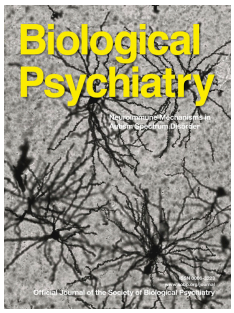
Accepted Manuscript



Understanding neurodevelopmental disorders: the promise of regulatory variation in the 3'UTRome

Kai Wanke, Paolo Devanna, Sonja C. Vernes

PII: S0006-3223(17)32191-1

DOI: 10.1016/j.biopsych.2017.11.006

Reference: BPS 13380

To appear in: *Biological Psychiatry*

Received Date: 7 February 2017

Revised Date: 2 November 2017

Accepted Date: 2 November 2017

Please cite this article as: Wanke K., Devanna P. & Vernes S.C., Understanding neurodevelopmental disorders: the promise of regulatory variation in the 3'UTRome, *Biological Psychiatry* (2017), doi: 10.1016/j.biopsych.2017.11.006.

TITLE: Understanding neurodevelopmental disorders: the promise of regulatory variation in the 3'UTRome

SHORT TITLE: 3'UTR variants and neurodevelopmental disorders

AUTHORS: Kai Wanke[1], Paolo Devanna[1] and Sonja C. Vernes[1, 2, 3].

AFFILIATIONS:

[1]Neurogenetics of Vocal Communication Group, Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands

[2]Donders Institute for Brain, Cognition and Behaviour, Nijmegen, The Netherlands

[3]Corresponding author.

CONTACT INFORMATION OF THE CORRESPONDING AUTHOR:

Mailing address:
P.O. Box 310, 6500 AH-Nijmegen
Street address:
Wundtlaan 1, 6525 XD Nijmegen, The Netherlands
Tel: +31 (0)24 3521 911
E-mail: sonja.vernes@mpi.nl

KEYWORDS:

Neurodevelopmental disorder; Neuropsychiatric disorders; Non-coding variation; 3' Untranslated Region (3'UTR); miRNAs; RNA-binding Proteins (RBP).

WORDCOUNT: 246 (Abstract); 3999 (Main Text)

NUMBER OF FIGURES: 2

NUMBER OF TABLES: 2

NUMBER OF SUPPLEMENTARY TABLES: 1

ABSTRACT

Neurodevelopmental disorders have a strong genetic component, but despite widespread efforts, the specific genetic factors underlying these disorders remain undefined for a large proportion of affected individuals. Given the accessibility of exome-sequencing, this problem has thus far been addressed from a protein-centric standpoint; however, protein-coding regions only make up ~1-2% of the human genome. With the advent of whole-genome sequencing we are in the midst of a paradigm shift as it is now possible to interrogate the entire sequence of the human genome (coding and non-coding) to fill in the missing heritability of complex disorders. These new technologies bring new challenges, as the number of non-coding variants identified per individual can be overwhelming, making it prudent to focus on non-coding regions of known function, for which the effects of variation can be predicted and directly tested to assess pathogenicity. The 3'UTRome is a region of the non-coding genome that perfectly fulfils these criteria and is of high interest when searching for pathogenic variation related to complex neurodevelopmental disorders. Herein, we review the regulatory roles of the 3'UTRome as binding sites for microRNAs, RNA binding proteins or during alternative polyadenylation. We detail existing evidence that these regions contribute to neurodevelopmental disorders and outline strategies for identification and validation of novel putatively pathogenic variation in these regions. This evidence suggests that studying the 3'UTRome will lead to the identification of new risk factors, new candidate disease genes and a better understanding of the molecular mechanisms contributing to NDDs.

INTRODUCTION

Neurodevelopmental disorders (NDDs) are estimated to affect more than 15% of the population (1) and can have a severe and persistent impact on quality of life, making them a major burden for society (2). The genetics of NDDs can involve rare, highly penetrant mutations in a single gene, in multiple genes, or common variation with small individual effect sizes (3). Heritability of this class of disorders is high and has been estimated from twin studies as 80-90% for Autism Spectrum Disorders (ASD) (4, 5), 76% for Attention Deficit Hyperactivity Disorder (ADHD) (5) and 81-85% for schizophrenia (SZ) (6), however the currently identified variation can only account for a proportion of the heritability of these disorders (reviewed in (7) for Mendelian and (8) for complex disorders). Comprehensive identification of the genetic factors underlying NDDs will reveal the molecular causes of these disorders and thus be central for diagnostics and the design of effective therapeutics. Until recently, the field has focused predominantly on the contribution of coding variation, however this protein-centric view has been challenged by a number of high profile studies pointing to key roles for non-coding variation in the aetiology of disorder (9-13) (reviewed in (5, 14)). Much of the non-coding portion of the genome is responsible for gene regulation (15) and thus can have significant effects on gene expression. Evidence supporting causative links between regulatory variants, perturbed gene expression, and NDDs is growing (13, 14, 16-18), suggesting that interrogating non-coding regulatory regions will make it possible to fill in some of the missing heritability of NDDs. In this review, we focus specifically on variation in one part of the non-coding regulatory genome - 3' untranslated regions (3'UTRs) – collectively known as the 3'UTRome.

3’UTRs play a crucial role in regulating protein expression and cellular localisation (19, 20), and studies are now emerging that have identified putatively pathogenic variation in 3’UTR regions in NDD cohorts (11, 21, 22) (and Table 1). Thus, the 3'UTRome represents a promising region to search for new genetic variation contributing to the pathogenesis of NDDs. Herein we detail molecular functions of 3'UTRs and perform a systematic review of the literature to summarise current evidence linking variation in these regions to the aetiology of NDDs. We focus the review on the best-described features of 3'UTRs: regulation by microRNAs, binding to RNA-Binding Proteins (RBPs) and modification by Alternative PolyAdenylation (APA). We outline approaches for identification and functional testing of

disease-associated variants, both in existing datasets and new patient cohorts. The collected data argue that identification and experimental testing of 3'UTR variation may identify new molecular mechanisms involved in NDDs and explain a proportion of the genetic architecture underlying these pathologies.

THE 3'UTROME IS A KEY FACTOR IN PROTEIN REGULATION

3'UTRs are transcribed as part of the mRNA expressed from all protein-coding genes but are not translated into protein because they are located after the stop codon (Figure 1). The length of 3'UTRs has increased during evolution and correlates positively with the overall complexity of the genome and organism (24). The length of human 3'UTR sequences can vary significantly, with a median length of ~1,200 nucleotides (23) and ranging from <80 (e.g. NM_001868.3) to more than 20,000 nucleotides (e.g. NM_001348800.1). Although 3'UTRs are less conserved than protein coding regions, they are more conserved than other non-coding regions such as promoters, 5'UTRs or introns (25).

3'UTRs play a key role in protein production and localization by modulating the amount of mRNA available for translation into protein and the transport of mRNA within the cell (Figure 1). There is evidence for additional functions of 3'UTRs including forming scaffolds for membrane proteins (26, 27), during termination of translation to support recycling of ribosomes (28), and acting independently of coding sequences following cleavage from their transcript (29-31). Below, we will explore the three best-described and most common mechanisms by which 3'UTRs contribute to protein regulation (Figure 1).

3'UTRs are bound by microRNAs to regulate translation of mRNAs into protein

3'UTRs influence the available pool of mRNAs that can be translated into protein by facilitating interactions with microRNA (miRNA) molecules. miRNAs are small non-coding RNAs that regulate protein production by binding to short motifs (6-8 nucleotides) known as miRNA binding sites (MBSs), found predominantly in the 3'UTR (Figure 1)(32, 33). This process reduces protein production by physically blocking translation or by inducing degradation of the transcript (32, 34, 35). Most protein coding genes are targeted by

miRNAs and on average each 3'UTR is estimated to have between 17-80 MBSs based on experimental data and *in silico* predictions (36, 37). Genetic variation in MBS sequences has the potential to interfere with miRNA-3'UTR interactions (either by destroying or creating MBSs) and thus alter the regulation of protein expression.

3'UTRs are bound by RNA-Binding Proteins (RBPs) to regulate translation or subcellular localisation of mRNAs

3'UTRs mediate interactions with RBPs to influence mRNA stability, translation and transport, ultimately regulating protein expression and subcellular localisation (Figure 1) (20). RBPs can regulate protein expression levels by binding to 3'UTR motifs (RBP sites) and either protecting mRNA molecules from degradation or speeding it up (38, 39). RBPs can facilitate transport of mRNAs within the cell by binding to 3'UTR regions, ensuring their localization to the correct subcellular compartment (19). RBPs can also act at the synapse to regulate the local translation of proteins, allowing synapses to dynamically respond to stimuli within an extremely short and physiologically relevant time frame (40).

3'UTRs length can vary to influence protein expression and localisation

Poly-adenylation sites are sequences that facilitate the addition of a polyA-tail at the end of an mRNA molecule (41). Each 3'UTR may have multiple possible poly-adenylation sites (Figure 1), which can be utilised in different cell types to produce Alternative Poly-Adenylation (APA) patterns from the same gene. APA produces mRNAs carrying 3'UTRs of different length and sequence, resulting in altered regulation of the mRNA (42).

NEUROBIOLOGICAL ROLES OF 3'UTR REGULATION

3'UTR regulation has important functions in the brain. miRNA regulation of 3'UTRs is involved in fundamental processes that shape the brain during early neurodevelopment including neuronal differentiation, neuronal migration and neurite outgrowth (43, 44). In the mature brain miRNA-3'UTR regulatory mechanisms contribute to synaptic plasticity (see (45-48) for reviews). RBP interactions with 3'UTRs are crucial for neuronal development and function (49, 50). RBPs such as the Human Antigen R (HuR) regulate the expression of

proteins that directly contribute to neocortical development and cortical layering (51). RBP-3'UTR interactions are also vital for normal synaptic function as they ensure that mRNAs are delivered to synapses and maintained there without degradation to facilitate local translation (52). Local translation allows individual synapses to produce independent responses to specific inputs, contributing to synaptic plasticity, neural network formation, and circuit maintenance (53), processes frequently disrupted in NDDs (49, 50, 54-56). Although the specific neuronal consequences of APA are not well characterised, genes expressed in the human and mouse brain have longer 3'UTRs on average than most other tissues (57, 58), and in the mouse 3'UTR length has been shown to differ across brain regions and developmental stages (59). Together these data may point to a requirement for increased mRNA regulation in neurons.

3'UTR VARIATION LINKED TO NEURODEVELOPMENTAL DISORDERS

Given the important regulatory mechanisms facilitated by 3'UTRs and the neurological processes in which they play a role, we propose that variation in the 3'UTRome is likely to contribute to the complex aetiology of NDDs. Indeed, it has already been shown that disruption of these regulatory pathways occurs in SZ and ASD where mutation or dysregulation of miRNAs has been observed (60-62). Variation within 3'UTRs can have subtle but important functional consequences for protein expression or localisation, suggesting that these variants are more likely to contribute to NDDs in a polygenic fashion, rather than as monogenic causes of disorder.

To explore links between 3'UTR variants and NDDs, we conducted a systematic review of the literature reporting 3'UTR variants in a selection of complex NDDs (ADHD, SZ, ASD and Intellectual Disability (ID); see Supplemental Information for detailed description of the search). In addition, we included four variants identified in a Specific Language Impairment (SLI) cohort from our own data published after the search was performed (22). Together, this identified 106 variants in 41 genes (see Supplementary Table 1); 18 of these variants have shown effects in functional testing (see Table 1). Given that most studies to date have targeted coding regions or filtered out non-coding variants, this represents a considerable

number of 3'UTR variants in NDDs. Below we highlight four examples from Table 1 that represent promising candidates that may contribute to disorder.

Fragile X Syndrome (FXS)

FXS is a form of ID almost exclusively caused by non-coding mutations in the 5'UTR of the *FMR1* gene that result in hypermethylation of its promoter and reduced gene expression (63, 64). In a small proportion of cases the disorder is caused by larger deletions of *FMR1* or by point mutations that alter the resulting FMRP protein (16, 64, 65). In rare cases, FXS patients do not have mutations affecting the 5'UTR or coding region of the *FMR1* gene. In these unexplained cases, regulatory variants affecting the 3'UTR make good candidates for pathogenicity. An *FMR1* candidate screen of 963 FXS patients identified 12 novel variants (5 unique; 7 recurrent) and 3 polymorphisms in the 3'UTR region (16). One 3'UTR variant (c.*746T>C; Table 1) was found in six patients but was not detected in the control population (N=1260). Two patients with the variant showed varying degrees of FXS symptoms, suggesting the variant may not be fully penetrant. Functional testing showed that this variant prevented the binding of HuR to the *FMR1* mRNA. This destabilized the transcript and led to reduced expression of the FMRP protein in a lymphoid patient cell line (66). This reduction in expression is reminiscent of the effect of the 5'UTR mutations described above, supporting the putative pathogenicity of this 3'UTR variant. However further studies are required to establish causality.

Rett syndrome (RTT)

RTT is a disorder on the autism spectrum that can also involve ID, speech loss, microcephaly and seizures. The majority of affected individuals are explained by causative mutations in the *MECP2* gene (67). However, approx. 35% of cases with classical RTT phenotype do not carry coding mutations in *MECP2* and are of unknown aetiology (68). *MECP2* has a very long and highly conserved 3'UTR region, suggesting an important function (68). The 3'UTR of *MECP2* has been screened in RTT patients that do not have coding changes in *MECP2* (69), identifying a number of variants (Table 1). A sub-set of these variants resulted in reduced expression of *MECP2* in patient cell lines (69). Given that the effect was a reduction in *MECP2* expression rather than complete loss, it is possible that these variants are not

causative on their own, and may co-occur with other variants in different genes associated with RTT phenotypes (e.g. *CDKL5* or *FOXG1*) (70). Further studies are needed to assess the impact of these variants on RTT pathogenesis.

Tourette's syndrome (TS)

TS is characterized by persistent, involuntary vocal and motor tics and may also include learning disabilities, features of ASD, ADHD or Obsessive-Compulsive Disorder (71). Multiple variants of small effect size are thought to underlie the pathology of TS (71), making it a good candidate for contributions from regulatory non-coding variation. A number of association studies and candidate sequencing studies have provided evidence that the *SLITRK1* gene contributes to the aetiology of TS (11). A single nucleotide change in the 3'UTR of *SLITRK1* (called 'Var321'; Table 1) was found in two independent cohorts of TS patients (11, 72), but not in >2,000 control individuals (11, 72) suggesting Var321 as a putative TS risk factor (73). Var321 is located within a MBS for a brain-expressed miRNA (miR-189). The presence of this variant resulted in stronger repression of the *SLITRK1* 3'UTR by miR-189 and was thus predicted to result in reduced functional protein in individuals carrying the change (11). Increased repression of *SLITRK1* resulting from Var321 could thus have similar effects to a disruptive protein coding mutation and contribute to TS pathogenicity.

Specific Language Impairment (SLI)

SLI is the failure to acquire normal language skills in the absence of other explanatory factors (74) and is thought to be caused by many genetic risk factors of small effect size (75). We recently identified four 3'UTR variants in a cohort of 43 severely affected SLI children (22). One of these variants was a common SNP (rs72727021; Table 1) significantly associated with a core feature of SLI - non-word repetition (NWR; a task where individuals repeat non-sense words that involves short-term phonological memory, vocal-motor planning and production). This SNP was within a MBS for mir-215 and had functional consequences, disrupting expression of the *ARHGEF39* gene in human cells and in post-mortem human cortex. This suggested the SNP as a novel SLI risk factor and provided the first example of a functional non-coding SNP that may contribute to SLI.

Taken together this evidence suggests that screening genetically complex disorders for

variation in 3'UTR regions is likely to reveal further regulatory variants involved in NDDs.

APPROACHES FOR IDENTIFYING AND ASSESSING POTENTIALLY PATHOGENIC 3'UTROME VARIATION

3'UTR variation can be identified by specifically screening known candidate genes, or via genome-wide approaches such as Genome Wide Association Studies (GWAS), Whole Exome Sequencing (WES) and Whole Genome Sequencing (WGS). Genome-wide studies have generated extensive data on genomic variation found in individuals with NDDs (76) and while most studies have focused on protein-coding variation, these datasets can be mined to identify non-coding 3'UTR variation (e.g. (22, 77)). Given the increasing evidence that 3'UTRome variation may contribute to NDD pathogenesis, mining existing and future datasets for these regulatory variants will be imperative for a comprehensive understanding of the genetic foundations of NDDs. In the remaining sections of this review, we outline simple yet effective strategies for identifying 3'UTRome variation from genomic datasets and methods for functional validation of these potentially pathogenic changes (summarised in Figure 2).

Genome-wide approaches capture variation in 3'UTRs

GWAS have been highly successful in the identification of common variants associated with NDDs. A sub-set of SNPs surveyed using standard GWAS-chips are located within 3'UTR regions (e.g. ~5% of SNPs on the Affimetrix Gene-Chip Human 100k (78)), making it possible to identify common variation within 3'UTR regions from sufficiently large GWAS datasets. WES studies have become a leading method for identifying pathogenic variation, with hundreds of thousands of exomes sequenced to date. By design, WES specifically captures protein-coding regions (79), however some 3'UTR sequence flanking the exons is also captured. Only a fraction of the 3'UTR sequence can be covered using this method, however we and others have shown that 3'UTR variation can be confidently called from WES data (22, 80). Together with the massive number of existing WES datasets, this makes WES a valuable, untapped resource for identifying pathogenic 3'UTR variation (22). New studies aiming to use WES to identify non-coding variants should choose the platform carefully as some

technologies have better coverage than others (81) and some newer WES platforms also specifically target parts of the non-coding genome (82).

WGS is rapidly becoming the dominant sequencing method given its advantages in genome coverage, sequence quality, and recent reductions in cost (83). Unlike other methods, WGS surveys the entire 3'UTRome making it a powerful way to obtain a comprehensive view of its contribution to disorders. Although only a small number of NDD WGS datasets currently exist (13, 76, 84-86), numbers are rapidly increasing. Estimates from the 1000 genomes project indicate that a typical genome has 4-5 million variants compared to the reference human genome (including SNPs, indels and structural variants) and based on 3'UTRome size this suggests that ~25,000 of these variants will fall in 3'UTR regions (87). Thus, appropriate experimental design (e.g. case-control, trios etc), variant filtering and prioritisation are crucial when exploiting any WGS dataset. We expect that most 3'UTR variants will have a small effect size, meaning that large-scale studies can require thousands of participants to identify risk variants (88), so data mining and integration will be powerful tools for the discovery of new 3'UTR variants. It is noteworthy that to date most 3'UTR variants have been identified in candidate approaches (Supplementary Table S1) where smaller cohort sizes (in the hundreds) were sufficient.

In the following section, we will discuss the tools and strategies that can be used to prioritise variants identified via any of the abovementioned methods, and predict and test the biological impact of 3'UTR variants.

Tools and databases to predict the pathogenicity of 3'UTRome variation

The effects of non-coding variants can be predicted based on their sequence and position within functional motifs. A summary of the wide range of bioinformatic tools available to predict MBSs, RBP sites and APA sites can be found in Supplementary Table S2. Most prediction tools are based on a combination of sequence homology to known motifs, evolutionary conservation of the putative motif (33, 89) and/or the thermodynamic properties of the possible interactions (90, 91). Some databases (e.g. PolyAsite for APA sites (92), CLIPdb for RBP sites (93) or miRGate for MBSs (94)) also include experimental data. Prediction algorithms are often revised based on newly obtained empirical information

about 3'UTR interactions but all *in silico* predictions come with false positive and negative rates that must be taken into account (95). The small size of most motifs, the large number of possible interactors and the effects of mRNA secondary structure make predictions challenging (96, 97). Combining multiple algorithms when assessing motifs is recommended (95, 98, 99) and is even possible within some programs such as miRWalk (100) and miRGen (101) (Supplementary Table S2). Using tools that combine bioinformatics with empirical data can also improve predictions (37, 102) (see Supplementary Table S2)

Functional Validation of 3'UTR variants

The functional validation of non-coding and coding variants alike is a powerful approach to determine which variants are most likely to contribute to pathology (103). 3'UTR variants can have different consequences depending on the type of change and the motif affected. As a result, there are a large number of techniques and approaches available to test their effects. One approach is to test functionality in human cell lines as they allow interrogation of the molecular effects of variants in an *in vitro* system. Luciferase reporter assays detect protein expression levels and thus are commonly used to measure regulation of a 3'UTR by miRNAs (22, 104). Direct interactions between 3'UTRs and miRNAs can be tested using methods such biotin tagged-mRNA-miRNA pulldowns (105), variations of (PAR-)CLIP (106, 107) by CLASH (36) or RNA-EMSA (108). The localization of RNA can be visualized using imaging techniques (109) and stability can be accessed via pulse-chase experiments (110). Cell lines cannot recapitulate complex, tissue specific effects that may occur in the human body making direct testing in patient material advisable, when possible. Patient samples may be in the form of blood or tissue from which expression levels could be directly measured (e.g. via RT-qPCR or Western Blot (111, 112)). Patient cell lines could also be established to observe regulatory effects (e.g. as induced pluripotent stem cells (113)). Gene expression can also be assessed in post-mortem brain tissue. Databases (such as GTEX and seeQTL) assess associations between common variants and gene expression making it possible to observe effects directly in the brain (114, 115). It should be noted that false positive/negative results are also possible during functional testing. Variants may show no effect in cell lines, despite having consequences in the brain, or conversely may be functional

in cell lines, but have no consequences in neural tissue. Given potential temporal and cell-type specificity, testing in neural tissue or cell types is advisable wherever possible.

CONCLUSIONS AND FUTURE DIRECTIONS

Investigating the coding genome to understand the genetic mechanisms underlying NDDs has already led to improved diagnosis, and driven the design and implementation of therapeutics (116, 117). However, we still only understand a fraction of the genetic factors contributing to NDDs, necessitating the need for further study and new approaches. To date, the contribution of non-coding variation to the pathogenesis of neurodevelopmental disorders has been severely understudied (118), despite the importance of these regions for regulation of gene expression. Given that non-coding DNA makes up ~98% of the 3 billion nucleotides of the human genome, methods like WGS can be expected to reveal millions of non-coding variants per person and interpreting the importance of variation in non-coding regions can be difficult. For these reasons, it is prudent to focus on regions of the non-coding genome for which the functions are well known and the effects of variation can be both predicted and tested. As outlined herein, the 3'UTRome clearly fits these criteria and thus represents a region of high interest for the identification of novel disease associated variation. This is further supported by the collected evidence we present of 106 published 3'UTR variants in 41 genes across NDD classifications (Supplementary Table S1).

With the increase in unbiased genome-wide screens such as WGS being performed, it is likely that many more 3'UTR variants will be linked to, and help explain the aetiology of these disorders. For this to happen however, it is crucial that these variants are not discarded in early filtering steps merely because they do not affect the structure of a protein, as is often the case. These non-coding changes can then be prioritized based on a number of factors including evolutionary conservation, frequency of occurrence in control populations (e.g. via 1000 Genome project, dbSNP, ClinVar), presence of a predicted functional motif at their position (Supplementary Table S2) or predicted pathogenicity (e.g. via CADD scores) (119). As with protein-coding variants, simply identifying variation is not enough (120). While *in silico* approaches make useful predictions about possible consequences of variation, none are completely accurate. Thus, direct lab-based testing of functional effects (such as outlined herein) is necessary to show that identified variants have

molecular consequences.

In this review, we have presented evidence that 3'UTRome variation can contribute to NDDs, and outlined clear approaches for further study of these regions in new and existing cohorts. This evidence suggests that studying the 3'UTRome will lead to the identification of new risk factors, new candidate disease genes and a better understanding of the fundamental biological mechanisms contributing to NDDs. However, we consider the 3'UTRome to be the 'tip of the iceberg' when considering the contribution of the non-coding genome to NDDs. Other non-coding regions are also becoming amenable to such approaches as we learn more about the regulatory genome. Promoter and enhancer regions are strong candidates for pathogenic variation and recent evidence has emerged that disruption of these regions may be prevalent in NDDs (13, 17, 121, 122). Linking variation in enhancers to effects on specific genes is challenging because enhancers can be megabases away from the genes they regulate. Despite this, rapid advances are being made by mapping chromatin architecture and interrogating genome-wide epigenomic data to predict the consequences of promoter and/or enhancer variation (123-126). A number of efforts are also underway to classify the effects of non-coding variation in a region-independent manner; the effect of common variation on gene expression can be assessed using tools that integrate transcriptome datasets (127-131), or via machine learning approaches incorporating multiple data types to predict consequences of non-coding variation (132). Similar Systems Biology approaches have been successfully applied in the investigation of NDDs and helped to unravel gene networks that are affected by regulatory changes (e.g. via miRNAs) (133-136). The integration of large datasets yields great potential to investigate the consequences of non-coding variants on all functional levels (RNA, protein, cell) and to increase our understanding for the consequences of subtle regulatory changes (137, 138).

In the future, integrating approaches to interrogate and functionally test *all* regions of the genome (protein-coding, promoter, enhancer, 3'UTR etc.) will exploit the full potential of next generation sequencing approaches to understand how *all* variation contributes to phenotypes, rather than cherry-picking protein-coding candidates. When this becomes routine, we will truly be in a post-genomic era and be closer to fully understanding the genetic factors contributing to complex phenotypes including, but not limited to, neurodevelopmental disorders.

ACKNOWLEDGEMENTS

This work was funded by a Marie Curie Career Integration Grant (PCIG12-GA-2012-333978) and by a Max Planck Research Group Grant both awarded to S.C.V. We thank members of the Neurogenetics of Vocal Communication group and Lot Snijders Blok for helpful discussions and feedback on the manuscript and Ksenia Lavrichenko for bioinformatic assistance. We are grateful to the anonymous reviewers for their constructive input.

FINANCIAL DISCLOSURE

All authors report no biomedical financial interests or potential conflicts of interest.

REFERENCES

1. Boyle CA, Boulet S, Schieve LA, Cohen RA, Blumberg SJ, Yeargin-Allsopp M, et al. (2011): Trends in the prevalence of developmental disabilities in US children, 1997-2008. *Pediatrics*. 127:1034-1042.
2. Dykens EM (2015): Family adjustment and interventions in neurodevelopmental disorders. *Curr Opin Psychiatry*. 28:121-126.
3. van Loo KM, Martens GJ (2007): Genetic and environmental factors in complex neurodevelopmental disorders. *Curr Genomics*. 8:429-444.
4. Freitag CM (2007): The genetics of autistic disorders and its clinical relevance: a review of the literature. *Mol Psychiatry*. 12:2-22.
5. Sadee W, Hartmann K, Seweryn M, Pietrzak M, Handelman SK, Rempala GA (2014): Missing heritability of common diseases and treatments outside the protein-coding exome. *Hum Genet*. 133:1199-1215.
6. Cardno AG, Marshall EJ, Coid B, Macdonald AM, Ribchester TR, Davies NJ, et al. (1999): Heritability estimates for psychotic disorders: the Maudsley twin psychosis series. *Arch Gen Psychiatry*. 56:162-168.
7. Chong JX, Buckingham KJ, Jhangiani SN, Boehm C, Sobreira N, Smith JD, et al. (2015): The Genetic Basis of Mendelian Phenotypes: Discoveries, Challenges, and Opportunities. *Am J Hum Genet.*. 97:199-215.
8. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, et al. (2009): Finding the missing heritability of complex diseases. *Nature*. 461:747-753.
9. Duan J, Shi J, Fiorentino A, Leites C, Chen X, Moy W, et al. (2014): A rare functional noncoding variant at the GWAS-implicated MIR137/MIR2682 locus might confer risk to schizophrenia and bipolar disorder. *Am J Hum Genet.* 95:744-753.
10. Spieler D, Kaffe M, Knauf F, Bessa J, Tena JJ, Giesert F, et al. (2014): Restless legs syndrome-associated intronic common variant in Meis1 alters enhancer function in the developing telencephalon. *Genome Res.* 24:592-603.
11. Abelson JF, Kwan KY, O'Roak BJ, Baek DY, Stillman AA, Morgan TM, et al. (2005): Sequence variants in SLITRK1 are associated with Tourette's syndrome. *Science*. 310:317-320.
12. Moncini S, Castronovo P, Murgia A, Russo S, Bedeschi MF, Lunghi M, et al. (2016): Functional characterization of CDK5 and CDK5R1 mutations identified in patients with non-syndromic intellectual disability. *J Hum Genet*. 61:283-293.
13. Turner TN, Hormozdiari F, Duyzend MH, McClymont SA, Hook PW, Iossifov I, et al. (2016): Genome Sequencing of Autism-Affected Families Reveals Disruption of Putative Noncoding Regulatory DNA. *Am J Hum Genet*. 98:58-74.
14. Zhang F, Lupski JR (2015): Non-coding genetic variants in human disease. *Hum Ml Genet*. 24:R102-110.
15. Gaiti F, Calcino AD, Tanurdzic M, Degnan BM (2017): Origin and evolution of the metazoan non-coding regulatory genome. *Dev Biol*. 427:193-202.
16. Collins SC, Bray SM, Suhl JA, Cutler DJ, Coffee B, Zwick ME, et al. (2010): Identification of novel FMR1 variants by massively parallel sequencing in developmentally delayed males. *Am J Hum Genet Part A*. 152A:2512-2520.
17. Yuen RK, Merico D, Cao H, Pellecchia G, Alipanahi B, Thiruvahindrapuram B, et al. (2016): Genome-wide characteristics of de novo mutations in autism. *NPJ Genom Med*. 1:160271-1602710.

18. Xiao X, Chang H, Li M (2017): Molecular mechanisms underlying noncoding risk variations in psychiatric genetic studies.*Mol Psychiatry.* 22:497-511.
19. Donnelly CJ, Fainzilber M, Twiss JL (2010): Subcellular communication through RNA transport and localized protein synthesis. *Traffic*. 11:1498-1505.
20. Eliscovich C, Buxbaum AR, Katz ZB, Singer RH (2013): mRNA on the move: the road to its biological destiny. *J Biol Chem*. 288:20361-20368.
21. McGowan H, Pang ZP (2015): Regulatory functions and pathological relevance of the MECP2 3'UTR in the central nervous system. *Cell Regen (Lond)*. 4:9.
22. Devanna P, Chen XS, Ho J, Gajewski D, Smith SD, Gialluisi A, et al. (2017): Next-gen sequencing identifies non-coding variation disrupting miRNA-binding sites in neurological disorders [published online ahead of print Mar 14; doi: 10.1038/mp.2017.30]. *Mol Psychiatry*.
23. Mayr C (2017): Regulation by 3'-Untranslated Regions. *Annu Rev Genet*.51:171-94
24. Mazumder B, Seshadri V, Fox PL (2003): Translational control by the 3'-UTR: the ends specify the means. *Trends Biochem Sci.* 28:91-98.
25. Duret L, Dorkeld F, Gautier C (1993): Strong conservation of non-coding sequences during vertebrates evolution: potential involvement in post-transcriptional regulation of gene expression. *Nucleic Acids Res*. 21:2315-2322.
26. Berkovits BD, Mayr C (2015): Alternative 3' UTRs act as scaffolds to regulate membrane protein localization. *Nature*. 522:363-367.
27. Chartron JW, Hunt KC, Frydman J (2016): Cotranslational signal-independent SRP preloading during membrane targeting. *Nature*. 536:224-228.
28. Panek J, Kolar M, Herrmannova A, Valasek LS (2016): A systematic computational analysis of the rRNA-3' UTR sequence complementarity suggests a regulatory mechanism influencing post-termination events in metazoan translation. *RNA.* 22:957-967.
29. Mercer TR, Wilhelm D, Dinger ME, Solda G, Korbie DJ, Glazov EA, et al. (2011): Expression of distinct RNAs from 3' untranslated regions. *Nucleic Acids Res*. 39:2393-2403.
30. Heasman J, Wessely O, Langland R, Craig EJ, Kessler DS (2001): Vegetal localization of maternal mRNAs is disrupted by VegT depletion. *Dev Biol*. 240:377-386.
31. Jenny A, Hachet O, Zavorszky P, Cyrklaff A, Weston MD, Johnston DS, et al. (2006): A translation-independent role of oskar RNA in early Drosophila oogenesis. *Development*. 133:2827-2833.
32. He L, Hannon GJ (2004): MicroRNAs: small RNAs with a big role in gene regulation. *Nat Rev Genet*. 5:522-531.
33. Lewis BP, Burge CB, Bartel DP (2005): Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell*. 120:15-20.
34. Hammond SM, Boettcher S, Caudy AA, Kobayashi R, Hannon GJ (2001): Argonaute2, a link between genetic and biochemical analyses of RNAi. *Science*. 293:1146-1150.
35. Valinezhad Orang A, Safaralizadeh R, Kazemzadeh-Bavili M (2014): Mechanisms of miRNA-Mediated Gene Regulation from Common Downregulation to mRNA-Specific Upregulation. *Int J Genomics*. 2014:970607.
36. Helwak A, Tollervey D (2014): Mapping the miRNA interactome by cross-linking ligation and sequencing of hybrids (CLASH). *NatProtoc*. 9:711-728.
37. Agarwal V, Bell GW, Nam JW, Bartel DP (2015): Predicting effective microRNA target sites in mammalian mRNAs. *eLife*. 4: e5005

38. Klein ME, Monday H, Jordan BA (2016): Proteostasis and RNA Binding Proteins in

Synaptic Plasticity and in the Pathogenesis of Neuropsychiatric Disorders. *Neural Plast*. 2016:3857934.
39. Keene JD (2007): RNA regulons: coordination of post-transcriptional events. *Nat Rev Genet*. 8:533-543.
40. Holt CE, Schuman EM (2013): The central dogma decentralized: new perspectives on RNA function and local translation in neurons. *Neuron*. 80:648-657.
41. Proudfoot NJ (2011): Ending the message: poly(A) signals then and now. *Genes Dev*. 25:1770-1782.
42. Miura P, Sanfilippo P, Shenker S, Lai EC (2014): Alternative polyadenylation in the nervous system: to what lengths will 3' UTR extensions take us? *Bioessays*. 36:766-777.
43. Han J, Kim HJ, Schafer ST, Paquola A, Clemenson GD, Toda T, et al. (2016): Functional Implications of miR-19 in the Migration of Newborn Neurons in the Adult Brain. *Neuron*. 91:79-89.
44. Sasaki Y, Gross C, Xing L, Goshima Y, Bassell GJ (2014): Identification of axon-enriched microRNAs localized to growth cones of cortical neurons. *DevNeurobiol*. 74:397-406.
45. Davis GM, Haas MA, Pocock R (2015): MicroRNAs: Not "Fine-Tuners" but Key Regulators of Neuronal Development and Function. *Front Neurol*. 6:245.
46. Diaz NF, Cruz-Resendiz MS, Flores-Herrera H, Garcia-Lopez G, Molina-Hernandez A (2014): MicroRNAs in central nervous system development. *Rev Neurosci*. 25:675-686.
47. Sun E, Shi Y (2015): MicroRNAs: Small molecules with big roles in neurodevelopment and diseases. *Exp Neurol*. 268:46-53.
48. Aksoy-Aksel A, Zampa F, Schratt G (2014): MicroRNAs and synaptic plasticity--a mutual relationship. *Philos Trans R Soc Lond B Biol Sci*. 369:pii: 20130515
49. Bryant CD, Yazdani N (2016): RNA-binding proteins, neural development and the addictions. *Genes Brain Behav*. 15:169-186.
50. Pilaz LJ, Silver DL (2015): Post-transcriptional regulation in corticogenesis: how RNA-binding proteins help build the brain. *Wiley Interdiscip Rev RNA*. 6:501-515.
51. Kraushar ML, Thompson K, Wijeratne HR, Viljetic B, Sakers K, Marson JW, et al. (2014): Temporally defined neocortical translation and polysome assembly are determined by the RNA-binding protein Hu antigen R. *Proc Natl Acad Scie U S A*. 111:E3815-3824.
52. Di Liegro CM, Schiera G, Di Liegro I (2014): Regulation of mRNA transport, localization and translation in the nervous system of mammals (Review). *Int J Mol Med*. 33:747-762.
53. Fernandez-Moya SM, Bauer KE, Kiebler MA (2014): Meet the players: local translation at the synapse. *Front Mol Neurosci*. 7:84.
54. de la Torre-Ubieta L, Won H, Stein JL, Geschwind DH (2016): Advancing the understanding of autism disease mechanisms through genetics. *Nat Med*. 22:345-361.
55. Sanders SJ, He X, Willsey AJ, Ercan-Sencicek AG, Samocha KE, Cicek AE, et al. (2015): Insights into Autism Spectrum Disorder Genomic Architecture and Biology from 71 Risk Loci. *Neuron*. 87:1215-1233.
56. Parikshak NN, Gandal MJ, Geschwind DH (2015): Systems biology and gene networks in neurodevelopmental and neurodegenerative disorders. *Nat Rev Genet.* 16:441-458.
57. Mayr C (2016): Evolution and Biological Roles of Alternative 3'UTRs. *Trends Cell Biol*. 26:227-237.
58. Miura P, Shenker S, Andreu-Agullo C, Westholm JO, Lai EC (2013): Widespread and extensive lengthening of 3' UTRs in the mammalian brain. *Genome Res*. 23:812-825.
59. Kocabas A, Duarte T, Kumar S, Hynes MA (2015): Widespread Differential Expression of Coding Region and 3' UTR Sequences in Neurons and Other Tissues. *Neuron*. 88:1149-

1156.
60. Geaghan M, Cairns MJ (2015): MicroRNA and Posttranscriptional Dysregulation in Psychiatry. *Biol Psychiatry*. 78:231-239.
61. Mishra PJ, Bertino JR (2009): MicroRNA polymorphisms: the future of pharmacogenomics, molecular epidemiology and individualized medicine. *Pharmacogenomics*. 10:399-416.
62. Roussos P, Mitchell AC, Voloudakis G, Fullard JF, Pothula VM, Tsang J, et al. (2014): A role for noncoding variation in schizophrenia. *Cell Rep*. 9:1417-1429.
63. Verkerk AJ, Pieretti M, Sutcliffe JS, Fu YH, Kuhl DP, Pizzuti A, et al. (1991): Identification of a gene (FMR-1) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome. *Cell*. 65:905-914.
64. Santoro MR, Bray SM, Warren ST (2012): Molecular mechanisms of fragile X syndrome: a twenty-year perspective. *Annu Rev Pathol*. 7:219-245.
65. Coffee B, Ikeda M, Budimirovic DB, Hjelm LN, Kaufmann WE, Warren ST (2008): Mosaic FMR1 deletion causes fragile X syndrome and can lead to molecular misdiagnosis: a case report and review of the literature. *Am J Med Genet A*. 146A:1358-1367.
66. Suhl JA, Muddashetty RS, Anderson BR, Ifrim MF, Visootsak J, Bassell GJ, et al. (2015): A 3' untranslated region variant in FMR1 eliminates neuronal activity-dependent translation of FMRP by disrupting binding of the RNA-binding protein HuR. *Proc Natl Acad Scie U S A*. 112:E6553-6561.
67. Amir RE, Van den Veyver IB, Wan M, Tran CQ, Francke U, Zoghbi HY (1999): Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2. *NatGenet*. 23:185-188.
68. Newnham CM, Hall-Pogar T, Liang S, Wu J, Tian B, Hu J, et al. (2010): Alternative polyadenylation of MeCP2: Influence of cis-acting elements and trans-acting factors. *RNA Biol.* 7:361-372.
69. Coutinho AM, Oliveira G, Katz C, Feng J, Yan J, Yang C, et al. (2007): MECP2 coding sequence and 3'UTR variation in 172 unrelated autistic patients. *Am J Med Genet Part B*. 144B:475-483.
70. Guerrini R, Parrini E (2012): Epilepsy in Rett syndrome, and CDKL5- and FOXG1-gene-related encephalopathies. *Epilepsia*. 53:2067-2078.
71. Pagliaroli L, Veto B, Aranyi T, Barta C (2016): From Genetics to Epigenetics: New Perspectives in Tourette Syndrome Research. *Front Neurosci*. 10:277.
72. O'Roak BJ, Morgan TM, Fishman DO, Saus E, Alonso P, Gratacos M, et al. (2010): Additional support for the association of SLITRK1 var321 and Tourette syndrome. *MolPsychiatry*. 15:447-450.
73. Yasmeen S, Melchior L, Bertelsen B, Skov L, Mol Debes N, Tumer Z (2013): Sequence analysis of SLITRK1 for var321 in Danish patients with Tourette syndrome and review of the literature. *Psychiatr Genet*. 23:130-133.
74. Rice ML (2016): Specific Language Impairment, Nonverbal IQ, Attention-Deficit/Hyperactivity Disorder, Autism Spectrum Disorder, Cochlear Implants, Bilingualism, and Dialectal Variants: Defining the Boundaries, Clarifying Clinical Conditions, and Sorting Out Causes. *J Speech Lang Hear Res*. 59:122-132.
75. Newbury DF, Monaco AP (2010): Genetic advances in the study of speech and language disorders. *Neuron*. 68:309-320.
76. Soden SE, Saunders CJ, Willig LK, Farrow EG, Smith LD, Petrikin JE, et al. (2014): Effectiveness of exome and genome sequencing guided by acuity of illness for diagnosis of

574 neurodevelopmental disorders. *Sci Transl Med*. 6:265ra168.
575 77. Liu C, Zhang F, Li T, Lu M, Wang L, Yue W, et al. (2012): MirSNP, a database of
576 polymorphisms altering miRNA target sites, identifies miRNA-related SNPs in GWAS SNPs
577 and eQTLs. *BMC Genomics*. 13:661.
578 78. Nicolae DL, Wen X, Voight BF, Cox NJ (2006): Coverage and characteristics of the
579 Affymetrix GeneChip Human Mapping 100K SNP set. *PLoS Genet*. 2:e67.
580 79. Ng SB, Turner EH, Robertson PD, Flygare SD, Bigham AW, Lee C, et al. (2009): Targeted
581 capture and massively parallel sequencing of 12 human exomes. *Nature*. 461:272-276.
582 80. An JY, Cristino AS, Zhao Q, Edson J, Williams SM, Ravine D, et al. (2014): Towards a
583 molecular characterization of autism spectrum disorders: an exome sequencing and systems
584 approach. *TranslPsychiatry*. 4:e394.
585 81. Shigemizu D, Momozawa Y, Abe T, Morizono T, Boroevich KA, Takata S, et al. (2015):
586 Performance comparison of four commercial human whole-exome capture platforms. *Sci*
587 *Rep.* 5:12742.
588 82. Warr A, Robert C, Hume D, Archibald A, Deeb N, Watson M (2015): Exome
589 Sequencing: Current and Future Perspectives. *G3 (Bethesda)*. 5:1543-1550.
590 83. Meynert AM, Ansari M, FitzPatrick DR, Taylor MS (2014): Variant detection sensitivity
591 and biases in whole genome and exome sequencing. *BMC Bioinformatics*. 15:247.
592 84. Jiang YH, Yuen RK, Jin X, Wang M, Chen N, Wu X, et al. (2013): Detection of clinically
593 relevant genetic variants in autism spectrum disorder by whole-genome sequencing. *Am J*
594 *Hum Genet*. 93:249-263.
595 85. Nemirovsky SI, Cordoba M, Zaiat JJ, Completa SP, Vega PA, Gonzalez-Moron D, et al.
596 (2015): Whole genome sequencing reveals a de novo SHANK3 mutation in familial autism
597 spectrum disorder. *PloS One*. 10:e0116358.
598 86. Yuen RK, Thiruvahindrapuram B, Merico D, Walker S, Tammimies K, Hoang N, et al.
599 (2015): Whole-genome sequencing of quartet families with autism spectrum disorder. *Nat*
600 *Med*. 21:185-191.
601 87. 1000 Genomes Project Consortium, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang
602 HM, et al. (2015): A global reference for human genetic variation. *Nature*. 526:68-74.
603 88. Sham PC, Purcell SM (2014): Statistical power and significance testing in large-scale
604 genetic studies. *Nat Rev Genet*. 15:335-346.
605 89. Paz I, Kosti I, Ares M, Jr., Cline M, Mandel-Gutfreund Y (2014): RBPmap: a web server
606 for mapping binding sites of RNA-binding proteins. *Nucleic Acids Res.* 42:W361-367.
607 90. John B, Enright AJ, Aravin A, Tuschl T, Sander C, Marks DS (2004): Human MicroRNA
608 targets. *PLoS Biol*. 2:e363.
609 91. Rehmsmeier M, Steffen P, Hochsmann M, Giegerich R (2004): Fast and effective
610 prediction of microRNA/target duplexes. *RNA*. 10:1507-1517.
611 92. Gruber AJ, Schmidt R, Gruber AR, Martin G, Ghosh S, Belmadani M, et al. (2016): A
612 comprehensive analysis of 3' end sequencing data sets reveals novel polyadenylation signals
613 and the repressive role of heterogeneous ribonucleoprotein C on cleavage and
614 polyadenylation. *Genome Res.* 26:1145-1159.
615 93. Yang YC, Di C, Hu B, Zhou M, Liu Y, Song N, et al. (2015): CLIPdb: a CLIP-seq database
616 for protein-RNA interactions. *BMC Genomics*. 16:51.
617 94. Andres-Leon E, Gonzalez Pena D, Gomez-Lopez G, Pisano DG (2015): miRGate: a
618 curated database of human, mouse and rat miRNA-mRNA targets. *Database (Oxford)*.
619 2015:bav035.
620 95. Witkos TM, Koscianska E, Krzyzosiak WJ (2011): Practical Aspects of microRNA Target

621 Prediction. *Curr Mol Med.* 11:93-109.
622 96. Pinzon N, Li B, Martinez L, Sergeeva A, Presumey J, Apparailly F, et al. (2017):
623 microRNA target prediction programs predict many false positives. *Genome Res.* 27:234-245.
624 97. Yue D, Liu H, Huang Y (2009): Survey of Computational Algorithms for MicroRNA
625 Target Prediction. *Curr Genomics.* 10:478-492.
626 98. Sethupathy P, Megraw M, Hatzigeorgiou AG (2006): A guide through present
627 computational approaches for the identification of mammalian microRNA targets.*Nat*
628 *Methods.* 3:881-886.
629 99. Fan X, Kurgan L (2015): Comprehensive overview and assessment of computational
630 prediction of microRNA targets in animals. *Brief Bioinform.* 16:780-794.
631 100. Dweep H, Gretz N (2015): miRWalk2.0: a comprehensive atlas of microRNA-target
632 interactions. *Nat Methods*. 12:697.
633 101. Megraw M, Sethupathy P, Corda B, Hatzigeorgiou AG (2007): miRGen: a database for
634 the study of animal microRNA genomic organization and function. *Nucleic Acids Res.*
635 35:D149-155.
636 102. Ding J, Li X, Hu H (2016): TarPmiR: a new approach for microRNA target site
637 prediction. *Bioinformatics*. 32:2768-2775.
638 103. Ipe J, Swart M, Burgess KS, Skaar TC (2017): High-Throughput Assays to Assess the
639 Functional Impact of Genetic Variants: A Road Towards Genomic-Driven Medicine.*Clin Transl*
640 *Sci.* 10:67-77.
641 104. Wang B, Doench JG, Novina CD (2007): Analysis of microRNA effector functions in
642 vitro. *Methods*. 43:91-104.
643 105. Tan SM, Lieberman J (2016): Capture and Identification of miRNA Targets by Biotin
644 Pulldown and RNA-seq. *Methods Mol Biol.*. 1358:211-228.
645 106. Hafner M, Lianoglou S, Tuschl T, Betel D (2012): Genome-wide identification of
646 miRNA targets by PAR-CLIP. *Methods*. 58:94-105.
647 107. Zhang C, Darnell RB (2011): Mapping in vivo protein-RNA interactions at single-
648 nucleotide resolution from HITS-CLIP data. *Nat Biotechnol.* 29:607-614.
649 108. Rio DC (2014): Electrophoretic mobility shift assays for RNA-protein complexes. *Cold*
650 *Spring Harb Protoc*. 2014:435-440.
651 109. Buxbaum AR, Haimovich G, Singer RH (2015): In the right place at the right time:
652 visualizing and understanding mRNA localization. *Nat Rev MolCell Biol*. 16:95-109.
653 110. Tani H, Akimitsu N (2012): Genome-wide technology for determining RNA stability in
654 mammalian cells: historical perspective and recent advantages based on modified
655 nucleotide labeling. *RNA Biol*. 9:1233-1238.
656 111. Murphy J, Bustin SA (2009): Reliability of real-time reverse-transcription PCR in
657 clinical diagnostics: gold standard or substandard? *Expert Rev Mol Diagn*. 9:187-197.
658 112. Hayes PC, Wolf CR, Hayes JD (1989): Blotting techniques for the study of DNA, RNA,
659 and proteins. *BMJ*. 299:965-968.
660 113. Malik N, Rao MS (2013): A review of the methods for human iPSC derivation.
661 *Methods Mol Biol*. 997:23-33.
662 114. Carithers LJ, Ardlie K, Barcus M, Branton PA, Britton A, Buia SA, et al. (2015): A Novel
663 Approach to High-Quality Postmortem Tissue Procurement: The GTEx Project. *Biopreserv*
664 *Biobank*. 13:311-319.
665 115. Xia K, Shabalin AA, Huang S, Madar V, Zhou YH, Wang W, et al. (2012): seeQTL: a
666 searchable database for human eQTLs. *Bioinformatics*. 28:451-452.
667 116. Carbone PS (2013): Moving from research to practice in the primary care of children

668 with autism spectrum disorders. *Acad Pediatr*. 13:390-399.
669 117. Healy A, Rush R, Ocain T (2011): Fragile X syndrome: an update on developing
670 treatment modalities. *ACS Chem Neurosci.* 2:402-410.
671 118. Spielmann M, Mundlos S (2013): Structural variations, the regulatory landscape of
672 the genome and their alteration in human disease. *Bioessays*. 35:533-543.
673 119. Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J (2014): A general
674 framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet*.
675 46:310-315.
676 120. Deriziotis P, Fisher SE (2013): Neurogenomics of speech and language disorders: the
677 road ahead. *Genome Biol*. 14:204.
678 121. Luo R, Sanders SJ, Tian Y, Voineagu I, Huang N, Chu SH, et al. (2012): Genome-wide
679 transcriptome profiling reveals the functional impact of rare de novo and recurrent CNVs in
680 autism spectrum disorders. *Am J Hum Genet*. 91:38-55.
681 122. Shi L, Zhang X, Golhar R, Otieno FG, He M, Hou C, et al. (2013): Whole-genome
682 sequencing in an autism multiplex family. *Mol Autism*. 4:8.
683 123. Lupianez DG, Spielmann M, Mundlos S (2016): Breaking TADs: How Alterations of
684 Chromatin Domains Result in Disease. *Trends Genet*. 32:225-237.
685 124. Allou L, Lambert L, Amsallem D, Bieth E, Edery P, Destree A, et al. (2012): 14q12 and
686 severe Rett-like phenotypes: new clinical insights and physical mapping of FOXG1-regulatory
687 elements. *Eur J Hum Genet*. 20:1216-1223.
688 125. Giorgio E, Robyr D, Spielmann M, Ferrero E, Di Gregorio E, Imperiale D, et al. (2015):
689 A large genomic deletion leads to enhancer adoption by the lamin B1 gene: a second path to
690 autosomal dominant adult-onset demyelinating leukodystrophy (ADLD). *Hum Mol Genet*.
691 24:3143-3154.
692 126. Farh KK, Marson A, Zhu J, Kleinewietfeld M, Housley WJ, Beik S, et al. (2015): Genetic
693 and epigenetic fine mapping of causal autoimmune disease variants. *Nature*. 518:337-343.
694 127. GTEX Consortium (2015): Human genomics. The Genotype-Tissue Expression (GTEx)
695 pilot analysis: multitissue gene regulation in humans. *Science*. 348:648-660.
696 128. Lappalainen T, Sammeth M, Friedlander MR, t Hoen PA, Monlong J, Rivas MA, et al.
697 (2013): Transcriptome and genome sequencing uncovers functional variation in humans.
698 *Nature*. 501:506-511.
699 129. Ramasamy A, Trabzuni D, Guelfi S, Varghese V, Smith C, Walker R, et al. (2014):
700 Genetic variability in the regulation of gene expression in ten regions of the human brain.
701 *Nat Neurosci*. 17:1418-1428.
702 130. Battle A, Mostafavi S, Zhu X, Potash JB, Weissman MM, McCormick C, et al. (2014):
703 Characterizing the genetic basis of transcriptome diversity through RNA-sequencing of 922
704 individuals. *Genome Res*. 24:14-24.
705 131. Gamazon ER, Wheeler HE, Shah KP, Mozaffari SV, Aquino-Michaels K, Carroll RJ, et al.
706 (2015): A gene-based association method for mapping traits using reference transcriptome
707 data. *Nat Genet*. 47:1091-1098.
708 132. Zhou J, Troyanskaya OG (2015): Predicting effects of noncoding variants with deep
709 learning-based sequence model. *Nat Methods*. 12:931-934.
710 133. Cristino AS, Williams SM, Hawi Z, An JY, Bellgrove MA, Schwartz CE, et al. (2014):
711 Neurodevelopmental and neuropsychiatric disorders represent an interconnected molecular
712 system. *Mol Psychiatry*. 19:294-301.
713 134. Parikshak NN, Luo R, Zhang A, Won H, Lowe JK, Chandran V, et al. (2013): Integrative
714 functional genomic analyses implicate specific molecular pathways and circuits in autism.

Cell. 155:1008-1021.
135. Willsey AJ, Sanders SJ, Li M, Dong S, Tebbenkamp AT, Muhle RA, et al. (2013): Coexpression networks implicate human midfetal deep cortical projection neurons in the pathogenesis of autism. *Cell*. 155:997-1007.
136. Ziats MN, Rennert OM (2014): Identification of differentially expressed microRNAs across the developing human brain. *Mol Psychiatry*. 19:848-852.
137. Geschwind DH, Konopka G (2009): Neuroscience in the era of functional genomics and systems biology. *Nature*. 461:908-915.
138. An JY, Claudianos C (2016): Genetic heterogeneity in autism: From single gene to a pathway perspective. *Neurosci Biobehav Rev*. 68:442-453.
139. Xu X, Breen G, Chen CK, Huang YS, Wu YY, Asherson P (2010): Association study between a polymorphism at the 3'-untranslated region of CLOCK gene and attention deficit hyperactivity disorder. *Behav Brain Funct*. 6:48.
140. Kissling C, Retz W, Wiemann S, Coogan AN, Clement RM, Hunnerkopf R, et al. (2008): A polymorphism at the 3'-untranslated region of the CLOCK gene is associated with adult attention-deficit hyperactivity disorder. *Am J Med Genet B.* 147B:333-338.
141. Ozburn AR, Purohit K, Parekh PK, Kaplan GN, Falcon E, Mukherjee S, et al. (2016): Functional Implications of the CLOCK 3111T/C Single-Nucleotide Polymorphism. *Front Psychiatry*. 7:67.
142. Tong J, McKinley LA, Cummins TD, Johnson B, Matthews N, Vance A, et al. (2015): Identification and functional characterisation of a novel dopamine beta hydroxylase gene variant associated with attention deficit hyperactivity disorder. *World J Biol Psychiatry*. 16:610-618.
143. Wu C, Gong Y, Sun A, Zhang Y, Zhang C, Zhang W, et al. (2013): The human MTHFR rs4846049 polymorphism increases coronary heart disease risk through modifying miRNA binding. *Nutr Metab Cardiovasc Dis*. 23:693-698.
144. Spellicy CJ, Northrup H, Fletcher JM, Cirino PT, Dennis M, Morrison AC, et al. (2012): Folate metabolism gene 5,10-methylenetetrahydrofolate reductase (MTHFR) is associated with ADHD in myelomeningocele patients. *PloS One*. 7:e51330.
145. Gong Y, Wu CN, Xu J, Feng G, Xing QH, Fu W, et al. (2013): Polymorphisms in microRNA target sites influence susceptibility to schizophrenia by altering the binding of miRNAs to their targets. *Eur Neuropsychopharmacol*. 23:1182-1189.
146. Wu S, Zhang R, Nie F, Wang X, Jiang C, Liu M, et al. (2016): MicroRNA-137 Inhibits EFNB2 Expression Affected by a Genetic Variant and Is Expressed Aberrantly in Peripheral Blood of Schizophrenia Patients. *EBioMedicine*. 12:133-142.
147. Begemann M, Grube S, Papiol S, Malzahn D, Krampe H, Ribbe K, et al. (2010): Modification of cognitive performance in schizophrenia by complexin 2 gene polymorphisms. *Arch Gen Psychiatry*. 67:879-888.
148. Rousseau P, Le Discorde M, Mouillot G, Marcou C, Carosella ED, Moreau P (2003): The 14 bp deletion-insertion polymorphism in the 3' UT region of the HLA-G gene influences HLA-G mRNA stability. *Hum Immunol*. 64:1005-1010.
149. Guerini FR, Bolognesi E, Chiappedi M, Ghezzo A, Canevini MP, Mensi MM, et al. (2015): An HLA-G(*)14bp insertion/deletion polymorphism associates with the development of autistic spectrum disorders. *Brain Behav Immun*. 44:207-212.
150. Mondal K, Ramachandran D, Patel VC, Hagen KR, Bose P, Cutler DJ, et al. (2012): Excess variants in AFF2 detected by massively parallel sequencing of males with autism spectrum disorder. *Hum Mol Genet*. 21:4356-4364.

762 **TABLES**

763 **Table 1: 3'UTR variants identified in NDDs that have shown functional effects.**

Disorder	Gene	Variant	Global MAF (1000 genomes)	ExAC (frequency)[1]	Effect *in vitro*	Refs
FXS	*FMR1*	c.*746T>C rs183130936	C=0.0021	C=0.0009688	Removes RBP (HuR) binding site & reduces transcript stability	(16, 66)
TS	*SLITRK1*	Var321 rs191284403	T=0.0006	N/A	Strengthens miR-189 binding and increases its repression in luciferase assays	(11)
SLI	*ARHGEF39*	rs72727021	C=0.0469	C= 0.07911	Disrupts miR-215 binding site to increase *ARHGEF39* expression in a luciferase assays and human brain tissue	(22)
RTT	*MECP2*	c.1832G>C	-	N/A	Reduces the amount of *MECP2* mRNA expressed in patient blood cells	(69)
		c.2015G>A	-	N/A		
		c.4017T>A	-	N/A		
		c.4417G>A	-	N/A		
ADHD	*CLOCK*	rs1801260	A= 0.2296	N/A	Alt allele (T) decreases expression of *CLOCK* mRNA in Mouse Embryonic Fibroblasts	(139-141)
	DBH	rs129882	T=0.2554	N/A	Alt allele (C) decreases expression in luciferase assays	(142)
	MTHFR	rs4846049	T=0.3716	N/A	Alt allele (T) increases repression by miR-149 in luciferase assays. Also reduced *MTHFR* protein in patient blood cells	(143, 144)
SZ	*RGS4*	rs10759	T=0.2831	N/A	The Alt allele (T) in miR-124 site, prevents repression	(145)

					in luciferase assays	
	EFNB2	rs550067317	G=0.0004	N/A	The Alt (C) allele reduces regulation by miR-137 in luciferase assays	(146)
	CPLX2	rs3822674	C=0.4948	C=0.4912	The Alt (C) allele increases regulation by miR-498 in luciferase assays	(147)
ASD	*HLA-G*	14 bp indel	-	N/A	Presence of the 14bp indel increases stability of the mRNA	(148, 149)
	AFF2	ChrX:148076068[C>T]	-	N/A	The variant reduces expression in luciferase assays in HEK cells and increases it in N2A cells	(150)
ID	*CDK5RI*	c.*397C>G	-	0	Reduces the expression in luciferase assays in HEK and SK-N-BE cells	(12)
		c.*1904_*1905del	-	N/A		
		c.*2099_*2101del	-	N/A	Reduces the expression in luciferase assays in HEK cells	

[1]N/A indicates that variant is in a region not covered by reads in ExAc database

768 **FIGURES**

769

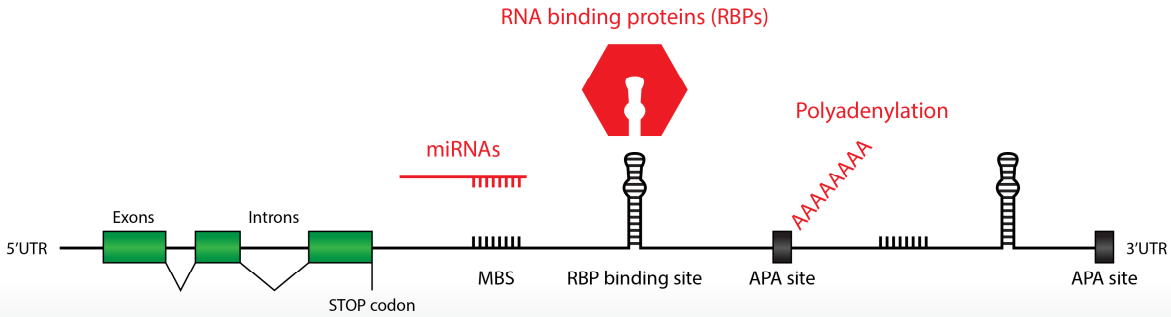
770 **Figure 1: Overview of the 3'UTR and its regulatory mechanisms.** 3'UTRs are untranslated
771 (non-coding) regions of an mRNA molecule, located after the stop codon of protein coding
772 genes. 3'UTRs contribute to regulation via multiple mechanisms, three of which are depicted
773 here; miRNA binding, RBP binding and APA. miRNAs interact with MBSs in the 3'UTR to
774 either block protein translation or degrade the mRNA molecule. RBPs interact with RBP
775 binding sites in the 3'UTR either regulate expression (via stabilising or degrading the
776 transcript), transport the mRNA to subcellular locations, or facilitating local translation of
777 proteins (e.g. at the synapse). APA is the mechanism by which one of multiple possible
778 polyadenylation sites may be used. APA determines the end point of the transcript and thus
779 the length of the 3'UTR. In turn, the length of the 3'UTR determines which and how many
780 regulatory elements (such as miRNA binding sites and RBP sites) are present to regulate the
781 transcript.

782
783

784
785

786 **Figure 2: Schematic workflow for the identification, analysis and validation of disease**
787 **associated variants in the 3'UTR.** Candidate variants can be identified using targeted or
788 high-throughput approaches. (01) Standard bioinformatic pipelines can be used to identify
789 variation in patient cohorts. (02) Using the known coordinates of all 3'UTR regions in the
790 human genome, variants located within the 3'UTRome can be identified (03) Different
791 algorithms (summarized in Supplementary Table S2) can be applied in order to identify
792 putative functional motifs in the 3'UTR that overlap with the variants found in step 02 and to
793 predict their functional outcome. Combining different algorithms at this step is
794 recommended in order to increase the accuracy of the prediction. (04) The information
795 obtained in step 03 can be used to design the appropriate experiments to test the functional
796 consequences of the variants. Taken together, the information obtained using this pipeline
797 will allow a description of the possible consequences of disease identified variants in the
798 3'UTRome to enhance our understanding of the molecular underpinnings of NDDs.

799



miRNAs regulate protein expression

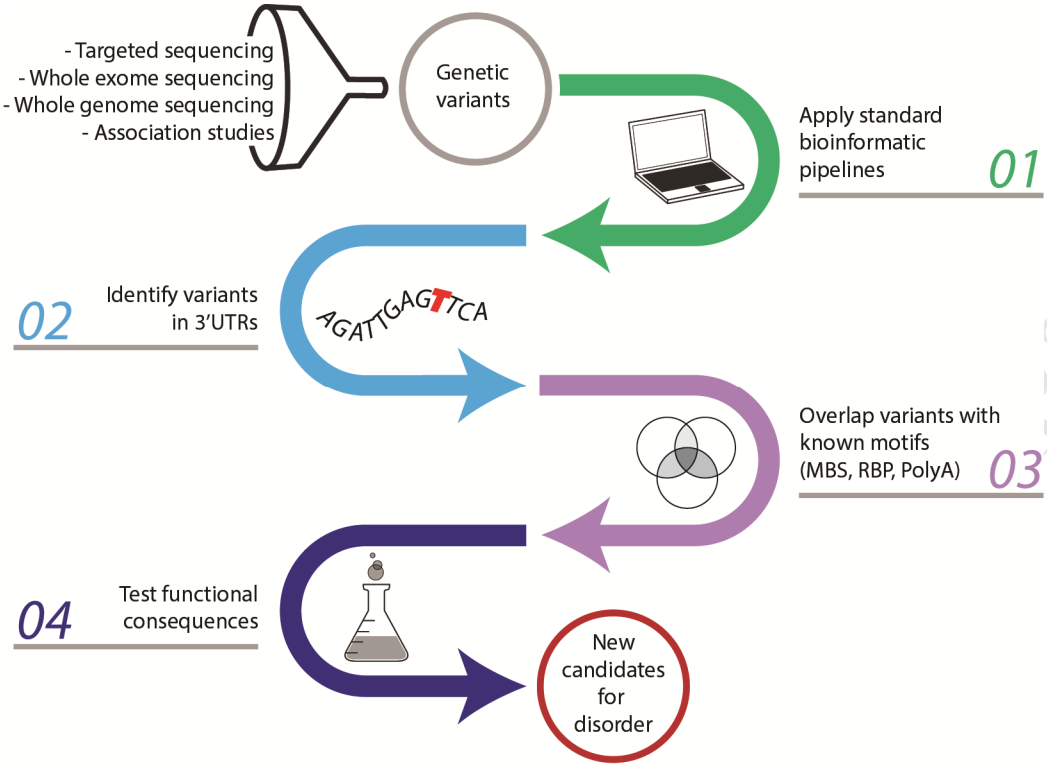
- Block protein translation
- Degrade mRNA transcript

RBPs regulate stability and sub-cellular localization of the mRNA transcript

- Stabilize/degrade mRNA transcript
- Transport mRNA
- Control local translation

APA sites regulate the 3'UTR length

- Determine presence of regulatory elements



Understanding Neurodevelopmental Disorders: The Promise of Regulatory Variation in the 3'UTRome

Supplemental Information

SUPPLEMENTARY DATA

Systematic review of literature

In order to identify NDD associated variants in 3'UTRs we performed a literature search with a combination of following Mesh Terms. In general, it was attempted to combine terms describing individual NDDs (ADHD, SZ, ASD, ID) with terms for next-generation sequencing (RNA-seq, WES, WGS) or GWAS and 3'UTRS. For a general search of variants in NDDs, the terms 'Non-coding RNA', 'Neurodevelopmental Disorder', '3'UTR' and 'Brain' were combined. The following paragraph shows all used combinations.

In total, 824 articles were obtained from the search. 41 of them were used to identify disease associated variants in 3'UTRs. The search was conducted on 08-08-2016 (specific disorders), 25-04-2016 (3'UTRs in general) and 01-11-2016 (disease + 3'UTR) using PubMed.

Variants that are located to 3'UTRs and that were correlated to a disorder were considered candidates for Table 1. The variation in study design, sample size and statistical method prevented a common statistical test for the variants. We therefore took every variant into account that was considered 'associated', 'linked' or 'correlated' by its corresponding study. The list was complemented with variants that were identified by studies in our own lab (ARHGEF39, BTN2A1, CENPJ, MTMR3) and variants that were found using the miRdSNP database. We extracted those SNPs from miRdSNP that were located in 20 bp distance of a MBS and also associated to ADHD, SZ or ASD. 11 variants in 10 genes were found this way.

For ADHD:

Search from **01-11-16**:

[ADHD] AND [3'UTRs]

Found 125 articles

Search from **08-08-16**:

[ADHD] AND [Next Generation Sequencing] AND [3'UTR]

[ADHD] AND [GWAS] AND [3'UTR].

Found 12 articles

For ASD:

Search from **01-11-16**:

[ASD] AND [3'UTRs]

Found 66 articles

Search from **08-08-16**:

[ASD] AND [Next Generation Sequencing] AND [3'UTR]

[ASD] AND [GWAS] AND [3'UTR].

Found 8 articles

Search from **01-11-16**:

[ID] AND [3'UTRs]

Found 180 articles

Search from **08-08-16**:

[ID] AND [Next Generation Sequencing] AND [3'UTR]

[ID] AND [GWAS] AND [3'UTR].

Found 18 articles

For SZ:

Search from **01-11-16**:

[Schizophrenia] AND [3'UTRs]

Found 152 articles

[Schizophrenia] AND [Next Generation Sequencing] AND [3'UTR]

[Schizophrenia] AND [GWAS] AND [3'UTR].

Found 14 articles

For 3'UTRs in general:

Search from **25-04-16**:

[Non-coding RNA] AND [Neurodevelopmental Disorder]

[3UTR AND Brain AND Mutation]

Found 249 articles

Supplementary Table S1: All 3'UTR variants identified via systematic review

Disorder	Gene	Variant	Genomic Coordinates (hg38)	RBPMap		Predicted MBS (miRSNP)	Reference	Known Candidate gene?	Functionally Validated?
				Input Coordinates for RBPMap	Outcome RBPs				
ADHD	SLC6A3 (DAT1)	rs28363170	chr5:1393747	chr5:1393737-1393757:-	ENOX, MBNL1, SRSF1, SRSF5	No prediction[1]	(1, 2)	Yes	Yes (3)
	CLOCK	rs1801260	chr4:55435202	chr4:55435192-55435212:--	DAZAP1, TRA2B	NA[3]	(4, 5)	Yes	Yes
	SNAP-25	rs3746544	chr20:10306436	chr20:10306426-10306446:-	FXR1, PABPC, RBM46, SNRNP70, SRSF1, SRSF5, SRSF7, TRA2B, YBX1	hsa-miR-3617, hsa-miR-3913-3p, hsa-miR-641	(6)	Yes	
		rs1051312	chr20:10306440	chr20:10306430-10306450:+	HuR, MATR3, PTBP1, SRSF1, SRSF3, SRSF5, TRA2B	hsa-miR-3646, hsa-miR-3664-3p, hsa-miR-510	(6-8)	Yes	
	MECP2	*c.1558insA	chrX:154028729	chrX:154028719-154028739:-	SRSF5	NA[2]	(9)	Yes	
	SLC9A9	rs1046706	chr3:143266258	chr3:143266248-143266268:+	DAZAP1, MSI1, PCBP3	NA[3]	(10)	Yes	
	DBH	rs129882	chr9:133658547	chr9:133658537-133658557:+	MBNL1, SRSF5, SRSF7	NA[3]	(11)	Yes	Yes
		rs1611115	chr9:133635393	chr9:133635383-	FMR1, HNRNPH1,	NA[3]	(11)	Yes	Yes

				133635403:+	HNRNPH2,LIN28A, SRSF1,SRSF10, SRSF2,SRSF5, TARDBP				
	MTHFR	rs4846049	chr1:11790308	chr1:11790298-11790318:+	SRSF2, TRA2B	NA[3]	(12)	Yes	Yes (13)
Schizophrenia	*PTPRA*	174620_174623 het_dupTGAT	No coordinates available	NA	NA	NA[2]	(14)	Yes	
	GZF1	rs7988	chr20:23372048	chr20:23372038-23372058:+	HNRNPL, HNRPLL, IGF2BP2, IGFBP3, KHDRBS2, KHDRBS3, PABPC1, PABPC4, PUM2, QKI, RBM46, RBMS3, SART3, SRSF3, SRSF5, ZCRB1	hsa-miR-3686	(7, 15)	No	
	CPLX2	rs3822674	chr5:175880253	chr5:175880243-175880263:-	CPEB2, CPEB4, HNRNPC, HNRNPCL1, MBNL1, MSI1, PTBP1, SRSF1, SRSF3, SRSF5, TUT1	hsa-miR-4287, hsa-miR-4685-3p, a-miR-498	(7, 16)	Yes	Yes
		rs56934064	chr5:175880290	chr5:175880280-175880300:+	HNRNPA1, HNRNPA1L2, HNRNPA2B1, HNRNPF, HNRNPH2	NA[3]	(16)	Yes	
	EFNB2	rs9520087	chr13:106491972	chr13:106491962-106491982:+	BRUNOL4, BRUNOL5, MBNL1, PTBP1, SRSF3, TARDBP	hsa-miR-155-3p, hsa-miR-4999-5p	(7, 17)	Yes	
		rs550067317	chr13:106491980	chr13:106491970-106491990:+	BRUNOL5, PTBP1, SRSF2, SRSF3, TARDBP	NA[3]	(18)	Yes	Yes
	DISC1/ DISC2	rs6675281	chr1:231818355	chr1:231818345-231818365:+	NOVA1, PABPC3, SRF3, SRSF5	hsa-miR-1294, hsa-miR-4710	(7, 19)	Yes	
	DISC1	rs821616	chr1:232008852	chr1:232008842-232008862:-	CUG-BP, MBNL1, SRSF1, SRSF5	hsa-miR-1243, hsa-miR-4423-3p	(7, 20)	Yes	
		r*s3737597*	Chr1:232037092	chr1:232037082-232037102:-	FXR1, HNRNPH1, PABPC3, PABPC5, PCBP1, PCBP3, RBM45, SRSF3, SRSF7, TRA2B	NA[3]	(21)	Yes	
	OLIG2	rs1059004	chr21:33028155	chr21:33028145-33028165:+	HNRNPK, PCBP1, SRSF2, SRSF3, SRSF5	hsa-miR-323a-5p, hsa-miR-3689d	(7, 22)	Yes	
		rs13046814	chr21:33029069	chr21:33029059-33029079:+	FUS, PCBP2, SRSF3, SRSF5	hsa-miR-2277-5p, hsa-miR-4639-3p, hsa-miR-744-5p	(7, 22)	Yes	
	AMACR	rs2278008	chr5:33989413	chr5:33989403-33989423:+	HNRPLL, MATR3, MBNL1, PTBP1, SRSF3, SRSF5	hsa-miR-942	(7, 23)	Yes	
	NFKBIA	rs8904	chr14:35402011	chr14:35402001-35402021:-	SRSF9, TARDBP, ZC3H10	hsa-miR-3121-5p	(7, 15)	Yes	
	LIF	rs737812	chr22:30243121	chr22:30243111-	CNOT4, PABPN1, PCBP1,	hsa-miR-373-5p,	(7, 24)	Yes	

				30243131:+	PCBP2, PTBP1, RBM6, SRSF3	hsa-miR-499b-5p, hsa-miR-616-5p, hsa-miR-617			
		rs929271	chr22:30242237	chr22:30242227-30242247:-	BRUNOL6,CUG-BP,HNRNPF, HNRNPU,MBNL1, PTBP1,SRSF2, SRSF5,TARDBP	NA[3]	(24)	Yes	
	DCLK1	rs9545297	chr13:35767531	chr13:35767521-35767541:+	BRUNOL6, CUG-BP, HNRNPU, HNRPLL, MBNL1, RBM41, SRSF3, TARDBP, TRA2B	NA[3]	(7, 25)	Yes	
	PLA2G12A	rs3087494	chr4:109710465	chr4:109710455-109710475:+	CPEB2, CPEB4, HNRNPA1L2, HNRNPCL1, HuR, MATR3, PTBP1, RALY, SRSF3, TIA1, TRA2B, U2AF2, ZC3H14	NA[3]	(26)	Yes	
	RGS4	rs10759	chr1:163076561	chr1:163076551-163076571:-	IGF2BP2, SRSF3	NA[3]	(27)	Yes	Yes
	COMT	rs165599	chr22:19969008	chr22:19968998-19969018:+	A1CF, KHDRBS2, KHDRBS3, MSI1, PABPC5, PUM2, RBM42, RBM46, RBMS3, ZCRB1	NA[3]	(28)	Yes	
	CTLA4	rs3087243	chr2:203874196	chr2:203874186-203874206:+	KHDRBS1, QKI, RBMS1, RBMS3, SRSF3	NA[3]	(29)	Yes	
Specific Language Disorder (SLI)	*ARHGEF39*	rs72727021	chr9:35661946	chr9:35661936-35661956:+	SRSF3, SRSF5	NA[3]	(30)	No	Yes
	BTN2A1	c.277 G>C	chr6:26468826	chr6:26468816-26468836:+	CUG-BP, PABPC5, PAPN1, RBM24, TARDBP, TRA2B	NA[2]	(30)	No	
	CENPJ	c.16G>T	chr13:24883161	chr13:24883151-24883171:-	QKI, SRSF1, SRSF3, TRA2B	NA[2]	(30)	No	
	MTMR3	c.70C>T	chr22:30025871	chr22:30025861-30025881:+	CNOT4, ENOX1	NA[2]	(30)	No	
Language and reading impairments	*CCDC136*	rs59197085	chr7:128820702	chr7:128820692-128820712:+	PTBP1, SRSF3, SRSF5	No prediction[1]	(31)	No	
	DYX1C1	rs57809907	chr15:55430684	chr15:55430674-55430694:+	A1CF, HuR, MBNL1, PTBP1, SRSF3, ZC3H14, ZCRB1, ZNF638	No prediction[1]	(32)	Yes	
Tourette's Syndrome	*SLITRK1*	Var321 (mutation)	chr13:83878728	chr13:83878718-83878738:+	SRSF3	NA[2]	(33)	Yes	Yes
Autism Spectrum Disorder (ASD)	*HLA-G*	14 bp indel	Chr6:1093208	chr6:1093198-1093218:+	SRSF5	NA[2]	(34)	Yes	
	AFF2	ChrX:148.076.068[C>T]	chrX:148994538	chrX:148994528-148994548:+	HNRNPA1, HNRNPA1L2, HNRNPA2B1, MBNL1,SRSF3, SRSF5,	NA[2]	(35)	Yes	Yes

					ZCRB1				
		ChrX:148075200[T>C]	chrX:148993670	chrX:148993660-148993680:+	HNRNPL, HNRPLL, RBM41, RBMS1, RBMS3, SRSF3, SRSF5, TUT1	NA[2]	(35)	Yes	
	STX1A	rs867500	chr7:73700110	chr7:73700100-73700120:+	HNRNPH2, RBM5, SRSF1, SRSF9, TARDBP	hsa-miR-1204	(36)	Yes	Yes
	APC	rs1804197	chr5:112844212	chr5:112844202-112844222:+	A1CF, FXR1, KHDRBS1, KHDRBS2, KHDRBS3, PABPC1, PABPC3, PABPC4, PABPC5, PABPN1, SART3, ZCRB1	hsa-miR-335-3p, hsa-miR-4282	(7, 37)	No	
	NLGN4X	chrX:5818136	chrX:5890095	chrX:5890085-5890105:-	CNOT4, KHDRBS1, KHDRBS2, KHDRBS3, PUM2, RBMS1, RBMS3, SRSF3	NA[2]	(38)	Yes	
		chrX:5820149–50	chrX:5892108	chrX:5892098-5892118:-	CPEB4, MBNL1, SRSF2, TIA1, TRA2B, U2AF2	NA[2]	(38)	Yes	
	SLC6A4	HTT-3′UTR-SNP	No coordinates available	NA	NA	NA[2]	(39)	Yes	
	MECP2	1461*139 G->A	chrX:154030343	chrX:154030333-154030353:-	HNRNPK, HNRPLL, IGF2BP2, PABPC3, RBM46, SNRPA, YBX1	NA[2]	(40)	Yes	
		c.1832G>C	chrX:154030065	chrX:154030055-154030075:-	CUG-BP, FMR1, FXR2, HNRPLL, MBNL1, SRSF2, SRSF7	NA[2]	(41)	Yes	Yes
		c.2015G>A	chrX:154029882	chrX:154029872-154029892:-	CNOT4, ENOX1, HNRNPA1, HNRNPH1, HNRNPH2, PABPN1, RBM24, SRSF10, SRSF2, TARDBP, TRA2A	NA[2]	(41)	Yes	Yes
		c.4017T>A	chrX:154027880	chrX:154027870-154027890:-	HNRNPK, HNRPLL, RBM42, RBM46, SRSF2, SRSF3	NA[2]	(41)	Yes	Yes
		c.4417G>A	chrX:154027480	chrX:154027470-154027490:-	NOVA1, SRSF5, SRSF7	NA[2]	(41)	Yes	Yes
		c.1655G>A	chrX:154030242	chrX:154030232-154030252:-	CPEB2, CPEB4, HNRNPC, HNRNPCL1, HuR, KHDRBS2, KHDRBS3, MBNL1, PABPC3, PTBP1, RALY, RBM41, SRSF3, SRSF5, TIA1, U2AF2, ZC3H14	NA[2]	(41)	Yes	
		c.2322T>G	chrX:154029575	chrX:154029565-154029749:-	DAZAP1, ENOX1, ESRP2, FMR1, FXR1, G3BP2, HNRNPA1, HNRNPA2B1,	NA[2]	(41)	Yes	

					HNRNPF, HNRNPH1, HNRNPH2, MBNL1, PTBP1, RBM28, RBM42, RBM5, SAMD4A, SRSF1, SRSF2, SRSF3, SRSF5, SRSF7, SRSF9, TARDBP				
		c.2829C>A	chrX:154029068	chrX:154029058-154029078:-	MBNL1, PCBP1, PCBP2, PTBP1, SRSF2, SRSF3, SRSF5	NA[2]	(41)	Yes	
		c.3198G>A	chrX:154028699	chrX:154028852-154028872:-	MBNL1, PCBP1, PCBP3, PTBP1, SRSF3, SRSF5	NA[2]	(41)	Yes	
		c.6037A>C	chrX:154025860	chrX:154025850-154025870:-	HNRNPA1L2,HNRNPA2B, MBNL1, PCBP2, SRSF5	NA[2]	(41)	Yes	
		c.6948ins(AT)	chrX:154024949	chrX:154024939:-154024959:-	A1CF, CPEB2, CPEB4, CUG-BP, DAZAP1, HNRNPC, HNRNPCL1, HuR, MBNL1, PCBP3, RALY, RBM41, SFPQ, SRSF2, SRSF3, TIA1, TRA2B, U2AF2, ZC3H14	NA[2]	(41)	Yes	
		c.9209C>T	chrX:154022688	chrX: 154022678-154022698:-	CUG-BP, HNRNPF, PTBP1, RBM38, SRSF1, SRSF5, TARDBP, TRA2B	NA[2]	(41)	Yes	
		c.9317A>C	chrX:154022580	chrX: 154022570-154022590:-	HNRNPK, PCBP1, PCBP2, PCBP3, PTBP1, SRSF3	NA[2]	(41)	Yes	
		T>C c.6809	chrX:154025007	chrX:154024997-154025017:-	MBNL1, SRSF3, TARDBP	NA[2]	(9)	Yes	
		G>C c.1638	chrX:154028649	chrX:154028639-154028659:-	CUG-BP, HNRNPA1, HNRNPA1L2, MBNL1, RBM28, ZC3H10	NA[2]	(9)	Yes	
		c.1470G > A	chrX:154030427	chrX:154030417-154030437:-	HNRNPK, HNRNPL, SRSF7, YBX1	NA[2]	(41)	Yes	
		c.2005G > A	chrX:154029892	chrX:154029882-154029902:-	CNOT4, ENOX1, FMR1, FXR2, HNRNPA1, HNRNPA2B1, HNRNPH2, RBM24, SRSF2, TARDBP	NA[2]	(41)	Yes	

		c.2228G > T	chrX:154029669	chrX:154029659-154029679:-	ENOX1, FMR1, FXR1, SRSF3, SRSF7	NA[2]	(41)	Yes	
		c.4118G > A	chrX:154027779	chrX:154027769-154027789:-	ESRP2, HNRNPA1, HNRNPA2B1, HNRNPF, HNRNPH1, HNRNPH2, PABPN1, SRSF2, TRA2A	NA[2]	(41)	Yes	
		c.4167G > A	chrX:154027730	chrX:154027720-154027740:-	CUG-BP, MBNL1, PTBP1, SRSF2, SRSF3	NA[2]	(41)	Yes	
		c.5119C > T	chrX:154026778	chrX:154026768-154026788:-	CUG-BP, HNRNPA1, MBNL1, SRSF2, SRSF5	NA[2]	(41)	Yes	
		c.5339G > C	chrX:154026558	chrX:154026548-154026568:-	SNRPA	NA[2]	(41)	Yes	
	OXTR	rs7632287	chr3:8749760	chr3:8749750-8749770:+	RBM41, SNRPA, SRSF3, SRSF7	NA[3]	(42)	Yes	
		rs237884	chr3:8751899	chr3:8751889-8751909:+	KHDRBS3, MBNL1, PUM2, SRSF5	NA[3]	(42)	Yes	
Intellectual Disability	*CDK5RI*	c.1005G>A	Chr17:32488404	chr17:32488394-32488414:+	MBNL1, PTBP1, SRSF2, SRSF5	NA[2]	(43)	Yes	
		c.1043G>A	Chr17:32488442	chr17:32488432-32488452:+	CUG-BP, MBNL1, SRSF2, TUT1	NA[2]	(43)	Yes	
		c.2160C>T	chr17:32489559	chr17:32489549-32489569:+	HuR, KHDRBS3, PTBP1, PUM2, RBMS1, SRSF3, TRA2B	NA[2]	(43)	Yes	
		c.2254C>G	chr17:32489653	chr17:32489643-32489663:+	CUG-BP, MBNL1, SRSF1, SRSF2	NA[2]	(43)	Yes	
		c.3452G>A	chr17:32490851	chr17: 32490841-32490861:+	CPEB2, CPEB4, HNRNPC, HNRNPCL1, HuR, MATR3, MBNL1, PTBP1, RALY, SRSF3, SRSF5, TIA1, U2AF2, ZC3H14, ZNF638	NA[2]	(43)	Yes	
		c.*71 G>A	chr17:32488916	chr17:32488906-32488926:+	CUG-BP,HNRNPA1, HNRNPA1L2, HNRNPA2B1, HNRNPF, SRSF2, SRSF5, TARDBP, TIA1	NA[2]	(44)	Yes	
		rs8192474	chr17:32488663	chr17:32488653-32488673:+	NOVA1, PCBP1, PCBP2, PCBP3, PTBP1, SRSF3, SRSF5	hsa-miR-1915-3p, hsa-miR-4441, hsa-miR-548ac	(44)	Yes	
		rs138054348	chr17:32490069	chr17:32490059-32490079:+	CUG-BP, HNRNPF, SRSF1, SRSF9	hsa-miR-4691-3p, hsa-miR-766-3p	(44)	Yes	
		rs735555	chr17:32490432	chr17:32490422-32490442:+	BRUNOL4, BRUNOL5, CPEB4, HuR, MBNL1, PTBP1, RBM38, RBM6, SRSF2, SRSF3, TARDBP,	NA[2]	(44)	Yes	

					ZC3H14				
		rs115744590	chr17:32490719	chr17:32490709-32490729:+	CPEB4, CUG-BP, MATR3, MBNL1, PTBP1, RBM41, SRSF1, SRSF2, SRSF3, SRSF5, TIA1, TRA2B, U2AF2, ZC3H14	hsa-miR-338-5p, hsa-miR-3680-3p	(44)	Yes	
		c.*397C>G	Chr17:32489242	chr17:32489232-32489252:+	A1CF, CUG-BP, HNRNPM, HuR, MBNL1, PTBP1, SFPQ	NA[3]	(44)	Yes	Yes
		c.*649_*659del	Chr17:32489494	chr17:32489484-32489504:+	CUG-BP, MBNL1, SRSF2, SRSF3	NA[3]	(44)	Yes	
		c.*1904_*1905del	Chr17:32490749	chr17:32490739-32490759:+	BRUNOL4, BRUNOL5, CPEB4, HNRNPC, HNRNPCL1, MATR3, PTBP1, RALY, RBM24, RBM38, SRSF1, SRSF9, TARDBP, TIA1	NA[3]	(44)	Yes	Yes
		c.*2099_*2101del	Chr17:32490944	chr17:32490934-32490954:+	KHDRBS1, KHDRBS2, KHDRBS3, NOVA1, PABPC1, PABPC4, PABPC5, PABPN1, SART3, TRA2B	NA[3]	(44)	Yes	Yes
	CDK5	rs9278	chr7:151053893	chr7:151053883-151053903:-	BRUNOL6, HNRNPF, HNRNPH2, MBNL1, SFPQ	hsa-miR-3064-5p, hsa-miR-3620, hsa-miR-3944-3p, hsa-miR-4269	(44)	Yes	
	FMR1	c.*746T>C	chrX:147949590	chrX:147949580-147949600:+	BRUNOL4, BRUNOL5, BRUNOL6, CPEB2, CPEB4, HNRNPC, HNRNPCL1, HNRNPM, HNRNPU, HuR, MBNL1, RALY, SRSF2, TIA1, TRA2B, U2AF2, ZC3H14, ZNF638	NA[2]	(45, 46)	Yes	Yes
		c.*1867G>A	chrX:147950711	chrX:147950701-147950721:+	DAZAP1, KHDRBS1, KHDRBS3, MSI1, PABPC1, PABPC5, RBMS1, RBMS3, SART3, SRSF2, TRA2B	NA[2]	(45)	Yes	
		c.*23T>C	chrX:147948867	chrX:147948857-147948877:+	A1CF, PCBP3, PTBP1, RBM41, RBMS1, RBMS3, SRSF3, TUT1, U2AF2	NA[2]	(47)	Yes	
		c.*2035C>T	chrX:147950879	chrX:147950869-147950889:+	CUG-BP, MBNL1, PTBP1, RBM42, SRSF3	NA[2]	(47)	Yes	
		10 unnamed variants	No coordinates available	NA	NA	NA	(47)	Yes	
	MECP2	1461*93G > A	chrX:154030436	chrX:154030426-	HNRNPK,SRSF3, YBX1	NA[2]	(48)	Yes	

				154030446:-					
Rett-Syndrome	*MECP2*	c.1461+98insA	chrX:154030436	chrX:154030426-154030446:-	HNRNPK,SRSF3, YBX1	NA[2]	(49, 50)	Yes	
		c.9961C>G	chrX:154021936	chrX: 154021926-154021946:-	A1CF, HNRNPC, HNRNPCL1, KHDRBS3, PABPC5, QKI, RBMS3, ZC3H14, ZCRB1	NA[2]	(49)	Yes	
		c.9964delC	chrX:154021933	chrX:154021923-154021943:-	A1CF, HNRNPC, HNRNPLC1, KHDRBS3, PABPC5, QKI, RBMS3, ZC3H14, ZCRB1	NA[2]	(49)	Yes	
		c.1461+9G>A	chrX:154030436	chrX:154030426-154030446:-	HNRNPK,SRSF3, YBX1	NA[2]	(49)	Yes	
		c.2595G>A	chrX:154029302	chrX: 154029292-154029312:-	SRSF1, SRSF5	NA[2]	(49)	Yes	
		c.1461+92C>G	chrX:154030436	chrX:154030426-154030446:-	HNRNPK,SRSF3, YBX1	NA[2]	(50)	Yes	
Anxiety and Depression	*MAP2K5*	rs41305272	chr15:67807105	chr15:67807095-67807115:+	CUG-BP, HNRNPA1, MBNL1, QKI, TARDBP, TRA2B, ZCRB1	NA[3]	(51, 52)	Yes	
	P2RX7	rs1653625	chr12:121185082	chr12:121185072-121185092:+	IGF2BP2, IGF2BP3, KHDRBS1, KHDRBS2, KHDRBS3, PABPC1, PABPC3, PABPC4, PABPC5, PABPN1, SART3	NA[3]	(53)	Yes	

[1]The SNP was not predicted to affect an existing MBS

[2]The variant was not reported as SNP with rs-identifier. As such it was not usable in the miRSNP program

[3]Not possible as MiRSNP webpage unavailable

Supplementary Table S2. Online tools to examine functional motifs in 3'UTRs

MBS	**ALGORITHMS TO PREDICT MBSs IN ANNOTATED 3'UTRs** Prone to false positive/negative predictions. Combination of different algorithms can increase the accuracy. Targetscan (http://www.targetscan.org/vert_71/) PITA (https://genie.weizmann.ac.il/pubs/mir07/mir07_data.html) DIANA-microT4 (http://diana.imis.athena-innovation.gr/DianaTools/index.php?r=microtv4/index) mirMap (http://mirmap.ezlab.org/) RNA22 (https://cm.jefferson.edu/rna22/Interactive/) miRANDA (http://www.microrna.org/microrna/home.do) microRNAmap (http://mirnamap.mbc.nctu.edu.tw/) RNAhybrid (https://bibiserv2.cebitec.unibielefeld.de/rnahybrid?id=rnahybrid_view_submission) MiRBridge (http://mirsystem.cgm.ntu.edu.tw/) PICTAR (http://pictar.mdc-berlin.de/) miRtarget (Access via miRWalk), microTar (http://tiger.dbs.nus.edu.sg/microtar/)
	ONLINE DATABASES OF PREDICTED MBSs IN ANNOTATED 3'UTRs Allows the combination of different algorithms in the prediction. Curated so may not always include the latest versions of the algorithms. miRDB (http://www.mirdb.org/miRDB/) miRWalk (http://zmf.umm.uni-heidelberg.de/apps/zmf/mirwalk2/) miRGate (http://mirgate.bioinfo.cnio.es/miRGate/) miRANDA (http://www.microrna.org/microrna/home.do)
	ALGORITHMS THAT COMBINE EMPIRICAL AND THEORETICAL MBS DATA The interaction between mRNA and miRNA can vary between tissue and cell types and might be affected by experimental setup. TarPmiR (http://hulab.ucf.edu/research/projects/miRNA/TarPmiR/) MEME-DREME (http://meme-suite.org/index.html)
	ONLINE DATABASES OF EXPERIMENTALLY VALIDATED mRNA-miRNA INTERACTIONS The interaction between mRNA and miRNA can vary between tissue and cell types and might be affected by experimental setup miRDB (http://www.mirdb.org/miRDB/) miRGate (http://mirgate.bioinfo.cnio.es/miRGate/) miRANDA (http://www.microrna.org/microrna/home.do)

	ONLINE DATABASES PREDICTING EFFECTS OF MBS VARIANTS Requires established SNPs with rs-numbers as input. Therefore rare mutations or rare variants may not be applicable. miRSNP (http://bioinfo.bjmu.edu.cn/mirsnp/search/) miRVaS (http://mirvas.bioinf.be/index.html) STARmiR (http://sfold.wadsworth.org/cgi-bin/starmir.pl)
	ONLINE DATABASES OF DISEASE ASSOCIATED SNPs IN OR NEAR MBSs Requires established SNPs with rs-numbers as input. Therefore rare mutations or rare variants may not be applicable. miRSNP (http://bioinfo.bjmu.edu.cn/mirsnp/search/) miRdSNP (http://mirdsnp.ccr.buffalo.edu/)
	ONLINE DATABASES OF miRNA EXPRESSION PATTERNS Show the expression of miRNAs at a given time in different tissues. The miRNA expression pattern might differ during development. miRANDA (http://www.microrna.org/microrna/home.do) miRGator (http://mirgator.kobic.re.kr/) miRWalk (http://zmf.umm.uni-heidelberg.de/apps/zmf/mirwalk2/)
RPB	**ALGORITHMS TO PREDICT RBP BINDING SITES** RBP binding sites are often very short motifs and depend on the secondary structure of the RNA. As such, they are difficult to predict based on sequence alone. Empirical data can be included to increase the accuracy but there may be biases based on cell type or experimental setup. RBPMap (http://rbpmap.technion.ac.il/) RBPDB (http://rbpdb.ccbr.utoronto.ca/) driMUST (http://drimust.technion.ac.il/index.html) MEMERIS (http://www.bioinf.uni-freiburg.de/~hiller/MEMERIS/) CIS-BP-RNA (http://cisbp-rna.ccbr.utoronto.ca/index.php) RBPmotif (http://www.rnamotif.org/) catRAPID (http://s.tartaglialab.com/page/catrapid_omics_group) ATtRACT (https://attract.cnic.es/) rMAPS (http://rmaps.cecsresearch.org) POSTAR (http://lulab.life.tsinghua.edu.cn/postar/)

	ONLINE DATABASES OF EXPERIMENTALLY VALIDATED mRNA-RBP INTERACTIONS The interaction between mRNA and RBP can vary between tissue and cell types and might be affected by experimental setup RBPDB (http://rbpdb.ccbr.utoronto.ca/) doRiNA (http://dorina.mdc-berlin.de/) CIS-BP-RNA (http://cisbp-rna.ccbr.utoronto.ca/index.php) RBP-Var (http://www.rbp-var.biols.ac.cn/) RAID (http://www.rna-society.org/raid/) CLIPdb (http://lulab.life.tsinghua.edu.cn/clipdb/) POSTAR (http://lulab.life.tsinghua.edu.cn/postar/)
	CURATED DATABASES OF RNA BINDING PROTEINS As a curated database, it is very reliable but might not always be completely up to date. READDB (http://darwin.soic.iupui.edu/)
PolyA	**ALGORITHMS TO PREDICT POLY-A SITES FROM USER PROVIDED SEQUENCES** Accurate for the prediction of different polyA sites, but should be complemented with experimental evidence to confirm the presence of different 3'UTRs. polyApred (http://www.imtech.res.in/raghava/polyapred/) poly(A) Signal (http://dnafsminer.bic.nus.edu.sg/PolyA.html) Dragon PolyA Spotter (DPS) (http://www.cbrc.kaust.edu.sa/dps/)
	ONLINE DATABASES OF EXPERIMENTALLY DETERMINED POLY-A AND APA EVENTS As curated databases, they are very reliable but might not always be completely up to date. APADB (http://tools.genxpro.net/apadb/) PACDB (http://harlequin.jax.org/pacdb/) PolyAsite (http://polyasite.unibas.ch/)
3'UTRs	**CURATED DATABASE OF MULTIPLE MOTIF TYPES IN UTR SEQUENCES** As a curated database, it is very reliable but might not always be completely up to date. UTRdb (http://utrdb.ba.itb.cnr.it/) AURA (http://aura.science.unitn.it/)

Supplementary References

1. Ettinger U, Merten N, Kambeitz J (2016): Meta-analysis of the association of the SLC6A3 3'-UTR VNTR with cognition. *Neuroscience and biobehavioral reviews*. 60:72-81.
2. Sano A, Kondoh K, Kakimoto Y, Kondo I (1993): A 40-nucleotide repeat polymorphism in the human dopamine transporter gene. *Human genetics*. 91:405-406.
3. VanNess SH, Owens MJ, Kilts CD (2005): The variable number of tandem repeats element in DAT1 regulates in vitro dopamine transporter density. *BMC genetics*. 6:55.
4. Xu X, Breen G, Chen CK, Huang YS, Wu YY, Asherson P (2010): Association study between a polymorphism at the 3'-untranslated region of CLOCK gene and attention deficit hyperactivity disorder. *Behavioral and brain functions : BBF*. 6:48.
5. Kissling C, Retz W, Wiemann S, Coogan AN, Clement RM, Hunnerkopf R, et al. (2008): A polymorphism at the 3'-untranslated region of the CLOCK gene is associated with adult attention-deficit hyperactivity disorder. *American journal of medical genetics Part B, Neuropsychiatric genetics : the official publication of the International Society of Psychiatric Genetics*. 147:333-338.
6. Forero DA, Arboleda GH, Vasquez R, Arboleda H (2009): Candidate genes involved in neural plasticity and the risk for attention-deficit hyperactivity disorder: a meta-analysis of 8 common variants. *Journal of psychiatry & neuroscience : JPN*. 34:361-366.
7. Bruno AE, Li L, Kalabus JL, Pan Y, Yu A, Hu Z (2012): miRdSNP: a database of disease-associated SNPs and microRNA target sites on 3'UTRs of human genes. *BMC genomics*. 13:44.
8. Kim JW, Biederman J, Arbeitman L, Fagerness J, Doyle AE, Petty C, et al. (2007): Investigation of variation in SNAP-25 and ADHD and relationship to co-morbid major depressive disorder. *American journal of medical genetics Part B, Neuropsychiatric genetics : the official publication of the International Society of Psychiatric Genetics*. 144B:781-790.
9. Shibayama A, Cook EH, Jr., Feng J, Glanzmann C, Yan J, Craddock N, et al. (2004): MECP2 structural and 3'-UTR variants in schizophrenia, autism and other psychiatric diseases: a possible association with autism. *American journal of medical genetics Part B, Neuropsychiatric genetics : the official publication of the International Society of Psychiatric Genetics*. 128B:50-53.
10. Markunas CA, Quinn KS, Collins AL, Garrett ME, Lachiewicz AM, Sommer JL, et al. (2010): Genetic variants in SLC9A9 are associated with measures of attention-deficit/hyperactivity disorder symptoms in families. *Psychiatric genetics*. 20:73-81.
11. Tong J, McKinley LA, Cummins TD, Johnson B, Matthews N, Vance A, et al. (2015): Identification and functional characterisation of a novel dopamine beta hydroxylase gene variant associated with attention deficit hyperactivity disorder. *The world journal of biological psychiatry : the official journal of the World Federation of Societies of Biological Psychiatry*. 16:610-618.

12. Spellicy CJ, Northrup H, Fletcher JM, Cirino PT, Dennis M, Morrison AC, et al. (2012): Folate metabolism gene 5,10-methylenetetrahydrofolate reductase (MTHFR) is associated with ADHD in myelomeningocele patients. *PloS one*. 7:e51330.

13. Wu C, Gong Y, Sun A, Zhang Y, Zhang C, Zhang W, et al. (2013): The human MTHFR rs4846049 polymorphism increases coronary heart disease risk through modifying miRNA binding. *Nutrition, metabolism, and cardiovascular diseases : NMCD*. 23:693-698.

14. Xing J, Wang C, Kimura H, Takasaki Y, Kunimoto S, Yoshimi A, et al. (2014): Resequencing and association analysis of PTPRA, a possible susceptibility gene for schizophrenia and autism spectrum disorders. *PloS one*. 9:e112531.

15. Arinami T, Ohtsuki T, Ishiguro H, Ujike H, Tanaka Y, Morita Y, et al. (2005): Genomewide high-density SNP linkage analysis of 236 Japanese families supports the existence of schizophrenia susceptibility loci on chromosomes 1p, 14q, and 20p. *American journal of human genetics*. 77:937-944.

16. Begemann M, Grube S, Papiol S, Malzahn D, Krampe H, Ribbe K, et al. (2010): Modification of cognitive performance in schizophrenia by complexin 2 gene polymorphisms. *Archives of general psychiatry*. 67:879-888.

17. Zhang R, Zhong NN, Liu XG, Yan H, Qiu C, Han Y, et al. (2010): Is the EFNB2 locus associated with schizophrenia? Single nucleotide polymorphisms and haplotypes analysis. *Psychiatry research*. 180:5-9.

18. Wu S, Zhang R, Nie F, Wang X, Jiang C, Liu M, et al. (2016): MicroRNA-137 Inhibits EFNB2 Expression Affected by a Genetic Variant and Is Expressed Aberrantly in Peripheral Blood of Schizophrenia Patients. *EBioMedicine*. 12:133-142.

19. Hodgkinson CA, Goldman D, Jaeger J, Persaud S, Kane JM, Lipsky RH, et al. (2004): Disrupted in schizophrenia 1 (DISC1): association with schizophrenia, schizoaffective disorder, and bipolar disorder. *American journal of human genetics*. 75:862-872.

20. Callicott JH, Straub RE, Pezawas L, Egan MF, Mattay VS, Hariri AR, et al. (2005): Variation in DISC1 affects hippocampal structure and function and increases risk for schizophrenia. *Proceedings of the National Academy of Sciences of the United States of America*. 102:8627-8632.

21. Saetre P, Agartz I, De Franciscis A, Lundmark P, Djurovic S, Kahler A, et al. (2008): Association between a disrupted-in-schizophrenia 1 (DISC1) single nucleotide polymorphism and schizophrenia in a combined Scandinavian case-control sample. *Schizophrenia research*. 106:237-241.

22. Georgieva L, Moskvina V, Peirce T, Norton N, Bray NJ, Jones L, et al. (2006): Convergent evidence that oligodendrocyte lineage transcription factor 2 (OLIG2) and interacting genes influence susceptibility to schizophrenia. *Proceedings of the National Academy of Sciences of the United States of America*. 103:12469-12474.

23. Bespalova IN, Durner M, Ritter BP, Angelo GW, Rossy-Fullana E, Carrion-Baralt J, et al. (2010): Non-synonymous variants in the AMACR gene are associated with schizophrenia. *Schizophrenia research*. 124:208-215.

24. Okahisa Y, Ujike H, Kunugi H, Ishihara T, Kodama M, Takaki M, et al. (2010): Leukemia inhibitory factor gene is associated with schizophrenia and working memory function. *Progress in neuro-psychopharmacology & biological psychiatry*. 34:172-176.

25. Havik B, Degenhardt FA, Johansson S, Fernandes CP, Hinney A, Scherag A, et al. (2012): DCLK1 variants are associated across schizophrenia and attention deficit/hyperactivity disorder. *PloS one*. 7:e35424.

26. Yang G, Xu H, Zhang H, Yu Q, Wu Y, Shi J, et al. (2016): Association between PLA2G12A Polymorphisms and Schizophrenia in a Han Chinese Population from Northeast China. *PloS one*. 11:e0159584.

27. Gong Y, Wu CN, Xu J, Feng G, Xing QH, Fu W, et al. (2013): Polymorphisms in microRNA target sites influence susceptibility to schizophrenia by altering the binding of miRNAs to their targets. *European neuropsychopharmacology : the journal of the European College of Neuropsychopharmacology*. 23:1182-1189.

28. Chien YL, Liu CM, Fann CS, Liu YL, Hwu HG (2009): Association of the 3' region of COMT with schizophrenia in Taiwan. *Journal of the Formosan Medical Association = Taiwan yi zhi*. 108:301-309.

29. Jones AL, Holliday EG, Mowry BJ, McLean DE, McGrath JJ, Pender MP, et al. (2009): CTLA-4 single-nucleotide polymorphisms in a Caucasian population with schizophrenia. *Brain, behavior, and immunity*. 23:347-350.

30. Devanna P, Chen XS, Ho J, Gajewski D, Smith SD, Gialluisi A, et al. (2017): Next-gen sequencing identifies non-coding variation disrupting miRNA-binding sites in neurological disorders. *Molecular psychiatry*. [published online ahead of print Mar 14; doi: 10.1038/mp.2017.30]

31. Gialluisi A, Newbury DF, Wilcutt EG, Olson RK, DeFries JC, Brandler WM, et al. (2014): Genome-wide screening for DNA variants associated with reading and language traits. *Genes, brain, and behavior*. 13:686-701.

32. Newbury DF, Paracchini S, Scerri TS, Winchester L, Addis L, Richardson AJ, et al. (2011): Investigation of dyslexia and SLI risk variants in reading- and language-impaired subjects. *Behavior genetics*. 41:90-104.

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

34. Guerini FR, Bolognesi E, Chiappedi M, Ghezzo A, Canevini MP, Mensi MM, et al. (2015): An HLA-G(*)14bp insertion/deletion polymorphism associates with the development of autistic spectrum disorders. *Brain, behavior, and immunity*. 44:207-212.

35. Mondal K, Ramachandran D, Patel VC, Hagen KR, Bose P, Cutler DJ, et al. (2012): Excess variants in AFF2 detected by massively parallel sequencing of males with autism spectrum disorder. *Human molecular genetics*. 21:4356-4364.

36. Nakamura K, Anitha A, Yamada K, Tsujii M, Iwayama Y, Hattori E, et al. (2008): Genetic and expression analyses reveal elevated expression of syntaxin 1A (STX1A) in high functioning autism. *The international journal of neuropsychopharmacology / official scientific journal of the Collegium Internationale Neuropsychopharmacologicum*. 11:1073-1084.

37. Zhou XL, Giacobini M, Anderlid BM, Anckarsater H, Omrani D, Gillberg C, et al. (2007): Association of adenomatous polyposis coli (APC) gene polymorphisms with autism spectrum disorder (ASD). *American journal of medical genetics Part B, Neuropsychiatric genetics : the official publication of the International Society of Psychiatric Genetics*. 144B:351-354.

38. Steinberg KM, Ramachandran D, Patel VC, Shetty AC, Cutler DJ, Zwick ME (2012): Identification of rare X-linked neuroligin variants by massively parallel sequencing in males with autism spectrum disorder. *Molecular autism*. 3:8.

39. Guhathakurta S, Sinha S, Ghosh S, Chatterjee A, Ahmed S, Gangopadhyay PK, et al. (2008): Population-based association study and contrasting linkage disequilibrium pattern reveal genetic association of SLC6A4 with autism in the Indian population from West Bengal. *Brain research*. 1240:12-21.

40. Xi CY, Ma HW, Lu Y, Zhao YJ, Hua TY, Zhao Y, et al. (2007): MeCP2 gene mutation analysis in autistic boys with developmental regression. *Psychiatric genetics*. 17:113-116.

41. Coutinho AM, Oliveira G, Katz C, Feng J, Yan J, Yang C, et al. (2007): MECP2 coding sequence and 3'UTR variation in 172 unrelated autistic patients. *American journal of medical genetics Part B, Neuropsychiatric genetics : the official publication of the International Society of Psychiatric Genetics*. 144B:475-483.

42. Harrison AJ, Gamsiz ED, Berkowitz IC, Nagpal S, Jerskey BA (2015): Genetic variation in the oxytocin receptor gene is associated with a social phenotype in autism spectrum disorders. *American journal of medical genetics Part B, Neuropsychiatric genetics : the official publication of the International Society of Psychiatric Genetics*. 168:720-729.

43. Venturin M, Moncini S, Villa V, Russo S, Bonati MT, Larizza L, et al. (2006): Mutations and novel polymorphisms in coding regions and UTRs of CDK5R1 and OMG genes in patients with non-syndromic mental retardation. *Neurogenetics*. 7:59-66.

44. Moncini S, Castronovo P, Murgia A, Russo S, Bedeschi MF, Lunghi M, et al. (2016): Functional characterization of CDK5 and CDK5R1 mutations identified in patients with non-syndromic intellectual disability. *Journal of human genetics*. 61:283-293.

45. Collins SC, Bray SM, Suhl JA, Cutler DJ, Coffee B, Zwick ME, et al. (2010): Identification of novel FMR1 variants by massively parallel sequencing in developmentally delayed males. *American journal of medical genetics Part A*. 152A:2512-2520.

46. Suhl JA, Muddashetty RS, Anderson BR, Ifrim MF, Visootsak J, Bassell GJ, et al. (2015): A 3' untranslated region variant in FMR1 eliminates neuronal activity-dependent translation of FMRP by disrupting binding of the RNA-binding protein HuR. *Proceedings of the National Academy of Sciences of the United States of America*. 112:E6553-6561.

47. Handt M, Epplen A, Hoffjan S, Mese K, Epplen JT, Dekomien G (2014): Point mutation frequency in the FMR1 gene as revealed by fragile X syndrome screening. *Molecular and cellular probes*. 28:279-283.

48. Ylisaukko-Oja T, Rehnstrom K, Vanhala R, Kempas E, von Koskull H, Tengstrom C, et al. (2005): MECP2 mutation analysis in patients with mental retardation. *American journal of medical genetics Part A*. 132A:121-124.

49. Santos M, Yan J, Temudo T, Oliveira G, Vieira JP, Fen J, et al. (2008): Analysis of highly conserved regions of the 3'UTR of MECP2 gene in patients with clinical diagnosis of Rett syndrome and other disorders associated with mental retardation. *Disease markers*. 24:319-324.

50. Fendri-Kriaa N, Mkaouar-Rebai E, Moalla D, Belguith N, Louhichi N, Zemni R, et al. (2010): Mutational analysis of the MECP2 gene in Tunisian patients with Rett syndrome: a novel double mutation. *Journal of child neurology*. 25:1042-1046.

51. Jensen KP, Kranzler HR, Stein MB, Gelernter J (2014): The effects of a MAP2K5 microRNA target site SNP on risk for anxiety and depressive disorders. *American journal of medical genetics Part B, Neuropsychiatric genetics : the official publication of the International Society of Psychiatric Genetics*. 165B:175-183.

52. Jensen KP, Covault J, Conner TS, Tennen H, Kranzler HR, Furneaux HM (2009): A common polymorphism in serotonin receptor 1B mRNA moderates regulation by miR-96 and associates with aggressive human behaviors. *Molecular psychiatry*. 14:381-389.

53. Rahman OA, Sasvari-Szekely M, Szekely A, Faludi G, Guttman A, Nemoda Z (2010): Analysis of a polymorphic microRNA target site in the purinergic receptor P2RX7 gene. *Electrophoresis*. 31:1790-1795.