Published in final edited form as:
Future Neurol. 2013 January 1; 8(1): 29–42. doi:10.2217/fnl.12.83.

The expanding genomic landscape of autism: discovering the 'forest' beyond the 'trees'

Valerie W Hu
Department of Biochemistry & Molecular Medicine, The George Washington University, School of Medicine & Health Sciences, 2300 Eye St., N.W., Washington, DC 20037, USA Tel.: +1 202 994 8431 Fax: +1 202 994 4415 valhu@gwu.edu

Abstract

Autism spectrum disorders are neurodevelopmental disorders characterized by significant deficits in reciprocal social interactions, impaired communication and restricted, repetitive behaviors. As autism spectrum disorders are among the most heritable of neuropsychiatric disorders, much of autism research has focused on the search for genetic variants in protein-coding genes (i.e., the 'trees'). However, no single gene can account for more than 1% of the cases of autism spectrum disorders. Yet, genome-wide association studies have often identified statistically significant associations of genetic variations in regions of DNA that do not code for proteins (i.e., intergenic regions). There is increasing evidence that such noncoding regions are actively transcribed and may participate in the regulation of genes, including genes on different chromosomes. This article summarizes evidence that suggests that the research spotlight needs to be expanded to encompass far-reaching gene-regulatory mechanisms that include a variety of epigenetic modifications, as well as noncoding RNA (i.e., the 'forest'). Given that noncoding RNA represents over 90% of the transcripts in most cells, we may be observing just the 'tip of the iceberg' or the 'edge of the forest' in the genomic landscape of autism.

Keywords

autism spectrum disorder phenotypes; 'dark matter' RNA; epigenetics; future therapeutic options; gene–environment interaction; genetics; new research paradigm

Autism spectrum disorders (ASDs) refer collectively to a group of neurodevelopmental disorders diagnosed on the basis of abnormal development and behaviors, which include social and communication deficits, as well as the manifestation of restricted interests and stereotyped repetitive behaviors [1]. The phenotypes of ASDs range from a total lack of language and social isolation characteristic of the most severe form (classic autism) described by Kanner [2], through to a more moderate phenotype known as pervasive developmental disorder – not otherwise specified, to the mildest variant, known as Asperger's syndrome, which is typically not associated with clinically significant language and cognitive deficits [3]. As ASDs are among the most heritable of all psychiatric disorders, much effort has been placed on understanding the genetics of autism, specifically identifying genetic variants and other chromosomal abnormalities that are responsible for the etiology of autism [4, 5]. However, unlike the genes that are causal for syndromic

Financial & competing interests disclosure
The author has no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.
No writing assistance was utilized in the production of this manuscript.

disorders (such as fragile X syndrome) that sometimes manifest with autistic symptoms (recently reviewed by Benvenuto *et al.* 6]), no single gene can be considered 'causal' for more than 1% of the cases of idiopathic autism. This observation may relate to the phenotypic heterogeneity of ASDs. In this article, I emphasize the need to look beyond the search for structural variants in protein-coding genes (the 'trees') to investigate a variety of gene-regulatory mechanisms (the 'forest') that may disrupt the expression of large batteries of genes responsible for the pervasive characteristics of ASDs (Figure 1).

Genetics of ASDs

To date, many different approaches have been used to investigate the genetics of ASDs, from gene or chromosome-focused linkage analyses and family-based association tests to the more recent genome-wide linkage and association analyses, which have interrogated hundreds of thousands of single nucleotide polymorphisms (SNPs) in thousands of probands and families. These studies have together resulted in the identification of hundreds of candidate genes for ASDs (the latest estimates approaching 400 genes [7]), with relatively few genes replicated among the different studies and cohorts. A recent genome-wide association (GWA) study of common SNPs illustrates the vexing problem of identifying highly significant and replicable risk alleles. In this study, most of the SNPs with a nominal p-value $<5\times10^{-6}$ in the discovery phase, including one that exhibited genome-wide significance at $p = 2.1\times10^{-8}$, actually decreased in significance when independent cohorts (some exceeding 1000 probands each) were combined [8], most likely due to the increased phenotypic heterogeneity in the combined cohort. Increased LOD scores (i.e., better p-values) were achieved in a number of genome-wide linkage analyses in which the affected individuals were stratified by subphenotypes [9–14]. However, additional studies using independent cohorts are needed to replicate the improvements in LOD scores achieved by subphenotyping efforts, and the replication studies need to apply the same criteria for subtyping ASD cases in order to be comparable to each other. The findings from a wide range of genetic studies on ASDs have been extensively covered in a number of excellent reviews [15–18], and thus will not be discussed further here.

Recently, five papers reported the results of whole exome sequencing in which genomic DNA enriched for exonic sequences (covering the protein-coding regions of genes) were subjected to next-generation (next-gen) sequencing methods, which determine the base sequences of hundreds of thousands of oligonucleotide molecules captured on exome arrays containing probes that are designed to recognize and hybridize to exonic sequences in the genome [7, 19–23]. The primary goal of such studies was to identify, in an unbiased manner, rare and *de novo* genetic mutations as well as copy number variants associated with ASDs. With a few exceptions, there was relatively little overlap among the genes identified by these studies. However, 13–14% of the genes identified in two of the studies [19, 20] overlapped with genes that were differentially expressed in lymphoblastoid cell lines (LCL) derived from a subtype of ASD characterized by severe language impairment [24]. The latter study suggests that incorporating ASD subphenotypes into genetic analyses might improve the ability to corroborate genetic variants among different cohorts. Indeed, Hu *et al.* also demonstrated in a meta-analysis of published GWA data that dividing the cases into four ASD subphenotypes (subtypes) based on cluster analyses of functional/behavioral severity scores from the Autism Diagnostic Interview – Revised [25], considered a 'gold-standard' measure for autism, revealed 18 novel subtype-associated SNPs, with ten of the SNPs associated with two or more ASD subtypes [26]. However, the ten shared SNPs exhibited distinctly different minor allele frequencies and odds ratios among the subtypes, indicative of the genetic differences among these subtypes. Interestingly, the differences in minor allele frequencies of these ten shared SNPs between subtypes were sufficient to obscure any significant SNP association with ASD when the four subtypes were combined

into one large case group. More importantly, all 18 SNPs were found to be in noncoding regions (i.e., promoter, intronic and intergenic regions), suggesting their involvement in gene regulation rather than overt structural aberrations in the associated proteins. For example, a SNP at the intron boundary of the serotonin receptor 4 gene (*HTR4*) may be associated with its expression level, which was observed in a separate study [24] to be only reduced in LCL from the ASD subtype distinguished by severe language impairment; the odds ratio for the SNP in this subtype was 1.44, but not in LCL from two other ASD subtypes (moderate and mild) for which the odds ratios were approximately 0.7. Thus, this study suggests that using gene expression as a quantitative trait may be a powerful way of identifying functionally relevant genetic loci. To date, there are no published studies on expression quantitative trait loci analyses involving ASDs. In summary, the abovementioned subphenotype-dependent meta-analysis of GWA data, coupled with the results of recent GWA studies showing that the most significant SNPs are outside of exons [8, 27, 28], suggest that noncoding regions (which may associate with coding variants through linkage disequilibrium) and epigenetic mechanisms of gene regulation should also be considered in the etiology of autism.

Beyond genetics: epigenetics & environment

Epigenetics refers to the study of potentially heritable changes in gene expression and phenotype that do not involve changes in the sequence of nucleotides in DNA. The most studied epigenetic mechanisms include DNA methylation and histone modifications, which together determine the availability of a segment of DNA to transcriptional activators, as well as microRNA expression, which has been shown to regulate gene expression at post-transcriptional levels [29, 30]. The relatively recent demonstration that epigenetic modifications to the genome can result in transgenerational (inherited) changes in organismic phenotype, including psychiatric conditions [31 has intensified the focus on epigenetic mechanisms operating in complex neurological disorders [30]. However, epigenetic studies on ASDs are not trivial because epigenetic regulation of gene expression is typically tissue-specific, and the availability of postmortem brain tissues, particularly from well-characterized donors with ASDs, is very limited, thus impeding robust correction for multiple testing in genome-wide epigenetics analyses. On the other hand, peripheral tissues or cell lines derived from well-phenotyped living individuals may be useful as an experimental surrogate to identify at least some of the epigenetically dysregulated genes that may play a role in the pathobiology of ASDs. In this section, studies on each of the common epigenetic mechanisms will be briefly reviewed within the context of ASD, and the way in which environment may contribute to epigenetic changes and disease susceptibility will be discussed.

DNA methylation

DNA methylation involves the attachment of a methyl group to cytosine(s) in DNA. When the cytosines occur in a proximal promoter region enriched in CpG dinucleotides (i.e., CpG island), methylation typically (but not always) results in suppression of gene expression. More recently, methylation in other regions of the genome outside of CpG islands, including gene bodies and more distal promoter regions, have been described, with varying consequences on gene expression [32–35]. The relevance of DNA methylation to neurodevelopmental disorders was first recognized in the context of syndromic disorders, such as fragile X and Rett syndromes, in which a fraction of individuals also manifest autism. The etiology of fragile X syndrome involves the expansion of a CGG trinucleotide repeat in the 5′ untranslated region of the *FMR1* gene, which is subsequently silenced by methylation of the CGG expanded promoter region, resulting in reduced expression of the protein product, FMRP [36–38]. Unlike the etiology for much of idiopathic autism, FMRP deficiency due to mutations in a single gene, *FMR1*, is causal for the disorder.

Rett syndrome, on the other hand, is caused by a mutation in the gene, *MECP2*, which recognizes methylated cytosines in DNA [39–42]. Although MeCP2 has been thought to silence genes by binding to CpG island regions and preventing transcription, only 6% of MeCP2 binding sites genome wide were found to be in CpG islands, and surprisingly, 63% of genes whose promoters were bound to MeCP2 in a neuronal cell line were actively transcribed rather than repressed [43]. This study, demonstrating that the majority of MeCP2 binding sites are intronic or intergenic, thus provides sufficient reason to re-examine the function of MeCP2, as well as DNA methylation in relation to gene regulation.

With respect to nonsyndromic or idiopathic autism, expression prof iling studies that reveal differential gene expression between monozygotic twins discordant for autism diagnosis and behavioral phenotype are highly suggestive of the operation of epigenetic factors in ASDs [44]. To investigate the possibility that differential methylation might be involved in idiopathic ASDs, Nguyen *et al.* used CpG island arrays to conduct global methylation analyses of LCL (a surrogate model) derived from discordant monozygotic twins and sib pairs [45]. The differentially methylated genes were then compared with differentially expressed genes that were identified by separate gene expression analyses of LCL from the same individuals to identify genes for which increased methylation was associated with decreased expression and *vice versa* 44, 46]. Importantly, the expression of one of the confirmed differentially methylated genes, *RORA*, was also found to be reduced in LCL from a separate cohort of unrelated ASD-diagnosed individuals with a phenotype dominated by severe language impairment [24], as well as in post-mortem brain tissues (frontal cortex and cerebellum) of two separate cohorts of affected donors in comparison to unaffected controls [45, 47]. These four separate studies, showing reduction of RORA in LCL and brain tissues of individuals with ASDs, demonstrate that at least some of the molecular changes in cell lines derived from peripheral tissues are ref lected in the autistic brain. Collectively, these studies illustrate the value of integrating data across large-scale genomic analyses (e.g., identifying differentially expressed genes that may be dysregulated by aberrant methylation), as well as the use of peripheral tissues as surrogate experimental models for a neurological disorder in an integrative ‘systems’ approach toward understanding the underlying pathobiology of ASDs, as described in more detail elsewhere [48, 49]. The relevance of *RORA*, to the pathobiology of autism will be addressed later in the context of gene–environment (G × E) interactions.

With respect to the regulation of genes associated with autism susceptibility by DNA methylation, Gregory *et al.* showed that reduced expression of the *OXTR* gene was the result of CpG island methylation in both peripheral blood mononuclear cells and brain tissues from individuals with ASD [50]. Interestingly, they were led to investigate methylation differences at this gene upon discovery of a copy number variant in *OXTR* in an autistic proband that was not present in his affected sibling, who instead exhibited enhanced methylation of *OXTR*. Thus, both genetic and epigenetic factors were found to contribute to reduced *OXTR* expression in siblings from the same family, and the methylation status of *OXTR* in peripheral blood cells was shown to reflect similar methylation in brain tissues, again supporting the use of the blood cells as an experimental surrogate for brain cells in this case.

Histone modifications & chromatin remodeling

The packaging of DNA in the nucleus and its availability for transcription is dictated by the core histone proteins around which DNA is wrapped and associated chromatin remodeling proteins, many of which are involved in posttranslational modifications of histone proteins [29]. The histone modifications occur on the protruding histone ‘tails’ (N-termini), which are accessible on the surface of chromatin to modifying enzymes, which generally include histone acetylases, deacetylases and methyltransferases. These modifications provide signals

for the recruitment of additional proteins that either activate or repress transcription at a particular locus.

At present, there is relatively limited information on the involvement of histone modifications in ASDs. However, a recent study, which explored a specific histone modification, trimethylation of H3K4me3, in neuronal and non-neuronal cells from the post-mortem prefrontal cortex of individuals with autism and controls, demonstrated excess spreading of this histone modification from the transcriptional start site of a number of autism-relevant genes in neurons from approximately 25% of the cases, with no differences in H3K4me3 modification noted in controls [51]. Furthermore, this modification was positively associated with the activation of the respective gene, thus demonstrating dysregulation of gene expression in autism by a specific histone modification. The fact that only a fraction of the cases exhibited these changes suggests that this particular form of epigenetic regulation may apply to a specific subphenotype of ASD, or a subset of individuals within a subphenotype.

MicroRNA-dependent gene regulation

Another epigenetic mechanism that has been studied to a limited extent with respect to ASDs is microRNA (miRNA) expression. MiRNAs are endogenous, single-stranded, noncoding RNA molecules approximately 20–22 nucleotides long that negatively regulate gene expression post-transcriptionally by inhibition of translation or by target mRNA degradation [52]. Recently, miRNAs have also been shown to play a role in gene activation [53, 54], thus broadening the scope and potential of gene regulation by miRNAs. The involvement of miRNAs in neurogenesis, neural differentiation and development is well documented [55–58], and there is increasing evidence for miRNA dysregulation in neurological diseases and psychiatric disorders [59–61]. To date, four studies have implicated a role for miRNAs in ASDs. Abu-Elneel *et al.* first reported 28 miRNAs that were differentially expressed in cerebellar tissues from at least one of 13 autistic individuals relative to a control set of 13 age-, gender-, post-mortem index- and hemisphere-matched controls [62]. Talebizadeh *et al.* also profiled miRNA expression, but in LCL from six individuals with ASD and six age- and sex-matched individuals, identifying nine differentially expressed miRNAs between cases and controls [63]. Although previously identified autism candidate genes were among the predicted targets for the miRNAs identified in both studies, none of the putative targets were directly validated in either study. Interestingly, four out of the nine miRNAs from the latter study overlapped with those identified by the former study, suggesting that LCL derived from peripheral tissues may be useful experimental models that reflect at least some of the molecular changes in the autistic brain.

Two subsequent studies coupled miRNA expression profiling with gene expression profiling in LCL from the same samples in an effort to reveal novel targets of miRNAs in ASDs. In a study of monozygotic twins and sibling pairs discordant for diagnosis of autism, Sarachana *et al.* identified 43 differentially expressed miRNAs whose expression profiles completely separated cases from controls [64]. Interestingly, 16 of the miRNAs had been previously reported to be involved in neural differentiation and function [65–67]. Putative gene targets *ID3* and *PLK2* of two real-time quantitative PCRconfirmed brain-specific miRNAs (hsa-miR-29b and hsa-miR-219–5p), were further validated by miRNA overexpression and knockdown assays, respectively. While *ID3* is involved in inhibiting neuronal differentiation and maintaining the neural stem cell pool [68, 69], *PLK2* is critical for neuronal differentiation [70] and homeostatic synaptic plasticity [71]. *In silico* functional analysis of all of the predicted gene targets whose transcripts showed opposite changes from those of the respective differentially expressed miRNAs revealed gene networks involved in embryonic development, synaptic development and function, circadian rhythm signaling, inf

lammation, androgen metabolism and digestive functions, all of which are relevant to dysregulated processes in ASDs. Similarly integrating miRNA and mRNA expression in LCL derived from case–control siblings, Ghahramani Seno *et al.* reported 12 miRNAs that were differentially expressed in at least 12 out of 24 sib-pair comparisons [72]. Although there was some overlap of differentially expressed miRNAs among the four studies, there was no consistent pattern of miRNA expression. It is postulated that this inconsistency is in part due to the phenotypic heterogeneity of the cases employed in the different studies. In support of this hypothesis, significant improvements in p-values (almost seven orders of magnitude for miR486–5p) were achieved and seven additional miRNAs were revealed when similarity of mRNA expression profiles were used to define cohorts [72]. Thus, this study further demonstrates the added value of subtyping subjects (in this case by similar gene-expression profiles) for large-scale genomic analyses.

G × E interactions

There is increasing evidence that environment may contribute to risk for ASDs [73–75]. In this review, the term ‘environment’ broadly refers to both intrinsic (e.g., hormones) as well as extrinsic (e.g., pesticides) factors, which are not part of the genome *per se*, but can affect gene expression and phenotype. Environmental factors frequently impact disease susceptibility through epigenetic modifications, resulting in an altered ‘epigenome’. However, little is known regarding gene-specific epigenetic modifications or G × E interactions that may increase risk for ASDs.

The schematic in Figure 2 describes an example in which the expression of a functionally relevant candidate gene for ASDs, *RORA*, is modulated by both epigenetic modification and physiological environment (i.e., level of sex hormones). As mentioned earlier, *RORA* was found to be differentially methylated in LCL from individuals with ASD relative to controls [45]. Increased methylation was associated with decreased expression, since treatment of LCL with a methylation inhibitor increased *RORA* expression. Interestingly, *RORA* was found to be regulated by male and female hormones but in opposite directions, with dihydrotestosterone suppressing and estradiol enhancing *RORA* expression [47]. RORA was also found to regulate *CYP19A1*, the gene for the aromatase enzyme that is responsible for conversion of testosterone to estradiol. The correlated decrease in RORA and aromatase protein expression in the post-mortem frontal cortex further suggests an inhibitory feedback mechanism in which reduction of RORA and aromatase proteins may lead to increased testosterone, which further suppresses *RORA* expression [47]. These studies thus reveal a plausible biological explanation for the elevated testosterone levels that have been proposed as a sex-dependent risk factor for autism, which is strongly biased toward males [76]. Conversely, higher estrogen levels in females may lead to increased *RORA* expression, which may buffer against ROR A deficiency. Aside from providing a possible mechanistic explanation for elevated testosterone levels and sex bias, ROR A deficiency can also impact a number of other physiological functions known to be impaired in ASDs. Based on studies with a *Rora*-deficient mouse model, these functions include Purkinje cell differentiation and survival [77, 78], cerebellar development [79], protection against neuroinflammation and oxidative stress [80, 81], regulation of circadian rhythm [82] and transcriptional regulation of a number of genes that function at the glutamatergic synapse [83], thus suggesting that *RORA* is a functionally relevant autism susceptibility gene. Interestingly, the *Rora*-deficient staggerer mice also exhibit some behaviors that are characteristic of autism, including perseverative tendencies and impaired discrimination learning [84]. Thus, RORA deficiency, which has been associated with at least one subtype of ASD [24], may contribute to some of the neuropathology and behaviors observed in individuals with ASDs.

Although we do not yet understand the factors causing increased methylation of *RORA*, the methylation status of the genome can be influenced by many factors, including

environmental pollutants, drugs and diet, which are discussed in a comprehensive review by LaSalle [85]. The demonstrated hormone sensitivity of *RORA* makes it particularly vulnerable to dysregulation by environmentally dispersed endocrine-disrupting compounds, such as bisphenol A (a plasticizer) and atrazine (an herbicide), which mimic or antagonize natural hormones, thus interfering with normal hormonal signaling processes [86, 87]. In fact, La Salle and colleagues have recently demonstrated persistent epigenetic interactions between exposure to an environmental pollutant and a mutation in the *Mecp2* gene (homologous Rett syndrome gene) in mice, which resulted in global hypomethylation of DNA in the brain of adult female offspring from perinatally exposed dams [88]. Again, the mechanism for the altered methylation pattern is not clear. Recent studies by James and colleagues have associated genes and metabolites in the folate and transmethylation metabolic pathways with global DNA hypomethylation, particularly in the mothers of children with autism who harbor a specific risk allele for the reduced folate carrier gene, *RFC1* 89]. These findings emphasize the impact of the intrauterine environment as a possible risk factor for ASD, since only the mothers (and not the affected children) possess the risk allele, and suggest that dietary correction of folate or methyl donor deficiency might lower risk for ASD in children of women with deficits in folate metabolism [75].

Interplay between genetic & epigenetic mechanisms

Although epigenetics refers to heritable changes that do not involve a change in DNA sequence, there is substantial crosstalk between genetic and epigenetic mechanisms. For example, mutations in DNA may alter normal sites of DNA methylation, and altered methylation patterns or histone modifications may predispose segments of DNA to hypermutation. A recent study by Li *et al.* reported that ‘methylation deserts’, regions of lowest methylation comprising approximately 1% of the genome in germline (sperm) cells, are ‘hotspots’ for SNPs and structural mutability [90]. Furthermore, copy number variants (i.e., microdeletions and duplications in DNA), including *de novo* copy number variants that, as a group, have been associated with autism [91–93], are enriched within these methylation-poor regions [90]. Notably, pseudogenes (a class of noncoding DNA to be discussed later) are also enriched approximately twofold in methylation deserts. Thus, the structural mutability associated with aberrant DNA methylation in germline cells, which may result from poor nutrition, exposure to environmental toxins or other insults affecting methyl donor availability, may lead to heritable genetic mutations giving rise to a variety of complex disorders in progeny.

Small noncoding miRNAs are no less protected from genetic mutation than protein-coding genes. A recent study by Sun *et al.* that investigated naturally occurring SNPs in 16 miRNA genes on the X chromosome of males found 24 point mutations in either the mature miRNAs or their precursors [94]. Analyses of these mutated miRNAs revealed that the SNPs disrupted miRNA processing as well as target recognition. An additional study by this group, which focused on miRNA SNPs associated with schizophrenia, identified eight ultra-rare variants in the precursor or mature miRNA genes that resulted in altered expression levels including gain/loss of function [95]. Thus, the effect of genetic mutations on miRNA biogenesis and function adds another layer of complexity to the interface between genetics and epigenetics and its impact on disease. Aside from SNPs in miRNA genes, genetic variation in other noncoding regions of the genome, for example, enhancer or repressor binding sites, which may be kilobases away from gene promoters, may also associate with altered gene regulation and disease susceptibility.

The unexplored noncoding ‘dark matter’ in the depths of the genomic ‘forest’

In humans, protein-coding regions of the genome account for less than 2% of the DNA in a cell, with the majority of the remaining approximately 98% of DNA initially considered ‘junk DNA’. Recently, a number of studies have suggested that approximately 70–90% of the DNA in human cells is actively transcribed [96–100]. The term genomic ‘dark matter’ was first used by Johnson *et al.* to refer to such ncRNA transcripts that fall outside of protein-coding genes and which, for the most part, were of unknown function [101].

However, the idea that ncRNA is indeed an important and functional contributor to gene regulation and the developmental complexity of higher organisms, including humans, was first advanced by Mattick [102]. His proposal countered the then-prevailing dogma of genetic information flow in biological systems from DNA to RNA to proteins, with proteins performing not only structural and catalytic functions, but also the majority of regulatory functions required for the coordinated and properly timed gene expression in the development of all organisms. The recent emergence of ncRNA as multifunctional gene-regulatory molecules thus requires a reassessment of the role of ncRNA in development as well as in diseases, especially phenotypically complex disorders such as ASDs.

Emerging functions of ‘dark matter’ ncRNA

Strictly speaking, the term ncRNA could apply to any RNA transcript that does not code for protein, including transfer RNA, ribosomal RNA, miRNA (discussed above), small nucleolar RNA, which are involved in RNA splicing and other RNA modifications, piwi RNA, which are associated with transposon silencing in germline cells, and long ncRNA (lncRNA), which are the most varied and enigmatic of the ncRNAs [103]. As functions have been identified, at least broadly, for the first five ncRNA species mentioned and reviewed elsewhere, this section will focus on lncRNAs functions of which are least understood and thus comprise the major part of ‘dark matter’ RNA [103, 104].

lncRNAs (≥200 nucleotides) originate from intronic as well as intergenic regions and are also transcribed as antisense transcripts to protein-coding genes. ‘Pseudogenes’, which may have arisen by faulty or incomplete duplication of a parental gene or by retrotransposition of a transcribed parental gene from which they are distinguished by the lack of open reading frames and the presence of poly(A) tails, also contribute to the lncRNA pool when transcribed. Because of their sequence similarity to the parent gene, pseudogene regions are also a rich source of natural antisense transcripts (NATs) that may regulate the parental gene. An example of this phenomenon is the demonstration that a lncRNA encoding a NAT to the *Oct4* pseudogene 5 regulates the transcription of *Oct4* in *trans* (i.e., on a separate chromosome) as well as the expression of *Oct4* pseudogenes 4 and 5 in *cis* in mice [105]. Interestingly, when this NAT was suppressed with small interfering RNAs, expression of the *Oct4* gene increased with the loss of silencing epigenetic marks at the *Oct4* promoter. Importantly, this study reveals the intricate inter-relationships between gene regulation by lncRNA and epigenetic modifications on histones. In addition, such epigenetic modifications can have long-lasting effects on gene expression, lasting 28 days in one instance involving small dsRNA-induced transcriptional gene silencing of *UbC*, which resulted in early increases in histone methylation followed later by DNA methylation at the *UbC* promoter [106]. Thus, ncRNA may induce epigenetic changes with potential for transgenerational transmission.

Evidence for involvement of lncRNA in ASDs

The first association of a lncRNA with ASDs was the report of a breakpoint in the chromosomal 7q31.3 region, which contains a ncRNA transcript (GenBank ID# CB338058) [107]. Although there was no follow-up study on this region, it is interesting that this lncRNA is on the opposite strand just upstream of the gene *KCND2*, which has recently been identified as a candidate gene for ASDs [108]. A more recent study [109] examined a noncoding intergenic region 5p14.1 on chromosome 5 containing SNPs, which were reported to be highly significant in an earlier GWA ana lysis [28]. Interestingly, Kerin *et al.* found that this region encoded a 3.9 kb ncRNA, which was an antisense transcript to MSNP1AS and which can regulate in *trans* the expression of the *MSN* gene on chromosome X in human cell lines [109]. However, although the antisense ncRNA is increased in post-mortem brain tissues from individuals with ASD, there was no difference in MSN protein levels, which were instead found to be negatively correlated with the difference between *MSN* and MSNP1AS expression levels. This complex pattern of gene regulation related to the expression of antisense ncRNA coupled with the fact that there are many more highly significant SNPs in noncoding regions that have been identified in GWA studies of ASDs suggest that this finding may just be the 'tip of the iceberg' with respect to the involvement of lncRNA in ASDs [109].

In a genome-wide expression profiling study of LCL published in 2009, we reported that 20 lncRNAs were the only shared differentially expressed transcripts (relative to unaffected controls) among three subgroups of individuals with ASDs who were subphenotyped according to cluster analyses of 123 severity scores from the ADI-R assessment [24, 25]. Based on expression studies reported in the Gene Expression Omnibus repository, the majority of these transcripts were found to be expressed in a variety of cell types and differentially expressed either in androgen-dependent versus -independent prostate cancer cell lines or in dihydrotestosterone (DHT)-treated versus untreated fibroblasts from individuals with androgen insufficiency syndrome [110, 111]. To investigate the implied involvement of androgen in the regulation of these transcripts as well as to examine expression in a neuronal cell model, we treated SH-SY5Y human neuroblastoma cells with DHT and found that DHT altered the expression of all seven of the transcripts analyzed. Most of the 20 transcripts are intronic and in antisense orientation to the respective parent genes [24]. Although we do not yet know the function of these noncoding transcripts, the expression of these transcripts in different tissues and their regulated response to DHT suggest that these transcripts may have biological function and are not mere artifacts of faulty transcription. On the other hand, the fact that all of the transcripts are decreased in all three subtypes of ASD relative to unaffected controls may suggest a problem with splicing, but one that uniformly affects only these intronic transcripts. In this regard, it is noted that androgen-responsive intronic transcripts were reported to associate with alternative usage of exons or expression levels of the parent protein-coding genes [112], but we did not detect differential expression of the parent genes of the intronic transcripts in our study. Thus, if these novel lncRNAs regulate gene expression, they may be acting in *trans*. Moreover, the androgen sensitivity of these shared lncRNA transcripts is intriguing, given the 4:1 bias toward males in ASDs.

Roles of lncRNA in neurological development, functions & diseases

Although little is known about the biological functions of lncRNAs in neural development, their importance within the CNS is strongly indicated by the specific expression of 849 out of 1328 lncRNAs in specific cell types as well as neuroanatomical and subcellular regions in the mouse brain [113]. In humans, the involvement of lncRNAs in CNS development and function is implicated by dysregulated expression of an increasing number of noncoding transcripts in various neuropsychiatric and neurodegenerative diseases [114, 115]. Aside

from those that were described in the preceding section as being associated with ASDs, two lncRNAs (FMR4 and ASFMR1) that are expressed from the fragile X mental retardation gene locus were found to associate with both fragile X syndrome and fragile X tremor ataxia syndrome [116, 117]. Alzheimer's disease is associated with dysregulated expression of two other lncRNAs, BC200 and BACE1-AS, which have been shown to regulate protein translation at postsynaptic dendritic microdomains [118] and expression of the amyloid-cleaving b-secretase, BACE1 [119], respectively. A recent study using next-gen RNA-sequencing analyses to exhaustively identify all transcripts whose expression levels were changed upon differentiation of induced pluripotent stem cells into neurons revealed over 1200 transcripts (out of a total of 5953) encoding lncRNAs and pseudogenes, a number of which were mapped to SNPs previously identified by GWA studies as being associated with schizophrenia, bipolar disorder or ASD [120]. The results of this study, in combination with that of Kerin *et al.* 109] and our previous findings from sequential quantitative trait and case–control association analyses of published GWA data that showed that 18 novel SNPs in noncoding regions were strongly associated with four subphenotypes of ASDs [26], suggest that many more noncoding elements will be found to play a significant role in the pathobiology, if not etiology, of complex neurological and psychiatric disorders.

Future directions in ASD research & therapy

The preceding sections of this article illustrate the enormous genomic complexity of ASDs and emphasize the need to not only study, but also integrate, whenever possible, findings obtained at different levels of analyses, including genetics, epigenetics, gene expression and G × E interactions. While much of the research effort on ASDs over the past 15 or more years has been focused on identifying structural genetic variants in protein-coding genes, which have indeed provided valuable insight into the neuropathology of ASDs (e.g., deficits in synapse formation and axon guidance, among others), it is becoming increasingly clear that noncoding elements of the genome and epigenetic mechanisms of gene regulation also merit intensive research.

The need to incorporate subphenotypes into autism research

As we move forward into the next phase of autism research, which will undoubtedly take advantage of the enormous capabilities of next-gen sequencing that, in turn, will produce massive amounts of genomic data, we must keep in mind the need to reduce the heterogeneity of the ASD population under study by subgrouping individuals according to clinical phenotypes, specific traits or even comorbidities, which has been demonstrated to improve LOD scores in genome-wide linkage analyses [9–14]. We have also shown that reducing heterogeneity among cases by subgrouping individuals with ASDs according to severity of symptoms assessed by the ADI-R [25] resulted in the identification of subtype-specific differentially expressed genes and associated SNPs from genome-wide gene expression [24] and GWA analyses [26], respectively. Among the differentially expressed genes unique to one of the subtypes of ASD (characterized by severe language impairment), we observed 15 genes involved in circadian rhythm, two of which immediately suggest therapeutic options for this subgroup. For example, a deficiency in AANAT, which is the rate-limiting enzyme in the biosynthesis of melatonin, suggests that melatonin supplements may help to correct sleep disturbances exhibited by many on the autism spectrum, as well as alleviate some of the ASD symptoms that may be related to circadian disruption. On the other hand, a deficiency in DPYD, which predisposes individuals to epilepsy, mental retardation and ASD [121], may indicate anticonvulsant medications as a first choice of pharmacologic therapy for individuals with this subtype of ASD. Thus, it is expected that the identification of subtype-specific genes/pathways and SNPs will promote the development and use of novel pharmacologic interventions targeted toward the specific

biological deficits associated with a given subphenotype of ASD as well as allow a more specific, pharmacogenomically informative diagnosis of ASDs.

Is a paradigm shift toward a systems view of ASDs needed?

While ASDs are def ined primarily as neurodevelopmental disorders, many individuals with an ASD manifest problems in peripheral organ systems, such as the immune system and gastrointestinal disorders. Indeed, the plethora and varied functions of genes shown to be dysregulated in gene expression studies (such as circadian rhythm genes and those involved in cytoskeletal arrangements, muscle function and inflammation) as well as those implicated by genetic studies suggest that ASDs may be more accurately considered a systems disorder with severe neurological and behavioral manifestations. This paradigm shift away from a purely neurocentric perspective to a systems view of ASD would result in two important implications for autism research and therapy: first, peripheral tissues, such as blood-derived cells, may reflect at least some of the molecular and pathway alterations contributing to the ASD phenotype (as shown for the reduction of RORA in both blood-derived and brain tissues [24, 44, 45, 47]), and thus be useful surrogate experimental models for investigating the molecular underpinnings of ASDs; second, therapeutics aimed at resolving some of the metabolic or systemic problems exhibited by individuals with ASDs (e.g., gastrointestinal problems and sleep disturbances) may also prove to be beneficial in ameliorating some of the ASD symptoms. The key to appropriate systems-based therapeutics would come from identifying the specific metabolic/biological deficits affecting a particular individual or subtype of ASD.

The promise of ncRNA in therapeutics of ASDs

As discussed in a preceding section, ncRNAs are involved in many gene regulatory functions, and altered expression of a gene by ncRNA can have long-lasting effects due to epigenetic modifications, including DNA methylation, which may be transmitted transgenerationally. Thus, while it would be difficult to alter the sequence of a protein-coding gene that exhibits a genetic mutation or polymorphism or to introduce a new functional gene by gene therapy methods, it would be much more feasible to regulate expression of a gene that may be deficient or overexpressed in a disorder using ncRNAs that can be targeted to a specific gene by means of sequence complementarity. However, the problem of delivery of ncRNAs to the proper tissues is a major obstacle that would need to be resolved. Although research in the field of ncRNA, especially lncRNA, is still in its infancy, the rapidly evolving next-gen RNA sequencing technologies, which will allow complete identification of all transcripts in both healthy and diseased tissues, promise to yield true gems as we venture deeper into the ‘forest’ of dark matter RNA.

Conclusion

In exploring the genomic landscape of ASDs, it is clear that we need to broaden our perspective of the etiology of autism from the prevailing view that focuses on the proximal ‘trees’ (i.e., exonic, copy number and rare variants in protein-coding genes) to examine the more expansive genomic ‘forest’ encompassing extensive gene-regulatory elements, including epigenetic modifications and ncRNAs, which, in turn, are both influenced by environment. Beyond the question of etiology, the more pressing issue for individuals with an ASD, as well as for clinicians, is that of effective and personalized therapies that are targeted toward specific biological deficits [49]. Identification of such deficits may benefit from adopting a ‘systems’ perspective of ASDs for which novel therapeutic strategies may be revealed through studies of more easily accessible peripheral tissues from well-defined phenotypic subgroups of affected individuals [48].

Acknowledgments

The author is grateful to Drs Daniel Fairbanks (Utah Valley University, Orem, UT, USA), Kevin Morris (The Scripps Research Institute, La Jolla, CA, USA) and Peter Szatmari (McMaster University, Ontario, Canada) for their helpful comments and thoughtful review of the manuscript.

The work that is described from the author's laboratory was supported by grants from NIMH (R21 MH073393), Autism Speaks (#2381) and a generous gift from The LIFE Foundation (Aspen, CO).

References

Papers of special note have been highlighted as:

- ■ of interest
- ■ ■ of considerable interest

1. American Psychological Association. Diagnostic and Statistical Manual of Mental Disorders. Washington, DC, USA: American Psychological Association; 1994.
2. Kanner L. Early infantile autism. J. Pediatr. 1944; 25(3):211–217.
3. Asperger H. [Autistic psychopaths in childhood]. Arch. Psychiatr. Nervenkrankheiten. 1944; 117(1): 76–136.
4. Folstein S, Rutter M. Infantile autism: a genetic study of 21 twin pairs. J. Child Psychol. Psychiatry. 1977; 18(4):297–321. [PubMed: 562353]
5. Bailey A, Le Couteur A, Gottesman I et al. Autism as a strongly genetic disorder: evidence from a British twin study. Psychol. Med. 1995; 25(1):63–77. [PubMed: 7792363]
6. Benvenuto A, Manzi B, Alessandrelli R, Galasso C, Curatolo P. Recent advances in the pathogenesis of syndromic autisms. Int. J Pediatr. 2009; 2009:198736. [PubMed: 19946417]
7. Iossifov I, Ronemus M, Levy D, et al. *De novo* gene disruptions in children on the autistic spectrum. Neuron. 2012; 74(2):285–299. [PubMed: 22542183]
8. Anney R, Klei L, Pinto D, et al. A genomewide scan for common alleles affecting risk for autism. Hum. Mol. Genet. 2010; 19(20):4072–4082. [PubMed: 20663923]
9. Liu XQ, Paterson AD, Szatmari P. Genomewide linkage analyses of quantitative and categorical autism subphenotypes. Biol. Psychiatry. 2008; 64(7):561–570. [PubMed: 18632090]
10. Cannon D, Miller J, Robison R, et al. Genome-wide linkage analyses of two repetitive behavior phenotypes in Utah pedigrees with autism spectrum disorders. Mol. Autism. 2010; 1(1):3. [PubMed: 20678246]
11. Coon H, Villalobos M, Robison R, et al. Genome-wide linkage using the social responsiveness Scale in Utah autism pedigrees. Mol. Autism. 2010; 1(1):8. [PubMed: 20678250]
12. Alarcon M, Yonan AL, Gilliam TC, Cantor RM, Geschwind DH. Quantitative genome scan and ordered-subsets analysis of autism endophenotypes support language QTLs. Mol. Psychiatry. 2005; 10(8):747–757. [PubMed: 15824743]
13. Chen GK, Kono N, Geschwind DH, Cantor RM. Quantitative trait locus analysis of nonverbal communication in autism spectrum disorder. Mol. Psychiatry. 2006; 11(2):214–220. [PubMed: 16189504]
14. Nurmi EL, Dowd M, Tadevosyan-Leyfer O, Haines JL, Folstein SE, Sutcliffe JS. Exploratory subsetting of autism families based on savant skills improves evidence of genetic linkage to 15q11-q13. J. Am Acad. Child Adolesc. Psychiatry. 2003; 42(7):856–863. [PubMed: 12819446]
15. State MW, Levitt P. The conundrums of understanding genetic risks for autism spectrum disorders. Nat. Neurosci. 2011; 14(12):1499–1506. [PubMed: 22037497]
16. Geschwind DH. Autism: many genes, common pathways? Cell. 2008; 135(3):391–395. [PubMed: 18984147]
17. Freitag CM, Staal W, Klauck SM, Duketis E, Waltes R. Genetics of autistic disorders: review and clinical implications. Eur. Child Adolescent Psychiatry. 2010; 19(3):169–178.

18. Grigorenko EL. Pathogenesis of autism: a patchwork of genetic causes. Future Neurol. 2009; 4(5): 591–599. [PubMed: 19953194]
19. O'Roak BJ, Vives L, Girirajan S, et al. Sporadic autism exomes reveal a highly interconnected protein network of *de novo* mutations. Nature. 2012; 484(7397):246–250. [PubMed: 22437503]
20. Sanders SJ, Murtha MT, Gupta AR, et al. *De novo* mutations revealed by whole-exome sequencing are strongly associated with autism. Nature. 2012; 484(7397):237–241. [PubMed: 22426000]
21. Neale BM, Kou Y, Liu L, et al. Patterns and rates of exonic *de novo* mutations in autism spectrum disorders. Nature. 2012; 484(7397):242–245. [PubMed: 22437497]
22. Chahrour MH, Yu TW, Lim ET, et al. Whole-exome sequencing and homozygosity analysis implicate depolarization-regulated neuronal genes in autism. PLoS Genet. 2012; 8(4):e1002635. [PubMed: 22511880]
23. O'Roak BJ, Deriziotis P, Lee C, et al. Corrigendum: exome sequencing in sporadic autism spectrum disorders identifies severe *de novo* mutations. Nat. Genet. 2011; 43:585–589. [PubMed: 21572417]
24. Hu VW, Sarachana T, Kim KS, et al. Gene expression profiling differentiates autism casecontrols and phenotypic variants of autism spectrum disorders: evidence for circadian rhythm dysfunction in severe autism. Autism Res. 2009; 2(2):78–97. [PubMed: 19418574] ■ In addition to identifying subtype-dependent differentially expressed genes relative to controls, this study reveals 20 long noncoding RNAs as the only differentially expressed transcripts shared by three subtypes of autism spectrum disorders.
25. Hu VW, Steinberg ME. Novel clustering of items from the autism diagnostic interviewrevised to define phenotypes within autism spectrum disorders. Autism Res. 2009; 2(2):67–77. [PubMed: 19455643]
26. Hu VW, Addington A, Hyman A. Novel autism subtype-dependent genetic variants are revealed by quantitative trait and subphenotype association analyses of published GWAS data. PLoS ONE. 2011; 6(4):e19067. [PubMed: 21556359] ■ Shows that decreasing phenotypic heterogeneity of cases increases the ability to detect statistically significant associations of novel single nucleotide polymorphisms with autism spectrum disorders from pre-existing genotype data.
27. Weiss LA, Arking DE, Daly MJ, et al. A genome-wide linkage and association scan reveals novel loci for autism. Nature. 2009; 461(7265):802–808. [PubMed: 19812673]
28. Wang K, Zhang H, Ma D, et al. Common genetic variants on 5p14.1 associate with autism spectrum disorders. Nature. 2009; 459(7246):528–533. [PubMed: 19404256]
29. Allis, CD.; Jenuwein, T.; Reinberg, D. Epigenetics. NY, USA: Cold Spring Harbor Press; 2006. p. 502
30. Kaminsky Z, Wang SC, Petronis A. Complex disease, gender and epigenetics. Ann. Med. 2006; 38(8):530–544. [PubMed: 17438668]
31. Guerrero-Bosagna C, Skinner MK. Environmentally induced epigenetic transgenerational inheritance of phenotype and disease. Mol. Cell Endocrinol. 2012; 354(1–2):3–8. [PubMed: 22020198]
32. Maunakea AK, Nagarajan RP, Bilenky M, et al. Conserved role of intragenic DNA methylation in regulating alternative promoters. Nature. 2010; 466(7303):253–257. [PubMed: 20613842]
33. Zhang D, Cheng L, Badner JA, et al. Genetic control of individual differences in genespecific methylation in human brain. Am. J Hum. Genet. 2010; 86(3):411–419. [PubMed: 20215007]
34. Oberdoerffer S. A conserved role for intragenic DNA methylation in alternative pre-mRNA splicing. Transcription. 2012; 3(3):1–4.
35. Irizarry RA, Ladd-Acosta C, Wen B, et al. The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. Nat. Genet. 2009; 41(2):178–186. [PubMed: 19151715]
36. Feng Y, Zhang F, Lokey LK, et al. Translational suppression by trinucleotide repeat expansion at *FMR1* . Science. 1995; 268(5211):731–734. [PubMed: 7732383]
37. Bardoni B, Mandel JL, Fisch GS. *FMR1* gene and fragile X syndrome. Am. J Med. Genet. 2000; 97(2):153–163. [PubMed: 11180223]

38. Li Z, Zhang Y, Ku L, Wilkinson KD, Warren ST, Feng Y. The fragile X mental retardation protein inhibits translation via interacting with mRNA. Nucleic Acids Res. 2001; 29(11):2276–2283. [PubMed: 11376146]
39. Amir RE, Zoghbi HY. Rett syndrome: methyl-CpG-binding protein 2 mutations and phenotype-genotype correlations. Am. J Med. Genet. 2000; 97(2):147–152. [PubMed: 11180222]
40. Kim SJ, Cook EH Jr. Novel *de novo* nonsense mutation of *MECP2* in a patient with Rett syndrome. Hum. Mutat. 2000; 15(4):382–383. [PubMed: 10737989]
41. Amir RE, Van den Veyver IB, Wan M, Tran CQ, Francke U, Zoghbi HY. Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2. Nat. Genet. 1999; 23(2):185–188. [PubMed: 10508514]
42. Van den Veyver IB, Zoghbi HY. Genetic basis of Rett syndrome. Ment. Retard Dev. Disabil. Res. Rev. 2002; 8(2):82–86. [PubMed: 12112732]
43. Yasui DH, Peddada S, Bieda MC, et al. Integrated epigenomic analyses of neuronal MeCP2 reveal a role for long-range interaction with active genes. Proc. Natl Acad. Sci. USA. 2007; 104(49): 19416–19421. [PubMed: 18042715] ■ Suggests re-examination of the role of MeCP2 and DNA methylation in gene regulation.
44. Hu VW, Frank BC, Heine S, Lee NH, Quackenbush J. Gene expression profiling of lymphoblastoid cell lines from monozygotic twins discordant in severity of autism reveals differential regulation of neurologically relevant genes. BMC Genomics. 2006; 7:118. [PubMed: 16709250]
45. Nguyen A, Rauch TA, Pfeifer GP, Hu VW. Global methylation profiling of lymphoblastoid cell lines reveals epigenetic contributions to autism spectrum disorders and a novel autism candidate gene *RORA*, whose protein product is reduced in autistic brain. FASEB J. 2010; 24(8):3036–3051. [PubMed: 20375269]
46. Hu VW, Nguyen A, Kim KS, et al. Gene expression profiling of lymphoblasts from autistic and nonaffected sib pairs: altered pathways in neuronal development and steroid biosynthesis. PLoS ONE. 2009; 4(6):e5775. [PubMed: 19492049]
47. Sarachana T, Xu M, Wu RC, Hu VW. Sex hormones in autism: Androgens and estrogens differentially and reciprocally regulate RORA, a novel candidate gene for autism. PLoS ONE. 2011; 6(2):e17116. [PubMed: 21359227] ■ Shows hormonal regulation of a functionally relevant autism candidate gene and offers a plausible explanation for both higher testosterone levels and male bias in autism.
48. Hu VW. From genes to environment: using integrative genomics to build a ‘systems level’ understanding of autism spectrum disorders. Child Dev. 2012 (Epub ahead of print).
49. Hu VW. A systems approach towards an understanding, diagnosis and personalized treatment of autism spectrum disorders. Pharmacogenomics. 2011; 12(9):1235–1238. [PubMed: 21919600]
50. Gregory SG, Connelly JJ, Towers AJ, et al. Genomic and epigenetic evidence for oxytocin receptor deficiency in autism. BMC Med. 2009; 7:62. [PubMed: 19845972]
51. Shulha HP, Cheung I, Whittle C, et al. Epigenetic signatures of autism: trimethylated H3K4 landscapes in prefrontal neurons. Arch. Gen. Psychiatry. 2012; 69(3):314–324. [PubMed: 22065254] ■ First study to demonstrate histone modifications contributing to gene activation in autism spectrum disorders.
52. Nilsen TW. Mechanisms of microRNAmediated gene regulation in animal cells. Trends Genet. 2007; 23(5):243–249. [PubMed: 17368621]
53. Li LC, Okino ST, Zhao H, et al. Small dsRNAs induce transcriptional activation in human cells. Proc. Natl Acad. Sci. USA. 2006; 103(46):17337–17342. [PubMed: 17085592]
54. Janowski BA, Younger ST, Hardy DB, Ram R, Huffman KE, Corey DR. Activating gene expression in mammalian cells with promoter-targeted duplex RNAs. Nat. Chem. Biol. 2007; 3(3): 166–173. [PubMed: 17259978]
55. Kawahara H, Imai T, Okano H. MicroRNAs in neural stem cells and neurogenesis. Front. Neurosci. 2012; 6:30. [PubMed: 22416227]
56. Meza-Sosa KF, Valle-García D, Pedraza-Alva G, Pérez-Martínez L. Role of microRNAs in central nervous system development and pathology. J. Neurosci. Res. 2012; 90(1):1–12. [PubMed: 21922512]

57. Bredy TW, Lin Q, Wei W, Baker-Andresen D, Mattick JS. MicroRNA regulation of neural plasticity and memory. Neurobiol. Learn. Mem. 2011; 96(1):89–94. [PubMed: 21524708]
58. Fiore R, Siegel G, Schratt G. MicroRNA function in neuronal development, plasticity and disease. Biochim. Biophys. Acta. 2008; 1779(8):471–478. [PubMed: 18194678]
59. Maes OC, Chertkow HM, Wang E, Schipper HM. MicroRNA: implications for Alzheimer disease and other human CNS disorders. Curr. Genomics. 2009; 10(3):154–168. [PubMed: 19881909]
60. Martino S, Di Girolamo I, Orlacchio A, Datti A, Orlacchio A. MicroRNA implications across neurodevelopment and neuropathology. J. Biomed. Biotechnol. 2009; 2009:654346. [PubMed: 19841678]
61. Bicker S, Schratt G. microRNAs: tiny regulators of synapse function in development and disease. J. Cell. Mol. Med. 2008; 12(5A):1466–1476. [PubMed: 18624757]
62. Abu-Elneel K, Liu T, Gazzaniga FS, et al. Heterogeneous dysregulation of microRNAs across the autism spectrum. Neurogenetics. 2008; 9(3):153–161. [PubMed: 18563458]
63. Talebizadeh Z, Butler MG, Theodoro MF. Feasibility and relevance of examining lymphoblastoid cell lines to study role of microRNAs in autism. Autism Res. 2008; 1(5):307.
64. Sarachana T, Zhou R, Chen G, Manji HK, Hu VW. Investigation of post-transcriptional gene regulatory networks associated with autism spectrum disorders by microRNA expression profiling of lymphoblastoid cell lines. Genome Med. 2010; 2(4):23. [PubMed: 20374639]
65. Sempere LF, Freemantle S, Pitha-Rowe I, Moss E, Dmitrovsky E, Ambros V. Expression profiling of mammalian microRNAs uncovers a subset of brain-expressed microRNAs with possible roles in murine and human neuronal differentiation. Genome Biol. 2004; 5(3):R13. [PubMed: 15003116]
66. Krichevsky AM, King KS, Donahue CP, Khrapko K, Kosik KS. A microRNA array reveals extensive regulation of microRNAs during brain development. RNA. 2003; 9(10):1274–1281. [PubMed: 13130141]
67. Liang Y, Ridzon D, Wong L, Chen C. Characterization of microRNA expression profiles in normal human tissues. BMC Genomics. 2007; 8:166. [PubMed: 17565689]
68. Jung S, Park RH, Kim S, et al. Id proteins facilitate self-renewal and proliferation of neural stem cells. Stem Cells Dev. 2010; 19(6):831–841. [PubMed: 19757990]
69. Jogi A, Persson P, Grvnfeld A, Pahlman S, Axelson H. Modulation of basic helix-loophelix transcription complex formation by Id proteins during neuronal differentiation. J. Biol. Chem. 2002; 277(11):9118–9126. [PubMed: 11756408]
70. Draghetti C, Salvat C, Zanoguera F, et al. Functional whole-genome analysis identifies polo-like kinase 2 and poliovirus receptor as essential for neuronal differentiation upstream of the negative regulator aB-crystallin. J. Biol. Chem. 2009; 284(46):32053–32065. [PubMed: 19700763]
71. Lee K, Lee Y, Rozeboom A, et al. Requirement for Plk2 in orchestrated Ras and Rap signaling, homeostatic structural plasticity, and memory. Neuron. 2011; 69(5):957–973. [PubMed: 21382555]
72. Ghahramani Seno MM, Hu P, Gwadry FG, et al. Gene and miRNA expression profiles in autism spectrum disorders. Brain Res. 2011; 1380:85–97. [PubMed: 20868653]
73. Hallmayer J, Cleveland S, Torres A, et al. Genetic heritability and shared environmental factors among twin pairs with autism. Arch. Gen. Psychiatry. 2011; 68(11):1095–1102. [PubMed: 21727249] ■ Suggests re-examination of the relative contributions of genetics and the environment to autism spectrum disorders.
74. Croen LA, Grether JK, Yoshida CK, Odouli R, Hendrick V. Antidepressant use during pregnancy and childhood autism spectrum disorders. Arch. Gen. Psychiatry. 2011; 68(11):1104–1112. [PubMed: 21727247]
75. Schmidt RJ, Tancredi DJ, Ozonoff S, et al. Maternal periconceptional folic acid intake and risk of autism spectrum disorders and developmental delay in the CHARGE (CHildhood Autism Risks from Genetics and Environment) case-control study. Am. J. Clin. Nutr. 2012; 96(1):80–89. [PubMed: 22648721]
76. Baron-Cohen S, Knickmeyer RC, Belmonte MK. Sex differences in the brain: implications for explaining autism. Science. 2005; 310(5749):819–823. [PubMed: 16272115]

77. Doulazmi M, Frédéric F, Lemaigre-Dubreuil Y, Hadj-Sahraoui N, Delhaye-Bouchaud N, Mariani J. Cerebellar Purkinje cell loss during life span of the heterozygous staggerer mouse (Rora+/Rora(sg)) is gender-related. J. Comp. Neurol. 1999; 411(2):267–273. [PubMed: 10404252]

78. Hadj-Sahraoui N, Frederic F, Zanjani H, Delhaye-Bouchaud N, Herrup K, Mariani J. Progressive atrophy of cerebellar Purkinje cell dendrites during aging of the heterozygous staggerer mouse (Rora+/sg). Dev. Brain Res. 2001; 126(2):201–209. [PubMed: 11248354]

79. Gold DA, Gent PM, Hamilton BA. RORa in genetic control of cerebellum development: 50 staggering years. Brain Res. 2007; 1140(1):19–25. [PubMed: 16427031]

80. Delerive P, Monté D, Dubois G, et al. The orphan nuclear receptor RORa is a negative regulator of the inflammatory response. EMBO Rep. 2001; 2(1):42–48. [PubMed: 11252722]

81. Boukhtouche F, Vodjdani G, Jarvis CI, et al. Human retinoic acid receptor-related orphan receptor a1 overexpression protects neurones against oxidative stress-induced apoptosis. J. Neurochem. 2006; 96(6):1778–1789. [PubMed: 16539693]

82. Sato TK, Panda S, Miraglia LJ, et al. A functional genomics strategy reveals rora as a component of the mammalian circadian clock. Neuron. 2004; 43(4):527–537. [PubMed: 15312651]

83. Gold DA, Baek SH, Schork NJ, et al. RORa coordinates reciprocal signaling in cerebellar development through sonic hedgehog and calcium-dependent pathways. Neuron. 2003; 40(6):1119–1131. [PubMed: 14687547]

84. Lalonde R, Strazielle C. Discrimination learning in *Rorasg* and *Grid2ho* mutant mice. Neurobiol. Learn. Mem. 2008; 90(2):472–474. [PubMed: 18583162]

85. LaSalle JM. A genomic point-of-view on environmental factors influencing the human brain methylome. Epigenetics. 2011; 6(7):862–869. [PubMed: 21617367]

86. Colborn T. Neurodevelopment and endocrine disruption. Environ. Health Perspect. 2004; 112(9):944–949. [PubMed: 15198913]

87. McLachlan JA. Functional toxicology: a new approach to detect biologically active xenobiotics. Environ. Health Perspect. 1993; 101(5):386–387. [PubMed: 8119246]

88. Woods R, Vallero RO, Golub MS, et al. Long-lived epigenetic interactions between perinatal PBDE exposure and *Mecp2* 308 mutation. Hum. Mol. Genet. 2012; 21(11):2399–2411. [PubMed: 22343140]

89. James SJ, Melnyk S, Jernigan S, et al. A functional polymorphism in the reduced folate carrier gene and DNA hypomethylation in mothers of children with autism. Am. J. Med. Genet. B Neuropsychiatr. Genet. 2010; 153B(6):1209–1220. [PubMed: 20468076]

90. Li J, Harris RA, Cheung SW, et al. Genomic hypomethylation in the human germline associates with selective structural mutability in the human genome. PLoS Genet. 2012; 8:e1002692. [PubMed: 22615578] ■ Demonstrates interplay between epigenetics and genetics on structural variation in DNA.

91. Sebat J, Lakshmi B, Malhotra D, et al. Strong association of *de novo* copy number mutations with autism. Science. 2007; 316(5823):445–449. [PubMed: 17363630]

92. Marshall CR, Noor A, Vincent JB, et al. Structural variation of chromosomes in autism spectrum disorder. Am. J. Hum. Genet. 2008; 82(2):477–488. [PubMed: 18252227]

93. Pinto D, Pagnamenta AT, Klei L, et al. Functional impact of global rare copy number variation in autism spectrum disorders. Nature. 2010; 466(7304):368–372. [PubMed: 20531469]

94. Sun G, Yan J, Noltner K, et al. SNPs in human miRNA genes affect biogenesis and function. RNA. 2009; 15(9):1640–1651. [PubMed: 19617315]

95. Feng J, Sun G, Yan J, et al. Evidence for X-chromosomal schizophrenia associated with microRNA alterations. PLoS ONE. 2009; 4(7):e6121. [PubMed: 19568434]

96. Birney E, Stamatoyannopoulos JA, Dutta A, et al. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature. 2007; 447(7146):799–816. [PubMed: 17571346]

97. Kapranov P, St Laurent G, Raz T, et al. The majority of total nuclear-encoded nonribosomal RNA in a human cell is ‘dark matter’ un-annotated RNA. BMC Biol. 2010; 8:149. [PubMed: 21176148]

98. Carninci P, Kasukawa T, Katayama S, et al. The transcriptional landscape of the mammalian genome. Science. 2005; 309(5740):1559–1563. [PubMed: 16141072]

99. Kapranov P, Drenkow J, Cheng J, et al. Examples of the complex architecture of the human transcriptome revealed by RACE and high-density tiling arrays. Genome Res. 2005; 15(7):987–997. [PubMed: 15998911]

100. Cawley S, Bekiranov S, Ng HH, et al. Unbiased mapping of transcription factor binding sites along human chromosomes 21 and 22 points to widespread regulation of noncoding RNAs. Cell. 2004; 116(4):499–509. [PubMed: 14980218]

101. Johnson JM, Edwards S, Shoemaker D, Schadt EE. Dark matter in the genome: evidence of widespread transcription detected by microarray tiling experiments. Trends Genet. 2005; 21(2): 93–102. [PubMed: 15661355] ■ Early demonstration of widespread transcription of noncoding DNA referred to as ‘dark matter’.

102. Mattick JS. Challenging the dogma: the hidden layer of non-protein-coding RNAs in complex organisms. Bioessays. 2003; 25(10):930–939. [PubMed: 14505360] ■ ■ Original proposal stressing the functional relevance of noncoding RNAs as regulatory elements in developmental processes of higher organisms.

103. Esteller M. Non-coding RNAs in human disease. Nat. Rev. Genet. 2011; 12(12):861–874. [PubMed: 22094949]

104. Eddy SR. Non-coding RNA genes and the modern RNA world. Nat. Rev. Genet. 2001; 2(12): 919–929. [PubMed: 11733745]

105. Hawkins PG, Morris KV. Transcriptional regulation of Oct4 by a long non-coding RNA antisense to Oct4-pseudogene 5. Transcription. 2010; 1(3):165–175. [PubMed: 21151833] ■ Shows that suppression of long noncoding RNA can lead to epigenetic changes resulting in altered expression of a gene on a different chromosome.

106. Hawkins PG, Santoso S, Adams C, Anest V, Morris KV. Promoter targeted small RNAs induce long-term transcriptional gene silencing in human cells. Nucleic Acids Res. 2009; 37(9):2984–2995. [PubMed: 19304753]

107. Scherer SW, Cheung J, MacDonald JR, et al. Human chromosome 7: DNA sequence and biology. Science. 2003; 300(5620):767–772. [PubMed: 12690205]

108. Casey JP, Magalhaes T, Conroy JM, et al. A novel approach of homozygous haplotype sharing identifies candidate genes in autism spectrum disorder. Hum. Genet. 2012; 131(4):565–579. [PubMed: 21996756]

109. Kerin T, Ramanathan A, Rivas K, Grepo N, Coetzee GW, Campbell DB. A noncoding RNA antisense to moesin at 5p14.1 in autism. Sci. Transl. Med. 2012; 4(128):128ra40. ■ Elegant study demonstrating that a transcript derived from an intergenic region containing single nucleotide polymorphisms with significant genomewide associations with autism spectrum disorders regulates a gene on another chromosome.

110. Holterhus PM, Deppe U, Werner R, et al. Intrinsic androgen-dependent gene expression patterns revealed by comparison of genital fibroblasts from normal males and individuals with complete and partial androgen insensitivity syndrome. BMC Genomics. 2007; 8:376. [PubMed: 17945006]

111. Zhao H, Kim Y, Wang P, et al. Genome-wide characterization of gene expression variations and DNA copy number changes in prostate cancer cell lines. Prostate. 2005; 63(2):187–197. [PubMed: 15486987]

112. Louro R, Nakaya HI, Amaral PP, et al. Androgen responsive intronic non-coding RNAs. BMC Biol. 2007; 5:4. [PubMed: 17263875]

113. Mercer TR, Dinger ME, Sunkin SM, Mehler MF, Mattick JS. Specific expression of long noncoding RNAs in the mouse brain. Proc. Natl Acad. Sci. USA. 2008; 105(2):716–721. [PubMed: 18184812]

114. Qureshi IA, Mattick JS, Mehler MF. Long non-coding RNAs in nervous system function and disease. Brain Res. 2010; 1338:20–35. [PubMed: 20380817]

115. Bian S, Sun T. Functions of noncoding RNAs in neural development and neurological diseases. Mol. Neurobiol. 2011; 44(3):359–373. [PubMed: 21969146]

116. Ladd PD, Smith LE, Rabaia NA, et al. An antisense transcript spanning the CGG repeat region of FMR1 is upregulated in premutation carriers but silenced in full mutation individuals. Hum. Mol. Genet. 2007; 16(24):3174–3187. [PubMed: 17921506]

117. Khalil AM, Faghihi MA, Modarresi F, Brothers SP, Wahlestedt C. A novel RNA transcript with antiapoptotic function is silenced in fragile X syndrome. PLoS ONE. 2008; 3(1):e1486. [PubMed: 18213394]

118. Mus E, Hof PR, Tiedge H. Dendritic BC200 RNA in aging and in Alzheimer's disease. Proc. Natl Acad. Sci. USA. 2007; 104(25):10679–10684. [PubMed: 17553964]

119. Faghihi MA, Modarresi F, Khalil AM, et al. Expression of a noncoding RNA is elevated in Alzheimer's disease and drives rapid feedforward regulation of b-secretase. Nat. Med. 2008; 14(7):723–730. [PubMed: 18587408]

120. Lin M, Pedrosa E, Shah A, et al. RNA-Seq of human neurons derived from iPS cells reveals candidate long non-coding RNAs involved in neurogenesis and neuropsychiatric disorders. PLoS ONE. 2011; 6(9):e23356. [PubMed: 21915259]

121. Van Kuilenburg ABP, Vreken P, Abeling NGGM, et al. Genotype and phenotype in patients with dihydropyrimidine dehydrogenase deficiency. Hum. Genet. 1999; 104(1):1–9. [PubMed: 10071185]

Executive summary

Introduction: why study genetics of autism?

- Autism spectrum disorders (ASDs) are among the most heritable of psychiatric disorders.
- However, no single gene can account for more than 1% of the cases of idiopathic autism, and the number of candidate genes associated with ASDs is rapidly rising to >400.
- One of the confounding factors in identifying causal variants in DNA is the clinical and phenotypic heterogeneity of ASDs, which undoubtedly reflects multiple etiologies.

Beyond genetics: the need to consider epigenetics & environment

- Discordance of autism diagnosis and phenotype among some sets of monozygotic twins suggests the contribution of epigenetic and environmental factors to autism risk.
- Epigenetic changes, including differences in DNA methylation, histone modification and microRNA expression, have been associated with ASDs.
- Environmental factors can affect risk for ASDs through direct, indirect and epigenetic mechanisms.
- Epigenetic modifications may be transmitted transgenerationally, thus contributing a nongenetic but heritable component to autism risk.
- There is crosstalk between genetic and epigenetic mechanisms.

The need to explore ‘dark matter’ (noncoding) RNA in the etiology of ASDs

- Much of DNA is transcribed, thus resulting in a large number of ncRNA transcripts, many of unknown functions.
- Long ncRNAs (exceeding 200 base pairs) have been identified in both genetic and gene expression studies of ASDs.
- Although gene regulatory functions are the suspected mechanisms for the effect of long ncRNAs, more research is needed to understand their roles in the etiology and pathogenesis of ASDs.

Future directions in ASD research & therapy

- While much of the recent focus of autism research has been on the search for structural variants in the protein-coding regions of genes (i.e., the ‘trees’), the spotlight needs to be expanded to encompass gene-regulatory mechanisms (the ‘forest’), which include transcribed regions outside of the reading frame in genes (e.g., introns), noncoding regulatory regions outside of the transcribed regions of genes (e.g., enhancers), epigenetic modifications and ncRNAs, which are all capable of inducing pervasive dysregulation of gene expression.
- Although next-generation sequencing of whole genomes will undoubtedly provide massive amounts of data on both coding and noncoding transcripts in healthy and diseased tissues, it will be necessary to study defined clinical/behavioral subphenotypes of ASDs in order to elicit biologically meaningful information from such large-scale genomic studies.

- A paradigm shift from a purely neurocentric view of ASDs to a ‘systems’ perspective will allow a more comprehensive unbiased identification of biological deficits in specific subphenotypes of ASDs, which may be revealed through studies using peripheral tissues.
- The identification of subtype-specific biological deficits is expected to reveal novel therapeutic targets as well as facilitate the development of more personalized pharmacogenomics approaches to treatment of ASDs.

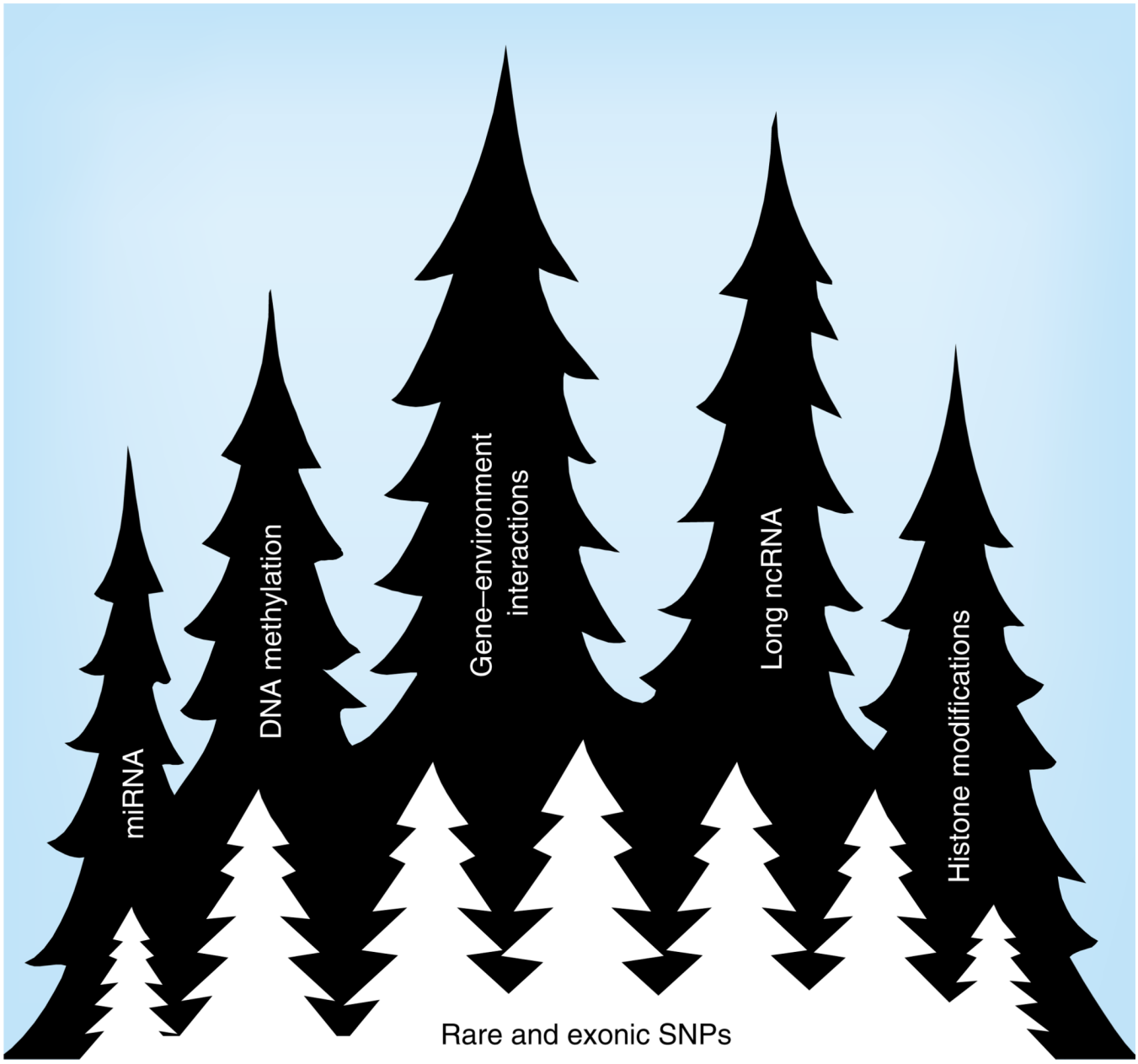

Figure 1. The expansive genomic landscape of autism

Much of the effort in autism research over the past 15 or more years has focused on identifying rare and exonic single nucleotide polymorphisms in protein-coding genes, with the number of genes rapidly increasing. This article takes the perspective that there is a vast genomic landscape yet to be explored in order to gain a better understanding of the gene regulatory mechanisms that may contribute to the etiology and pathogenesis of autism.

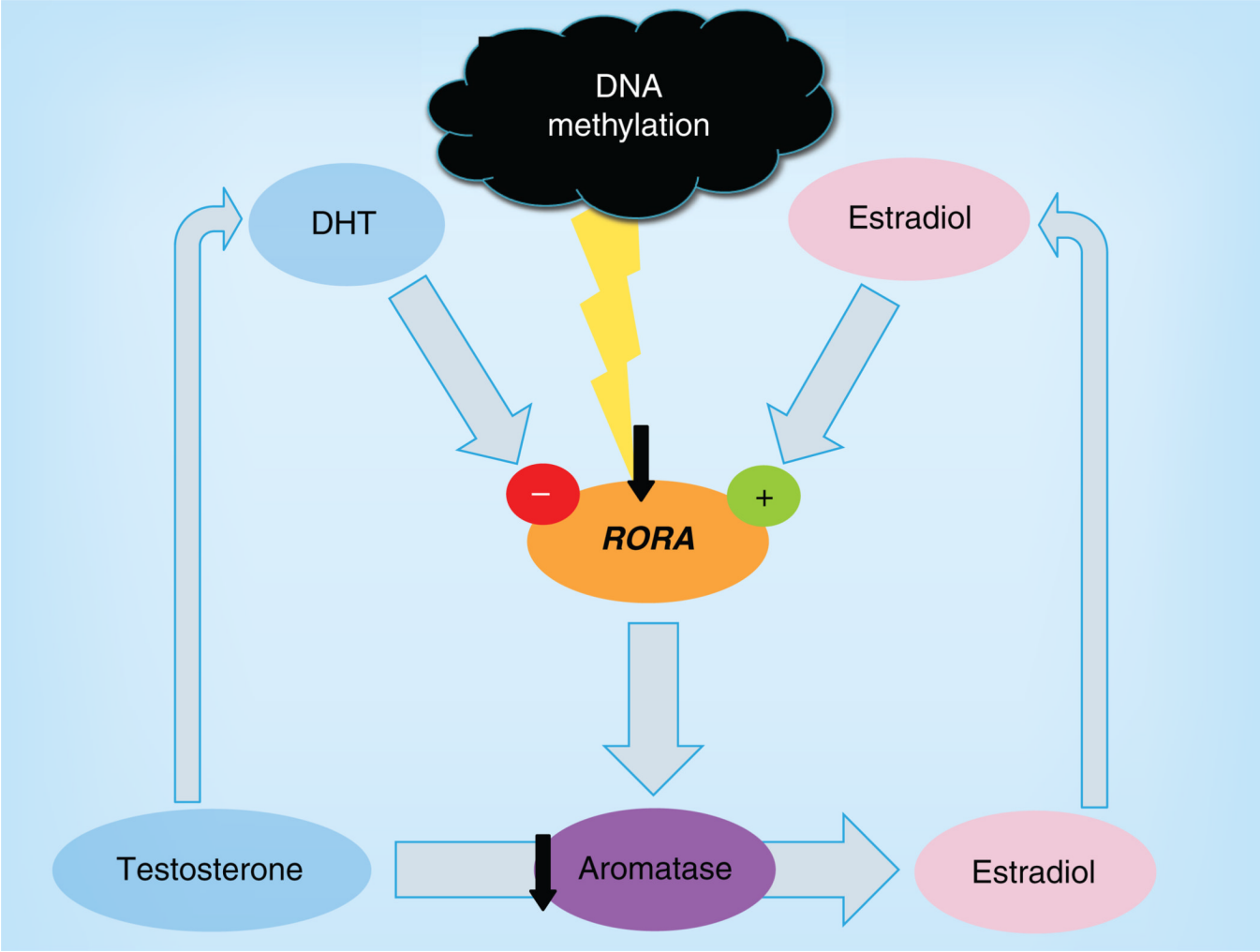

Figure 2. A summary of experimental findings that show how both epigenetic as well as environmental factors (i.e., sex hormones) can contribute to dysregulation of *RORA*, a functionally relevant autism susceptibility gene

DNA methylation decreases expression of *RORA* in lymphoblastoid cell lines [45]. Reflecting the decreased expression in lymphoblastoid cell lines, a decrease in RORA protein is observed in post-mortem brain tissues of individuals with autism spectrum disorders relative to age-matched unaffected controls [45, 47], and is correlated with a decrease in aromatase protein in the frontal cortex. A decrease in aromatase is expected to lead to a build up of its substrate, testosterone, which we show can further suppress *RORA* expression through feedback inhibition by DHT, a potent metabolite of testosterone [47]. Aromatase deficiency also results in lower estradiol, a positive regulator of *RORA* expression [47], thus further exacerbating RORA deficiency. We do not yet know the original underlying cause for the increased methylation of *RORA* in lymphoblastoid cell lines or for reduced RORA protein in brain tissues.

ASD: Autism spectrum disorder; DHT: Dihydrotestosterone.

Adapted from [47], provided by Tewarit Sarachana.