ANALYSIS OF A-TO-I RNA EDITING IN SCHIZOPHRENIA RISK GENE MIR137

Harish Pudukodu

Honors Thesis ENHS
Department of Environmental Sciences and Engineering
Gillings School of Global Public Health
The University of North Carolina at Chapel Hill
December 7th, 2016

Approved:

[Signature]

Dr. James J. Crowley, Department of Genetics

[Signature]

Dr. Louise M. Ball, Department of Environmental Sciences and Engineering

[Signature]

Dr. Shuntai Zhou, Department of Biochemistry and Biophysics
ABSTRACT

Schizophrenia is a complex psychiatric disorder that likely emerges as a result of interacting genetic and environmental risk factors. A prominent hypothesis for the etiology of schizophrenia is that schizophrenia emerges as a result of pre- and/or perinatal viral infection-induced inflammation in genetically predisposed individuals. This gene-environment interaction may be mediated by genetic regulators such as microRNA 137 (miR137), a regulatory RNA with a set of related SNPs that is highly associated with schizophrenia, and epigenetic phenomena such as RNA editing, which has been previously associated with schizophrenia and infection. Here I interrogate this gene-environment interaction by quantifying RNA editing of miR137, assessing the functional impact of this editing, and predicting its effects on schizophrenia risk in conjunction with a risk SNP. High-throughput RNA sequencing coupled to the PrimerID methodology revealed that miR137 is edited at a statistically significant level (p < 0.05) in both fetal human brain and adult mouse brain. Site-directed mutagenesis and luciferase assays demonstrated that A-to-I modification of certain pertinent bases in miR137 impart a large decrease in affinity to a target sequence. Finally, mathematical modeling and simulations based on empirical findings suggest that this RNA editing and a MIR137 schizophrenia-risk SNP significantly modulate inflammation-based schizophrenia risk burden (p < 2 x 10^{-16}). These findings demonstrate that infection, the MIR137 gene, and RNA editing likely interact to promote schizophrenia risk, and thus this study serves to elucidate the cooperative influences of genetic and environmental risk factors on schizophrenia.
ACKNOWLEDGEMENTS

I would like to thank Dr. Shuntai Zhou, Bryant Su, Dr. Kensuke Sakamoto, Dr. Ron Swanstrom, Dr. Louise Ball, and Dr. James Crowley for their contributions to this work. This research was supported by K01MH094406 (PI: Dr. James Crowley). This research was funded in part by a Summer Undergraduate Research Fellowship from the Office of Undergraduate Research at the University of North Carolina at Chapel Hill.

TABLE OF CONTENTS

Abstract .................................................................................................................................................. 2
Acknowledgements .................................................................................................................................. 3
Table of Contents .................................................................................................................................... 3
List of Tables and Figures ....................................................................................................................... 4
Introduction ........................................................................................................................................... 5
Materials and Methods .......................................................................................................................... 7
  Computational RNA Analyses .............................................................................................................. 7
  RNA Preparation and Sequencing ....................................................................................................... 8
  Mutagenesis, Luciferase, and RT-PCR Experiments ......................................................................... 9
  Mathematical Modeling of MIR137 System Dynamics .................................................................... 10
  Statistical Analyses ............................................................................................................................ 11
Results .................................................................................................................................................... 11
  Pre-miR137 is a predicted target of hADAR1 .................................................................................. 11
Adult mouse and fetal human brains exhibit statistically significant levels of A-to-I MIR137 editing
A-to-I editing of MIR137 significantly changes target affinity
MIR137 SNP rs1625579 modulates inflammatory burden in a manner significantly dependent on miR137 A-to-I editing
Discussion and Conclusions
MIR137 is subjected to A-to-I RNA editing
A-to-I RNA editing of MIR137 changes target affinity
Schizophrenia risk burden from infection is modulated by a SNP in MIR137 and miR137 editing rate
References
Tables and Figures
Appendix
Supplementary Tables and Figures
Supplementary Computational Methods

LIST OF TABLES AND FIGURES
Tables and Figures
Figure 1
Table 1
Table 1a
INTRODUCTION

The Psychiatric Genetics Consortium released a landmark paper in 2014, which identified numerous risk loci that are associated with schizophrenia[1]. One of the most highly associated haplotypes is positioned around the MIR137 locus, which codes for a microRNA that targets other schizophrenia-associated genes such as CACNA1C and TCF4[1-42]. This discovery has prompted considerable functional work to determine the effects of various key MIR137 single nucleotide polymorphisms (SNPs) on the expression of miR137 and downstream targets[2, 3, 5, 43]. This work has led to the identification of gene-expression effects resulting from MIR137 SNPs, such as the downregulation of miR137 and the upregulation of its schizophrenia-
associated targets[2, 3, 5, 43]. This suggests that MIR137 plays a central role in the etiology of schizophrenia.

It has been discovered in recent years that microRNAs are subject to A-to-I editing by the various forms of the adenosine deaminase acting on RNA (ADAR) enzyme[44-52]. This phenomenon, which causes specific adenosine bases in double stranded RNA to be converted to inosine, has been determined to have functional impacts on microRNA biogenesis and subsequent targeting, even leading to altered sets of target genes for a given microRNA[44-51]. In prior high-throughput RNA sequencing experiments, our group has identified possible non-background levels of A-to-I editing in the miR137 sequence (data not shown). I hypothesize that such A-to-I editing of miR137 and/or the interaction between the ADAR enzymes and miR137 may be involved in the etiology of schizophrenia.

The phenomenon of RNA editing has actually been associated with schizophrenia in a number of ways[53-55]. For example, the age of onset of schizophrenia is associated with expression levels of the ADAR1-A isoform and expression levels of multiple ADAR isoforms are significantly altered in dorsolateral prefrontal cortex tissue samples from patients with schizophrenia[53]. Of note is the fact that ADAR1-A is the interferon-inducible isoform of ADAR1, which offers an interesting connection to the viral hypothesis of schizophrenia[56-65]. This hypothesis suggests that schizophrenia may develop in genetically predisposed individuals following an intense immune response from the mother, likely triggered by a viral infection, while the individual is still in the womb[56-65]. The altered expression of ADAR1-A in the brain of patients with schizophrenia, coupled to its interferon-inducible (and thus infection-inducible), lends support to the viral hypothesis of schizophrenia and simultaneously offers an etiological clue.
Apart from ADAR1-A’s standard RNA editing role in the nucleus, it also acts as an inducing agent for Dicer, an important enzyme in the biogenesis pathway of miRNAs[66, 67]. More specifically, ADAR1-A is capable of increasing Dicer’s $V_{\text{max}}$ four-fold, and thus enables Dicer to produce mature miRNAs at a faster rate[66]. Interestingly, it seems that ADAR1-A preferentially performs the latter cytoplasmic function because of its tendency to localize to the cytoplasm in greater proportion compared to other ADAR forms[68].

Given the facts that miR137 may experience A-to-I RNA editing, miR137 almost certainly interacts with ADAR1-A via Dicer, and ADAR1-A is induced by interferon (which is released during viral infection), I hypothesize that infection-mediated interferon release and schizophrenia-associated SNPs in $MIR137$ interact with ADAR1-A via A-to-I editing of miR137 and Dicer-mediated biogenesis to influence schizophrenia risk. This hypothesis can be parsed into three specific aims: (1) to determine if miR137 is subjected to A-to-I RNA editing in fetal human and adult mouse brain, (2) to determine if such editing has any functional consequences in vitro, and (3) to determine if a schizophrenia-associated $MIR137$ SNP modulates marginal schizophrenia risk changes due to infection-mediated interferon release and subsequent ADAR1-A induction. I tested the first specific aim using a high-throughput RNA sequencing methodology called PrimerID[69, 70], I tested the second specific aim using site-directed mutagenesis and luciferase assay data, and I tested the third specific aim using a predictive modeling approach implemented in MATLAB.

**Materials and Methods**

*Computational RNA Analyses*

Prediction of A-to-I editing likelihood by hADAR1 was performed using InosinePredict, which use previously described methods to predict RNA editing activity for dsRNA
In short, the algorithm has been trained on hADARI/dsRNA reaction data to assign editing likelihood scores at adenosine bases in a custom dsRNA sequence based on nearby ribonucleotide identities[71]. The sequence for the human precursor miR137 molecule was found using the UCSC Genome Browser[72] and this sequence was entered into the InosinePredict web application to produce A-to-I edit prediction data.

Secondary structure prediction for pre-miR137 was performed using RNAstructure, which utilizes thermodynamic models and nearest neighbor parameters to predict RNA secondary structure[73]. Visual output from this software was analyzed systematically to identify plausible and implausible sites of A-to-I RNA editing. Implausible sites were deemed to be all unpaired bases in the miR137 secondary structure and plausible sites were deemed to be paired adenosine bases in the miR137 secondary structure.

**RNA Preparation and Sequencing**

We used a previously described protocol to construct Primer ID MiSeq sequencing libraries[70]. In brief, cDNA was made using primers containing a block of random nucleotides (Primer ID). The cDNA primer sequences were 5’-

GTGACTGGAGTTCCAGACGTGTGCTCTTCCGATCTNNNNNNNNNTAGCTGCCGCTG
GTACTCTCCTCGA -3’ and 5’-

GTGACTGGAGTTCCAGACGTGTGCTCTTCCGATCTNNNNNNNNNTACCTGCCGCTG
GTACTCTCCTCGA -3’ for human and mouse sample, respectively. After purification, cDNA was amplified using two rounds of PCR and MiSeq indexed primers were incorporated. The primer sequence of the first round PCR forward primer was 5’-

GCCTCCCTCGGCACATCAGAGATGTGTATAAGAAGACAGNNNCCCTCTGACTCTCTTC
GGTGA-3’ for both human and mouse samples. Purified and pooled libraries were sequenced using Illumina MiSeq 150 bp pair-end sequencing. We used the Illumina bcl2fastq pipeline (v.1.8.4) for the initial processing of sequencing data, followed by the previously described Primer ID consensus pipeline[70] to create template consensus sequences (available at https://github.com/SwanstromLab/PID). We analyzed mutations at each nucleotide positions of the template consensus sequences.

**Mutagenesis, Luciferase, and RT-PCR Experiments**

Plasmid constructs for each combination of edit were created using site-directed mutagenesis and standard cloning procedures. For the luciferase assays, Hek293 cells were split into a 96 well plate with 3*10^4 cells per well with 100 µL of media per well. Approximately 18 hours after splitting, a mixture of pLuc137TRGTx4, Renilla, optiMEM, Lipofectamine 2000, and the desired mutagenic sequence (or the wild variation or the control pCMVPL) was made and incubated for 20-30 minutes. After incubation, these mixtures were added to each of the wells. This left triplicate wells for each transfected DNA sequence. After 2 days of growth, the media was dumped, the cells were washed once, and then lysed passively for 15 minutes. Finally, a dual-luciferase assay system was used to measure both Firefly and Renilla Luciferase levels.

For the quantitative RT-PCR assays, Hek293 were split into 3 12 well plates with 2*10^5 cells per well and 1 mL of media per well. 28 hours later, the mutagenic and control plasmids were mixed with lipofectamine 2000 and incubated for 20-30 minutes. These mixtures were randomly added to wells of each of the plates, leaving triplicate wells for each transfected DNA sequence again. After 3 days, Trizol Reagent was added to each of the wells and RNA was purified using the Thermofish Trizol Reagent protocol. cDNA was then created using the
purified RNA using a NCode™ VILO™ miRNA cDNA Synthesis Kit. Finally, some of these cDNA was used to run and RT-PCR using a SYBR Green Master Mix.

**Mathematical Modeling of MIR137 System Dynamics**

The temporal dynamics of interferon (IFN), ADAR1-A, miR137 (MIR), TCF4 mRNA (TM) and protein (TP), dendritic cells (DC), and inflammation burden (IB) in response to an infection were modeled using a system of coupled ordinary differential equations (ODEs) implemented in MATLAB. This modeling system encapsulated the interactions among these relevant physiological entities, as they relate to inflammation, through standard biochemical kinetics paradigms such as mass-action kinetics and Michaelis-Menten kinetics. Infection was modeled as entrance of pathogen, which incorporated a custom function to specify square-wave input.

The ODEs used in this system are as follows:

\[
\frac{d\text{Pathogen}}{dt} = \text{SquareWave}(t, \text{pathogen period}, \text{pathogen length}, \text{pathogen high}, \text{pathogen low}) - (\text{pathogen clearance rate}) \times \text{Pathogen}
\]

\[
\frac{d\text{IFN}}{dt} = (\text{IFN release parameter}) \times \text{Pathogen} \times \text{DC} - (\text{IFN degradation rate}) \times \text{IFN}
\]

\[
\frac{d\text{ADAR1}}{dt} = (\text{ADAR1 - A synthesis parameter}) \times \text{IFN} - (\text{ADAR1 - A degradation rate}) \times \text{ADAR1}
\]

\[
\frac{d\text{MIR}}{dt} = (\text{basal miR137 transcription rate}) \times \left(1 - 0.3 \times (\text{binary SNP presence})\right) \times \left(1 + 2.98 \times \left(\frac{\text{ADAR1}}{20}\right)\right) - (\text{miR137 degradation rate}) \times \text{MIR}
\]

\[
E = \text{Proportion of MIR that is edited} = (\text{basal editing rate}) \times \left(1 + \frac{0.6 \times \text{ADAR1}}{\text{ADAR1B}}\right)
\]

\[
\frac{d\text{TM}}{dt} = (\text{basal TM transcription rate}) - (\text{TM degradation rate}) \times \text{TM} - 0.16 \times E \times \text{MIR} \times \text{TM} \times (\text{Targeting V}_{\text{max}}) \times \left(1 - E\right) \times \text{MIR} \times \text{TM} \times (\text{Targeting V}_{\text{max}}) \times \left(\frac{\text{Targeting K}_{\text{m}}}{\text{Targeting K}_{\text{m}} + \text{TM}}\right)
\]

\[
\frac{d\text{TP}}{dt} = \frac{\text{TM} \times (\text{Protein V}_{\text{max}})}{(\text{Protein K}_{\text{m}} + \text{TM})} - \frac{\text{TP} \times (\text{Degradation V}_{\text{max}})}{(\text{Degradation K}_{\text{m}} + \text{TP})}
\]

\[
\frac{d\text{DC}}{dt} = (\text{basal DC production rate}) + (\text{TCF4 stimulation factor}) \times \text{TP} - (\text{DC degradation rate}) \times \text{DC}
\]

\[
\frac{d\text{IB}}{dt} = (\text{inflammation factor}) \times \text{IFN}
\]

Sensitivities to parameters were calculated via one-by-one variation of parameter values and subsequent comparisons of final IB.
Statistical Analyses

Edit frequency significance testing was performed in Microsoft Excel, haplotype frequency analyses were performed in MATLAB, luciferase and RT-PCR data analyses were performed in JMP and R, and statistical analyses of the mathematical modeling results were performed in R. Edit and haplotype frequency analyses employed the Poisson distribution to model probabilities of individual edit events, while the remaining statistical analyses utilized standard Gaussian procedures.

RESULTS

Pre-miR137 is a predicted target of hADAR1

Based on consensus sequence data derived from preliminary, exploratory high-throughput RNAseq experiments performed by our group, we hypothesized that the precursor miR137 molecule would experience A-to-I RNA editing at three sites: Bases 29, 69, and 74 (Figure 1). To initially assess the biochemical and physical plausibility of this potential editing activity, we utilized a published hADAR1 activity prediction tool known as InosinePredict[71]. InosinePredict uses predictive algorithms developed from hADAR1/dsRNA reaction data to assign a score to each base of a queried dsRNA molecule, reflecting the likelihood of A-to-I editing by hADAR1[71]. We analyzed the likelihood of editing of MIR137 using the pre-miR137 sequence fed into the InosinePredict tool.

The InosinePredict algorithm predicted that many, but not all, of the adenosine bases in the precursor miR137 molecule would be edited at a significant level by hADAR1 (Figure 1a). Of note, the three initially hypothesized editing sites in pre-miR137 were among the high likelihood hADAR1 target sites in the sequence (Figure 1a).
Because the InosinePredict tool assumes a perfectly paired dsRNA molecule, the predicted A-to-I editing sites were assessed for structural feasibility of editing. To this end, we used a published web server-based software known as RNAstructure to predict the secondary structure of pre-miR137[73]. Secondary structure prediction via RNAstructure revealed that some of the predicted hADAR1 sites are not biologically plausible A-to-I editing sites because those bases are not paired in a double-stranded manner (Figure 1b). These implausible sites of predicted editing were ignored in downstream analyses.

**Adult mouse and fetal human brains exhibit statistically significant levels of A-to-I MIR137 editing**

Following our computational interrogation of the plausibility and likelihood of A-to-I editing in pre-miR137, we sought to determine if the phenomenon is truly exhibited in brain tissue from adult mice and fetal humans. To accomplish this, we utilized a published method called PrimerID[70] in conjunction with high-throughput RNA sequencing. PrimerID enables one to parse out the contribution to nucleotide-level read variability from original biological variation (as opposed to errors introduced from library preparation and sequencing steps) using randomized transcript barcodes[70], and thus this system allowed us to determine the true frequency of A-to-I editing in our samples.

Whole RNA samples from adult mouse brain and fetal human brain tissues were used for library preparation and subsequent sequencing. Read results from high-throughput RNAseq were aligned to mouse and human genomes (depending on the sample origin) and consensus reads were constructed. Base-level changes were statistically analyzed for significance using a Poisson distribution to determine sites in MIR137 that experienced significant A-to-I editing. A multitude
of sites, many of which corresponded directly with predicted locations, were identified as significant A-to-I editing sites in both adult mouse brain and fetal human brain (Supplementary Table 1). Among these sites, the highest rates of editing were determined to be at the hypothesized editing sites (positions 29, 69, and 74) in both mouse and human (Table 1a).

To interrogate the potential interdependence of A-to-I editing at the high frequency sites in fetal human and adult mouse, haplotypes and associated frequencies were determined from the sequence read data. A MATLAB script was written to 1) predict the frequencies of edit combinations based on the assumption of independent editing at each site, 2) calculate the true frequencies of edit combinations from haplotype data, and 3) calculate fold differences between predicted and observed frequencies. This analysis revealed that RNA editing at each high frequency site was only loosely coupled in adult mouse brain and was uncoupled in fetal human brain (Table 1b).

Of the significant A-to-I editing sites identified from the analysis of the RNAseq data, we were interested in identifying the sites that are conserved across mouse and human for enriched editing frequency. To this end, editing frequencies were normalized across the two test species and compared. Of the twelve total significant editing sites identified across both species, only editing at positions 69 and 74 (two of the initially hypothesized editing sites) were conserved and enriched in both fetal human brain and adult mouse brain (Figure 2). Both of these positions in the precursor miR137 molecule are present in the mature miRNA.

**A-to-I editing of MIR137 significantly changes target affinity**

To determine possible functional consequences of A-to-I editing of pre-miR137 on mature miR137’s targeting capacity, a series of mutagenesis experiments were performed. All
combinations of the three edits (WT, a29g, a69g, a74g, a29g+a69g, a29g+a74g, a69g+a74g, all three a-to-g) were made using site-directed mutagenesis in otherwise identical miR137 expression constructs. Subsequently, a luciferase assay system was designed to determine binding activity by each of the miR137 forms. A reporter plasmid was constructed that included a luciferase expression cassette coupled to multiple repeats of the antisense sequence of mature miR137.

Each of the miR137 constructs (along with a positive and negative control) was tested using this luciferase assay system, where miRNA binding to the antisense repeats would lead to reduced luciferase activity. All of the miR137 edited forms, except for the single a74g form, led to increased luciferase activity compared to the WT form of miR137 (Figure 3). This outcome could be caused either by correlated differences in expression of miR137 from each construct or by truly biological differences in targeting behavior caused by RNA editing.

To test the possibility that the luciferase data could be explained by expression differences, each of the luciferase experiments were coupled to RT-PCR experiments to determine expression levels of each construct. Regression of the luciferase data by the RT-PCR data revealed that the luciferase data were not significantly associated with the RT-PCR expression data (Supplementary Figure 1). As such, the increased luciferase activities observed for the majority of the edited forms of miR137 are not due to differences in expression levels of each construct. This suggests that the modification of A-to-I in select positions of miR137 imparts altered targeting activity to the mature microRNA.
**MIR137 SNP rs1625579 modulates inflammatory burden in a manner significantly dependent on miR137 A-to-I editing**

To investigate the role of miR137 A-to-I editing in the potential gene-environment interaction system posited by the viral hypothesis of schizophrenia, a mathematical modeling framework was constructed (Figure 4) and simulated using MATLAB. This modeling framework consisted of a set of coupled nonlinear ordinary differential equations (ODEs), which were solved using MATLAB’s ode45 ODE solver. The parameterization of the model allowed for the modification of production/elimination rates, binding affinities, the basal editing rate, and presence of a single well-studied MIR137 SNP: rs1625579, which has been shown to reduce miR137 expression[2, 43].

Schizophrenia risk was modeled as inflammatory burden, which was calculated as the time-integral of interferon (IFN) level. This methodology is supported by numerous findings that indicate that inflammation and dysregulated general immune response are likely etiological factors for schizophrenia[56, 58, 61-65]. IFN release was modeled using brain dendritic cells, which are targets of TCF4-mediated induction[25]. These inclusions, along with the other model structures, resulted in biologically plausible dynamics for the various species of interest (e.g. miR137 and TCF4 protein levels) (Supplementary Figure 2).

To test the influence of the SNP on inflammatory burden (and thus schizophrenia risk), the model was simulated with and without the presence of the risk SNP under a periodic infection spike input (Figure 5). These simulations revealed that the presence of the SNP causes larger marginal changes in inflammatory burden in response to infection, which accumulate over time (Figure 5). To assess whether this influence of the risk SNP on inflammatory burden significantly depends on RNA editing, the model was simulated with and without the presence of
the risk SNP and with different basal RNA editing rates under a near-continuous infection input. For each RNA editing rate tested, an inflammatory burden fold change was calculated (inflammatory burden with the SNP over the inflammatory burden without the SNP). These fold changes were analyzed via linear regression against the various RNA editing rate values, and this analysis revealed that the basal RNA editing rate significantly affects the inflammatory burden fold change due to the SNP ($p < 2 \times 10^{-16}$) (Figure 6). This indicates that the phenomenon of miR137 A-to-I RNA editing plays a significant role in schizophrenia risk upon infection.

**DISCUSSION AND CONCLUSIONS**

*MIR137 is subjected to A-to-I RNA editing*

To test the hypothesis that miR137 is subjected to A-to-I RNA editing by an ADAR enzyme (likely ADAR1 due to its prevalence in the brain[44, 46, 47, 49, 66, 71, 74, 75]), the precursor to mature miR137 (pre-miR137) was analyzed using predictive bioinformatics and the editing rate was estimated using high throughput RNA-sequencing and PrimerID[69, 70]. The computational analyses suggested that certain bases of pre-miR137 were likely to be subjected to A-to-I RNA editing by ADAR1 at a significant rate (Figure 1) and the RNA-sequencing results recapitulated these predictions in both fetal human brain and adult mouse brain (Table 1). Upon further analysis of these RNA-sequencing data, it was revealed that the sites with the highest relative editing rates in adult mouse brain are conserved in fetal human brain and that these sites are present in the final mature miR137 (Figure 2), which suggests that these conserved edits have functional consequences. All of this evidence taken together leads to the conclusion that there is a high likelihood that miR137 is subjected to A-to-I editing *in vivo*.

It is important to note that the raw editing rates in the fetal human samples are far lower than those in the adult mouse samples (Table 1). This is likely because RNA editing rates
increase with neural development \cite{76} and the fetal human brain tissue used in this study was inferior in quality to the adult mouse brain tissue used in this study. Given that primates typically exhibit higher RNA editing rates than nonprimates \cite{77}, one would expect that high quality adult human brain tissue would present with higher rates of miR137 A-to-I editing. Further molecular transcriptomic work must be done to test this prediction and determine the true rates of miR137 A-to-I editing in adult human brains.

**A-to-I RNA editing of MIR137 changes target affinity**

Given the likelihood of *in vivo* A-to-I editing of miR137, we were interested in determining whether such editing would have functional consequences with respect to miR137 targeting affinity. To address this question, miR137 constructs with point mutations (induced by site directed mutagenesis) corresponding to edit combinations were assayed for antisense target binding activity in a custom luciferase reporter-based system. This set of experiments revealed that many of the A-to-I (A-to-G) edit combinations led to lower binding affinity between miR137 and the antisense target (**Figure 3**).

The results of the above assay suggest that miR137 A-to-I editing has meaningful functional consequences *in vivo* in the form of altered target affinity and subsequent changes in gene expression networks, but further research must be done to confirm this finding and elucidate its mechanistic nature. Given that this assay employed antisense repeats as pseudo-targets, it may still be plausible that miR137 targeting capacity may be unchanged *in vivo* with respect to real target sequences. To address this concern, future work must employ real target sequences (derived from real target mRNAs) to better assess binding affinity changes. Additionally, future *in situ* investigations may be useful to determine the precise mechanistic
consequences of miR137 RNA editing (at the biochemical and cellular levels). That being said, the presence of this luciferase reporter assay-based *in vitro* data coupled to the previously discussed evidence for editing conservation suggests that there is a true *in vivo* consequence to the miR137 A-to-I editing identified in this study.

*Schizophrenia risk burden from infection is modulated by a SNP in MIR137 and miR137 editing rate*

Because of the IFN-inducible nature of ADAR1[44-50, 66, 71, 77], the association between RNA editing and ADAR levels with schizophrenia[53, 70], the strong association between MIR137 SNPs and schizophrenia, and the likely interaction between ADARs and miR137 discovered in this study, it is likely that miR137 (and its associated genetic variation) and RNA editing would play central roles in the development of schizophrenia risk from viral infection, as posited by the viral hypothesis of schizophrenia. Furthermore, the role of RNA editing in schizophrenia etiology is likely two-fold because of ADAR1’s ability to both edit double stranded RNA molecules and promote Dicer activity[66]. To test the prediction that miR137 A-to-I editing and MIR137 genetic variation (implemented here as the rs1625579 SNP) interact to influence schizophrenia risk upon viral infection, a mathematical modeling framework based on literature-derived biochemical kinetics and genetics data was constructed and simulated. These computational experiments confirmed the aforementioned prediction: presence of the rs1625579 SNP considerably modified inflammatory burden (*Figure 5*) and miR137 editing significantly influenced this SNP effect (*Figure 6*).

These findings lend credence to the viral hypothesis of schizophrenia and offer a plausible partial mechanism for the etiology of schizophrenia, but additional empirical work
must be done to solidify the connections between viral infections, miR137 A-to-I editing, and schizophrenia. Schizophrenia is a highly complicated and polygenic disease that almost certainly arises from varied gene-environment interactions[1]. Given this inherent mass of complexity, it is unlikely that any one study will provide more than a fraction of the information necessary to understand schizophrenia. That being said, the findings in the present study related to the influence of gene-environment interactions on schizophrenia risk allow for the development of a basic schizophrenia risk framework, which would assist public health specialists and healthcare practitioners as they work to prevent and treat this debilitating disorder (Figure 7). The present study suggests that schizophrenia risk arises from genetic risk elements, environmental risk elements, and the interaction among those genetic and environmental factors (Figure 7). Further integrative and interdisciplinary work is crucial for the development of a quantitatively precise model of schizophrenia risk, but the conceptual model posited here would likely be a useful guiding framework for subsequent model development.

REFERENCES

27. Olde Loohuis, N.F., et al., The schizophrenia risk gene MIR137 acts as a hippocampal gene network node orchestrating the expression of genes relevant to nervous system development and function. Prog Neuropsychopharmacol Biol Psychiatry, 2016.
45. Cui, Y., T. Huang, and X. Zhang, RNA editing of microRNA prevents RNA-induced silencing complex recognition of target mRNA. Open Biol, 2015. 5(12).
Tables and Figures

Figure 1

The precursor to mature miR137 is predicted to undergo A-to-I RNA editing by ADAR1 at significant levels. (A) This graph displays the output from InosinePredict, which represents relative ADAR1-based editing likelihoods at the various positions along pre-miR137. The bars colored in red represent sites that were previously hypothesized to be editing sites based on in-house data and the purple base numbers represent implausible editing sites. (B) This figure corresponds to the secondary structure of pre-miR137, as predicted by RNAstructure.
Table 1

Table 1a

<table>
<thead>
<tr>
<th>Index</th>
<th>Position in pre-miR137</th>
<th>Raw Frequency Table</th>
<th>Edit Frequency in Adult Mouse</th>
<th>Edit Frequency in Fetal Human</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>29</td>
<td>0.032675289</td>
<td>0.00046463</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>69</td>
<td>0.056820286</td>
<td>0.00092926</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>74</td>
<td>0.080326753</td>
<td>0.00139389</td>
<td></td>
</tr>
</tbody>
</table>

Table 1b

<table>
<thead>
<tr>
<th>SPECIES</th>
<th>Strict Combination Table</th>
<th>Actual Combination Frequency</th>
<th>Predicted Combination Frequency</th>
<th>Fold difference (Actual/Predicted)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADULT MOUSE</td>
<td>Strict Indexed Edit Combination</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.024506467</td>
<td>0.028373157</td>
<td>0.863720137</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.039482641</td>
<td>0.049658991</td>
<td>0.795075381</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.061946903</td>
<td>0.073364702</td>
<td>0.844369306</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>0.002722941</td>
<td>0.001677432</td>
<td>1.623278902</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>0.004765146</td>
<td>0.002478188</td>
<td>1.922834589</td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>0.012933969</td>
<td>0.004337351</td>
<td>2.981997362</td>
<td></td>
</tr>
<tr>
<td>123</td>
<td>0.000680735</td>
<td>0.000146511</td>
<td>4.646291942</td>
<td></td>
</tr>
<tr>
<td>FETAL HUMAN</td>
<td>Strict Indexed Edit Combination</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.00046463</td>
<td>0.000463551</td>
<td>1.002327258</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.00092926</td>
<td>0.000927534</td>
<td>1.001861331</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.00139389</td>
<td>0.001391948</td>
<td>1.001395403</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>0</td>
<td>4.31E-07</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>0</td>
<td>6.47E-07</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>0</td>
<td>1.29E-06</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>123</td>
<td>0</td>
<td>6.02E-10</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>

**MicroRNA 137 is subjected to A-to-I editing in vivo at multiple base positions.** (A) This table displays total editing rates by species at the three hypothesized edit sites in miR137, as determined from high throughput RNA-sequencing data. (B) This table portrays the results of further analysis of the RNA-sequencing data to determine frequencies of distinct edit combinations. Strict combinations refer to edits solely at the enumerated sites (e.g. “1” corresponds to editing at only site 1).
A-to-I RNA editing is conserved at relatively high levels at two distinct positions present in mature miR137. This graph illustrates the normalized editing rates (based on species-specific averages) at all of the observed edit sites for each species. Positions 69 and 74 exhibit relatively high normalized editing in fetal human brain and adult mouse brain.
**Figure 3**

A-to-G modifications to miR137 impart changes in target affinity *in vitro*. This figure displays the results of luciferase assays performed to quantify miR137 antisense target binding affinity for different edit combinations. Luciferase activity is proportional to target gene expression, so target affinity is inversely related to luciferase activity. The numeric combinations and All correspond to various edit combinations, Control corresponds to a positive control (binds strongly to the target), PL corresponds to a negative control (does not bind to the target), and WT corresponds to WT miR137.
A molecular model of interactions that contribute to schizophrenia risk. This schematic depicts the biochemical model constructed and subsequently simulated using MATLAB. The curved lines represent RNA molecules, the brown circles represent A-to-I edits, standard arrows represent promotion, and barred arrows represent inhibition.
The presence of the rs1625579 risk SNP increases schizophrenia-related inflammatory burden. This plot portrays the inflammatory burden output of the IFN-miR137-ADAR1 mathematical model with and without the rs1625579 risk SNP. The orange line corresponds to the results in the presence of the risk SNP and the blue line corresponds to the results in the absence of the risk SNP.
A-to-I editing of miR137 significantly influences the effect of the rs1625579 risk SNP on schizophrenia-related inflammatory burden. This figure represents the output of the statistical analysis on inflammatory burden fold change under different miR137 editing rate conditions.

The analysis was performed using a standard linear regression approach, where inflammatory burden fold change was modeled using basal editing rate as the predictor (while all other factors were held constant). The basal editing rate significantly affected inflammatory burden fold change (p < 2 x 10^{-16}).
Figure 7

A holistic schizophrenia risk model framework that incorporates genetic risk, environmental risk, and gene-environment interaction. This figure represents a holistic conceptual risk model framework for schizophrenia posited based on the findings of this study. The enclosed terms represent causal or intermediary entities, while the arrows represent causal relationships. The term “GxE” is used here as shorthand for gene-environment interaction.
**APPENDIX**

**Supplementary Table 1**

<table>
<thead>
<tr>
<th>Position</th>
<th>Reference Base</th>
<th>Consensus Count</th>
<th>Tissue</th>
<th>Error Base</th>
<th>Error Base Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>29</td>
<td>A</td>
<td>8609</td>
<td>Fetal Human Brain</td>
<td>G</td>
<td>4</td>
</tr>
<tr>
<td>41</td>
<td>A</td>
<td>8609</td>
<td>Fetal Human Brain</td>
<td>G</td>
<td>4</td>
</tr>
<tr>
<td>43</td>
<td>A</td>
<td>8609</td>
<td>Fetal Human Brain</td>
<td>G</td>
<td>3</td>
</tr>
<tr>
<td>44</td>
<td>A</td>
<td>8609</td>
<td>Fetal Human Brain</td>
<td>G</td>
<td>4</td>
</tr>
<tr>
<td>50</td>
<td>A</td>
<td>8609</td>
<td>Fetal Human Brain</td>
<td>G</td>
<td>4</td>
</tr>
<tr>
<td>53</td>
<td>A</td>
<td>8609</td>
<td>Fetal Human Brain</td>
<td>G</td>
<td>5</td>
</tr>
<tr>
<td>61</td>
<td>A</td>
<td>8609</td>
<td>Fetal Human Brain</td>
<td>G</td>
<td>4</td>
</tr>
<tr>
<td>69</td>
<td>A</td>
<td>8609</td>
<td>Fetal Human Brain</td>
<td>G</td>
<td>8</td>
</tr>
<tr>
<td>71</td>
<td>A</td>
<td>8609</td>
<td>Fetal Human Brain</td>
<td>G</td>
<td>4</td>
</tr>
<tr>
<td>72</td>
<td>A</td>
<td>8609</td>
<td>Fetal Human Brain</td>
<td>G</td>
<td>12</td>
</tr>
<tr>
<td>74</td>
<td>A</td>
<td>8609</td>
<td>Fetal Human Brain</td>
<td>G</td>
<td>12</td>
</tr>
<tr>
<td>29</td>
<td>A</td>
<td>1469</td>
<td>Adult Mouse Brain</td>
<td>G</td>
<td>48</td>
</tr>
<tr>
<td>61</td>
<td>A</td>
<td>1469</td>
<td>Adult Mouse Brain</td>
<td>G</td>
<td>14</td>
</tr>
<tr>
<td>68</td>
<td>A</td>
<td>1469</td>
<td>Adult Mouse Brain</td>
<td>G</td>
<td>4</td>
</tr>
<tr>
<td>69</td>
<td>A</td>
<td>1469</td>
<td>Adult Mouse Brain</td>
<td>G</td>
<td>82</td>
</tr>
<tr>
<td>71</td>
<td>A</td>
<td>1469</td>
<td>Adult Mouse Brain</td>
<td>G</td>
<td>3</td>
</tr>
<tr>
<td>72</td>
<td>A</td>
<td>1469</td>
<td>Adult Mouse Brain</td>
<td>G</td>
<td>11</td>
</tr>
<tr>
<td>74</td>
<td>A</td>
<td>1469</td>
<td>Adult Mouse Brain</td>
<td>G</td>
<td>118</td>
</tr>
</tbody>
</table>

Multiple sites on miR137 are edited *in vivo* at levels significantly higher than background changes. This table displays the base-modification read data from the high throughput RNA-sequencing experiment that met significance criteria under a Poisson distribution. Reference Base refers to the WT base and Error Base refers to the base read by the sequencer.
Supplementary Figure 1

Response hsaACTB

Effect Tests

<table>
<thead>
<tr>
<th>Source</th>
<th>Nparm</th>
<th>DF</th>
<th>Sum of Squares</th>
<th>F Ratio</th>
<th>Prob &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plasmid</td>
<td>9</td>
<td>9</td>
<td>37.553475</td>
<td>1.4211</td>
<td>0.2474</td>
</tr>
</tbody>
</table>

Least Squares Means Table

<table>
<thead>
<tr>
<th>Level</th>
<th>Least Sq Mean</th>
<th>Std Error</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>a119g</td>
<td>28.194555</td>
<td>0.9890734</td>
<td>28.1946</td>
</tr>
<tr>
<td>a119g-a124g</td>
<td>29.129404</td>
<td>0.9890734</td>
<td>29.1294</td>
</tr>
<tr>
<td>a124g</td>
<td>28.179814</td>
<td>1.2113626</td>
<td>28.1798</td>
</tr>
<tr>
<td>a79g</td>
<td>29.026718</td>
<td>0.9890734</td>
<td>29.0267</td>
</tr>
<tr>
<td>a79g-a119g</td>
<td>27.558155</td>
<td>0.9890734</td>
<td>27.5582</td>
</tr>
<tr>
<td>a79g-a119g-a124g</td>
<td>28.604111</td>
<td>0.9890734</td>
<td>28.6041</td>
</tr>
<tr>
<td>a79g-a124g</td>
<td>30.555158</td>
<td>0.9890734</td>
<td>30.5552</td>
</tr>
<tr>
<td>None</td>
<td>26.241238</td>
<td>0.9890734</td>
<td>26.2412</td>
</tr>
<tr>
<td>pCMVmiR137</td>
<td>27.264446</td>
<td>0.9890734</td>
<td>27.2644</td>
</tr>
<tr>
<td>pCMVPL</td>
<td>27.797063</td>
<td>0.9890734</td>
<td>27.7971</td>
</tr>
</tbody>
</table>

Response hsaGAPDH202

Effect Tests

<table>
<thead>
<tr>
<th>Source</th>
<th>Nparm</th>
<th>DF</th>
<th>Sum of Squares</th>
<th>F Ratio</th>
<th>Prob &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plasmid</td>
<td>9</td>
<td>9</td>
<td>50.368116</td>
<td>1.5206</td>
<td>0.2109</td>
</tr>
</tbody>
</table>

Least Squares Means Table

<table>
<thead>
<tr>
<th>Level</th>
<th>Least Sq Mean</th>
<th>Std Error</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>a119g</td>
<td>30.210362</td>
<td>1.1076074</td>
<td>30.2104</td>
</tr>
<tr>
<td>a119g-a124g</td>
<td>31.547063</td>
<td>1.1076074</td>
<td>31.5471</td>
</tr>
<tr>
<td>a124g</td>
<td>30.011102</td>
<td>1.3565365</td>
<td>30.0111</td>
</tr>
<tr>
<td>a79g</td>
<td>31.265220</td>
<td>1.1076074</td>
<td>31.2652</td>
</tr>
<tr>
<td>a79g-a119g</td>
<td>29.465520</td>
<td>1.1076074</td>
<td>29.4655</td>
</tr>
<tr>
<td>a79g-a119g-a124g</td>
<td>30.359163</td>
<td>1.1076074</td>
<td>30.3592</td>
</tr>
<tr>
<td>a79g-a124g</td>
<td>32.977412</td>
<td>1.1076074</td>
<td>32.9774</td>
</tr>
<tr>
<td>None</td>
<td>28.171171</td>
<td>1.1076074</td>
<td>28.1712</td>
</tr>
<tr>
<td>pCMVmiR137</td>
<td>29.506629</td>
<td>1.1076074</td>
<td>29.5066</td>
</tr>
<tr>
<td>pCMVPL</td>
<td>29.198672</td>
<td>1.1076074</td>
<td>29.1987</td>
</tr>
</tbody>
</table>

Response Pre-miR137

Effect Tests

<table>
<thead>
<tr>
<th>Source</th>
<th>Nparm</th>
<th>DF</th>
<th>Sum of Squares</th>
<th>F Ratio</th>
<th>Prob &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plasmid</td>
<td>9</td>
<td>9</td>
<td>172.79586</td>
<td>1.5763</td>
<td>0.1928</td>
</tr>
</tbody>
</table>
The luciferase reporter-based target affinity data does not represent differences in construct expression. This figure presents the output of a regression of the luciferase data against RT-PCR data for each construct. All of the p-values exceed 0.05, which suggests that the luciferase data is not explained by variation in expression of the various plasmids used in the luciferase experiments.
Supplementary Figure 2

The IFN-miR137-ADAR1 mathematical model exhibits biologically plausible input-output dynamics across time. This graph displays the results of a simulation of the IFN-miR137-ADAR1 mathematical model for a periodic infection input. Pathogen levels (corresponding to the periodic infection input) are in blue.

Supplementary Computational Methods

The square-wave pathogen input function was constructed in MATLAB as follows:

```matlab
function [out] = SquareWave (t, period, length, high, low)
if (mod(t,period) <= length)
    out = high;
else
    out = low;
end
```

The ODE system employed in this study was implemented in MATLAB as follows:

```matlab
function [dydt] = IFNmir137Model(t,y,parameters)
% Order of Quantities: Pathogen, IFN, ADAR1-A Level, Total miR137, TCF4
% mRNA, TCF4 protein, Dendritic Cells, Inflammation Burden.
% Order of parameters: Pathogen period, pathogen length, pathogen high,
% pathogen low, pathogen clearance rate, IFN release rate per DC per pathogen unit,
```
% IFN degradation rate, ADAR1-A synthesis rate per IFN, ADAR1-A degradation rate,
% basal miR137 transcription rate, binary rs1625579
% presence, miR137 degradation rate, basal
% editing rate, ADAR1-B level, TCF4 transcription rate,
% TCF4 degradation rate, miR137-TCF4 targeting Vmax, miR137-TCF4 targeting
% Km, TCF4 protein production Vmax, TCF4 protein production Km, TCF4
% protein degradation Vmax, TCF4 protein degradation Km, DC basal production
% rate, DC replication rate per TCF4 protein,
% DC degradation rate, Inflamation factor.
dydt = zeros(8,1);
dydt(1) =
    SquareWave(t,parameters(1),parameters(2),parameters(3),parameters(4))  -
    parameters(5)*y(1);
dydt(2) = parameters(6)*y(1)*y(7) -
    parameters(7)*y(2);
dydt(3) = parameters(8)*y(2) -
    parameters(9)*y(3);
dydt(4) = parameters(10)*(1 - 0.3*parameters(11))*(1 + 2.98*(y(3)/20)) -
    parameters(12)*y(4);
E = parameters(13)*((1 + (0.6*y(4)))/parameters(14));
dydt(5) = parameters(15) -
    parameters(16)*y(5) -
    0.16*(E*y(4))*y(5)*parameters(17)/(parameters(18)+y(5)) -
    (1-E)*y(4)*y(5)*parameters(17)/(parameters(18)+y(5));
dydt(6) =
    (parameters(19)*y(5))/(parameters(22)+y(6));
dydt(7) = parameters(23) + parameters(24)*y(6) -
    parameters(25)*y(7);
dydt(8) = parameters(26)*y(2);
end

The sensitivity analysis method was implemented in MATLAB as follows:

function [out] = SNPcompare
(t,y,tspan,y0,want,test,parameters,varySet,lower,upper,steps)

% Initialize inputs and outputs
set = varySet;
curr = parameters(set);
currL = curr*(1-(lower/100));
currU = curr*(1+(upper/100));
for idx = 1:length(curr)
    vec = linspace(currL(idx),currU(idx),steps);
    if idx == 1
        mat = vec;
    else
        mat = combvec(mat,vec);
    end
end
mat = mat';
out = [mat, zeros(size(mat,1),1)];
progressbar;

% Scan through parameter values and compute the output metrics
for idx = 1:size(mat,1)
    parms = parameters;
    parms(set) = mat(idx,:);
    parms(test) = 0;
    [~,yn] = ode45(@(t,y) IFNmir137Model(t,y,parms),tspan,y0);
yS0 = yn(end,want);
parms(test) = 1;
[~,yn] = ode45(@(t,y) IFNmir137Model(t,y,parms),tspan,y0);
yS1 = yn(end,want);
out(idx,end) = yS1/yS0;
end
end