Autism-associated SNPs in the clock genes *npas2*, *per1* and the homeobox gene *en2* alter DNA sequences that show characteristics of microRNA genes.

Brad Nicholas [1], Michael J Owen [2], Dawn C Wimpory[3/4] and Thomas Caspari [1]

Correspondence: Brad Nicholas
North West Cancer Research Fund Institute,
Bangor University,
Gwynedd, LL57 2DG,
United Kingdom.

E-mail: b.nicholas@bangor.ac.uk

[1] North West Cancer Research Fund Institute, Bangor University, U.K.
[2] Department of Psychological Medicine, Cardiff University, U.K.
[3] School of Psychology, Bangor University, UK
[4] Specialist Children's Services, North West Wales NHS Trust, Bangor, U.K.

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 bioinformatics analysis of the intronic regions which 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 A-C 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*.

Key words: Autistic Disorder; Polymorphism, Single Nucleotide; MicroRNAs; Biological Clocks; Genes, Homeobox; Cerebellum; Computational Biology; Behavioural Genetics; RNA; Child Development Disorders.

Introduction

A number of genes are considered to contribute to the heritability (1) of the neuro-developmental disorder autism (2). Numerous candidate gene studies have been deployed (reviewed by Yang and Gill (3) 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, Matthews, Turner et al. (4)); cytogenetic studies of affected individuals (for example, Wolff, Clifton, Karr et al. (5)) and genetic studies of disorders that show co-morbidity with autism, (reviewed, for example, by Zafeiriou, Ververi and Vargiami (6)). 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 (7), the association of the intronic SNPs rs1861972 and rs1861973 with autistic disorder (8) has been reinvestigated a number of times with overall positive but complex results (9, 10) that support evidence for abnormal cerebellar development in autism (11). The reported association of *per1* and *npas2* with autistic disorder (12) is noteworthy given that specific sleep anomalies (13), altered circadian rhythm (14) and increased measures of oxidative stress are reported to be associated with autism (13, 15). *Per1* and *npas2* regulate sleep in mammal (16, 17) and *npas2* effects redox signaling (18, 19). The protein products of these genes interact (20), 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 (21) and schizophrenia (22), both indicated to share genetic overlap with autism (23). 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 (p*er1*) with autism, no other autism linked functional changes in these genes were established (8) (12).

Recent findings in the field of small RNAs show that the introns of protein coding genes may harbour 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 (24). MicroRNA genes within the introns of protein coding genes are *a priori* transcribed concurrently along with the host gene and the genes *en2, npas2 and per1* are all expressed in brain regions relevant to autism (25, 20, 16). In this study we sought to investigate whether the intron regions containing the autism-associated SNPs: rs1861972 (*en2*); rs1861973 (*en2*); rs1811399 (*npas2*) and, rs885747 (p*er1*) show characteristics of human microRNA genes which might be affected by the autism-associated SNPs.

MicroRNAs regulate development (26). They are abundant in the mammalian central nervous system (27) and are reported to play a role in Fragile-X sydrome (28), Tourette's syndrome (29) and perhaps schizophrenia (30, 31) and autism (32). Mammalian mature microRNAs are short (21 to 23 nucleotide) RNA molecules that together with proteins of

the RISC complex effect gene silencing (33). The precursors of microRNAs are derived from intergenic microRNA genes (34) or microRNA genes within introns of protein coding genes. Indeed some 50% of human microRNA genes appear to be intronic (35). 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 (36), (37) . The ~60-100 nt long pre-microRNA hairpins are transported to the cytoplasm by EXPORTIN-5. Here the loop and tail 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 (38) or rapid degradation of the transcript of the target gene (39).

MicroRNAs regulate cellular processes that are relevant to the study of autism e.g.: developmental timing (26, 40); cell death (41); the patterning of the nervous system (42); and, the survival of Purkinje neurons (43), a cell type affected in cases of autism (11). SNPs in microRNA genes can profoundly affect the target specificity and gene silencing power of microRNAs (44, 45). We wondered, therefore, whether intronic SNPs that are associated with autistic disorder may 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 (p*er1*) 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.

Materials and Methods

(Autism enriched alleles are shown in bold throughout)

We carried out a bioinformatics 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 http://rnaanalyzer.bioapps.biozentrum.uni-wuerzburg.de (67). 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, Cuppen and Plasterk (46) and Stark, Lin and Kheradpour et al. (36). 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. Pri-microRNA-like structures were screened for DROSHA processing sites (required to generate a pre-microRNA from a pri-microRNA) using the bioinformatics tool Microprocessor SMV *in silico* DROSHA described by Helvik, Snøve and Sætrom (47) .

To determine what would be the predicted target genes of a mature microRNA, derived from any such hairpins found harbouring 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. (36) and Berezikov et al. (46) 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 (48)). 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 (http://microrna.sanger.ac.uk) was used (49). 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 (36). Thus if candidate seeds are considered (by a heptomer 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 (36) (see Figure 1). 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 (36). BLASTN and SSEARCH tools were also used to consider sequence conservation. Searches were performed using the web based tools in MirBASE (http://microrna.sanger.ac.uk (49)). Thus a number of arguments were deployed against the notion that hairpins containing autism-associated SNPs are *bona fide* microRNA genes.

Insert Figure 1 about here

The two SNPs (rs1861972 and rs1861973) in *en2*, reported by Benayed, Gharani, Rossman et al. (8) 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. (8) 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 40). Rs885747(**C**/G), that is reported to be associated with autism (12) and rs885953(G/C), that has not been reported to have 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 publicly available software (http://microrna.sanger.ac.uk/ (49)), 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, do the common targets have relevance to autism?

Results

Analysis of 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 (8). These SNPs are situated in the 3' half of the single large intron that divides the two exons of *en2* (7q 31.1). The two SNPs are 151 nucleotides apart and were analysed together in a 300 nucleotide genomic DNA sequence fragment (Ensembl 2008 release 40). Two other SNPs are present in this sequence fragment; rs35529773 (C/-) for which linkage 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 2A) 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-rs1861973**C**/T (Figure 2A). 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 (47) gave a PPV >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.

Insert Figure 2 about here

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 closely base paired stretch in 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 (50)). The overlapping seed sequences of miR-10, miR-339 and miR-504 each contain the autism-associated SNP (see Figure 3). Intriguingly, miR-10 is reported to play a role in hind brain development (51) that is in keeping with the function of the *en2* gene in the cerebellum development (52) and with reports of cerebellum anomalies in autism (11). 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 (see Figure 3).

Insert Figure 3 about here.

Analysis of the intron region of *npas2* containing the autism-associated SNP rs1811399 (**C**/A)

A 300 nucleotide DNA sequence fragment (ensemble 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 2B. 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 (36). 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 2C 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. 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 (47), we were able to detect a DROSHA processing site that would derive a pre-microRNA-like structure shown in Figure 1B. 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 (S.I. F2). In the 5' and 3' arms, seed-match trend peaks coincided with base pairing mismatches and were taken as likely mature microRNA start points. The two best candidate seeds, one from the 5' arm and one from the 3'arm, were used for investigating the predicted targets of this hairpin using the web based tool TargetScan (49).

Analysis of intron 12 of of *per1* containing the autism-associated SNP rs885747 (**C**/G)

Intron 12 of *per1* contains two SNPs: rs885747, that Nicholas, Rudrasingham and Nash et al. (12) found to be associated with autistic disorder (**C** allele) and rs885953 that was not analysed in their experiment. There is remarkable conservation amongst primates of the predicted hairpin structure of the RNA transcripts of this intron (see SI File 1, *per1*). In *H. sapiens* this hairpin structure is strongly affected by the allele combination at these two loci (see Figure 2C). 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 (45), we considered that this intron may best be analysed in terms of the structural features of mirtrons (53, 54). Mirtrons are mircoRNAs transcribed from small introns that bye-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 (see Figure 4). 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. 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 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.

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* (see S.I. F.2) included a number of autism candidate genes*: Rai1*, *gabrb3*, *gabrb2*, *shank3, nrxn3*, *reln*, *pitx1*, *shank3-interacting-protein-1*, *a2bp1, stk39* and *dlx1* (3, 6). 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 1.

Discussion

The above findings suggest that introns 1, 2 and 12 of the genes *en2*, *npas2* and *per1* respectively, may harbour 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. (8). Further comparing our results we see that in all of the samples presented by Benayed et al. (AGRE 1, AGRE 2, NIMH and the DSP siblings) (8), 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. 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, Korvatska, Allen-Brady et al. (9) confirmed association of rs1861972 with autism (rs1861973 was not analysed in their study). However, according to Brune et al. (9) 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. (9). 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. (9) and Benayed et al. (8), 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 (see table 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 (12) 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 shown in Figure 1B 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* (55, 56) and the Down syndrome related kinase *Dyrk1a* (6, 57-59). If this microRNA exists and is active in autism then our model would also implicate rs885953, a SNP which is not currently fully characterized (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 (1). 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 (63). *Map2k4* (17p11.2) is linked to the cellular response to oxidative stress (60) and along with *npas2*, may be implicated in oxidative stress induced apoptosis of dopaminergic neurons (61). *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 (62). *Dab2ip* (9q33.2) transduces TRAF2-induced ASK1-JNK activation (63) and thus plays a central role in the oxidative stress response pathway (64) which is reported to be affected in autism (15). *Mtmr4* (17q22) is a lipid phosphatase for phosphatidinositol-3-phosphate (PtdIns3P) (65) and disrupted phosphatidylinositol signalling is also reported in autism (66).

It is not currently possible to verify the existence of microRNA genes on the basis of bioinformatics 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.

Acknowledgements

We gratefully acknowledge the collaboration of Autism Cymru and the European Social Fund for funding our Principal Investigator, Brad Nicholas, throughout the duration of this research. We 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). We also particularly wish to thank Professor William Fraser for his enthusiasm and advice, together with Professor Elizabeth Newson, OBE, for her original and inspiring clinical insight.

References

1. Pickles A, Bolton P, Macdonald H, Bailey AL, Couteur A, Sim CH *et al.* Latent-class analysis of recurrence risks for complex phenotypes with selection and measurement error: a twin and family history study of autism. *Am J Hum Genet* 1995; **3**: 717-726.

2. APA (ed). *Diagnostic and Statistical Manual of Mental Disorders, (DSM-IV)*: Washington, DC, 1994.

3. Yang MS, Gill M. A review of gene linkage, association and expression studies in autism and an assessment of convergent evidence. *International Journal of Developmental Neuroscience* 2007; **25**: 69-85.

4. Palferman S, Matthews N, Turner M, Moore J, Hervas A, Aubin A *et al.* A genomewide screen for autism: Strong evidence for linkage to chromosomes 2q, 7q, and 16p. *American Journal of Human Genetics* 2001; **69**: 570-581.

5. Wolff DJ, Clifton K, Karr C, Charles J. Pilot assessment of the subtelomeric regions of children with autism: Detection of a 2q deletion. *Genetics in Medicine* 2002; **4**: 10-14.

6. Zafeiriou DI, Ververi A, Vargiami E. Childhood autism and associated comorbidities. *Brain & Development* 2007; **29**: 257-272.

7. Sgaier SK, Lao Z, Villanueva MP, Berenshteyn F, Stephen D, Turnbull RK *et al.* Genetic subdivision of the tectum and cerebellum into functionally related regions based on differential sensitivity to engrailed proteins. *Development* 2007; **134**: 2325-2335.

8. Benayed R, Gharani N, Rossman I, Mancuso V, Lazar G, Kamdar S *et al.* Support for the homeobox transcription factor gene ENGRAILED 2 as an autism spectrum disorder susceptibility locus. *American Journal of Human Genetics* 2005; **77**: 851-868.

9. Brune CW, Korvatska E, Allen-Brady K, Cook EH, Dawson G, Devlin B *et al.* Heterogeneous association between engrailed-2 and autism in the CPEA network. *American Journal of Medical Genetics Part B-Neuropsychiatric Genetics* 2008; **2:** 187-193.

10. Wang L JM, Yue W, Tang F, Qu M, Ruan Y, Lu T, Zhang H, et al. Association of the ENGRAILED 2 (EN2) gene with autism in Chinese Han population. *Am J Med Genet B Neuropsychiatr Genet* 2008; **4**: 434-438.

11. Kemper TL BM. Neuropathology of infantile autism. *Mol Psychiatry* 2002; **7**: S12-3.

12. Nicholas B, Rudrasingham V, Nash S, Kirov G, Owen MJ, Wimpory DC. Association of Per1 and Npas2 with autistic disorder: support for the clock genes/social timing hypothesis. *Molecular Psychiatry* 2007; **12**: 581-592.

13. Limoges E, Mottron L, Bolduc C, Berthiaume C, Godbout R. Atypical sleep architecture and the autism phenotype. *Brain* 2005; **128**: 1049-1061.

14. Corbett BA, Mendoza S, Abdullah M, Wegelin JA, Levine S. Cortisol circadian rhythms and response to stress in children with autism. *Psychoneuroendocrinology* 2006; **31**: 59-68.

15. James SJ, Cutler P, Melnyk S, Jernigan S, Janak L, Gaylor DW *et al.* Metabolic biomarkers of increased oxidative stress and impaired methylation capacity in children with autism. *American Journal of Clinical Nutrition* 2004; **80**: 1611-1617.

16. Sun ZS, Albrecht U, Zhuchenko O, Bailey J, Eichele G, Lee CC. RIGUI, a putative mammalian ortholog of the Drosophila period gene. *Cell* 1997; **90**: 1003-1011.

17. Franken P, Dudley CA, Estill SJ, Barakat M, Thomason R, O'Haran BF *et al.* 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* 2006; **103**: 7118-7123.

18. Rutter J, Reick M, Wu LC, McKnight SL. Regulation of clock and NPAS2 DNA binding by the redox state of NAD cofactors. *Science* 2001; **293**: 510-514.

19. Dioum EM, Rutter J, Tuckerman JR, Gonzalez G, Gilles-Gonzalez MA, McKnight SL. NPAS2: A gas-responsive transcription factor. *Science* 2002; **298**: 2385-2387.

20. Reick M, Garcia JA, Dudley C, McKnight SL. NPAS2: An analog of clock operative in the mammalian forebrain. *Science* 2001; **293**: 506-509.

21. Yang S, Van Dongen HP, Wang K, Berrettini W, M. B. Assessment of circadian function in fibroblasts of patients with bipolar disorder. *Mol Psychiatry* 2008 Feb 26. [Epub ahead of print].

22. Aston C JL, Sokolov BP. Microarray analysis of postmortem temporal cortex from patients with schizophrenia. *J Neurosci Res* 2004; **77**: 858-866.

23. Rzhetsky A, Wajngurt D, Park N, Zheng T. Probing genetic overlap among complex human phenotypes. *Proceedings of the National Academy of Sciences of the United States of America* 2007; **104**: 11694-11699.

24. Lin SL, Miller JD, Ying SY. IntronicMicroRNA (microRNA). *Journal of Biomedicine and Biotechnology* 2006 **4**: 26818.

25. Millen KJ, Wurst W, Herrup K, AL. J. Abnormal embryonic cerebellar development and patterning of postnatal foliation in two mouse Engrailed-2 mutants. *Development* 1994; **3**: 695-706.

26. Lee RC, Feinbaum RL, Ambros V. The C-elegans heterochronic gene LIN-4 encodes small RNAs with antisense complementarity to LIN-14. *Cell* 1993; **75**: 843-854.

27. Bak M, Silahtaroglu A, Moller M, Christensen M, Rath MF, Skryabin B *et al.* MicroRNA expression in the adult mouse central nervous system. *Rna-a Publication of the Rna Society* 2008; **14**: 432-444.

28. Lin SL, Chang SJE, Ying SY. First in vivo evidence of microRNA-induced fragile X mental retardation syndrome. *Molecular Psychiatry* 2006; **11**: 616-617.

29. Abelson JF, Kwan KY, O'Roak BJ, Baek DY, Stillman AA, Morgan TM *et al.* Sequence variants in SLITRK1 are associated with Tourette's syndrome. *Science* 2005; **310**: 317-320.

30. Perkins DO, Jeffries CD, Jarskog LF, Thomson JM, Woods K, Newman MA *et al.* microRNA expression in the prefrontal cortex of individuals with schizophrenia and schizoaffective disorder. *Genome Biology* 2007; **2**: R27

31. Hansen T OL, Lindow M, Jakobsen KD, Ullum H, Jonsson E, Andreassen OA, Djurovic S, Melle I, Agartz I, Hall H, Timm S, Wang AG, Werge T. Brain expressed microRNAs implicated in schizophrenia etiology. *PLoS ONE* 2007; **2**: e873.

32. Abu-Elneel K, Liu T GF, Nishimura Y, Wall DP, Geschwind DH, Lao K *et al.* Heterogeneous dysregulation of microRNAs across the autism spectrum. *Neurogenetics* 2008; **9**: 153-161.

33. Lewis BB, CB; Bartel, DP. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* 2005; **120**: 15-20.

34. Saini HK, Griffiths-Jones S, Enright AJ. Genomic analysis of human microRNA transcripts. *Proceedings of the National Academy of Sciences of the United States of America* 2007; **104**: 17719-17724.

35. Rodriguez A, Griffiths-Jones S, Ashurst JL, Bradley A. Identification of mammalian microRNA host genes and transcription units. *Genome Research* 2004; **14**: 1902-1910.

36. Stark A, Kheradpour P, Parts L, Brennecke J, Hodges E, Hannon GJ *et al.* Systematic discovery and characterization of fly microRNAs using 12 Drosophila genomes. *Genome Research* 2007; **17**: 1865-1879.

37. Stark A, Lin MF, Kheradpour P, Pedersen JS, Parts L, Carlson JW *et al.* Discovery of functional elements in 12 Drosophila genomes using evolutionary signatures. *Nature* 2007; **450**: 219-232.

38. Wightman B, Ha I, Ruvkun G. Posttranscriptional regulation of the heterochronic gene LIN-14 by LIN-4 mediates temporal pattern-formation in C-elegans. *Cell* 1993; **75:** 855-862.

39. Giraldez AJ, Mishima Y, Rihel J, Grocock RJ, Van Dongen S, Inoue K *et al.* Zebrafish MiR-430 promotes deadenylation and clearance of maternal mRNAs. *Science* 2006; **312**: 75-79.

40. Reinhart BJ, Weinstein EG, Rhoades MW, Bartel B, Bartel DP. MicroRNAs in plants. *Genes & Development* 2002; **16**: 1616-1626.

41. Jovanovic M, Hengartner MO. microRNAs and apoptosis: RNAs to die for. *Oncogene* 2006; **25**: 6176-6187.

42. Kosik KS. The neuronal microRNA system. *Nature Reviews Neuroscience* 2006; **7**: 911-920.

43. Schaefer A, O'Carroll D, Tan CL, Hillman D, Sugimori M, Llinas R *et al.* Cerebellar neuro degeneration in the absence of microRNAs. *Journal of Experimental Medicine* 2007; **204**: 1553-1558.

44. Saunders MA, Liang H, Li WH. Human polymorphism at microRNAs and microRNA target sites. *Proceedings of the National Academy of Sciences of the United States of America* 2007; **104**: 3300-3305.

45. Duan RH, Pak CH, Jin P. Single nucleotide polymorphism associated with mature miR-125a alters the processing of pri-microRNA. *Human Molecular Genetics* 2007; **9**: 1124-1131.

46. Berezikov E, Cuppen E, Plasterk RHA. Approaches to microRNA discovery. *Nature Genetics* 2006; **38**: S2-S7.

47. Helvik SA, Snove O, Saetrom P. Reliable prediction of Drosha processing sites improves microRNA gene prediction. *Bioinformatics* 2007; **23**: 142-149.

48. Michael MZ, O' Connor SM, van Holst Pellekaan NG, Young GP, RJ. J. Reduced accumulation of specific microRNAs in colorectal neoplasia. *Molecular Cancer Research* 2003; **12**: 882 -891.

49. Griffiths-Jones S, Grocock RJ, van Dongen S, Bateman A, Enright AJ. miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Research* 2006; **34**: D140-D144.

50. Xiaomin Yu, Qing Zhou, Sung-Chou Li, Qibin Luo, Yimei Cai, Wen-chang Lin *et al.* The Silkworm (Bombyx mori) microRNAs and Their Expressions in Multiple Developmental Stages. *PL0S ONE* 2008; **3**: e2997.

51. Woltering JM, Durston DA. MiR-10 represses HoxB1a and HoxB3a in zebrafish. *PLoS ONE* 2008 2008; **1:** e1396.

52. Sgaier SK LZ, Villanueva MP, Berenshteyn F, Stephen D, Turnbull RK, Joyner AL. Genetic subdivision of the tectum and cerebellum into functionally related regions based on differential sensitivity to engrailed proteins. *Development* 2007; **134**: 2325-2335.

53. Ruby JG, Jan CH, Bartel DP. Intronic microRNA precursors that bypass Drosha processing. *Nature* 2007; **448**: 83-U87.

54. Berezikov E, Chung WJ, Willis J, Cuppen E, Lai EC. Mammalian mirtron genes. *Molecular Cell* 2007; **28**: 328-336.

55. Bi WM, Yan J, Shi X, Yuva-Paylor LA, Antalffy BA, Goldman A *et al.* Rai1 deficiency in mice causes learning impairment and motor dysfunction, whereas Rai1 heterozygous mice display minimal behavioral phenotypes. *Human Molecular Genetics* 2007; **16**: 1802-1813.

56. Potocki L, Bi WM, Treadwell-Deering D, Carvalho CMB, Eifert A, Friedman EM *et al.* 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* 2007; **80**: 633-649.

57. Altafaj X, Dierssen M, Baamonde C, Marti E, Visa J, Guimera J *et al.* Neurodevelopmental delay, motor abnormalities and cognitive deficits in transgenic mice overexpressing Dyrk1A (minibrain), a murine model of Down's syndrome. *Human Molecular Genetics* 2001; **10**: 1915-1923.

58. Fotaki V, Dierssen M, Alcantara S, Martinez S, Marti E, Casas C *et al.* Dyrk1A haploinsufficiency affects viability and causes developmental delay and abnormal brain morphology in mice. *Molecular and Cellular Biology* 2002; **22**: 6636-6647.

59. Dowjat WK, Adayev T, Kuchna I, Nowicki K, Palminiello S, Hwang YW *et al.* Trisomy-driven overexpression of DYRK1A kinase in the brain of subjects with Down syndrome. *Neuroscience Letters* 2007; **413**: 77-81.

60. Wang X, Martindale JL, Liu Y, NJ. H. The cellular response to oxidative stress: influences of mitogen-activated protein kinase signalling pathways on cell survival. *Biochem J* 1998; **15**: 291-300.

61. Anantharam V, Lehrmann E, Kanthasamy A, Yang Y, Banerjee P, Becker KG *et al.* Microarray analysis of oxidative stress regulated genes in mesencephalic dopaminergic neuronal cells: relevance to oxidative damage in Parkinson's disease. *Neurochem Int* 2007; **6**: 834-847.

62. Ageta H MA, Migishima R, Kida S, Tsuchida K, Yokoyama M, Inokuchi K. Activin in the brain modulates anxiety-related behavior and adult neurogenesis. *PLoS ONE* 2008; **3**: e1869.

63. Zhang H ZR, Luo Y, D'Alessio A, Pober JS, Min W. AIP1/DAB2IP, a novel member of the Ras-GAP family, transduces TRAF2-induced ASK1-JNK activation. *J Biol Chem* 2004; **43**: 44955-44965.

64. Shen HM, Lin Y, Choksi S, Tran J, Jin T, Chang L *et al.* Essential roles of receptor-interacting protein and TRAF2 in oxidative stress-induced cell death. *Mol Cell Biol* 2004; **13**: 5914-5922.

65. Lorenzo O, Urbé S, MJ C. Systematic analysis of myotubularins: heteromeric interactions, subcellular localisation and endosome related functions. *J Cell Sci* 2006; **14**: 2953-2959.

66. Serajee FJ, Nabi R, Zhong H, AH. MH. Association of INPP1, PIK3CG, and TSC2 gene variants with autistic disorder: implications for phosphatidylinositol signalling in autism. *J Med Genet* 2003; **11**: e119.

67. Bengert B. and Dandekar T.
A software tool-box for analysis of regulatory RNA elements
Nucl. Acids. Res. 2003 **31**: 3441-3445.

Table 1: Table showing the predicted targets of the candidate microRNAs in *en2, npas2 and per1*

Host gene	***Npas2***		***En2***		***Per1***
Arm of candidate microRNA	5'	3'	3'	3'	5'
Seed of candidate mature microRNA	UCUGGAG	ACAGUCA	ACCCTGT hsa-mir-10	TTACAGC	GGACAGG
number of predicted targets	246	235	178	564	159
Autism relevant targets of the candidate microRNAs	**GABRB2** *MECP2* **NLGN2** ARID1A-	**GABRB3** **PITX1** **NRXN3** *RAI1* **RELN-** **STK39** TLK1- MITF+ ProSAPiP1	**BDNF** ***SHANK3^*** *NF1* FLT1+	**A2BP1^** **AUTS2** **DLX1** **GABRB3** **HTR2A** *RAI1* ARID1A- SCHIP1+ TLK1-	*DYRK1A* *RAI1* FLOT2- KIF1A^
Targets common to candidate microRNAs from each gene	ACVR1B DAB2IP MAP2K4 MTMR4 RAI1				

Table 1

Table 1 shows the seed sequences of the candidate mature microRNAs in *en2*, *npas2* and *per1* in relation to their autism relevant target genes. 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.

Titles and Legends to Figures

Figure 1: Candidate microRNA from intron 2 of *npas2* and verified human microRNA hsa-mir-10b for comparison

Figure 1 shows the candidate microRNA from intron 2 of *npas2* (containing the autism associated SNP rs1811399) as compared with the verified human microRNA hsa-mir-10b. Bold black arrows indicate the position of predicted DROSHA binding 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 while inset B and C show the predicted structures of the *npas2* candidate microRNA and has-mir-10b, respectively, after DROSHA cleavage.

Figure 2 :The effect of autism associated SNPs on RNA hairpin structures in intronic transcrips of *en2, npas2 and per1*.

Figure 2A shows the predicted secondary structures of RNA transcripts from the region of intron 1 of *en2* containing the autism-associated SNPs rs1861972 and rs1861973. *In silico* analysis predicts that different transcript structures are determined by the various allele combinations of rs1861972, rs1861973, rs3824067 and rs35529773. The long hairpin (candidate microRNA) is, in all cases, disrupted by the presence of the autism-associated haplotype rs1861972(**A**)-rs1861973(**C**). 2B shows the predicted secondary structure of the RNA transcript from the region of intron 2 of *npas2* containing the autism-associated SNP rs1811399. A long hairpin is predicted with the A allele which is

disrupted by the autism enriched **C** allele of rs1811399. 2C shows the predicted secondary structure of the RNA transcript of intron 12 of *per1*. This intron contains two SNPs: rs885953 and autism-associated rs885747. The C-**C** and G-G haplotypes preclude formation of a mirtron-like structure which is permitted by the presence of the C-G and G-**C** haplotypes.

Figure 3: Candidate microRNA from *En2* intron1 containing the autism-associated SNP rs1861973.

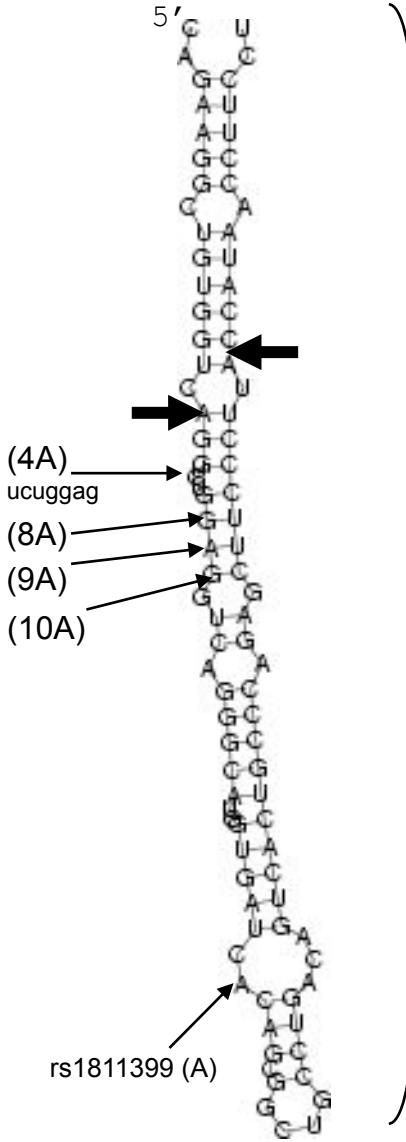
Figure 3 shows the *en2* candidate microRNA that contains the autism-associated SNP rs1861973. Candidate seed vs. number of targets are shown for the 5' and 3' arms in boxes A and B respectively. Note the seed sequences of three known human microRNAs are represented in the candidate mature microRNA region of the 3' arm.

Figure 4: Predicted RNA secondary structures of the transcript of intron 12 *per1* containing the autism associated SNP rs885747

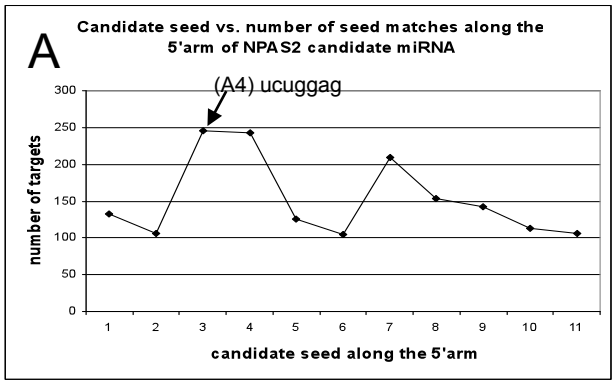
Figure 4 shows the predicted folding of the complete transcript of human *per1* intron 12. The 4 combinations of allele of the two SNPs in this intron determine different RNA secondary structures, two of which are mirtron-like. The graphs in the bottom half of the figure show number of seed matches vs candidate seed for the 5' and 3' arms of the mirtron-like structures. The shaded region in the graph relating to the 5'arm (bottom left) delineates seed matches that would be excluded as candidate seeds in the G-**C** haplotype. This region in the C-G haplotype would be single stranded and upstream of the double

stranded mirtron-like region. A seed match peak coinciding with a likely start site (black arrow C-G haplotype) would be relocated in the G-**C** haplotype structure to an unlikely start position, and thus lost as a candidate seed. The grey arrow indicates a likely start position for the G-**C** haplotype 5'candidate microRNA. This start point has a low seed match number and is thus a less likely candidate mature microRNA. For the 3' arm, the shaded region in the graph (bottom right) relating to the C-G haplotype indicates seed matches which would be lost in the G-**C** haplotype by their relocation to the loop region of the G-**C** haplotype structure.

Candidate miRNA
from *npas2* intron 2

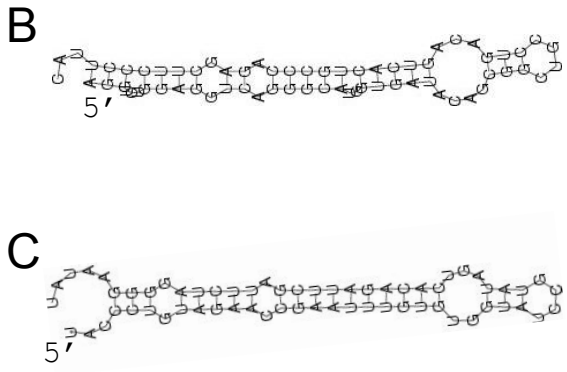


Stem length 95nt
energy -39.70 kcal/mol
(containing candidate
miRNA)

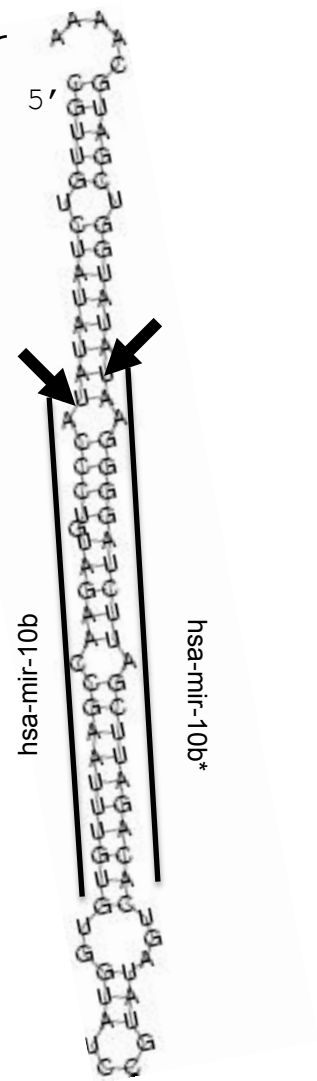


npas2 fragment

cagaaggcugugguc aggucuggaggucagggcauggugaucacag
cg***gcug***ccugacagucacugcccagagcuucccuua ccauaaccuuccu



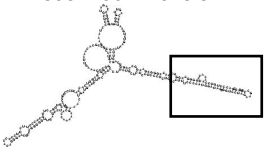
Example of a verified
human microRNA
(hsa-mir-10b)



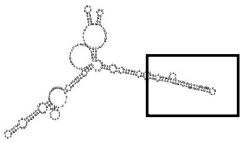
Stem length 91nt
energy -37.20 kcal/mol
mature hsa-mir-10b
(shown as black bars)

A

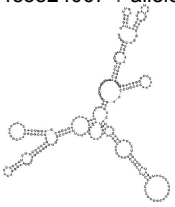
G-T haplotype with rs3824067 A allele



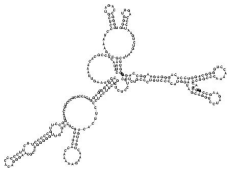
A-T haplotype with Rs3824067 A allele



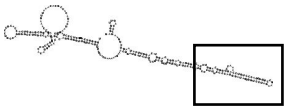
G-C haplotype with rs3824067 T allele



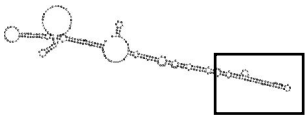
A-C haplotype with rs3824067 A allele



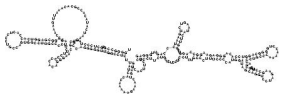
G-T haplotype with rs3824067 A allele and rs35529773 deletion



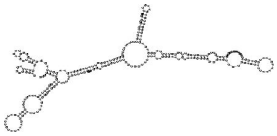
A-T haplotype with rs3824067 A allele and rs35529773 deletion



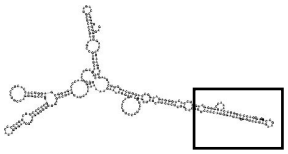
G-C haplotype with rs3824067 A allele and rs35529773 deletion



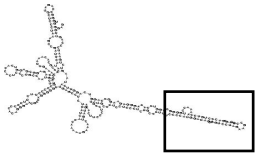
A-C haplotype with rs3824067 A allele and rs35529773 deletion



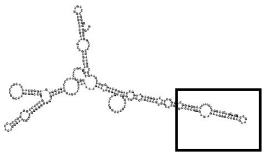
G-T haplotype with rs3824067 T allele



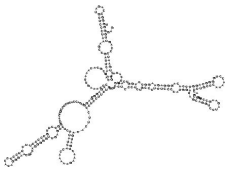
A-T haplotype with rs3824067 T allele



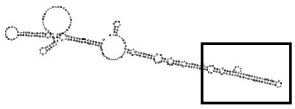
G-C haplotype with rs3824067 T allele



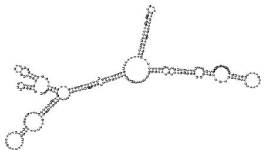
A-C haplotype with rs3824067 T allele



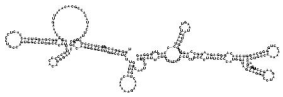
G-T haplotype with rs3824067 T allele and rs35529773 deletion



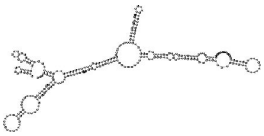
A-T haplotype with rs3824067 T allele and rs35529773 deletion



G-C haplotype with rs3824067 T allele and rs35529773 deletion



A-C haplotype with rs3824067 T allele and rs35529773 deletion



B

Candidate miRNA in *npas2* containing rs1811399 A allele

Disruption of the candidate miRNA by the **C** allele of rs1811399

C

Mirtron-like structures

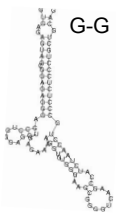
G-C

C-G

rs885953 - rs885747 G-C haplotype in *per1* candidate mirtron

rs885953 - rs885747 C-G haplotype in *per1* candidate mirtron

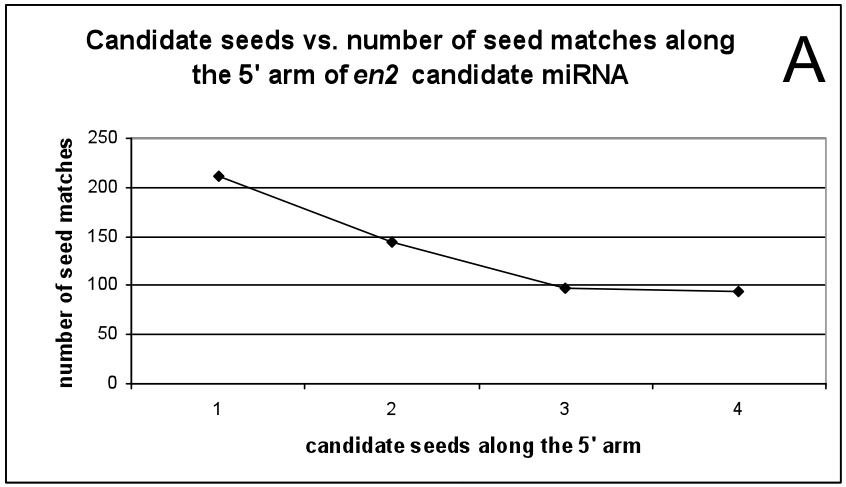
Disrupted hairpins

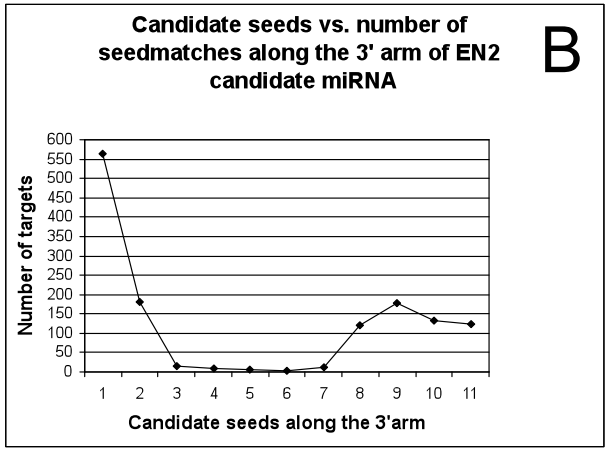


G-G

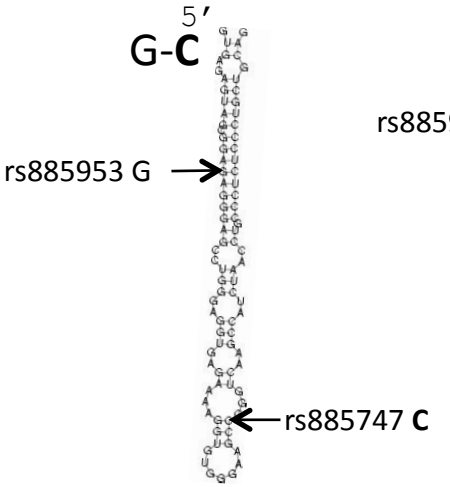
C-C

G-G and C-C haplotypes Precludes candidate mirtron formation

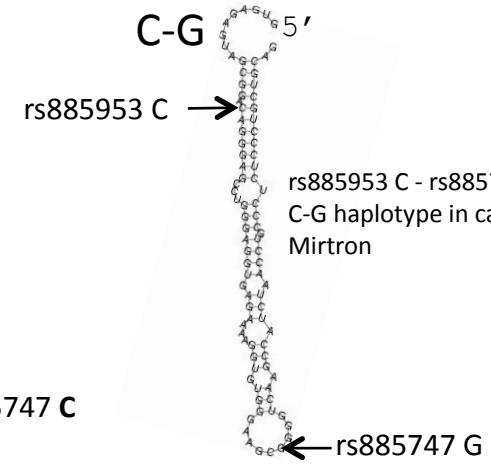




Mirtron-like structures

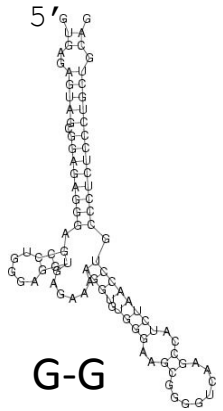


rs885953 G - rs885747 **C**
G-**C** haplotype in candidate Mirtron



rs885953 C - rs885747 G
C-G haplotype in candidate Mirtron

Disrupted hairpins



G-G and C-**C** haplotypes
Precludes candidate Mirtron formation