptogenesis. Proteins containing proline rich regions are an ill-defined



class in terms of their sequence conservation. Nevertheless, poly-proline regions like those in PER1 have been implicated in strong and rapid molecular binding processes of low specificity typical of the vesicle-associated proteins, the RNA polymerase II pre-initiation complex, and the SH3 domain binding proteins that regulate hormone receptors.

The presence of the protein transduction and histamine receptor domains together with the indication that the proline rich extensin signature may be involved in vesicle trafficking has resonance with recent implication of synaptic vesicle recycling in the clock function of the SCN (Deery et al. 2009) and with circadian rhythms of vesicle size and location in *Drosophila* motor terminals (Ruiz et al.). Could PER1 (long) have some kind of SNARE or actin attachment function?

Valproate and lithium are both used in the treatment of autism (Gillberg 1991; Myers 2007) and both drugs have powerful effects on circadian rhythms where lithium is an inhibitor of GSK3B (Figure 1-3) (Cordeiro et al. 2004; Dokucu et al. 2005; Tsujino et al. 2007). Peculiarly, valproate administration in a narrow time window during pregnancy can also cause autism (Moore et al. 2000) and this effect has been modeled in rat (Schneider and Przewlocki 2005).

Both valproate and lithium interfere with synaptic vesicle associated proteins (Cordeiro et al. 2004) and the effect of valproate on vesicle trafficking is highly conserved (Bellringer et al. 1988; Miyatake et al. 2007). In a yeast-model of valproate action, the acid inhibits the action of vps45. *VPS45* is highly expressed in the brain and is indicated to facilitate vesicle transport from the Golgi complex to synaptic vesicles (El-Husseini et al. 1997). Valproate and lithium both affect trafficking of large dense-core

vesicles (LDCVs) (Cordeiro et al. 2004) and anomalies in dense-core granules, which closely resemble LDCVs are found in blood platelets of autism cases. Further the D-Box and E-box containing neurobeachin (*NBEA*) and secretory carrier membrane protein 5 (SCAMP5) have each been genetically implicated in autism and regulate secretion of LDCVs in mouse beta-TC3 cells (Castermans et al. 2010). Inhibition of GSK3B activated migration of PER1, that affects vesicle trafficking via clock-controlled genes, might also contribute an effect on the cell's vesicle transport capacity if the long isoform of PER1 does have vesicle interacting function.

## **References**

- Abe, M., Herzog, E. D., Yamazaki, S., Straume, M., Tei, H., Sakaki, Y., Menaker, M., and Block, G. D. (2002). "Circadian rhythms in isolated brain regions." *Journal of Neuroscience*, 22(1), 350-356.
- Abelson, J. F., Kwan, K. Y., O'Roak, B. J., Baek, D. Y., Stillman, A. A., Morgan, T. M., Mathews, C. A., Pauls, D. A., Rasin, M. R., Gunel, M., Davis, N. R., Ercan-Sencicek, A. G., Guez, D. H., Spertus, J. A., Leckman, J. F., Dure, L. S., Kurlan, R., Singer, H. S., Gilbert, D. L., Farhi, A., Louvi, A., Lifton, R. P., Sestan, N., and State, M. W. (2005). "Sequence variants in SLITRK1 are associated with Tourette's syndrome." *Science*, 310(5746), 317-320.
- Abu-Elneel, K., Liu, T., Gazzaniga, F. S., Nishimura, Y., Wall, D. P., Geschwind, D. H., Lao, K. Q., and Kosik, K. S. (2008). "Heterogeneous dysregulation of microRNAs across the autism spectrum." *Neurogenetics*, 9(3), 153-161.
- Ageta, H., Murayama, A., Migishima, R., Kida, S., Tsuchida, K., Yokoyama, M., and Inokuchi, K. (2008). "Activin in the brain modulates anxiety-related behavior and adult neurogenesis." *Plos One*, 3(4).
- Ai, X., and Parthun, M. R. (2004). "The nuclear Hat1p/Hat2p complex: A molecular link between type B histone acetyltransferases and chromatin assembly." *Molecular Cell*, 14(2), 195-205.
- Akashi, M., and Takumi, T. (2005). "The orphan nuclear receptor ROR alpha regulates circadian transcription of the mammalian core-clock Bmal1." *Nature Structural & Molecular Biology*, 12(5), 441-448.
- Akshoomoff, N., Pierce, K., and Courchesne, E. (2002). "The neurobiological basis of autism from a developmental perspective." *Development and Psychopathology*, 14(3), 613-634.

- Al-Reefy, S., Osman, H., Jiang, W., and Mokbel, K. (2010). "Evidence for a pro-apoptotic function of RACK1 in human breast cancer." *Oncogene*, 29(41), 5651-5651.
- Alarcon, M., Abrahams, B. S., Stone, J. L., Duvall, J. A., Perederiy, J. V., Bomar, J. M., Sebat, J., Wigler, M., Martin, C. L., Ledbetter, D. H., Nelson, S. E., Cantor, R. M., and Geschwind, D. H. (2008). "Linkage, association, and gene-expression analyses identify CNTNAP2 as an autism-susceptibility gene." *American Journal of Human Genetics*, 82(1), 150-159.
- Alarcon, M., Yonan, A. L., Gilliam, T. C., Cantor, R. M., and Geschwind, D. H. (2005). "Quantitative genome scan and Ordered-Subsets Analysis of autism endophenotypes support language QTLs." *Molecular Psychiatry*, 10(8), 747-757.
- Allada, R., White, N. E., So, W. V., Hall, J. C., and Rosbash, M. (1998). "A mutant *Drosophila* homolog of mammalian Clock disrupts circadian rhythms and transcription of period and timeless." *Cell*, 93(5), 791-804.
- Alt, S., Ringo, J., Taly, B., Bray, W., and Dowse, H. (1998). "The period gene controls courtship song cycles in *Drosophila melanogaster*." *Animal Behaviour*, 56, 87-97.
- Altafaj, X., Dierssen, M., Baamonde, C., Marti, E., Visa, J., Guimera, J., Oset, M., Gonzalez, J. R., Florez, J., Fillat, C., and Estivill, X. (2001). "Neurodevelopmental delay, motor abnormalities and cognitive deficits in transgenic mice overexpressing Dyrk1A (minibrain), a murine model of Down's syndrome." *Human Molecular Genetics*, 10(18), 1915-1923.
- Altshuler, D. L., Durbin, R. M., Abecasis, G. R., Bentley, D. R., Chakravarti, A., Clark, A. G., Collins, F. S., De la Vega, F. M., Donnelly, P., Egholm, M., Flicek, P., Gabriel, S. B., Gibbs, R. A., Knoppers, B. M., Lander, E. S.,

Lehrach, H., Mardis, E. R., McVean, G. A., Nickerson, D., Peltonen, L., Schafer, A. J., Sherry, S. T., Wang, J., Wilson, R. K., Deiros, D., Metzker, M., Muzny, D., Reid, J., Wheeler, D., Li, J., Jian, M., Li, G., Li, R., Liang, H., Tian, G., Wang, B., Wang, J., Wang, W., Yang, H., Zhang, X., Zheng, H., Ambrogio, L., Bloom, T., Cibulskis, K., Fennell, T. J., Jaffe, D. B., Shefler, E., Sougnez, C. L., Gormley, N., Humphray, S., Kingsbury, Z., Koko-Gonzales, P., Stone, J., McKernan, K. J., Costa, G. L., Ichikawa, J. K., Lee, C. C., Sudbrak, R., Borodina, T. A., Dahl, A., Davydov, A. N., Marquardt, P., Mertes, F., Niefeld, W., Rosenstiel, P., Schreiber, S., Soldatov, A. V., Timmermann, B., Tolzmann, M., Affourtit, J., Ashworth, D., Attiya, S., Bachorski, M., Buglione, E., Burke, A., Caprio, A., Celone, C., Clark, S., Conners, D., Desany, B., Gu, L., Guccione, L., Kao, K., Kebbel, A., Knowlton, J., Labrecque, M., McDade, L., Mealmaker, C., Minderman, M., Nawrocki, A., Niazi, F., Pareja, K., Ramenani, R., Riches, D., Song, W., Turcotte, C., Wang, S., Dooling, D., Fulton, L., Fulton, R., et al. (2010). "A map of human genome variation from population-scale sequencing." *Nature*, 467(7319), 1061-1073.

Alvarez, J. D., Chen, D. C., Storer, E., and Sehgal, A. (2003). "Non-cyclic and developmental stage-specific expression of circadian clock proteins during murine spermatogenesis." *Biology of Reproduction*, 69(1), 81-91.

Amaral, D. G., Schumann, C. M., and Nordahl, C. W. (2008). "Neuroanatomy of autism." *Trends in Neurosciences*, 31(3), 137-145.

Ambros, V., and Lee, R. C. (2004). "Identification of microRNAs and other tiny noncoding RNAs by cDNA cloning." *Methods in molecular biology (Clifton, N.J.)*, 265, 131-58.

- Amelio, A. L., Miraglia, L. J., Conkright, J. J., Mercer, B. A., Batalov, S., Cavett, V., Orth, A. P., Busby, J., Hogenesch, J. B., and Conkright, M. D. (2007). "A coactivator trap identifies NONO (p54 (nrb)) as a component of the cAMP-signaling pathway." *Proceedings of the National Academy of Sciences of the United States of America*, 104(51), 20314-20319.
- Amir, R. E., Van den Veyver, I. B., Wan, M., Tran, C. Q., Francke, U., and Zoghbi, H. Y. (1999). "Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2." *Nature Genetics*, 23(2), 185-188.
- An, J. J., Gharami, K., Liao, G. Y., Woo, N. H., Lau, A. G., Vanevski, F., Torre, E. R., Jones, K. R., Feng, Y., Lu, B., and Xu, B. (2008). "Distinct role of long 3' UTR BDNF mRNA in spine morphology and synaptic plasticity in hippocampal neurons." *Cell*, 134(1), 175-187.
- Anantharam, V., Lehrmann, E., Kanthasamy, A., Yang, Y., Banerjee, P., Becker, K. G., Freed, W. J., and Kanthasamy, A. G. (2007). "Microarray analysis of oxidative stress regulated genes in mesencephalic dopaminergic neuronal cells: Relevance to oxidation damage in Parkinson's disease." *Neurochemistry International*, 50(6), 834-847.
- Andersen, I. M., Kaczmarek, J., McGrevv, S. G., and Malow, B. A. (2008). "Melatonin for insomnia in children with autism spectrum disorders." *Journal of Child Neurology*, 23(5), 482-485.
- Anderson, G. M. (2002). "Genetics of childhood disorders: XLV. Autism, part 4: Serotonin in autism." *Journal of the American Academy of Child and Adolescent Psychiatry*, 41(12), 1513-1516.
- Anderson, G. M., Gutknecht, L., Cohen, D. J., Brailly-Tabard, S., Cohen, J. H. M., Ferrari, P., Roubertoux, P. L., and Tordjman, S. (2002). "Serotonin

transporter promoter variants in autism: functional effects and relationship to platelet hyperserotonemia." *Molecular Psychiatry*, 7(8), 831-836.

Anne, J., Ollo, R., Ephrussi, A., and Mechler, B. M. (2007). "Arginine methyltransferase Capsuleen is essential for methylation of spliceosomal Sm proteins and germ cell formation in *Drosophila*." *Development*, 134(1), 137-146.

Annunziato, A. T., and Seale, R. L. (1983). "Histone deacetylation is required for the maturation of newly replicated chromatin." *Journal of Biological Chemistry*, 258(20), 2675-2684.

Antoch, M. P., Gorbacheva, V. Y., Vykhovanets, O., Toshkov, I. A., Kondratov, R. V., Kondratova, A. A., Lee, C., and Nikitin, A. Y. (2008). "Disruption of the circadian clock due to the Clock mutation has discrete effects on aging and carcinogenesis." *Cell Cycle*, 7(9), 1197-1204.

APA. (1994). "Diagnostic and Statistical Manual of Mental Disorders - Fourth Edition (DSM-IV)." Washington D.C: American Psychiatric Association.

Aschoff, J. (1965). "Circadian rhythms in man - a self-sustained oscillator with an inherent frequency underlies human 24-hour periodicity." *Science*, 148(3676), 1427-&.

Aschoff, J., Fatransk.M, Giedke, H., Doerr, P., Stamm, D., and Wisser, H. (1971). "Human circadian rhythms in continuous darkness - entrainment by social cues." *Science*, 171(3967), 213-&.

Asher, G., Gatfield, D., Stratmann, M., Reinke, H., Dibner, C., Kreppel, F., Mostoslavsky, R., Alt, F. W., and Schibler, U. (2008). "SIRT1 regulates circadian clock gene expression through PER2 deacetylation." *Cell*, 134(2), 317-328.

- Asher, G., and Schibler, U. (2006). "A CLOCK-less clock." *Trends in Cell Biology*, 16(11), 547-549.
- Aston, C., Jiang, L. X., and Sokolov, B. P. (2004). "Microarray analysis of postmortem temporal cortex from patients with schizophrenia." *Journal of Neuroscience Research*, 77(6), 858-866.
- Auranen, M., Vanhala, R., Varilo, T., Ayers, K., Kempas, E., Ylisaukko-oja, T., Sinsheimer, J. S., Peltonen, L., and Jarvela, I. (2002). "A genomewide screen for autism-spectrum disorders: Evidence for a major susceptibility locus on chromosome 3q25-27." *American Journal of Human Genetics*, 71(4), 777-790.
- Bachevalier, J., and Loveland, K. A. (2006). "The orbitofrontal-amygdala circuit and self-regulation of social-emotional behavior in autism." *Neuroscience and Biobehavioral Reviews*, 30(1), 97-117.
- Bailey, A., and Int Mol Genet Study Autism, C. (1998). "A full genome screen for autism with evidence for linkage to a region on chromosome 7q." *Human Molecular Genetics*, 7(3), 571-578.
- Bailey, A., Lecouteur, A., Gottesman, I., Bolton, P., Simonoff, E., Yuzda, E., and Rutter, M. (1995). "Autism as a strongly genetic disorder - evidence from a British twin study." *Psychological Medicine*, 25(1), 63-77.
- Bak, M., Silaharoglu, A., Moller, M., Christensen, M., Rath, M. F., Skryabin, B., Tommerup, N., and Kauppinen, S. (2008). "MicroRNA expression in the adult mouse central nervous system." *Rna-a Publication of the Rna Society*, 14(3), 432-444.
- Balsalobre, A., Brown, S. A., Marcacci, L., Tronche, F., Kellendonk, C., Reichardt, H. M., Schutz, G., and Schibler, U. (2000). "Resetting of circadian time peripheral tissues by glucocorticoid signaling." *Science*,



289(5488), 2344-2347.

Balsalobre, A., Damiola, F., and Schibler, U. (1998). "A serum shock induces circadian gene expression in mammalian tissue culture cells." *Cell*, 93(6), 929-937.

Barbeau, E. B., Mendrek, A., and Mottron, L. (2009). "Are autistic traits autistic?" *British Journal of Psychology*, 100, 23-28.

Barchuk, A. R., Cristino, A. S., Kucharski, R., Costa, L. F., Simoes, Z. L. P., and Maleszka, R. (2007). "Molecular determinants of caste differentiation in the highly eusocial honeybee *Apis mellifera*." *Bmc Developmental Biology*, 7.

Barman, H. K., Takami, Y., Ono, T., Nishijima, H., Sanematsu, F., Shibahara, K., and Nakayama, T. (2006). "Histone acetyltransferase 1 is dispensable for replication-coupled chromatin assembly but contributes to recover DNA damages created following replication blockage in vertebrate cells." *Biochemical and Biophysical Research Communications*, 345(4), 1547-1557.

Barnard, A. R., and Nolan, P. M. (2008). "When clocks go bad: Neurobehavioural consequences of disrupted circadian timing." *Plos Genetics*, 4(5).

Barnes, J. W., Tischkau, S. A., Barnes, J. A., Mitchell, J. W., Burgoon, P. W., Hickok, J. R., and Gillette, M. U. (2003). "Requirement of mammalian *Timeless* for circadian rhythmicity." *Science*, 302(5644), 439-442.

Baron-Cohen, S. (2002). "The extreme male brain theory of autism." *Trends in Cognitive Sciences*, 6(6), 248-254.

Baron-Cohen, S., Ring, H. A., Bullmore, E. T., Wheelwright, S., Ashwin, C., and Williams, S. C. R. (2000). "The amygdala theory of autism." *Neuroscience and Biobehavioral Reviews*, 24(3), 355-364.

- Baron-Cohen, S., Leslie, A. M., and Frith, U. (1985). "Does the autistic-child have a theory of mind." *Cognition*, 21(1), 37-46.
- Bauman, M., and Kemper, T. L. (1985). "Histoanatomic observations of the brain in early infantile-autism." *Neurology*, 35(6), 866-874.
- Bauman, M. L., and Kemper, T. L. (1986). "Developmental cerebellar abnormalities - a consistent finding in early infantile-autism." *Neurology*, 36(4), 190-190.
- Bauman, M. L., and Kemper, T. L. (1988). "Limbic and cerebellar abnormalities - consistent findings in infantile-autism." *Journal of Neuropathology and Experimental Neurology*, 47(3), 369-369.
- Bauman, M. L., and Kemper, T. L. (2005). "Neuroanatomic observations of the brain in autism: a review and future directions." *International Journal of Developmental Neuroscience*, 23, 183-187.
- Beaulieu, J. M., Gainetdinov, R. R., and Caron, M. G. (2009). "Akt/GSK3 Signaling in the action of psychotropic drugs." *Annual Review of Pharmacology and Toxicology*, 49, 327-347.
- Beaver, L. M., and Giebultowicz, J. M. (2004). "Regulation of copulation duration by period and timeless in *Drosophila melanogaster*." *Current Biology*, 14(16), 1492-1497.
- Bebko, J. M., Weiss, J. A., Demark, J. L., and Gomez, P. (2006). "Discrimination of temporal synchrony in intermodal events by children with autism and children with developmental disabilities without autism." *Journal of Child Psychology and Psychiatry*, 47(1), 88-98.
- Bedford, M. T., and Richard, S. (2005). "Arginine methylation: An emerging regulator of protein function." *Molecular Cell*, 18(3), 263-272.

- Bellringer, M. E., Rahman, K., and Coleman, R. (1988). "Sodium valproate inhibits the movement of secretory vesicles in rat hepatocytes." *Biochemical Journal*, 249(2), 513-519.
- Ben Shalom, D. (2003). "Memory in autism: Review and synthesis." *Cortex*, 39(4-5), 1129-1138.
- Benayed, R., Choi, J., Matteson, P. G., Gharani, N., Kamdar, S., Brzustowicz, L. M., and Millonig, J. H. (2009). "Autism-associated haplotype affects the regulation of the homeobox gene, ENGRAILED 2." *Biological Psychiatry*, 66(10), 911-917.
- Benayed, R., Gharani, N., Rossman, I., Mancuso, V., Lazar, G., Kamdar, S., Bruse, S. E., Tischfield, S., Smith, B. J., Zimmerman, R. A., DiCiccio-Bloom, E., Brzustowicz, L. M., and Millonig, J. H. (2005). "Support for the homeobox transcription factor gene ENGRAILED 2 as an autism spectrum disorder susceptibility locus." *American Journal of Human Genetics*, 77(5), 851-868.
- Bengert, P., and Dandekar, T. (2003). "A software tool-box for analysis of regulatory RNA elements." *Nucleic Acids Research*, 31(13), 3441-3445.
- Benson, L. J., Phillips, J. A., Gu, Y., Parthun, M. R., Hoffman, C. S., and Annunziato, A. T. (2007). "Properties of the type B histone acetyltransferase Hat1 - H4 tail Interaction, site preference, and involvement in DNA repair." *Journal of Biological Chemistry*, 282(2), 836-842.
- Berezikov, E., Chung, W. J., Willis, J., Cuppen, E., and Lai, E. C. (2007). "Mammalian mirtron genes." *Molecular Cell*, 28, 328-336.
- Berezikov, E., Cuppen, E., and Plasterk, R. H. A. (2006). "Approaches to microRNA discovery." *Nature Genetics*, 38, S2-S7.

- Berger, S. L. (2007). "The complex language of chromatin regulation during transcription." *Nature*, 447(7143), 407-412.
- Berson, D. M., Dunn, F. A., and Takao, M. (2002). "Phototransduction by retinal ganglion cells that set the circadian clock." *Science*, 295(5557), 1070-1073.
- Beveridge, N. J., Gardiner, E., Carroll, A. P., Tooney, P. A., and Cairns, M. J. (2010). "Schizophrenia is associated with an increase in cortical microRNA biogenesis." *Molecular Psychiatry*, 15(12), 1176-1189.
- Bi, W. M., Yan, J., Shi, X., Yuva-Paylor, L. A., Antalffy, B. A., Goldman, A., Yoo, J. W., Noebels, J. L., Armstrong, D. L., Paylor, R., and Lupski, J. R. (2007). "Rai1 deficiency in mice causes learning impairment and motor dysfunction, whereas Rai1 heterozygous mice display minimal behavioral phenotypes." *Human Molecular Genetics*, 16(15), 1802-1813.
- Billeter, J.C., Rideout, E.J., Dornan, A.J., Goodwin S.F. (2006) "Control of male sexual behavior in *Drosophila* by the sex determination pathway." *Current Biology* 16(17), R766-76.
- Bird, A. P. (1986). "CpG-rich islands and the function of DNA methylation." *Nature*, 321(6067), 209-213.
- Bittel, D. C., and Butler, M. G. (2005). "Prader-Willi syndrome: clinical genetics, cytogenetics and molecular biology." *Expert Reviews in Molecular Medicine*, 7(14), 1-20.
- Borgs, L., Beukelaers, P., Vandenbosch, R., Nguyen, L., Moonen, G., Maquet, P., Albrecht, U., Belachew, S., and Malgrange, B. (2009). "Period 2 regulates neural stem/progenitor cell proliferation in the adult hippocampus." *Bmc Neuroscience*, 10.

- Boucher, J. (2000). "Time parsing, normal language acquisition, and language-related developmental disorders." *New Directions in Language Development and Disorders*, 13-23.
- Boucher, J. (2001). "Lost in a sea of time: Time parsing and autism." *Time and memory*. C. Hoerl and T. Cormack. Oxford, Clarendon Press.
- Boucher, J., and Lewis, V. (1989). "Memory impairments and communication in relatively able autistic-children." *Journal of Child Psychology and Psychiatry and Allied Disciplines*, 30(1), 99-122.
- Boucher, J., Pons, F., Lind, S., and Williams, D. (2007). "Temporal cognition in children with autistic spectrum disorders: Tests of diachronic thinking." *Journal of Autism and Developmental Disorders*, 37(8), 1413-1429.
- Bourgeron, T. (2007). "The possible interplay of synaptic and clock genes in autism spectrum disorders." *Cold Spring Harbor Symposia on Quantitative Biology*, 72, 645-654.
- Bourgeron, T. (2009). "A synaptic trek to autism." *Current Opinion in Neurobiology*, 19(2), 231-234.
- Bowler, D. M., Gardiner, J. M., and Grice, S. J. (2000). "Episodic memory and remembering in adults with Asperger syndrome." *Journal of Autism and Developmental Disorders*, 30(4), 295-304.
- Brock, J., Brown, C. C., Boucher, J., and Rippon, G. (2002). "The temporal binding deficit hypothesis of autism." *Development and Psychopathology*, 14(2), 209-224.
- Brown, C., Gruber, T., Boucher, J., Rippon, G., and Brock, J. (2005). "Gamma abnormalities during perception of illusory figures in autism." *Cortex*, 41(3), 364-376.

- Brown, S. A., Ripperger, J., Kadener, S., Fleury-Olela, F., Vilbois, F., Rosbash, M., and Schibler, U. (2005). "PERIOD1-associated proteins modulate the negative limb of the mammalian circadian oscillator." *Science*, 308(5722), 693-696.
- Brown, S. A., Zimbrunn, G., Fleury-Olela, F., Preitner, N., and Schibler, U. (2002). "Rhythms of mammalian body temperature can sustain peripheral circadian clocks." *Current Biology*, 12(18), 1574-1583.
- Broyd, S. J., Demanuele, C., Debener, S., Helps, S. K., James, C. J., and Sonuga-Barke, E. J. S. (2009). "Default-mode brain dysfunction in mental disorders: A systematic review." *Neuroscience and Biobehavioral Reviews*, 33(3), 279-296.
- Brune, C. W., Korvatska, E., Allen-Brady, K., Cook, E. H., Dawson, G., Devlin, B., Estes, A., Hennelly, M., Hyman, S. L., McMahon, W. M., Munson, J., Rodier, P. M., Schellenberg, G. D., Stodgell, C. J., and Coon, H. (2008). "Heterogeneous association between engrailed-2 and autism in the CPEA network." *American Journal of Medical Genetics Part B-Neuropsychiatric Genetics*, 147B(2), 187-193.
- Buckner, R. L., and Vincent, J. L. (2007). "Unrest at rest: Default activity and spontaneous network correlations." *Neuroimage*, 37, 1091-1096.
- Bukelis, I., Porter, F. D., Zimmerman, A. W., and Tierney, E. (2007). "Smith-Lemli-Opitz syndrome and autism spectrum disorder." *American Journal of Psychiatry*, 164, 1655-1661.
- Bunger, M. K., Wilsbacher, L. D., Moran, S. M., Clendenin, C., Radcliffe, L. A., Hogenesch, J. B., Simon, M. C., Takahashi, J. S., and Bradfield, C. A. (2000). "Mop3 is an essential component of the master circadian pacemaker in mammals." *Cell*, 103(7), 1009-1017.

- Bur, I. M., Cohen-Solal, A. M., Carmignac, D., Abecassis, P. Y., Chauvet, N., Martin, A. O., van der Horst, G. T. J., Robinson, I., Maurel, P., Mollard, P., and Bonnefont, X. (2009). "The circadian clock components CRY1 and CRY2 are necessary to sustain sex dimorphism in mouse liver metabolism." *Journal of Biological Chemistry*, 284(14), 9066-9073.
- Burbach, J. P. H., and van der Zwaag, B. (2009). "Contact in the genetics of autism and schizophrenia." *Trends in Neurosciences*, 32(2), 69-72.
- Busino, L., Bassermann, F., Maiolica, A., Lee, C., Nolan, P. M., Godinho, S. I. H., Draetta, G. F., and Pagano, M. (2007). "SCF(Fbx13) controls the oscillation of the circadian clock by directing the degradation of cryptochrome proteins." *Science*, 316(5826), 900-904.
- Buxbaum, J. D., Silverman, J. M., Smith, C. J., Greenberg, D. A., Kilifarski, M., Reichert, J., Cook, E. H., Fang, Y., Song, C. Y., and Vitale, R. (2002). "Association between a GABRB3 polymorphism and autism." *Molecular Psychiatry*, 7(3), 311-316.
- Buxbaum, J. D., Silverman, J. M., Smith, C. J., Kilifarski, M., Reichert, J., Hollander, E., Lawlor, B. A., Fitzgerald, M., Greenberg, D. A., and Davis, K. L. (2001). "Evidence for a susceptibility gene for autism on chromosome 2 and for genetic heterogeneity." *American Journal of Human Genetics*, 68(6), 1514-1520.
- Buzsaki, G., and Draguhn, A. (2004). "Neuronal oscillations in cortical networks." *Science*, 304(5679), 1926-1929.
- Cahill, L., and McGaugh, J. L. (1998). "Mechanisms of emotional arousal and lasting declarative memory." *Trends in Neurosciences*, 21(7), 294-299.
- Cai, S. T., Lee, C. C., and Kohwi-Shigematsu, T. (2006). "SATB1 packages densely looped, transcriptionally active chromatin for coordinated

- expression of cytokine genes." *Nature Genetics*, 38(11), 1278-1288.
- Cambray, S., Pedraza, N., Rafel, M., Gari, E., Aldea, M., and Gallego, C. (2009). "Protein Kinase KIS Localizes to RNA Granules and Enhances Local Translation." *Molecular and Cellular Biology*, 29(3), 726-735.
- Campanero, M. R., Armstrong, M. I., and Flemington, E. K. (2000). "CPG methylation as a mechanism for the regulation of E2F activity." *Proceedings of the National Academy of Sciences of the United States of America*, 97(12), 6481-6486.
- Campbell, D. B., Li, C., Sutcliffe, J. S., Persico, A. M., and Levitt, P. (2008). "Genetic evidence implicating multiple genes in the MET receptor tyrosine kinase pathway in autism spectrum disorder." *Autism Research*, 1(3), 159-168.
- Campesan, S., Dubrova, Y., Hall, J. C., and Kyriacou, C. P. (2001). "The nonA gene in *Drosophila* conveys species-specific behavioral characteristics." *Genetics*, 158(4), 1535-1543.
- Cantor, R. M., Kono, N., Duvall, J. A., Alvarez-Retuerto, A., Stone, J. L., Alarcon, M., Nelson, S. F., and Geschwind, D. H. (2005). "Replication of autism linkage: Fine-mapping peak at 17q21." *American Journal of Human Genetics*, 76(6), 1050-1056.
- Cao, Q., Gery, S., Dashti, A., Yin, D., Zhou, Y., Gu, J., and Koeffler, H. P. (2009). "A role for the clock gene *Per1* in prostate cancer." *Cancer Research*, 69(19), 7619-7625.
- Cardone, L., Hirayama, J., Giordano, F., Tarnaru, T., Palvimo, J. J., and Sassone-Corsi, P. (2005). "Circadian clock control by SUMOylation of BMAL1." *Science*, 309(5739), 1390-1394.



- Carroll, L. S., and Owen, M. J. (2009). "Genetic overlap between autism, schizophrenia and bipolar disorder." *Genome medicine*, 1(10), 102.
- Castermans, D., Volders, K., Crepel, A., Backx, L., De Vos, R., Freson, K., Meulemans, S., Vermeesch, J. R., Schrandt-Stumpel, C., De Rijk, P., Del-Favero, J., Van Geet, C., Van De Ven, W. J. M., Steyaert, J. G., Devriendt, K., and Creemers, J. W. M. (2010). "SCAMP5, NBEA and AMISYN: three candidate genes for autism involved in secretion of large dense-core vesicles." *Human Molecular Genetics*, 19(7), 1368-1378.
- Cermakian, N., Monaco, L., Pando, M. P., Dierich, A., Sassone-Corsi, P. (2001). "Altered behavioral rhythms and clock gene expression in mice with a targeted mutation in the Period1 gene." *The EMBO Journal*, (15), 3967-74.
- Chahrour, M., Jung, S. Y., Shaw, C., Zhou, X. B., Wong, S. T. C., Qin, J., and Zoghbi, H. Y. (2008). "MeCP2, a key contributor to neurological disease, activates and represses transcription." *Science*, 320(5880), 1224-1229.
- Chakrabarti, S., and Fombonne, E. (2001). "Pervasive developmental disorders in preschool children." *Jama-Journal of the American Medical Association*, 285(24), 3093-3099.
- Chen, S. T., Choo, K. B., Hou, M. F., Yeh, K. T., Kuo, S. J., and Chang, J. G. (2005). "Deregulated expression of the PER1, PER2 and PER3 genes in breast cancers." *Carcinogenesis*, 26(7), 1241-1246.
- Chen, W. G., Chang, Q., Lin, Y. X., Meissner, A., West, A. E., Griffith, E. C., Jaenisch, R., and Greenberg, M. E. (2003). "Derepression of BDNF transcription involves calcium-dependent phosphorylation of MeCP2." *Science*, 302(5646), 885-889.
- Chen, Z., and McKnight, S. L. (2007). "A conserved DNA damage response pathway responsible for coupling the cell division cycle to the circadian

- and metabolic cycles." *Cell Cycle*, 6(23), 2906-2912.
- Chen, Z., Odstrcil, E. A., Tu, B. P., and McKnight, S. L. (2007). "Restriction of DNA replication to the reductive phase of the metabolic cycle protects genome integrity." *Science*, 316(5833), 1916-1919.
- Cheng, H. Y. M., Papp, J. W., Varlamova, O., Dziema, H., Russell, B., Curfman, J. P., Nakazawa, T., Shimizu, K., Okamura, H., Impey, S., and Obrietan, K. (2007). "microRNA modulation of circadian-clock period and entrainment." *Neuron*, 54(5), 813-829.
- Cheng, Y. Z., Gvakharia, B., and Hardin, P. E. (1998). "Two alternatively spliced transcripts from the *Drosophila* period gene rescue rhythms having different molecular and behavioral characteristics." *Molecular and Cellular Biology*, 18(11), 6505-6514.
- Chilov, D., Hofer, T., Bauer, C., Wenger, R. H., and Gassmann, M. (2001). "Hypoxia affects expression of circadian genes PER1 and CLOCK in mouse brain." *Faseb Journal*, 15(14), 2613-2622.
- Chu, L. W., Zhu, Y., Yu, K., Zheng, T. Z., Chokkalingam, A. P., Stanczyk, F. Z., Gao, Y. T., and Hsing, A. W. (2008). "Correlation between circadian gene variants and serum levels of sex steroids and insulin-like growth factor-I." *Cancer Epidemiology Biomarkers & Prevention*, 17(11), 3268-3273.
- Chura, L. R., Lombardo, M. V., Ashwin, E., Auyeung, B., Chakrabarti, B., Bullmore, E. T., and Baron-Cohen, S. (2010). "Organizational effects of fetal testosterone on human corpus callosum size and asymmetry." *Psychoneuroendocrinology*, 35(1), 122-132.
- Churbanov, A., Vorechovský, I., Hicks, C. (2010). "A method of predicting changes in human gene splicing induced by genetic variants in context of cis-acting elements." *BMC Bioinformatics*, 11, 22.

- Citri, Y., Colot, H. V., Jacquier A. C., Yu, Q., Hall J. C., Baltimore D., Rosbash M. (1987). "A family of unusually spliced biologically active transcripts encoded by a *Drosophila* clock gene." *Nature*, 326(6108), 42-7.
- Colhoun, E. H., and Smith, M. V. (1960). "Neurohormonal properties of royal jelly." *Nature*, 188(4753), 854-855.
- Collins, B. H., Rosato, E., and Kyriacou, C. P. (2004). "Seasonal behavior in *Drosophila melanogaster* requires the photoreceptors, the circadian clock, and phospholipase C." *Proceedings of the National Academy of Sciences of the United States of America*, 101(7), 1945-1950.
- Comery, T. A., Harris, J. B., Willems, P. J., Oostra, B. A., Irwin, S. A., Weiler, I. J., and Greenough, W. T. (1997). "Abnormal dendritic spines in fragile X knockout mice: Maturation and pruning deficits." *Proceedings of the National Academy of Sciences of the United States of America*, 94(10), 5401-5404.
- Cook, E. H., Courchesne, R., Lord, C., Cox, N. J., Yan, S., Lincoln, A., Haas, R., Courchesne, E., and Leventhal, B. L. (1997). "Evidence of linkage between the serotonin transporter and autistic disorder." *Molecular Psychiatry*, 2(3), 247-250.
- Cook, E. H., Courchesne, R. Y., Cox, N. J., Lord, C., Gonen, D., Guter, S. J., Lincoln, A., Nix, K., Haas, R., Leventhal, B. L., and Courchesne, E. (1998). "Linkage-disequilibrium mapping of autistic disorder, with 15q11-13 markers." *American Journal of Human Genetics*, 62(5), 1077-1083.
- Cook, E. H., Kim, S., Cox, N., Courchesne, R., Lord, C., Corsello, C., Akshoomoff, N., Guter, S., Leventhal, B. L., and Courchesne, E. (2001). "Transmission disequilibrium mapping in the serotonin transporter gene (SLC6A4) region in autistic disorder." *American Journal of Human*

*Genetics*, 69(4), 572-572.

Cook, E. H., Jr., and Leventhal, B. L. (1996). "The serotonin system in autism."

*Current Opinion in Pediatrics*, 8(4), 348-354.

Corbett, B. A., Mendoza, S., Abdullah, M., Wegelin, J. A., and Levine, S. (2006).

"Cortisol circadian rhythms and response to stress in children with autism."

*Psychoneuroendocrinology*, 31(1), 59-68.

Cordeiro, M. L., Gundersen, C. B., and Umbach, J. A. (2004). "Convergent

effects of lithium and valproate on the expression of proteins associated with large dense core vesicles in NGF-differentiated PC12 cells."

*Neuropsychopharmacology*, 29(1), 39-44.

Correia, C., Coutinho, A. M., Almeida, J., Lontro, R., Lobo, C., Miguel, T. S.,

Martins, M., Gallagher, L., Conroy, J., Gill, M., Oliveira, G., and Vicente, A.

M. (2009). "Association of the alpha 4 integrin subunit gene (ITGA4) with autism." *American Journal of Medical Genetics Part B-Neuropsychiatric*

*Genetics*, 150B(8), 1147-1151.

Courchesne, E., Carper, R., and Akshoomoff, N. (2003). "Evidence of brain

overgrowth in the first year of life in autism." *Jama-Journal of the American Medical Association*, 290(3), 337-344.

Courchesne, E., Townsend, J., Akshoomoff, N. A., Saitoh, O., Yeungcourchesne,

R., Lincoln, A. J., James, H. E., Haas, R. H., Schreibman, L., and Lau, L.

(1994). "Impairment in shifting attention in autistic and cerebellar patients."

*Behavioral Neuroscience*, 108(5), 848-865.

Courchesne, E., Townsend, J., and Saitoh, O. (1994). "The brain in infantile-

autism - posterior-fossa structures are abnormal." *Neurology*, 44(2), 214-223.

- Crespi, B. (2008). "Genomic imprinting in the development and evolution of psychotic spectrum conditions." *Biological Reviews*, 83(4), 441-493.
- Crosio, C., Heitz, E., Allis, C. D., Borrelli, E., and Sassone-Corsi, P. (2003). "Chromatin remodeling and neuronal response: multiple signaling pathways induce specific histone H3 modifications and early gene expression in hippocampal neurons." *Journal of Cell Science*, 116(24), 4905-4914.
- Crown, C. L., Feldstein, S., Jasnow, M. D., Beebe, B., and Jaffe, J. (2002). "The cross-modal coordination of interpersonal timing: Six-week-olds infants' gaze with adults' vocal behavior." *Journal of Psycholinguistic Research*, 31(1), 1-23.
- Crumbley, C., and Burris, T. P. (2011). "Direct regulation of CLOCK expression by REV-ERB." *Plos One*, 6(3).
- Crumbley, C., Wang, Y., Kojetin, D. J., and Burris, T. P. (2010). "Characterization of the core mammalian clock component, NPAS2, as a REV-ERB alpha/ROR alpha Target Gene." *Journal of Biological Chemistry*, 285(46), 35386-35392.
- Curtis, A. M., Seo, S. B., Westgate, E. J., Rudic, R. D., Smyth, E. M., Chakravarti, D., FitzGerald, G. A., and McNamara, P. (2004). "Histone acetyltransferase-dependent chromatin remodeling and the vascular clock." *Journal of Biological Chemistry*, 279(8), 7091-7097.
- Danilova, N., Krupnik, V. E., Sugden, D., and Zhdanova, I. V. (2004). "Melatonin stimulates cell proliferation in zebrafish embryo and accelerates its development." *Faseb Journal*, 18(2), 751-+.
- Dardente, H., and Cermakian, N. (2007). "Molecular circadian rhythms in central and peripheral clocks in mammals." *Chronobiology International*, 24(2),

195-213.

- Darlington, T. K., Wager-Smith, K., Ceriani, M. F., Staknis, D., Gekakis, N., Steeves, T. D. L., Weitz, C. J., Takahashi, J. S., and Kay, S. A. (1998). "Closing the circadian loop: CLOCK-induced transcription of its own inhibitors *per* and *tim*." *Science*, 280(5369), 1599-1603.
- Das, C., Lucia, M. S., Hansen, K. C., and Tyler, J. K. (2009). "CBP/p300-mediated acetylation of histone H3 on lysine 56." *Nature*, 459(7243), 113-U123.
- Dauwalder, B., Tsujimoto, S., Moss, J., and Mattox, W. (2002). "The *Drosophila* takeout gene is regulated by the somatic sex-determination pathway and affects male courtship behavior." *Genes & Development*, 16(22), 2879-2892.
- Davies, W., Isles, A. R., and Wilkinson, L. S. (2005). "Imprinted gene expression in the brain." *Neuroscience and Biobehavioral Reviews*, 29(3), 421-430.
- Davis, S., Mirick, D. K., and Stevens, R. G. (2001). "Night shift work, light at night, and risk of breast cancer." *Journal of the National Cancer Institute*, 93(20), 1557-1562.
- DeBaun, M. R., Niemitz, E. L., McNeil, D. E., Brandenburg, S. A., Lee, M. P., and Feinberg, A. P. (2002). "Epigenetic alterations of H19 and LIT1 distinguish patients with Beckwith-Wiedemann syndrome with cancer and birth defects." *American Journal of Human Genetics*, 70(3), 604-611.
- DeBruyne, J. P., Weaver, D. R., and Reppert, S. M. (2007). "CLOCK and NPAS2 have overlapping roles in the suprachiasmatic circadian clock." *Nature Neuroscience*, 10(5), 543-545.
- Deery, M. J., Maywood, E. S., Chesham, J. E., Sladek, M., Karp, N. A., Green, E. W., Charles, P. D., Reddy, A. B., Kyriacou, C. P., Lilley, K. S., and

Hastings, M. H. (2009). "Proteomic Analysis Reveals the Role of Synaptic Vesicle Cycling in Sustaining the Suprachiasmatic Circadian Clock."

*Current Biology*, 19(23), 2031-2036.

del Gaudio, D., Fang, P., Scaglia, F., Ward, P. A., Craigen, W. J., Glaze, D. G., Neul, J. L., Patel, A., Lee, J. A., Irons, M., Berry, S. A., Pursley, A. A., Grebe, T. A., Freedenberg, D., Martin, R. A., Hsich, G. E., Khera, J. R., Friedman, N. R., Zoghbi, H. Y., Eng, C. M., Lupski, J. R., Beaudet, A. L., Cheung, S. W., and Roa, B. B. (2006). "Increased MECP2 gene copy number as the result of genomic duplication in neurodevelopmentally delayed males." *Genetics in Medicine*, 8(12), 784-792.

Demir, E., and Dickson, B. J. (2005). "fruitless splicing specifies male courtship behavior in *Drosophila*." *Cell*, 121(5), 785-794.

Diernfellner, A. C. R., Schafmeier, T., Mellow, M. W., and Brunner, M. (2005). "Molecular mechanism of temperature sensing by the circadian clock of *Neurospora crassa*." *Genes & Development*, 19(17), 1968-1973.

Dilella, A. G., Marvit, J., Brayton, K., and Woo, S. L. C. (1987). "An amino-acid substitution involved in phenylketonuria is in linkage disequilibrium with DNA haplotype-2." *Nature*, 327(6120), 333-336.

Dioum, E. M., Rutter, J., Tuckerman, J. R., Gonzalez, G., Gilles-Gonzalez, M. A., and McKnight, S. L. (2002). "NPAS2: A gas-responsive transcription factor." *Science*, 298(5602), 2385-2387.

Doi, M., Hirayama, J., and Sassone-Corsi, P. (2006). "Circadian regulator CLOCK is a histone acetyltransferase." *Cell*, 125(3), 497-508.

Dokucu, M. E., Yu, L. P., and Taghert, P. H. (2005). "Lithium- and valproate-induced alterations in circadian locomotor behavior in *Drosophila*." *Neuropsychopharmacology*, 30(12), 2216-2224.

- Dowjat, W. K., Adayev, T., Kuchna, I., Nowicki, K., Palminiello, S., Hwang, Y. W., and Wegiel, J. (2007). "Trisomy-driven overexpression of DYRK1A kinase in the brain of subjects with Down syndrome." *Neuroscience Letters*, 413(1), 77-81.
- Duan, R. H., Pak, C. H., and Jin, P. (2007). "Single nucleotide polymorphism associated with mature miR-125a alters the processing of pri-miRNA." *Human Molecular Genetics*, 16(9), 1124-1131.
- Dudbridge, F. (2003). "Pedigree disequilibrium tests for multilocus haplotypes." *Genetic Epidemiology*, 25(2), 115-121.
- Duffield, G. E., Watson, N. P., Mantani, A., Peirson, S. N., Robles-Murguía M., Loros J. J., Israel, M. A., Dunlap, J. C. (2009) "A role for *Id2* in regulating photic entrainment of the mammalian circadian system." *Current Biology*. 19(4), 297-304.
- Dunn, L. M and Dunn, L. L. (1997). "Peabody Picture and Vocabulary Test –Third Edition". Circle Pines, Minnesota: American Guidance Service.
- Duong, H. A., Robles, M. S., Knutti, D., and Weitz, C. J. (2011). "A molecular mechanism for circadian clock negative feedback." *Science*, 332(6036), 1436-1439.
- Durand, C. M., Betancur, C., Boeckers, T. M., Bockmann, J., Chaste, P., Fauchereau, F., Nygren, G., Rastam, M., Gillberg, I. C., Anckarsater, H., Sponheim, E., Goubran-Botros, H., Delorme, R., Chabane, N., Mouren-Simeoni, M. C., de Mas, P., Bieth, E., Roge, B., Heron, D., Burglen, L., Gillberg, C., Leboyer, M., and Bourgeron, T. (2007). "Mutations in the gene encoding the synaptic scaffolding protein SHANK3 are associated with autism spectrum disorders." *Nature Genetics*, 39(1), 25-27.



- Dworniczak, B., Aulehlscholz, C., Kalaydjieva, L., Bartholome, K., Grudda, K., and Horst, J. (1991). "Aberrant splicing of phenylalanine-hydroxylase messenger-RNA - the major cause for phenylketonuria in parts of southern Europe." *Genomics*, 11(2), 242-246.
- Eberharter, A., Lechner, T., GoralikSchramel, M., and Loidl, P. (1996). "Purification and characterization of the cytoplasmic histone acetyltransferase B of maize embryos." *Febs Letters*, 386(1), 75-81.
- Ebisawa, T., Uchiyama, M., Kajimura, N., Mishima, K., Kamei, Y., Katoh, M., Watanabe, T., Sekimoto, M., Shibui, K., Kim, K., Kudo, Y., Ozeki, Y., Sugishita, M., Toyoshima, R., Inoue, Y., Yamada, N., Nagase, T., Ozaki, N., Ohara, O., Ishida, N., Okawa, M., Takahashi, K., and Yamauchi, T. (2001). "Association of structural polymorphisms in the human period3 gene with delayed sleep phase syndrome." *Embo Reports*, 2(4), 342-346.
- Ecker, C., Marquand, A., Mourao-Miranda, J., Johnston, P., Daly, E. M., Brammer, M. J., Maltezos, S., Murphy, C. M., Robertson, D., Williams, S. C., and Murphy, D. G. M. (2010). "Describing the brain in autism in five dimensions-magnetic resonance imaging-assisted diagnosis of autism spectrum disorder using a multiparameter classification approach." *Journal of Neuroscience*, 30(32), 10612-10623.
- El-Husseini, A. E. D., Guthrie, H., Snutch, T. P., and Vincent, S. R. (1997). "Molecular cloning of a mammalian homologue of the yeast vesicular transport protein vps45." *Biochimica Et Biophysica Acta-Biomembranes*, 1325(1), 8-12.
- Elia, M., Ferri, R., Musumeci, S. A., Del Gracco, S., Bottitta, M., Scuderi, C., Miano, G., Panerai, S., Bertrand, T., and Grubar, J. C. (2000). "Sleep in subjects with autistic disorder: a neurophysiological and psychological

study." *Brain & Development*, 22(2), 88-92.

Etchegaray, J. P., Lee, C., Wade, P. A., and Reppert, S. M. (2003). "Rhythmic histone acetylation underlies transcription in the mammalian circadian clock." *Nature*, 421(6919), 177-182.

Etchegaray, J. P., Yang, X. M., DeBruyne, J. P., Peters, A., Weaver, D. R., Jenuwein, T., and Reppert, S. M. (2006). "The polycomb group protein EZH2 is required for mammalian circadian clock function." *Journal of Biological Chemistry*, 281(30), 21209-21215.

Faghihi, M. A., and Wahlestedt, C. (2009). "Regulatory roles of natural antisense transcripts." *Nature Reviews Molecular Cell Biology*, 10(9), 637-643.

Feillet, C. A., Mendoza, J., Albrecht, U., Pevet, P., and Challet, E. (2008). "Forebrain oscillators ticking with different clock hands." *Molecular and Cellular Neuroscience*, 37(2), 209-221.

Feldman, R. (2006). "From biological rhythms to social rhythms: Physiological precursors of mother-infant synchrony." *Developmental Psychology*, 42(1), 175-188.

Feldman, R., Greenbaum, C. W., and Yirmiya, N. (1999). "Mother-infant affect synchrony as an antecedent of the emergence of self-control." *Developmental Psychology*, 35(1), 223-231.

Feldman, R., Greenbaum, C. W., Yirmiya, N., and Mayes, L. C. (1996). "Relations between cyclicity and regulation in mother-infant interaction at 3 and 9 months and cognition at 2 years." *Journal of Applied Developmental Psychology*, 17(3), 347-365.

Feldstein, S., Konstantareas, M., Oxman, J., and Webster, C. D. (1982). "The chronography of interactions with autistic speakers - an initial report." *Journal of Communication Disorders*, 15(6), 451-460.

- Felsenfeld, G., and Groudine, M. (2003). "Controlling the double helix." *Nature*, 421(6921), 448-453.
- Fernandez, A., Llacuna, L., Fernandez-Checa, J. C., and Colell, A. (2009). "Mitochondrial cholesterol loading exacerbates amyloid beta peptide-induced inflammation and neurotoxicity." *Journal of Neuroscience*, 29(20), 6394-6405.
- Fernandez, M. P., Berni, J., and Ceriani, M. F. (2008). "Circadian remodeling of neuronal circuits involved in rhythmic behavior." *Plos Biology*, 6(3), 518-524.
- Ferreira, M. A. R., O'Donovan, M. C., Meng, Y. A., Jones, I. R., Ruderfer, D. M., Jones, L., Fan, J., Kirov, G., Perlis, R. H., Green, E. K., Smoller, J. W., Grozeva, D., Stone, J., Nikolov, I., Chambert, K., Hamshere, M. L., Nimgaonkar, V. L., Moskvina, V., Thase, M. E., Caesar, S., Sachs, G. S., Franklin, J., Gordon-Smith, K., Ardlie, K. G., Gabriel, S. B., Fraser, C., Blumenstiel, B., Defelice, M., Breen, G., Gill, M., Morris, D. W., Elkin, A., Muir, W. J., McGhee, K. A., Williamson, R., MacIntyre, D. J., MacLean, A. W., Clair, D. S., Robinson, M., Van Beck, M., Pereira, A. C. P., Kandaswamy, R., McQuillin, A., Collier, D. A., Bass, N. J., Young, A. H., Lawrence, J., Ferrier, I. N., Anjorin, A., Farmer, A., Curtis, D., Scolnick, E. M., McGuffin, P., Daly, M. J., Corvin, A. P., Holmans, P. A., Blackwood, D. H., Gurling, H. M., Owen, M. J., Purcell, S. M., Sklar, P., Craddock, N., and Wellcome Trust Case Control, C. (2008). "Collaborative genome-wide association analysis supports a role for ANK3 and CACNA1C in bipolar disorder." *Nature Genetics*, 40(9), 1056-1058.
- Fishel, M. L., Vasko, M. R., and Kelley, M. R. (2007). "DNA repair in neurons: So if they don't divide what's to repair?" *Mutation Research-Fundamental and*

*Molecular Mechanisms of Mutagenesis*, 614(1-2), 24-36.

Fombonne, E. (1999). "The epidemiology of autism: a review." *Psychological Medicine*, 29(4), 769-786.

Fombonne, E. (2005). "Epidemiology of autistic disorder and other pervasive developmental disorders." *Journal of Clinical Psychiatry*, 66, 3-8.

Fotaki, V., Dierssen, M., Alcantara, S., Martinez, S., Marti, E., Casas, C., Visa, J., Soriano, E., Estivill, X., and Arbones, M. L. (2002). "Dyrk1A haploinsufficiency affects viability and causes developmental delay and abnormal brain morphology in mice." *Molecular and Cellular Biology*, 22(18), 6636-6647.

Fradin, D., Cheslack-Postava, K., Ladd-Acosta, C., Newschaffer, C., Chakravarti, A., Arking, D. E., Feinberg, A., and Fallin, M. D. (2010). "Parent-Of-Origin effects in autism identified through genome-wide linkage analysis of 16,000 SNPs." *Plos One*, 5(9).

Franken, P., Dudley, C. A., Estill, S. J., Barakat, M., Thomason, R., O'Haran, B. F., and McKnight, S. L. (2006). "NPAS2 as a transcriptional regulator of non-rapid eye movement sleep: Genotype and sex interactions." *Proceedings of the National Academy of Sciences of the United States of America*, 103(18), 7118-7123.

Franken, P., Lopez-Molina, L., Marcacci, L., Schibler, U., and Tafti, M. (2000). "The transcription factor DBP affects circadian sleep consolidation and rhythmic EEG activity." *Journal of Neuroscience*, 20(2), 617-625.

Frazer, K. A., Ballinger, D. G., Cox, D. R., Hinds, D. A., Stuve, L. L., Gibbs, R. A., Belmont, J. W., Boudreau, A., Hardenbol, P., Leal, S. M., Pasternak, S., Wheeler, D. A., Willis, T. D., Yu, F., Yang, H., Zeng, C., Gao, Y., Hu, H., Hu, W., Li, C., Lin, W., Liu, S., Pan, H., Tang, X., Wang, J., Wang, W., Yu,

J., Zhang, B., Zhang, Q., Zhao, H., Zhao, H., Zhou, J., Gabriel, S. B., Barry, R., Blumenstiel, B., Camargo, A., Defelice, M., Faggart, M., Goyette, M., Gupta, S., Moore, J., Nguyen, H., Onofrio, R. C., Parkin, M., Roy, J., Stahl, E., Winchester, E., Ziaugra, L., Altshuler, D., Shen, Y., Yao, Z., Huang, W., Chu, X., He, Y., Jin, L., Liu, Y., Shen, Y., Sun, W., Wang, H., Wang, Y., Wang, Y., Xiong, X., Xu, L., Waye, M. M. Y., Tsui, S. K. W., Wong, J. T.-F., Galver, L. M., Fan, J.-B., Gunderson, K., Murray, S. S., Oliphant, A. R., Chee, M. S., Montpetit, A., Chagnon, F., Ferretti, V., Leboeuf, M., Olivier, J.-F., Phillips, M. S., Roumy, S., Sallee, C., Verner, A., Hudson, T. J., Kwok, P.-Y., Cai, D., Koboldt, D. C., Miller, R. D., Pawlikowska, L., Taillon-Miller, P., Xiao, M., Tsui, L.-C., Mak, W., Song, Y. Q., Tam, P. K. H., Nakamura, Y., Kawaguchi, T., Kitamoto, T., Morizono, T., Nagashima, A., Ohnishi, Y., Sekine, A., et al. (2007). "A second generation human haplotype map of over 3.1 million SNPs." *Nature*, 449(7164), 851-U3.

Freitag, C. M. (2007). "The genetics of autistic disorders and its clinical relevance: a review of the literature." *Molecular Psychiatry*, 12(1), 2-22.

Friedman, J. I., Vrijenhoek, T., Markx, S., Janssen, I. M., Van der Vliet, W. A., Faas, B. H. W., Knoers, N. V., Cahn, W., Kahn, R. S., Edelman, L., Davis, K. L., Silverman, J. M., Brunner, H. G., Van Kessel, A. G., Wijmenga, C., Ophoff, R. A., and Veltman, J. A. (2008). "CNTNAP2 gene dosage variation is associated with schizophrenia and epilepsy." *Molecular Psychiatry*, 13(3), 261-266.

Friez, M. J., Jones, J. R., Clarkson, K., Lubs, H., Abuelo, D., Bier, J. A. B., Pai, S., Simensen, R., Williams, C., Giampietro, P. F., Schwartz, C. E., and Stevenson, R. E. (2006). "Recurrent infections, hypotonia, and mental

retardation caused by duplication of MECP2 and adjacent region in Xq28." *Pediatrics*, 118(6), E1687-E1695.

- Fu, L. N., Pelicano, H., Liu, J. S., Huang, P., and Lee, C. C. (2002). "The circadian gene *Period2* plays an important role in tumor suppression and DNA damage response in vivo." *Cell*, 111(1), 41-50.
- Fuks, F., Burgers, W. A., Brehm, A., Hughes-Davies, L., and Kouzarides, T. (2000). "DNA methyltransferase Dnmt1 associates with histone deacetylase activity." *Nature Genetics*, 24(1), 88-91.
- Garbarino-Pico, E., and Green, C. B. (2007). "Posttranscriptional regulation of mammalian circadian clock output." *Cold Spring Harbor Symposia on Quantitative Biology*, 72, 145-156.
- Garcia, J. A., Zhang, D., Estill, S. J., Michnoff, C., Rutter, J., Reick, M., Scott, K., Diaz-Arrastia, R., and McKnight, S. L. (2000). "Impaired cued and contextual memory in NPAS2-deficient mice." *Science*, 288(5474), 2226-2230.
- Gauger, N. A., and Sancar, A. (2005). "Cryptochrome, circadian cycle, cell cycle checkpoints, and cancer." *Cancer Research*, 65(15), 6828-6834.
- Gekakis, N., Staknis, D., Nguyen, H. B., Davis, F. C., Wilsbacher, L. D., King, D. P., Takahashi, J. S., and Weitz, C. J. (1998). "Role of the CLOCK protein in the mammalian circadian mechanism." *Science*, 280(5369), 1564-1569.
- Gentile, A., Trusolino, L., and Comoglio, P. M. (2008). "The Met tyrosine kinase receptor in development and cancer." *Cancer and Metastasis Reviews*, 27(1), 85-94.
- Gery, S., Komatsu, N., Baldjyan, L., Yu, A., Koo, D., and Koeffler, H. P. (2006). "The circadian gene *Per1* plays an important role in cell growth and DNA damage control in human cancer cells." *Molecular Cell*, 22(3), 375-382.

- Geschwind, D. H., Sowinski, J., Lord, C., Iversen, P., Shestack, J., Jones, P., Ducat, L., Spence, S. J., and Comm, A. S. (2001). "The Autism Genetic Resource Exchange: A resource for the study of autism and related neuropsychiatric conditions." *American Journal of Human Genetics*, 69(2), 463-466.
- Gharani, N., Benayed, R., Mancuso, V., Brzustowicz, L. M., and Millonig, J. H. (2004). "Association of the homeobox transcription factor, ENGRAILED 2, 3, with autism spectrum disorder." *Molecular Psychiatry*, 9(5), 474-484.
- Gillberg, C. (1991). "The treatment of epilepsy in autism." *Journal of Autism and Developmental Disorders*, 21(1), 61-77.
- Giraldez, A. J., Mishima, Y., Rihel, J., Grocock, R. J., Van Dongen, S., Inoue, K., Enright, A. J., and Schier, A. F. (2006). "Zebrafish MiR-430 promotes deadenylation and clearance of maternal mRNAs." *Science*, 312(5770), 75-79.
- Glickman, G. (2010). "Circadian rhythms and sleep in children with autism." *Neuroscience and Biobehavioral Reviews*. 34(5), 755-68.
- Gold, D. A., Gent, P. M., and Hamilton, B. A. (2007). "ROR alpha in genetic control of cerebellum development: 50 staggering years." *Brain Research*, 1140, 19-25.
- Gonsalvez, G. B., Rajendra, T. K., Tian, L. P., and Matera, A. G. (2006). "The Sm-protein methyltransferase, *dart5*, is essential for germ-cell specification and maintenance." *Current Biology*, 16(11), 1077-1089.
- Gorbacheva, V. Y., Kondratov, R. V., Zhang, R. L., Cherukuri, S., Gudkov, A. V., Takahashi, J. S., and Antoch, M. P. (2005). "Circadian sensitivity to the chemotherapeutic agent cyclophosphamide depends on the functional status of the CLOCK/BMAL1 transactivation complex." *Proceedings of the*

*National Academy of Sciences of the United States of America*, 102(9), 3407-3412.

- Gori, F., Divieti, P., and Demay, M. B. (2001). "Cloning and characterization of a novel WD-40 repeat protein that dramatically accelerates osteoblastic differentiation." *Journal of Biological Chemistry*, 276(49), 46515-46522.
- Gowen, E., and Miall, R. C. (2005). "Behavioural aspects of cerebellar function in adults with Asperger syndrome." *Cerebellum*, 4(4), 279-289.
- Grechez-Cassiau, A., Rayet, B., Guillaumond, F., Teboul, M., and Delaunay, F. (2008). "The circadian clock component BMAL1 is a critical regulator of p21(WAF1/CIP1) expression and hepatocyte proliferation." *Journal of Biological Chemistry*, 283(8), 4535-4542.
- Greenacre, M. L., Ritchie, M. G., Byrne, B. C., and Kyriacou, C. P. (1993). "Female song preference and the period gene in *Drosophila*." *Behavior Genetics*, 23(1), 85-90.
- Grice, S. J., Spratling, M. W., Karmiloff-Smith, A., Halit, H., Csibra, G., de Haan, M., and Johnson, M. H. (2001). "Disordered visual processing and oscillatory brain activity in autism and Williams Syndrome." *Neuroreport*, 12(12), 2697-2700.
- Griffiths-Jones, S., Grocock, R. J., van Dongen, S., Bateman, A., and Enright, A. J. (2006). "miRBase: microRNA sequences, targets and gene nomenclature." *Nucleic Acids Research*, 34, D140-D144.
- Gronostajski, R. M. (2000). "Roles of the NFI/CTF gene family in transcription and development." *Gene*, 249(1-2), 31-45.
- Grossberg, S., and Seidman, D. (2006). "Neural dynamics of autistic behaviors: Cognitive, emotional, and timing substrates." *Psychological Review*, 113(3), 483-525.



- Guantieri, V., Pepe, A., Zordan, M., Kyriacou, C. P., Costa, R., and Tamburro, A. M. (1999). "Different period gene repeats take 'turns' at fine-tuning the circadian clock." *Proceedings of the Royal Society of London Series B-Biological Sciences*, 266(1435), 2283-2288.
- Guo, X., Williams, J. G., Schug, T. T., and Li, X. (2010). "DYRK1A and DYRK3 promote cell survival through phosphorylation and activation of SIRT1." *Journal of Biological Chemistry*, 285(17), 13223-13232.
- Hacker, C., Kirsch, R. D., Ju, X. S., Hieronymus, T., Gust, T. C., Kuhl, C., Jorgas, T., Kurz, S. M., Rose-John, S., Yokota, Y., and Zenke, M. (2003). "Transcriptional profiling identifies Id2 function in dendritic cell development." *Nature Immunology*, 4(4), 380-386.
- Haig, D., and Westoby, M. (1989). "parent-specific gene-expression and the triploid endosperm." *American Naturalist*, 134(1), 147-155.
- Haist, F., Adamo, M., Westerfield, M., Courchesne, E., and Townsend, J. (2005). "The functional neuroanatomy of spatial attention in autism spectrum disorder." *Developmental Neuropsychology*, 27(3), 425-458.
- Hamann, S. (2001). "Cognitive and neural mechanisms of emotional memory." *Trends in Cognitive Sciences*, 5(9), 394-400.
- Hampp, G., Ripperger, J. A., Houben, T., Schmutz, I., Blex, C., Perreau-Lenz, S., Brunk, I., Spanagel, R., Ahnert-Hilger, G., Meijer, J. H., and Albrecht, U. (2008). "Regulation of monoamine oxidase a by circadian-clock components implies clock influence on mood." *Current Biology*, 18(9), 678-683.
- Hansen, T., Olsen, L., Lindow, M., Jakobsen, K. D., Ullum, H., Jonsson, E., Andreassen, O. A., Djurovic, S., Melle, I., Agartz, I., Hall, H., Timm, S., Wang, A. G., and Werge, T. (2007). "Brain expressed microRNAs

- implicated in schizophrenia etiology." *Plos One*, 2(9).
- Happé, F., and Frith, U. (2006). "The weak coherence account: Detail-focused cognitive style in autism spectrum disorders." *Journal of Autism and Developmental Disorders*, 36(1), 5-25.
- Harding, H. P., and Lazar, M. A. (1993). "The orphan receptor REV-ERBA-alpha activates transcription via a novel response element." *Molecular and Cellular Biology*, 13(5), 3113-3121.
- Hastings, M. H., Reddy, A. B., and Maywood, E. S. (2003). "A clockwork web: Circadian timing in brain and periphery, in health and disease." *Nature Reviews Neuroscience*, 4(8), 649-661.
- Hattar, S., Liao, H. W., Takao, M., Berson, D. M., and Yau, K. W. (2002). "Melanopsin-containing retinal ganglion cells: Architecture, projections, and intrinsic photosensitivity." *Science*, 295(5557), 1065-1070.
- Hayashizaki, Y., and Carninci, P. (2006). "Genome network and FANTOM3: Assessing the complexity of the transcriptome." *Plos Genetics*, 2(4), 492-497.
- Hayes, K. R., Baggs, J. E., and Hogenesch, J. B. (2005). "Circadian clocks are seeing the systems biology light." *Genome Biology*, 6(5).
- Helfrich-Forster, C. (2000). "Differential control of morning and evening components in the activity rhythm of *Drosophila melanogaster* - Sex-specific differences suggest a different quality of activity." *Journal of Biological Rhythms*, 15(2), 135-154.
- Helvik, S. A., Snove, O., and Saetrom, P. (2007). "Reliable prediction of Drosha processing sites improves microRNA gene prediction." *Bioinformatics*, 23(2), 142-149.

- Henningsson, S., Jonsson, L., Ljunggren, E., Westberg, L., Gillberg, C., Rastam, M., Anckarsater, H., Nygren, G., Landen, M., Thuresson, K., Betancur, C., Leboyer, M., Eriksson, E., and Melke, J. (2009). "Possible association between the androgen receptor gene and autism spectrum disorder." *Psychoneuroendocrinology*, 34(5), 752-761.
- Herman, J. G., Graff, J. R., Myohanen, S., Nelkin, B. D., and Baylin, S. B. (1996). "Methylation-specific PCR: A novel PCR assay for methylation status of CpG islands." *Proceedings of the National Academy of Sciences of the United States of America*, 93(18), 9821-9826.
- Hesling, I., Dilharreguy, B., Peppe, S., Amirault, M., Bouvard, M., and Allard, M. (2010). "The integration of prosodic speech in high functioning autism: A preliminary fMRI study." *Plos One*, 5(7).
- Hirayama, J., Cardone, L., Doi, M., and Sassone-Corsi, P. (2005). "Common pathways in circadian and cell cycle clocks: Light-dependent activation of Fos/AP-1 in zebrafish controls CRY-1a and WEE-1." *Proceedings of the National Academy of Sciences of the United States of America*, 102(29), 10194-10199.
- Hirayama, J., Cho, S., and Sassone-Corsi, P. (2007a). "Circadian control by the reduction/oxidation pathway: Catalase represses light-dependent clock gene expression in the zebrafish." *Proceedings of the National Academy of Sciences of the United States of America*, 104, 15747-15752.
- Hirayama, J., Miyamura, N., Uchida, Y., Asaoka, Y., Honda, R., Sawanobori, K., Todo, T., Yamamoto, T., Sassone-Corsi, P., and Nishina, H. (2009). "Common light signaling pathways controlling DNA repair and circadian clock entrainment in zebrafish." *Cell Cycle*, 8(17), 2794-2801.

- Hirayama, J., Sahar, S., Grimaldi, B., Tamaru, T., Takamatsu, K., Nakahata, Y., and Sassone-Corsi, P. (2007b). "CLOCK-mediated acetylation of BMAL1 controls circadian function." *Nature*, 450, 1086-U13.
- Hirsch, M. R., Gaugler, L., Deagostinibazin, H., Ballycuif, L., and Goriadis, C. (1990). "Identification of positive and negative regulatory elements governing cell-type-specific expression of the neural cell-adhesion molecule gene." *Molecular and Cellular Biology*, 10(5), 1959-1968.
- Hoffman, A. E., Zheng, T. Z., Ba, Y., and Zhu, Y. (2008). "The circadian gene NPAS2, a putative tumor suppressor, is involved in DNA damage response." *Molecular Cancer Research*, 6(9), 1461-1468.
- Hogenesch, J. B., Gu, Y. Z., Jain, S. J., and Bradfield, C. A. (1998). "The basic-helix-loop-helix-PAS orphan MOP3 forms transcriptionally active complexes with circadian and hypoxia factors." *Proceedings of the National Academy of Sciences of the United States of America*, 95(10), 5474-5479.
- Hogenesch, J. B., Gu, Y. Z., Moran, S. M., Shimomura, K., Radcliffe, L. A., Takahashi, J. S., and Bradfield, C. A. (2000). "The basic helix-loop-helix-PAS protein MOP9 is a brain-specific heterodimeric partner of circadian and hypoxia factors." *Journal of Neuroscience*, 20(13), art. no.-RC83.
- Holliday, R., and Pugh, J. E. (1975). "DNA modification mechanisms and gene activity during development." *Science*, 187(4173), 226-232.
- Honma, S., Kawamoto, T., Takagi, Y., Fujimoto, K., Sato, F., Noshiro, M., Kato, Y., and Honma, K. (2002). "Dec1 and Dec2 are regulators of the mammalian molecular clock." *Nature*, 419(6909), 841-844.
- Horsthemke, B., and Wagstaff, J. (2008). "Mechanisms of imprinting of the Prader-Willi/Angelman region." *American Journal of Medical Genetics Part*

A, 146A(16), 2041-2052.

- Hosoda, H., Kato, K., Asano, H., Ito, M., Kato, H., Iwamoto, T., Suzuki, A., Masushige, S., and Kida, S. (2009). "CBP/p300 is a cell type-specific modulator of CLOCK/BMAL1-mediated transcription." *Mol Brain*, 2(1), 34.
- Hotchkiss, R. S., Strasser, A., McDunn, J. E., and Swanson, P. E. (2009). "Mechanisms of disease: Cell death." *New England Journal of Medicine*, 361(16), 1570-1583.
- Howard, M. A., Cowell, P. E., Boucher, J., Broks, P., Mayes, A., Farrant, A., and Roberts, N. (2000). "Convergent neuroanatomical and behavioural evidence of an amygdala hypothesis of autism." *Neuroreport*, 11(13), 2931-2935.
- Hu, L. J., Lu, F., Wang, Y. H., Liu, Y. Y., Liu, D. S., Jiang, Z., Wan, C. M., Zhu, B., Gan, L., Wang, Y. Q., and Wang, Z. R. (2006). "RACK1, a novel hPER1-interacting protein." *Journal of Molecular Neuroscience*, 29(1), 55-63.
- Hu, V. W., Frank, B. C., Heine, S., Lee, N. H., and Quackenbush, J. (2006). "Gene expression profiling of lymphoblastoid cell lines from monozygotic twins discordant in severity of autism reveals differential regulation of neurologically relevant genes." *Bmc Genomics*, 7.
- Hu, V. W., Sarachana, T., Kim, K. S., Nguyen, A., Kulkarni, S., Steinberg, M. E., Luu, T., Lai, Y. L., and Lee, N. H. (2009). "Gene expression profiling differentiates autism case-controls and phenotypic variants of autism spectrum disorders: Evidence for circadian rhythm dysfunction in severe autism." *Autism Research*, 2(2), 78-97.
- Hu, V. W., and Steinberg, M. E. (2009). "Novel clustering of items from the Autism Diagnostic Interview-Revised to define phenotypes within autism

spectrum disorders." *Autism Research*, 2(2), 67-77.

Huang, J. C., Svoboda, D. L., Reardon, J. T., and Sancar, A. (1992). "Human nucleotide excision nuclease removes thymine dimers from DNA by incising the 22nd phosphodiester bond 5' and the 6th phosphodiester bond 3' to the photodimer." *Proceedings of the National Academy of Sciences of the United States of America*, 89(8), 3664-3668.

Hung, H. C., Maurer, C., Kay, S. A., and Weber, F. (2007). "Circadian transcription depends on limiting amounts of the transcription co-activator *nejire*/CBP." *Journal of Biological Chemistry*, 282, 31349-31357.

Hyman, S. E. (2008). "A glimmer of light for neuropsychiatric disorders." *Nature*, 455(7215), 890-893.

Iguchi-iriga, S. M. M., and Schaffner, W. (1989). "CpG methylation of the cAMP-responsive enhancer promoter sequence TGACGTCA abolishes specific factor binding as well as transcriptional activation." *Genes & Development*, 3(5), 612-619.

Impey, S., Davare, M., Lasiek, A., Fortin, D., Ando, H., Varlamova, O., Obrietan, K., Soderling, T. R., Goodman, R. H., and Wayman, G. A. (2010). "An activity-induced microRNA controls dendritic spine formation by regulating Rac1-PAK signaling." *Molecular and Cellular Neuroscience*, 43(1), 146-156.

Inui, N., and Asama, K. (2003). "Timing of bimanual rhythmic finger tapping in adolescents with mental retardation or autism." *Journal of Human Movement Studies*, 45(1), 59-80.

Isles, A. R., Davies, W., and Wilkinson, L. S. (2006). "Genomic imprinting and the social brain." *Philosophical Transactions of the Royal Society B-Biological Sciences*, 361(1476), 2229-2237.

- Jackson, V., Granner, D., and Chalkley, R. (1976). "Deposition of histone onto replicating chromosome - newly synthesized histone is not found near replication fork." *Proceedings of the National Academy of Sciences of the United States of America*, 73(7), 2266-2269.
- Jaffe, J., Beebe, B., Feldstein, S., Crown, C. L., and Jasnow, M. D. (2001). "Rhythms of dialogue in infancy: Coordinated timing in development - Introduction." *Monographs of the Society for Research in Child Development*, 66(2), 1-+.
- James, S. J., Cutler, P., Melnyk, S., Jernigan, S., Janak, L., Gaylor, D. W., and Neubrandner, J. A. (2004). "Metabolic biomarkers of increased oxidative stress and impaired methylation capacity in children with autism." *American Journal of Clinical Nutrition*, 80(6), 1611-1617.
- James, S. J., Melnyk, S., Jernigan, S., Cleves, M. A., Halsted, C. H., Wong, D. H., Cutler, P., Bock, K., Boris, M., Bradstreet, J. J., Baker, S. M., and Gaylor, D. W. (2006). "Metabolic endophenotype and related genotypes are associated with oxidative stress in children with autism." *American Journal of Medical Genetics Part B-Neuropsychiatric Genetics*, 141B(8), 947-956.
- James, S. J., Melnyk, S., Jernigan, S., Hubanks, A., Rose, S., and Gaylor, D. W. (2008). "Abnormal transmethylation/transsulfuration metabolism and DNA hypomethylation among parents of children with autism." *Journal of Autism and Developmental Disorders*, 38(10), 1966-1976.
- Jasnow, M., Crown, C. L., Feldstein, S., Taylor, L., Beebe, B., and Jaffe, J. (1988). "Coordinated interpersonal timing of downs-syndrome and nondelayed infants with their mothers - evidence for a buffered mechanism of social-interaction." *Biological Bulletin*, 175(3), 355-360.

- Jeltsch, A. (2006). "Molecular enzymology of mammalian DNA methyltransferases." *DNA Methylation: Basic Mechanisms*, 301, 203-225.
- Ji, Y., Qin, Y., Shu, H., and Li, X. (2010). "Methylation analyses on promoters of mPer1, mPer2, and mCry1 during perinatal development." *Biochemical and Biophysical Research Communications*, 391(4), 1742-1747.
- Jin, X., Shearman, L. P., Weaver, D. R., Zylka, M. J., de Vries, G. J., Reppert, S. M. (1999). "A molecular mechanism regulating rhythmic output from the suprachiasmatic circadian clock." *Cell*, 96(1), 57-68.
- Johnston, R. J., and Hobert, O. (2003). "A microRNA controlling left/right neuronal asymmetry in *Caenorhabditis elegans*." *Nature*, 426(6968), 845-849.
- Jones, P. L., Veenstra, G. J. C., Wade, P. A., Vermaak, D., Kass, S. U., Landsberger, N., Strouboulis, J., and Wolffe, A. P. (1998). "Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription." *Nature Genetics*, 19(2), 187-191.
- Jordan, C., Li, H. H., Kwan, H. C., and Francke, U. (2007). "Cerebellar gene expression profiles of mouse models for Rett syndrome reveal novel MeCP2 targets." *Bmc Medical Genetics*, 8.
- Jovanovic, M., and Hengartner, M. O. (2006). "miRNAs and apoptosis: RNAs to die for." *Oncogene*, 25(46), 6176-6187.
- Just, M. A., Cherkassky, V. L., Keller, T. A., Kana, R. K., and Minshew, N. J. (2007). "Functional and anatomical cortical underconnectivity in autism: Evidence from an fMRI study of an executive function task and corpus callosum morphometry." *Cerebral Cortex*, 17(4), 951-961.
- Kadener, S., Vilella, A., Kula, E., Palm, K., Pyza, E., Botas, J., Hall, J. C., and Rosbash, M. (2006). "Neurotoxic protein expression reveals connections



between the circadian clock and mating behavior in *Drosophila*."

*Proceedings of the National Academy of Sciences of the United States of America*, 103(36), 13537-13542.

Kageyama, R., Masamizu, Y., and Niwa, Y. (2007). "Oscillator mechanism of notch pathway in the segmentation clock." *Developmental Dynamics*, 236(6), 1403-1409.

Kaminsky, Z., Wang, S. C., and Petronis, A. (2006). "Complex disease, gender and epigenetics." *Annals of Medicine*, 38(8), 530-544.

Kang, T. H., Lindsey-Boltz, L. A., Reardon, J. T., and Sancar, A. (2010). "Circadian control of XPA and excision repair of cisplatin-DNA damage by cryptochrome and HERC2 ubiquitin ligase." *Proceedings of the National Academy of Sciences of the United States of America*, 107(11), 4890-4895.

Kang, T. H., Reardon, J. T., Kemp, M., and Sancar, A. (2009). "Circadian oscillation of nucleotide excision repair in mammalian brain." *Proceedings of the National Academy of Sciences of the United States of America*, 106(8), 2864-2867.

Kang, T. H., and Sancar, A. (2009). "Circadian regulation of DNA excision repair Implications for chrono-chemotherapy." *Cell Cycle*, 8(11), 1665-1667.

Kao, H. T., Buka, S. L., Kelsey, K. T., Gruber, D. F., and Porton, B. (2010). "The correlation between rates of cancer and autism: An exploratory ecological investigation." *Plos One*, 5(2).

Kelly, T. J., Qin, S., Gottschling, D. E., and Parthun, M. R. (2000). "Type B histone acetyltransferase Hat1p participates in telomeric silencing." *Molecular and Cellular Biology*, 20(19), 7051-7058.

- Kemper, T. L., and Bauman, M. L. (2002). "Neuropathology of infantile autism." *Molecular Psychiatry*, 7, S12-S13.
- Kennaway, D. J., Owens, J. A., Voultios, A., and Varcoe, T. J. (2006). "Functional central rhythmicity and light entrainment, but not liver and muscle rhythmicity, are Clock independent." *American Journal of Physiology-Regulatory Integrative and Comparative Physiology*, 291(4), R1172-R1180.
- Kim, A. H., Reimers, M., Mahera, B., Williamson, V., McMichael, O. I., McClay, J. L., van den Oord, E. J. C. G., Riley, B. P., Kendler, K. S., and Vladimirov, V. I. (2010). "MicroRNA expression profiling in the prefrontal cortex of individuals affected with schizophrenia and bipolar disorders." *Schizophrenia Research*, 124(1-3), 183-191.
- Kim, S. J., Cox, N., Courchesne, R., Lord, C., Corsello, C., Akshoomoff, N., Guter, S., Leventhal, B. L., Courchesne, E., and Cook, E. H. (2002). "Transmission disequilibrium mapping at the serotonin transporter gene (SLC6A4) region in autistic disorder." *Molecular Psychiatry*, 7(3), 278-288.
- Kim, V. N., and Nam, J. W. (2006). "Genomics of microRNA." *Trends in Genetics*, 22(3), 165-173.
- Kim, Y. K., and Kim, V. N. (2007). "Processing of intronic microRNAs." *Embo Journal*, 26(3), 775-783.
- Kirov, G., Rujescu, D., Ingason, A., Collier, D. A., O'Donovan, M. C., and Owen, M. J. (2009). "Neurexin 1 (NRXN1) deletions in schizophrenia." *Schizophrenia Bulletin*, 35(5), 851-854.
- Kishino, T., Lalande, M., and Wagstaff, J. (1997). "UBE3A/E6-AP mutations cause Angelman syndrome." *Nature Genetics*, 15(1), 70-73.

- Kiyama, A., Isojima, Y., and Nagai, K. (2006). "Role of Per1-interacting protein of the suprachiasmatic nucleus in NGF mediated neuronal survival." *Biochemical and Biophysical Research Communications*, 339(2), 514-519.
- Kleff, S., Andrulis, E. D., Anderson, C. W., and Sternglanz, R. (1995). "Identification of a gene encoding a yeast histone H4 acetyltransferase." *Journal of Biological Chemistry*, 270(42), 24674-24677.
- Klein, M. E., Liroy, D. T., Ma, L., Impey, S., Mandel, G., and Goodman, R. H. (2007). "Homeostatic regulation of MeCP2 expression by a CREB-induced microRNA." *Nature Neuroscience*, 10, 1513-1514.
- Knickmeyer, R. C., and Baron-Cohen, S. (2006). "Fetal testosterone and sex differences in typical social development and in autism." *Journal of Child Neurology*, 21(10), 825-845.
- Ko, H. W., Kim, E. Y., Chiu, J., Vanselow, J. T., Kramer, A., and Edery, I. (2010). "A hierarchical phosphorylation cascade that regulates the timing of PERIOD nuclear entry reveals novel roles for proline-directed kinases and GSK-3 beta/SGG in circadian clocks." *Journal of Neuroscience*, 30(38), 12664-12675.
- Kobrynski, L. J., and Sullivan, K. E. (2007). "Velocardiofacial syndrome, DiGeorge syndrome: the chromosome 22q11.2 deletion syndromes." *Lancet*, 370, 1443-1452.
- Kocerha, J., Faghihi, M. A., Lopez-Toledano, M. A., Huang, J., Ramsey, A. J., Caron, M. G., Sales, N., Willoughby, D., Elmen, J., Hansen, H. F., Orum, H., Kauppinen, S., Kenny, P. J., and Wahlestedt, C. (2009). "MicroRNA-219 modulates NMDA receptor-mediated neurobehavioral dysfunction." *Proceedings of the National Academy of Sciences of the United States of America*, 106(9), 3507-3512.

- Koekkoek, S. K. E., Hulscher, H. C., Dortland, B. R., Hensbroek, R. A., Elgersma, Y., Ruigrok, T. J. H., and De Zeeuw, C. I. (2003). "Cerebellar LTD and learning-dependent timing of conditioned eyelid responses." *Science*, 301(5640), 1736-1739.
- Kokura, K., Kaul, S. C., Wadhwa, R., Nomura, T., Khan, M. M., Shinagawa, T., Yasukawa, T., Colmenares, C., and Ishii, S. (2001). "The ski protein family is required for MeCP2-mediated transcriptional repression." *Journal of Biological Chemistry*, 276(36), 34115-34121.
- Kondratov, R. V., and Antoch, M. P. (2007). "Circadian proteins in the regulation of cell cycle and genotoxic stress responses." *Trends in Cell Biology*, 17(7), 311-317.
- Kondratov, R. V., Kondratova, A. A., Gorbacheva, V. Y., Vykhovanets, O. V., and Antoch, M. P. (2006). "Early aging and age-related pathologies in mice deficient in BMAL1, the core component of the circadian clock." *Genes & Development*, 20(14), 1868-1873.
- Konopka, R. J., and Benzer, S. (1971). "Clock mutants of *Drosophila-melanogaster*." *Proceedings of the National Academy of Sciences of the United States of America*, 68(9), 2112-&.
- Konopka, R. J., Kyriacou, C. P., and Hall, J. C. (1996). "Mosaic analysis in the *Drosophila* CNS of circadian and courtship-song rhythms affected by a period clock mutation." *Journal of Neurogenetics*, 11(1-2), 117-&.
- Kornberg, R. D., and Thomas, J. O. (1974). "Chromatin structure - oligomers of histones." *Science*, 184(4139), 865-868.
- Kosik, K. S. (2006). "The neuronal microRNA system." *Nature Reviews Neuroscience*, 7, 911-920.

- Kotani, S., Kawahara, S., and Kirino, Y. (2003). "Purkinje cell activity during learning a new timing in classical eyeblink conditioning." *Brain Research*, 994(2), 193-202.
- Kovanen, L., Saarikoski, S. T., Aromaa, A., Lönnqvist, J., Partonen T. (2010) "ARNTL (BMAL1) and NPAS2 gene variants contribute to fertility and seasonality." *PLoS One*, 5(4), e10007.
- Kowalska, E., and Brown, S. A. (2007). "Peripheral clocks: Keeping up with the master clock." *Cold Spring Harbor Symposia on Quantitative Biology*, 72, 301-305.
- Kozlov, S. V., Bogenpohl, J. W., Howell, M. P., Wevrick, R., Panda, S., Hogenesch, J. B., Muglia, L. J., Van Gelder, R. N., Herzog, E. D., and Stewart, C. L. (2007). "The imprinted gene *Magel2* regulates normal circadian output." *Nature Genetics*, 39(10), 1266-1272.
- Kubicek, L. (1980). "Organization in two mother–infant interactions involving a normal infant and his fraternal twin brother who was later diagnosed as autistic." *High-Risk Infants and Children. T. Field, S. Goldberg, D. Stern and A. Sostek. New York, Academic Press: 99–110.*
- Kucharski, R., Maleszka, J., Foret, S., and Maleszka, R. (2008). "Nutritional control of reproductive status in honeybees via DNA methylation." *Science*, 319(5871), 1827-1830.
- Kuczewski, N., Porcher, C., Lessmann, V., Medina, I., and Gaiarsa, J. L. (2009). "Activity-Dependent Dendritic Release of BDNF and Biological Consequences." *Molecular Neurobiology*, 39(1), 37-49.
- Kulkarni, S. J., Steinlauf, A. F., and Hall, J. C. (1988). "The dissonance mutant of courtship song in *Drosophila-melanogaster* - isolation, behavior and

cytogenetics." *Genetics*, 118(2), 267-285.

Kulman, G., Lissoni, P., Rovelli, F., Roselli, M. G., Brivio, F., and Sequeri, P.

(2000). "Evidence of pineal endocrine hypofunction in autistic children."

*Neuroendocrinology Letters*, 21(1), 31-34.

Kumaki, Y., Ukai-Tadenuma, M., Uno, K. D., Nishio, J., Masumoto, K., Nagano,

M., Komori, T., Shigeyoshi, Y., Hogenesch, J. B., and Ueda, H. R. (2008).

"Analysis and synthesis of high-amplitude Cis-elements in the mammalian circadian clock." *Proceedings of the National Academy of Sciences of the United States of America*, 105(39), 14946-14951.

Kume, K., Zylka, M. J., Sriram, S., Shearman, L. P., Weaver, D. R., Jin, X. W.,

Maywood, E. S., Hastings, M. H., Reppert, S. M. (1999) "mCRY1 and

mCRY2 are essential components of the negative limb of the circadian clock feedback loop." *Cell* 98 (2)2,193-205.

Kurabayashi, N., Hirota, T., Sakai, M., Sanada, K., and Fukada, Y. (2010).

"DYRK1A and Glycogen Synthase Kinase 3 beta, a dual-kinase mechanism directing proteasomal degradation of CRY2 for circadian timekeeping." *Molecular and Cellular Biology*, 30(7), 1757-1768.

Kyriacou, C. P., Greenacre, M. L., Ritchie, M. G., Byrne, B. C., and Hall, J. C.

(1992). "Genetic and molecular analysis of the love song preferences of *Drosophila* females." *American Zoologist*, 32(1), 31-39.

Kyriacou, C. P., and Hall, J. C. (1980). "Circadian-rhythm mutations in

*Drosophila-melanogaster* affect short-term fluctuations in the males courtship song." *Proceedings of the National Academy of Sciences of the United States of America-Biological Sciences*, 77(11), 6729-6733.

Kyriacou, C. P., Oldroyd, M., Wood, J., Sharp, M., and Hill, M. (1990). "Clock mutations alter developmental timing in *Drosophila*." *Heredity*, 64, 395-

401.

- Lachman, H. M., Pedrosa, E., Petruolo, O. A., Cockerham, M., Papolos, A., Novak, T., Papolos, D. F., and Stopkova, P. (2007). "Increase in GSK3 beta gene copy number variation in bipolar disorder." *American Journal of Medical Genetics Part B-Neuropsychiatric Genetics*, 144B(3), 259-265.
- Ladd, A. N., and Cooper, T. A. (2002). "Finding signals that regulate alternative splicing in the postgenomic era." *Genome Biology*, 3(11), ISSN 1474-760X(print)|1474-760X(electronic).
- Lagier, S., Panzanelli, P., Russo, R. E., Nissant, A., Bathellier, B., Sassoe-Pognetto, M., Fritschy, J. M., and Lledo, P. M. (2007). "GABAergic inhibition at dendrodendritic synapses tunes gamma oscillations in the olfactory bulb." *Proceedings of the National Academy of Sciences of the United States of America*, 104(17), 7259-7264.
- Lalande, M., and Calciano, M. A. (2007). "Molecular epigenetics of Angelman syndrome." *Cellular and Molecular Life Sciences*, 64(7-8), 947-960.
- Langner, G. (1992). "Periodicity coding in the auditory-system." *Hearing Research*, 60(2), 115-142.
- Le Hellard, S., Muhleisen, T. W., Djurovic, S., Ferno, J., Ouriaghi, Z., Mattheisen, M., Vasilescu, C., Raeder, M. B., Hansen, T., Strohmaier, J., Georgi, A., Brockschmidt, F. F., Melle, I., Nenadic, I., Sauer, H., Rietschel, M., Nothen, M. M., Werge, T., Andreassen, O. A., Cichon, S., and Steen, V. M. (2010). "Polymorphisms in SREBF1 and SREBF2, two antipsychotic-activated transcription factors controlling cellular lipogenesis, are associated with schizophrenia in German and Scandinavian samples." *Molecular Psychiatry*, 15(5), 463-472.

- Lee, C., Etchegaray, J. P., Cagampang, F. R. A., Loudon, A. S. I., and Reppert, S. M. (2001). "Posttranslational mechanisms regulate the mammalian circadian clock." *Cell*, 107(7), 855-867.
- Lee, G. H., Bahn, J. H., and Park, J. H. (2006). "Sex- and clock-controlled expression of the neuropeptide F gene in *Drosophila*." *Proceedings of the National Academy of Sciences of the United States of America*, 103(33), 12580-12585.
- Lee, M. T., and Kim, J. (2008). "Self Containment, a Property of Modular RNA Structures, Distinguishes microRNAs." *Plos Computational Biology*, 4(8).
- Lee, R. C., Feinbaum, R. L., and Ambros, V. (1993). "The *C-elegans* heterochronic gene LIN-4 encodes small RNAs with antisense complementarity to LIN-14." *Cell*, 75(5), 843-854.
- Leloup, J. C., and Goldbeter, A. (2011). "Modelling the dual role of Per phosphorylation and its effect on the period and phase of the mammalian circadian clock." *let Systems Biology*, 5(1), 44-U75.
- Levenson, J. M., and Sweatt, J. D. (2006). "Epigenetic mechanisms: a common theme in vertebrate and invertebrate memory formation." *Cellular and Molecular Life Sciences*, 63(9), 1009-1016.
- Levine, J. D., Funes, P., Dowse, H. B., and Hall, J. C. (2002). "Resetting the circadian clock by social experience in *Drosophila melanogaster*." *Science*, 298(5600), 2010-2012.
- Lewis, B. P., Burge, C. B., and Bartel, D. P. (2005). "Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets." *Cell*, 120(1), 15-20.
- Lim, J., Hao, T., Shaw, C., Patel, A. J., Szabo, G., Rual, J. F., Fisk, C. J., Li, N., Smolyar, A., Hill, D. E., Barabasi, A. L., Vidal, M., and Zoghbi, H. Y.



- (2006). "A protein-protein interaction network for human inherited ataxias and disorders of Purkinje cell degeneration." *Cell*, 125(4), 801-814.
- Lin, S. L., Chang, S. J. E., and Ying, S. Y. (2006a). "First in vivo evidence of microRNA-induced fragile X mental retardation syndrome." *Molecular Psychiatry*, 11(7), 616-617.
- Lin, S. L., Miller, J. D., and Ying, S. Y. (2006b). "Intronic MicroRNA (miRNA)." *Journal of Biomedicine and Biotechnology*.
- Lintas, C., and Persico, A. M. (2010). "Neocortical RELN promoter methylation increases significantly after puberty." *Neuroreport*, 21(2), 114-118.
- Liu, D., Gharavi, R., Pitta, M., Gleichmann, M., and Mattson, M. P. (2009). "Nicotinamide prevents NAD(+) depletion and protects neurons against excitotoxicity and cerebral ischemia: NAD(+) consumption by SIRT1 may endanger energetically compromised neurons." *Neuromolecular Medicine*, 11(1), 28-42.
- Looby, P., and Loudon, A. S. I. (2005). "Gene duplication and complex circadian clocks in mammals." *Trends in Genetics*, 21(1), 46-53.
- Lord, C., Risi, S., Lambrecht, L., Cook, E. H., Leventhal, B. L., DiLavore, P. C., Pickles, A., and Rutter, M. (2000). "The Autism Diagnostic Observation Schedule-Generic: A standard measure of social and communication deficits associated with the spectrum of autism." *Journal of Autism and Developmental Disorders*, 30(3), 205-223.
- Lord, C., Rutter, M., and Lecouteur, A. (1994). "Autism Diagnostic Interview-Revised - A revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders." *Journal of Autism and Developmental Disorders*, 24(5), 659-685.

- Lorenzo, O., Urbe, S., and Clague, M. J. (2006). "Systematic analysis of myotubularins: heteromeric interactions, subcellular localisation and endosome-related functions." *Journal of Cell Science*, 119(14), 2953-2959.
- Loudon, A. S. I., Wayne, N. L., Krieg, R., Iranmanesh, A., Veldhuis, J. D., and Menaker, M. (1994). "Ultradian endocrine rhythms are altered by a circadian mutation in the syrian-hamster." *Endocrinology*, 135(2), 712-718.
- Lovenberg, T. W., Roland, B. L., Wilson, S. J., Jiang, X. X., Pyati, J., Huvar, A., Jackson, M. R., and Erlander, M. G. (1999). "Cloning and functional expression of the human histamine H-3 receptor." *Molecular Pharmacology*, 55(6), 1101-1107.
- Low, K. H., Lim, C., Ko, H. W., and Ederyl, I. (2008). "Natural variation in the splice site strength of a clock gene and species-specific thermal adaptation." *Neuron*, 60(6), 1054-1067.
- Luger, K., Mader, A. W., Richmond, R. K., Sargent, D. F., and Richmond, T. J. (1997). "Crystal structure of the nucleosome core particle at 2.8 angstrom resolution." *Nature*, 389(6648), 251-260.
- Lugli, G., Larson, J., Martone, M. E., Jones, Y., and Smalheiser, N. R. (2005). "Dicer and eIF2c are enriched at postsynaptic densities in adult mouse brain and are modified by neuronal activity in a calpain-dependent manner." *Journal of Neurochemistry*, 94(4), 896-905.
- Lunter, G., Ponting, C. P., and Hein, J. (2006). "Genome-wide identification of human functional DNA using a neutral indel model." *Plos Computational Biology*, 2(1), 2-12.
- Ma, D. Q., Salyakina, D., Jaworski, J. M., Konidari, I., Whitehead, P. L., Andersen, A. N., Hoffman, J. D., Slifer, S. H., Hedges, D. J., Cukier, H. N.,

- Griswold, A. J., McCauley, J. L., Beecham, G. W., Wright, H. H., Abramson, R. K., Martin, E. R., Hussman, J. P., Gilbert, J. R., Cuccaro, M. L., Haines, J. L., and Pericak-Vance, M. A. (2009). "A Genome-wide Association Study of Autism Reveals a Common Novel Risk Locus at 5p14.1." *Annals of Human Genetics*, 73, 263-273.
- Ma, D. Q., Whitehead, P. L., Menold, M. M., Martin, E. R., Ashley-Koch, A. E., Mei, H., Ritchie, M. D., DeLong, G. R., Abramson, R. K., Wright, H. H., Cuccaro, M. L., Hussman, J. P., Gilbert, J. R., and Pericak-Vance, M. A. (2005). "Identification of significant association and gene-gene interaction of GABA receptor subunit genes in autism." *American Journal of Human Genetics*, 77(3), 377-388.
- Manoli, D. S., Foss, M., Vellella, A., Taylor, B. J., Hall, J. C., and Baker, B. S. (2005). "Male-specific fruitless specifies the neural substrates of *Drosophila* courtship behaviour." *Nature*, 436(7049), 395-400.
- Mansour, H. A., Monk, T. H., and Nimgaonkar, V. L. (2005). "Circadian genes and bipolar disorder." *Annals of Medicine*, 37(3), 196-205.
- Mansour, H. A., Wood, J., Logue, T., Chowdari, K. V., Dayal, M., Kupfer, D. J., Monk, T. H., Devlin, B., and Nimgaonkar, V. L. (2006). "Association study of eight circadian genes with bipolar I disorder, schizoaffective disorder and schizophrenia." *Genes Brain and Behavior*, 5(2), 150-157.
- Mark, M., Ghyselinck, N. B., and Chambon, P. (2009). "Function of retinoic acid receptors during embryonic development." *Nucl Recept Signal*, 7, e002.
- Margolis, S. S., Salogiannis, J., Lipton, D. M., Mandel-Brehm, C., Wills, Z. P., Mardinly, A. R., Hu, L., Greer, P. L., Bikoff, J. B., Ho, H.-Y. H., Soskis, M. J., Sahin, M., Greenberg, M. E. (2010) "EphB-mediated degradation of the RhoA GEF ephexin5 relieves a developmental brake on excitatory

synapse formation." *Cell* 143, 442-455.

- Marshall, C. R., Noor, A., Vincent, J. B., Lionel, A. C., Feuk, L., Skaug, J., Shago, M., Moessner, R., Pinto, D., Ren, Y., Thiruvahindrapduram, B., Fiebig, A., Schreiber, S., Friedman, J., Ketelaars, C. E. J., Vos, Y. J., Ficiocioglu, C., Kirkpatrick, S., Nicolson, R., Sloman, L., Summers, A., Gibbons, C. A., Teebi, A., Chitayat, D., Weksberg, R., Thompson, A., Vardy, C., Crosbie, V., Luscombe, S., Baatjes, R., Zwaigenbaum, L., Roberts, W., Fernandez, B., Szatmari, P., and Scherer, S. W. (2008). "Structural variation of chromosomes in autism spectrum disorder." *American Journal of Human Genetics*, 82(2), 477-488.
- Martinowich, K., Hattori, D., Wu, H., Fouse, S., He, F., Hu, Y., Fan, G. P., and Sun, Y. E. (2003). "DNA methylation-related chromatin remodeling in activity-dependent Bdnf gene regulation." *Science*, 302(5646), 890-893.
- Matsuki, T., Kiyama, A., Kawabuchi, M., Okada, M., and Nagai, K. (2001). "A novel protein interacts with a clock-related protein, rPer1." *Brain Research*, 916(1-2), 1-10.
- Matsuo, T., Yamaguchi, S., Mitsui, S., Emi, A., Shimoda, F., and Okamura, H. (2003). "Control mechanism of the circadian clock for timing of cell division in vivo." *Science*, 302(5643), 255-259.
- Matsuura, T., Sutcliffe, J. S., Fang, P., Galjaard, R. J., Jiang, Y. H., Benton, C. S., Rommens, J. M., and Beaudet, A. L. (1997). "De novo truncating mutations in E6-AP ubiquitin-protein ligase gene (UBE3A) in Angelman syndrome." *Nature Genetics*, 15(1), 74-77.
- Mayer, W., Smith, A., Fundele, R., and Haaf, T. (2000). "Spatial separation of parental genomes in preimplantation mouse embryos." *Journal of Cell Biology*, 148(4), 629-634.

- McCarthy, S. E., Makarov, V., Kirov, G., Addington, A. M., McClellan, J., Yoon, S., Perkins, D. O., Dickel, D. E., Kusenda, M., Krastoshevsky, O., Krause, V., Kumar, R. A., Grozeva, D., Malhotra, D., Walsh, T., Zackai, E. H., Kaplan, P., Ganesh, J., Krantz, I. D., Spinner, N. B., Roccanova, P., Bhandari, A., Pavon, K., Lakshmi, B., Leotta, A., Kendall, J., Lee, Y. H., Vacic, V., Gary, S., Iakoucheva, L. M., Crow, T. J., Christian, S. L., Lieberman, J. A., Stroup, T. S., Lehtimaki, T., Puura, K., Haldeman-Englert, C., Pearl, J., Goodell, M., Willour, V. L., DeRosse, P., Steele, J., Kassem, L., Wolff, J., Chitkara, N., McMahon, F. J., Malhotra, A. K., Potash, J. B., Schulze, T. G., Nothen, M. M., Cichon, S., Rietschel, M., Leibenluft, E., Kustanovich, V., Lajonchere, C. M., Sutcliffe, J. S., Skuse, D., Gill, M., Gallagher, L., Mendell, N. R., Craddock, N., Owen, M. J., O'Donovan, M. C., Shaikh, T. H., Susser, E., DeLisi, L. E., Sullivan, P. F., Deutsch, C. K., Rapoport, J., Levy, D. L., King, M. C., Sebat, J., and Wellcome Trust Case Control, C. (2009). "Microduplications of 16p11.2 are associated with schizophrenia." *Nature Genetics*, 41(11), 1223-U85.
- McNamara, P., Seo, S. B., Rudic, R. D., Sehgal, A., Chakravarti, D., and FitzGerald, G. A. (2001). "Regulation of CLOCK and MOP4 by nuclear hormone receptors in the vasculature: A humoral mechanism to reset a peripheral clock." *Cell*, 105(7), 877-889.
- Meehan, R. R. (2003). "DNA methylation in animal development." *Seminars in Cell & Developmental Biology*, 14(1), 53-65.
- Mehnert, K. I., Beramendi, A., Elghazali, F., Negro, P., Kyriacou, C. P., and Cantera, R. (2007). "Circadian changes in *Drosophila* motor terminals." *Developmental Neurobiology*, 67(4), 415-421.

- Mehnert, K. I., and Cantera, R. (2008). "A peripheral pacemaker drives the circadian rhythm of synaptic boutons in *Drosophila* independently of synaptic activity." *Cell and Tissue Research*, 334(1), 103-109.
- Meins, M., Lehmann, J., Gerresheim, F., Herchenbach, J., Hagedorn, M., Hameister, K., and Epplen, J. T. (2005). "Submicroscopic duplication in Xq28 causes increased expression of the MECP2 gene in a boy with severe mental retardation and features of Rett syndrome." *Journal of Medical Genetics*, 42(2).
- Melke, J., Botros, H. G., Chaste, P., Betancur, C., Nygren, G., Anckarsater, H., Rastam, M., Stahlberg, O., Gillberg, I. C., Delorme, R., Chabane, N., Mouren-Simeoni, M. C., Fauchereau, F., Durand, C. M., Chevalier, F., Drouot, X., Collet, C., Launay, J. M., Leboyer, M., Gillberg, C., Bourgeron, T., and Study, P. (2008). "Abnormal melatonin synthesis in autism spectrum disorders." *Molecular Psychiatry*, 13(1), 90-98.
- Mercer, R. E., Kwolek, E. M., Bischof, J. M., van Eede, M., Henkelman, R. M., and Wevrick, R. (2009). "Regionally reduced brain volume, altered serotonin neurochemistry, and abnormal behavior in mice null for the circadian rhythm output gene *Magel2*." *American Journal of Medical Genetics Part B-Neuropsychiatric Genetics*, 150B(8), 1085-1099.
- Meunier, N., Belgacem, Y. H., and Martin, J.-R. (2007). "Regulation of feeding behaviour and locomotor activity by takeout in *Drosophila*." *Journal of Experimental Biology*, 210(8), 1424-1434.
- Michael, M. Z., O'Connor, S. M., Pellekaan, N. G. V., Young, G. P., and James, R. J. (2003). "Reduced accumulation of specific microRNAs in colorectal neoplasia." *Molecular Cancer Research*, 1(12), 882-891.

- Millen, K. J., Wurst, W., Herrup, K., and Joyner, A. L. (1994). "abnormal embryonic cerebellar development and patterning of postnatal foliation in 2 mouse *Engrailed-2* mutants." *Development*, 120(3), 695-706.
- Millward, C., Powell, S., Messer, D., and Jordan, R. (2000). "Recall for self and other in autism: Children's memory for events experienced by themselves and their peers." *Journal of Autism and Developmental Disorders*, 30(1), 15-28.
- Minshew, N. J., and Williams, D. L. (2007). "The new neurobiology of autism - Cortex, connectivity, and neuronal organization." *Archives of Neurology*, 64(7), 945-950.
- Mistlberger, R. E., and Skene, D. J. (2004). "Social influences on mammalian circadian rhythms: animal and human studies." *Biological Reviews*, 79(3), 533-556.
- Miura, S., Nozawa, M., and Nei, M. (2011). "Evolutionary changes of the target sites of two MicroRNAs encoded in the Hox gene cluster of *Drosophila* and other insect species." *Genome Biology and Evolution*, 3, 129-139.
- Miyamoto, N., Izumi, H., Noguchi, T., Nakajima, Y., Ohmiya, Y., Shiota, M., Kidani, A., Tawara, A., and Kohno, K. (2008). "Tip60 is regulated by circadian transcription factor clock and is involved in cisplatin resistance." *Journal of Biological Chemistry*, 283(26), 18218-18226.
- Miyatake, M., Kuno, T., Kita, A., Katsura, K., Takegawa, K., Uno, S., Nabata, T., and Sugiura, R. (2007). "Valproic acid affects membrane trafficking and cell-wall integrity in fission yeast." *Genetics*, 175(4), 1695-1705.
- Miyatake, T., and Kanmiya, K. (2004). "Male courtship song in circadian rhythm mutants of *Bactrocera cucurbitae* (Tephritidae : Diptera)." *Journal of Insect Physiology*, 50(1), 85-91.

- Miyazaki, K., Mapendano, C. K., Fuchigami, T., Kondo, S., Ohta, T., Kinoshita, A., Tsukamoto, K., Yoshiura, K. I., Niikawa, N., and Kishino, T. (2009). "Developmentally dynamic changes of DNA methylation in the mouse Snurf/Snrpn gene." *Gene*, 432(1-2), 97-101.
- Mizzen, C. A., Yang, X. J., Kokubo, T., Brownell, J. E., Bannister, A. J., OwenHughes, T., Workman, J., Wang, L., Berger, S. L., Kouzarides, T., Nakatani, Y., and Allis, C. D. (1996). "The TAF(II)250 subunit of TFIID has histone acetyltransferase activity." *Cell*, 87(7), 1261-1270.
- Moessner, R., Marshall, C. R., Sutcliffe, J. S., Skaug, J., Pinto, D., Vincent, J., Zwaigenbaum, L., Fernandez, B., Roberts, W., Szatmari, P., and Scherer, S. W. (2007). "Contribution of SHANK3 mutations to autism spectrum disorder." *American Journal of Human Genetics*, 81(6), 1289-1297.
- Moller, A. R. (1983). "Frequency-selectivity of phase-locking of complex sounds in the auditory-nerve of the rat." *Hearing Research*, 11(3), 267-284.
- Moore, S. J., Turnpenny, P., Quinn, A., Glover, S., Lloyd, D. J., Montgomery, T., and Dean, J. C. S. (2000). "A clinical study of 57 children with fetal anticonvulsant syndromes." *Journal of Medical Genetics*, 37(7), 489-497.
- Moore, T., and Haig, D. (1991). "Genomic imprinting in mammalian development - a parental tug-of-war." *Trends in Genetics*, 7(2), 45-49.
- Mor, E., Cabilly, Y., Goldshmit, Y., Zalts, H., Modai, S., Edry, L., Elroy-Stein, O., and Shomron, N. (2011). "Species-specific microRNA roles elucidated following astrocyte activation." *Nucleic Acids Research*, 39(9), 3710-3723.
- Moran, C. N., and Kyriacou, C. P. (2009). "Functional neurogenomics of the courtship song of male *Drosophila melanogaster*." *Cortex*, 45(1), 18-34.
- Moretti, P., and Zoghbi, H. Y. (2006). "MeCP2 dysfunction in Rett syndrome and related disorders." *Current Opinion in Genetics & Development*, 16(3),



276-281.

Morgan, H. D., Santos, F., Green, K., Dean, W., and Reik, W. (2005). "Epigenetic reprogramming in mammals." *Human Molecular Genetics*, 14, R47-R58.

Morse, D., and Sassone-Corsi, P. (2002). "Time after time: inputs to and outputs from the mammalian circadian oscillators." *Trends in Neurosciences*, 25(12), 632-637.

Mottron, L. (2004). "Matching strategies in cognitive research with individuals with high-functioning autism: Current practices, instrument biases, and recommendations." *Journal of Autism and Developmental Disorders*, 34(1), 19-27.

Mount, S. M. (2000). "Genomic sequence, splicing, and gene annotation." *American Journal of Human Genetics*, 67(4), 788-792.

Mukhopadhyay, N. K., Ferdinand, A. S., Mukhopadhyay, L., Cinar, B., Lutchman, M., Richie, J. P., Freeman, M. R., and Liu, B. C. S. (2006). "Unraveling androgen receptor interactomes by an array-based method: Discovery of proto-oncoprotein c-Rel as a negative regulator of androgen receptor." *Experimental Cell Research*, 312(19), 3782-3795.

Mulle, J. G., Dodd, A. F., McGrath, J. A., Wolyniec, P. S., Mitchell, A. A., Shetty, A. C., Sobreira, N. L., Valle, D., Rudd, M. K., Satten, G., Cutler, D. J., Pulver, A. E., and Warren, S. T. (2010). "Microdeletions of 3q29 Confer High Risk for Schizophrenia." *American Journal of Human Genetics*, 87(2), 229-236.

Mullenders, J., Fabius, A. W. M., Madiredjo, M., Bernards, R., and Beijersbergen, R. L. (2009). "A Large scale shRNA barcode screen identifies the circadian clock component ARNTL as putative regulator of the p53 tumor suppressor pathway." *Plos One*, 4(3).

- Murray, L. and C. Trevarthen (1985). "Emotional regulation of interactions between two-month olds and their mothers." *Social Perception in Infants. Norwood, NJ., Ablex.*
- Murrell, A., Heeson, S., and Reik, W. (2004). "Interaction between differentially methylated regions partitions the imprinted genes Igf2 and H19 into parent-specific chromatin loops." *Nature Genetics*, 36(8), 889-893.
- Myakishev, M. V., Khripin, Y., Hu, S., and Hamer, D. H. (2001). "High-throughput SNP genotyping by allele-specific PCR with universal energy-transfer-labeled primers." *Genome Research*, 11(1), 163-169.
- Myers, S. M. (2007). "The status of pharmacotherapy For autism spectrum disorders." *Expert Opinion on Pharmacotherapy*, 8(11), 1579-1603.
- Nadel, J., Carchon, I., Kervella, C., Marcelli, D., and Reserbat-Plantey, D. (1999). "Expectancies for social contingency in 2-month-olds." *Developmental Science*, 2(2), 164-173.
- Nagarajan, R. P., Patzel, K. A., Martin, M., Yasui, D. H., Swanberg, S. E., Hertz-Picciotto, I., Hansen, R. L., Van de Water, J., Pessah, I. N., Jiang, R., Robinson, W. P., and LaSalle, J. M. (2008). "MECP2 Promoter methylation and X chromosome inactivation in autism." *Autism Research*, 1(3), 169-178.
- Nakahata, Y., Kaluzova, M., Grimaldi, B., Sahar, S., Hirayama, J., Chen, D., Guarente, L. P., and Sassone-Corsi, P. (2008). "The NAD(+)-dependent deacetylase SIRT1 modulates CLOCK-mediated chromatin remodeling and circadian control." *Cell*, 134(2), 329-340.
- Nakahata, Y., Sahar, S., Astarita, G., Kaluzova, M., and Sassone-Corsi, P. (2009). "Circadian control of the NAD(+) salvage pathway by CLOCK-SIRT1." *Science*, 324(5927), 654-657.

- Nan, X. S., Ng, H. H., Johnson, C. A., Laherty, C. D., Turner, B. M., Eisenman, R. N., and Bird, A. (1998). "Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex." *Nature*, 393(6683), 386-389.
- Newson, E. (1984). "The social development of the young autistic child. Autism - Its nature, Implications and Consequences." London: National Autistic Society.
- Nguyen, A., Rauch, T. A., Pfeifer, G. P., and Hu, V. W. (2010). "Global methylation profiling of lymphoblastoid cell lines reveals epigenetic contributions to autism spectrum disorders and a novel autism candidate gene, RORA, whose protein product is reduced in autistic brain." *Faseb Journal*, 24(8), 3036-3051.
- Nicholas, B., Rudrasingham, V., Nash, S., Kirov, G., Owen, M. J., and Wimpory, D. C. (2007). "Association of Per1 and Npas2 with autistic disorder: support for the clock genes/social timing hypothesis." *Molecular Psychiatry*, 12(6), 581-592.
- Nir, I., Meir, D., Zilber, N., Knobler, H., Hadjez, J., and Lerner, Y. (1995). "Brief report: Circadian melatonin, thyroid-stimulating hormone, prolactin, and cortisol levels in serum of young adults with autism." *Journal of Autism and Developmental Disorders*, 25(6), 641-654.
- O'Donovan, M. C., Craddock, N., Norton, N., Williams, H., Peirce, T., Moskvina, V., Nikolov, I., Hamshere, M., Carroll, L., Georgieva, L., Dwyer, S., Holmans, P., Marchini, J. L., Spencer, C. C. A., Howie, B., Leung, H. T., Hartmann, A. M., Moller, H. J., Morris, D. W., Shi, Y. Y., Feng, G. Y., Hoffmann, P., Propping, P., Vasilescu, C., Maier, W., Rietschel, M., Zammit, S., Schumacher, J., Quinn, E. M., Schulze, T. G., Williams, N. M.,

- Giegling, I., Iwata, N., Ikeda, M., Darvasi, A., Shifman, S., He, L., Duan, J., Sanders, A. R., Levinson, D. F., Gejman, P. V., Cichon, S., Nothen, M. M., Gill, M., Corvin, A., Rujescu, D., Kirov, G., Owen, M. J., and Mol Genetics Of Schizophrenia, C. (2008). "Identification of loci associated with schizophrenia by genome-wide association and follow-up." *Nature Genetics*, 40(9), 1053-1055.
- Orlova, K. A., and Crino, P. B. (2010). "The tuberous sclerosis complex." *Year in Neurology* 2, 1184, 87-105.
- Oswald, J., Engemann, S., Lane, N., Mayer, W., Olek, A., Fundele, R., Dean, W., Reik, W., and Walter, J. (2000). "Active demethylation of the paternal genome in the mouse zygote." *Current Biology*, 10(8), 475-478.
- Ott, C. J., Suszko, M., Blackledge, N. P., Wright, J. E., Crawford, G. E., and Harris, A. (2009). "A complex intronic enhancer regulates expression of the CFTR gene by direct interaction with the promoter." *Journal of Cellular and Molecular Medicine*, 13(4), 680-692.
- Ozturk, N., Lee, J. H., Gaddameedhi, S., and Sancar, A. (2009). "Loss of cryptochrome reduces cancer risk in p53 mutant mice." *Proceedings of the National Academy of Sciences of the United States of America*, 106(8), 2841-2846.
- Pagni, M., Ioannidis, V., Cerutti, L., Zahn-Zabal, M., Jongeneel, C. V., and Falquet, L. (2004). "MyHits: a new interactive resource for protein annotation and domain identification." *Nucleic Acids Research*, 32, W332-W335.
- Pagni, M., Ioannidis, V., Cerutti, L., Zahn-Zabal, M., Jongeneel, C. V., Hau, J., Martin, O., Kuznetsov, D., and Falquet, L. (2007). "MyHits: improvements to an interactive resource for analyzing protein sequences." *Nucleic Acids*

*Research*, 35, W433-W437.

Palferman, S., Matthews, N., Turner, M., Moore, J., Hervas, A., Aubin, A., Wallace, S., Michelotti, J., Wainhouse, C., Paul, A., Thompson, E., Gupta, R., Garner, C., Murin, M., Freitag, C., Ryder, N., Cottington, E., Parr, J., Pickles, A., Rutter, M., Bailey, A., Barnby, G., Lamb, J. A., Marlow, A., Scudder, P., Monaco, A. P., Baird, G., Cox, A., Docherty, Z., Warburton, P., Green, E. P., Abbs, S. J., Le Couteur, A., McConachie, H. R., Berney, T., Kelly, T. P., De Vries, P. J., Bolton, P. F., Green, J., Gilchrist, A., Whittacker, J., Bolton, B., Packer, R., Maestrini, E., Blasi, F., Van Engeland, H., De Jonge, M. V., Kemner, C., Klauck, S. M., Beyer, K. S., Epp, S., Poustka, A., Benner, A., Goethe, J. W., Poustka, F., Ruhl, D., Schmotzer, G., Boolte, S., Feineis-Matthews, S., Fombonne, E., Rogee, B., Fremolle-Kruck, J., Pienkowski, C., Tauber, M. T., Pedersen, L., Nielsen, K. B., Eriksen, G., Haracopos, D., Cotterill, R. M. J., Tsiantis, J., Papanikolaou, K., Lord, C., Corsello, C., Guter, S., Leventhal, B., Cook, E., Smalley, S. L., Bailey, J., McGough, J., Levitt, J., Pauls, D., Volkmar, F., and Weeks, D. E. (2001). "A genomewide screen for autism: Strong evidence for linkage to chromosomes 2q, 7q, and 16p." *American Journal of Human Genetics*, 69(3), 570-581.

Panda, S., Antoch, M. P., Miller, B. H., Su, A. I., Schook, A. B., Straume, M., Schultz, P. G., Kay, S. A., Takahashi, J. S., and Hogenesch, J. B. (2002). "Coordinated transcription of key pathways in the mouse by the circadian clock." *Cell*, 109(3), 307-320.

Peppe, S., McCann, J., Gibbon, F., O'Hare, A., and Rutherford, M. (2007). "Receptive and expressive prosodic ability in children with high-functioning autism." *Journal of Speech Language and Hearing Research*, 50(4), 1015-

1028.

- Perkins, D. O., Jeffries, C. D., Jarskog, L. F., Thomson, J. M., Woods, K., Newman, M. A., Parker, J. S., Jin, J. P., and Hammond, S. M. (2007). "microRNA expression in the prefrontal cortex of individuals with schizophrenia and schizoaffective disorder." *Genome Biology*, 8(2).
- Petit, E., Herault, J., Martineau, J., Perrot, A., Barthelemy, C., Hameury, L., Sauvage, D., Lelord, G., and Muh, J. P. (1995). "Association study with 2 markers of a human homeogene in infantile-autism." *Journal of Medical Genetics*, 32(4), 269-274.
- Phelps, E. A. (2004). "Human emotion and memory: interactions of the amygdala and hippocampal complex." *Current Opinion in Neurobiology*, 14(2), 198-202.
- Philippe, A., Martinez, M., Guilloud-Bataille, M., Gillberg, C., Rastam, M., Sponheim, E., Coleman, M., Zappella, M., Aschauer, H., van Malldergerme, L., Penet, C., Feingold, J., Brice, A., and Leboyer, M. (1999). "Genome-wide scan for autism susceptibility genes." *Human Molecular Genetics*, 8(5), 805-812.
- Pickles, A., Bolton, P., Macdonald, H., Bailey, A., Lecouteur, A., Sim, C. H., and Rutter, M. (1995). "Latent-class analysis of recurrence risks for complex phenotypes with selection and measurement error - a twin and family history study of autism." *American Journal of Human Genetics*, 57(3), 717-726.
- Pinto, D., Pagnamenta, A. T., Klei, L., Anney, R., Merico, D., Regan, R., Conroy, J., Magalhaes, T. R., Correia, C., Abrahams, B. S., Almeida, J., Bacchelli, E., Bader, G. D., Bailey, A. J., Baird, G., Battaglia, A., Berney, T., Bolshakova, N., Bolte, S., Bolton, P. F., Bourgeron, T., Brennan, S., Brian,

J., Bryson, S. E., Carson, A. R., Casallo, G., Casey, J., Chung, B. H. Y., Cochrane, L., Corsello, C., Crawford, E. L., Crossett, A., Cytrynbaum, C., Dawson, G., de Jonge, M., Delorme, R., Drmic, I., Duketis, E., Duque, F., Estes, A., Farrar, P., Fernandez, B. A., Folstein, S. E., Fombonne, E., Freitag, C. M., Gilbert, J., Gillberg, C., Glessner, J. T., Goldberg, J., Green, A., Green, J., Guter, S. J., Hakonarson, H., Heron, E. A., Hill, M., Holt, R., Howe, J. L., Hughes, G., Hus, V., Iglizzi, R., Kim, C., Klauck, S. M., Kolevzon, A., Korvatska, O., Kustanovich, V., Lajonchere, C. M., Lamb, J. A., Laskawiec, M., Leboyer, M., Le Couteur, A., Leventhal, B. L., Lionel, A. C., Liu, X. Q., Lord, C., Lotspeich, L., Lund, S. C., Maestrini, E., Mahoney, W., Mantoulan, C., Marshall, C. R., McConachie, H., McDougle, C. J., McGrath, J., McMahon, W. M., Merikangas, A., Migita, O., Minshew, N. J., Mirza, G. K., Munson, J., Nelson, S. F., Noakes, C., Noor, A., Nygren, G., Oliveira, G., Papanikolaou, K., Parr, J. R., Parrini, B., Paton, T., Pickles, A., Pilorge, M., et al. (2010). "Functional impact of global rare copy number variation in autism spectrum disorders." *Nature*, 466(7304), 368-372.

Piotrowska, K., Wianny, F., Pedersen, R. A., and Zernicka-Goetz, M. (2001).

"Blastomeres arising from the first cleavage division have distinguishable fates in normal mouse development." *Development*, 128(19), 3739-3748.

Piotrowska, K., and Zernicka-Goetz, M. (2001). "Role for sperm in spatial

patterning of the early mouse embryo." *Nature*, 409(6819), 517-521.

Piven, J., Bailey, J., Ranson, B. J., and Arndt, S. (1997). "An MRT study of the corpus callosum in autism." *American Journal of Psychiatry*, 154(8), 1051-1056.

- Potocki, L., Bi, W. M., Treadwell-Deering, D., Carvalho, C. M. B., Eifert, A., Friedman, E. M., Glaze, D., Krull, K., Lee, J. A., Lewis, R. A., Mendoza-Londono, R., Robbins-Furman, P., Shaw, C., Shi, X., Weissenberger, G., Withers, M., Yatsenko, S. A., Zackai, E. H., Stankiewicz, P., and Lupski, J. R. (2007). "Characterization of Potocki-Lupski syndrome (dup(17)(p11.2p11.2)) and delineation of a dosage-sensitive critical interval that can convey an autism phenotype." *American Journal of Human Genetics*, 80(4), 633-649.
- Preitner, N., Damiola, F., Molina, L. L., Zakany, J., Duboule, D., Albrecht, U., and Schibler, U. (2002). "The orphan nuclear receptor REV-ERB alpha controls circadian transcription within the positive limb of the mammalian circadian oscillator." *Cell*, 110(2), 251-260.
- Puppe, J., Drost, R., Liu, X. L., Joosse, S. A., Evers, B., Cornelissen-Steijger, P., Nederlof, P., Yu, Q., Jonkers, J., van Lohuizen, M., and Pietersen, A. M. (2009). "BRCA1-deficient mammary tumor cells are dependent on EZH2 expression and sensitive to Polycomb Repressive Complex 2-inhibitor 3-deazaneplanocin A." *Breast Cancer Research*, 11(4).
- Puri, V., McQuillin, A., Datta, S., Choudhury, K., Pimm, J., Thirumalai, S., Krasucki, R., Lawrence, J., Quested, D., Bass, N., Crombie, C., Fraser, G., Walker, N., Moorey, H., Ray, M. K., Sule, A., Curtis, D., Clair, D. S., and Gurling, H. (2008). "Confirmation of the genetic association between the U2AF homology motif (UHM) kinase 1 (UHMK1) gene and schizophrenia on chromosome 1q23.3." *European Journal of Human Genetics*, 16(10), 1275-1282.
- Qin, J., Jia, M. X., Wang, L. F., Lu, T. L., Ruan, Y., Liu, J., Guo, Y. Q., Zhang, J. S., Yang, X. L., Yue, W. H., and Zhang, D. (2009). "Association study of



SHANK3 gene polymorphisms with autism in Chinese Han population."

*Bmc Medical Genetics*, 10.

Qin, S., and Parthun, M. R. (2002). "Histone H3 and the histone acetyltransferase

Hat1p contribute to DNA double-strand break repair." *Molecular and Cellular Biology*, 22(23), 8353-8365.

Qin, S., and Parthun, M. R. (2006). "Recruitment of the type B histone

acetyltransferase Hat1p to chromatin is linked to DNA double-strand breaks." *Molecular and Cellular Biology*, 26(9), 3649-3658.

Ramocki, M. B., and Zoghbi, H. Y. (2008). "Failure of neuronal homeostasis

results in common neuropsychiatric phenotypes." *Nature*, 455(7215), 912-918.

Rankin, L., and Belz, G. T. (2011). "Diverse Roles of Inhibitor of Differentiation 2

in Adaptive Immunity." *Clinical & Developmental Immunology*.

Raven, J. (1996). "Progressive Matrices: A Perceptual Test of Intelligence.

Individual Form." Oxford: Oxford Psychologists Press Ltd.

Raymond, G. V., Bauman, M. L., and Kemper, T. L. (1996). "Hippocampus in

autism: A Golgi analysis." *Acta Neuropathologica*, 91(1), 117-119.

Reick, M., Garcia, J. A., Dudley, C., and McKnight, S. L. (2001). "NPAS2: An

analog of clock operative in the mammalian forebrain." *Science*, 293(5529), 506-509.

Reinhart, B. J., Weinstein, E. G., Rhoades, M. W., Bartel, B., and Bartel, D. P.

(2002). "MicroRNAs in plants." *Genes & Development*, 16(13), 1616-1626.

Reppert, S. M., and Weaver, D. R. (2002). "Coordination of circadian timing in

mammals." *Nature*, 418(6901), 935-941.

Richdale, A. L., and Prior, M. R. (1992). "Urinary cortisol circadian-rhythm in a

group of high-functioning children with autism." *Journal of Autism and*

*Developmental Disorders*, 22(3), 433-447.

Richdale, A. L., and Prior, M. R. (1995). "The sleep-wake rhythm in children with autism." *European Child & Adolescent Psychiatry*, 4(3), 175-186.

Riley, B., Thiselton, D., Maher, B. S., Bigdeli, T., Wormley, B., McMichael, G. O., Fanous, A. H., Vladimirov, V., O'Neill, F. A., Walsh, D., and Kendler, K. S. (2010). "Replication of association between schizophrenia and ZNF804A in the Irish Case-Control Study of Schizophrenia sample." *Molecular Psychiatry*, 15(1), 29-37.

Ripperger, J. R. A., and Schibler, U. (2006). "Rhythmic CLOCK-BMAL1 binding to multiple E-box motifs drives circadian Dbp transcription and chromatin transitions." *Nature Genetics*, 38(3), 369-374.

Ritchie, M. G., Halsey, E. J., and Gleason, J. M. (1999). "Drosophila song as a species-specific mating signal and the behavioural importance of Kyriacou & Hall cycles in *D-melanogaster* song." *Animal Behaviour*, 58, 649-657.

Ritvo, E. R., Freeman, B. J., Scheibel, A. B., Duong, T., Robinson, H., Guthrie, D., and Ritvo, A. (1986). "Lower purkinje-cell counts in the cerebella of 4 autistic subjects - initial findings of the UCLA-NSAC autopsy research report." *American Journal of Psychiatry*, 143(7), 862-866.

Riva, A., and Kohane, I. S. (2002). "SNPper: retrieval and analysis of human SNPs." *Bioinformatics*, 18(12), 1681-1685.

Rivas, E., and Eddy, S. R. (2000). "Secondary structure alone is generally not statistically significant for the detection of noncoding RNAs." *Bioinformatics*, 16(7), 583-605.

Robnett, C. C., Vaughan, A. G., Knapp, J. M., and Baker, B. S. (2010). "Sex and the single cell. ii. There is a time and place for sex." *Plos Biology*, 8(5).

- Robles, M. S., Boyault, C., Knutti, D., Padmanabhan, K., and Weitz, C. J. (2010). "Identification of RACK1 and Protein Kinase C alpha as Integral Components of the Mammalian Circadian Clock." *Science*, 327(5964), 463-466.
- Rodriguez, A., Griffiths-Jones, S., Ashurst, J. L., and Bradley, A. (2004). "Identification of mammalian microRNA host genes and transcription units." *Genome Research*, 14(10A), 1902-1910.
- Rose, G. J., and Capranica, R. R. (1984). "Processing amplitude-modulated sounds by the auditory midbrain of 2 species of toads - matched temporal filters." *Journal of Comparative Physiology*, 154(2), 211-219.
- Ruby, J. G., Jan, C. H., and Bartel, D. P. (2007). "Intronic microRNA precursors that bypass Drosha processing." *Nature*, 448, 83-U7.
- Ruiz, S., Ferreiro, M. J., Casanova, G., Olivera, A., and Cantera, R. (2010). "Synaptic vesicles in motor synapses change size and distribution during the day." *Synapse*, 64(1), 14-19.
- Rutila, J. E., Suri, V., Le, M., So, W. V., Rosbash, M., and Hall, J. C. (1998). "CYCLE is a second bHLH-PAS clock protein essential for circadian rhythmicity and transcription of *Drosophila* period and timeless." *Cell*, 93(5), 805-814.
- Rutter, J., Reick, M., Wu, L. C., and McKnight, S. L. (2001). "Regulation of clock and NPAS2 DNA binding by the redox state of NAD cofactors." *Science*, 293(5529), 510-514.
- Ryner, L. C., Goodwin, S. F., Castrillon, D. H., Anand, A., Villella, A., Baker, B. S., Hall, J. C., Taylor, B. J., and Wasserman, S. A. (1996). "Control of male sexual behavior and sexual orientation in *Drosophila* by the fruitless gene." *Cell*, 87(6), 1079-1089.

- Rzhetsky, A., Wajngurt, D., Park, N., and Zheng, T. (2007). "Probing genetic overlap among complex human phenotypes." *Proceedings of the National Academy of Sciences of the United States of America*, 104(28), 11694-11699.
- Saini, H. K., Griffiths-Jones, S., and Enright, A. J. (2007). "Genomic analysis of human microRNA transcripts." *Proceedings of the National Academy of Sciences of the United States of America*, 104, 17719-17724.
- Sakai, T., and Ishida, N. (2001). "Circadian rhythms of female mating activity governed by clock genes in *Drosophila*." *Proceedings of the National Academy of Sciences of the United States of America*, 98(16), 9221-9225.
- Sakai, T., Tamura, T., Kitamoto, T., and Kidokoro, Y. (2004). "A clock gene, period, plays a key role in long-term memory formation in *Drosophila*." *Proceedings of the National Academy of Sciences of the United States of America*, 101(45), 16058-16063.
- Salmond, C. H., Ashburner, J., Connelly, A., Friston, K. J., Gadian, D. G., and Vargha-Khadem, F. (2005). "The role of the medial temporal lobe in autistic spectrum disorders." *European Journal of Neuroscience*, 22(3), 764-772.
- Salz, H. K. (2011). "Sex determination in insects: a binary decision based on alternative splicing." *Current Opinion in Genetics & Development*, 21(4), 395-400.
- Samuels, I. S., Karlo, J. C., Faruzzi, A. N., Pickering, K., Herrup, K., Sweatt, J. D., Saitta, S. C., and Landreth, G. E. (2008). "Deletion of ERK2 mitogen-activated protein kinase identifies its key roles in cortical neurogenesis and cognitive function." *Journal of Neuroscience*, 28(27), 6983-6995.

- Sancar, A. (1996). "DNA excision repair." *Annual Review of Biochemistry*, 65, 43-81.
- Sancar, A., Lindsey-Boltz, L. A., Kang, T. H., Reardon, J. T., Lee, J. H., and Ozturk, N. (2010). "Circadian clock control of the cellular response to DNA damage." *Febs Letters*, 584(12), 2618-2625.
- Sancar, A., Lindsey-Boltz, L. A., Unsal-Kacmaz, K., and Linn, S. (2004). "Molecular mechanisms of mammalian DNA repair and the DNA damage checkpoints." *Annual Review of Biochemistry*, 73, 39-85.
- Sanchez, S. E., Petrillo, E., Beckwith, E. J., Zhang, X., Rugnone, M. L., Hernando, C. E., Cuevas, J. C., Herz, M. A. G., Depetris-Chauvin, A., Simpson, C. G., Brown, J. W. S., Cerdan, P. D., Borevitz, J. O., Mas, P., Fernanda Ceriani, M., Kornblihtt, A. R., and Yanovsky, M. J. (2010). "A methyl transferase links the circadian clock to the regulation of alternative splicing." *Nature*, 468(7320), 112-+.
- Santangelo, S. L., and Tsatsanis, K. (2005). "What is known about autism - Genes, brain, and behavior." *American Journal of Pharmacogenomics*, 5(2), 71-92.
- Sarachana, T., Zhou, R., Chen, G., Manji, H. K., and Hu, V. W. (2010). "Investigation of post-transcriptional gene regulatory networks associated with autism spectrum disorders by microRNA expression profiling of lymphoblastoid cell lines." *Genome Med*, 2(4), 23.
- Sato, T. K., Panda, S., Miraglia, L. J., Reyes, T. M., Rudic, R. D., McNamara, P., Naik, K. A., Fitzgerald, G. A., Kay, S. A., and Hogenesch, J. B. (2004). "A functional genomics strategy reveals rora as a component of the mammalian circadian clock." *Neuron*, 43(4), 527-537.

- Saunders, M. A., Liang, H., and Li, W. H. (2007). "Human polymorphism at microRNAs and microRNA target sites." *Proceedings of the National Academy of Sciences of the United States of America*, 104, 3300-3305.
- Sawyer, L. A., Hennessy, J. M., Peixoto, A. A., Rosato, E., Parkinson, H., Costa, R., and Kyriacou, C. P. (1997). "Natural variation in a *Drosophila* clock gene and temperature compensation." *Science*, 278(5346), 2117-2120.
- Schaefer, A., O'Carroll, D., Tan, C. L., Hillman, D., Sugimori, M., Llinas, R., and Greengard, P. (2007). "Cerebellar neuro degeneration in the absence of microRNAs." *Journal of Experimental Medicine*, 204(7), 1553-1558.
- Schildberger, K. (1984). "Temporal selectivity of identified auditory neurons in the Cricket brain." *Journal of Comparative Physiology*, 155(2), 171-185.
- Schmitz, C., Martineau, J., Barthelemy, C., and Assaiante, C. (2003). "Motor control and children with autism: deficit of anticipatory function?" *Neuroscience Letters*, 348(1), 17-20.
- Schmitzova, J., Klaudiny, J., Albert, S., Schroder, W., Schreckengost, W., Hanes, J., Judova, J., and Simuth, J. (1998). "A family of major royal jelly proteins of the honeybee *Apis mellifera* L." *Cellular and Molecular Life Sciences*, 54(9), 1020-1030.
- Schneegans, T., Borgmeyer, U., Hentschke, M., Gronostajski, R. M., Schachner, M., and Tilling, T. (2009). "Nuclear factor I-A represses expression of the cell adhesion molecule L1." *Bmc Molecular Biology*, 10.
- Schneider, T., and Przewlocki, R. (2005). "Behavioral alterations in rats prenatally exposed to valproic acid: Animal model of autism." *Neuropsychopharmacology*, 30(1), 80-89.
- Scott, J. A., Schumann, C. M., Goodlin-Jones, B. L., and Amaral, D. G. (2009). "A comprehensive volumetric analysis of the cerebellum in children and

adolescents with autism spectrum disorder." *Autism Research*, 2(5), 246-257.

Sears, L. L., Finn, P. R., and Steinmetz, J. E. (1994). "Abnormal classical eye-blink conditioning in autism." *Journal of Autism and Developmental Disorders*, 24(6), 737-751.

Segawa, M., Shinomiya, N., Hosaka, S., and Nomura, Y. (1981). "Disturbance of sleep-wakefulness rhythm in early infantile-autism." *Folia Psychiatrica Et Neurologica Japonica*, 35(3), 400-401.

Serajee, F. J., Nabi, R., Zhong, H., and Mahbubul Huq, A. H. M. (2003). "Association of INPP1, PIK3CG, and TSC2 gene variants with autistic disorder: implications for phosphatidylinositol signalling in autism." *Journal of medical genetics*, 40(11), e119.

Sgaiier, S. K., Lao, Z. M., Villanueva, M. P., Berenshteyn, F., Stephen, D., Turnbull, R. K., and Joyner, A. L. (2007). "Genetic subdivision of the tectum and cerebellum into functionally related regions based on differential sensitivity to engrailed proteins." *Development*, 134(12), 2325-2335.

Shao, Y. J., Raiford, K. L., Wolpert, C. M., Cope, H. A., Ravan, S. A., Ashley-Koch, A. A., Abramson, R. K., Wright, H. H., DeLong, R. G., Gilbert, J. R., Cuccaro, M. L., and Pericak-Vance, M. A. (2002). "Phenotypic homogeneity provides increased support for linkage on chromosome 2 in autistic disorder." *American Journal of Human Genetics*, 70(4), 1058-1061.

Shav-Tal, Y., and Zipori, D. (2002). "PSF and p54(nrb)/NonO - multi-functional nuclear proteins." *Febs Letters*, 531(2), 109-114.

Shearman, L. P., Jin, X. W., Lee, C. G., Reppert, S. M., and Weaver, D. R. (2000). "Targeted disruption of the mPer3 gene: Subtle effects on

circadian clock function." *Molecular and Cellular Biology*, 20(17), 6269-6275.

Shen, H. M., Lin, Y., Choksi, S., Tran, J., Jin, T., Chang, L. F., Karin, M., Zhang, J. K., and Liu, Z. G. (2004). "Essential roles of receptor-interacting protein and TRAF2 in oxidative stress-induced cell death." *Molecular and Cellular Biology*, 24(13), 5914-5922.

Shi, J., Gershon, E. S., and Liu, C. (2008). "Genetic associations with schizophrenia: Meta-analyses of 12 candidate genes." *Schizophrenia Research*, 104(1-3), 96-107.

Shimano, H., Yahagi, N., Amemiya-Kudo, M., Hasty, A. H., Osuga, J., Tamura, Y., Shionoiri, F., Iizuka, Y., Ohashi, K., Harada, K., Gotoda, T., Ishibashi, S., and Yamada, N. (1999). "Sterol regulatory element-binding protein-1 as a key transcription factor for nutritional induction of lipogenic enzyme genes." *Journal of Biological Chemistry*, 274(50), 35832-35839.

Shoubridge, C., Fullston, T., and Gecz, J. (2010). "ARX Spectrum Disorders: Making Inroads into the Molecular Pathology." *Human Mutation*, 31(8), 889-900.

Skuse, D. H. (1999). "Genomic imprinting of the X chromosome: A novel mechanism for the evolution of sexual dimorphism." *Journal of Laboratory and Clinical Medicine*, 133(1), 23-32.

Skuse, D. H., James, R. S., Bishop, D. V. M., Coppin, B., Dalton, P., AamodtLeeper, G., BacareseHamilton, M., Creswell, C., McGurk, R., and Jacobs, P. A. (1997). "Evidence from Turner's syndrome of an imprinted X-linked locus affecting cognitive function." *Nature*, 387(6634), 705-708.

Slater, G. S., and Birney, E. (2005). "Automated generation of heuristics for biological sequence comparison." *Bmc Bioinformatics*, 6.



- Smyk, M., Obersztyn, E., Nowakowska, B., Nawara, M., Cheung, S. W., Mazurczak, T., Stankiewicz, P., and Bocian, E. (2008). "Different-sized duplications of Xq28, including MECP2, in three males with mental retardation, absent or delayed speech, and recurrent infections." *American Journal of Medical Genetics Part B-Neuropsychiatric Genetics*, 147B(6), 799-806.
- So, W. V., Sarov-Blat, L., Kotarski, C. K., McDonald, M. J., Allada, R., and Rosbash, M. (2000). "Takeout, a novel *Drosophila* gene under circadian clock transcriptional regulation." *Molecular and Cellular Biology*, 20(18), 6935-6944.
- Song, D. L., and Joyner, A. L. (2000). "Two Pax2/5/8-binding sites in Engrailed2 are required for proper initiation of endogenous mid-hindbrain expression." *Mechanisms of Development*, 90(2), 155-165.
- Sorbello, V., Fusco, L., Sfiligoi, C., Scafoglio, C., Ponzzone, R., Biglia, N., Weisz, A., Sismondi, P., and De Bortoli, M. (2003). "Quantitative real-time RT-PCR analysis of eight novel estrogen-regulated genes in breast cancer." *International Journal of Biological Markers*, 18(2), 123-129.
- Sparrow, S.S., Balla, D. A., and Cicchetti, D. V. (1984). "The Vineland Adaptive Behavior Scale." Circle Pines, Minnesota: American Guidance Service.
- Stark, A., Kheradpour, P., Parts, L., Brennecke, J., Hodges, E., Hannon, G. J., and Kellis, M. (2007a). "Systematic discovery and characterization of fly microRNAs using 12 *Drosophila* genomes." *Genome Research*, 17, 1865-1879.
- Stark, A., Lin, M. F., Kheradpour, P., Pedersen, J. S., Parts, L., Carlson, J. W., Crosby, M. A., Rasmussen, M. D., Roy, S., Deoras, A. N., Ruby, J. G.,

Brennecke, J., Hodges, E., Hinrichs, A. S., Caspi, A., Park, S. W., Han, M. V., Maeder, M. L., Polansky, B. J., Robson, B. E., Aerts, S., van Helden, J., Hassan, B., Gilbert, D. G., Eastman, D. A., Rice, M., Weir, M., Hahn, M. W., Park, Y., Dewey, C. N., Pachter, L., Kent, W. J., Haussler, D., Lai, E. C., Bartel, D. P., Hannon, G. J., Kaufman, T. C., Eisen, M. B., Clark, A. G., Smith, D., Celniker, S. E., Gelbart, W. M., and Kellis, M. (2007b).

"Discovery of functional elements in 12 *Drosophila* genomes using evolutionary signatures." *Nature*, 450, 219-232.

Stickgold, R. (2005). "Sleep-dependent memory consolidation." *Nature*, 437(7063), 1272-1278.

Stokkan, K. A., Yamazaki, S., Tei, H., Sakaki, Y., and Menaker, M. (2001). "Entrainment of the circadian clock in the liver by feeding." *Science*, 291(5503), 490-493.

Subauste, J. S., Katz, R. W., and Koenig, R. J. (1994). "DNA-binding specificity and function of Retinoid-X Receptor-alpha." *Journal of Biological Chemistry*, 269(48), 30232-30237.

Sun, G., Yan, J., Noltner, K., Feng, J., Li, H., Sarkis, D. A., Sommer, S. S., and Rossi, J. J. (2009). "SNPs in human miRNA genes affect biogenesis and function." *Rna-a Publication of the Rna Society*, 15(9), 1640-1651.

Sun, M., Hurst, L. D., Carmichael, G. G., and Chen, J. J. (2005). "Evidence for a preferential targeting of 3'-UTRs by cis-encoded natural antisense transcripts." *Nucleic Acids Research*, 33(17), 5533-5543.

Sun, Z. S., Albrecht, U., Zhuchenko, O., Bailey, J., Eichele, G., and Lee, C. C. (1997). "RIGUI, a putative mammalian ortholog of the *Drosophila* period gene." *Cell*, 90(6), 1003-1011.

- Sykes, N. H., Toma, C., Wilson, N., Volpi, E. V., Sousa, I., Pagnamenta, A. T., Tancredi, R., Battaglia, A., Maestrini, E., Bailey, A. J., Monaco, A. P., and Imgsac. (2009). "Copy number variation and association analysis of SHANK3 as a candidate gene for autism in the IMGSA collection." *European Journal of Human Genetics*, 17(10), 1347-1353.
- Szelag, E., Kowalska, J., Galkowski, T., and Poppel, E. (2004). "Temporal processing deficits in high-functioning children with autism." *British Journal of Psychology*, 95, 269-282.
- Takano, A., Uchiyama, M., Kajimura, N., Mishima, K., Inoue, Y., Kamei, Y., Kitajima, T., Shibui, K., Katoh, M., Watanabe, T., Hashimoto, Y., Nakajima, T., Ozeki, Y., Hori, T., Yamada, N., Toyoshima, R., Ozaki, N., Okawa, M., Nagai, K., Takahashi, K., Isojima, Y., Yamauchi, T., and Ebisawa, T. (2004). "A missense variation in human casein kinase I epsilon gene that induces functional alteration and shows an inverse association with circadian rhythm sleep disorders." *Neuropsychopharmacology*, 29(10), 1901-1909.
- Talebizadeh, Z., Butler, M. G., and Theodoro, M. F. (2008). "Feasibility and Relevance of Examining Lymphoblastoid Cell Lines to Study Role of microRNAs in Autism." *Autism Research*, 1(4), 240-250.
- Tantam, D., Holmes, D., and Cordess, C. (1993). "NONVERBAL EXPRESSION IN AUTISM OF ASPERGER TYPE." *Journal of Autism and Developmental Disorders*, 23(1), 111-133.
- Taunton, J., Hassig, C. A., and Schreiber, S. L. (1996). "A mammalian histone deacetylase related to the yeast transcriptional regulator Rpd3p." *Science*, 272(5260), 408-411.

- Tierney, E., Bukelis, I., Thompson, R. E., Ahmed, K., Aneja, A., Kratz, L., and Kelley, R. I. (2006). "Abnormalities of cholesterol metabolism in autism spectrum disorders." *American Journal of Medical Genetics Part B-Neuropsychiatric Genetics*, 141B(6), 666-668.
- Tierney, E., Nwokoro, N. A., Porter, F. D., Freund, L. S., Ghuman, J. K., and Kelley, R. I. (2001). "Behavior phenotype in the RSH/Smith-Lemli-Opitz syndrome." *American Journal of Medical Genetics*, 98(2), 191-200.
- Toh, K. L., Jones, C. R., He, Y., Eide, E. J., Hinz, W. A., Virshup, D. M., Ptacek, L. J., and Fu, Y. H. (2001). "An hPer2 phosphorylation site mutation in familiar advanced sleep phase syndrome." *Science*, 291(5506), 1040-1043.
- Toichi, M., and Kamio, Y. (2002). "Long-term memory and levels-of-processing in autism." *Neuropsychologia*, 40(7), 964-969.
- Toichi, M., and Kamio, Y. (2003). "Long-term memory in high-functioning autism: Controversy on episodic memory in autism reconsidered." *Journal of Autism and Developmental Disorders*, 33(2), 151-161.
- Tokuhiro, S., Yamada, R., Chang, X. T., Suzuki, A., Kochi, Y., Sawada, T., Suzuki, M., Nagasaki, M., Ohtsuki, M., Ono, M., Furukawa, H., Nagashima, M., Yoshino, S., Mabuchi, A., Sekine, A., Saito, S., Takahashi, A., Tsunoda, T., Nakamura, Y., and Yamamoto, K. (2003). "An intronic SNP in a RUNX1 binding site of SLC22A4, encoding an organic cation transporter, is associated with rheumatoid arthritis." *Nature Genetics*, 35(4), 341-348.
- Tordjman, S., Anderson, G. M., Pichard, N., Charbuy, H., and Touitou, Y. (2005). "Nocturnal excretion of 6-sulphatoxymelatonin in children and adolescents with autistic disorder." *Biological Psychiatry*, 57(2), 134-138.

- Tordjman, S., Gutknecht, L., Carlier, M., Spitz, E., Antoine, C., Slama, F., Carsalade, V., Cohen, D. J., Ferrari, P., Roubertoux, P. L., and Anderson, G. M. (2001). "Role of the serotonin transporter gene in the behavioral expression of autism." *Molecular Psychiatry*, 6(4), 434-439.
- Torres-Padilla, M. E., Parfitt, D. E., Kouzarides, T., and Zernicka-Goetz, M. (2007). "Histone arginine methylation regulates pluripotency in the early mouse embryo." *Nature*, 445(7124), 214-218.
- Townsend, J., Courchesne, E., Covington, J., Westerfield, M., Harris, N. S., Lyden, P., Lowry, T. P., and Press, G. A. (1999). "Spatial attention deficits in patients with acquired or developmental cerebellar abnormality." *Journal of Neuroscience*, 19(13), 5632-5643.
- Townsend, J., Westerfield, M., Leaver, E., Makeig, S., Jung, T. P., Pierce, K., and Courchesne, E. (2001). "Event-related brain response abnormalities in autism: evidence for impaired cerebello-frontal spatial attention networks." *Cognitive Brain Research*, 11(1), 127-145.
- Trevarthen, C. (2000). "Autism as a neurodevelopmental disorder affecting communication and learning in early childhood: prenatal origins, post natal course and effective educational support." *Prostaglandins Leukotrienes and Essential Fatty Acids*, 63(1-2), 41-46.
- Trevarthen, C., and Aitken, K. J. (2001). "Infant intersubjectivity: Research, theory, and clinical applications." *Journal of Child Psychology and Psychiatry*, 42(1), 3-48.
- Trevarthen, C., and Daniel, S. (2005). "Disorganized rhythm and synchrony: Early signs of autism and Rett syndrome." *Brain & Development*, 27, S25-S34.

- Trikalinos, T. A., Karvouni, A., Zintzaras, E., Ylisaukko-oja, T., Peltonen, L., Jarvela, I., and Ioannidis, J. P. A. (2006). "A heterogeneity-based genome search meta-analysis for autism-spectrum disorders." *Molecular Psychiatry*, 11(1), 29-36.
- Trivers, R. L. (1974). "Parent-offspring conflict." *American Zoologist*, 14(1), 249-264.
- Tsankova, N., Renthal, W., Kumar, A., and Nestler, E. J. (2007). "Epigenetic regulation in psychiatric disorders." *Nature Reviews Neuroscience*, 8(5), 355-367.
- Tsujino, N., Nakatani, Y., Seki, Y., Nakasato, A., Nakamura, M., Sugawara, M., and Arita, H. (2007). "Abnormality of circadian rhythm accompanied by an increase in frontal cortex serotonin in animal model of autism." *Neuroscience Research*, 57(2), 289-295.
- Tudor, M., Akbarian, S., Chen, R. Z., and Jaenisch, R. (2002). "Transcriptional profiling of a mouse model for Rett syndrome reveals subtle transcriptional changes in the brain." *Proceedings of the National Academy of Sciences of the United States of America*, 99(24), 15536-15541.
- Ubeda, F., and Gardner, A. (2010). "A model for genomic imprinting in the social brain: Juveniles." *Evolution*, 64(9), 2587-2600.
- Uchida, Y., Hirayama, J., and Nishina, H. (2010). "A common origin: Signaling similarities in the regulation of the circadian clock and DNA damage responses." *Biological & Pharmaceutical Bulletin*, 33(4), 535-544.
- Ueda, H. R., Chen, W. B., Adachi, A., Wakamatsu, H., Hayashi, S., Takasugi, T., Nagano, M., Nakahama, K., Suzuki, Y., Sugano, S., Iino, M., Shigeyoshi, Y., and Hashimoto, S. (2002). "A transcription factor response element for gene expression during circadian night." *Nature*, 418(6897), 534-539.

- Ueda, H. R., Hayashi, S., Chen, W. B., Sano, M., Machida, M., Shigeyoshi, Y., Iino, M., and Hashimoto, S. (2005). "System-level identification of transcriptional circuits underlying mammalian circadian clocks." *Nature Genetics*, 37(2), 187-192.
- Unsal-Kacmaz, K., Mullen, T. E., Kaufmann, W. K., and Sancar, A. (2005). "Coupling of human circadian and cell cycles by the timeless protein." *Molecular and Cellular Biology*, 25(8), 3109-3116.
- Utge, S. J., Soronen, P., Loukola, A., Kronholm, E., Ollila, H. M., Pirkola, S., Porkka-Heiskanen, T., Partonen, T., and Paunio, T. (2010). "Systematic analysis of circadian genes in a population-based sample reveals association of TIMELESS with depression and sleep disturbance." *Plos One*, 5(2).
- Van Esch, H., Bauters, M., Ignatius, J., Jansen, M., Raynaud, M., Hollanders, K., Lugtenberg, D., Bienvenu, T., Jensen, L. R., Gecz, J., Moraine, C., Marynen, P., Fryns, J. P., and Froyen, G. (2005). "Duplication of the MECP2 region is a frequent cause of severe mental retardation and progressive neurological symptoms in males." *American Journal of Human Genetics*, 77(3), 442-453.
- Verkerk, A., Pieretti, M., Sutcliffe, J. S., Fu, Y. H., Kuhl, D. P. A., Pizzuti, A., Reiner, O., Richards, S., Victoria, M. F., Zhang, F. P., Eussen, B. E., Vanommen, G. J. B., Blondin, L. A. J., Riggins, G. J., Chastain, J. L., Kunst, C. B., Galjaard, H., Caskey, C. T., Nelson, D. L., Oostra, B. A., and Warren, S. T. (1991). "Identification of a gene (FMR-1) Containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile-x syndrome." *Cell*, 65(5), 905-914.

Verreault, A., Kaufman, P. D., Kobayashi, R., and Stillman, B. (1997).

"Nucleosomal DNA regulates the core-histone-binding subunit of the human Hat1 acetyltransferase." *Current Biology*, 8(2), 96-108.

Villella, A., Ferri, S. L., Krystal, J. D., and Hall, J. C. (2005). "Functional analysis of fruitless gene expression by transgenic manipulations of *Drosophila* courtship." *Proceedings of the National Academy of Sciences of the United States of America*, 102(46), 16550-16557.

Vitzthum, V. J., Worthman, C. M., Beall, C. M., Thornburg, J., Vargas, E., Villena, M., Soria, R., Caceres, E., and Spielvogel, H. (2009). "Seasonal and circadian variation in salivary testosterone in rural bolivian men." *American Journal of Human Biology*, 21(6), 762-768.

Vorstman, J. A. S., Morcus, M. E. J., Duijff, S. N., Klaassen, P. W. J., Heineman-de Boer, J. A., Beemer, F. A., Swaab, H., Kahn, R. S., and van Engeland, H. (2006). "The 22q11.2 deletion in children: High rate of autistic disorders and early onset of psychotic symptoms." *Journal of the American Academy of Child and Adolescent Psychiatry*, 45(9), 1104-1113.

Waddington, C. H. (1953). "Epigenetics and evolution." *Symposia of the Society for Experimental Biology*, 7, 186-199.

Wade, P. A., Geggion, A., Jones, P. L., Ballestar, E., Aubry, F., and Wolffe, A. P. (1999). "Mi-2 complex couples DNA methylation to chromatin remodelling and histone deacetylation." *Nature Genetics*, 23(1), 62-66.

Wall, D. P., Pivovarov, R., Tong, M., Jung J. Y., Fusaro V. A., DeLuca T. F., Tonellato P. J. 2010. "Genotator: a disease-agnostic tool for genetic annotation of disease". *BMC Medical Genomics*, (3), 50.

Wang, H., Liu, F., Millette, C. F., and Kilpatrick, D. L. (2002). "Expression of a novel, sterol-insensitive form of sterol regulatory element binding protein 2



(SREBP2) in male germ cells suggests important cell- and stage-specific functions for SREBP targets during spermatogenesis." *Molecular and Cellular Biology*, 22(24), 8478-8490.

Wang, H., San Agustin, J. T., Witman, G. B., and Kilpatrick, D. L. (2004). "Novel role for a sterol response element binding protein in directing spermatogenic cell-specific gene expression." *Molecular and Cellular Biology*, 24(24), 10681-10688.

Wang, K., Zhang, H. T., Ma, D. Q., Bucan, M., Glessner, J. T., Abrahams, B. S., Salyakina, D., Imielinski, M., Bradfield, J. P., Sleiman, P. M. A., Kim, C. E., Hou, C. P., Frackelton, E., Chiavacci, R., Takahashi, N., Sakurai, T., Rappaport, E., Lajonchere, C. M., Munson, J., Estes, A., Korvatska, O., Piven, J., Sonnenblick, L. I., Retuerto, A. I. A., Herman, E. I., Dong, H. M., Hutman, T., Sigman, M., Ozonoff, S., Klin, A., Owley, T., Sweeney, J. A., Brune, C. W., Cantor, R. M., Bernier, R., Gilbert, J. R., Cuccaro, M. L., McMahon, W. M., Miller, J., State, M. W., Wassink, T. H., Coon, H., Levy, S. E., Schultz, R. T., Nurnberger, J. I., Haines, J. L., Sutcliffe, J. S., Cook, E. H., Minshew, N. J., Buxbaum, J. D., Dawson, G., Grant, S. F. A., Geschwind, D. H., Pericak-Vance, M. A., Schellenberg, G. D., and Hakonarson, H. (2009). "Common genetic variants on 5p14.1 associate with autism spectrum disorders." *Nature*, 459(7246), 528-533.

Wang, L. F., Jia, M. X., Yue, W. H., Tang, F. L., Qu, M., Ruan, Y., Lu, T. L., Zhang, H., Yan, H., Liu, J., Guo, Y. Q., Zhang, J., Yang, X. L., and Zhang, D. (2008). "Association of the ENGRAILED 2 (EN2) gene with autism in Chinese Han population." *American Journal of Medical Genetics Part B-Neuropsychiatric Genetics*, 147B(4), 434-438.

- Wang, X.-J. (2010). "Neurophysiological and computational principles of cortical rhythms in cognition." *Physiological Reviews*, 90(3), 1195-1268.
- Wang, X. T., Martindale, J. L., Liu, Y. S., and Holbrook, N. J. (1998). "The cellular response to oxidative stress: influences of mitogen-activated protein kinase signalling pathways on cell survival." *Biochemical Journal*, 333, 291-300.
- Wang, Y. M., Medvid, R., Melton, C., Jaenisch, R., and Blelloch, R. (2007). "DGCR8 is essential for microRNA biogenesis and silencing of embryonic stem cell self-renewal." *Nature Genetics*, 39(3), 380-385.
- Wayman, G. A., Davare, M., Ando, H., Fortin, D., Varlamova, O., Cheng, H. Y. M., Marks, D., Obrietan, K., Soclerling, T. R., Goodman, R. H., and Impey, S. (2008). "An activity-regulated microRNA controls dendritic plasticity by down-regulating p250GAP." *Proceedings of the National Academy of Sciences of the United States of America*, 105(26), 9093-9098.
- Wechsler, D. (1997). "The Wechsler Adult Intelligence Scale, 3rd edn." San Antonio, California: Psychological Corporation, Harcourt Brace.
- Weiler, I. J., Irwin, S. A., Klintsova, A. Y., Spencer, C. M., Brazelton, A. D., Miyashiro, K., Comery, T. A., Patel, B., Eberwine, J., and Greenough, W. T. (1997). "Fragile X mental retardation protein is translated near synapses in response to neurotransmitter activation." *Proceedings of the National Academy of Sciences of the United States of America*, 94(10), 5395-5400.
- Weiss, L. A., Arking, D. E., Johns, H., and Autism, C. (2009). "A genome-wide linkage and association scan reveals novel loci for autism." *Nature*, 461(7265), 802-U62.

- Weksberg, R., Shuman, C., and Beckwith, J. B. (2010). "Beckwith-Wiedemann syndrome." *European Journal of Human Genetics*, 18(1), 8-14.
- Welsh, J. P., Ahn, E. S., and Placantonakis, D. G. (2005). "Is autism due to brain desynchronization?" *International Journal of Developmental Neuroscience*, 23(2-3), 253-263.
- Wheeler, D. A., Kyriacou, C. P., Greenacre, M. L., Yu, Q., Rutila, J. E., Rosbash, M., and Hall, J. C. (1991). "Molecular transfer of a species-specific behavior from *Drosophila-simulans* to *Drosophila-melanogaster*." *Science*, 251(4997), 1082-1085.
- Wightman, B., Ha, I., and Ruvkun, G. (1993). "Posttranscriptional regulation of the heterochronic gene LIN-14 by LIN-4 mediates temporal pattern-formation in *C-elegans*." *Cell*, 75(5), 855-862.
- Williams, D. L., Goldstein, G., and Minshew, N. J. (2006). "The profile of memory function in children with autism." *Neuropsychology*, 20(1), 21-29.
- Williams, H. J., Craddock, N., Russo, G., Hamshere, M. L., Moskvina, V., Dwyer, S., Smith, R. L., Green, E., Grozeva, D., Holmans, P., Owen, M. J., and O'Donovan, M. C. (2011). "Most genome-wide significant susceptibility loci for schizophrenia and bipolar disorder reported to date cross-traditional diagnostic boundaries." *Human Molecular Genetics*, 20(2), 387-391.
- Williams, H. J., Norton, N., Dwyer, S., Moskvina, V., Nikolov, I., Carroll, L., Georgieva, L., Williams, N. M., Morris, D. W., Quinn, E. M., Giegling, I., Ikeda, M., Wood, J., Lencz, T., Hultman, C., Lichtenstein, P., Thiselton, D., Maher, B. S., Malhotra, A. K., Riley, B., Kendler, K. S., Gill, M., Sullivan, P., Sklar, P., Purcell, S., Nimgaonkar, V. L., Kirov, G., Holmans, P., Corvin, A., Rujescu, D., Craddock, N., Owen, M. J., O'Donovan, M. C., Mol Genetics, S., Isc, Plus, S., and Group. (2011). "Fine mapping of ZNF804A

and genome-wide significant evidence for its involvement in schizophrenia and bipolar disorder." *Molecular Psychiatry*, 16(4), 429-441.

Williamson, M. P. (1994). "The structure and function of proline-rich regions in proteins." *Biochemical Journal*, 297, 249-260.

Wimpory, D., Chadwick, P., and Nash, S. (1995). "Musical interaction therapy for children with autism - an evaluative case-study with 2-year follow-up - brief report." *Journal of Autism and Developmental Disorders*, 25(5), 541-552.

Wimpory, D., and Nash, S. (1999). "Musical interaction play therapy." *Child Language and Teaching Therapies* **15**: 17–28.

Wimpory, D., Nicholas, B., and Nash, S. (2002). "Social timing, clock genes and autism: a new hypothesis." *Journal of Intellectual Disability Research*, 46, 352-358.

Wimpory, D. C., Hobson, R. P., Williams, J. M. G., and Nash, S. (2000). "Are infants with autism socially engaged? A study of recent retrospective parental reports." *Journal of Autism and Developmental Disorders*, 30(6), 525-536.

Wiznitzer, M. (2004). "Autism and tuberous sclerosis." *Journal of Child Neurology*, 19(9), 675-679.

Wolff, D. J., Clifton, K., Karr, C., and Charles, J. (2002). "Pilot assessment of the subtelomeric regions of children with autism: Detection of a 2q deletion." *Genetics in Medicine*, 4(1), 10-14.

Woltering, J. M., and Durston, A. J. (2008). "MiR-10 Represses HoxB1a and HoxB3a in Zebrafish." *Plos One*, 3(1).

Wysocka, J., Myers, M. P., Laherty, C. D., Eisenman, R. N., and Herr, W. (2003). "Human Sin3 deacetylase and trithorax-related Set1/Ash2 histone H3-K4 methyltransferase are tethered together selectively by the cell-proliferation

- factor HCF-1." *Genes & Development*, 17(7), 896-911.
- Xiao, J. H., Davidson, I., Ferrandon, D., Rosales, R., Vigneron, M., Macchi, M., Ruffenach, F., and Chambon, P. (1987). "One cell-specific and 3 ubiquitous nuclear proteins bind in vitro to overlapping motifs in the domain-B1 of the SV40 enhancer." *Embo Journal*, 6(10), 3005-3013.
- Xu, S., Witmer, P. D., Lumayag, S., Kovacs, B., and Valle, D. (2007). "MicroRNA (miRNA) transcriptome of mouse retina and identification of a sensory organ-specific miRNA cluster." *Journal of Biological Chemistry*, 282(34), 25053-25066.
- Yamakawa, H., Oyama, S., Mitsuhashi, H., Sasagawa, N., Uchino, S., Kohsaka, S., and Ishiura, S. (2007). "Neuroligins 3 and 4X interact with syntrophin-gamma 2, and the interactions are affected by autism-related mutations." *Biochemical and Biophysical Research Communications*, 355(1), 41-46.
- Yan, J., Feng, J. N., Goldman, D., Cook, E. H., Craddock, N., Jones, I. R., Heston, L. L., and Sommer, S. S. (2004). "Mutation scanning of the androgen receptor gene in patients with psychiatric disorders reveals highly conserved variants in alcoholic and phobia patients." *Psychiatric Genetics*, 14(1), 57-60.
- Yang, M., Lee, J.-E., Padgett, R. W., and Edery, I. (2008a). "Circadian regulation of a limited set of conserved microRNAs in *Drosophila*." *Bmc Genomics*, 9.
- Yang, M. S., and Gill, M. (2007). "A review of gene linkage, association and expression studies in autism and an assessment of convergent evidence." *International Journal of Developmental Neuroscience*, 25(2), 69-85.
- Yang, P., Shu, B.-C., Hallmayer, J. F., and Lung, F.-W. (2010). "Intronic single nucleotide polymorphisms of Engrailed homeobox 2 modulate the disease vulnerability of autism in a Han Chinese population." *Neuropsychobiology*,

62(2), 104-115.

Yang, S., Van Dongen, H. P. A., Wang, K., Berrettini, W., and Bucan, M. (2009).

"Assessment of circadian function in fibroblasts of patients with bipolar disorder." *Molecular Psychiatry*, 14(2), 143-155.

Yang, S. Z., Wang, K., Valladares, O., Hannenhalli, S., and Bucan, M. (2007).

"Genome-wide expression profiling and bioinformatics analysis of diurnally regulated genes in the mouse prefrontal cortex." *Genome Biology*, 8(11).

Yang, X. L., Xie, J., Niu, B., Hu, X. N., Gao, Y., Xiang, Q., Zhang, Y. H., Guo, Y.,

and Zhang, Z. G. (2005). "Structure analysis of the protein transduction domain of human Period1 and its mutant analogs." *Journal of Molecular Graphics & Modelling*, 23(5), 389-394.

Yin, L., and Lazar, M. A. (2005). "The orphan nuclear receptor Rev-erb alpha

recruits the N-CoR/histone deacetylase 3 corepressor to regulate the circadian Bmal1 gene." *Molecular Endocrinology*, 19(6), 1452-1459.

Yin, L., Wang, J., Klein, P. S., and Lazar, M. A. (2006). "Nuclear receptor Rev-

erba is a critical lithium-sensitive component of the circadian clock." *Science*, 311(5763), 1002-1005.

Yirmiya, N., Pilowsky, T., Nemanov, L., Arbelle, S., Feinsilver, T., Fried, I., and

Ebstein, R. P. (2001). "Evidence for an association with the serotonin transporter promoter region polymorphism and autism." *American Journal of Medical Genetics*, 105(4), 381-386.

Yu, S., Pritchard, M., Kremer, E., Lynch, M., Nancarrow, J., Baker, E., Holman,

K., Mulley, J. C., Warren, S. T., Schlessinger, D., Sutherland, G. R., and Richards, R. I. (1991). "Fragile-x genotype characterized by an unstable region of DNA." *Science*, 252(5009), 1179-1181.

- Yu, X., Zhou, Q., Li, S.-C., Luo, Q., Cai, Y., Lin, W.-c., Chen, H., Yang, Y., Hu, S., and Yu, J. (2008). "The silkworm (*Bombyx mori*) microRNAs and their expressions in multiple developmental stages." *Plos One*, 3(8).
- Yuan, J., Pu, M. T., Zhang, Z. G., and Lou, Z. K. (2009). "Histone H3-K56 acetylation is important for genomic stability in mammals." *Cell Cycle*, 8(11), 1747-1753.
- Yujnovsky, I., Hirayama, J., Doi, M., Borrelli, E., and Sassone-Corsi, P. (2006). "Signaling mediated by the dopamine D2 receptor potentiates circadian regulation by CLOCK : BMAL1." *Proceedings of the National Academy of Sciences of the United States of America*, 103(16), 6386-6391.
- Zafeiriou, D. I., Ververi, A., and Vargiami, E. (2007). "Childhood autism and associated comorbidities." *Brain & Development*, 29(5), 257-272.
- Zald, D. H. (2003). "The human amygdala and the emotional evaluation of sensory stimuli." *Brain Research Reviews*, 41(1), 88-123.
- Zhang, H. F., Zhang, R., Luo, Y., D'Alessio, A., Pober, J. S., and Min, W. (2004). "AIP1/DAB2IP, a novel member of the Ras-GAP family, transduces TRAF2-induced ASK1-JNK activation." *Journal of Biological Chemistry*, 279(43), 44955-44965.
- Zhao, X., Leotta, A., Kustanovich, V., Lajonchere, C., Geschwind, D. H., Law, K., Law, P., Qiu, S., Lord, C., Sebat, J., Ye, K., and Wigler, M. (2007). "A unified genetic theory for sporadic and inherited autism." *Proceedings of the National Academy of Sciences of the United States of America*, 104(31), 12831-12836.
- Zhao, Z., Tavoosidana, G., Sjolinder, M., Gondor, A., Mariano, P., Wang, S., Kanduri, C., Lezcano, M., Sandhu, K. S., Singh, U., Pant, V., Tiwari, V., Kurukuti, S., and Ohlsson, R. (2006). "Circular chromosome conformation

capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions." *Nature Genetics*, 38(11), 1341-1347.

- Zhou, Y. D., Barnard, M., Tian, H., Li, X., Ring, H. Z., Francke, U., Shelton, J., Richardson, J., Russell, D. W., and McKnight, S. L. (1997). "Molecular characterization of two mammalian bHLH-PAS domain proteins selectively expressed in the central nervous system." *Proceedings of the National Academy of Sciences of the United States of America*, 94(2), 713-718.
- Zhu, Y., Stevens, R. G., Hoffman, A. E., FitzGerald, L. M., Kwon, E. M., Ostrander, E. A., Davis, S., Zheng, T. Z., and Stanford, J. L. (2009). "Testing the Circadian Gene Hypothesis in Prostate Cancer: A Population-Based Case-Control Study." *Cancer Research*, 69(24), 9315-9322.
- Zhu, Y., Stevens, R. G., Leaderer, D., Hoffman, A., Holford, T., Zhang, Y. W., Brown, H. N., and Zheng, T. Z. (2008). "Non-synonymous polymorphisms in the circadian gene NPAS2 and breast cancer risk." *Breast Cancer Research and Treatment*, 107(3), 421-425.
- Zwaigenbaum, L., Bryson, S., Rogers, T., Roberts, W., Brian, J., and Szatmari, P. (2005). "Behavioral manifestations of autism in the first year of life." *International Journal of Developmental Neuroscience*, 23(2-3), 143-152.