

1 **Association study of schizophrenia with variants in miR-137 binding sites**

2

3 Running title:

4 **MiR-137 binding sites and schizophrenia**

5

6 ^{1,2}David Curtis, ^{1,3,4}Warren Emmett

7 1. UCL Genetics Institute, University College London

8 2. Centre for Psychiatry, Barts and the London School of Medicine and Dentistry

9 3. The Francis Crick Institute

10 4. Department of Molecular Neuroscience, UCL Institute of Neurology

11

12 Correspondence: David Curtis, Darwin Building, Gower Street, London WC1E 6BT, UK.

13 Phone 00 44 20 7679 2212. Fax 00 44 20 7791 5201. d.curtis@ucl.ac.uk

14

15 **Conflict of interest**

16 The authors declare they have no conflict of interest.

17

18

19 **Abstract**

20 There is strong cumulative evidence for the involvement of miR-137 and its targets in the
21 aetiology of schizophrenia. Here we test whether variants, especially rare variants, in miR-
22 137 binding sites are associated with schizophrenia in an exome-sequenced sample of 4225
23 cases and 5834 controls. A weighted burden test using 372 variants was significant at
24 $p=0.024$. The sample size is too small to implicate individual variants or genes but overall
25 this finding provides further support for the hypothesis that disruption of miR-137 binding
26 sites can increase the risk of schizophrenia, perhaps by leading to over-expression of the
27 target gene. These findings could be followed up by genotyping these variants in larger
28 samples and by experimentally testing whether they do indeed effect expression. When
29 carrying out exome sequencing it is important to include UTRs so that disruption of
30 microRNA bindings sites can be detected.

31

32 **Keywords**

33 miR-137, schizophrenia, microRNA, exome, UTR, binding site

34

35 **Introduction**

36 MicroRNAs can bind to specific sites in the 3' UTR of target genes and modulate their
37 expression. As discussed recently (Olde Loohuis et al. 2017) markers for both the gene for
38 miR-137 and for the genes it targets demonstrate association with schizophrenia (Kwon et
39 al. 2013; Schizophrenia Working Group of the Psychiatric Genomics Consortium 2014).
40 Experimentally, reducing or increasing miR-137 expression in rat hippocampal neurons
41 identifies sets of regulated genes involved in neurodevelopmental processes and neuronal
42 maturation (Olde Loohuis et al. 2017). Here, we investigate whether variants in the binding
43 sites of miR-137 are more commonly found in exome-sequenced schizophrenia cases than
44 controls.

45

46 **Material and methods**

47 The data analysed consisted of whole exome sequence variants downloaded from dbGaP
48 from a Swedish schizophrenia association study containing 4968 cases and 6245 controls
49 (Genovese et al. 2016). As described elsewhere (Curtis 2017), the dataset was subjected to
50 QC procedures including the removal of subjects who appeared to have a substantial
51 Finnish component to their ancestry to leave a sample of 4225 cases and 5834 controls. As
52 reported in the original paper, the subjects were sequenced in 12 waves. For all but the first
53 wave the Agilent SureSelect Human All Exon v.2 Kit was used for the hybrid-capture
54 procedure whereas for the first wave an earlier version was used. Predicted binding sites for
55 miR-137 were obtained from microRNA.org (Betel et al. 2010). Excluding the Y
56 chromosome, there were 8145 predicted binding sites, of which 1139 were covered by
57 SureSelect, Variants in these regions were extracted and analysed by SCOREASSOC,
58 which performs a weighted burden test such that very rare variants are weighted 10 times
59 higher than a common variant with MAF=0.5 (Curtis 2012). Each subject receives a score
60 consisting of the weighted sum of the variant alleles possessed by that subject at all sites
61 and a t test is used to compare the scores between cases and controls. The predicted
62 effects of variants were annotated using VEP (McLaren et al. 2016).

63

64 **Results**

65 In the regions covered, 372 variants were found which passed QC procedures. On average,
66 scores were higher for cases than controls ($t=2.3$, 10057 df, $p=0.024$), indicating that
67 subjects with schizophrenia are more likely to have variants, especially rare variants, in miR-
68 137 binding sites than controls. The results are presented in full in Supplementary Table 1.
69 Many variants only occurred in one or two subjects and for others there were mostly only
70 small differences in frequencies between cases and controls so the effect could not be
71 assigned to specific variants. However it may be worth noting the results for two variants. A
72 variant at 10:106027165, rs7589, was heterozygous in 5 out of 4224 cases and none out of
73 5833 controls. It lies in the 3' UTR of some transcripts of GSTO1. A variant at 19:58773876
74 was heterozygous in 17 out of 4205 cases and 13 out of 5814 controls, OR=1.8. It lies in the
75 3'UTR of some transcripts of ZNF544.

76

77 **Discussion**

78 Our results provide some further support for the hypothesis that abnormalities in miR-137
79 functionality could be involved in the aetiology of schizophrenia. Disruption of a microRNA
80 binding site could lead to increased gene expression, providing a mechanism for a dominant,
81 gain of function effect. With the sample sizes used it is not possible to assign risk to
82 individual variants and the two which we refer to are too rare to have been imputed in large
83 GWAS samples (Schizophrenia Working Group of the Psychiatric Genomics Consortium
84 2014). All five subjects with the variant in the GSTO1 site have schizophrenia. The gene
85 codes for an omega class glutathione S-transferase and a study showed that patients with
86 schizophrenia have reduced glutathione levels in cerebrospinal fluid (Do et al. 2000) but an
87 association study of schizophrenia with markers for GSTO1 and other glutathione related
88 genes was negative (Matsuzawa et al. 2009). ZNF544 belongs to the C2H2-type zinc-finger
89 family and is involved in gene transcription. In a genome-wide association study of ADHD
90 traits an intronic SNP, rs260461, was significant at $p=10^{-5}$ (Lasky-Su et al. 2008) and in a

91 methylome -wide study a CpG island near ZNF544 was found to be hyper-methylated at
92 birth in subjects with a high trajectory for ADHD symptoms (Walton et al. 2017).

93

94 In order to follow up these findings, genotyping could be carried out in the larger datasets
95 which have been recruited for association studies. The variants could be introduced into cell
96 cultures in order to determine whether they do indeed affect gene expression and if so in
97 which tissues. Only a minority of binding sites were covered by the capture procedure and
98 we recommend that exome sequencing studies should routinely include UTRs so that
99 variants potentially affecting microRNA binding can be detected.

100

101 **Acknowledgments**

102 The datasets used for the analysis described in this manuscript were obtained from dbGaP at
103 <http://www.ncbi.nlm.nih.gov/gap> through dbGaP accession number *phs000473.v2.p2*. Samples
104 used for data analysis were provided by the Swedish Cohort Collection supported by the NIMH grant
105 R01MH077139, the Sylvan C. Herman Foundation, the Stanley Medical Research Institute and The
106 Swedish Research Council (grants 2009-4959 and 2011-4659). Support for the exome sequencing
107 was provided by the NIMH Grand Opportunity grant RCMH089905, the Sylvan C. Herman
108 Foundation, a grant from the Stanley Medical Research Institute and multiple gifts to the Stanley
109 Center for Psychiatric Research at the Broad Institute of MIT and Harvard. The work of W.E. is
110 supported by the Wellcome Trust (103760/Z/14/Z) and the Francis Crick Institute, which receives its
111 core funding from Cancer Research UK (FC001002), the UK Medical Research Council (FC001002),
112 and the Wellcome Trust (FC001002);

113

114 **Conflict of interest**

115 The authors declare they have no conflict of interest.

116

117 **References**

118 Betel, D. et al., 2010. Comprehensive modeling of microRNA targets predicts functional non-

- 119 conserved and non-canonical sites. *Genome biology*, 11(8), p.R90.
- 120 Curtis, D., 2012. A rapid method for combined analysis of common and rare variants at the
121 level of a region, gene, or pathway. *Adv Appl Bioinform Chem*, 5, p.1.
- 122 Curtis, D., 2017. Construction of an exome-wide risk score for schizophrenia based on a
123 weighted burden test. *bioRxiv*, 145961.
- 124 Do, K.Q. et al., 2000. Schizophrenia: glutathione deficit in cerebrospinal fluid and prefrontal
125 cortex in vivo. *The European journal of neuroscience*, 12(10), pp.3721–8.
- 126 Genovese, G. et al., 2016. Increased burden of ultra-rare protein-altering variants among
127 4,877 individuals with schizophrenia. *Nature Neuroscience*, 19(11), pp.1433–1441.
- 128 Kwon, E., Wang, W. & Tsai, L.-H., 2013. Validation of schizophrenia-associated genes
129 CSMD1, C10orf26, CACNA1C and TCF4 as miR-137 targets. *Molecular psychiatry*,
130 18(1), pp.11–2.
- 131 Lasky-Su, J. et al., 2008. Genome-wide association scan of quantitative traits for attention
132 deficit hyperactivity disorder identifies novel associations and confirms candidate gene
133 associations. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*,
134 147B(8), pp.1345–1354.
- 135 Matsuzawa, D. et al., 2009. Association study between the genetic polymorphisms of
136 glutathione-related enzymes and schizophrenia in a Japanese population. *American
137 Journal of Medical Genetics Part B: Neuropsychiatric Genetics*, 150B(1), pp.86–94.
- 138 McLaren, W. et al., 2016. The Ensembl Variant Effect Predictor. *Genome biology*, 17(1),
139 p.122.
- 140 Olde Loohuis, N.F.M. et al., 2017. The schizophrenia risk gene MIR137 acts as a
141 hippocampal gene network node orchestrating the expression of genes relevant to
142 nervous system development and function. *Progress in Neuro-Psychopharmacology
143 and Biological Psychiatry*, 73, pp.109–118.
- 144 Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014. Biological
145 insights from 108 schizophrenia-associated genetic loci.
- 146 Walton, E. et al., 2017. Epigenetic profiling of ADHD symptoms trajectories: a prospective,

147 methylome-wide study. *Molecular Psychiatry*, 22(2), pp.250–256.

148