In silico Analysis of miR-137 Transcription Inhibition in Homozygous (T/T) Schizophrenic Patients: a pilot study.

Stephany C. Esmaile[1], Bruno G. Sousa[2], Carlos A. G. Blaha[2], Jonas I. N. Oliveira[1]

[1]Departamento de Biofísica e Farmacologia, Universidade Federal do Rio Grande do Norte, Natal, Rio Grande do Norte, Brazil.

[2]Department of Molecular and Cell Biology, Universidade Federal do Rio Grande do Norte, Natal, Rio Grande do Norte, Brazil.

Corresponding Author:

Stephany Esmaile[1]

Campus Universitário Lagoa Nova, Natal, Rio Grande do Norte, 59078-970, Brazil

Email address: sce.neuroscience@gmail.com

Abstract

MicroRNA miR-137 single nucleotide polymorphism (rs1625579 SNP) is strongly associated with the worsening of schizophrenia symptoms and is involved in miR-137 gene suppression. MicroRNA miR-137 regulates synaptogenesis, neural plasticity and suppresses a variety of cancer types. Based on *in silico* predictions of the current MIR137 Host Gene with and without the SNP, it can be hypothesized that the mutation reversibly inhibits miR-137 gene transcription by steric hindrance due to an alteration on DNA conformation, stability, electrostatic potential, and transcription factor binding sites.

Keywords: MIR137HG; miR-137; in silico; Schizophrenia.

Introduction

Schizophrenia is considered a "synapse disease" caused by synaptic connectivity dysfunction of the brain and by an imbalance of neurotransmitters, especially dopamine (Siegert, 2010). It affects approximately 1% of the worldwide population and multiple candidate genes contribute to 80% of chances to develop the symptoms (Han et al. 2015; Williams et al. 2010). Additionally, monozygotic twins differ in methylation pattern of their genes highlighting the involvement of epigenetic mechanisms, which can lead to disorders of inhibitory activity and psychosis in one of the twins but not in the other (Melka et al. 2015). However the exact aetiology and genetic mechanisms of schizophrenia are still unknown.

Several robust and replicable genetic findings have been reported for Schizophrenia and the results indicate a considerable association between small non-coding RNAs (miRNA) and schizophrenia (Sullivan et al. 2012). Indeed, the nervous system has the broadest spectrum of miRNA expression of all human tissues, approximately 70% of all known human miRNAs (Krichevsky et al. 2003). These miRNAs may act as potent regulators of gene expression related to adult neurogenesis, neuronal maturation, brain function and plasticity (Forero et al. 2010). For instance, the miR-137 regulates the translation of a subset of proteins for optimising synaptic transmission and neuronal plasticity in mature neurons (Smrt et al. 2010; Han et al. 2015). In 2010, this microRNA was first identified as a candidate gene for schizophrenia susceptibility by a large-scale genome-wide association study (Potkin et al. 2010), being its Single Nucleotide Polymorphism (SNP), namely rs1625579, strongly associated with schizophrenia (Ripke et al. 2011).

Individuals carrying rs1625579 SNP present a single mutation by transversion of guanine (G) for thymine (T) in both alleles of the MIR137HG (T/T genotype), typically having hyperactivated brains, a schizophrenic phenotype well established in the clinical literature as a measure of functional inefficiency, when compared to G/G or G/T genotypes, less frequent in these patients. Supporting evidence also show that rs1625579 SNP may decrease human miR-137 gene expression (Van Erp et al. 2013; Cummings et al. 2012; Whalley et al. 2012; Kuswanto et al. 2015) which is located in 1p21.3 chromosome (Willemsen et al. 2011). Besides, it is correlated with dysregulation

in synaptogenesis and changes in the integrity of nerve fibers (Cousijin et al. 2014; Patel, et al. 2015; Rose et al. 2014).

Regarding the meaningful roles of miR-137 in Central Nervous System, the characterization of its gene with and without mutation is important for elucidation of the role of miR-137 in schizophrenia. Therefore, the purpose of this study was to predict promoter regions, DNA stability, CpG islands, transcription factor binding sites (TFBS), helix twist, minor groove width, roll and nucleosome position of the current version of MIR137HG.

Materials & Methods

The upgraded version of MIR137HG sequence (GRCh38.p7) was obtained from The Single Nucleotide Polymorphism Database (dbSNP) with code rs1625579 SNP to locate this mutation in MIR137HG sequence (*Fig. S1*). Since miR-137 (available in MirBase website) is read in the negative strand, it had to undergo a shift of uracil (U) by thymine (T) and be converted to its complement sequence to be found in MIR137HG, which is presented in dbSNP with its positive strand. The promoter regions are those preceding a gene responsible for its transcription initiation. Thus, a mutation in promoters may severely compromise the expression of the gene of interest (Newburger, 1994). The Promoter Prediction 2.0 Server program allows the prediction of eukaryotic promoters sequences of up to 1,500,000 bp in FASTA format (">" + Enter + following text without spaces) (Knudsen, 1999).

The energy stability of a macromolecule such as DNA can also be altered by a mutation. An unstable molecule tends to have its conformation modified and thus may make certain genes more or less accessible to regulatory agents. The more negative the energy value of the DNA covalent bonds, the more stable it is. Instability can even change the binding sites of transcription factors, which may stimulate or repress gene transcription (Latchman, 1996; Meysman, 2011). The distribution of CpG islands (dinucleotide 5'-CG-3 '), in turn, gives a broad prediction of areas suitable for methylation (Bell et al, 2011) and how conserved the region is, as well (Elliott et al. 2015). The GPminer program (Lee et al. 2012) provides such information about DNA stability and CpG islands in an intuitive way.

Transcription factor binding sites (TFBS) are regions in DNA where transcription factors may bind more or less intensely and thus regulate transcription even outside the core-binding site of a gene (Levo et al. 2015). Patch 1.0 software, which is based on TRANSFAC® 6.0 public database (Hashimoto, 2011) was used for this prediction. DNAshape software of the Rohs Lab was used to evaluate DNA parameters such as helix twist, propeller twist, minor groove width, and roll. The helical twist is the orientation of a base pair with respect to the helix axis. In the most common free B-form DNA, the helical twist is 36°, meaning that there are 10.5 bp/turn of helix. However, may have more extreme twists (up and down) to change its secondary structure, function and accessibility, The negative supercoiling shifts the helical periodicity from ~10.5 bp (free

B-form DNA) to ~11 bp, whereas the positive supercoiling shifts it toward ~10 bp (Strick, 1998). Therefore, the helical periodicity in genomic sequences contains information about the local DNA curvature and the global character of chromosomal supercoiling, and this can affect the regulation of transcription events (Kravatskaya et al. 2011).

The minor groove width depends on the existing hydrogen bonds in the DNA molecule and electrostatic potential of the minor groove. The greater the diameter, the more positive the electrostatic potential, since more hydrogens are exposed to the DNA deoxyribose backbone. The smaller the width, the more negative the electrostatic potential (Bishop, 2011). It is worth noting that this change in the electrostatic potential generates a modification in the preference of ligands for these regions, which may include epigenetic agents.

The parameter roll, or DNA winding, is based on the distance, in degrees, between base pairs, which can make a groove more or less accessible for drugs, for example. The greater the distance, the more negative the roll value; the shorter the distance, the more positive it is (Grasel, 2015). Based on the calculations of DNA shape program, loss of DNA function by extreme deformations (and the expected values) can be based on these criteria: (i) Helical Twist >45°(35° in B-DNA); (ii) Minor groove width >8.5Å or <1.5Å (5.8Å in B-DNA); (iii) Roll >20°(0° in B-DNA) (Zhou, 2013).

Nucleosomes cores are the most accessible regions to the transcription machinery and consist of approximately 147 bps of DNA turning around a histone octamer of pairs of histone H2A, H2B, H3 and H4 bound to linker histone H1. Between nucleosomes, DNA linkers ranging from 0 to 80 bp in length depending on the tissue and species (Routh, 2008). It is noteworthy that the same nucleosomes (composed by the same DNA sequence and same proteins) may differ in conformational dynamics (Ngo, 2015). In this study, the nucleosome positioning was predicted with RECON program interface, which calculates the most likely nucleosome distribution along the sequence of interest in which values greater than zero are more reliable in precision compared to negative values, and predictions matching the value +1 mean 100% chances of being in a nucleosome region (Levitsky , 2004).

Results & Discussion

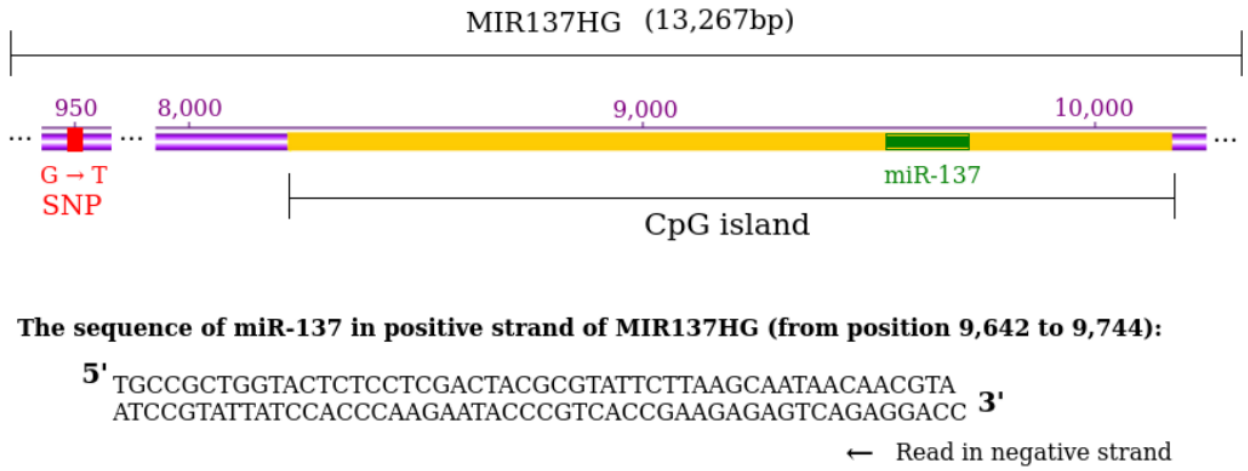
The MIR137HG consists of 13,267 bp and the mutation is located at position 950 of the positive strand, being more than 8,000 bp far from miR-137 gene, which is located between positions 9,642 and 9,744 and read in the negative strand. Regarding the prediction of promoter regions, the nearest one was the position 3,500 (its equivalent was position 9,767 in the positive strand) being only 24bp far from miR-137 first nucleotide in the negative strand (*Fig. 1*). The GPminer showed an increase in DNA energy of the mutated sequence with thymine, which implies less stability (*Fig. 2*). The CpG island result was expected since miR-137 gene is located in this region and, as a microRNA gene, it is strongly conserved during Evolution (Hobert, 2008).

Transcription factor binding sites (TFBS) for the wild sequence were for Glycoprotein Hormone Alpha (GHA) and Aldehyde Dehydrogenase 2 (ADH2) in position 949 of the positive strand including position 950 for this binding (*Fig. 3*). However, these two TFs were not present in the mutated sequence, but a third one appeared at position 948 of the positive strand: Pituitary-Specific Positive Transcription Factor 1 or POU1F1A (also including position 950 for this binding). In this way, TFBS were different around position 950 for the wild and mutated sequence, showing that a distinct nature, proportion, and distribution of TFs might contribute for inhibiting miR-137 transcription (Latchman, 1996; Meysman, 2011). No significant changes were detected in the rest of MIR137HG regarding this parameter.

Helical Twist, Minor Groove Width and Roll were predicted with position 950 centered and an interval from position 450 to 1450 of MIR137HG. The angulation of the helical twist in the region with thymine tends to be closer to the standard value of 35° than the region with guanine, although neither of them exceeded 45°. Accordingly, extreme values up and down were more homogeneous in the mutated sequence, which implies that this variation of DNA angle may contribute to a different signaling around position 950. The mutation did not lead to a deformation but could indicate a sufficient condition to change its secondary structure, therefore making the region around thymine950 more accessible to regulatory machinery.

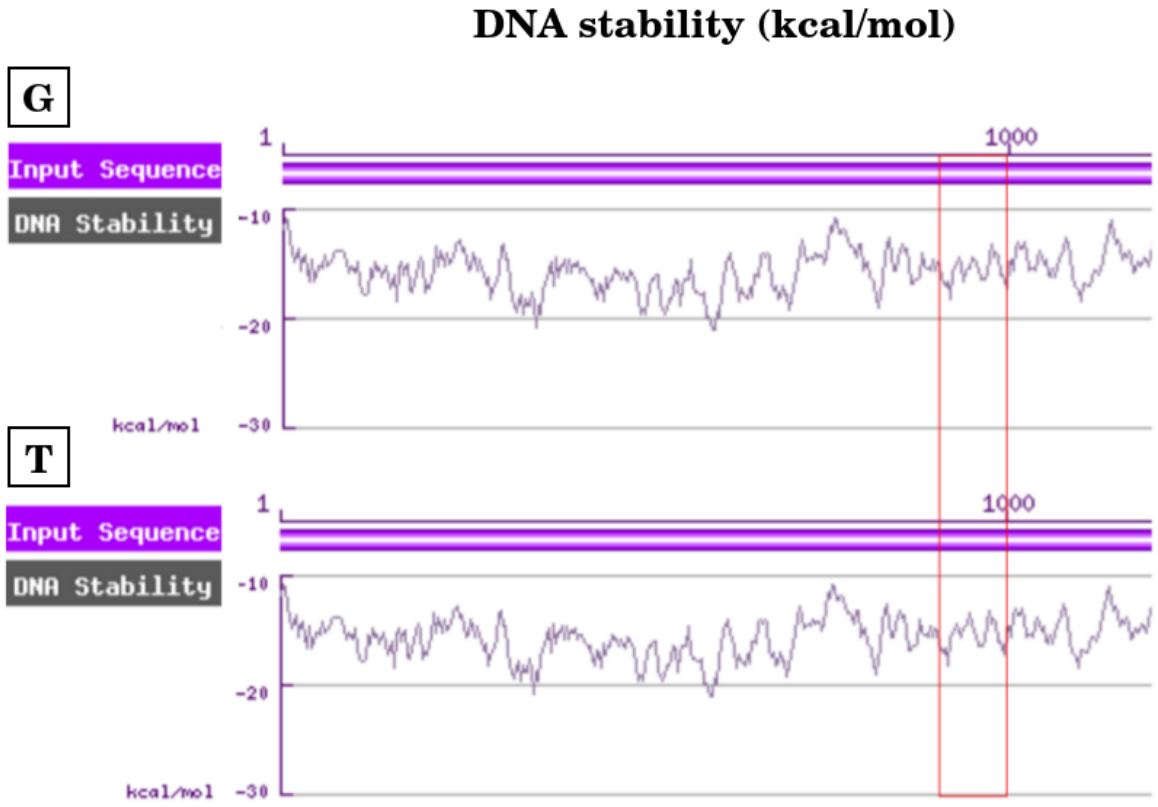
Both sequences had the width of minor groove slightly below the standard value of 5.8Å and above the minimum acceptable value of 1.5Å. For this reason, the mutated sequence may acquire a more negative electrostatic potential (Morávek, 2002). The roll parameter was 4° for the wild sequence and 0° with thymine950, which means a less tensioned molecule. The nucleosome distribution along MIR137HG sequence pointed that the position at which the mutation occurs (950) has a low probability of being into a nucleosome (-0.158352), as well as the sequence of miR-137 from 9642 (-0.035731) to 9744 (-0.981327).

As far as we know, this is the first *in silico* study that compares DNA parameters between wild and polymorphic sequences of MIR137HG.



178

179 *Figure 1: The miR-137 gene inserted in MIR137HG and the distance in base pair of it from the SNP.*



180

181 *Figure 2: Comparison of DNA stability between wild (G) and mutant (T) sequences.*

182

(a)

Position	Score	Probability
1100	0.503	Average
1800	0.641	Average
3500	0.666	Average
4700	0.654	Average
5800	0.592	Average

(b)

Parameters	G	T
TFBS*	949(-):GHA and ADH2	948(-):POU1F1A
Helical twist	← 35°→	→ 35°←
Minor groove width	5.5Å	4.5Å
Roll	4°	0°
Nucleosome positioning	Position 450 and interval between 9642 and 9744 with negative values for both sequences	

*TFBS (Transcription factor binding sites

Figure 3: (a) Promotor region prediction. (b) Comparison of DNA parameters between wild and mutant sequences.

Conclusion

Both SNP and miR-137 gene are part of a linker region of MIR137HG, being naturally more accessible or more easily regulated by intrinsic/extrinsic agents compared to nucleosome regions. Based on these predictions, the substitution of guanine for thymine may not globally alter MIR137HG. However, small and punctual changes may be sufficient for a reversible spatial inhibition of miR-137 transcription. The region around thymine950 may be less tensioned, less stable, more accessible, with negative electrostatic potential in the minor groove and a different transcription factors binding sites distribution, allowing distinct regulatory pathways. Since the mutation is more than 8,000bp far from miR-137 gene, a transcriptional suppression by steric hindrance can be hypothesized. We hope further analysis in Functional Genomics will contribute to elucidate the relationship between rs1625579 SNP and Schizophrenia.

References

Bell A, Bell D, Weber RS, El-Naggar and Adel K (2011). CpG island methylation profiling in human salivary gland adenoid cystic carcinoma. Cancer, 117:2898-909.

Bishop EPP, Rohs R, Parker, SCJ, West SM, Liu P, Mann RS, Honig B and Tullius TD (2011). A map of minor groove shape and electrostatic potential from hydroxyl radical cleavage patterns of DNA. ACS Chemical Biology, 6:1314-20.

Cousijin H, Eissing M, Fernández G, Fisher SE, Franke B, Zwiers M, Harrison PJ and Arias-Vásquez A (2014). No effect of schizophrenia risk genes MIR137, TCF4, and ZNF804A on macroscopic brain structure. Elsevier, 159:329-33.

Cummings EDG, Hargreaves A, Moore S, Fahey C, Dinan TG, Mcdonald C, O'callaghan E, O'neill FA, Waddington JL, Murphy KC, Morris DW, Gill M, Corvin A (2012). Mood congruent psychotic symptoms and specific cognitive deficits in carriers of the novel schizophrenia risk variant at MIR-137. Elsevier, 532:33-38.

Elliott G, Hong C, Xing X, Zhou X, Li L, Coarfa C, Bell RJA, Maire CL, Ligon, KL, Sigaroudinia M, Gascard P, Tlsty TD, Harris RA, Schalkwyk LC, Bilenky M, Mill J, Farnham PJ, Kellis, M, Marra MA, Milosavljevic A, Hirst M, Stormo GD, Wang T, Costello JF (2015) Intermediate DNA methylation is a conserved signature of genome regulation. Nature Communications, 6:6363.

Forero, D. A., Van Der Ven, K., Callaerts, P., & Del-Favero, J. (2010). miRNA genes and the brain: implications for psychiatric disorders a. Human Mutation, 31(11), 1195-1204.

Grasel FDA, De Oliveira TE, Netz PA (2015). Investigation of the interaction of 2-(2′-hydroxyphenyl)-benzoxazoles and their derivatives with B-DNA by docking and molecular dynamics. BMC Evolutionary Biology, 15:1.

Han J, Sarkar A, Gage FH (2015). MIR137: big impacts from small changes. Nature Medicine, 18:931.

Hashimoto R, Ohi K, Yasuda Y, Fukumoto M, Yamamori H, Takahashi H, Iwase M, Okochi T, Kazui H, Saitoh O, Tatsumi M, Iwata N, Ozaki N, Kamijima K, Kunugi H, Takeda M (2011). Variants of the RELA Gene are Associated with Schizophrenia and their Startle Responses Neuropsychopharmacology, 36:1921.

Hobert O (2008). Regulation by Transcription Factors and MicroRNAs. Science, Nova Iorque, 319:1785-1786.

Karki R, Pandya D, Elston RC, Ferlini C. (2015). Defining "mutation" and "polymorphism" in the era of personal genomics. BMC Medical Genomics, 8:1.

Knudsen, S. (1999). Promoter2.0: for the recognition of PolII promoter sequences. Bioinformatics, 15:356-361.

Kravatskaya, G. I., Kravatsky, Y. V., Chechetkin, V. R., & Tumanyan, V. G. (2011). Coexistence of different base periodicities in prokaryotic genomes as related to DNA curvature, supercoiling, and transcription. Genomics, 98(3), 223-231.

Krichevsky, A. M., King, K. S., Donahue, C. P., Khrapko, K., & Kosik, K. S. (2003). A microRNA array reveals extensive regulation of microRNAs during brain development. Rna, 9(10), 1274-1281.

Kuswanto CN, Sum MY, Qiu A, Sitoh Y-Y, Liu J, Sim K (2015). The Impact of Genome Wide Supported MicroRNA-137 (MIR137) Risk Variants on Frontal and Striatal White Matter Integrity, Neurocognitive Functioning, and Negative Symptoms in Schizophrenia. American Journal of Medical Genetics Part B: Neuropsychatric Genetics, 168:317-326.

Latchman DS (1996). Inhibitory transcription factors. International Journal of Biochemistry and Cell Biology, 28:965-974.

Lee T-Y, Chang W-C, Hsu J, Chang T-H, Shien D-M (2012). GPMiner: an integrated system for mining combinatorial cis-regulatory elements in mammalian gene group. BMC Genomics, 13:S3.

Levitsky VG (2004). RECON: a program for prediction of nucleosome formation potential. Elsevier, 32:W346-W349.

Levo M, Zalckvar E, Sharon E, Dantas MAC, Kalma Y, Lotam-Pompan M, Weinberger A, Yakhini Z, Rohs R, Segal E (2015). Unraveling determinants of transcription factor binding outside the core binding site. Genome Research, 25:1018-1029.

Melka MG, Castellani CA, O'Reilly RS, Shiva M (2015). Insights into the origin of DNA methylation differences between monozygotic twins discordant for schizophrenia. Journal of Molecular Psychiatry, 3:7.

Morávek Z, Neidle S, Schneider B (2002). Protein and drug interactions in the minor groove of DNA. Nucleic acids research 30:1182-91.

Meysman P, Dang T, Laukens K, De Smet R, Wu Y, Marchal, K. & Engelen, K. (2011). Use of structural DNA properties for the prediction of transcription-factor binding sites in Escherichia coli. Nucleic Acids Research, 39:e6-e6.

Morávek Z, Neidle, S, Schneider, B (2002). Protein and drug interactions in the minor groove of DNA. Nucleic acids research, 301182-91.

Newburger P, Skalnik DG, Hopkins P, Eklund E, Curnutte J (1994). Mutations in the promoter region of the gene for GP91-PHOX in X-linked chronic granulomatous-disease with decreased expression of cytochrome b558. Journal of Clinical Investigation, 94:1205-1211.

Ngo, TTM, Ha, T (2015). Nucleosomes undergo slow spontaneous gaping. Nucleic Acids Research, 43:3964-71.

Patel VS, Kelly S, Wright C, Gupta CN, Arias-Vasquez A, Perrone-Bizzozero N, Ehrlich S, Wang L, Bustillo JR, Morris D, Corvin A, Cannon DM, Mcdonald C, Donohoe G, Calhoun VD, Turner JA (2015). MIR137HG risk variant rs1625579 genotype is related to corpus callosum volume in schizophrenia. Neuroscience Letters, 602:44-49.

Perez-Iratxeta C, Andrade-Navarro MA, Wren JD (2007). Evolving research trends in bioinformatics. Briefings in bioinformatics, 8:88-95.

Potkin, S. G., Macciardi, F., Guffanti, G., Fallon, J. H., Wang, Q., Turner, J. A., ... & Xie, X. (2010). Identifying gene regulatory networks in schizophrenia. Neuroimage, 53(3), 839-847.

Ripke, S., Sanders, A. R., Kendler, K. S., Levinson, D. F., Sklar, P., Holmans, P. A., ... & Scolnick, E. (2011). Genome-wide association study identifies five new schizophrenia loci. Nature genetics, 43(10), 969.

Rose EJ, Morris DW, Fahey C, Cannon D, Mcdonald C, Scanlon C, Kelly S, Gill M, Corvin A, Donohoe G (2014). The miR-137 schizophrenia susceptibility variant rs1625579 does not predict variability in brain volume in a sample of schizophrenic patients and healthy individuals. American Journal of Medical Genetics Part B: Neuropsychiatric Genetics, 165:467-471.

Routh A, Sandin S, Rhodes D (2008). Nucleosome repeat length and linker histone stoichiometry determine chromatin fiber structure. Proceedings of the National Academy of Sciences of the United States of America, 105:8872-7.

Siegert S, Seo J, Kwon EJ, Rudenko A, Cho S, Wang W, Flood Z, Martorell AJ, Ericsson M, Mungenast AE, Tsai L-H (2010). The schizophrenia risk gene product miR-137 alters presynaptic plasticity. Nature Neuroscience, 18:1008-1016.

Smrt RD, Szulwach KE, Pfeiffer RL, Li X, Guo W, Pathania M, Teng Z-Q, Luo Y, Peng J, Bordey A, Jin P, Zhao X (2010). MicroRNA miR-137 regulates neuronal maturation by targeting ubiquitin ligase Mind Bomb-1. National Institutes of Health, 28:1060-1070.

Strick TR, Allemand J-F, Bensimon D, Croquette V (1998). Behavior of Supercoiled DNA. Biophysical Journal, 74:2016-2028.

Sullivan, P. F., Daly, M. J., & O'donovan, M. (2012). Genetic architectures of psychiatric disorders: the emerging picture and its implications. Nature Reviews Genetics, 13(8), 537.

Thiaville MM, Stoeck A, Chen L, Wu R-C, Magnani L, Oidtman J, Shih I-M, Lupien M, Wang T-Li, Yu J (2012). Identification of PBX1 Target Genes in Cancer Cells by Global Mapping of PBX1 Binding Sites (Pbx1 Direct Target Genes in Ovarian Cancer Cells). PLoS ONE, 7:e36054.

Van Erp TGM, Guella I, Vawter MP, Turner JB, Gregory G, Mccarthy G, Greve DN, Glover GH, Calhoun VD, Lim KO, Bustillo JR, Belger A, Ford JM, Mathalon DH, Diaz M,

309 Preda A, Nguyen D, Macciardi F, Potkin, SG (2013). Schizophrenia miR-137 Locus Risk
310 Genotype Is Associated with Dorsolateral Prefrontal Cortex Hyperactivation. Elsevier,
311 75:398-405.

312 Whalley HC, Papmeyer M, Romaniuk L, Sprooten E, Johnstone EC, Hall J, Lawrie SM,
313 Evans KL, Blumberg HP, Sussmann JE, Mcintosh AM (2012). Impact of a microRNA
314 MIR137 Susceptibility Variant on Brain Function in People at High Genetic Risk of
315 Schizophrenia or Bipolar Disorder. Nature Publishing Group (NPG), 37:2720.

316 Williams, H. J., Norton, N., Dwyer, S., Moskvina, V., Nikolov, I., Carroll, L., ... & Giegling,
317 I. (2011). Fine mapping of ZNF804A and genome-wide significant evidence for its
318 involvement in schizophrenia and bipolar disorder. Molecular psychiatry, 16(4), 429.

319 Willemsen, M. H., Vallès, A., Kirkels, L. A., Mastebroek, M., Loohuis, N. O., Kos, A., ... &
320 Holder-Espinasse, M. (2011). Chromosome 1p21. 3 microdeletions comprising DPYD
321 and MIR137 are associated with intellectual disability. Journal of medical genetics,
322 48(12), 810-818.

323 Zhou, T., Yang, L., Lu, Y., Dror, I., Dantas Machado, A. C., Ghane, T., ... & Rohs, R.
324 (2013). DNAshape: a method for the high-throughput prediction of DNA structural
325 features on a genomic scale. Nucleic acids research, 41(W1), W56-W62.