



Ethanol/alcohol pathways along with immune, cation and development may play a role in autism spectrum disorder

Vishal Saxena^{1,2,3,4,5}

Correspondence: vishal@mit.edu

¹Department of Medicine, Massachusetts General Hospital, Boston, Massachusetts, USA.

²Department of Neurology, Massachusetts General Hospital, Charlestown, Massachusetts, USA.

³Harvard Medical School, Boston, Massachusetts, USA.

⁴Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA.

⁵Children's Hospital, Boston, Massachusetts, USA.

Abstract

Introduction: Autism spectrum disorder or ASD was studied via copy number variations or CNVs and via windows straddling LoGS loci. We found strong concordance between our prior findings relating to development. Further, alcohol pathways were discovered.

Methods: CNVs previously described in autism were studied and use of pathway overlaps in deducing mechanism is described.

Results: CNVs related to autism were split into those falling within LoGS loci windows and those outside. Those inside related to immune function while those outside to alcohol pathways. Further, the enrichment of genes within windows centered on the loci showed mechanistic insights into autism.

Discussion: Our analysis provides a mechanistic framework to define detailed mechanisms underlying autism. The found ethanol/alcohol pathways rank topmost in the analysis conducted and provide another mechanistic insight into autism. Further the CNVs that contain such alcohol dehydrogenase or detoxifying genes show a loss in copy number as we would predict because alcohol has been known to affect the developing brain.

Keywords: Autism spectrum disorder, alcohol, immune, cation, development

Introduction

Autism is an important disorder whose incidence recently has been reported to be rising. However, its etiology is still unknown although several avenues are currently under exploration.

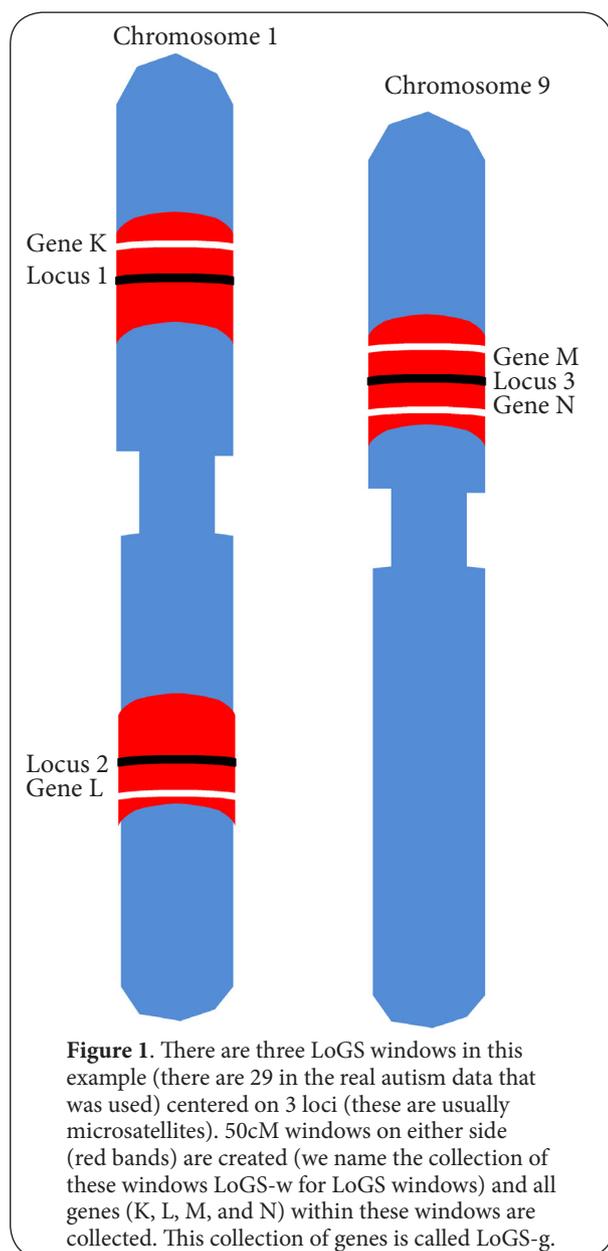
Over the recent few years, pathway analysis has emerged as a powerful tool in the elucidation of disease mechanism, especially diseases that have a complex etiology. In contrast to simple Mendelian disorders, complex disorders are not characterizable through a single gene defect but rather their etiology is a result of the interplay between several genes which may sometimes not even be in the same spatial domain [1].

Genetic linkage studies have proven quite successful in identifying genes involved in simple disorders such as Huntington's disease, cystic fibrosis and others. However, it has been thought that these analyses were not as successful in identifying etiologic factors involved in complex disorders and the guiding rationale for such pessimism has been the lack of coherence in the loci identified (even though the discovery of several loci should come as no surprise given that complex disorders are by definition not driven by a single gene defect).

A population based rationale underpins our analysis. Autism is clearly diagnosable with great accuracy and the symptom set that defines it is common across different people with this condition (albeit with different levels of severity). We therefore rationalized that different parts of a population affected by

autism or any other (especially complex but generally any) condition will point in some way to a common mechanism. Such a mechanism can coalesce around a pathway, or even at a network or systems level. Thus, by looking for connectivity across pathways, we have a way to obtain different contributions or signals that lead to this condition. Since a genetic perspective is likely to be common within families, we thus have different pedigrees or families giving us different ways to identify such a set of signals. When each signal is seen in isolation, it shows a weak strength and is not necessarily discernible, because in a complex disorder, subtle signals from several interacting factors such as genes work together in producing the disease. However, when such signals are taken from different pedigrees and families, the signals add up in a sense giving us a much larger signal. For a pathway based analysis such a signal increase is proportional to the square root of the size of the studied pathway [2].

In a recent paper [3], we applied a new methodology called Linkage ordered Gene Sets (LoGS) that takes as its input loci identified via pedigree analyses by several prior investigators and looks for coherence at the pathway level. The rationale for such a view is that complex disorders such as autism (or rather autism spectrum disorder) are unified in their range of phenotypes (given by levels of severity) with very strong concordance between experts diagnosing individuals with



this disorder. Such phenotypic concordance in phenotypes (albeit within different severity levels described by a range within autism spectrum disorder) in our view is driven by a shared mechanism in these individuals. Each pedigree in a linkage analysis or each CNV analysis may point to some subtle mediator of such conditions while other pedigrees may point to other mediators. Because of phenotypic concordance, we have reasoned that such mediators will converge either at the pathway level or groups of pathways level.

One of the inputs in our LoGS analysis is the group of genes that reside within 50cM distance of all loci linked to autism at a LOD significance of 3 or higher (LOD stands for logarithm of the odds and is a measure of the strength of genetic linkage

where a score of 3 is considered quite significant for genetic linkage). Using such genes (and as explained in our recent paper [3]), LoGS derives a ranked set of pathways. We wanted to analyze these genes separately of LoGS to see if further mechanistic insights could be gleaned from such genes.

The results of pathway analysis are often presented as a ranked list and no attempt is generally made to show connectivities across the pathways found. By looking at what unites these pathways, mechanisms in a disease may be inferable. We analyzed in this paper CNVs in autism and also genes derived via windows straddling LoGS loci in autism (see methods). We found a bifurcation in the mechanisms inferable from the CNVs that have in the literature shown a link to autism.

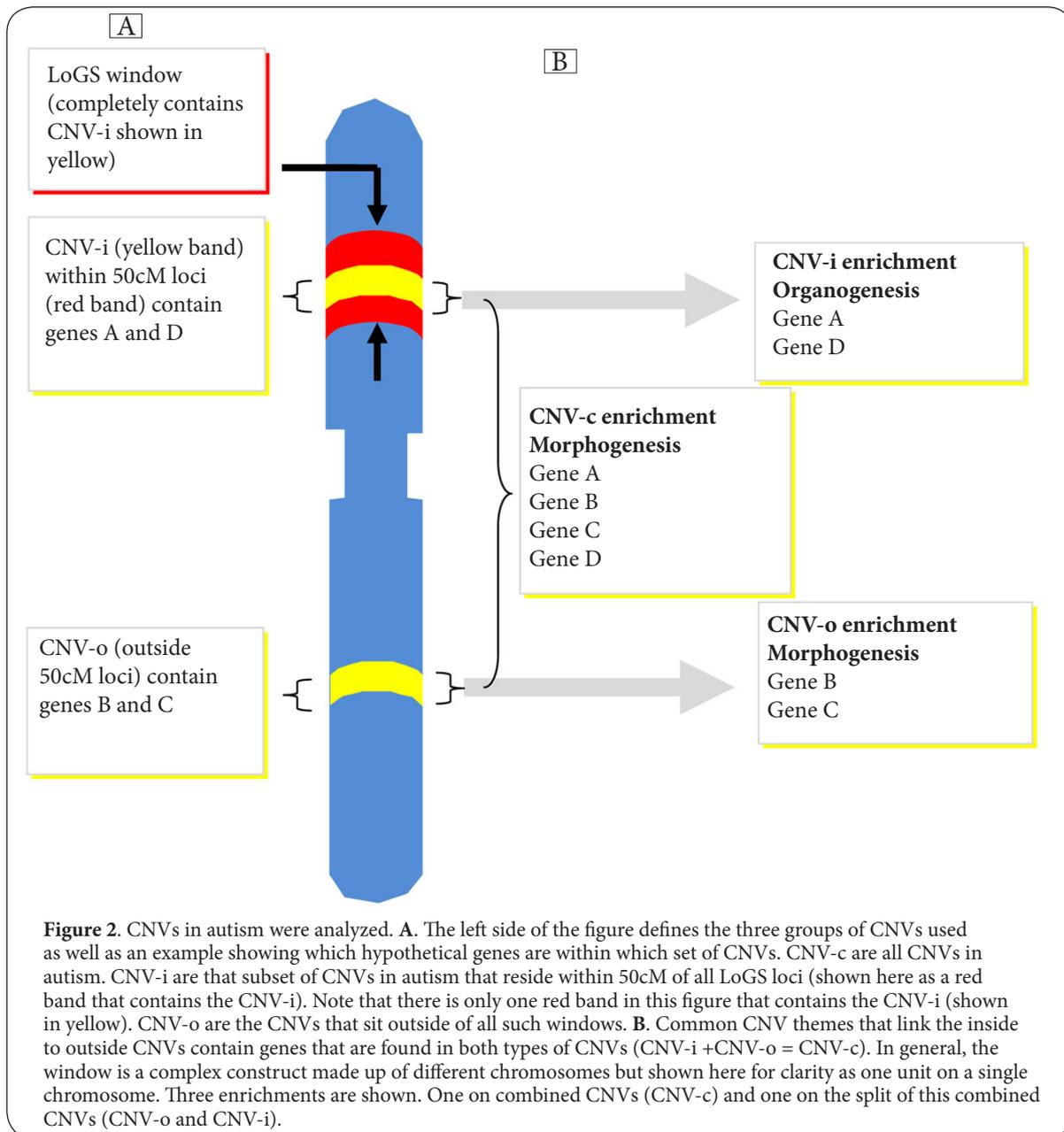
So that some of the terms are clear at the outset, we are overall looking at CNVs and identified genomic loci in autism. In a recent paper [3], we have used windows in the genome (see Figure 1). We explain these windows in more detail in the methods. Briefly, we take loci implicated in autism and throw windows of 50cM on either side of such loci. These windows collectively are termed LoGS-w and the genes within them are LoGS-g. We also take CNVs in autism. See Figure 2A for clarification on the CNVs. All CNVs in autism taken together are labeled as CNV-c (for combined). These CNVs can be split into two groups: those that reside within LoGS-w (CNV-i) and those CNVs that sit outside LoGS-w (CNV-o). Both of these groups taken together define CNV-c. We are thus essentially collecting four overall regions in the genome (each region can be spread over several chromosomes): The LoGS-w, CNV-c, CNV-i, and CNV-o. All genes within each region are collected and EASE enrichments performed (EASE finds overrepresented categories of genes [5]). Some of these regions are mutually exclusive (such as CNV-i, and CNV-o) while other regions can overlap (CNV-i by definition falls within LoGS-w). Therefore some genes may find themselves in more than one genomic region or window type.

Methods

Obtaining the 50cM windows used (or the LoGS-w)

In our previous paper, we introduced LoGS, a method that takes various loci implicated in autism and then throws a net of 50cM on either side of all such loci to generate first genes ranked by distance from such loci and then pathways ranked by their distance from such loci. To better understand such windows, we briefly describe LoGS to better contextualize such windows.

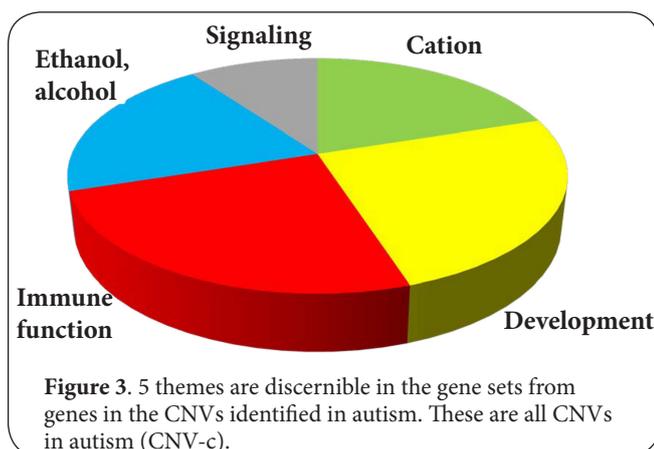
Pedigree analysis (or even genomewide association analysis) takes families (or data with cases and controls) and obtains locations on the genome that link (or associate) with a disease in question (such as autism). Such analyses work quite successfully in Mendelian disorders, where a single gene is involved. However, in complex disorders, by definition, we may have the involvement of several genes (in addition to the environment). Each pedigree analysis may point



to a different location on the genome. A combined signal taking into account these various regions will shed light on relevant mechanisms in the disease. Such a combined signal should exist, because the various signals (or locations) in the genome should aggregate into a common mechanism, since all diseases (complex or simple) have common symptoms and progression. We rationalize that such a common signal occurs at the level of pathways (or groups of pathways or themes). We thus, rather than looking for the most proximal genes to a locus, look for the most proximal pathways from all loci implicated in autism. To derive such pathways, we take all genes that could have any linkage to the loci implicated

in autism. Linkage ends at 50cM from any location.

LoGS coalesces all implicated loci and then ranks pathways from such coalesced loci. The first step in LoGS is to take all implicated loci above a threshold significance of a LOD score of 3 or higher and then to define windows of 50cM on either side of such loci (the loci are usually microsatellites and we take for each microsatellite its point location defined by the average of the start and end points). All genes are identified within such windows. Next for each gene, its distance is divided by the LOD score of the closest locus to that gene. All genes are then ranked and a pathway rank is obtained. For further details along with the listing of loci used, see [3].



In this paper, we used the genes from all windows described above but we did not obtain the enriched pathways via a distance metric but rather via overrepresentation of pathways within the collective set of those genes.

EASE enrichment on the CNV-i and CNV-o

We took the CNVs that have been implicated in autism (and that were used in our prior study [3]) and then searched for those CNVs in this group that were within LoGS-w. These gave us the CNV-i. We then took the genes that were only in CNV-i and enriched them using EASE (Expression Analysis Systematic Explorer), a software for finding overrepresented pathways in a given set of genes. EASE finds significance of pathways within a group of genes using the hypergeometric distribution [5]. We took all CNVs that showed any overlap with the 50cM windows from each locus.

We then repeated this analysis on the CNVs that were outside of the 50cM LoGS windows (CNV-o).

Overlaps analysis on the top 10 pathways of the LoGS-w

All genes within 50cM of all autism related loci (these genes are called LoGS-g) were input into EASE and the most enriched pathways ranked. The top ranked pathway because it was large was used as the baseline and the remaining 9 pathways were ranked according to their percent overlap with this pathway (cellular process). This percent overlap is defined as number of genes overlapping divided by the size of the other pathway. For example, the overlap of cell communication is 1022 and the size of this gene set is 1022, giving a percent overlap of 100%. When more than one gene set had 100% overlap, the larger gene set was given the higher priority of overlap with cellular process. In this way, successive overlaps were obtained with cellular process. A graphical display of such overlaps helps to visualize the sequence of overlaps and such overlaps help define a mechanism in autism with increasing levels of detail.

Separately, genes that overlapped across all 10 gene sets were identified (GRIK1 and GRIN1).

Top 50 pathways from LoGS-w grouped into 5 themes

We looked at the top 50 pathways to see if we could discern gene sets that show repeating themes. This was done to see if more global patterns of pathway overlap could show similarity to the top 10 pathways or possibly shed further insights into mechanism. Five such themes were identified by inspection (not all pathways were collapsible into such themes). We then ran the analysis (similar to the top 10 analysis above) to find genes that were common. We also looked for the most common pathway repeats. Such maximum repeats were obtained by cation/calcium pathways.

Results

Five themes occur in the CNV pathway analysis

We have recently [3] described the top 20 results from EASE run over genes from CNVs identified (CNV-c) in autism (there were 2210 genes). EASE is a software platform to find over-represented themes or pathways in a given group of genes. We took the results previously described [3] and looked for common themes and we found 5 overall themes that the top 20 pathways fit into: Immune function, development, cation, signaling, and ethanol. Please see (Supplementary Table S1) in Saxena [3] which shows the top 20 themes from the CNV genes while (Figure 3) outlines the overall groups of pathways found.

CNVs outside LoGS windows (CNV-o) point to alcohol pathways while those within (CNV-i) point to immune pathways

We found in our recent paper a strong concordance in pathways discovered from CNVs and ranked via our novel LoGS analysis [3]. However, the concordance unifies two ways to obtain significance. EASE looks for enrichment in pathways based on a given set of genes. This is a probabilistic argument that does not rely on location while LoGS is a location based method. That there is concordance between the two results suggests that two different measures of significance match and that the most significant EASE gene sets are also very proximal to the LoGS loci (please see methods section "Obtaining the 50cM windows used" as well as [3] for more details on LoGS). However, there are many more genes (outside of the top gene sets) that are also within CNVs and the location of those genes does not have to fall within the LoGS loci windows.

Thus, we next tested to see what fraction of the CNV-c's previously described in autism [4] overlap with LoGS-w and what fraction sit outside them to see if we can discern differences in mechanism (CNVs in general may not reside within 50cM of previously described loci in autism). We found that out of 356 CNVs, 141 are within LoGS-w while the remaining 215 are outside LoGS-w. We next looked for the genes within each group and looked for enrichment of these genes for possible mechanism in autism. There are 1556 genes in the CNV-o and there are 546 genes in CNV-i (see Supplementary Table S1 for lists of these genes).

Table 1. CNV genes outside of the 50cM windows of LoGS loci. Shown are the top 20 categories from EASE.

Pathway	GO	P
1 Ethanol Metabolism	BP	5.10 ⁻⁶
2 Ethanol Oxidation	BP	5.10 ⁻⁶
3 Copper ion homeostasis	BP	3.10 ⁻⁵
4 Alcohol dehydrogenase activity/zinc-dependent	MF	3.10 ⁻⁶
5 Heavy metal sensitivity/resistance	BP	10 ⁻⁴
6 Alcohol dehydrogenase activity	MF	10 ⁻⁴
7 MAP kinase activity	MF	4.10 ⁻⁴
8 Organogenesis	BP	5.10 ⁻⁴
9 Transition metal ion binding	MF	10 ⁻³
10 Morphogenesis	BP	10 ⁻³
11 Copper/cadmium binding	MF	2.10 ⁻³
12 Oxidoreductase activity	MF	2.10 ⁻³
13 Oxidoreductase activity/acting on the CH-CH group of donors	MF	2.10 ⁻³
14 Protein Modification	BP	2.10 ⁻³
15 Cell Communication	BP	3.10 ⁻³
16 JNK Kinase	BP	3.10 ⁻³
17 Copper Ion Binding	MF	3.10 ⁻³
18 SAP Kinase Activity	MF	4.10 ⁻³
19 Transition metal ion homeostasis	BP	4.10 ⁻³
20 Alcohol Metabolism	BP	4.10 ⁻³

P = EASE score P value. GO = gene ontology category;
 BP = biological process; MF = molecular function;
 CC = cellular component (there are no CC categories in this set).

Table 2. CNV genes inside the 50cM windows of LoGS loci. Shown are the top 20 categories from EASE.

Pathway	GO category	P
1 Interferon-alpha/beta receptor binding	MF	2.10 ⁻¹²
2 Hematopoietin/interferon-class (D200-domain) cytokine receptor binding	MF	2.10 ⁻¹¹
3 Cytokine activity	MF	4.10 ⁻⁹
4 Response to virus	BP	10 ⁻⁷
5 Receptor binding	MF	7.10 ⁻⁷
6 Antiviral response protein activity	MF	3.10 ⁻⁶
7 Extracellular	CC	3.10 ⁻⁵
8 Cell-cell signaling	BP	3.10 ⁻⁵
9 Signal transducer activity	MF	2.10 ⁻⁴
10 Chemokine receptor binding	MF	2.10 ⁻³
11 Chemokine activity	MF	2.10 ⁻³
12 Neurogenesis	BP	3.10 ⁻³
13 Chemoattractant activity	MF	3.10 ⁻³
14 G-protein-coupled receptor binding	MF	3.10 ⁻³
15 Calcium ion homeostasis	BP	3.10 ⁻³
16 Tight junction	CC	3.10 ⁻³
17 Extracellular space	CC	6.10 ⁻³
18 Organogenesis	BP	7.10 ⁻³
19 Intercellular junction	CC	8.10 ⁻³
20 Defense response	BP	10 ⁻²

P = EASE score P value. GO = gene ontology category;
 BP = biological process; MF = molecular function;
 CC = cellular component.

Enriched pathways as above were obtained using EASE and the results of finding overrepresented pathways are shown in (Figure 4). We found a bifurcation in the themes these two

groups of CNVs show with all immune pathways now moving into CNV-i solely while all alcohol pathways (and heavy metal pathways) have moved into the CNV-o (with immune pathways scoring highest in the CNV-i group while alcohol pathways score highest in the CNV-o group). Our prior immune gene sets are very similar but not completely the same and all immune gene sets have moved up in their ranks in the CNV-i (compared with their ranks in CNV-c). There are also now 4 additional immune pathways in the top 20 (bringing the total to 9 of 20). The alcohol pathways in the bifurcated analysis also show a rise in rank and also show an increase in number from 4 to 5 within the top 20. The top 20 ranked pathways outside the LoGS loci and within the LoGS loci windows are presented in (Tables 1 and 2) respectively with more details in (Supplementary Table S2 and Supplementary Table S3) respectively.

Common pathways linking CNV-i to CNV-o

We rationalized that since autism is one disorder, there must be some links between the pathways obtained from CNV-o and from CNV-i. To find such pathways, we looked for pathways in the common CNV (CNV-c) results (in other words, all genes within all CNVs were analyzed via EASE and the pathways obtained—see Supplementary Table S1 in Saxena [3]) that had genes in CNV-o and CNV-i pathways (See Figure 2). Of the top 20 CNV-c pathways, 10 pathways showed significant overlap across both CNV-i and CNV-o, while both the viral pathways showed one overlap each with the CNV-o gene sets (while having similarly named themed sets in CNV-i). The overlapping genes from these common viral pathways are TLR3 from the response to virus gene set in CNV-c matching with cell communication in CNV-o and RNF4 from the antiviral response protein activity matching with transition metal ion binding gene set in CNV-o.

The pathways that showed large overlaps are: organogenesis, morphogenesis, cell-cell signaling, Neurogenesis, skeletal development, metal ion homeostasis, map kinase activity, development, cytokine activity, and cation homeostasis. We note that these pathways overall relate to development and cations. The gene sets showing no overlap are: Interferon-alpha/beta receptor binding, Hematopoietin/interferon-class, ethanol oxidation, ethanol metabolism, copper ion homeostasis, Alcohol dehydrogenase activity (zinc-dependent), Heavy metal sensitivity/resistance, and Alcohol dehydrogenase activity. Thus, a majority of the immune pathways and all the alcohol pathways show no overlap (although in the outside group, CNV-o, we have a new alcohol pathway—alcohol metabolism—and it does show overlap with morphogenesis). Figure 5 summarizes these findings. To find pathways that link outside to inside CNVs we simply looked for those common CNV pathways that had members in both the outside and inside groups. To find those gene sets within the CNVs inside the LoGS windows (CNV-i) and outside (CNV-o) that linked to the other, we looked for overlaps with the CNV-c pathways.

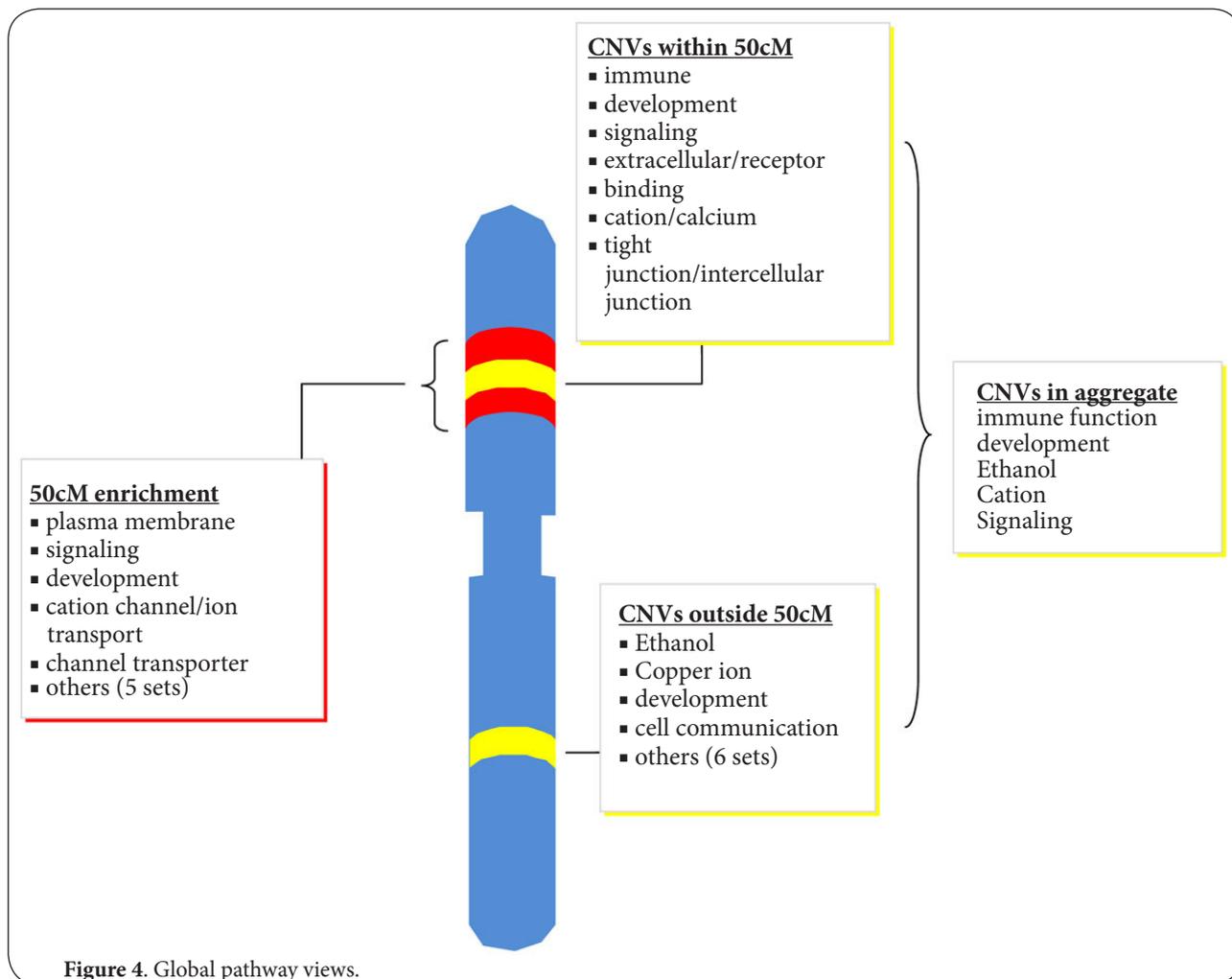


Figure 4. Global pathway views.

The cations we are discovering in the two groups (CNV-o and CNV-i) relate to different types of activities. One set of cations is involved in copper and other heavy/transition metal function, while the other set (in the LoGS loci CNVs) is involved in calcium function and/or signaling. This separation becomes clearer with the identification of CNV-i and CNV-o.

Further, we note that several gene sets are connecting across the two CNV groups from the combined group (as just mentioned above). However, many genes are overlapping from one gene set in one group of CNVs to another set in the same group of CNVs showing an interlinking within or intra of that CNV group (for example, IL11 is found in several pathways relating to immune function as well as development).

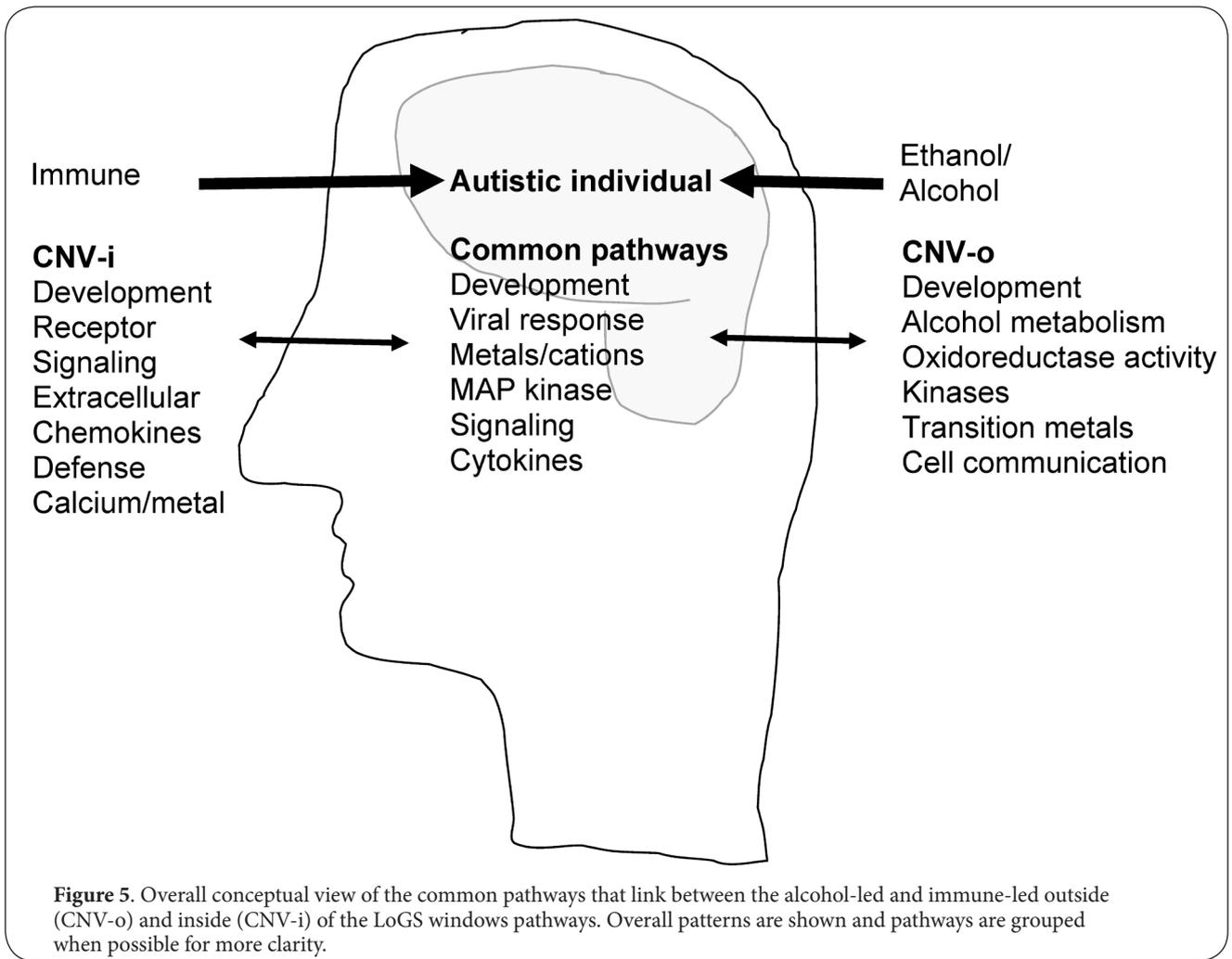
We have used the term 'overlap' in two different ways in our various analyses. In one case, overlap refers to the same genes falling into two different gene sets or pathways. That is the way overlap is used in the 50cM LoGS loci windows (LoGS-w) analysis. However, to find connectivity across the two sets of CNVs (CNV-o and CNV-i), we looked for overlaps of pathways not genes. This is because the two sets of CNVs will not have by definition overlapping genes. However, they

can have the same pathways. For example, morphogenesis is a global pathway from which some genes fall within the CNV-o group while a different set of genes fall within the CNV-i set (see [Figure 2](#)).

Development themes sit alongside cation and signaling themes in the top 10 results of LoGS-w windows

As in the CNV themes, we again found that development, cation, and signaling themes predominate in our top 10 list of pathways in the LoGS-w. There are 6905 genes in LoGS-w. The top 10 output of running these genes in EASE is presented in ([Table 3](#)) while a more detailed output is presented in ([Supplementary Table S4](#)).

Further inspection of the results from EASE shows that the top ranked pathway is not only very significant but is also very large and a quick inspection indicated that many of the genes in this pathway were overlapping with subsequent pathways. An inspection of these results also showed that many of the themes are repeating in different forms (for example cation related themes). This overlap over such seemingly disparate pathways suggested a connectivity between these pathways



and suggested a mechanism via which autism may occur (for example, we found both plasma membrane and signaling in our top 10 results. However, in general, not all plasma membrane pathways will necessarily relate to signaling and more importantly not all signaling occurs via the plasma membrane and thus an overlap 'fixes' literally for us the mechanism that is inferable from one pathway to the next so that signaling now is actually inferable as occurring through the plasma membrane).

Overlaps between such seemingly disparate themes can help us deduce mechanisms relevant to autism and can also show the progression of mechanism from one pathway to the next. We thus next conducted several analyses to find this connectivity across these pathways.

Genes overlapping with all gene sets in the top 10 LoGS-w enrichment results

We first wanted to see if any genes overlap across all the pathways in the top 10 results run on LoGS-w from EASE (see [Table 3](#)). We found the genes GRIK1 and GRIN1 are common

across all the pathways tested. Both these genes are glutamate receptors.

Interestingly, as an aside (because this note is not completely driven through an unbiased perspective but rather because we previously have identified immune function), one of the genes in our top 10 overlap analysis is CD86, an immune function gene that overlaps with everything except 'integral to plasma membrane' and 'cation'.

Cation pathways are important in their frequency in LoGS-w enrichment ranks

In the top 50 LoGS-w pathways from EASE, there are 14 pathways relating to cation/metal (including calcium). The cation pathway in the top 10 shows high overlap with the cell growth and cell communication pathways.

Top 10 connectivity map from LoGS-w

Next, we wanted to see how much overlap there is across all enriched pathways in the top 10 of LoGS-w enrichment as that would give us an indication of the level of connectivity

Table 3. EASE enrichment on 50cM windows. The top 10 categories in EASE are shown.

	Pathway	GO	P
1	Cellular process	BP	10 ⁻²⁶
2	Plasma membrane	CC	4.10 ⁻²²
3	Integral to plasma membrane	CC	10 ⁻¹⁹
4	Cell communication	BP	10 ⁻¹⁷
5	Signal transducer activity	MF	5.10 ⁻¹⁶
6	Signal transduction	BP	6.10 ⁻¹⁴
7	Morphogenesis	BP	2.10 ⁻¹¹
8	Organogenesis	BP	3.10 ⁻¹¹
9	Cell growth and/or maintenance	BP	2.10 ⁻⁹
10	Cation channel activity	MF	4.10 ⁻¹⁰

P = EASE score P value.

GO = gene ontology category; BP = biological process;

MF = molecular function; CC = cellular component.

Table 4. Percent overlap of a given gene set with the cellular process gene set.

Gene set	% overlap with cellular process	calculation
Cell process	-	-
Cell growth	100%	1163
Cell comm	100%	1022
Signal transduction	100%	803
Cation	100%	111
Int to Plasma mem.	87.75934	423/482
Plasma mem.	84.45441	565/669
Signal transducer act	81.86528	632/772
Organogenesis	0.652941	222/340
Morphogenesis	0.646277	243/376

across these pathways. This analysis presented itself via inspection where we saw that the highest ranked pathway (cellular process) was quite large and many of its genes showed significant overlap with the remaining pathways within the top 10. We thus starting with this pathway looked for percentage overlaps (number of genes overlapping divided by the total number of genes in the other pathway) with all pathways in the top 10. These percentage overlaps are presented in (Table 4). We found that cell communication, cell growth, signal transduction, and cation were completely subsumed within the cell process pathway. In other words, these pathways provide increasing levels of details (albeit at relatively lower significances) and mechanistic insights compared with cellular process. We also found significant overlaps between each of these pathways. The results of this analysis are presented in the (Supplementary figure S1).

The rationale for this analysis is that the largest pathway (which also is the most significant in our case) provides via its enrichment the most generic pathway description and therefore the highest level view that is obtainable. Subsequent pathways that overlap with this pathway provide finer details and pathways that overlap less with it are downstream mechanistic mediators.

Our connectivity order therefore based on Table 4 is:

Cellular process-->cell growth --> cell communication-->signal transduction-->cation-->integral to PM-->PM-->signal transducer activity-->organogenesis-->morphogenesis.

We also ran a top 50 grouped into 5 pathways overlaps as explained below, but the connectivity becomes a little more tricky since by combining the gene sets we lose some measure of significance. We also didn't use the most significant pathway precisely because it is too general to provide us with a way to connect it with other themes via inspection. Thus, we only present for this case the overall overlaps as explained next and do not describe a sequential pathway connectivity. However, the overlaps do provide justification for the idea that even among the top 50, there are significant overlaps across groups of pathways.

Top 5 themes in the top 50 pathways from LoGS-w

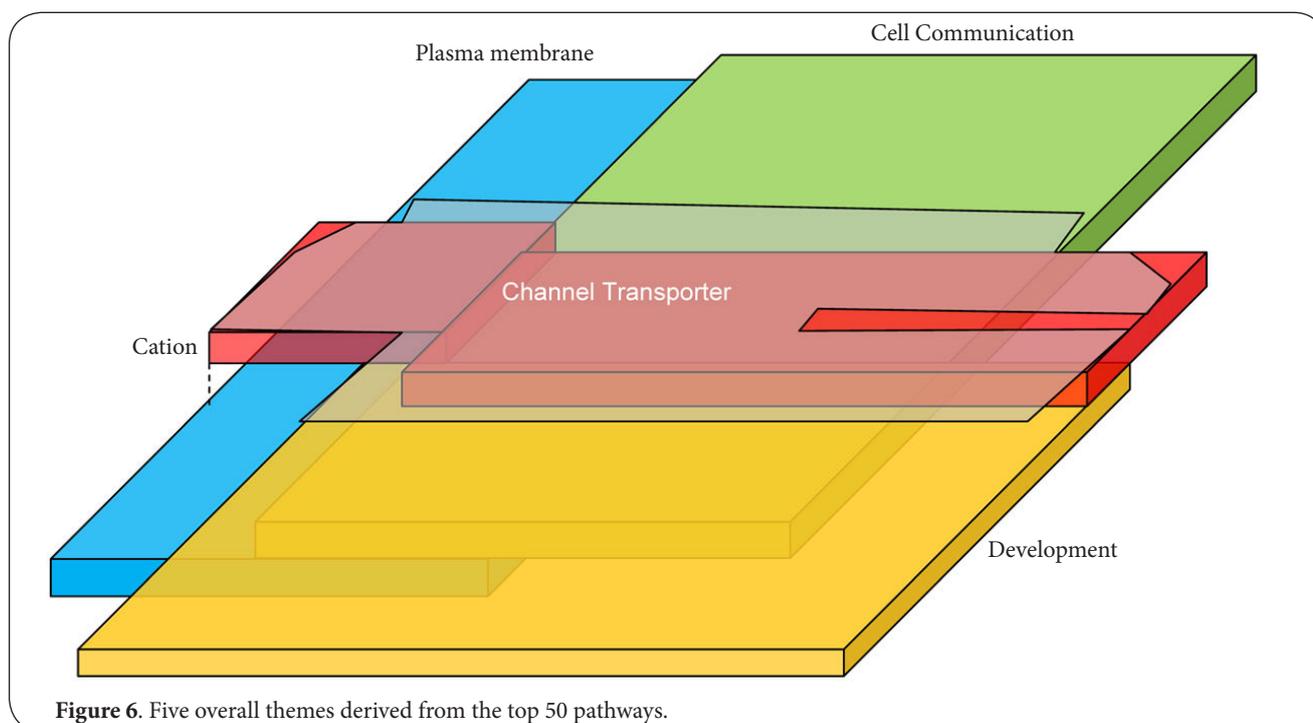
Next to get a global view of connectivity, we searched for repeating themes in the top 50 results from EASE run on LoGS-w. We found 5 overall repeating themes: Plasma membrane, cell communication, development, cation, and channel transporter. To conduct an overlap analysis, all genes for each category were placed together under that category (not all of the 50 pathways were grouped in this 'by inspection' analysis—the top 10 overlap analysis presented above however did use all 10 pathways in its overlap analysis) Figure 6 describes an overall view of such pathways overlaps while (Supplementary figure S2) describes the details of the building of the pathway overlaps.

We also ran an analysis to see what genes are common across all five groups. These genes are: SRI, P2RX5, KCNN3, GRIN1, GRIK1, CHRM3, CHRM2, CHRM1, ATP2A2, and ACCN1.

Discussion

This paper presents novel findings in autism relating to alcohol pathway dysfunction along with providing mechanistic insights into autism. To glean mechanism in autism, we looked for overlap across ranked pathways obtained via the use of EASE. This analysis shows that alcohol may be another mediator of autism with all loci relating to alcohol showing up outside LoGS-w. This bifurcation of results (the most significant pathways being driven by immune function within CNV-i and alcohol outside in CNV-o) shows two possible global pathways that might lead to autism. Their linking could have profound importance—in other words, environmental alcohol exposure along with viral infection in the context of perturbed immune genes that we describe and have previously described could be possible clues to the rising incidence of autism as we discuss below.

Across all of our analyses, what stands out is that development consistently shows up (LoGS analysis [3], CNV-c enrichment, CNV-o, CNV-i, and the LoGS-g enrichment). Also, each group of CNVs (CNV-i and CNV-o) showed more overlap across pathways within that group versus with the other group, although development themes showed significant



overlap (but not of the same genes obviously since the two groups of genes will be mutually exclusive between CNV-o and CNV-i). Thus, for example, alcohol showed overlap with transition metal group and copper also showed overlap with the transition metal group. Because overall the two groups of CNVs find common pathways that link them, this shows that there is unity in the mechanism that might underlie the development or the progression of, or the response to, autism.

What is interesting is that of the themes identified via the common CNV analysis, all five show up in our very unbiased LoGS analysis in our LoGS publication [3]. In other words, every gene that we identified in this analysis has been subjected to filters that were independent of our current or prior findings and yet repeatedly we find the same themes. For example, in our common CNV (or CNV-c) analysis (in addition to development themes) signaling themes are prominent in the top 20 pathways. The LoGS analysis found the same themes playing out. This is important because it provides a cross-validation built into our analysis. The CNV work is independently conducted of the 50cM linkage windows work and yet they find very similar themes.

Even though the CNV analysis has far fewer genes than our 50cM LoGS loci genes enrichment, they both point to essentially the same mechanisms at play.

Intriguingly, both the top 10 overlaps analysis and the top 50 overlap analysis of the LoGS-g enrichment identified genes involved in brain function.

Our overlap analysis on the genes from the LoGS-w windows in the top 10 (along with the top 50) gene sets shows two glutamate receptors that may be important in autism, GRIK1

and GRIN1. Both these receptors are involved in synaptic plasticity. GRIK1 has been shown to have some effect on subsequent cognitive ability during brain development [6]. Further genes were discovered via the grouped 5 pathway analysis on the top 50 pathways. Most of these genes have some role in the brain. The sorcin gene, SRI is known to be expressed in the caudate putamen nucleus in the rat brain [7]. P2RX5 (also known as P2X5) is expressed in rat brain during development [8]. Deignan et al., in talking of KCNN3 (which is also known as SK3) say that "SK3-containing channels influenced AP [action potential] frequency" [9]. SK channels such as SK3 (or KCNN3) in dopamine neurons are involved in the activity of dopamine neurons in response to glutamate release in the synapse [10]. Three muscarinic cholinergic receptor genes were found: CHRM1, CHRM2, and CHRM3. Each of these has a role in the brain. "The clinical study supports the proposal that CHRM1 sequence is associated with alterations in some aspects of executive function" [11]. CHRM2 may have a role in alcoholism [12] as well as intelligence, drug dependence and affective disorders [13]. CHRM3 is also expressed in the brain [14]. ACCN1 was shown to have increased expression in lymphoblastoid cells lines of individuals with fragile X syndrome [15]. Interestingly, these genes sit outside the CNV-i.

Both the glutamate receptors we have found, GRIK1 and GRIN1 overlap with all 10 of the top pathways and thus have a role in cation signaling.

TLR3 is a gene that is found in the response to virus gene set in the CNV-c gene set and was also found in the cell communication gene set in the CNV-o group has previously

been shown to have a role in inhibiting memory retention [16]. Human neurons express TLR3 [17]. Also, according to the foregoing reference, TLR-3 knock outs show enhanced memory. Intriguingly, a check with the Marshall autism CNV database shows a gain in copy number of the CNVs containing TLR3 as one would predict based on the prior reading of the literature [4]. RNF4 is found in the developing brain and reproductive system (it binds androgen receptor) [18]. RNF4 also shows a gain according to the autism CNV database [4].

The alcohol view obtained via the CNVs that sit outside the windows straddling the LoGS loci (CNV-o) could derive from the following perspective. It may be that the loci that have been described in autism did not pick up such alcohol related loci, possibly because alcohol relevant cases of autism were not as common in the past and thus such families that might now be affected by autism because of the increasing use of alcohol among women would never have had autism affected probands to be recruited into such studies. Alcohol use is known to be rising among young women [19,20]. It might also be that there are few microsatellite markers close to alcohol loci, or they might be so far away that recombinations will dampen or destroy the signal.

Although one possible explanation could be rising use of alcohol as described above, another link may occur via the rising incidence of diabetes. It is known that there is endogenous production of alcohol. One way for this is via yeast infection in the gut [21]. The liver enzyme that detoxifies alcohol is ADH (this enzyme in its various isoforms is present abundantly in the alcohol pathways we have in the CNVs outside the LoGS loci windows). It has also been reported that diabetes as well as obesity in pregnant women may be a causation for autism [22]. Further yeast infections are more prevalent in type 2 diabetes [23,24]. It is possible therefore that a rise in gut flora with type 2 diabetes or obesity could be one mechanism for the rise in endogenous alcohol production. Thus, the rising tide of obesity and diabetes could be one reason for the rise in autism. The increasing levels of alcohol use among young women would via this view have an additive effect. Such flora could potentially also increase when women are under immunosuppressant therapy for various conditions.

Another clue to this type of a mechanism is found by looking at another gene that we found in our LoGS loci analysis (this gene was found in an unbiased search). CHRM2 is known to be a risk factor for alcohol dependence. There have been earlier reports of a possible link between autism and alcohol exposure during pregnancy [25] and thus our CNV results shed an additional unbiased view on this perspective.

Others have found a link between alcohol use during pregnancy and lower IQ at age 8. The genes mentioned are: ADH1A, ADH1B, ADH1C, ADH4, and ADH7. Further, genetic variants of these genes were predisposing to low IQ in the child [26]. Our analysis contains all of the foregoing genes except ADH1B. A quick check in the autism database however shows ADH1B as also being listed in the region that contains

the other genes listed. Also, as we would expect, this region containing all these genes shows a copy number loss. Thus, this reinforces our assertion that alcohol/ethanol increase in the body of women with genetic backgrounds (with a lower ability to detoxify alcohol) who are also pregnant may predispose such women to having a child with autism.

One of the trace element pathways that repeatedly shows up in the CNV-o group relates to copper. There are also several pathways relating to heavy metal sensitivity and resistance with one pathway relating to both copper and cadmium. Increased alcohol in utero has been linked to deficiency in copper during development [27]. Copper deficiency causes abnormal fetal and neonatal brain development [28,29]. Further, reductions in trace elements such as copper can predispose the body to viral infections [30]. Interestingly, cadmium, a pollutant, can predispose individuals to increased alcohol use [31]. Ethanol itself increases cadmium in the urine and decreases it in the blood [31]. Autistic children have been previously shown to have slightly decreased levels of cadmium in the blood [32]. Individuals with autism also show higher levels of cadmium (but lower levels of calcium) in hair samples [33]. Others have reported similar findings for cadmium and calcium along with lower levels of copper in hair samples of autistic individuals [34].

It should also be noted that even though the genes in both CNV-o and CNV-i are separate, there are pathways that have genes from both groups that were previously discovered. For example, both CCL2 and CCL7 (these genes are in our CNV-i group and interestingly have been described by others as having a link to autism [35]) were in the CNV-c gene set "Metal ion homeostasis". This gene set had genes from both types of CNVs (CNV-i and CNV-o). Thus, there is cross-talk between the two sets of CNVs **Figure 5** as described earlier shows overall themes that show such links.

The fact that there is a lot of overlap across the top 10 (and the top 50) LoGS-w pathways shows that the themes are interlinked. The finding that the same genes have a function in development and cation and signaling and plasma membrane suggests not only multiple roles for such genes but also shows that such themes have connectivity (thus if we find signaling and plasma membrane sharing several genes, then we can conclude that the signaling is occurring through the plasma membrane).

Our current analysis complements our findings in LoGS by providing a possible mechanistic link across pathways. This paper shows that in addition to ranking pathways to discern mechanisms in disease, it may also help to look for pathway connectivities via overlaps (or other rationales) across populations to derive possible mechanisms related to the disease in question. Our analysis substantially validates our prior work with LoGS and gives another layer of detail that could help derive the mechanism underlying the development of autism. Further, although immune function was found in the CNV analysis that was conducted, it was not seen in the LoGS-w enrichment analysis. This justifies the running of LoGS

as this new analysis is not sensitive enough to pick up gene sets that may not be as significant within such a large group of genes as the 50cM LoGS loci group because the new analysis does not specifically search for the most proximal pathways. This new analysis however when run in tandem with LoGS can provide important mechanistic insights into the condition. Further, the 50cM loci regions helped to create a bifurcation in the types of CNVs identified. Our current analysis shows an important contribution from analyzing all genes that are within 50cM of all linkage loci implicated in autism. The rationale for this flowed from the way the enrichments of the 50cM window genes grouped themselves with high amounts of overlap (thus any analysis where such an overlap is evident could be subjected to our linked analysis).

Additional files

[Supplement figure S1](#)
[Supplement figure S2](#)
[Supplement Table S1](#)
[Supplement Table S2](#)
[Supplement Table S3](#)
[Supplement Table S4](#)

Competing interests

The author declares no competing interests.

Publication history

Editors: Craig A Erickson, Cincinnati Children's Hospital Medical Center, USA.

Ambuja S Bale, National Center for Environmental Assessment, USA.

Received: 02-May-2013 Revised: 15-Jun-2013

Re-Visited: 17-Jun-2013 Accepted: 19-Jun-2013

Published: 23-Jul-2013

References

1. Imai J, Katagiri H, Yamada T, Ishigaki Y, Suzuki T, Kudo H, Uno K, Hasegawa Y, Gao J, Kaneko K, Ishihara H, Nijima A, Nakazato M, Asano T, Minokoshi Y and Oka Y. **Regulation of pancreatic beta cell mass by neuronal signals from the liver.** *Science*. 2008; **322**:1250-4. | [Article](#) | [PubMed](#)
2. Mootha VK, Lindgren CM, Eriksson KF, Subramanian A, Sihag S, Lehar J, Puigserver P, Carlsson E, Ridderstrale M, Laurila E, Houstis N, Daly MJ, Patterson N, Mesirov JP, Golub TR, Tamayo P, Spiegelman B, Lander ES, Hirschhorn JN, Altshuler D and Groop LC. **PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes.** *Nat Genet*. 2003; **34**:267-73. | [Article](#) | [PubMed](#)
3. Saxena V, Ramdas S, Ochoa CR, Wallace D, Bhide P and Kohane I. **Structural, genetic, and functional signatures of disordered neuro-immunological development in autism spectrum disorder.** *PLoS One*. 2012; **7**:e48835. | [Article](#) | [PubMed Abstract](#) | [PubMed Full Text](#)
4. Marshall CR, Noor A, Vincent JB, Lionel AC, Feuk L, Skaug J, Shago M, Moessner R, Pinto D, Ren Y, Thiruvahindrapuram B, Fiebig A, Schreiber S, Friedman J, Ketelaars CE, Vos YJ, Ficocioglu C, Kirkpatrick S, Nicolson R, Sloman L, Summers A, Gibbons CA, Teebi A, Chitayat D, Weksberg R, Thompson A, Vardy C, Crosbie V, Luscombe S, Baatjes R, Zwaigenbaum L, Roberts W, Fernandez B, Szatmari P and Scherer SW. **Structural variation of chromosomes in autism spectrum disorder.** *Am J Hum Genet*. 2008; **82**:477-88. | [Article](#) | [PubMed Abstract](#) | [PubMed Full Text](#)
5. Hosack DA, Dennis G, Jr., Sherman BT, Lane HC and Lempicki RA. **Identifying biological themes within lists of genes with EASE.** *Genome Biol*. 2003; **4**:R70. | [Article](#) | [PubMed Abstract](#) | [PubMed Full Text](#)
6. Docherty SJ, Davis OS, Kovas Y, Meaburn EL, Dale PS, Petrill SA, Schalkwyk LC and Plomin R. **A genome-wide association study identifies multiple loci associated with mathematics ability and disability.** *Genes Brain Behav*. 2010; **9**:234-47. | [Article](#) | [PubMed Abstract](#) | [PubMed Full Text](#)
7. Meyers MB, Puri TS, Chien AJ, Gao T, Hsu PH, Hosey MM and Fishman GI. **Sorcín associates with the pore-forming subunit of voltage-dependent L-type Ca²⁺ channels.** *J Biol Chem*. 1998; **273**:18930-5. | [Article](#) | [PubMed](#)
8. Cheung KK, Chan WY and Burnstock G. **Expression of P2X purinoceptors during rat brain development and their inhibitory role on motor axon outgrowth in neural tube explant cultures.** *Neuroscience*. 2005; **133**:937-45. | [Article](#) | [PubMed](#)
9. Deignan J, Lujan R, Bond C, Riegel A, Watanabe M, Williams JT, Maylie J and Adelman JP. **SK2 and SK3 expression differentially affect firing frequency and precision in dopamine neurons.** *Neuroscience*. 2012; **217**:67-76. | [Article](#) | [PubMed](#)
10. Ji H, Hougaard C, Herrik KF, Strobaek D, Christophersen P and Shepard PD. **Tuning the excitability of midbrain dopamine neurons by modulating the Ca²⁺ sensitivity of SK channels.** *Eur J Neurosci*. 2009; **29**:1883-95. | [Article](#) | [PubMed](#)
11. Scarr E, Sundram S, Deljo A, Cowie TF, Gibbons AS, Juzva S, Mackinnon A, Wood SJ, Testa R, Pantelis C and Dean B. **Muscarinic M1 receptor sequence: preliminary studies on its effects on cognition and expression.** *Schizophr Res*. 2012; **138**:94-8. | [Article](#) | [PubMed](#)
12. Luo X, Kranzler HR, Zuo L, Wang S, Blumberg HP and Gelernter J. **CHRM2 gene predisposes to alcohol dependence, drug dependence and affective disorders: results from an extended case-control structured association study.** *Hum Mol Genet*. 2005; **14**:2421-34. | [Article](#) | [PubMed](#)
13. Dick DM, Aliev F, Kramer J, Wang JC, Hinrichs A, Bertelsen S, Kuperman S, Schuckit M, Nurnberger J, Jr. Edenberg HJ, Porjesz B, Begleiter H, Hesselbrock V, Goate A and Bierut L. **Association of CHRM2 with IQ: converging evidence for a gene influencing intelligence.** *Behav Genet*. 2007; **37**:265-72. | [Article](#) | [PubMed](#)
14. Weston-Green K, Huang XF, Lian J and Deng C. **Effects of olanzapine on muscarinic M3 receptor binding density in the brain relates to weight gain, plasma insulin and metabolic hormone levels.** *Eur Neuropsychopharmacol*. 2012; **22**:364-73. | [Article](#) | [PubMed](#)
15. Bittel DC, Kibiriyeva N and Butler MG. **Whole genome microarray analysis of gene expression in subjects with fragile X syndrome.** *Genet Med*. 2007; **9**:464-72. | [Article](#) | [PubMed](#)
16. Okun E, Griffioen K, Barak B, Roberts NJ, Castro K, Pita MA, Cheng A, Mughal MR, Wan R, Ashery U and Mattson MP. **Toll-like receptor 3 inhibits memory retention and constrains adult hippocampal neurogenesis.** *Proc Natl Acad Sci U S A*. 2010; **107**:15625-30. | [Article](#) | [PubMed Abstract](#) | [PubMed Full Text](#)
17. Lafon M, Megret F, Lafage M and Prehaud C. **The innate immune facet of brain: human neurons express TLR-3 and sense viral dsRNA.** *J Mol Neurosci*. 2006; **29**:185-94. | [Article](#) | [PubMed](#)
18. Galili N, Nayak S, Epstein JA and Buck CA. **Rnf4, a RING protein expressed in the developing nervous and reproductive systems, interacts with Gscl, a gene within the DiGeorge critical region.** *Dev Dyn*. 2000; **218**:102-11. | [Article](#) | [PubMed](#)
19. Gilmore IT. **Excessive Drinking in Young Women: Not just a "lifestyle disease".** *BMJ*. 2008; **336**:952-3. | [Article](#) | [PubMed Abstract](#) | [PubMed Full Text](#)
20. Stojek M and Fischer S. **Impulsivity and motivations to consume alcohol: a prospective study on risk of dependence in young adult women.** *Alcohol Clin Exp Res*. 2013; **37**:292-9. | [Article](#) | [PubMed](#)
21. Dahshan A and Donovan K. **Auto-brewery syndrome in a child with short gut syndrome: case report and review of the literature.** *J Pediatr Gastroenterol Nutr*. 2001; **33**:214-5. | [Article](#) | [PubMed](#)
22. Krakowiak P, Walker CK, Bremer AA, Baker AS, Ozonoff S, Hansen RL and Hertz-Picciotto I. **Maternal metabolic conditions and risk for autism and other neurodevelopmental disorders.** *Pediatrics*. 2012; **129**:e1121-8. | [Article](#) | [PubMed Abstract](#) | [PubMed Full Text](#)

23. Araujo SM, Fontes CJ, Leite Junior DP and Hahn RC. **Fungal agents in different anatomical sites in Public Health Services in Cuiaba, state of Mato Grosso, Brazil.** *Rev Inst Med Trop Sao Paulo.* 2012; **54**:5-10. | [Article](#) | [PubMed](#)
24. Zarei-Mahmoudabadi A, Zarrin M, Ghanatir F and Vazirianzadeh B. **Candiduria in hospitalized patients in teaching hospitals of Ahvaz.** *Iran J Microbiol.* 2012; **4**:198-203. | [PubMed Abstract](#) | [PubMed Full Text](#)
25. Fombonne E. **Is exposure to alcohol during pregnancy a risk factor for autism?** *J Autism Dev Disord.* 2002; **32**:243. | [Article](#) | [PubMed](#)
26. Lewis SJ, Zuccolo L, Davey Smith G, Macleod J, Rodriguez S, Draper ES, Barrow M, Alati R, Sayal K, Ring S, Golding J and Gray R. **Fetal alcohol exposure and IQ at age 8: evidence from a population-based birth-cohort study.** *PLoS One.* 2012; **7**:e49407. | [Article](#) | [PubMed Abstract](#) | [PubMed Full Text](#)
27. Keen CL, Uriu-Hare JY, Hawk SN, Jankowski MA, Daston GP, Kwik-Urbe CL and Rucker RB. **Effect of copper deficiency on prenatal development and pregnancy outcome.** *Am J Clin Nutr.* 1998; **67**:1003S-1011S. | [Article](#) | [PubMed](#)
28. Georgieff MK. **Nutrition and the developing brain: nutrient priorities and measurement.** *Am J Clin Nutr.* 2007; **85**:614S-620S. | [Article](#) | [PubMed](#)
29. Gybina AA, Tkac I and Prohaska JR. **Copper deficiency alters the neurochemical profile of developing rat brain.** *Nutr Neurosci.* 2009; **12**:114-22. | [Article](#) | [PubMed Abstract](#) | [PubMed Full Text](#)
30. Smith AD, Botero S and Levander OA. **Copper deficiency increases the virulence of amyocarditic and myocarditic strains of coxsackievirus B3 in mice.** *J Nutr.* 2008; **138**:849-55. | [Article](#) | [PubMed](#)
31. Brzoska MM, Galazyn-Sidorczuk M and Dzwilewska I. **Ethanol consumption modifies the body turnover of cadmium: a study in a rat model of human exposure.** *J Appl Toxicol.* 2012. | [Article](#) | [PubMed](#)
32. Adams JB, Audhya T, McDonough-Means S, Rubin RA, Quig D, Geis E, Gehn E, Loresto M, Mitchell J, Atwood S, Barnhouse S and Lee W. **Toxicological status of children with autism vs. neurotypical children and the association with autism severity.** *Biol Trace Elem Res.* 2013; **151**:171-80. | [Article](#) | [PubMed](#)
33. Blaurock-Busch E, Amin OR, Dessoki HH and Rabah T. **Toxic Metals and Essential Elements in Hair and Severity of Symptoms among Children with Autism.** *Maedica (Buchar).* 2012; **7**:38-48. | [Article](#) | [PubMed](#)
34. Al-Farsi YM, Waly MI, Al-Sharbaty MM, Al-Shafae MA, Al-Farsi OA, Al-Khaduri MM, Gupta I, Ouhtit A, Al-Adawi S, Al-Said MF and Deth RC. **Levels of heavy metals and essential minerals in hair samples of children with autism in Oman: a case-control study.** *Biol Trace Elem Res.* 2013; **151**:181-6. | [Article](#) | [PubMed](#)
35. Jyonouchi H, Geng L, Streck DL and Toruner GA. **Children with autism spectrum disorders (ASD) who exhibit chronic gastrointestinal (GI) symptoms and marked fluctuation of behavioral symptoms exhibit distinct innate immune abnormalities and transcriptional profiles of peripheral blood (PB) monocytes.** *J Neuroimmunol.* 2011; **238**:73-80. | [Article](#) | [PubMed](#)

Citation:

Saxena V. **Ethanol/alcohol pathways along with immune, cation and development may play a role in autism spectrum disorder.** *J Med Disord.* 2013; **1**:1. <http://dx.doi.org/10.7243/2053-3659-1-1>