

## ORIGINAL ARTICLE

## Neurodevelopmental and neuropsychiatric disorders represent an interconnected molecular system

AS Cristino<sup>1</sup>, SM Williams<sup>1</sup>, Z Hawi<sup>1,2</sup>, J-Y An<sup>1</sup>, MA Bellgrove<sup>1,2</sup>, CE Schwartz<sup>3</sup>, L da F Costa<sup>4</sup> and C Claudianos<sup>1</sup>

Many putative genetic factors that confer risk to neurodevelopmental disorders such as autism spectrum disorders (ASDs) and X-linked intellectual disability (XLID), and to neuropsychiatric disorders including attention deficit hyperactivity disorder (ADHD) and schizophrenia (SZ) have been identified in individuals from diverse human populations. Although there is significant aetiological heterogeneity within and between these conditions, recent data show that genetic factors contribute to their comorbidity. Many studies have identified candidate gene associations for these mental health disorders, albeit this is often done in a piecemeal fashion with little regard to the inherent molecular complexity. Here, we sought to abstract relationships from our knowledge of systems level biology to help understand the unique and common genetic drivers of these conditions. We undertook a global and systematic approach to build and integrate available data in gene networks associated with ASDs, XLID, ADHD and SZ. Complex network concepts and computational methods were used to investigate whether candidate genes associated with these conditions were related through mechanisms of gene regulation, functional protein–protein interactions, transcription factor (TF) and microRNA (miRNA) binding sites. Although our analyses show that genetic variations associated with the four disorders can occur in the same molecular pathways and functional domains, including synaptic transmission, there are patterns of variation that define significant differences between disorders. Of particular interest is DNA variations located in intergenic regions that comprise regulatory sites for TFs or miRNA. Our approach provides a hypothetical framework, which will help discovery and analysis of candidate genes associated with neurodevelopmental and neuropsychiatric disorders.

*Molecular Psychiatry* (2014) **19**, 294–301; doi:10.1038/mp.2013.16; published online 26 February 2013

**Keywords:** neurodevelopmental disorder; neuropsychiatric disorder; mental health disorder; gene networks; systems biology

## INTRODUCTION

Neurodevelopmental and neuropsychiatric disorders are complex traits that result from multiple genetic determinants interacting in the context of poorly understood environmental factors to give rise to clinically diverse phenotypes.<sup>1–3</sup> Affected individuals harbour different risk alleles in a heterogeneous genetic background that make candidate disorder genes difficult to detect. In spite of this difficulty there are now hundreds of candidate genes with DNA copy number variations or single nucleotide polymorphisms (SNPs) characterised from clinically diagnosed individuals. The power of modern genetic screening approaches has led to the identification of candidate genes associated with autism spectrum disorders (ASDs),<sup>4–8</sup> X-linked intellectual disability (XLID),<sup>9,10</sup> attention deficit hyperactivity disorder (ADHD)<sup>11–13</sup> and schizophrenia (SZ).<sup>14–20</sup> More recently, whole-genome sequencing (exome capture) coupled with computational approaches that integrate protein interaction information have helped to build hypotheses concerning molecular pathways and processes that are likely to underpin these disorders.<sup>6,7,21</sup>

ASDs are a group of clinically diverse neurodevelopmental disorders (1–2% of the population) with a significant genetic heterogeneity.<sup>7,8,22–24</sup> Although large twin studies have shown monozygotic concordance rates with high heritability estimates of ~90%,<sup>25</sup> the underlying genetic determinants remain largely unknown. Only three genetic loci (5p14.113, 5p15.214

and 7q31–q35) have statistically significant support with ASDs, suggesting common variation will account for only a small proportion of the heritability in ASD.<sup>7,26</sup> In contrast, recent copy number variation and SNP data show many rare variants occur in key neurological molecules that function in the synaptic junctions of neurons. These molecules include members of the neuroligin-neurexin complex (NRXN1, CNTNAP2, NLGN1, NLGN3, NLGN4X, NLGN4Y, LRRTM1 and LRRTM2<sup>27,28</sup> and interacting proteins (SHANK3, PSD95 SHANK2, SHANK1, SYNGAP1, DLGAP2, FOXP1, GRIN2B, SCN1A and LAMC3<sup>7,22</sup>), suggesting that synapse development and function represents a major pathogenetic hub for ASDs and related disorders.<sup>27</sup>

XLID accounts for 5–10% of intellectual disability in males. There are over 150 XLID syndromes, including fragile X syndrome and Rett syndrome and numerous non-syndromal XLID disorders many, of which are caused by 102 genes on the X chromosome.<sup>29</sup> Among these are genes that also contribute to ASD (that is, *NLGN4*, *RPL10* and *RAB39*). Although familial linkage studies have facilitated a better understanding of the biological basis of XLID, the genetic basis for autosomal intellectual disability remains poorly defined.<sup>10</sup>

ADHD is a neuropsychiatric condition of childhood. Although underlying molecular mechanisms are poorly understood, genetic influences are recognised as important aetiological components of ADHD. Large twin studies show heritability estimates of 75–90%<sup>30</sup> for this disorder. A polygenic model is consistent with

<sup>1</sup>Queensland Brain Institute, The University of Queensland, Brisbane, QLD, Australia; <sup>2</sup>School of Psychology and Psychiatry, Monash University, Melbourne, VIC, Australia; <sup>3</sup>Greenwood Genetic Center, Greenwood, SC, USA and <sup>4</sup>Instituto de Física de São Carlos, Universidade de São Paulo, São Paulo, Brazil. Correspondence: Professor C Claudianos or AS Cristino, Queensland Brain Institute, The University of Queensland, QBI Building 79, Upland Rd, Brisbane, QLD 4072, Australia.

E-mail: c.claudianos@uq.edu.au or a.cristino@uq.edu.au

Received 4 September 2012; revised 14 December 2012; accepted 2 January 2013; published online 26 February 2013

the high prevalence of ADHD (2–10%) and high concordance in monozygotic twins, but modest risk to first-degree relatives. Pathogenetic models of ADHD have traditionally focused on molecules involved in neurotransmission and catecholamine synaptic dysfunction,<sup>31,32</sup> including dopamine transporter *DAT1* (*SLC6A3*), dopamine receptors *DRD4*, *DRD5* and synaptosomal protein SNAP-25.<sup>11</sup> More recently neural developmental genes including cadherin 13 (*CDH13*) and cGMP-dependent protein kinase I (*PRKG1*)<sup>33–36</sup> have been associated with ADHD.

SZ is similarly a highly heritability disorder (~80%) with monozygotic twin concordance rates estimated to be as high as 40–65%.<sup>37,38</sup> Similar to ASD and ADHD, there have been significant efforts to identify possible common and rare genetic variants that might explain susceptibility to this disorder. Recent genome-wide association studies<sup>39</sup> have confirmed a substantial polygenic component associated with SZ. Not unlike ASD and ADHD, SZ is likely to encompass a broad genetic aetiology. Increasing evidence suggests the onset of neurological symptoms probably occur when a threshold of cumulative genetic liability is reached.<sup>40</sup> A compelling feature concerning SZ is the number of genetic factors that are shared with other disorders. DNA variants identified in SZ have also been associated with bipolar disorder,<sup>41</sup> ASD,<sup>23,42</sup> mental retardation,<sup>43,44</sup> and ADHD.<sup>12</sup> A notable example of genetic comorbidity is the documented structural variation in the neurexin-1 (*NRXN1*) gene that increases risk for both ASD and SZ.<sup>45</sup>

Here, we propose a novel biological systems approach, a hypothetical 'gene network model' that can be used to analyse candidate genes and predict the association of genetic screening data with ASD, XLID, ADHD and SZ. This is based on observed differences in gene distributions and functional patterns that are informative for each disorder. The gene network model was

successfully validated using cohort data from six recent disorder studies.

## MATERIALS AND METHODS

A diagram of our computational approach is depicted in Supplementary Figure S1. Detailed description of genes, biological databases and computational and statistical methods used in this study are presented in the Supplementary Information.

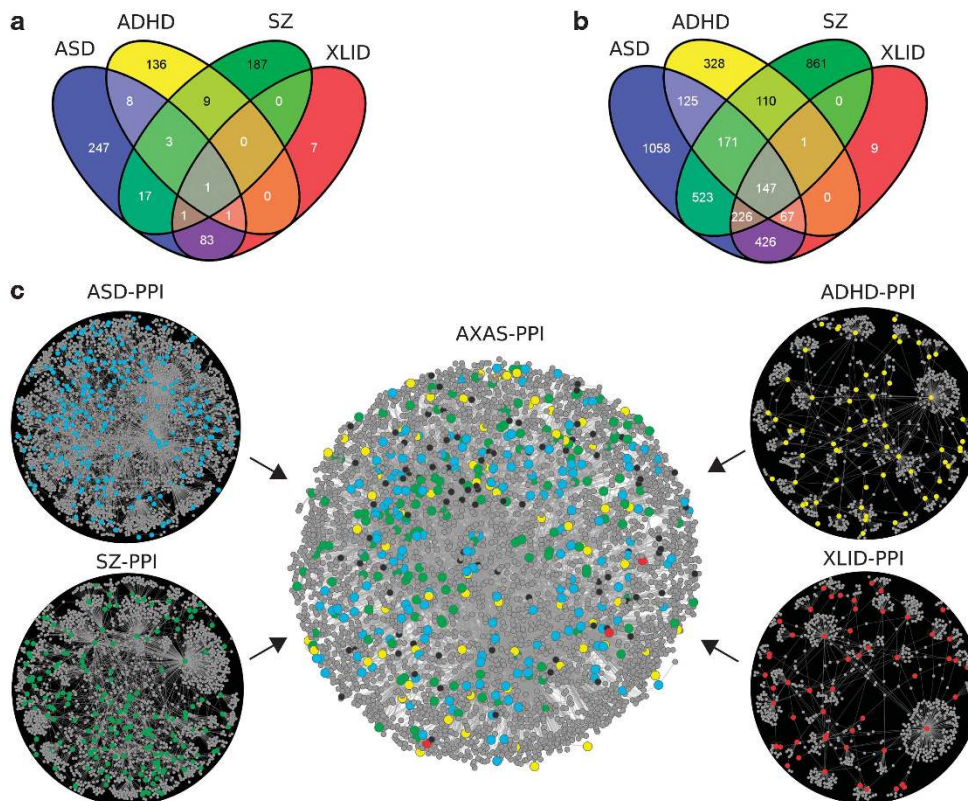
## RESULTS

### Primary database of neurodevelopmental and neuropsychiatric disorder genes

Our approach used current information concerning genes that have a documented association with neurodevelopmental and neuropsychiatric disorders. We used publicly available molecular data to create a comprehensive database of primary candidate genes associated with ASD, XLID, ADHD and SZ. For candidate genes to be included in the analysis, we required evidence of DNA variation including SNPs, insertion and deletions and larger copy number variations. A primary database of 700 genes comprising 361 genes for ASD,<sup>4,7</sup> 93 for XLID,<sup>9,10</sup> 158 for ADHD<sup>11,12,33–36</sup> and 218 for SZ<sup>14–20</sup> provided *a priori* information for our analyses (Supplementary Table S1). There were 125 genes found to be associated with more than one disorder (Figure 1a).

### Protein–protein interaction networks of neurodevelopmental and neuropsychiatric disorders

A computational analysis was designed to build a functional map of the neurodevelopmental and neuropsychiatric disorder genes based on the fundamental features of protein–protein interaction



**Figure 1.** (a) Venn diagram of primary candidate genes associated with four disorders: blue (ASD), green (SZ), yellow (ADHD) and red (XLID). (b) Venn diagram of primary candidate genes and their first degree interacting neighbours in the PPI network. (c) PPI networks of all four disorders showing primary candidate genes and their adjacent neighbours. The AXAS-PPI network is the union of all four disorder PPI networks. Those genes that overlapped more than one disorder are marked as black circles (Supplementary Table S1 and S2).

(PPI) and gene regulation (that is, identifying transcription factor (TF) and microRNA (miRNA) regulatory sites). First, a database of nonredundant PPIs was created by joining the BioGRID (biological general repository for interaction data sets)<sup>46</sup> and HPRD (human protein reference database) databases<sup>47–49</sup> (Supplementary Table S2).

The BioGRID database (version 3.1.77) includes 9115 proteins and 37 748 interactions between proteins, and the HPRD (release 9) comprises 9563 proteins and 38 774 interactions. The union of the two major PPI networks yields a nonredundant network with 11 028 proteins and 58 256 interactions. This indicates that more than 80% of these proteins and 65% of the interactions are commonly found in both databases (Supplementary Table S2).

The PPI network is represented as a graph where the nodes are proteins and the edges are interactions between proteins. PPI networks for ASD, XLID, ADHD and SZ were created by retrieving all possible interactions from the PPI database between primary candidate genes and their respective first-order interactors (Table 1). A total of 534 out of 700 (76%) encoded primary candidate genes were found in the merged PPI database. We built a neurodevelopmental and neuropsychiatric disorder PPI network named AXAS to reflect the precise origins of the data (ASD, XLID, ADHD and SZ). The AXAS–PPI network was created by joining these 534 encoded proteins with 3413 first-order interacting proteins. Many proteins of the AXAS–PPI network were found to interact with primary candidate genes from more than one disorder (Figure 1b).

The AXAS–PPI network reconstructed from 534 candidate genes represents 35% (3946 out of 11 028; see Table 1 and Figure 1c) of the merged BioGRID–HPRD PPI network. Strikingly, the AXAS–PPI network encompasses ~20% of encoded genes found in the human genome. We also validated the neurological context of the AXAS–PPI network using available transcriptome data for the whole brain.<sup>50</sup> A nucleotide sequence comparison of 3946 protein-coding genes found that 92% of genes in the AXAS–PPI network are expressed in the human brain (Supplementary Table S3).

All four individual disorder PPI networks, as well as the entire AXAS–PPI network were characterised using six structural properties of complex networks (Table 1). We further tested whether the distributions of the structural properties were similar or different between AXAS–PPI networks and PPI networks constructed from an equivalent number of randomly sampled proteins from the merged PPI database (Table 1 and Supplementary Figure S2). Kolmogorov–Smirnov tests showed that the structural properties of average degree, density and clustering coefficient were significantly higher in the AXAS–PPI networks than in the random

networks, while the average path length was smaller in the disorder PPI networks (Table 1). This control analysis confirms that genes previously associated with neurodevelopmental and neuropsychiatric disorders are significantly more interconnected than expected by chance. A functional enrichment analysis of the 534 primary candidate genes revealed that specific biological processes are overrepresented in each of these disorders. This higher-order functional pattern provides a basis to discriminate between these polygenic disorders (Table 2 and Supplementary Table S4). There is no single functional domain that clearly characterises any one disorder.

Another common property of biological networks is community structure.<sup>51,52</sup> We found that the AXAS–PPI network is highly modular ( $Q = 0.61$ ), being subdivided into 30 structural communities (modules). The AXAS–PPI modules were ranked in order of the number of proteins that contribute to each module; M1 is the largest with 700 proteins, whereas M30 is the smallest with only three proteins. Most of the AXAS–PPI modules (21 out of 30) also show a significant enrichment of proteins that are functionally clustered (Supplementary Table S5). To illustrate these features, the AXAS–PPI network was summarised as a simple graph showing the contribution of modules to the four disorders (Figure 2). Interestingly M2 (synaptic transmission, signal transduction and phosphorylation) interacts with 12 other modules and is the main protein hub or functional focus of the network. M2 significantly integrates key biological processes, including the regulation of transcription and biosynthetic processes (M1), cell cycle (M3), cell–cell adhesion and development (M4) and cell–cell communication and differentiation and protein transport (M5) (Supplementary Table S5). Although primary candidate genes associated with ASD, XLID, ADHD and SZ contribute to most of the AXAS–PPI modules; they are not equally distributed in 9 out of 13 modules (Supplementary Table S6).

#### cis-Regulatory networks of neurodevelopmental and neuropsychiatric disorders

We analysed the upstream control region (genomic DNA containing cis-regulatory elements) and downstream untranslated regions (3′-UTR) of putative mRNA transcripts in the 30 AXAS–PPI modules. The aim was to test the hypothesis that genes in the same functional module or disorder share a common pattern of regulatory elements. In addition to identifying candidate genes that are TFs or containing miRNAs, we characterised the number and distribution of TF binding and miRNA target sites associated with genes of the AXAS–PPI network. This upstream control region analysis identified overrepresented DNA motifs in the upstream

**Table 1.** Structural properties of PPI networks

Disorder	N	n (% of N)	Number of components	Number of nodes (main network)	Number of edges (main network)	Average degree (D; P-value <sup>a</sup> )	Clustering coefficient (D; P-value <sup>a</sup> )	Average path length (D; P-value <sup>a</sup> )	Density (D; P-value <sup>a</sup> )
ASD	361	281 (77%)	39	2743 (2647)	4562 (4496)	3.2 0.51; P < 0.01	0.087 0.15; P < 0.01	4.3 0.60; P < 0.01	0.0013 0.29; P < 0.01
ADHD	158	92 (58%)	28	949 (819)	1050 (942)	2.1 0.60; P < 0.01	0.017 0.20; P < 0.01	5.5 0.12; P < 0.01	0.0028 0.99; P < 0.01
SZ	218	189 (86%)	18	2039 (1981)	2830 (2784)	2.7 0.44; P < 0.01	0.045 0.16; P < 0.01	4.5 0.39; P < 0.01	0.0014 0.31; P < 0.01
XLID	93	72 (77%)	19	876 (824)	945 (909)	2.1 0.16; P < 0.01	0.005 0.08; P < 0.05	5.3 0.24; P < 0.01	0.0027 0.46; P < 0.01
Total	700	534 (76%)	45	4052 (3946)	7738 (7662)	3.7 0.34; P < 0.01	0.078 0.12; P < 0.01	4.4 0.69; P < 0.01	0.0013 0.81; P < 0.01

Abbreviations: ASD, autism spectrum disorder; ADHD, attention deficit hyperactivity disorder; SZ, schizophrenia; XLID, X-linked intellectual disability. N = number of candidate genes. n = number of candidate genes that contribute to PPI networks. <sup>a</sup>Kolmogorov–Smirnov significance test compared disorder PPI networks with arbitrary PPI networks created from a random selection of genes from the merged BioGRID–HPRD database.

**Table 2.** Functional analysis of primary candidate disorder genes

Disorder	Number of genes	GO-biological process	GO ID	P-value		
ASD	281	Nervous system development	GO:0007399	2.9E – 17		
		Synaptic transmission	GO:0007268	1.7E – 05		
		Blood circulation	GO:0008015	4.5E – 05		
		Amino acid derivative metabolic process	GO:0006575	4.4E – 04		
		Learning or memory	GO:0007611	5.5E – 04		
		Catecholamine metabolic process	GO:0006584	6.4E – 04		
		Regulation of cell communication	GO:0010646	4.8E – 03		
		G-protein signalling, Coupled to cyclic nucleotide second messenger	GO:0007187	8.4E – 03		
		XLID	72	Nervous system development	GO:0007399	8.2E – 05
				ADHD	92	Synaptic transmission
Catecholamine metabolic process	GO:0006584	2.1E – 07				
G-protein signalling, Coupled to cyclic nucleotide second messenger	GO:0007187	3.4E – 06				
Regulation of synaptic transmission	GO:0050804	5.5E – 04				
Learning or memory	GO:0007611	2.0E – 03				
Positive regulation of cellular process	GO:0048522	6.2E – 03				
Cell migration	GO:0016477	9.2E – 03				
SZ	189	Synaptic transmission	GO:0007268			2.7E – 14
		Nervous system development	GO:0007399	3.3E – 11		
		Signal transduction	GO:0007165	7.0E – 10		
		Regulation of transport	GO:0051049	2.9E – 08		
		Positive regulation of cellular process	GO:0048522	4.1E – 08		
		Learning or memory	GO:0007611	6.8E – 08		
		Regulation of synaptic transmission	GO:0050804	8.1E – 08		
		Regulation of secretion	GO:0051046	1.4E – 07		
		Dopamine metabolic process	GO:0042417	1.4E – 06		
		Regulation of apoptosis	GO:0042981	1.1E – 05		
		Regulation of cell communication	GO:0010646	1.3E – 05		
		Glutamate signalling pathway	GO:0007215	9.3E – 04		
		Inflammatory response	GO:0006954	5.7E – 03		
Blood circulation	GO:0008015	8.2E – 03				

Abbreviations: ASD, autism spectrum disorder; ADHD, attention deficit hyperactivity disorder; SZ, schizophrenia; XLID, X-linked intellectual disability; GO, Gene Ontology.

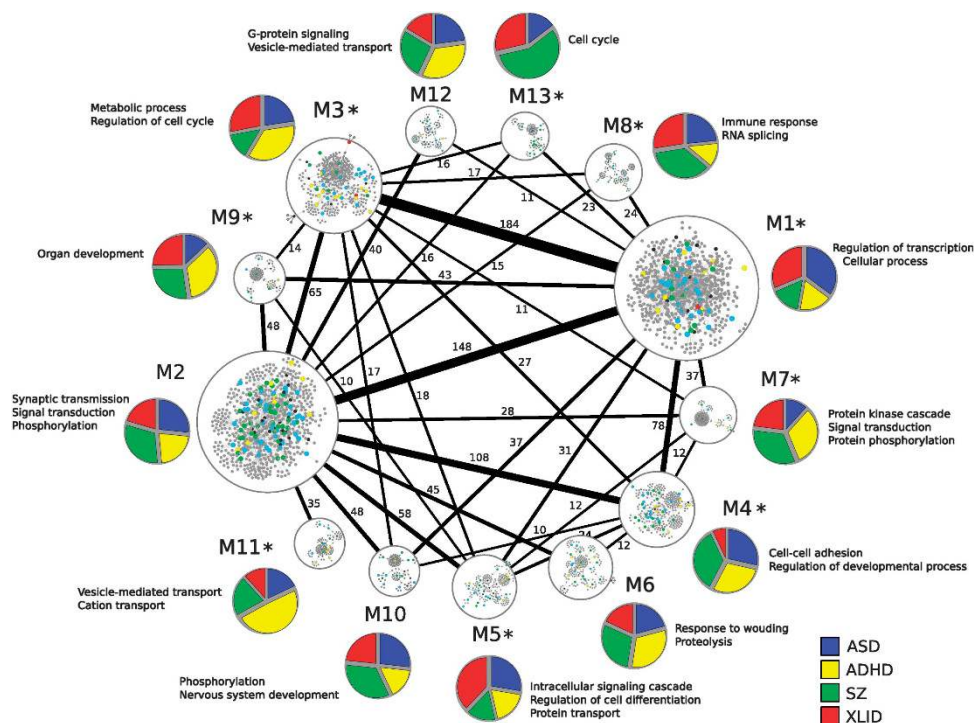
regions of genes that group in the same biological processes and disorder (Supplementary Figure S3). We used functional criteria defined by the Gene Ontology consortium to identify TF binding

sites in the 13 most populated AXAS–PPI modules (Figure 2 and Supplementary Table S5). A total of 401 functional groups were identified among the hierarchical levels in the Gene Ontology database (Supplementary Table S5), including 52 DNA motifs that were found to be enriched in these functional groups (Supplementary Table S7). Of the 52, 38 show similarity to TF binding sites previously described in the TRANSFAC<sup>53</sup> or JASPAR<sup>54</sup> databases. The remaining 14 are novel motifs that are likely to function as binding sites for unknown TFs or other regulatory molecules. Interestingly, more than 80% of the TFs that bind to these enriched DNA target sites are proteins found in the AXAS–PPI network (Supplementary Table S7). This indicates that TFs, which are part the AXAS–PPI network also regulate many of the genes in this network.

The 3′-UTR analysis was designed to identify enriched miRNA target sites of genes in the AXAS–PPI modules. We used available sequence data from the miRBase database<sup>55</sup>, which contains 1223 human miRNA sequences (miRBase version 16). We identified 621 miRNAs that have target sites enriched in the AXAS–PPI network, including 154 that were enriched in the 534 primary candidates genes (1.5-fold enriched;  $P$ -value < 0.01; Supplementary Table S8). A total of 683 miRNAs target a greater percentage of genes in the AXAS–PPI network than in the whole-genome (Supplementary Figure S4). We also found 3.9% (149) of the genes in the AXAS–PPI network contain miRNAs embedded within their predicted transcripts, and that this occurrence is nearly twice that observed for the whole-human genome (2.0%;  $P$ -value < 0.01). Furthermore, although the AXAS–PPI network only represents ~20% of encoded genes in the human genome, more than half of the known miRNAs (56%) target transcripts of this network. We also show that the number of DNA motifs enriched in the upstream control regions is positively correlated with the number of miRNA target sites in the 3′-UTRs (Pearson's correlation = 0.64;  $t = 4.67$ ;  $df = 32$ ;  $P$ -value < 0.01; Supplementary Figure S5). The correlation between increased transcriptional and post-transcriptional control suggests that the molecular network that underpins neurodevelopmental and neuropsychiatric disorders is likely to be evolving in a concerted manner.

SNP loci associated with neurodevelopmental and neuropsychiatric disorders are enriched with regulatory elements. On the basis of analysis of data from recent autism<sup>26,56</sup> and SZ genome-wide association studies<sup>41,57–61</sup> (Supplementary Table S9), we created a database comprising 4850 unique SNPs associated with ASD and SZ, and found that a number of these SNPs map to TF binding sites enriched in the AXAS–PPI network (Supplementary Table S7 and S9). The most overrepresented DNA motifs in the genome-wide association data sets were binding sites for STAT1 and STAT6 (signal transducer and activator of transcription) TF proteins. SNP loci associated with STAT1 and STAT6 binding sites were overrepresented in seven out of eight genome-wide association data sets (Supplementary Table S9). Our analyses also highlight that TF binding sites for SREBF1 (sterol regulatory element binding factor 1) and CREB (cAMP response element binding) are exclusively associated with autism and SZ, respectively.

There are 789 of 4839 nonredundant SNPs (16%) located near or within 435 genes in the AXAS–PPI network (Supplementary Table S10). Only 4% of 789 SNPs (32) occur within coding regions, whereas 81% (636) are located within intronic regions and include SNPs that map to alternate promoter regions of genes. The remaining 15% were located within 10 kb of the 5′-upstream (54) or 3′-downstream (67) regions of AXAS–PPI genes. We also compared SNP data with TF binding sites data validated by chromatin immunoprecipitation-sequencing at ENCODE (Encyclopedia of DNA Elements)<sup>62</sup>, and found that 113 SNP loci (14%) were associated with at least one TF binding site. Interestingly,



**Figure 2.** Graphical representation of the 13 most populated protein modules of the AXAS-PPI network. The size of each module is proportional to the number of proteins. Weighted lines with numbers represent protein interactions between modules. Distributions of primary candidate genes are shown as a pie chart for each module. The proteins and intraconnections of each module are described in Supplementary Table S5. Modules with significant differences in the frequency of primary candidate genes are indicated by \* (Supplementary Table S6).

we found 12 autism and 13 SZ SNPs associated with seven validated TF binding sites (STAT1, STAT3, YY1, SP1, ETS1, E2F1 and USF2) enriched in our AXAS-PPI network (Supplementary Table S11). Most of these SNPs are associated with genes involved in synaptic transmission, with multiple different SNPs often occurring in the same gene. These genes include *NTRK3* (neurotrophic tyrosine receptor kinase, type 3), *NRXN1*, *SLC25A12* (solute carrier family 25, mitochondrial carrier, Aralar, member 12), *GRID1* (glutamate receptor, ionotropic, delta 1) and *SDC2* (syndecan 2). Approximately 50% of the 789 SNPs map to three functional modules: M2 (25%, synaptic transmission), M1 (14%, regulation of transcription), and M4 (11%, cell-cell adhesion) (Supplementary Figure S5A). Our approach offers the potential to characterise SNPs according to molecular function, with an example being shown for *NTRK3* and *NRXN1* (Supplementary Figure S5B and S5C). These functions only become apparent when integrating information concerning gene organisation, regulatory mechanisms and PPI.

#### Cross-validation of the AXAS-PPI network

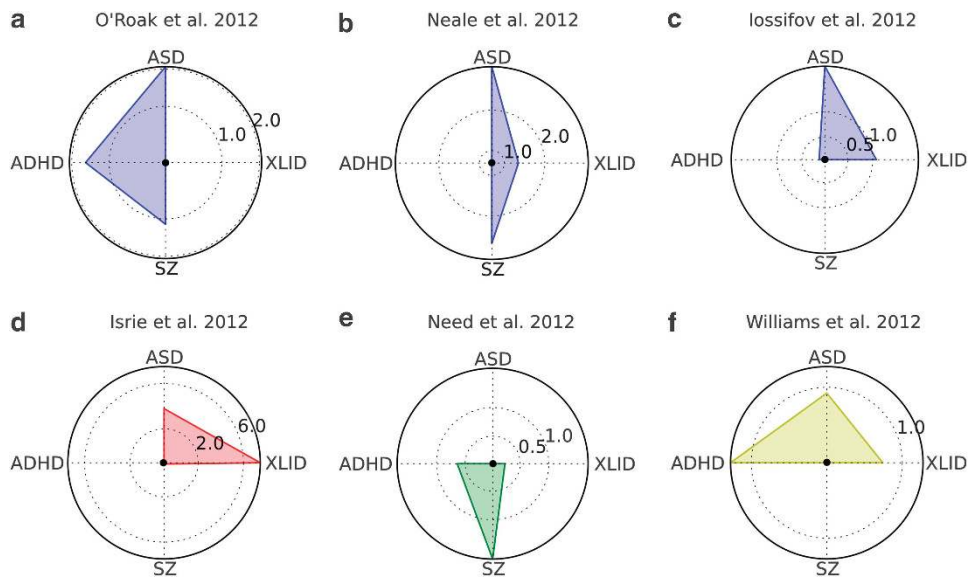
We have tested our gene network model using genetic data from six recent studies; ASD,<sup>5,6,63</sup> XLID,<sup>64</sup> ADHD<sup>65</sup> and SZ<sup>66</sup> (Supplementary Table S12). A binomial test and a standardised score (Z-score) were used to assess how these data sets would be distributed across the disorders represented in our AXAS-PPI network. A compelling result of the analysis was a demonstrated capacity to correctly predict the association of these data with ASD, XLID, ADHD and SZ, based on the highest Z-score (Figure 3; Supplementary Table S13). Our analysis also confirms cohort screening data can vary significantly in genetic composition between studies of the same disorder. Interestingly, two data sets (ASD<sup>5</sup> and XLID<sup>64</sup>) show significant Z-scores for more than one disorder. Although cross-validation analysis shows the AXAS-PPI

network model can potentially assess risk from large genetic screens, to assess individual risk, we would need to build a reliable classifier based on the statistical representation of gene variants across different populations.

#### DISCUSSION

The AXAS-PPI network indicates that there are up to 4000 genes that may contribute to neurodevelopmental and neuropsychiatric disorders (Figure 1). Therefore, the number of polygenic combinations including *de novo* variants<sup>67,68</sup> that potentially contribute to a neurological deficit, will be extraordinarily large. However, how can this diversity manifest with high incidence (that is, ~1–2% for ASD, ~1% for SZ) in the human population. The answer may lie in the ‘small world’ properties of the AXAS-PPI network, which is characterised by a small clustering coefficient and average shortest path length (Table 1). Greater than 90% of the ~4000 genes function with less than 4.0 degrees of separation between any two genes (average path length equals 4.4). Evidently, DNA variation is anchored by functional proximity (many genes contributing to the same pathways and processes) and neurodevelopmental and neuropsychiatric disorders are not unlike a journey where there are various paths to the same destination.

For this reason we can look to higher order interactions including biochemical pathways and biological processes to examine the nature of neurodevelopmental and neuropsychiatric disorders. We observed that genes associated with ASD, XLID, ADHD and SZ are mainly distributed in 13 modules, whose encoded proteins are predominantly involved in the regulation of transcription, synaptic transmission, cell-cell communication, intracellular signalling pathways, cell cycle, metabolic processes and nervous system development. However, the number of genes associated with ASD, XLID, ADHD and SZ are not equally



**Figure 3.** Binomial distribution radar plots with standardised Z-scores that examine association of genetic data with the AXAS–PPI model. The longest axis from the origin indicates the most significant association. All six genetic screening data were correctly predicted: ASD (**a**, **b** and **c**); XLID (**d**), SZ (**e**) and ADHD (**f**); see also Supplementary Table S12 and S13.

distributed in 9 out of 13 modules. Some functional domains such as synaptic transmission (M2), proteolysis (M6), phosphorylation (M10) and G-protein signalling (M12) seem to be commonly affected in all four disorders yet other modules, such as regulation of transcription (M1), vesicle-mediated transport (M11) and protein kinase signal transduction (M7) are overrepresented in ASD, ADHD and SZ, respectively (Figure 2).

The great majority of proteins (92%) in the AXAS–PPI network were found to be expressed in adult human brain and are likely to be regulated by similar mechanisms. We addressed this possibility by analysing the regulatory elements in the upstream control region and 3′-UTR of disorder genes in AXAS–PPI modules. A number of TF and miRNA binding sites were found to be significantly enriched in the non-coding regions of these genes. These analyses confirm that genes found in functional modules of the AXAS–PPI network are likely to be coregulated. It is also worth noting that genes with alternatively spliced 3′-UTRs are potentially regulated by different miRNAs. This is particularly important for neurodevelopmental and neuropsychiatric disorders considering that alternative splicing of genes is more prevalent in the brain and liver than in other tissues.<sup>50</sup> An inherent capacity to target many specific genes or variant transcripts in parallel makes miRNAs a powerfully configurable regulatory mechanism that has evolved for dynamic control of genes, biochemical pathways and biological processes.<sup>69,70</sup> Although miRNAs can be thought of as fine-tuning controls,<sup>69</sup> transcript levels of some genes in the brain might be under a dynamic balance that is delicate enough that even a slight change mediated by a miRNA or miRNA target mutation could translate into an observable phenotype. The importance of transcript levels might be demonstrated in certain genes, where haploinsufficiency can be enough to cause phenotypic effects.<sup>71</sup>

In summary our hypothetical framework can be used to examine the molecular basis of neurodevelopmental and neuropsychiatric disorders, and to assist with the selection of candidate genes for further analysis. A noteworthy outcome of this approach is the potential to classify, which disorder would arise from a particular set of genetic variations, with the eventual possibility of new diagnostic tools and therapeutic strategies. However we are mindful of genetic heterogeneity in different populations, sparse statistical representation and potential

problems with clinical/behavioural diagnosis. These issues will need to be addressed in order to build a more specific and sensitive diagnostic tool capable of assessing risk for individual genotypes.

#### CONFLICT OF INTEREST

The authors declare no conflict of interest.

#### ACKNOWLEDGEMENTS

AS Cristino was supported by a University of Queensland Post-doctoral Fellowship. C Claudianos received funding support from the ARC (110100292) and NHMRC of Australia (1008125), MA Bellgrove from the NHMRC (569636; 569533; 1002458) and L da F Costa from FAPESP (2011/50761-2) and CNPq (304351/2009-1) Brazil.

#### REFERENCES

- 1 Eapen V. Genetic basis of autism: is there a way forward? *Curr Opin Psychiatry* 2011; **24**: 226–236.
- 2 Najmabadi H, Hu H, Garshasbi M, Zemojtel T, Abedini SS, Chen W *et al*. Deep sequencing reveals 50 novel genes for recessive cognitive disorders. *Nature* 2011; **478**: 57–63.
- 3 Plomin R, Owen MJ, McGuffin P. The genetic basis of complex human behaviors. *Science* 1994; **264**: 1733–1739.
- 4 Basu SN, Kollu R, Banerjee-Basu S. AutDB: a gene reference resource for autism research. *Nucleic acids research* 2009; **37**(Database issue): D832–D836.
- 5 Neale BM, Kou Y, Liu L, Ma'ayan A, Samocha KE, Sabo A *et al*. Patterns and rates of exonic de novo mutations in autism spectrum disorders. *Nature* 2012; **485**: 242–245.
- 6 O'Roak BJ, Vives L, Girirajan S, Karakoc E, Krumm N, Coe BP *et al*. Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. *Nature* 2012; **485**: 246–250.
- 7 Pinto D, Pagnamenta AT, Klei L, Anney R, Merico D, Regan R *et al*. Functional impact of global rare copy number variation in autism spectrum disorders. *Nature* 2010; **466**: 368–372.
- 8 Miles JH. Autism spectrum disorders—a genetics review. *Genet Med* 2011; **13**: 278–294.
- 9 Chirazzi P, Schwartz CE, Gecz J, Neri G. XLMR genes: update 2007. *Eur J Hum Genet* 2008; **16**: 422–434.
- 10 Lubs HA, Stevenson RE, Schwartz CE. Fragile X and X-linked intellectual disability: four decades of discovery. *Am J Hum Genet* 2012; **90**: 579–590.
- 11 Gizer IR, Ficks C, Waldman ID. Candidate gene studies of ADHD: a meta-analytic review. *Hum genet* 2009; **126**: 51–90.

- 12 Williams NM, Zaharieva I, Martin A, Langley K, Mantripragada K, Fossdal R et al. Rare chromosomal deletions and duplications in attention-deficit hyperactivity disorder: a genome-wide analysis. *Lancet* 2010; **376**: 1401–1408.
- 13 Franke B, Neale BM, Faraone SV. Genome-wide association studies in ADHD. *Hum genet* 2009; **126**: 13–50.
- 14 Jia P, Sun J, Guo AY, Zhao Z. SZGR: a comprehensive schizophrenia gene resource. *Mol Psychiatry* 2010; **15**: 453–462.
- 15 Allen NC, Bagade S, McQueen MB, Ioannidis JP, Kavvoura FK, Khoury MJ et al. Systematic meta-analyses and field synopsis of genetic association studies in schizophrenia: the SzGene database. *Nat Genet* 2008; **40**: 827–834.
- 16 Lewis CM, Levinson DF, Wise LH, DeLisi LE, Straub RE, Hovatta I et al. Genome scan meta-analysis of schizophrenia and bipolar disorder, part II: Schizophrenia. *Am J Hum Genet* 2003; **73**: 34–48.
- 17 Ng MY, Levinson DF, Faraone SV, Suarez BK, DeLisi LE, Arinami T et al. Meta-analysis of 32 genome-wide linkage studies of schizophrenia. *Mol Psychiatry* 2009; **14**: 774–785.
- 18 Ross CA, Margolis RL, Reading SA, Pletnikov M, Coyle JT. Neurobiology of schizophrenia. *Neuron* 2006; **52**: 139–153.
- 19 Sun J, Jia P, Fanous AH, Webb BT, van den Oord EJ, Chen X et al. A multi-dimensional evidence-based candidate gene prioritization approach for complex diseases-schizophrenia as a case. *Bioinformatics* 2009; **25**: 2595–6602.
- 20 Sun J, Kuo PH, Riley BP, Kendler KS, Zhao Z. Candidate genes for schizophrenia: a survey of association studies and gene ranking. *Am J Med Genet B Neuropsychiatr Genet* 2008; **147B**: 1173–1181.
- 21 Ben-David E, Shifman S. Networks of neuronal genes affected by common and rare variants in autism spectrum disorders. *PLoS Genet* 2012; **8**: e1002556.
- 22 O'Roak BJ, Deriziotis P, Lee C, Vives L, Schwartz JJ, Girirajan S et al. Exome sequencing in sporadic autism spectrum disorders identifies severe de novo mutations. *Nature genetics* 2011; **43**: 585–589.
- 23 Sebat J, Lakshmi B, Malhotra D, Troge J, Lese-Martin C, Walsh T et al. Strong association of de novo copy number mutations with autism. *Science* 2007; **316**: 445–449.
- 24 Voineagu I. Gene expression studies in autism: moving from the genome to the transcriptome and beyond. *Neurobiol Dis* 2012; **45**: 69–75.
- 25 Bailey A, Le Couteur A, Gottesman I, Bolton P, Simonoff E, Yuzda E et al. Autism as a strongly genetic disorder: evidence from a British twin study. *Psychol med* 1995; **25**: 63–77.
- 26 Anney R, Klei L, Pinto D, Regan R, Conroy J, Magalhaes TR et al. A genome-wide scan for common alleles affecting risk for autism. *Hum mol genet* 2010; **19**: 4072–4082.
- 27 Clarke RA, Lee S, Eapen V. Pathogenetic model for Tourette syndrome delineates overlap with related neurodevelopmental disorders including Autism. *Transl Psychiatry* 2012; **2**: e158.
- 28 Sudhof TC. Neuroligins and neuexins link synaptic function to cognitive disease. *Nature* 2008; **455**: 903–911.
- 29 Stevenson RE, Schwartz CE, Rogers RC. *Atlas of X-linked intellectual disability syndromes*. 2nd edn. Oxford University Press: New York; USA, 2012, 368 pp.
- 30 Levy F, Hay DA, McStephen M, Wood C, Waldman I. Attention-deficit hyperactivity disorder: a category or a continuum? Genetic analysis of a large-scale twin study. *J Am Acad Child Adolesc Psychiatry* 1997; **36**: 737–744.
- 31 Gainetdinov RR, Wetsel WC, Jones SR, Levin ED, Jaber M, Caron MG. Role of serotonin in the paradoxical calming effect of psychostimulants on hyperactivity. *Science* 1999; **283**: 397–401.
- 32 Giros B, Jaber M, Jones SR, Wightman RM, Caron MG. Hyperlocomotion and indifference to cocaine and amphetamine in mice lacking the dopamine transporter. *Nature* 1996; **379**: 606–612.
- 33 Lasky-Su J, Anney RJ, Neale BM, Franke B, Zhou K, Maller JB et al. Genome-wide association scan of the time to onset of attention deficit hyperactivity disorder. *Am J Med Genet B Neuropsychiatr Genet* 2008; **147B**: 1355–1358.
- 34 Mick E, Todorov A, Smalley S, Hu X, Loo S, Todd RD et al. Family-based genome-wide association scan of attention-deficit/hyperactivity disorder. *J Am Acad Child Adolesc Psychiatry* 2010; **49**: 898–905e893.
- 35 Neale BM, Lasky-Su J, Anney R, Franke B, Zhou K, Maller JB et al. Genome-wide association scan of attention deficit hyperactivity disorder. *Am J Med Genet B Neuropsychiatr Genet* 2008; **147B**: 1337–1344.
- 36 Neale BM, Medland S, Ripke S, Anney RJ, Asherson P, Buitelaar J et al. Case-control genome-wide association study of attention-deficit/hyperactivity disorder. *J Am Acad Child Adolesc Psychiatry* 2010; **49**: 906–920.
- 37 Cardno AG, Gottesman, II. Twin studies of schizophrenia: from bow-and-arrow concordances to star wars Mx and functional genomics. *Am J Med Genet* 2000; **97**: 12–17.
- 38 Sullivan PF, Kendler KS, Neale MC. Schizophrenia as a complex trait: evidence from a meta-analysis of twin studies. *Arch Gen Psychiatry* 2003; **60**: 1187–1192.
- 39 Lee SH, DeCandia TR, Ripke S, Yang J, Sullivan PF, Goddard ME et al. Estimating the proportion of variation in susceptibility to schizophrenia captured by common SNPs. *Nat Genet* 2012; **44**: 247–250.
- 40 Gottesman, II, Gould TD. The endophenotype concept in psychiatry: etymology and strategic intentions. *Am J Psychiatry* 2003; **160**: 636–645.
- 41 Purcell SM, Wray NR, Stone JL, Visscher PM, O'Donovan MC, Sullivan PF et al. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* 2009; **460**: 748–752.
- 42 Weiss LA, Arking DE, Daly MJ, Chakravarti A. A genome-wide linkage and association scan reveals novel loci for autism. *Nature* 2009; **461**: 802–808.
- 43 Mefford HC, Sharp AJ, Baker C, Itsara A, Jiang Z, Buysse K et al. Recurrent rearrangements of chromosome 1q21.1 and variable pediatric phenotypes. *The New England journal of medicine* 2008; **359**: 1685–1699.
- 44 Sharp AJ, Mefford HC, Li K, Baker C, Skinner C, Stevenson RE et al. A recurrent 15q13.3 microdeletion syndrome associated with mental retardation and seizures. *Nature genetics* 2008; **40**: 322–328.
- 45 Voineskos AN, Lett TA, Lerch JP, Tiwari AK, Ameis SH, Rajji TK et al. Neurexin-1 and frontal lobe white matter: an overlapping intermediate phenotype for schizophrenia and autism spectrum disorders. *PLoS one* 2011; **6**: e20982.
- 46 Stark C, Breitkreutz BJ, Reguly T, Boucher L, Breitkreutz A, Tyers M. BioGRID: a general repository for interaction datasets. *Nucleic Acids Res* 2006; **34**(Database issue): D535–D539.
- 47 Kandasamy K, Keerthikumar S, Goel R, Mathivanan S, Patankar N, Shafreen B et al. Human Proteinpedia: a unified discovery resource for proteomics research. *Nucleic Acids Res* 2009; **37**(Database issue): D773–D781.
- 48 Mathivanan S, Ahmed M, Ahn NG, Alexandre H, Amanchy R, Andrews PC et al. Human Proteinpedia enables sharing of human protein data. *Nat Biotechnol* 2008; **26**: 164–167.
- 49 Mathivanan S, Pandey A. Human Proteinpedia as a resource for clinical proteomics. *Mol Cell Proteomics* 2008; **7**: 2038–2047.
- 50 Pan Q, Shai O, Lee LJ, Frey BJ, Blencowe BJ. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat Genet* 2008; **40**: 1413–1415.
- 51 Costa LdF OliveiraON, Travieso G, Rodrigues FA, Villas-Boas PR, Antiquera L et al. Analyzing and modeling real-world phenomena with complex networks: a survey of applications. *Advances in Physics* 2011; **60**: 329–412.
- 52 Girvan M, Newman ME. Community structure in social and biological networks. *Proc Natl Acad Sci USA* 2002; **99**: 7821–7826.
- 53 Wingender E, Chen X, Hehl R, Karas H, Lieblich I, Matys V et al. TRANSFAC: an integrated system for gene expression regulation. *Nucleic Acids Res* 2000; **1**: 316–319.
- 54 Vlieghe D, Sandelin A, De Bleser PJ, Vleminckx K, Wasserman WW, van Roy F et al. A new generation of JASPAR, the open-access repository for transcription factor binding site profiles. *Nucleic Acids Res* 2006; **34**(Database issue): D95–D97.
- 55 Kozomara A, Griffiths-Jones S. miRBase: integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Res* 2011; **39**(Database issue): D152–D157.
- 56 Wang K, Zhang H, Ma D, Bucan M, Glessner JT, Abrahams BS et al. Common genetic variants on 5p14.1 associate with autism spectrum disorders. *Nature* 2009; **459**: 528–533.
- 57 Ripke S, Sanders AR, Kendler KS, Levinson DF, Sklar P, Holmans PA et al. Genome-wide association study identifies five new schizophrenia loci. *Nat Genet* 2011; **43**: 969–976.
- 58 Shi J, Levinson DF, Duan J, Sanders AR, Zheng Y, Pe'er I et al. Common variants on chromosome 6p22.1 are associated with schizophrenia. *Nature* 2009; **460**: 753–757.
- 59 Shifman S, Johannesson M, Bronstein M, Chen SX, Collier DA, Craddock NJ et al. Genome-wide association identifies a common variant in the reelin gene that increases the risk of schizophrenia only in women. *PLoS Genet* 2008; **4**: e28.
- 60 Stefansson H, Ophoff RA, Steinberg S, Andreassen OA, Cichon S, Rujescu D et al. Common variants conferring risk of schizophrenia. *Nature* 2009; **460**: 744–747.
- 61 Wang KS, Liu XF, Aragam N. A genome-wide meta-analysis identifies novel loci associated with schizophrenia and bipolar disorder. *Schizophr Res* 2010; **124**: 192–199.
- 62 Birney E, Stamatoyannopoulos JA, Dutta A, Guigo R, Gingeras TR, Margulies EH et al. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* 2007; **447**: 799–816.
- 63 Iossifov I, Ronemus M, Levy D, Wang Z, Hakker I, Rosenbaum J et al. De novo gene disruptions in children on the autistic spectrum. *Neuron* 2012; **74**: 285–299.
- 64 Isrie M, Froyen G, Devriendt K, de Ravel T, Frys JP, Vermeesch JR et al. Sporadic male patients with intellectual disability: contribution of X-chromosome copy number variants. *Eur J Med Genet* 2012; **55**: 577–585.
- 65 Williams NM, Franke B, Mick E, Anney RJ, Freitag CM, Gill M et al. Genome-wide analysis of copy number variants in attention deficit hyperactivity disorder: the role of rare variants and duplications at 15q13.3. *Am J Psychiatry* 2012; **169**: 195–204.

- 66 Need AC, McEvoy JP, Gennarelli M, Heinzen EL, Ge D, Maia JM *et al*. Exome sequencing followed by large-scale genotyping suggests a limited role for moderately rare risk factors of strong effect in schizophrenia. *Am J Hum Genet* 2012; **91**: 303–312.
- 67 Kong A, Frigge ML, Masson G, Besenbacher S, Sulem P, Magnusson G *et al*. Rate of de novo mutations and the importance of father's age to disease risk. *Nature* 2012; **488**: 471–475.
- 68 Sanders SJ, Murtha MT, Gupta AR, Murdoch JD, Raubeson MJ, Willsey AJ *et al*. De novo mutations revealed by whole-exome sequencing are strongly associated with autism. *Nature* 2012; **485**: 237–241.
- 69 Bartel DP. MicroRNAs: target recognition and regulatory functions. *Cell* 2009; **136**: 215–233.
- 70 Kim VN, Nam JW. Genomics of microRNA. *Trends Genet* 2006; **22**: 165–173.
- 71 Betancur C. Etiological heterogeneity in autism spectrum disorders: more than 100 genetic and genomic disorders and still counting. *Brain research* 2011; **1380**: 42–77.



This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/3.0/>

Supplementary Information accompanies the paper on the Molecular Psychiatry website (<http://www.nature.com/mp>)