bioRxiv preprint doi: https://doi.org/10.1101/042218; this version posted March 6, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Title: Genetic heterogeneity in autism: from single gene to a pathway perspective

Joon Yong An[1], Charles Claudianos[1,2]*

1. Queensland Brain Institute, University of Queensland, Building 79, St Lucia, 4072, Brisbane, QLD, Australia
2. Institute of Cognitive and Clinical Neuroscience, School of Psychological Sciences, Faculty of Medicine, Nursing and Health Sciences, Monash University, Clayton, VIC 3800, Australia

* Correspondence to C. Claudianos (charles.claudianos@monash.edu)

bioRxiv preprint doi: https://doi.org/10.1101/042218; this version posted March 6, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Abstract:

The extreme genetic heterogeneity of autism spectrum disorder (ASD) represents a major challenge. Recent advances in genetic screening and systems biology approaches have extended our knowledge of the genetic etiology of ASD. In this review, we discuss the paradigm shift from a single gene causation model to pathway perturbation model as a guide to better understand the pathophysiology of ASD. We discuss recent genetic findings obtained through next-generation sequencing (NGS) and examine various integrative analyses using systems biology and complex networks approaches that identify convergent patterns of genetic elements associated with ASD. This review provides a summary of the genetic findings of family-based genome screening studies.

Keywords

Autism Spectrum disorders (ASD), neurodevelopment, genome sequencing, systems biology, brain gene expression

bioRxiv preprint doi: https://doi.org/10.1101/042218; this version posted March 6, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

1. Introduction

Autism spectrum disorder (ASD) is a highly heritable neurodevelopmental disorder that affects up to 1-2% of the population (Autism et al., 2012; Devlin and Scherer, 2012). It is characterized by impaired social interactions, communication deficits and repetitive behavior. Moreover, ASD is often comorbid with other phenotypic and clinical features, such as motor impairments, sleep disturbance, and sensory abnormalities. Not surprisingly, ASD is characterized by both genetic and phenotypic heterogeneity, and there is a high level of variation between individuals diagnosed with autism.

Over the last five years, technological advances have facilitated identification of causal DNA variants associated with ASD. Advances in next-generation sequencing (NGS), whole-exome sequencing (WES), and whole-genome sequencing (WGS) have enabled the identification of large numbers of rare single nucleotide variants (SNVs), including small insertion/deletion mutations (indels) that are associated with ASD. Furthermore, family-based sequencing (trio and sibling families) has helped to disentangle the causal relationship of *de novo* and inherited variants (Ku et al., 2013; Stein et al., 2013), such that, we now understand there are hundreds of genetic variations that affect a wide range of molecular functions associated with ASD (Betancur, 2011a; Buxbaum et al., 2012; Stein et al., 2013). More importantly, these genes are not randomly distributed but converge in functionally relevant biological processes such as synapse development and transcriptional regulation in the brain. Thus, the challenge that arises is to find and precisely identify potentially hundreds of heterogeneous DNA variations that converge in functional pathways associated with ASD. This requires refocusing molecular related morbidity paradigms from single or a few unrelated coding genes to functional

bioRxiv preprint doi: https://doi.org/10.1101/042218; this version posted March 6, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

networks of genes that contribute to key neurodevelopmental and neurological processes.

Systems biology and complex networks approaches are now being used to build the relationships between biological elements among heterogeneous components. This has been important to focus on a level of resolution that allows us to 'see the forest for the trees'. We are mindful that 'nothing in biology makes sense except in the light of evolution' (Dobzhansky, 1973), and that evolutionary constraint underpins the fidelity of genes and molecules that function in biological process. Importantly, this has led to a shift from an exclusive reliance on statistical genetics to more appropriately understanding the evolutionary preconfigured nature of a biological system and complex molecular networks that underpin human development. The application of the systems paradigm in mental health research allows us to examine the relevance of combinations of genetic elements that putatively impact on biological functions associated with ASD.

Much of the scientific momentum regarding systems approaches comes from traction arising from many different omics data (Mitra et al., 2013), increasing our capacity to generate new models, for hypothesis-driven analyses of genetic data (Barabasi et al., 2011). This has led to the development of integrative approaches involving multiple data sources, model construction, and complex network analyses. Here, we provide an overview of recent WES and WGS studies for large ASD cohorts. We are mindful, the systems paradigm does not advocate abandoning statistical approaches but rather refines the search space for probability analysis to be appropriately applied. This

bioRxiv preprint doi: https://doi.org/10.1101/042218; this version posted March 6, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

review summarizes various integrative approaches that were made possible by advancements in systems biology methods.

2. Recent genetic findings in ASD

2.1. *De novo* variants

A major genetic finding was the identification of *de novo* variants (DNVs) associated with ASD. DNVs are DNA variants that occur spontaneously in the parental germline during meiosis (Conrad et al., 2011). Similar to inherited variants, DNVs comprise a wide range of genomic alterations, including chromosomal variations, copy-number variations (CNVs; >1kb), small insertions/deletions (indels; 2-1,000bp) and single nucleotide polymorphisms (SNPs; 1bp). On average, approximately 40 DNVs are thought to arise in an individual's genome (Conrad et al., 2011). Most DNVs appear to be correlated with paternal age (Flatscher-Bader et al., 2011; Hultman et al., 2011; Malaspina et al., 2001; McGrath et al., 2014), as aging germ-line cells involved in spermatogenesis introduce errors during replication, thereby increasing the rate of DNVs (Kong et al., 2012; O'Roak et al., 2011). In sporadic cases of ASD (i.e., ASD probands with unaffected parents), DNVs have been reported to have a large genetic effect, as they have not been subject to natural selection processes in previous generations (Crow, 2000; Eyre-Walker and Keightley, 2007).

A number of previous large-scale studies focused on copy number variations (CNVs) using SNP genotyping arrays and identified *de novo* events associated with ASD (Levy et al., 2011; Marshall et al., 2008; Pinto et al., 2010; Sanders et al., 2011; Sebat et al., 2007; Szatmari et al., 2007). The results highlight several loci where recurrent *de novo* CNVs,

bioRxiv preprint doi: https://doi.org/10.1101/042218; this version posted March 6, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

including 1q21.1, 7q11.23, 15q11-13, 16p11.2, 17q12, and 22q11.2. These CNVs are larger duplications or deletions that encompass multiple genes or regulatory regions and are expected to impart a substantial risk for certain phenotypes. Many of these *de novo* CNVs were shown to affect important synaptic genes, such as *NRXN1* (Szatmari et al., 2007), *NLGN3* (Sanders et al., 2011), *SHANK3* (Marshall et al., 2008), *SHANK2* (Berkel et al., 2010), and *SYNGAP1* (Pinto et al., 2010), suggesting synaptic dysfunction or perturbation of synapse development may underlie the pathophysiology of ASD. Indeed, subsequent studies using animal models and transcriptomics data have elucidated functional changes associated with these CNVs (Durand et al., 2012; Golzio et al., 2012; Luo et al., 2012; Won et al., 2012). *De novo* CNVs have an increased frequency in individuals with ASD compared with unaffected controls, in contrast to inherited CNVs, which occur at similar rates between cases and controls. The enriched occurrence of *de novo* CNVs in ASD cases (accounting for only 5-10% of total ASD cases) and low co-occurrence in unaffected controls (~1%) (Carter and Scherer, 2013; Geschwind, 2011) indicate inherited CNVs are often not causally related and that a minor percentage of *de novo* CNVs are not fully penetrant in ASD (Bucan et al., 2009).

Despite some success showing CNVs being causatively associated with ASD, the SNP genotyping array approach is restricted by the large genomic intervals between SNP markers and cannot be used to reliably investigate SNPs or indels that occur in this intervening space. In contrast recent advances using massive-throughput genome sequencing,, allows point mutations and indels (DNA variations) to be examined on the basis of hundreds of millions of sequence reads. For example, high-fidelity whole-exome sequencing (WES) focuses on genetic information of protein-coding transcripts and

bioRxiv preprint doi: https://doi.org/10.1101/042218; this version posted March 6, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

their nearby regulatory regions (Bamshad et al., 2011). Alternatively, whole-genome sequencing (WGS) provides a more comprehensive readout across the entire genome but it is computationally difficult to analyze and currently still significantly more expensive than WES.

Since the introduction of WES and WGS, increasing evidence has linked DNVs with ASD. Recent family-based studies identified thousands of DNVs in a large number of ASD probands and their siblings (**Table 1**). Loss-of-function variants (LoF variants) – those annotated as nonsense variants, frameshifts or splice site changes – occur at twice the frequency in probands (roughly 0.21 per child) than in unaffected siblings (roughly 0.12 per child) (Iossifov et al., 2014). For this reason, *de novo* LoF variants are considered to provide one of the major genetic contributions to sporadic cases (Hoischen et al., 2014; Ronemus et al., 2014). Importantly, two recent large exome studies, involving 3,871 and 2,508 cases revealed multiple *de novo* LoF variants occur in coding genes – *ADNP*, *ANK2*, *ARID1B*, *CHD8*, *DYRK1A*, *GRIN2B*, *KATNAL2*, *POGZ*, *SCN2A*, and *TBR1* (Iossifov et al., 2014; Poultney et al., 2014).

De novo missense variants are the most abundant type of DNV (approximately 60%) identified in WES and WGS studies. Although they are thought to account for approximately 10% of ASD cases (Ronemus et al., 2014), it is difficult to determine whether *de novo* missense variants show greater frequency in ASD probands than in unaffected siblings. Recent studies reported the rate of *de novo* missense variants was increased in probands versus siblings (O'Roak et al., 2012; Sanders et al., 2012), but another study did not detect this difference (Iossifov et al., 2012). Nevertheless, there

bioRxiv preprint doi: https://doi.org/10.1101/042218; this version posted March 6, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

are a few factors that should be taken into consideration when estimating their contributions. Firstly, most *de novo* missense variants result in deleterious amino acid substitutions (**Table 1**), and their individual effects must be considered in terms of structural modification and functional consequence (binding, dimerization, catalysis, post-translational modification) to the protein or isoform protein . Secondly, a large number of *de novo* missense variants found in the same gene are required to establish a significant association with ASD (Sanders et al., 2012). Thirdly, *de novo* missense variants have been found to be overrepresented during early neurodevelopment and prenatal development (Parikshak et al., 2013).

Interestingly, there are numerous recurring variants observed in ASD probands. These include various types of DNVs that occur in coding regions of SCN2A (Sodium Channel, Voltage-Gated, Type II, Alpha subunit) were found in four different studies (An et al., 2014; Iossifov et al., 2012; Jiang et al., 2013; Sanders et al., 2012). We also know many of the recurrent DNVs have been associated with other neurodevelopmental disorders, such as intellectual disabilities, epileptic encephalopathy's and schizophrenia, suggesting there is high molecular comorbidity between these disorders (Iossifov et al., 2014; Li et al., 2015; Poultney et al., 2014). Cristino *et al*. recently demonstrated there is significant molecular overlap between neurodevelopmental and neuropsychiatric disorders (Cristino et al., 2014a). Apart from recurrent genes containing DNVs, it is noteworthy that there is an increased frequency of DNVs occurring in several gene families. These include voltage-gated calcium channel (*CACNA*), catenin (*CTNN*), and chromodomain helicase DNA binding protein (*CHD*) gene families that appear to contribute to specific ASD phenotypes (Bernier et al., 2014; Pinggera et al., 2015;

bioRxiv preprint doi: https://doi.org/10.1101/042218; this version posted March 6, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Turner et al., 2015). This increased frequency of DNVs in specific gene families highlights related genes with overlapping function contribute to biological pathways and processes associated with synapse development and function, neuronal development, and chromatin modulation (**Figure 1**).

2.2. Rare inherited variants

In contrast to DNVs, inherited variants account for the greatest number of genetic alterations in human genomes and can be classified as either common or rare on the basis of their frequency in the population. Although both common and rare inherited variants may contribute to the development of ASD, rare variants are expected to have a larger effect than individual common variants (State and Levitt, 2011; Stein et al., 2013). Importantly, recent, WES studies have successfully identified a range of rare inherited variants associated with ASD (An et al., 2014; Lim et al., 2013; Toma et al., 2013; Yu et al., 2013).

Rare complete knockouts (homozygous and compound variants) have typically been associated with Mendelian disorders because they more obviously result in deleterious phenotypes. Lim *et al*. recently reported rare complete knockouts in ASD cases using WES analysis of 933 ASD cases and 869 controls to examine (Lim et al., 2013). Among the 933 ASD cases and 869 controls examined there were 91 rare complete knockouts were found. These occur at twice the frequency in cases (n=62) than in controls (n=29), and the authors concluded this incidence accounts for ~3% of ASD cases. This disparity was not observed for rare heterozygous variants. Of particular interest was the high occurrence of rare complete knockouts associated with the X-chromosome in cases

bioRxiv preprint doi: https://doi.org/10.1101/042218; this version posted March 6, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

compared with the controls. The predominant of hemizygous variants found in males or homozygous variants in females are estimated to account for another 2% of ASD cases. In another study by Yu *et al*., heterozygous mapping and linkage analysis was used to examine consanguinity in familial ASD cases (Yu et al., 2013). The authors found rare LoF and missense variants associated with *AMT*, *PEX7* and *SYNE1* genes. They also examined protein-altering variants (missense, nonsense, splice site, and frameshift mutations) present at low frequencies in the population. The analysis of 70 genes previously associated with ASD identified an excessive number of monogenic, autosomal recessive or X-linked variants (Betancur, 2011b). They found five families affected by rare homozygous, compound heterozygous, or hemizygous variants. For example, a nonsense hemizygous variant occurring in *NLGN4X* was identified in one family; this variant was inherited from an unaffected mother in an affected male child. Another family showed the same inheritance pattern for a nonsense hemizygous variant of *MECP2*. Although some rare homozygous variants can explain the etiology of some families with clinical phenotypes associated with certain syndromes (e.g., phenylketonuria, dysmorphic features or microcephaly) others occur in unaffected siblings or cases that do not show clear phenotypes, suggesting hypomorphic variants or other genes might be involved.

Two WES studies have analyzed the role of rare heterozygous variants in ASD. In some multiplex ASD families rare heterozygous truncating variants – nonsense and frameshift variants – are more frequently inherited from parents than non-truncating variants. More importantly, there is an inverse correlation between the number of rare heterozygous truncating variants and non-verbal intelligence quotient (NVIQ) has been

bioRxiv preprint doi: https://doi.org/10.1101/042218; this version posted March 6, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

observed (Toma et al., 2013), confirming previous observation concerning an increased number of DNVs and decreased NVIQ (O'Roak et al., 2012). However, this study did not ascertain the contribution of rare heterozygous missense variants. More recent evidence showed that rare heterozygous missense variants appear to contribute to the etiology of ASD. Recently, An *et al*. (2014) investigated significance of rare heterozygous variants identified by stringent filtering of sequence data. Rare variants were defined as those that a) showed a minor allele frequency (MAF) of <1% or were unknown in the 1,000 Genomes Project and NHLBI Exomes database and b) were predicted to be deleterious by SIFT and PolyPhen2. This analysis identified ~125 rare variants per individual. These variants were significantly enriched in genes that were previously associated with ASD and correlated with the level of rarity (i.e. MAF or prediction scores). Moreover, inherited variants from parents with broader autism phenotype (BAP), a weak phenotype of ASD (Taylor et al., 2013), were enriched in ASD-associated genes compared with inherited variants from parents without BAP. Taken together, this evidence indicates that rare heterozygous variants make a cumulative contribution to ASD.

3. Convergent molecular aberrations underlying ASD

The extreme genetic heterogeneity found in ASD represents a huge challenge in attempting to understand the etiology and development of autism (Betancur, 2011b; Geschwind, 2011). In the absence of any specific gene responsible for the majority of ASD cases, the most common genes only account for approximately 2% of ASD (Abrahams and Geschwind, 2008). Furthermore, it has been shown that candidate genes cover a wide range of functions, including cell-cell adhesion, synaptic activities

bioRxiv preprint doi: https://doi.org/10.1101/042218; this version posted March 6, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

and neurotransmission, transcriptional regulation, and activity-dependent protein translation, and possibly contribute to a number of common clinical features that accompany ASD (Abrahams and Geschwind, 2008), which are generally devalued in the diagnostic processes and in the discovery of candidate genes (Jeste and Geschwind, 2014; Stessman et al., 2014). For this reason, ASD should be considered as a combination of phenotypes, not as a disorder caused by any individual gene (Geschwind, 2011; Mitchell, 2012). It is therefore essential to examine the contribution of the many candidate ASD genes with respect to functions that contribute to diagnostic behaviors in order to elucidate the numerous subtypes of ASD.

This situation has led to the emergence of new approaches for detecting molecular convergence associated with ASD. There are now a number of studies that have attempted to resolve whether putative candidate genes converge in specific molecular or biological processes. These studies are based on the premise that convergent pathway(s) and the co-expression or functional proximity of candidate genes display distinctive patterns of interaction and provide a *segue* for systems biology analyses.

3.1. Interconnection of ASD genes

Many studies have investigated the convergence of ASD-associated CNVs in biological pathways. Gilman *et al*. developed the Network-Based Analysis of Genetic associations (NETBAG) technique for this purpose (Gilman et al., 2011). This method was designed to identify combinations of genes contributing to any one disorder, based on the hypothesis that these genes contribute to phenotypes typical of a disorder. NETBAG used high-confidence datasets of ASD-associated CNVs (Levy et al., 2011) and showed

bioRxiv preprint doi: https://doi.org/10.1101/042218; this version posted March 6, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

these were enriched in pathways related to synaptogenesis, cell-cell adhesion, axon guidance, neuronal motility, GTPase signaling and the actin cytoskeleton. This method was later applied to examine a large sample set, and helped to identify a network of 113 *de novo* CNVs, enriched in functions associated with chromatin and transcription regulation, *MAPK* signaling, and synaptic development (Pinto et al., 2014). Interestingly, putative functional perturbations associated with CNVs are more pronounce in females than in males, consistent with the idea of highly penetrance CNVs occurring in females (Pinto et al., 2014). By considering duplication and deletion variations, Noh et al. have further elucidated the phenotypic consequences of CNVs (Noh et al., 2013).

Importantly, pathways relevant to synaptic functions have repeatedly been associated with ASD pathology (Betancur et al., 2009; Toro et al., 2010). Sakai *et al*. (Sakai et al., 2011) highlight how CNVs converge in neuronal post-synaptic processes in ASD. Specifically, they used a yeast two-hybrid system to examine protein-protein interactions. They found genes from ASD cases had greater connectivity compared to non-ASD individuals. Many key genes (or hub genes) that interact with a number of other genes (protein-protein interactions) are more frequently associated with CNVs found in ASD cases. For example, *SHANK3* is connected with a number of postsynaptic genes (e.g., *PSD95*, *TSC1*, *ACTN1*, and *HOMER3*). *In vivo* evidence (Becker et al., 2014; Tsai et al., 2012) confirms the post-synaptic complex is an important contributor to ASD pathophysiology (Berg and Geschwind, 2012).

This integrative approach provides a basis for constructing more comprehensive models for conceptualizing the molecular relationship between ASD and other neurodevelopmental and neuropsychiatric disorders. Cristino *et al*. (2014a) examined

bioRxiv preprint doi: https://doi.org/10.1101/042218; this version posted March 6, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

the functional patterns of candidate genes associated with ASD, X-linked intellectual disability (XLID), attention deficit hyperactivity disorder (ADHD) and schizophrenia (SZ). In this study, candidate disorder genes and their first-degree interacting proteins were used to create a comprehensive protein-protein interaction network, which was collectively referred to as the AXAS-PPI network. Interestingly, the AXAS-PPI network was shown to have significantly more connections with increased structural properties (e.g., in terms of the average degree, density and clustering coefficient) compared to networks created from an equivalent number of randomly selected genes in the genome. This confirmed the AXAS network reflects meaningful biological signal, confirming neighboring candidate genes function in the related biological processes and pathways. Closer examination revealed this AXAS network is divided into 30 functional modules, with unique patterns of gene distributions across the four disorders and this can be used to analyze the relationship of DNA variations with different disorders. For instance, the module for synapse development and signal transduction was shown to be the largest with disorder genes being equally distributed. Synapse development and signal transduction may therefore be key biological processes common to ASD, XLID, ADHD and SZ. This is consistent with the number of studies that highlight the importance of synaptic genes (*DLG3, DLG4, GRIN1, GRIN2B, PRKCB, STX1B, STXBP1*, and *SYNGAP1*) in ASD, epilepsy, intellectual disabilities and schizophrenia (Hormozdiari et al., 2015).

Complex network models are now proving to be powerful tools for analyzing genetic screening data and abstracting the molecular basis of neurodevelopmental and neuropsychiatric disorders. An *et al*. (2014) utilized the AXAS model to analyze WES

bioRxiv preprint doi: https://doi.org/10.1101/042218; this version posted March 6, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

sequencing data and identify patterns of candidate genes that are putatively overrepresented in functional processes and pathways associated with ASD. They found approximately 100 rare variants per individual case were distributed across a number of different biological pathways and processes. For instance, rare variants from four ASD probands were clustered in the *L1CAM* pathway including combinations of rare inherited variants and DNVs. This result showed there was molecular convergence of predominantly heterozygous variants in families with ASD. Their data suggests combinations of weak heterozygous alleles converge in common biological processes and likely contribute to increase threshold of liability associated with ASD. Again, the clustering phenomena were observed in pathways related to synaptic function (e.g. the neurexin trans-synaptic complex). Indeed, by examining the distribution of rare variants among interacting proteins involved in synapse development and function, they confirm previously reported pattern of pervasive rare variants (Kenny et al., 2013).

The interconnection of ASD genes has been further investigated by considering the expression of alternative spliced genes. Corominas *et al*. (2014) built a PPI network, the Autism Splice-form Interaction Network (ASIN), based on alternatively spliced brain-expressed isoforms and genetic associations (Corominas et al., 2014). The background interactions of the ASIN consisted of experimentally validated isoforms generated from high-throughput transcriptomics sequencing, but the data suffered from a lack of spatiotemporal evidence due to the use of only two adult brain samples (one from an 18-year-old male and another from a 66-year-old female). Interestingly, the interacting partners of the ASIN were enriched for rare *de novo* CNVs. This is consistent with findings in other network-based studies (Cristino et al., 2014a; Liu et al., 2014), and

bioRxiv preprint doi: https://doi.org/10.1101/042218; this version posted March 6, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

implicates functional proximity as a potential measure to identify other candidate ASD genes. For instance, *de novo* LoF variants in GATAD2B have been observed in patients with limited speech (de Ligt et al., 2012), and due to its physical interaction with FOXP2, impairment of GATAD2B may explain this language development issue. Another study proposed a composite gene model by integrating brain-specific exon expression and allele frequencies in the general population (Uddin et al., 2014). This model assumed the integrity of gene expression to be modulated by selective pressure on particular isoform proteins, thus ranking the importance of genes that confer susceptibility. In this model, genes showing high exonic expression in the brain and a low allele frequency significantly overlapped with known ASD candidate genes relevant to fragile-X protein targets.

3.2. Transcriptional convergence on ASD and neurodevelopment

Co-expression networks are widely used for investigating the relationship between genes in specific spatial and temporal contexts. These networks are emerging as a powerful tool and incorporate recent high-resolution transcriptional profiles associated with human brain development (Hawrylycz et al., 2012; Kang et al., 2011; Miller et al., 2014).

The most common approach to look at transcriptome data is Weighted-Gene Co-expression Network Analysis (WGCNA) (Langfelder and Horvath, 2008). WGCNA has been used to survey modules of highly correlated genes typical of specific brain regions and developmental stages. Voineagu *et al*. analyzed postmortem brains to demonstrate common pathophysiological abnormalities associated with cortical areas in ASD cases (Voineagu et al., 2011). They again substantiated numerous synaptic and immune genes

bioRxiv preprint doi: https://doi.org/10.1101/042218; this version posted March 6, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

were found to be perturbed in the frontal and temporal cortex of ASD cases. WGCNA was further applied to examine the global differences in the organization of the brain transcriptome. This analysis revealed salient differences in regional gene expression between autism cases and controls. In the frontal and temporal cortex, gene expression patterns varied significantly among the controls but not in ASD cases, suggesting there are specific regional abnormalities in the brain. Furthermore, co-expression modules similarly confirm synaptic genes were upregulated, whereas immune-related genes were downregulated. The synaptic gene module was found to be enriched with genome-wide association signals, subsequently confirmed by others (Ben-David and Shifman, 2012). Similarly, in a recent study by Gupta *et al*. (2014), the authors used WGCNA to identify the negative correlation between the module of activated M2-state microglia genes and that of synaptic transmission genes from 47 ASD postmortem brains (Gupta et al., 2014).

Several studies have now combined DNA variations discovered from genome sequencing with co-expression network analysis to demonstrate there is a convergence of genetic variations in the perinatal/postnatal cerebellar cortex and layer 5/6 cortical projection neurons (Parikshak et al., 2013; Willsey et al., 2013). Parikshak *et al*. (2013) also demonstrated ASD candidate genes associated with the superficial cortical layers during early neurodevelopment. The analysis of co-expression networks found that DNVs in ASD cases were enriched in the transcriptional regulation module but common inherited variants were enriched in the synaptic transmission module. Furthermore, co-expression modules highlighted the temporal and spatial difference contributes to developmental trajectories. From the prenatal period to infancy, the level of genes

bioRxiv preprint doi: https://doi.org/10.1101/042218; this version posted March 6, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

involved in transcriptional regulation decreased, whereas the levels of genes involved in synaptic transmission increased. In the adult cortex, rare *de novo* and common inherited variants were clustered in superficial cortical layers 2–4 and upper layer glutamatergic neurons. A study by Willsey *et al*. (2013) demonstrated the gene expression changes associated with *de novo* LoF variants in deep projection neurons of the mid-fetal frontal cortex. Furthermore, these variants were enriched at the inner cortical plate and the cortical glutamatergic neurons. Similarly, Li et al. reported cortical disconnections in ASD cases (Li et al., 2014). Combining genomic and RNA sequencing data, they found rare nonsynonymous variants occur in genes involved in oligodendrocyte development, which is the major cell type of the corpus callosum. Taken together, these observations suggest genetically-driven developmental *diaschisis* occurs, whereby early dysfunction may disrupt the maturation of cortical circuits in the brain and contribute to ASD phenotypes (Wang et al., 2014).

The above examples indicate co-expression network analysis can be integrated into a risk prediction model for ASD. The Detecting Association With Networks (DAWN) model proposed by Liu *et al*. (Liu et al., 2014) determined 'hot spots' of genetic associations within a co-expression network. In addition to the statistical assessment of *de novo* and inherited variants (He et al., 2013) showed DAWN was capable of prioritizing possible high-risk genes in the mid-fetal pre-frontal and motor-somatosensory neocortex. This approach was able to provide a rigorous assessment and identification of causal genes that includes evidence from targeted sequencing. Another feature of this model was the identification of sub-networks centered on *FOXP1*

bioRxiv preprint doi: https://doi.org/10.1101/042218; this version posted March 6, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

and *PTEN* clearly demonstrate convergence of ASD-associated genes in neural organization and language development.

4. Concluding remarks

In this review, we provide an overview of published ASD studies based on genome screening and systems biology. Recent advances using pedigree-based genome screening have enabled more rigorous examination of genetic variations. WES and WGS studies have led to the discovery of rare *de novo* and inherited variants with putative greater penetrance than predicted common variants. Moreover, there is great utility with these approaches to interpret biological significance of genetic variations 1) examining causative DNA variations based on an actual nucleotide position (Ku et al., 2013), 2) estimating effect of DNA variations on gene expression in key pathways (Deriziotis et al., 2014), 3) utilizing information of minor allele frequency and evolutionary conservation to characterize rare variations associated with ASD (An et al., 2014).

It is now clear that there is no single ASD-causing gene. Intensive research over the last five years has confirmed there is an extensive genetic heterogeneity associated with ASD. This is supported by understanding that DNA variations often converge in biological pathways and functional processes and are shared across a broad range of neurodevelopmental and neuropsychiatric disorders or specific to only one disorder (Sullivan and Posthuma, 2015). Therefore, systems biology approaches will be a valuable tool to dissect specific differences between disorders (Cristino et al., 2014a;

bioRxiv preprint doi: https://doi.org/10.1101/042218; this version posted March 6, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Parikshak et al., 2013; Willsey et al., 2013), and identify functional distribution of causal DNA variants in individuals with ASD (An et al., 2014).

There are now an impressive range of technologies that can be used investigate functional consequences of genetic variations. Induced pluripotent stem cells (iPSCs) are currently being used to characterize the effect of genetic variations in neuronal cell lineages involved in early neurodevelopment (Griesi-Oliveira et al., 2014). Generating human neurons from iPSCs has potential utility comparing healthy and affected individuals that will help characterize the role of causal ASD genes. The use of clustered regularly interspaced short palindromic repeats (CRISPR)/Cas9 gene editing system with iPS cells provides a tractable method for *in vitro* analysis of DNA variations whereby a number of overlaid genetic elements can be examined additively or in a subtractive fashion to assess relative contribution to risk genes to ASD phenotypes (Bernier et al., 2014). Similarly, these gene editing technologies can be used *in vivo* and aside from rodent models, high throughput small model species (zebra fish, fly, honeybee and worm) are now being used to validate the biological relevance of conserved genes and human genetic variations (Burne et al., 2011), including their contributions to complex behaviors, such as learning and memory (Cristino et al., 2014b).

In conclusion, systems biology approaches have facilitated an important paradigm shift from single gene to pathway etiology associated with ASD. Nevertheless, we are mindful of the enormous task of analyzing non-coding regulatory information with respect to DNA variations. Determining the functional impact of coding and non-coding regulatory

bioRxiv preprint doi: https://doi.org/10.1101/042218; this version posted March 6, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

DNA variations remains a significant hurdle for developing a truly integrated omic systems approach. Resources, such as ENCODE (Encode Project Consortium, 2012), FANTOM (Fantom Consortium, 2014), and mirBase (Kozomara and Griffiths-Jones, 2011), need to be integrated into current analysis pipelines to achieve a comprehensive predictive systems model for ASD.

bioRxiv preprint doi: https://doi.org/10.1101/042218; this version posted March 6, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Acknowledgement

We thank Sarah Williams and Aoife Larkin for critical reading of the manuscript. J-Y An was supported by a University of Queensland PhD scholarship. C.C. was supported by funding from the Australian Research Council (FT110100292), National Health and Medical Research Council (APP1008125).

Tables and Figures

Table 1. *De novo* variants identified through exome and genome sequencing

Study	Method	Probands	Siblings		LoF*	Missense**	Others***	Total		LoF*	Missense**	Others***	Total
O'Roak 2012	WES	209	50	DNVs in probands	38	154(92)	68	260	DNVs in siblings	3	31(21)	16	50
Sanders 2012	WES	225	200		15	125(87)	32	172		5	82(54)	38	125
Neale 2012	WES	175	0		18	101(NA)	50	169		NA	NA	NA	NA
Iossifov 2012	WES	343	343		61	209(NA)	92	362		30	203(NA)	81	314
Jiang 2013	WGS	32	0		5	28(13)	9	42		NA	NA	NA	NA
An 2014	WES	48	0		3	18(11)	8	29		NA	NA	NA	NA
De Rubeis 2014	WES	3,871	NA		307	2,186 (828)	NA	1,135		NA	NA	NA	NA
Iossifov 2014	WES	2,508	1,911		383	1,667 (NA)	646	2,696		178	1,140 (NA)	490	1,808

* LoF variants are nonsense variants, frameshifts or splice-site-changing variants
**The numbers in brackets indicate the number of missense variants classified as putatively deleterious by functional prediction algorithms (e.g., SIFT, PolyPhen2).
*** Including synonymous and non-frameshift variants

Table 2. Network approaches in ASD studies

Study	Network type	Alias	Node source	Interaction source	Highlighted convergent pathways
Gilman 2011	PPI network	NETBAG (NEtwork-Based Analysis of Genetic associations)	De novo copy-number variations associated with ASD	BiGG, BIND, BioGRID, DIP, HPRD, InNetDB, IntAct, MINT, MIPS	• Actin filament organization (GO:0007015) • Axon (GO:0030424) • Cell maturation (GO:0048469) • Learning and/or memory (GO:0007611) • Synapse (GO:0045202)
Sakai 2011	PPI network	NA	ASD syndromic and ASD-associated genes	A yeast two-hybrid screen of a human complementary DNA library	• Cytoskeleton (GO:0005856) • Postsynaptic density (GO:0014069) • Small GTPase-mediated signal transduction (GO:0007264) • Metabolic glutamate receptor signaling pathway (GO:0007216)
Cristino 2014a	PPI network	AXAS (Autism, X-linked intellectual disabilities, Attention deficit hyperactivity disorder, and Schizophrenia PPI network)	Genes associated with ASD, ADHD, SZ and XLID, and their first-degree interacting genes in the whole PPI network	BioGRID, HPRD	• Regulation of transcription (GO:0006355) • Synaptic transmission (GO:0007268) • Signal transduction (GO:0007165) • Cell-cell adhesion (GO:0007155) • Intracellular signaling cascade (GO:0035556)
Corominas 2014	PPI network	ASIN (Autism Spliceform Interaction	Syndromic ASD genes, *de novo* and recurrent CNVs, other variants	BioGRID, DIP, IntAct, MINT, MIPS, PDB	Not provided

		Network)	found in cytogenic screening, animal studies or gene-expression studies		
An 2014	PPI network	NA	Rare inherited and *de novo* variants	BioGRID, HPRD	• Neuron migration (GO:0001764) • Neurotrophin signaling pathway (KEGG:04722) • L1CAM interactions (REACTOME:22205) • Trafficking of AMPA receptors (REACTOME:18307)
Voineague 2011	Co-expression network	NA	Co-expressed genes in autism postmortem brains	Co-expressed module	• Synaptic transmission (GO:0007268) • Neuropeptide hormone activity (GO:0005184) • Cell-cell adhesion (GO:0007155) • Immune response (GO:0006955)
Parikshak 2013	Co-expression network	NA	Co-expressed genes in BrainSpan developmental RNA-seq data	BioGRID, InWeb	• Nucleic acid metabolic process (GO:0090304) • Zinc ion binding (GO:0008270) • Synaptic transmission (GO:0007268) • Cation-transporting ATPase activity
Willsey 2013	Co-expression network	NA	High-confidence de novo LoF variant genes and their co-expressed genes in brain transcriptomics	Co-expressed module	• Cortical projection neurons during mid-fetal periods

			data(Kang et al., 2011)		
Liu 2014	Co-expression network	DAWN (Detecting Association With Networks)	Risk genes defined by Transmission and de novo association (TADA) model(He et al., 2013), and their co-expressed genes in brain transcriptomics data(Kang et al., 2011)	BioGRID, HPRD, IntAct, KEA, KEGG, MINT, MIPS, SNAVI	• Transcriptional regulation • Chromatin regulation • Regulation of translation • Cell adhesion/migration • Transcriptional regulation RNA polymerase II • Protein scaffolding • Ligase activity
Li 2014	PPI network	NA	Interacting genes in BioGRID(Stark et al., 2006)	BioGRID	• Synaptic transmission (GO:0007268)
Hormozdiari 2014	PPI network	MAGI (Merging Affected Genes into Integrated networks)	*De novo* variant genes associated with ASD, epilepsy, and schizophrenia	HPRD, StringDB	• Long-term potentiation (KEGG: 04720) • Synaptic transmission (GO:0007268) • Regulation of transmission of nerve impulse (GO:0051970) • Regulation of neuronal synaptic plasticity (GO:0048168) • SNARE interaction and vesicular transport pathways (KEGG: 04130)
Gupta 2014	Co-expression network	NA	Co-expressed genes in autism postmortem brains	Co-expressed module	• Type I Interferon pathway (GO:0060337) • Defense response to virus (GO:0051607) • Cytokine-mediated signaling pathway

		(GO:0019221) • GABA-related ion channel activity (GO:0022851)

* Interaction sources include databases, including BiGG (Schellenberger et al., 2010), BIND (Bader et al., 2003), BioGRID (Stark et al., 2006), DIP (Xenarios et al., 2002), HPRD (Prasad et al., 2009), KEA (Lachmann and Ma'ayan, 2009), KEGG (Kanehisa and Goto, 2000), InNetDB (Xia et al., 2006), IntAct (Kerrien et al., 2012), InWeb (Rossin et al., 2011), MINT (Zanzoni et al., 2002), MIPS (Mewes et al., 2002), PDB (Berman et al., 2003), SNAVI (Ma'ayan et al., 2009), StringDB (Szklarczyk et al., 2011)

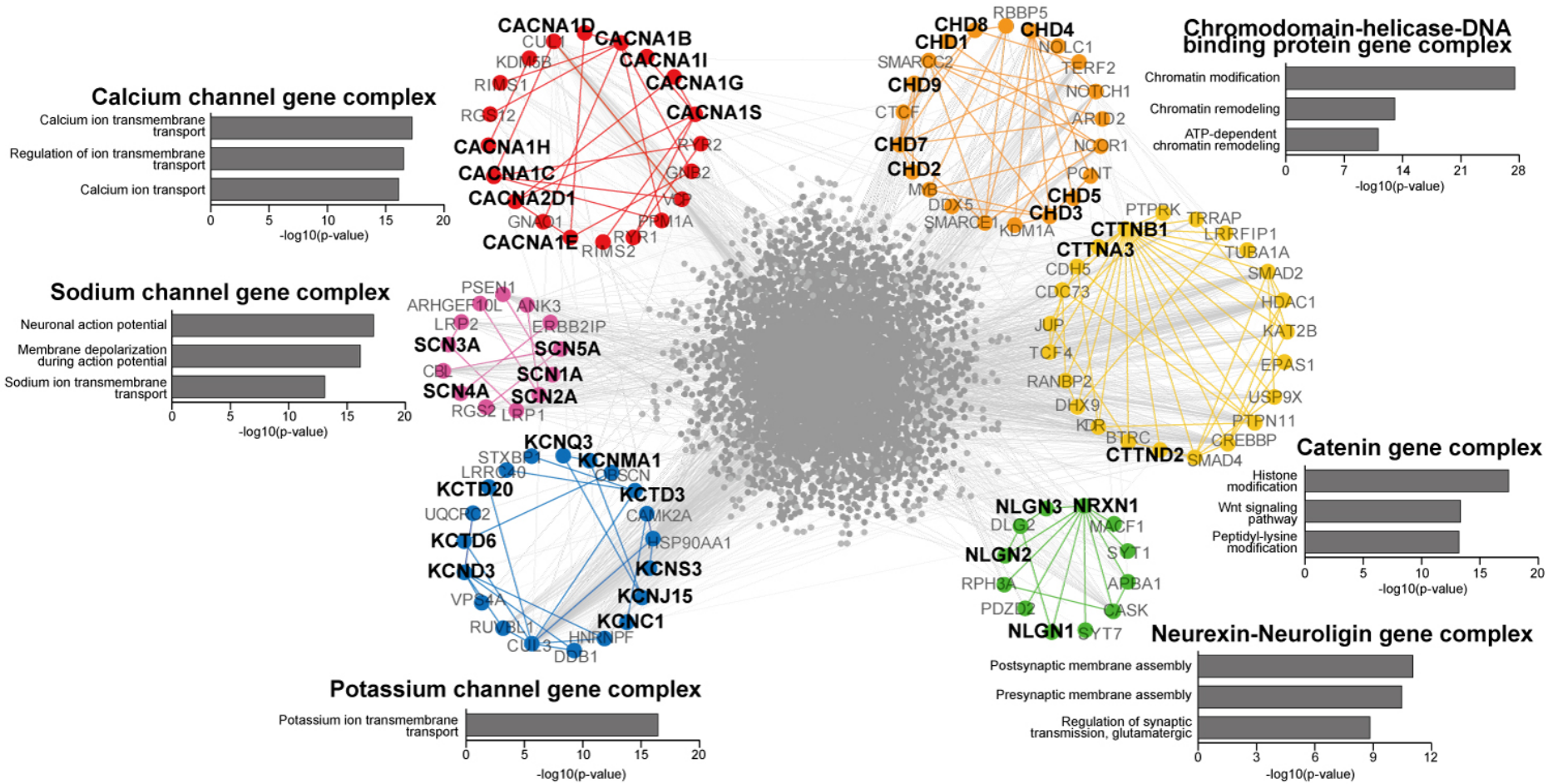

Figure 1. Molecular pathways of *de novo* variants (DNVs) associated with ASD. The protein-protein interaction network plot illustrates six gene families that DNVs were frequently occurred in ASD cases. Interactions were retrieved from HPRD and BioGrid. The sub-network contains DNVs of gene complex and their first-degree interactors. Gene set enrichment analysis was performed to the sub-network genes using the Gene-Ontology database. Bar graphs show top significant pathways associated with the sub-network genes. The y-axis in the bar graph is an enrichment score, provided in a log10-transformed p-value.

bioRxiv preprint doi: https://doi.org/10.1101/042218; this version posted March 6, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

References

Abrahams, B.S., Geschwind, D.H., 2008. Advances in autism genetics: on the threshold of a new neurobiology. Nature Reviews Genetics 9, 341-355.

An, J.Y., Cristino, A.S., Zhao, Q., Edson, J., Williams, S.M., Ravine, D., Wray, J., Marshall, V.M., Hunt, A., Whitehouse, A.J., Claudianos, C., 2014. Towards a molecular characterization of autism spectrum disorders: an exome sequencing and systems approach. Translational psychiatry 4, e394.

Autism, Developmental Disabilities Monitoring Network Surveillance Year Principal, I., Centers for Disease, C., Prevention, 2012. Prevalence of autism spectrum disorders--Autism and Developmental Disabilities Monitoring Network, 14 sites, United States, 2008. Morbidity and mortality weekly report. Surveillance summaries 61, 1-19.

Bader, G.D., Betel, D., Hogue, C.W., 2003. BIND: the Biomolecular Interaction Network Database. Nucleic acids research 31, 248-250.

Bamshad, M.J., Ng, S.B., Bigham, A.W., Tabor, H.K., Emond, M.J., Nickerson, D.A., Shendure, J., 2011. Exome sequencing as a tool for Mendelian disease gene discovery. Nat Rev Genet 12, 745-755.

Barabasi, A.L., Gulbahce, N., Loscalzo, J., 2011. Network medicine: a network-based approach to human disease. Nat Rev Genet 12, 56-68.

Becker, J.A., Clesse, D., Spiegelhalter, C., Schwab, Y., Le Merrer, J., Kieffer, B.L., 2014. Autistic-Like Syndrome in Mu Opioid Receptor Null Mice is Relieved by Facilitated mGluR4 Activity. Neuropsychopharmacology : official publication of the American College of Neuropsychopharmacology 39, 2049-2060.

bioRxiv preprint doi: https://doi.org/10.1101/042218; this version posted March 6, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Ben-David, E., Shifman, S., 2012. Networks of neuronal genes affected by common and rare variants in autism spectrum disorders. PLoS genetics 8, e1002556.

Berg, J.M., Geschwind, D.H., 2012. Autism genetics: searching for specificity and convergence. Genome biology 13, 247.

Berkel, S., Marshall, C.R., Weiss, B., Howe, J., Roeth, R., Moog, U., Endris, V., Roberts, W., Szatmari, P., Pinto, D., Bonin, M., Riess, A., Engels, H., Sprengel, R., Scherer, S.W., Rappold, G.A., 2010. Mutations in the SHANK2 synaptic scaffolding gene in autism spectrum disorder and mental retardation. Nature genetics 42, 489-491.

Berman, H., Henrick, K., Nakamura, H., 2003. Announcing the worldwide Protein Data Bank. Nature structural biology 10, 980.

Bernier, R., Golzio, C., Xiong, B., Stessman, H.A., Coe, B.P., Penn, O., Witherspoon, K., Gerdts, J., Baker, C., Vulto-van Silfhout, A.T., Schuurs-Hoeijmakers, J.H., Fichera, M., Bosco, P., Buono, S., Alberti, A., Failla, P., Peeters, H., Steyaert, J., Vissers, L.E., Francescatto, L., Mefford, H.C., Rosenfeld, J.A., Bakken, T., O'Roak, B.J., Pawlus, M., Moon, R., Shendure, J., Amaral, D.G., Lein, E., Rankin, J., Romano, C., de Vries, B.B., Katsanis, N., Eichler, E.E., 2014. Disruptive CHD8 Mutations Define a Subtype of Autism Early in Development. Cell 158, 263-276.

Betancur, C., 2011a. Etiological heterogeneity in autism spectrum disorders: more than 100 genetic and genomic disorders and still counting. Brain research 1380, 42-77.

Betancur, C., 2011b. Etiological heterogeneity in autism spectrum disorders: More than 100 genetic and genomic disorders and still counting. Brain research 1380, 42-77.

Betancur, C., Sakurai, T., Buxbaum, J.D., 2009. The emerging role of synaptic cell-adhesion pathways in the pathogenesis of autism spectrum disorders. Trends in neurosciences 32, 402-412.

bioRxiv preprint doi: https://doi.org/10.1101/042218; this version posted March 6, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Bucan, M., Abrahams, B.S., Wang, K., Glessner, J.T., Herman, E.I., Sonnenblick, L.I., Alvarez Retuerto, A.I., Imielinski, M., Hadley, D., Bradfield, J.P., Kim, C., Gidaya, N.B., Lindquist, I., Hutman, T., Sigman, M., Kustanovich, V., Lajonchere, C.M., Singleton, A., Kim, J., Wassink, T.H., McMahon, W.M., Owley, T., Sweeney, J.A., Coon, H., Nurnberger, J.I., Li, M., Cantor, R.M., Minshew, N.J., Sutcliffe, J.S., Cook, E.H., Dawson, G., Buxbaum, J.D., Grant, S.F., Schellenberg, G.D., Geschwind, D.H., Hakonarson, H., 2009. Genome-wide analyses of exonic copy number variants in a family-based study point to novel autism susceptibility genes. PLoS genetics 5, e1000536.

Burne, T., Scott, E., van Swinderen, B., Hilliard, M., Reinhard, J., Claudianos, C., Eyles, D., McGrath, J., 2011. Big ideas for small brains: what can psychiatry learn from worms, flies, bees and fish? Molecular psychiatry 16, 7-16.

Buxbaum, J.D., Daly, M.J., Devlin, B., Lehner, T., Roeder, K., State, M.W., Autism Sequencing, C., 2012. The autism sequencing consortium: large-scale, high-throughput sequencing in autism spectrum disorders. Neuron 76, 1052-1056.

Carter, M.T., Scherer, S.W., 2013. Autism spectrum disorder in the genetics clinic: a review. Clinical genetics 83, 399-407.

Conrad, D.F., Keebler, J.E., DePristo, M.A., Lindsay, S.J., Zhang, Y., Casals, F., Idaghdour, Y., Hartl, C.L., Torroja, C., Garimella, K.V., Zilversmit, M., Cartwright, R., Rouleau, G.A., Daly, M., Stone, E.A., Hurles, M.E., Awadalla, P., Genomes, P., 2011. Variation in genome-wide mutation rates within and between human families. Nature genetics 43, 712-714.

Corominas, R., Yang, X., Lin, G.N., Kang, S., Shen, Y., Ghamsari, L., Broly, M., Rodriguez, M., Tam, S., Trigg, S.A., Fan, C., Yi, S., Tasan, M., Lemmens, I., Kuang, X., Zhao, N., Malhotra, D., Michaelson, J.J., Vacic, V., Calderwood, M.A., Roth, F.P., Tavernier, J., Horvath, S., Salehi-Ashtiani, K., Korkin, D., Sebat, J., Hill, D.E., Hao, T., Vidal, M., Iakoucheva, L.M., 2014.

bioRxiv preprint doi: https://doi.org/10.1101/042218; this version posted March 6, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Protein interaction network of alternatively spliced isoforms from brain links genetic risk factors for autism. Nature communications 5, 3650.

Cristino, A.S., Barchuk, A.R., Freitas, F.C., Narayanan, R.K., Biergans, S.D., Zhao, Z., Simoes, Z.L., Reinhard, J., Claudianos, C., 2014b. Neuroligin-associated microRNA-932 targets actin and regulates memory in the honeybee. Nature communications 5, 5529.

Cristino, A.S., Williams, S.M., Hawi, Z., An, J.Y., Bellgrove, M.A., Schwartz, C.E., Costa Lda, F., Claudianos, C., 2014a. Neurodevelopmental and neuropsychiatric disorders represent an interconnected molecular system. Molecular psychiatry 19, 294-301.

Crow, J.F., 2000. The origins patterns and implications of human spontaneous mutation. Nature Reviews Genetics 1, 40-47.

de Ligt, J., Willemsen, M.H., van Bon, B.W., Kleefstra, T., Yntema, H.G., Kroes, T., Vulto-van Silfhout, A.T., Koolen, D.A., de Vries, P., Gilissen, C., del Rosario, M., Hoischen, A., Scheffer, H., de Vries, B.B., Brunner, H.G., Veltman, J.A., Vissers, L.E., 2012. Diagnostic exome sequencing in persons with severe intellectual disability. The New England journal of medicine 367, 1921-1929.

Deriziotis, P., O'Roak, B.J., Graham, S.A., Estruch, S.B., Dimitropoulou, D., Bernier, R.A., Gerdts, J., Shendure, J., Eichler, E.E., Fisher, S.E., 2014. De novo TBR1 mutations in sporadic autism disrupt protein functions. Nature communications 5, 4954.

Devlin, B., Scherer, S.W., 2012. Genetic architecture in autism spectrum disorder. Current opinion in genetics & development 22, 229-237.

Dobzhansky, T., 1973. Nothing in Biology Makes Sense except in the Light of Evolution. The American Biology Teacher 35, 125-129.

bioRxiv preprint doi: https://doi.org/10.1101/042218; this version posted March 6, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Durand, C.M., Perroy, J., Loll, F., Perrais, D., Fagni, L., Bourgeron, T., Montcouquiol, M., Sans, N., 2012. SHANK3 mutations identified in autism lead to modification of dendritic spine morphology via an actin-dependent mechanism. Molecular psychiatry 17, 71-84.

Encode Project Consortium, 2012. An integrated encyclopedia of DNA elements in the human genome. Nature 489, 57-74.

Eyre-Walker, A., Keightley, P.D., 2007. The distribution of fitness effects of new mutations. Nat Rev Genet 8, 610-618.

Fantom Consortium, 2014. A promoter-level mammalian expression atlas. Nature 507, 462-470.

Flatscher-Bader, T., Foldi, C.J., Chong, S., Whitelaw, E., Moser, R.J., Burne, T.H., Eyles, D.W., McGrath, J.J., 2011. Increased de novo copy number variants in the offspring of older males. Translational psychiatry 1, e34.

Geschwind, D.H., 2011. Genetics of autism spectrum disorders. Trends in cognitive sciences 15, 409-416.

Gilman, S.R., Iossifov, I., Levy, D., Ronemus, M., Wigler, M., Vitkup, D., 2011. Rare De Novo Variants Associated with Autism Implicate a Large Functional Network of Genes Involved in Formation and Function of Synapses. Neuron 70, 898-907.

Golzio, C., Willer, J., Talkowski, M.E., Oh, E.C., Taniguchi, Y., Jacquemont, S., Reymond, A., Sun, M., Sawa, A., Gusella, J.F., Kamiya, A., Beckmann, J.S., Katsanis, N., 2012. KCTD13 is a major driver of mirrored neuroanatomical phenotypes of the 16p11.2 copy number variant. Nature 485, 363-367.

Griesi-Oliveira, K., Acab, A., Gupta, A.R., Sunaga, D.Y., Chailangkarn, T., Nicol, X., Nunez, Y., Walker, M.F., Murdoch, J.D., Sanders, S.J., Fernandez, T.V., Ji, W., Lifton, R.P., Vadasz, E., Dietrich, A., Pradhan, D., Song, H., Ming, G.L., Gu, X., Haddad, G., Marchetto, M.C., Spitzer,

bioRxiv preprint doi: https://doi.org/10.1101/042218; this version posted March 6, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

N., Passos-Bueno, M.R., State, M.W., Muotri, A.R., 2014. Modeling non-syndromic autism and the impact of TRPC6 disruption in human neurons. Molecular psychiatry.

Gupta, S., Ellis, S.E., Ashar, F.N., Moes, A., Bader, J.S., Zhan, J., West, A.B., Arking, D.E., 2014. Transcriptome analysis reveals dysregulation of innate immune response genes and neuronal activity-dependent genes in autism. Nature communications 5, 5748.

Hawrylycz, M.J., Lein, E.S., Guillozet-Bongaarts, A.L., Shen, E.H., Ng, L., Miller, J.A., van de Lagemaat, L.N., Smith, K.A., Ebbert, A., Riley, Z.L., Abajian, C., Beckmann, C.F., Bernard, A., Bertagnolli, D., Boe, A.F., Cartagena, P.M., Chakravarty, M.M., Chapin, M., Chong, J., Dalley, R.A., Daly, B.D., Dang, C., Datta, S., Dee, N., Dolbeare, T.A., Faber, V., Feng, D., Fowler, D.R., Goldy, J., Gregor, B.W., Haradon, Z., Haynor, D.R., Hohmann, J.G., Horvath, S., Howard, R.E., Jeromin, A., Jochim, J.M., Kinnunen, M., Lau, C., Lazarz, E.T., Lee, C., Lemon, T.A., Li, L., Li, Y., Morris, J.A., Overly, C.C., Parker, P.D., Parry, S.E., Reding, M., Royall, J.J., Schulkin, J., Sequeira, P.A., Slaughterbeck, C.R., Smith, S.C., Sodt, A.J., Sunkin, S.M., Swanson, B.E., Vawter, M.P., Williams, D., Wohnoutka, P., Zielke, H.R., Geschwind, D.H., Hof, P.R., Smith, S.M., Koch, C., Grant, S.G., Jones, A.R., 2012. An anatomically comprehensive atlas of the adult human brain transcriptome. Nature 489, 391-399.

He, X., Sanders, S.J., Liu, L., De Rubeis, S., Lim, E.T., Sutcliffe, J.S., Schellenberg, G.D., Gibbs, R.A., Daly, M.J., Buxbaum, J.D., State, M.W., Devlin, B., Roeder, K., 2013. Integrated model of de novo and inherited genetic variants yields greater power to identify risk genes. PLoS genetics 9, e1003671.

Hoischen, A., Krumm, N., Eichler, E.E., 2014. Prioritization of neurodevelopmental disease genes by discovery of new mutations. Nature neuroscience 17, 764-772.

Hormozdiari, F., Penn, O., Borenstein, E., Eichler, E.E., 2015. The discovery of integrated gene networks for autism and related disorders. Genome research 25, 142-154.

bioRxiv preprint doi: https://doi.org/10.1101/042218; this version posted March 6, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Hultman, C.M., Sandin, S., Levine, S.Z., Lichtenstein, P., Reichenberg, A., 2011. Advancing paternal age and risk of autism: new evidence from a population-based study and a meta-analysis of epidemiological studies. Molecular psychiatry 16, 1203-1212.

Iossifov, I., O'Roak, B.J., Sanders, S.J., Ronemus, M., Krumm, N., Levy, D., Stessman, H.A., Witherspoon, K.T., Vives, L., Patterson, K.E., Smith, J.D., Paeper, B., Nickerson, D.A., Dea, J., Dong, S., Gonzalez, L.E., Mandell, J.D., Mane, S.M., Murtha, M.T., Sullivan, C.A., Walker, M.F., Waqar, Z., Wei, L., Willsey, A.J., Yamrom, B., Lee, Y.H., Grabowska, E., Dalkic, E., Wang, Z., Marks, S., Andrews, P., Leotta, A., Kendall, J., Hakker, I., Rosenbaum, J., Ma, B., Rodgers, L., Troge, J., Narzisi, G., Yoon, S., Schatz, M.C., Ye, K., McCombie, W.R., Shendure, J., Eichler, E.E., State, M.W., Wigler, M., 2014. The contribution of de novo coding mutations to autism spectrum disorder. Nature 515, 216-221.

Iossifov, I., Ronemus, M., Levy, D., Wang, Z., Hakker, I., Rosenbaum, J., Yamrom, B., Lee, Y.-h., Narzisi, G., Leotta, A., Kendall, J., Grabowska, E., Ma, B., Marks, S., Rodgers, L., Stepansky, A., Troge, J., Andrews, P., Bekritsky, M., Pradhan, K., Ghiban, E., Kramer, M., Parla, J., Demeter, R., Fulton, L.L., Fulton, R.S., Magrini, V.J., Ye, K., Darnell, J.C., Darnell, R.B., Mardis, E.R., Wilson, R.K., Schatz, M.C., McCombie, W.R., Wigler, M., 2012. De Novo Gene Disruptions in Children on the Autistic Spectrum. Neuron 74, 285-299.

Jeste, S.S., Geschwind, D.H., 2014. Disentangling the heterogeneity of autism spectrum disorder through genetic findings. Nature reviews. Neurology.

Jiang, Y.H., Yuen, R.K., Jin, X., Wang, M., Chen, N., Wu, X., Ju, J., Mei, J., Shi, Y., He, M., Wang, G., Liang, J., Wang, Z., Cao, D., Carter, M.T., Chrysler, C., Drmic, I.E., Howe, J.L., Lau, L., Marshall, C.R., Merico, D., Nalpathamkalam, T., Thiruvahindrapuram, B., Thompson, A., Uddin, M., Walker, S., Luo, J., Anagnostou, E., Zwaigenbaum, L., Ring, R.H., Wang, J., Lajonchere, C., Wang, J., Shih, A., Szatmari, P., Yang, H., Dawson, G., Li, Y., Scherer, S.W.,

bioRxiv preprint doi: https://doi.org/10.1101/042218; this version posted March 6, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

2013. Detection of Clinically Relevant Genetic Variants in Autism Spectrum Disorder by Whole-Genome Sequencing. American journal of human genetics.

Kanehisa, M., Goto, S., 2000. KEGG: kyoto encyclopedia of genes and genomes. Nucleic acids research 28, 27-30.

Kang, H.J., Kawasawa, Y.I., Cheng, F., Zhu, Y., Xu, X., Li, M., Sousa, A.M., Pletikos, M., Meyer, K.A., Sedmak, G., Guennel, T., Shin, Y., Johnson, M.B., Krsnik, Z., Mayer, S., Fertuzinhos, S., Umlauf, S., Lisgo, S.N., Vortmeyer, A., Weinberger, D.R., Mane, S., Hyde, T.M., Huttner, A., Reimers, M., Kleinman, J.E., Sestan, N., 2011. Spatio-temporal transcriptome of the human brain. Nature 478, 483-489.

Kenny, E.M., Cormican, P., Furlong, S., Heron, E., Kenny, G., Fahey, C., Kelleher, E., Ennis, S., Tropea, D., Anney, R., Corvin, A.P., Donohoe, G., Gallagher, L., Gill, M., Morris, D.W., 2013. Excess of rare novel loss-of-function variants in synaptic genes in schizophrenia and autism spectrum disorders. Molecular psychiatry.

Kerrien, S., Aranda, B., Breuza, L., Bridge, A., Broackes-Carter, F., Chen, C., Duesbury, M., Dumousseau, M., Feuermann, M., Hinz, U., Jandrasits, C., Jimenez, R.C., Khadake, J., Mahadevan, U., Masson, P., Pedruzzi, I., Pfeiffenberger, E., Porras, P., Raghunath, A., Roechert, B., Orchard, S., Hermjakob, H., 2012. The IntAct molecular interaction database in 2012. Nucleic acids research 40, D841-846.

Kong, A., Frigge, M.L., Masson, G., Besenbacher, S., Sulem, P., Magnusson, G., Gudjonsson, S.A., Sigurdsson, A., Jonasdottir, A., Jonasdottir, A., Wong, W.S., Sigurdsson, G., Walters, G.B., Steinberg, S., Helgason, H., Thorleifsson, G., Gudbjartsson, D.F., Helgason, A., Magnusson, O.T., Thorsteinsdottir, U., Stefansson, K., 2012. Rate of de novo mutations and the importance of father's age to disease risk. Nature 488, 471-475.

bioRxiv preprint doi: https://doi.org/10.1101/042218; this version posted March 6, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Kozomara, A., Griffiths-Jones, S., 2011. miRBase: integrating microRNA annotation and deep-sequencing data. Nucleic acids research 39, D152-157.

Ku, C.S., Polychronakos, C., Tan, E.K., Naidoo, N., Pawitan, Y., Roukos, D.H., Mort, M., Cooper, D.N., 2013. A new paradigm emerges from the study of de novo mutations in the context of neurodevelopmental disease. Molecular psychiatry 18, 141-153.

Lachmann, A., Ma'ayan, A., 2009. KEA: kinase enrichment analysis. Bioinformatics 25, 684-686.

Langfelder, P., Horvath, S., 2008. WGCNA: an R package for weighted correlation network analysis. BMC bioinformatics 9, 559.

Levy, D., Ronennus, M., Yamrom, B., Lee, Y.-H., Leotta, A., Kendall, J., Marks, S., Lakshmi, B., Pai, D., Ye, K., Buja, A., Krieger, A., Yoon, S., Troge, J., Rodgers, L., Lossifov, I., Wigler, M., 2011. Rare De Novo and Transmitted Copy-Number Variation in Autistic Spectrum Disorders. Neuron 70, 886-897.

Li, J., Cai, T., Jiang, Y., Chen, H., He, X., Chen, C., Li, X., Shao, Q., Ran, X., Li, Z., Xia, K., Liu, C., Sun, Z.S., Wu, J., 2015. Genes with de novo mutations are shared by four neuropsychiatric disorders discovered from NPdenovo database. Molecular psychiatry.

Li, J., Shi, M., Ma, Z., Zhao, S., Euskirchen, G., Ziskin, J., Urban, A., Hallmayer, J., Snyder, M., 2014. Integrated systems analysis reveals a molecular network underlying autism spectrum disorders. Molecular systems biology 10, 774.

Lim, E.T., Raychaudhuri, S., Sanders, S.J., Stevens, C., Sabo, A., MacArthur, D.G., Neale, B.M., Kirby, A., Ruderfer, D.M., Fromer, M., Lek, M., Liu, L., Flannick, J., Ripke, S., Nagaswamy, U., Muzny, D., Reid, J.G., Hawes, A., Newsham, I., Wu, Y., Lewis, L., Dinh, H., Gross, S., Wang, L.S., Lin, C.F., Valladares, O., Gabriel, S.B., dePristo, M., Altshuler, D.M., Purcell, S.M., Project, N.E.S., State, M.W., Boerwinkle, E., Buxbaum, J.D., Cook, E.H., Gibbs,

bioRxiv preprint doi: https://doi.org/10.1101/042218; this version posted March 6, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

R.A., Schellenberg, G.D., Sutcliffe, J.S., Devlin, B., Roeder, K., Daly, M.J., 2013. Rare complete knockouts in humans: population distribution and significant role in autism spectrum disorders. Neuron 77, 235-242.

Liu, L., Lei, J., Sanders, S.J., Willsey, A.J., Kou, Y., Cicek, A.E., Klei, L., Lu, C., He, X., Li, M., Muhle, R.A., Ma'ayan, A., Noonan, J.P., Sestan, N., McFadden, K.A., State, M.W., Buxbaum, J.D., Devlin, B., Roeder, K., 2014. DAWN: a framework to identify autism genes and subnetworks using gene expression and genetics. Molecular autism 5, 22.

Luo, R., Sanders, S.J., Tian, Y., Voineagu, I., Huang, N., Chu, S.H., Klei, L., Cai, C., Ou, J., Lowe, J.K., Hurles, M.E., Devlin, B., State, M.W., Geschwind, D.H., 2012. Genome-wide transcriptome profiling reveals the functional impact of rare de novo and recurrent CNVs in autism spectrum disorders. American journal of human genetics 91, 38-55.

Ma'ayan, A., Jenkins, S.L., Webb, R.L., Berger, S.I., Purushothaman, S.P., Abul-Husn, N.S., Posner, J.M., Flores, T., Iyengar, R., 2009. SNAVI: Desktop application for analysis and visualization of large-scale signaling networks. BMC systems biology 3, 10.

Malaspina, D., Harlap, S., Fennig, S., Heiman, D., Nahon, D., Feldman, D., Susser, E.S., 2001. Advancing paternal age and the risk of schizophrenia. Archives of general psychiatry 58, 361-367.

Marshall, C.R., Noor, A., Vincent, J.B., Lionel, A.C., Feuk, L., Skaug, J., Shago, M., Moessner, R., Pinto, D., Ren, Y., Thiruvahindrapduram, B., Fiebig, A., Schreiber, S., Friedman, J., Ketelaars, C.E.J., Vos, Y.J., Ficicioglu, C., Kirkpatrick, S., Nicolson, R., Sloman, L., Surnmers, A., Gibbons, C.A., Teebi, A., Chitayat, D., Weksberg, R., Thompson, A., Vardy, C., Crosbie, V., Luscombe, S., Baatjes, R., Zwaigenbaum, L., Roberts, W., Fernandez, B., Szatmari, P., Scherer, S.W., 2008. Structural variation of chromosomes in autism spectrum disorder. American journal of human genetics 82, 477-488.

bioRxiv preprint doi: https://doi.org/10.1101/042218; this version posted March 6, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

McGrath, J.J., Petersen, L., Agerbo, E., Mors, O., Mortensen, P.B., Pedersen, C.B., 2014. A comprehensive assessment of parental age and psychiatric disorders. JAMA psychiatry 71, 301-309.

Mewes, H.W., Frishman, D., Guldener, U., Mannhaupt, G., Mayer, K., Mokrejs, M., Morgenstern, B., Munsterkotter, M., Rudd, S., Weil, B., 2002. MIPS: a database for genomes and protein sequences. Nucleic acids research 30, 31-34.

Miller, J.A., Ding, S.L., Sunkin, S.M., Smith, K.A., Ng, L., Szafer, A., Ebbert, A., Riley, Z.L., Royall, J.J., Aiona, K., Arnold, J.M., Bennet, C., Bertagnolli, D., Brouner, K., Butler, S., Caldejon, S., Carey, A., Cuhaciyan, C., Dalley, R.A., Dee, N., Dolbeare, T.A., Facer, B.A., Feng, D., Fliss, T.P., Gee, G., Goldy, J., Gourley, L., Gregor, B.W., Gu, G., Howard, R.E., Jochim, J.M., Kuan, C.L., Lau, C., Lee, C.K., Lee, F., Lemon, T.A., Lesnar, P., McMurray, B., Mastan, N., Mosqueda, N., Naluai-Cecchini, T., Ngo, N.K., Nyhus, J., Oldre, A., Olson, E., Parente, J., Parker, P.D., Parry, S.E., Stevens, A., Pletikos, M., Reding, M., Roll, K., Sandman, D., Sarreal, M., Shapouri, S., Shapovalova, N.V., Shen, E.H., Sjoquist, N., Slaughterbeck, C.R., Smith, M., Sodt, A.J., Williams, D., Zollei, L., Fischl, B., Gerstein, M.B., Geschwind, D.H., Glass, I.A., Hawrylycz, M.J., Hevner, R.F., Huang, H., Jones, A.R., Knowles, J.A., Levitt, P., Phillips, J.W., Sestan, N., Wohnoutka, P., Dang, C., Bernard, A., Hohmann, J.G., Lein, E.S., 2014. Transcriptional landscape of the prenatal human brain. Nature 508, 199-206.

Mitchell, K.J., 2012. What is complex about complex disorders? Genome biology 13, 237.

Mitra, K., Carvunis, A.R., Ramesh, S.K., Ideker, T., 2013. Integrative approaches for finding modular structure in biological networks. Nat Rev Genet 14, 719-732.

Noh, H.J., Ponting, C.P., Boulding, H.C., Meader, S., Betancur, C., Buxbaum, J.D., Pinto, D., Marshall, C.R., Lionel, A.C., Scherer, S.W., Webber, C., 2013. Network topologies and

bioRxiv preprint doi: https://doi.org/10.1101/042218; this version posted March 6, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

convergent aetiologies arising from deletions and duplications observed in individuals with autism. PLoS genetics 9, e1003523.

O'Roak, B.J., Deriziotis, P., Lee, C., Vives, L., Schwartz, J.J., Girirajan, S., Karakoc, E., Mackenzie, A.P., Ng, S.B., Baker, C., Rieder, M.J., Nickerson, D.A., Bernier, R., Fisher, S.E., Shendure, J., Eichler, E.E., 2011. Exome sequencing in sporadic autism spectrum disorders identifies severe de novo mutations. Nature genetics 43, 585-589.

O'Roak, B.J., Vives, L., Girirajan, S., Karakoc, E., Krumm, N., Coe, B.P., Levy, R., Ko, A., Lee, C., Smith, J.D., Turner, E.H., Stanaway, I.B., Vernot, B., Malig, M., Baker, C., Reilly, B., Akey, J.M., Borenstein, E., Rieder, M.J., Nickerson, D.A., Bernier, R., Shendure, J., Eichler, E.E., 2012. Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. Nature 485, 246-U136.

Parikshak, N.N., Luo, R., Zhang, A., Won, H., Lowe, J.K., Chandran, V., Horvath, S., Geschwind, D.H., 2013. Integrative functional genomic analyses implicate specific molecular pathways and circuits in autism. Cell 155, 1008-1021.

Pinggera, A., Lieb, A., Benedetti, B., Lampert, M., Monteleone, S., Liedl, K.R., Tuluc, P., Striessnig, J., 2015. CACNA1D De Novo Mutations in Autism Spectrum Disorders Activate Cav1.3 L-Type Calcium Channels. Biological psychiatry 77, 816-822.

Pinto, D., Delaby, E., Merico, D., Barbosa, M., Merikangas, A., Klei, L., Thiruvahindrapuram, B., Xu, X., Ziman, R., Wang, Z., Vorstman, J.A., Thompson, A., Regan, R., Pilorge, M., Pellecchia, G., Pagnamenta, A.T., Oliveira, B., Marshall, C.R., Magalhaes, T.R., Lowe, J.K., Howe, J.L., Griswold, A.J., Gilbert, J., Duketis, E., Dombroski, B.A., De Jonge, M.V., Cuccaro, M., Crawford, E.L., Correia, C.T., Conroy, J., Conceicao, I.C., Chiocchetti, A.G., Casey, J.P., Cai, G., Cabrol, C., Bolshakova, N., Bacchelli, E., Anney, R., Gallinger, S., Cotterchio, M., Casey, G., Zwaigenbaum, L., Wittemeyer, K., Wing, K.,

bioRxiv preprint doi: https://doi.org/10.1101/042218; this version posted March 6, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Wallace, S., van Engeland, H., Tryfon, A., Thomson, S., Soorya, L., Roge, B., Roberts, W., Poustka, F., Mouga, S., Minshew, N., McInnes, L.A., McGrew, S.G., Lord, C., Leboyer, M., Le Couteur, A.S., Kolevzon, A., Jimenez Gonzalez, P., Jacob, S., Holt, R., Guter, S., Green, J., Green, A., Gillberg, C., Fernandez, B.A., Duque, F., Delorme, R., Dawson, G., Chaste, P., Cafe, C., Brennan, S., Bourgeron, T., Bolton, P.F., Bolte, S., Bernier, R., Baird, G., Bailey, A.J., Anagnostou, E., Almeida, J., Wijsman, E.M., Vieland, V.J., Vicente, A.M., Schellenberg, G.D., Pericak-Vance, M., Paterson, A.D., Parr, J.R., Oliveira, G., Nurnberger, J.I., Monaco, A.P., Maestrini, E., Klauck, S.M., Hakonarson, H., Haines, J.L., Geschwind, D.H., Freitag, C.M., Folstein, S.E., Ennis, S., Coon, H., Battaglia, A., Szatmari, P., Sutcliffe, J.S., Hallmayer, J., Gill, M., Cook, E.H., Buxbaum, J.D., Devlin, B., Gallagher, L., Betancur, C., Scherer, S.W., 2014. Convergence of genes and cellular pathways dysregulated in autism spectrum disorders. American journal of human genetics 94, 677-694.

Pinto, D., Pagnamenta, A.T., Klei, L., Anney, R., Merico, D., Regan, R., Conroy, J., Magalhaes, T.R., Correia, C., Abrahams, B.S., Almeida, J., Bacchelli, E., Bader, G.D., Bailey, A.J., Baird, G., Battaglia, A., Berney, T., Bolshakova, N., Boelte, S., Bolton, P.F., Bourgeron, T., Brennan, S., Brian, J., Bryson, S.E., Carson, A.R., Casallo, G., Casey, J., Chung, B.H.Y., Cochrane, L., Corsello, C., Crawford, E.L., Crossett, A., Cytrynbaum, C., Dawson, G., de Jonge, M., Delorme, R., Drmic, I., Duketis, E., Duque, F., Estes, A., Farrar, P., Fernandez, B.A., Folstein, S.E., Fombonne, E., Freitag, C.M., Gilbert, J., Gillberg, C., Glessner, J.T., Goldberg, J., Green, A., Green, J., Guter, S.J., Hakonarson, H., Heron, E.A., Hill, M., Holt, R., Howe, J.L., Hughes, G., Hus, V., Igliozzi, R., Kim, C., Klauck, S.M., Kolevzon, A., Korvatska, O., Kustanovich, V., Lajonchere, C.M., Lamb, J.A., Laskawiec, M., Leboyer, M., Le Couteur, A., Leventhal, B.L., Lionel, A.C., Liu, X.-Q., Lord, C., Lotspeich, L., Lund, S.C., Maestrini, E., Mahoney, W., Mantoulan, C., Marshall, C.R., McConachie, H., McDougle, C.J., McGrath, J.,

bioRxiv preprint doi: https://doi.org/10.1101/042218; this version posted March 6, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

McMahon, W.M., Merikangas, A., Migita, O., Minshew, N.J., Mirza, G.K., Munson, J., Nelson, S.F., Noakes, C., Noor, A., Nygren, G., Oliveira, G., Papanikolaou, K., Parr, J.R., Parrini, B., Paton, T., Pickles, A., Pilorge, M., Piven, J., Ponting, C.P., Posey, D.J., Poustka, A., Poustka, F., Prasad, A., Ragoussis, J., Renshaw, K., Rickaby, J., Roberts, W., Roeder, K., Roge, B., Rutter, M.L., Bierut, L.J., Rice, J.P., Salt, J., Sansom, K., Sato, D., Segurado, R., Sequeira, A.F., Senman, L., Shah, N., Sheffield, V.C., Soorya, L., Sousa, I., Stein, O., Sykes, N., Stoppioni, V., Strawbridge, C., Tancredi, R., Tansey, K., Thiruvahindrapduram, B., Thompson, A.P., Thomson, S., Tryfon, A., Tsiantis, J., Van Engeland, H., Vincent, J.B., Volkmar, F., Wallace, S., Wang, K., Wang, Z., Wassink, T.H., Webber, C., Weksberg, R., Wing, K., Wittemeyer, K., Wood, S., Wu, J., Yaspan, B.L., Zurawiecki, D., Zwaigenbaum, L., Buxbaum, J.D., Cantor, R.M., Cook, E.H., Coon, H., Cuccaro, M.L., Devlin, B., Ennis, S., Gallagher, L., Geschwind, D.H., Gill, M., Haines, J.L., Hallmayer, J., Miller, J., Monaco, A.P., Nurnberger, J.I., Jr., Paterson, A.D., Pericak-Vance, M.A., Schellenberg, G.D., Szatmari, P., Vicente, A.M., Vieland, V.J., Wijsman, E.M., Scherer, S.W., Sutcliffe, J.S., Betancur, C., 2010. Functional impact of global rare copy number variation in autism spectrum disorders. Nature 466, 368-372.

Poultney, C.S., Samocha, K., Kou, Y., Liu, L., Walker, S., Singh, T., Klei, L., Kosmicki, J., Biscaldi, M., Bolton, P.F., Brownfeld, J.M., Cai, J., Campbell, N.G., Crawford, E.L., Crooks, L., Curran, S.R., Dawson, G., Duketis, E., Fernandez, B.A., Gallagher, L., Geller, E., Guter, S.J., Jimenez Gonzalez, P., Kilpinen, H., Klauck, S.M., Kolevzon, A., Lee, I., Lei, J., Lehtimaki, T., Lin, C.F., Ma'ayan, A., Marshall, C.R., McInnes, A.L., Neale, B., Owen, M.J., Ozaki, N., Parellada, M., Parr, J.R., Purcell, S., Puura, K., Rajagopalan, D., Rehnstrom, K., Reichenberg, A., Sabo, A., Sachse, M., Sanders, S.J., Schafer, C., Schulte-Ruther, M., Skuse, D., Stevens, C., Szatmari, P., Tammimies, K., Valladares, O., Voran, A., Wang, L.S., Weiss,

bioRxiv preprint doi: https://doi.org/10.1101/042218; this version posted March 6, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

L.A., Jeremy Willsey, A., Yu, T.W., Yuen, R.K., Study, D.D.D., Homozygosity Mapping Collaborative for, A., Consortium, U.K., Autism Sequencing, C., Lehner, T., Palotie, A., Schellenberg, G.D., Sklar, P., State, M.W., Sutcliffe, J.S., Walsh, C.A., Scherer, S.W., Zwick, M.E., Roeder, K., 2014. Synaptic, transcriptional and chromatin genes disrupted in autism. Nature 515, 209-215.

Prasad, T.S., Kandasamy, K., Pandey, A., 2009. Human Protein Reference Database and Human Proteinpedia as discovery tools for systems biology. Methods in molecular biology 577, 67-79.

Ronemus, M., Iossifov, I., Levy, D., Wigler, M., 2014. The role of de novo mutations in the genetics of autism spectrum disorders. Nat Rev Genet 15, 133-141.

Rossin, E.J., Lage, K., Raychaudhuri, S., Xavier, R.J., Tatar, D., Benita, Y., International Inflammatory Bowel Disease Genetics, C., Cotsapas, C., Daly, M.J., 2011. Proteins encoded in genomic regions associated with immune-mediated disease physically interact and suggest underlying biology. PLoS genetics 7, e1001273.

Sakai, Y., Shaw, C.A., Dawson, B.C., Dugas, D.V., Al-Mohtaseb, Z., Hill, D.E., Zoghbi, H.Y., 2011. Protein Interactome Reveals Converging Molecular Pathways Among Autism Disorders. Science translational medicine 3.

Sanders, S.J., Ercan-Sencicek, A.G., Hus, V., Luo, R., Murtha, M.T., Moreno-De-Luca, D., Chu, S.H., Moreau, M.P., Gupta, A.R., Thomson, S.A., Mason, C.E., Bilguvar, K., Celestino-Soper, P.B., Choi, M., Crawford, E.L., Davis, L., Wright, N.R., Dhodapkar, R.M., DiCola, M., DiLullo, N.M., Fernandez, T.V., Fielding-Singh, V., Fishman, D.O., Frahm, S., Garagaloyan, R., Goh, G.S., Kammela, S., Klei, L., Lowe, J.K., Lund, S.C., McGrew, A.D., Meyer, K.A., Moffat, W.J., Murdoch, J.D., O'Roak, B.J., Ober, G.T., Pottenger, R.S., Raubeson, M.J., Song, Y., Wang, Q., Yaspan, B.L., Yu, T.W., Yurkiewicz, I.R., Beaudet, A.L., Cantor, R.M., Curland, M., Grice,

bioRxiv preprint doi: https://doi.org/10.1101/042218; this version posted March 6, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

D.E., Gunel, M., Lifton, R.P., Mane, S.M., Martin, D.M., Shaw, C.A., Sheldon, M., Tischfield, J.A., Walsh, C.A., Morrow, E.M., Ledbetter, D.H., Fombonne, E., Lord, C., Martin, C.L., Brooks, A.I., Sutcliffe, J.S., Cook, E.H., Jr., Geschwind, D., Roeder, K., Devlin, B., State, M.W., 2011. Multiple recurrent de novo CNVs, including duplications of the 7q11.23 Williams syndrome region, are strongly associated with autism. Neuron 70, 863-885.

Sanders, S.J., Murtha, M.T., Gupta, A.R., Murdoch, J.D., Raubeson, M.J., Willsey, A.J., Ercan-Sencicek, A.G., DiLullo, N.M., Parikshak, N.N., Stein, J.L., Walker, M.F., Ober, G.T., Teran, N.A., Song, Y., El-Fishawy, P., Murtha, R.C., Choi, M., Overton, J.D., Bjornson, R.D., Carriero, N.J., Meyer, K.A., Bilguvar, K., Mane, S.M., Sestan, N., Lifton, R.P., Guenel, M., Roeder, K., Geschwind, D.H., Devlin, B., State, M.W., 2012. De novo mutations revealed by whole-exome sequencing are strongly associated with autism. Nature 485, 237-U124.

Schellenberger, J., Park, J.O., Conrad, T.M., Palsson, B.O., 2010. BiGG: a Biochemical Genetic and Genomic knowledgebase of large scale metabolic reconstructions. BMC bioinformatics 11, 213.

Sebat, J., Lakshmi, B., Malhotra, D., Troge, J., Lese-Martin, C., Walsh, T., Yamrom, B., Yoon, S., Krasnitz, A., Kendall, J., Leotta, A., Pai, D., Zhang, R., Lee, Y.-H., Hicks, J., Spence, S.J., Lee, A.T., Puura, K., Lehtimaeki, T., Ledbetter, D., Gregersen, P.K., Bregman, J., Sutcliffe, J.S., Jobanputra, V., Chung, W., Warburton, D., King, M.-C., Skuse, D., Geschwind, D.H., Gilliam, T.C., Ye, K., Wigler, M., 2007. Strong association of de novo copy number mutations with autism. Science 316, 445-449.

Stark, C., Breitkreutz, B.J., Reguly, T., Boucher, L., Breitkreutz, A., Tyers, M., 2006. BioGRID: a general repository for interaction datasets. Nucleic acids research 34, D535-539.

bioRxiv preprint doi: https://doi.org/10.1101/042218; this version posted March 6, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

State, M.W., Levitt, P., 2011. The conundrums of understanding genetic risks for autism spectrum disorders. Nature neuroscience 14, 1499-1506.

Stein, J.L., Parikshak, N.N., Geschwind, D.H., 2013. Rare inherited variation in autism: beginning to see the forest and a few trees. Neuron 77, 209-211.

Stessman, H.A., Bernier, R., Eichler, E.E., 2014. A genotype-first approach to defining the subtypes of a complex disease. Cell 156, 872-877.

Sullivan, P.F., Posthuma, D., 2015. Biological pathways and networks implicated in psychiatric disorders. Current Opinion in Behavioral Sciences 2, 58-68.

Szatmari, P., Paterson, A.D., Zwaigenbaum, L., Roberts, W., Brian, J., Liu, X.-Q., Vincent, J.B., Skaug, J.L., Thompson, A.P., Senman, L., Feuk, L., Qian, C., Bryson, S.E., Jones, M.B., Marshall, C.R., Scherer, S.W., Vieland, V.J., Bartlett, C., Mangin, L.V., Goedken, R., Segre, A., Pericak-Vance, M.A., Cuccaro, M.L., Gilbert, J.R., Wright, H.H., Abramson, R.K., Betancur, C., Bourgeron, T., Gillberg, C., Leboyer, M., Buxbaum, J.D., Davis, K.L., Hollander, E., Silverman, J.M., Hallmayer, J., Lotspeich, L., Sutcliffe, J.S., Haines, J.L., Folstein, S.E., Piven, J., Wassink, T.H., Sheffield, V., Geschwind, D.H., Bucan, M., Brown, W.T., Cantor, R.M., Constantino, J.N., Gilliam, T.C., Herbert, M., LaJonchere, C., Ledbetter, D.H., Lese-Martin, C., Miller, J., Nelson, S., Samango-Sprouse, C.A., Spence, S., State, M., Tanzi, R.E., Coon, H., Dawson, G., Devlin, B., Estes, A., Flodman, P., Klei, L., McMahon, W.M., Minshew, N., Munson, J., Korvatska, E., Rodier, P.M., Schellenberg, G.D., Smith, M., Spence, M.A., Stodgell, C., Tepper, P.G., Wijsman, E.M., Yu, C.-E., Roge, B., Mantoulan, C., Wittemeyer, K., Poustka, A., Felder, B., Klauck, S.M., Schuster, C., Poustka, F., Boelte, S., Feineis-Matthews, S., Herbrecht, E., Schmoetzer, G., Tsiantis, J., Papanikolaou, K., Maestrini, E., Bacchelli, E., Blasi, F., Carone, S., Toma, C., Van Engeland, H., de Jonge, M., Kemner, C., Koop, F., Langemeijer, M., Hijimans, C., Staal, W.G., Baird, G., Bolton, P.F., Rutter, M.L., Weisblatt,

bioRxiv preprint doi: https://doi.org/10.1101/042218; this version posted March 6, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

E., Green, J., Aldred, C., Wilkinson, J.-A., Pickles, A., Le Couteur, A., Berney, T., McConachie, H., Bailey, A.J., Francis, K., Honeyman, G., Hutchinson, A., Parr, J.R., Wallace, S., Monaco, A.P., Barnby, G., Kobayashi, K., Lamb, J.A., Sousa, I., Sykes, N., Cook, E.H., Guter, S.J., Leventhal, B.L., Salt, J., Lord, C., Corsello, C., Hus, V., Weeks, D.E., Volkmar, F., Tauber, M., Fombonne, E., Shih, A., 2007. Mapping autism risk loci using genetic linkage and chromosomal rearrangements. Nature genetics 39, 319-328.

Szklarczyk, D., Franceschini, A., Kuhn, M., Simonovic, M., Roth, A., Minguez, P., Doerks, T., Stark, M., Muller, J., Bork, P., Jensen, L.J., von Mering, C., 2011. The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. Nucleic acids research 39, D561-568.

Taylor, L.J., Maybery, M.T., Wray, J., Ravine, D., Hunt, A., Whitehouse, A.J., 2013. Brief Report: Do the Nature of Communication Impairments in Autism Spectrum Disorders Relate to the Broader Autism Phenotype in Parents? Journal of autism and developmental disorders.

Toma, C., Torrico, B., Hervas, A., Valdes-Mas, R., Tristan-Noguero, A., Padillo, V., Maristany, M., Salgado, M., Arenas, C., Puente, X.S., Bayes, M., Cormand, B., 2013. Exome sequencing in multiplex autism families suggests a major role for heterozygous truncating mutations. Molecular psychiatry.

Toro, R., Konyukh, M., Delorme, R., Leblond, C., Chaste, P., Fauchereau, F., Coleman, M., Leboyer, M., Gillberg, C., Bourgeron, T., 2010. Key role for gene dosage and synaptic homeostasis in autism spectrum disorders. Trends in Genetics 26, 363-372.

Tsai, P.T., Hull, C., Chu, Y., Greene-Colozzi, E., Sadowski, A.R., Leech, J.M., Steinberg, J., Crawley, J.N., Regehr, W.G., Sahin, M., 2012. Autistic-like behaviour and cerebellar dysfunction in Purkinje cell Tsc1 mutant mice. Nature 488, 647-651.

bioRxiv preprint doi: https://doi.org/10.1101/042218; this version posted March 6, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Turner, T.N., Sharma, K., Oh, E.C., Liu, Y.P., Collins, R.L., Sosa, M.X., Auer, D.R., Brand, H., Sanders, S.J., Moreno-De-Luca, D., Pihur, V., Plona, T., Pike, K., Soppet, D.R., Smith, M.W., Cheung, S.W., Martin, C.L., State, M.W., Talkowski, M.E., Cook, E., Huganir, R., Katsanis, N., Chakravarti, A., 2015. Loss of delta-catenin function in severe autism. Nature 520, 51-56.

Uddin, M., Tammimies, K., Pellecchia, G., Alipanahi, B., Hu, P., Wang, Z., Pinto, D., Lau, L., Nalpathamkalam, T., Marshall, C.R., Blencowe, B.J., Frey, B.J., Merico, D., Yuen, R.K., Scherer, S.W., 2014. Brain-expressed exons under purifying selection are enriched for de novo mutations in autism spectrum disorder. Nature genetics 46, 742-747.

Voineagu, I., Wang, X., Johnston, P., Lowe, J.K., Tian, Y., Horvath, S., Mill, J., Cantor, R.M., Blencowe, B.J., Geschwind, D.H., 2011. Transcriptomic analysis of autistic brain reveals convergent molecular pathology. Nature 474, 380-384.

Wang, S.S., Kloth, A.D., Badura, A., 2014. The Cerebellum, Sensitive Periods, and Autism. Neuron 83, 518-532.

Willsey, A.J., Sanders, S.J., Li, M., Dong, S., Tebbenkamp, A.T., Muhle, R.A., Reilly, S.K., Lin, L., Fertuzinhos, S., Miller, J.A., Murtha, M.T., Bichsel, C., Niu, W., Cotney, J., Ercan-Sencicek, A.G., Gockley, J., Gupta, A.R., Han, W., He, X., Hoffman, E.J., Klei, L., Lei, J., Liu, W., Liu, L., Lu, C., Xu, X., Zhu, Y., Mane, S.M., Lein, E.S., Wei, L., Noonan, J.P., Roeder, K., Devlin, B., Sestan, N., State, M.W., 2013. Coexpression networks implicate human midfetal deep cortical projection neurons in the pathogenesis of autism. Cell 155, 997-1007.

Won, H., Lee, H.R., Gee, H.Y., Mah, W., Kim, J.I., Lee, J., Ha, S., Chung, C., Jung, E.S., Cho, Y.S., Park, S.G., Lee, J.S., Lee, K., Kim, D., Bae, Y.C., Kaang, B.K., Lee, M.G., Kim, E., 2012. Autistic-like social behaviour in Shank2-mutant mice improved by restoring NMDA receptor function. Nature 486, 261-265.

bioRxiv preprint doi: https://doi.org/10.1101/042218; this version posted March 6, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Xenarios, I., Salwinski, L., Duan, X.J., Higney, P., Kim, S.M., Eisenberg, D., 2002. DIP, the Database of Interacting Proteins: a research tool for studying cellular networks of protein interactions. Nucleic acids research 30, 303-305.

Xia, K., Dong, D., Han, J.D., 2006. IntNetDB v1.0: an integrated protein-protein interaction network database generated by a probabilistic model. BMC bioinformatics 7, 508.

Yu, T.W., Chahrour, M.H., Coulter, M.E., Jiralerspong, S., Okamura-Ikeda, K., Ataman, B., Schmitz-Abe, K., Harmin, D.A., Adli, M., Malik, A.N., D'Gama, A.M., Lim, E.T., Sanders, S.J., Mochida, G.H., Partlow, J.N., Sunu, C.M., Felie, J.M., Rodriguez, J., Nasir, R.H., Ware, J., Joseph, R.M., Hill, R.S., Kwan, B.Y., Al-Saffar, M., Mukaddes, N.M., Hashmi, A., Balkhy, S., Gascon, G.G., Hisama, F.M., LeClair, E., Poduri, A., Oner, O., Al-Saad, S., Al-Awadi, S.A., Bastaki, L., Ben-Omran, T., Teebi, A.S., Al-Gazali, L., Eapen, V., Stevens, C.R., Rappaport, L., Gabriel, S.B., Markianos, K., State, M.W., Greenberg, M.E., Taniguchi, H., Braverman, N.E., Morrow, E.M., Walsh, C.A., 2013. Using whole-exome sequencing to identify inherited causes of autism. Neuron 77, 259-273.

Zanzoni, A., Montecchi-Palazzi, L., Quondam, M., Ausiello, G., Helmer-Citterich, M., Cesareni, G., 2002. MINT: a Molecular INTeraction database. FEBS letters 513, 135-140.