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

Kai A. Wanke, Paolo Devanna, and Sonja C. Vernes

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% to 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 noncoding) to fill in the missing heritability of complex disorders. These new technologies bring new challenges, as the number of noncoding variants identified per individual can be overwhelming, making it prudent to focus on noncoding 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 noncoding genome that perfectly fulfills 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 or 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 neurodevelopmental disorders.

Keywords: Neurodevelopmental disorder, Neuropsychiatric disorders, Noncoding variation, miRNAs, RBP, RNA-binding proteins, 3’ Untranslated region, 3’UTR

https://doi.org/10.1016/j.biopsych.2017.11.006

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 or 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% to 90% for autism spectrum disorder (ASD) (4,5), 76% for attention-deficit/hyperactivity disorder (5), and 81% to 85% for schizophrenia (6); however, the currently identified variation can only account for a proportion of the heritability of these disorders [reviewed in Chong et al. (7) for Mendelian disorders and Manolio et al. (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 noncoding variation in the etiology of disorder (9–13) [reviewed in Sadee et al. (5) and Zhang and Lupski (14)]. Much of the noncoding 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 noncoding 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 noncoding 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 localization (19,20), and studies are now emerging that have identified putatively pathogenic variation in 3’UTR regions in NDD cohorts (11,21,22) (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 summarize current evidence linking variation in these regions to the etiology of NDDs. We focus the review on the best-described features of 3’UTRs: regulation by microRNAs, interaction with 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 in 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 messenger RNA (mRNA) expressed from all protein-coding genes, but they 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 (23). The length of human 3'UTR sequences can vary significantly, with a median length of ~1200 nucleotides (24) 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 noncoding 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).

---

Table 1. 3' UTR Variants Identified in NDDs That Have Shown Functional Effects

<table>
<thead>
<tr>
<th>Disorder</th>
<th>Gene</th>
<th>Variant</th>
<th>Global MAF (1000 Genomes Project)</th>
<th>ExAC (Frequency)</th>
<th>Effect In Vitro</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>FXS</td>
<td>FMR1</td>
<td>c.746T&gt;C</td>
<td>C = 0.0021</td>
<td>C = 0.0009688</td>
<td>Removes RBP (HuR) binding site and reduces transcript stability</td>
<td>(16,66)</td>
</tr>
<tr>
<td>TS</td>
<td>SLITRK1</td>
<td>Var321</td>
<td>T = 0.0006</td>
<td>N/A</td>
<td>Strengthens miR-189 binding and increases its repression in luciferase assays</td>
<td>(11)</td>
</tr>
<tr>
<td>SLI</td>
<td>ARHGEF39</td>
<td>rs72727021</td>
<td>C = 0.0469</td>
<td>C = 0.07911</td>
<td>Disrupts miR-215 binding site to increase ARHGEF39 expression in a luciferase assay and human brain tissue</td>
<td>(22)</td>
</tr>
<tr>
<td>RTT</td>
<td>MECP2</td>
<td>c.1832G&gt;C</td>
<td>--</td>
<td>N/A</td>
<td>Reduces the amount of MECP2 mRNA expressed in patient blood cells</td>
<td>(69)</td>
</tr>
<tr>
<td>ADHD</td>
<td>CLOCK</td>
<td>rs1801260</td>
<td>A = 0.2296</td>
<td>N/A</td>
<td>Att allele (T) decreases expression of CLOCK mRNA in mouse embryonic fibroblasts</td>
<td>(139–141)</td>
</tr>
<tr>
<td>DBH</td>
<td>rs129882</td>
<td>T = 0.2554</td>
<td>N/A</td>
<td></td>
<td>Att allele (C) decreases expression in luciferase assays</td>
<td>(142)</td>
</tr>
<tr>
<td>MTHFR</td>
<td>rs4846049</td>
<td>T = 0.3716</td>
<td>N/A</td>
<td></td>
<td>Att allele (T) increases repression by miR-149 in luciferase assays; also reduced MTHFR protein in patient blood cells</td>
<td>(143,144)</td>
</tr>
<tr>
<td>SZ</td>
<td>RGS4</td>
<td>rs10759</td>
<td>T = 0.2831</td>
<td>N/A</td>
<td>Att allele (T) in miR-124 site prevents repression in luciferase assays</td>
<td>(145)</td>
</tr>
<tr>
<td>EFN B2</td>
<td>rs550067317</td>
<td>G = 0.0004</td>
<td>N/A</td>
<td></td>
<td>Att (C) allele reduces regulation by miR-137 in luciferase assays</td>
<td>(146)</td>
</tr>
<tr>
<td>CPLX2</td>
<td>rs3822674</td>
<td>C = 0.4948</td>
<td>C = 0.4912</td>
<td></td>
<td>Att (C) allele increases regulation by miR-498 in luciferase assays</td>
<td>(147)</td>
</tr>
<tr>
<td>ASD</td>
<td>HLA-G</td>
<td>14-bp indel</td>
<td>--</td>
<td>N/A</td>
<td>Presence of the 14-bp indel increases stability of the mRNA</td>
<td>(148,149)</td>
</tr>
<tr>
<td>AFF2</td>
<td>ChrX:148076068[C&gt;T]</td>
<td>--</td>
<td>N/A</td>
<td></td>
<td>The variant reduces expression in luciferase assays in HEK cells and increases it in N2A cells</td>
<td>(150)</td>
</tr>
<tr>
<td>ID</td>
<td>CDK5RI</td>
<td>14-bp indel</td>
<td>--</td>
<td>N/A</td>
<td>Reduces the expression in luciferase assays in HEK cells</td>
<td>(12)</td>
</tr>
<tr>
<td></td>
<td>C2099T&gt;G</td>
<td>14-bp indel</td>
<td>--</td>
<td>N/A</td>
<td>Reduces the expression in luciferase assays in HEK cells</td>
<td>(12)</td>
</tr>
<tr>
<td></td>
<td>C2099T&gt;G</td>
<td>14-bp indel</td>
<td>--</td>
<td>N/A</td>
<td>Reduces the expression in luciferase assays in HEK cells</td>
<td>(12)</td>
</tr>
</tbody>
</table>

ADHD, attention-deficit/hyperactivity disorder; ASD, autism spectrum disorder; FXS, fragile X syndrome; HEK, human embryonic kidney; HuR, human antigen R protein; ID, intellectual disability; MAF, minor allele frequency; mRNA, messenger RNA; NDD, neurodevelopmental disorder; N2A, neuro 2a; RBP, RNA-binding protein; RTT, Rett syndrome; SK-N-BE, human neuroblastoma cell line; SLI, specific language impairment; SZ, schizophrenia; 3'UTR, 3' untranslated region; TS, Tourette's syndrome.

N/A indicates that variant is in a region not covered by reads in Exome Aggregation Consortium (ExAC) database.
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 mediate interactions with microRNA (miRNA) molecules, located after the stop codon of protein coding genes. 3'UTRs contribute to regulation via multiple mechanisms, three of which are depicted here: miRNA binding, RNA-binding protein (RBP) binding, and alternative polyadenylation (APA). MicroRNAs (miRNAs) interact with miRNA binding sites (MBSs) in the 3'UTR to either block protein translation or degrade the mRNA molecule. RBPs interact with RBP binding sites in the 3'UTR either regulate expression (via stabilizing or degrading the transcript), transport the mRNA to subcellular locations, or facilitate local translation of proteins (e.g., at the synapse). APA is the mechanism by which one of multiple possible polyadenylation sites may be used. APA determines the endpoint of the transcript and thus the length of the 3'UTR. In turn, the length of the 3'UTR determines which and how many regulatory elements (such as miRNA binding sites and RBP sites) are present to regulate the transcript.

Figure 1. Overview of the 3' untranslated region (3'UTR) and its regulatory mechanisms. 3'UTRs are untranslated (noncoding) regions of a messenger RNA (mRNA) molecule, located after the stop codon of protein coding genes. 3'UTRs contribute to regulation via multiple mechanisms, three of which are depicted here: miRNA binding, RNA-binding protein (RBP) binding, and alternative polyadenylation (APA). MicroRNAs (miRNAs) interact with miRNA binding sites (MBSs) in the 3'UTR to either block protein translation or degrade the mRNA molecule. RBPs interact with RBP binding sites in the 3'UTR either regulate expression (via stabilizing or degrading the transcript), transport the mRNA to subcellular locations, or facilitate local translation of proteins (e.g., at the synapse). APA is the mechanism by which one of multiple possible polyadenylation sites may be used. APA determines the endpoint of the transcript and thus the length of the 3'UTR. In turn, the length of the 3'UTR determines which and how many regulatory elements (such as miRNA binding sites and RBP sites) are present to regulate the transcript.

3'UTRs Are Bound by RNA-Binding Proteins to Regulate Translation or Subcellular Localization of mRNAs

3'UTRs mediate interactions with RBPs to influence mRNA stability, translation, and transport, ultimately regulating protein expression and subcellular localization (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'UTRs, 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 Localization

Polyadenylation 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 polyadenylation sites (Figure 1), which can be utilized in different cell types to produce 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 [for reviews, see (45–48)]. RBP interactions with 3'UTRs are crucial for
neuronal development and function (49,50). RBPs such as the human antigen R protein 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 characterized, 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 etiology of NDDs. Indeed, it has already been shown that disruption of these regulatory pathways occurs in schizophrenia 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 localization, 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 (attention-deficit/hyperactivity disorder, schizophrenia, ASD, and intellectual disability; see the Supplement 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 Supplemental Table S1); 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 noncoding 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

Fragile X syndrome (FXS) is a form of intellectual disability almost exclusively caused by noncoding 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 fragile X mental retardation protein (FMR1) (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 (five unique; seven recurrent) and three polymorphisms in the 3'UTR region (16). One 3'UTR variant (c.*746T>C; Table 1) was found in 6 patients but was not detected in the control population (n = 1260). Two patients with the variant showed varying degrees of FXS symptoms, suggesting that the variant may not be fully penetrant. Functional testing showed that this variant prevented the binding of human antigen R protein to the FMR1 mRNA. This destabilized the transcript and led to reduced expression of the FMR1 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

Rett syndrome (RTT) is a disorder on the autism spectrum that can also involve intellectual disability, speech loss, microcephaly, and seizures. The majority of affected individuals are explained by causative mutations in the MECP2 gene (67). However, approximately 35% of cases with classical RTT phenotype do not carry coding mutations in MECP2 and are of unknown etiology (68). MECP2 has a very long and highly conserved 3'UTR region, suggesting a unique 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 subset 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

Tourette’s syndrome (TS) is characterized by persistent, involuntary vocal and motor tics and may also include learning disabilities, features of ASD, attention-deficit/hyperactivity disorder, 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 noncoding variation. A number of association studies and candidate sequencing studies have provided evidence that the SLITRK1 gene contributes to the etiology 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 >2000 control individuals (11,72), suggesting Var321 as a putative TS risk factor (73). Var321 is located within an 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 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 single nucleotide polymorphism (SNP) (rs72727021; Table 1) significantly associated with SLI—nonword repetition (a task in which individuals repeat nonsense words that involves short-term phonological memory as well as vocal-motor planning and production). This SNP was within an MBS for miR-215 and had functional consequences, altering expression of the ARHGEF39 gene in human cells and in postmortem human cortex. This suggested the SNP as a novel SLI risk factor and provided the first example of a functional noncoding 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 (GWASs), 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 noncoding 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 (summarized in Figure 2).

**Genome-wide Approaches Capture Variation in 3’UTRs**

GWASs have been highly successful in the identification of common variants associated with NDDs. A subset of SNPs surveyed using standard GWAS chips are located within 3’UTRs [e.g., ~5% of SNPs on the Affimmetrix 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 noncoding 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 noncoding 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 to 5 million variants compared with 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), variant filtering, and prioritization 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 (Supplemental Table S1) where smaller cohort sizes (in the hundreds) were sufficient.

In the following section, we discuss the tools and strategies that can be used to prioritize variants identified via any of the above-mentioned 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 noncoding 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 MBSSs, RBP sites, and APA sites can be found in Supplemental 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 MBSSs (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) (Supplemental Table S2). Using tools that combine bioinformatics with empirical data can also improve predictions (37,102) (see Supplemental Table S2).

Functional Validation of 3’UTR Variants

The functional validation of noncoding 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 as biotin-tagged mRNA-miRNA pulldowns (105), variations of photoactivatable ribonucleoside–enhanced crosslinking and immunoprecipitation (PAR-CLIP) (106,107), crosslinking, ligating, and sequencing of hybrids (CLASH) (36), or RNA electrophoretic mobility shift assay (EMSA) (108). The localization of RNA can be visualized using imaging techniques (109), and stability can be assessed 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 reverse transcriptase quantitative polymerase chain reaction (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 postmortem 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 or false 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 has driven the design and implementation of therapeutics (116,117). However, we still understand only a fraction of the genetic factors contributing to NDDs, necessitating the need for further study and new approaches. To date, the contribution of noncoding 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 noncoding DNA makes up ~98% of the 3 billion nucleotides of the human genome, methods like WGS can be expected to reveal millions of noncoding variants per person, and interpreting the importance of variation in noncoding regions can be difficult. For these reasons, it is prudent to focus on regions of the noncoding 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 (Supplemental 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 etiology 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 noncoding changes can then be prioritized based on a number of factors including evolutionary conservation, frequency of occurrence in control populations (e.g., via 1000 Genomes Project, dbSNP, ClinVar), presence of a predicted functional motif at their position (Supplementary Table S2), or predicted pathogenicity (e.g., via Combined Annotation Dependent Depletion [CADD] scores) (119). As with protein-coding variants, simply identifying variation is not enough (120). Although 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 those 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 noncoding genome to NDDs. Other noncoding 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 noncoding 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 noncoding 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 noncoding 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 postgenomic era and closer to fully understanding the genetic factors contributing to complex phenotypes including, but not limited to, neurodevelopmental disorders.

ACKNOWLEDGMENTS AND DISCLOSURES
This work was supported by Marie Curie Career Integration Grant No. PCIG12-GA-2012-333978 (to SCV) and a Max Planck Research Group Grant (to SCV).

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.

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

ARTICLE INFORMATION
From the Neurogenetics of Vocal Communication Group (KAW, PD, SCV), Max Planck Institute for Psycholinguistics; Language and Genetics Department (KAW), Max Planck Institute for Psycholinguistics, Nijmegen; and the Donders Institute for Brain, Cognition and Behaviour (SCV), Nijmegen, the Netherlands.

Address correspondence to Sonja C. Vernes, Ph.D., Max Planck Institute for Psycholinguistics, Wundtlaan 1, 6525 XD Nijmegen, the Netherlands; E-mail: sonja.vernes@mpi.nl.

Received Feb 7, 2017; revised and accepted Nov 2, 2017.

Supplementary material cited in this article is available online at https://doi.org/10.1016/j.biopsycho.2017.11.006.

REFERENCES
3’UTR Variants and Neurodevelopmental Disorders


3'UTR Variants and Neurodevelopmental Disorders


