Genomics 93 (2009) 120–129

Contents lists available at ScienceDirect

Genomics

journal homepage: www.elsevier.com/locate/ygeno

Comparative analysis of neurological disorders focuses genome-wide search for autism genes

D.P. Wall *, F.J. Esteban, T.F. DeLuca, M. Huyck, T. Monaghan, N. Velez de Mendizabal, J. Goñí, I.S. Kohane

Center for Biomedical Informatics, Harvard Medical School, Boston, MA 02115, USA

ARTICLE INFO

Article history:
Received 30 March 2008
Accepted 23 September 2008
Available online 12 November 2008

Keywords:
Autism spectrum disorder
ASD
Autism
Autism network
Autism sibling disorders
Gene set enrichment
Process enrichment
Comparative network analysis
Systems biology

ABSTRACT

The behaviors of autism overlap with a diverse array of other neurological disorders, suggesting common molecular mechanisms. We conducted a large comparative analysis of the network of genes linked to autism with those of 432 other neurological diseases to circumscribe a multi-disorder subcomponent of autism. We leveraged the biological process and interaction properties of these multi-disorder autism genes to overcome the across-the-board multiple hypothesis corrections that a purely data-driven approach requires. Using prior knowledge of biological process, we identified 154 genes not previously linked to autism of which 42% were significantly differentially expressed in autistic individuals. Then, using prior knowledge from interaction networks of disorders related to autism, we uncovered 334 new genes that interact with published autism genes, of which 87% were significantly differentially regulated in autistic individuals. Our analysis provided a novel picture of autism from the perspective of related neurological disorders and suggested a model by which prior knowledge of interaction networks can inform and focus genome-scale studies of complex neurological disorders.

Published by Elsevier Inc.

Introduction

Autism is a complex multigenic disorder with a wide range of phenotypes. Although it is clear that the disorder is highly heritable, the molecular agents responsible remain elusive and it remains unclear whether the genetic component is a combination of a few common variants, or of many rare variants [1]. More than 100 genes have been tied to autism, each of which is involved in numerous biological processes and in a variety of different molecular interactions. It becomes daunting if at all manageable for a single researcher to encompass the complexity of this autism gene space, and perhaps for this reason, integration of this space into a productive set of hypotheses has taken a backseat to investigations of single genes or mechanisms. To date, these efforts have not delivered highly accurate markers or proven targets for therapeutic intervention.

As a result, autism remains a behavioral or symptomatic diagnosis rather than a molecular diagnosis. The behavioral manifestations include social anxiety and gaze avoidance, repetitive movements and behaviors, hypersensitivity to touch, reduced coordination, delayed speech and echolalia. Interestingly, several of these symptoms overlap with other neurological disorders, including Tuberous Sclerosis [2,3], Hypotonia [4], Rett Syndrome and Fragile X syndrome. These behavioral similarities suggest that there might be shared molecular mechanisms, at least in part. In support of this suggestion, the causative genes for Fragile X and Rett Syndrome have been linked to autism [5–7]. Interestingly even though disorders like Fragile X are monogenic, their behaviors can vary. For example, in Fragile X, behaviors range from relatively mild learning disabilities to mental retardation, speech impairments, and echolalia. The reason for this range is most likely due in part to a network of interactions between the genes linked to monogenic disorders and their direct or indirect binding partners. In such a case, mutations in the causative gene alter its interactions with neighboring genes that are required to perform specific biological functions, and the effects of these alterations become compounded when binding partners downstream of the causative agent are also mutated or otherwise dysfunctional.

Therefore, it is the interaction network and set of biological processes it performs rather than a single gene that effects symptoms indicative of classically monogenic disorders [8–10]. Moreover, this suggests a testable hypothesis: disorders with behavioral similarities to autism may have many genes in common with autism. For example, this would imply significant overlap in the biological processes disordered in instances of Fragile X with the behaviorally overlapping non-Fragile X causes of autism. By quantifying this overlap at the level of molecular physiology we aim to obtain a more complete understanding of the spectrum of behaviors indicative of autism. Eventually, the comparison of autism to other disorders of the central nervous system using symptoms,

* Corresponding author.
E-mail address: dpwall@hms.harvard.edu (D.P. Wall).

0888-7543/$ – see front matter. Published by Elsevier Inc.
doi:10.1016/j.ygeno.2008.09.015

genes, and entire processes may yield powerful insights that have an immediate impact our understanding of the disorder's etiology. In this investigation, we compared autism to 432 other neurological disorders at the levels of biological processes and gene networks. By this comparative approach, we were able to leverage information from related disorders to predict new genes of possible importance to the etiology of autism. The figure of merit that we use to validate our predictions is the capability of finding significance in high-throughout studies, in this instance, transcriptome-scale expression profiling. We determined that the focus provided by prior knowledge yields the specificity to overcome the multiple hypothesis testing corrections that purely data-driven approaches entail.

Results

The multi-disorder component of the autism network

Using OMIM and GeneCards we generated gene lists for 433 neurological disorders listed by NINDS as of December 2006. We selected a subset of the disorders that had 3 or more genes in common with autism. By converting the gene lists into a matrix of gene presence and absence we were able to generate a disorder phylogeny that grouped autism together with 13 related disorders, including Microcephaly, Mental Retardation, Ataxia, and Seizure Disorder (Fig.1). We focused on the members of this autism sibling group for subsequent analyses.

We used the tool STRING to construct gene networks for each member of the autism sibling group in order to investigate genetic overlaps with autism (summary of edge information available online in Supplementary Table 1). Of the 127 genes in our candidate list for autism, 66 have also been linked to at least one other autism sibling disorder (Table 1). This multi-disorder gene set (MDAG) formed a highly connected subcomponent of the complete autism network (Fig. 2), suggesting that the genes in the MDAG share biological function. To test this, we used the Explain™ System from BioBase, which contains an abundance of manually reviewed information, to identify significant overrepresentation of MDAG genes in biological processes. A total of 12 biological processes had significant enrichment following Bonferroni multiple test correction (Table 2).

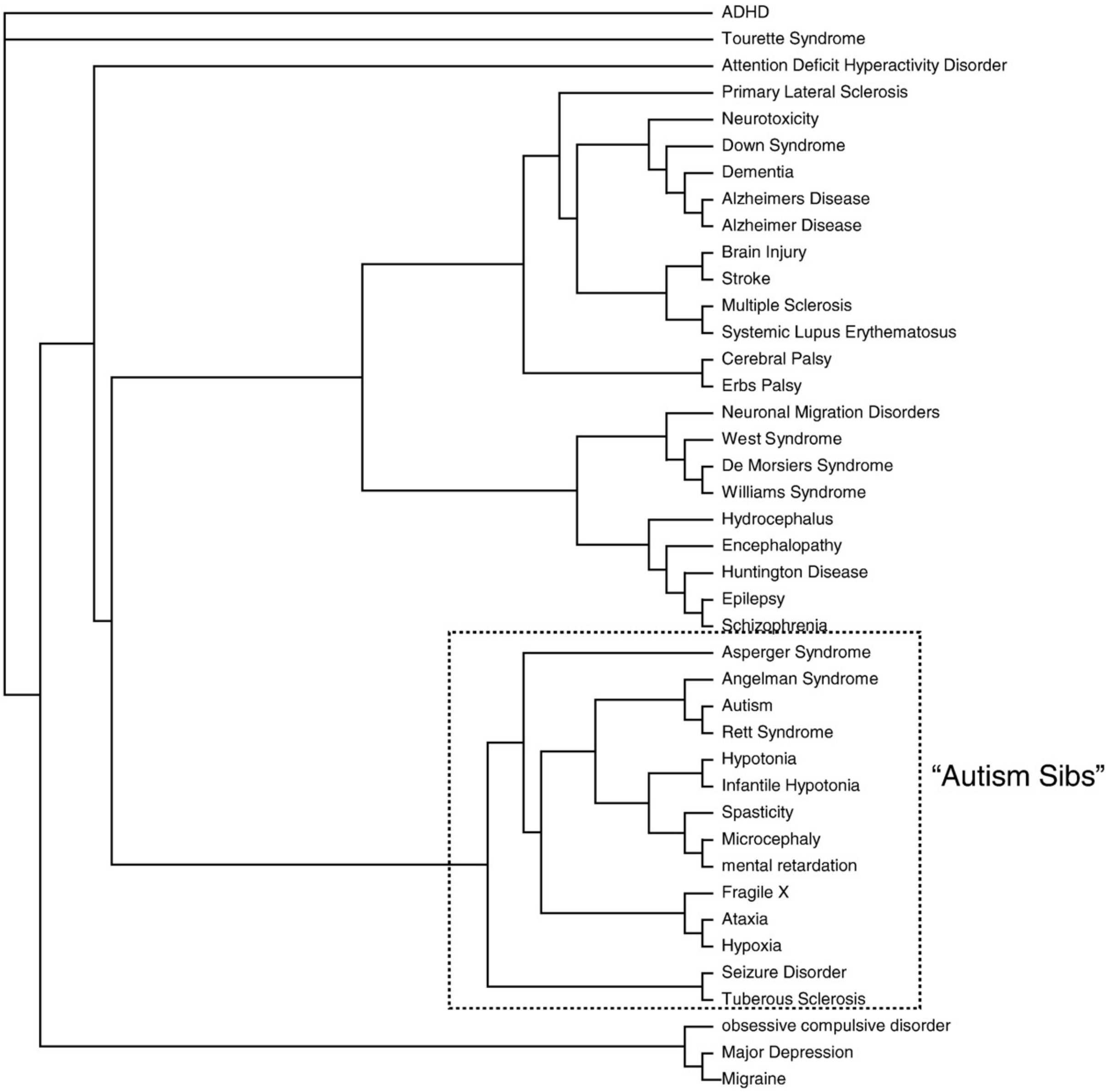

Fig. 1. Maximum-parsimony based phylogeny of autism and related neurological disorders. The group containing autism is highlighted and referred to in the text as "Autism sibling disorders" or "Autism sibling group."

Table 1
Multi-disorder autism gene set (MDAG). The 66 MDAG genes are highlighted in orange within the autism network (Fig. 2) and are found in at least one autism sibling disorder (Fig. 1)

Gene	Neurological disorders
ABAT	Tuberous sclerosis, autism, hypotonia, mental retardation
ACADL	Autism, hypotonia
ADA	Hypoxia, autism
ADM	Hypoxia, autism
ADSL	Microcephaly, autism, hypotonia, mental retardation
ALDH5A1	Ataxia, seizure disorder, autism, hypotonia, mental retardation
APOE	Hypoxia, autism, tuberous sclerosis
ATP10A	Angelman syndrome, microcephaly, ataxia, autism, hypotonia
ARX	Microcephaly, spasticity, mental retardation
ASPG1	Autism, asperger syndrome
ASPG2	Autism, asperger syndrome
BTD	Ataxia, hypotonia, mental retardation
CACNA1D	Autism, Rett syndrome
CD69	Ataxia, autism
	Infantile hypotonia, Angelman ayndrome, ataxia, Rett syndrome
CDKL5	Microcephaly, autism, hypotonia, mental retardation
CHRNA4	Autism, mental retardation
CHRNA7	Autism, mental retardation
DAB1	Autism, mental retardation
	Seizure disorder, hypoxia, tuberous sclerosis, microcephaly, autism
DCX	Mental retardation
DGCR	Autism, mental retardation
DHCR7	Microcephaly, autism, hypotonia, mental retardation
DPYD	Microcephaly, ataxia, autism, mental retardation
EXT1	Autism, mental retardation
EXT2	Autism, mental retardation
	Fragile X, infantile hypotonia, ataxia, Rett syndrome, microcephaly
FMR1	Autism, hypotonia, mental retardation
FOXP2	Ataxia, autism
FXR1	Fragile X, autism, mental retardation
GABRA5	Angelman syndrome, autism
GABRB3	Angelman syndrome, ataxia, autism, mental retardation
GABRG2	Ataxia, seizure disorder, autism
GATA3	Hypoxia, autism
GLO1	Ataxia, autism
GRIN2A	Hypoxia, autism
GRPR	Autism, Rett syndrome, mental retardation
HOXA1	Asperger syndrome, autism, Rett syndrome, mental retardation
KIF1A	Autism, tuberous sclerosis
MAGEL2	Angelman syndrome, autism, hypotonia, mental retardation
MAOA	Autism, mental retardation
MAP2	Tuberous sclerosis, hypoxia, autism, Rett syndrome, mental retardation
MBD1	Autism, Rett syndrome
MBD2	Autism, Rett syndrome
	Fragile X, infantile hypotonia, seizure disorder, Angelman syndrome, spasticity, ataxia, Asperger syndrome, Rett syndrome, microcephaly,
MECP2	Tuberous sclerosis, autism, hypotonia, mental retardation
MED12	Fragile X, autism, mental retardation
MET	Hypoxia, autism
MTF1	Hypoxia, autism
NDN	Angelman syndrome, hypoxia, Rett syndrome, autism, hypotonia, mental retardation
NDNL2	Angelman syndrome, autism, seizure disorder, spasticity, tuberous sclerosis, microcephaly, autism
NF1	Mental metardation
NLGN3	Asperger syndrome, autism, mental retardation
NLGN4Y	Asperger syndrome, Angelman syndrome, autism, mental retardation
NLGN4X	Asperger syndrome, autism, mental retardation
NTF4	Autism, mental retardation
PAX3	Autism, mental retardation
PTEN	Ataxia, seizure disorder, hypoxia, autism
RELN	Ataxia, tuberous sclerosis, autism, hypotonia, mental retardation
SCN1A	Ataxia, autism
SDC2	Autism, mental retardation
SLC40A1	Ataxia, autism
SLC6A4	Fragile X, hypoxia, Asperger syndrome, tuberous sclerosis, autism, Rett syndrome, mental retardation
SNRPN	Angelman syndrome, Rett syndrome, microcephaly, autism, hypotonia, mental retardation
SSTR5	Autism, tuberous sclerosis
TH	Hypoxia, spasticity, autism, Rett syndrome, infantile hypotonia
TPH1	Autism, Rett syndrome, mental retardation
TSC1	Tuberous sclerosis, autism, mental retardation
	Angelman syndrome, spasticity, ataxia, hypoxia, Rett syndrome,
UBE3A	Microcephaly, autism, hypotonia, mental retardation
VLDLR	Fragile X, ataxia, hypoxia, autism

Biological process-driven search for new autism genes

One possible reason for the large extent of behavioral overlap between autism and many of the disorders in the autism sibling group may be the result of context specific dysregulation of any or all of the processes for which the MDAG is enriched. Thus, other autism sibling disorder-linked genes that are known to be involved in any of the 12 significantly enriched processes but that have not yet been implicated in autism may represent viable autism candidates. To address this hypothesis, we mined the gene lists of the autism sibling disorders and identified a non-redundant set of 154 process-based candidates (PBC). The process "transmission of nerve impulse" was not found among the genes in the autism sibling disorders. All other enriched processes yielded 2 or more unique predictions all of which are involved in at least two of the autism sibling disorders, but not found in our original autism candidate list (Table 3; complete list of 154 process-based candidates is available as online Supplementary Table 2).

To empirically test the importance of the PBC in autism, we asked whether any exhibited significantly different gene expression in autistic patients in comparison to controls. As we were only concerned with testing the PBC, we performed multiple test correction on just these 154 hypotheses. Specifically we calculated q values, an FDR-based measure of significance, and discovered that 64 of the 154, 42%, were significantly differentially regulated with $q \leq 0.05$ (Table 3). By recalculating q values for 1000 randomly constructed gene sets of size 154 (drawing from the entire set of genes sampled in the microarray experiment used), we determined that this frequency of significant features was unlikely to occur by chance ($p < 0.01$).

Intersecting the autism and sibling disorder networks: network-driven search for new autism candidates

Next, using data derived from STRING [11], we constructed the network of genes for every autism sibling disorder to study the interactions surrounding members of the MDAG, specifically focusing on direct neighbors to MDAG genes that were not present in our original autism gene list (e.g., as shown in Fig. 3). This analysis found 334 genes that were directly connected to a member of the MDAG, but not known to be of importance in autism. Of these, 198 occurred in 1 autism sibling disorder, 83 occurred in 2, 35 in 3, 9 in 4 disorders and 9 in 5 with 5 being the maximum extent of overlap (Table 4 summarizes the top 40 genes after ranking on the number of disorders to which the genes have been linked; a full list is available online as Supplementary Table 3). These network-based candidates contained several genes that have some pre-established association to neurological dysfunction. For example L1CAM's functions include guidance of neurite outgrowth in development, neuronal cell migration, axon bundling, synaptogenesis, myelination, and neuronal cell survival and it is among a family of genes that were recently shown to have roles in neurological dysfunction [12]. In addition, BDNF has been associated with autism [13], and SLC6A8 has established ties to autism via dysfunction of creatine transporter activities [14].

We used the same mRNA expression data as above to test if the 334 network-based candidates were differential expressed within autistics when compared to control samples. Correcting only for the 334 hypotheses being tested, we found that 289 had q values less than 0.05 (all NBC q values available in Supplementary Table 3). That is, ~87% of the genes predicted via our comparative analysis of the disease networks turned out to be significantly differentially regulated in autism in comparison to controls. To determine if this percentage could

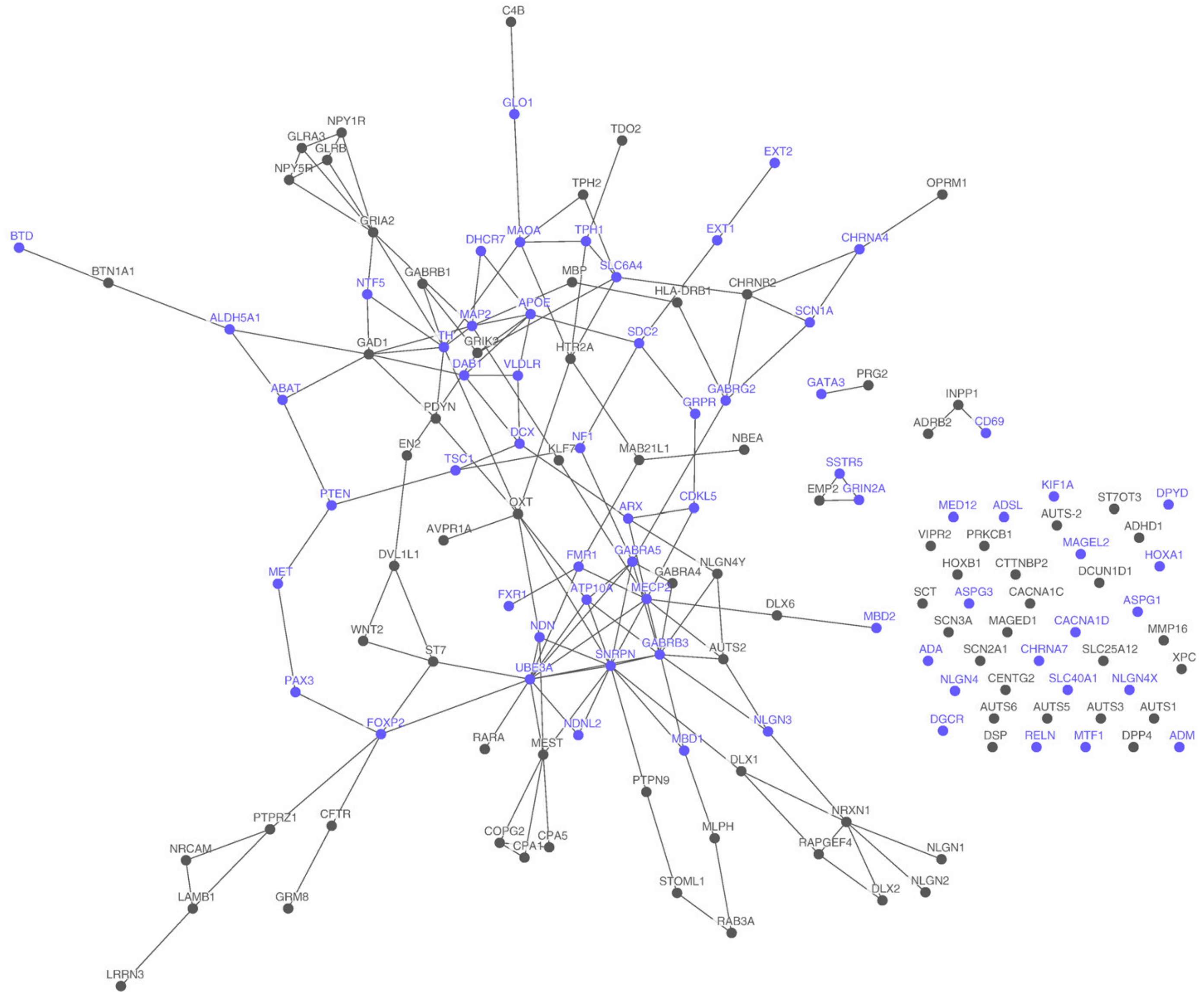

Fig. 2. The complete network of autism candidate genes. The autism network with all multi-disorder autism genes (MDAG) highlighted. These are genes that occur in one or more of the autism sibling disorders, which are circumscribed in Fig. 1.

arise by chance or represent a bias of our methods, we constructed 1000 groups of 334 genes by randomly sampling from the complete set of autism expression data and recalculated q values. The mean number of significant features within these randomly constructed genes sets was 31, indicating that our observed value of 289 was highly unlikely to occur by chance ($p<0.01$).

Intersection of the network and process-based approaches to prioritize genes for further study in autism

We can intersect our two computational approaches to triangulate on the set of genes that were independently predicted and verified by both strategies and reduce the size of the intersection by filtering out those genes that occur in 2 or less autism sibling disorders. This is predicated on the assumption that genes with multiple independent associations to neurological disorders are more likely to have an impact on normal neurological development and function. A total of 9 genes satisfied these criteria and were 'SLC16A2', 'SLC6A8', 'OPHN1', 'FXN', 'AR', 'L1CAM', 'FLNA', 'MYO5A', 'PAFAH1B1.' All were found to be differentially expressed in one of the two tests for differential expression and all occur in 3 or more autism sibling disorders. We performed a process enrichment analysis to determine whether these 9 genes were enriched for a particular set of biological processes. Although no process was significant following multiple test correction, a total of 14 processes had uncorrected p values below 0.05 and included cell and cytoskeletal organization and biogenesis, cell motility, and cell communication (Table 5). That the two analytical strategies had appreciable overlap was not entirely surprising, as we should expect genes to interact if they are involved in the same biological processes. Nonetheless, triangulation on the same genes using independent approaches and different data sources provides an internal consistency check and suggests that further investigations of these 9 genes are warranted. Recent work suggested that malformations in cytoskeletal organization may contribute to neuronal migration and lead to neurological impairments [15]. It is possible that similar defects are occurring in autistic patients via dysregulation of these 9 genes or combinations thereof. Further study, such as experiments to determine whether single gene perturbations entail coordinated changes among any of these genes would help to confirm this hypothesis.

Discussion

In this study, we conducted a comparative analysis of autism and 432 other neurological disorders listed by the National

Table 2
Biological processes for which the multi-disorder component of the autism gene set (MDAG) were enriched

Biological process	*P* value	MDAG genes
Transmission of nerve impulse	3.00E-11	*ABAT, ALDH5A1, APOE, CHRNA4, CHRNA7, GABRA5, GABRB3, GABRG2, GATA3, GRIN2A, MAOA, MET, NF1, NTF4, SCN1A, SLC6A4, TH, TPH1, TSC1*
Nervous system development	3.29E-11	*ALDH5A1, APOE, ARX, BTD, CHRNA4, DAB1, DCX, FMR1, FOXP2, GABRA5, GATA3, GRIN2A, HOXA1, MAP2, MECP2, MET, NDN, NF1, PAX3, PTEN, RELN, TSC1, UBE3A, VLDLR*
Synaptic transmission	7.68E-10	*ABAT, ALDH5A1, APOE, CHRNA4, CHRNA7, GABRA5, GABRB3, GABRG2, GATA3, GRIN2A, MAOA, MET, NF1, NTF4, SLC6A4, TH, TPH1*
Cell–cell signaling	3.12E-09	*ABAT, ADM, ALDH5A1, APOE, CHRNA4, CHRNA7, GABRA5, GABRB3, GABRG2, GATA3, GRIN2A, MAOA, MET, NF1, NTF4, SCN1A, SLC6A4, SSTR5, TH, TPH1, TSC1*
Brain development	2.64E-06	*ARX, DAB1, DCX, FOXP2, GABRA5, HOXA1, MET, NF1, RELN, TSC1, UBE3A*
Generation of neurons	2.43E-05	*APOE, ARX, DAB1, DCX, MAP2, MECP2, MET, NDN, NF1, NTF4, PTEN, RELN, VLDLR*
Regulation of cell proliferation	2.45E-04	*ADM, ARX, CHRNA7, DHCR7, FOXP2, GRPR, MECP2, MET, NDN, NF1, PAX3, PTEN, SSTR5, TSC1*
Cell migration	3.93E-03	*ARX, DAB1, DCX, MET, NDN, NF1, PAX3, PTEN, RELN, VLDLR*
Homeostasis	1.90E-02	*ADM, APOE, ARX, CHRNA4, CHRNA7, GRIN2A, MBD1, NDN, NF1, SCN1A, SLC40A1, SSTR5, TH*
Cell morphogenesis	1.94E-02	*APOE, ARX, ATP10A, DCX, MAP2, MECP2, NDN, PTEN, RELN, TSC1*
Ion transport	2.74E-02	*ARX, CACNA1D, CHRNA4, CHRNA7, GABRA5, GABRB3, GABRG2, GRIN2A, MECP2, MET, SCN1A, SLC40A1, TSC1*
Cell differentiation	4.35E-02	*ADM, APOE, ARX, DAB1, DCX, DHCR7, EXT2, FXR1, GATA3, GLO1, GRIN2A, MAP2, MECP2, MET, NDN, NF1, NTF4, PAX3, PTEN, RELN, TSC1, VLDLR*

Identities of the MDAG genes overrepresented in the processes as well as the corrected *p* value for the enrichment scores are provided. Enrichment was calculated in the ExPlain^{Tm} 2.3 Tool from BioBase (www.biobase-international.com) and *p* values were adjusted by the Bonferroni method.

Institute of Neurological Disorders and Stroke (NINDS). By focusing on a set of disorders that appeared to be most closely related to autism (autism sibling disorders, Fig. 1), we found that more than half of the published autism genes have implicated in related neurological disorders. This confirmed that there is molecular overlap and suggested that these disorders may share molecular mechanisms that could be informative to our understanding of the genetic etiology of autism. The multi-disorder component of the autism network (MDAG) was highly connected and enriched for a small number of biologically informative processes, including synaptic transmission, and central nervous system development.

Motivated by these findings we devised two analytical strategies to test whether information from related disorders could provide meaningful focus to the genome-wide search for autism gene candidates. The first, a process-based strategy, was predicated on the assumption that processes for which the MDAG genes were enriched are generally important for neurological dysfunction. It is further predicated on the hypothesis that genes involved in these processes that have been linked to one or more autism sibling disorder, but that have not yet been implicated in autism, should be autism gene candidates. We tested this hypothesis using available whole-genomic expression data from 17 autistics with early onset and 12 controls from the general population (data from [16]) and found that 42% of the predictions were under significant differential expression in autistic individuals. The fact that they have been implicated in neurological dysfunction and involved in what appear to be important autism processes together makes these genes appealing new leads for the understanding the molecular pathology of autism.

The second strategy was grounded in the now mainstream understanding that protein interaction networks can provide valuable and often serendipitous leads for disease causing agents [17–22]. In our network-based strategy, rather than look at the entire protein interaction network, we filtered the set of all interactions to MDAG genes such that they included only those proteins present in the list of autism sibling disorders, but absent from out list of published autism candidates. This strategy uncovered 334, of which 87% were found to be significantly differentially expressed in autistic when compared to controls. This network-based strategy revealed that there are a large number of differences in the interaction networks of these related neurological disorders, even one step removed from those genes that are shared among them (the MDAG). While the differences may reflect real mechanistic differences between autism and its sibling disorders, given the high number found to be differentially regulated here, it is more likely that at least a fraction of them represent key gaps in our understanding of autism.

In both analytical strategies, we were able to use the prior knowledge from external resources, in this case from biological processes and interaction networks, to yield focused sets of genes hypothesized to be under differential regulation in autistics. From the methodological perspective, it bears emphasis that in the absence of such prior knowledge many of the genes measured on the small number of patients in this study had False Discovery Rates> 0.5. This a frequent circumstance in instances of weak signals and large background noise in many transcriptome-level experiments [23–25]. In contrast, with the application of prior knowledge, the majority of the genes tested had an FDR<0.05. This inversion of the usual specificity problem at the genome-scale points to a promising fusion of knowledge-driven and data-driven methods.

Finally, although the networks evaluated herein should be considered preliminary, how they intersect and in particular how the networks of related CNS-associated disorders overlap with autism may be an important way to begin to understand the genetics of different parts of the phenotypic spectrum of autism. Fig. 4 represents a first attempt at circumscribing subcomponents of the entire autism network that largely correspond to single related neurological disorders. This disorder-centric view of autism may eventually provide more direct clues to the genetic basis of the spectrum of behaviors indicative of autism. Qualitatively this figure reveals clusters of autism genes a majority of which are linked to a single autism-related neurological disorder, namely Mental Retardation, Tuberous Sclerosis, Angelman Syndrome, and Fragile X. This hints at the possibility that further comparative analysis may provide a way to understand the genotype–phenotype map for autism's diverse symptom spectrum. Also these circumscriptions could help to serve as a map to research done on related neurological disorders that may be directly relevant to our understanding of the etiology of autism. Future work, including the evaluation of more neurological disorders within Fig. 1 and disorders other than those that affect the CNS may help to rank and reorder genes that have been implicated in autism to date, and possibly reveal new genes worth investigating.

Materials and methods

Diseases and gene lists

We downloaded a complete set of 433 neurological disorders from the National Institute of Neurological Disorders and Strokes

Table 3
The 64 process-based candidates found to be significantly differentially regulated in autistic individuals when compared to controls

Gene	P value	Q value	Processes	Disorders
FN1	0	0	Cell migration	Hypoxia, Asperger syndrome
AFF2	0.0121	0.02294	Brain development	Fragile X, mental retardation
ANGPT1	0.0025	0.02294	Cell differentiation	Hypoxia
ATXN3	0.0124	0.02294	Nervous system development, synaptic transmission, transmission of nerve impulse	Ataxia, hypotonia
BMP2	0.0091	0.02294	Cell–cell signaling	Hypoxia
CHL1	0.0016	0.02294	Cell differentiation, nervous system development	Ataxia, microcephaly, mental retardation
DYRK1A	0.0157	0.02294	Nervous system development	Hypotonia, mental retardation
EDNRB	0.0043	0.02294	Nervous system development	Microcephaly, hypoxia
FHL1	0.0037	0.02294	Cell differentiation	Fragile X
FXN	0.0034	0.02294	Synaptic transmission, transmission of nerve impulse	Ataxia, fragile X, mental retardation
GNPTAB	0.005	0.02294	Cell differentiation	Hypotonia
GPM6B	0.0067	0.02294	Cell differentiation	Rett syndrome
HIF1A	0.0007	0.02294	Homeostasis	Hypoxia
ITGB1	0.0023	0.02294	Cell migration	Hypoxia
KCNMA1	0.015	0.02294	Ion transport, synaptic transmission, transmission of nerve impulse	Ataxia, hypoxia
MYOD1	0.0029	0.02294	Cell differentiation	Hypoxia
NRP1	0.0132	0.02294	Cell differentiation, cell–cell signaling	Hypoxia
OPHN1	0.007	0.02294	nervous system development	Ataxia, hypotonia, mental retardation
PAFAH1B1	0.0092	0.02294	Cell differentiation, ion transport, nervous system development	Microcephaly, mental retardation, tuberous sclerosis, hypotonia, spasticity tuberous sclerosis
RELN	0.0146	0.02294	Brain development	Ataxia, major depression, tuberous sclerosis, hypotonia, mental retardation
SDHD	0.0015	0.02294	Ion transport	Angelman syndrome
SLC1A1	0.0028	0.02294	Ion transport, synaptic transmission, transmission of nerve impulse	Ataxia, hypoxia
TGFB2	0.0137	0.02294	Generation of neurons, cell morphogenesis	Hypoxia
TP53	0.0079	0.02294	Cell differentiation	Ataxia, fragile X, hypoxia
ZIC2	0.0126	0.02294	Brain development, cell differentiation	Microcephaly
ANGPT2	0.0186	0.02376	Cell differentiation	Hypoxia
AR	0.0189	0.02376	Ion transport, cell–cell signaling	Ataxia, fragile X, hypoxia
NOS1	0.0182	0.02376	cell–cell signaling	Major depression, hypoxia
SIX3	0.0211	0.0252	Brain development	Microcephaly
CACNA1A	0.0224	0.02524	Ion transport, nervous system development, synaptic transmission, transmission of nerve impulse	Ataxia, major depression, mental retardation
PURA	0.0228	0.02524	Cell differentiation	Fragile X
ESR2	0.0241	0.02568	Cell–cell signaling	Hypoxia
ITGB2	0.0249	0.02568	Cell–cell signaling	Hypoxia
CREBBP	0.0266	0.02619	Homeostasis	Ataxia, hypoxia, mental retardation
ZEB2	0.0271	0.02619	Nervous system development	Microcephaly, mental retardation
ATP2A2	0.0285	0.0267	Ion transport	Hypoxia, mental retardation
S100B	0.0369	0.03335	Nervous system development	Major depression, hypoxia, mental retardation
HBA2	0.0405	0.03381	Ion transport	Major depression, hypoxia, tuberous sclerosis, mental retardation
PPT1	0.0403	0.03381	Brain development, nervous system development	Major depression, mental retardation, spasticity
TAL1	0.0399	0.03381	Cell differentiation	Hypoxia
FGF13	0.0495	0.03781	Nervous system development, cell–cell signaling	Hypoxia, mental retardation
MPZ	0.053	0.03781	Synaptic transmission, transmission of nerve impulse	Ataxia, infantile hypotonia, hypotonia, mental retardation
MYO5A	0.052	0.03781	Ion transport	Fragile X, hypotonia, mental retardation
PAX5	0.0533	0.03781	Cell differentiation	Microcephaly
RPS6KA3	0.0503	0.03781	Nervous system development	Ataxia, microcephaly, hypoxia, hypotonia, mental retardation
SLC1A3	0.0514	0.03781	Ion transport, synaptic transmission, transmission of nerve impulse	Ataxia
TSPAN32	0.0498	0.03781	Cell–cell signaling	Angelman syndrome
FOXG1	0.0571	0.03926	Brain development	Microcephaly
PAX8	0.0577	0.03926	Cell differentiation	Mental retardation
SIAH1	0.0619	0.04127	Cell differentiation	Hypoxia
PHEX	0.0637	0.04164	Cell–cell signaling	Hypotonia, mental retardation
ABCD1	0.0684	0.04385	Ion transport	Hypotonia, spasticity
SLC11A2	0.0707	0.04447	Ion transport	Ataxia, hypoxia
EDN2	0.0746	0.04489	Cell–cell signaling	Hypoxia
EIF2B2	0.0781	0.04489	Nervous system development	Ataxia, spasticity
ETFDH	0.078	0.04489	Ion transport	Hypotonia, mental retardation
PMP22	0.0776	0.04489	Synaptic transmission, transmission of nerve impulse	Ataxia, mental retardation, Infantile hypotonia, hypotonia, spasticity
RHOB	0.0741	0.04489	Cell differentiation	Hypoxia
AMH	0.0835	0.04563	Cell–cell signaling	Hypoxia
FLT1	0.0876	0.04563	Cell differentiation	Hypoxia
GPI	0.0866	0.04563	Nervous system development	Ataxia, hypoxia
GRIA3	0.0868	0.04563	Ion transport	Seizure disorder, Rett syndrome, mental retardation
SIM2	0.085	0.04563	Cell differentiation, nervous system development	Hypoxia, mental retardation
TTR	0.0847	0.04563	ion transport	Ataxia, fragile X, major depression

Student *t*-test *p* values and FDR-based *q* values indicate significance. The gene ontology biological processes and autism sibling disorders in which these genes are implicated are listed. The complete list of 154 process-based candidates is available as online Supplementary Table 2.

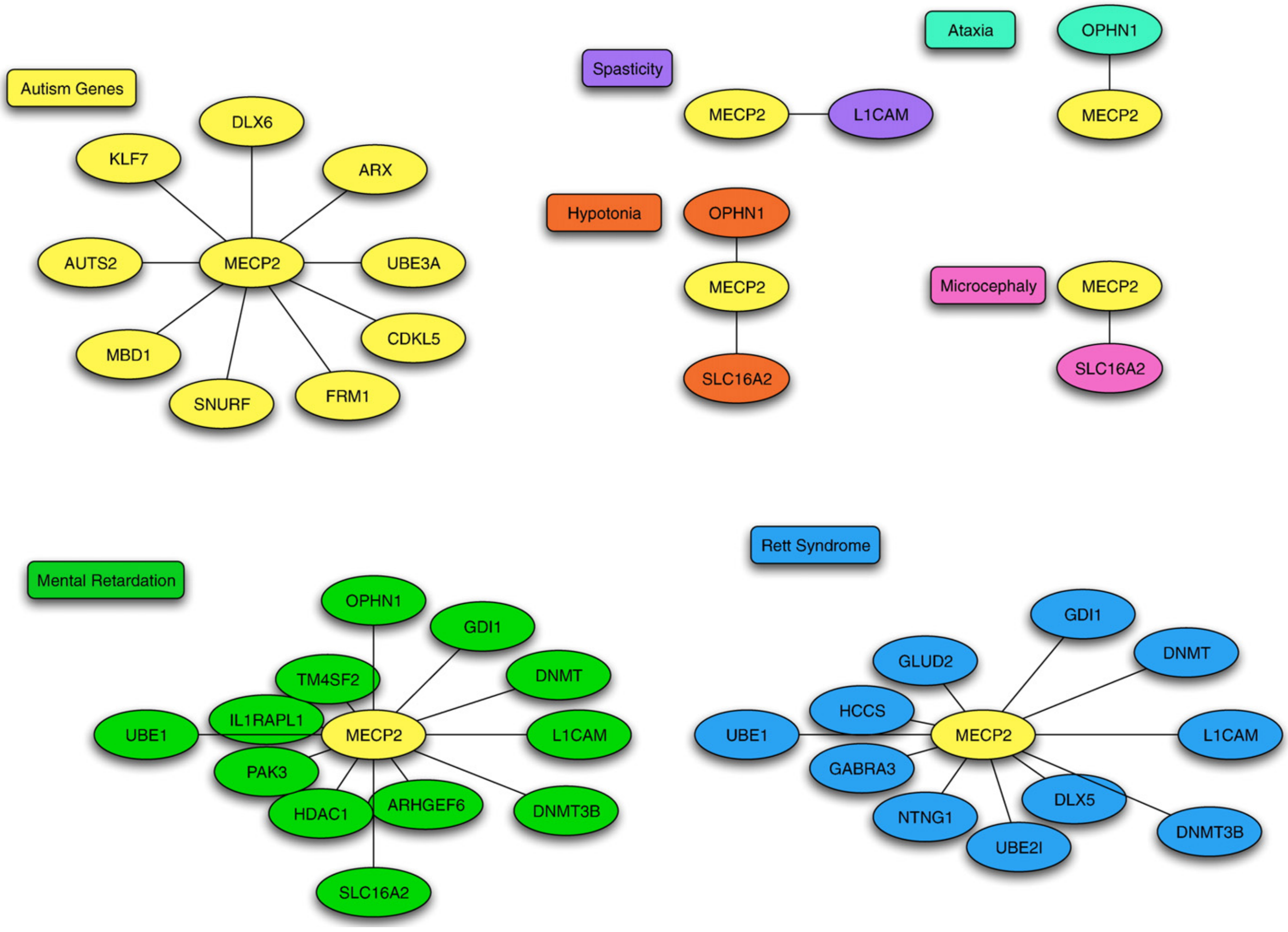

Fig. 3. Example of network-driven based strategy for identifying new autism gene candidates.

(NINDS) online database. NINDS treats disorders typically considered to be part of the Autism Spectrum Disorder as separate disorders. To minimize biases in our analysis, we opted to retain this circumscription "as-is". As a consequence, throughout this manuscript we use the term "autism" rather than "autism spectrum disorder". We then generated lists of candidate genes for each disorder by taking the union of genes returned from OMIM [26] and GeneCards [27]. A candidate is hereafter simply defined as a gene that is listed in either or both of these databases as associated with the disease term. Candidates may be based on linkage studies in families, linkage disequilibrium, or other sources (e.g. as described in ftp://ftp.ncbi.nih.gov/repository/OMIM/genemap.key). We computed the intersection of each disease gene list with the list for autism and ranked the results in descending order of number of shared genes. This allowed us to circumscribe a list of neurological disorders with the greatest number of genes in common with autism, resulting in a set of 40 diseases with possible molecular similarities to autism.

Disease relationship tree

The seed lists provided by OMIM and GeneCards were combined and transformed into a matrix of binary gene presence/absence with respect to each disease. The matrix was then analyzed using maximum parsimony in PAUP* [28] to reconstruct the relationships among the 41 neurological disorders. Distance based clustering approaches (neighbor joining and UPMGA) produced equivalent results.

Molecular network reconstruction

For each disorder and associated gene list, we used STRING (Search Tool for Retrieval of Interacting Genes/Proteins) version 6.3 [11], downloaded in December 2006, to construct networks from 5 separate lines of evidence: Conserved neighborhoods, Co-occurrence, Co-expression, Databases, and Text mining. The evidence involves: (1) synteny derived from SwissProt and Ensembl, (2) phylogenetic profiles derived from COG database [29,30], (3) coregulation of genes measured using microarrays imported from ArrayProspector [31], (4) validated small-scale interactions, protein complexes, and annotated pathways from BIND [32], KEGG [33] and MIPS [34], and (5) co-mention of gene names from PubMed abstracts. The networks were generated using the default settings in STRING, with a medium confidence score of 0.4. The lists of edges returned for each disorder were then imported into a relational database for subsequent analysis.

Biological process enrichment

Gene symbol identities corresponding to the PBC list were loaded into the ExPlain™ 2.3 Tool (www.biobase-international.com), which performs a Fisher's exact test to generate a p value for all biological processes containing 2 or more genes. To account for multiple testing, all p values were corrected using the Bonferroni adjustment [35]; each p value was adjusted by the number of biological processes in GO, which at the time of writing was 14648.

Table 4
The top 40 network-based candidates (NBC) sorted on the number of autism sibling disorders (see Fig. 1) in which the genes are implicated

Gene	*P* value	*Q* value	MDAG interactor	Disorders
CREBBP	0.0266	0.011509	*SLC6A4*	Ataxia, hypoxia, mental retardation
HPRT1	0.0353	0.012024	*FMR1,ADSL*	Ataxia, fragile X, mental retardation
RPS6KB1	0.0402	0.012267	*PTEN*	Ataxia, hypoxia, tuberous sclerosis
MYO5A	0.052	0.013362	*FMR1*	Fragile X, hypotonia, mental retardation
LAMA2	0.0557	0.013695	*DCX*	Microcephaly, hypotonia, mental retardation
MAP1B	0.0561	0.013695	*FMR1,FXR1,MAP2*	Fragile X, tuberous sclerosis, mental retardation
HNRNPK	0.0577	0.013969	*FMR1*	Fragile X, hypoxia, mental retardation
TTR	0.0847	0.016765	*APOE*	Ataxia, fragile X, major depression
GRIA3	0.0868	0.016897	*NTF4*	Seizure disorder, Rett syndrome, mental retardation
EIF4E	0.1003	0.018382	*MAP2*	Ataxia, hypoxia, tuberous sclerosis
ALAS2	0.108	0.019111	*MAOA,SLC40A1*	Ataxia, hypoxia, mental retardation
MTHFR	0.1179	0.020015	*APOE*	Microcephaly, major depression, mental retardation
GH1	0.1297	0.021108	*NF1*	Seizure disorder, Hypoxia, mental retardation
PIK3R1	0.1459	0.023355	*PTEN,TSC1,MET*	Ataxia, Hypoxia, tuberous sclerosis
DBH	0.149	0.023482	*TPH1,MAOA,TH,TSC1*	Major depression, tuberous sclerosis, mental retardation
CYFIP1	0.1764	0.024964	*FMR1,FXR1*	Angelman syndrome, fragile X, mental retardation
PTS	0.2004	0.026444	*PTEN*	Seizure disorder, hypotonia, mental retardation
EDN3	0.2304	0.029733	*ADSL,PAX3*	Microcephaly, hypoxia, mental retardation
TFRC	0.2655	0.033006	*SLC40A1,CD69*	Ataxia, major depression, hypoxia
MAPK14	0.2887	0.034661	*GLO1,GATA3*	Ataxia, hypoxia, Rett syndrome
SLC6A8	0.296	0.034976	*FMR1*	Fragile X, hypotonia, mental retardation
TBX1	0.3729	0.041067	*PAX3*	Major depression, Asperger syndrome, mental retardation
ATL1	0.4141	0.044273	*FMR1,GABRB3*	Ataxia, mental retardation, spasticity
L1CAM	0.4366	0.046007	*FMR1,MECP2,DCX*	Mental retardation, Rett syndrome, spasticity
ATM	0.0097	0.009728	*PTEN*	Ataxia, microcephaly, hypoxia, mental retardation
OGDH	0.0062	0.009728	*ABAT*	Hypoxia, Rett syndrome, hypotonia, mental retardation
PQBP1	0.013	0.009728	*SLC6A4*	Ataxia, microcephaly, mental retardation, spasticity
DMD	0.043	0.012724	*DCX*	Hypoxia, infantile hypotonia, mental retardation
PSEN1	0.0519	0.013362	*APOE*	Ataxia, major depression, hypoxia, spasticity
BDNF	0.0643	0.014489	*MAP2,GRIN2A,NTF4,TH,SLC6A4*	Major depression, hypoxia, Rett syndrome, mental retardation
FRAP1	0.0951	0.017973	*PTEN,TSC1*	Ataxia, seizure disorder, hypoxia, tuberous sclerosis
FLNA	0.2635	0.033006	*DCX,TSC1*	Microcephaly, hypoxia, tuberous sclerosis, mental retardation
NP	0.4431	0.046524	*ADSL*	Ataxia, seizure disorder, hypotonia, spasticity microcephaly, mental retardation
PAFAH1B1	0.0092	0.009728	*ARX,DAB1,DCX,TSC1*	Tuberous sclerosis, hypotonia, spasticity microcephaly, seizure disorder
HADHA	0.0204	0.011078	*ACADL*	Infantile hypotonia, hypotonia, mental retardation
RPS6KA3	0.0503	0.013362	*ARX,GRPR*	Ataxia, microcephaly, hypoxia, hypotonia, mental retardation
RB1	0.0662	0.014495	*PTEN,NF1*	Ataxia, major depression, hypoxia, tuberous sclerosis, mental retardation
EMX2	0.1834	0.025706	*DCX,TSC1*	Microcephaly, mental retardation, tuberous sclerosis, hypotonia, spasticity
ATRX	0.3561	0.039603	*SNRPN*	Ataxia, microcephaly, mental retardation, hypotonia, spasticity
SLC17A5	0.3569	0.039603	*TH*	Ataxia, mental retardation, infantile hypotonia, hypotonia, spasticity

NBC genes are directly connected to a multi-disorder autism gene, but not yet implicated in autism. The MDAG gene(s) to which the NBC gene interacts is provided. Student *t*-test *p* value and FDR-based *q* value are provided to indicate the extent to which these genes are differentially regulated in autistic individuals when compared to controls. A total of 289 NBC genes were found to be differentially expressed. A complete list of the NBC genes is available online in Supplementary Table 3.

Expression analysis

From Gene Expression Omnibus (GEO) we downloaded GSE6575 [16,36]. This dataset consisted of 17 samples of autistic patients without regression, 18 patients with regression, 9 patients with mental retardation or developmental delay, and 12 typically developing children from the general population; total RNA was extracted from whole blood samples using the PaxGene Blood RNA System according the manufacturer's specifications and run on Affymetrix U133plus2.0. For the purposes of the present study, we elected to use only the 35 autistic patient samples and 12 control samples from the general population. All preprocessing and expression analyses were done with the Bioinformatics Toolbox Version 2.6 (For Matlab R2007a+). GCRMA was used for background adjustment and control probe intensities were used to estimate non-specific binding [37]. Housekeeping genes, gene expression data with empty gene symbols, genes with very low absolute expression values and genes with a small variance across samples were removed from the preprocessed dataset. We then conducted a preliminary analysis to determine the difference in signal between the two groups of autistic individuals, autistics with and without regression (early onset autism). When compared to the 12 control samples using a *t*-test, we learned that the *p* value distribution for

Table 5
Enriched biological processes for 9 genes found by both the process-and network-based analytical strategies

Gene ontology biological process	*P* value	Genes
Establishment of localization	1.02E-04	*FLNA, SLC16A2, PAFAH1B1, AR, MYO5A, OPHN1, FXN, SLC6A8, FLNA, SLC16A2, PAFAH1B1, AR, MYO5A*
Localization	1.05E-04	*OPHN1, FXN, SLC6A8*
Nervous system development	0.001619529	*FLNA, PAFAH1B1, L1CAM, OPHN1*
Cell organization and biogenesis	0.005608511	*FLNA, PAFAH1B1, AR, MYO5A, OPHN1*
Locomotion	0.00673254	*FLNA, PAFAH1B1, OPHN1*
Localization of cell	0.00673254	*FLNA, PAFAH1B1, OPHN1*
Cell motility	0.00673254	*FLNA, PAFAH1B1, OPHN1*
Cytoskeleton organization and biogenesis	0.018371497	*FLNA, PAFAH1B1, MYO5A*
Cell communication	0.021262073	*FLNA, PAFAH1B1, AR, OPHN1, FXN, SLC6A8*
Cell differentiation	0.028541906	*PAFAH1B1, L1CAM, OPHN1*
Cell–cell signaling	0.031948026	*AR, FXN, SLC6A8*
Transport	0.047207875	*SLC16A2, AR, MYO5A, FXN, SLC6A8*

The *p* value is based on Fisher's exact test and was not corrected for multiple tests. All 9 genes occur in 3 or more autism sibling disorders and were found to be significantly differentially expressed in autistic patients.

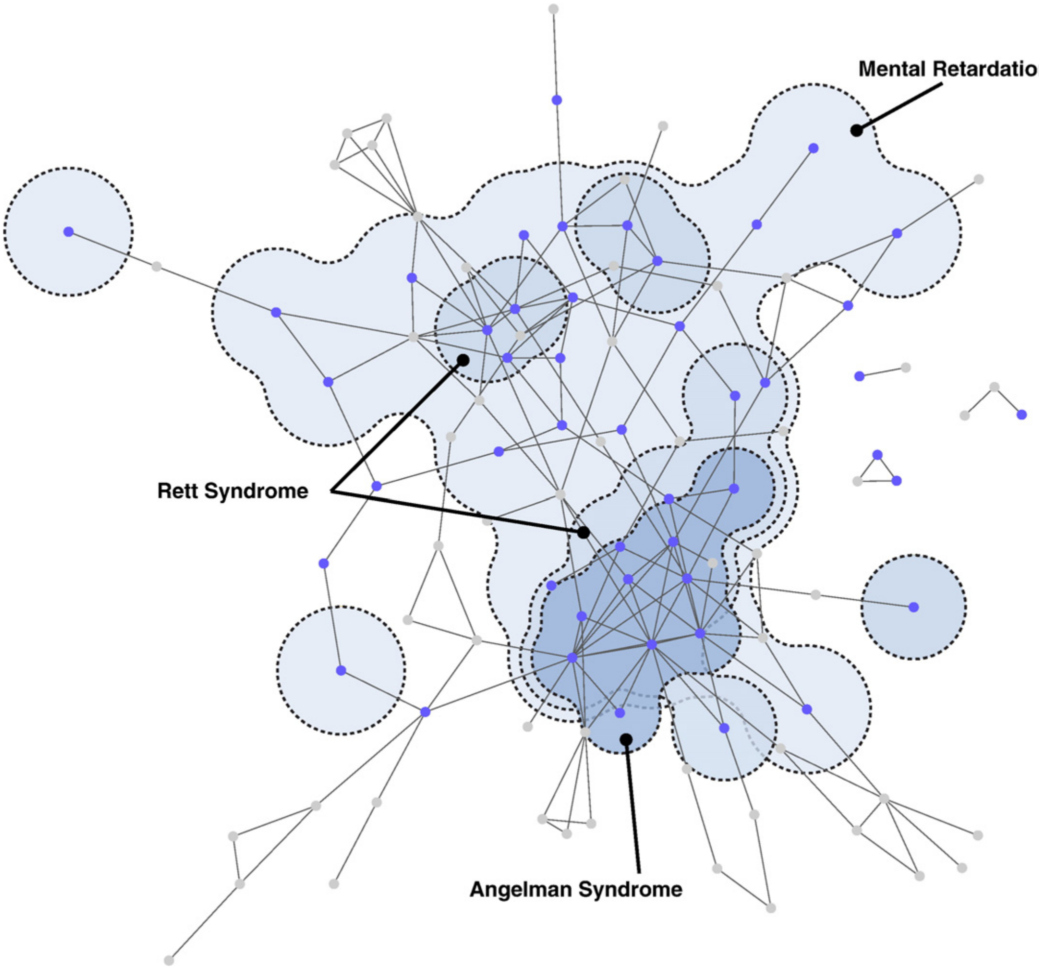

Fig. 4. The autism network redrawn in light of its intersection with the autism sibling disorders. Several obvious delineations are highlighted, namely Mental retardation, Angelman's syndrome, and Rett syndrome.

the autistic patients with regression was flat and therefore non-informative. Thus, throughout the present study we used only the 17 samples from autistic individuals without regression (also referred to as early onset autism). Correction for multiple tests was done by calculating q values, a measure of significant in terms of the false discovery rate [38].

Acknowledgment

This project was supported in part by a grant from the National Science Foundation DBI 0543480 awarded to Dennis P. Wall. We are grateful to two anonymous reviewers for improving the quality of the manuscript, to J. P. Gregg for early access to microarray data, and to the Kunkel Lab, Ally Eran, and the Boston Autism Consortium for helpful discussions.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ygeno.2008.09.015.

References

[1] J. Sebat, et al., Strong association of De Novo copy number mutations with autism, Science 316 (5823) (2007) 445–449.

[2] L. Marcotte, P.B. Crino, The neurobiology of the tuberous sclerosis complex, Neuromolecular Med. 8 (4) (2006) 531–546.

[3] V. Wong, Study of the relationship between tuberous sclerosis complex and autistic disorder, J. Child Neurol. 21 (3) (2006) 199–204.

[4] M.A. Manning, et al., Terminal 22q deletion syndrome: a newly recognized cause of speech and language disability in the autism spectrum, Pediatrics 114 (2) (2004) 451–457.

[5] P. Moretti, H.Y. Zoghbi, MeCP2 dysfunction in Rett syndrome and related disorders, Curr. Opin. Genet. Dev. 16 (3) (2006) 276–281.

[6] R.J. Hagerman, Lessons from fragile X regarding neurobiology, autism, and neurodegeneration, J. Dev. Behav. Pediatr. 27 (1) (2006) 63–74.

[7] K. Shinahara, et al., Single-strand conformation polymorphism analysis of the FMR1 gene in autistic and mentally retarded children in Japan, J. Med. Invest. 51 (1–2) (2004) 52–58.

[8] L. Sam, Y. Liu, J. Li, C. Friedman, Y.A. Lussier, Discovery of protein interaction networks shared by diseases, Pac. Symp. Biocomput. (2007) 76–87.

[9] J. Loscalzo, I.S. Kohane, A.L. Barabasi, Human disease classification in the postgenomic era: a complex systems approach to human pathobiology, Mol. Syst. Biol. 3 (2007) 124.

[10] J. Lim, et al., A protein–protein interaction network for human inherited ataxias and disorders of Purkinje cell degeneration, Cell 125 (4) (2006) 801–814.

[11] C. von Mering, et al., STRING: known and predicted protein–protein associations, integrated and transferred across organisms, Nucleic Acids Res. 33 (Database issue) (2005) D433–D437.

[12] K. Gerrow, A. El-Husseini, Cell adhesion molecules at the synapse, Front Biosci. 11 (2006) 2400–2419.

[13] K. Nishimura, et al., Genetic analyses of the brain-derived neurotrophic factor (BDNF) gene in autism, Biochem. Biophys. Res. Commun. 356 (1) (2007) 200–206.

[14] P. Poo-Arguelles, et al., X-Linked creatine transporter deficiency in two patients with severe mental retardation and autism, J. Inherit Metab. Dis. 29 (1) (2006) 220–223.

[15] J.E. Crandall, et al., Dopamine receptor activation modulates GABA neuron migration from the basal forebrain to the cerebral cortex, J. Neurosci. 27 (14) (2007) 3813–3822.

[16] J.P. Gregg, et al., Gene expression changes in children with autism, Genomics 91 (1) (2008) 22–29.

[17] J.Y. Chen, C. Shen, A.Y. Sivachenko, Mining Alzheimer disease relevant proteins from integrated protein interactome data, Pac. Symp. Biocomput. (2006) 367–378.

[18] M.E. Cusick, et al., Interactome: gateway into systems biology, Hum. Mol. Genet. Spec No. 2 (2005) R171–R181.

[19] F. Giorgini, P.J. Muchowski, Connecting the dots in Huntington' s disease with protein interaction networks, Genome Biol. 6 (3) (2005) 210.

[20] H. Goehler, et al., A protein interaction network links GIT1, an enhancer of Huntington aggregation, to Huntington' s disease, Mol. Cell 15 (6) (2004) 853–865.

[21] S. Humbert, F. Saudou, The ataxia-ome: connecting disease proteins of the cerebellum, Cell 125 (4) (2006) 645–647.

[22] M.G. Kann, Protein interactions and disease: computational approaches to uncover the etiology of diseases, Brief Bioinform. 8 (5) (2007) 333–346.

[23] O. Alter, P.O. Brown, D. Botstein, Singular value decomposition for genome-wide expression data processing and modeling, Proc Natl Acad Sci U S A. 97 (2000) 10101–10106.

[24] A.J. Butte, J. Ye, H.U. Häring, M. Stumvoll, M.F. White, I.S. Kohane, Determining significant fold differences in gene expression analysis, Pac Symp Biocomput. (2001) 6–17.

[25] W.P. Kuo, T.K. Jenssen, A.J. Butte, L. Ohno-Machado, I.S. Kohane, Analysis of matched mRNA measurements from two different microarray technologies, Bioinformatics 18 (2002) 405–412.

[26] http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=OMIM.

[27] http://www.genecards.org/index.shtml.

[28] D.L. Swofford, PAUP* Phylogenetic Analysis Using Parsimony (*and Other Methods), Sinauer Associates, Sunderland, Massachusetts, 2002.

[29] R.L. Tatusov, E.V. Koonin, D.J. Lipman, A genomic perspective on protein families, Science 278 (5338) (1997) 631–637.

[30] R.L. Tatusov, et al., The COG database: a tool for genome-scale analysis of protein functions and evolution, Nucleic Acids Res. 28 (1) (2000) 33–36.

[31] L.J. Jensen, et al., ArrayProspector: a web resource of functional associations inferred from microarray expression data, Nucleic Acids Res. 32 (Web Server issue) (2004) W445–W448.

[32] G.D. Bader, et al., BIND—The Biomolecular Interaction Network Database, Nucleic Acids Res. 29 (1) (2001) 242–245.

[33] M. Kanehisa, et al., The KEGG resource for deciphering the genome, Nucleic Acids Res. 32 (Database issue) (2004) D277–D280.

[34] http://mips.gsf.de/.

[35] T.K. Rice, N.J. Schork, D.C. Rao, Methods for handling multiple testing, Adv. Genet. 60 (2008) 293–308.

[36] http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE6575.

[37] Z. Wu, et al., A model based background adjustment for oligonucleotide expression arrays, J. Amer. Stat. Assoc. 99 (468) (2004) 909–917.

[38] J.D. Storey, R. Tibshirani, Statistical significance for genomewide studies, Proc. Natl. Acad. Sci. U. S. A. 100 (16) (2003) 9440–9445.